THE MECHANICAL SYSTEMS DESIGN HANDBOOK

Modeling, Measurement, and Control

OSITA D. I. NWOKAH
YILDIRIM HURMUZLU

CRC PRESS

THE MECHANICAL SYSTEMS DESIGN HANDBOOK

Modeling, Measurement, and Control

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

The Electrical Engineering Handbook Series

Series Editor

Richard C. Dorf

University of California, Davis

Titles Included in the Series

The Avionics Handbook, Cary R. Spitzer

The Biomedical Engineering Handbook, *2nd Edition,* Joseph D. Bronzino

The Circuits and Filters Handbook, Wai-Kai Chen

The Communications Handbook, Jerry D. Gibson

The Control Handbook, William S. Levine

The Digital Signal Processing Handbook, Vijay K. Madisetti & Douglas Williams

The Electrical Engineering Handbook, *2nd Edition,* Richard C. Dorf

The Electric Power Engineering Handbook, Leo L. Grigsby

The Electronics Handbook, Jerry C. Whitaker

The Engineering Handbook, Richard C. Dorf

The Handbook of Formulas and Tables for Signal Processing, Alexander D. Poularikas

The Industrial Electronics Handbook, J. David Irwin

The Measurement, Instrumentation, and Sensors Handbook, John G. Webster

The Mechanical Systems Design Handbook, Osita D.I. Nwokah

The RF and Microwave Handbook, Mike Golio

The Mobile Communications Handbook, *2nd Edition*, Jerry D. Gibson

The Ocean Engineering Handbook, Ferial El-Hawary

The Technology Management Handbook, Richard C. Dorf

The Transforms and Applications Handbook, *2nd Edition,* Alexander D. Poularikas

The VLSI Handbook, Wai-Kai Chen

The Mechatronics Handbook, Robert H. Bishop

The Computer Engineering Handbook, Vojin Oklobdzija

Forthcoming Titles

The Circuits and Filters Handbook, *2nd Edition,* Wai-Kai Chen

The Handbook of Ad hoc Wireless Networks, Mohammad Ilyas

The Handbook of Optical Communication Networks, Mohammad Ilyas

The Handbook of Nanoscience, Engineering, and Technology, William A. Goddard, Donald W. Brenner, Sergey E. Lyshevski, and Gerald J. Iafrate

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

THE MECHANICAL SYSTEMS DESIGN HANDBOOK

Modeling, Measurement, and Control

OSITA D. I. NWOKAH
YILDIRIM HURMUZLU
Southern Methodist University
Dallas, Texas

CRC PRESS

Boca Raton London New York Washington, D.C.

www.TechnicalBooksPDF.com

Library of Congress Cataloging-in-Publication Data

The Mechanical systems design handbook : modeling, measurement, and control / edited by Osita D.I. Nwokah, Yildirim Hurmuzlu.
p. cm. -- (The Electrical engineering handbook series)
Includes bibliographical references and index.
ISBN 0-8493-8596-2 (alk. paper)
1. Production engineering. 2. Manufacturing processes. I. Nwokah, Osita D. I. II. Hurmuzlu, Yildirim. III. Series.

TS176 .M42 2001
658.5--dc21

2001043150

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the authors and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

All rights reserved. Authorization to photocopy items for internal or personal use, or the personal or internal use of specific clients, may be granted by CRC Press LLC, provided that $1.50 per page photocopied is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA The fee code for users of the Transactional Reporting Service is ISBN 0-8493-8596-2/02/$0.00+$1.50. The fee is subject to change without notice. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press LLC for such copying.

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

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 CRC Press Web site at www.crcpress.com

© 2002 by CRC Press LLC

No claim to original U.S. Government works
International Standard Book Number 0-8493-8596-2
Library of Congress Card Number 2001043150
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

www.TechnicalBooksPDF.com

Preface

This handbook is targeted as a reference for the use of engineers and scientists in industry. We have compiled a collection of selected topics that are directly related to the design and control of mechanical systems. The main motivation for the book is to present a practical overview of fundamental issues associated with design and control of mechanical systems. The reader will find four sections in the handbook: (1) Manufacturing, (2) Vibration Control, (3) Aerospace Systems, and (4) Robotics. Although the sections are arranged in a certain order, each contribution can stand alone to represent its subject. Thus, people can read the handbook in any order they see fit.

The late Professor Osita Nwokah envisioned this project. Unfortunately, he could not see it through to completion. Professor Nwokah was the chairman of the mechanical engineering department at Southern Methodist University and a distinguished member of the control community when he passed away on April 20, 1999. It was important to me to finish one of Professor Nwokah's last projects.

The reader will find a broad range of thoroughly covered important topics by well-known experts in their respective fields. Section I encompasses control issues related to manufacturing systems including several topics from precision manufacturing to machine vibrations. Section II deals with active vibration control including a diverse spectrum of topics such as suspension systems and piezoelectric networks. Section III touches upon aerospace systems, and the authors have presented a detailed analysis of tensegrity structures. Section IV covers robotics and is an encyclopedic review of most issues related to the control and design of robotic systems.

It has been a pleasure to work with the four section editors, each a renowned international expert in his respective area. They, in turn, recruited very competent people who wrote chapters that, in my view, are individually important contributions to the design and control of mechanical systems. I also thank the people at CRC Press whose energy and constant support were essential to the completion of this handbook. I especially thank Nora Konopka who has spent numerous hours developing and producing this handbook.

Yildirim Hurmuzlu
Dallas, Texas

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Editors

Yildirim Hurmuzlu currently serves as the Chairman of the Department of Mechanical Engineering at Southern Methodist University in Dallas, Texas. He has been with the department since 1987, and has served as assistant, associate, and full professor. Dr. Hurmuzlu's research interests are in the field of dynamic systems and controls, with particular emphasis on robotics and biomechanics. His research has been supported by the National Science Foundation, Whitaker Foundation, and Texas National Laboratory Commission, and industrial corporations such as Bell Helicopter, Raytheon, Saudi Aramco, and Alcatel Corp. He has authored more than 50 articles in journals and conference proceedings and has organized sessions at national and international conferences. Dr. Hurmuzlu is an associate editor of the *ASME Journal of Dynamic Systems Measurement and Control*. He has also served as the chairman of IEEE Dallas–Fort Worth Control Systems Society and the ASME DSC biomechanics panel.

Osita Nwokah was a leading international authority on the application of multivariable design methods for the control of high-performance, high-bypass ratio turbomachinery. As a graduate student at the University of Manchester Institute of Science and Technology (UMIST), Manchester, England, he was a member of the team that wrote the initial control algorithms for the regulation of the Rolls Royce Concordce Olympus 925 Engines using the inverse Nyquist array in 1971. After moving to the United States, Dr. Nwokah continued this line of work and developed fundamental methodologies to combine the inverse Nyquist array with the quantitative feedback theory (QFT) design method of Horowitz. At the time of his death, Dr. Nwokah was studying multivariable control design and implementation for the RASCAL Helicopter for NASA and U.S. Army at NASA Ames RC, Moffet Field, California.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Contributors

Rajesh Adhikari
Department of Mechanical and Aerospace Engineering
University of California, San Diego
La Jolla, CA

Yusuf Altintas
Department of Mechanical Engineering
The University of British Columbia
Vancouver, B.C., Canada

Antal K. Bejczy
Jet Propulsion Lab
California Institute of Technology
Pasadena, CA

Branislav Borovać
Faculty of Technical Sciences
University of Novi Sad
Novi Sad, Yugoslavia

Frederic Bossens
Université Libre de Bruxelles
Brussels, Belgium

Waileung Chan
Department of Mechanical and Aerospace Engineering
University of California, San Diego
La Jolla, CA

Kourosh Danai
Department of Mechanical and Industrial Engineering
University of Massachusetts
Amherst, MA

Darren M. Dawson
Electrical and Computer Engineering
Clemson University
Clemson, SC

Richard J. Furness
Advanced Manufacturing Technology Development
Ford Motor Company
Detroit, MI

Martin Hägele
Fraunhofer Institute
Stuttgart, Germany

David E. Hardt
Professor of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA

J. William Helton
Department of Mathematics
University of California, San Diego
La Jolla, CA

Martin Hosek
University of Connecticut
Storrs, CT

S. Jack Hu
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

Yildirim Hurmuzlu
Department of Mechanical Engineering
Southern Methodist University
Dallas, TX

Kenji Inoue
Department of Systems and Human Science
Osaka University
Osaka, Japan

Nader Jalili
Department of Mechanical Engineering
Clemson University
Clemson, SC

Elijah Kannatey-Asibu, Jr.
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Branko Karan
Mihajlo Pupin Institute
Belgrade, Yugoslavia

Dusko M. Katić
Mihajlo Pupin Institute
Belgrade, Yugoslavia

David Kazmer
Department of Mechanical and Industrial Engineering
University of Massachusetts
Amherst, MA

P. P. Khargonekar
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI

Nenad M. Kircanski
University of Toronto
Toronoto, Ontario, Canada

Yoram Koren
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

Willi Kortüm
German Aerospace Research Establishment
Wessling, Germany

Thomas R. Kurfess
The George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA

Robert G. Landers
Department of Mechanical Engineering and Mathematics
University of Missouri
Rolla, MO

Nicolas Loix
Micromega Dynamics
Angleur, Belgium

M. G. Mehrabi
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

D. L. Mingori
Department of Mechanical and Aerospace Engineering
University of California
Los Angeles, CA

Siddharth P. Nagarkatti
Lucent Technologies
Sturbridge, MA

Osita D. I. Nwokah
Department of Mechanical Engineering
Southern Methodist University
Dallas, TX

Nejat Olgac
Department of Mechanical Engineering
University of Connecticut
Storrs, CT

Jean-Paul Pinaud
Department of Mechanical and Aerospace Engineering
University of California
La Jolla, CA

Veljko Potkonjak
University of Belgrade
Belgrade, Yugoslavia

A. Preumont
Université Libre de Bruxelles
Brussels, Belgium

Rolf Dieter Schraft
Fraunhofer Institute
Stuttgart, Germany

Bruno Siciliano
Universita degli Studi di Napoli Frederico II
Naples, Italy

Robert E. Skelton
Department of Mechanical and Aerospace Engineering
University of California
La Jolla, CA

Dragan Stokić
ATB–Institute für Angewandte Systemtechnik
Bremen, Germany

Dragoljub Šurdilović
Fraunhofer Institute
Stuttgart, Germany

Masaharu Takano
Department of Industrial Engineering
Kansai University
Osaka, Japan

D. M. Tilbury
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

A. Galip Ulsoy
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

Michael Valášek
Czech Technical University
Prague, Czech Republic

Miomir Vukobratović
Mihajlo Pupin Institute
Belgrade, Yugoslavia

Kon-Well Wang
Structural Dynamics and Controls Lab
Pennsylvania State University
University Park, PA

Derek Yip-Hoi
Department of Mechanical Engineering
University of Michigan
Ann Arbor, MI

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Contents

SECTION I Manufacturing

1 Manufacturing Systems and Their Design Principles
- 1.1 Introduction
- 1.2 Major Manufacturing Paradigms and Their Objectives
- 1.3 Significance of Functionality/Capacity Adjustments in Modern Manufacturing Systems
- 1.4 Critical Role of Computers in Modern Manufacturing
- 1.5 Design Principles of Modern Manufacturing Systems
- 1.6 Future Trends and Research Directions

Selected References

2 Computer-Aided Process Planning for Machining

Abstract
- 2.1 Introduction
- 2.2 What Is Computer-Aided Process Planning (CAPP)?
- 2.3 Review of CAPP Systems
- 2.4 Drivers of CAPP System Development
- 2.5 Characteristics of CAPP Systems
- 2.6 Integrating CAD with CAPP: Feature Extraction
- 2.7 Integrating CAPP with Manufacturing
- 2.8 CAPP for New Domains
- 2.9 Conclusions

References

3 Discrete Event Control of Manufacturing Systems
- 3.1 Introduction
- 3.2 Background on the Logic Control Problems
- 3.3 Current Industrial Practice
- 3.4 Current Trends
- 3.5 Formal Methods for Logic Control
- 3.6 Further Reading

Acknowledgments

References

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

4 Machine Tool Dynamics and Vibrations
4.1 Introduction
4.2 Chatter Vibrations in Cutting
4.3 Analytical Prediction of Chatter Vibrations in Milling
References

5 Machine Tool Monitoring and Control
5.1 Introduction
5.2 Process Monitoring
5.3 Process Control
5.4 Conclusion
References

6 Process Monitoring and Control of Machining Operations
6.1 Introduction
6.2 Force/Torque/Power Generation
6.3 Forced Vibrations and Regenerative Chatter
6.4 Tool Condition Monitoring and Control
6.5 Other Process Phenomena
6.6 Future Direction and Efforts
Acknowledgments
References

7 Forming Processes: Monitoring and Control
7.1 Introduction: Process and Control Objectives
7.2 The Plant or Load: Forming Physics
7.3 Machine Control
7.4 Machine Control: Force or Displacement?
7.5 Process Resolution Issues: Limits to Process Control
7.6 Direct Shape Feedback and Control
7.7 Summary
References

8 Assembly and Welding Processes and Their Monitoring and Control
8.1 Assembly Processes
8.2 Monitoring and Control of Resistance Welding Process
8.3 Monitoring and Control of Arc Welding Processes
References

9 Control of Polymer Processing
9.1 Introduction
9.2 Process Description

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

9.3 Process Variability
9.4 Modeling
9.5 Process Control
9.6 Conclusions
References

10 Precision Manufacturing
10.1 Deterministic Theory Applied to Machine Tools
10.2 Basic Definitions
10.3 Motion
10.4 Sources of Error and Error Budgets
10.5 Some Typical Methods of Measuring Errors
10.6 Conclusion
10.7 Terminology
References

SECTION II Vibration Control

11 Active Damping of Large Trusses
Abstract
11.1 Introduction
11.2 Active Struts
11.3 Active Tendon Control
11.4 Active Damping Generic Interface
11.5 Microvibrations
11.6 Conclusions
Acknowledgment
References

12 Semi-Active Suspension Systems
12.1 Introduction
12.2 Semi-Active Suspensions Design
12.3 Adjustable Suspension Elements
12.4 Automotive Semi-Active Suspensions
12.5 Application of Control Techniques to Semi-Active Suspensions
12.6 Practical Considerations and Related Topics
References

13 Semi-Active Suspension Systems II
13.1 Concepts of Semi-Active Suspension Systems
13.2 Control Design Methodology

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

13.3 Properties of Semi-Active Suspensions: Performance Indexes
13.4 Examples of Practical Applications
References

14 Active Vibration Absorption and Delayed Feedback Tuning
14.1 Introduction
14.2 Delayed Resonator Dynamic Absorbers
14.3 Multiple Frequency ATVA and Its Stability
Acknowledgments
References

15 Vibration Suppression Utilizing Piezoelectric Networks
15.1 Introduction
15.2 Passive and Semi-Active Piezoelectric Networks for Vibration Absorption and Damping
15.3 Active-Passive Hybrid Piezoelectric Network Treatments for General Modal Damping and Control
15.4 Active-Passive Hybrid Piezoelectric Network Treatments for Narrowband Vibration Suppression
15.5 Nonlinear Issues Related to Active-Passive Hybrid Piezoelectric Networks
15.6 Summary and Conclusions
Acknowledgments
References

16 Vibration Reduction via the Boundary Control Method
16.1 Introduction
16.2 Cantilevered Beam
16.3 Axially Moving Web
16.4 Flexible Link Robot Arm
16.5 Summary
Acknowledgments
References

SECTION III Dynamics and Control of Aerospace Systems

17 An Introduction to the Mechanics of Tensegrity Structures
Abstract
17.1 Introduction
17.2 Planar Tensegrity Structures Efficient in Bendin
17.3 Planar Class K Tensegrity Structures Efficient in Compression

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

17.4 Statics of a 3-Bar Tensegrity
17.5 Concluding Remarks
Acknowledgment
Appendix 17.A Nonlinear Analysis of Planar Tensegrity
Appendix 17.B Linear Analysis of Planar Tensegrity
Appendix 17.C Derivation of Stiffness of the C4T1i Structure
References

18 The Dynamics of the Class 1 Shell Tensegrity Structure
Abstract
18.1 Introduction
18.2 Tensegrity Definitions
18.3 Dynamics of a Two-Rod Element
18.4 Choice of Independent Variables and Coordinate Transformations
18.5 Tendon Forces
18.6 Conclusion
Acknowledgment
Appendix 18.A Proof of Theorem 18.1
Appendix 18.B Algebraic Inversion of the Q Matrix
Appendix 18.C General Case for (n, m) = (i, 1)
Appendix 18.D Example Case (n,m) = (3,1)
Appendix 18.E Nodal Forces
References

SECTION IV Robotics

19 Robot Kinematics
19.1 Introduction
19.2 Description of Orientation
19.3 Direct Kinematics
19.4 Inverse Kinematics
19.5 Differential Kinematics
19.6 Differential Kinematics Inversion
19.7 Inverse Kinematics Algorithms
19.8 Further Reading
References

20 Robot Dynamics
20.1 Fundamentals of Robot Dynamic Modeling
20.2 Recursive Formulation of Robot Dynamics
20.3 Complete Model of Robot Dynamics

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

20.4 Some Application of Computer-Aided Dynamics
20.5 Extension of Dynamic Modeling — Some Additional Dynamic Effects
Appendix: Calculation of Transformation Matrices
References

21 Actuators and Computer-Aided Design of Robots
21.1 Robot Driving Systems
21.2 Computer-Aided Design
References

22 Control of Robots
22.1 Introduction
22.2 Hierarchical Control of Robots
22.3 Control of a Single Joint of the Robot
22.4 Control of Simultaneous Motion of Several Robot Joints
References

23 Control of Robotic Systems in Contact Tasks
23.1 Introduction
23.2 Contact Tasks
23.3 Classification of Robotized Concepts for Constrained Motion Control
23.4 Model of Robot Performing Contact Tasks
23.5 Passive Compliance Methods
23.6 Active Compliant Motion Control Methods
23.7 Contact Stability and Transition
23.8 Synthesis of Impedance Control at Higher Control Levels
23.9 Conclusion
References

24 Intelligent Soft-Computing Techniques in Robotics
24.1 Introduction
24.2 Connectionist Approach in Robotics
24.3 Neural Network Issues in Robotics
24.4 Fuzzy Logic Approach
24.5 Neuro-Fuzzy Approach in Robotics
24.6 Genetic Approach in Robotics
24.7 Conclusion
References

25 Teleoperation and Telerobotics
25.1 Introduction
25.2 Hand Controllers

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

25.3 FRHC Control System
25.4 ATOP Computer Graphics
25.5 ATOP Control Experiments
25.6 Anthropomorphic Telerobotics
25.7 New Trends in Applications
Acknowledgment
References

26 Mobile Robotic Systems
26.1 Introduction
26.2 Fundamental Issues
26.3 Dynamics of Mobile Robots
26.4 Control of Mobile Robots
References

27 Humanoid Robots
27.1 Zero-Moment Point — Proper Interpretation
27.2 Modeling of Biped Dynamics and Gait Synthesis
27.3 Control Synthesis for Biped Gait
27.4 Dynamic Stability Analysis of Biped Gait
27.5 Realization of Anthropomorphic Mechanisms and Humanoid Robots
27.6 Conclusion
References

28 Present State and Future Trends in Mechanical Systems Design for Robot Application
28.1 Introduction
28.2 Industrial Robots
28.3 Service Robots
References

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

I

Manufacturing

A. Galip Hulsoy

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

1

Manufacturing Systems and Their Design Principles

1.1 Introduction
1.2 Major Manufacturing Paradigms and Their Objectives
1.3 Significance of Functionality/Capacity Adjustments in Modern Manufacturing Systems
1.4 Critical Role of Computers in Modern Manufacturing
1.5 Design Principles of Modern Manufacturing Systems
Product Design and Design for Manufacturability • Process Planning and System Design of Manufacturing Systems • Software/Hardware Architecture and Communications in Manufacturing Systems • Monitoring and Control of Manufacturing Systems
1.6 Future Trends and Research Directions

M. G. Mehrabi
University of Michigan

A. Galip Ulsoy
University of Michigan

Yoram Koren
University of Michigan

1.1 Introduction

Manufacturing has always been the key to success among nations in the world economy (Figure 1.1). A responsive manufacturing system working in harmony with the rest of an enterprise has a major impact on its competitiveness; it plays a vital role in the successful introduction of new products or continuous improvements of existing products in response to demands of the market (Cohen, 1987).

A wide variety of items are produced by manufacturing firms, depending upon the market demands they may be custom made or mass produced. Manufacturing systems used for their production are designed and tailored to specific requirements. Consequently, several manufacturing techniques are adopted to address new market demands.

This chapter is devoted to a high-level overview of manufacturing techniques, their objectives and design principles. In this regard, some of the available manufacturing techniques are explained and their achievements, advantages, and limitations are discussed. Due to the significant impact of computers on manufacturing, an effort is made to introduce the role of computers and information technology in modern manufacturing systems. In this regard, applications and functions of computers in various stages of product design, generation of the sequence of operations and process planning, control of the machines and monitoring of the processes (on/off line), automation, networking and communication systems, and quality control of the production systems are explained. Later in the chapter, the design principles of manufacturing systems and their components

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

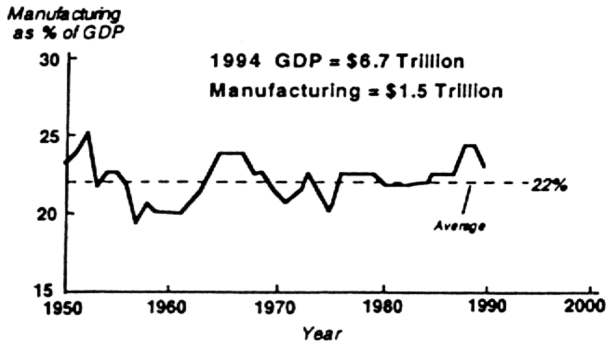


FIGURE 1.1 Despite assertions that the U.S. is becoming a service industry, manufacturing has consistently accounted for about 22% of GDP. (Source: U.S. Bureau of Labor Statistics.)

are presented as well as some of the issues related to their enabling technologies and barriers. The chapter concludes with a discussion of some of the future directions in manufacturing systems.

1.2 Major Manufacturing Paradigms and Their Objectives

New technological developments and market demands have major impacts on manufacturing. As a result, several shifts in the focus of manufacturing processes can be observed, which can be conveniently divided into three major epochs: (1) precomputer numerical control, (2) computer numerical control (CNC), and (3) knowledge epochs (Mehrabi and Ulsoy, 1997; Mehrabi, Ulsoy, and Koren, 1998). In the pre-CNC epochs (before the 1970s), the emphasis was on increased production rate; little demand existed for product variations and the market was characterized by local competition. Mass production uses dedicated lines designed for production of a specific part; it uses transfer line technology with fixed tooling and automation. The objective is to cost-effectively produce one specific part type at high volumes and with the required quality.

The emphasis on cost-effective production was supplemented with a focus on improved product quality in the CNC epoch (the 1970s and 1980s). Manufacturing was dramatically affected by the invention of CNC machines as they provide more accurate control and means for better quality. Japanese production techniques such as Kaizen (continuous improvement); just-in-time (JIT) (elimination/minimization of inventory as the ideal goal to reduce costs); lean manufacturing (efficiently eliminate waste, reduce cost, and improve quality control; and total quality management (TQM) (increased and faster communications with customers to meet their requirements) attracted considerable attention. Furthermore, CNC machines provided necessary tools for easier integration/automation which, in turn, contributed to manufacturing of a product family on the same system. Consequently, flexible manufacturing systems (FMSs) were introduced to address changes in work orders, production schedules, part programs, and tooling for the production of a family of parts. The economic objective of an FMS (see Figure 1.2) is to make possible the cost-effective manufacture of several types of parts that can change over time, with shortened changeover time, on the same system at the required volume and quality. It has a fixed hardware and fixed (but programmable) software (see Figure 1.3). In terms of design, the system possesses an integral architecture (hardware/software), i.e., the boundaries between the components and their functionalities are often difficult to identify and are tightly linked together. This type of architecture does not allow for reconfiguration changes to be made. Therefore, an FMS has limited capabilities for upgrading, add-ons, customization, and changes in production capacity.

In the knowledge epoch (i.e., starting in the 1990s), focus shifted to the responsiveness of a manufacturing system characterized by intensified global competition, the fast pace of technological innovations, and enormous progress in computer and information technology (Jaikumar, 1993; Mehrabi

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

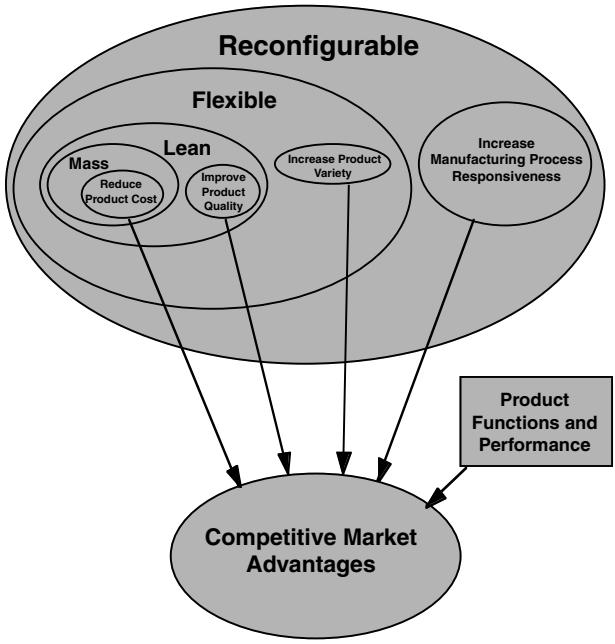


FIGURE 1.2 Economic goals for various manufacturing paradigms.

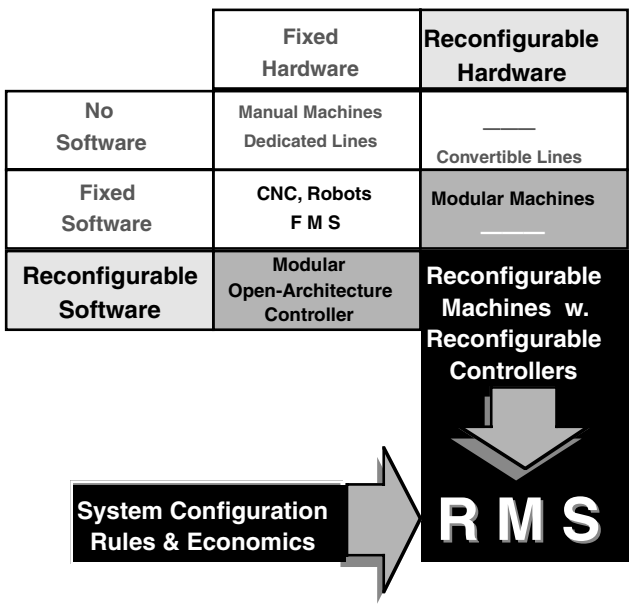


FIGURE 1.3 Key hardware and software features of manufacturing systems.

and Ulsoy, 1997; Mehrabi, Ulsoy, and Koren, 1998). Rapid progress was made in areas such as management information systems, development of software/application programs for various specific purposes, advances in communication systems (hardware and software), and penetration of computer technology in various fields (Gyorki, 1989). Therefore, global competition and information technology are the driving forces behind recent changes in manufacturing. These conditions

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

TABLE 1.1 Summary of Definitions

Systems (Machining/Manufacturing)	Definitions
Machining System	One or more machine tools and tooling, and auxiliary equipment (e.g., material handling, control, communications) that operate in a coordinated manner to produce parts at the required volumes and quality.
Dedicated Machining System (DMS)	A machining system designed for production of a specific part, and uses transfer line technology with fixed tooling and automation.
Flexible Manufacturing System (FMS)	A machining system configuration with fixed hardware and fixed, but programmable, software to handle changes in work orders, production schedules, part programs, and tooling for several types of parts.
Reconfigurable Manufacturing System (RMS)	A machining system that can be created by incorporating basic process modules, both hardware and software, that can be rearranged or replaced quickly and reliably. Reconfiguration will allow adding, removing, or modifying specific process capabilities, controls, software, or machine structure to adjust production capacity in response to changing market demands or technologies. This type of system will provide customized flexibility for a particular part family, and will be open-ended, so that it can be improved, upgraded, and reconfigured, rather than replaced.

Note: A part family is defined as one or more part types with similar dimensions, geometric features, and tolerances, such that they can be produced on the same, or similar, production equipment.

require a responsive manufacturing system that can be rapidly designed, able to convert quickly to the production of new product models, able to adjust capacity quickly, able to integrate process technology, and able to produce an increased variety of products in unpredictable quantities. Agile manufacturing (Goldman, Nagel, and Preiss, 1995) was introduced as a new approach to respond to rapid change due to competition. It brings together individual companies to form an enterprise of manufacturers and their suppliers linked via advanced networks of computers and communication systems. Agile manufacturing, however, does not deal with production system technology or operations.

More recently, reconfigurable manufacturing systems (RMSs) were introduced (Koren and Ulsoy, 1997; Mehrabi and Ulsoy, 1997) to respond to the new market-oriented manufacturing environment. In terms of design, an RMS has a modular structure (software and hardware) that allows ease of reconfiguration as a strategy to adapt to market demands (see Table 1.1). Open-architecture control systems are one of the key enabling technologies of an RMS, and have the ability to integrate/remove new software/hardware modules without affecting the rest of the system. Another key enabling technology is modular machines (Moon and Kota, 1998; Garro and Martin, 1993). System design tools are also needed to properly configure a system from these software and hardware building blocks (see Figure 1.3). This means an RMS has the ability to be converted quickly to the production of new models, to be adjusted rapidly to exact capacity requirements as the market grows and product changes, and to integrate new technology. The objective of an RMS is to provide the functionality and capacity that is needed, when it is needed. Thus, a given RMS configuration can be dedicated or flexible, and can change as needed. An RMS goes beyond the economic objectives of an FMS by permitting: (1) reduction of lead time for launching new systems and reconfiguring existing systems, and (2) the rapid manufacturing modification and quick integration of new technology and/or new functions into existing systems.

1.3 Significance of Functionality/Capacity Adjustments in Modern Manufacturing Systems

Due to the globalization of economies, responsiveness is becoming the cornerstone of manufacturing competitiveness. Therefore, rapid, controlled-cost response to market demands is the key to the success of manufacturing companies. This section is devoted to discussion of the abilities of

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

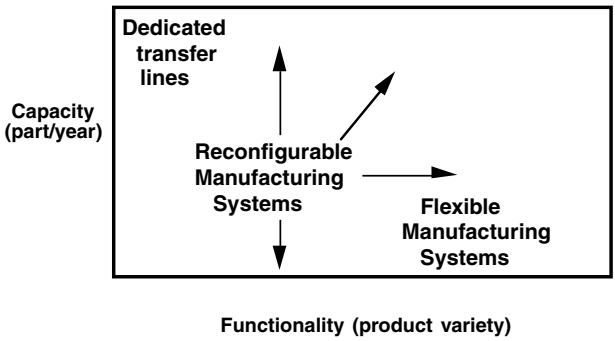


FIGURE 1.4 Mapping several types of manufacturing systems in capacityfunctionality coordinates.

available manufacturing systems in terms of the rapid adjustment of capacity and functionality in response to the market demands. Figure 1.4 provides mapping of the available manufacturing systems in capacity-functionality coordinates. As is shown, dedicated transfer lines typically have high capacity but limited functionality (Koren and Ulsoy, 1997). They are cost effective as long as they produce a limited number of part types and demand exceeds supply. But with saturated markets and the increasing pressure of global competition, situations exist where the dedicated lines do not operate at their full capacity, which creates a loss. Flexible systems, on the other hand, are built with all the flexibility and functionality available, including some cases that may not be needed at installation time. In these cases, capital lies idle on the shop floor and a major portion of the capital investment is wasted. These two types of waste will be eliminated with RMS technology. In the first case, the RMS allows the addition of the extra capacity when required, and in the second case, adds functionality when needed. Referring again to the capacity vs. functionality trade-off in Figure 1.4, the RMSs may, in many cases, occupy a middle ground between DMSs and FMSs. This also raises the possibility of various types of RMSs, with different granularity of the RMS modules that evolve from either DMSs or FMSs, respectively. For example, an RMS can be designed with a CNC machine tool as the basic building block. This would require an evolution of current FMSs through lower-cost, higher-velocity CNC machine tools with modular tooling that also have in-process measurement systems to assure consistent product quality. On the other hand, an RMS can be designed with drive system modules, rather than CNC machines, as the basic building blocks. This would represent an evolution of RMSs from DMSs and require, for example, modular machine tool components and distributed controllers with high bandwidth communication.

1.4 Critical Role of Computers in Modern Manufacturing

A number of steps are involved in manufacturing a part from its conceptualization to production. They include product design, process planning, production system design, and process control. Computers are used extensively in all these stages to make the entire process easier and faster. Potential benefits of using computers in manufacturing include reduced costs and lead times in all engineering design stages, improved quality and accuracy, minimization of errors and their duplication, more efficient analysis tool, and accurate control and monitoring of the machines/processes, etc. Some of the applications of computers in manufacturing are shown in Figure 1.5. In computer-aided design (CAD), computers are used in the design and analysis of the products and processes. They play a critical role in reducing lead time and cost at the design stages of the products/process. Also, computers may be utilized to plan, manage, and control the operations of a manufacturing system: computer-aided manufacturing (CAM) (Bedworth, Handerson, and Wolfe, 1991). In CAM, computers are either used directly to control and monitor the machines/processes (in real-time) or used off-line to support manufacturing operations such as computer-aided process planning (CAPP)

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

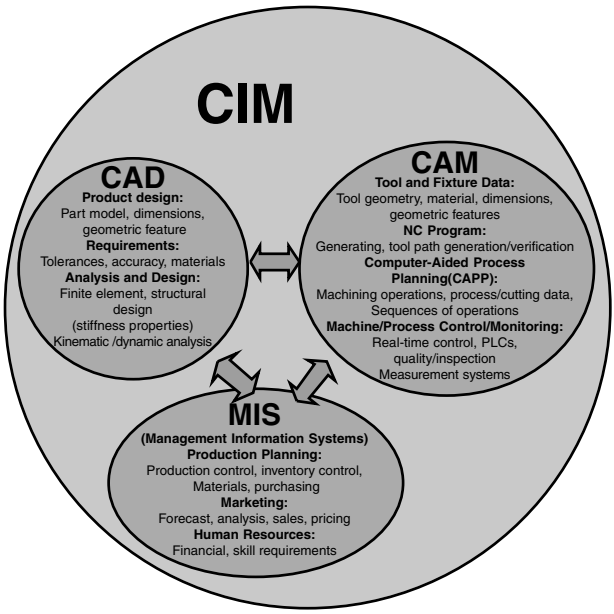


FIGURE 1.5 Applications of computer technology in manufacturing.

or planning of required materials. At higher levels, computers are utilized in support of management. They play a critical role in all stages of decision making and control of financial operations by processing and analyzing data and reporting the results (management information systems, MIS) (Hollingam, 1987). Computers facilitate integration of CAD, CAM, and MIS (computer-integrated manufacturing, CIM) (Vajpayee, 1995) (see Figure 1.5). They provide an effective communication interface among engineers, design, management, production workers, and project groups to improve efficiency and productivity of the entire system.

1.5 Design Principles of Modern Manufacturing Systems

Manufacturing is a complex process that begins with evaluating the market and investigating the demands for a product, and ends with delivery of the actual product. Successful marketing should take into account the factors that affect current and future demands for a product. It provides management with appropriate inputs for decision making and directing resources of a company toward production of a part that is needed in the market. This sets the stage for product design and manufacturing as described in the following sections.

1.5.1 Product Design and Design for Manufacturability

At the product design stage, designers and product engineers generate new ideas and study various aspects of design. Also, production engineers investigate the availability of the resources and capabilities of the production system. CAD systems are extensively used at this stage for rapid design and revisions of a product (Groover and Zimmers, 1984). Designs for manufacturability (DFM) and assembly are used to emphasize the significance of the links between design of a product and its manufacturing (Beckert, 1990). Design for manufacturing focuses on appropriate product design, process planning, and manufacturing to ensure optimum results (Vajpayee, 1995). It emphasizes the importance of quality and its relation with the machines/processes accuracy of machined (produced) parts tolerances, and correction of a product defect at the design stage (as opposed to after production) and its significant impact on cost of a product.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

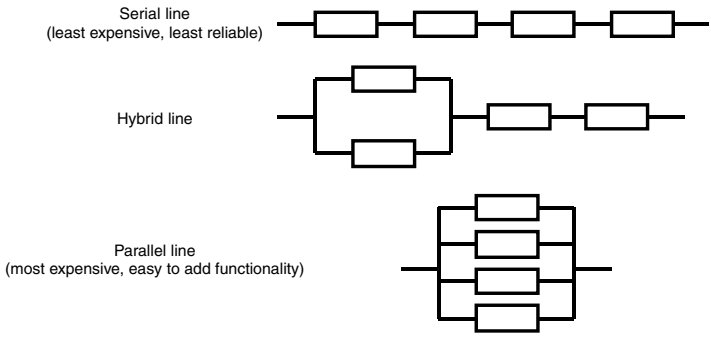


FIGURE 1.6 Several possible configurations with four machines.

1.5.2 Process Planning and System Design of Manufacturing Systems

Once a product design is completed, it is produced by using machines and other equipment (e.g., material handling) and resources. Computers are used extensively to identify optimal machining configurations by taking into account the cost, quality, and reliability of the entire system (see Figure 1.6), control the activities of planning and distributing the sequence of operations among the machines, and to specify machining parameters such as feed, speed, etc., computer-aided process planning (CAPP) (Bedworth, Handerson, and Wolfe, 1991; Vajpayee, 1995).

Two basic approaches to CAPP exist, variant and regenerative. The variant technique is used mostly for process planning of a family of products. With this technique, group technology (GT) is used to create and classify the plans (for a family of parts), and store them in a database. For the next design, the required plans are retrieved from the database already created for this family of parts (Groover and Zimmers, 1984). With the regenerative method, process plans are produced for every new product and as such, no database of plans exists (Gyorki, 1989; Vajpayee, 1995). It is more sophisticated than the variant method and has the advantage of facilitating integration of process planning stage with product design while the needs for human experts are minimized or totally eliminated.

1.5.3 Software/Hardware Architecture and Communications in Manufacturing Systems

An integral part of a manufacturing system is the software required to handle tasks at various levels such as control, monitoring, and communications among mechanical, electrical, and electronic components (low level) as well as higher level tasks such as process planning, user interface, process control, data collection/report from the process, etc. Therefore, the structure and functionality of the control software are very critical and directly affect the performance of the entire system. The controllers of the machines, networking and data communication between CNC controller/PLC (programmable logic controllers) or PLC/PLC, have been through proprietary networks (similar situation as with controllers); i.e., related control systems, communication systems, protocols, and software/hardware are not open to users or other vendors (Aronson, 1997; Altintas and Munasinghe, 1996). Therefore, further system enhancements, integration of sensors, and new technologies are severely restricted. Open-architecture principles and systems are introduced to accommodate these features (see Figure 1.7).

Another critical issue in the design of modern intelligent manufacturing systems is communication. Let us consider a set of sensors/devices communicating with a central computer/controller. Traditionally, they should be hard-wired to the central controller/PLC; therefore, the costs associated with wiring, connections, control cabinet, space, labor, maintenance, and trouble shooting are quite

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

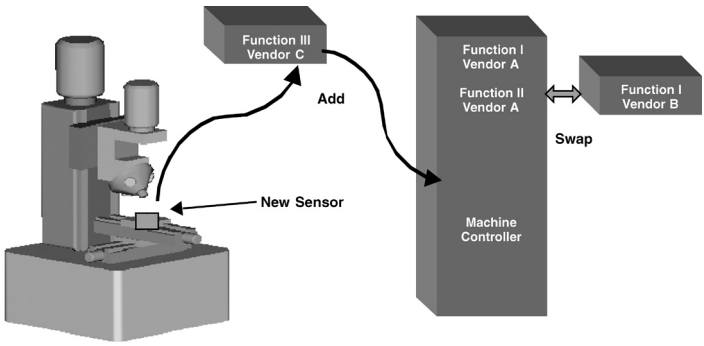


FIGURE 1.7 Open-architecture principle in machine tool control systems.

high. With a proper communication system, the same sensor/device is connected to a network (locally) which takes care of all data reporting and condition monitoring of the entire manufacturing system.

Recent developments in built-in intelligent control devices and communication networks, such as Devicenet, address some of these issues (Proctor and Albus, 1997; Proctor and Micholski, 1993). In the Devicenet network, local devices have built-in intelligence (with little cost) and their communication capabilities are enhanced. Therefore, control decisions/actions are made locally and the entire control system for manufacturing is decentralized. Also, progress is made in the development of standard terminology for message and instruction sets, such as manufacturing message specification (MMS), which is necessary for shop floor communication.

1.5.4 Monitoring and Control of Manufacturing Systems

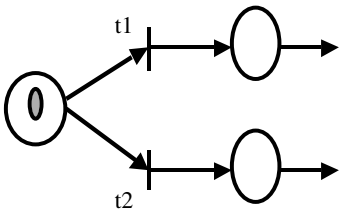
One of the key factors in evaluating product quality is precision in machining. To achieve that, the cutting operation is tightly controlled by using real-time data collected from sensors located at different locations of the workpiece, tool, and machine. Also, some measurements are made for process monitoring purposes with the objective of preventing irrepairable damages to the workpiece and the machine. In general, real-time measurements of the following variables are required: dimensional errors, quality of surface finish, thermal deformations during machining, and dynamic deformations of the workpiece; chatter vibration, cutting force, condition of the chip, and identification of the cutting for process monitoring; thermal deformation, dynamic deformation of the machine elements, and structural vibration of the machine tool and wear, failure, and thermal deformations of the tool (Rangwala and Dornfeld, 1990; Li and Elbestawi, 1996).

Currently, commercially available controllers of CNC machines have been equipped with proprietary control systems; i.e., the users do not have access to the controller and further modifications/enhancements of the system (by the users) are either impossible or very costly. This has significantly hindered the applications of efficient control algorithms, addition of new sensors for process improvement/monitoring purposes, and has suppressed the automation of the entire production system. PC-based control systems (Koren et al., 1998; Hollenback, 1996) are the answer to the limitations mentioned above; they are very suitable for operating in an open-architecture environment (see Figure 1.7).

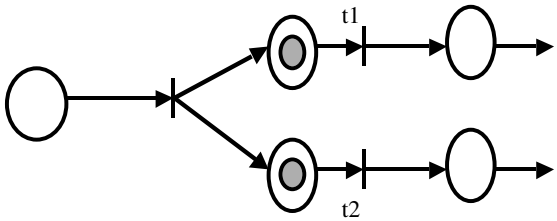
The same view is valid for programmable logic controllers (PLCs). To date, PLCs have been used in industrial automation to control and monitor discrete event systems. The functionality of PLCs can be enhanced, however, by proper implementation of available I/O boards (and compatible software) on a much more compact and industrial PC platform such as PC/104. This offers the advantage of integrating the functional logic (discrete) of PLCs and machine-tools' motion control (continuous) by utilizing modeling capabilities of Petri nets (Park et al., 1998) (see Figure 1.8).

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com



(a) Conflict: if t1 fires, t2 is not enabled and visa versa



(b) Concurrency: t1and t2 can be fired independently

FIGURE 1.8 Examples of modeling capabilities of Petri nets.

1.6 Future Trends and Research Directions

It is very difficult to forecast long-term trends for manufacturing systems, because the changes are happening at a very fast pace. However, it is possible to extrapolate future trends from the current situation by analyzing and specifying the key drivers behind the changes. Certainly, availability and distribution of information play an important role in this transition and are considered key drivers. In this regard, the need for improvements and standardization of various components (such as data interfaces, protocols, communication systems, etc.) exists so that data can be transferred to the desired location at a faster rate (Agility Forum, 1997).

There are many research efforts underway; however, we are still at the beginning of a new era of modern manufacturing systems, and there are many barriers to their advancement. Advances in manufacturing will not occur without the proper machine tools and equipment. Machine tools are undergoing some fundamental changes in terms of their structure (modular structure) and components (controllers, hardware/software, spindles, tooling, sensors, etc.). Therefore, new theories, design concepts, and methodologies should be developed for these purposes (Garro and Martin, 1993; Lee, 1997; Moon and Kota, 1998). These changes are fundamental to the success of future manufacturing systems.

Selected References

Agility Forum, 1997, *Next-Generation Manufacturing: A Framework for Action*, Bethlehem, PA.

Altintas, Y. and W.K. Munasinghe, 1996, A hierarchical open-architecture CNC system for machine tools, *Annals of the CIRP*, 43, 1, 349–354.

Aronson, R.B., 1997, Operation plug-and-play is on the way, *Manufacturing Engineering*, 118, 3, 108–112.

Beckert, B., 1990, Integrated manufacturing: New wizards of management, *Industry Week*, 239, 6, 60–84.

Bedworth, D.D., Handerson, M.R., and P. Wolfe, 1991, *Computer Integrated Design and Manufacturing*, McGraw-Hill, New York.

Cohen, S.S., 1987, *Manufacturing Matters: The Myth of the Post-Industrial Economy*, Basic Books, New York.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Garro, O., and P. Martin, 1993, Towards new architecture of machine tools, *International Journal of Production Research*, 31, 10, 2403–2414.

Goldman, S.L., Nagel, R.N., and K. Preiss, 1995, *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*, Van Nostrand Reinhold, New York.

Groover, M. and E.W. Zimmers, 1984, *CAD/CAM, Computer-Aided Design and Manufacturing*, Prentice Hall, Englewood Cliffs, New Jersey.

Gyorki, J.R., 1989, How to succeed CIM, *Machine Design*, 42, 1, 99–105.

Hollenback, D., 1996, PCs provide the foundation for open architecture, *Control Engineering*, 75–78.

Hollingum, J., 1987, *Implementing an Information Strategy in Manufacturing*, IFS Publication, UK.

Jaikumar, R., 1993, 200 years to CIM, *IEEE Spectrum,* 26–27.

Koren,Y., Jovane, F., and G. Pritschow (Eds.), 1998, *Open Architecture Control Systems*, ITIA Series, Italy.

Koren, Y. and A.G. Ulsoy, 1997, Reconfigurable Manufacturing Systems, Engineering Research Center for Reconfigurable Machining Systems (ERC/RMS) Report #1, The University of Michigan, Ann Arbor.

Lee, G.H., 1997, Reconfigurability consideration design of components and manufacturing systems, *International Journal of Advanced Manufacturing Technology*, 13, 5, 376–386.

Li, S., and M.A. Elbestawi, 1996, Tool condition monitoring in machining by fuzzy neural networks, *ASME Journal of Dynamic Systems, Measurement, and Control*, 118, 4, 665–672.

Mehrabi, M.G., Ulsoy, A.G., and Y. Koren, 1998, Reconfigurable manufacturing systems: Key to future manufacturing, *Proceedings of the 1998 Japan–U.S. Symposium on Flexible Automation*, Otsu, Japan, 677–682.

Mehrabi, M.G. and A.G. Ulsoy, 1997, State-of-the-Art in Reconfigurable Manufacturing Systems, ERC/RMS Report #2, Vol. I, Engineering Research Center for Reconfigurable Machining Systems (ERC/RMS), The University of Michigan, Ann Arbor.

Moon, Y. and S. Kota, Generalized kinematic modeling method for reconfigurable machine tools, *ASME DETC 98*, Atlanta, GA, Paper number MECH-5946, Sept. 1998.

Park, I., D. Tilbury, and P.P. Khargonekar, 1998, A Formal Implementation of Logic Controllers for Machining Systems Using Petri Nets and Sequential Function Charts, presented at the 1998 Japan–U.S. Symposium on Flexible Automation, Otsu, Japan.

Proctor, F.M. and J.S. Albus, 1997, Open-architecture controllers, *IEEE Spectrum,* 34, 6, 60–64.

Proctor, F.M. and J. Micholski, 1993, Enhanced Machine Controller Architecture Overview, NISTIR-5331, NIST Tech. Rep., Gaitherburg, MD.

Rangwala, S. and D.A. Dornfeld, 1990, Sensor integration using neural networks for intelligent tool condition monitoring, *ASME Journal of Engineering for Industry*, 112, 219–228.

Vajpayee, K. 1995, *Principles of Computer-Integrated Manufacturing*, Prentice Hall, Englewood Cliffs, New Jersey.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

2

Computer-Aided Process Planning for Machining

2.1 Introduction
2.2 What Is Computer-Aided Process Planning (CAPP)?
2.3 Review of CAPP Systems
Variant Planning • Generative Planning • Hybrid Planning • Artificial Intelligence (AI) Approaches • Object-Oriented Approaches • Part Geometry • Part Specification Input
2.4 Drivers of CAPP System Development
Design Automation • Manufacturing Automation • Extension of Planning Domains; New Planning Domains • Market Conditions • Summary of Drivers
2.5 Characteristics of CAPP Systems
2.6 Integrating CAD with CAPP: Feature Extraction
What Are Features? • Feature Recognition • Discussion
2.7 Integrating CAPP with Manufacturing
NC Tool-Path Generation • Manufacturing Data and Knowledge
2.8 CAPP for New Domains
Parallel Machining
2.9 Conclusions

Derek Yip-Hoi
University of Michigan

Abstract

This chapter presents an overview of the research work in computer-aided process planning (CAPP) during the past 2 decades. This has been driven primarily by the need to automate the mapping of design information and intent from computer-aided design (CAD) systems to instructions for driving automated manufacturing equipment. While the concept of CAPP extends over all manufacturing domains, we summarize those developments primarily in the machining domain. As part of CAPP research, we also discuss developments in the area of feature recognition. Features are fast becoming the mechanism through which higher level design information is embodied and manipulated within the computer-aided engineering (CAE) environment. Feature recognition is one mechanism by which this higher level of abstraction is constructed and related to the underlying geometry. Finally, we briefly introduce a new area of research in CAPP, parallel machining.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

2.1 Introduction

The past decade has seen an explosion in the use of computers throughout all engineering diciplines. This is particularly true in the activities that span the life cycle of discrete product development. Commercial viability of computer-based tools has occurred at either end of the product life cycle, i.e., in product design and in manufacturing. In product design, previously expensive CAD systems are now affordable and run on ever cheaper and more computationally powerful PCs, which makes this technology more widely accessible to an evergrowing number of users. In addition, the sophistication of these systems has increased dramatically. Whereas the initial first-generation CAD system was primarily concerned with wireframe modeling and automated drafting, current third-generation systems are incorporating features technology built on top of powerful geometric/solid modeling engines (second-generation systems).

As explosive as the CAD side of product development has been, so has that in manufacturing automation. With the advent of cheaper computers and controllers, an increasing percentage of machines used in the modern factory is software controlled and interconnected through networks. This greatly reduces the length of time during which a machine tool or robot can theoretically be reprogrammed for a new task, thus increasing productivity. Practically, these increases are yet to be realized because of the lead time required to convert design information into programs to drive these machines. Computer-aided process planning (CAPP) systems enable shorter lead times and enhanced productivity in the automated factory.

In the following sections, we discuss research developments in CAPP systems during the past 2 decades. While much research has been done, commercialization of this technology is yet to be realized in the same way that other CAE technologies have experienced.

2.2 What Is Computer-Aided Process Planning (CAPP)?

In this section we introduce the topic of CAPP, and review important components of this technology.

Chang and Wysk (1985) define process planning as "machining processes and parameters that are to be used to convert (machine) a workpiece from its initial form to a final form predetermined from an engineering drawing." Implicit in their definition is the selection of machining resources (machine and cutting tools), the specification of setups and fixturing, and the generation of operation sequences and numerical control (NC) code. Traditionally, the task of process planning is performed by a human process planner with acquired expertise in machining practices who determines from a part's engineering drawings what the machining requirements are.

Manual process planning has many drawbacks. In particular, it is a slow, repetitive task that is prone to error. With industry's emphasis on automation for improved productivity and quality, computerized CAD and computer-aided manufacturing (CAM) systems which generate the data for driving computer numerical control (CNC) machine tools, are the state-of-the-art. Manual process planning in this context is a bottleneck to the information flow between design and manufacturing.

CAPP is the use of computerized software and hardware systems for automating the process planning task. The objective is to increase productivity and quality by improving the speed and accuracy of process planning through automation of as many manual tasks as possible. CAPP will increase automation and promote integration among the following tasks:

1. Recognition of machining features and the construction of their associated machining volumes from a geometric CAD model of the part and workpiece
2. Mapping machining volumes to machining operations
3. Assigning operations to cutting tools
4. Determining setups and fixturing

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

5. Selecting suitable machine tools
6. Generating cost-effective machining sequences
7. Determining the machining parameters for each operation
8. Generating cutter location data and finally NC machine code

Traditionally, CAPP has been approached in two ways. These two approaches are variant process planning and generative process planning. In the following section we discuss these and other issues in a review of work in this field.

2.3 Review of CAPP Systems

The immense body of work done in the field of CAPP makes it impossible to discuss each development in detail within the confines of this chapter. We, therefore, direct the reader to Alting and Zhang (1989), CAM-I (1989), and Kiritsis (1995) for detailed surveys of the state-of-the-art in CAPP. Eversheim and Schneewind (1993) and ElMaraghy (1993) provide good perspectives on the future developments of CAPP. It is worth mentioning that although the surveys by Alting and Zhang (1989) and CAM-I (1989) are over 12 years old, they came at a time when most of the basic foundation for CAPP system development had already been laid. Although new researchers have entered the field, these surveys still provide valuable insight to the problem. Kiritsis (1995) provides a later survey that focuses on systems that are knowledge based. He also classifies the feature recognition approach that is used for each reviewed CAPP system. The perspectives proposed by Eversheim et al. (1993) and ElMaraghy (1993) are directed toward a second generation of CAPP systems. The characteristics of these second generation systems are summarized in Section 2.5.

Figure 2.1 is a chronology of CAPP system developments through the 1980s until 1995, showing some of the more well-known contributions. In addition to indicating the year when each initiative began, the figure also lists the characteristics of each system. These characteristics include among others, the planning methodology adopted and the planning domain that is targeted. In the following sections we discuss a subset of the most important characteristics.

2.3.1 Variant Planning

The variant planning approach was the first to be adopted by CAPP system developers. This approach, as the name implies, creates a process plan as a variant of an existing plan. The most common technique used to implement this approach is group technology (GT). GT uses similarities between parts to classify them into part families. When applied to machining process planning, a part family consists of a set of parts that have similar machining requirements. In addition to part family classes, two other ingredients are necessary for variant process planning: a coding scheme for describing parts, and a generic process plan for each part family.

Whenever a process plan is needed for a new part, the part in question is mapped to a part code. This code is then compared with a code associated with each part family class. If a match is found, the plan for the matched family is retrieved. It is then modified to suit the new part.

The variant approach has obvious disadvantages. The most glaring is the dependence for success on the existence of a family with which a match can be made. This means that new parts with significantly different characteristics than any found in the database must be planned from scratch. Another major disadvantage of the variant approach is the cost involved in creating and maintaining databases for the part families. Due to these problems, variant systems are normally adopted only when a well-defined part family class structure exists, and it is expected that new parts will generally conform closely to the characteristics of these classes.

Variant systems developed in-house have been widely implemented throughout industry. Examples include CAPP, (Link, 1976) GENPLAN, (Tulkoff, 1981), and GTWORK (Joshi et al., 1994).

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

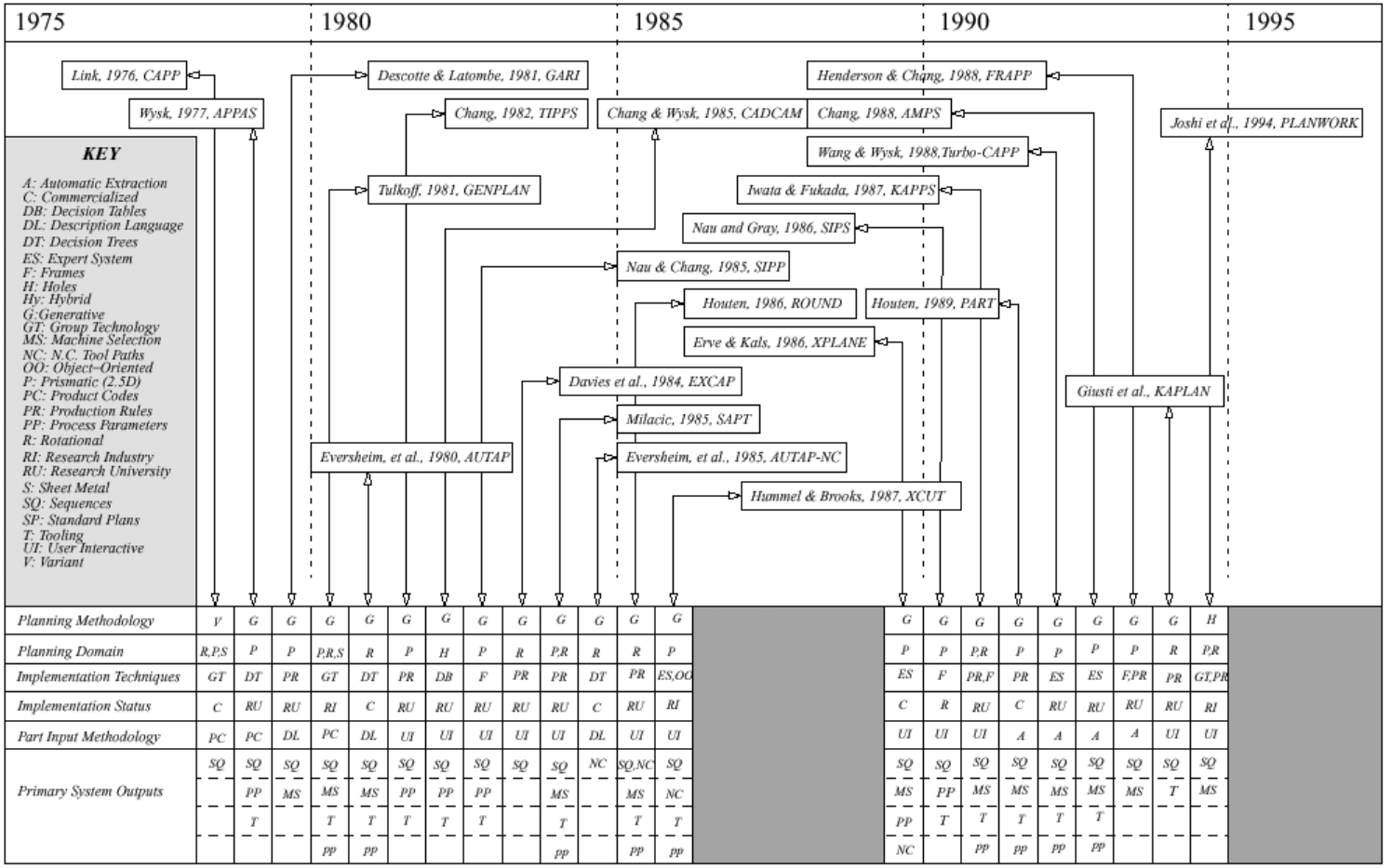


FIGURE 2.1 CAPP system development chronology.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

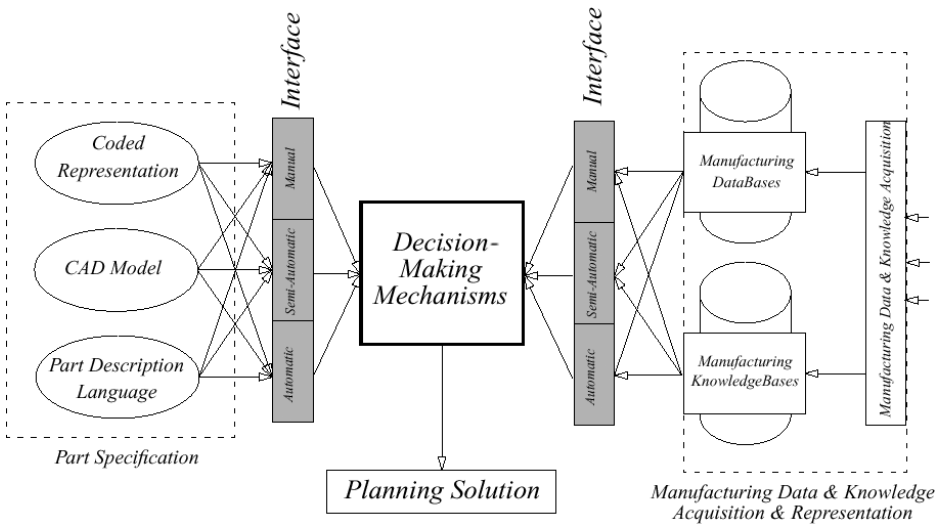


FIGURE 2.2 Components of a generative CAPP system.

2.3.2 Generative Planning

Generative planning creates unique process plans from scratch for each new part, utilizing algorithmic techniques, process knowledge, process data, and the geometric and technological specifications of the part. In contrast to the variant approach, generative planning does not use a generic family plan as the starting point. Experiential knowledge is applied through the use of techniques such as decision tables, decision trees, or production rules which can be customized to fit specific planning environments. The key components of a generative CAPP system are illustrated in Figure 2.2. They are

- *Part Specification Input*: See Section 2.3.7.
- *Manufacturing Data and Knowledge Acquisition and Representation:* In the machining domain this refers to the data and knowledge that are commonly applied by human process planners in planning machining operations. In this context, examples of manufacturing data are the machining process parameters stored in a database or derived from formulae constructed from machinability experiments. Examples of machining knowledge are the rules that match machining requirements based on part specifications to process capabilities.
- *Decision-Making Mechanisms:* These are the techniques used to generate a process plan given the part specifications and the available manufacturing data and knowledge. Examples of these mechanisms include hard-coded procedural algorithms, decision trees and tables, and production rules. The actual decision-making mechanism is likely to be a hybrid combination of different types of reasoning mechanisms.

Generative process planning systems are not necessarily fully automatic. Chang (1990) used the term automatic process planning to define systems with (1) an automated CAD interface, and (2) a complete and intelligent planning mechanism. Because these are the two major high-level tasks in planning, these systems eliminate human decision making. The current state-of-the-art is such that no CAPP system, either research or commercial, can claim to be fully automatic.

A major advantage of generative CAPP systems over variant systems is that they can provide a planning solution for a part for which no explicit manufacturing history exists, i.e., no variant of the part has an existing plan which may be retrieved and modified. Another advantage is the generation of more consistent process plans. While these advantages seem to weigh heavily in favor

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

of generative planning solutions, the practical problems to be overcome are formidable. The computerization of manufacturing knowledge (its acquisition, representation, and utilization), in particular, is difficult. A high level of expertise is currently required to build and maintain knowledge bases. Cost effectiveness and confidence in such systems are not yet at a state where commercialization is viable. Examples of generative CAPP systems are APPAS (Wysk, 1977),TIPPS (Chang, 1982), EXCAP (Davies et al., 1988), SIPS (Nau and Gray, 1986), XPLANE (Erve and Kals, 1986) XCUT (Hummel and Brooks, 1986; 1988; Brooks et al., 1987), and PART (Houten and Erve, 1988; 1989a; 1989b; Houten et al., 1990).

2.3.3 Hybrid Planning

While fully generative process planning is the goal of CAPP system development, in the interim, systems that combine the variant and generative planning approaches are useful. We refer to these as hybrid planners. Another term used to refer to this approach is semi-generative plannign (Alting and Zhang, 1989). A hybrid planner, for example, might use a variant, GT-based approach to retrieve an existing process plan, and generative techniques for modifying this plan to suit the new part (Joshi et al., 1994).

One important aspect of hybrid planning is user interaction. As generative CAPP systems become more and more automatic, the amount of work a process planner needs to do will decrease. However, this trend should not lead to a process planning system that removes the human planner from the roles of arbitrator and editor. The human planner should always have the ability to modify and influence the CAPP system's decisions. This leads to a hybrid planning approach where two parallel planning streams exist. The first utilizes generative planning techniques, and the second a user-interaction approach. User interaction acts either to bypass generative planning functions or becomes part of feedback loops in an evaluate-and-update cycle. In this way, the user always has control over the planner and makes the final decisions when conflicts arise that cannot be resolved automatically.

2.3.4 Artificial Intelligence (AI) Approaches

Since the early 1980s, AI techniques have found widespread application in CAPP work. They have been applied both at the feature recognition stage and in capturing best machining practices for the purposes of operation selection and sequencing, resource selection, and process plan evaluation. Expert systems have been the main AI tool used in CAPP work. These systems combine domain data, knowledge (rules), and an inference mechanism for drawing conclusions about a planning problem. Expert systems are based on nonprocedural programming in contrast to the procedural approach of more conventional programming languages such as Basic, Fortran, or C. This makes them especially suited for domains where algorithms are difficult to structure and where high uncertainty exists.

Knowledge representation schemes used in expert systems include production rules, frames, semantic nets, predicate logic, and neural networks. Of these, the most commonly used are production rules and frames. CAPP systems that use production rules include GARI (Descotte and Latombe, 1981) (one of the first AI-based CAPP systems), TIPPS (Chang, 1982), SAPT (Milacic, 1985; 1988), XCUT (Hummel and Brooks, 1986), Turbo-CAPP (Wang and Wysk, 1987), Hi-Mapp (Berenji and Khoshnevis, 1986), and FRAPP (Henderson and Chang, 1988). Systems that use frames include SIPP (Nau and Gray, 1986), Hi-Mapp (Berenji and Khoshnevis, 1986), FRAPP (Henderson and Chang, 1988) and QTC (Chang et al., 1988).

2.3.5 Object-Oriented Approaches

Object-oriented programming is often associated with artificial intelligence. They provide a technique by which data and methods can be encapsulated within an object. Encapsulation masks the

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

inner workings of the object behind an interface through which the objects communicate with each other and the rest of the world. Inheritance allows objects to be ordered hierarchically such that they inherit data and methods from their ancestors.

One of the most powerful features of object-oriented programming is the ability to separate the calling program or application from the inner workings of objects. The calling program interacts with objects through the use of message handlers (member functions in the case of C++). This interface allows objects to be changed without the need to modify the application program in which the objects are used. This is particularly useful in situations where objects are changing or evolving, as is usually the case in the CAPP domain.

Object-oriented programming has been integrated into expert system shells. CLIPS™ (C Language Integrated Production System* (Giarrantano and Riley, 1989) is an example of this. COOL™ (CLIPS' Object-Oriented Language) allows the knowledge engineer to represent data as objects and manipulate these objects within production rules. This is a great help in structuring and managing the knowledge base. XCUT (Hummel and Brooks, 1986) is an example of a CAPP system which uses a rule-based expert system with an embedded object-oriented language. Other researchers who have utilized the object-oriented paradigm include Turner and Anderson (1988), Lee et al. (1991), and Yut and Chang (1994).

2.3.6 Part Geometry

Almost all CAPP research work in the machining domain focuses on either rotational or prismatic (2.5D milled) part geometries. Systems that generate plans for rotational parts include MICROPLAN (Philips et al., 1986), DMAP (Wong et al., 1986), ROUND (Houten, 1986), and EXCAP (Davies et al., 1988). Examples of systems that generate plans for prismatic parts include GARI (Descotte and Latombe, 1981), TIPPS (Chang, 1982) SAPT (Milacic, 1985) Hi-Mapp (Berenji and Khoshnevis, 1986), SIPS (Nau and Gray, 1986), XCUT (Brooks et al., 1987) and PART (Houten and Erve, 1988; 1989a; 1989b; Houten et al., 1990).

2.3.7 Part Specification Input

The front end to a generative planning system is designed to input the part specification. Various approaches have been adopted for this step. Some approaches use coding schemes similar to those found in many variant planning systems to describe the part. One example is that adopted by Wysk (1977) as part of the APPAS generative planning system. The coding scheme in this work is called COFORM (Rose, 1977) and is used to generate a coded description of each individual machined surface of a part. The surface's coded attributes are subsequently used to drive process selection in the generative planner.

Another approach to part specification input is through the use of a part description language which translates the basic part geometry into a higher level format that can be used by the process planning system. Technological information (surface finishes, tolerances) also can be included. Examples of this approach to part input can be found in GARI (Descotte and Latombe, 1981) and AUTAP-NC (Eversheim and Holtz, 1982). One of the problems encountered in using part description languages and codes in the earlier systems was that the information for each part needed to be prepared manually. This was both time consuming and prone to error. With CAD systems, it is now possible to write a translator to automatically or interactively create the part description file.

The widespread use of solid modeling in CAD now makes this the preferred choice for part specification input. However, because part modeling and planning tools (e.g., expert system shells) generally are not designed to work as an integrated environment, the information within CAD

*CLIPS™ and COOL™ are components of an expert system shell developed at the Software Technology Branch of the Lyndon B. Johnson Space Center.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

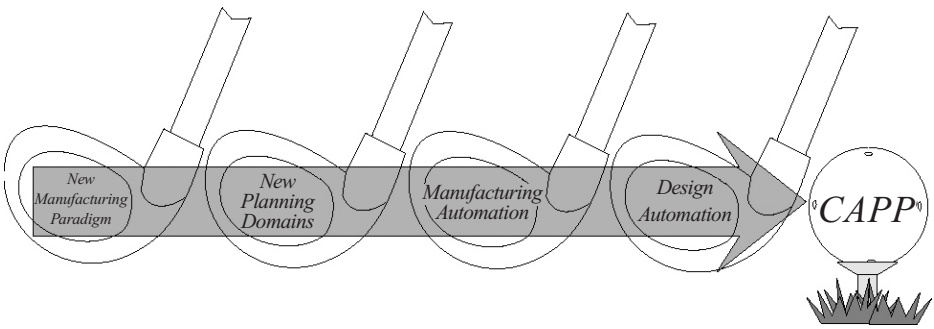


FIGURE 2.3 Drivers of CAPP.

models must still be translated to some representation within the planning environment (e.g., frame or object instances). A truly integrated system will allow the planning mechanisms (rules or methods) to directly interrogate the CAD model.

2.4 Drivers of CAPP System Development

In the previous section we reviewed work in CAPP. In this section we briefly discuss the drivers of CAPP system development. This discussion shows that continual advances in design and manufacturing automation, the emergence of new planning domains, and ever-changing market conditions call for new and improved CAPP tools. As illustrated in Figure 2.3, developments in CAPP are driven primarily by

- Design automation
- Manufacturing automation
- Extension of planning domains; new planning domains
- Market conditions

2.4.1 Design Automation

Design automation closely parallels advances in computer hardware and software. In particular, design automation is driven by advances in CAD. The growth of CAD software development remained strong throughout the 1990s. The following trends are largely responsible for this growth:

- More computing power for less cost
- The use of solid modeling as an integral part of CAD systems
- CAD software migration from UNIX systems to PC platforms
- Feature-based CAD systems

The result of these trends is that powerful CAD systems are now available to a much wider range of end-users than ever before. With a large proportion of CAD systems being links in the production cycle, a corresponding increase in the need to convert CAD product models quickly and easily into manufacturing data exists.

2.4.2 Manufacturing Automation

As with design automation, trends in manufacturing automation are geared toward improving the speed, efficiency, predictability, reliability, and quality of manufacturing processes. Machining systems in particular are an example of this trend. The mill/turn is one machining system that

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

represents the state-of-the-art in manufacturing automation. At the same time, severe restrictions exist on the utilization of this type of complex machining system because of the lack of automated process planning tools. This work is, in fact, an example of how advances in manufacturing automation are driving CAPP system development.

2.4.3 Extension of Planning Domains; New Planning Domains

Developments in CAPP are always driven by the introduction of new planning domains and the extension of old ones. Most of the work to date in CAPP has focused on process planning for machining. New planning domains, on the other hand, arise when new processes are created. An example of a new process is layered manufacturing. This process creates parts a layer or slice at a time. Researchers are looking at a broad range of issues which can be regarded as process planning for this new domain. They include adaptive slicing, locating the optimal part orientation, and the generation of support structures.

2.4.4 Market Conditions

What is eventually manufactured is dictated to a large extent by demand. The market conditions that reflect demand usher in new manufacturing paradigms from time to time. These paradigm shifts are the manufacturing sector adapting to market forces so as to remain viable and competitive. According to analysts (e.g., Pine, 1993), the mass production system that characterized manufacturing from the 1960s through the 1980s is giving way to a new paradigm, one of mass customization, in which traditional, standardized products are replaced by those customized to individual consumer needs and preferences. This leads to the fragmentation of homogeneous markets with subsequent reductions in product development time and overall life cycles.

CAPP is a crucial piece of the puzzle in creating a manufacturing environment that is responsive to mass customization. An ability to create customizable CAD models (using features and parametric modeling, for example) needs to be matched with an ability to generate manufacturing data for those models just as quickly. Without efficient CAPP systems for mapping design specifications to manufacturing instructions, design and manufacturing environments that are separately responsive to customized production are largely unresponsive when integrated.

2.4.5 Summary of Drivers

From the above discussion, the following can be said about the drivers of CAPP system development:

- Advances in design and manufacturing automation continue to call for better CAPP tools.
- CAPP development is needed for extensions to existing domains (machining) and to provide automation for new domains.
- The move toward mass customization in manufacturing requires CAPP systems that are compatible with tools in design and manufacturing environments that are responsive to customized product development.

Figure 2.4 illustrates the view of CAPP as both an interface and a bottleneck between CAD and CAM. While it is likely that CAPP will remain the weakest of the three, the drivers we have discussed are challenging CAPP system developers to make the bottleneck as wide as possible.

2.5 Characteristics of CAPP Systems

In the previous section we looked at the drivers of CAPP system development. In this section we present a set of CAPP system characteristics that are required if these systems are to become viable, integrated parts of production environments. We do this by first presenting our perspectives on

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

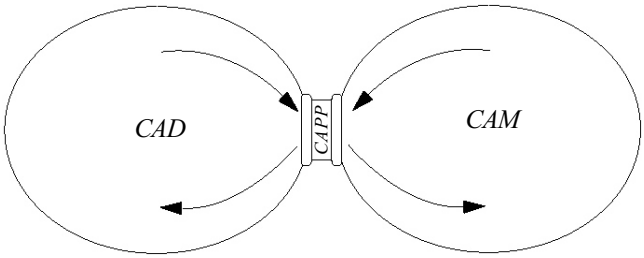


FIGURE 2.4 CAPP bottleneck between CAD and CAM.

CAPP systems based on experiences from research in the field. These perspectives along with their relevance to the key characteristics of CAPP systems are presented in Table 2.1.

A major problem that has affected the evolution of CAPP systems toward commercialization is that many systems have been implemented using a prototype philosophy. With this approach a tendency exists to neglect important practical concerns which greatly affect the nature of the conceptual and implemented models. Because the ultimate goal is to provide an end-user with a practical CAPP solution, these concerns must be addressed if these systems are to become commercially viable. The perspectives presented in Table 2.1 address many of these concerns.

Table 2.2 brings this discussion full circle. It summarizes the characteristics presented in Table 2.1 (plus a few others) and indicates the effect(s) of the characteristic. These effects in turn address the perspectives presented in Table 2.1.

2.6 Integrating CAD with CAPP: Feature Extraction

A considerable amount of research effort has been invested in integrating CAPP with CAD. A major component of this task is the extraction of machining features from a CAD representation of the product. This is an essential step in improving the speed at which design information is converted into manufacturing instructions during process planning. This section reviews some of the important research contributions in this field.

2.6.1 What Are Features?

The term feature is now commonly used in engineering jargon. The first use of the term was, however, in the context of process planning. One of the earliest definitions of a feature can be found in CAM-I:41 A specific geometric configuration formed on the surface, edge, or corner of a workpiece.

The use of the term workpiece in the definition shows the relation to the machining domain. Other researchers who have linked their definition of a feature to the manufacturing domain include CAM-I (1986), Chang et al. (1988), Henderson (1984), Hummel and Brooks (1986), Turner and Anderson (1988), and Vandenbrande (1990).

Since its inception in the process planning domain features, technology has evolved to encompass a much broader range of definitions. The following terms are examples of some definitions that are relevant to this work (for a more comprehensive list of feature terms, see Shah (1991):

Form Feature: First used in the process planning domain. Form features are defined based on their geometry and not their function. Examples of form features include holes, slots, steps, and pockets.

Manufacturing Feature: A feature that is meaningful within a manufacturing domain. Although the machining domain is the most common, researchers also have looked at other domains including features in sheet metal manufacture.

Machining Feature: A feature that is generated by a machining process.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

Volumetric Feature: A volumetric feature consists of a connected solid entity that corresponds to a removal (sub-) volume for a particular manufacturing process. This definition is relevant to the machining domain.

Surface Feature: A surface feature is a collection of workpiece faces that result from machining (i.e., subtracting) a volumetric feature (Vandenbrande, 1990).

Precision Feature: This may refer to reference or datum surfaces from which dimensions or tolerances are specified, or to the actual dimensions or tolerances themselves.

Many different ways of using the concept of features exist in engineering design and manufacture. Although a number of attempts have been made to create feature taxonomies, e.g., CAM-I (1986), no standard has yet been adopted by the research community. This is problematic because the lack of standardization works against integration. For example, having a standard set of design and manufacturing features would allow researchers to develop generic methodologies for mapping between the two domains. This would help to integrate CAPP with feature-based CAD.

For machining process planning, machining features are of primary interest. Figure 2.5 illustrates how they are related to the broader view of features. Machining features are just one of many different types of manufacturing features as can be seen from Figure 2.5(a). Other types of manufacturing features include casting, welding, and sheet metal features. Manufacturing features themselves are a subclass of the basic feature class. Other subclasses at the same level include design features and assembly features.

Two ways of representing a machining feature are illustrated in Figure 2.5(b). The first representation defines the feature by the machined surfaces that are left on the part after the machining process, a slotting operation in this example. The second representation defines the feature by the actual volume that is removed by the machining process, referred to as a machining volume. The two representations are, in fact, interdependent; by removing a machining volume associated with a machining feature, its machined surfaces are generated. The machined surfaces representation is, however, more general because as indicated in the figure, more than one machinable volume may generate the same machined surfaces (e.g., S1 or S1').

2.6.2 Feature Recognition

The area of feature extraction has received much attention over the past 2 decades. We discuss in the following sections relevant developments that have taken place in the field, including a chronology of feature extraction work since 1980 when research in this field was first published. This chronology classifies the feature extraction methodologies into one of several categories. The more important contributions are discussed.

The purpose of feature recognition in the context of machining process planning is to identify machining features in a CAD model. Research work in feature recognition can be classified into the following areas:

- Volume decomposition
- Alternating sums of volume
- Graph-based recognition
- Syntactic pattern recognition
- Knowledge-based feature recognition
- User-interactive recognition
- Recognition from CSG representations
- Recognition from 2D drawings
- Hybrid feature recognition
- Recognition of alternate feature sets

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

TABLE 2.1 Perspectives on CAPP System Characteristics

Perspective	Comments	CAPP System Characteristics									
		User Friendly	Customizable	Robust	Extendable	Complete	Adaptable	Integratable	Teachable	Modular	Efficient
1. An ability to generate, compare, and record multiple process plans to a given part input.	The user should be able to generate multiple feasible mappings of the part to manufacturing operation sets. This is facilitated by the paradigms for interpreting the part and applying machining practices.					•					
2. An ability to learn in a quick and efficient way that is controlled by the end-user.	Two areas where learning capabilities can be utilized are in the part interpretation stage (matching volume extraction) and in the application of manufacturing practice rules.		•		•				•		
3. The system should evolve during use to provide planning that is adapted to the application.	Due to this feature, the unique quality of a CAPP system becomes the information it has acquired during use within a particular environment. This will obviously vary from user to user. The "local knowledge" makes the system more user friendly after it has fully evolved.		•				•				
4. CAPP systems should demonstrate definite time savings and provide consistently equivalent or better plans than those generated by human planners.	This implies that the system should be easy to use and can perform computationally in a manner that is acceptable to the planner. It is worth noting that most systems in use today demonstrate savings of less than 15% over manually prepared plans.			•							•
5. The CAPP system should assimilate information from various stages of the product life cycle, most importantly from the shop floor.	Process plans must often be modified by shop-floor personnel during a test period when the part is brought into production. The reasoning used to make these changes is often lost. Integrating this knowledge into the accumulated knowledge within the process planning tools can lead to future plans utilizing this knowledge at the planning stage.				•			•	•		

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

6. The philosophy of a CAPP system as a black-box is unacceptable to most end-users.	To most process planners the inability to understand how a solution is generated and to control and influence the generation, leads to skepticism. The more the system is understood and tailored by those who use it, the more accepted it will be.	•	•							
7. The CAPP system should be independent of any specific design or manufacturing system.	The purpose of this is to make the system usable by the largest range of end-users who as a group may have a wide variety of CAD/CAM systems which must be integrated with process planning.							•		
8. CAPP systems should provide tools which aid synthesis and analysis in addition to tools which seek to automate and simulate.	While automation may promote planning efficiency, planning diversity comes from allowing the end-user to investigate a wide range of feasible planning solutions. Efficient synthesis and analysis tools give impetus to the planner to explore new approaches to machining.	•				•				•
9. CAPP systems should be more holistic in their approach to planning.	CAPP system research and commercialization have focused primarily on machining processes even though few mechanical parts are produced solely by machining. A holistic system that can combine many processes within one planning environment generates more complete solutions.				•	•				
10. CAPP systems should support planing on different levels.	There are many activities for which an initial, nondetailed (high-level) process plan might be useful: bidding for jobs, and for equipment procurement and facility planning.					•				
11. The CAPP system should be cost effective to purchase, operate, and maintain.	Because much manufacturing work is out-sourced today, CAPP systems must be affordable to smaller manufacturers.		•						•	

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

TABLE 2.2 CAPP System Characteristics and Their Effects

Characteristic	Effects
Complete	• Provides a complete manufacturing solution for the part in question. • Meets all the end-user's requirements. • Facilitates the generation of multiple solutions.
Extendable	• New technologies can be merged into the system. • The system can be extended by the end-user or a third-party software developer.
Adaptable	• The system can be used by many different types of end-users.
User Inclusive	• Utilizes human expertise and computer efficiency in correct proportions. • Promotes synthesis and analysis in addition to automation and simulation.
User Friendly	• Easy to implement and maintain. • Easy to use.
Teachable	• Allows the expertise of the end-user to be incorporated into the system. • The system can act as an archiving tool for the end-user's expertise. • The system can be used to train new process planners.
Customizable	• The system (and its cost) can be tailored to the end-user's requirements.
Modular	• Facilitates extendability, adaptability, customizability, and cost effectiveness.
Robust	• Provides consistently "correct" (by the end-user's standard) solutions. • Reduces human error.
Efficient	• Solutions are generated in a more timely fashion than by conventional planning. • The work load for a process planner generating a solution is reduced.
Integratable	• Implementation is not computer hardware or software specific.
Cost Effective	• The system in a customized form suits the budget of a wide range of end-users.

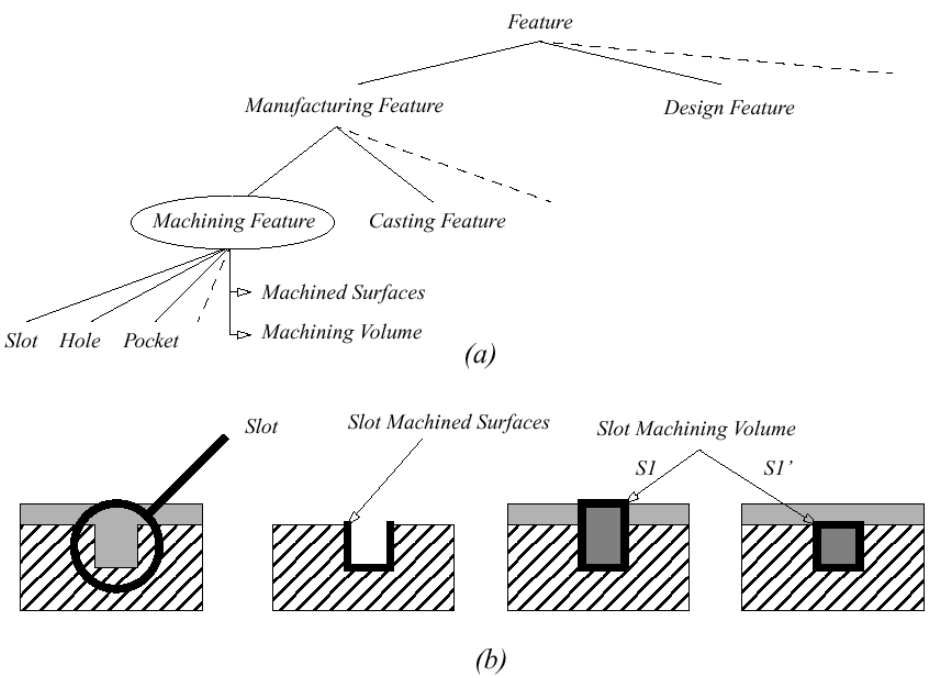


FIGURE 2.5 Machining features.

Figure 2.6 presents a chronology of feature recognition work during the past 2 decades. The figure shows the year in which the research was published as well as the category (from above) into which the work falls. It can be seen from the figure that graph-based recognition, syntactic pattern recognition and knowledge-based approaches have received the widest attention. In the following sections, some of the categories mentioned above are discussed.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

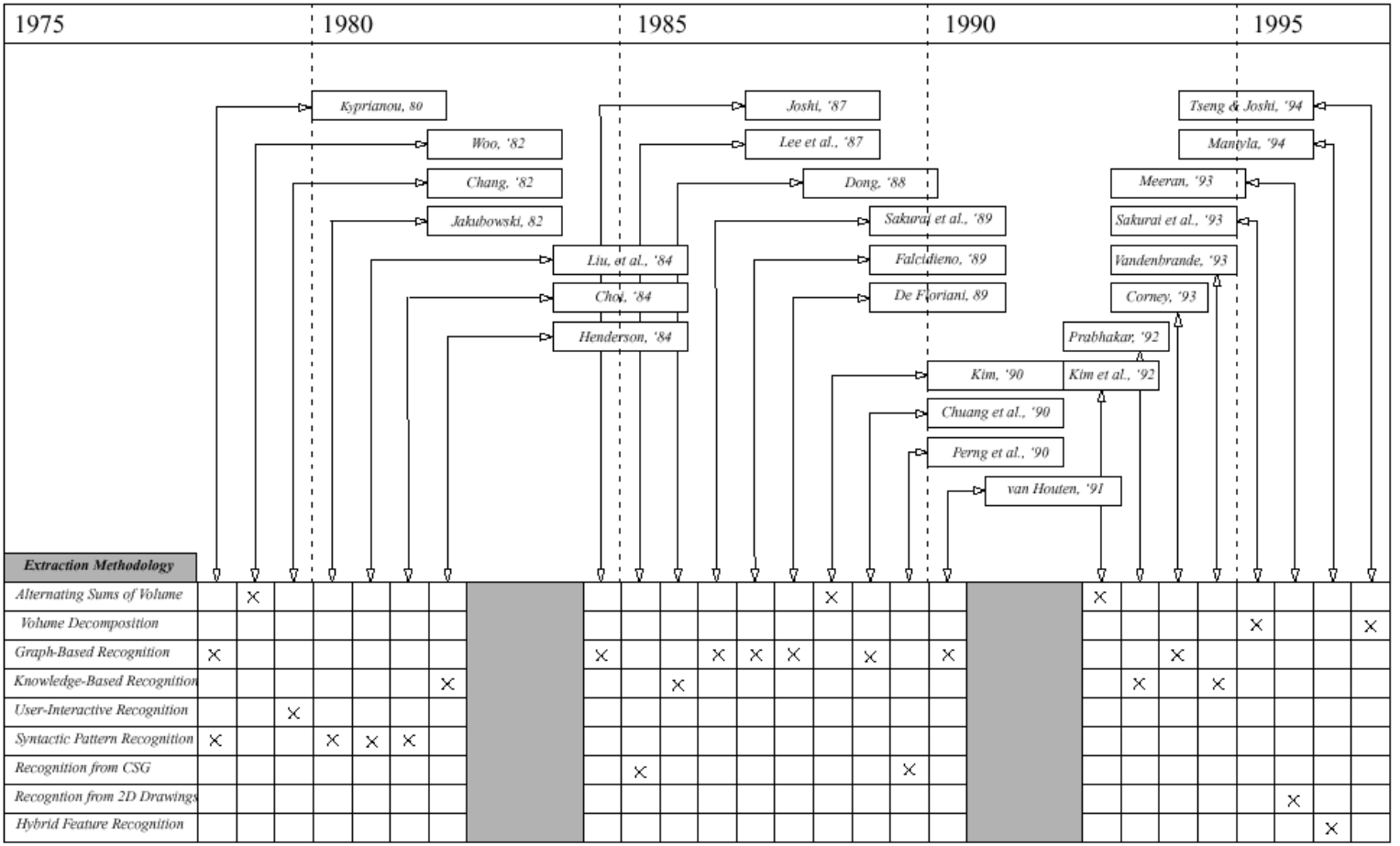


FIGURE 2.6 Chronology of feature recognition work.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

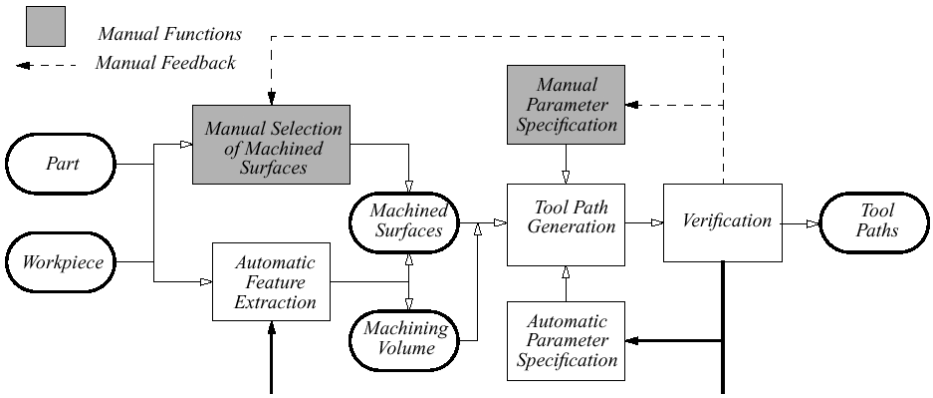


FIGURE 2.7 Volume decomposition approach to feature recognition.

2.6.2.1 Volume Decomposition

Volume decomposition approaches seek to break up the Δ*Volume** into machining volumes. This process is illustrated in Figure 2.7. One of the most well-known approaches using volume decomposition is that adopted by Sakurai and Chin (1994). In their approach, the Δ*Volume* is decomposed by extending planar and curved faces of the part into minimal cells. These cells are recombined to form maximal volumes. By subtracting these volumes in different orders, alternate volume decompositions can be generated. Tseng and Joshi (1994) have also adopted a similar decomposition process.

One advantage of the approach adopted by Sakurai and Chin (1994) is that the definition and use of maximal volumes permit their algorithm to generate all feature interpretations. They argue that this provides an opportunity "to find the optimal or near optimal feature interpretations" (Sakurai and Chin, 1994). This follows because their method is purely algorithmic as opposed to the more common heuristic approaches.

One concern about their approach is that it is driven purely by the geometry of the part. Although this enables them to create a decomposition without having to specify a feature type and domain, a priori, it raises the question as to whether or not all maximal volumes generated can be mapped to a feature within a given domain, in particular the machining domain. A second concern is whether their approach can be extended to surfaces which generate closed halfspaces, quadrics, for example.

2.6.2.2 Alternating Sums of Volume

A similar approach to volume decomposition, first proposed by Woo (1982) and known as the alternating sum of volumes (ASV), recursively subtracts the part from its convex hull until the null set is reached. Woo represented the resulting decomposition as a series of convex volumes with alternating signs. This approach was not always successful for two reasons: (1) When the convex hull at successive iterations was the same, the algorithm cycled, and (2) the algorithm did not always generate a usable decomposition from a machining perspective. A third shortcoming is similar to that of Sakuari's approach to volume decomposition: ASV is driven purely by part geometry. This is even more critical in the ASV approach, because the algorithm generates nonintersecting convex volumes, i.e., precedences are generated using only the part geometry. For machining volume decompositions, this is unacceptable since machining practices need to be considered in determining precedences. Finally, parts with curved surfaces must first be mapped to a polyhedral representation for the convex hull operator. Kim and Wilde (1992), Waco and Kim (1993), and Kim (1994) have extended the ASV approach by introducing modifications that eliminate cycling and generate machinable convex volumes.

*The Δ *Volume* (delta volume) is a term commonly used in feature recognition to refer to the stock that must be removed from a workpiece to generate the final machined part.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

2.6.2.3 Graph-Based Recognition

Graph-based feature recognition has received much attention. The basic philosophy of the approach is to represent the part model and features as graphs, and to perform feature recognition by finding subgraphs of the part graph that match feature graphs.

The first step in this approach is to represent the part as a graph. Because a boundary representation (B-rep) solid model is itself a graph, some researchers have worked directly from this representation (Sakurai and Gossard, 1990). Others have mapped the part from its original representation to other graph representations. These include Attributed Adjacency Graphs (AAG) (Joshi, 1987), Face Edge Graphs (FEG) (DeFloriani, 1989), Face Adjacency Hypergraphs (FAH) (Falcidieno and Giannini, 1989), Vertex Edge graphs (V-E) (Chuang and Henderson, 1990), and Aspect Face Edge Graphs (AFEG) (Corney, 1993). Often the arcs in these graphs are supplemented with additional geometric information. For example, Joshi (1987) tags the arcs in the AAG with 0 if the edge is concave and 1 if the edge is convex.

A decomposition step is normally performed to break up the part graph into a number of smaller subgraphs. These subgraphs may be equivalent to protrusions or depressions in the part. This decomposition step is geared to creating a more computationally manageable problem. Graph matching subsequently is performed on these subgraphs using the feature graphs as templates.

The two main problems with graph-based techniques are (1) the computational complexity of the problem as the size of the part and the number of features and their complexity increase, and (2) the problem of feature interactions which can create phantom features and mask the presence of true features.

2.6.2.4 Syntactic Pattern Recognition

Syntactic pattern recognition is closely related to graph-based techniques. In syntactic pattern recognition, a language is developed with which to represent the part model. The resulting representation is then "parsed" using a feature grammar. Features are recognized by finding combinations of literals of the language within a part representation that conform to the rules of the grammar.

One of the first applications of this approach to feature recognition was by Kyprianou (1980), who used a faceset data structure to represent the part. His algorithm first mapped the B-rep of a part to a series of facesets for depressions and protrusions of the part. These facesets are then analyzed using a feature grammar to generate a part code for the one in question. Kyprianou's work in syntactic pattern recognition is acknowledged by many as ground breaking in the field of feature recognition. It can be argued that syntactic pattern recognition is a formalization of many of the other recognition methodologies. Henderson (1984) uses such an argument in his work. Other researchers who have used this approach include Choi (1982), Jakubowski (1982), and Liu and Srinivasan (1984).

The main limitation to syntactic pattern recognition is the difficulty in developing 3D feature grammars that are general and robust enough to model features of the complexity and diversity found in design and manufacturing. Concern about the computational complexity of shape grammars also exists. Finally, customization of the recognition process requires feature grammars that must be adaptable to different applications.

2.6.2.5 Knowledge-Based Feature Recognition

One of the earliest applications of knowledge-based expert systems to the problem of feature recognition can be attributed to Henderson (1984). His approach uses feature production rules created in the logic programming language Prolog, to interrogate the part. The part itself is first converted from a B-rep into a series of Prolog facts which convey geometric and topological information about the part. The successful execution of a feature rule returns information about the feature from the part facts. This information is used to construct a feature volume which is subtracted from the Δ *Volume*. The recognition process continues until the Δ *Volume* is the null set.

Vandenbrande (1990) attempted to overcome some of the shortcomings of previous work using AI techniques. One of the primary problems he addressed was that of interacting features. He

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

critiqued Joshi's work (Joshi, 1987) as being based on "rigid feature definitions that rely mainly on face adjacency information." When feature interactions occur, the rigidity of these definitions limits the feature-matching algorithm. This observation is the main motivation behind his work: To implement a strategy that is based on feature hints not rigid feature definitions, which can handle features incompletely represented in the object's B-rep due to feature interactions. To do this he proposed a "hint generation and testing" methodology which he implemented using an expert system shell that integrated object-oriented and rule-based programming.

Vandenbrande's approach and others based on feature hints and knowledge about feature recognition are promising. They recognize that feature recognition is a complex problem in human reasoning and approaches that handle inexact and uncertain data have a greater chance of being successful. For example, because these approaches are driven only by hints of features and not a complete representation of the feature embedded within the model, they are more robust in handling feature interactions.

2.6.2.6 User-Interactive Approaches

User-interactive recognition approaches rely on the user to select constitutive geometric elements of a feature (edges or faces) through a graphical interface. These can be viewed as hints similar to those identified automatically in other approaches (Vandenbrande, 1990). The user may be required to either select all trace elements of a feature in the model, or select a minimal set from which the other elements may be identified. Thus, a user-interactive approach need not necessarily be brute force. Rather, by minimizing the level of work that the user must do in selecting feature hints, the system can be designed to behave intelligently. User-interactive methodologies also may be coupled with automatic recognition to extend the domain of the latter.

Although some researchers have implemented user-interactive recognition, it has almost exclusively been done within the context of CAPP system development (Chang, 1982; Brooks et al., 1987; Giusti et al., 1989). The focus of these approaches has been to provide an integrated environment more than to develop an intelligent, user-interactive methodology.

2.6.3 Discussion

As is clear from the chronology in Figure 2.6, feature recognition is a problem that has been addressed by many researchers over the past 2 decades. The focus of this work has been primarily on rotational and 2.5D (prismatic) geometries. While many researchers have solved subsets of these domains, no one work provides a provably complete methodology for automatic feature extraction in either domain.

At the same time, while limitations to current solutions exist, this research highlights the inherent complexity of the problem when the objective is to develop a practical solution. A major complicating factor lies in the definition of a feature itself. Three approaches to feature definition are possible. The first is to create a standard set of features that can be used by all CAE system developers. While such a standardization is useful, deciding on a feature set broad enough to cover the requirements of all possible contexts just within the machining domain is difficult, if not impossible. The second approach attempts to address this open-endedness by proposing that features be user defined, i.e., the feature recognition methodology utilizes a feature set created by and customized to the needs of each end-user. The difficulty with this approach is that it requires representation methodologies that are generic, modular, and customizable, yet implementable in the sense that they can be integrated with the underlying recognition algorithms. The third approach is a hybrid combination of predefined feature sets and user-defined features. This approach offers the best of both worlds. It recognizes that there is a standard feature set that is applicable to many machining contexts while providing a mechanism for extending the set when the situation requires it.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

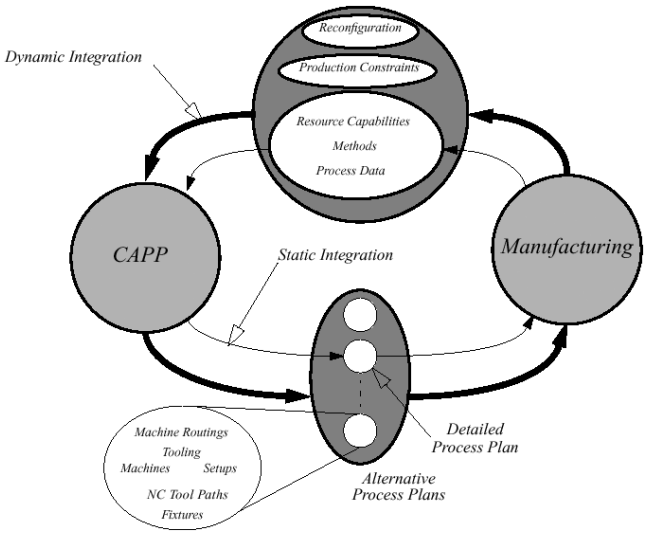


FIGURE 2.8 Integration of CAPP with manufacturing.

Recent commercialization efforts in CAPP have resulted in a number of viable feature extractors. Once such extractor comes with the PART™ CAPP system.* This system evolved from the work of Houten at the University of Twente, The Netherlands, in the mid to late 1980s (Houten, 1988; 1989a; 1989b; 1990). In PART, feature extraction is based on the hybrid notion of features. The current version of the system comes with a standard library of around 60 features to which others can be added by the user. PART also has an editing facility that allows the user to modify automatically generated results. The inclusion of this capability underscores the assertion made previously that a provably correct, fully automatic feature extraction methodology has yet to be developed.

2.7 Integrating CAPP with Manufacturing

As important as the task of integrating CAPP with automated design systems, is the task of integration with manufacturing. Figure 2.8 illustrates the relationship. CAPP provides the information to drive the manufacturing processes, yet at the same time it relies upon an understanding of the manufacturing facility (resource data, methods, process data, etc.) to constrain the planning task so that the plans created are relevant to the manufacturing context. Traditionally, this interaction has been based on a static view of manufacturing, i.e., CAPP provides a single detailed process plan for a part on a given facility that is static both in configuration and capability. This is indicated in the figure as the inner loop (thinner line). There is now, however, much interest in considering the dynamic nature of this integration by requiring CAPP systems to generate alternative process plans that conform to changing production constraints (product mix, annual volume, machine utilization) and reconfiguration of the machining facility. This is represented by the outer loop (heavier line) in the figure. Examples of research work in this area include ElMaraghy and ElMaraghy (1993), Chryssolouris et al. (1984), Lenderink and Kals (1993), and Zhang (1993).

Two components of integration that warrant special attention are NC tool path generation and machining methods. These are discussed briefly in the next sections.

*PART™ is a commercial CAPP system originally developed at the University of Twente, The Netherlands. The system was commercialized by CDC as part of the ICEM system, but has since been acquired by Technomatix Inc.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

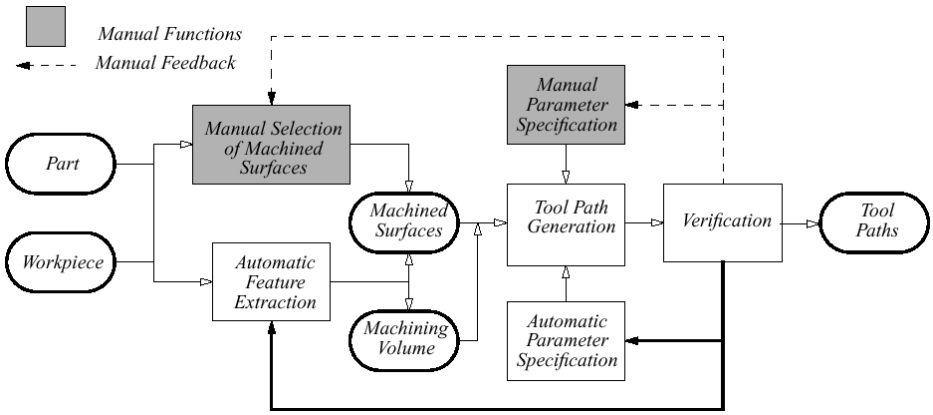


FIGURE 2.9 Role of machining feature recognition in NC tool-path generation.

2.7.1 NC Tool-Path Generation

Much work has been done on the problem of NC tool-path generation. Many commercial NC tool-path generation systems, stand alone (SmartCAM™, Gibbs™) and integrated within CADCAM systems (Bravo™, Unigraphics™, Pro-Manufacture™) are now available. While researchers are still investigating new techniques for improving tool path generation (Sarma, 1996), another major challenge that needs to be addressed is improving the ease with which these tools are used. NC part programming remains a time-intensive task.

Machining feature recognition has the potential of doing just this by driving the automation of part programming. This is illustrated in Figure 2.9. The figure also shows manual functions which give the user the ability to override any decisions made by the system as well as automatic feedback links from verification. The tasks for which automation can be helpful are

- Selection of machining surfaces
- Specification of generation parameters
- Feedback of changes from verification for the reselection of machining surfaces
- Feedback of changes based on verification for the respecification of generation parameters

As can be seen from the figure, feature extraction has the potential to eliminate the timely and error-prone task of machining surface selection because, by definition, the surfaces of the machining feature are identified. In addition, extraction procedures can be further automated to create the machining volume associated with the surface feature. This volume is useful in automating the identification and avoidance of interference geometry. Another useful output generated from feature extraction is the precedences between features. These precedences can be used to automatically merge the tool paths created for each feature into a single "tape" for machining the part.

2.7.2 Manufacturing Data and Knowledge

Machining methods (also referred to as machining practices) provide CAPP with the knowledge, expertise, and procedures that a human process planner uses. These methods may be based on sound scientific principles, experimental results, experience, or preferences established within a particular machining context. They also may be generic and applicable over a wide range of machining problems or specific to a single one.

The challenges in using machining methods within CAPP fall into the following categories:

- Identification and retrieval
- Implementation

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

- Maintenance
- Customization

Identification and retrieval are concerned with understanding how a human process planner applies experience and techniques to make decisions when generating process plans: What decisions are being made? What characteristics of the situation are being recognized by the planner that trigger these decisions? The main challenge here stems from the fact that human planners do not necessarily follow a consistent strategy in applying methods. The process often requires complex trade-offs of information from several sources. When one of these sources is experience, the basis of the applied method can be difficult to verify. Thus, identification and retrieval of methods are not just a bookkeeping task. Rather, it requires the cultivation of an attitude toward process planning based on a sound methodology for applying machining methods.

Methods implementation requires an approach that is general enough to capture information from very different sources while at the same time is simple enough to provide a maintainable, noncorruptible environment. Rule-based expert systems have been the most commonly adopted implementation strategy among CAPP system developers.

Because the need to update or add new methods always exists as more information becomes available or as new methods are applied to more applications, maintenance of the knowledge base becomes a key concern. As changes are made, the integrity of the information needs to be preserved. One problem occurs when new methods are added that conflict with old ones. The system needs to include a strategy for resolving such conflicts. One approach that has been used extensively with expert systems is to place the onus on a knowledgable engineer to avert such problems. However, as the size of the knowledge base grows, the cost of employing dedicated personnel for this task becomes prohibitive.

Finally, creating off-the-shelf CAPP systems with the methods included is a difficult if not impossible task. This is because it is unlikely that the system developer can capture all the desired methods from all potential users during system development. Thus, while a system may come with some generic, widely accepted methods, it must include a facility to allow new methods customized to each context to be added to the system.

2.8 CAPP for New Domains

Even though formidible problems remain in the development of commercially viable CAPP systems, researchers have continued to broaden the applicability of this technology to new domains. An example of such research is in the domain of parallel machining.

2.8.1 Parallel Machining

Parallel machining is the simultaneous removal of material from a workpiece by multiple cutting tools on a single machine tool or machining system. This concept has been in existence for some time. Examples of parallel machining are found on transfer line machines in automobile production for machining powertrain components, multi-spindle plano-milling machines for the simultaneous machining of casting surfaces and multi-turret (4-axis)* lathes. In these instances, parallel machining is preferred to sequential machining because higher production rates can be realized due to the reduction of cutting times.

The application of parallel machining in these examples suffers from one major drawback: a limitation in the range of parts which can be machined due to the dedication of the machining resources to specific tasks. Transfer lines are a prime example of this. Transfer line machines are constructed with the aim of mass producing components from a single engine model. The machining

*The term *4-axis* lathe is commonly used in industry to refer to a lathe with two turrets. Each turret is positioned by movements along an independent pair of orthogonal axes (x-axis and z-axis).

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

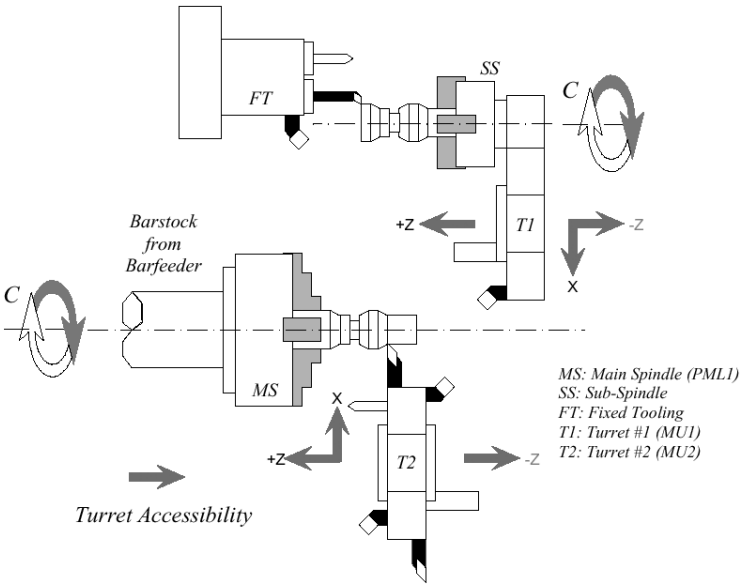


FIGURE 2.10 Examples of parallel machines.

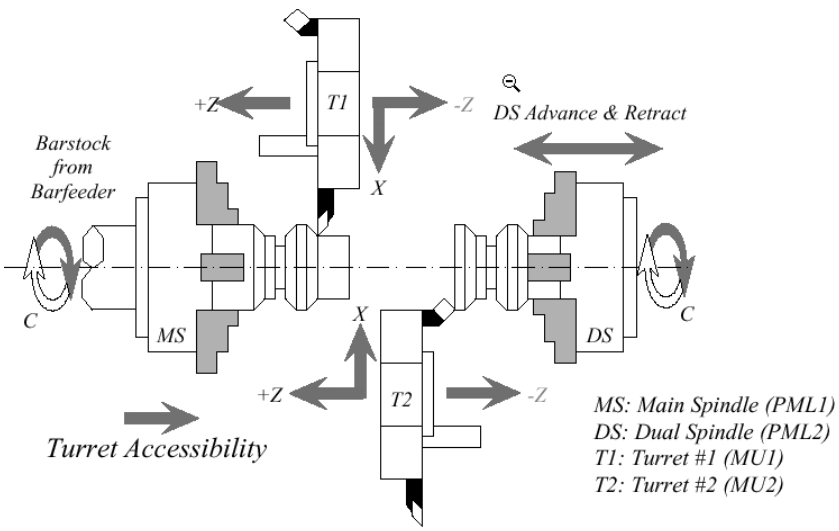


FIGURE 2.11 Example of a dual-spindle mill/turn.

elements are dedicated to this task. Once this model is taken out of production, the machines must be stripped down and retooled. This contributes to high setup costs when switching lines from an old engine model to a new one.

One class of machine tools that combines the advantages of parallel machining with the flexibility of nondedicated tooling afforded by computer numerical control (CNC) is the mill/turn (also referred to as a turning center) (Figures 2.10 and 2.11). Parallel machining on mill/turns takes two forms:

- Multiple machining operations performed simultaneously on a single part
- Multiple operations performed simultaneously on multiple parts

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

A secondary feature of mill/turns which adds to their flexibility is, as their name implies, the ability to perform both turning and milling operations in the same setup. This contrasts greatly with conventional machining practice which dictates that turning and milling operations be performed in separate setups on different machines. The resulting elimination of setups on mill/turns has obvious advantages in reducing the machining time per part and in increasing part accuracy by reducing work handling.

For the capabilities of mill/turns to be fully exploited, a CAPP system for the mill/turn domain must be developed. This presents problems of a different nature than those encountered for conventional CAPP systems. In particular, the presence of multiple tool- and work-holding devices raises the question of the efficient utilization of the machine tool. Considerations of the effect of parallel machining on tool wear and part quality also must be addressed. A greater need for collision checking and avoidance planning due to the simultaneous motions of multiple turrets is necessary. Currently, the complexity of process planning for this domain results in conservative process plans which underutilize the machine tool's resources.

2.8.1.1 CAPP for Parallel Machining

While a great body of work exists in the area of CAPP for the sequential machining domain, research about CAPP for the parallel machining domain is relatively new. One example of prior work in this domain is by Levin and Dutta (1992). In their work, they outline their experiences in implementing their version of a CAPP system for parallel machining (PMPS). Within PMPS, a Giffler-Thompson algorithm which generates active–delay type schedules was used to sequence machining operations. An active schedule is one in which no operation can be started any earlier without either delaying some other operation or violating a technological constraint. A delay type schedule allows a resource such as a machine tool turret to be idle instead of performing an operation. The author surmises that these two characteristics are highly applicable for process planning in this domain.

While the Giffler-Thompson algorithm is intuitively easy to understand and equally easy to implement, it is difficult to determine how good the final schedule is. In fact, because it uses a one-step look-ahead strategy, the plans are likely to be myopic in nature. Nevertheless, this work does discuss in detail the nuances of process planning for parallel machining and provides a good foundation for this research.

New approaches to scheduling for mill/turns using Genetic Algorithms have been developed by Yip-Hoi (1997). This dissertation work also makes contributions to defining an architectural framework for a CAPP system for parallel machining as well as developing numerous geometric modeling and feature extraction tools to assist the process planner in generating process plans for this domain.

2.9 Conclusions

This chapter presents a overview of research work in the area of computer-aided process planning. This field has generated much attention over the past 20 years as researchers have tried to bridge the gap between automated design and manufacturing. We have presented some of the key enablers and characteristics of CAPP systems. We also have discussed research in feature recognition, which is one of the key underlying technologies of CAPP.

Despite the efforts outlined, and extensions to new CAPP domains, fewer commercially viable CAPP systems are available than CAD or CAM systems. This is in large part due to the complexity of interpreting CAD models of complex engineered products and the difficulties in identifying and capturing machining practices that are customized to the end-user's requirements. Current trends such as the increasing use of features in CAD/CAM systems and the explosion in information engineering techniques prompted by internet development are likely to spur on a second generation of CAPP systems that will attempt to address current deficiencies.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

References

Alting, L. and Zhang, H., Computer-aided process planning: The state-of-the-art survey, *International Journal of Production Research*, 27, 4, 553–585, 1989.

Berenji, H.R. and Khoshnevis, B., Use of artifical intelligence in automated process planning, *Computers in Mechanical Engineering*, 5, 2, 47–55, September 1986.

Brooks, S.L., Hummel, K.E., and Wolf, M.L., XCUT: A rule-based expert system for the automated process planning of machined parts, PED Vol. 25, *Intelligent and Integrated Manufacturing Analysis and Synthesis*, presented at the ASME Winter Annual Meeting, Boston, Massachusetts, December 1987, 181–194.

CAM-I, Illustrated Glossary of Workpiece Form Features, Revised, Report #R-80-PPP- 02.1, CAM-I Inc., Arlington, Texas, 1981.

CAM-I, Part Features for Process Planning, Report #R-86-PPP-01, CAM-I Inc., Arlington, Texas, 1986.

CAM-I, The State of the Art in Computer-Aided Process Planning, Report #DR-89-PP-01, CAM-I Inc., Arlington Texas, 1989.

CAM-I, A Conceptual Model for Process Planning, Report #R-90-PP-01, CAM-I Inc., Arlington, Texas, 1990.

Chang, T.C., TIPPS – A Totally Integrated Process Planning System, Ph.D. dissertation, Blacksburg Virginia Polytechnic Institute, 1982.

Chang, T.C. and Wysk, R.A., *An Introduction to Automated Process Planning Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

Chang, T.C., Anderson, D.C., and Mitchell, O.R., QTC – An Integrated Design/Manufacturing/Inspection System for Prismatic Parts, *Proceedings of the ASME Computers in Engineering Conference*, San Francisco, July, 1988, 417–425.

Chang, T.C., *Expert Process Planning for Manufacture*, 1st ed. Addison-Wesley, New York, 1990.

Choi, B.K., CAD/CAM Compatible, Tool-Oriented Process Planning for Machining Center,. Ph.D. dissertation, Purdue University, Lafayette, 1982.

Chryssolouris, G., Chan, S., and Cobb, W., Decision making on the factory floor: An integrated approach to process planning and scheduling, *Robotics and Computer-Integrated Manufacturing*, 1, 3/4, 315–319, 1984.

Chuang, S.H. and Henderson, M.R., Three-dimensional pattern recognition using vertex classification and vertex-edge graphs, *Computer-Aided Design*, 22, 6, 377–387, 1990.

Corney, R.C., Graph-Based Feature Recognition, Ph.D. dissertation, Heriot-Watt University, Edinburg, 1993.

Davies, B.J., Darbyshire, I.L., Wright, A.J., and Zhang, K.F., IKBS process planning system for rotational parts, *Intelligent Manufacturing Systems II*, edited by V.R. Milacic, Elsevier Science Publishers B.V., Amsterdam, 1988, 27–39.

De Floriani, L., Feature extraction from boundary models of three-dimensional objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 8, 785–798, 1989.

Descotte, Y. and Latombe, J.C., GARI: A problem solver that plans how to machine mechanical parts, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Vancouver, August 1981, 766–772.

Dong, X., Geometric Feature Extraction for Computer-Aided Process Planning, Ph.D. dissertation, Rensselaer Polytechnic Institute, Albany, 1988.

ElMaraghy, H.A., Evolution and future perspectives of CAPP, *Annals of CIRP*, 42, 2, 1–13, 1993.

ElMaraghy, H.A. and ElMaraghy, W.H., Bridging the gap between process planning and production planning and control, *Manufacturing Systems*, 22, 1, 5–11, 1993.

Erve, A.H. van 't, Kals, H.J.J., XPLANE, a generative computer sided process planning system, *Annals of CIRP*, 35, 1, 325–329, 1986.

Eversheim, W. and Holtz, B., Computer aided programming of NC-machine tools by using the system AUTAP-NC, *Annals of CIRP*, 31, 1, 1982.

Eversheim, W. and Schneewind, J., Computer-aided process planning – State of the art and future development, *Robotics and Computer-Integrated Manufacturing*, 10, 1/2, 65–70, 1993.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Falcidieno, B. and Giannini, F., Automatic recognition and representation of shape-based features in a geometric modeling syste, *Computer Vision, Graphics, and Image Processing*, 48, 93–123, 1989.
Giarratano, J. and Riley, G., Expert Systems Principles and Programming, PWS-KENT Publishing Company, Boston, 1989.
Giarratano, J., CLIPS User's Guide, Lyndon B. Johnson Space Center Information Systems Directorate Software Technology Branch, May 1993.
Giusti, F., Santochi, M., and Dini, G., KAPLAN: A knowledge-based approach to process planning of rotational parts, *Annals of CIRP*, 38, 1, 481–484.
Henderson, M.R., Extraction of Feature Information from Three-Dimensional CAD Date, Ph.D. dissertation, Purdue University, 1984.
Henderson, M.R. and Chang, G.J., FRAPP: Automated feature recognition and process planning from solid model data, *Proceedings of the ASME International Computers in Engineering Conference and Exhibition*, July 1988, 529–536.
Houten, F.J. A.M. van., Strategy in generative planning of turning processes, *Annals of CIRP*, 35, 1, 331–336, 1986.
Houten, F.J.A.M. van and Erve, van 't A.H., PART, a parallel approach to computer aided process planning. Presented at CAPE-4, Edinburg, November 1988. Web URL: http://utwpu9.wb.utwente.nl/projects/part/part-doc/cape_edinburgh_1988.ps.
Houten, F.J.A.M. van and Erve, van 't A.H., PART, a feature based CAPP System. Presented at the 21st CIRP International Seminar on Manufacturing Systems, Stockholm, 1989. Web URL: http://utwpu9.wb.utwente.nl/projects/part/part-doc/cirp_stockholm_1989.ps.
Houten, F.J.A.M. van and Erve, van 't A.H., PART, a CAPP system with a flexible architecture. Presented at the 2nd CIRP International Workshop on CAPP, Hannover, September 1989. Web URL: http://utwpu9.wb.utwente.nl/projects/part/part-doc/hannover_sept_1989.ps.
Houten, F.J.A.M. van and Erve, van 't A.H., Boogert, R.M., Nauta, J.M., Kals, H.J.J., PART, selection of machining methods and tools. Presented at the 22nd CIRP International Seminar on Manufacturing Systems, Enschede, June 1990. Web URL:http://utwpu9.wb.utwente.nl/ projects/part/part-doc/cirp_twente_1990.ps.
Hummel, K.E. and Brooks, S.L., Symbolic representation of for an automated process planning system, *Proceedings of the ASME Winter Annual Meeting*, Anaheim, California, 233–243, 1986.
Hummel, K.E. and Brooks, S.L., Using hierarchically structured problem-solving knowledge in a rule-based process planning system, *Expert Systems and Intelligent Manufacturing*, edited by M. D. Oliff, Elsevier, Amsterdam, 120–137, 1988.
Iwata, K. and Fukuda, Y., Representation of know-how and its application of machining reference surface in computer aided process planning, *Annals of CIRP*, 35, 1, 321–324, 1986.
Jakubowski, R., Syntactic characterization of machine parts shapes, *Cybernetics and Systems: An International Journal,* 13, 1–24, 1982.
Joshi, S.B., CAD Interface for Automated Process Planning, Ph.D. dissertation, Purdue University, Lafayette, Indiana, 1987.
Joshi, S.B., Hoberecht, W.C., Lee, J., Wysk, R.A., and Barrick, D.C., Design, development and implementation of an integrated group technology and computer aided process planning system, *IIE Transactions*, 26, 4, 2–18, July 1994.
Kim, Y.S., *Volumetric Feature Recognition Using Convex Decomposition. Advances in Feature Based Manufacturing*, 1st ed. Edited by J.J. Shah, M. Mantyla and D.S. Nau, Elsevier, Amsterdam, 1994, 39–63.
Kim, Y.S. and Wilde, D.J., Local cause of non-convergence in a convex decomposition using convex hulls, *ASME Journal of Mechanical Design,* 114, 468–476, September 1992.
Kiritsis, D., A review of knowledge-based expert systems for process planning. methods and problems, *International Journal of Advanced Manufacturing Technology*, 10, 240–262, 1995.
Kyprianou, L.K., Shape Classification in Computer Aided Design, Ph.D. thesis, Christ's College, University of Cambridge, Cambridge, U.K., July 1980.
Laako, T. and Mantyla, M., *Incremental feature modeling, Advances in Feature Based Manufacturing*, 1st ed., Edited by J.J. Shah, M. Mantyla and D.S. Nau, Elsevier, Amsterdam, 1994, 455–480.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

Lee I.B.H. and L.B.S., Nee, A.Y.C., IKOOPPS: An intelligent knowledged-based object-oriented process planning system for the manufacture of progressive dies, *Expert Systems*, 8, 1, 19–33, 1991.
Lenderink, A. and Kals, H.J.J., The integration of process planning and machine loading in small batch part manufacturing, *Robotics and Computer-Integrated Manufacturing*, 10, 1/2, 89–98, 1993.
Levin, J. and Dutta, D., Computer-aided process planning for parallel machines, *Journal of Manufacturing Systems*, 11, 2, 79–92, 1992.
Link, C.H., CAPP, CAM-I Automated Process Planning System, *Proceedings of the 1976 NC Conference*, CAM-I Inc., Arlington, Texas.
Liu, R.C. and Srinivasan, R., Generative process planning using syntactic pattern recognition. *Computers in Mechanical Engineering*, March 1984, 63–66.
Meeran, S. and Pratt, M.J., Automated feature recognition from 2D drawings, *Computer Aided Design*, 25 1, 7–17, 1993.
Milacic, V.R., SAPT–Expert system for manufacturing process planning, *Proceedings of the ASME Winter Annual Meeting*, Miami Beach, Florida, November 1985, 43–53.
Milacic, V.R., SAPT: Expert system based on hybrid concept of group technology, *Intelligent Manufacturing Systems II*, edited by Milacic, V.R., Elsevier, Amsterdam, 1988, 49–51.
Nau, D.A. and Chang, T.C., A knowledge-based approach to generative process planning, PED No. 19, *Computer Aided/Intelligent Process Planning,* edited by Liu, C.R. and Chang, T.C., ASME Winter Annual Meeting, Miami Beach, Florida, November 1985, 253–263.
Nau, D.A. and Gray, M., SIPS: An application of hierarchical knowledge clustering to process planning, PED Vol. 21, *Integrated and Intelligent Manufacturing,* edited by Liu, C.R. and Chang, T.C., ASME Winter Annual Meeting, Anaheim, California, December 1986, 219–225.
Philips, R.H., Arunthavanathan, V., and Zhou, X.D., Microplan: A microcomputer based expert system for generative process planning, *Proceedings of ASME Winter Annual Meeting,* Anaheim, California, December 1986, 263–273.
Pine, J.B., *Mass Customization The New Frontier in Business Competition*, Harvard Business School Press, Boston, Massachusetts, 1993.
Rose, D.W., Coding for Manufacturing COFORM, unpublished Master's thesis, Purdue University, Lafayette, Indiana, 1977.
Sakurai, H. and Gossard, D.C., Recognizing shape features in solid models, *IEEE Computer Graphics and Applications,* 10, 5, 22–32, 1990.
Sakurai, H. and Chin, C., Definition and recognition of volume features for process planning. *Advances in Feature Based Manufacturing,* 1st ed. Edited by J.J. Shah, M. Mantyla and D.S. Nau, Elsevier, Amsterdam, 1994, 65–79.
Sarma, R., NC Tool Path Synthesis, Ph.D. dissertation, University of Michigan, 1996.
Shah, J., Assessment of features technology, *Computer-Aided Design*, 23, 5, 331–343, 1991.
Tseng, Y. and Joshi, S.B., Recognizing multiple interpretations of interacting machining features. *Computer-Aided Design*, 26, 9, 667–688, 1994.
Tulkoff, J., Lockheed's GENPLAN, *Proceedings of the annual meeting and technical conference of the Numerical Control Society*, May 1981, 417–421.
Turner, G.P. and Anderson, D.C., An object oriented approach to interactive, feature based design for quick turnaround manufacturing, *Proceedings of the ASME Computers in Engineering Conference*, San Francisco, July 1988, 551–555.
Vandebrande, J.H., Automatic Recognition of Machinable Features in Solid Models, Ph.D. dissertation, University of Rochester, 1990.
Waco, D.L. and Kim, Y.S., Considerations in positive to negative conversion for machining features using convex decomposition, *Proceedings of the ASME Computers in Engineering Conference*, Albuquerque, 1993, 35–46.
Wang, H.B. and Wysk, R.A., Intelligent reasoning for process planning, *Computers in Industry*, 8, 293–309, 1987.
Wong, C.L., Bagchi, A., and Ahluwalia, R.A., DMAP: A computer integrated system for design and manufacturing of axisymmetric parts, *Proceedings of the ASME Winter Annual Meeting*, Anaheim, California, December 1986, 327–338.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

Woo, T., Feature Extraction by Volume Decomposition, Technical Report No. 82-4, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, 1982.

Wysk, R.A., An Automated Process Planning and Selection Program: APPAS, Ph.D. dissertation, Purdue University, Lafayette, Indiana, 1977.

Yip-Hoi, D., Methodologies for Computer-Aided Planning for Parallel Machining, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.

Yut, G.A. and Chang, T.C., A five-step object-oriented architecture for process planning, *International Journal of Production Research*, 32, 2, 545–562, 1994.

Zhang, H., *IPPM: A prototype to integrate process planning and job shop scheduling functions, CIRP Annals*, 42, 1, 513–518, 1993.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

3

Discrete Event Control of Manufacturing Systems

3.1 Introduction
3.2 Background on the Logic Control Problems
Logic Control Definition • Control Modes • Logic Control Specification • Tasks of a Logic Control Programmer
3.3 Current Industrial Practice
Programmable Logic Controllers • Relay Ladder Logic • Sequential Function Charts
3.4 Current Trends
Issues with Current Practice • PC-Based Control • Distributed Control • Simulation
3.5 Formal Methods for Logic Control
Important Criteria for Control • Discrete Event Systems • Finite State Machines • Petri Nets
3.6 Further Reading

D. M. Tilbury
University of Michigan

P. P. Khargonekar
University of Florida

3.1 Introduction

A (discrete part) manufacturing process, whether it be machining or assembly, consists of a sequence of steps that must occur to transform the raw materials into finished parts. A manufacturing system is a set of machines (and humans) along with associated control and information systems protocols that implement the manufacturing process. The steps in the process, often called "operations," are assigned to certain machines. The machines are arranged in a line, and as the part moves along the line, the specified operations are performed on it; at the end of the line, it becomes a finished product. The line of machines may be a physical arrangement, or a virtual "line" where the machines are grouped into cells and an operator or computer guides the parts through the appropriate sequence of machines.

Automated manufacturing systems must perform the same sequence of operations repeatedly. There are two distinct types of control systems in a typical automated manufacturing system: continuous control and discrete event control. Continuous control systems regulate continuous variables such as position, velocity, etc.* Discrete event control correctly sequences the system

*In current technology, continuous control is often implemented using digital computers. In this sense, this type of control is discrete-time digital control. This discrete-time control should not be confused with discrete event control.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

operations: do one step after another, perform a specified sequence in the event of a failure, etc. In actual operation, these two types of control systems work concurrently. In this presentation, we will focus on the discrete event control and neglect the interactions between discrete event controllers and the continuous controllers.

In a discrete event framework, the behavior of a manufacturing system is described by a sequence of events, such as the flip of a switch, the push of a button, or the start or end of an operation. These events take the system from one discrete state to another. The state of the manufacturing system is one of a finite set of states, rather than a collection of continuous variables. For example, the discrete event model of a robot gripper may have four states: open, closing, closed, opening; whereas, the continuous model of the gripper would contain position, velocity, and force variables to indicate how wide the gripper is open, how fast it is moving, and the force exerted by the gripper in the closed position.

Because the capital equipment cost for an automated manufacturing system is extremely high, many of these systems typically operate 2 or 3 shifts each day, and 6 or 7 days a week, making reliability extremely important. Thus, in addition to controlling the manufacturing system when it is working well, the discrete event controller must be able to handle various errors. For example, if one machine breaks, the machine before it should stop sending it parts, or if the coolant tank is empty, the spindle should stop drilling. When errors do occur, the discrete event controller should notify an operator by producing some type of error message.

In this chapter, we discuss the problem of discrete event control related to manufacturing systems, how industry currently solves these control problems, current trends in the area, and formal methods that can be used to design and analyze the discrete event control systems used in manufacturing.

3.2 Background on the Logic Control Problems

Discrete event control problems encountered in manufacturing systems consist of the logic and sequence coordination, error recovery, and manual control. These problems are simple in the small view, but extremely complex in the overall picture due to the large number of events that must be coordinated, each with its own input and/or output. For example, a transfer line machining system with ten machining stations can easily contain 10,000 discrete I/O points. Even for such complex manufacturing systems, with thousands of inputs and outputs, the discrete event control is typically written in a low-level programming language. This creates large, unwieldy programs that, although they are intuitive at a very low level, are difficult both to implement and to maintain.

3.2.1 Logic Control Definition

The discrete event control for a manufacturing system controls all of the activity at the machine level as well as the coordination between machines (including material handling). The discrete event controller is also responsible for machine services, such as lubrication and coolant.

Both the discrete event behavior of a manufacturing system and the discrete event controller for the system can be modeled as discrete event systems. Because of the overwhelming complexity of most industrial manufacturing systems, however, the entire possible behavior of the system is rarely described. Typically, only the desired or controlled behavior is specified. In any case, the existing formal methods for analyzing such a combined discrete event system are limited by the computational complexity of dealing with large numbers of states.

A simple block diagram of a manufacturing system with a logic controller is shown in Figure 3.1. The logic controller governs the sequence of the manufacturing process. It controls the system so that the events occur in the specified order in the process, and generate an error event and stops the process in case something goes awry.

Inputs to a discrete event control system consist of proximity and limit switches that indicate the state of the manufacturing system as well as buttons and switches controlled by the operator.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

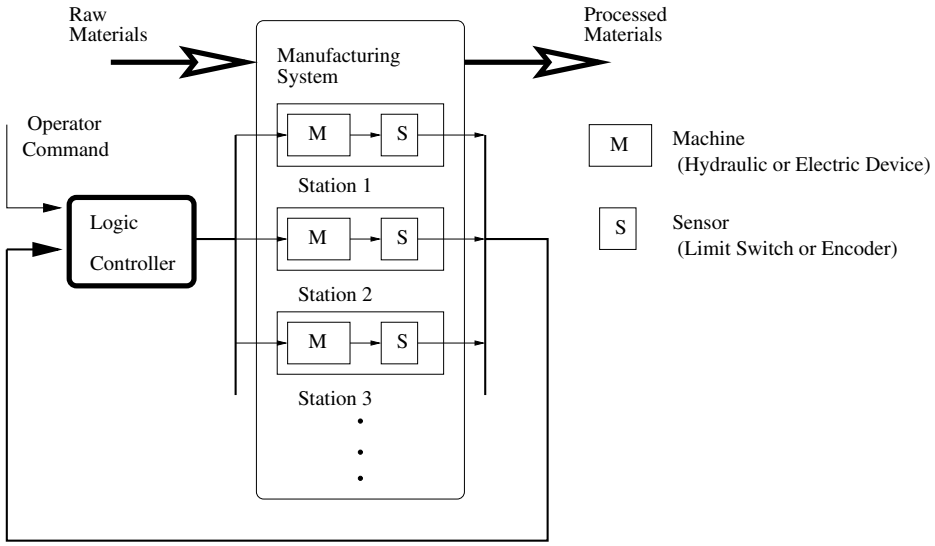


FIGURE 3.1 A block diagram of a manufacturing system with logic control. Raw materials (unfinished parts) enter the system, the machines in the system perform some operations on the parts (such as machining, assembly, etc.), and processed materials (finished or semi-finished parts) leave the system. The logic controller coordinates the operations of the various machines. It is preprogrammed to execute the proper sequence, and also takes some inputs from a human operator. Sensors attached to each machine provide feedback to the logic controller.

The outputs are on/off signals that control valves, motors, and relays as well as lights on the operator interface panel.

3.2.2 Control Modes

The discrete event control for a manufacturing system typically has several different modes. In normal operation, when the system is producing parts, the control operates in the automatic cycle or auto mode. This mode requires little or no operator supervision or intervention, and is the simplest mode to specify and implement. In the event that an error occurs, an operator is usually required to help get the system back to normal operation. The manual modes allow the operator to step through the operation one task at a time or retract slides to allow access to change a tool. Other modes allow the entire operation sequence to be performed only once, or provide diagnostics.

For example, consider a machining system operating in the auto mode. At some point, the tool on the drilling station may break while the system is drilling. The part being worked on will need to be removed, and the machine returned to its default or home position to be ready for the next part. To accomplish this, the operator will first put the machine into manual mode, and will push a sequence of buttons to turn off the power to the spindle, retract the slide, unclamp the part, etc. Then he or she will reach into the machine and physically remove the damaged part and replace the broken tool; hardwired safety interlocks will ensure that the machine cannot operate while the operator is inside the enclosure. Another sequence of buttons will need to be pressed to reset the machine to its home position, and then the operator can switch the machine to the auto mode again. A flow chart depicting this switching of control modes is shown in Figure 3.2.

3.2.3 Logic Control Specification

The sequencing behavior of a manufacturing system can be specified in many different ways. The process plan specifies the operations that must be done to a part to transform it from raw material to a finished product. This plan is generated from the part definition along with the chosen manufacturing

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

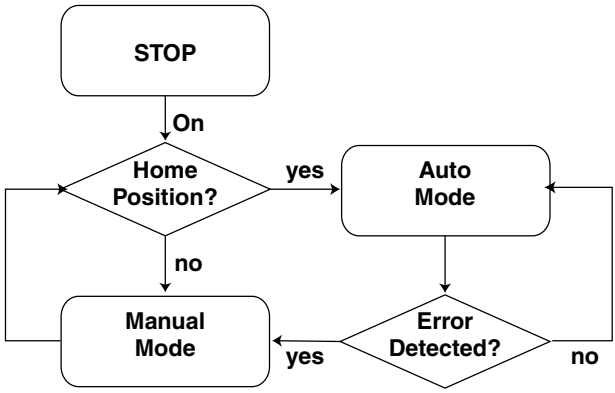


FIGURE 3.2 A flow chart indicating the transitions between auto mode and manual mode. In the manual mode, operator pushbuttons are enabled that can help the operator get the machine back into the home position. The auto mode can only begin when the machine is correctly configured. An error will cause the machine to exit the auto mode and go to the manual mode; an operator is required to fix the machine and help return it to the home position.

processes. If there is only a single sequence in the process, an ordered list of operations will suffice for the logic control specification. Often, however, many tasks must take place simultaneously. The interrelationships and sequential dependencies between these tasks may be specified using a timing bar chart. The tasks to be performed are listed on the vertical axis, and the time taken for each task is represented by a horizontal bar, with the horizontal axis representing time. Dependencies between tasks are indicated by dotted arrows.

A transfer line is a manufacturing system used for high-volume machining operations, for example, automotive engine blocks. Generally, a transfer line is composed of 4 to 12 machining stations; the operation of the system is governed by event sequences within the stations as well as dependencies across the stations. In devising control algorithms for such a machining system, it is necessary to consider not only the sequence of each station but also the correlated sequences of the whole system. An example of a transfer line is shown in Figure 3.3. The system has 15 stations, consisting of 4 mills, 3 clamps, a cradle, and a rotating table. Not all stations are used; the extra space is needed to provide access to the machines for maintenance and repair. The engine blocks move through the machine via a transfer bar from station 1 to station 15. At station 6, they are reoriented.

The timing bar chart shown in Figure 3.4 represents part of the behavior of the high-volume transfer line shown in Figure 3.3. In a transfer line, all of the individual stations must synchronize their operations to the transfer mechanism. Thus, each station has the same amount of time to finish its operation. The total time for operation and transfer is called the cycle time of the transfer line. The causal dependencies of the sequences are represented using the time axis, and the dotted arrows correlate the sequences which depend on each other physically. The timing information of each operation comes from the specifications of the continuous control loops that govern the underlying continuous-time mechanical systems. The timing bar chart shows at a glance the time taken by each task within the cycle time, the time dependencies of tasks, and the total cycle time.

The timing bar chart thus has all the information needed to describe the sequences of tasks that must be performed, and it represents the specification of the operations for the desired process. It is limited by the fact that it only includes the specification for the normal operation of a system, the automatic cycle, or auto mode. The specifications for the other modes of the system (manual, diagnostics, etc.) are rarely described precisely; the control programmer uses experience and intuition to write the logic control for these other modes. Because of this imprecise specification, and the impossibility of foreseeing every possible error that may occur, the logic for the manual modes often requires significant modification during the testing and debug phase.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

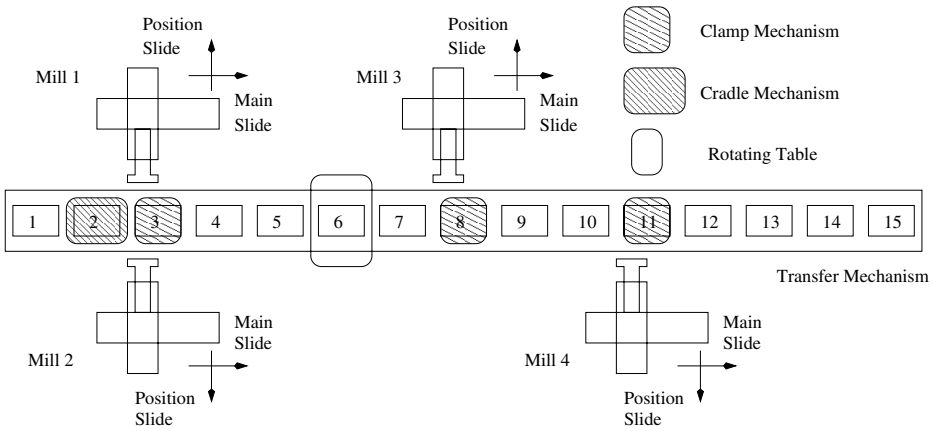


FIGURE 3.3 Sketch of a high-volume transfer line for engine block surface milling. Engine blocks move through the transfer line from station 1 to station 15. The system is composed of four milling machines, a transfer mechanism, and fixture mechanisms. The clamp mechanisms are fixtures for the milling machines and the cradle mechanism prevents interference between mill 2 and the engine block in location number 2. The transfer bar mechanism moves each engine block to next location in each cycle motion. The milling machines start to work after the engine blocks are located properly by the transfer bar mechanism, the cradle mechanism, and the clamp mechanisms.

3.2.4 Tasks of a Logic Control Programmer

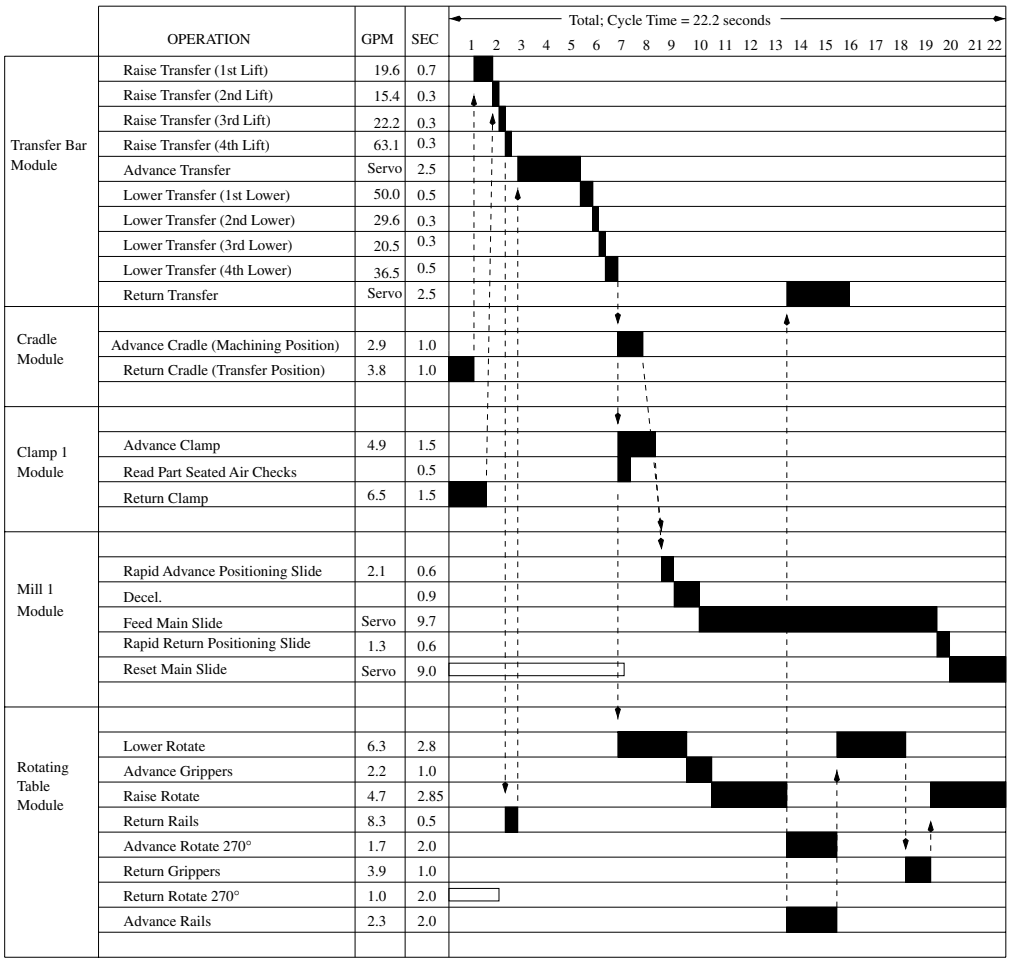
A major task in the design of manufacturing systems is the design and programming of the logic controllers. A logic control programmer starts from the mechanical definition of the machine and the tasks that it must perform. The inputs to the mechanical system (valves to control coolant and lubrication, motor drives, etc.) are identified, and a set of outputs (limit switches, proximity sensors, etc.) are determined. The total number of inputs and outputs for the system must be known before the control hardware can be specified. It is not uncommon for a machine tool to have 1000 or more I/O points; the complexity is considerable.

Each input and output must be assigned a unique address. Oftentimes, one controller is used for several machines. Even if the logic program is the same for each machine and can be written once and copied, the I/O addresses must be changed — a laborious process. A table of the I/O is maintained to guide the programmer as well as the electrician who will wire everything up.

Once all of the I/O are available, the logic control program must be written. A logic control program may be written as a sequence of if/then rules, or as a flow chart. For example, a logical statement may be "if the part is in place, then engage the clamp." The part is considered to be in place if the appropriate proximity sensor is active, and the clamp is engaged by turning on a solenoid. This statement is implemented in a low-level language as "if the memory location P contains a 1, then write a 1 to the memory location S." It is common for variables to be referred to by their memory locations and not by names; thus, the I/O table must be accurate and up-to-date. Logic control programs may also be written in a flow-chart type program to emphasize the sequential nature of the tasks.

Although each logical statement may be relatively simple, tens of thousands of such statements will be required to make the machine work properly. Also, the logic control program must implement all of the control modes, and it must prevent damage from occurring to the machine. For example, if a drill is extended, the "open clamp" command should be disabled. Other things that must be considered when writing the logic control program include supplying lubrication to a spindle and coolant to a machining operation, checking for availability of hydraulic fluids, as well as all the operator interfaces.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com



	OPERATION	GPM	SEC
Transfer Bar Module	Raise Transfer (1st Lift)	19.6	0.7
	Raise Transfer (2nd Lift)	15.4	0.3
	Raise Transfer (3rd Lift)	22.2	0.3
	Raise Transfer (4th Lift)	63.1	0.3
	Advance Transfer	Servo	2.5
	Lower Transfer (1st Lower)	50.0	0.5
	Lower Transfer (2nd Lower)	29.6	0.3
	Lower Transfer (3rd Lower)	20.5	0.3
	Lower Transfer (4th Lower)	36.5	0.5
	Return Transfer	Servo	2.5
Cradle Module			
	Advance Cradle (Machining Position)	2.9	1.0
	Return Cradle (Transfer Position)	3.8	1.0
Clamp 1 Module			
	Advance Clamp	4.9	1.5
	Read Part Seated Air Checks		0.5
	Return Clamp	6.5	1.5
Mill 1 Module			
	Rapid Advance Positioning Slide	2.1	0.6
	Decel.		0.9
	Feed Main Slide	Servo	9.7
	Rapid Return Positioning Slide	1.3	0.6
	Reset Main Slide	Servo	9.0
Rotating Table Module			
	Lower Rotate	6.3	2.8
	Advance Grippers	2.2	1.0
	Raise Rotate	4.7	2.85
	Return Rails	8.3	0.5
	Advance Rotate 270°	1.7	2.0
	Return Grippers	3.9	1.0
	Return Rotate 270°	1.0	2.0
	Advance Rails	2.3	2.0

FIGURE 3.4 A portion of the timing bar chart for the transfer line system shown in Figure 3.3. Each operation that must be performed by the system is listed on the left-hand side of the table; the horizontal axis indicates time. The solid lines indicate the amount of time taken by each operation, and the dotted lines indicate causal dependencies between operations. Note that all operations are synchronized to the transfer bar mechanism. The total cycle time is 22.2 seconds.

In any automated manufacturing system, safety is always a primary concern. Typically, the safety circuitry is not programmed into the logic controller but hardwired using relays. An "emergency stop" switch (big red button) is always available; when it is engaged, the machine will stop immediately. The logic control programmer is often responsible for specifying the emergency control logic to be wired by an electrician.

3.3 Current Industrial Practice

Logic controllers for manufacturing systems run on proprietary control systems known as PLCs, or programmable logic controllers.

3.3.1 Programmable Logic Controllers

PLCs are specialized computing devices designed for logic control. They combine a general-purpose microprocessor with discrete I/O capabilities, and are able to handle the thousands of inputs and

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

outputs that are necessary to control a manufacturing system. There are several manufacturers of PLCs, each with their own software tools for programming and slightly different interpretations of the standard languages. Code written for PLCs is not generally portable; a program written for an Allen-Bradley PLC will not run on a Modicon PLC without modification.

PLCs typically operate by reading all of the inputs to a system, then computing all of the logic, then writing all of the outputs. This "scan time" depends on the number of inputs and outputs as well as the complexity of the logic, and may not be repeatable from scan to scan. In addition, the same logical program implemented in a different language or even in the same language on a different platform may require a different scan time. For this reason, it is difficult to achieve guaranteed and repeatable real-time performance with PLCs.

In the early days of automated manufacturing, hardwired relays were used to control the logical behavior of the machines. The logic control "program" was an electromechanical circuit, and programming was done by electricians. When the first microprocessors became available, they were used to replace the unreliable relays. A programming language called "relay ladder logic" was developed to program these early logic controllers. Its graphical interface mimicked the appearance of relays, to make the transition from hardwiring to software easier.

3.3.2 Relay Ladder Logic

Almost 30 years after it was developed, ladder logic remains the industry standard for logic control. Ladder logic is similar to assembly language, the lowest-level programming language commonly used. This makes it easier to implement ladder logic on a microprocessor than it would be to implement a higher-level language. In addition, low-level languages such as assembly and ladder logic give the programmer full control over the instructions being executed on the processor. Programs written in these low-level languages can be made to run very efficiently.

A sample ladder logic program is shown in Figure 3.5. The main elements of ladder logic are normally open contacts, normally closed contacts, and output coils. The relay contacts switch from open to closed or vice versa if the corresponding input terminal or memory location contains a "high" voltage or a "1." Each rung of the ladder implements a simple "if/then" statement. If all of the relays in a rung are closed, then the output coil will be activated. In many implementations of ladder logic, an animated display can tell the programmer or operator which signals are high and which rungs are active, allowing for efficient low-level debugging.

However, because ladder logic is a low-level programming language, the programs for even a relatively small system rapidly become unwieldy (the printout may be several inches high). There is very little support for subroutines or procedures, and no sense of variable "scope." Because all variables are global, it is relatively easy for one part of a large program to mistakenly overwrite or change a variable used by another part of the program. In addition, no facility exists for structured data; only bits and registers are allowed.

Ladder logic has many disadvantages; programs written in ladder logic take longer to develop, are harder to maintain, and are less reusable than equivalent programs written in a higher-level language (such as C++). The most common method for reuse of ladder logic code is to copy the rungs of the ladder from an old program and paste them into a new program. The data I/O address must still be changed to match those of the current project. Databases and libraries can be developed to automate this process, but it is still tedious.

Several alternatives to ladder logic have been proposed. A new standard, the IEC 1131-3,[12,14] includes five distinct languages. One is the familiar ladder diagram; others include structured text, function block diagrams, instruction list, and sequential function charts. Although these languages are based on familiar languages, they have more support for subroutines, parameter passing, limited scope, and strongly typed variables. The standard is intended to allow software written for one brand of PLC to be able to be run on other brands of PLCs.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

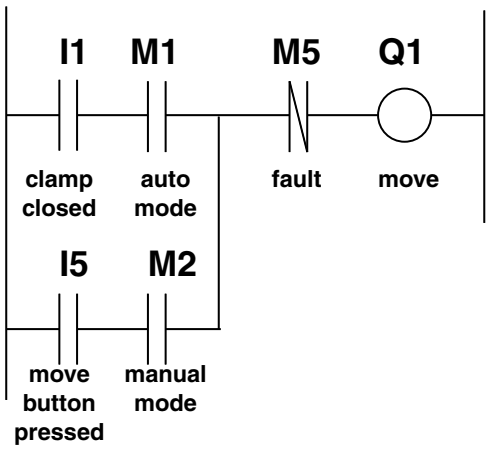


FIGURE 3.5 A sample of relay ladder logic. There are three types of elements: normally open contacts, normally closed contacts, and commands. The I1 and I5 are input signals from a clamp proximity switch and a pushbutton, respectively; the signals M1, M2, and M5 represent memory locations; and Q1 represents an output that may go to a solenoid or a memory location. The ladder diagram implements the following logical statement: "If (((I1 and M1) or (I5 and M2)) and not M5) then Q1;" or equivalently "If the clamp is closed and the system is in auto mode, or the move button is pressed and the system is in manual mode, and there is no fault, then move."

3.3.3 Sequential Function Charts

Sequential Function Chart (SFC) is one of the IEC 1131-3 languages for logic controllers.[12] It is based on Grafcet which was inspired from Petri nets, and thus logic controllers designed using Petri nets (see 3.5.4) can be easily implemented using SFC. Logic control programs can also be written directly in SFC.

Sequential Function Chart and Grafcet are both commonly used in industry along with the ladder diagram.[4,5] SFC programs have two types of nodes: steps and transitions. Steps are represented by squares and initial steps are represented by a double square. The steps in Grafcet can have only one token; in other words, the marking of a step is a Boolean representation. In SFC, a set of simultaneously firable transitions can be fired. It can be shown that a special class of Petri nets (safe marked graphs) is equivalent to SFC.

3.4 Current Trends

3.4.1 Issues with Current Practice

Because logic control programs must be implemented in proprietary programming languages, there is little ability to reuse code (or even library functions) from one project to the next unless the same brand of hardware is used. Even if the same hardware is used, and some code can be reused, the hardware is not inexpensive. Because there is a relatively small market for PLCs, they are expensive compared to more general-purpose computers (such as PCs) with similar performance. Hardware add-ons, such as video cards and networking cards, must be developed for each proprietary architecture and contribute significantly toward the overall cost of a PLC system.

Another major expense associated with discrete event control in a manufacturing system comes from the required electrical wiring. Each limit switch or proximity sensor must have power, and its output must be connected to the PLC. With hundreds or even thousands of I/O points on a typical machine, the labor needed to initially set up this wiring results in a high cost. Additionally,

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

such a mass of wires is extremely difficult to debug — and often the wires do get crossed or connected to the wrong terminal. The PLC and its associated I/O are typically housed in an electrical cabinet near the machine, along with the power supplies, transformers, and motor drives. The floor space consumed by this cabinet is significant.

Most PLC programming languages are fairly low level, requiring many lines of code to implement simple functions. The development time for such programs is relatively long. Some code can be reused mostly through copying and pasting previously written code. Because of the low-level language, all variables are referred to by either their I/O address or their memory address. Thus, if the same function is going to be performed on a different part of the same machine, the same code can be reused, but all of the variable names need to be changed.

In current practice, the logic programs are written while the machine is being built, and are verified on the machine during the ramp-up phase. No method for formally testing the program for correctness exists (although simple tests can be done to find inputs not used or conflicts in the logic program). Some work is currently being done to automatically convert ladder logic into a more formal discrete event system formalism for verification purposes.[24] However, current verification algorithms for discrete event formalisms test all possible combinations of states. With large systems, the number of combinations of states grows too large to feasibly test every combination.

3.4.2 PC-Based Control

There is currently a great deal of interest in moving away from standard logic controllers implemented as ladder logic on a PLC. Both hardware and software are changing. The drivers for this change include price and flexibility. As noted earlier, most PLC systems are proprietary, and even ladder logic programs are not interchangeable between brands. As special-purpose computing devices, PLCs have a relatively small market size. The competition is based on software and support; the hardware commands premium prices. The most likely successor of the PLC is an industrialized version of the desktop PC, which benefits from a large market share to drive down prices for microprocessors, memory, communication peripherals, etc. Because of this intense competition, PCs have much more computational power at a lower cost than PLCs. As the market moves toward general-purpose PCs, programming languages and development tools designed for conventional software will become available. There will certainly be ladder logic implementations on a PC, but more varied programming languages, more powerful and easier to use, will also become viable options. PC-based control will allow the continuous and discrete event control to be integrated on the same computer platform.

3.4.3 Distributed Control

Traditionally, the I/O for an entire machine was brought back to a centralized PLC. Now, distributed systems are being implemented. In a distributed system, a group of smaller PLCs each control a region or subsystem of the machine, and these PLCs communicate and cooperate to control the entire machine. These distributed systems are easier to wire up, and can be designed and debugged in a modular manner.

In some instances, all of the sensors and actuators for a machine may be connected to a sensor or control network. Instead of two or three wires for each sensor, there is one cable which brings both power and a network connection to each sensor. The sensor information is then transmitted to the PLC over the network. Control networks, or sensor networks, are high-bandwidth networks optimized for sending small, periodic packets of information, as opposed to data networks which send large, asynchronous packets of information.[15,21] Currently, these networks are used only to replace the wiring; in the future, each device may also have some embedded intelligence and be able to glean information off the network to determine appropriate control actions.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

3.4.4 Simulation

Although some simulation packages have recently become available, control systems for machining systems are typically not verified before they are implemented. A relatively long "cycle and debug" stage in the development process is used to fix most of the problems with the control code. In the past, a transfer line could be expected to build the same parts for 10 or more years. With reduced product lifecycles, the lifetime has been reduced to 5 years or less. Currently, the control system cannot be tested until all of the machinery is in place in the factory setting.

Simulation of the control system combined with the mechanical machine is becoming more common in industry, but is time consuming in terms of both operator setup and computer time. For an unfamiliar system, this may be warranted, but many systems are built as variations of previously built ones, and a reasonable degree of confidence in the correctness of the approach exists.

Several simulation environments are available for production systems, both from universities[20] and commercially.[6,25] A simulation of the manufacturing equipment can be built, and an interface built to the control system. Then the control system can "control" the simulation. Depending on the fidelity and accuracy of the simulation, the control software can be sufficiently tested before it is deployed on the plant floor. Performance can be predicted, and problems with collisions and timing discovered. Some environments provide simple 2-D line graphics; others use 3-D or even virtual reality to animate the manufacturing process.

In addition to control analysis and testing, these simulations have other advantages such as enabling process improvements by the manufacturing engineers (and subsequent changes to the control program) and operator training in a virtual environment. Because the control software and the manufacturing system are so complex, formal verification methods typically fail. However, in a simulation environment, many different test cases can be examined quickly, and some problems can be identified and fixed before they occur on the plant floor.

3.5 Formal Methods for Logic Control

Even though logic controllers are very important in the manufacturing industry, a standard integrated tool does not yet exist that is sufficiently simple to use, powerful, versatile and with which it is possible to carry out systematic analysis and design of discrete event control systems.

3.5.1 Important Criteria for Control

Logic controllers for manufacturing systems must satisfy a given set of criteria. The most important is performing the given task. The task may be defined as a single sequence of events or as an intertwined sequence such as a timing bar chart. It must not be possible for a logic controller to get stuck in a state from which it cannot move; this is formalized as the definition of deadlock-free. The systems must also be reversible, meaning that from any state, they can always return to the initial state with a suitable sequence of events. The time taken to complete one entire cycle of the operation is called the cycle time of the system; this time is often specified in advance (if not, it should be as short as possible while maintaining the desired part quality). In addition to performing the specified task in the automatic mode, the logic controller should contain some diagnostics to detect errors or problems when they occur, and either inform the operator or possibly take action to correct them. The manual modes must allow the operator enough flexibility to control the machine through a pushbutton interface.

3.5.2 Discrete Event Systems

A discrete event system is defined as a dynamic system whose evolution through the state space is defined by the occurence of instantaneous discrete events.[3] Examples of discrete events are the push of a button by an operator, the triggering of a limit switch, the activation of a solenoid, a tool breaking. An event occurs at some discrete moment in time rather than over a time interval.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

There has been quite a bit of academic research into the area of discrete event systems, primarily as they relate to computer programming. Most of the control-related research has been based on the framework set up by Ramadge and Wonham,[22,23] in which a controller is defined entirely as a mechanism for the prevention of unacceptable behavior, but not as a method of forcing a discrete event system to complete a task. There has been some research into the concept of forcing events,[1,9] but it is not yet complete. Very little extension of the formal academic theory to real manufacturing systems has occurred. The reason generally given is that as a system grows in complexity, the complexity of a discrete event system associated with the system grows at an exponential rate. This quickly leads to intractable controller verification.

A variety of representations of discrete event systems exist; a few of them are described in this section. Languages are the most general way to express a discrete event system. Any discrete event system can be described using a language, but generally languages are difficult to work with. Finite state machines are a common expression of a discrete event system and have a more well-defined structure than languages; however, they can still be quite complex. Basic descriptions of these representations can be found in Kumar and Garg.[13] Petri nets are another formalism for describing discrete event systems. All of these representations have their advantages and disadvantages. As a rule, the more general the representation of a discrete event system, the more difficult it is to prove desired properties about the system.

The most general representation of a discrete event system is in terms of language theory. The set of events which can occur in a system is denoted by the set $\Sigma = \{\sigma_1, \sigma_2, \ldots\}$. The basis of language theory is the "string," which represents one possible sequence of events which can occur in a discrete event system.

Definition 3.5.1 (Languages) A *string* is an ordered list of events, representing a possible sequence of events in a discrete event system. A *language* is a (possibly infinite) set of strings, representing all possible sequences of events that may occur in a discrete event system.

Given a set of events Σ, the language consisting of all possible strings with elements in this set is denoted Σ^*. Other languages with the same event set are subsets of this. Two strings s and t can be combined by concatenation; $s.t$ denotes the list of events in s immediately followed by the list of events in t.

Thus, a discrete event system with event set Σ has a language L which is a subset of Σ^*, i.e., $L \subseteq \Sigma^*$. Even if Σ is a finite set (which is not necessary), L is often an infinite set. This complexity of enumeration makes language theory difficult to work with from a computational point of view.

Although a language can describe any discrete event system, it is difficult to prove desirable properties of a system from its language definition. For this reason, other modeling formalisms such as finite state machines and Petri nets are more popular than language theory.

3.5.3 Finite State Machines

A finite state machine is a special type of discrete event system in which the event set Σ contains only a finite number of events. In addition, the language of the discrete event system must be describable in terms of the evolution of a state machine with finitely many states.

Definition 3.5.2 (Finite State Machine) A *finite state machine* is a quintuple, $S = \{X, \Sigma, \alpha, x_0, X_m\}$, where:

X = The finite set of all states in the FSM
Σ = The set of all events recognized by the FSM
α = The transition function; α: $X \times \Sigma \rightarrow X$
x_0 = The initial state
X_m = A set of marked states

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

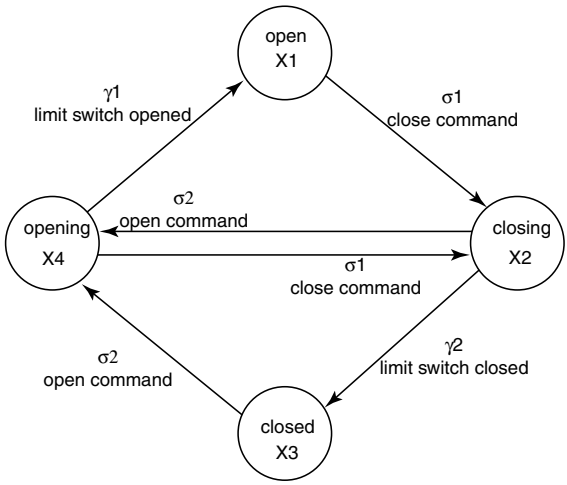


FIGURE 3.6 A finite state machine modeling a gripper or clamp. The system has four states ($X = \{x_1, x_2, x_3, x_4\}$) and four events ($\Sigma = \{\sigma_1, \sigma_2, \gamma_1, \gamma_2\}$).

The transition function α is not generally defined for all possible event/state pairs. At any state, only a subset of the events in Σ can happen. The function α is generally extended recursively to map the set of all states $\times$ strings to the set of states as follows. For a string s, a state x, and an event σ, if $\alpha(x, s) = x'$ then $\alpha(x, s.\sigma) = \alpha(x', \sigma)$. This is equivalent to the state reached if all the events in the string are executed sequentially starting from the state x.

The marked states typically represent some desired final states that the finite state machine should reach. If only cyclic behavior is desired, then the initial state may be the only marked state. In other cases, more than one marked state may be used to denote different execution models.

The language admitted by the finite state machine S is denoted $\mathcal{L}(S)$. This is the set of all strings admitted by the finite state machine. The marked language $\mathcal{L}_m(S)$ of the finite state machine can also be defined as the set of all strings which take the initial state to a marked state.

$$\mathcal{L}(S) = \{s \in \Sigma^*: \alpha(x_0, s) = x \in X\}$$

$$\mathcal{L}_m(S) = \{s \in \Sigma^*: \alpha(x_0, s) = x \in X_m\}$$

A finite state machine is generally visualized as a set of nodes representing the states connected by a set of arrows labeled with events representing the transitions. The finite state machine shown in Figure 3.6 can be used as a model of a gripper or clamp. There are four states representing the discrete state of the gripper (x_1 = open, x_2 = closing, x_3 = closed, and x_4 = opening). There are four events in the system. Two of them, σ_1 = close and σ_2 = open represent commands that tell the gripper to change state. The other two, γ_2 = closed and γ_1 = opened, represent limit switches that trip when the gripper has completed its state transition. Not every event is allowed at every state. The state transition function α can be specified by enumeration; it is given in Table 3.1.

TABLE 3.1 The State Transition Function α for the Finite State Machine in Figure 3.6.

α	σ_1	σ_2	γ_1	γ_2
x_1	x_2	–	–	–
x_2	–	x_4	x_3	–
x_3	–	x_4	–	–
x_4	x_2	–	–	x_1

States are listed along the left-hand side, events across the top, and the entries in the table indicate the state that results after an event occurs. Entries marked with a – indicate that the corresponding event cannot occur when the system is in that state.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

3.5.3.1 Combinations of Finite State Machines

When broken down into small pieces such as the four-state system described above, discrete event models of manufacturing systems are relatively simple. The complexity arises when many small pieces are put together to form the discrete event model of the entire system. The formal method for combining finite state machines is through parallel composition. For a finite state machine $S = \{X_S, \Sigma_S, \alpha_S, x_{S_0}, X_{S_m}\}$, let the set of events which can occur at a state $x \in X$ be denoted $\Sigma_S(x)$.

$$\Sigma_S(x) = \{\sigma \in \Sigma_S : \alpha(x, \sigma) \in X\}$$

Parallel composition for two finite state machines is then defined as follows.

Definition 3.5.3 (Parallel Composition, ||) Given two finite state machines, A and B:

$$A = \{X_A, \Sigma_A, \alpha_A, x_{A0}, X_{Am}\}$$

$$B = \{X_B, \Sigma_B, \alpha_B, x_{B0}, X_{Bm}\}$$

The *parallel composition* of A and B is defined as:

$$A \,\|\, B = \{(X_A \times X_B), (\Sigma_A \cup \Sigma_B), \alpha, (x_{A0}, x_{B0}), X_{Am} \times X_{Bm}\}$$

where the transition function α of the parallel composition A || B is defined as:

$$\alpha((x_A, x_B), \alpha) = \begin{cases} (\alpha_A(x_A, \sigma), \alpha_B(x_B, \sigma)) & \text{if } \sigma \in \Sigma_A(x_A) \cap \Sigma_B(x_B) \\ (\alpha_A(x_A, \sigma), x_B) & \text{if } \sigma \in \Sigma_A(x_A) \text{ and } \sigma \notin \Sigma_B \\ (x_A, \alpha_B(x_B, \sigma)) & \text{if } \sigma \in \Sigma_B(x_B) \text{ and } \sigma \notin \Sigma_A \\ \text{undefined} & \text{otherwise} \end{cases}$$

In other words, the parallel composition of two machines is equivalent to running both machines simultaneously, with the restriction that events which are elements of both event sets must occur concurrently in both machines. Although the number of events in the combined state machine is (at most) the sum of the number of events in Σ_A and Σ_B, the number of states in the parallel composition is the *product* of the number of states in each state machine *A* and *B*. This leads to the state explosion property as many finite state machines are combined.

3.5.3.2 Supervisory Control of Discrete Event Systems

The most prevalent framework for supervisory control of discrete event systems is that of Ramadge and Wonham.[22,23] In this formalism the set of events Σ is divided into two subsets, labeled Σ_c and Σ_u, called the "controllable" and "uncontrollable" events, respectively. All events must be in one of the two sets, thus $\Sigma = \Sigma_c \cup \Sigma_\sigma$ and $\Sigma_c \cap \Sigma_\sigma = \varnothing$. As suggested by the name, controllable events, can be disabled by the supervisor, which means that any transition labeled with a controllable event can be removed at will. For example, in the simple finite state machine of Figure 3.6, the events $\Sigma_c = \{\gamma_1, \gamma_2\}$ which represent the open and close commands would be considered controllable. The events $\Sigma_u = \{\gamma_1, \gamma_2\}$ which represent the trippings of the two limit switches are influenced by a physical process and would generally be considered uncontrollable.

The supervisor consists of another state machine, which operates on the same set of events as the finite state machine being controlled. Associated with each state in the supervisor finite state machine is a set of controllable events that are enabled when the supervisor is in that state. Only events in this set can be executed by the machine being controlled. At all times every uncontrollable

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

event must be enabled. For example, consider the state machine shown in Figure 3.6 representing a robot gripper combined with a similar one representing the robot moving from one location to another (say to pick and place a part). Once the robot is positioned over the part to be grasped, the supervisor would enable σ_1, representing the close command. Once the event γ_2 has occurred, signaling that the gripper is closed, the supervisor would enable the event corresponding to commanding the robot to move to the desired destination, and so forth. The supervisor keeps track of the state of the system and enables events appropriate to that state.

Another notion of supervisory control of discrete event systems has also been proposed.[1,9] In this framework, controlled events are those that can be forced by the supervisor onto the plant, and will only happen then. Uncontrollable events are those the plant can force on the supervisor. Technically, these notions are equivalent;[2] either can be used to design and analyze finite state machine controllers.

3.5.3.3 Verification of Closed-Loop Behavior

Specification of the desired behavior of control system in the finite state machine framework is generally given in terms of the language L that should be admitted by the finite state machine. The language may be enumerated or specified by another finite state machine. The controller finite state machine will act to disable controllable events that should not happen in certain states.

For a plant P and a controller C, both finite state machines, the controlled (closed-loop) behavior is defined as the parallel composition of the two, $P \parallel C$. The controller should interact with the plant in such a manner that the parallel combination can always reach a marked state; this is formalized by the definition of non-blocking.

Definition 3.5.4 (Non-blocking) A finite state machine S is *non-blocking* if a marked state can be reached from every state in the machine. That is, for every state $x \in$ X, there exists a sequence of events $s = \sigma_1 \sigma_2 \ldots$ such that $\alpha\,(x, s) = x_m \in X_m$.

The existence of a non-blocking controller/plant combination that allows all possible uncontrollable events to occur can be determined based on the finite state machine representing the plant and the desired language L.[22] If a non-blocking controller exists, it can be constructed in a straightforward manner as a finite state machine.

Most approaches for verification of finite state machines rely on enumerating all of the states and events to guarantee that an undesirable state is never reached. Even though there are finitely many states and finitely many events, the number of states grows exponentially as more state machines are combined together using parallel composition. Thus, although techniques exist for constructing a supervisory controller given the finite state machine of the plant P and the specified closed-loop behavior (the language L), the large size of the resulting state space limits the sizes of systems that can be handled.

3.5.4 Petri Nets

Petri nets, as graphical and mathematical tools, provide a powerful environment for modeling, formal analysis, and design of discrete event systems. Historically, Carl Adam Petri first developed Petri nets in 1962 as a net-like mathematical tool for the study of communication. Since that time, they have found many uses in a wide variety of applications such as communication protocols, manufacturing systems, and software development. A good survey on properties, analysis, and applications of Petri nets can be found in References 4, 7, 16, and 27.

Petri nets have been used as an analysis tool for event-based systems that are characterized as being concurrent, synchronized, and distributed. Petri nets enable the qualitative and quantitative analysis of an event-based system. The modeled system can be verified to be correct from the qualitative analysis, and the efficiency of the modeled system can be determined from the quantitative analysis.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

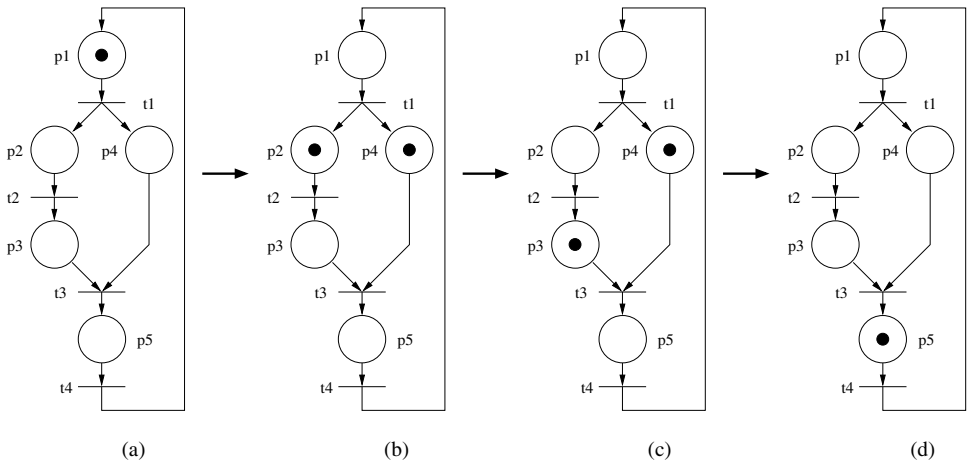


FIGURE 3.7 Marking evolution of an ordinary Petri net: (a) initial marking, (b) firing transition t_1, (c) firing transition t_2, (d) firing transition t_3. In the initial marking (a), there is one token in p_1. Transition t_1 is enabled because all places leading to it are marked; it is the only transition enabled. After transition t_1 fires, the marking becomes that shown in (b). Each place leading out of transition t_1 gets a token. Now transition t_2 is enabled (transition t_3 cannot fire until both places p_3 and p_4 are marked). If more than one transition is enabled at a time, the Petri net exhibits nondeterministic behavior. After t_4 fires, the Petri net returns to its initial marking (a). The Petri net is thus said to be reversible.

Petri net models are used to analyze three important properties of a discrete event system: liveness, safeness, and reversibility. The meanings of these properties for a Petri net for a logic controller are summarized in Reference 26 and are discussed later in this section. By analyzing these properties of the Petri net model, the functional correctness of the logic generated from the Petri net model can be assured.

3.5.4.1 Graphical Representation of Petri Nets

As stated above, a Petri net is a mathematical formalism which has a simple graphical representation. Petri nets consist of two types of nodes: places represented by circles, and transitions represented by bars. Nodes are connected by directed arcs. The dynamics of a Petri net are determined by its initial marking and marking evolution rule. A marking assigns to each place a nonnegative integer and the integer value is graphically represented by the number of tokens in each circle (place). The number of tokens in a place represents the local state of the place and the state of the whole system is defined by the collection of local states of the places. A pictorial example of the evolution of an ordinary Petri net is given in Figure 3.7.

A Petri net and its evolution rule can be represented formally by the following definitions.

Definition 3.5.5 (Petri Nets) A *Petri net* is a four-tuple, $\langle P, T, F, W \rangle$ where:

$Px = x\{p_1, p_2, \ldots, p_n\}$, *a finite non-empty set of places*
$Tx = x\{t_1, t_2, \ldots, t_m\}$, *a finite non-empty set of transitions*
$F \subset (P \times T) \cup (T \times P)$, *the flow relation (set of directed arcs)*
$Wx{:}xF \rightarrow Z^+$, *the weight function which assigns an integer weight to each arc*

A Petri net is termed *ordinary* if all the arc weights are one. A *marking M* of a Petri net *N* is the assignment of a nonnegative integer to each place. It is an *n*-dimensional state-vector of the Petri net system. A Petri net with the given initial marking is denoted by $\langle N, M_0 \rangle$. The state or marking in a Petri net evolves according to the following transition (evolution) rules:

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

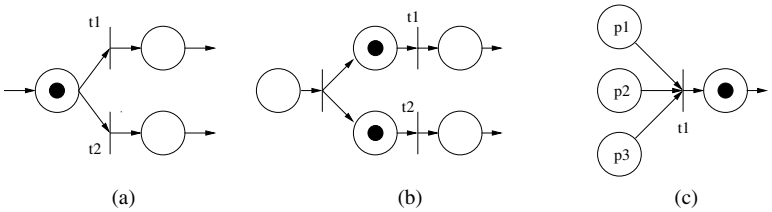


FIGURE 3.8 Some modeling capabilities of Petri nets: (a) conflict: if t_1 fires, t_2 is not enabled and vice versa; (b) concurrency: t_1 and t_2 can be fired independently; (c) synchronization: t_1 synchronizes p_1, p_2, and p_3.

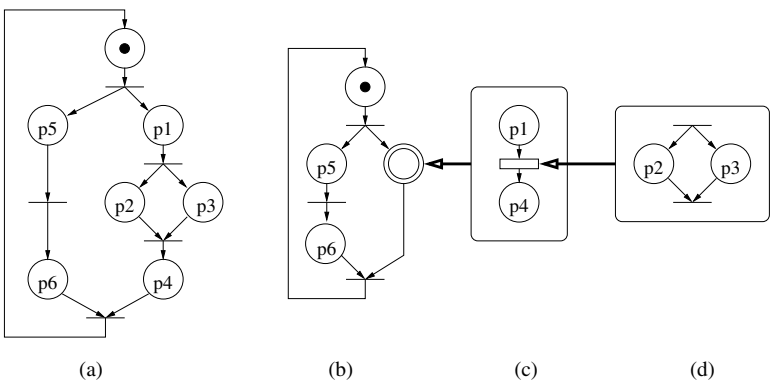


FIGURE 3.9 Hierarchical representation of a Petri net. The original Petri net is shown in (a). The hierarchical reduction (b) uses a double circle place to encapsulate the right branch of the Petri net. In (c), the internal structure of the double circle is shown with the original places p_1 and p_4 with the box transition as another hierarchical level. The internal structure of the box transition is shown in (d) with the original places p_2 and p_3.

1. A transition t is enabled if each input place p of t is marked with at least as many tokens as the weight of the arc joining them.
2. An enabled transition may or may not fire depending on whether or not the transition (event) actually takes place.
3. A firing of an enabled transition t removes $w(p, t)$ tokens from each input place p of t, and adds $w(t, p)$ tokens to each output place p of t.

The use of Petri nets in modeling manufacturing systems has several practical features. It can easily model causal dependencies, conflicts, synchronization, mutual exclusion, and concurrency. Some of these modeling capabilities are shown in Figure 3.8. Petri nets also have a locality property on places and transitions which enables hierarchical and modular constructions of complicated systems; a hierarchical representation of a Petri net is shown in Figure 3.9.

In Section 3.2, an example of a timing bar diagram for a transfer line was given. For simplicity, consider only the behavior of mill 1. The Petri net shown in Figure 3.10 describes its behavior as specified by the timing bar chart of Figure 3.4. Here places represent operations and transitions are enabled at the end of operations. A place with the notation "W" represents a waiting place; these places are very useful in modeling synchronization among operations. When the mill Petri net is combined with the clamp Petri net, as shown in the figure, the mill operation "rapid advance" cannot occur until the clamp has advanced due to the synchronizing transition. The two waiting places in the clamp Petri net indicate synchronizations with other parts of the machining system. The tokens are shown in their initial places, representing the starting moment of the timing bar chart.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

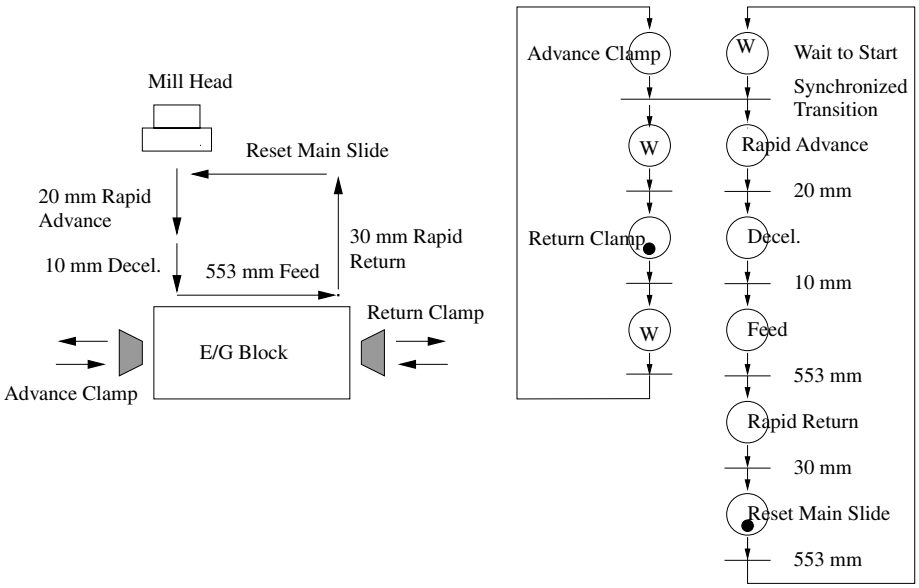


FIGURE 3.10 A Petri net implementation of the controller for mill 1 along with the corresponding clamp.

3.5.4.2 Analysis of Petri Net Models

As mentioned earlier, the qualitative properties that are especially important in Petri net models for manufacturing systems are liveness, boundedness or safeness, and reversibility. The formal definitions of these properties are omitted here, but their general meaning in logic controllers in manufacturing systems can be summarized as follows:[17,26]

1. Boundedness or safeness guarantees the stable behavior of the system without any overflow. The safeness property of the places which represent operations indicates there is no attempt to request execution of an ongoing operation. Another important implication of safeness is the Boolean representation of places, which enables a direct conversion from a Petri net to SFC as shown in Figure 3.11.
2. Liveness is equivalent to absence of deadlocks. This property guarantees that all transitions can be firable and that all operations or conditions represented by places can happen.
3. Reversibility characterizes the recoverability of the initial state from any reachable state of the system. It implies the cyclic behavior of a system and that it will perform its function repeatedly.

Petri net models of logic controllers can be formally analyzed to verify that the boundedness, liveness, and reversibility properties are satisfied. This verification process can guarantee that the corresponding manufacturing system exhibits the desired behavior. There are three approaches to the analysis of these qualitative properties: analysis by enumeration, analysis using linear algebraic techniques, and analysis by transformation.[8,16] The enumeration methods are based on the construction of the reachability graph or the coverability graph of the Petri net. The linear algebraic techniques use the state transition equation to represent the evolution of a Petri net and derive some invariant structures. The transformation method is based on simple reduction rules that preserve the important properties of Petri nets (boundedness, liveness, and reversibility); some simple reduction rules are presented graphically in Figure 3.12. The transformation procedure is iterative and applies the reduction rules until the reduced Petri net becomes irreducible. Generally, the first two techniques are limited by the complexity of the system. Although reduced Petri nets are irreducible, they may not be simple to analyze. One of the first two methods, however, can then

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

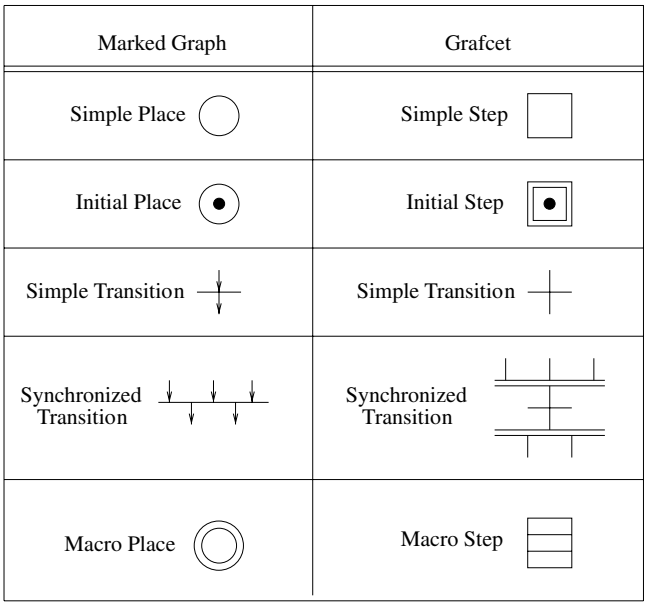


FIGURE 3.11 The conversion rules between a marked graph Petri net and Grafcet or SFC.

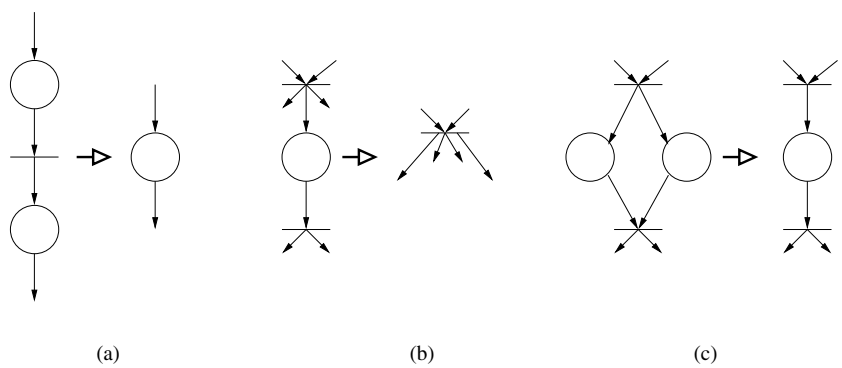


FIGURE 3.12 Simple reductions of Petri net models that preserve the properties of liveness, safeness (or boundedness), and reversibility: (a) fusion of series places, (b) fusion of series transitions, (c) fusion of parallel places.

be applied to the reduced model. In other words, these techniques are complementary and not exclusive.

Petri nets models can be categorized into several subclasses based on their structural characteristics. Analysis techniques are well developed for some subclasses of Petri nets, and the properties of Petri net models can easily be verified using these powerful structural results. For example, a Petri net is said to be *strongly connected* if there exists a directed path (sequence of places and transitions) from every place to every transition, and from every transition to every place. Many properties of Petri nets rely on the definition of *directed circuit*, a sequence of connected places and transitions with the final place being the same as the initial place.

The behavior of many manufacturing systems, including the high-volume transfer line shown in Figure 3.3, can be represented by a subclass of Petri nets called *marked graphs*. In a marked graph, each place p has exactly one input transition and exactly one output transition. Although transitions can have multiple input and output places, the marked graph formulation does not allow for

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

contention for operations; the systems have no conflict and are decision free. The simpler structure of the marked graph allows many theoretical results to be derived.[8] For example, a marked graph is live if and only if its initial marking contains at least one token on each directed circuit. A live marked graph is safe if and only if there is exactly one token on each directed circuit. An initial marking of the marked graph which results in the graph being both live and safe can be found if and only if the graph is strongly connected. And finally, a marked graph is reversible if and only if it is live. Therefore, the verification procedure of safeness, liveness, and reversibility properties of Petri nets representing a manufacturing system and its logic control can be simplified if the system can be modeled using a marked graph or one of the other subclasses of Petri nets with well-understood structural behaviors.

3.6 Further Reading

Industry is currently moving toward open-architecture control systems for manufacturing automation. There are several national and international efforts to define, formalize, and institute an open-architecture standard. OSACA, which began as a European consortium, has focused on developing an open interface standard for proprietary control systems.[19] This open standard allows integration and communication between different systems; OSACA-compliant commercial systems are currently being produced. The North American effort, OMAC, grew out of a specification issued by the "Big 3" auto manufacturers in 1994.[18] The entire control system, from the interface to the factory network down to the servo control algorithm, must be open and user-modifiable. Although this architecture gives much more freedom and flexibility to the end-user, technical and business issues remain to be addressed before it becomes practical. Numerous control system companies promote varying degrees of openness in their products. For up-to-date information, the reader is encouraged to consult the web sites for the various open control consortia as well as the National Industrial Automation Show and Conference (in conjunction with *National Manufacturing Week*) and the International Automotive Manufacturing Conference sponsored by the Society of Automotive Engineers.

This chapter discussed only a few of the PLC languages currently in use in industry. The language of choice in a given factory depends on the industry (automotive, chemical, etc.) as well as the geographic location. An international standard, the IEC 1131, attempts to unify the many languages in use to enable conversion between them. More information can be found in the standard[12] and in textbooks.[14]

The logic controllers discussed in this chapter are typically implemented using digital computers. The field of real-time computer systems is focused on issues of operating systems, networks, applications programming, formal analysis, and design of algorithms with a focus on real-time issues. This area, which is directly relevant to the topic discussed here, is a very active area of research and development. The reader is referred to proceedings of the IEEE Real-Time Systems Symposium for recent developments in this field.

Within the field of control theory, there has been a lot of work on the theory of discrete event systems. This work is focused on fundamental concepts of controllability, observability, controller synthesis, etc. for discrete event systems. In this approach, the system to be controlled is modeled as a finite state machine with discrete event inputs and outputs, and closed-loop specifications are given in terms of the language generated by the machine. The reader is referred to recent books[3,13] for background in this area. The *IEEE Transactions on Automatic Control,* the *Journal of Discrete Event Systems,* and the *Proceedings of the IEEE Conference on Decision and Control should* be consulted for the latest developments in this field.

In this chapter, we have intentionally ignored the interactions between the logic controller and the servo controllers used to control continuous variables such as position, velocity, etc. Hybrid systems is an emerging field of research emplasizing systems that contain both continuous variables

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

and discrete variables. The work in this area is focused on defining appropriate frameworks for analyzing and designing such hybrid systems. The interested reader is referred to the *Proceedings of the Workshop on Hybrid Systems*[10,11] for the latest developments.

Some of the formal methods for discrete event control described in this chapter are also used to control the scheduling of flexible manufacturing systems. Flexible systems, which produce many different types of parts on the same set of machines, are much more complex than the systems described thus far. In addition to the logic control for each machine, a supervisor or scheduler must determine which parts to send to which machine at which time. The supervisor tries to optimize the overall performance of the manufacturing system, but with unknown part mixes and potential machine breakdowns, the problem can become intractable.

Acknowledgments

We were introduced to the problems in discrete event control of manufacturing systems through the Engineering Research Center for Reconfigurable Machining Systems at the University of Michigan. We thank the director and co-director of this center, Yoram Koren and A. Galip Ulsoy, for giving us the opportunity to work in this exciting area. We also acknowledge our students who have worked with us on these problems: Eric Endsley, Dhrubajyoti Kalita, Morrison Lucas, and Euisu Park. In particular, Euisu Park supplied many of the figures used in this chapter. Credit also goes to the control engineers at Lamb Technicon, including Bryan Graham, Mike Griffin, and Matt VanGilder, who have shared their expertise in logic control with us.

References

1. S. Balemi, G. J. Hoffman, P. Gyugyi, H. Wong-Toi, and G. F. Franklin, Supervisory control of a rapid thermal multiprocessor, *IEEE Transactions on Automatic Control*, 38(7), 1040–1059, July 1993.
2. G. Barrett and S. Lafortune, Bisimulation, the supervisory control problem and strong model matching for finite state machines, *Journal of Discrete Event Dynamical Systems*, 8(4), 1998.
3. C. G. Cassandras and S. L. Lafortune, *Introduction to Discrete Event Systems,* Kluwer, Boston, 1999.
4. R. David, Grafcet: A powerful tool for specification of logic controllers, *IEEE Transactions on Control Systems Technology*, 3(3), 253–268, September 1995.
5. R. David and H. Alla, Petri nets for modeling of dynamic systems — A survey, *Automatica*, 30(2), 175–202, 1994.
6. Deneb Robotics, http://www.deneb.com.
7. A. A. Desrochers and R. Y. Al-Jaar, *Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis*, IEEE Press, Piscataway, NJ, 1995.
8. F. Dicesare, G. Harhalakis, J. M. Proth, M. Silva, and F. B. Vernadat, *Practice of Petri Nets in Manufacturing*, Chapman & Hall, New York, 1993.
9. C. H. Golaszewski and P. J. Ramadge, Control of discrete event processes with forced events. In *Proceedings of the IEEE Conference of Decision and Control*, 247–251, December 1987.
10. Hybrid Systems: Computation and control, The First International Workshop, HSCC'98, Berkeley, California, Springer, April 1998.
11. International Conference on Hybrid Systems, *Lecture Notes in Computer Science*, Springer, 1994, 1995, 1996.
12. International Electrotechnical Commission (IEC), *Programmable Controllers Programming Languages, IEC Standard 1131, Part 3*, 1993.
13. R. Kumar and V. K. Garg, *Modeling and Control of Logical Discrete Event Systems*, Kluwer, Boston, 1995.
14. R. W. Lewis, *Programming Industrial Control Systems Using IEC 1131-3*, Institution of Electrical Engineers, London, 1995.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

15. F.-L. Lian, J. R. Moyne, and D. M. Tilbury, Performance evaluation of control networks: Ethernet, ControlNet, and DeviceNet, *IEEE Control Systems Magazine*, 21(3), 66-83, February 2001.
16. T. Murata, Petri nets: Properties, analysis and applications, *Proceedings of the IEEE*, 77(5), 541–580, April 1989.
17. Y. Narahari and N. Viswanadham, A Petri net approach to the modeling and analysis of flexible manufacturing systems, *Annals of Operations Research*, 3, 449–472, 1985.
18. Open modular architecture controls, http://www.arcweb.com/omac/.
19. Open system architecture for controls within automation systems, http://www.osaca.org.
20. G. Pritschow, A. Storr, and T. Jost, Simulation-based testing environment for master control systems, *Production Engineering*, IV(1), 51–54, 1997.
21. R. S. Raji, Smart networks for control, *IEEE Spectrum*, 31(6), 49–55, June 1994.
22. P. J. G. Ramadge and W. M. Wonham, Supervisory control of a class of discrete event processes, *SIAM Journal of Control and Optimization*, 25(1), 206–230, January 1987.
23. P. J. G. Ramadge and W. M. Wonham, The control of discrete event systems, *Proceedings of the IEEE*, 77(1), 81–98, January 1989.
24. M. Rausch and B. Krogh, Formal verification of PLC programs. In *Proceedings of the American Control Conference*, 234–238, 1998.
25. Sirius Systems, http://www.sirius.com.
26. M. C. Zhou, F. Dicesare, and A. A. Desrochers, A hybrid methodology for synthesis of Petri net models for manufacturing systems, *IEEE Transactions on Robotics and Automation*, 8(3), 350–361, June 1992.
27. R. Zurawski and M. C. Zhou, Petri nets and industrial applications: A tutorial, *IEEE Transactions on Industrial Electronics*, 41(6), 567–582, December 1994.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

4

Machine Tool Dynamics and Vibrations

4.1 Introduction
Mechanical Structure • Drives • Controls
4.2 Chatter Vibrations in Cutting
Stability of Regenerative Chatter Vibrations in Orthogonal Cutting
4.3 Analytical Prediction of Chatter Vibrations in Milling
Dynamic Milling Model • Chatter Stability Lobes

Yusuf Altintas
The University of British Columbia

4.1 Introduction

The accuracy of a machined part depends on the precision motion delivered by a machine tool under static, dynamic, and thermal loads. The accuracy is evaluated by measuring the discrepancy between the desired part dimensions identified on a part drawing and the actual part achieved after machining operations. The cutting tool deviates from a desired tool path due to errors in positioning the feed drives, thermal expansion of machine tool and workpiece structures, static and dynamic deformations of machine tool and workpiece, and misalignment of machine tool drives and spindle during assembly. Because the parts to be machined will vary depending on the end-user, the builder must design the machine tool structure and control of drives to deliver maximum accuracy during machining.

A machine tool system has three main groups of parts: mechanical structures, drives, and controls.

4.1.1 Mechanical Structure

The structure consists of stationary and moving bodies. The stationary parts carry moving bodies, such as table and spindle drives. They must be designed to carry large weights and absorb vibrations transmitted by the moving and rotating parts. The stationary parts are generally made of cast iron, concrete, and composites, which have high damping properties. The contact interface between the stationary and moving bodies can be selected from steel alloys that allow surface hardness in order to minimize wear.

4.1.2 Drives

In machine tools moving mechanisms are grouped into spindle and feed drives. The spindle drive provides sufficient angular speed, torque, and power to a rotating spindle shaft, which is held in

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

the spindle housing with roller or magnetic bearings. Spindle shafts with a medium-speed range are connected to the electric motor via belts. There may be a single-step gear reducer and a clutch between the electric motor and spindle shaft. High-speed spindles have electric motors built into the spindle in order to reduce the inertia and friction produced by the motor–spindle shaft coupling. The feed drives carry the table or the carriage. In general, the table is connected to the nut, and the nut houses a lead screw. The screw is connected to the drive motor either directly or via a gear system depending on the feed speed, inertia, and torque reduction requirements. High-speed machine tools may employ linear direct motors and drives without the feed screw and nut, thus avoiding excessive inertia and friction contact elements. The rotating parts such as feed screws and spindles are usually made of steel alloys, which have high elasticity, a surface-hardening property, and resistance against fatigue and cracks under dynamic, cyclic loads.

4.1.3 Controls

The control parts include servomotors, amplifiers, switches, and computers. The operator controls the motion of the machine from an operator panel of the CNC system.

Readers are referred to machine design handbooks and texts for the basics of designing stationary, linearly moving, and rotating shafts.[1] The principles of machine tool control can be found in dedicated texts.[2,3] The fundamentals of machine tool vibrations, which are unique to metal cutting, are covered in this handbook.

4.2 Chatter Vibrations in Cutting

Machine tool chatter vibrations occur due to a self-excitation mechanism in the generation of chip thickness during machining operations. One of the structural modes of the machine tool–workpiece system is excited initially by cutting forces. A wavy surface finish left during the previous revolution in turning, or by a previous tooth in milling, is removed during the succeeding revolution or tooth period and also leaves a wavy surface due to structural vibrations.[4] Depending on the phase shift between the two successive waves, the maximum chip thickness may exponentially grow while oscillating at a chatter frequency which is close to, but not equal to, a dominant structural mode in the system. The growing vibrations increase the cutting forces and may chip the tool and produce a poor, wavy surface finish. The self-excited chatter vibrations may be caused by mode coupling or regeneration of the chip thickness.[5] Mode-coupling chatter occurs when there are vibrations in two directions in the plane of cut. Regenerative chatter occurs due to phase differences between the vibration waves left on both sides of the chip, and occurs earlier than mode-coupling chatter in most machining cases. Hence, the fundamentals of regenerative chatter vibrations are explained in the following section using a simple, orthogonal cutting process as an example.

4.2.1 Stability of Regenerative Chatter Vibrations in Orthogonal Cutting

Consider a flat-faced orthogonal grooving tool fed perpendicular to the axis of cylindrical shaft held between the chuck and the tail stock center of a lathe (see Figure 4.1). The shaft is flexible in the direction of feed, and it vibrates due to feed cutting force (F_f). The initial surface of the shaft is smooth without waves during the first revolution, but the tool starts leaving wavy surface behind due to vibrations of the shaft in the feed direction y which is in the direction of radial cutting force (F_f). When the second revolution starts, the surface has waves both inside the cut where the tool is cutting (i.e., inner modulation, $y(t)$) and outside surface of the cut due to vibrations during the previous revolution of cut (i.e., outer modulation, $y(t - T)$). The resulting dynamic chip thickness $h(t)$ is no longer constant, but varying as a function of vibration frequency and the speed of the workpiece,

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

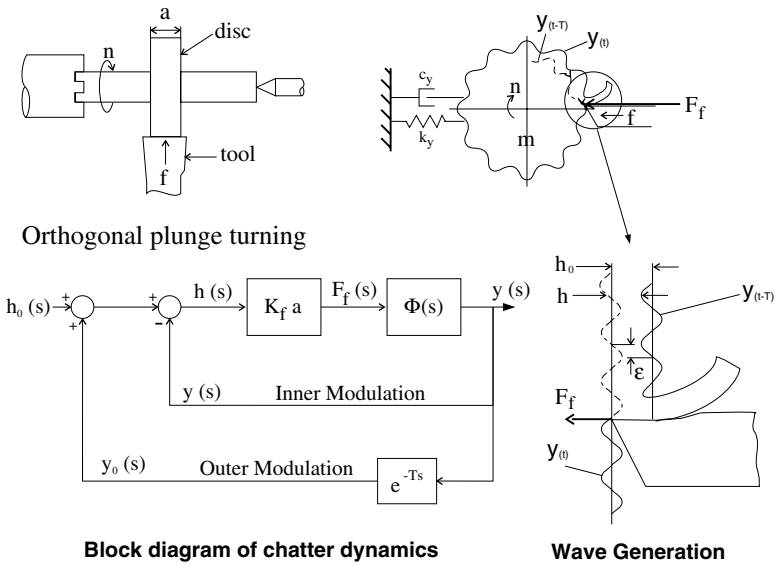


FIGURE 4.1 Mechanism of chatter vibrations in a plunge turning process.

$$h(t) = h_0 - [y(t) - y(t-T)] \tag{4.1}$$

where h_0 is the intended chip thickness which is equal to the feed rate of the machine. Assuming that the workpiece is approximated as a single degree-of-freedom system in the radial direction, the equation of motion of the system can be expressed as

$$\left.\begin{aligned} m_y y(t) + c_y y(t) + k_y y(t) = F_f(t) &= K_f a h(t) \\ &= K_f a[h_0 + y(t-T) - y(t)] \end{aligned}\right\} \tag{4.2}$$

where the feed cutting force $F_f(t)$ is proportional to the cutting constant in the feed direction (K_f), width of cut a, and the dynamic chip load $h(t)$. Because the forcing function on the right-hand side depends on the present and past solutions of vibrations ($y(t)$, $y(t-T)$) on the left side of the equation, the chatter vibration expression is a delay differential equation. The jumping of the tool due to excessive vibrations, and the influence of vibration marks left on the surface during the previous revolutions may further complicate the computation of exact chip thickness. The cutting constant K_f may change depending on the magnitude of instantaneous chip thickness and the orientation of the vibrating tool or workpiece, which is additional difficulty in the dynamic cutting process. When the flank face of the tool rubs against the wavy surface left behind, additional process damping is added to the dynamic cutting process which attenuates the chatter vibrations. The whole process is too complex and nonlinear to model correctly with analytical means, hence time-domain numerical methods are widely used to simulate the chatter vibrations in machining. However, a clear understanding of chatter stability is still important and best explained using a linear stability theory. The stability of chatter vibrations is analyzed using linear theory by Tobias,[6] Tlusty,[4] and Merritt.[7]

The chatter vibration system can be represented by the block diagram shown in Figure 4.1, where the parameters of the dynamic cutting process are shown in a Laplace domain. Input to the system is the desired chip thickness h_0, and the output of the feedback system is the current vibration $y(t)$ left on the inner surface. In the Laplace domain, $y(s) = \mathcal{L}y(t)$, and the vibration imprinted on the

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

outer surface during the previous revolution is $e^{-sT}y(s) = \mathcal{L}y(t-T)$ where T is the spindle period. The dynamic chip thickness in the Laplace domain is

$$h(s) = h_0 - y(s) + e^{-sT}y(s) = h_0 + (e^{-sT} - 1)y(s) \tag{4.3}$$

which produces dynamic cutting force,

$$F_f(s) = K_f\, a h(s) \tag{4.4}$$

The cutting force excites the structure and produces the current vibrations $y(s)$,

$$y(s) = F_f(s)\Phi(s) = K_f\, a h(s)\ \Phi(s) \tag{4.5}$$

where $\Phi(s)$ is the transfer function of the single degree of workpiece structure,

$$\Phi(s) = \frac{y(s)}{F_f(s)} = \frac{\omega_n^2}{k_y\left(s^2 + 2\zeta\omega_n s + \omega_n^2\right)}$$

Substituting $y(s)$ into $h(s)$ yields,

$$h(s) = h_0 + (e^{-sT} - 1)K_f\, a h(s)\Phi(s)$$

and the resulting transfer function between the dynamic and reference chip loads becomes,

$$\frac{h(s)}{h_0(s)} = \frac{1}{1 + (1 - e^{-sT})K_f\, a\Phi(s)} \tag{4.6}$$

The stability of the above close-loop transfer function is determined by the roots (s) of its characteristic equation, i.e.,

$$1 + (1 - e^{-sT)}\, K_f\, a\Phi(s) = 0$$

Let the root of the characteristic equation is $s = \sigma + j\omega_c$. If the real part of the root is positive ($\sigma > 0$), the time domain solution will have an exponential term with positive power (i.e., $e^{+|\sigma|t)}$. The chatter vibrations will grow indefinitely, and the system will be unstable. A negative real root ($\sigma < 0$) will suppress the vibrations with time (i.e., $e^{-|\sigma|t)}$, and the system is stable with chatter vibration-free cutting. When the real part is zero ($s = j\omega_c$), the system is critically stable, and the workpiece oscillates with constant vibration amplitude at chatter frequency ω_c. For critical borderline stability analysis ($s = j\omega_c$), the characteristic function becomes,

$$1 + (1 - e^{-j\omega_c T})K_f a_{lim}\Phi(j\omega_c) = 0 \tag{4.7}$$

where a_{lim} is the maximum axial depth of cut for chatter vibration free machining. The transfer function can be partitioned into real and imaginary parts, i.e., $\Phi(j\omega_c) = G + jH$. Rearranging the characteristic equation with real and complex parts yields,

$$\{1 + K_f a_{lim}[G(1 - \cos\omega_c T) - H\sin\omega_c T]\} + j\{K_f a_{lim}[G\sin\omega_c T + H(1 - \cos\omega_c T)]\} = 0$$

Both real and imaginary parts of the characteristic equation must be zero. If the imaginary part is considered first,

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

$$G\sin\omega_c T + H(1-\cos\omega_c T) = 0$$

and

$$\tan\psi = \frac{H(\omega_c)}{G(\omega_c)} = \frac{\sin\omega_c T}{\cos\omega_c T - 1} \tag{4.8}$$

where ψ is the phase shift of the structure's transfer function. Using the trigonometric identity $\cos\omega_c T = \cos^2(\omega_c T/2) - \sin^2(\omega_c T/2)$ and $\sin\omega_c T = 2\sin(\omega_c T/2)\cos(\omega_c T/2)$,

$$\tan\psi = \frac{\cos(\omega_c T/2)}{-\sin(\omega_c T/2)} = \tan[(\omega_c T)/2 - (3\pi)/2]$$

and

$$\omega_c T = 3\pi + 2\psi, \quad \psi = \tan^{-1}\frac{H}{G} \tag{4.9}$$

The spindle speed (n[*rev/s*]) and the chatter vibration frequency (ω_c) have a relationship which affects the dynamic chip thickness. Let's assume that the chatter vibration frequency is ω_c[*rad/s*] or f_c[*Hz*]. The number of vibration waves left on the surface of the workpiece is

$$f_c[\text{Hz}]\cdot T[\text{sec.}] = \frac{f_c}{n} = k + \frac{\epsilon}{2\pi} \tag{4.10}$$

where k is the integer number of waves and $\epsilon/2\pi$ is the fractional wave generated. The angle ϵ represents the phase difference between the inner and outer modulations. Note that if the spindle and vibration frequencies have an integer ratio, the phase difference between the inner and outer waves on the chip surface will be zero or 2π, hence the chip thickness will be constant albeit the presence of vibrations. In this case, the inner ($y(t)$) and outer ($y(t - T)$) waves are parallel to each other and there will be no chatter vibration. If the phase angle is not zero, the chip thickness changes continuously. Considering k integer number of full vibration cycles and the phase shift,

$$2\pi f_c T = 2k\pi + \epsilon \tag{4.11}$$

where the phase shift between the inner and outer waves is $\epsilon = 3\pi + 2\psi$. The corresponding spindle period (T[*sec*]) and speed (n[*rev/min*]) is found,

$$T = \frac{2k\pi + \epsilon}{2\pi f_c} \rightarrow n = \frac{60}{T} \tag{4.12}$$

The critical axial depth of the cut can be found by equating the real part of the characteristic equation to zero,

$$1 + K_f a_{lim}[G(1-\cos\omega_c T - H\sin\omega_c T] = 0$$

or

$$a_{lim} = \frac{-1}{K_f G[(1-\cos\omega_c T - (H/G)\sin\omega_c T]}$$

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

Substituting $H/G = (\sin \omega_c T)/(\cos \omega_c T - 1)$ and rearranging the above equation yields,

$$a_{lim} = \frac{-1}{2K_f G(\omega_c)} \tag{4.13}$$

Note that since the depth of cut is a physical quantity, the solution is valid only for the negative values of the real part of the transfer function ($G(\omega_c)$). The chatter vibrations may occur at any frequency where $G(\omega_c)$ is negative. If a_{lim} is selected using the minimum value of $G(\omega_c)$, the avoidance of chatter is guaranteed at any spindle speed. The expression indicates that the axial depth of cut is inversely proportional to the flexibility of the structure and cutting constant of the workpiece material. The harder the work material is, the larger the cutting constant K_f will be, thus reducing the chatter vibration-free axial depth of cut. Similarly, flexible machine tool or workpiece structures will also reduce the axial depth of cut or the *productivity.*

The above stability expression was first obtained by Tlusty.[4] Tobias[6] and Merrit[7] presented similar solutions. Tobias presented stability charts indicating chatter vibration-free spindle speeds and axial depth of cuts. Assuming that the transfer function of the structure at the cutting point (Φ) and cutting constant K_f are known or measured, the procedure of plotting the stability lobes can be summarized in the following:

- Select a chatter frequency (ω_c) at the negative real part of the transfer function.
- Calculate the phase angle of the structure at ω_c, Equation (4.8).
- Calculate the critical depth of cut from Equation (4.13).
- Calculate the spindle speed from Equation (4.12) for each stability lobe $k = 0, 1, 2, \ldots$.
- Repeat the procedure by scanning the chatter frequencies around the natural frequency of the structure.

If the structure has multiple degrees of freedom, an oriented transfer function of the system in the direction of chip thickness must be considered for Φ. In that case, the negative real part of the complete transfer function around all dominant modes must be scanned using the same procedure outlined for the orthogonal cutting process.

4.3 Analytical Prediction of Chatter Vibrations in Milling

The rotating cutting force and chip thickness directions, and intermittent cutting periods complicate the application of orthogonal chatter theory to milling operations. The following analytical chatter prediction model was presented by Altintas and Budak,[8,9] and provides practical guidance to machine tool users and designers for optimal process planning of depth of cuts and spindle speeds in milling operations.

4.3.1 Dynamic Milling Model

Milling cutters can be considered to have 2-orthogonal degrees of freedom as shown in Figure 4.2. The cutter is assumed to have N number of teeth with a zero helix angle. The cutting forces excite the structure in the feed (X) and normal (Y) directions, causing dynamic displacements x and y, respectively. The dynamic displacements are carried to rotating tooth number (j) in the radial or chip thickness direction with the coordinate transformation of $v_j = -x \sin \phi_j - y \cos \phi_j$ where ϕ_j is the instantaneous angular immersion of tooth (j) measured clockwise from the normal (Y) axis. If the spindle rotates at an angular speed of Ω (*rad/s*) the immersion angle varies with time as $\phi_j(t) = \Omega t$. The resulting chip thickness consists of static part ($s_t \sin \phi_j$), which is due to rigid body motion of the cutter, and the dynamic component caused by the vibrations of the tool at the present and

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

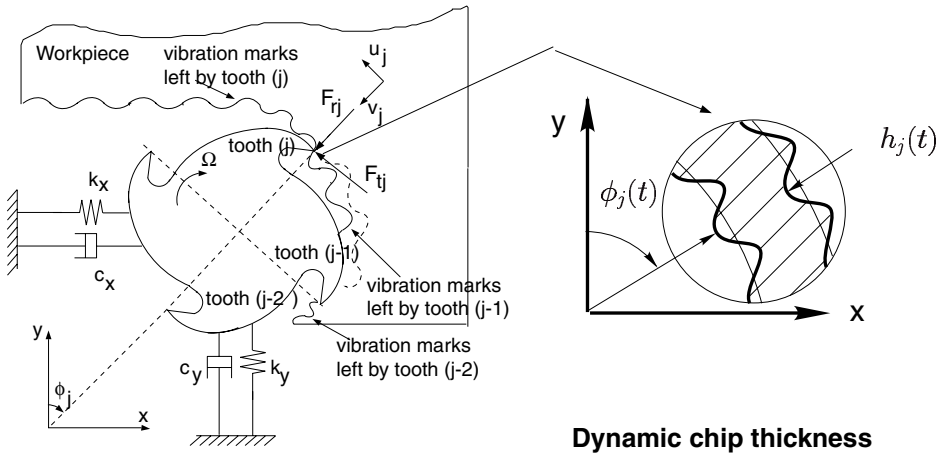


FIGURE 4.2 Mechanism of chatter in milling.

previous tooth periods. Because the chip thickness is measured in the radial direction (v_j), the total chip load can be expressed by,

$$h(\phi_j) = [s_t \sin \phi_j + (v_{j,0} - v_j)]g(\phi_j) \tag{4.14}$$

where s_t is the feed rate per tooth and ($v_{j,0}$, v_j) are the dynamic displacements of the cutter at the previous and present tooth periods, respectively. $g(\phi_j)$ is zero when the tool is out of cut, and unity otherwise

$$\left.\begin{aligned} g(\phi_j) &= 1 \leftarrow \phi_{st} < \phi_j < \phi_{ex} \\ g(\phi_j) &= 0 \leftarrow \phi_j < \phi_{st} \text{ or } \phi_j > \phi_{ex}. \end{aligned}\right\} \tag{4.15}$$

where ϕ_{st}, ϕ_{ex} are start and exit immersion angles of the cutter to and from the cut, respectively. Henceforth, the static component of the chip thickness ($s_t \sin \phi_j$) is dropped from the expressions because it does not contribute to the dynamic chip load regeneration mechanism. Substituting v_j into (4.14) yields,

$$h(\phi_j) = [\Delta x \sin \phi_j + \Delta y \cos \phi_j]g(\phi_j) \tag{4.16}$$

where $\Delta x = x - x_0$, $\Delta y = y - y_0$. (x, y) and (x_0, y_0) represent the dynamic displacements of the cutter structure at the present and previous tooth periods, respectively. The tangential (F_{tj}) and radial (F_{rj}) cutting forces acting on the tooth j is proportional to the axial depth of cut (a) and chip thickness (h),

$$F_{tj} = K_t a h(\phi_j), \quad F_{rj} = K_r F_{tj} \tag{4.17}$$

where cutting coefficients K_t and K_r are constant. Resolving the cutting forces in the x and y directions,

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

$$F_{xj} = -F_{tj}\cos\phi_j - F_{rj}\sin\phi_j$$
$$F_{yj} = +F_{tj}\sin\phi_j - F_{rj}\cos\phi_j \tag{4.18}$$

and summing the cutting forces contributed by all teeth, the total dynamic milling forces acting on the cutter are found as

$$F_x = \sum_{j=0}^{N-1} F_{x_j}(\phi_j) \quad ; \quad F_y = \sum_{j=0}^{N-1} F_{y_j}(\phi_j) \tag{4.19}$$

where $\phi_j = \phi + j\phi_p$, and cutter pitch angle is $\phi_p = 2\pi / N$. Substituting the chip thickness (4.16) and tooth forces (4.7) into (4.18), and rearranging the resulting expressions in matrix form yields,

$$\begin{Bmatrix} F_x \\ F_y \end{Bmatrix} = \frac{1}{2} a K_t \begin{bmatrix} a_{xx} & a_{xy} \\ a_{yx} & a_{yy} \end{bmatrix} \begin{Bmatrix} \Delta x \\ \Delta y \end{Bmatrix} \tag{4.20}$$

where time-varying directional dynamic milling force coefficients are given by

$$a_{xx} = \sum_{j=0}^{N-1} -g_j[\sin 2\phi_j + K_r(1-\cos 2\phi_j)]$$

$$a_{xy} = \sum_{j=0}^{N-1} -g_j[(1+\cos 2\phi_j) + K_r \sin 2\phi_j]$$

$$a_{yx} = \sum_{j=0}^{N-1} g_j[(1-\cos 2\phi_j) - K_r \sin 2\phi_j]$$

$$a_{yy} = \sum_{j=0}^{N-1} g_j[\sin 2\phi_j - K_r(1+\cos 2\phi_j)]$$

Considering that the angular position of the parameters changes with time and angular velocity, Equation (4.20) can be expressed in time domain in a matrix form as[10,11]

$$\{F(t)\} = \frac{1}{2} a K_t [A(t)]\{\Delta(t)\} \tag{4.21}$$

As the cutter rotates, the directional factors vary with time, which is the fundamental difference between milling and operations like turning, where the direction of the force is constant. However, like the milling forces, $[A(t)]$ is periodic at tooth passing frequency $\omega = N\Omega$ or tooth period $T = 2\pi/\omega$, thus can be expanded into Fourier series.

$$[A(t)] = \sum_{r=-\infty}^{\infty} [A_r] e^{ir\omega t}, \qquad [A_r] = \frac{1}{T}\int_0^T [A(t)]\, e^{-ir\omega t} dt \tag{4.22}$$

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

The number of harmonics (r) of the tooth-passing frequency (ω) to be considered for an accurate reconstruction of $[A(t)]$ depends on the immersion conditions and the number of teeth in the cut. If the most simplistic approximation, the average component of the Fourier series expansion, is considered, i.e., $r = 0$,

$$[A_0] = \frac{1}{T}\int_0^T [A(t)dt. \tag{4.23}$$

Because $[A_0]$ is valid only between the entry (ϕ_{st}) and exit (ϕ_{ex}) angles of the cutter (i.e., $g_j(\phi_j) = 1$), and $\phi_j = \Omega\, t$ and $\phi_p = \Omega\, T$, it becomes equal to the average value of $[A(t)]$ at cutter pitch angle $\phi_p = 2\pi / N$.

$$[A(0)] = \frac{1}{\phi_p}\int_{\phi_{st}}^{\phi_{ex}} [A(\phi)]d\phi = \frac{N}{2\pi}\begin{bmatrix} \alpha_{xx} & \alpha_{xy} \\ \alpha_{yx} & \alpha_{yy} \end{bmatrix} \tag{4.24}$$

where the integrated functions are given as

$$\alpha_{xx} = \tfrac{1}{2}\left[\cos 2\phi - 2K_r\phi + K_r \sin 2\phi\right]_{\phi_{st}}^{\phi_{ex}}$$

$$\alpha_{xy} = \tfrac{1}{2}\left[-\sin 2\phi - 2\phi + K_r \cos 2\phi\right]_{\phi_{st}}^{\phi_{ex}}$$

$$\alpha_{yx} = \tfrac{1}{2}\left[-\sin 2\phi + 2\phi + K_r \cos 2\phi\right]_{\phi_{st}}^{\phi_{ex}}$$

$$\alpha_{yy} = \tfrac{1}{2}\left[-\cos 2\phi - 2K_r\phi - K_r \sin 2\phi\right]_{\phi_{st}}^{\phi_{ex}}$$

The average directional factors are dependent on the radial cutting constant (K_r) and the width of cut bound by entry (ϕ_{st}) and exit (ϕ_{ex}) angles. The dynamic milling expression (4.21) is reduced to the following

$$\{F(t)\} = \frac{1}{2} a K_t [A_0]\{\Delta(t)\} \tag{4.25}$$

where $[A_0]$ is a time-invariant but immersion-dependent directional cutting coefficient matrix. Because the average cutting force-per-tooth period is independent of the helix angle, $[A_0]$ is valid for helical end mills as well.

4.3.2 Chatter Stability Lobes

Transfer function matrix ($[\Phi\ (i\omega)]$) identified at the cutter–workpiece contact zone,

$$[\Phi(i\omega)] = \begin{bmatrix} \Phi_{xx}(i\omega) & \Phi_{xy}(i\omega) \\ \Phi_{yx}(i\omega) & \Phi_{yy}(i\omega) \end{bmatrix} \tag{4.26}$$

where $\Phi_{xx}(i\omega)$ and $\Phi_{yy}(i\omega)$ are the direct transfer functions in the x and y directions, and $\Phi_{xy}(i\omega)$ and $\Phi_{yx}(i\omega)$ are the cross-transfer functions. The vibration vectors at the present time (t) and previous tooth period ($t - T$) are defined as,

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

$$\{r\} = \{x(t)\ y(t)\}^T\ ;\{r_0\} = \{x(t-T)\ y(t-T)\}^T.$$

Describing the vibrations at the chatter frequency ω_c in the frequency domain using harmonic functions,

$$\left.\begin{aligned}\{r(i\omega_c)\} &= [\Phi(i\omega)]\{F\}e^{i\omega_c t}\\ \{r_0(i\omega_c)\} &= e^{-i\omega_c t}\{r(i\omega_c)\}\end{aligned}\right\} \tag{4.27}$$

and substituting $\{\Delta\} = \{(x - x_0\}\ (y - y_0)\}^T$ gives,

$$\begin{aligned}\{\Delta(i\omega_c)\} &= \{r(i\omega_c)\} - \{r_0(i\omega_c)\}\\ &= [1 - e^{-i\omega_c T}]e^{i\omega_c t}[\Phi(i\omega_c)]\{F\}\end{aligned}$$

where $\omega_c T$ is the phase delay between the vibrations at successive tooth periods T. Substituting $\{\Phi(i\omega_c)\}$ into the dynamic milling Equation (4.25) gives

$$\{F\}e^{i\omega_c t} = \frac{1}{2}aK_t[1 - e^{-i\omega_c T}][A_0][\Phi(i\omega_c)]\{F\}e^{i\omega_c t}$$

which has a nontrivial solution if its determinant is zero,

$$\det[[I] - \frac{1}{2}K_t a(1 - e^{i\omega_c T})[A_0][\Phi(i\omega_c)]] = 0$$

which is the characteristic equation of the closed-loop dynamic milling system. The notation is further simplified by defining the oriented transfer function matrix as

$$[\Phi_0(i\omega_c)] = \begin{bmatrix}\alpha_{xx}\Phi_{xx}(i\omega_c) + \alpha_{xy}\Phi_{yx}(i\omega_c) & \alpha_{xx}\Phi_{xy}(i\omega_c) + \alpha_{xy}\Phi_{yy}(i\omega_c)\\ \alpha_{yx}\Phi_{xx}(i\omega_c) + \alpha_{yy}\Phi_{yx}(i\omega_c) & \alpha_{yx}\Phi_{xy}(i\omega_c) + \alpha_{yy}\Phi_{yy}(i\omega_c)\end{bmatrix} \tag{4.28}$$

and the eigenvalue of the characteristic equation as

$$\Lambda = -\frac{N}{4\pi}aK_t(1 - e^{-i\omega_c T}). \tag{4.29}$$

The resulting characteristic equation becomes,

$$\det[[I] + \Lambda[\Phi_0(i\omega_c)]] = 0 \tag{4.30}$$

The eigenvalue of the above equation can easily be solved for a given chatter frequency ω_c, static cutting coefficients (K_t, K_r) which can be stored as a material-dependent quantity for any milling cutter geometry, radial immersion (ϕ_{st}, ϕ_{ex}), and transfer function of the structure (4.28). If two orthogonal degrees-of-freedom in feed (X) and normal (Y) directions are considered (i.e., $\Phi_{xy} = \Phi_{yx} = 0.0$), the characteristic equation becomes just a quadratic function

$$a_0\Lambda^2 + a_1\Lambda + 1 = 0 \tag{4.31}$$

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

where

$$a_0 = \Phi_{xx}(i\omega_c)\Phi_{yy}(i\omega_c)(\alpha_{xx}\alpha_{yy} - \alpha_{xy}\alpha_{yx})$$

$$a_1 = \alpha_{xx}\Phi_{xx}(i\omega_c) + \alpha_{yy}\Phi_{yy}(i\omega_c)$$

Then, the eigenvalue Ω is obtained as

$$\Lambda = -\frac{1}{2a_0}(a_1 \pm \sqrt{a_1^2 - 4a_0}). \tag{4.32}$$

As long as the plane of cut (x, y) is considered, the characteristic equation is still a simple quadratic function regardless of the number of modes considered in the machine tool structure. Indeed, the actual transfer function measurements of the machine dynamics can be used at each frequency. Because the transfer functions are complex, the eigenvalue has a real and an imaginary part, $\Lambda = \Lambda_R + i\Lambda_I$. Substituting the eigenvalue and $e^{-i\omega_c T} = \cos\omega_c T - i\sin\omega_c T$ in Equation (4.29) gives the critical axial depth of cut at chatter frequency ω_c,

$$a_{lim} = -\frac{2\pi}{NK_t}\left[\frac{\Lambda_R(1-\cos\omega_c T) + \Lambda_I \sin\omega_c T}{(1-\cos\omega_c T)} + i\frac{\Lambda_I(1-\cos\omega_c T) - \Lambda_R \sin\omega_c T}{(1-\cos\omega_c T)}\right] \tag{4.33}$$

Because a_{lim} is a real number, the imaginary part of the Equation (4.33) must vanish,

$$\Lambda_I(1-\cos\omega_c T) - \Lambda_R \sin\omega_c T = 0 \tag{4.34}$$

By substituting,

$$\kappa = \frac{\Lambda_I}{\Lambda_R} = \frac{\sin\omega_c T}{1-\cos\omega_c T} \tag{4.35}$$

into the real part of the Equation (4.33) (imaginary part vanishes), the final expression for chatter-free axial depth of cut is found as

$$a_{\lim} = -\frac{2\pi\Lambda_R}{NK_t}(1+\kappa^2) \tag{4.36}$$

Therefore, given the chatter frequency (ω_c), the chatter limit in terms of the axial depth of cut can directly be determined from Equation (4.36).

The corresponding spindle speeds are also found in a manner similar to the chatter in orthogonal cutting presented in the previous section.

From Equation 4.35,

$$\kappa = \tan\psi = \frac{\cos(\omega_c T/2)}{\sin(\omega_c T/2)} = \tan[\pi/2 - (\omega_c T/2)] \tag{4.37}$$

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

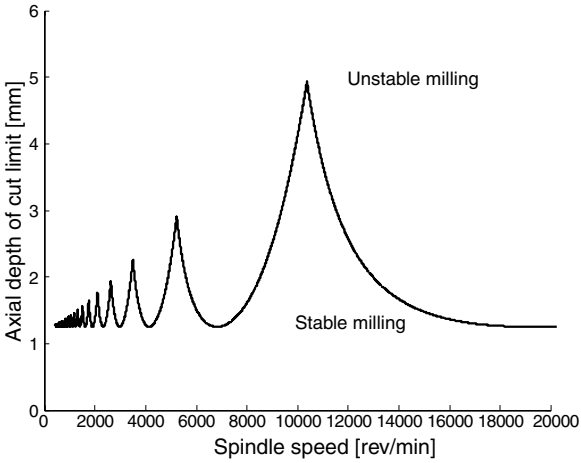


FIGURE 4.3 Stability lobes for a half immersion down milling of Al7075-T6 material with a bullnose cutter having two edges, 31.75 shank diameter and 4.7625-mm corner radius. The feed per tooth was $s_t = 0.050$ *mm*/rev in cutting tests.

and the phase shift of the eigenvalue is $\psi = \tan^{-1}\kappa$, and $\in = \pi - 2\psi$ is the phase shift between inner and outer modulations (present and previous vibration marks). Thus, if k is the integer number of full vibration waves (i.e., lobes) imprinted on the cut arc,

$$\omega_c T = \in + 2k\pi \tag{4.38}$$

Again, care must be taken in calculating the phase shift (ψ) from the real (Λ_R) and imaginary (Λ_I) parts of the eigenvalue. The spindle speed n(*rev/min*) is simply calculated by finding the tooth-passing period T(s),

$$T = \frac{1}{\omega_c}(\in + 2k\pi) \rightarrow n = \frac{60}{NT} \tag{4.39}$$

In summary, the transfer functions of the machine tool system are identified, and the dynamic cutting coefficients are evaluated from the derived Equation (4.24) for a specified cutter, workpiece material, and radial immersion of the cut. Then the stability lobes are calculated as follows:[8]

- Select a chatter frequency from transfer functions around a dominant mode.
- Solve the eigenvalue Equation (4.31).
- Calculate the critical depth of cut from Equation (4.36).
- Calculate the spindle speed from Equation (4.39) for each stability lobe $k = 0, 1, 2, \ldots$.
- Repeat the procedure by scanning the chatter frequencies around all dominant modes of the structure evident on the transfer functions.

A sample stability lobe for a vertical machining center milling Aluminum 7075 alloy with a four-fluted helical end mill is shown in Figure 4.3. The measured transfer function parameters of the machine at the tool tip are given as follows: $\omega_{nx} = \{452.8, 1448\}H\ z$; $\zeta_x = \{0.12, 0.017\}$, $k_x = \{124.7E + 6, (-)\ 6595.6E + 6\}N/m$; $\omega_{ny} = \{516, 1407\}H\ z$; $\zeta_x = \{0.024, 0.0324\}$, $k_y = \{(-)\ 2.7916E + 10, 3.3659E + 9\}N/m$ in the feed (x) and normal (y) directions, respectively. The stability lobes are predicted analytically with the theory given here, as well as using a time domain numerical solution which takes a considerable amount of computation time. The analytical method agrees well with the numerical solutions. The machine tool exhibits severe chatter vibrations when the

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

spindle speed is set to 9500 rev/min. The cutting force amplitudes are large, and the chatter occurs at 1448 Hz, which is the second bending mode of the spindle. When the speed and, therefore, productivity are increased to 14,000 rev/min, the chatter disappears and the force is dominated by the regular tooth-passing frequency of 467 Hz. The finish surface becomes acceptable, and the cutting force magnitude drops at the chatter vibration-free spindle speed and depth of cut.

References

1. F. Koenigsberger and J. Tlusty, *Machine Tool Structures, Vol. I: Stability against Chatter,* Pergamon Press, Oxford, 1967.
2. Y. Koren, *Computer Control of Manufacturing Systems,* McGraw Hill, New York, 1983.
3. Y. Altintas, *Manufacturing Automation: Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design,* Cambridge University Press, Cambridge, 2000.
4. J. Tlusty and M. Polacek, The stability of machine tools against self-excited vibrations in machining, *International Research in Production Engineering,* ASME, 465–474, 1963.
5. S.A. Tobias and W. Fishwick, *Theory of Regenerative Chatter,* The Engineer, London, 1958.
6. S.A Tobias, *Machine Tool Vibrations,* Blackie and Sons Ltd., London, 1965.
7. H.E. Merrit, Theory of self-excited machine tool chatter, *Transactions of ASME Journal of Engineering for Industry,* 87, 447–454, 1965.
8. Y. Altintas and E. Budak, Analytical prediction of stability lobes in milling, *Annals of the CIRP,* 44(1), 357–362, 1995.
9. E. Budak and Y. Altintas, Analytical prediction of chatter stability conditions for multi-degree of systems in milling. Part i: Modelling, Part ii: Applications, *Transactions of ASME Journal of Dynamic Systems, Measurement and Control,* 120, 22–36, 1998.
10. R.E. Hohn, R. Sridhar, and G.W. Long, A stability algorithm for a special case of the milling process, *Transactions of ASME Journal of Engineering for Industry,* 325–329, May 1968.
11. I. Minis, T. Yanushevsky, R. Tembo, and R. Hocken, Analysis of linear and nonlinear chatter in milling, *Annals of the CIRP,* 39, 459–462, 1990.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

5

Machine Tool Monitoring and Control

5.1 Introduction
5.2 Process Monitoring
Tool Wear Estimation • Tool Breakage Detection • Chatter Detection
5.3 Process Control
Control for Process Regulation • Control for Process Optimization
5.4 Conclusion

Kourosh Danai
University of Massachusetts, Amherst

5.1 Introduction

Machine tool monitoring and control are essential for automated manufacturing. Monitoring is necessary for detection of a process anomaly to prevent machine damage by stopping the process, or to remove the anomaly by adjusting the process inputs (feeds and speeds). A process anomaly may be gradual such as tool/wheel wear, may be abrupt such as tool breakage, or preventable such as excessive vibration/chatter. Knowledge of tool wear is necessary for scheduling tool changes; detection of tool breakage is important for saving the workpiece and/or the machine; and identifying chatter is necessary for triggering corrective action. One difficulty in machine tool monitoring stems from the limited sensing capability afforded by the harsh manufacturing environment. Sensors can seldom be placed at the point of interest, and when located at remote locations they do not provide the clarity of measurement necessary for reliable monitoring. This limited sensing capability is often compensated for by using multiple sensors to enhance reliability. Another difficulty in machine tool monitoring is the absence of accurate analytical models to account for changes in the measured variables by variations in the cutting conditions. Such changes are often attributed to process anomalies by the monitoring system, which result in false alarms.

Machine tool control is motivated by two objectives: (1) process regulation, so as to preempt excessive forces, correct a process anomaly, or reduce contouring errors; and (2) process optimization, for the purpose of improving the quality of the part or reducing operation time based on feedback from the process.

The aim of this chapter is to provide a conceptual survey of machine tool monitoring and control. As such, no attempt has been made to acknowledge all the research in this area, and the citations are included mainly to provide representative examples of various approaches.

5.2 Process Monitoring

Process monitoring is generally performed through the analysis of process measurements. For this purpose, a process variable or a set of variables (e.g., force, power, acoustic emission, feed motor

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

current) is measured and processed on-line to be compared against its expected value. Any deviation from this expected value is attributed to a process anomaly. Expected values of measurements are either determined according to an analytical model of the process[1] or established empirically.[2] The advantage of using analytical models is that they account for changes in the machine inputs such as feeds and speeds. The disadvantage of analytical models is that they are often not accurate and need to be calibrated for the process. Establishing the expected values of measurements empirically is simpler and more straightforward. However, the empirical values are only suitable for particular operations and cannot be extrapolated to others. To provide a representative sample of approaches used in this area, tool wear estimation, tool breakage detection, and chatter identification are discussed as the most investigated topics in machine tool monitoring.

5.2.1 Tool Wear Estimation

Flank wear directly influences the size and quality of the surface.[3] Flank wear can affect fatigue endurance limit by affecting surface finish, lubrication retention capability by changing the distribution of heights and slopes of the surface,[4] and other tribological aspects[5,6] by affecting the topography of the machined surface. Therefore, information about the state of flank wear is sought to plan tool changes in order to avoid scrapping or manipulating the feed and cutting speed in-process to control tool life.[7]

Methods used for flank wear estimation can be classified as either direct or indirect.[8] Direct methods measure flank wear either in terms of material loss from the tool[9] or by observing the worn surface using optical methods.[10] Direct methods are generally more reliable, although they are not convenient for in-process use in a harsh manufacturing environment. Indirect methods, on the other hand, estimate the flank wear by relating it to a measured variable such as the change in size of the workpiece,[11] cutting force,[12] temperature,[13] vibration,[14] or acoustic emissions.[15] The ideal measured variable in the indirect method is one that is insensitive to process inputs. For example, noncontact methods have been recently developed for surface roughness measurement,[16,17] which will undoubtedly have an impact on on-line estimation of tool wear.

Among the measurements used for indirect flank wear estimation, acoustic emission (AE) and the cutting force have been the most popular due to their sensitivity to tool wear and reliability of measurement. The cutting force generally increases with flank wear due to an increase in the contact area of the wear land with the workpiece. Zorev[18] and De Filippi and Ippolito[19] were among the first who demonstrated the direct effect of flank wear on the cutting force, which motivated separation of the cutting force signal into two components, one associated with the unworn tool and the other associated with tool wear. The unworn tool component is usually estimated at the beginning of the cut with a new tool, and then subtracted from the measured force to estimate the wear affected component. This method can provide relatively accurate estimates of flank wear so long as the cutting variables (feed, speed, and depth of cut) remain unchanged. However, when the cutting variables change, due to such factors as the geometric requirements of the part or manipulation of the operating parameters, the identification of the wear affected component becomes difficult. In such cases, either the effect of the manipulated cutting variable on the cutting force is estimated by a model[1] and separated to identify the wear affected component,[10,20] or the wear affected component is estimated from small cutting segments where the cutting variables remain unchanged.[21] In either case, recursive parameter estimation techniques, which require persistent excitation of the cutting force to guarantee parameter convergence, are used for identification purposes. The requirement for persistent excitation is relaxed,[12] by measuring the cutting force during the transient at the beginning of the cut when the tool engages the workpiece. During this transient, the sharp tool chip formation component, which is proportional to the cross-sectional area of the cut normal to the main cutting velocity, takes a wide range of values, from zero to the steady-state value (product of the feed and depth of cut). The method uses the variations of the cross-sectional area of the cut during this short time interval when flank wear is essentially constant

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

to tune the model and estimate its parameters. It has been shown in laboratory experiments that the residual force components in the axial and tangential directions increase linearly with the wear land width, which can be used to estimate flank wear.[12]

Similar to the cutting force signal, acoustic emission has been studied extensively for flank wear estimation, where various statistical properties of the AE signal have been shown to correlate with flank wear.[15] To define more clearly the effect of flank wear, statistical pattern classification of AE signal in frequency domain has been utilized as well.[22,23]

Despite the considerable effort toward estimation of flank wear from a single variable, single sensor measurements do not seem to be robust to varying cutting conditions. This has motivated integration of multiple measurements through artificial neural networks.[24,25] Artificial neural networks have the ability to represent patterns of fault signatures by complex decision regions without reliance on the probabilistic structure of the patterns. Thus, they are powerful tools for fault detection/diagnosis. Generally, a neural network is trained to identify the tool wear pattern by supervised learning from samples of measurements taken at various levels of tool wear. Therefore, the ability of neural networks to form reliable wear patterns depends not only on their topology, but the extent of their training. In cases such as machining where adequate data are not available to select the topology of the network or to provide the tool wear patterns for a wide range of cutting conditions and material/tool combinations, these networks are not practical.

A remedy to supervised learning is the application of unsupervised neural networks[26] that can form pattern clusters of data without a known target for each input vector. These networks use prototype vectors to characterize each category, and then classify input vectors within each category according to their similarity to these prototype vectors. While there is a need to provide data from each category to these networks in order to form the prototype vectors, the demand for training is considerably less. Therefore, unsupervised networks have better potential for on-line utility in machine tool monitoring. A comprehensive demonstration of unsupervised neural networks in tool failure monitoring is provided by Li et al.,[27] who applied an array of adaptive resonance theory (ART2) networks[28] to detect tool wear, tool breakage, and chatter using vibration and AE measurements.

5.2.2 Tool Breakage Detection

Fracture is the dominant mode of failure for more than one quarter of all advanced tooling material. Therefore, on-line detection of tool breakages is crucial to the realization of fully automated machining. Ideally, a tool breakage detection system must be able to detect failures rapidly to prevent damage to the workpiece, and must be reliable to eliminate unnecessary downtime due to false alarms.

Several measurements have been reported as good indicators of tool breakage.[29] Among these, the cutting force,[30] acoustic emission,[31,32] spindle motor current,[33] feed motor current,[34] and machine tool vibration[35,36] have been investigated extensively for their sensitivity to tool breakage. In general, to utilize a measurement for tool breakage detection, two requirements need to be satisfied. First, the measurement must reflect tool breakage under diverse cutting conditions (e.g., variable speeds, feeds, coolant on/off, workpiece material). Second, the effect of tool breakage on the measurement (tool breakage signature) must be uniquely distinguishable, so that other process irregularities such as hard spots will not be confused with tool breakage. The tool breakage signature is commonly in the form of an abrupt change, in excess of a threshold value. Despite considerable effort,[37,38] reliable signatures of tool breakage that are robust to diverse cutting conditions have not yet been found from individual measurements.

To extract more information from individual measurements to improve the reliability of tool breakage signatures, pattern classification techniques have been utilized. One of the earliest efforts was by Sata et al.[39] who related features of the cutting force spectrum such as its total power, the power in the very low frequency range, and the power at the highest spectrum peak and its frequency to chip formation, chatter, and a built-up edge. It was shown that the cutting force measurement

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

alone provides sufficient information for unique identification of the above phenomena. Another important work in this category is by Kannatey-Asibu and Emel[22] who applied statistical pattern classification to identify chip formation, tool breakage, and chip noise from acoustic emission measurements. They reported a success rate of 90% for tool breakage detection. The only drawback to spectrum-based tool breakage detection is the computational burden associated with obtaining the spectrum, which often precludes its on-line application.

The alternative to single-sensor-based pattern classification is the multi-sensor approach using artificial neural networks for establishing the breakage patterns.[24] However, as already mentioned for tool wear estimation, the utility of neural networks for tool breakage detection is limited by their demand for expensive training. A pattern classifier that requires less training than artificial neural networks is the multi-valued influence matrix (MVIM) method[40] which has a fixed structure and has been shown to provide robust detection of tool breakages in turning with limited training.[41]

Unsupervised neural networks have also been proposed for tool breakage detection in machining.[42] The two predominant methods of unsupervised learning presently available for neural networks are Kohonen's feature mapping and adaptive resonance theory (ART2).[28] Kohonen's method of feature mapping establishes the decision regions for normal and abnormal categories through prototype vectors that represent the centers of measurement clusters belonging to these categories. Classification is based on the Euclidean distance between the measurements and each of the prototype vectors. While Kohonen's method forms the prototype vectors far enough from each other to cope with variations in the tool breakage signature, it requires one or more sets of measurements at tool breakage to establish the prototype vector for the abnormal category. The other method of unsupervised learning, the adaptive resonance theory (ART2), classifies the measurements as normal unless they are sufficiently different. When applied to tool breakage detection, it does not require any samples of measurements to be taken at tool breakage. ART2, however, may not cope effectively with varying levels of noise associated with different sensors, and may classify multiples of a prototype within the same category, so it may produce misclassification. A hybrid of the above pattern classifiers is the single category-based classifier (SCBC)[43] that performs detection by comparing each set of measurements against their corresponding prototype values for their normal category and detects tool breakage when the measurements are sufficiently different from their normal prototypes. Another variant of ART2 applied to tool breakage detection is a network consisting of an array of ART2 networks, each classifying the pattern associated with an individual sensor.[27]

5.2.3 Chatter Detection

Chatter is the self-excited vibration of the machine tool that reflects the instability of the cutting process. Chatter is often a serious limitation to achieving higher rates of removal, as it adversely affects the surface finish, reduces dimensional accuracy, and may damage the tool and machine. Therefore, machine tool chatter needs to be detected rapidly and corrected before it damages the workpiece, tool, or the machine.

Several variables have been studied for detection of chatter. These include the cutting force signal, displacement or acceleration of a point in the vicinity of the tool–workpiece interface, or the sound emitted from the machine. Delio et al.[44] claim that sensor placement and the frequency response limitations of the transducer are the two major difficulties in detection of chatter. They also claim that sound provides the most reliable and robust signature for chatter. While chatter has been investigated extensively, most of the efforts have been directed toward prediction of chatter rather than its detection. The approaches used for chatter detection mirror those employed for tool breakage detection, except that analysis is performed primarily in frequency domain where the effect of vibration is most pronounced.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

5.3 Process Control

The advent of open-architecture control provides a natural framework for implementation of control systems in machine tools.[45] Machine tool control is generally performed at two levels: (1) servo-control to execute the command motion dictated by interpolators for following a prespecified contour, or (2) supervisory control to continually adjust the process variables for the purpose of either regulating the process against disturbances/detected anomalies, or optimizing performance.[46] Process regulation is often incorporated as the next step to process monitoring, whereby the controller attempts to correct, if possible, the detected anomaly. Process optimization, on the other hand, is implemented to enhance productivity based on an assessment of process and part quality constraints.

5.3.1 Control for Process Regulation

Control for process regulation has been attempted for one of the following reasons: maintaining constant power or force, safeguarding against chatter, or correcting machine tool errors. The most regulated process variable in machining has been the cutting force, mainly for its ease of measurement on-line, and its reflection of process anomalies such as tool breakage and chatter. While there have been differences in format and the underlying models used, most of the controllers designed for force regulation have used a dynamic model of the cutting force with respect to the manipulated variable (i.e., feed or speed) and have employed parameter estimation to adapt the model to changing process conditions.[47-53] Within this category, Furness et al.[54] regulated the torque in drilling to avoid possible chipping of the drill tips, stall of the spindle motor, thermal softening of the tool, or torsional failure of the drill.

Among the first to design a controller for elimination of chatter were Nachtigal and Cook[55] who used the cutting force signal as feedback to control the position of the tool for increased stability. They designed their controller on a fixed model of the machine tool–workpiece dynamics. As a next step and to account for parameter uncertainty in that model, Mitchell and Harrison[56] integrated an observer in their control system to estimate the cutting tool motion on-line for feedback to the control system. Active control of chatter is, by and large, an identification problem, because once the presence of chatter is detected, the solution seems to be straightforward.[44,57]

Another active area of research in process regulation is error correction. The accuracy of a machined part is generally attributed to geometric and kinematic errors of the machine spindle, thermal effects, and static and dynamic loading of the drives.[58] Therefore, considerable effort has been directed toward error compensation by modifying the tool position. Two fundamental approaches have been used for reducing contouring errors:[46] (1) by reducing the tracking error of individual axes, and (2) by reducing contour error which is defined as the error between the actual and desired tool path. As in force-regulation problems, a common approach used in many of these systems is utilization of parameter estimation to update the servo-models in the presence of variable loading and friction (e.g., see Tsao and Tomizuka[59]). The literature on tool error compensation is quite extensive and is not surveyed here in the interest of space. Interested readers are referred to Koren[46] or Tung et al.[60] for specific examples and an overview of the research in this area.

5.3.2 Control for Process Optimization

The adaptation of process variables for the purpose of enhancing process efficiency is addressed within the area of control for process optimization.[1] Process efficiency is generally defined in terms of reduced* production cost or cycle time. Under deterministic conditions (no modeling uncertainty

*Control for process optimization has also been referred to as adaptive control optimization (ACO) in the manufacturing engineering literature.[46]

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

and noise), there would be no need for a controller, as the optimal process inputs (feeds and speeds) could be determined by nonlinear programming.[61] In view of the highly complex nature of machining processes, however, the process inputs need to be changed iteratively in response to measurements of process and part quality constraints. This interactive approach to process optimization is adopted to enable the control system to maintain constraint satisfaction despite modeling uncertainty arising from (1) the diversity of machining conditions due to variations in material properties, tool/wheel type, and lubrication, (2) the stochastic nature of these processes caused by material inhomogeneity, workpiece misalignment, and measurement noise, and (3) process time variability due to tool wear.

The first attempt at control for process optimization was the Bendix system,[62] which was designed to continually maximize the machining removal rate through changes in both the feedrate and spindle speed in response to feedback measurements of cutting torque, tool temperature, and machine vibration. The Bendix System, however, was limited in applicability due to the need to estimate tool wear based on an accurate model. A subsequent advancement in control for process optimization was the Optimal Locus Approach,[63,64] which made it possible to forego estimation of tool wear. In this approach, the locus of the optimal points associated with various levels of tool wear is computed, and the optimal point is sought where process and part quality constraints become tight. The Optimal Locus Approach can avoid estimation of tool wear by using the tightness of constraints as the measure for optimality, but it still needs to rely on the accuracy of the process model for computing the optimal locus and determining *a priori* which constraints are tight at the optimum. Because the success of this approach depends on the premise that modeling uncertainty will have negligible effect on the accuracy of the optimal locus, it will produce suboptimal results when this premise is violated. A similar approach in drilling, but with several more constraints, was demonstrated by Furness et al.[65] by locating the feasible region of the process according to the pair of constraints active during each of the three drilling phases. In this application, the constraints were considered to be stationary, due to the absence of tool wear in short-duration drilling cycles.

One approach to coping with modeling uncertainty in process optimization is to calibrate (e.g., by parameter estimation) the closed-form solution of the optimal process inputs. This approach has been implemented in cylindrical plunge grinding where each cycle is moved closer to its minimum time based on a closed-form solution of the optimization problem according to a monotonicity analysis.[66] In this method, parameter estimation is used to cope with modelling uncertainty and process variability by continually updating the estimated optimal conditions using parameters estimated from the preceding grinding cycle. The basic requirement for this system is the availability of a relatively accurate model of the process that can be updated using parameter estimation. Such accurate modeling is possible for a few machining processes, but its extension to less-understood processes is difficult.

Another approach that uses an iterative strategy to process optimization but does not require accurate process models is the method of Recursive Constraint Bounding (RCB).[67] Like the Optimal Locus Approach, RCB assesses optimality from the tightness in the constraints using measurements of process and part quality after each workpiece has been finished (cycle). It also uses the model of the process to find the optimal point. However, unlike the Optimal Locus Approach, RCB assumes the model to be uncertain when determining which constraints are to be tight at the optimum and selecting the machine settings for each process cycle. It obtains the machine settings by solving a customized nonlinear programming (NLP) problem, and allows for uncertainty by incorporating conservatism into the NLP problem. This conservatism is tailored according to the severity of modeling uncertainty associated with each constraint. The repeated minimization of the objective function with a progressively less conservative model has been shown to lead to bound constraints and optimal machine settings.[68]

Empirical modeling using neural networks has also been proposed for coping with modeling uncertainty in process optimization.[69,70] In one case, separate neural networks are used to represent tool wear and the process, respectively, as a function of process variables (i.e., feed and speed),

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

and the optimal point of the process is determined according to the neural network model and the estimate of tool wear.[69] In another approach, an iterative method to process optimization is adopted by using a neural network trained as an inverse process model to provide increasingly more optimal process variables.[70] One of the inputs to this neural network is an estimate of a cost function obtained from measurements of cutting force and vibration. Neural network modeling is appealing from the point of view of coping with process uncertainty; however, it has limited utility in manufacturing due to the expense associated with obtaining training data.

5.4 Conclusion

Machine tool monitoring and control provide the bridge between machining research and the production line. Nevertheless, despite years of research and the multitude of success stories in the laboratory, only a small amount of this technology has been transferred to production. It may be argued that the slowness in technology transfer is due to the complexity of machining processes and their incompatibility with the sensing technology. This is supported by the fact that most of the monitoring systems developed are specific to isolated problems, and cannot be integrated with other solutions to provide an effective monitoring system for all the process anomalies of concern. Similarly, it may be argued that most control systems developed in the laboratory use impractical or expensive transducers that are not suitable for the harsh production environment.

While complexity and sensing limitations are important impediments to technology transfer in monitoring, they are minor compared to the cultural barrier imposed by the stringent manufacturing environment. For implementation in production, monitoring and control systems need to be either retrofitted to the existing machine tools or incorporated into new machine tools. The first option will almost never happen because the savings from these systems rarely justify the loss from production downtime. The second option, while more plausible, has not broadly occurred either, mainly due to the cost competitiveness of the machine tool market. Three requirements need to be satisfied for inclusion of monitoring and control in machine tools: (1) the underlying sensors need to be nonintrusive and inexpensive, (2) the monitoring system needs to be comprehensive to detect every process anomaly possible in operation, and (3) both monitoring and control need to be perfectly reliable and robust to process variations. It is basically impossible to satisfy the above conditions, particularly the third one.

A compromise position is to incorporate monitoring and control for specific operations, based on the sensing capability already available on the machine tool. The presence of open-architecture control systems will be a significant boost to this solution, mainly due to the versatility these systems offer in software development and trouble shooting.

References

1. Danai, K. and Ulsoy, A. G., 1987, A dynamic state model for on-line tool wear estimation in turning, *ASME Journal of Engineering for Industry*, 109, 4, 396–399.
2. Du, R., Elbestawi, M. A., and Wu, S. M., 1995, Automated monitoring of manufacturing processes, Part 1: Monitoring methods, and Part 2: Applications, *ASME Journal of Engineering for Industry,* 117, 121–132.
3. Jetly, S., 1984, Measuring cutting tool wear on-line: some practical considerations, *Manufacturing Engineering*, July, 55–60.
4. Whitehouse, D. J., 1978, Surfaces — a link between manufacture and function, *Proceedings of the Institution of Mechanical Engineers*, 179–188.
5. Tonder, K., 1987, Effects of skew unidirectional striated roughness on hydrodynamic lubrication, *Wear*, 115, 19.
6. Wilson, W. R. D. and Sheu, S., 1988, Influence of surface topography on viscoplastic asperity lubrication, *Wear*, 124, 311.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

7. Koren, Y. and Ulsoy, A. G., 1989, Adaptive control in machining, in *Metals Handbook*, ASM International, Cleveland, Ohio.
8. Cook, N. H., 1980, Tool wear sensors, *Wear*, 62, 49–57.
9. Cook, N. H. and Subramanian, K., 1978, Micro-isotope tool wear sensor, *CIRP Annals*, 27, 1, 73–78.
10. Park, J.-J. and Ulsoy, A. G., 1993, On-line flank wear estimation using an adaptive observer and computer vision, Part 1: Theory, Part 2: Experiment, *ASME Journal of Engineering for Industry,* 115, 30–43.
11. El Gomayel, J. I. and Bregger, K. D., 1986, On-line tool wear sensing for turning operations, *ASME Journal of Engineering for Industry*, 108, 44–47.
12. Nair, R., Danai, K., and Malkin, S., 1992, Turning process identification through force transients, *ASME Journal of Engineering for Industry*, 114, 1, 1–7.
13. Groover, M. P., Karpovich R. J., and Levy, E. K., 1977, A study of the relationship between remote thermocouple temperature and tool wear in machining, *International Journal of Product Research*, 25, 2, 129–141.
14. Martin, P., Mutels B., and Draiper, J. P., 1975, Influence of lathe tool wear on the vibrations sustained in cutting, 16th International Machine Tool Design and Research Conference.
15. Kannatey-Asibu, Jr., E. and Dornfeld, D. A., 1982, A study of tool wear in metal cutting using statistical analysis of acoustic emission, *Wear*, 76, 2, 247–261.
16. Coker, S. A., Oh, S. J., and Shin, Y. C., In-process monitoring of surface roughness utilizing ultrasound, *ASME Journal for Manufacturing Scientists and Engineers,* 120, 197–200.
17. Bradley, C., Bohlmann, J., and Kurada, S., 1998, A fiber optic sensor for surface roughness measurement, *ASME Journal for Manufacturing Scientists and Engineers,* 120, 359–367.
18. Zorev, N. N., 1966, Mechanics of contact on the clearance surface, in *Metal Cutting Mechanics*, Shaw, M. C. (Ed.), 129–180, Pergamon Press, Oxford, England.
19. De Filippi, A. and Ippolito, R., 1969, Adaptive control in turning: cutting forces and tool wear relationships for P10, P20, P30 carbides, *CIRP Annals*, 17, 377–379.
20. Danai, K. and Ulsoy, A. G., 1987, An adaptive observer for on-line tool wear estimation in turning, Part I: Theory, Part II: Results, *Mechanical Systems and Signal Processing*, 1, 2, 211–240.
21. Koren, Y., Ko, T., Ulsoy, A. G., and Danai, K., 1991, Flank wear estimation under varying cutting conditions, *ASME Journal of Dynamic Systems, Measurments, and Control*, 113, 2, 300–307.
22. Kannatey-Asibu, E. and Emel, E., 1987, Linear discriminant function analysis of acoustic emission signals for cutting tool monitoring, *Mechanical Systems and Signal Processing*, 4, 333–347.
23. Houshmand, A. A. and Kannatey-Asibu, E., 1989, Statistical process control of acoustic emission for cutting tool monitoring, *Mechanical Systems and Signal Processing*, 3, 4, 405–424.
24. Rangwala, S. and Dornfeld, D., 1990, Sensor integration using neural networks for intelligent tool condition monitoring, *ASME Journal of Engineering for Industry*, 112, 219–228.
25. Govekar, E. and Grabec, I., 1994, Self-organizing neural network application to drill wear classification, *ASME Journal of Engineering for Industry*, 116, 233–238.
26. Leem, C. S., Dornfeld, D. A., and Dreyfus, S. E., 1995, A customized neural network for sensor fusion in on-line monitoring for cutting tool wear, *ASME Journal of Engineering for Industry*, 117, 152–159.
27. Li, X. Q., Wong, Y. S., and Nee, A. Y. C., 1998, A comprehensive idenitification of tool failure and chatter using a parallel multi-Art2 neural network, *ASME Journal for Manufacturing Scientists and Engineers,* 120, 433–442.
28. Hertz, J., Krogh, A., and Palmer, R. G., Eds., 1991, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA.
29. Tlusty, J. and Andrews, G. C., 1983, A critical review of sensors for unmanned machining, *Annals of the CIRP*, 32, 2, 563–572.
30. Altintas, Y. and Yellowley, I., 1987, In-process detection of tool failure in milling using cutting force models, in *Sensors for Manufacturing*, ASME, New York, 1–16.
31. Moriwaki, T., 1980, Detection for tool fracture by acoustic emission measurement, *Annals of the CIRP*, 29, 1, 35–40.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

32. Lan, M. S. and Dornfeld, D. A., 1984, In-process tool fracture detection, *ASME Journal of Engineering Materials and Technology*, 106, April, 111–118.
33. Matsushima, K., Bertok, P., and Sata, T., 1982, In-process detection of tool breakage by monitoring the spindle motor current of a machine tool, in *Measurement and Control for Batch Manufacturing*, ASME, New York, 145–154.
34. Altintas, Y., 1997, Prediction of cutting forces and tool breakage in milling from feed drive current measurements, *ASME Journal for Manufacturing Scientists and Engineers,* 119, 386–392.
35. Grieshaber, D., Ramalingam, R., and Frohrib, D., 1987, On real-time tool fracture in milling, *Proceedings of the 15th NAMRC*, May, 477–484.
36. Hayashi, S. R., Thomas, C. E., and Wildes, D. G., 1988, Tool break detection by monitoring ultrasonic vibrations, *Annals of the CIRP*, 37, 1, 61–64.
37. Lan, M. and Naerheim, Y., 1986, In-process detection of tool breakage in milling, *ASME Journal of Engineering for Industry*, 108, August, 191–197.
38. Altintas, Y., Yellowley, I., and Tlusty, J., 1988, The detection of tool breakage in milling operations, *ASME Journal of Engineering for Industry*, 110, 3, 271–277.
39. Sata, T., Matsushima, K., Nagakura, T., and Kono, E., 1973, Learning and recognition of the cutting states by the spectrum analysis, *Annals of the CIRP*, 22, 41–42.
40. Danai, K. and Chin, H., 1991, Fault diagnosis with process uncertainty, *ASME Journal of Dynamic Systems, Measurement and Control*, 113, 3, 339–343.
41. Colgan, J., Chin, H., Danai, K., and Hayashi, S., 1994, Tool breakage detection in turning: a multi-sensor method, *ASME Journal of Engineering for Industry*, 116, 1, 117–123.
42. Tansel, I. N. and McLaughlin, C., 1991, On-line monitoring of tool breakage with unsupervised neural networks, *Transactions of NAMRC*, SME, 364–370.
43. Jammu, V. B. and Danai, K., 1993, Unsupervised neural network for tool breakage detection in turning, *Annals of the CIRP*, 42, 1, 67–70.
44. Delio, T., Tlusty, J., and Smith, S., 1992, Use of audio signals for chatter detection and control, *ASME Journal for Manufacturing Scientists and Engineers,* 119, 146–157.
45. Schofield, S. and Wright, P., 1998, Open architecture controllers for machine tools, Part 1: Design principles, *ASME Journal for Manufacturing Scientists and Engineers,* 120, 417–424.
46. Koren, Y., 1997, Control of machine tools, *ASME Journal for Manufacturing Scientists and Engineers,* 119, 749–755.
47. Masory, O. and Koren, Y., 1985, Stability analysis of a constant force adaptive control system for turning, *ASME Journal of Engineering for Industry*, 107, 1, 295–300.
48. Daneshmend, L. K. and Pak, H. A., 1986, Model reference adaptive control of feed force in turning, *ASME Journal of Dynamic Systems, Measurement, and Control*, 108, 1, 215–222.
49. Lauderbaugh, L. K. and Ulsoy, A. G., 1988, Dynamic modeling for control of the milling process, *ASME Journal of Engineering for Industry*, 110, 4, 367–375.
50. Tomizuka, M. and Zhang, S., 1988, Modeling and conventional adaptive PI control of a lathe cutting process, *ASME Journal of Dynamic Systems, Measurement, and Control*, 110, December, 350–354.
51. Rober, S. J. and Shin, Y. C., 1996, Control of cutting force for end milling processes using an extended model reference adaptive control scheme, *ASME Journal for Manufacturing Scientists and Engineers,* 118, 339–347.
52. Hsu, P.-L. and Fann, W.-R., 1996, Fuzzy adaptive control of machining processes with a self-learning algorithm, *ASME Journal for Manufacturing Scientists and Engineers,* 118, 522–530.
53. Liang S. Y. and Perry, S. A., 1994, In-Process Compensation for Milling Cutter Runout via Chip Load Manipulation, *ASME Journal of Engineering for Industry,* 116, 153–160.
54. Furness, R. J., Ulsoy, A. G., and Wu, C. L., 1996, Feed, speed, and torque controllers for drilling, *ASME Journal for Manufacturing Scientists and Engineers,* 118, 2–9.
55. Nachtigal, C. L. and Cook, N. H., 1970, Active control of machine-tool chatter, *ASME Journal of Basic Engineering*, 92, 2, 238–244.
56. Mitchell, E. E. and Harrison, E., 1977, Design of a hardware observer for active machine tool control, *ASME Journal of Dynamic Systems, Measurement, and Control*, 99, 227–232.

© 2002 by CRC Press LLC
www.TechnicalBooksPDF.com

57. Subramanian, T. L., DeVries, M. F., and Wu, S. M., 1976, An investigation of computer control of machining chatter, *ASME Journal of Engineering for Industry*, 98, 1209–1214.
58. Li, C. J. and Li, S. Y., 1992, On-line roundness error compensation via P-integrator learning control, *ASME Journal of Engineering for Industry,* 114, 476–480.
59. Tsao, T.-C. and Tomizuka, M., 1987, Adaptive zero phase error tracking algorithm for digital control, *ASME Journal Dynamic Systems, Measurement, and Control*, 109, 349–354.
60. Tung, E. D., Tomizuka, M., and Urushisaki, Y., 1996, High-speed end milling using a feedforward control architecture, *ASME Journal for Manufacturing Scientists and Engineers,* 118, 178–187.
61. Ermer, D. S., 1997, A century of optimizing machining operations, *ASME Journal for Manufacturing Scientists and Engineers,* 119, 817–822.
62. Centner, R., 1964, Final report on development of adaptive control technique for numerically controlled milling machining, USAF Tech. Documentary Report ML-TDR-64-279.
63. Amitay, G., Malkin S., and Koren, Y., 1981, Adaptive control optimization of grinding, *ASME Journal of Engineering for Industry*, 103, 1, 102–111.
64. Koren, Y., 1989, The optimal locus approach with machining applications, *ASME Journal of Dynamic Systems, Measurement, and Control*, 111, 1, 260–267.
65. Furness, R. J., Ulsoy, A. G., and Wu, C. L., 1996, Supervisory control of drilling, *ASME Journal for Manufacturing Scientists and Engineers,* 118, 10–19.
66. Xiao, G., Malkin S., and Danai, K., 1993, Autonomous system for multistage cylindrical grinding, *ASME Journal of Dynamic Systems, Measurement, and Control*, 115, 4, 667–672.
67. Ivester, R. W. and Danai, K., 1996, Intelligent control of machining under modeling uncertainty, *CIRP Manufacturing Systems*, 25, 1, 73–79.
68. Ivester, R., Danai, K., and Malkin, S., 1997, Cycle time reduction in machining by recursive constraint bounding, *ASME Journal for Manufacturing Scientists and Engineers,* 119, 2, 201–207.
69. Ko, T. J. and Cho, D. W., 1998, Adaptive optimization of face milling operations using neural networks, *ASME Journal for Manufacturing Scientists and Engineers,* 120, 443–451.
70. Azouzi, R. and Guillot, M., 1998, On-line optimization of the turning process using an inverse process neurocontroller, *ASME Journal for Manufacturing Scientists and Engineers,* 120, 101–108.

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

6

Process Monitoring and Control of Machining Operations

Robert G. Landers
University of Missouri at Rolla

A. Galip Ulsoy
University of Michigan

Richard J. Furness
Ford Motor Company

6.1 Introduction
6.2 Force/Torque/Power Generation
Cutting Force Models • Force/Torque/Power Monitoring • Force/Torque/Power Control
6.3 Forced Vibrations and Regenerative Chatter
Regenerative Chatter Detection • Regenerative Chatter Suppression
6.4 Tool Condition Monitoring and Control
Tool Failure • Tool Wear
6.5 Other Process Phenomena
Burr Formation • Chip Formation • Cutting Temperature Generation
6.6 Future Direction and Efforts

6.1 Introduction

Machining operations (e.g., drilling, milling) are shape transformation processes in which metal is removed from a stock of material to produce a part. The objective of these operations is to produce parts with specified quality as productively as possible. Many phenomena that are detrimental to this objective occur naturally in machining operations. In this chapter, we present techniques for monitoring and controlling the process phenomena that arise due to the interaction of the cutting tool and the workpiece (e.g., force generation, chatter, tool failure, chip formation).

Process monitoring is the manipulation of sensor measurements (e.g., force, vision, temperature) to determine the state of the processes. The machine tool operator routinely performs monitoring tasks; for example, visually detecting missing and broken tools and detecting chatter from the characteristic sound it generates. Unmanned monitoring algorithms utilize filtered sensor measurements that, along with operator inputs, determine the process state (Figure 6.1). The state of complex processes is monitored by sophisticated signal processing of sensor measurements that typically involve thresholding or artificial intelligence (AI) techniques.[1] For more information on sensors for process monitoring, the reader is referred to References 2 and 3.

Process control is the manipulation of process variables (e.g., feed, speed, depth-of-cut) to regulate the processes. Machine tool operators perform on-line and off-line process control by adjusting feeds and speeds to suppress chatter, initiate an emergency stop in response to a tool breakage event, rewrite a part program to increase the depth-of-cut to minimize burr formation, etc. Off-line process control is performed at the process planning stage; typically by selecting

© 2002 by CRC Press LLC

www.TechnicalBooksPDF.com

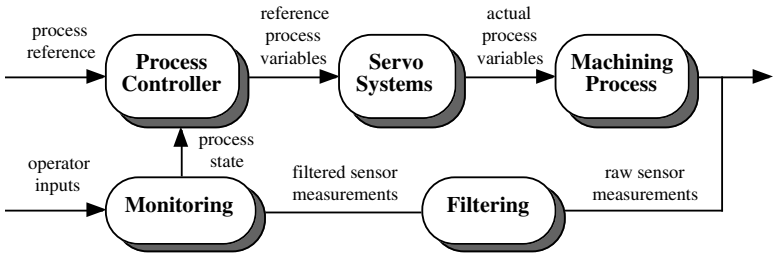


FIGURE 6.1 Process feedback control system.

process variables from a machining handbook or the operator's experience. Computer-aided process planning[4] is a more sophisticated technique which, in some cases, utilizes process models off-line to select process variables. The drawbacks of off-line planning are dependence on model accuracy and the inability to reject disturbances. Adaptive control techniques,[5] which include adaptive control with optimization, adaptive control with constraints, and geometric adaptive control, view processes as constraints and set process variables to meet productivity or quality requirements. A significant amount of research in AI techniques such as fuzzy logic, neural networks, knowledge base, etc. which require very little process information has also been conducted.[6]

This chapter concentrates on model-based process control techniques. A block diagram of a typical process feedback control system is shown in Figure 6.1. A process reference, set from productivity and quality considerations, and the process state are fed to the controller that adjusts the desired process variables. These references are input to the servo controllers that drive the servo systems (e.g., slides and spindles) that produce the actual process variables. Sensor measurements of the process are then filtered and input to the monitoring algorithms.

The trend toward making products with greater quality faster and cheaper has lead manufacturers to investigate innovative solutions such as process monitoring and control technology. Figure 6.2 shows the results of one study that clearly illustrates the benefits of process monitoring and control. A trend toward more frequent product changes has driven research in the area of reconfigurable machining systems.[7] Process monitoring technology will be critical to the cost-effective ramp-up of these systems, while process control will provide options to the designer who reconfigures the machining system. While process control has not made significant headway in industry, currently companies exist that specialize in developing process monitoring packages. Process monitoring and control technology will have a greater impact in future machining systems based on open-architecture systems[8] that provide the software platform necessary for the cost-effective integration of this technology.

The rest of the chapter is divided into six sections. The following three sections discuss force/torque/power generation, forced vibrations and regenerative chatter, and tool condition monitoring and control, respectively. The next section discusses burr and chip formation and cutting temperatures. These discussions focus on the development of models for, and the design of, process monitoring and control techniques. The last section provides future research directions. This chapter is not intended to provide an exhaustive overview of research in process monitoring and control; rather, relevant issues and major techniques are presented.

6.2 Force/Torque/Power Generation

The contact between the cutting tool and the workpiece generates significant forces. These forces create torques on the spindle and drive motors, and these torques generate power that is drawn from the motors. Excessive forces and torques cause tool failure, spindle stall (an event which is typically detected by monitoring the spindle speed), undesired structural deflections, etc. The cutting forces, torques, and power directly affect the other process phenomena; therefore, these quantities

© 2002 by CRC Press LLC

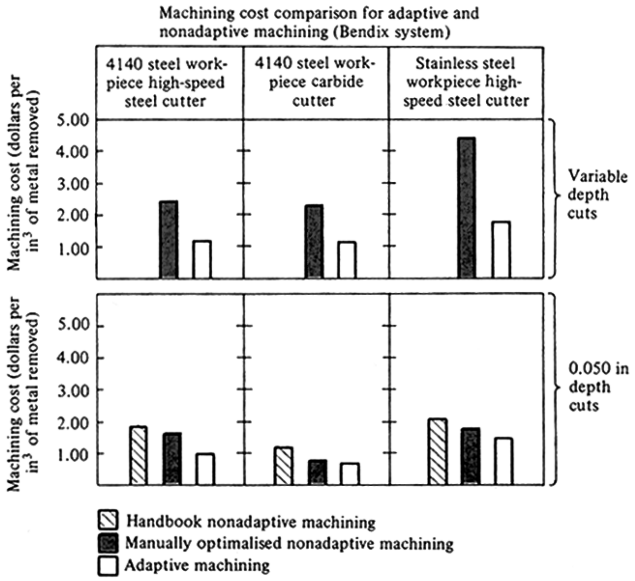


FIGURE 6.2 Machining cost comparison of adaptive and nonadaptive machining operations. (From Koren, Y. *Computer Control of Manufacturing Systems,* McGraw Hill, New York, 1983. With permission.)

are often monitored as an indirect measurement of other process phenomena and are regulated so that productivity is maximized while meeting machine tool and product quality constraints.

6.2.1 Cutting Force Models

A tremendous amount of effort has occurred in the area of cutting-force modeling over the past several decades. However, these models tend to be quite complex and experimentation is required to calibrate their parameters because an analytical model based on first principles is still not available. The models used for controller design are typically simple; however, the models used for simulation purposes are more complex and incorporate effects such as tooth and spindle runout, structural vibrations and their impact on the instantaneous feed, the effect of the cutting tool leaving the workpiece due to vibrations, intermittent cutting, tool geometry, etc. Two models that relate the actual process variables to the cutting force and are suitable for force control design are given below.

The structure of the static cutting force is

$$F = Kd^{\beta}V^{\gamma}f^{\alpha} \tag{6.1}$$

where F is the cutting force, K is the gain, d is the depth-of-cut, V is the cutting speed, f is the feed, and α, β, and γ are coefficients describing the nonlinear relationships between the force and the process variables. The model parameters in Equation (6.1) depend on the workpiece and cutting tool materials, coolant, etc. and must be calibrated for each different operation. Static models are used when considering a maximum or average force *per spindle revolution.* Such models are suitable for interrupted operations (e.g., milling) where, in general, the chip load changes throughout the spindle revolution and the number of teeth engaged in the workpiece constantly changes during steady operation (see Figure 6.3).

The structure of the first-order cutting force, assuming a zero-order hold equivalent, is

$$F = Kd^{\beta}V^{\gamma}\frac{1+a}{z+a}f^{\alpha} \tag{6.2}$$

© 2002 by CRC Press LLC

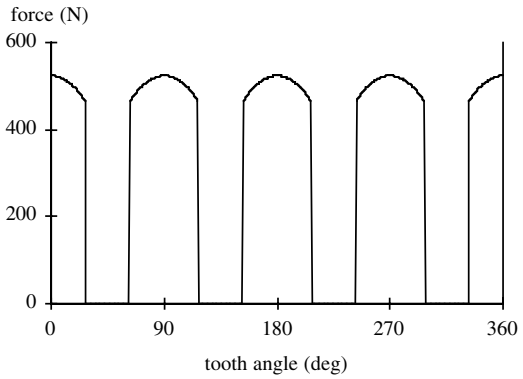


FIGURE 6.3 Simulated cutting force response for an interrupted face milling operation (four teeth, entry and exit angles of –/+ 27°). (From: Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

where a is the discrete-time pole which depends upon the time constant and the sample period, and z is the discrete-time forward shift operator. The time constant, in turn, is sensitive to the spindle speed because a full chip load is developed in approximately one tool revolution.[9] In addition to the other model parameters, a must be calibrated for each different operation. First-order models are typically employed when considering an instantaneous force that is sampled several times per spindle revolution. Such models are suitable for uninterrupted operations (e.g., turning) where, typically, a single tool is continuously engaged with the workpiece and the chip load remains constant during steady operation.

6.2.2 Force/Torque/Power Monitoring

Load cells are often attached to the machine structure to measure cutting forces. Expensive dynamometers are often used in laboratory settings for precise measurements; however, they are impractical for industrial applications. Forces in milling operations were predicted from the current of the feed axis drive.[10] This technique is only applicable if the tooth-passing frequency is lower than the servo bandwidth and the friction forces are low or can be accounted for accurately. Torque is typically monitored on the spindle unit(s) with strain gauge devices. Again, expensive dynamometers may be used, but are cost prohibitive in industrial applications. Power from the spindle and axis motors is typically monitored using Hall-effect sensors. These sensors may be located in the electrical cabinet making them easy to install and guard from the process. Due to the large masses these motors drive, the signal typically has a small bandwidth.

6.2.3 Force/Torque/Power Control

Although the three major process variables (i.e., f, d, and V) affect the cutting forces, the feed is typically selected as the variable to adjust for regulation. Typically, the depth-of-cut is fixed from the part geometry and the force–speed relationship is weak (i.e., $\gamma \approx 0$); therefore, these variables are not actively adjusted for force control. References are set in roughing passes to maximize productivity, while references are set in finishing passes to maximize quality. References in roughing passes are due to such constraints as tool failure and maximum spindle power, and references in finishing passes are due to such constraints as surface finish and tool deflections (which lead to inaccuracies in the workpiece geometry).

Most force control technology is based on adaptive techniques;[11] however, model-based techniques have recently been gaining attention.[12] Adaptive techniques consider a linear relationship between the force and the feed and view changes in process variables and other process phenomena

© 2002 by CRC Press LLC

as changes in the cutting-force parameters. Model-based techniques directly incorporate the nonlinear model and the effects of other process phenomena must be estimated. Robust control techniques[13] have also gained recent attention. These techniques incorporate the cutting-force model and require bounds on the model's parameters. Regardless of the control approach, saturation limits must be set on the commanded feed. A lower saturation of zero is typical because a negative feed will disengage the cutting tool from the workpiece; however, a nonzero lower bound may be set due to process constraints. An upper bound is set due to process or machine tool servo constraints.

Two machining force controllers are designed and implemented next for the following static cutting force

$$F = 0.76 d^{0.65} f^{0.63} \tag{6.3}$$

where $\gamma = 0$ and F is a maximum force per spindle revolution in a face milling operation. For control design, the model is augmented with an integral state to ensure constant reference tracking and constant disturbance rejection.

A model-based design is now applied.[12] The control variable is $u = f^{0.63}$ and the design model (with an integral state) is

$$F(z) = \theta \frac{1}{z-1} u(z) \tag{6.4}$$

where $\theta = 0.76 d^{0.65}$ is the gain. Note that the nonlinear model-based controller utilizes process information (in this case, depth-of-cut) to directly account for known process changes. The model reference control (MRC) approach is applied and the control law is

$$u(z) = \frac{1}{z-1} \frac{1+b_0}{\theta} \left[F_r(z) - F(z) \right] \tag{6.5}$$

where F_r is the reference force and b_0 is calculated given a desired closed-loop time constant and sample period. The commanded feed is calculated from the control variable as

$$f = \exp\left[\frac{\ln(u)}{0.63} \right] \tag{6.6}$$

Therefore, the lower saturation on the control variable is chosen to have a small non-negative value. The experimental results for the nonlinear model-based controller are shown in Figure 6.4.

Next, an adaptive force controller is designed. The control design model, including an integral state, is

$$F(z) = \theta \frac{1}{z-1} f(z) \tag{6.7}$$

where θ is the gain and is assumed to be unknown. The MRC approach is applied and the control law is

$$f(z) = \frac{1}{z-1} \frac{1+b_0}{\hat{\theta}} \left[F_r(z) - F(z) \right] \tag{6.8}$$

The term $\hat{\theta}$ is an estimate of the gain. In this example, the common recursive least squares technique is employed.[14] At the i^{th} time iteration, the estimate is calculated as

© 2002 by CRC Press LLC

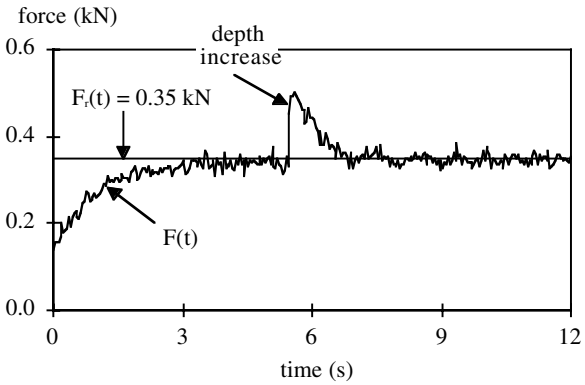


FIGURE 6.4 Force response, nonlinear model-based force controller. (From Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

$$\hat{\theta}(i) = \hat{\theta}(i-1) + K(i)\varepsilon(i) \tag{6.9}$$

where

$$K(i) = \frac{P(i-1)f(i)}{\left[1 + f(i)P(i-1)f(i)\right]} \tag{6.10}$$

$$P(i) = \left[1 - K(i)f(i)\right]P(i-1) \tag{6.11}$$

$$\varepsilon(i) = F(i) - f(i)\hat{\theta}(i-1) \tag{6.12}$$

The parameter P is known as the covariance and the parameter ε is known as the residual. Estimating the model parameters on-line is a strong method of accounting for model inaccuracies; however, the overall system becomes much more complex, and chaotic behavior may result.

The experimental results for the adaptive controller are shown in Figures 6.5 and 6.6. Both approaches successfully regulate the cutting force while accounting for process changes in very different ways. The adaptive technique is useful when an accurate model is not available, but is more complex compared to the model-based approach.

6.3 Forced Vibrations and Regenerative Chatter

The forces generated when the tool and workpiece come into contact produce significant structural deflections. Regenerative chatter is the result of the unstable interaction between the cutting forces and the machine tool–workpiece structures, and may result in excessive forces and tool wear, tool failure, and scrap parts due to unacceptable surface finish.

The feed force for an orthogonal cutting process (e.g., turning thin-walled tubes) is typically described as

$$F(t) = Kd\left[f_n + x(t) - x(t-\tau)\right] \tag{6.13}$$

© 2002 by CRC Press LLC

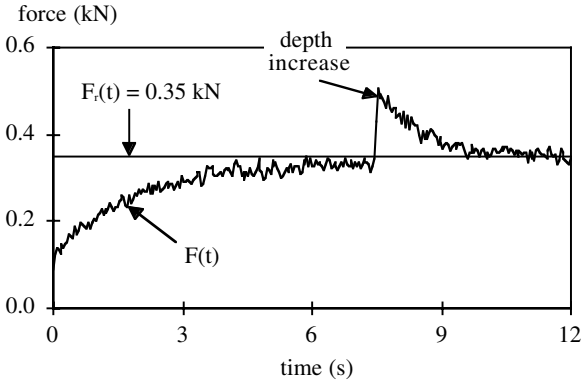


FIGURE 6.5 Force response, an adaptive force controller. (From Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

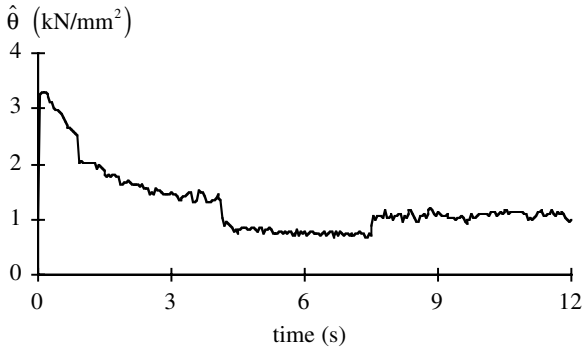


FIGURE 6.6 Force model gain estimate, an adaptive force controller. (From Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

where f_n is the nominal feed, x is the displacement of the tool in the feed direction, and τ is the time for one tool revolution. The assumption is that the workpiece is much more rigid than the tool, and the force is proportional to the instantaneous feed and the depth-of-cut and does not explicitly depend upon the cutting speed. The instantaneous chip load is a function of the nominal feed, the current tool displacement, and the tool displacement at the previous tool revolution. Assuming a simple model, the vibration of the tool structure may be described by

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = F(t) \tag{6.14}$$

where m, c, and k are the effective mass, damping, and stiffness, respectively, of the tool structure. The stability of the closed-loop system formed by equations combining (6.13) and (6.14) may be examined to generate the so-called stability lobe diagram (Figure 6.7) and select appropriate process variables.

Another cause of unacceptable structural deflections, known as forced vibrations, arises when an input frequency (e.g., tooth-passing frequency) is close to a resonant structural frequency. The resulting large relative deflections between the cutting tool and workpiece lead to inaccuracies in

© 2002 by CRC Press LLC

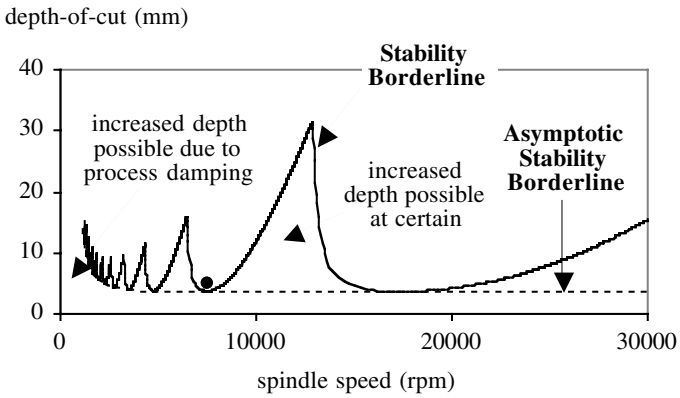


FIGURE 6.7 Stability lobe diagram. The tool structure's natural frequency is 12,633 Hz. Operating point (d = 5 mm, N_s = 7500 rpm) denoted by dark circle is used in the simulations in Figures 6.10 and 6.11.

the workpiece geometry. An example of forced vibrations may be found in Reference 15. When the tooth-passing frequency is close to a dominant structural frequency, productivity may be increased (see Figure 6.7); however, forced vibrations will occur. Therefore, the designer must make a trade-off between controlling regenerative chatter and inducing forced vibrations

In this section, common techniques for on-line chatter detection and suppression are presented.

6.3.1 Regenerative Chatter Detection

Regenerative chatter is easily detected by an operator because of the loud, high-pitched noise it produces and the distinctive "chatter marks" it leaves on the workpiece surface. However, automatic detection is much more complicated. The most common approach is to threshold the spectral density of a process signal such as sound,[16] force,[17] etc. An example in which the force signal is utilized for chatter detection (see Figure 6.8) demonstrates that chatter frequency occurs near a dominant structural frequency. Note that the tooth-passing frequency contains significant energy. In this application, the lower frequencies may be ignored by the chatter detection algorithm; however, if the operation is performed at a higher spindle speed, the force signal has to be filtered at the tooth-passing frequency. Also, the impact between the cutting tool and workpiece will cause structural vibrations that must not be allowed to falsely trigger the chatter detection algorithm.

These thresholding algorithms all suffer from the lack of an analytical method to select the threshold value. This value is typically selected empirically and will not be valid over a wide range of cutting conditions. A more general signal was proposed by Bailey et al.[18] An accelerometer signal mounted on the machine tool structure close to the cutting region was processed to calculate the so-called variance ratio

$$R = \sigma_s^2 / \sigma_n^2 \tag{6.15}$$

where σ_s and σ_n are the variances of the accelerometer signal in low and high frequency ranges, respectfully. A value of $R << 1$ indicates chatter.

6.3.2 Regenerative Chatter Suppression

Chatter is typically suppressed by adjusting the spindle speed to lie in one of the stability lobe pockets, as shown in Figure 6.7.[19] Feed has been shown to have a monotonic effect on the marginally stable depth-of-cut (see Figure 6.9) and is sometimes the variable of choice by machine tool

© 2002 by CRC Press LLC

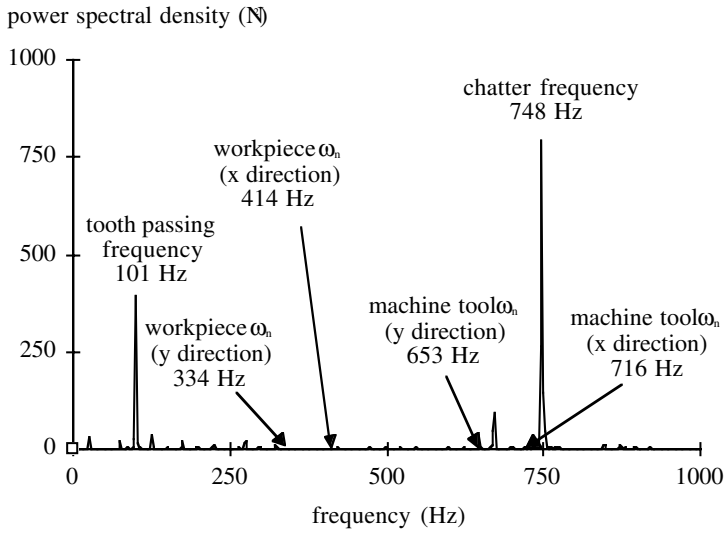


FIGURE 6.8 Power spectrum of force signal during chatter. (From Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

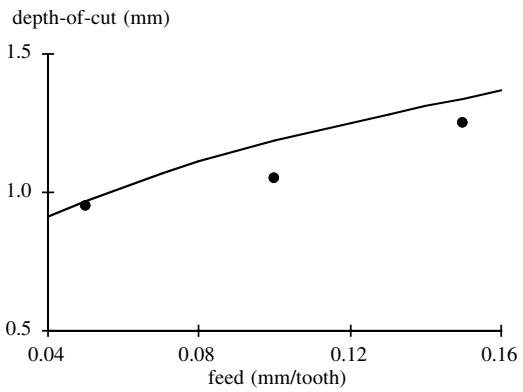


FIGURE 6.9 Theoretical prediction (solid line) vs. experimental data (circles) demonstrating the feed effect on chatter. (From Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

operators.[20] The tool position may also be adjusted (e.g., depth-of-cut decreased) to suppress chatter, and while it is guaranteed to work (see Figure 6.7), this approach is typically not employed because the part program must be rewritten and productivity is drastically decreased.

Spindle speed variation (SSV) is another technique for chatter suppression.[15] The spindle speed is varied about some nominal value, typically in a sinusoidal manner. Figures 6.10 and 6.11 demonstrate how varying the spindle speed sinusoidally with an amplitude of 50% of the nominal value and at a frequency of 6.25 Hz will suppress chatter that occurs when a constant spindle speed at the nominal value is utilized (see Figure 6.7). Although SSV is a promising technique, little theory exists to guide the designer to the optimal variation and, in some cases, SSV may create chatter which will not occur when using a constant spindle speed. Further, it can be seen in Figure 6.11b that SSV will cause force fluctuations even though the chatter is suppressed.

© 2002 by CRC Press LLC

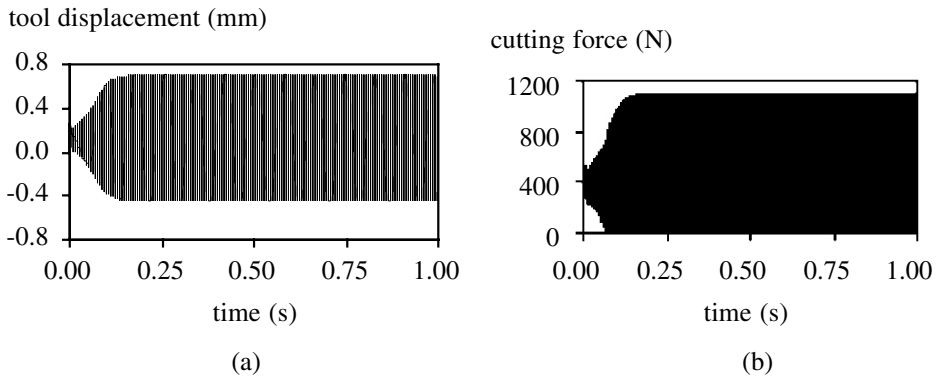


FIGURE 6.10 Simulated responses of force and structural displacements for constant speed machining. Cutting conditions given in Figure 6.7.

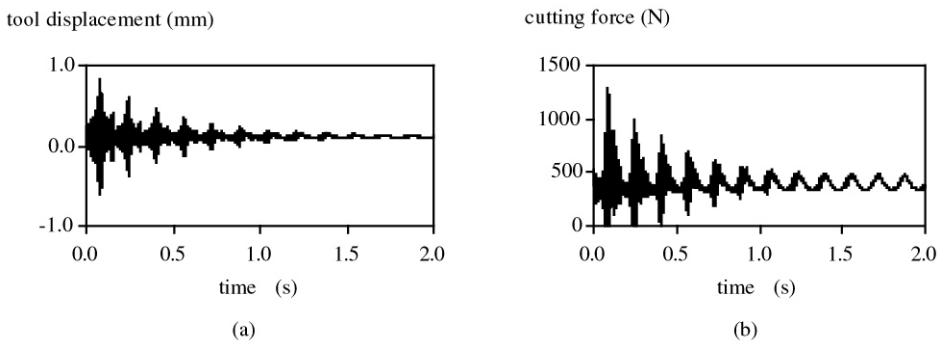


FIGURE 6.11 Simulated responses of force and structural displacements for variable speed machining. Cutting conditions given in Figure 6.7.

6.4 Tool Condition Monitoring and Control

Some of the most common monitoring techniques concentrate on tool condition monitoring. Vision sensors and probes are used to detect missing cutting tools in a tool magazine and to ensure the correct tool is being used. Vision and force sensors are also used to detect tool–workpiece collisions or tool–tool collisions in parallel machining operations. If a collision is detected, an emergency stop is typically initiated and the part program must be rewritten. The monitoring and control of the more complicated tool condition phenomena (i.e., tool failure and tool wear) are discussed next.

6.4.1 Tool Failure

A tool has failed when it can no longer perform its designated function. This event may occur when a significant portion of the tool breaks off, the tool shaft or cutting teeth severely fracture, or a significant portion of one or more teeth chip. Broken tools drastically decrease productivity by creating unnecessary tool changes, wasting tools, and creating scrap parts, and possibly injuring operators.

The simplest way to detect a failed tool is to use a probe or vision system to inspect the cutting tool. While this inspection is typically performed off-line, some techniques are being developed

© 2002 by CRC Press LLC

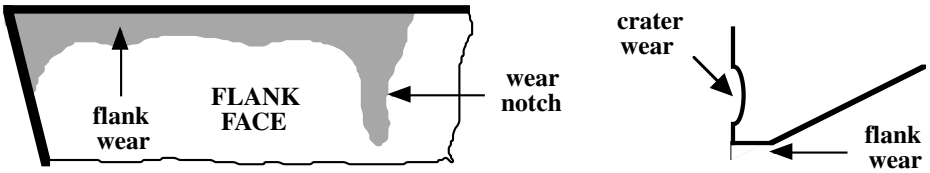


FIGURE 6.12 Illustration of different types of tool wear.

for on-line detection;[21] however, chip and coolant interference is still a major obstacle to overcome. Many sensors have been used to indirectly detect tool failure, including acoustic emission, force, sound, vibration, etc. In these indirect methods, the signal magnitude, root mean square value, or the magnitude of the power spectrum, among others, are inspected, typically via thresholding. One example is given in Altintas[10] where the residual of a first-order adaptive auto-regressive time series filter of the average (during a tooth pass) drive current was monitored to detect insert chippage. Creating a static threshold value is difficult to do in complex machining operations; therefore, dynamic limits are often set to account for entry and exit conditions, changes in process variables, etc. For operations where the feed rate is not adjusted, these limits may be correlated with time; however, in general, these limits should be correlated with position. Pattern recognition techniques may also be utilized. If a signal is compared to a stored pattern, then breakage may be determined independent of the signal magnitude. Comparison to teach-in signals (i.e., an average of several signals in similar operations where breakage did not occur) is another technique. Currently, little theory exists to guide the user in setting these limits.

When a tool failure event has been detected, an emergency stop is typically initiated. A significant amount of time is spent not only changing the cutting tool and workpiece, but also restarting the machine tool or machining line. This loss of productivity can be avoided by an intelligent reaction to the tool failure event. For example, the cutting tool may be moved to the tool change position and vision may be utilized to examine the workpiece surface to verify whether or not the workpiece must be scrapped. As another example, if a tooth chips in a milling cutter, optical techniques may be used to determine if the workpiece and tool are undamaged and, if so, the feed can be decreased and cutting may continue.

There have been some studies to detect the onset of tool failure. In Rice and Wu,[22] the energy release rate of an acoustic emission signal was monitored in interrupted cutting tests to determine the advancement of a fracture event. If a tool does fail, steps must be taken to ensure that failure does not happen again. Typically, a process parameter, i.e., the feed is adjusted; however, a reference force may also be adjusted if a force control scheme is being employed.

6.4.2 Tool Wear

The contact between the cutting tool and the chips causes the shape of the tool to change (Figure 6.12). This phenomenon, known as tool wear, has a major influence in machining economics, affects the final workpiece dimensions, and will lead to eventual tool failure. A typical tool-wear curve is shown in Figure 6.13. The tool wears rapidly in the initial phase and then levels off to a constant rate during the steady phase. From an economic point of view, the designer would like to use the tool until just before it enters the accelerated wear phase during which the tool will eventually fail.

The three main tool-wear mechanisms include abrasion between the cutting tool and workpiece, which is always present; adhesion of the chips or workpiece to the cutting tool, which removes cutting tool material and is more active as the cutting temperature increases; and diffusion of the cutting tool atoms to the chips or workpiece, which is typically active during the accelerated tool-wear phase.

The most well-known equation describing tool wear was developed by F. W. Taylor early in the twentieth century.[23] This equation, known as Taylor's tool equation, is

© 2002 by CRC Press LLC

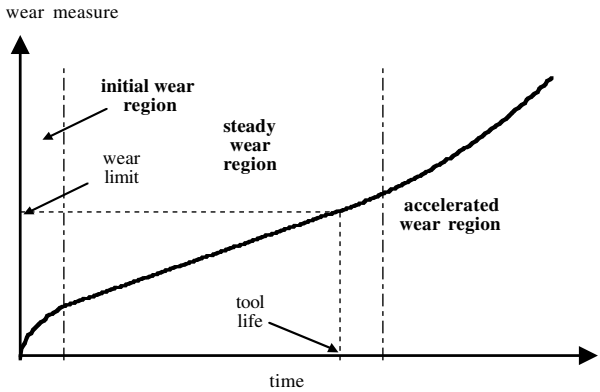


FIGURE 6.13 Typical tool wear history.

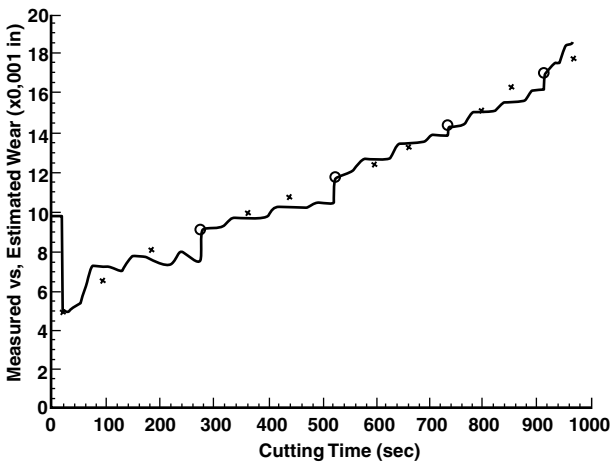


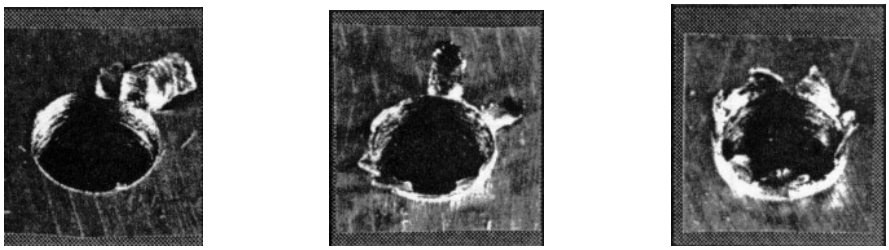
FIGURE 6.14 Estimated (solid line) vs. measured (crosses) flank wear. The circles are vision measurements used to recalibrate the adaptive observer. (From Park, J.J. and Ulsoy, A.G., *ASME Journal of Engineering for Industry*, 115, 37, 1993. With permission.)

$$Vt_l^n = C \tag{6.16}$$

where t_l is the tool lifetime and C and n are empirically determined constants. Modified Taylor equations include the effects of feed rate and depth-of-cut, as well as interaction effects between these variables. Increased testing is required to determine the extra model coefficients; however, these models are applicable over a wider range of cutting conditions. Models relating tool wear and cutting forces have also been developed.[24,25] See Kendall[26] for more information regarding cutting tool-wear mechanisms and modeling.

The most reliable way to monitor tool wear is by direct visual inspection. Indirect techniques utilizing such measurements as acoustic emission, force, temperature, vibration, etc. have also been developed, or the final part geometry may be measured. Similar to tool breakage monitoring, these indirect signals are typically processed to expose those characteristics that are highly correlated

© 2002 by CRC Press LLC



(a) (b) (c)

FIGURE 6.15 Exit burrs in a through-hole drilling operation and their burr ratings: (a) 1, (b) 3, (c) 5. (From Furness, R.J., Ulsoy, A.G., and Wu, C.L., *ASME Journal of Engineering for Industry*, 118, 10, 1996. With permission.)

with tool wear. Again, cutting tests are required to determine this correlation. In Park and Ulsoy,[25] a hybrid tool-wear monitoring technique was investigated. An adaptive observer was applied to estimate wear on-line and a vision system was used intermittently (e.g., between parts) to recalibrate the observer (Figure 6.14). The reader is referred to Dan and Mathew[27] for an overview of tool-wear monitoring.

The two main issues in tool-wear regulation are to compensate for tool wear and to control the tool-wear rate. As the tool wears, the workpiece dimension may become out of tolerance; thus, the tool position must be adjusted (typically through the part program) to compensate for the tool wear. From an economic point of view, it is desirable to regulate the tool-wear rate so that the tool life corresponds to the scheduled tool change period in mass production, or to maximize tool life in job-shop situations.

6.5 Other Process Phenomena

6.5.1 Burr Formation

Small, undesirable metal fragments left on the workpiece after the machining operation is complete are known as burrs (Figure 6.15). Burrs cause improper part mating, accelerated device wear, and decreased device performance. Because it is typically impossible to avoid the formation of burrs, the designer should strive to reduce the complexity of subsequent deburring operations by minimizing the burr strength and ensuring the burrs form at easily accessible workpiece locations.

The three major burr types (poisson, roll-over, and tear) form due to workpiece plastic deformation. When the cutting-tool edge extends over a workpiece edge, material is compressed and may flow laterally forming a poisson burr. Roll-over burrs form when the cutting tool exits the workpiece and the chip bends over the edge instead of being cut. If a chip is torn from the workpiece, instead of being sheared off, some material from the chip will be left on the workpiece. The material is known as a tear burr. The reader is referred to Gillespie[28] for greater detail concerning burr models. Burr measurement is typically performed off-line by measuring the average height, base thickness, and toughness. Burr location and its accessibility are also important to note.

Process variables are known to have a strong effect on the physical characteristics of burrs. If the depth-of-cut in a face milling operation is too small, the cutting tool will push the material over the side of the workpiece and form a large, strong burr on the workpiece edge. In Furness, Ulsoy, and Wu,[29] a feed controller regulated the feed at 0.051 mm/rev as the tool exited the workpiece in a through-hole drilling operation to obtain an acceptable burr rating. The burr rating depended on burr thickness and peak height, percentage of the hole's circumference with an attached burr, and qualitative assessment of the relative ease of removal. Without adequate models, one is left to empirical techniques or AI methods to predict, and hence control, burr formation.

© 2002 by CRC Press LLC

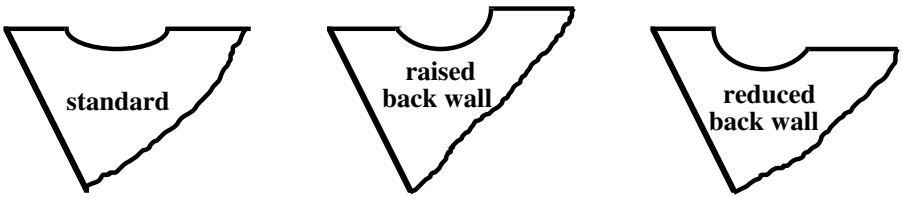


FIGURE 6.16 Illustration of common chip breakers.

6.5.2 Chip Formation

The three major chip formation types are discontinuous, continuous, and continuous with built-up edge (BUE).[30] Discontinuous chips arise when the operation continuously forms and fractures chips because of the workpiece's inability to undergo large amounts of plastic deformation, while continuous chips do not fracture but form continuous ribbons. Continuous chips with BUE form when part of the chip welds to the tool due to high cutting temperatures and pressures. Continuous chips (with and without BUE) will interfere with the normal interaction between the tool and workpiece and cause poor surface finish, as will discontinuous chips that do not clear the cutting zone. Therefore, chip control is the proper formation of chips that clear the cutting zone and are directed toward the chip conveyor system for efficient removal.

Research of the chip formation process goes back nearly a century, starting most notably with Taylor.[23] Theories have been developed to predict shear plane angle, chip velocity, etc. mainly for two-dimensional cases. More recently, chip curling and chip breaking models have been emphasized. These models, however, are not widely applicable. Currently, computational mechanics (i.e., finite element methods) and artificial intelligence (AI) methods have been applied. See van Luttervelt, et al.[31] for a comprehensive overview of the current status of machining modeling.

High-speed filming techniques have been used to directly monitor chip formation. Indirect methods include force, acoustic emission, and infrared emission measurements, and sensor fusion based on AI techniques.

Chip formation control is typically achieved through the design of chip breakers (Figure 6.16). The grooves cause an otherwise continuous chip to curl and fracture. Small amplitude, high-frequency variations in the feed are a relatively new technique for ensuring chip fracture. This variation is accomplished using a passive device attached to the cutting tool and may also be accomplished by varying the feed rate on-line; however, the variation frequency will be limited by the bandwidth of the servo system. The use of process parameters has also been investigated. While chip curling is typically independent of process variables, thicker chips formed from relatively large feeds break more easily than do thinner chips.[32] Due to the complexity and incomplete knowledge of chip formation, a database approach to selecting chip breakers and process variables is the most reliable method for chip control. See Jawahir and van Luttervelt[33] for a comprehensive overview of research in this area.

6.5.3 Cutting Temperature Generation

Friction between the cutting tool and workpiece generates significant temperature in the cutting zone. The cutting temperature affects the tool wear rate and workpiece surface integrity, and contributes to thermal deformation.

The most basic temperature models estimate steady-state cutting temperatures and typically have the following nonlinear relationship with the process variables[34]

$$T = aV^b f^c \tag{6.17}$$

© 2002 by CRC Press LLC

where T is the workpiece temperature and a, b, and c are empirically determined constants. A comparison with experimental results shows most models are qualitatively correct, but quantitatively overestimate cutting temperatures and are unable to estimate cutting temperatures in operations with discontinuous chip formation.[35] The use of thermocouples and infrared data to measure cutting temperatures was investigated; however, cutting temperature measurements are rarely utilized in industrial settings.[35]

Similar to burr and chip formation, cutting temperature generation has received little attention from the control community. One investigation was performed by D'Errico, Calzavarini, and Settineri.[36] Using a simple static nonlinear relationship between cutting temperature and cutting velocity similar to Equation (6.17), with $c = 0$, a self-tuning regulator was developed to control the cutting temperature via adjustment of the cutting velocity.

6.6 Future Directions and Efforts

This chapter has presented the major techniques for monitoring and controlling the phenomena arising from the interaction of the cutting tool and the workpiece in machining operations. It can be readily seen that advances in the modeling of cutting mechanics are required; in particular, analytical models based on first principles applicable to a wide variety of cutting conditions must be developed. Currently, models are determined empirically and typically contain nonlinear terms that account for unmodeled effects. Further, the cost-effective design of process monitoring and control technology will require simulation tools that simulate not only cutting mechanics and monitoring and control modules, but also the machine tool structure and servo mechanisms. A comprehensive simulator will allow the designer to investigate process monitoring and control technology in a realistic environment (i.e., one with the appropriate complexities).

The biggest obstacles facing the implementation of process monitoring technology are low reliability, limited applicability, and the need for experimentation to determine threshold values, characteristic patterns, etc. Advances in models based on first principles and the increased use of sophisticated signal processing techniques will be required to overcome these obstacles. Other issues in process monitoring include the use of increasingly sophisticated sensors and the placement of these sensors in harsh machining environments. Advances in sensor technology to integrate the sensors with the machine tool or cutting tool and research into using computer numerical control (CNC)-integral sensors (e.g., drive current) will address these issues.

Currently, the largest research effort in process monitoring is the Intelligent Manufacturing Systems (IMS) project Sensor Fused Intelligent Monitoring System for Machining (SIMON) which is an international, industry-driven project with the goal of developing a practical monitoring system that can reliably identify actual cutting conditions according to information obtained from a sensor-fused system.[37] Another development in the field of process monitoring is a mapping theory to facilitate the cost-effective design of modular monitoring packages.[38] Given the machining operation, the so-called fault space (e.g., chippage, tool deformation) is generated. The characteristics of these faults are mapped to those of the required sensor and used to select the correct sensor package. The monitoring package will then be applied in the ramp-up phase of a machining system.

As process monitoring techniques become more reliable, process control will become more prevalent. During the ramp-up phase of a machining system, process controllers will provide an effective means of determining near-optimal process variables for complex operations. The part program can be modified to incorporate the new process variable time histories and then process controllers may be utilized in the production phase to reject disturbances. While process control is not widely implemented in industry today, a substantial amount of work has been done in research laboratories. This research has almost always been concerned with regulating a single process via a single process variable. Future research will be concerned with utilizing multiple process variables to control a single process and implementing multiple process controllers simultaneously in a single operation.

© 2002 by CRC Press LLC

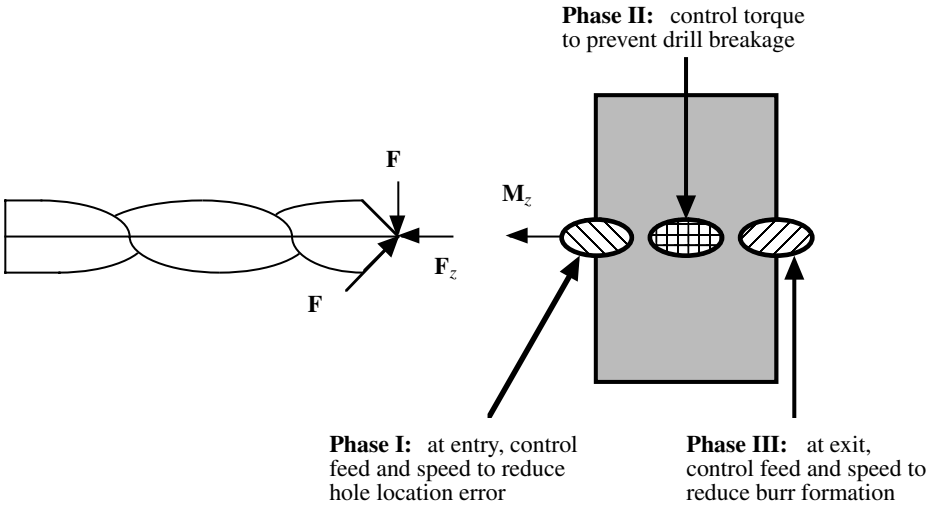


FIGURE 6.17 Illustration of an off-line supervisory control implementation in a through-hole drilling operation.

The concept of implementing multiple process controllers has lead to research in supervisory control.[29,39,40] The supervisory control of a through-hole drilling operation was investigated in Furness, Ulsoy, and Wu.[29] The objective was to maximize operation productivity subject to a set of machine, process, and quality constraints. Machine constraints included a maximum spindle speed and feed rate. Process constraints included a maximum torque to avoid drill breakage and cutting torque limitations, a maximum force to avoid buckling, and a minimum tool life to maintain a constant tool replacement period. Quality constraints included a maximum hole location error and minimum burr formation. The process controllers were supervised using an off-line optimization technique where the controller configuration depended on workpiece location (see Figure 6.17). The experimental results for the supervisory controller compared to other controller configurations are shown in Table 6.1.

A state-based, on-line supervisory controller was developed in Landers and Ulsoy.[40] A state supervisor monitored the operation including discrete events (e.g., tool–workpiece contact, chatter) and continuous signals (e.g., force model parameter estimates). Given the operation state, an operation supervisor configured the monitoring and control modules (i.e., turned them off and on, reset them, etc.). Experimental results for a face milling operation are shown in Figure 6.18. The force controller and chatter detector were turned on when the tool and workpiece came into contact. As the tool became fully engaged in the workpiece, chatter developed. The chatter suppressor rewrote the part program to add an additional tool pass and implemented a feed hold for five tool revolutions to allow the vibrations to die out. The force controller was then reset and machining continued. The force controller and chatter detector were turned off as the tool exited the workpiece and were again implemented as the second tool pass began.

TABLE 6.1 Comparison of Drilling Control Strategies[41]

	No Controller	Feed/Speed Controller	Torque/Speed Controller	Supervisory Controller
Machining time (s)	11.11	11.28	9.79	11.71
Burr rating	2.93	2.94	2.26	1.58
Hole location quality (in)	4.43 E-3	4.53 E-3	6.28 E-3	4.25 E-3
Event stoppages (%)	25	15	0	0

Source: Ulsoy, A.G. and Koren, Y., *ASME Journal of Dynamic Systems, Measurement, and Control,* 115, 301, 1993. With permission.

© 2002 by CRC Press LLC

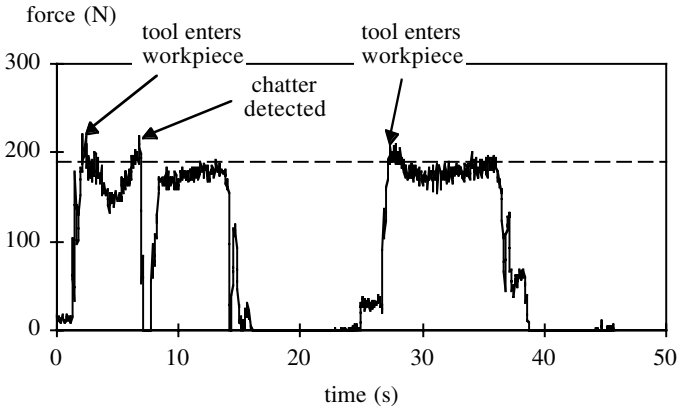


FIGURE 6.18 Force history results using a supervisory controller during a face milling operation. (From Landers, R.G., Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components, Ph.D. dissertation, University of Michigan, Ann Arbor, 1997.)

Acknowledgments

The authors gratefully acknowledge Yuan-Hung (Kevin) Ma and Yowjie Chien for their assistance in preparing Figures 6.7, 6.10, and 6.11 and Figures 6.2, 6.14, and 6.15, respectively, and the National Science Foundation Engineering Research Center for Reconfigurable Machining Systems (Grant EEC95-92125) at The University of Michigan for its financial support.

References

1. Du, R., Elbestawi, M. A., and Wu, S. M., Automated monitoring of manufacturing processes, Part 1: Monitoring methods, *ASME Journal of Engineering for Industry*, 117, 121, 1995.
2. Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., Konig, W., and Teti, R., Tool condition monitoring (TCM) — the status of research and industrial application, *Annals of the CIRP*, 44, 541, 1995.
3. Jemielniak, K., Commercial tool condition monitoring systems, in *5th International Conference on Monitoring & Automatic Supervision in Manufacturing*, Warsaw University of Technology, Warsaw, 1998, 59.
4. ElMaraghy, H. A., Evolution and future perspectives of CAPP, *Annals of the CIRP*, 42, 739, 1993.
5. Koren, Y., Adaptive control systems for machining, *Manufacturing Review*, 2, 6, 1989.
6. Rangwala, S. and Dornfeld, D. A., Learning and optimization of machining operations using computing abilities of neural networks, *IEEE Transactions on Systems, Man, and Cybernetics*, 19, 299, 1989.
7. Koren, Y. and Ulsoy, A. G., Reconfigurable manufacturing systems, Technical Report #1, NSF Engineering Research Center for Reconfigurable Machining Systems, University of Michigan, Ann Arbor, 1998.
8. Pritschow, G., Daniei, C. H., Jurghans, G., and Sperling, W., Open systems controllers — a challenge for the future of the machine tool industry, *Annals of the CIRP,* 42, 449, 1993.
9. Koren, Y. and Masory, O., Adaptive control with process estimation, *Annals of the CIRP*, 30, 373, 1981.
10. Altintas, Y., Prediction of cutting forces and tool breakage in milling from feed drive current measurements, *ASME Journal of Engineering for Industry*, 114, 386, 1992.
11. Ulsoy, A. G., Koren, Y., and Rasmussen, F., Principal developments in the adaptive control of machine tools, *ASME Journal of Dynamic Systems, Measurement, and Control*, 105, 107, 1983.
12. Landers, R. G. and Ulsoy A. G., Machining force control including static, nonlinear effects, in *Japan–USA Symposium on Flexible Automation*, ASME, New York, 1996, 983.

© 2002 by CRC Press LLC

13. Rober, S. J., Shin, Y. C., and Nwokah, O. D. I., A digital robust controller for cutting force control in the end milling process, *ASME Journal of Dynamic Systems, Measurement, and Control*, 119, 146, 1997.
14. Åström, K. J. and Wittenmark, B., *Adaptive Control*, Addison-Wesley, New York, 1995, 2.
15. Radulescu, R., Kapoor, S. G., and DeVor, R. E., An investigation of variable spindle speed face milling for tool-work structures with complex dynamics, Part 2: Physical explanation, *ASME Journal of Manufacturing Science and Engineering*, 119, 273, 1997.
16. Smith, S. and Delio, T., Sensor-based chatter detection and avoidance by spindle speed selection, *ASME Journal of Dynamic Systems, Measurement, and Control*, 114, 486, 1992.
17. Landers, R. G., *Supervisory Machining Control: A Design Approach Plus Force Control and Chatter Analysis Components*, Ph.D. dissertation, Department of Mechanical Engineering and Applied Mechanics, University of Michigan, Ann Arbor, 1997.
18. Bailey, T., Ruget, Y., Spence, A., and Elbestawi, M. A., Open-architecture controller for die and mold machining, in *Proceedings of the American Control Conference*, 1, IEEE, Piscataway, 1995, 194.
19. Delio, T., Tlusty, J., and Smith, S., Use of audio signals for chatter detection and control, *ASME Journal of Engineering for Industry*, 114, 146, 1992.
20. Landers, R. G. and Ulsoy A. G., Chatter analysis of machining systems with nonlinear force processes, in *ASME International Mechanical Engineering Congress and Exposition*, DSC 58, ASME, New York, 1996, 183.
21. Jones, S. D., Mori, K., and Ryabov, O., Cutting tool sensor and requirements for reducing process variation, in *Japan–USA Symposium on Flexible Automation*, ASME, New York, 1996, 991.
22. Rice, J. A. and Wu, S. M., On the feasibility of catastrophic cutting tool fracture prediction via acoustic emission analysis, *ASME Journal of Engineering for Industry*, 115, 390, 1993.
23. Taylor, F. W., On the art of cutting tools, *Transactions ASME*, 28, 1907.
24. Koren, Y., Ko, T. R., Ulsoy, A. G., and Danai, K., Flank wear estimation under varying cutting conditions, *ASME Journal of Dynamic Systems, Measurement, and Control*, 113, 300, 1991.
25. Park, J. J. and Ulsoy, A. G., On-line flank wear estimation using an adaptive observer and computer vision, Part 2: Experiment, *ASME Journal of Engineering for Industry*, 115, 37, 1993.
26. Kendall, L. A., Tool wear and tool life, in *Metals Handbook: Machining*, ASM International, Metals Park, Ohio, 1989, 16.
27. Dan, L. and Mathew, J., Tool wear and failure monitoring techniques for turning — a review, *International Journal of Machine Tools and Manufacture*, 30, 579, 1990.
28. Gillespie, L. K., *Deburring Capabilities and Limitations*, Society of Manufacturing Engineers, Dearborn, MI, 1976.
29. Furness, R. J., Ulsoy, A. G., and Wu, C. L., Supervisory control of drilling, *ASME Journal of Engineering for Industry*, 118, 10, 1996.
30. DeVries, W. R., *Analysis of Material Removal Processes*, Springer-Verlag, New York, 1992.
31. van Luttervelt, C. A., Childs, T. H. C., Jawahir, I. S., Klocke, F., and Venuvinod, P. K., The state of the art of modeling in machining processes, *Annals of the CIRP*, 47, 587, 1998.
32. Rotberg, J. and Ber, A., Chip control in cut-off tools, *Annals of the CIRP,* 40, 73, 1991.
33. Jawahir, I. S. and van Luttervelt, C. A., Recent developments in chip control research and applications, *Annals of the CIRP,* 42, 659, 1993.
34. Chu, T. H. and Wallbank, J., Determination of the temperature of a machined surface, *ASME Journal of Manufacturing Science and Engineering*, 120, 259, 1998.
35. Stephenson, D. A., Assessment of steady-state metal cutting temperature models based on simultaneous infrared and thermocouple data, *ASME Journal of Engineering for Industry*, 113, 121, 1991.
36. D'Errico, G. E., Calzavarini, R., and Settineri, L., Experiments on self-tuning regulation of cutting temperature in turning process, in *Proceedings of the IEEE Conference on Control Applications*, IEEE, Piscataway, 1994, 1165.

© 2002 by CRC Press LLC

37. Kaever, M. and Weck, M., Intelligent process monitoring for rough milling operations based on digital drive currents and machine integrated sensors, in *ASME International Mechanical Engineering Congress and Exposition*, MED 6-1, ASME, New York, 1997, 97.
38. Kannatey-Asibu, E., New concepts on multi-sensor monitoring for reconfigurable machining systems, in *ASME International Mechanical Engineering Congress and Exposition*, MED8, ASME, New York, 1998, 589.
39. Teltz, R. and Elbestawi, M. A., Hierarchical, knowledge-based control in turning, *ASME Journal of Dynamic Systems, Measurement, and Control*, 115, 122, 1993.
40. Landers, R. G. and Ulsoy, A. G., Supervisory machining control: Design approach and experiments, *Annals of the CIRP*, 47, 301, 1998.
41. Ulsoy, A. G. and Koren, Y., Control of machining processes, *ASME Journal of Dynamic Systems, Measurement, and Control*, 115, 301, 1993.
42. Koren, Y., *Computer Control of Manufacturing Systems*, McGraw Hill, New York, 1983.

© 2002 by CRC Press LLC

7

Forming Processes: Monitoring and Control

7.1 Introduction: Process and Control Objectives
Process Control Issues • The Process: Material Diagram • The Machine Control Diagram
7.2 The Plant or Load: Forming Physics
Mechanics of Deformation: Machine Load Dynamics • Mechanics of Forming: Bending, Stretching, and Springback
7.3 Machine Control
Sensors
7.4 Machine Control: Force or Displacement?
7.5 Process Resolution Issues: Limits to Process Control
Process Resolution Enhancement
7.6 Direct Shape Feedback and Control
7.7 Summary

David E. Hardt
Massachusetts Institute of Technology

7.1 Introduction: Process and Control Objectives

Forming of metallic materials is the process of choice when complex net shapes with high levels of productivity are desired. Myriad processes, ranging from job-shop metal bending machines to very high speed stamping and forging presses are available. In all cases, the processes involve plastic deformation of the workpiece, and the resulting strong forces required to create plastic stresses. In this chapter, the problem of controlling such processes is considered from both the viewpoint of controlling the forming equipment and the deformation process itself. Several unique aspects of forming processes arise when considering control system design:

1. The process or plant transfer function becomes a static block with variable gain and severe hysteresis.
2. The plant (the forming process) is inherently variable owing to the sensitivity to the workpiece material properties.
3. An inherent lack of process degrees of freedom with respect to controlling overall part shape exists.

Metal forming can be divided into sheet-forming processes and bulk-forming processes (typically forging). The major difference is that the latter involves a complex three-dimensional flow of the material, while the former tends to be dominated by plane strain conditions, and the process is not intended to change material thickness, only the curvatures. In what follows, the sheet-forming processes are used as model processes, but much of what is developed applies to bulk-forming as well.

© 2002 by CRC Press LLC

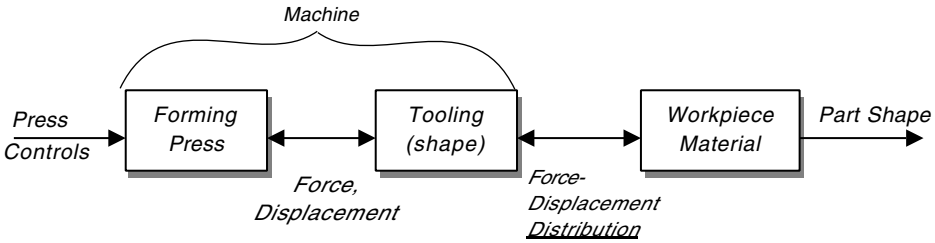


FIGURE 7.1 Basic block diagram for forming.

7.1.1 Process Control Issues

The objective of all sheet-forming processes is to alter the curvature of the material to achieve a target shape. In so doing, the material also may be intentionally stretched to aid in reducing shape errors and to induce strain hardening for strength properties. Accordingly, the control objective for the process is to achieve the desired shape, and (from a manufacturing point of view) to achieve this shape with rapid setup (flexibility) and minimal part-to-part variation (quality).

Application of control principles can have a great impact on all three: shape fidelity, variation reduction, and rapid changeover or setup. This control is accomplished either through the use of machine or process feedback to achieve higher accuracy and repeatability or by facilitating more mechanically complex machines to enhance process flexibility and control degrees of freedom. In all cases, the properties of control loops: tracking changing inputs (i.e., new part shapes), rejecting disturbances, and decreasing sensitivity to process parameter changes (e.g., tool–workpiece friction, constitutive property changes) are perfect matches to forming processes.

To help see this connection at a phenomenological level, it is useful to develop a set of block diagrams for these processes.

7.1.2 The Process: Material Diagram

A simple block diagram of the process is shown in Figure 7.1. Here the plant comprises:

- The forming machine or press, which provides the forming energy (force displacement)
- The tooling that takes this *lumped* energy and *distributes* it over the face of the tool–workpiece interface
- The workpiece material that plastically deforms according to the force or displacement field

In each block a set of constitutive properties determines how the energy or power variable pairs of each element relate to each other. For the machine blocks these properties would typically be the stiffness, mass, and damping of the machine as well as the overall geometry. For the workpiece, the set includes the large strain properties of the material and its initial geometry, which will affect how the distributed forces and displacements, and moments and curvatures are related. As will be seen, these material constitutive properties are the largest components of process variability in forming.

7.1.3 The Machine Control Diagram

In practice, the most common type of control used with forming processes is simple feedback of the machine outputs (herein referred to as *machine contro*l). As with any mechanical process, these outputs will be displacement or force, and control will involve application of servo-control technology to the actuators of the machine, whether eletrohydraulic or electromechanical. As shown schematically in Figure 7.2, closing this loop affords good regulation of these quantities, and will reject disturbances that enter the machine loop. These could include variations in the net force–displacement curve of the load (the workpiece) and variations in the machine properties such as friction and

© 2002 by CRC Press LLC

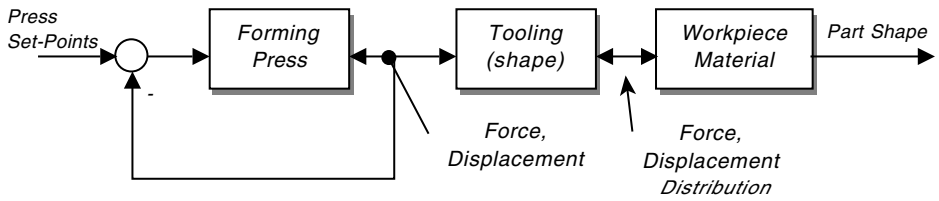


FIGURE 7.2 Closed-loop machine control for regulating force or displacement.

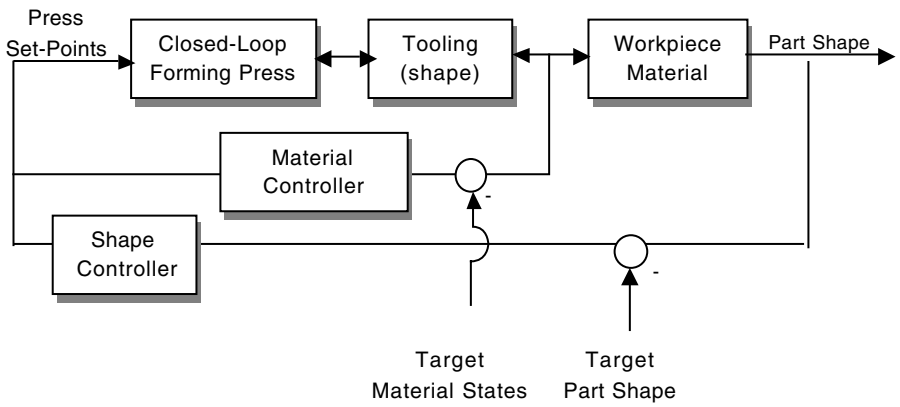


FIGURE 7.3 Material feedback and shape feedback control loops.

actuator nonlinearities and drift. It also can allow for a rapid change of set-points as production demands change. However, it cannot change the force–displacement distribution, and it leaves the part shape (which is the process output) outside the control loop.

Further stages of control can be attempted by actual measurement of forces and displacements at the tool (material control) and direct measurement of the resulting part shape (shape control). However, as shown in Figure 7.3, the only variables that can be manipulated are the press set-points, which are restricted to the limited number of actuator degrees of freedom. This, in turn, limits the process resolution, which is discussed below as the ultimate limit on process control effectiveness.

Many mechanical systems issues are involved in forming press control, but it is equally evident that even with precise control of force and displacement of the press, the resulting shape will still be a strong function of the tooling and the material itself.

To appreciate the latter aspect of forming processes it is necessary to consider the physics of forming as viewed in a control system's context.

7.2 The Plant or Load: Forming Physics

7.2.1 Mechanics of Deformation: Machine Load Dynamics

To consider the control of forming processes it is important to have at least a general understanding of the mechanics of the load as seen by a forming machine. Here a simple input–output description of forming is developed that can be shown to cover the basic phenomena of any forming process.

While a detailed model of the deformation process is well beyond the scope of this chapter, the basic phenomena of forming can be summarized by the classical unidirectional tensile stress–strain or force–displacement diagram. If we consider the simplest forming operation, that of stretching

© 2002 by CRC Press LLC

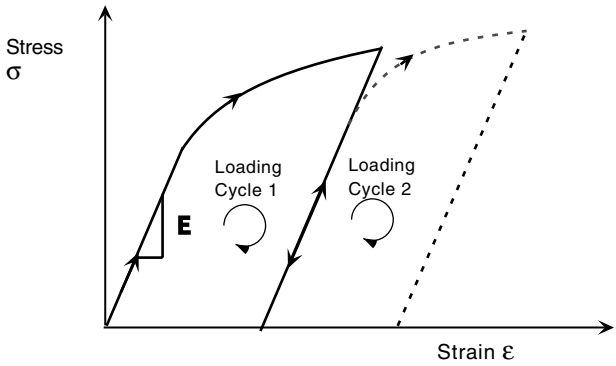


FIGURE 7.4 Cyclic loading stress–strain curve showing hysteresis and load-dependent offset.

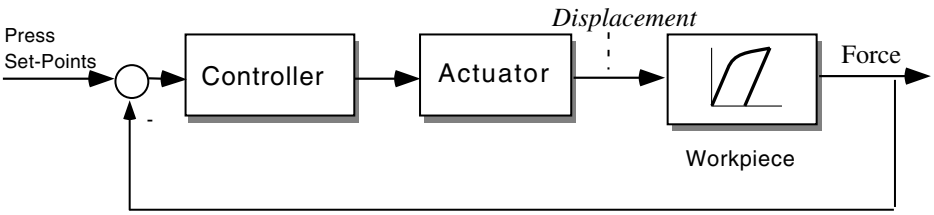


FIGURE 7.5 Simple tensile force control loop.

a bar of metal from an initial shape to a longer one, the force–displacement relationship of the workpiece is given by the constitutive stress–strain curve of the material. As shown in Figure 7.4 the curve includes not only the loading portion of the process, but also the unloading.

When looked at from a control system's perspective, the material appears to be a static block with nonlinear behavior. This arises from a power law-like plastic region, a hysteresis-like behavior arising from the elastic unloading behavior, and a history-dependent reference point owing to the permanent plastic deformation after loading beyond yield.

Because of the low mass of the material relative to the machine and tooling, the dynamics of the material block are usually ignored. However, the deformation process involves very low damping, and unless there is considerable sliding friction between the workpiece and tool, the contribution to overall system damping is minimal.

The variable slope in Figure 7.4 illustrates that if the sheet deformation process is within a control loop, the level of strain and its history can cause the gain of this element to vary widely, because the slope of the elastic region of the curve is typically more than an order of magnitude greater than the equivalent slope of the post-yield curve (the plastic modulus). Consider the impact of this on a closed-loop force controller for a simple tensile deformation. As shown in Figure 7.5, the actuator is providing a displacement output, and the tensile force generated in the material is measured and fed back to the controller. Figure 7.4 is the gain model for the workpiece block, and it indicates that the overall loop gain will be highly variable over the entire range of deformation, and will depend as well upon whether the displacement is increasing or decreasing.

7.2.2 Mechanics of Forming: Bending, Stretching, and Springback

Because all forming involves curvature change, some type of bending is always present. One of the most common and simplest forming processes is brakeforming, which is essentially three-point bending (see Figure 7.6). At any given cross-section along the arc length of the part, stress and strain distributions can be approximated by those of pure bending.

© 2002 by CRC Press LLC

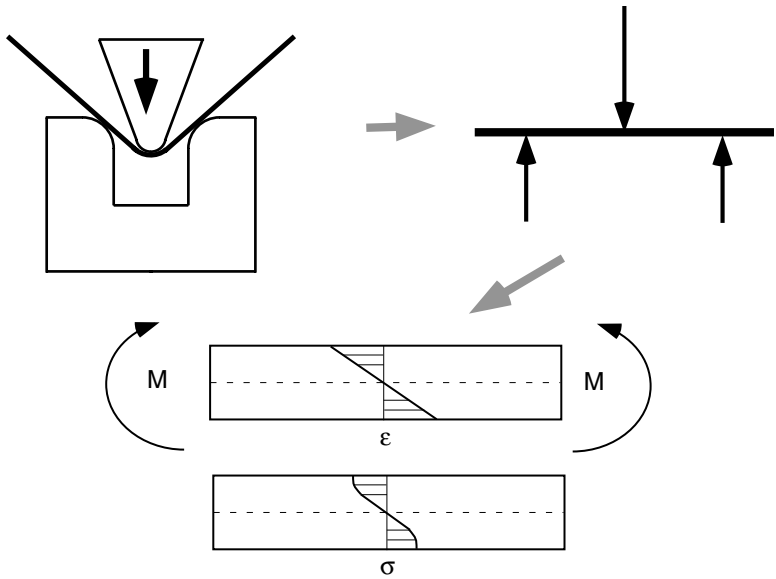


FIGURE 7.6 Simple brakeforming. Approximated as three-point bending with resulting stress and strain distributions.

With the resulting bi-directional stress distribution about the neutral axis, release of the forming loads leads to elastic unbending of the material. This curvature "springback" is the key source of error in forming processes, because it causes a difference between the curvature of the part when loaded to a known displacement and the final unloaded curvature.

To help reduce this springback and to achieve beneficial strain-hardening of the workpiece, the ends of the material are either constrained not to move or allowed to slip under a frictional force to provide an additive tensile force in the plane of the part. This process is shown in Figure 7.7 where it can be seen that the resulting stress distribution is now more uniform. As the tensile strain increases, the stress distribution becomes all positive and nearly constant. (For an idealized material that does not strain harden it will be constant.) As a result, the elastic unbending or springback of the part from the loaded curvature is greatly reduced. Consequently, for precision forming operations, or for operations where very small curvatures are involved (as with the stretch forming process used in aerospace) an intentional tensile force is added. Also, for three-dimensional forming problems, this tensile "bias" is also necessary to prevent in-plane buckling.

From the above it is obvious that for sheet forming, springback is the main source of errors, and variation in the springback will be the main source of process uncertainty. If we consider the simple bending example of Figure 7.6, the bending constitutive relationship can be written in terms of the moment–curvature relationship for the sheet. In the elastic region this is given by the simple relationship:

$$M = \frac{1}{EI} K \tag{7.1}$$

where
M = pure bending moment
K = resulting sheet curvature
E = modulus of elasticity
I = area moment of inertia for the sheet,

and for a rectangular cross-section,

© 2002 by CRC Press LLC

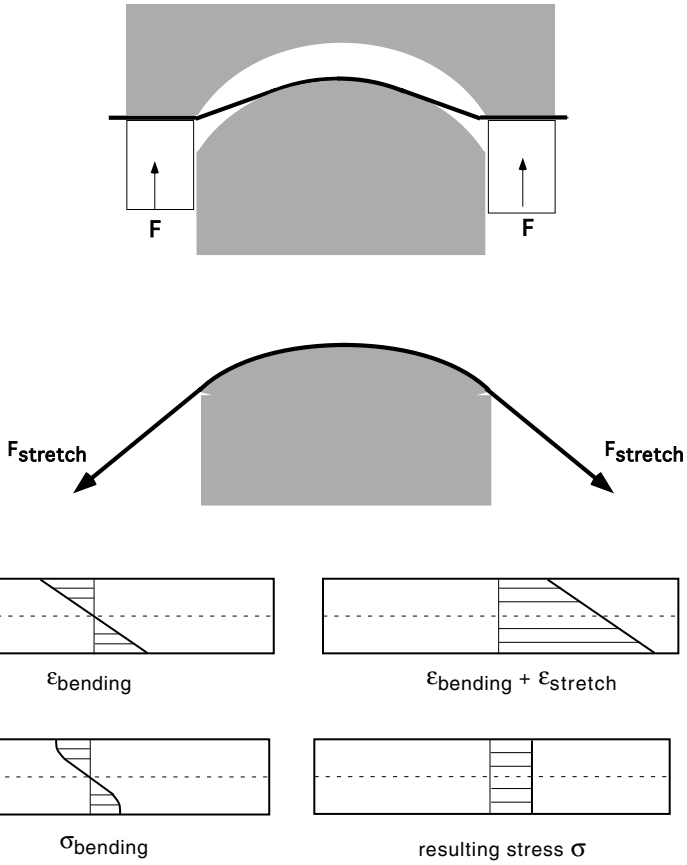


FIGURE 7.7 Simple two-dimensional draw forming with a blankholder and stretch forming. Notice the effect of adding stretch: the resulting stress distribution can become nearly uniform for a mildly strain-hardening material.

$$I = \frac{1}{12} bh^3 \tag{7.2}$$

where
b = width of the sheet
h = thickness of the sheet

As the beam curvature K increases, the bending moment will increase, and eventually the beam will begin to yield. When yielding occurs, the bending moment required for incrementally higher curvatures will decrease, and a moment–curvature relationship such as shown in Figure 7.8 will emerge. Just as with the tension example of Figure 7.4, the beam, when loaded to a maximum moment M_L, will elastically unload along a line of slope EI. The curvature springback ΔK will, as shown in the figure, be determined by the magnitude of this moment and the slope.

Consider now a very simple process where a sheet is formed between a matched set of cylindrical tools (see Figure 7.9). We are interested in the final curvature (K_U) of the part after the sheet is removed from the tools. The matched tools impose a fixed loaded curvature K_L on the sheet, which will load the sheet as shown in the figure. The amount of springback $\Delta K = K_L – K_U$ will depend on the maximum moment M_{max} and the slope EI according to

$$\Delta K = \frac{M_Y}{EI} \tag{7.3}$$

© 2002 by CRC Press LLC

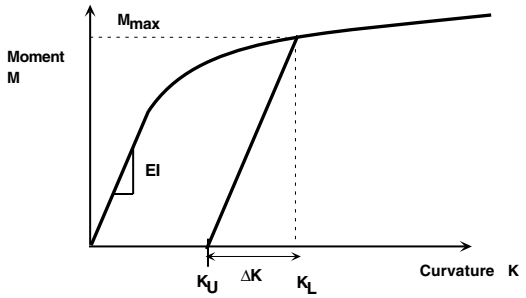


FIGURE 7.8 Generic moment curvature diagram showing curvature springback ΔK after unloading from the loaded curvature K_L.

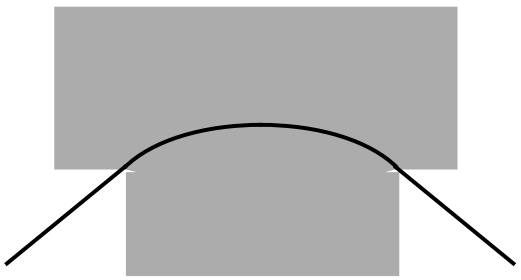


FIGURE 7.9 Simple matched tool forming over a cylinder. No edge constraint is used so the sheet sees only a bending moment if no interface friction is assumed.

Because the tooling imposes a fixed (input) curvature, the maximum moment (output) is determined by the constitutive relationship of the material, most importantly the yield stress and the thickness. The modulus E is most nearly constant, but the moment of inertia I varies with thickness to the 3rd power. Not surprisingly, in practice it is found the most sensitive parameters with respect to springback are the thickness, the yield stress, and the post-yield (strain-hardening) properties of the sheet.

7.2.2.1 Material Variations

The most common variations in sheet material are the thickness, yield stress, and plastic flow properties. The thickness can vary owing to rolling mill variations, and while some stock (such as aluminum beverage can stock) can be rolled to very low variations (~0.0002 in.), larger material can vary considerably. In some thicker material, and up into plates of thickness > 0.5 in., material specifications often call for only maintaining a minimum thickness for minimum service strength, but have a very broad tolerance on maximum thickness.

Perhaps more insidious from a process control perspective is variation of the constitutive properties of the sheet. If we imagine a linearly strain-hardening material, there are (at least) three parameters of concern: the elastic modulus E, the yield stress σ_Y, and the equivalent plastic modulus E_P. Because the modulus E depends primarily on the crystalline structure of the material, it is nearly constant for a given material independent of the particular alloy or working history. However, both σ_Y and E_P are very sensitive to the chemistry, heat-treating, and cold working history of the piece. Variations in σ_Y of up to 20% from supplier to supplier for a given alloy have been reported, although these quantities vary less within a given mill run or heat of material.

7.2.2.2 Machine Variation

Machine variations in forming are typical of most machine tools except that the loads and corresponding structural distortions are greater than most other processes. Forming loads of 10^3 or even

© 2002 by CRC Press LLC

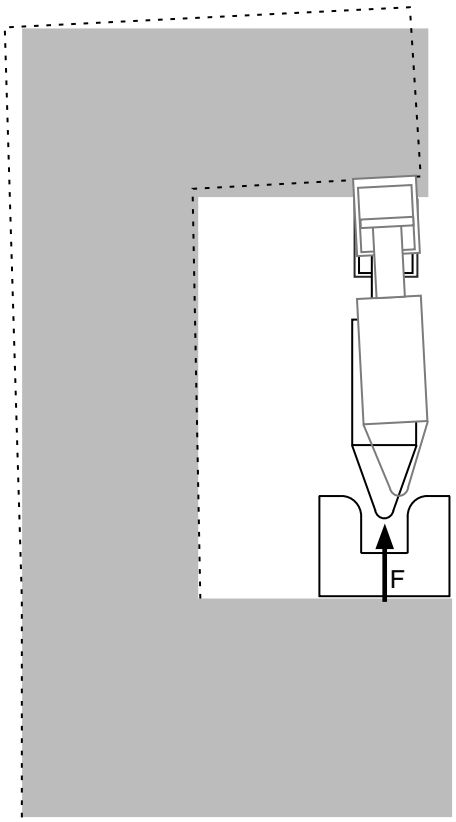


FIGURE 7.10 Simple closed-frame press shows the effect of sensor location on tool displacement control. $Y_{tooling} < A_{ctuator}$ because of stretching of the frame under the influence of the forming load F.

10^4 tons are not unusual with sheet and can be far greater for bulk forming. The elastic frames of the machine will deform with load, changing the relationship of the actuator displacements to the actual displacement of the tool–sheet interface.

Consider the situation shown in Figure 7.10. This shows the "C" frame typical of a pressbrake or stretch-forming machine. Clearly, the frame opening will stretch under load, and if the displacement sensor is collocated with the actuators, a load-dependent bias will always occur. It is also possible for the frame to bend as shown in the figure, further distorting the actuator–frame–tool geometry.

A similar collocation problem occurs with force measurement because of friction in the actuators and machine ways. If the forming force is measured at the actuator, or if as is often done, it is measured using the cylinder pressure in a hydraulic system, the actual forming force transmitted to the tooling will be attenuated by any static or sliding friction present. In general, it is wise to place the force sensor in or very near the tooling to avoid this problem.

7.2.2.3 Material Failure during Forming

In addition to controlling a process to achieve repeatable shape fidelity, it is also important that forming process control avoids situations where the workpiece will fail. Failure of sheet for bulk-forming processes is a complex phenomenon, and often failure avoidance can be no more than observing certain force or displacement limits on the machine.

Most failures occur either because of excessive tension in the sheet, causing it to tear, or excessive in-plane compression (from compound curvature shapes) which causes the sheet to wrinkle if unrestrained. Both forms of failure are difficult to detect. Tearing is preceded by localization of

© 2002 by CRC Press LLC

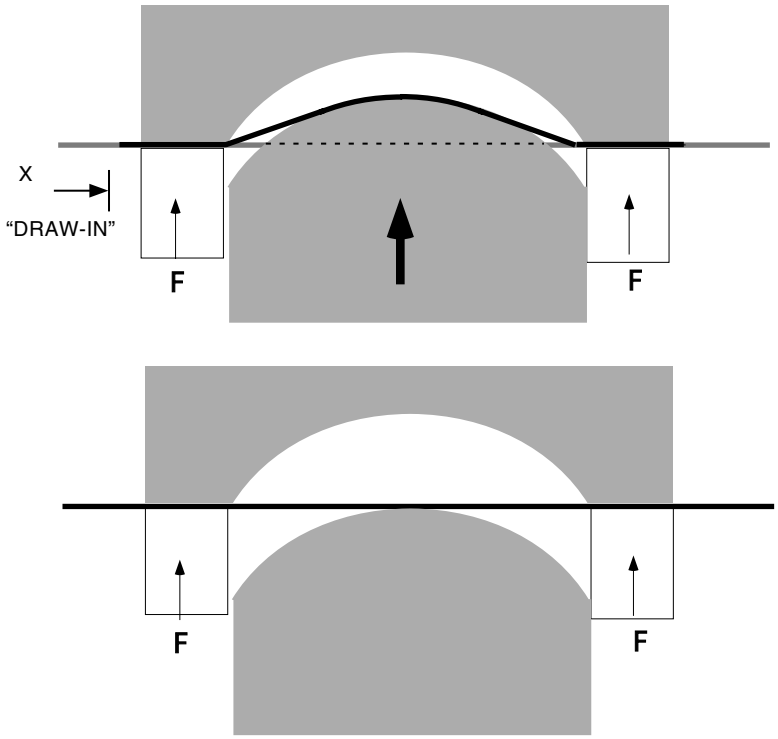


FIGURE 7.11 Simple draw forming with a frictional blankholder. As the tools move together, the sheet is drawn in an amount Δx.

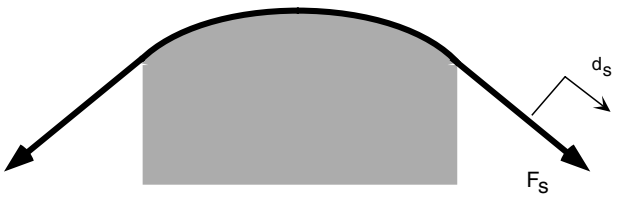


FIGURE 7.12 A stretch-forming process instrumented to measure force and displacement of the sheet during forming.

strain with attendant local thinning, and failure then occurs because of the resulting stress concentration. Wrinkling or buckling failure is even subtler because it often shows no detectable change in the force–displacement characteristics of the process. Instead, it can be thought of as an uncontrolled material flow (bucking) out of plane caused by in-plane compressive forces.

Active control to avoid failure is a complex topic both with respect to the mechanics of failure[1] and use of control to avoid these limits.[2–4] However, we can consider a simple example, that of stretch forming as shown in Figure 7.12. Here the stretch actuators are monitoring force (F_s) and displacement (d_s). As the process progresses, the resulting F–d curve for the actuators mimics the stress–strain characteristics of the sheet. By watching this curve develop, it is possible to determine the state of deformation and, for example, discover how close one is to the ultimate tensile strength of the material. In a more general case, the F–d data can be used as a process signature for which nominal trajectories are determined. Then, variations from these trajectories can be used to diagnose incipient failure.

© 2002 by CRC Press LLC

In some processes, such as the draw forming commonly used in automobile part production and in aerospace stretch forming, it is possible to measure the strain of the material directly using surface mounted gauges,[5] or by measuring the movement of the edge of the sheet as it is drawn into the tool.[6] In either case, the strain in the sheet can be used to estimate proximity to failure limits and control the process accordingly.

7.3 Machine Control

Historically, forming machines were used as a purely mechanical means to provide the large forces necessary, whether by using a slider crank or knuckle-type mechanism, or even more crudely, using high-momentum drop presses, to create the forming forces. However, with the advent of low-cost servo-control technology, most presses are now controlled by either motor-driven high-load leadscrews, or direct-acting linear hydraulic actuators with proportional servo valves.

The motor-driven leadscrews have the advantage of being mechanically simple, quieter, and often less expensive than hydraulics. In addition, the leadscrew, if the pitch is high enough, can isolate the actuator from the forming load in such a way as to nearly decouple the actuator dynamics from that of the load. However, leadscrew systems are typically limited to lower loads, owing to limits of the screw threads and nuts, and to lower velocities owing to the high pitches and wear on heavily loaded screw surfaces. Therefore, the vast majority of modern forming machines are hydraulically actuated and use either proportional servo-control of the actuators or a simple form of on–off control.

7.3.1 Sensors

As discussed above, there are many opportunities to measure either the forming machine or the workpiece itself. Because the most important constitutive relationship to forming is stress–strain or force–displacement, the latter two quantities are most often measured. In general, it is most practical to locate such measurements on the machine itself, independent of any part-specific tooling and the workpiece. However, as shown in Figure 7.10, it is always preferable to locate sensors as near to the workpiece as possible to mitigate the effects of machine distortion.

7.3.1.1 On Machine

For hydraulically actuated machines, the pressure in the cylinders can be measured and used as a surrogate force measurement if the cylinder area is known. For double-acting cylinders this area will be different depending upon the movement direction, and the cylinder seal friction as well as machine-bearing friction will add errors to this measurement. Load cells can be located either near the actuator–tool interface or in the machine frame itself. The cell must not add too significantly to machine compliance but must be sensitive enough to give useful force resolution over a large range for forces.

Displacements are most typically measured using cable-connected rotary sequential encoders. This allows for remote location of the encoder, and the cable can be stretched over long distances to ensure the correct displacement is measured. Such encoders commonly have resolutions far better than 0.001" and are noise free (except for quantization errors at very low displacements). The major design concern is that the cable be protected if it is near the forming region.

7.3.1.2 On Sheet

The ideal feedback measurement for forming would be the stress and strain fields throughout the sheet, preferably on each surface. With this information the local springback could be determined and failure prevented. Unfortunately, in-process measurements of stresses and strains are impractical. However, certain strains and correlates to strain can be measured. For example, in processes where substantial sections of the material remain free of surface pressures, optical or mechanical strain measurement devices could be inserted. Again, in practice, this has limited viability, but some

© 2002 by CRC Press LLC

examples have been tested in the aerospace industry[5] using surface mounted linear variable differential transducers (LVDTs). Optical measurement of surface strains is done regularly in material testing using video capture and measurement of circle grids on the surface of the sheet,[7] but it has not been used in volume production In this case, the surface strains can be used to directly control the extent of forming and, as was discussed in the earlier section on the process mechanics, controlling strains instead of stresses leads to a far more robust process.

In the draw-forming process, like that shown in Figure 7.11, the sheet is pulled against the frictional blankholder as the punch ascends into the die. The edge displacement of this sheet can be measured at one or more places, and if combined with knowledge of the punch displacement, can be used as an indirect indicator of strain.[2,6] However, for all but the simplest geometries this estimate will be crude at best. This measurement can be accomplished again with LVDTs but they are difficult to protect in the industrial environment. Instead, optical methods are preferred, though none are in practice at this time.

7.3.1.3 On Final Part

The ultimate measurement for control of forming processes is the actual final contour of the part. This allows full closure of the process loop as shown in Figure 7.3. All of the disturbances that enter the system, including material variations, press variations, and even machine controller variations (provided they are not entirely uncorrelated random signals) will be reflected in this measurement. However, such measurements have yet to be practical on an in-process basis, and are at best limited to use after the actual forming is complete. In addition, if complete part shapes are required, three-dimensional surface measurements are very time consuming, and can often take 10 to 100 times longer than the actual part processing time. This extended delay makes such measurements useless for in-process control, and they are better used for process diagnosis or some form of statistical process control.

New optical methods are under development[8] that may allow immediate post-process measurement, and with this innovation the delay may be short enough to allow effective part-to-part compensation. However, even if the measurement is made, for a general three-dimensional case the issue of limited control degrees of freedom or process resolution limits confounds full implementation of such a scheme.

7.4 Machine Control: Force or Displacement?

Each actuator in a forming machine can be placed rather easily under force or displacement feedback control. The design question then becomes: which is best? Of course, the answer depends upon the details of the process at hand, but there are some general observations that can be helpful in approaching this problem.

Consider the typical stress–strain curve in Figure 7.13. The implications of this curve are that at high strains (typical of forming) large variations in displacement cause small changes in force, and conversely, small variations in stress cause large variations in strain. This implies that we can most accurately relate both springback and incipient failure to strain, and it suggests that it is most logical to control displacement if given the choice. In addition, if the properties of the material change as shown in Figure 7.13, controlling the strain (displacement) would also be less sensitive to this variation than controlling the stress (force).

Indeed, it is best to control the true sheet strain if possible, but as discussed above it is usually not feasible. The substitute is to control displacement of the tooling and try to relate that to strain. Herein lie several problems. First of all, the single lumped machine displacement variable must be related to a specific point strain, and on complex three-dimensional parts, the strain field can be highly varied. Second, the machine will always have uncertainties caused by both the frame deflections mentioned earlier and by mechanical backlash in the frame and actuators. Third, in processes such as stretch forming, the sheet is loaded manually and the force–displacement "zero

© 2002 by CRC Press LLC

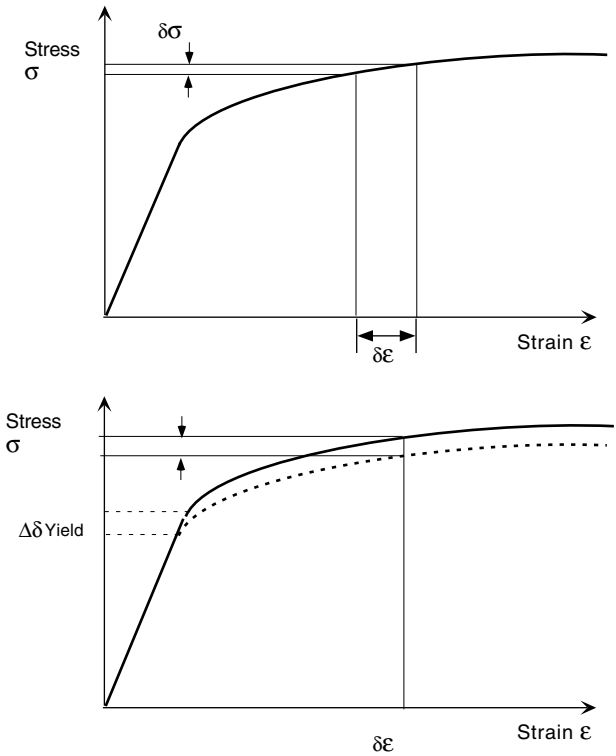


FIGURE 7.13 Stress–strain sensitivity at high strains.

point" may be highly variable. For these reasons alone displacement control is prone to large errors, despite its apparent robustness with respect to force and material property variations.

However, when looking at specific classes of processes, the question becomes a bit easier to answer. For the brakeforming process shown in Figure 7.6, none of the above concerns is present, and indeed all such machines are displacement controlled to give a more robust performance when material properties change. However, from the geometry in Figure 7.6 it should be apparent that changes in the thickness of the material introduce a displacement bias. (A novel method for in-process determination of the thickness is possible using both force and displacement measurements. By tracking the initial F–d curve, the actual zero point can be extrapolated from the data and used to determine appropriate command bias.)

In contrast to brakeforming, consider again the matched tool-forming process shown in Figure 7.9. In this case, displacement control would be very dangerous if the exact thickness of the material is unknown or the tooling locations had some uncertainty. In simple terms, the problem is between the extremes of never fully forming the part or bottoming the tooling and creating excessive tool surface forces. Therefore, for this process, active force control or displacement control into a compliant cushion (effectively a form of force control) is preferred.

7.5 Process Resolution Issues: Limits to Process Control

If we consider controlling part shape to be the ultimate goal of our process, then it is important to evaluate the ultimate ability of the process to vary the shape under some form of process control. This requires that the resolution of the process — the relationship between the actuator degrees of freedom and the degrees of freedom required by the part shape — be determined.

© 2002 by CRC Press LLC

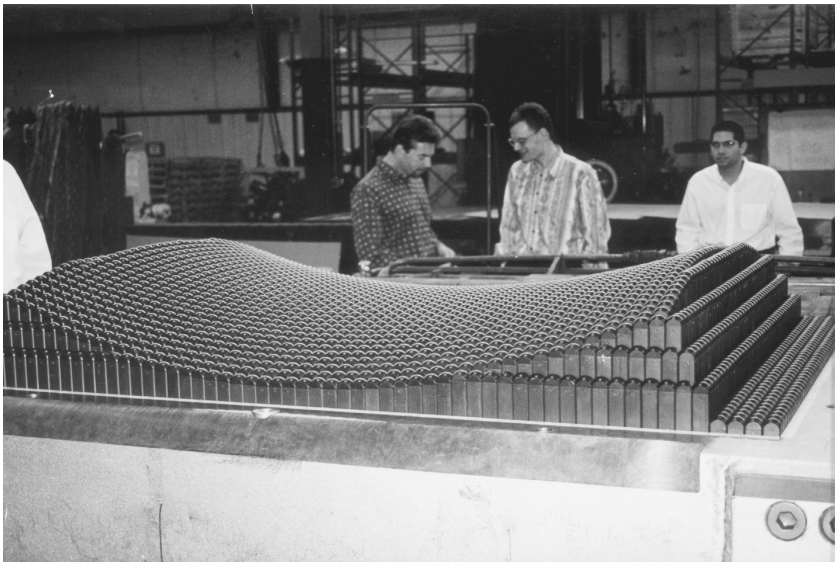


FIGURE 7.14 Photo of a prototype reconfigurable stretch-forming tool. The tool is comprised of > 2600 individual servo-driven pins with spherical ends.

There is a natural diffusion of the lumped forming energy provided by the tooling to the inherently distributed energy necessary to create a general three-dimensional deformation. This physical fact emphasizes the most important control impediment in forming processes. Because control is most easily exerted on lumped power variables on the machine (e.g., actuator forces or velocities/displacements) the effect of this control is diffused over the entire workpiece by the tooling. As a result, the effect of the lumped controls on the final part shape is indeterminate and well outside the control loop. Instead, the control system is merely providing a highly consistent level of bulk energy to the tool, which will, in turn, distribute the energy according to the local constitute relationships of the tool and workpiece. The only solution to this dilemma is to add the energy distributor degrees of freedom (the tool) to the control system. This can be done only by adding spatial degrees of freedom such as programmable or movable die surfaces, or by taking three-dimensional parallel forming processes and doing them in a series of two-dimensional stages. The former has been accomplished, for example, by using discrete tools whose elements can be moved in real time, and the latter is exemplified by processes such as roll forming.

7.5.1 Process Resolution Enhancement

It is worthwhile to close with some leading edge examples of how control can be extended beyond the classical machine servo controllers commonly found on production machinery to include some reflection of the sheet-forming process itself. Perhaps the two most interesting examples are attempts to control the strain in a complex three-dimensional draw-forming process and attempts to use a tool whose shape can be rapidly reprogrammed between forming cycles.

The process resolution discussion makes it clear that the main degrees of freedom with respect to part shape are contained in the tool shape itself. If the tool can be changed only by actual addition or subtraction of material, this can hardly be called process control. However, if the tool surface is in some way programmable, then the process resolution can be greatly increased. An example of such a tool[10–12] is shown in Figure 7.14 where the tool surface is comprised of many individually controllable "pins" that form a discrete surface. This surface is then smoothed by a polymer pad and can be used to form commercially acceptable parts.

© 2002 by CRC Press LLC

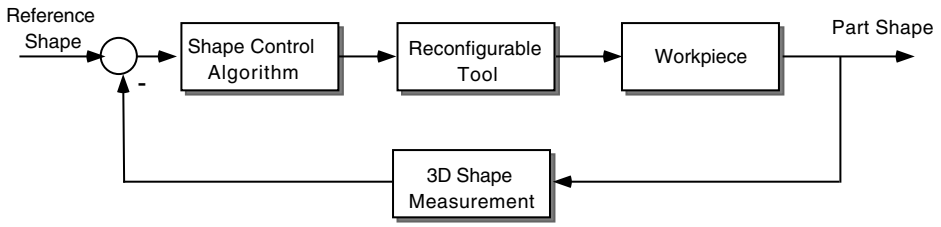


FIGURE 7.15 Shape control system using a reconfigurable tool and spatial frequency controller.

Other forms of resolution enhancement have been proposed. These include a sheet blankholder (see Figure 7.7) that is either broken into independently controllable segments so that the frictional restraining force can have several discrete values around the periphery of the sheet, or a deformable blankholder with variable displacement supports[10] that allow a continuously variable (but spatially band-limited) blankholder pressure distribution.

7.6 Direct Shape Feedback and Control

The special case shown in Figure 7.3 of direct feedback of part shape has recently found pre-commercial application to stretch forming in the aerospace industry.[11] In this system the reconfigurable tool of Figure 7.14 is combined with a novel three-dimensional shape-sensing device and a spatial frequency-based control law[11–13] to actuate the tool until shape errors are minimized (see Figure 7.15). The actual control system has a minimum one forming cycle delay built in because the part cannot be measured until after forming.

7.7 Summary

Control of metal-forming processes has advanced considerably with the advent of inexpensive computer servo controls. However, the inherent sensitivity of the process to variations in the constitutive properties of the workpiece materials prevents simple servo control of machine variables from fully controlling the process output. Such control does, however, greatly reduce the process variability, and with good production control of material and proper maintenance of the machine and tooling, highly consistent and accurate parts can be produced at high rates. To move to the next level of control where either the strains or final shapes are actively controlled involves a large jump in sensing, actuation, and control law technology that has yet to emerge on the production floor.

References

1. Cao, J. and Boyce, M. C., A predictive tool for delaying wrinkling and tearing failures in sheet metal forming, *Journal of Engineering Materials and Technology (Transactions of the ASME)* (U.S.A.), 119(4), 354–365, October 1997.
2. Jalkh, P., Cao, J., Hardt, D., and Boyce, M. C., Optimal forming of aluminum 20008-T4 conical cups using force trajectory control, *Society of Automotive Engineers International,* 11, 1993.
3. Traversin, M. and Kergen, R., Closed-loop control of the blankholder force in deep-drawing: Finite-element modeling of its effects and advantages, *Journal of Materials Processing Technology,* 50(1–4), 306–317, 1 March 1995.
4. Majlessi, S. A. and Obermeyer, E. J., A review of recent advances in the application of blankholder force towards improving the forming limits of sheet metal parts, *Journal of Materials Processing Technology*, 75(1), 222–234, 1998.
5. Parris, A. N., Precision Stretch Forming for Precision Assembly, Ph.D. thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1996.

© 2002 by CRC Press LLC

6. Fenn, R. and Hardt, D. E., Real-time control of sheet stability during forming, *ASME Journal of Engineering for Industry*, December 1993.
7. Manthey, D.W. and D. Lee, Recent developments in a vision-based surface strain measurement system, *Journal of Metals*, 23(10), 46–49, July 1995.
8. Mermelstein, M. B., Feldkun, D. L., and Shirley, L. G., Video rate surface profiling with acousto-optic accordian fringe interferometry, *Journal of Optical Engineering,* 39(1), 106–113, 2000.
9. West, J. S., Adaptive Stroke Reversal Control in Brakeforming, SM thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1980.
10. Siegert, K., Hohnhaus, J., and Wagner, S., Combination of hydraulic multipoint cushion system and segment-elastic blankholders, *Developments in Sheet Metal Stamping, Proceedings of the 1998 SAE International Congress & Exposition,* February 23–26, 1998, Detroit, MI.
11. Valjavec, M. and Hardt, D. E., Closed-loop shape control of stretch formed sheet metal parts using a reconfigurable discrete die press, Sheet Metal 1998*, Proceedings of the 6th International Conference on Sheet Metal*. Twente, The Netherlands, April 1998.
12. Hardt, D. E. and Webb R. D., A transfer function description of sheet metal forming for process control, *Transactions of the ASME,* 113, 44–52, 1991.
13. Boyce, M. C., Hardt, D. E., Ousterhout, K. B., Karafillis, A., and Eigen, G., A CAD driven flexible forming system for three-dimensional sheet metal parts, *Proceedings of the SAE Congress, Symposium on Sheet Forming*, Detroit, March 1993.

© 2002 by CRC Press LLC

8

Assembly and Welding Processes and Their Monitoring and Control

8.1 Assembly Processes
Monitoring of KPCs • Monitoring of KCCs
8.2 Monitoring and Control of Resistance Welding Process
Monitoring • Control
8.3 Monitoring and Control of Arc Welding Processes
Modeling for Arc Length Control • Weld Bead Geometry Control • Weld Material Properties • Monitoring of Arc Welding and Laser Welding

S. Jack Hu
University of Michigan

Elijah Kannatey-Asibu, Jr.
University of Michigan

Assembly is a very important part of most product realization processes. Components fabricated through machining, forming, etc. will be assembled together to form higher level of assemblies or the final products. An assembly process generally includes part positioning (or mating) followed by part joining. Part positioning can be accomplished using fixtures or robots. Part joining methods include mechanical fasteners, shrink and expansion fits, welding, and adhesives. Because an assembly process is the place where quality variation from the individual components could accumulate, it is critical to monitor and diagnose assembly and joining problems quickly and effectively.

This chapter provides an overview of various approaches available for monitoring assembly and joining processes, in particular, resistance spot welding and arc welding processes; Section 8.1 describes techniques in the monitoring of assembly processes using examples from automotive body assembly processes; Section 8.2 describes the monitoring and control of resistance spot-welding processes; and Section 8.3 presents techniques in the monitoring and control of gas metal arc welding processes.

8.1 Assembly Processes

There are two types of assembly processes (Mantripragada, 1998). Type I assemblies are comprised of machined or molded parts that have their matting features fully defined by their respective fabrication processes prior to assembly, for example, the insertion of a peg into a hole. Mating of part features is the main function of the assembly process. Type II assemblies are those where some or all of the assembly features and/or their relative locations are defined during assembly. These types of assembly processes include, for example, automotive and aircraft body assemblies where part mating is accomplished using fixtures during the assembly process.

© 2002 by CRC Press LLC

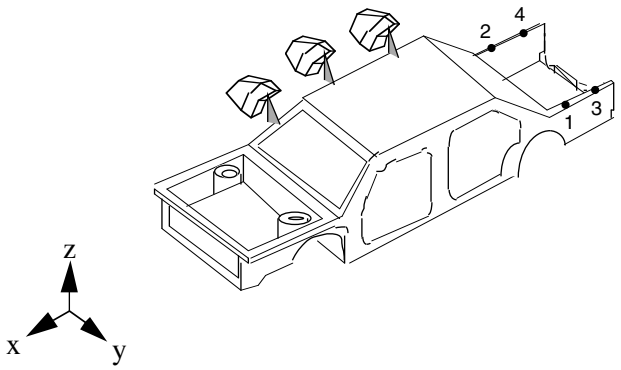


FIGURE 8.1 A schematic of an optical coordinate measuring machine checking body dimensions.

Monitoring of an assembly process can be accomplished by either directly monitoring the quality characteristics of the assembled products (i.e., key product characteristics or KPCs), or monitoring the processes characteristics that control the assembly process (key control characteristics or KCCs), i.e., fixtures and welding machines. Examples of KPC monitoring include inspection of an assembly on coordinate measuring machines. In automotive body assembly, the KPCs in a car body are the sizes and shapes of the openings. Figure 8.1 shows schematically an in-line optical coordinate measuring machine that is checking the dimensions of a car body assembly.

8.1.1 Monitoring of KPCs

In automotive body assembly, the critical KPCs are the sizes and shapes of the body openings, e.g., doors, trunk opening, etc. Their sizes and shapes influence the downstream panel fitting processes, which, in turn, influence the quality and functionality of the final vehicle. For example, width and straightness are the critical product characteristics for the trunk opening. The indices for the width and straightness of the decklid opening are defined as (Roan and Hu, 1994):

$$I_1 = y_1 + y_2,\ I_2 = y_3 + y_4$$

$$I_3 = y_1 - y_3,\ I_4 = y_2 - y_4$$

where $I_1$and I_2 are width indices, I_3 and I_4 are straightness indices, and y_is are the measured deviations from design nominal dimensions. Because multiple product characteristics are to be monitored at the same time, the simultaneous confidence interval (Johnson & Wichern, 1992) approach can be used to establish control limits for the KPCs.

8.1.2 Monitoring of KCCs

As mentioned before, an assembly process can be monitored using the key control characteristics, such as the fixturing and joining processes. Monitoring the torque in a fastening operation provides such a direct approach to assembly monitoring. However, there are situations in which process measurements are not readily available. In such a case, when only the product characteristics are measured, various transformation techniques can be used to relate KPCs to KCCs. For example, principal component analysis can be used to relate dimensional measurements on automotive bodies to various fixturing faults (Hu and Wu, 1992; Ceglarek and Shi, 1996), then process monitoring can be accomplished using the resulting principal components.

The basic idea behind principal component analysis is to find the interrelationship between variables by taking the combination of them to produce uncorrelated variables. The principal components, z_i, are represented as linear combinations of the n original correlated variables, y_i, as

© 2002 by CRC Press LLC

$$\begin{Bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{Bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ . & . & \cdots & . \\ a_{n1} & & & a_{nn} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{Bmatrix}$$

where the a_{ij} are the j-th elements of the i-th eigenvectors of the covariance matrix C of the original correlated variable y_i.

An example of assembly monitoring using principal components is shown in Figure 8.2. Here measurements are made on the cross-car deviation of the roof after assembly. Figure 8.2(a) shows these dimensions. Figure 8.2(b) shows the principal components, z_i's. Because z_i's are not correlated with each other, standard process control charts, such as x-bar and R charts, can be used as tools for monitoring (DeVor et al., 1992).

8.2 Monitoring and Control of Resistance Welding Process

The resistance welding process is a very popular joining technique used in the manufacture of such items as automobiles, furniture, and appliances. For example, in a typical steel auto body, there are from 3000 to 5000 weld spots. Because of the extensive use of resistance spot welding, even a small improvement would bring significant economic benefits. This potential payoff has attracted a significant amount of research in both the resistance spot-welding field in general and the specific field of resistance spot-welding monitoring and control.

Resistance welding is the process of welding two or more metal parts together in a localized area by applying heat and pressure. The heat is provided by the resistance furnished by the metal parts to the flow of current through the electrode tips. The pressure is also provided by these same electrodes through pneumatic cylinders or servo drives. The schematics of a resistance welding machine are shown in Figure 8.3.

Many models of resistance spot welding were based on two coupled partial differential equations (Matushita, 1993): an electrical equation

$$\nabla \cdot \left(1/\rho_1 \nabla V \right) = 0$$

and a thermal equation

$$C\sigma \frac{\partial T}{\partial t} = \nabla \cdot (K \nabla T) + \rho_2 \delta^2$$

where $\rho 1$ is the electrical resistivity of the workpiece, V is the electrical potential, K is the thermal conductivity, ∇ is the gradient, C is the specific heat, σ is the workpiece mass density, and δ is the current density. To handle the complexity of solving these partial differential equations, most researchers have resorted to finite difference methods or finite elements methods. Unfortunately, these models and methods are not computable on-line, therefore, not suitable for on-line monitoring and control.

The difficulty of generating simple dynamic models from the first principles has led researchers to use ad hoc techniques for monitoring and control. Because weld quality, whether defined as a weld attribute such as butt diameters from peel test, or strength, such as tensile strength of the weld, is not directly measurable, identifying variables with a high correlation with nugget size would be desirable. Variables studied so far include thermal emission, ultrasound, acoustic emission, thermal expansion, temperature, voltage, current, energy, resistance, force, and residual stress. The most commonly used variables are current (I), dynamic resistance (DR), and electrode displacement (D).

© 2002 by CRC Press LLC

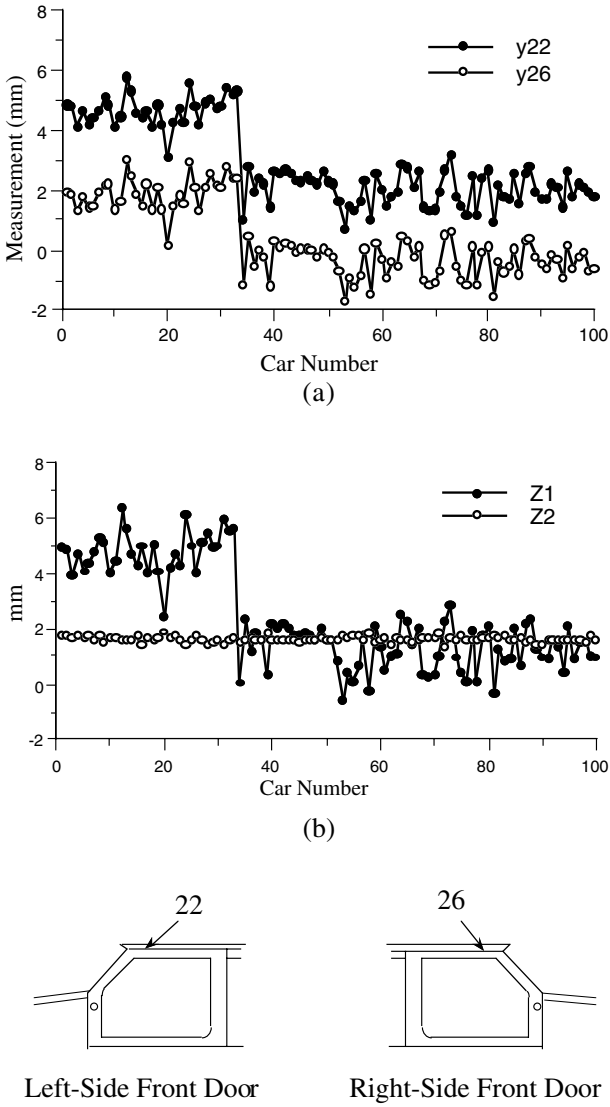


FIGURE 8.2 Monitoring of principal components.

8.2.1 Monitoring

The possible importance of electrode head displacement was recognized early in a 1942 U.K. patent. Waller (1964) reasoned that weld quality was related to maximum displacement and thus took maximum displacement as a sign of weld quality. Needham proposed a controller that shuts off the current when the weld displacement reaches approximately 80% of a predetermined maximum value. In other words, it is a closed-loop weld schedule around the displacement measurement. Jantoa (1975) suggested using a zero rate of expansion as the signal that a complete weld had been made. Kuchar et al. (1982) use a finite element model (FEM) model to create ideal electrode displacement curves and then design a classical controller to track them. After this, several research groups (Cho et al., 1985, Wood et al., 1985, Chang et al., 1989) also studied tracking control of displacement signals. Adaptive control techniques have also been studied (Chang et al., 1989, Haefner et al., 1991).

A displacement curve as shown in Figure 8.4 has been suggested by various researchers (Gedeon et al., 1987). Here the displacement curve is divided into different regions and process monitoring

© 2002 by CRC Press LLC

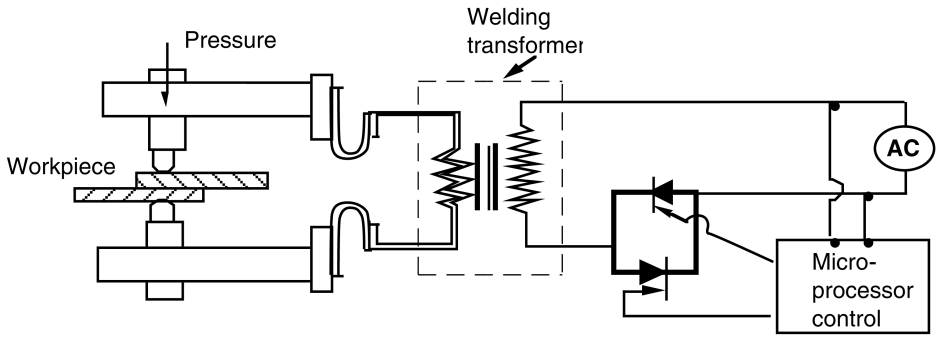


FIGURE 8.3 Resistance welding process.

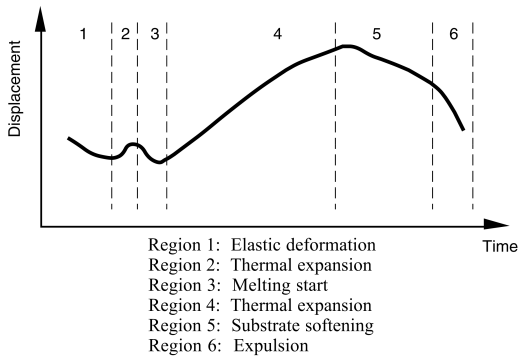


FIGURE 8.4 Monitoring of resistance welding process using electrode displacement.

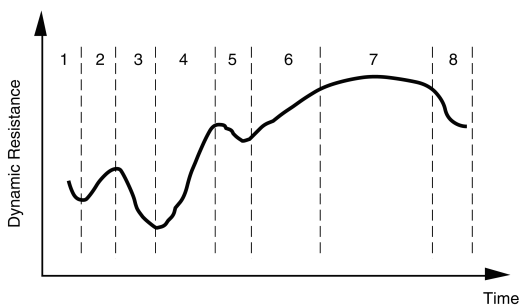


FIGURE 8.5 Monitoring of resistance welding process using dynamic resistance.

is accomplished by detecting changes of the curve from region to region. However, the magnitude of the displacement curve will be modulated by machine stiffness and weld force. Therefore, there is no ideal displacement curve unless the welding force is maintained at a constant level and the curve is calibrated for each machine.

The rationale behind using dynamic resistance as a feedback signal has taken a very similar approach to that of electrode displacement. The dynamic resistance curves provide excellent information and were believed to be much easier to instrument than force or displacement (Figure 8.5). However, for coated steels, it was difficult to relate dynamic resistance with nugget information. One of the early dynamic resistance-based controllers was presented by Towey (1968).

© 2002 by CRC Press LLC

The idea was that the resistance drop was related to the size of the nugget and thus, by looking for a predetermined resistance drop, they could get the desired size nugget. Dickinson et al. (1980) divided the dynamic resistance curve into the following stages: surface breakdown, asperity collapse, heating of the workpieces, molten nugget formation, nugget growth, and mechanical collapse. In 1987, Gould found that neither poor fit-up nor use of sealer at the faying surface adversely affected the resistance-based control algorithms.

Monitoring systems based on other indirect signals also have been developed. For example, one of the earliest acoustic/ultrasonic monitoring systems was devised by Burbank et al. in 1965. Vahavilos (1981) studied acoustic emission as a feedback signal for weld quality control. While good performance was claimed, this controller appears to have been unsuccessful in production environments. The biggest obstacles seem to be the availability of sensors suitable for a shop-floor environment, and lack of a real-time signal-processing device that can handle the huge amount of data coming from the sensors.

Currently, process monitoring for resistance spot welding has focused on a multivariate approach. For example, Hao, Osman, Boomer, and Newton studied the characterization of resistance spot welding of aluminum. Both single-phase alternating current (AC) and medium-frequency direct current (MFDC) are used. From the recorded weld data file, a large number of features are extracted to monitor the nugget growth. Li et al. (1998) used principal component analysis to extract features and then neural networks to classify fault and predict nugget growth.

8.2.2 Control

Two major difficulties exist with spot-welding control: First, there is no direct way to sense nugget diameter (or strength) in real time. All the variables that can be sensed in real time have been shown to be at best weakly linked to nugget diameter and strength. Many of the available sensors are also found to be unsuitable under a production environment. Second, a sufficiently good model of the process, in a form useful for control design, is difficult to develop.

To circumvent the first difficulty, two control approaches are usually taken: (1) open-loop control (weld schedule, table lookup); and (2) feedback and control of indirect welding variables such as current, displacement, force, acoustic emission, etc. In the first approach, the system is vulnerable to any external disturbances (e.g., power fluctuation, poor fit-up, etc.). In the second approach, the system is vulnerable to any external disturbances whose effect on nugget size/strength is undetectable from the feedback signal. The second approach seems to be more promising for generating consistent welds if we can identify the right signal/sensor to close the loop.

Current was used in the earliest attempts as a signal for resistance spot welding (RSW) control for two main reasons: First, there is a close relationship between current and total energy input to the welding process. Second, current is directly controllable and is often used as the control input. The assumption behind current control is that if the resistance across the two electrodes is constant, then controlling electrical current (I) will provide direct control of the heat generated. Later on, it was realized that resistance between electrodes (R) is not constant (it changes with temperature, pressure, etc.). Variation to current control was adapted. For example, current density (current divided by electrode face area) was attempted to compensate for electrode wear. As an electrode wears, a current stepper in the weld control system will increase the current to try to maintain constant current density.

The paper by Kuchar (1982) discusses a closed-loop multivariable control system using an axisymmetric finite element model. The outputs from the FEM model are predicted nugget size and corresponding electrode displacement for quality welds. Measured electrode displacement is then compared with the ideal displacement curve and the error is used for feedback control. The controller adjusts the electrode force, current, and voltage to bring the actual displacement close to the ideal displacement curve. Tsai et al. (1991) also studied the correlation between the expansion displacements among the electrodes during welding to the weld nugget quality.

© 2002 by CRC Press LLC

Haefner, Carey, Bernstein, Overton, and D'Andrea (Haefner et al., 1991) developed a system incorporating adaptive control technology for the process. This paper relates thermal growth to nugget formation by deriving the thermal growth from the electrode displacement measurement. This real-time adaptive strategy adjusts for long-term electrode wear and provides a short-term weld-to-weld control to compensate for fit-up and surface oxide variations. Schumacher et al. (1984) developed an adaptive control system that could weld different low-carbon and high-strength steels, or a series of different welds in the same steel.

Recently, the research focus on spot-welding control seems to have shifted toward intelligent control, or more specifically, neural network/fuzzy logic/expert system-based control systems. One of the unique features of these systems, compared with traditional control design methods, is that they generally do not require an explicit system model, and the control algorithm can be based on rules or other forms of knowledge. Examples include Jou et al. (1994) and Shriver et al. (1998). Because these techniques are relatively new, most of the proposed methods were not implemented as control algorithms. They either involve proof-of-concept type of study, or are designed to generate weld parameter suggestions, instead of controlling the weld process directly.

8.3 Monitoring and Control of Arc Welding Processes

Welding processes often encounter disturbances that effectively change the process outputs, resulting in a weld of undesirable characteristics. Such disturbances may include thermal distortion, workpiece fit-up, geometrical variations in workpieces, robot motion errors, and the effects of fixturing equipment. To achieve the desired weld characteristics while the process is subjected to disturbances, it is necessary to use feedback control. The three principal stages of process control involve modeling, sensing, and control (Cook et al., 1989; Kannatey-Asibu, Jr., 1997).

At the core of feedback control are the process inputs and outputs. The primary inputs in the case of gas metal arc welding, for example, are the arc current/arc voltage, traverse velocity (welding speed), and electrode wire feed rate (Cook, 1980; Dornfeld et al., 1982). The secondary inputs include shielding gas flow, torch positioning and orientation, torch weaving or oscillation, and mode of metal transfer. Non-manipulatable inputs include workpiece and electrode material properties, workpiece geometry, and joint configuration. The primary outputs are usually difficult to measure in real time, i.e., while the process is going on, and without destroying the part, while the secondary outputs are more easily measured on-line, but not after the process. The primary outputs include penetration, bead width, reinforcement (collectively, the bead cross-sectional area), hardness, strength, microstructure, residual stresses, and discontinuities (cracks, inclusions, porosity, etc.). The secondary outputs include peak temperatures (temperature distribution), cooling rate, arc length, acoustic emission, arc geometry, arc motion, and pool motion.

In this section, we focus on modeling and sensing of arc welding processes for control, even though control schemes are discussed in other chapters, and with specific emphasis on welding processes in Cook (1989), Suzuki et al. (1991), and Tomizuka et al. (1980). The discussion starts with modeling for feedback control of arc length followed by models for control of weld bead geometry and weld material properties. Various techniques for monitoring the welding process are then outlined.

8.3.1 Modeling for Arc Length Control

Control of arc length is useful for wire feed welding systems such as gas metal arc welding. Arc length variations for these systems can result from variations in power line voltage, groove geometry, etc. and can affect porosity and other forms of discontinuity. Feedback control of arc length using wire feed as input normally involves a constant current power source. With such a power source, the system is not self-regulatory, and therefore significant variations in arc length can occur unless it is under closed-loop control.

© 2002 by CRC Press LLC

The simplest model of arc length dynamics describing the characteristics of the gas metal arc welding system is based on the assumption that the rate of correction of the welding wire tip is proportional to displacement from its equilibrium position or operating point. In other words, the rate of change of arc length is proportional to the change in arc length and is expressed (Muller, Greene, Rothschild, 1951) as

$$\frac{dl}{dt} + \frac{1}{\tau} l = 0 \tag{8.1}$$

where l = change in arc length, and τ = proportionality constant.

Using the melting rate relationship (Lesnewich, 1958; Halmoy, 1979; 1981), a more complete form of Equation (8.1) which incorporates the control input is given (Kannatey-Asibu, Jr., 1987; Wu and Richardson, 1989) by

$$\frac{dl}{dt} = -K_5 l - r\omega \tag{8.2}$$

where $K_5 = K_0\, mn$, m = arc voltage — arc length characteristics slope, n = absolute value of the inverse of the power source characteristics slope, K_0 = constant, r = transmission ratio from the wire drive motor to the wire speed, l = arc length, t = time, and ω = drive motor rotational speed.

The corresponding transfer function is

$$L(S) = -\frac{K_w}{\tau_w s + 1} \Omega(S) \tag{8.3}$$

where $\tau_w = 1/K_5$ is the weld process time constant, $K_w = r/K_5$ is the weld process gain, and L(S) and Ω(S) are the Laplace transforms of the arc length and motor angular speed, respectively.

If the wire-feed drive motor is modeled as a first-order system, then the overall system transfer function becomes

$$L(S) = -\frac{K_w K_m}{(\tau_w S + 1)(\tau_m S + 1)} E_m(S) \tag{8.4}$$

where E_m is the input voltage to the drive motor, τ_m the motor time constant, and K_{m} the motor gain.

8.3.2 Weld Bead Geometry Control

One of the important characteristics of a weldment is the geometry of the weld bead as defined by its cross-sectional area, but in simpler terms the bead width and depth of penetration. The models developed in this and the next section may also be applicable to conduction mode laser welding.

The dynamics of the weld pool for full penetration autogenous welding, i.e., when there is no filler metal being added, can be obtained by considering the idealized configuration when the weld pool is assumed to be isothermal and at the melting point of the material (Hardt et al., 1985; Bates and Hardt, 1985). The pool walls are assumed to be vertical, conduction heat transfer is considered to be the principal mode, and the dynamics of weld pool volume resulting from melting are considered to overshadow thermal dynamics of the solid material. For an idealized cylindrical geometry, the heat balance for the system is

$$Q_{in} = Q_c + \rho L_h \frac{dV_0}{dt} \tag{8.5}$$

© 2002 by CRC Press LLC

where Q_{in} is the net heat input from the source to the weld pool and is given by ηEI for arc welding; Q_c is the heat flow by conduction from the weld pool to the base material; ρ is the density of the molten pool; L_h the latent heat of fusion; V_0 the pool volume; η heat transfer efficiency; E arc voltage; and I the welding current.

Using Fourier's law, the conduction term can be expressed as

$$Q_c = -2\pi khr\frac{dT}{dr} \tag{8.6}$$

where k is the thermal conductivity, h the plate thickness, r the pool radius, and T is the temperature. Expressing the volume V_0 in terms of the radius and height of the pool, Equation (8.5) then reduces to

$$Q_{in} = 2\pi\rho L_h hr\frac{dr}{dt} - 2\pi khr\frac{dT}{dr} \tag{8.7}$$

This is a nonlinear equation for the dynamics of the pool radius. In this form, the equation is not suitable for use in simple feedback control. A form more suitable for simple control can be obtained by lumping variables together as follows:

$$\eta EI = A(r,h)\frac{dr}{dt} + B\left(k,h,\frac{dT}{dr}\right)r \tag{8.8}$$

The result is a nonlinear first-order model of the process. However, if the parameters A and B are assumed to be constant, then the Laplace transform of the equation can be taken to obtain the following transfer function of the system:

$$\frac{R(S)}{I(S)} = \frac{K}{\tau_p S + 1} \tag{8.9}$$

where $K = hE/B$ is the process gain, $\tau_p = A/B$ is the process time constant, and $R(S)$ and $I(S)$ are the Laplace transforms of the pool radius and welding current, respectively.

8.3.3 Weld Material Properties

Another primary output of the welding process is the microstructure, which determines the weld material properties. Again, we are faced with the problem that this output is not directly measurable in real time, i.e., it is unobservable. Thus, feedback control that involves direct measurement of this parameter as an output cannot be implemented. However, closed-loop control of the temperature field, along with an open-loop microstructure and material properties output would significantly mitigate the impact of disturbances.

In this regard, the appropriate inputs for the process are the heat input Q_{in}, and traverse velocity, V. The outputs are the bead cross-sectional area NS, heat-affected zone size HAZ, and centerline cooling rate CR.

8.3.3.1 Bead Size

The dynamic relationship between the bead size NS and either the heat input Q_{in} or welding velocity V is modeled as first order (Doumanidis and Hardt, 1989):

$$\frac{NS(S)}{V(S)} = \frac{K_a}{\tau_a S + 1} \tag{8.10}$$

© 2002 by CRC Press LLC

8.3.3.2 Heat-Affected Zone Size

Because the heat-affected zone is given by the difference between two isotherms, the solidification temperature T_s (for a pure material) and the temperature at which a phase change occurs T_h, with each being described by a first-order behavior, the heat-affected zone is expected to exhibit a non-minimum phase second-order behavior. Thus,

$$\frac{HAZ(S)}{Q_{in}(S)} = \frac{K_1}{\tau_1 S+1} - \frac{K_2}{\tau_2 S+1} = \frac{K_b(\tau_b S+1)}{(\tau_1 S+1)(\tau_2 S+1)} \tag{8.11}$$

8.3.3.3 Cooling Rate

The centerline cooling rate response to a step change in either Q_{in} or V is best described by an overdamped second-order behavior:

$$\frac{CR(S)}{Q_{in}(S)} = \frac{K_c}{(\tau_\alpha S+1)(\tau_\beta S+1)} \tag{8.12}$$

Having outlined some of the basic models that constitute the basis for weld process control, we now discuss some of the more common sensor systems for monitoring process outputs.

8.3.4 Monitoring of Arc Welding and Laser Welding

The hostile nature of the process environment (high temperatures and spatter) presents difficulties in the development of reliable sensors. The principal parameters that need to be monitored during laser welding, for example, include the weld pool geometry (width and penetration); discontinuities (cracking, porosity, etc.); microstructure (strength); residual stresses; peak temperatures; and cooling rates. Among the most commonly used sensors are acoustic emission, audible sound (acoustic sensing), infrared/ultraviolet detectors, and optical (vision) sensors. A brief overview of commercially available systems is presented first, followed by an outline of each of the principal sensor systems.

8.3.4.1 Commercially Available Systems

Most of the systems currently available commercially in the United States for monitoring welding processes maintain process inputs such as current, voltage, wire feed rate (in the case of arc welding), and gas flow rate within some desirable range. Two of the key systems include the Computer Weld Technology (formerly CRC-Evans) Arc Data Monitor (ADM) and Jetline Engineering's Archcon Weld Monitor. The LWM 900 is marketed by JURCA Optoelektronik in Germany, for monitoring CO_2 laser welding processes. As opposed to the ADM and Archon systems, the LWM 900 indirectly monitors the process output by detecting the ultraviolet and infrared radiation emitted by the welding plasma and glowing metal spatter, respectively. It analyzes the amplitude and frequency of the detected signals. The PMS10 plasma monitoring system by Thyssen also detects plasma radiation and analyzes it by considering the plasma interrupts that are grouped into three categories, plasma flashes grouped into two categories, and average plasma intensity. The groupings for the first two cases are based on the duration of the signal. These parameters are then used to detect porosity formation and incomplete penetration.

8.3.4.2 Acoustic Emission

One sensor type that has been extensively investigated for weld process monitoring is acoustic emission (AE). AE refers to stress waves that are generated as a result of the rapid release of elastic strain energy within a material due to a rearrangement of its internal structure. It is also sometimes referred to as stress wave emission. The resulting stress waves propagate through the structure and

© 2002 by CRC Press LLC

produce small displacements on the surface of the structure. These are detected by sensors which convert the displacements into electrical signals. AE is an active phenomenon, because it is generated by the process under investigation. In addition, AE signals are well suited for real-time or continuous monitoring because they are generated while the phenomenon is undergoing change. Two types of transducers are normally used for AE signal detection: piezoelectric transducers and capacitive transducers.

Investigations into AE generation during electron beam welding indicate that an increase in the intensity of energy input increases the AE signal intensity (Dickhaut and Eisenblatter, 1975). Continuous signals have been associated with smooth weld beads, while burst signals apparently correlate with surface markings on nonuniform weld beads. Defect-related signals, especially cracks, have been found to be of greater amplitude than the continuous AE signals (Fang et al., 1996; Jolly, 1969; Wehrmeister, 1977). However, the presence of other undesired signal sources made the detection of the actual crack signals rather difficult (Prine, 1978). Most of the difficulty was caused by the method of signal analysis used at the time, the ring-down count. In recent years, signal processing of acoustic emission signals has been extended from traditional count and count rate analyses to the more reliable pattern recognition analysis that also enables different signal sources to be identified (Liu and Kannatey-Asibu, 1990).

Acoustic emission, too, has found application in the location of the focal point during laser welding, being maximum when the focal point coincides with the work surface (Orlick et al., 1991), and also in laser spot welding (Hamann et al., 1989; Weeter and Albright, 1987).

Precautions that need to be taken when applying conventional AE instrumentation to welding include (a) protecting the transducer from the high temperatures of welding environments and providing a highly reliable acoustic contact between the transducer and the structure; (b) positioning the transducer with respect to the material being welded and the source location; and (c) protecting the instrumentation from electromagnetic interferences resulting from arc welding equipment (Nechaev, 1978).

8.3.4.3 Audible Sound

Most manufacturing processes naturally emit sound, and an experienced human operator can use these operational sounds to determine whether or not the process is functioning normally. This indicates that the sound emitted by the process contains information that can be used to monitor the system. Audible sound sensors detect low-frequency (5 to 20 kHz) signals generated during processing (Mombo-Caristan et al., 1991), and involve microphones directed toward the process area. An advantage of audible sound monitoring is that it is noncontact, and also reduces the risk of instrumentation damage. Another advantage is the relatively lower frequency range, which makes it easier to digitize and analyze the signals.

Various methods have been investigated for analyzing sound signals generated during welding. These include statistical approaches which show that there is a narrow band of audible sound emission near 4.5 kHz for good welds, with no narrow band being observed for poor welds, but where the spectrum spreads out with a significantly lower amplitude (Gu and Duley, 1994, 1996). Neural network and linear discriminant functions also have been used to monitor on-line arc welding quality and classify the signals as acceptable or unacceptable (Matteson et al., 1993). Time-frequency analysis of audible sound signals emanating from the weld also indicates that the spectrum of a good weld can be differentiated from the spectrum of a bad weld (Farson et al., 1991, 1996).

8.3.4.4 Acoustic Nozzle and Acoustic Mirror

Airborne signals sensed by mounting a piezoelectric transducer on the focusing optic have been compared with AE signals from a piezoelectric transducer mounted on the workpiece. The results indicate airborne signals are capable of monitoring weld defects (Hamann et al., 1989; Jon, 1985). Signals from the laser welding process have also been monitored using the acoustic nozzle and the acoustic mirror (Li and Steen, 1992; Steen and Weerasinghe, 1986). With the acoustic nozzle, the

© 2002 by CRC Press LLC

transducer is mounted on the focusing assembly nozzle, while with the acoustic mirror the transducer is mounted on the reflecting mirror. Experimental results indicate that signal strength is a function of penetration depth, incident power, and plasma density. Additional results indicate that signal amplitudes increase dramatically when the keyhole forms.

8.3.4.5 Infrared/Ultraviolet Sensors

The infrared-ultraviolet (IR/UV) detection technique analyzes radiation emitted from the process zone in two wavelength bands: the infrared band in which most of the radiation from the hot material is considered to be concentrated, and the ultraviolet band in which the plasma radiation is considered to be concentrated (Chen et al., 1991; Lewis and Dixon, 1985). A typical sensor used for infrared radiation is a germanium photodiode fitted with a silicon filter having a spectral range from 1.0 to 1.9 mm. The ultraviolet radiation may be measured with a gallium phosphide (GaP) photodiode with a spectral range from 0.19 to 0.52 mm. Even though the signal intensity is generally observed to depend on the viewing distance, its characteristics are found to be independent of the arrangement used when viewing at two fixed wavebands. Both the ultraviolet and infrared signal intensities, however, increase with laser power, while increasing shielding gas flow rate reduces the signal intensities, probably due to a reduction in plasma volume.

Spatial temperature gradients in the vicinity of the weld pool can be detected using infrared thermography. An ideal weld should result in regular and repeatable patterns of the temperature gradients. Imperfections in the welding process, however, result in a discernible change in the thermal profiles. Chin et al. (1983, 1989), Boillot et al. (1985), Khan et al. (1984), and Nishar et al. (1994) showed that the average weld pool diameter can be obtained from a line scan across the center of the pool profile, and is given by the inflections around the peak temperature. When the heat source is shifted to one side of the joint center, the thermal image becomes distorted in shape, consisting then of halfmoon shapes. This asymmetrical temperature distribution is caused by the excess energy which is deposited on one side of the joint relative to the other, and the contact resistance at the joint, which reduces heat flow across the joint, resulting in higher temperatures on the side with excess energy. The heat source can then be moved in the appropriate direction until the two radii are equal. A variation in the seam also causes a shift in the shapes of the isotherms.

In addition to being used for joint tracking, the temperature isotherms can also be used to identify geometrical variations encountered in the welding process such as in the joint opening and mismatches. For example, a variation in the joint opening causes an indentation in the isothermal lines corresponding to a decrease from the peak temperatures of the metal surrounding the opening. Impurities in the weld pool appear as cold spots in the thermograms.

8.3.4.6 Weld Pool Oscillation

The weld pool, being a fluid system, oscillates when subjected to appropriate excitation, and the nature of the oscillation is determined by the pool's geometric configuration as well as its physical properties (Renwick and Richardson, 1983; Sorensen and Eagar, 1990; Xiao and den Ouden, 1993). For a stationary weld pool of infinite depth, the natural frequency of the pool is related to its geometry and properties if the fluid is assumed to be inviscid and incompressible, with flow being irrotational:

$$\omega_n^2 = \frac{7.66g}{W} + \frac{4.49\gamma}{W^3\rho}$$

while that of a pool of finite depth D is

$$\omega_n^2 = \left[\frac{7.66g}{W} + \frac{4.49\gamma}{W^3\rho}\right]\tanh\left(\frac{7.66D}{W}\right)$$

© 2002 by CRC Press LLC

where g = acceleration due to gravity, γ = surface tension, W = width of the weld pool, and ρ = density of the weld pool. This may be used to characterize arc and conduction-mode laser welding systems.

8.3.4.7 Optical Sensing

Optical sensing (vision) is often used for monitoring weld pool geometry, observing flow on the free pool surface, and chevron formation during welding. It is also useful for monitoring the kerf size during laser cutting and laser material interactions in general (Denney and Metzbower, 1991).

The basic components of an optical sensing system include the sensor, illumination source, object, transmission elements, and finally the processor. The sensing elements may be, for example, silicon photodiodes or lateral effect diodes. The lateral effect diode behaves like a resistor with a photogenerated current induced along its length by an incident light. The detector elements are normally very light sensitive, and thus may saturate easily. Attenuation of the signal is often necessary, and caution needs to be exercised in this regard because improper attenuation can introduce distortion and interference effects. The wavelength response is typical of the spectral response of the silicon which falls in the range 0.19 to 1.10 mm.

In the case of welding, for example, the sensed objects include the joint to be welded, weld pool, under bead, and bead surface. Some of the problems associated with optical sensing include the extreme brightness of the plasma plume compared to that of the molten pool (high contrast), and dependence of the intensity on processing conditions. Spatter, fumes, and flux also may obscure the object to some extent. As a result of these problems, separate illumination is often used to counteract the effect of plasma plume illumination, maintain a stable intensity that is appropriate for the sensor, enhance contrast, and provide a brightness level that is suitable for the sensor. This increases the system resolution. The separate illumination may be in the form of either structured light or general illumination, i.e., nonstructured light. A structured light is a pattern of lines or a grid of light projected onto the object to help provide information on the three-dimensional shape of the object based on the apparent distortion of the pattern.

The general illumination could come from an auxiliary high-intensity light source. One application of general illumination would involve lighting the object with a narrow bandwidth laser beam, with the beam bandwidth selected to be in the region where, based on the spectral characteristics of the detector, the detector's sensitivity is high. All light on the detector is then filtered except for the narrow bandwidth of the auxiliary beam, thereby subduing the effect of the bright light from the plume. An enhancement of this technique involves the use of both diffused and focused light (Voelkel and Mazumder, 1990).

There are two main forms of optical sensing systems: linear array systems and two-dimensional array systems. The linear systems may consist of a column of, for instance, up to 2048 pixels or individual sensing elements in a line, while the two-dimensional system may have 500×500 elements.

One principal advantage of the linear array sensor is the rapid processing of information. The resolution is limited by the size of the field of view and the spacing of the sensing elements. Moving the sensor along the joint provides information on the joint profile. Periodic scanning of the array yields the light intensity detected by each sensing element. Objects of interest can be identified using various techniques, but in the simplest case, a threshold light intensity may be defined for the object, such as the edge of a weld pool, and used to identify the pool edges. A line scan camera has been used to measure the width of the weld puddle (Vroman and Brandt, 1976; Nomura et al., 1976).

The two-dimensional array detector monitors a sizeable area simultaneously, and is thus suited for two-dimensional objects such as the weld pool. The sensor in this case is normally a solid-state video camera with an array say, 500×500 charge injection device or charge coupled device light sensitive elements. The output of each element or pixel may be an 8-bit digitized video. The output from the camera may be immediately dumped into a memory buffer for analysis.

© 2002 by CRC Press LLC

The pool width may be identified by analyzing the output of a row of elements located across the weld pool. The pool area will require the entire two-dimensional image. The output may be processed by averaging each pixel's signal with a given number of pixels on either side. The waveform may then be numerically differentiated by finding the difference between each adjoining pixel, and again averaging the resulting signal. From this processed signal, the weld pool edges would be given, for example, by the second zero crossings (Kovacevic et al., 1995; Richardson et al., 1982).

For viewing the pool and/or the joint, the camera may be positioned at any desirable location, but a convenient configuration involves having the camera's optical axis coincident with the beam axis, providing an image of the weld pool and surrounding area (Richardson et al., 1984).

8.3.4.8 Multi-Sensor Systems

In recent years multi-sensor systems have been investigated for monitoring manufacturing processes. Utilizing multi-sensor integration incorporates the advantages of different sensors into one system. Furthermore, incorporating modularity permits the selection of the combination of sensors most appropriate for a particular application. An integrated system consisting of an acoustic mirror for back reflection, acoustic nozzle for airborne emissions, plasma charge sensor for plasma monitoring, and a dual wavelength infrared and ultraviolet sensing of the weld region has been investigated for laser welding. (Steen, 1992) The results indicate that the acoustic mirror, acoustic nozzle, and plasma charge sensor can monitor keyhole formation while the infrared/ultraviolet sensor can monitor the temperature and size of the weld pool and the stability of the keyhole. Other sensor combinations have been investigated (Parthasarathi et al., 1992).

8.3.4.9 Seam Tracking

A weld-seam tracking system that senses the arc voltage (GTAW) or current (GMAW) while oscillating the welding torch from one sidewall extremity of the joint to the other has been developed using the melting rate equation and relationships that exist between the arc voltage, current, and torch-to-work spacing, Cook (1983). Seam tracking also can be implemented using infrared and vision systems.

References

Bates, B. E. and Hardt, D.E., 1985, A real-time calibrated thermal model for closed-loop weld bead geometry control, *ASME Journal of Dynamic Systems Measuremetn and Control*, 107, 25–33.

Boillot, J. P., Cielo, P., Begin, G., Michel, C., Lessard, M., Fafard, P., Villemure, D., 1985, Adaptive welding by fiber optic thermographic sensing: An analysis of thermal and instrumental considerations, *Welding Journal*, 64, 209s–217s.

Ceglarek, D., and Shi, J., 1996, Fixture failure diagnosis for autobody assembly using pattern recognition, *ASME Journal of Engineering for Industry,* 118, 1, 55–66.

Chang, H. S. and Cho, H. S., 1989, An interactive learning control system for resistance spot welding process, *Transactions of ASME*, 111, 129–135.

Chen, H., Li, L., Brookfield, D., Williams, K., and Steen, W., 1991, Laser process monitoring with dual wavelength optical sensors, *ICALEO'91*, 113–122.

Chin, B. A., Madsen, N. H., and Goodling, J. S., 1983, Infrared thermography for sensing the arc welding process, *Welding Journal*, 62, 227s–234s.

Chin, B. A., Nagarajan, S. and Chen, W. H., 1989, Infrared sensing for adaptive arc welding, *Welding Journal*, 68, 462s–466s.

Cho, H. S., and Chun, D. W., 1985, A microprocessor-based electrode movement controller for spot weld quality assurance, *IEEE Transactions on Industrial Electronics*, IEEE-32, No. 3.

Cook, G.E., 1980, Feedback control of process variables in arc welding, *Proceedings, 1980 Joint Automatic Control Conference*, San Francisco, CA.

© 2002 by CRC Press LLC

Cook, G. E., 1983, Through-the-Arc Sensing for Arc Welding, *10th NSF Conference on Production Research and Technology*, Detroit, Michigan, 141–151.

Cook, G. E., Anderson, K., and Barrett, R. J., 1989, Feedback and Adaptive Control in Welding, *2nd International Conference on Trends in Welding Research*, Gatlinburg, Tennessee, David, S., and Vitek, J. M., editors, 891–903.

Denney, P. E., and Metzbower, E. A., 1991, Synchronized laser-video camera system study of high power laser material interactions, *ICALEO*, 84–93.

DeVor, R.E., Chang, T., and Sutherland, J., 1992, *Statistical Quality Design and Control, Contemporary Concepts and Methods*, Macmillan, New York.

Dickinson, D. W., Franklin, J. E., and Stanya, A., 1980, Characterization of spot welding behavior by dynamic electrical resistance monitoring, *Welding Journal*, 59(6), 170-s–176-s.

Dickhaut, E. and Eisenblatter, J., 1975, Acoustic emission measurements during electron beam welding of nickel-base alloys, *Journal of Engineering Power Transactions ASME*, 97, 47–52.

Dornfeld, D.A., Tomizuka, M., and Langari, G., 1982, Modeling and adaptive control of arc-welding processes, *Measurement and Control for Batch Manufacturing*, Hardt, D. E., ed., 65–75.

Doumanidis, C., and Hardt, D.E., 1989, A model for in-process control of thermal properties during welding, *ASME Journal of Dynamic Systems, Measurement, and Control*, 111, 40–50.

Fang, C.-K., Kannatey-Asibu, Jr., E., and Barber, J., 1996, Far-field initial response of acoustic emission from cracking in a weldment, *ASME Journal of Manufacturing Science and Engineering*, 119, 281–289.

Farson, D. F., Fang, K. S., and Kern, J., 1991, Intelligent laser welding control, *ICALEO*, 104–112.

Farson, D., Hillsley, K., Sames, J., and Young, R., 1996, Frequency-time characteristics of air-borne signals from laser welds, *Journal of Laser Applications*, 8, 33–42.

Gedeon, S. A., Sorenson, C. D., Ulrich, K. T., and Eagar, T. W., 1987, Measurement of dynamic electrical and mechanical properties of resistance spot welding, *Welding Journal*, 65, 12, 378s–385s.

Gould, J. E., 1987, An examination of nugget development during spot welding, using both experimental and analytical techniques, Welding Research Supplement, *Welding Journal*.

Gu, H., and Duley, W.W., 1994, Acoustic emission and optimized CO2 laser welding of steel sheets, *ICALEO'94*, 79, 77–85.

Gu, H., and Duley, W.W., 1996, Statistical approach to acoustic monitoring of laser welding, *Journal of Physics D: Applied Physics*, 29, 3, 556–560.

Haefner, K., Carey, B., Bernstein, B., Overton, K., and Andrea, M. D., 1991, Real-time adaptive spot welding control, *Journal of Dynamic Systems, Measurement and Control.*

Halmoy, E., 1979, Wire melting rate, droplet temperature, and effective anode melting potential, *Proceedings International Conference on Arc Physics and Weld Pool Behavior,* The Welding Institute, *Cambridge*, 49–57.

Halmoy, E., April 1981, Dynamics of gas metal arc welding, Presented at the American Welding Society Annual Meeting, Cleveland, Ohio.

Hamann, C., Rosen, H.-G., and LaBiger, B., 1989, Acoustic emission and its application to laser spot welding, *SPIE High Power Lasers and Laser Machining Technology*, 1132, 275–281.

Han, Z., Orozco, J., Indacochea, J. E., Chen, C. H., 1989, Resistance spot welding: A heat transfer study, Welding Research Supplement, *Welding Journal.*

Hao, M., Osman, K. A., Boomer, D. R., and Newton, C. J., 1996, Developments in characterization of resistance spot welding of aluminum, *Welding Journal*, 75, 1, 1s–8s.

Hardt, D.E., Garlow, D.A., Weinert, J.B., 1985, A model of full penetration arc-welding for control system design, *ASME Journal for Dynamic Systems, Measurement and Control,* 107, 40–46.

Hu, S. J., and Wu, S. M., 1992, Identifying root causes of variation in automotive body assembly using principal component analysis, *Transactions of NAMRI,* XX, 311–316.

Janota, M., 1975, Control of current and time on the basis of weld nugget, *Proceedings of the IIW.*

Johnson, R. A., and Wichern, D. W., 1992, *Applied Multivariate Statistical Analysis*, Prentice Hall, New York.

Jolly, W. D., 1969, Acoustic emission exposes cracks during welding, *Welding Journal*, 48, 21–27.

Jon, M. C., 1985, Non-contact acoustic emission monitoring of laser beam welding, *Welding Journal*, 64, 43–48.

© 2002 by CRC Press LLC

Jou, M., Messler, R., and Li, C. J., 1994, A fuzzy logic control system for resistance spot welding based on a neural network model, IEEE IAS Meeting, 95CB35862, 1757–1763.

Kannatey-Asibu, Jr., E., 1987, Analysis of the GMAW process for microprocessor control of arc length, *ASME Journal of Engineering for Industry*, 109, 172–176.

Kannatey-Asibu, Jr., E., 1997, Milestone developments in welding and joining processes, *ASME Journal of Manufacturing Science and Engineering,* 119, 801–810.

Khan, M. A., Madsen, N. H., Chin, B. A., Ballard, P., and Lin, T. T., 1984, Infrared thermography as a control for the welding process, *Welding Research Progress*, XXXIX, 2, 28–40.

Kovacevic, R., Zhang, Y. M., and Ruan, S., 1995, Sensing and control of weld pool geometry for automated GTA welding, *ASME Journal of Engineering for Industry*, 117, 210–222.

Kuchar, N. R., Cohen, R. K., Nied, H. A., and Godwin, S. J., Nov. 14, 1982, A closed-loop control system for resistance spot welding, *ASME Winter Annual Meeting.*

Lesnewich, A., August, 1958, Control of melting rate and metal transfer in gas-shielded metal-arc-welding. Part I — Control of electrode melting rate*, The Welding Journal,* 37, 343–353-S.

Lewis, G. K., and Dixon, R. D., 1985, Plasma monitoring of laser beam welds, *Welding Journal*, 64, 49s–54s.

Li, L., and Steen, W. M., 1992, Non-contact acoustic emission monitoring during laser processing, *ICALEO'92*, 719–728.

Li, W., Hu, S. J., and Ni, J., 1998, A model for on-line quality prediction of resistance spot welding, Accepted by *ASME Journal of Manufacturing Science and Engineering.*

Liu, X., and Kannatey-Asibu, Jr., E., 1990, Classification of AE signals for martensite formation from welding, *Welding Journal*, 69, No. 10, 389s–394s.

Mantripragada, R., 1998, Assembly Oriented Design: Concepts, Algorithms and Computational Tools, Ph.D. Thesis, Massachusetts Institute of Technology.

Matteson, A., Morris, R., and Tate, R., 1993, Real-time GMAW quality classification using an artificial neural network with airborne acoustic signals as inputs, *Proceedings of the 12th International Conference on Offshore Mechanics and Arctic Engineering*, III-A, 273–278.

Matushita, 1993, A real time method for contact area calculation, *Proceedings of the IIW.*

Mombo-Caristan, J.-C., Koch, M., and Prange, W., 1991, Seam geometry monitoring for tailored welded blanks, *ICALEO*, 123–132.

Muller, A., Greene, W.J., and Rothschild, G.R., August 1951, Characteristics of the inert-gas-shielded metal-arcs*, The Welding Journal*, 30, 717–727.

Nechaev, V. V., 1978, An acoustic emission transducer for inspecting welding quality, *A translation from Defektoskopiya*, No. 11, 21–27.

Needham, J. C., Benton, D. B., Hannah, M. D., and Newlin, R. G., 1965, Automotive quality control in resistance spot welding mild steel, *Welding Journal,* 44, 4, 168s.

Nied, H. A., 1984, The finite element modeling of the resistance spot welding process, Welding Research Supplement, *Welding Journal.*

Nishar, D. V., Schiano, J. L., Perkins, W. R., and Weber, R. A., 1994, Adaptive control of temperature in arc welding, *IEEE Control Systems,* 94, 4–24.

Nomura, H., Yoshida, T., and Tohno, K., June 1976, Control of weld penetration, *Metal Construction,* 244–246.

Orlick, H., Morgenstern, H., and Meyendorf, N., 1991, Process monitoring in welding and solid state lasers by sound emission analysis, *Welding and Cutting*, 12, 15–18.

Parthasarathi, S., Khan, A.A., and Paul, A.J., 1992, Intelligent laser processing of materials, *ICALEO'92*, 75, 708–718.

Prine, D. W., 1978, A Two Channel Microprocessor Controlled Acoustic Emission Monitor for In-Process Weld Monitoring, *Proceedings 24th Annual ISA Conference*, Albuquerque, New Mexico.

Renwick, R. J., and Richardson, R. W., 1983, Experimental investigation of GTA weld pool oscillations, *Welding Journal*, 62, 29s–35s.

Richardson, R. W., Gutow, A. A., and Rao, S. H., 1982, A vision based system for arc weld pool size control, *Measurement and Control for Batch Manufacturing*, ASME Booklet, Hardt, D.E., Ed., 65–75.

© 2002 by CRC Press LLC

Richardson, R. W., Gutow, D. A., Anderson, R. A., and Farson, D. F., 1984, Coaxial arc weld pool viewing for process monitoring and control, *Welding Journal*, 63, 43s–50s.

Roan, C. and Hu, S. J., July 1994, Multivariate monitoring and classification of dimensional faults for automotive body assembly, *First S. M. Wu Symposium on Manufacturing Sciences*, Beijing, China.

Schumacher, B. W., Cooper, J. C., and Dilay, W., 1984, Resistance spot welding control that automatically selects the welding schedule for different types of steel, *Ford Motor Company Research Report.*

Shriver, J., Hu, S. J., and Peng, H., 1998, Resistance spot welding: A neural network approach to modeling, *Proceedings of the ASME,* MED 8, 201–211.

Sorensen, C. D., and Eagar, T. W., 1990, Measurement of oscillations in partially penetrated weld pools through spectral analysis, *ASME Journal of Dynamic Systems, Measurement and Control*, 112, 463–468.

Steen, W. M., and Weerasinghe, V. M., 1986, In Process Beam Monitoring, *SPIE Laser Processing: Fundamentals, Applications, and Systems Engineering*, 668, 37–44.

Steen, W. M., 1992, Adaptive control of laser material processing, *Proceedings of LAMP*, 1, 439–444.

Suzuki, A., Hardt, D. E., Valavani, L., 1991, Application of adaptive control theory to on-line GTA weld geometry regulation, *ASME Journal of Dynamic Systems, Measurement and Control,* 113, 93–103.

Tomizuka M., Dornfeld, D., and Purcell, M., 1980, Application of microcomputers to automatic weld quality control, *ASME Journal of Dynamic Systems, Measurement and Control,* 102, 62–68.

Towey, M., and Andrews, D. R., October 1968, Instantaneous resistance during spot welding formation as a parameter for an automatic control system, *Welding and Metal Fabrication*, 383–392.

Tsai, C. L., Dai, W. L., Dickinson, D. W., and Papritan, J. C., 1991, Analysis and development of a real-time control methodology in resistance spot welding.

Vahavilos, S.J., Carlos, M.F., and Slykhouse, S.J., 1981, Adaptive spot weld feedback control loop via acoustic emission, *Material Evaluation*, 39, 10, 1057–1060.

Voelkel, D. D., and Mazumder, J., 1990, Visualization and dimensional measurement of the laser weld pool, *ICALEO*, 422–429.

Vroman, A. R., and Brandt, H., 1976, Feedback control of GTA welding using puddle width measurements, *Welding Journal*, 55, 742–749.

Waller, D. N., 1964, Head movement as a means of resistance welding quality control, *British Welding Journal,* 11, 118–122.

Weeter, L., and Albright, C., 1987, The effect of full penetration on laser-induced stress-wave emissions during laser spot welding, *Materials Evaluation*, 45, 353–357.

Wehrmeister, A. E., 1977, Acoustic emission monitoring of multipass submerged-arc welding, *Materials Evaluation*, 35, 45–47.

Wood, R. T., Bauer, L. W., Bedard, J. F., Bernstein, B. M., Czechowski, J., D'Andrea, M. M., and Hogle, R. A., 1985, Closed-loop control systemm for three-phase resistance spot welding, *Welding Journal*, 64, 12, 26–30.

Wu, G.-D., and Richardson, R.W., 1989, The dynamic response of self-regulation of the welding arc, *2nd International Conference on Trends in Welding Research*, Gatlinburg, Tennessee, David, S., and Vitek, J. M., editors, 929–933.

Xiao, Y. H., and den Ouden, G., 1993, Weld pool oscillation during GTA welding of mild steel, *Welding Journal,* 72, 428s–434s.

© 2002 by CRC Press LLC

9

Control of Polymer Processing

David Kazmer
University of Massachusetts, Amherst

Kourosh Danai
University of Massachusetts, Amherst

9.1 Introduction
9.2 Process Description
9.3 Process Variability
9.4 Modeling
9.5 Process Control
Machine Control • State-Variable Control • Set-Point Control
9.6 Conclusions

9.1 Introduction

Process control is recognized as an important means of improving the performance and consistency of thermoplastic parts. However, no single control strategy or system design is universally accepted, and manufacturing systems continue to produce defective components during production. This chapter provides an overview of modeling, measurement, and control strategies in polymer processing, and discusses some of the difficulties posed by their complex and distributed nature.

Most plastic parts are fabricated by thermoforming, extrusion, or injection molding. In thermoforming and its variants (vacuum forming, blow molding, male forming, drape forming, plug-assist forming, etc.) a continuous sheet of material is heated first until it becomes pliable (elastic modulus of approximately 0.5 Mpa), and then it is expanded at strain rates of approximately 100% per second to assume the shape of an evacuated mold. The hot sheet is then cooled by conduction of heat to the mold, which itself is cooled with conditioned recirculated water. The resulting part typically exhibits thickness distributions from 10 to 90% of the initial sheet thickness, with mold cycle times varying from 15 seconds to 5 minutes per part.

Unlike thermoforming, which is a cyclic process, extrusion is a continuous and steady-state process. In extrusion, solid thermoplastic pellets are fed into a rotating screw to be compacted into a tightly packed solid bed. The thermal energy for melting comes from the mechanical power of the motor that is consumed to rotate the screw. The tapered flight on the screw geometry is designed to match the rate of dissipative melting to present minimum flow restriction and smooth flow. The resulting homogeneous melt is then forced at a constant rate through a complex profile die designed such that the material exits the die at uniform temperature and velocity. The continuous extruded part is fed through a series of cooling molds to maintain and set the part geometry, after which sections are cut to length while the extrusion process continues. Extrusion rates of approximately 20 feet per minute are typical. While the majority of extruded parts are simple round or square tubing, the process is capable of producing intricate profiles such as window casings and structural members.

© 2002 by CRC Press LLC

Injection molding consists of several stages: plastication, injection, packing, cooling, and ejection. It is the most complex of the above processes and capable of producing very complex components to tight specifications. Injection molding embodies the extrusion process for generating polymer melt, yet has faster time dynamics than thermoforming, over a greater temperature and pressure range. In injection molding and its variants (coinjection, injection compression, gas assist molding, etc.), thermoplastic pellets are fed into a rotating screw and melted. With a homogeneous melt collected in front of the screw, the screw is moved axially at a controlled, time-varying velocity to drive the melt into an evacuated cavity. Once the melt is solidified and the molded component is sufficiently rigid to be removed, the mold is opened and the part is ejected while the next cycle's thermoplastic melt is plasticized by the screw. Cycle times range from less than 4 seconds for compact discs to more than 3 minutes for automotive instrument panels. In order to present a general overview of issues involved in control of polymer processing in this chapter, we focus on modeling and control strategies applied to injection molding.

9.2 Process Description

Control of injection molding is significantly challenged by the nonlinear behavior of the polymeric materials, dynamic and coupled process physics, and convoluted interactions between the mold geometry and final product quality attributes. A system's view of a conventional injection molding process is presented in Figure 9.1. The machine parameters are indicated on the left side of the figure and some common molded part measures of quality are listed on the right. In this figure, the process is decomposed into five distinct but coupled stages. The output of each stage not only directly determines the initial conditions of the next stage, but also influences some of the final qualities of the molded part.

Every stage of the injection molding process is complex and warrants detailed discussion regarding its behavior. Plastication of the polymer melt is accomplished through simultaneous shearing by rotation of an internal screw and heating by an externally heated barrel. As shown in Figure 9.1, the plastication inputs include barrel temperature, screw rotation rate, screw plastication pressure, and shot size. This list is simplified in that most inputs are vectors rather than scalar quantities. For instance, barrel temperature is specified at several locations, because multiple heater bands along the length of the injection unit control the temperature of the plasticized melt. Each local segment of the barrel is typically equipped with a type J or K thermocouple embedded in the barrel steel, and the power to each heater band is individually controlled through a closed-loop programmable logic controller utilizing proportional–integral–derivative (PID) control.[1] The resulting melt quality and residence time can directly affect the quality of the molded part as unplasticized pellets and/or degraded material can reduce the structural integrity and aesthetics of the molded component.

The purpose of the injection stage is to completely fill the mold cavity with the polymer melt. This goal is achieved by driving forward the screw used for plastication at velocities of the order of 100 cm/sec according to a selected time-velocity profile. The velocity profile is selected such that the melt travels at relatively uniform velocity while converging and diverging in the mold cavity. During polymer injection, contact of the hot polymer melt with the cold mold wall results in the immediate generation of a frozen skin. Thermal conduction to the mold is then balanced against thermal convection of the melt. This thermal equilibrium stabilizes the growth of the frozen layer, which reduces the flow conductance of the melt. If too low a velocity is selected, the melt front will prematurely solidify. If too high a velocity is selected, the resin may degrade or cause excessive mold deflection and flash. The relationship between the screw velocity profile and melt front velocity is convoluted by the compressibility and acceleration dynamics of the melt. The specification of the time-velocity profile is so difficult, in fact, that most molders utilize the same profile (slow at start, fast in the middle, and slow at the end) for all molding applications. The distributed nature of the melt flow, and velocities changing with both time and position, also preclude

© 2002 by CRC Press LLC

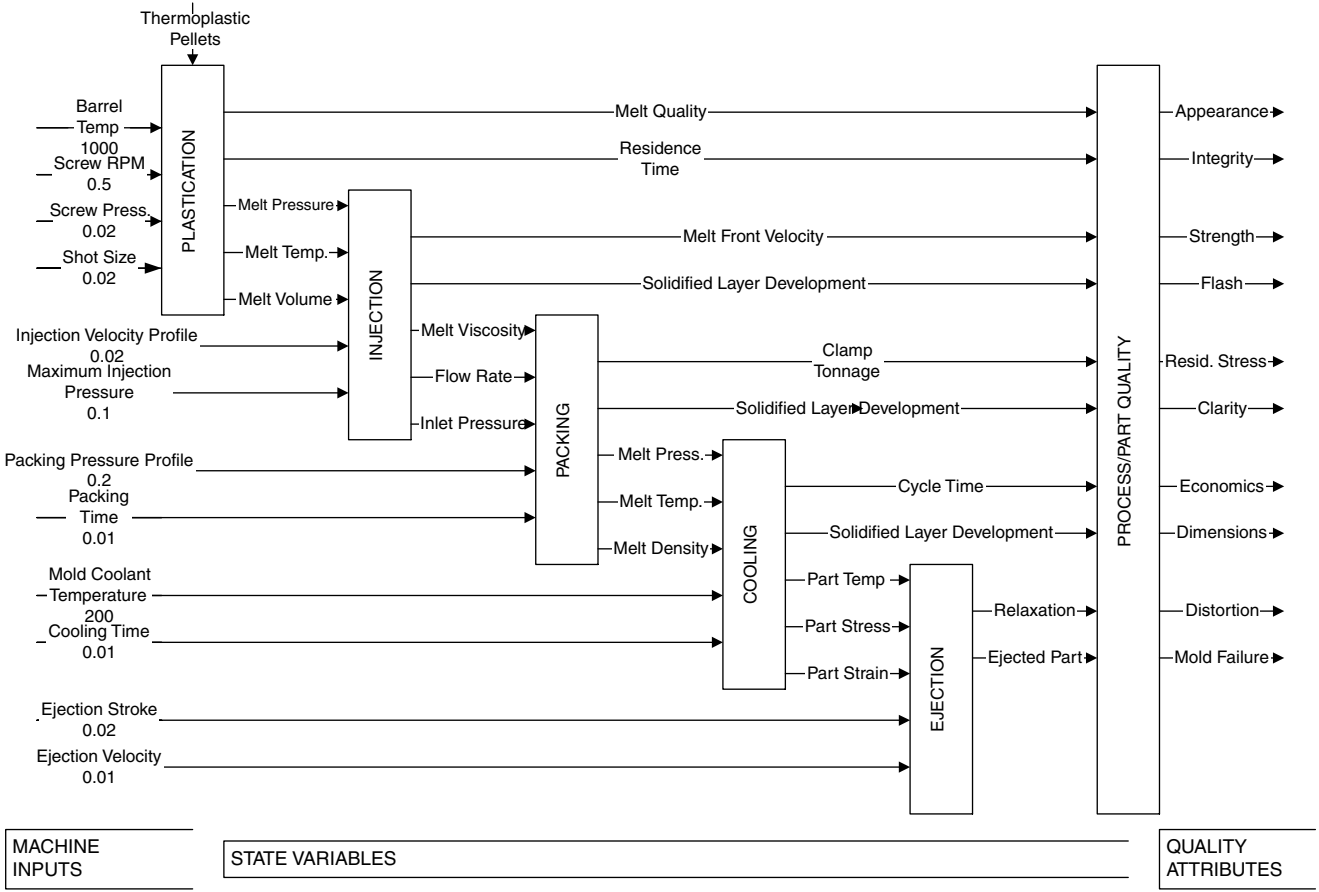


FIGURE 9.1 System's view of the injection molding process.

© 2002 by CRC Press LLC

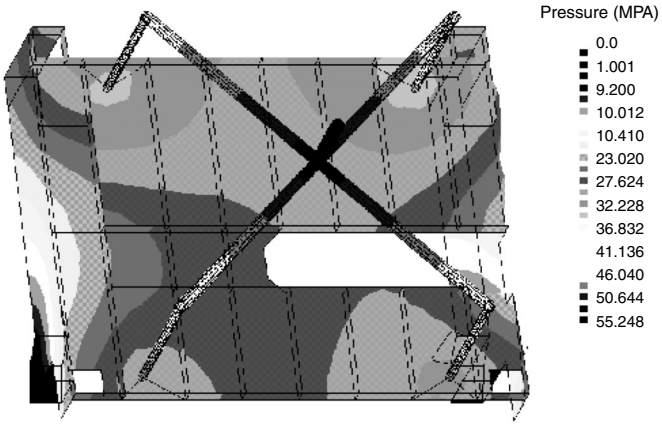


FIGURE 9.2 Pressure distribution of a typical molding at the end of the injection stage.

simultaneous control of the melt flow at different positions. Considering that the injection stage provides the initial conditions for the packing stage, the absence of complete controllability of the melt flow would result in uncontrolled melt viscosity, solidified layer distribution, and temperature/pressure contours (see Figure 9.2).

Due to volumetric shrinkage during cooling of the melt, additional material must be forced into the mold cavity during the packing stage to obtain satisfactory parts. For pack pressure control, the hydraulic pressure behind the screw is adjusted through a high-speed servo valve to decrease or increase the melt pressure at the inlet to the mold. The pressure feedback for control may be provided by a pressure transducer mounted at the mold inlet, or it may be calculated by multiplying the hydraulic pressure by a screw intensification ratio. Pressure is maintained and additional material is forced into the mold cavity until the part has solidified. However, part solidification is an internal variable to the molding process that cannot be measured directly. To determine the correct packing time, multiple molding trials with various packing times must be performed and the molded parts weighed. It should be noted that part weight is also dependent on melt temperature and pressure, so a change in machine inputs may result in inaccurate packing times.

After packing, the polymer melt is solidified but is too soft for part ejection. As such, coolant is recirculated at a controlled temperature through the mold to remove heat. The cooling stage predominates the molding cycle, requiring approximately half of the cycle to complete. Production economics dictate shorter cycle times, but shorter cooling times may lead to excessive part shrinkage and warpage.

9.3 Process Variability

Process variability in injection molding further complicates process control. The sources of variability are attributed to the thermoplastic resin, the injection molding machine, and environmental factors. Product inconsistencies among a batch of molded parts are most frequently assigned to lot-to-lot variations in material properties. Small changes in viscosity, density, or composition may occur when regrind is mixed with virgin material, a material is used after it has been stored over an extended period of time, or a switch is made between different batches of the same material grade.[2] Small changes in material properties can lead to inconsistencies in part weight, part dimensions, aesthetics, strength, etc.

The second source of variability is process machinery. Molding machines of different injection cylinder and clamp design will have very different machine dynamics, and provide different levels of molded part quality for the same process set points. Even identical machines from the same manufacturer can induce significant quality variation as a result of differences in their controllers

© 2002 by CRC Press LLC

TABLE 9.1 Magnitude of Process Variation by Machine Input

Control Quality	Low (Class 9)	High (Class 1)
Melt temperature (C)	5	1
Mold temperature (C)	8	2
Injection time (sec)	0.17	0.04
Pack pressure (Mpa)	0.5	0.1
Pack time (sec)	0.02	0.09
Cooling time (sec)	0.86	0.20

and varying amounts of wear in the melt and hydraulic delivery systems. Finally, parts molded from the same press may vary due to internal controller variations relating to the shot size, injection velocity, switchover point, pack pressure, etc. Hunkar[3] has characterized and described a machine evaluation methodology that quantifies the process consistency of any molding machine. The plastics industry is adopting this methodology, which categorizes machines into capability classes from 1 to 9 with predefined variances as shown in Table 9.1.

The third source of variability is human and environmental interaction with the process. For instance, process engineers have different definitions of "optimal"[4] and can induce product inconsistency through the modification of standard process set points such as injection velocity, pack pressure, back pressure, cooling time, and ejection set-up. Press operators directly determine cycle time and part handling, and may influence some process settings. The physical environment also will introduce variation. For instance, outdoor temperature may affect the effectiveness of evaporative coolers that determine the temperature of the plant water. Indoor temperature can likewise have a significant effect on the mold wall temperature as well as the post-molding behavior of the molded parts. Humidity can effect the dryness of the polymeric material entering the barrel, thus introducing further quality inconsistencies.

9.4 Modeling

As previously discussed, the primary barrier to control of injection molding stems from the distributed nature of the polymeric material. This demands models that can represent the state of the material both spatially and temporally. For example, state variables such as the melt velocity, melt pressure, and melt temperature are not only functions of time but are inhomogeneous both through the thickness and across the mold.

Fundamental research of the injection molding process began with Spencer's empirical investigation of melt flow advancement.[5] Harry and Parrott later utilized a finite difference form of the heat equation to predict the melt flow advancement along a long, narrow strip for a specific material and injection pressure.[6] Williams and Lord[7] advanced the simulation of the injection molding process by discretizing both the length and thickness dimension to track the melt front propagation while simultaneously performing heat transfer calculations. This was the first analysis to consider the dynamic buildup of a solidified skin layer as well as the polymer's complex non-Newtonian (shear dependent) rheological behavior. Based on these analyses, sophisticated simulations were soon introduced for use in part design and process troubleshooting.[8] More advanced numerical schemes based on the hybrid finite element/finite difference method were then introduced to simulate melt propagation in arbitrarily complex three-dimensional geometries,[9,10] such as those presented in Figure 9.2. Continuing research seeks to predict the residual stresses,[11-13] fiber orientation,[14,15] and other properties of the final molded product.[11,16,17] These simulation softwares are now standard tools in the design of thermoplastic parts, as well as verification of various control strategies.

The modeling advances in injection molding, however, have not yet significantly impacted control of these processes. The primary reason is the unsuitability of the developed mechanistic models for control analysis and design. Although there have been applications of these mechanistic models in

© 2002 by CRC Press LLC

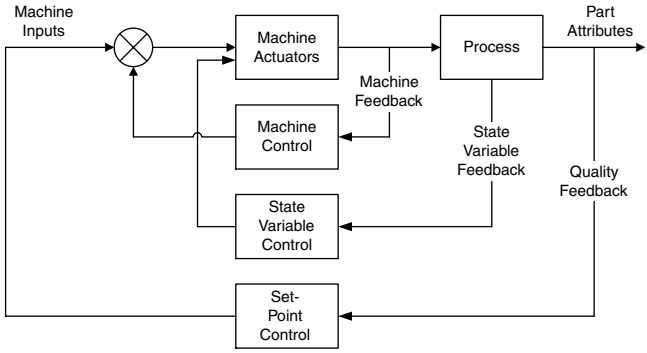


FIGURE 9.3 System diagram of injection molding control.

controls,[18-20] by and large, they have not been used directly in control. As an alternative, models in the form of a time series or auto-regressive moving average (ARMA) have been developed empirically for control design.[21,22] In such cases, the state of the material at a point only within the mold is modeled and controlled. Another approach used for representing the melt behavior is neural network modeling,[54,55] where the distributed nature of the melt can be represented by multi-input/multi-output patterns.

9.5 Process Control

A fundamental difficulty in control of injection molding is that none of the final molded part properties can be ascertained within the molding cycle. Instrumentation does not yet exist, and may never exist, to yield information about aesthetics or structural integrity prior to opening the mold and ejection of the part. Therefore, part quality is satisfied through a combination of on-line state-variable control (through continuous control of the melt state) and off-line cycle-to-cycle adjustment of the machine set points. These two modes of control give injection molding the characteristic of both a continuous and discrete process.

An overview of injection molding control is shown in Figure 9.3. At the innermost level, only the machine actuators are regulated. This level of control will ensure proper execution of the programmed machine inputs (see Figure 9.1). At the second level, state variables such as melt temperature and melt pressure are controlled to track prespecified profiles. This will provide more precise control of the state of the melt. At the outermost level, the machine inputs are adjusted to improve the quality of the part through better set points given feedback of part quality.

The logic behind the control strategy in Figure 9.3 can be explained by an example. Consider the specification of the packing pressure profile as a machine input for control of the part width in Figure 9.2. In this case, the machine actuator will be the hydraulic servo valve to the injection cylinder, and machine control will ensure a specified packing pressure at the melt inlet. However, the packing pressure will be nonuniformly distributed in the mold, as shown in Figure 9.2. This motivates state-variable control to regulate the cavity pressure more precisely based on feedback of measured pressure inside the mold. In this case, the input to the hydraulic servo valve will be augmented to provide the additional level of precision. While this additional level of control ensures realization of the specified cavity pressure, it still may not lead to a satisfactory molded part because of a poorly specified cavity pressure. Set point control is incorporated to adjust the specified cavity pressure. Each of these control levels is discussed next.

9.5.1 Machine Control

Prior to the 1970s, the majority of molding machines utilized open-loop control for most subsystems. For example, heater wattage was set to achieve a prespecified barrel temperature, or the

© 2002 by CRC Press LLC

servo valve spool position was set to provide a specified screw velocity and pressure profile. Since the advent of programmable logic control, the majority of machine input variables have become individually controlled via single-input/single-output PID algorithms. Among the machine inputs listed in Figure 9.1, the melt temperature, the packing pressure profile, and the injection velocity profile are considered the most important to control.

The first modern computer-controlled injection molding machine was described by Carl Ma in 1974 while employed at Cincinnati Milacron.[23] Ma's work led to the development of modern control systems for injection molding machines and enabled current closed-loop control systems for ram velocity and injection pressure.[24] In theory, machine control algorithms are simple enough to enable the molder to properly tune them. In practice, molders find controller tuning difficult, so controller parameters are rarely changed from their factory defaults. Poor or infrequent controller tuning results in reductions in process capability because one set of controller parameters will not be appropriate for all molding applications. For example, an increase in polymer viscosity would increase the resistance to flow and would increase the load on the screw, as would a decrease in melt temperature. Each of these cases would require a different set of controller parameters. In an effort to improve control performance, more sophisticated control methods than PID have been investigated. For example, Pandelidis and Agrawal demonstrated the application of linear quadratic control to tracking ram velocity.[25] Tsai and Lu developed a multivariable self-tuning predictive controller for improving set-point tracking performance, disturbance rejection, and robustness of a temperature control system for an extruder barrel.[22]

9.5.2 State-Variable Control

While machine control is important, it is the polymer state (pressure, temperature, and morphology) which directly determines the molded part quality.[26] As such, recent technological developments have rightly focused on closing the loop between the machine parameters and the polymer state. If achieved, these advanced control strategies will provide increased molded part quality and consistency.

The dichotomy between the machine inputs and state variables is illustrated in Figure 9.1, where every input variable that utilizes closed-loop control has been identified with a numeric subscript that quantifies the approximate time response of the controlled parameter in seconds. Also indicated in this figure, is the role of state variables as intermediate variables between the machine inputs and the final part quality attributes. A fundamental difficulty in injection molding control is the lack of models to define the relationships from inputs to state variables and from state variables to outputs. For example, melt temperature is known to be affected by barrel temperature, screw rotational speed, and melt. However, only 20 to 50% of the energy required for melting originates from the barrel heaters, and the exact relation to melt temperature is a function of polymer properties and screw/barrel design. Similarly, melt temperature is widely accepted as affecting cycle time and part dimensions, but the precise one-to-many relationships are generally not available prior to molding. Although the void for mechanistic relationships is often filled with empirical or heuristic models in state-variable control, empirical modeling has not been adopted by industry due to the cost of experimentation.

The two dominant variables defining the state of the melt are temperature and pressure. Typical strategies used for melt temperature control are discussed in References 27 and 28. The main effort in these studies has been to identify the control method that can best achieve a prespecified melt temperature. In addition to the lack of a systematic method for specifying the melt temperature, melt temperature control suffers from the absence of reliable sensors for melt temperature measurement. Intrusive thermocouple probes placed in the viscous melt stream fail quickly,[29] and infrared pyrometers do not calibrate automatically with changes in resin color, filler content, or emissivity.[30] A review of temperature sensors available for injection molding is provided in Reference 31.

© 2002 by CRC Press LLC

Another fundamental state variable that can be regulated during the cycle is cavity pressure. Closed-loop control of cavity pressure could automatically compensate for variations in melt viscosity and injection pressure to achieve a consistent process and uniform set of product attributes.[31] Mann introduced one of the first pressure control schemes by using modulated pressure relief valves,[32] and Abu Fara developed a process control model by relating the cavity pressure response to open-loop perturbations. Srinivasan later used these models to propose a learning controller for closed-loop cavity pressure control.[34] Adaptive control methods have also been proposed to track cavity pressure profile, usually at one location in the mold.[35-37]

Like melt temperature control, cavity pressure control suffers from the lack of a systematic method for determining the pressure profile. In addition, it is handicapped by the absence of appropriate actuators for distributed pressure control, as conventional molding machines are equipped with only one actuator (the screw) which does not allow simultaneous cavity pressure control at multiple points in the mold. A step toward solving this problem has been the development of dynamic melt flow regulators that allow control of the flow and pressure of the polymer melt at multiple points in the mold.[38] Similar concepts regarding dynamic thermal actuation are discussed in Reference 39.

Further advancements in state-variable control are becoming possible through development of remote smart sensors. Packing time, for example, is currently controlled open-loop, using a fixed time delay specified by the machine operator. Thomas et al.[40] have developed new sensors that infer the solidification of polymer in the mold, and have devised a closed-loop strategy where pack time is automatically controlled based on feedback from a solidification sensor. Using this strategy, the pack time can be set once in reference to the sensor signal, making it possible to provide a minimum pack time for each part under changing processing conditions.

9.5.3 Set-Point Control

The adjustment of machine inputs is a discrete control process, where the molded part quality attributes from the cycle just completed are utilized to determine the magnitude of the machine inputs for the next molding cycle. Ideally, these set points should be specified to produce parts with acceptable part quality attributes, which for an injection molded part would typically be size, surface topography, and/or mechanical properties (e.g., tensile strength, flexural strength). However, the molding process is typically over-constrained, so a trade-off needs to be made between multiple quality objectives and cost in the specification of the set points.

The traditional approach to machine input selection (tuning) in the plastics industry has been trial and error. For this, shots are taken during start-up and part quality attributes are measured after each shot to evaluate the acceptability of produced parts. The process engineer then uses his/her knowledge of the process to select the machine inputs in such a way as to improve the quality of the part from shot to shot. This tuning exercise is repeated until the specifications for part quality are satisfied. The main drawback of the traditional tuning approach is its inefficiency due to its ad hoc nature. An alternative to the traditional trial and error approach is the use of expert systems where corrective guidelines are presented in the form of if–then rules.[41-44] The main shortcoming of expert systems is that a generalized set of rules may not be applicable across a broad range of part geometries, material properties, and machine dynamics.

The predominant practice for set-point specification in large job operations is to develop an empirical model based on data obtained from a set of designed experiments.[45] Based on the empirical model, an optimization may be performed to find the set of machine inputs that best maximizes the molded part quality. Design of experiments (DOE)-based methods offer a systematic approach to tuning that can also be used for mold qualification,[46-48] but they often require significant investment in training and technology.

Alternative approaches have been utilized to relate machine inputs to the observed part quality attributes. Woll and Cooper trained a backpropagation network (BPN) as an inverse model relating

© 2002 by CRC Press LLC

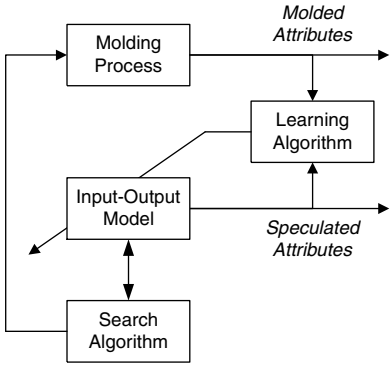


FIGURE 9.4 Diagram of the virtual search method of tuning.

discretized patterns of cavity pressure as inputs to the corresponding values of holding pressure and barrel temperature that had produced them via simulation as outputs. The values of holding pressure and barrel temperature were then adjusted from cycle to cycle by comparing the actual cavity pressure pattern with a desired pattern, using the learned patterns as baselines.[49] A similar approach was utilized by Demirci et al. to determine the inlet flow rate to the mold given the current position of the flow front during the filling stage.[50] This control scheme was based on a neural network that was trained with data obtained from a mechanistic model. The network was trained to estimate the position of the next flow front as output given the present position of the flow front and the inlet flow rate as inputs. Using this network as a forward model, a search was conducted to determine the inlet flow rate to the mold, based on the present position of the flow front and its desired next position. With this strategy, one could specify a desired flow progression scheme and the controller would iteratively take corrective actions to realize this scheme. The drawback of the above approaches is the considerable time they require to develop the underlying models off-line.

A similar approach to the above methods for set-point control is the virtual search method (VSM) that also uses a forward model and search to determine the machine inputs;[51] however, VSM has the advantage of not requiring an off-line model by developing the input-output (I-O) model concurrent with the process. The block diagram of VSM is shown in Figure 9.4. It consists of an I-O model that estimates the corresponding changes to the part attributes, a search algorithm that determines prospective changes to the machine inputs for the next part, and a learning algorithm to update the I-O model after each cycle based on part quality measurements. VSM exhausts the search based on the current I-O model and refers to the process in order to (1) test the feasibility of the best set of inputs obtained from the I-O model, and (2) to update the I-O model using the measurements of part quality attributes obtained from the process. According to this scheme, the I-O model is updated only when it no longer provides guidance toward the feasible region, thus, enabling efficient utilization of the I-O model to its fullest capacity before updating it. VSM's interleaved approach to tuning and model development has been shown to require fewer process iterations than DOE methods, which require a comprehensive model of the process over a broad range of machine inputs.

9.6 Conclusions

The polymer processing industry utilizes sophisticated control algorithms for machine control. However, two significant barriers prevent 100% quality assurance and true cost minimization. First, the relationships between the machine input variables and final quality attributes are not precisely known. Second, these processes are largely over-constrained, such that improvement in one part quality attribute is not feasible without reducing other quality attributes or increasing cost. In theory,

© 2002 by CRC Press LLC

more accurate process simulations could eliminate the need for costly molding trials and mold tooling iterations uncertainty of material properties and the variability of the process.

Several development issues need to be addressed toward meaningful control of polymer processing. First, more comprehensive models that can provide an accurate estimate of part quality attributes for various sets of machine inputs, material properties, and mold configurations have to be developed. Second, robust and miniaturized sensors should be developed to provide feedback about the state of the melt inside the mold. Third, advanced actuators need to be developed that can provide the multi-degrees of freedom required for control of the melt in a distributed manner. The ultimate aim is a machine that will produce no scrap material at increased production rates, and will require less labor skill, less energy, and minimal maintenance.

References

1. Anon., Information guide to temperature controls for injection molding/extrusion, *Plastics Design & Processing*, 15, 17–19, 1975.
2. A. J. Poslinski, C. A. Deborski, and S. Aslam, Effects of small changes in the melt viscosity on the filling and packing stages of injection molding, presented at Annual Technical Conference — ANTEC, Conference Proceedings, San Francisco, California, 1994.
3. D. B. Hunkar, Managing the future with dynamic statistics, presented at Annual Technical Conference — ANTEC, Conference Proceedings, Boston, Massachusetts, 1995.
4. N. M. Morris and W. B. Rouse, Effects of type of knowledge upon human problem solving in a process control task, *IEEE Transactions on Systems, Man and Cybernetics*, SMC-15, 698–707, 1985.
5. R. Spencer and R. Dillon, The viscous flow of molten polystyrene, *Journal of Colloidal Science*, 3, 163, 1948.
6. D. H. Harry and R. G. Parrott, Numerical simulation of injection mold filling, 10, 209–214, 1970.
7. H. A. Lord and G. Williams, Mold filling studies for the injection molding of thermoplastic materials: Transient flow of plastic materials in the cavities of injection molding dies, *Canadian Controls and Instrumentation,* 318–328, 1975.
8. C. A. Hieber and S. F. Shen, Flow analysis of the non-isothermal two-dimensional filling process in injection molding, *Israel Journal of Technology*, 16, 248–254, 1978.
9. V. W. Wang, C. A. Hieber, and K. K. Wang, Filling of an arbitrary three-dimensional thin cavity, *Journal of Polymer Engineering*, 7, 21, 1986.
10. M. R. Kamal, E. Chu, P. G. Lafleur, and M. E. Ryan, Computer simulation of injection mold filling for viscoelastic melts with fountain flow, *Polymer Engineering and Science*, 26, 190–196, 1986.
11. R. Y. Chang and B. D. Tsaur, Experimental and theoretical studies of shrinkage, warpage, and sink marks of crystalline polymer injection molded parts, *Polymer Engineering and Science*, 35, 1222, 1995.
12. W. F. Zoetelief, L. F. A. Douven, and A. J. I. Housz, Residual thermal stresses in injection molded products, *Polymer Engineering and Science*, 36, 1886–1896, 1996.
13. S.-J. Liu, Modeling and simulation of thermally induced stress and warpage in injection molded thermoplastics, *Polymer Engineering and Science*, 36, 807–818, 1996.
14. J. J. McGrath and J. M. Wille, Determination of 3D fiber orientation distribution in thermoplastic injection molding, *Composites Science and Technology*, 53, 133–143, 1995.
15. V. Verleye, A. Couniot, and F. Dupret, Prediction of fiber orientation in complex injection molded parts, presented at the American Society of Mechanical Engineers Winter Annual Meeting, Applied Mechanics Division, New Orleans, Louisiana, 1993.
16. S.-C. Chen, S.-Y. Hu, R. D. Chien, and J.-S. Huang, Integrated simulations of structural performance, molding process, and warpage for gas-assisted injection-molded parts. I. Analysis of part structural performance, *Journal of Applied Polymer Science*, 68, 417–428, 1998.
17. B. Friedrichs, M. Horie, and Y. Yamaguchi, Simulation and analysis of birefringence in magneto-optical discs. Part A: formulation, *Journal of Materials Processing and Manufacturing Science*, 5, 95–113, 1996.

© 2002 by CRC Press LLC

18. M. Rafizadeh, W. I. Patterson, and M. R. Kamal, Physically-based model of thermoplastics injection molding for control applications, *International Polymer Processing*, 11, 352, 1996.
19. D. Kazmer, J. Rowland, and G. Sherbelis, Foundations of intelligent process control for injection molding, *Journal of Injection Molding Technology*, 1, 44–56, 1997.
20. P. D. Coates, A. R. Haynes, and R. G. Speight, In-line characterization of polymer deformation in melt and solid phase processing, *Polymer*, 35, 3831–3843, 1994.
21. T. C. Bulgrin and T. H. Richards, Application of advanced control theory to enhance molding machine performance, *IEEE Transactions on Industry Applications*, 31, 1350–1357, 1995.
22. C.-C. Tsai and C.-H. Lu, Multivariable self-tuning temperature control for plastic injection molding process, *IEEE Transactions on Industry Applications*, 34, 310–318, 1998.
23. C. Y. W. Ma, A design approach to a computer-controlled injection-molding machine, *Polymer Engineering and Science*, 11, 768–772, 1974.
24. J. E. O'Bryan, Injection molding machines meet automated manufacturing demands. Proportional valves, microprocessors, and closed-loop control keep plastics molders competitive, *Hydraulics & Pneumatics*, 42, 95, 1989.
25. I. O. Pandelidis and A. R. Agrawal, Optimal anticapatory control of ram velocity in injection molding, *Polymer Engineering and Science*, 28, 147–156, 1988.
26. P. D. Coates and R. G. Speight, Towards intelligent process control of injection moulding of polymers, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 209, 357–367, 1995.
27. R. Dubay, A. C. Bell, and Y. P. Gupta, Control of plastic melt temperature: A multiple input multiple output model predictive approach, *Polymer Engineering and Science*, 37, 1550–1563, 1997.
28. V. G. Gomes, W. I. Patterson, and M. R. Kamal, Injection molding study: Evaluation of alternative control strategies for melt temperature, *Polymer Engineering and Science*, 26, 867–876, 1986.
29. T. W. McCullough and M. A. Spalding, Predicting actual temperature distributions in a polymer stream using an adjustable-depth, exposed-tip thermocouple assembly, *Journal of Reinforced Plastics and Composites*, 16, 1622–1630, 1997.
30. G.-Y. Lai and J. X. Rietveld, Role of polymer transparency and temperature gradients in the quantitative measurement of process stream temperatures during injection molding via IR pyrometry, *Polymer Engineering and Science*, 36, 1755–1768, 1996.
31. U. Langkamp, Pressure and temperature sensors, *Kunststoffe Plast Europe*, 86, 1804–1812, in German, 1996.
32. J. W. Mann, Process parameter control: The key to optimization, *Plastics Engineering*, 30, 25–27, 1974.
33. M. R. Kamal, W. I. Patterson, N. Conley, D. Abu Fara, and G. Lohfink, Dynamics and control of pressure in the injection molding of thermoplastics, *Polymer Engineering and Science*, 27, 1403–1410, 1987.
34. K. Srinivasan and T. Brinivasan, Learning control of melt pressure in injection molding processes, presented at the American Society of Mechanical Engineers Winter Annual Meeting, Dynamic Systems and Control Division (Publication) DSC, Atlanta, Georgia, 1991.
35. F. Gao, I. A. N. Patterson, and M. R. Kamal, Self-tuning cavity pressure control of injection molding filling, *Advances in Polymer Technology*, 13, 111–120, 1994.
36. C.-P. Chiu, M.-C. Shih, and J.-H. Wei, Dynamic modeling of the mold filling process in an injection molding machine, *Polymer Engineering and Science*, 31, 1417–1424, 1991.
37. R. E. Nunn and C. P. Grolman, Closed loop cavity pressure control in injection molding, *Journal of Reinforced Plastics and Composites*, 9, 2121, 1991.
38. D. O. Kazmer and P. Barkan, Multi-cavity pressure control in the filling and packing stages of the injection molding process, *Polymer Engineering and Science*, 37, 1865–1879, 1997.
39. K. M. B. Jansen, Heat transfer in injection moulding systems with insulation layers and heating elements, *International Journal of Heat and Mass Transfer*, 38, 309–316, 1995.
40. C. L. Thomas, A. A. Tseng, A. J. Bur, and J. L. Rose, Solidification sensing for closed loop control of injection molding hold time, *Advances in Polymer Technology*, 15, 151–163, 1996.

© 2002 by CRC Press LLC

41. Y. J. Huh, S. Han, and K. K. Wang, Knowledge-based synthesis system for chip encapsulation, presented at Annual Technical Conference — ANTEC, Conference Proceedings, Boston, Massachusetts, 1995.
42. K. Shelesh-Nezhad and E. Siores, Intelligent system for plastic injection molding process design, *Journal of Materials Processing Technology*, 63, 458–462, 1997.
43. I. Catic, M. Slavica, M. Sercer, and G. Baric, Expert system aided troubleshooting in polymer engineering, presented at Annual Technical Conference, Indianapolis, Indiana, 1996.
44. A. Bernhardt, G. Bertacchi, and A. Vignale, Rationalization of molding machine intelligent setting & control, presented at Annual Technical Conference, Atlanta, Georgia, 1998.
45. J. T. Luftig and V. S. Jordan, *Design of Experiments in Quality Engineering*, McGraw-Hill, New York, 1998.
46. P. J. Meheran and P. J. Blyskal, Design of experiments for the injection molding process (qualifying materials to produce equivalent parts), presented at Annual Technical Conference, Boston, Massachusetts, 1995.
47. C. Liu and L. T. Manzione, Process studies in precision injection molding. I: Process parameters and precision, *Polymer Engineering and Science*, 36, 1–9, 1996.
48. J. C. Viana, P. Kearney, and A. M. Cunha, Improving impact strength of injection molded plates through molding conditions optimization: A design of experiments approach, presented at Annual Technical Conference, Atlanta, Georgia, 1998.
49. S. L. B. Woll and D. J. Cooper, Pattern-based closed-loop quality control for the injection molding process*, Polymer Engineering and Science*, 37, 801–812, 1997.
50. H. H. Demirci, J. P. Coulter, and S. I. Guceri, Numerical and experimental investigation of neural network-based intelligent control of molding processes, *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 119, 88–94, 1997.
51. R. Ivester and K. Danai, Automatic tuning and regulation of injection molding by the virtual search method, *ASME Journal of Manufacturing Science and Engineering,* 120, 323–329, 1998.

© 2002 by CRC Press LLC

10

Precision Manufacturing

10.1 Deterministic Theory Applied to Machine Tools
10.2 Basic Definitions
10.3 Motion
Rigid Body Motion and Kinematic Errors • Sensitive Directions • Amplification of Angular Errors, The Abbe Principle
10.4 Sources of Error and Error Budgets
Sources of Errors • Determination and Reduction of Thermal Errors • Developing an Error Budget
10.5 Some Typical Methods of Measuring Errors
Linear Displacement Errors • Spindle Error Motion — Donaldson Reversal • Straightness Errors — Straight Edge Reversal • Angular Motion — Electronic Differential Levels
10.6 Conclusion
10.7 Terminology

Thomas R. Kurfess
Georgia Institute of Technology

International competition and ever improving technology have forced manufacturers to increase quality as well as productivity. Often the improvement of quality is realized via the enhancement of production system precision. This chapter discusses some of the basic concepts in precision system design including definitions, basic principles of metrology and performance, and design concepts for precision engineering.

This chapter is concerned with the design and implementation of high precision systems. Due to space limitations, only a cursory discussion of the most basic and critical issues pertaining to the field of precision engineering is addressed. In particular, this chapter is targeted at the area of precision machine tool design. These concepts have been used to design some of the most precise machines ever produced, such the Large Optics Diamond Turning Machine (LODTM) at the Lawrence Livermore National Laboratory which has a resolution of 0.1 μin. (10^{-7} inches). However, these ideas are quite applicable to machine tools with a wide range of precision and accuracy. The first topic discussed is the Deterministic Theory, which has provided guidelines over the past 30 years that have yielded the highest precision machine tools ever realized and designed. Basic definitions followed by a discussion of typical errors are presented as well as developing an error budget. Finally, fundamental principles to reduce motion and measurement errors are discussed.

10.1 Deterministic Theory Applied to Machine Tools

The following statement is the basis of the Deterministic Theory: "Automatic machine tools obey cause and effect relationships that are within our ability to understand and control and that there

© 2002 by CRC Press LLC

is nothing random or probabilistic about their behavior" (Dr. John Loxham). Typically, the term random implies that the causes of the errors are not understood and cannot be eradicated. Typically, these errors are quantified statistically with a normal distribution or at best, with a known statistical distribution. The reality is that these errors are *apparently* nonrepeatable errors that the design engineers have decided to quantify statistically rather than completely understand. Using statistical approaches to evaluate results is reasonable when sufficient resources using basic physical principles and good metrology are not available to define and quantify the variables causing errors.[1] It must be understood that in all cases, machine tool errors that appear random are not random; rather, they have not been completely addressed in a rigorous fashion. It is important that a machine's precision and accuracy are defined early in the design process. These definitions are critical in determining the necessary depth of understanding that must be developed with respect to machine tools errors. For example, if it is determined that a machine needs to be accurate to 1 μm, then understanding its errors to a level of 1 nm may not be necessary. However, apparently, random errors of 1 μm are clearly unacceptable for the same machine.

Under the deterministic approach, errors are divided into two categories: repeatable or systematic errors and apparent nonrepeatable errors. Systematic errors are those errors that recur as a machine executes specific motion trajectories. Typical causes of systematic errors are linear slideways not being perfectly straight or improper calibration of measurement systems. These errors repeat consistently every time. Typical sources of apparent nonrepeatable errors are thermal variations, variations in procedure, and backlash. It is the apparent nonrepeatable errors that camouflage the true accuracy of machine tools and cause them to appear to be random. If these errors can be eliminated or controlled, a machine tool should be capable of having repeatability that is limited only by the resolution of its sensors. Figure 10.1 presents some of the factors affecting workpiece accuracy.[2]

10.2 Basic Definitions

This section presents a number of definitions related to precision systems. Strict adherence to these definitions is necessary to avoid confusion during the ensuing discussions. The following definitions are taken from ANSI B5.54.-1991.[5]

Accuracy: A quantitative measure of the degree of conformance to recognized national or international standards of measurement.

Repeatability: A measure of the ability of a machine to sequentially position a tool with respect to a workpiece under similar conditions.

Resolution: The least increment of a measuring device; the least significant bit on a digital machine.

The target shown in Figure 10.2 is an excellent approach to visualizing the concepts of accuracy and repeatability. The points on the target are the results of shots at the target's center or the bullseye. Accuracy is the ability to place all of the points near the center of the target. Thus, the better the accuracy, the closer the points will be to the center of the target. Repeatability is the ability to consistently cluster or group the points at the same location on the target. (Precision is often used as a synonym for repeatability; however, it is a nonpreferred, obsolete term.) Figure 10.3 shows a variety of targets with combinations of good and poor accuracy and repeatability. Resolution may be thought of as the size of the points on the target. The smaller the points, the higher the resolution.[3,4]

Error: The difference between the actual response of a machine to a command issued according to the accepted protocol of the machine's operation and the response to that command anticipated by the protocol.

Error motion: The change in position relative to the reference coordinate axes, or the surface of a perfect workpiece with its center line coincident with the axis of rotation. Error motions are specified as to location and direction and do not include motions due to thermal drift.

© 2002 by CRC Press LLC

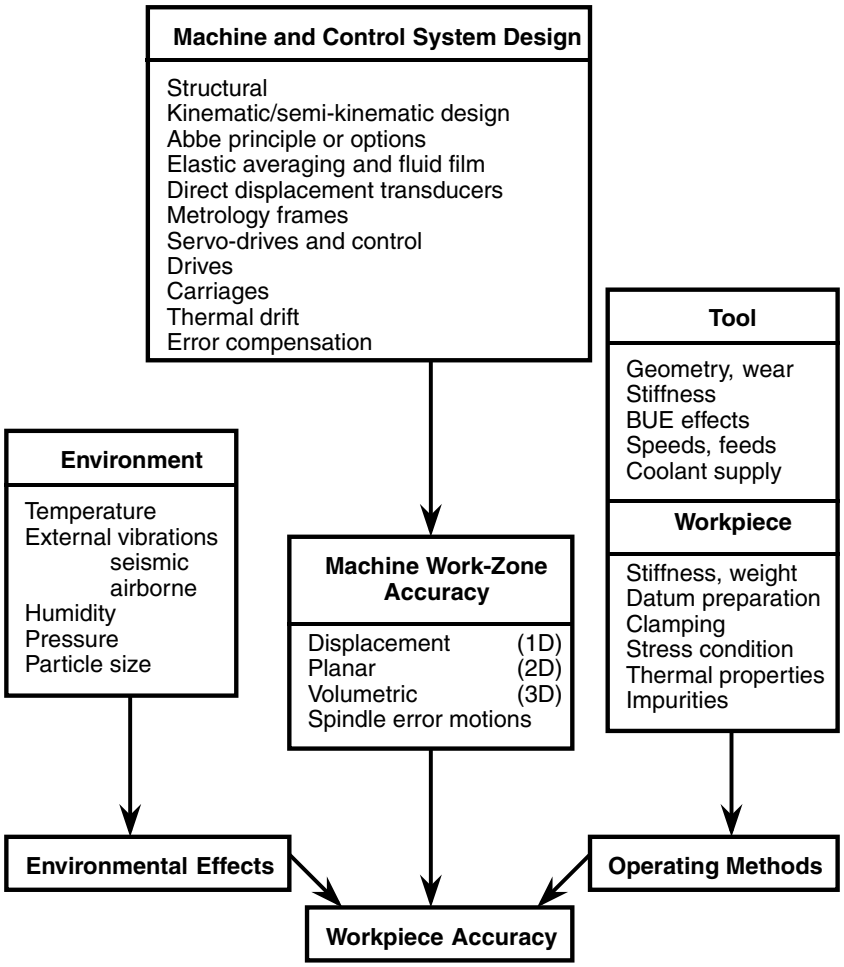


FIGURE 10.1 Some of the factors affecting workpiece accuracy.

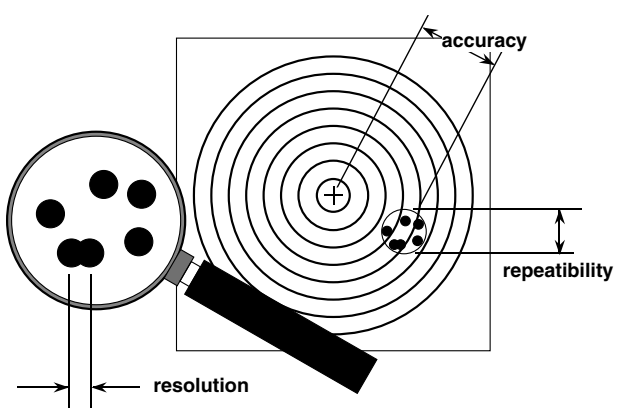


FIGURE 10.2 Visualization of accuracy, repeatability, and resolution. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

© 2002 by CRC Press LLC

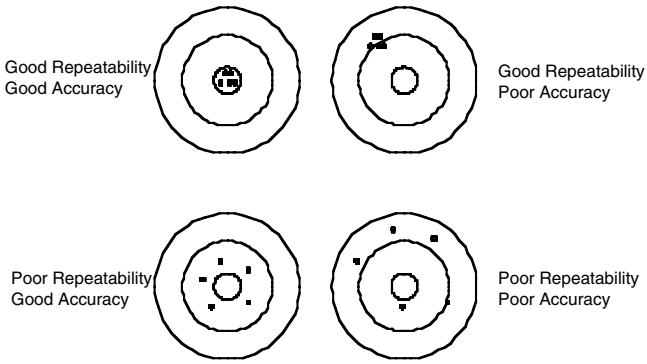


FIGURE 10.3 A comparison of good and poor accuracy and repeatability.

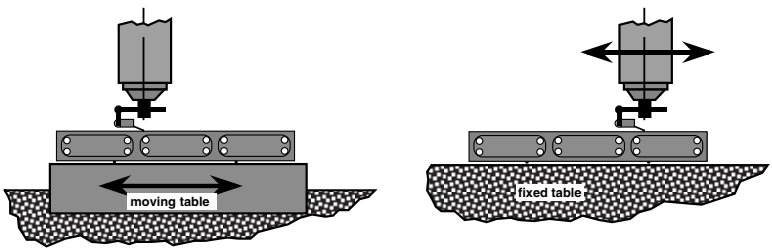


FIGURE 10.4 Slideway straightness relationships. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

Error motion measurement: A measurement record of error motion which should include all pertinent information regarding the machine, instrumentation, and test conditions.

Radial error motion: The error motion of the rotary axis normal to the Z reference axis and at a specified angular location (see Figure 10.4).[5]

Runout: The total displacement measured by an instrument sensing a moving surface or moved with respect to a fixed surface.

Slide straightness error: The deviation from straight line movement that an indicator positioned perpendicular to a slide direction exhibits when it is either stationary and reading against a perfect straightedge supported on the moving slide, or moved by the slide along a perfect straightedge that is stationary.

10.3 Motion

This chapter treats machine tools and their moving elements (slides and spindles) as being completely rigid, even though they do have some flexibility. Rigid body motion is defined as the gross dynamic motions of extended bodies that undergo relatively little internal deformation. A rigid body can be considered to be a distribution of mass rigidly fixed to a rigid frame.[6] This assumption is valid for average-sized machine tools. As a machine tool becomes larger, its structure will experience larger deflections, and it may become necessary to treat it as a flexible structure. Also, as target tolerances become smaller, compliance must be considered. For example, modern ultra-rigid production class machine tools may possess stiffnesses of over 5 million pounds per inch. While this may appear to be large, the simple example of a grinding machine that typically applies 50 lbs. of force can demonstrate that compliance can cause unacceptable inaccuracies. For this example, the 50 lbs. of force will yield a 10 μin. deflection during the grinding process, which is a large portion of the acceptable tolerance of such machine tools. These deflections are ignored in this section. Presented in this section is a fundamental approach to linking the various rigid body error motions of machine tools.

© 2002 by CRC Press LLC

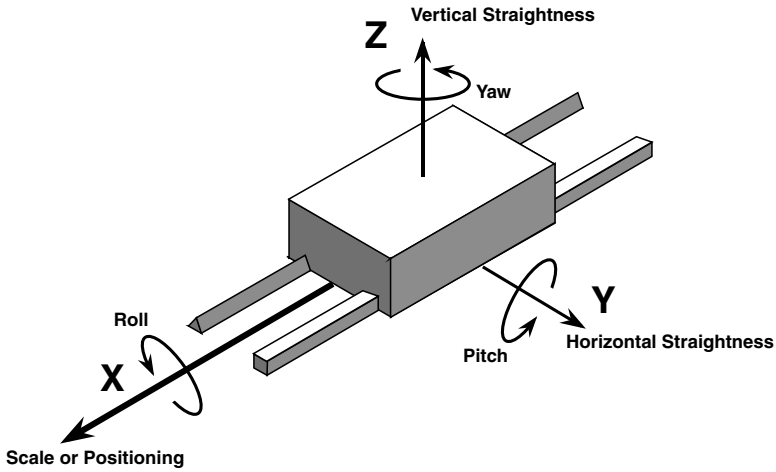


FIGURE 10.5 Slide and carriage rigid body relationships. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

10.3.1 Rigid Body Motion and Kinematic Errors

There are six degrees of freedom defined for a rigid body system, three translational degrees of freedom along the X, Y, and Z axes, as well as three rotational degrees of freedom about the X, Y, and Z axes. Figure 10.5 depicts a linear slide that is kinematically designed to have a single translational degree of freedom along the X axis. The other five degrees of freedom are undesired, treated as errors, and often referred to as kinematic errors.[7]

There are two straightness errors and three angular errors that must be considered for the slide and carriage system shown in Figure 10.5. In addition, the ability of the slide to position along its desired axis of motion is measured as scale errors. These definitions are given below:

Angular errors: Small unwanted rotations (about the X, Y, and Z axes) of a linearly moving carriage about three mutually perpendicular axes.

Scale errors: The differences between the position of the read-out device (scale) and those of a known reference linear scale (along the X axis).

Straightness errors: The nonlinear movements that an indicator sees when it is either (1) stationary and reading against a perfect straightedge supported on a moving slide or (2) moved by the slide along a perfect straightedge which is stationary (see Figure 10.5).[5] Basically, this translates to small unwanted motion (along the Y and Z axes) perpendicular to the designed direction of motion.

While slides are designed to have a single translational degree of freedom, spindles and rotary tables are designed to have a single rotational degree of freedom. Figure 10.6 depicts a single degree-of-freedom rotary system (a spindle) where the single degree of freedom is rotation about the Z axis. As with the translational slide, the remaining five degrees of freedom for the rotary system are considered to be errors.[8] As shown in Figure 10.6, two radial motion (translational) errors exist, one axial motion error, and two tilt motion (angular) errors. A sixth error term for a spindle exists only if it has the ability to index or position angularly. The definitions below help to describe spindle error motion:

Axial error motion: The translational error motion collinear with the Z reference axis of an axis of rotation (about the Z axis).

Face motion: The rotational error motion parallel to the Z reference axis at a specified radial location (along the Z axis).

© 2002 by CRC Press LLC

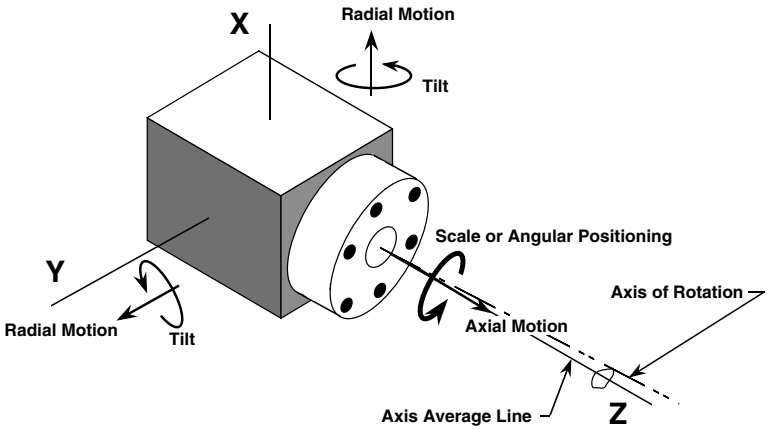


FIGURE 10.6 Spindle rigid body relationships. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

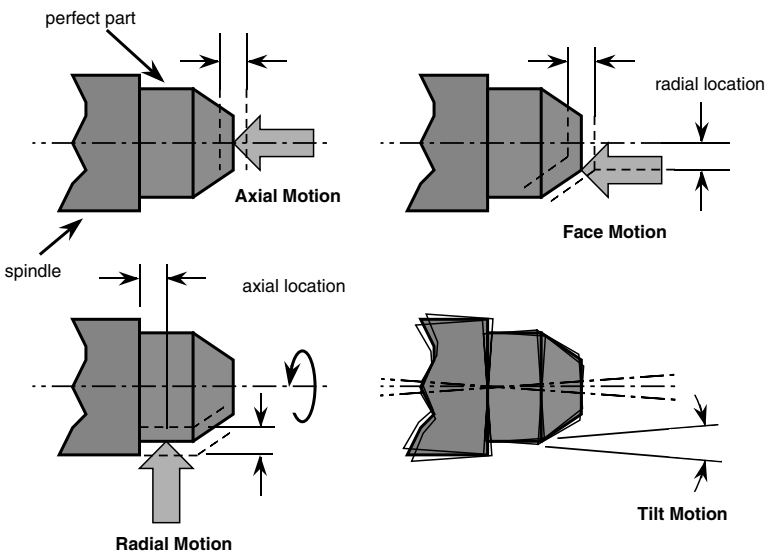


FIGURE 10.7 Spindle error motion. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

Radial error motion: The translational error motion in a direction normal to the Z reference axis and at a specified axial location (along the X and Y axes).

Tilt error motion: The error motion in an angular direction relative to the Z reference axis (about the X and Y axes).

Figure 10.7 is a plan view of a spindle with an ideal part demonstrating the spindle errors that are discussed. Both the magnitude and the location of angular motion must be specified when addressing radial and face motion.[9]

As previously stated, runout is defined as the total displacement measured by an instrument sensing against a moving surface or moved with respect to a fixed space. Thus, runout of the perfect part rotated by a spindle is the combination of the spindle error motion terms depicted in Figure 10.7 and the centering error relative to the spindle axis of rotation.[9]

Typically, machine tools consist of a combination of spindles and linear slides. Mathematical relationships between the various axes of multi-axis machine tools must be developed. Even for a

© 2002 by CRC Press LLC

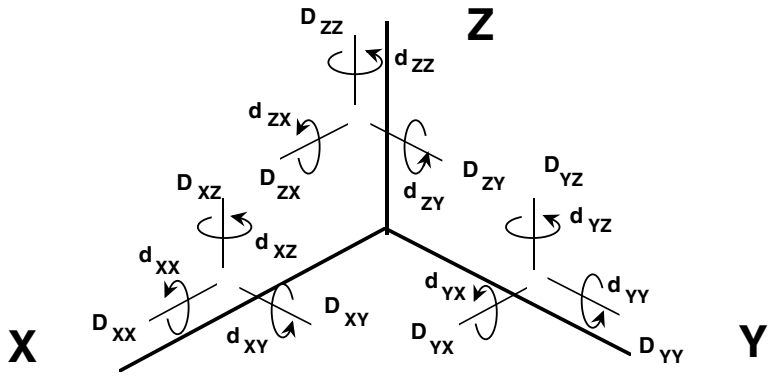


FIGURE 10.8 Error terms for a machine tool with three orthogonal axes. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

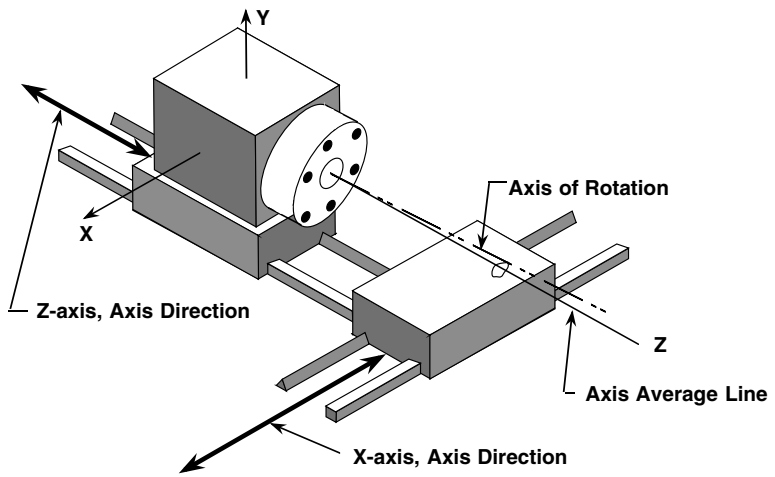


FIGURE 10.9 Typical machine tool with three desired degrees of freedom, the lathe. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

simple three-axis machine, the mathematical definition of its kinematic errors can become rather complex. Figure 10.8 presents the error terms for positioning a machine tool (without a spindle) having three orthogonal linear axes. There are six error terms per axis totaling 18 error terms for all three axes. In addition, three error terms are required to completely describe the axes relationships (e.g., squareness) for a total of 21 error terms for this machine tool. Figure 10.9 shows a simple lathe where two of the axes are translational and the third is the spindle rotational axis.

The following definitions are useful when addressing relationships between axes:

Squareness: A planar surface is "square" to an axis of rotation if coincident polar profile centers are obtained for an axial and face motion polar plot at different radii. For linear axes, the angular deviation from 90° measured between the best-fit lines drawn through two sets of straightness data derived from two orthogonal axes in a specified work zone (expressed as small angles).

Parallelism: The lack of parallelism of two or more axes (expressed as a small angle).

For machines with fixed angles other than 90°, an additional definition is used:

Angularity: The angular error between two or more axes designed to be at fixed angles other than 90°.

© 2002 by CRC Press LLC

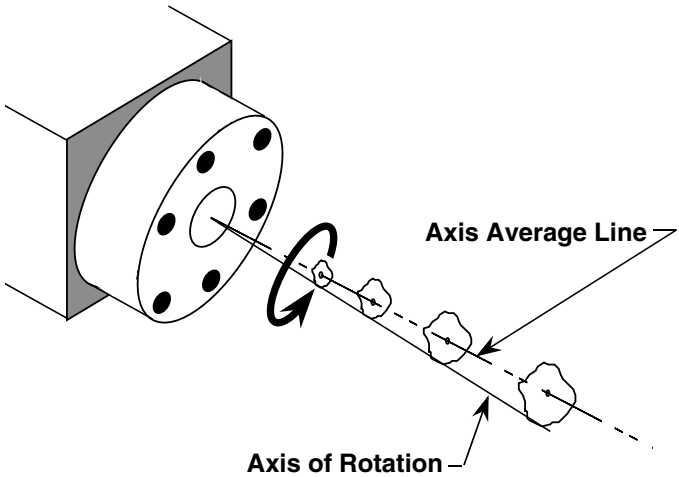


FIGURE 10.10 Determination of axis average line. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

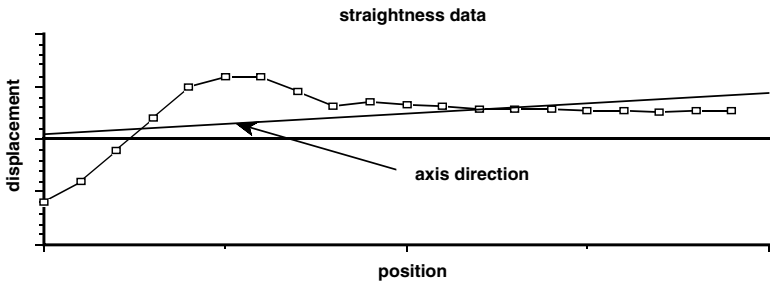


FIGURE 10.11 Determination of axis direction. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

The rotor of a spindle rotates about the average axis line as shown in Figure 10.7. Average axis line (shown in Figure 10.10) as defined in ANSI B5.54-1992[5] is

Average axis line: For rotary axes it is the direction of the best-fit straight line (axis of rotation) obtained by fitting a line through centers of the least-squared circles fit to the radial motion data at various distances from the spindle face.

The actual measurement of radial motion data is discussed later in this chapter.

Just as spindles must have a defined theoretical axis about which they rotate, linear slides must have a specific theoretical direction along which they traverse. In reality, of course, they do not track this axis perfectly. This theoretical axial line is the slide's equivalent of the average axis line for a spindle and is termed the axis direction:

Axis direction: The direction of any line parallel to the motion direction of a linearly moving component. The direction of a linear axis is defined by a least-squares fit of a straight line to the appropriate straightness data.

The best fit is necessary because the linear motion of a slide is never perfect. Figure 10.11 presents typical data used in determining axis direction in one plane. The position indicated on the horizontal scale is the location of the slide in the direction of the nominal degree of freedom. The displacement on the vertical scale is the deviation perpendicular to the nominal direction. The axis direction is the best-fit line to the straightness data points plotted in the figure. It should be noted

© 2002 by CRC Press LLC

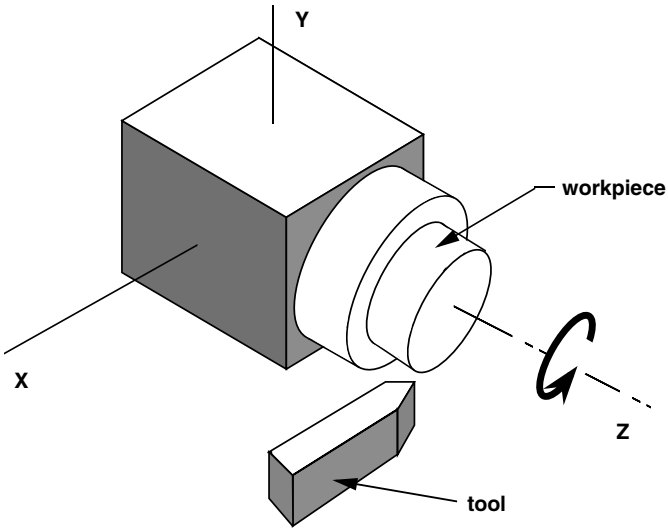


FIGURE 10.12 Sketch of a lathe configuration. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

that these data are plotted for two dimensions; however, three-dimensional data may be used as well (if necessary). Measurement of straightness data is discussed later in this chapter.

10.3.2 Sensitive Directions

Of the six error terms associated with a given axis, some will affect the machine tool's accuracy more than others. These error terms are associated with the sensitive directions of the machine tool. The other error terms are associated the machine's nonsensitive directions. Although six error terms are associated with an individual axis, certain error components typically have a greater effect on the machine tool's accuracy than others. Sensitivities must be well understood for proper machine tool design and accuracy characterization.

The single-point lathe provides an excellent example of sensitive and nonsensitive directions. Figure 10.12 and 10.13 depict a lathe and its sensitive directions. The objective of the lathe is to turn the part to a specified radius, R, using a single point tool. The tool is constrained to move in the X–Z plane of the spindle. It is clear that if the tool erroneously moves horizontally in the X–Z plane, the error will manifest itself in the part shape and be equal to the distance of the erroneous move. If the tool moves vertically, the change in the size and shape of the part is relatively small. Therefore, it can be said that the accuracy is sensitive to the X and Z axes nonstraightness in the horizontal plane but nonsensitive to the X and Z nonstraightness in the vertical plane (the Y direction in Figure 10.12). The error, S, can be approximated for motion in the vertical (nonsensitive) direction by using the equation:

$$S \approx \frac{1}{8}\frac{\varepsilon^2}{R}; \quad \varepsilon << R$$

Sensitive directions do not necessarily have to be fixed. While the lathe in Figure 10.13 has a fixed sensitive direction, other machine tools may have rotating sensitive directions. Figure 10.14 depicts a lathe which has a fixed sensitive direction (fixed cutting tool position relative to the spindle) and a milling machine with a rotating cutting tool that has a rotating sensitive direction. Because the sensitive direction of the mill rotates with the boring bar, it is constantly changing directions.[3,4]

© 2002 by CRC Press LLC

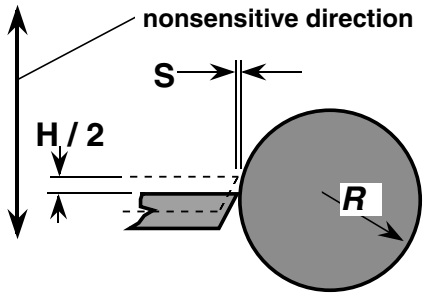


FIGURE 10.13 Sensitive direction for a lathe. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

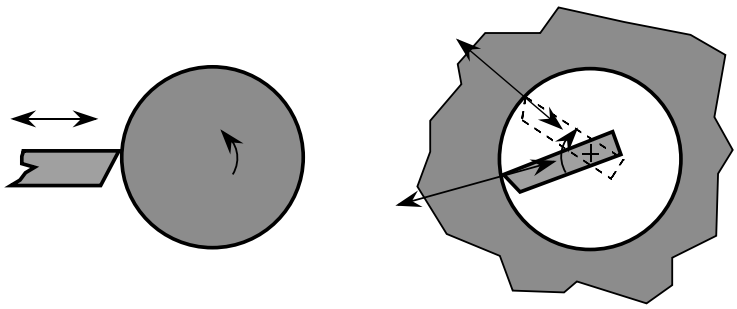


FIGURE 10.14 Fixed and rotating sensitive directions. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

10.3.3 Amplification of Angular Errors, The Abbe Principle

One of the most common errors affecting a machine's ability to accurately position a linear slide is Abbe error. Abbe error is a result of the slide's measuring scales (used for position feedback) not being in line with the functional point where positioning accuracy is desired. The resulting linear error at the functional point is caused by the angular motion of the slide that occurs due to nonstraightness of the guide ways. The product of the offset distance (from the measuring system to the functional point) and the angular motion that the slide makes when positioning from one point to another yields the magnitude of the Abbe error. Dr. Ernst Abbe (a co-founder of Zeiss Inc.) was the first person to mention this error.[10] He wrote, "If errors in parallax are to be avoided, the measuring system must be placed coaxially with the axis along which the displacement is to be measured on the workpiece." This statement has since been named "The Abbe Principle." It has also been called the first principle of machine tool design and dimensional metrology. The Abbe Principle has been generalized to cover those situations where it is not possible to design systems coaxially. The generalized Abbe Principle reads: "The displacement measuring system should be in line with the functional point whose displacement is to be measured. If this is not possible, either the slideways that transfer the displacement must be free of angular motion or angular motion data must be used to calculate the consequences of the offset."[11]

While the Abbe Principle is straightforward conceptually, it can be difficult to understand at first. However, a variety of examples exist that clearly show the effects of Abbe error. An excellent illustration of the Abbe Principle is to compare the vernier caliper with the micrometer. Both of these instruments measure the distance between two points, and are thus considered two point measurement instruments. Figure 10.15 shows these two instruments measuring a linear distance, D. The graduations for the caliper are *not* located along the same line as the functional axis of measurement. Abbe error is generated if the caliper bar is bent causing the slide of the caliper to

© 2002 by CRC Press LLC

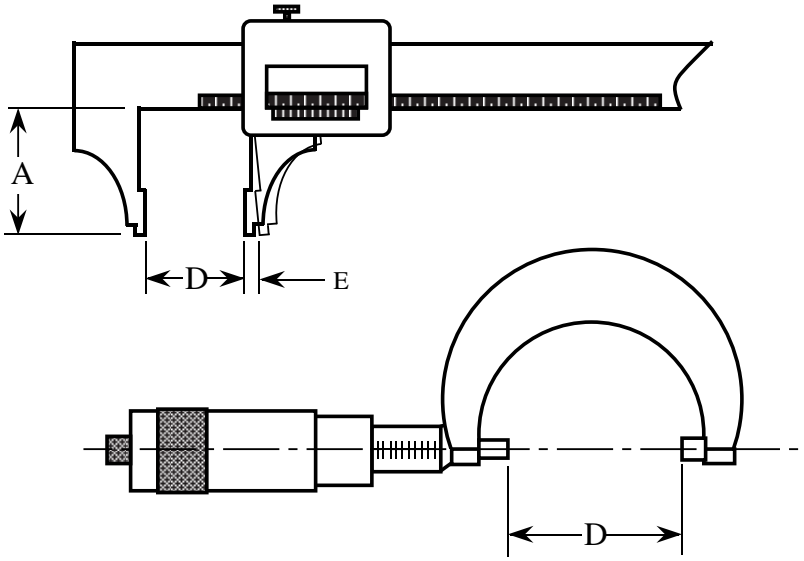


FIGURE 10.15 Micrometer and caliper comparison for Abbe offsets and errors. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

move through an angle θ when measuring D as shown in Figure 10.15. The distance, A, between the measurement graduations and the point of measurement is called the Abbe offset. In general, the Abbe error, E, is given by

$$E = A\sin(\theta)$$

Since the angle, θ, will be very small for most situations, the Abbe error can be accurately approximated as the product of the Abbe offset and the angle expressed in radians. Since most angular errors are measured in arc seconds, it is perhaps easier to remember that 1 arc sec is equal to approximately 4.8 μin/inch so the calculation becomes:

$$Abbe\ \ error\ \ (\mu in) = [Abbe\ \ offset\ \ (in)] \bullet [angular\ \ error\ \ (\sec)] \bullet [4.8(\mu in/in)]$$

The screw and graduated drum used in a micrometer are coaxially located to the distance being measured. Therefore, angular errors will have no effect on the measured distance, as the Abbe offset is zero. Thus, the micrometer obeys the Abbe Principle and is typically considered more accurate than the caliper.

Another excellent example of Abbe error is the height gauge shown in Figure 10.16. Here the slide of the gauge has a uniform angular motion of 10 arc sec error (that is exaggerated in the figure). This is the equivalent of a 100 μin. nonstraightness over the length of the slide. The probe arm of length 10 in. amplifies and transforms this angular error into a linear error in the height measurement by the following relationship

$$E = A\sin(\theta) = [10(in)] \bullet \sin[10(\sec)] = [10(in)] \bullet \sin[10(\sec)] = 0.000485\,in$$

Using the approximate relationship that 1 arc sec is equal to approximately 4.8 μin/inch, the error may also be computed as

$$E = [10(in)] \bullet [10(\sec)] \bullet [4.8(\mu in/in)] = 480\,\mu in = 0.000480\,in$$

© 2002 by CRC Press LLC

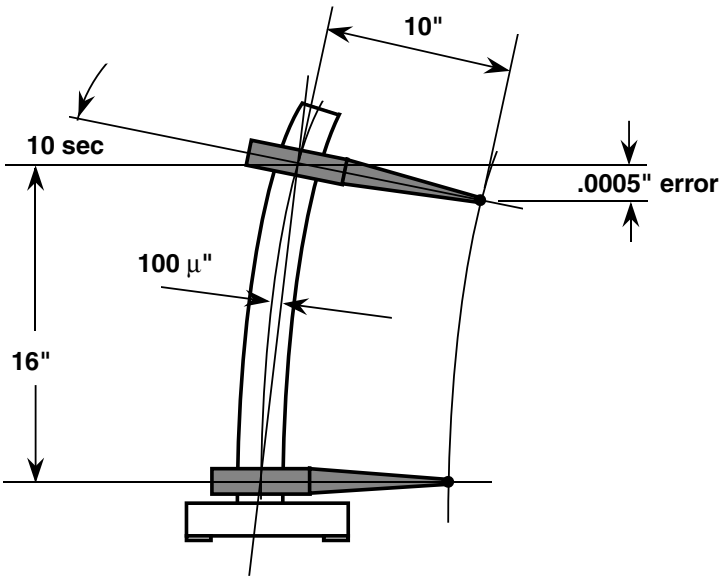


FIGURE 10.16 Abbe error for a height gauge. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

Thus, an error of approximately 485 μin is realized due to the angular motion error of the probe arm as it traverses the length of the height gauge slide.[3]

10.3.3.1 Reducing Abbe Error

There are three methods that may be used to reduce the effects of Abbe error. The first is to reduce the Abbe offset as much as possible. For example, placing measurement instrumentation coaxially with the points being measured, or placing templates for tracer lathes in the plane of the tool motion. Such modifications will eliminate Abbe error completely.

In many cases, machine designers are forced to place measurement devices at some distance from the functional measurement axis. The retro-fitting of machine tools with glass scales is an excellent example. In the retro-fit case, the replacement of the wheel gauge s (with typical resolutions of 0.0001 in.) on a machine tool with glass scale linear encoders that have an order of magnitude better resolution may cause the machine's positioning accuracy to be worse than the original design due to larger Abbe offsets for the glass scales. Such a retro-fit does not obey the Abbe Principle; however, the engineers effecting the retro-fit may not have an alternative to increasing the Abbe offset since it may be difficult to find a location to mount the linear scales that is close to the working volume.

Besides reducing the Abbe offset, designers may employ the two other methods to reduce the effects of Abbe error: (1) use slideways that are free of angular motion, or (2) use angular motion data to calculate the consequences of the offset (map out the Abbe error). Either of these two methods may be used to correct for Abbe error. However, slideways will never be completely free of angular motion, and tighter angular motion specifications can be expensive. Using angular motion data to correct the Abbe errors requires more calculations in the machine controller; however, with modern controllers these additional calculations are easily executed. Still, the best option is to minimize Abbe offsets before attempting to correct for them.[11]

10.3.3.2 The Bryan Principle

There is a corollary to the Abbe Principle that addresses angular error when determining straightness, known as the Bryan Principle. The Bryan Principle states that "The straightness measuring system should be in line with the functional point whose straightness is to be measured. If this is not possible, [two options are available] either the slideways that transfer the straightness must be

© 2002 by CRC Press LLC

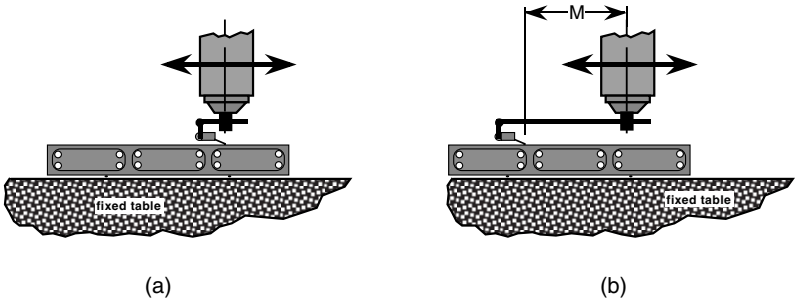


FIGURE 10.17 Visualization of the Bryan Principle for straightness measurements. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

free of angular motion or angular motion data must be used to calculate the consequences of the offset."[11] Either of these two options may be used to improve straightness measurements; however, they may require expensive modifications to the machine tool and its controller. As with the Abbe Principle, it is always best, if possible, to comply with the Bryan Principle and design machines with zero offsets.[11] Figure 10.17 demonstrates this principle with a fixed table straightness test. The set-up presented in Figure 10.17(a) obeys the Bryan Principle since the probe tip is located in line with the spindle axis. However, Figure 10.17(b) does not obey the Bryan Principle since the probe tip is at a distance, M, from the spindle axis.

10.4 Sources of Error and Error Budgets

As stated in the previous section, before any advanced mapping or control techniques should be employed to improve a machine tool's accuracy and repeatability, the designer should attempt to design the best machine possible. Of course, time, economic constraints, and physics will prevent the design engineer from achieving perfection and at best, the various error that a machine possesses will be greatly reduced. An error budget is the realization that a perfect machine without error cannot be constructed. The error budget is an attempt to separate and quantify a machine tool's errors into its basic components. These error components are then budgeted such that the combination of the various acceptable errors does not exceed the total desired error of the machine tool.

The error budget is developed before the machine is designed, and may be modified during the design process if the target accuracies cannot be achieved by redistributing the error components until a technically feasible and economically viable design is reached. By redistributing the errors, the design team can still maintain the target acceptable while allowing the contributions of the various individual errors to change. The error budget is both a guideline when designing a machine and a tool to determine the machine's final accuracy when the design process is complete. The error budget provides a set of goals to the design team and identifies the errors that are the most significant and those on which the most resources must be expended. This section briefly describes some of the major considerations in developing an error budget. The process is a long and tedious one as every component of error must be identified and quantified either precisely or statistically.

10.4.1 Sources of Errors

Generally, the sources of errors may be broken into four categories: geometric errors, dynamic errors, workpiece effects, and thermal errors. This section presents a brief discussion of these errors providing some insight into their causes and possible methods to reduce their effects.[4]

10.4.1.1 Geometric Errors

Geometric errors manifest themselves in both translational and rotational errors on a machine tool. Typical causes of such errors are lack of straightness in slideways, nonsquareness of axes, angular

© 2002 by CRC Press LLC

errors, and static deflection of the machine tool. Angular errors are, perhaps, the least understood and most costly of the various geometric errors. They are enhanced and complicated by the fact that they are typically amplified by the linear distance between the measurement device and the point of measurement (Abbe error). They are also the errors that can result in the largest improvement with simple design modifications like reducing the Abbe offset. With proper procedures, instrumentation and careful metrology, many errors can be identified, predicted and held within the desired level of the error budget.

10.4.1.2 Dynamic Errors

Dynamic errors are typically caused by machine tool vibration (or chatter). They are generated by exciting resonances within the machine tool's structure. Current research is investigating the prediction of vibrations in machine tools; however, from a practical perspective, this is quite difficult. Usually, a machine tool is built and its resonant frequencies are determined experimentally. The machine's controller can then be programmed to avoid combinations of feeds and speeds that may excite its various resonances. Typically, the best one can do during the design phases of machine tools is to design a structure that is stiff, light weight, and well damped.

10.4.1.3 Workpiece Effects

The workpiece can affect a machine tool's accuracy and precision in two manners: deflection during the cutting process and inertial effects due to motion. Deflection may be addressed by reducing the overall compliance of the machine tool. This is a relatively simple and well understood solution. It should be noted that most machine tool's are quite rigid by design, and it is usually the fixturing that provides the largest amount of compliance. For example, a lathe is typically a massive machine with an extremely rigid bed. However, the cutting tool or tool holder are often held in place by only a few small screws. Clearly, the stiffness of these components is small in comparison to the lathe bed, and are thus the weak point in the machine's structural loop. It is typically these weak points that yield the largest amount of stiffness increase with the least effort and design modification (i.e., it is easier to change a tool holder design, and typically more beneficial, than changing the design of the machine bed).[12]

Inertial effects of the workpiece, however, are not as simple to address. They become more pronounced with the increased speed that is associated with higher production rates. They are one of the critical limiting factors in high speed machining and typically their severity increases nonlinearly with respect to speed. Inertial effects may manifest themselves in several manners including asymmetry about a rotating axis and overshoot on a linear slide. If the part is asymmetric and is being turned on a lathe, the asymmetry may cause periodic spindle deviations reducing accuracy. A typical solution to these rotary problems is to balance the spindle with the workpiece mounted on it. For high speed spindles operating at over 200,000 rpm, balance levels under 3 mg are necessary for precision grinding operations. Other inertial effects are seen as large parts are moved rapidly in high production rate machines. Because of the high velocities and large masses of the workpieces, the machine tools may overshoot their target point. Basically, the machine's brakes are not powerful enough to stop the part at the desired position without overshooting that position. Proper design of servo systems as well as reasonable trajectories (smooth acceleration and velocity profiles) can substantially reduce inertial errors. Also, position probes used in conjunction with the machine tool can inform the controller if the workpiece is, indeed, tracking the proper trajectory.

10.4.1.4 Thermal Errors

Thermal errors are probably the most significant set of factors that cause apparent nonrepeatable errors in a machine tool. These errors result from fluctuating temperatures within and around the machine tool. They also result from nonfluctuating conditions at constant temperatures other than 20°C. Although deviations in machine tool geometry from thermal causes may be theoretically calculated, in practice such an analysis is difficult at best to successfully achieve even in the simplest of machine tools. Thus, proper thermal control is required.

© 2002 by CRC Press LLC

For typical machine tools, thermal errors may be caused by a wide variety of fluctuating heat sources including motors, people, coolant, bearings, and the cutting process. Furthermore, variations in the temperature of the environment may cause substantial thermal errors. For example, temperatures in a machine shop may vary from 95° F in the summer to 65° F or cooler in the winter. For higher precision machine tools, sources such as overhead lighting and sunlight may substantially contribute to thermal errors. Even windows or skylights in a machine shop may permit sunlight to shine on such machines during a specific time of day causing them to expand more than the tolerances that they are supposed to hold.

Thermal errors may be reduced substantially by proper procedure and design. For example, errors due to motors and bearings heating-up during use are reduced by warming-up the machine tool before it is used. Typically, high precision machine tools such as grinders are not shut down unless they are not being used for a substantial period of time. The grinding wheels for such grinders are kept spinning at their operational speed continuously, even when the machine is idle. This insures that the grinder's spindle motor and bearings as well as the grinding wheel are at a constant temperature. To further eliminate thermal effects, coolant temperature as well as environmental temperature should be controlled. The target temperature for the machine tool's environment and coolant is typically 20°C (68°F) which is the national (and international) temperature at which all distance measurements are made.[13]

Finally, there are several design techniques that may be employed to reduce thermal effects, including reducing the thermal capacitance of the machine tool. This permits the entire machine tool to thermally equilibrate rapidly rather than have thermal gradients, thus reducing the amount of time required for the system to warm-up. The use of materials with similar coefficients of thermal expansion (C_{te}), or the kinematic isolation of materials with different C_{te} will reduce thermally induced stresses in the system. For example, glass scales having a low C_{te} are often fixed at both ends to steel machines having a higher C_{te}. When the temperature of the machine varies, the steel structure will deform more from the thermal variations than the glass scale. Since the scale is significantly less rigid than the steel structure, the scale may undergo deformation as scale and structure deform at different rates. This could generate an error in the measurement system. A solution to this problem is to fix the scale at one end, and mount the other end of the scale such that there is compliance in the scale's sensitive direction. When the two bodies change size at different rates the stresses are then mostly absorbed by the compliant mount.[3,14]

10.4.2 Determination and Reduction of Thermal Errors

The environment in which the machine tool operates has a significant effect on the performance of the machine tool. Typically, in high precision applications thermal effects are the largest single source of errors (Bryan, 1968).[15] Figure 10.18 is a block diagram depicting various sources of thermal disturbances that influence machine tools. As stated in ANSI B5.54, "Thermally caused errors due to operating a machine tool in a poor environment cannot be corrected for by rebuilding the machine tool, nor are they grounds for rejection of a machine tool during acceptance test unless the machine is specified to operate in that particular environment."[5] Furthermore, thermal error cannot be completely eliminated by enhanced control algorithms or the addition of sensors. The reality is that it is simpler and more cost effective to limit thermal effects than to attempt to compensate for them. This section briefly discusses basic concepts of thermal behavior characterization, and simple methods to limit errors caused by varying thermal conditions.

To quantify the effects of thermal errors on a machine tool's performance, the Thermal Error Index (TEI) is used. The TEI is the summation, without regard to sign, of the estimates of all thermally induced measurement errors, expressed as a percentage of the working tolerance or total permissible error. The TEI and its computation are thoroughly explained in ANSI B89.6.2-1973.[13] The computational procedures account for uncertainties in the quantification of various parameters such as expansion coefficients and the differential expansions of various materials when machines

© 2002 by CRC Press LLC

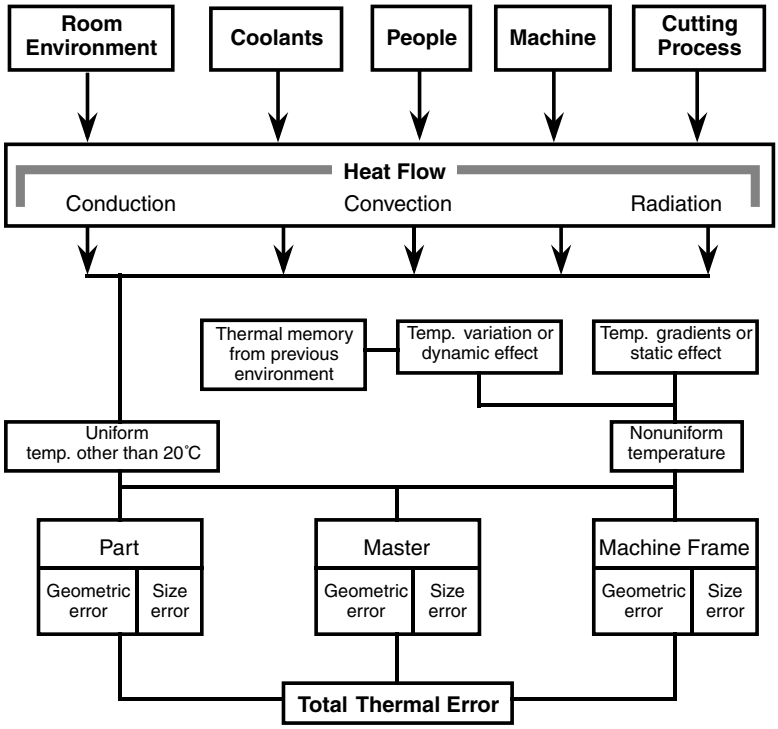


FIGURE 10.18 Factors thermally affecting machine tool environment. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

are operated at temperatures other than 20°C. The ANSI standard B5.54-1992[5] provides a method of using the TEI to develop contractual agreements for the purchasing and selling of machine tools and manufactured parts. It states that calibration, part manufacture and part acceptance procedures are valid if all pertinent components of the system are at 20°C, or it can be shown that the TEI is a reasonable and acceptable percentage of the working tolerance.

An important value used in the computation of the TEI is the temperature variation error (TVE). The TVE is the maximum possible measurement error induced solely by the deviation of the environment from average conditions. In particular this applies to repeatability, linear displacement accuracy and telescoping ball bar performance measurement results. The TVE may be determined from measurements using a standard drift test. Figure 10.19 presents a schematic for a three-axis drift test using three orthogonally positioned air bearing LVDTs (linear variable differential transformer).[5] Once the set-up in Figure 10.19 is established, the LVDT signals are sampled and recorded over an extended time period (typically 24 hours). The results are used to quantify the amount of error motion that is generated along three orthogonal directions via thermal drift over a long period in time. The error recorded in a drift test is often used to provide a bound on the repeatability of a machine tool since a machine's repeatability clearly cannot be smaller amount of drift that it experiences.

There are several factors that must be considered if the machine tool and environment are to be thermally controlled. The first is the temperature of the machine's environment. The defined standard temperature at which machine tools should be calibrated is 20°C (68°F). Proper temperature control of the ambient air around the machine tool is critical in high precision operations. This includes temperature control of the environment as well as providing sufficient circulation to remove any excess heat generated by the system. Even seemingly small heat sources such as lights and sunlight can substantially add to thermal errors.

© 2002 by CRC Press LLC

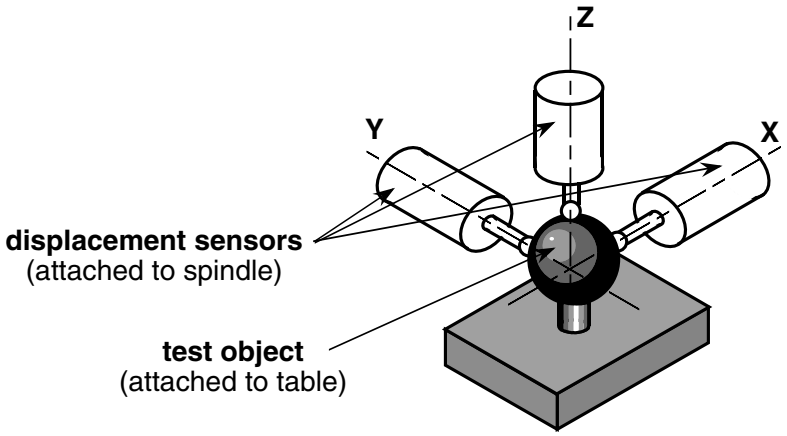


FIGURE 10.19 Three-axis drift test. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

It is also critical to control the temperature of the coolant. Variations of coolant temperature of more than 30°F are typical in many plants depending on the time of the year and even the time of the day, in particular if a central coolant system is used. Furthermore, a constant flow rate of coolant should be supplied to the machine tool to eliminate any type of time dependent thermal gradients in the machine. In fact, some machine tools are *oil showered* specifically to engulf the machine tool in temperature controlled oil. Even the composition of the coolant is critical for temperature control. Water based coolants will evaporate, cooling the machine more than expected depending on environmental conditions. This will cause changing thermal gradients over time and yield thermal errors in the machine tool (Bryan et al., 1982).[16] However, water-based coolants are currently preferred over oil-based ones because they are not as harmful (toxic) the environment and are not nearly as flammable. Thus, evaporation effects of water-based coolant should be considered if it becomes necessary to use them in high precision operations.

Clearly, thermal gradients will exist in a machine tool; however, it is important that these gradients remain constant with respect to time. For example, a large electric motor on a lathe will generate heat. Ideally, it is best to remove this heat. However, in reality sections of the machine tool that are nearest to the motor will have a higher temperature. Thus, from a spatial perspective, thermal gradients exist. However, as long as those gradients do not change in a temporal fashion, the machine tool's repeatability will not be significantly affected by the thermal gradients.[3]

10.4.3 Developing an Error Budget

The error budget is based on the behavior of individual components of the machine tool as well as their interactions with other components. Since no machine is perfect, error exists in positioning the cutting tool relative to the workpiece. This error is called the tool positioning error (TPE). The error budget is concerned with determining the effect of system variations (systematic and non-systematic) on the TPE. The error budget should contain as many of the sources of error as possible. The effects of each source of error on the TPE must also be well understood. Large error components that are in highly sensitive directions (thus, contributing greatly to the TPE) should be primary concerns. Other error components with lower sensitivities may be too small to be considered until larger TPE components have been reduced.

To properly use an error budget, two tasks must be undertaken:

1. Determine the sources of error within the machine tool and its environment.
2. Determine how those sources of error combine to affect the TPE.

© 2002 by CRC Press LLC

This chapter is limited to a brief discussion on the identification and combination of errors that affect the machine tool. However, extensive research has been conducted on these issues, and it is recommended that an engineer be familiar with the literature before using an error budget.[4] This section is concerned with combining error components that affect a machine tool to yield their overall effect on the TPE in a particular direction.

The errors discussed in the previous section may be placed into three categories when developing an error budget:

1. Random, which under apparently equal conditions at a given position, does not always have the same value and can only be expressed statistically.
2. Systematic, which always has the same value and sign at a given position and under given circumstances. (Wherever systematic errors have been established, they may be used for correcting the value measured.)
3. Hysteresis is a systematic error (which in this instance is separated out for convenience). It is usually highly reproducible, has a sign depending on the direction of the approach, and a value partly dependent on the travel. (Hysteresis errors may be used for correcting the measured value if the direction of the approach is known and an adequate pretravel is made.)[17]

Systematic errors, e_{sys}, may be considered vector quantities possessing both magnitude and direction that may be added in a vector sense. That is to say that all systematic errors of a machine tool along a particular axis may be summed together to yield the total systematic error. Because the errors do possess direction (positive or negative in a specified direction), individual errors may either increase the total system error or actually reduce the error via cancellation.

Random errors, however, must be treated via a statistical approach. The portions of an error budget that represent random errors are *always* additive. That is to say they will always make the error larger because the sign of their direction as well as the magnitude of the error is a random quantity. The assumption here is that nature will work against the machine designer and generate error components that increase the overall machine error. One cannot assume that one will be lucky and have a random error component reduce the overall system error.

Root mean square (RMS) error is often used to quantify random errors where the random errors tend to average together. The combined random RMS error is computed as the geometric sum of the individual RMS errors. Thus, for N random error components, the total RMS error is given by

$$\left(RMS_{tot}\right)_i = \sqrt{\left(\sum_{j=1}^{N}\left(RMS_j\right)^2\right)}$$

where RMS_j is the j^{th} component of random error in the i^{th} direction. This results in a total overall error of

$$\left(e_{RMS}\right)_i = \left|\sum\left(e_{sys}\right)_i + \sum\left(e_{hyst}\right)_i\right| + \left(RMS_{tot}\right)_i$$

where $(e_{sys})_i$ and $(e_{hyst})_i$ are the systematic error and hysteresis error of the system along the i^{th} axis. The absolute values about the systematic and hysteresis error make them positive quantities which are added to the always positive quantity of the random error. This reflects the fact that random error can only increase the total error; however, systematic errors may cancel each other.

Quite often, random errors are described in terms of a total peak-to-valley amplitude, PV. PV_j may be considered the separation of two parallel lines containing the j^{th} error signal. PV_j is related to RMS_j by the following equation

$$PV_j = \left(K_j\right)\left(RMS_j\right)$$

© 2002 by CRC Press LLC

where K_j is a scalar quantity that depends on the error signal's probability distribution. The values of K_j for uniform and $\pm 2\sigma$ Normal (Gaussian) distributions are 3.46 and 4, respectively. Typically, the value for the uniform distribution ($K_j = 3.46$) is used, since individual error traces are not generally normally distributed. If there are some central tendencies for the distribution, the uniform assumption will be conservative.[18] Using the relationship for PV_i given above, the total random error generated by combining N random error components in the i^{th} direction is given by

$$\left(e_{PV,rand}\right)_i = \frac{1}{2\sqrt{3}}\left[\sum_{j=1}^{N}\left(PV_j\right)^2\right]_i$$

The total error in the i^{th} direction for the peak-to-valley scenario is

$$\left(e_{PV}\right)_i = \left|\sum\left(e_{sys}\right)_i + \sum\left(e_{hyst}\right)_i\right| + \left(e_{PV,rand}\right)_i$$

It should be noted that these error values are based on a probabilistic estimation. Therefore, the actual error may be smaller or larger than the estimated value. Depending on the value that is used for K_j, the designer may estimate the probability of the error estimate being either too small or too large. It must be remembered that the above equations only provide for an estimation of the error and cannot provide a precise quantity, only a bound with a given probability. However, using these relationships with a K_j for a uniform distribution is the procedure that is practiced by many designers.

10.5 Some Typical Methods of Measuring Errors

Multi-axis machine tools have a wide variety of parametric error sources that may be determined using a broad spectrum of approaches. This section presents a few of the most common and important techniques for addressing scale errors, straightness errors, and radial motion of a spindle (or rotary table). The techniques discussed are not the only techniques available to qualify machine tools; however, they are a set of powerful tools that are relatively easy to implement and quite useful.

Before the various procedures for error measurement are described, it is worth while to discuss the laser measurement system, one of the most versatile measurement systems available to the metrologist. The laser measurement system may be used to measure linear displacement, angular displacement, straightness, squareness, and parallelism. The laser measurement system, often referred to as a laser interferometer, consists of the following components:

1. The laser head that is the laser beam's source.
2. A tripod or stand on which the laser head is mounted.
3. An air sensor to measure the temperature, humidity, and barometric pressure of the ambient air.
4. A material sensor to measure the temperature of the machine tool's measurement system.
5. A linear interferometer that actually performs the interference measurements.
6. A linear retro-reflector (or measurement corner cube) to reflect the laser beam off of the point being tracked.
7. A reference corner cube to split and recombine the beam generating the beam interference needed for the interferometer.

Figure 10.20 is a drawing of a laser interferometer and its components set-up for a linear displacement test.

Figure 10.21 is a schematic of the basic operational configuration of a laser measurement system. The beam originates in the laser head and is sent through the reference corner cube where it is

© 2002 by CRC Press LLC

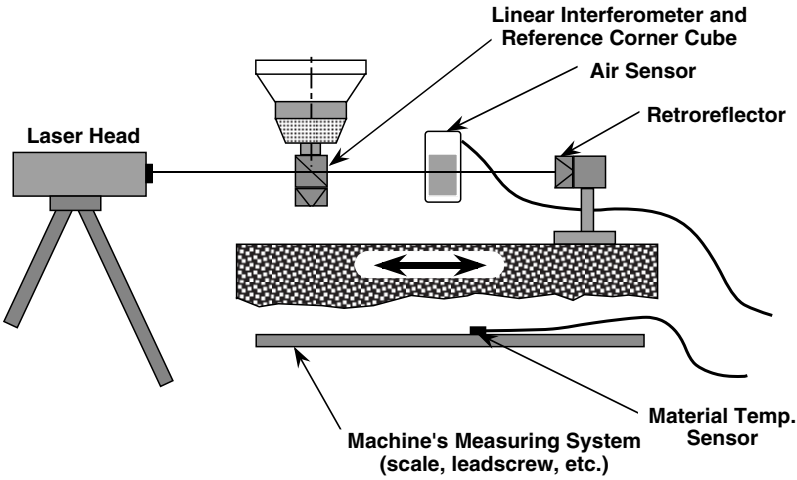


FIGURE 10.20 Laser interferometer set-up for linear displacement test. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

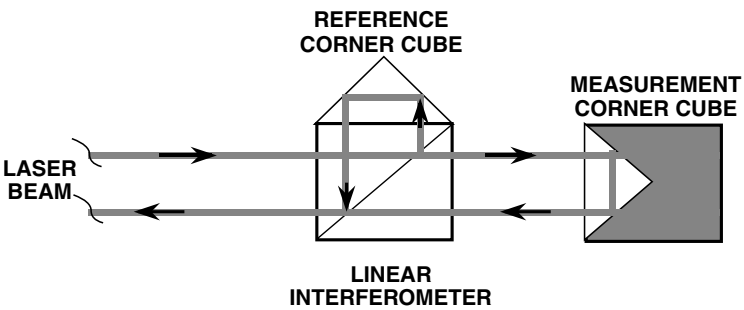


FIGURE 10.21 Laser path.

split. Part of the beam continues to the measurement corner cube where it is reflected back towards the reference corner cube. The two beams are then recombined in the reference corner cube where they may combine (interfere) in a constructive or destructive manner. The combined beam then continues to the interferometer. The interferometer measures the amount of interference between the two beams, and determines the distance traveled between the initial location of the measurement corner cube and its current position.

Laser interferometers typically use either a single or multiple frequency Helium-Neon gas laser. The interferometer simply counts the number of wavelengths that the slide traverses between two points. Thus, the laser interferometer can only measure relative displacements as opposed to absolute distances. It can only inform the operator as to the number of wavelengths of light between two points. The wave length of the laser is typically stabilized and known to better than 0.05 parts per million.

There are three basic guidelines in setting-up the laser measurement system:

1. Choose the correct set-up to measure the desired parameter (e.g., distance) and verify the directional signs (±) of the system.
2. Approximate the machine tool's working conditions as closely as possible. For example, make sure that the machine tool is at its operational temperature. Machine tool scales may be made of material that will change length as their temperature varies. This change in length directly affects their position output.
3. Minimize potential error sources such as environmental compensation, dead path, and alignment.

© 2002 by CRC Press LLC

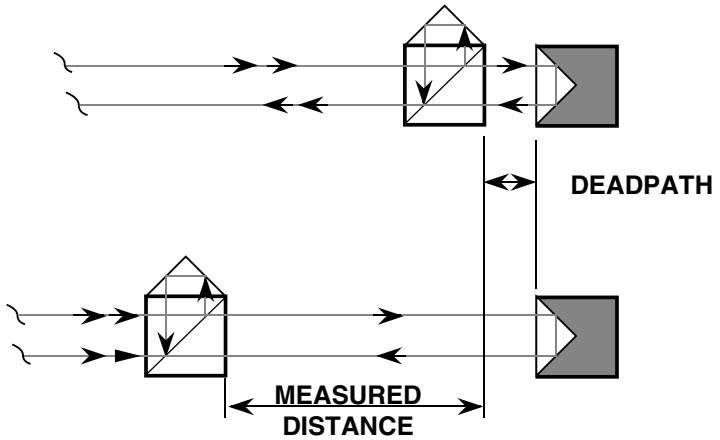


FIGURE 10.22 Laser interferometer dead path. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

The potential error sources from the environment are variations in air temperature, humidity, and barometric pressure. These affect the wavelength of light in the atmosphere. The wavelength of the laser light will vary one part per million for each

1°C (2°F) change in air temperature
2.5 mm (0.1 inch) Hg change in absolute barometric pressure
30% change in relative humidity

As a comparison, if the machine's scales are made of steel they will expand or contract one part per million for every 0.09°C (0.16°F). The accuracy of the laser interferometer is directly determined by how accurately the ambient conditions are known.

Typically, laser interferometers come equipped with environmental measurement systems that are capable of tracking the temperature, barometric pressure, and humidity during a test. This information is used to electronically alter the displacement values, compensating for the change in the velocity of light in air under the measured conditions. Thus, proper compensation can eliminate most environmental effects on the system. There is, however, an area known as the dead path where compensation for the velocity of light error is not applied. The dead path, shown in Figure 10.22, is the distance between the measurement corner cube and the reference corner cube when the laser interferometer is nulled or reset. The compensation for the velocity of light error is applied only to the portion of the path where displacement is measured as shown in Figure 10.22. To minimize the dead path error, the unused laser path must be minimized by placing the reference corner cube as close to the measurement corner cube as possible. The interferometer should then be reset, and the set of distance measurements made by moving the reference corner cube away from the measurement corner cube. Changes in the ambient environmental conditions during the measurement will only be considered for the measured distance and not for the dead path. However, if the dead path is small and the measurements are made over a short time period, the ambient conditions typically will not change enough to generate significant velocity of light errors. Dead path error may be further reduced by having a well controlled environment.

Misalignment of the laser beam to the linear axis of motion of the machine tool will result in an error between the measured distance and the actual distance. This error is typically called cosine error and is depicted in Figure 10.23. If the axis of motion is misaligned with the laser beam by an angle, θ, then the measured distance, $L_{measured}$, is related to the actual machine distance, $L_{machine}$ by the following equation

$$L_{measured} = L_{machine} \cos(\theta)$$

© 2002 by CRC Press LLC

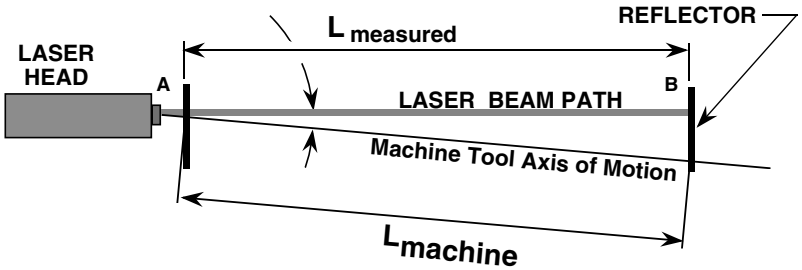


FIGURE 10.23 Geometry for cosine error. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

Cosine error will always result in the measured value being less than the actual machine value when the machine and the reference are perfect ($L_{measured} < L_{machine}$).

The significant advantage in using the laser measurement system is that it is dependent only on a well-calibrated wavelength of its laser source for measurement. That is to say, its measurement standard is related directly to fundamental characteristics defined in physics. Linear measurement devices such as micrometers, are typically calibrated to gauge blocks that are calibrated against other gauge blocks that eventually can be traced back to a formal calibration at a calibration laboratory such as the one at NIST (National Institute of Standards and Technology). Thus, the traceability of an individual measurement can be established. However, the laser measurement system need only be traced back to the wavelength of light; thus, it is a powerful tool in the metrologists' arsenal. When properly used, the laser measurement system is a powerful tool that is useful for determining many types of errors. It is very important to understand the basic theory of the laser measurement system's operation and the correct procedures before using it. If employed improperly, it can easily generate erroneous results that may not be at all obvious.

10.5.1 Linear Displacement Errors

As previously stated the linear displacement error is the difference between where the machine's scale indicates that a carriage is and where the carriage is actually located. To determine linear displacement error an accurate external reference device for measuring travel distance must be used. Typically, a laser measurement system is employed for this task. This section is concerned with the use of the laser measurement system to measure linear displacement.

The determination of linear displacement errors is accomplished by a simple comparison of the linear scale output to that of the laser interferometer at different locations along a particular machine tool slideway. The set-up for such a measurement is shown in Figure 10.20. The laser measurement system should be set-up in accordance with the procedures previously outlined, minimizing errors such as dead path errors, cosine errors, and environmental errors. The table of the machine tool is then moved in increments of a given amount along the length of the slideway. At each interval, the table is brought to a stop, and the distance traveled is computed from data gathered from the machine's scales. The scale distance is compared to the distance measured using the laser interferometer. The difference between the two distances is the linear displacement error. These measurements and comparisons are repeated several times along the entire length of the slideway, mapping the scale errors for the slideway. It should be noted that the linear displacement error includes not only the machine's scale errors, but the Abbe errors due to angular motion of the carriage.

10.5.2 Spindle Error Motion — Donaldson Reversal

As was discussed earlier, when a spindle or rotary table rotates, it has some error motion in the radial direction termed radial motion. It is important to measure the amount of radial motion in order to characterize spindle performance and understand the amount of error contributed by the

© 2002 by CRC Press LLC

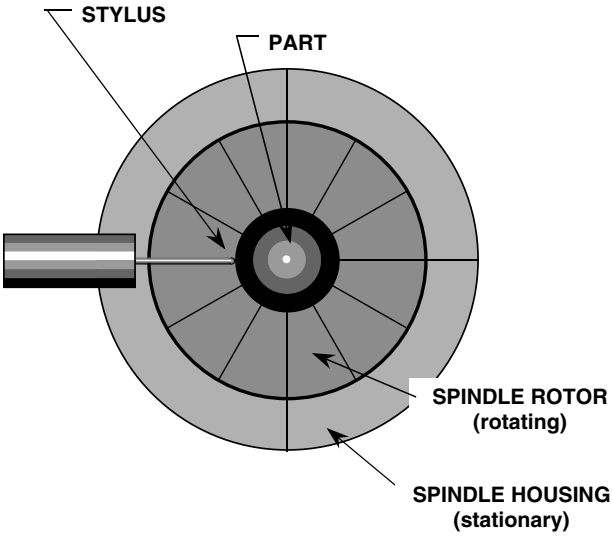


FIGURE 10.24 Radial motion set-up. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

spindle or rotary table to the machine tool's total error. To measure the radial motion of a spindle or rotary table, a precision ball is centered on the axis of rotation of the table and rotated. A probe is placed on the surface of the ball and radial deviations of the probe tip are recorded (see Figure 10.24). If the precision ball was perfect and it was perfectly centered, the signal from the probe would be the radial motion of the table.

Unfortunately, the precision ball is not a perfect sphere and the resulting probe signal is a combination of the radial motion of the spindle and the imperfections in the ball. Donaldson developed a method for completely separating gauge ball nonroundness from spindle radial motion.[19] This method has been termed Donaldson ball reversal. All that is needed for ball reversal is:

1. A spindle with radial motion that is approximately an order of magnitude less than the value of roundness desired (this is a rule of thumb).
2. An accurate indicator (preferably electronic).
3. Recording media (polar chart or a computer).

The following assumptions are made:

1. The radial motion is repeatable.
2. The indicator accurately measures displacement.

There are two set-ups for ball reversal that are shown in Figure 10.25. In the first set-up, the ball is mounted on the spindle with point B of the ball located at point A on the spindle. The stylus of the probe is located at point B on the ball. The spindle is then rotated 360° and the motion of the stylus is recorded. The signal from the stylus, $T_1(\theta)$ is given by the sum of the nonroundness of the gauge ball, $P(\theta)$, and the radial motion of the spindle, $S(\theta)$

$$T_1(\theta) = P(\theta) + S(\theta)$$

The spindle is then rotated back 360° and the gauge ball is relocated on the spindle such that point B is rotated 180°, and is at a position opposite to point A on the spindle. The probe is also positioned opposite point A and brought into contact with the gauge ball at point B. The spindle is once again rotated 360° and the data from the probe are recorded. The signal from the probe, $T_2(\theta)$ is

© 2002 by CRC Press LLC

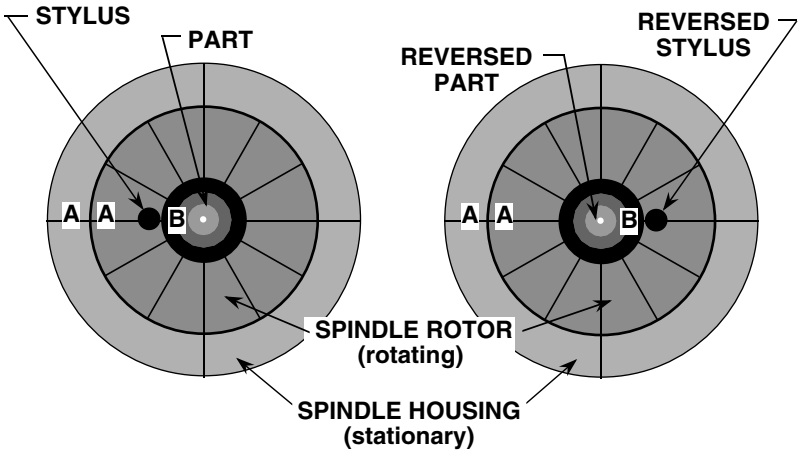


FIGURE 10.25 Donaldson ball reversal set-up. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

$$T_2(\theta) = P(\theta) - S(\theta)$$

From the two data sets, $T_1(\theta)$ and $T_2(\theta)$, the spindle radial motion may be computed as

$$S(\theta) = \frac{T_1(\theta) - T_2(\theta)}{2}$$

and the gauge ball nonroundness may be computed as

$$P(\theta) = \frac{T_1(\theta) + T_2(\theta)}{2}$$

This set of simple linear combinations of $T_1(\theta)$ and $T_2(\theta)$ provides information on both the ball and the spindle without using secondary or intermediate standards. The method is also independent of the errors in either the precision ball or the spindle. Thus, it is considered a self-checking method.[19]

If the spindle does not use rolling elements (e.g., an air aerostatic or hydrostatic bearing) then the spindle does not need to be rotated backwards 360° degrees between the two set-ups. Rotating the spindle back 360° between set-ups is necessary to insure that all of the rolling elements exactly repeat the same motions each time the data are taken. Furthermore, if the spindle is being used as a rotary axis, then is should only be used for the 360° measured by the reversal method. If the use of a rotary table with rolling element bearings exceeds the test rotation range, then the measured radial motion of the table, $S(\theta)$, will not correctly represent the radial motion of the table outside of the original 360° range. If more rotation than 360° is necessary, then the reversal should be done for the entire range of rotation that will be used.

10.5.3 Straightness Errors — Straight Edge Reversal

As was discussed earlier, when a machine table moves along a slideway, it experiences straightness errors along the slide perpendicular to the axis of travel. The straightness errors must be measured to determine the amounts and directions of error that the slideway nonstraightness is contributing to the overall machine tool error. To measure the nonstraightness of a slideway, a straight edge is placed on the machine table parallel to the axis direction. A probe is placed normal to the surface of the straight edge and deviations of the probe tip are recorded (see Figure 10.26). The resulting

© 2002 by CRC Press LLC

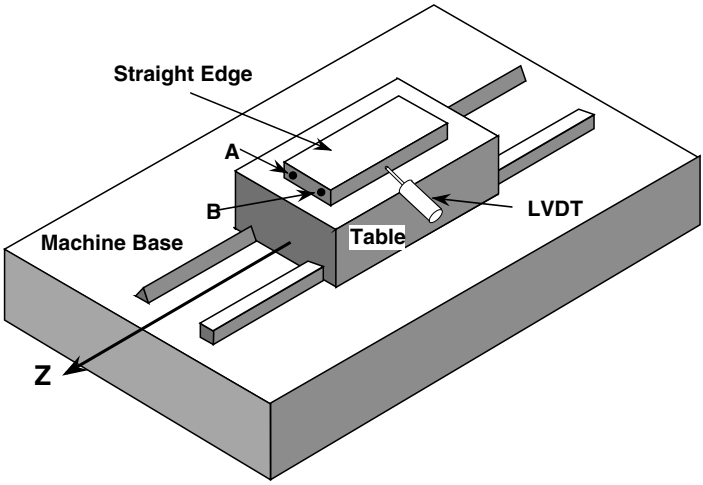


FIGURE 10.26 Nonstraightness measurement (first set-up). (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

probe signal is the nonstraightness of the slideway, the nonstraightness of the straight edge, and the nonparallelism of the straight edge to the axis. (If the slideway was perfectly straight and the straight edge was also perfect, the signal from the probe would be a straight line.)

In a fashion similar to Donaldson ball reversal, a method termed straight edge reversal can be used to separate the nonstraightness of the straight edge from the nonstraightness of the slideway. All that is needed for straight edge reversal is:

1. A straight edge that has a length equal to the length of the slideway to be measured.
2. An accurate indicator (preferably electronic).
3. Recording media (strip chart or a computer).

The following assumptions are made:

1. The slideway straightness error is repeatable.
2. The indicator accurately measures displacement.

There are two set-ups for straight edge reversal. The first is shown in Figure 10.26, and the second is shown in Figure 10.27. In the first set-up, the straight edge is mounted on the table with a three point kinematic mount. Point B of the straight edge is located at the front of the table and point A at the rear of the table. The stylus of the probe is located on the side of the straight edge nearest to point B. The table is then moved along the entire length of the slideway and the motion of the stylus is recorded. The signal from the stylus, $T_1(Z)$ is given by the sum of the nonstraightness of the straight edge, P(Z), and the nonstraightness of the slideway, S(Z)

$$T_1(Z) = P(Z) + S(Z)$$

The table is then positioned back to its original starting point and the straight edge is relocated (flipped) on the table such that point B is at the rear of the table and point A is at the front of the table as shown in Figure 10.27. The probe is also moved to the rear of the table such that it is in contact with the side of the straight edge that is nearest point B. The table is once again moved along the entire length of the slideway and the data from the probe are recorded. The signal from the probe, $T_2(Z)$ is

$$T_2(Z) = P(Z) - S(Z)$$

© 2002 by CRC Press LLC

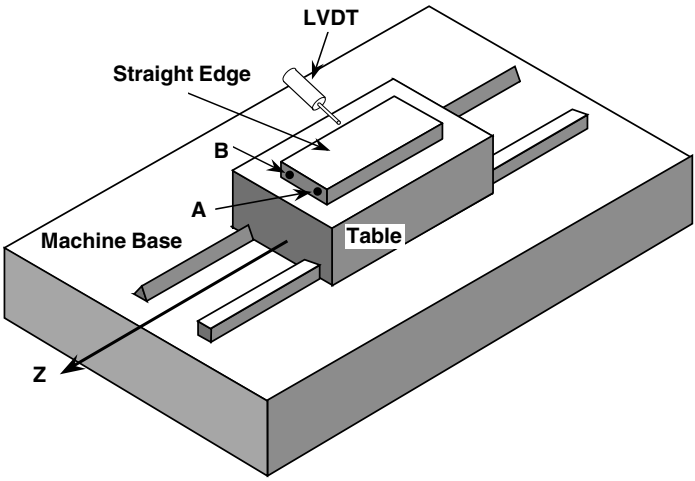


FIGURE 10.27 Second set-up for straight edge reversal. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.*)*

From the two data sets, $T_1(Z)$ and $T_2(Z)$, the slideway nonstraightness error may be computed as

$$S(Z) = \frac{T_1(Z) - T_2(Z)}{2}$$

and the straight edge nonstraightness error may be computed as

$$P(Z) = \frac{T_1(Z) + T_2(Z)}{2}$$

This set of simple linear combinations of $T_1(Z)$ and $T_2(Z)$ provides information on both the straight edge and the slideway without using secondary or intermediate standards. The method is also independent of the errors in either the straight edge or the slideway. Thus, it is considered a self-checking method.[20]

10.5.4 Angular Motion — Electronic Differential Levels

The angular motion about axes may be determined using a variety of tools including laser measurement system, autocollimator, and electronic differential levels. This section presents angular motion measurement using a set of electronic levels. This technique is simple and the levels are relatively inexpensive in comparison to a laser measurement system or autocollimator. Since electronic levels use gravity as a reference, they are limited to angular motion about axes in a horizontal plane. Thus, roll and pitch errors may be determined for axes in the horizontal plane, and pitch and yaw may be determined for vertical axes.

The electronic level is an instrument that measures small angles using the direction of gravity as a reference. A typical set-up for determining the pitch of an axis is shown in Figure 10.28. The two levels, A and B, are used differentially in one plane yielding the angular motion of one level relative to the other level. Level A is located in the tool location, and level B is located where the workpiece is mounted. These locations insure that the angular motions computed will be those that are experienced between the workpiece and the tool. To perform the measurement, the table is moved along its entire length, stopping at fixed distances along the length of the slideway. It is important that the table is brought to a complete stop at each point where the readings are taken. This permits the levels to stabilize so that accurate data can be recorded. The two levels are then read and

© 2002 by CRC Press LLC

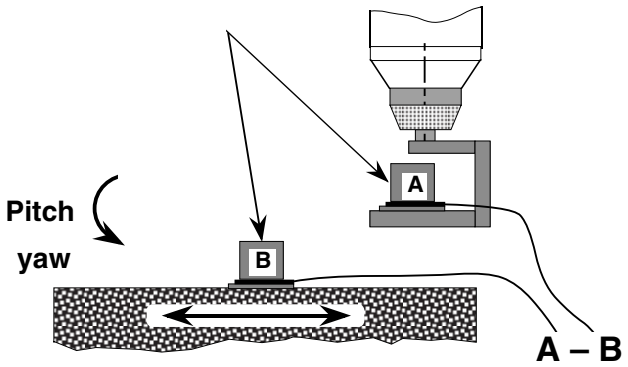


FIGURE 10.28 Pitch measurement set-up using electronic levels. (From Dorf, R. and Kusiak, A., *Handbook of Design, Manufacturing, and Automation*, John Wiley, New York, 1994. With permission.)

the value from B is subtracted from A resulting in the relative angular motion between the two levels. The table is then moved to the next location where another reading is taken. This procedure is repeated until the angular motion for the entire axis is mapped. The roll of the slide may be measured by simply rotating the two levels 90° about the vertical axis and repeating the procedure. Yaw measurement requires the use of either a laser angular interferometer or an autocollimator.

10.6 Conclusion

Precision manufacturing is continuously changing as technological advances and consumer demands push machine accuracy, resolution, and repeatability to ever improving levels. This chapter has presented some of the basic ideas, principles, and tools used to design high precision manufacturing systems. There are a plethora of other concepts available to engineers designing precision machine tools or metrology systems, and the reader is encouraged to make use of the references provided throughout this chapter. In conclusion, adherence to fundamental principles and the combination of good design, metrology, and practice are necessary to realize machine tools of the highest precision.

10.7 Terminology

Accuracy is formally defined as quantitative measure of the degree of conformance to recognized national or international standards of measurement.

Angular errors are small unwanted rotations (about X, Y, and Z axes) of a linearly moving carriage about three mutually perpendicular axes.

Angularity is the angular error between two or more axes designed to be at fixed angles other than 90°.

Average axis line for rotary axes is the direction of the "best fit" straight line (axis of rotation) obtained by fitting a line through centers of the least squared circles fit to the radial motion data at various distances from the spindle face.

Axial error motion is the translational error motion collinear with the Z reference axis of an axis of rotation (about the Z axis).

Axis direction is the direction of any line parallel to the motion direction of a linearly moving component. The direction of a linear axis is defined by a least squares fit of a straight line to the appropriate straightness data.[5]

Error is defined as the difference between the actual response of a machine to a command issued according to the accepted protocol of the machine's operation and the response to that command anticipated by the protocol.

© 2002 by CRC Press LLC

Error motion is the change in position relative to the reference coordinate axes, or the surface of a perfect workpiece with its center line coincident with the axis of rotation. Error motions are specified as to location and direction and do not include motions due to thermal drift.

Error motion measurement is a measurement record of error motion which should include all pertinent information regarding the machine, instrumentation, and test conditions.

Face motion is the rotational error motion parallel to the Z reference axis at a specified radial location (along the Z axis).

Parallelism is the lack of parallelism of two or more axes (expressed as a small angle).

Radial error motion is the error motion of rotary axis normal to the Z reference axis and at a specified angular location (see Figure 10.5).[5]

Radial error motion is the translational error motion in a direction normal to the Z reference axis and at a specified axial location (along the X and Y axes).

Repeatability is formally defined as a measure of the ability of a machine to sequentially position a tool with respect to a workpiece under similar conditions.

Resolution is the least increment of a measuring device; the least significant bit on a digital machine.

Runout is the total displacement measured by an instrument sensing a moving surface or moved with respect to a fixed surface.

Scale errors are the differences between the position of the readout device (scale) and that of a known reference linear scale (along the X axis).

Slide straightness error is the deviation from straight line movement that an indicator positioned perpendicular to a slide direction exhibits when it is either stationary and reading against a perfect straight edge supported on the moving slide, or moved by the slide along a perfect straight edge which is stationary.

Squareness is a plane surface that is "square" to an axis of rotation if coincident polar profile centers are obtained for an axial and face motion polar plot at different radii. For linear axes, the angular deviation from 90° measured between the best fit lines drawn through two sets of straightness data derived from two orthogonal axes in a specified work zone (expressed as small angles).

Straightness errors are the nonlinear movements that an indicator sees when it is either (1) stationary and reading against a perfect straightedge supported on a moving slide or (2) moved by the slide along a perfect straight edge which is stationary (see Figure 10.5).[5] Basically, this translates to small unwanted motion (along the Y and Z axes) perpendicular to the designed direction of motion.

Tilt error motion is the error motion in an angular direction relative to the Z reference axis (about the X and Y axes).

References

1. J. B. Bryan, The Power of Deterministic Thinking in Machine Tool Accuracy, UCRL-91531, September 1984.
2. C. Evans, *Precision Engineering: An Evolutionary View*, Cranfield Press, Bedford, U.K., 1989.
3. D. L. Carter, J. B. Bryan, H. Hauschildt, and C. Chung, Notes from Machine Tool Accuracy Workshop, Lawrence Livermore National Laboratory and the Society of Manufacturing Engineers in cooperation with The American Society of Mechanical Engineers, Livermore, CA, January 1993.
4. A. H. Slocum, *Precision Machine Design*, Prentice Hall, Englewood Cliffs, NJ, 1992.
5. ANSI B5.54-1991, Methods for Performance Evaluation of Computer Numerically-Controlled Machining Centers, American Society of Mechanical Engineers, 1991.
6. S. H. Crandall, D. C. Karnopp, E. F. Kurtz, Jr., and D. C. Pridmore-Brown, *Dynamics of Mechanical and Electromechanical Systems*, Robert E. Krieger Publishing Co., Malabar, FL, 1982.
7. J. B. Bryan and D. L. Carter, How straight is straight? *American Machinist*, 61–65, December 1989.

© 2002 by CRC Press LLC

8. J. B. Bryan, R. R. Clouser, and E. Holland, Spindle accuracy, *American Machinist*, 149–164, December 4, 1967.
9. ANSI B89.3.4M-1985, Axes of Rotation, American Society of Mechanical Engineers, 1985.
10. E. Abbe, Measuring instruments for physicists, *Journal for Instrumental Information*, 10, 446–448, 1890.
11. J. B. Bryan, The Abbe Principle revisited: An updated interpretation, *Precision Engineering*, 1, 3, 129–132, 1989.
12. H. E. Jenkins and T. R. Kurfess, Dynamic stiffness implications for a multi-axis grinding system, *Journal of Vibration and Control*, 3, 3, 297–313, 1997.
13. ANSI B89.6.2-1973, Temperature and Humidity Environment for Dimensional Measurement, American Society of Mechanical Engineers, 1973.
14. J. B. Bryan and E. R. McCure, Heat vs. tolerances, *American Machinist*, Special Report No. 605, June 5, 1967.
15. J. B. Bryan, International status of thermal error research, *Annals of the C.I.R.P.*, XVI, 203–215, 1968.
16. J. B. Bryan, D. L. Carter, R. W. Clouser, and J. H. Hamilton, An order of magnitude improvement in thermal stability with use of liquid shower on a general purpose measuring machine, ASME Technical Paper IQ82-936, presented at the Precision Machining Workshop, St. Paul, MN, 1982.
17. C.I.R.P. Scientific Committee for Metrology and Interchangability's, A proposal for defining and specifying the dimensional uncertainty of multiaxis measuring machines, *Annals of the C.I.R.P.*, 27, 2, 623–630, 1978.
18. R. R. Donaldson, Error budgets, Technology of Machine Tools, a Survey of the State of the Art by the Machine Tool Task Force, R. J. Hocken working group chairman, UCRL-52960-5, October 1980.
19. R. R. Donaldson, A simple method for separating spindle error from test ball roundness error, *Annals of the C.I.R.P.*, 21, 1, 125–126, 1972.
20. J. B. Bryan and D. L. Carter, How straight is straight? *American Machinist*, 61–65, December 1989.

© 2002 by CRC Press LLC

II

Vibration Control

Nejat Olgac

© 2002 by CRC Press LLC

11

Active Damping of Large Trusses

A. Preumont
Université Libre de Bruxelles

Frederic Bossens
Université Libre de Bruxelles

Nicolas Loix
Micromega Dynamics

11.1 Introduction
11.2 Active Struts
Open-Loop Dynamics of an Active Truss • Integral Force Feedback • Modal Damping • Experimental Results
11.3 Active Tendon Control
Active Damping of Cable Structures • Modal Damping • Active Tendon Design • Experimental Results
11.4 Active Damping Generic Interface
11.5 Microvibrations
11.6 Conclusions

Abstract

This chapter reviews various ways of damping large space trusses. The first part discusses the use of active struts consisting of a piezoelectric actuator collocated with a force sensor. The guaranteed stability properties of the integral force feedback are reviewed and the practical significance of the modal fraction of strain energy is stressed. The second part explains the concept of active tendon control of trusses; the similarity of this concept with the previous one is pointed out. The third part describes an active damping generic interface based on a Stewart platform architecture with piezoelectric legs. The similarity with the previous concepts is emphasized. Finally, the damping of microvibrations is briefly discussed.

11.1 Introduction

The development of future generations of ultralight and large space structures will probably not be possible without active damping enhancement of the structures and active isolation of the scientific payloads that are sensitive to vibrations. Interferometric missions are an example particularly stringent geometric stability requirements.[1,2] This chapter addresses the problem of active damping of large trusses with three different concepts: (i) active strut, (ii) active tendon, and (iii) generic interface. In all cases, the same control architecture is used: a collocated piezoelectric actuator and force sensor connected by a local controller with an integral force feedback (IFF).

11.2 Active Struts

The first concept is the most natural; it consists of replacing some passive bars in the truss by active struts (Figure 11.1). The active struts consist of a piezoelectric linear actuator (or another type of

© 2002 by CRC Press LLC

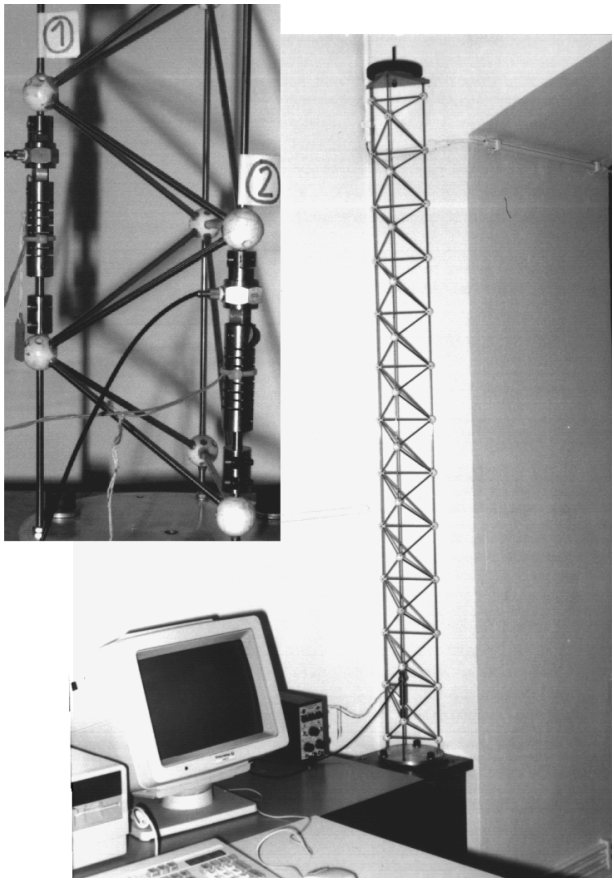


FIGURE 11.1 Active truss with piezoelectric struts (ULB).

linear displacement actuator such as magnetostrictive) co-linear with a force transducer. This concept was first demonstrated in the late 1980s.[3-6]

11.2.1 Open-Loop Dynamics of an Active Truss

Consider the active truss of Figure 11.2. when a voltage V is applied to an unconstrained linear piezoelectric actuator, it produces an expansion δ.

$$\delta = d_{33} n V = g_a V \tag{11.1}$$

where d_{33} is the piezoelectric coefficient, n is the number of piezoelectric ceramic elements in the actuator; g_a is the actuator gain. This equation neglects the hysteresis of the piezoelectric expansion. If the actuator is placed in a truss, its effect on the structure can be represented by equivalent piezoelectric loads acting on the passive structure. As for thermal loads, the pair of self-equilibrating piezoelectric loads applied axially to both ends of the active strut (Figure 11.2) has a magnitude equal to the product of the stiffness of the active strut, K_a, by the unconstrained piezoelectric expansion δ:

$$p = K_a \delta \tag{11.2}$$

© 2002 by CRC Press LLC

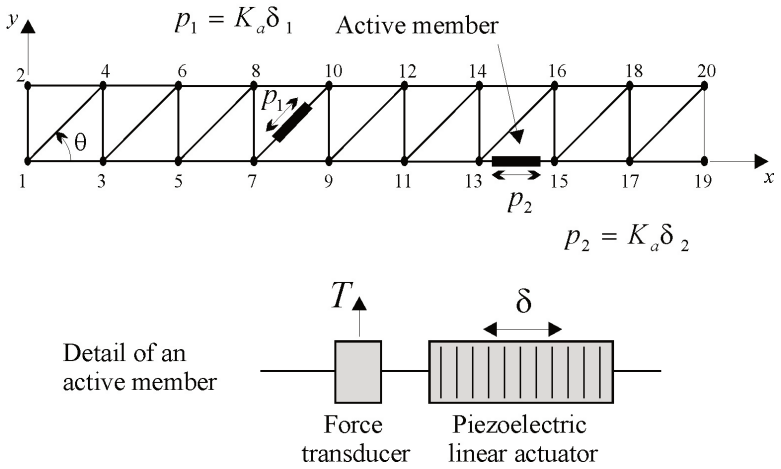


FIGURE 11.2 Active truss. The active struts consist of a piezoelectric linear actuator colinear with a force transducer.

Assuming no damping, the equation governing the motion of the structure excited by a single actuator is

$$M\ddot{x} + Kx = bp = bK_a\delta \tag{11.3}$$

where b is the influence vector of the active strut in the global coordinate system. The nonzero components of b are the direction cosines of the active bar. As for the output signal of the force transducer, it is given by

$$y = T = K_a\delta_e \tag{11.4}$$

where δ_e is the elastic extension of the active strut, equal to the difference between the total extension of the strut and its piezoelectric component δ. The total extension is the projection of the displacements of the end nodes on the active strut, $\Delta = b^T x$. Introducing this into Equation (11.4), we get

$$y = T = K_a(b^T x - \delta) \tag{11.5}$$

Note that because the sensor is located in the same strut as the actuator, the same influence vector b appears in the sensor Equation (11.5) and the equation of motion (11.3). If the force sensor is connected to a charge amplifier of gain g_s, the output voltage v_o is given by

$$v_0 = g_s T = g_s K_a(b^T x - \delta) \tag{11.6}$$

Note the presence of a feedthrough component from the piezoelectric extension δ. Upon transforming into modal coordinates, the frequency response function (FRF) $G(\omega)$ between the voltage V applied to the piezo and the output voltage of the charge amplifier can be written:[7]

$$\frac{v_0}{V} = G(\omega) = g_s g_a K_a \left\{ \sum_{i=1}^{n} \frac{\nu_i}{1 - \omega^2 / \Omega_i^2} - 1 \right\} \tag{11.7}$$

© 2002 by CRC Press LLC

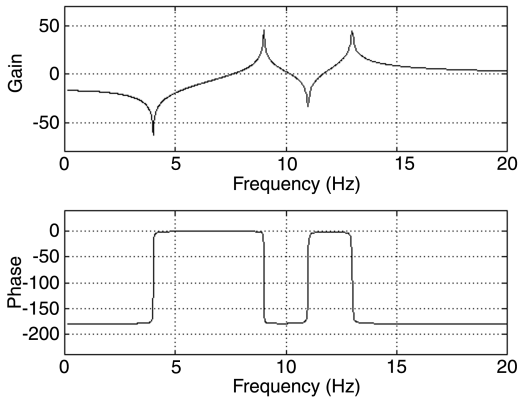


FIGURE 11.3 Open-loop FRF $G(\omega)$ of the active truss (a small damping is assumed).

where Ω_i are the natural frequencies, and we define

$$\nu_i = \frac{K_a(b^T\phi_i)^2}{\mu_i\Omega_i^2} = \frac{K_a(b^T\phi_i)^2}{\phi_i^T K\phi_i} \tag{11.8}$$

The numerator and the denominator of this expression represent, respectively, twice the strain energy in the active strut and twice the total strain energy when the structure vibrates according to mode i. $\nu_i(\geq 0)$ is, therefore, called the *modal fraction of strain energy* in the active strut. From Equation (11.7), we see that ν_i determines the residue of mode i, which is the amplitude of the contribution of mode i in the transfer function between the piezo actuator and the force sensor. It can, therefore, be regarded as a compound index of controllability and observability of mode i. ν_i is readily available from commercial finite element programs and can be used to select the proper location of the active strut in the structure: the best location is that with the highest ν_i for the modes that we wish to control.[5]

11.2.2 Integral Force Feedback

The FRF (Equation 11.7) has alternating poles and zeros (Figure 11.3) on the imaginary axis (or near if the structural damping is taken into account); on the other hand, $G(\omega)$ has a feedthrough component and some roll-off must be added to the compensator to achieve stability. It is readily established from the root locus (Figure 11.4) that the positive integral force feedback (IFF):

$$gD(s) = \frac{-g}{K_a s} \tag{11.9}$$

is unconditionally stable for all values of g. The negative sign in Equation (11.9) is combined with the negative sign in the feedback loop (Figure 11.5) to produce a positive feedback.

In practice, it is not advisable to implement plain integral control, because it would lead to saturation. A forgetting factor can be introduced by slightly moving the pole of the compensator from the origin to the negative real axis, leading to

$$gD(s) = \frac{-g}{K_a(s+\varepsilon)} \tag{11.10}$$

© 2002 by CRC Press LLC

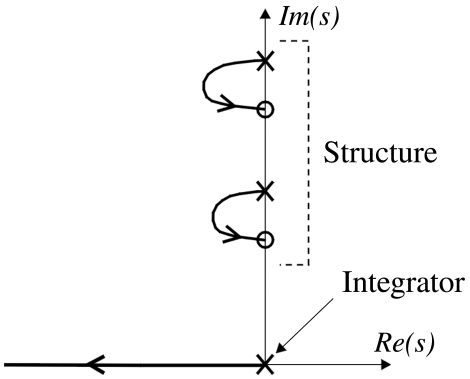


FIGURE 11.4 Root locus of the integral force feedback.

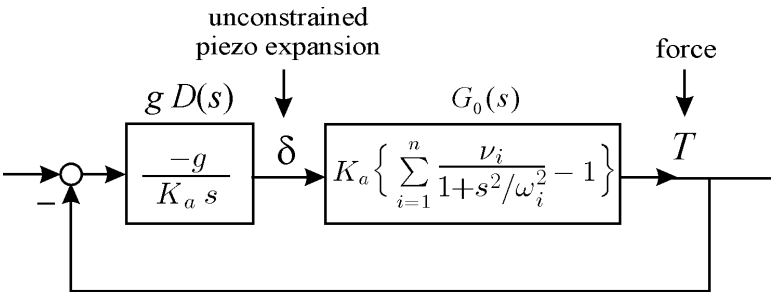


FIGURE 11.5 Block diagram of the integral force feedback.

This does not affect the general shape of the root locus and prevents saturation. Note that piezoelectric force sensors have a built-in high-pass filter.

11.2.3 Modal Damping

Combining the structure Equation (11.3), the sensor Equation (11.5), and the controller Equation (11.9), the closed-loop characteristic equation reads

$$\left[Ms^2 + K - \frac{g}{s+g}\left(bK_a b^T\right)\right]x = 0 \tag{11.11}$$

From this equation, we can deduce the open-loop transmission zeros, which coincide with the asymptotic values of the closed-loop poles as $g \to \infty$. Taking the limit, we get

$$\left[Ms^2 + (K - bK_a b^T)\right]x = 0 \tag{11.12}$$

which states that the zeros (i.e., the anti-resonance frequencies) coincide with the poles (resonance frequencies) of the structure where the active strut has been removed (corresponding to the stiffness matrix $K\text{-}bK_a b^T$).

To evaluate the modal damping, Equation (11.11) must be transformed in modal coordinates with the change of variables $x = \Phi z$. Assuming that the mode shapes have been normalized according to $\Phi^T M \Phi = I$ and taking into account that $\Phi^T K \Phi = diag(\Omega_i^2) = \Omega^2$, we have

© 2002 by CRC Press LLC

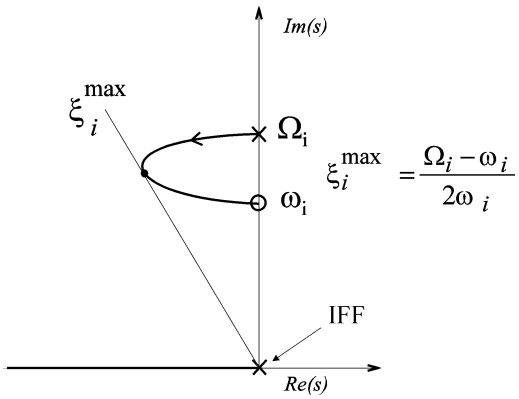


FIGURE 11.6 Root locus of the closed-loop pole for the IFF.

$$\left[Is^2 + \Omega^2 - \frac{g}{s+g} \Phi^T (bK_a b^T) \Phi \right] z = 0 \tag{11.13}$$

The matrix $\Phi^T(bK_a b^T)\Phi$ is, in general, fully populated. If we assume that it is diagonally dominant, and if we neglect the off-diagonal terms, it can be rewritten

$$\Phi^T (bK_a b^T) \Phi \approx \text{diag}(\nu_i \Omega_i^2) \tag{11.14}$$

where ν_i is the fraction of modal strain energy in the active member when the structure vibrates according to mode i; ν_i is defined by Equation (11.8). Substituting Equation (11.14) into Equation (11.13), we find a set of decoupled equations

$$s^2 + \Omega_i^2 - \frac{g}{s+g} \nu_i \Omega_i^2 = 0 \tag{11.15}$$

and, after introducing

$$\omega_i^2 = \Omega_i^2 (1 - \nu_i) \tag{11.16}$$

it can be rewritten

$$s^2 + \Omega_i^2 - \frac{g}{s+g} \left(\Omega_i^2 - \omega_i^2 \right) = 0 \tag{11.17}$$

By comparison with Equation (11.11), we see that the transmission zeros (the limit of the closed-loop poles as $g \to \infty$) are $\pm j\omega_i$. The characteristic equation can be rewritten

$$1 + g \frac{(s^2 + \omega_i^2)}{s(s^2 + \Omega_i^2)} = 0 \tag{11.18}$$

The corresponding root locus is shown in Figure 11.6. The depth of the loop in the left half plane depends on the frequency difference $\Omega_i - \omega_i$, and the maximum modal damping is given by

© 2002 by CRC Press LLC

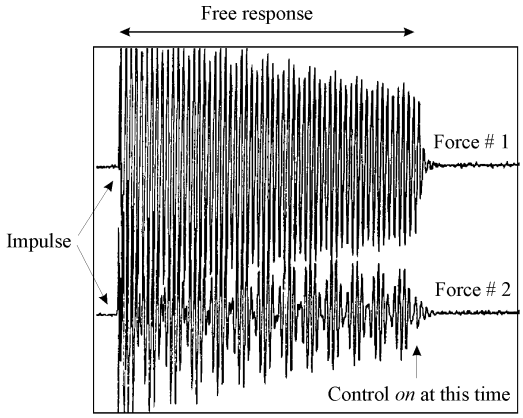


FIGURE 11.7 Force signal from the two active struts during the free response after impulsive load.

$$\xi_i^{\max} = \frac{\Omega_i - \omega_i}{2\omega_i} \tag{11.19}$$

It is obtained for $g = \Omega_i \sqrt{\Omega_i / \omega_i}$. For small gains, it can be shown that

$$\xi_i = \frac{g\nu_i}{2\Omega_i} \tag{11.20}$$

This interesting result tells us that, for small gains, the active damping ratio in a given mode is proportional to the fraction of modal strain energy in the active element. This result is very useful for the design of active trusses; the active struts should be located to maximize the fraction of modal strain energy ν_i in the active members for the critical vibration modes. The preceding results have been established for a single active member. If several active members are operating with the same control law and the same gain g, this result can be generalized under similar assumptions. It can be shown that each closed-loop pole follows a root locus governed by Equation (11.18) where the pole Ω_i is the natural frequency of the open-loop structure and the zero ω_i is the natural frequency of the structure where the active members have been removed.

11.2.4 Experimental Results

Figures 11.7 and 11.8 illustrate typical results obtained with the test structure of Figure 11.1. The modal damping ratio of the first two modes is larger than 10%. Note that in addition to being simple and robust, the control law can be implemented in an analog controller, which performs better in microvibrations.

11.3 Active Tendon Control

The use of cables to achieve lightweight spacecrafts is not new; it can be found in Herman Oberth's early books[17,18] on astronautics. In terms of weight, the use of guy cables is probably the most efficient way to stiffen a structure. They also can be used to prestress a deployable structure and eliminate the geometric uncertainty due to the gaps. One further step consists of providing the cables with active tendons to achieve active damping in the structure. This approach has been developed in References 7–12.

© 2002 by CRC Press LLC

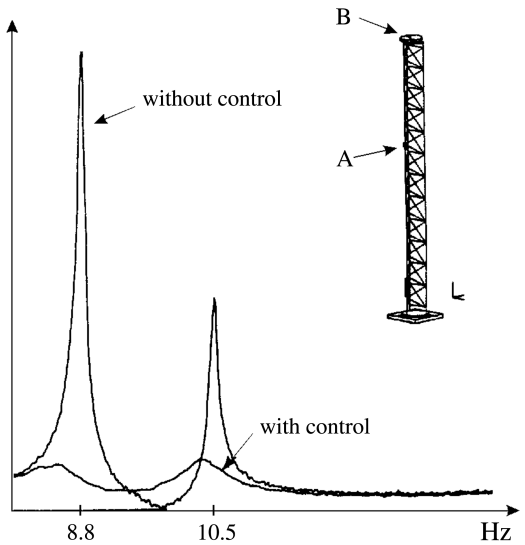


FIGURE 11.8 FRF between a force in A and an accelerometer in B, with and without control.

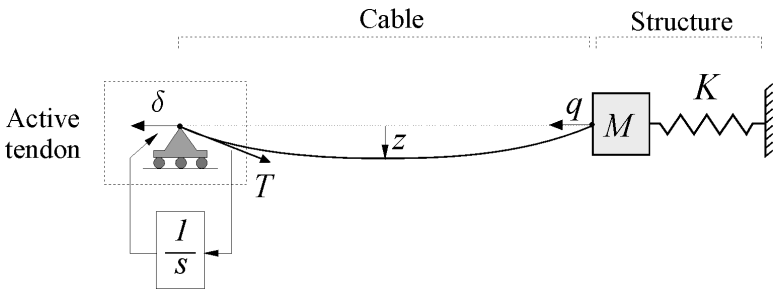


FIGURE 11.9 Active damping of cable structures.

11.3.1 Active Damping of Cable Structures

When using a displacement actuator and a force sensor, the (positive) integral force feedback Equation (11.9) belongs to the class of "energy absorbing" controls: indeed, if

$$\delta \sim \int T dt \tag{11.21}$$

the power flow from the control system is $W = -T\dot{\delta} \sim -T^2 \leq 0$. This means that the control can only extract energy from the system. This applies to nonlinear structures as well; all the states which are controllable and observable are asymptotically stable for all positive gains (infinite gain margin). The control concept is represented schematically in Figure 11.9 where the spring-mass system represents an arbitrary structure. Note that the damping introduced in the cables is usually very low, but experimental results have confirmed that it always remains stable, even at the parametric resonance, when the natural frequency of the structure is twice that of the cables. Whenever possible, however, the tension in the cables should be adjusted in such a way that their first natural frequency is above the frequency range where the global modes must be damped.

© 2002 by CRC Press LLC

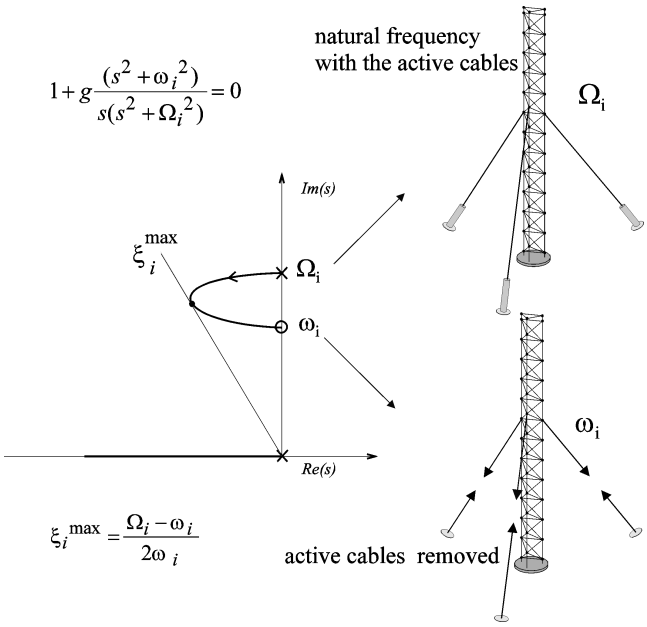


FIGURE 11.10 Root locus of the closed-loop poles.

11.3.2 Modal Damping

If we assume that the dynamics of the cables can be neglected and that their interaction with the structure is restricted to the tension in the cables, and that the global mode shapes are identical with and without the cables, one can develop an approximate linear theory for the closed-loop system. The following results that follow closely those obtained in the foregoing section (we assume no structural damping) can be established:

The open-loop poles are $\pm j\Omega_i$ where Ω_i are the natural frequencies of the structure including the active cables and the open-loop zeros are $\pm j\omega_i$ where ω_i are the natural frequencies of the structure where the active cables have been removed.

If the same control gain is used for every local control loop, as g goes from *0* to ∞, the closed-loop poles follow the root locus defined by Equation (11.18) and (Figure 11.10). Equations (11.19) and (11.20) also apply in this case.

11.3.3 Active Tendon Design

Figure 11.11 shows two possible designs of the active tendon: the first one (bottom left) is based on a linear piezoactuator from PI and a force sensor from B&K; a lever mechanism (top view) is used to transform the tension in the cable into a compression in the piezo stack, and amplifies the translational motion to achieve about 100 μm. This active element is identical to that in an active strut. In the second design (bottom center and right), the linear actuator is replaced by an amplified actuator from CEDRAT Research, also connected to a B&K force sensor and flexible tips. In addition to being more compact, this design does not require an amplification mechanism and tension of the flexible tips produces a compression in the piezo stack at the center of the elliptical structure.

11.3.4 Experimental Results

Figure 11.12 shows the test structure; it is representative of a scale model of the JPL-Micro-Precision-Interferometer[1] which consists of a large trihedral passive truss of about 9 m. The free-floating condition

© 2002 by CRC Press LLC

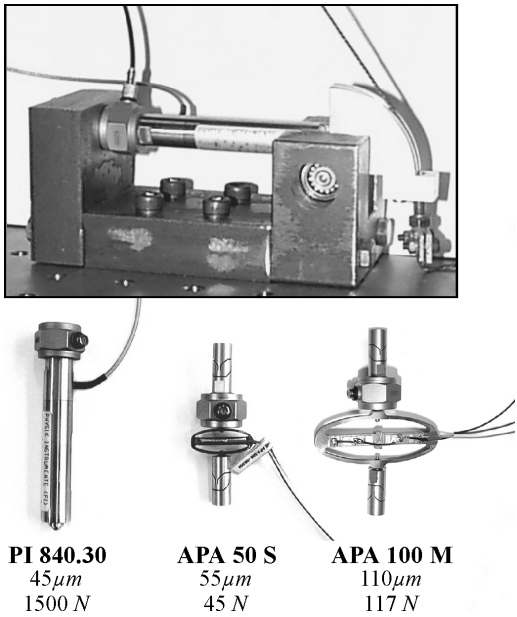
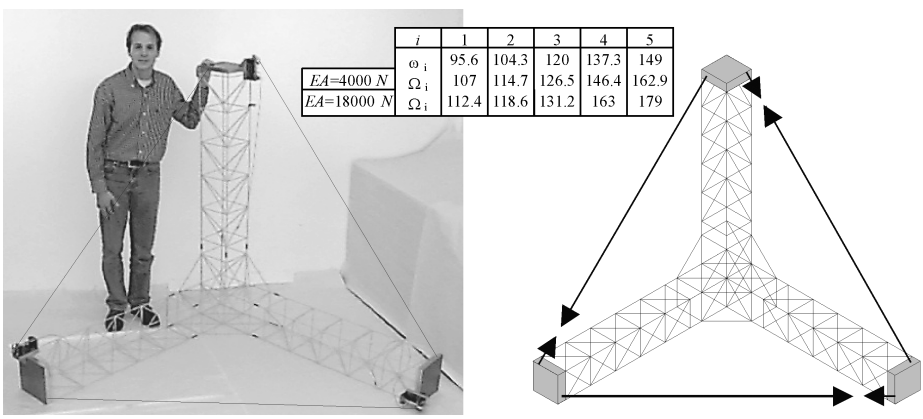


FIGURE 11.11 Three different designs of active tendon or active strut (ULB).



	i	1	2	3	4	5
	ω_i	95.6	104.3	120	137.3	149
EA=4000 *N*	Ω_i	107	114.7	126.5	146.4	162.9
EA=18000 *N*	Ω_i	112.4	118.6	131.2	163	179

FIGURE 11.12 Free floating truss with active tendons.

during the test is simulated by hanging the structure from the ceiling of the lab with soft springs. In this study, two different types of cables have been used: a fairly soft cable of 1-mm diameter of polyethylene ($EA \approx 4000N$) and a stiffer one of synthetic fiber *Dynema*™ ($EA \approx 18000N$). In both cases, the tension in the cables was chosen to set the first cable mode at 400 rad/sec or more, far above the first five flexible modes for which active damping is sought. The table inset in Figure 11.12 gives the measured natural frequencies ω_i (without cables) and Ω_i (with cables), for the two sets of cables.

Figure 11.13 compares the experimental closed-loop poles obtained for increasing gain g of the control with the root locus prediction of Equation (11.18). The results are consistent with the analytical predictions, although a larger scatter is observed with stiffer cables. Note, however, that the experimental results tend to exceed the root locus predictions. Figure 11.14 compares typical FRF with and without control. An analytical study was conducted[11] to investigate the possibility of using three Kevlar cables of *2* mm diameter connecting the tips of the three trusses of the JPL-MPI. Using

© 2002 by CRC Press LLC

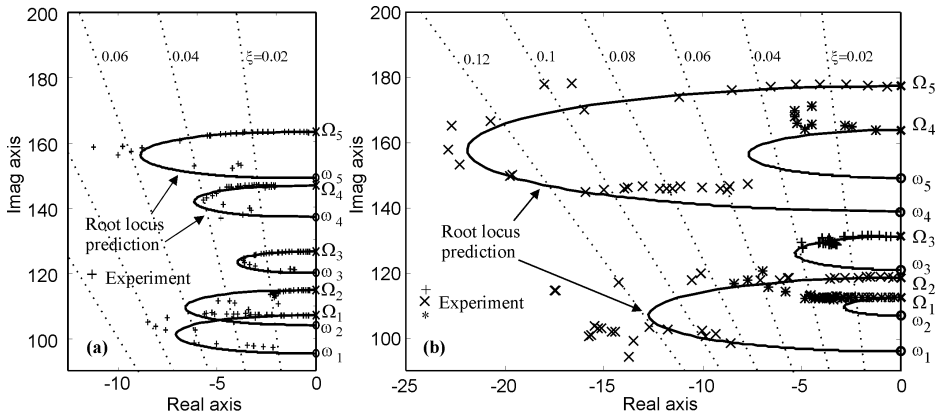


FIGURE 11.13 Experimental poles vs. root-locus prediction for the flexible modes of the free floating truss. (a) *EA = 4000 N*; (b) *EA = 18000 N*.

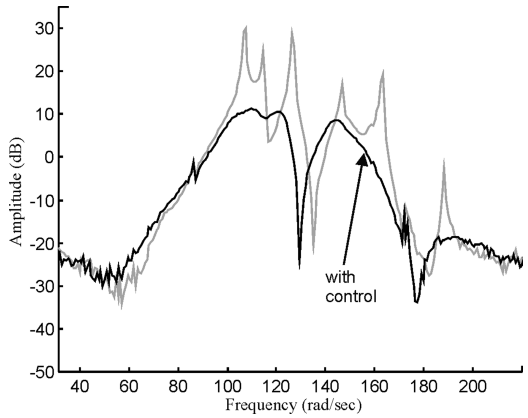


FIGURE 11.14 Typical FRF with and without control (*EA = 4000N*).

the root locus technique of Figure 11.10, a damping ratio between 14 and 21% was predicted in the first three flexible modes.

11.4 Active Damping Generic Interface

The active strut discussed in Section 11.2 can be developed into a generic six degrees-of-freedom interface which can be used to connect arbitrary substructures. Such an interface is shown in Figure 11.15; it consists of a Stewart platform with cubic architecture.[13] Each leg consists of an active strut similar to that shown at the center of Figure 11.11: a piezotranslator of the amplified design collocated with a force sensor, and connected to the base plates by flexible tips acting like spherical joints. The cubic architecture provides a uniform control capability in all directions, a uniform stiffness in all directions, and minimizes the cross-coupling among actuators (which are mutually orthogonal). The control is decentralized with the same gain for all loops. Figure 11.16 shows the generic interface mounted between a truss and the supporting structure. Figure 11.17 shows the evolution of the first two closed-loop poles when we increase the gain of the decentralized controller; the continuous line shows the root locus prediction of Equation (11.18); Ω_i are the open-loop natural frequencies, while ω_i are the high-gain asymptotes of the closed-loop poles. Figure 11.18 shows a typical FRF of the structure of Figure 11.16, with and without control of the Stewart platform.

© 2002 by CRC Press LLC

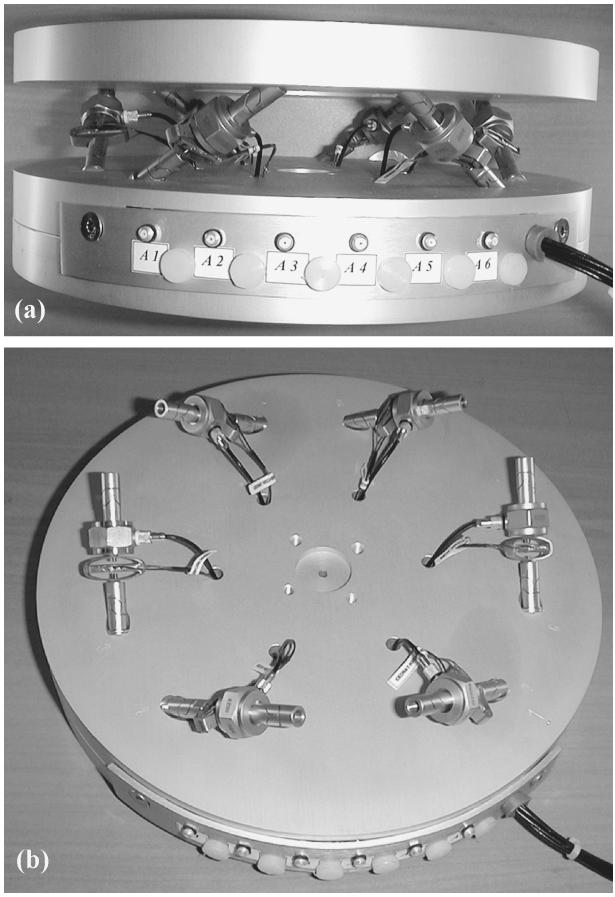


FIGURE 11.15 Stewart platform with piezoelectric legs as generic damping interface (a) general view; (b) the upper base plate removed.

11.5 Microvibrations

The performance and robustness of the control strategy have been experimentally demonstrated; however, most of the results presented in the literature have been obtained for vibration amplitudes in a range from millimeter to micron. For applications to precision space structures with optical payloads, it is essential that these results be confirmed for submicron vibrations,[14-16] despite the nonlinear behaviour of the actuator (hysteresis of the piezo).

It turns out that the performance limit of the control system is related to the sensitivity of the force sensor. This is illustrated in Figure 11.19, which shows the Lissajou plots δ vs. T (active tendon displacement vs. dynamic tension in the strut) for two sensors with different sensitivities. Because the control algorithm produces a 90° phase shift between the piezo extension and the force measurement, the theoretical shape of the plot is an ellipse; the area corresponds to the energy dissipation in the control system during one cycle. Figures 11.19 (a) and (c) on the left side have been obtained with a standard sensor (B&K 8200, 4 pc/N). The curve becomes more noisy as the vibration amplitude is reduced and the dynamic force approaches the sensitivity limit of the sensor. On the other hand, Figures 11.19 (b), (d), and (f), on the right side have been obtained with a more sensitive sensor (280000 pc/N). In this case, the Lissajou plots keep the right shape even for very small vibration amplitudes (the limit of this experiment actually came from background vibration).

© 2002 by CRC Press LLC

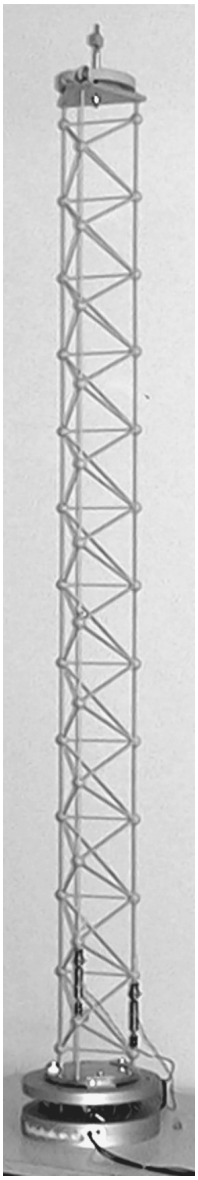


FIGURE 11.16 Generic active damping interface acting as a support of a truss.

11.6 Conclusions

This chapter reviewed various ways of damping large space trusses. The active strut consisting of a piezoelectric actuator collocated with a force sensor was described first. Next, three different ways to use this active strut to achieve active damping were reviewed: first by integrating the active strut as a member of the truss, next by using it as an active tendon in a cable structure, and finally by using the struts as the legs of a Stewart platform. The similarity between the various concepts has been pointed out when a decentralized controller is used, and it was shown that the closed-loop poles can be predicted with a root-locus technique. Finally, the damping of microvibrations was briefly discussed.

© 2002 by CRC Press LLC

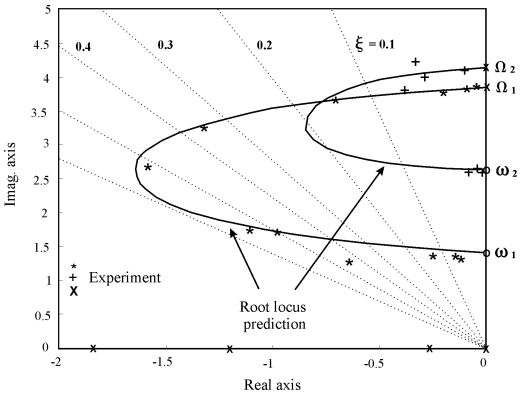


FIGURE 11.17 Experimental poles and root locus prediction from Equation (11.18).

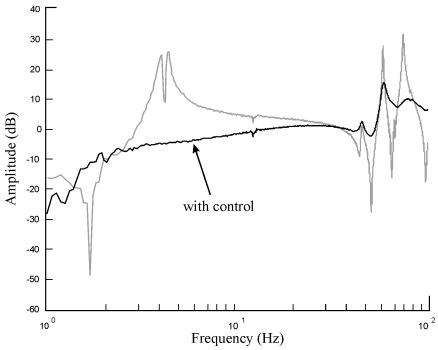


FIGURE 11.18 Typical FRF with and without control.

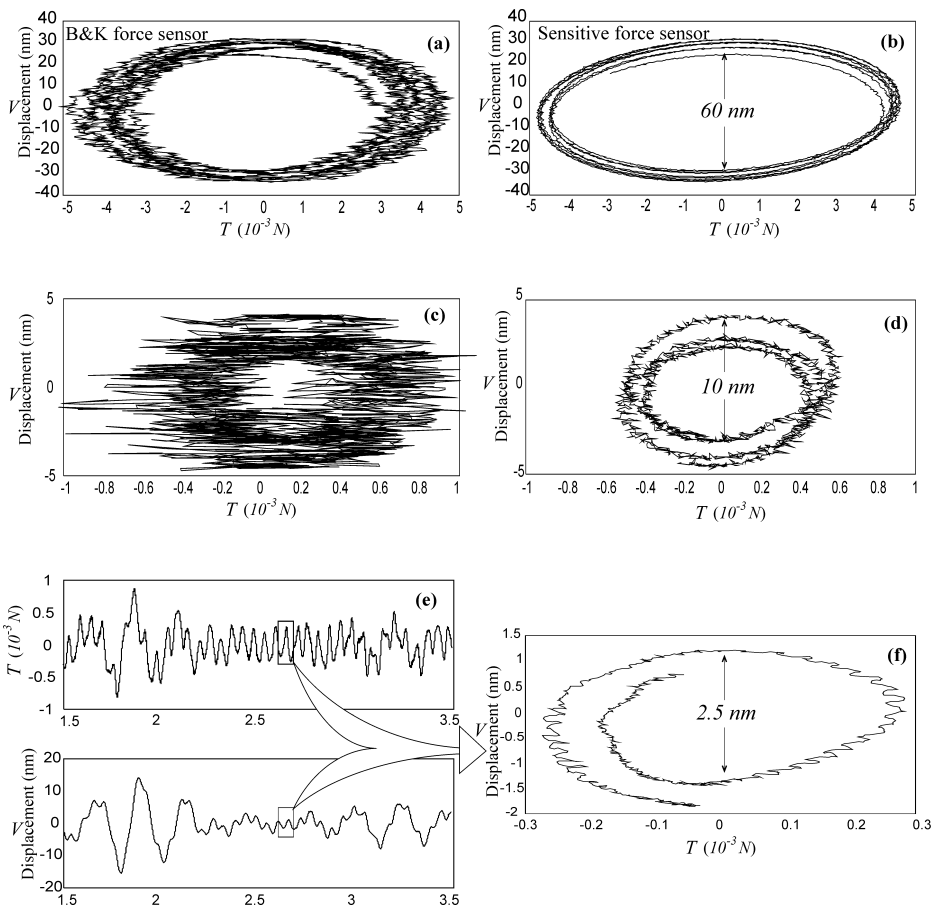


FIGURE 11.19 "Lissajou plots" δ vs. T in the microvibration range.

© 2002 by CRC Press LLC

Acknowledgment

This study was partly supported by the Inter University Attraction Pole IUAP IV-24 on Intelligent Mechatronics Systems. The authors wish to thank Ahmed Abu Hanieh for his contribution to the design, manufacture, and testing of the Stewart platform.

References

1. Neat, G.W., Abramovici, A., Melody, J.M., Calvet, R.J., Nerheim, N.M., and O'Brien, J.F., Control technology readiness for spaceborne optical interferometer missions, *Proceedings SMACS-2,* Toulouse, 13–32 (1997).
2. Mallory, G.J.W., Saenz-Otero, A., and Miller, D.W., Origins test bed: Capturing the dynamics and control of future space-based telescopes, *Optical Engineering,* 39, 6, 1665–1676 (2000).
3. Fanson, J.L., Blackwood, C.H., and Chu, C.C., Active member control of a precision structure, *Proceedings of the 30th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference,* AIAA, Washington, D.C., 1480–1494 (1989).
4. Chen, G.S., Lurie, B.J., and Wada, B.K., Experimental studies of adaptive structure for precision performance, *Proceedings of the 30th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference,* AIAA, Washington, D.C.,1462–1472 (1989).
5. Preumont, A., Dufour, J.P., and Malekian, Ch., Active damping by a local force feedback with piezoelectric actuators, *AIAA, Journal of Guidance*, 15, 2, 390–395 (1992).
6. Peterson, L.D., Allen, J.J., Lauffer, J.P., and Miller, A.K., An experimental and analytical synthesis of controlled structure design, SDM Conference*,* AIAA paper 89-1170-CP (1989).
7. Preumont, A., *Vibration Control of Active Structures: An Introduction*, Kluwer Academic Publishers, Dordrecht (1997).
8. Achkire, Y., *Active Tendon Control of Cable-Stayed Bridges*, Ph.D. dissertation, Active Structures Laboratory, Université Libre de Bruxelles, Belgium, May 1997.
9. Achkire, Y. and Preumont, A., Active tendon control of cable-stayed bridges, *Earthquake Engineering and Structural Dynamics*, 25, 6, 585–597, June 1996.
10. Preumont, A. and Achkire, Y., Active damping of structures with guy cables, *AIAA, Journal of Guidance, Control, and Dynamics*, 20, 2, 320–326, March-April 1997.
11. Preumont, A., Achkire, Y., and Bossens, F., Active tendon control of large trusses, *AIAA Journal*, 38, 3, 493–498, March 2000.
12. Preumont, A. and Bossens, F., Active tendon control of vibration of truss structures: Theory and experiments, *Journal of Intelligent Material Systems and Structures,* 11, 2, 91–99, 2000.
13. Geng, Z.J. and Haynes, L.S., Six degrees-of-freedom active vibration control using the Stewart platforms, *IEEE Transactions on Control Systems Technology*, 2, 1, 45–53, 1994.
14. Peterson, L.D., Lake, M.S., and Hardaway, L.M.R., Micron accuracy deployment experiments (MADE): A space station laboratory for actively controlled precision deployable structures technology, *Proceedings of the Space Technology and Applications International Forum*, Albuquerque (NM), February 1999.
15. Ingham, M.D. and Crawley, E. F., Microdynamic characterization of modal parameters for deployable space structure, *AIAA Journal*, 39, 2, 331–338, February 2001.
16. Hardaway, L.M.R. and Peterson, L.D., Microdynamics of a precision deployable optical truss, *SPIE Paper No. 3785-01, Proceedings of the SPIE Annual Meeting*, Denver (Colorado), July 1999.
17. Oberth, H., *Man into Space*. *New Projects for Rocket and Space Travel* (translated from German by G.P.H. DeFreville), Weidenfeld and Nicolson, London, 1957.
18. Walters, H.B., *Hermann Oberth: Father of Space Travel*, Macmillan, New York, 1962.

© 2002 by CRC Press LLC

12

Semi-Active Suspension Systems

12.1 Introduction
Vibration Isolation vs. Vibration Absorption • Classification of Suspension Systems • Why Semi-Active Suspension?

12.2 Semi-Active Suspensions Design
Introduction • Semi-Active Vibration Absorption Design • Semi-Active Vibration Isolation Design

12.3 Adjustable Suspension Elements
Introduction • Variable Rate Dampers • Variable Rate Spring Elements • Other Variable Rate Elements

12.4 Automotive Semi-Active Suspensions
Introduction • An Overview of Automotive Suspensions • Semi-Active Vehicle Suspension Models • Semi-Active Suspension Performance Characteristics • Recent Advances in Automotive Semi-Active Suspensions

12.5 Application of Control Techniques to Semi-Active Suspensions
Introduction • Semi-Active Control Concept • Optimal Semi-Active Suspension • Other Control Techniques

12.6 Practical Considerations and Related Topics

Nader Jalili
Clemson University

12.1 Introduction

Semi-active (SA) suspensions are those which otherwise passively generated damping or spring forces modulated according to a parameter tuning policy with only a small amount of control effort. SA suspensions, as their name implies, fill the gap between purely passive and fully active suspensions and offer the reliability of passive systems, yet maintain the versatility and adaptability of fully active devices. Because of their low energy requirement and cost, considerable interest has developed during recent years toward practical implementation of these systems. This chapter presents the basic theoretical concepts for SA suspensions' design and implementation, followed by an overview of recent developments and control techniques. Some related practical developments ranging from vehicle suspensions to civil and aerospace structures are also reviewed.

12.1.1 Vibration Isolation vs. Vibration Absorption

In most of today's mechatronic systems a number of possible devices, such as reaction or momentum wheels, rotating devices, and electric motors are essential to the systems' operations. These devices,

© 2002 by CRC Press LLC

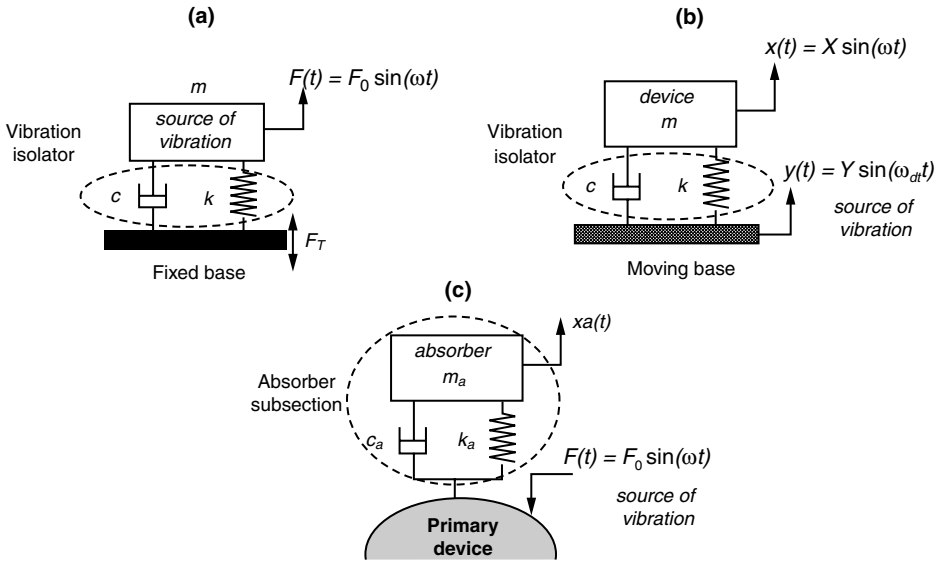


FIGURE 12.1 Schematic of (a) force transmissibility for foundation isolation, (b) displacement transmissibility for protecting device from vibration of the base, and (c) application of vibration absorber for suppressing primary system vibration.

however, can also be sources of detrimental vibrations that may significantly influence the mission performance, effectiveness, and accuracy of operation. Several techniques are utilized to either limit or alter the vibration response of such systems. Vibration isolation suspensions and vibration absorbers are quoted in the literature as the two most commonly used techniques for such utilization.

In vibration isolation either the source of vibration is isolated from the system of concern (also called "force transmissibility, see Figure 12.1a), or the device is protected from vibration of its point of attachment (also called displacement transmissibility, see Figure 12.1b). Unlike the isolator, a vibration absorber consists of a secondary system (usually mass–spring–damper trio) added to the primary device to protect it from vibrating (see Figure 12.1c). By properly selecting absorber mass, stiffness, and damping, the vibration of the primary system can be minimized.[1]

12.1.2 Classification of Suspension Systems

Passive, active, and semi-active are referred to in the literature as the three most common classifications of suspension systems (either as isolators or absorbers), see Figure 12.2.[2] A suspension system is said to be active, passive, or semi-active depending on the amount of external power required for the suspension to perform its function. A passive suspension consists of a resilient member (stiffness) and an energy dissipator (damper) to either absorb vibratory energy or load the transmission path of the disturbing vibration[3] (Figure 12.2a). It performs best within the frequency region of its highest sensitivity. For wideband excitation frequency, its performance can be improved considerably by optimizing the suspension parameters.[4-6] However, this improvement is achieved at the cost of lowering narrowband suppression characteristics.

The passive suspension has significant limitations in structural applications where broadband disturbances of highly uncertain nature are encountered. To compensate for these limitations, active suspension systems are utilized. With an additional active force introduced as a part of suspension subsection, $u(t)$ in Figure 12.2b, the suspension is then controlled using different algorithms to make it more responsive to source of disturbances.[2,7-9] A combination of active/passive treatment is intended to reduce the amount of external power necessary to achieve the desired performance characteristics.[10]

© 2002 by CRC Press LLC

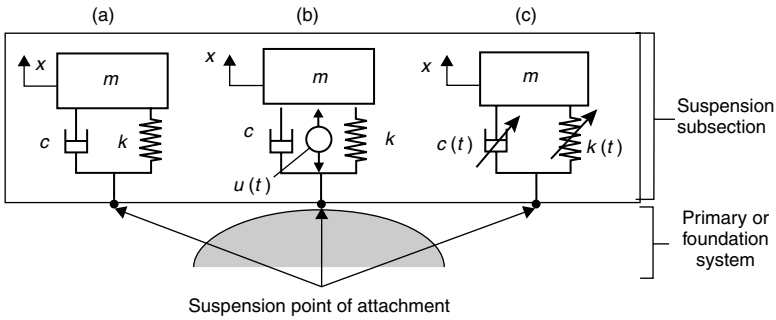


FIGURE 12.2 A typical primary structure equipped with three versions of suspension systems: (a) passive, (b) active, and (c) semi-active configuration.

12.1.3 Why Semi-Active Suspension?

In the design of a suspension system, the system is often required to operate over a wideband load and frequency range which is impossible to meet with a single choice of suspension stiffness and damping. If the desired response characteristics cannot be obtained, active suspension may provide an attractive alternative vibration control for such broadband disturbances. However, active suspensions suffer from control-induced instability in addition to the large control effort requirement. This is a serious concern that prevents common usage in most industrial applications. On the other hand, passive suspensions are often hampered by a phenomenon known as "de-tuning." De-tuning implies that the passive system is no longer effective in suppressing the vibration as it was designed to do. This occurs because of one of the following reasons: (1) the suspension structure may deteriorate and its structural parameters can be far from the original nominal design, (2) the structural parameters of the primary device itself may alter, or (3) the excitation frequency and/or nature of disturbance may change over time.

Semi-active (also known as adaptive-passive) suspension addresses these limitations by effectively integrating a tuning control scheme with tunable passive devices. For this, active force generators are replaced by modulated variable compartments such as a variable rate damper and stiffness, see Figure 12.2c.[11-13] These variable components are referred to as "tunable parameters" of the suspension system, which are retailored via a tuning control and thus result in semi-actively inducing optimal operation. Much attention is being paid to these suspensions for their low energy requirement and cost. Recent advances in smart materials and adjustable dampers and absorbers have significantly contributed to the applicability of these systems.[14-16]

12.2 Semi-Active Suspensions Design

12.2.1 Introduction

SA suspensions can achieve most of the performance characteristics of fully active systems, thus allowing for a wide class of applications. The idea of SA suspension is very simple: to replace active force generators with continually adjustable elements which can vary and/or shift the rate of energy dissipation in response to an instantaneous condition of motion. This section presents basic understanding and fundamental principles and design issues for SA suspension systems, which are discussed in the form of a vibration absorber and vibration isolator.

12.2.2 Semi-Active Vibration Absorption Design

With a history of almost a century,[17] vibration absorbers have proven to be useful vibration suppression devices, widely used in hundreds of diverse applications. It is elastically attached to

© 2002 by CRC Press LLC

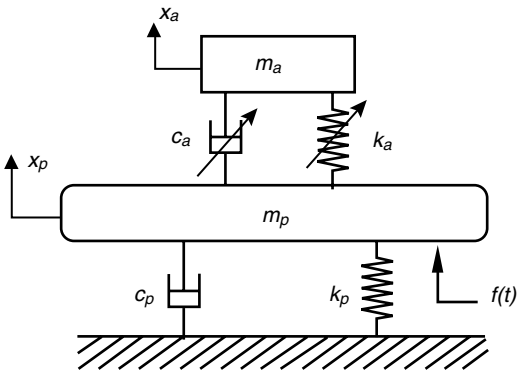


FIGURE 12.3 Application of a semi-active abosrber to SDOF primary system with adjustable stiffness k_a and damping c_a.

the vibrating body to alleviate detrimental oscillations from its point of attachment (see Figure 12.2). The underlying proposition for an SA absorber is to properly adjust the absorber parameters so that it absorbs the vibratory energy within the frequency interval of interest.

To explain the underlying concept, a single-degree-of-freedom (SDOF) primary system with a SDOF absorber attachment is considered (Figure 12.3). The governing dynamics are expressed as

$$m_a\ddot{x}_a(t)+c_a\dot{x}_a(t)+k_a x_a(t)=c_a\dot{x}_p(t)+k_a x_p(t) \tag{12.1}$$

$$m_p\ddot{x}_p(t)+\left(c_p+c_a\right)\dot{x}_p(t)+\left(k_p+k_a\right)x_p(t)-c_a\dot{x}_a(t)-k_a x_a(t)=f(t) \tag{12.2}$$

where $x_p(t)$ and $x_a(t)$ are the respective primary and absorber displacements, $f(t)$ is the external force, and the rest of the parameters including adjustable absorber stiffness k_a and damping c_a are defined per Figure 12.3. The transfer function between the excitation force and primary system displacement in Laplace domain is then written as

$$TF(s)=\frac{X_p(s)}{F(s)}=\left\{\frac{m_a s^2+c_a s+k_a}{H(s)}\right\} \tag{12.3}$$

where

$$H(s)=\left\{m_p s^2+(c_p+c_a)s+k_p+k_a\right\}(m_a s^2+c_a s+k_a)-(c_a s+k_a)^2 \tag{12.4}$$

and $X_a(s)$, $X_p(s)$, and $F(s)$ are the Laplace transformations of $x_a(t)$, $x_p(t)$, and $f(t)$, respectively.

The steady-state displacement of the system due to harmonic excitation is then

$$\left|\frac{X_p(j\omega)}{F(j\omega)}\right|=\left|\frac{k_a-m_a\omega^2+jc_a\omega}{H(j\omega)}\right| \tag{12.5}$$

where ω is the disturbance frequency and $j=\sqrt{-1}$. Utilizing adjustable properties of the SA unit (i.e., variable rate damper and spring), an appropriate parameter tuning scheme is selected to minimize the primary system's vibration subject to external disturbance $f(t)$.

© 2002 by CRC Press LLC

12.2.2.1 Harmonic Excitation

When excitation is tonal, the absorber is generally tuned at the disturbance frequency. For complete attenuation, the steady state $|X_p(j\omega)|$ must equal zero. Consequently, from Equation (12.5), the ideal stiffness and damping of SA absorber are adjusted as

$$k_a = m_a\,\omega^2, \quad c_a = 0 \tag{12.6}$$

Note that this tuned condition is only a function of absorber elements (m_a, k_a, and c_a). That is, the absorber tuning does not need information from the primary system and hence its design is stand-alone. For tonal applications, theoretically zero damping in an absorber subsection results in improved performance. In practice, however, damping is incorporated to maintain a reasonable trade-off between the absorber mass and its displacement. Hence, the design effort for this class of applications is focused on having precise tuning of an absorber to the disturbance frequency and controlling damping to an appropriate level. Referring to Snowdon,[18] it can be proven that the absorber, in the presence of damping, can be most favorably tuned and damped if adjustable stiffness and damping are selected as

$$k_{opt} = \frac{m_a m_p^2 \omega^2}{(m_a + m_p)^2}, \quad c_{opt} = m_a \sqrt{\frac{3k_{opt}}{2(m_a + m_p)}} \tag{12.7}$$

12.2.2.2 Broadband Excitation

In broadband vibration control, the absorber subsection is generally designed to add damping to and change the resonant characteristics of the primary structure to maximally dissipate vibrational energy over a range of frequencies. The objective of SA suspension design is, therefore, to adjust the *absorber parameters* to minimize the peak magnitude of the frequency transfer function ($FTF(\omega) = |TF(s)|_{s=j\omega}$) over the absorber variable suspension parameters $\mathbf{p} = \{c_a \;\; k_a\}^T$. That is, we seeк **p** to

$$\min_{\mathbf{p}} \left\{ \max_{\omega_{\min} \le \omega \le \omega_{\max}} \{|FTF(\omega)|\} \right\} \tag{12.8}$$

Alternatively, one may select the mean square displacement response (MSDR) of the primary system for vibration suppression performance. That is, the absorber variable parameters' vector **p** is selected such that the MSDR

$$E\{(\bar{x}_p)^2\} = \int_0^\infty \{FTF(\omega)\}^2 S(\omega) d\omega \tag{12.9}$$

is minimized over a desired wideband frequency range. $S(\omega)$ is the power spectral density of the excitation force $f(t)$, and FTF was defined earlier.

This optimization is subjected to some constraints in **p** space, where only positive elements are acceptable. Once the optimal absorber suspension properties, c_a and k_a, are determined they can be implemented using adjustment mechanisms on the spring and the damper elements. This is viewed as a semi-active adjustment procedure as it introduces no added energy to the dynamic structure. The conceptual devices for such adjustable suspension elements will be discussed later in 12.3.

© 2002 by CRC Press LLC

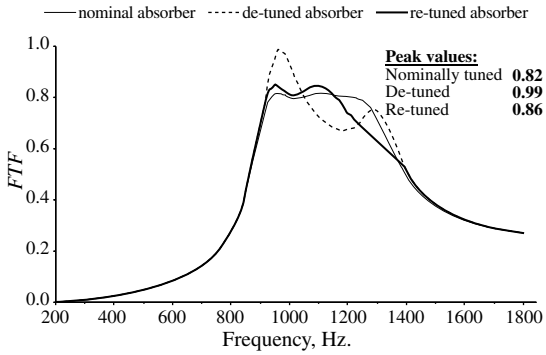


FIGURE 12.4 Frequency transfer functions (FTF) for nominal absorber (thin-solid); de-tuned absorber (thin-dotted); and re-tuned absorber (thick-solid) settings. (From N. Jalili and N. Olgac, 2000, *Journal of Guidance, Control, and Dynamics,* 23 (6), 961–990. With permission.)

12.2.2.3 Simulations

To better recognize the effectiveness of the SA absorber over the passive and optimum passive absorber settings, a simple example case is presented. For the simple system shown in Figure 12.3, the following nominal structural parameters (marked by over score) are taken:

$$\begin{aligned} \bar{m}_p &= 5.77\ kg, \quad \bar{k}_p = 251.132\times10^6\ N/m, \quad \bar{c}_p = 197.92\ kg/s \\ \bar{m}_a &= 0.227\ kg, \quad \bar{k}_a = 9.81\times10^6\ N/m, \quad \bar{c}_a = 355.6\ \ kg/s \end{aligned} \tag{12.10}$$

These are from an actual test setting which is optimal by design. That is, the peak of FTF is minimized (see thinner line in Figure 12.4). When the primary stiffness and damping increase 5% (for instance, during the operation), the FTF of the primary system deteriorates considerably (dashed line in Figure 12.4), and the absorber is no longer an optimum one for the present primary. When the absorber is optimized based on optimization problem (12.8), the re-tuned setting is reached as

$$k_a = 10.29\times10^6\ N/m, \quad c_a = 364.2\ \ kg/s \tag{12.11}$$

which yields a much better frequency response (see darker line in Figure 12.4).

The SA absorber effectiveness is better demonstrated at different frequencies by a frequency sweep test. For this, the excitation amplitude is kept fixed at unity and its frequency changes every 0.15 seconds from 1860 to 1970 Hz. The primary response with nominally tuned, with de-tuned, and with re-tuned absorber settings are given in Figures 12.5a, b, and c, respectively.

12.2.3 Semi-Active Vibration Isolation Design

The parameter tuning control scheme for an SA isolator is similar to that of an SA vibration absorber, with the only difference being in the derivation of the transfer function. The classical isolator system shown in Figure 12.1a and b consists of a rigid body of mass m, linear spring k, and viscous damping c. Conversely, for a vibration absorber, the function of the isolator is to reduce the amplitude of motion transmitted from a moving support to the body (Figure 12.1b), or to reduce the magnitude of the force transmitted from the body to the foundation to an acceptable level (Figure 12.1a).

The transfer functions between isolated mass displacement and base displacement or transmitted force to foundation and excitation force are expressed as

© 2002 by CRC Press LLC

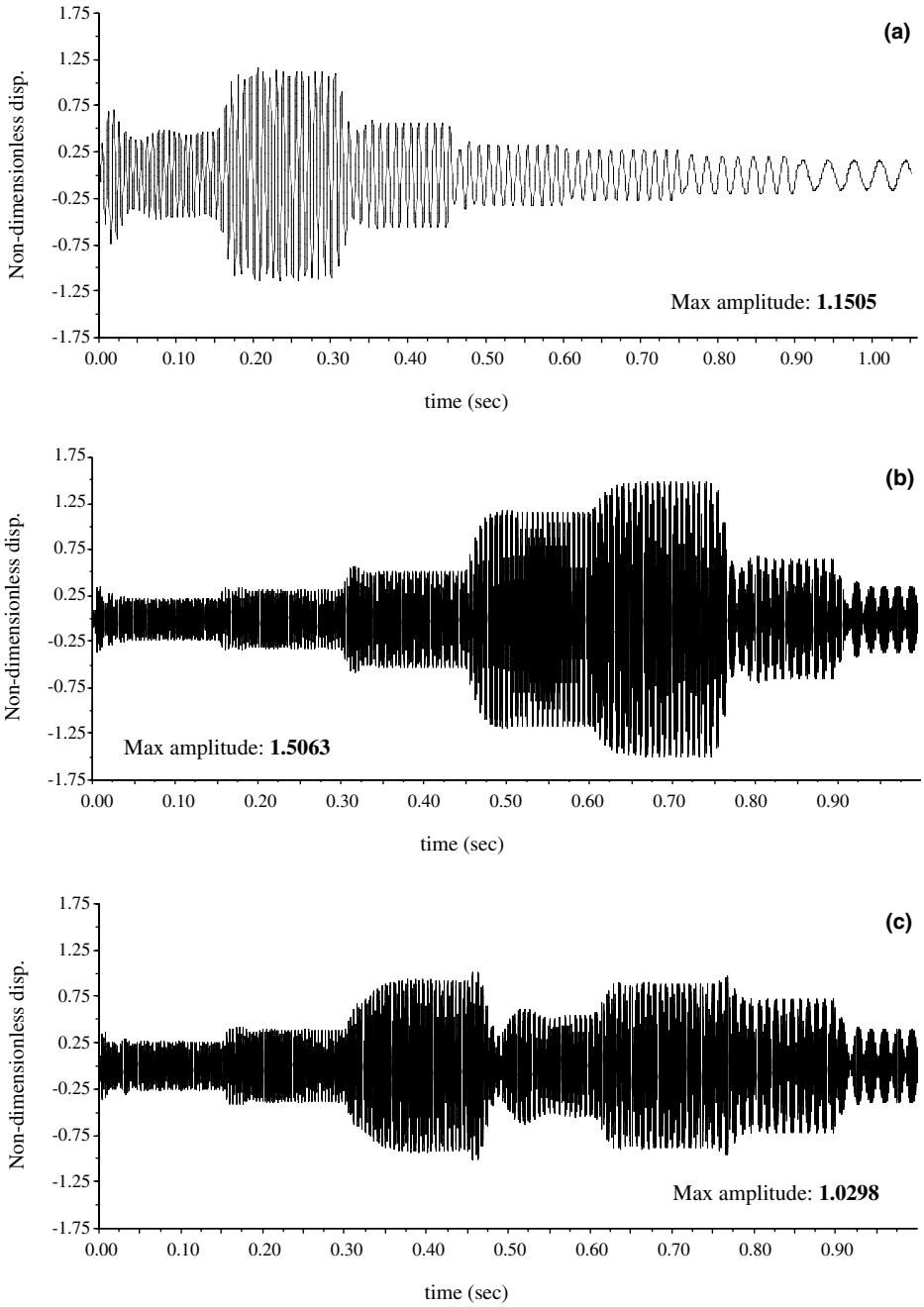


FIGURE 12.5 Frequency sweep each 0.15 with frequency change of [1860, 1880, 1900, 1920, 1930, 1950, 1970] Hz: (a) nominally tuned absorber, (b) de-tuned absorber, and (c) re-tuned absorber settings. (From N. Jalili and N. Olgac, 2000, *Journal of Guidance, Control, and Dynamics,* 23 (6), 961–990. With permission.)

$$\frac{F_T}{F_0} = \frac{X(s)}{Y(s)} = \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{12.12}$$

$$\frac{X(s)}{F(s)} = \frac{1/m}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{12.13}$$

© 2002 by CRC Press LLC

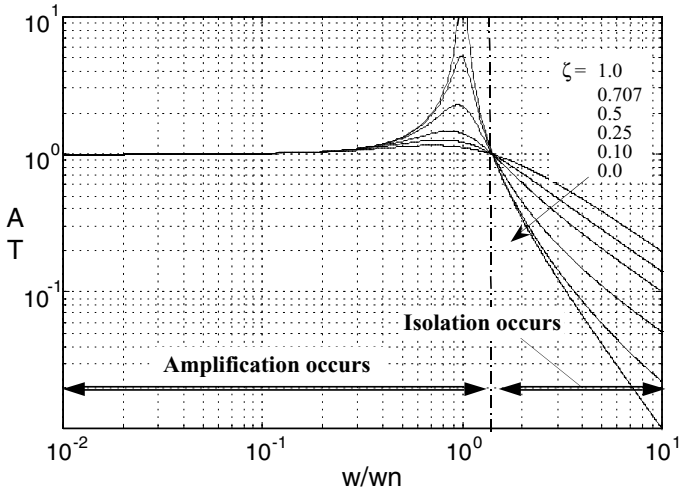


FIGURE 12.6 Frequency response plot of transmissibility T_A for the semi-active suspension as a function of variable damping ratio.

where $\zeta = c / 2\sqrt{km}$ is the damping ratio, $\omega_n = \sqrt{k/m}$ is the natural frequency, and F_T is the amplitude of the transmitted force to the foundation (see Figure 12.1a).

Figure 12.6 shows the transmissibility T_A ($T_A = |F_T / F_0| = |X / Y|$) as a function of the frequency ratio and the damping ratio ζ, where the low frequency range in which the mass displacement essentially follows the base excitation, $X = Y$, is separated from the high-frequency range of isolation, $X < Y$. Near resonance, the T_A is determined completely by the value of the damping ratio. A fundamental problem is that while a high value of the damping ratio suppresses the resonance, it also compromises the isolation for the high-frequency region ($\omega > \omega_n$).

Similar to optimum vibration absorber, an optimal transfer function for the isolator can be obtained as

$$TF(s) = \frac{X}{Y} = \frac{\omega_n^2}{s^2 + 2\zeta_{opt}\,\omega_{opt}\,s + \omega_n^2} \tag{12.14}$$

where $\zeta_{opt} = \sqrt{2}/2$, and ω_{opt} depends upon the weighting factor between mean square acceleration and mean square rattle space in the criterion function used for optimization (similar to problem (12.8) except with transfer function (12.14).[20] The frequency response plot of this transfer function as shown in Figure 12.7 indicates that the damping values sufficient to control the resonance have no adverse effect on high-frequency isolation.

12.2.3.1 Variable Natural Frequency

Similar to an SA absorber, an SA isolator can be utilized for disturbances with time-varying frequency. The variation of natural frequency (which is a function of suspension stiffness) with the transmissibility T_A, in the absence of damping, is given as

$$\omega_n = \omega\sqrt{T_A / (1 + T_A)},\ 0 \le T_A \le 1 \tag{12.15}$$

© 2002 by CRC Press LLC

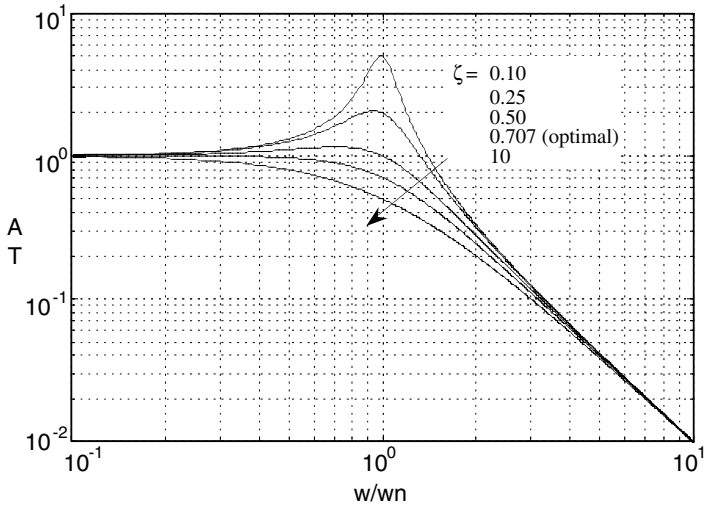


FIGURE 12.7 Frequency response plot of transmissibility T_A for optimum semi-active suspension as a function of variable damping ratio.

With variable disturbance frequency, ω, and desired transmissibility T_A, the natural frequency (or the suspension stiffness k) can be changed in accordance with Equation (12.15) to arrive at optimal performance operation.[21]

12.3 Adjustable Suspension Elements

12.3.1 Introduction

Adjustable suspension elements typically are comprised of a variable rate damper and stiffness. Significant efforts have been devoted to the development and implementation of such devices for a variety of applications. Examples of such devices include electro-rheological (ER),[22-24] magneto-rheological (MR)[25,26] fluid dampers, variable orifice dampers,[27,28] controllable friction braces,[29] controllable friction isolators,[30] and variable stiffness and inertia devices.[12,31-34] The conceptual devices for such adjustable properties are briefly reviewed in this section.

12.3.2 Variable Rate Dampers

A common and very effective way to reduce transient and steady-state vibration is to change the amount of damping in the SA suspension. Considerable design work of semi-active damping was done in the 1960s through 1980s[35,36] for vibration control of civil structures such as buildings and bridges[37] and for reducing machine tool oscillations.[38] Since then, SA dampers have been utilized in diverse applications ranging from trains[39] and other off-road vehicles[40] to military tanks.[41] During recent years considerable interest in improving and refining the SA concept has arisen in industry.[42,43] Recent advances in smart materials have led to the development of new SA dampers, which are widely used in different applications.

In view of these SA dampers, electro-rheological (ER) and magneto-rheological (MR) fluids probably serve as the best potential hardware alternatives for the more conventional variable-orifice hydraulic dampers.[44,45] From a practical standpoint, the MR concept appears more promising for

© 2002 by CRC Press LLC

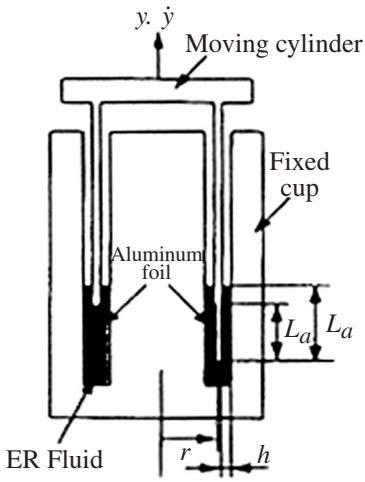


FIGURE 12.8 A schematic configuration of an ER damper. (From S. B. Choi, 1999, *ASME Journal of Dynamic Systems, Measurement and Control,* 121, 134–138. With permission.)

suspension because it can operate, for instance, on a vehicle's battery voltage, whereas the ER damper is based on high-voltage electric fields. Due to their importance in today's SA damper technology, we briefly review their operation and fundamental principles.

12.3.2.1 Electro-Rheological (ER) Fluid Dampers

ER fluids are materials which undergo significant instantaneous reversible changes in material characteristics when subjected to electric potentials (Figure 12.8). The most significant change is associated with complex shear moduli of the material, and hence ER fluids can be usefully exploited in SA suspensions where variable rate dampers are utilized. The idea of applying an ER damper to vibration control was initiated in automobile suspensions, followed by other applications.[46,47]

The flow motion of an ER fluid-based damper can be classified by shear mode, flow mode, and squeeze mode. However, the rheological property of ER fluid is evaluated in the shear mode.[23] Under the electrical potential, the constitutive equation of a ER fluid damper has the form of Bingham plastic[48]

$$\tau = \eta\,\dot{\gamma} + \tau_y(E), \quad \text{and} \quad \tau_y(E) = \alpha\,E^{\beta} \tag{12.16}$$

where τ is the shear stress, η is the fluid viscosity, $\dot{\gamma}$ is shear rate, and $\tau_y(E)$ is yield stress of the ER fluid which is a function of the electric field E. The coefficients α and β are intrinsic values, which are functions of particle size, concentration, and polarization factors.

Consequently, the variable damping force in shear mode can be obtained as

$$F_{ER} = 4\pi r L_d \left\{ \eta\,\dot{y}/h + \alpha E^{\beta}.\mathrm{sgn}(\dot{y}) \right\} \tag{12.17}$$

where h is the electrode gap, L_d is the electrode length of the moving cylinder, r is the mean radius of the moving cylinder, $\dot{y}$ is the transverse velocity of the ER damper, and $\mathrm{sgn}(\cdot)$ represents the signum function (Figure 12.8). As a result, the ER fluid damper provides an adaptive viscous and frictional damping for use in SA systems.[24,49]

© 2002 by CRC Press LLC

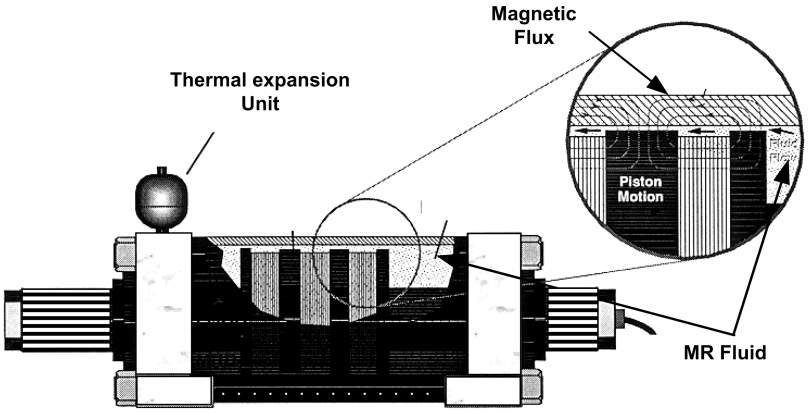


FIGURE 12.9 A schematic configuration of an MR damper.

12.3.2.2 Magneto-Rheological (MR) Fluid Dampers

MR fluids are the magnetic analogs of ER fluid and typically consist of micron-sized, magnetically polarizable particles dispersed in a carrier medium such as mineral or silicon oil. When a magnetic field is applied, particle chains form and the fluid becomes a semisolid, exhibiting plastic behavior similar to that of ER fluids (Figure 12.9). Transition to rheological equilibrium can be achieved in a few milliseconds, providing devices with high bandwidth.[25,26,50]

Similar to Bigham's plasticity model of (12.16), the behavior of controllable fluid is represented by

$$\tau = \eta\,\dot{\gamma} + \tau_y(H) \tag{12.18}$$

where H is the magnetic field. Most devices that use MR fluids can be classified as having either fixed poles (pressure-driven flow mode) or relatively movable poles (direct shear mode). In a manner like ER dampers, the variable force developed by an MR damper in direct-shear mode is

$$F_{MR} = \eta\,A\,\dot{y}/h + \tau_y(H)A \tag{12.19}$$

where $\dot{y}$ is the relative pole velocity, $A = Lw$ is the shear (pole) area, and the rest of the parameters are similar to those in the ER notations used in Figure 12.8.

12.3.3 Variable Rate Spring Elements

In contrast to studies of variable dampers, those of SA springs or time-varying stiffness have been geared for vibration isolation applications,[51] for structural controls, and for vibration attenuation (Reference 2 and references therein). The variable stiffness is a promising practical complement to SA damping, because, based on the discussion in Section 12.2, both the suspension damping and stiffness should change to optimally adapt to different conditions. Clearly, suspension stiffness has a significant influence on optimum operation (even more over the damping element[52]).

Unlike the variable rate damper, changing the effective stiffness requires high energy.[32] Semi-active or low-power implementation of variable stiffness techniques suffers from a limited frequency range, complex implementation, high cost, etc.[12,33,34] Therefore, in practice, both absorber damping and stiffness are concurrently adjusted to reduce the required energy.

© 2002 by CRC Press LLC

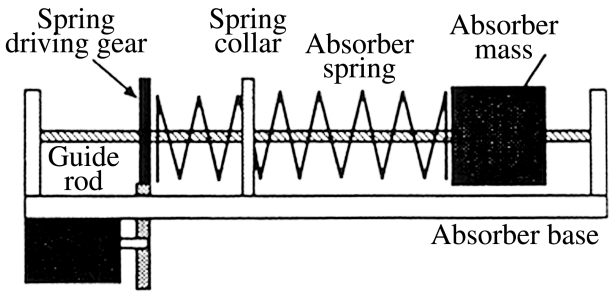


FIGURE 12.10 The application of a variable stiffness vibration absorber to a four-DOF building. (From M.A. Franchek, M.W. Ryan, and R.J. Bernhard, 1995, *Journal of Sound and Vibration,* 189(5), 565–585. With permission.)

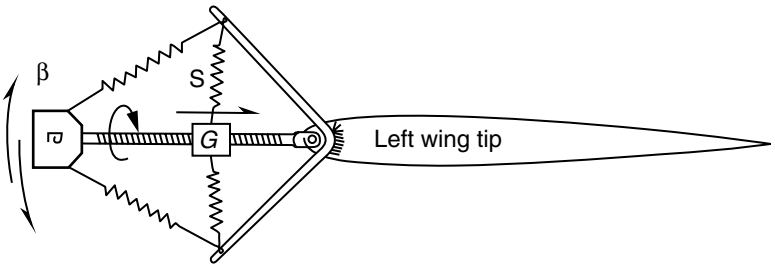


FIGURE 12.11 A semi-active flutter control using adjustable pitching stiffness. (From H. J. Liu, Z. C. Yang, and L. C. Zhao, 2000, *Journal of Sound and Vibration,* 229(1), 199–205. With permission.)

12.3.3.1 Variable Rate Stiffness (Direct Methods):

The primary objective is to directly change the spring stiffness to optimize a vibration suppression characteristic such as Equation (12.8) or (12.9). Different techniques can be utilized from traditional variable leaf-spring to smart-spring utilizing magnetostrictive materials. A tunable stiffness vibration absorber was utilized for a four-DOF building (Figure 12.10), where a spring is threaded through a collar plate and attached to the absorber mass from one side and to the driving gear from the other side.[34] Thus, the effective number of coils, N, can be changed resulting in a variable spring stiffness k_a.

$$k_a = \frac{d^4 G}{8D^3 N} \tag{12.20}$$

where d is the spring wire diameter, D is the spring diameter, and G is the modulus of shear rigidity.

12.3.3.2 Variable Rate Effective Stiffness (Indirect Methods):

In most SA applications, directly changing the stiffness may not always be possible or may require a large amount of control effort. For such cases, alternatives methods are utilized to change the effective tuning ratio ($\tau = \sqrt{k_a / m_a} / \omega_{primary}$), thus resulting in a tunable resonance frequency.

In Liu[53] a semi-active flutter suppression scheme was proposed using differential changes of the external store stiffness. As shown in Figure 12.11, the motor drives the guide screw to rotate with slide block G moving along it, thus changing the restoring moment and resulting in a change of store-pitching stiffness. Using a double-ended cantilever beam carrying intermediate lumped masses, a semi-active vibration absorber was recently introduced,[54] where the position of moving masses was adjustable (see Figure 12.12). Figure 12.13 shows an SA absorber with an adjustable

© 2002 by CRC Press LLC

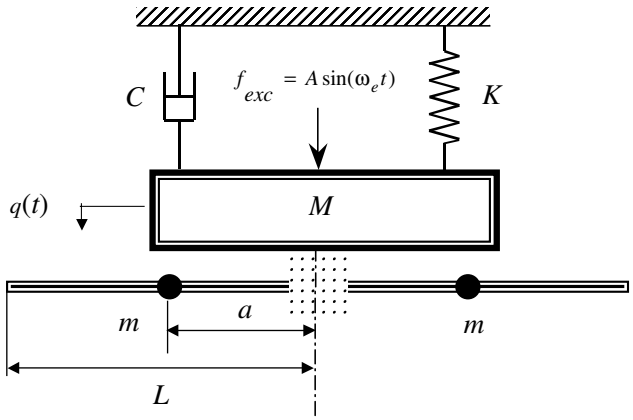


FIGURE 12.12 A typical primary system equipped with the double-ended cantilever absorber with adjustable tuning ration through moving masses *m*. (From N. Jalili, 2000, *Proceedings of 2000 International Mechanical Engineering Congress and Exposition,* Orlando, FL. With permission.)

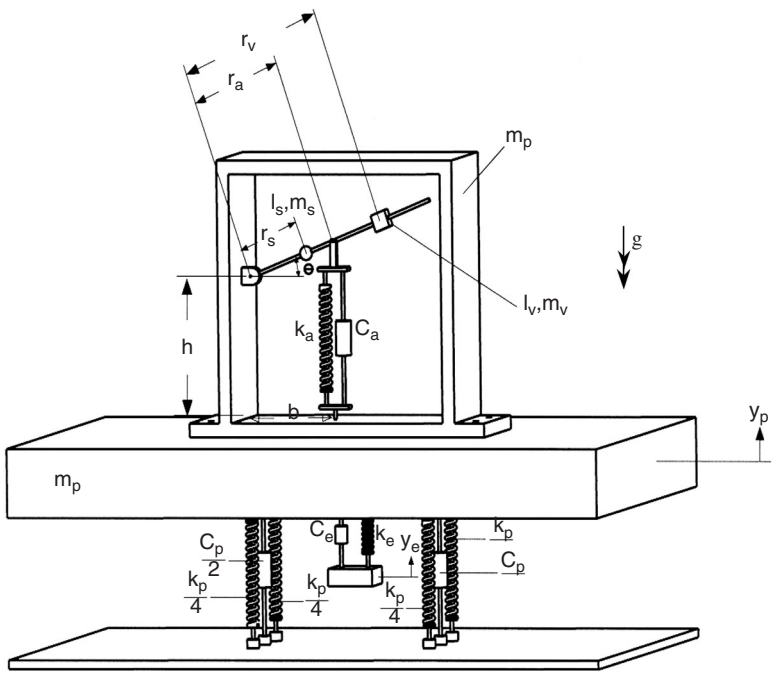


FIGURE 12.13 Schematic of the adjustable effective inertia vibration absorber. (From N. Jalili, B. Fallahi, and Z. K. Kusculuoglu, 2001, *International Journal of Modelling and Simulation,* 21(2), 148–154. With permission.)

effective inertia mechanism.[55] The SA absorber consists of a rod carrying a moving block and a spring and damper mounted on a casing. The position of the moving block, r_v, on the rod is adjustable, which provides a tunable resonance frequency.

12.3.4 Other Variable Rate Elements

Recent advances in smart materials have led to the development of new SA suspensions using indirect influence on the suspension elements. A semi-active piezoelectric network was utilized[16]

© 2002 by CRC Press LLC

for structural vibration control. The variable resistance and inductance in an external RL circuit are used as real-time adaptable control parameters.

Another class of adjustable suspensions is the so-called hybrid treatment.[56] The hybrid design has two modes: active and passive. With the aim of lowering the control effort, relatively small vibrations are reduced in the active mode, while the passive mode is used for large oscillations. Analogous to hybrid treatment, the semi-automated approach combines semi-active and active suspensions to benefit the advantages of individual schemes while eliminating their shortfalls.[57] By altering the adjustable structural properties (in a semi-active unit) and control parameters (in an active unit), a search is conducted to minimize an objective function subject to certain constraints, which may reflect performance characteristics.

12.4 Automotive Semi-Active Suspensions

12.4.1 Introduction

Earlier studies on SA suspensions focused on automobile-related applications. One notable reason is that the importance of energy dissipation in suspension systems is recognized most in automotive suspensions, where ride comfort and vehicle handling are encountered. For this reason, a section is devoted to the application of SA systems to automotive suspension. The objectives here are to briefly review the fundamental design aspects in automobile semi-active suspension and present some recent developments in this area.

12.4.2 An Overview of Automotive Suspensions

Advanced vehicle suspension systems such as adaptive, semi-active, and active have been used extensively in most conventional ground transport fleets. Due to slow response time in adaptive systems and high energy consumption and cost in active suspensions, they are unlikely to survive in the future market. Recently, much attention is being paid to controllable active or semi-active elements.[58-60]

Due to the large forces and velocities involved in suspension systems, it is important to minimize the actuator power requirement for practical and economical reasons.[36] For the actuator in semi-active suspension systems, multistage dampers and continuously variable dampers,[36] or variable lever ratio systems and modulated transformers are being utilized. These suspensions are called low bandwidth or fast load lever systems and often incorporate semi-active dampers which produce high-frequency controllable forces with low power requirements.

In vehicle suspensions, physical actuator limitations or cost considerations may render an elegant design concept totally impractical. For this reason, interest has surfaced in exploring the possibility of improving suspension performance by modulating the characteristics of essentially passive elements such as springs and dampers. SA suspensions represent a compromise between performance improvement and simplicity of implementation.

12.4.3 Semi-Active Vehicle Suspension Models

Different models are used for the design of a SA suspension. These models range from the simplest one, a single DOF quarter car model which allows for only one-dimensional vertical or heave motion, to very complex with many DOFs.[60,61] To illustrate the theoretical concepts and avoid disturbing the focus of the subject, we briefly discuss using a simple quarter car (SQC) model (Figure 12.14), which may be achieved by linear damping and spring stiffness variations. Although this is a simple model, it is quite suitable to study the performance of vehicle suspension in both bounce motion and tire deflection.[62]

© 2002 by CRC Press LLC

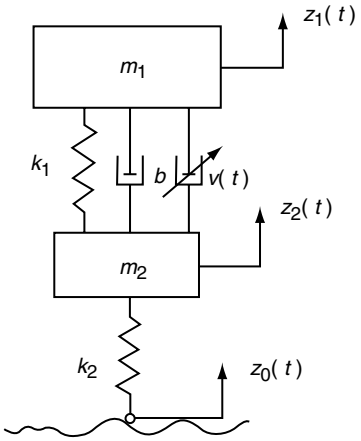


FIGURE 12.14 An SQC model of vehicle suspension system.

The governing equations of motion for the sprung and unsprung masses are

$$\begin{aligned} m_1\ddot{z}_1 + k_1(z_1 - z_2) + b(\dot{z}_1 - \dot{z}_2) &= 0 \\ m_2\ddot{z}_2 + k_1(z_2 - z_1) + b(\dot{z}_2 - \dot{z}_1) + k_2(z_2 - z_0) &= 0 \end{aligned} \tag{12.21}$$

where m_1 is a quarter of the body mass (sprung), m_2 is the mass of the wheel, b and k_1 are the adjustable damping and stiffness of the suspension, and the rest of the parameters are defined in Figure 12.14.

Figure 12.15 shows such an adjustable damper, whereby the check valves assure that for both directions of piston motion, the hydraulic fluid flows the same way through a solenoid-controlled blow-off valve, thus resulting in variable damping. To demonstrate the effect of suspension element variations on ride comfort, the frequency response of body velocity (as a measure of ride comfort) is shown in Figure 12.16. The adjustable damper and stiffness are optimized with respect to ride comfort, suspension rattle space, and road handling. A performance characteristic is then constructed to perform this optimization.

12.4.4 Semi-Active Suspension Performance Characteristics

It is important to recognize that automobile suspension must perform several tasks in addition to isolating the body from vibration induced by road unevenness.[59] The body attitude, the attitude of each wheel with respect to road surface, dynamic normal force variations at each wheel, and many other criteria must be controlled. Although the focus here is on vibration isolation of suspension systems, a good design should allow for meeting several conflicting requirements.

An optimal SA control problem is, therefore, formulated (for the SQC model of Figure 12.14) to briefly highlight the design procedure. For the performance index (PI) in the design of vehicle suspension, sprung mass acceleration, suspension travel, and tire spring excursion can be incorporated. Sprung mass acceleration is a measure of body isolation, i.e., passenger ride comfort. Suspension travel or rattle space is typically a design constraint for limiting rigid body motion of the vehicle. Tire spring stroke (or equivalently, dynamic tire force) is an indicator of road-holding ability. Accordingly, a PI of the following form can be selected:

$$PI = \frac{1}{2T} E \int_0^T \left\{ \gamma_1 \ddot{z}_2^2 + \gamma_2 (z_1 - z_0)^2 + \gamma_3 (z_2 - z_1)^2 \right\} dt \tag{12.22}$$

© 2002 by CRC Press LLC

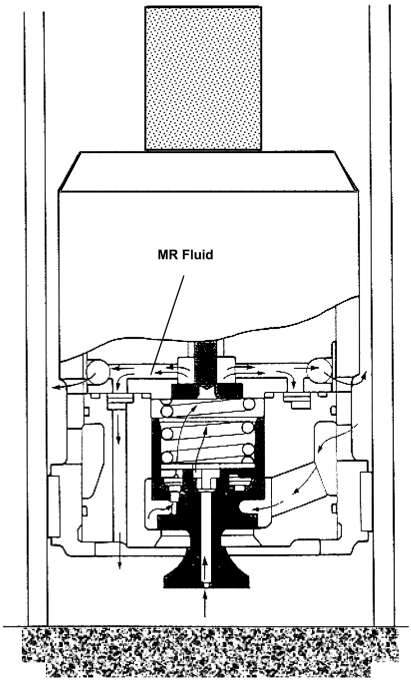


FIGURE 12.15 Schematic design of the Nissan electro-hydraulic valve in the piston of a semi-active damper.

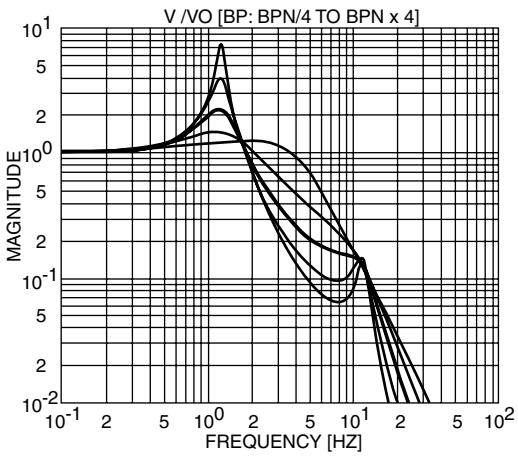


FIGURE 12.16 Variations in frequency response of body velocity for SQC model with variable damper. (From D. Karnopp, 1995, *ASME Transactions, Special 50th Anniversary, Design Issue,* 117, 177–185. With permission.)

where E denotes the expectations necessary because of the random road disturbance input z_0; T is a sufficient large endtime; and γ_1, γ_2, and γ_3 are weighting factors for the penalized variables.

Given the linear system described by Equation (12.21), a control sequence $U(t)$ can be chosen to minimize the PI given in Equation (12.22), under the passivity constraint[13]

$$U(t)\left[\,\dot{z}_1(t)-\dot{z}_2(t)\,\right]\geq 0,\ 0\leq t\leq T \tag{12.23}$$

© 2002 by CRC Press LLC

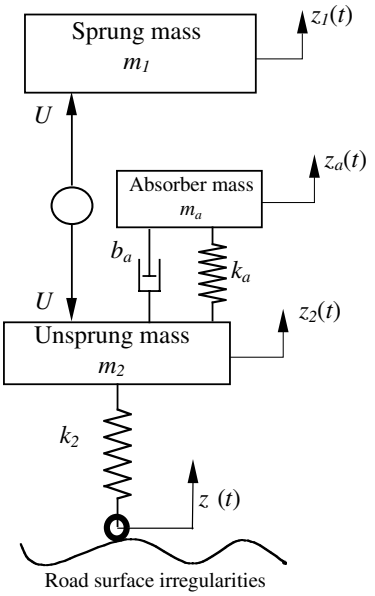


FIGURE 12.17 A two-DOF vehicle model with dynamic vibration absorber.

In addition, because the vehicle structure can tolerate only bounded suspension forces it is required that

$$| U(t) | \le U_M,\ 0 \le t \le T \tag{12.24}$$

where $U_M > 0$ is the maximal allowed force. Many exact (numerical) and approximate (analytical) solutions to this problem exist. We leave the details to Hrovat, Margolis, and Hubbard[13] and Hrovat.[61]

12.4.5 Recent Advances in Automotive Semi-Active Suspensions

The SA concept has been applied to a broad class of ground transport fleets, ranging from tractors and other farm vehicles to high-speed ground transportation vehicles. The SA suspension concept goes back to the early 1970s[35] in the form of variable, controllable damping. Although the focus here is on vibration isolation through vehicle suspension design, it is worthwhile mentioning that a few applications of vibration absorber with the aim of improving ride comfort have been used (see Figure 12.17).[64]

Some developments include SA suspension with variable stiffness,[65] electro-hydro-pneumatic slow-active suspension,[66] SA suspension using ER fluid mount,[67] fast load-lever suspension with a variable lever rate,[68] SA gas suspension for off-road vehicles,[40] SA suspension for passenger trains,[39] and SA suspension using a piston-controlled disk valve.[28]

12.5 Application of Control Techniques to Semi-Active Suspensions

12.5.1 Introduction

As discussed in the preceding section, the SA suspension generates forces passively, but these forces are modulated continuously in accordance with some prescribed control law with only small

© 2002 by CRC Press LLC

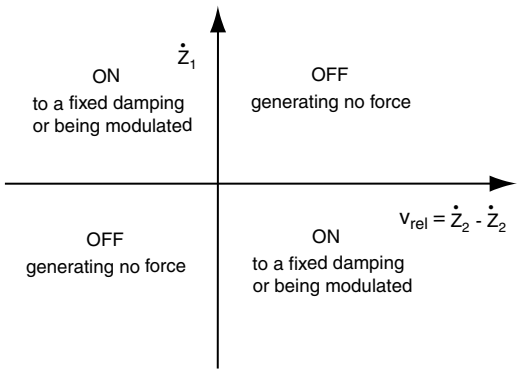


FIGURE 12.18 On-off semi-active control decision.

amount of external power. In other words, SA suspension is basically a device with time-varying controllable damping and spring.

The concept of SA control[36] has been developed and demonstrated to be a viable suspension alternative. Although not rigorously proven, damper and stiffness can be treated much like active force generators for the purpose of controller design. That is, the SA damper or spring is modulated according to the same control policy and same sate measurement as its fully active force generator counterpart. Obviously, the sign of the damper or spring force is dictated by the relative motion across it, and thus cannot be specified. This section briefly reviews the control techniques for SA suspensions.

12.5.2 Semi-Active Control Concept

The elementary SA controller design is the so-called on-off SA strategy, which was first proposed by Margolis, Tylee, and Hrovat.[69] It switches the damper off whenever sprung and unsprung masses move in the same direction and unsprung mass has a larger velocity. In any other situations the damper is set to the on position. The schematic of the conceptual control law is shown in Figure 12.18.

A somewhat more sophisticated approach is to change the damping from soft to firm and visa versa through a manual or slow adaptive control. This is referred to as the on-off skyhook control policy, whereby the damper forces are controlled like the configuration shown in Figure 12.19. Mathematically, the on-off skyhook control policy can be described as

$$\begin{aligned} \dot{z}_1(\dot{z}_1 - \dot{z}_2) \geq 0, \quad & c = \text{high damping} \\ \dot{z}_1(\dot{z}_1 - \dot{z}_2) < 0, \quad & c = \text{low damping} \end{aligned} \tag{12.25}$$

The combination of relative velocity damping forces and skyhook components is very effective in damping body response without detrimental effects (refer to Figure 12.16) on isolation for the frequencies between the body resonance frequency and the wheel hop frequency.[13] The frequency response is demonstrated in Figure 12.20, where significant improvement is attained over the conventional variable damping configuration of Figure 12.16.

During recent years considerable interest in the on-off SA concept has developed. Further improvements and refinements of the concept were reported (see Reference 60 and references therein]. Recent developments in multivariable control design methodology and microprocessor implementation of modern control algorithms have opened a new era for the design of externally controlled passive systems for use in SA suspensions.

© 2002 by CRC Press LLC

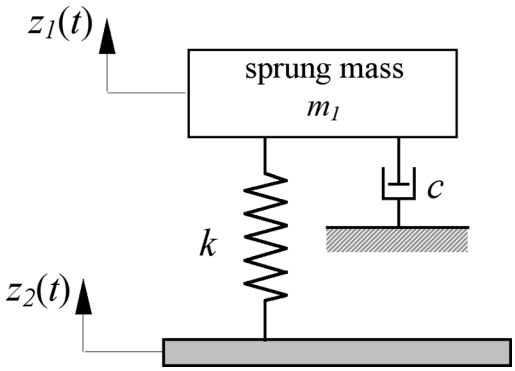


FIGURE 12.19 Schematic of skyhook damper arrangement.

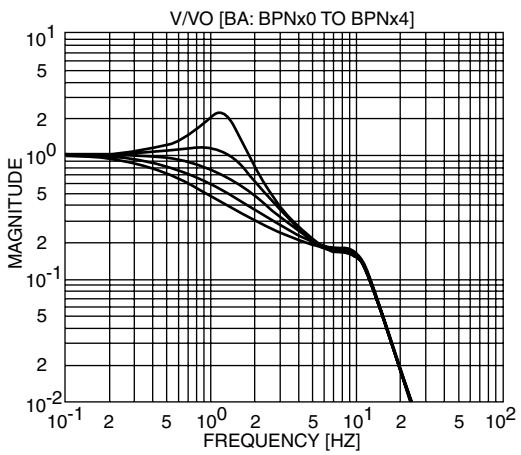


FIGURE 12.20 Variations in frequency response of body velocity for SQC model with combination of variable damper and skyhook damping. (From D. Karnopp, 1995, *ASME Transactions, Special 50th Anniversary, Design Issue,* 117, 177–185. With permission.)

12.5.3 Optimal Semi-Active Suspension

The continuously variable SA policy represents the next step up in sophistication. It requires that the SA actuator continuously reproduce a linear quadratic (LQ) optimal control skyhook damping force whenever this is possible in view of the passivity constraint.[13] When this is not possible, the damper is simply turned off. The continuously variable SA policy was subsequently extended to a more complex model, which led to so-called clipped SA control.[60] The optimal SA control law was first studied in Hrovat.[70] It was later proved that the clipped SA policy may often be very close to being optimal but not always.

The fundamental concepts of optimum SA are similar to the optimum automotive suspension systems discussed in 12.4.4. Simple, mostly LQ-based optimal control concepts give useful insights about the performance characteristics and other requirements.[60,70]

12.5.4 Other Control Techniques

As a result of substantial ongoing theoretical advances in the areas of adaptive and nonlinear controls,[71,72] it is expected that there will be future applications of these techniques in advanced

© 2002 by CRC Press LLC

suspension design. For practical implementation, however, it is preferable to simplify these strategies, thus leading to simpler software implementations. For instance, suboptimal policy neglecting some performance requirements can serve as an example of such simplifications. Some recent developments in control techniques for SA suspensions include fuzzy reasoning,[73] adaptive SA,[74] SA suspension with observer design,[75] and many others.

12.6 Practical Considerations and Related Topics

SA suspensions can achieve most of the performance characteristics of fully active systems, thus allowing for a wide class of applications. The idea of SA suspension is very simple: to replace active force generators with continually adjustable elements which can vary and/or shift the rate of the energy dissipation in response to instantaneous condition of motion.

The fundamental principles of SA suspension were formulated here. Many important areas are related directly or indirectly to the main theme of this chapter. These include practical implementation of SA suspensions, nonlinear control schemes, actual hardware implementation, actuator bandwidth requirements, reliability, and cost. Furthermore, in the process of designing an SA suspension, in practice, several critical criteria must be considered. These include weight, size, shape, center-of-gravity, types of dynamic disturbances, allowable system response, ambient environment, and service life.

SA suspensions provide vibration suppression solutions for tonal and broadband applications with a small amount of control and relatively low cost. However, using conventional technologies to build a practical SA suspension under the constraints of weight, size, and cost is quite a design challenge. Furthermore, the design of SA suspensions involves many mechanical and electrical components that put a limit on the tuning range of the resonance frequency of the device.

References

1. Inman, D. J., 1994, *Engineering Vibration*, Prentice-Hall, Englewood Cliffs, NJ.
2. Sun, J. Q., Jolly, M. R., and Norris, M. A., 1995, Passive, adaptive, and active tuned vibration absorbers — A survey, *ASME Transactions, Special 50th Anniversary, Design Issue*, 117, 234–242.
3. Korenev, B. G. and Reznikov, L. M., 1993, *Dynamic Vibration Absorbers: Theory and Technical Applications*, John Wiley & Sons, Chichester.
4. Puksand, H., 1975, Optimum conditions for dynamic vibration absorbers for variable speed systems with rotating and reciprocating unbalance, *International Journal of Mechanical Engineering Education*, 3, 145–152.
5. Warburton, G. B. and Ayorinde, E. O., 1980, Optimum absorber parameters for simple systems, *Earthquake Engineering and Structural Dynamics*, 8, 197–217.
6. Esmailzadeh, E. and Jalili, N., 1998, Optimal design of vibration absorbers for structurally damped Timoshenko beams, *ASME Journal of Vibration and Acoustics*, 120(4), 833–841.
7. Soong, T. T. and Constantinou, M. C., 1994, *Passive and Active Structural Control in Civil Engineering*, Springer-Verlag, Wien.
8. Olgac, N. and Holm-Hansen, B., 1994, A novel active vibration absorption technique: Delayed resonator, *Journal of Sound and Vibration*, 176, 93–104.
9. Margolis, D., 1998, Retrofitting active control into passive vibration isolation systems, *ASME Journal of Vibration and Acoustics,* 120, 104–110.
10. Lee-Glauser, G. J., Ahmadi, G., and Horta, L. G., 1997, Integrated passive/active vibration absorber for multistory buildings, *ASCE Journal of Structural Engineering*, 123(4), 499–504.
11. Franchek, M. A., Ryan, M. W., and Bernhard, R. J., 1995, Adaptive-passive vibration control, *Journal of Sound and Vibration*, 189(5), 565–585.
12. Nemir, D., Lin, Y., and Osequeda, R. 1994, Semi-active motion control using variable stiffness, *ASCE Journal of Structural Engineering*, 120(4), 1291–1306.

© 2002 by CRC Press LLC

13. Hrovat, D., Margolis, D. L., and Hubbard, M., 1988, An approach toward the optimal semi-active suspension, *ASME Journal of Dynamic Systems, Measurement and Control*, 110, 288–296.
14. Shaw, J., 1998, Adaptive vibration control by using magnetostrictive actuators, *Journal of Intelligent Material Systems and Structures,* 9, 87–94.
15. Garcia, E., Dosch, J., and Inman, D. J., 1992, The application of smart structures to the vibration suppression problem, *Journal of Intelligent Material Systems and Structures,* 3, 659–667.
16. Wang, K. W., Lai, J. S., and Yu, W. K., 1996, An energy-based parametric control approach for structural vibration suppression via semi-active piezoelectric networks, *ASME Journal of Vibration and Acoustics*, 118, 505–509.
17. Frahm, H., 1911, Devices for damping vibrations of bodies, U.S. Patent # 989,958.
18. Snowdon, J. C., 1968, *Vibration and Shock in Damped Mechanical Systems*, John Wiley & Sons, New York, NY.
19. Jalili, N. and Olgac, N., 2000, Identification and re-tuning of optimum delayed feedback vibration absorber, *AIAA Journal of Guidance, Control, and Dynamics,* 23(6), 961–970.
20. Karnopp, D., 1995, Active and semi-active vibration isolation, *ASME Transactions, Special 50th Anniversary, Design Issue*, 117, 177–185.
21. Esmailzadeh, E., 1978, Vibration isolation system with variable natural frequency, *IJMEE*, 6(3), 125–129.
22. Petek, N. K., 1992, Shock absorbers uses electrorheological fluid*, Automotive Engineering*, 100 (6), 27–30.
23. Choi, S. B., 1999, Vibration control of flexible structures using ER dampers, *ASME Journal of Dynamic Systems, Measurement and Control*, 121, 134–138.
24. Wang, K. W., Kim, Y. S., and Shea, D. B., 1994, Structural vibration control via electrorheological-fluid-based actuators with adaptive viscous and frictional damping, *Journal of Sound and Vibration*, 177(2), 227–237.
25. Spencer, B. F., Yang, G., Carlson, J. D., and Sain, M. K., 1998, Smart dampers for seismic protection of structures: A full-scale study, *Proceedings 2nd World Conference on Structural Control*, Kyoto, Japan, June 28–July 1.
26. Kim, K. and Jeon, D., 2000, Vibration suppression in an MR fluid damper suspension system, *Journal of Intelligent Material Systems and Structures*, 10(10), 779–786.
27. Mizuno, T., Kobori, T., Hirai, J., Matsunaga, Y., and Niwa, N., 1992, Development of adjustable hydraulic dampers for seismic response control of large structure, *ASME PVP Conference*, 229, 163–170.
28. Sun, Y. and Parker, G. A., 1993, A position controlled disc valve in vehicle semi-active suspension systems, *Control Eng. Practice*, 1(6), 927–935.
29. Dowell, D. J. and Cherry, S., 1994, Semi-active friction dampers for seismic response control of structures, *Proceedings 5th U.S. National Conference on Earthquake Engineering*, 1, 819–828.
30. Feng, Q. and Shinozuka, M., 1990, Use of a variable damper for hybrid control of bridge response under earthquake, *Proceedings of U.S. National Workshop on Structural Control Research*, USC Publication No. CE-9013, 107–112.
31. Giliomee, C. L. and Els, P. S., 1998, Semi-active hydropneumatic spring and damper system, *Journal of Terramechanics,* 35, 109–117.
32. Walsh, P. L. and Lamnacusa, J. S., 1992, A variable stiffness vibration absorber for minimization of transient vibrations, *Journal of Sound and Vibration*, 158(2), 195–211.
33. Nagarajaiah, 1997, Semi-active control of structures, *Proceedings of Structural Congress, ASCE*, Portland, OR, 1574–1578.
34. Franchek, M. A., Ryan, M. W., and Bernhard, R. J., 1995, Adaptive passive vibration control, *Journal of Sound and Vibration*, 189(5), 565–585.
35. Crosby, M. and Karnopp, D. C., 1973, The active damper- A new concept for shock and vibration control, *Shock Vibration Bulletin*, *Part H*, Washington, D.C.
36. Karnopp, D. C., Crodby, M. J., and Harwood, R. A., 1974, Vibration control using semi-active force generators, *Journal of Engineering for Industry*, 96(2), 619–626, May.

© 2002 by CRC Press LLC

37. Hrovat, D., Barker, P., and Rabins, M., 1983, Semi-active vs. passive or active tuned mass dampers for structural control, *Journal of Engineering Mechanics*, 109, 691–705.
38. Tanaka, N. and Kikushima, Y., 1992, Impact vibration control using a semi-active damper, *Journal of Sound and Vibration*, 158(2), 277–292.
39. Stribersky, A., Muller, H., and Rath, B., 1998, The development of an integrated suspension control technology for passenger trains, *Proceedings of the Institute for Mechanical Engineers*, 212, Part F, 33–41.
40. Horton, D. N. and Crolla, D. A., 1986, Theoretical analysis of a semi-active suspension fitted to an off-road vehicle, *Vehicle System Dynamics*, 15, 351–372.
41. Miller, L. R. and Nobles, C. M., 1988, The Design and Development of a Semi-Active Suspension for Military Tank, SAE Paper No. 881133.
42. Karnopp, D., 1990, Design principles for vibration control systems using semi-active dampers, *ASME Journal of Dynamic Systems, Measurement and Control*, 112(3), 448–455.
43. Emura, J., Kakizaki, S., Yamaoka, F., and Nakamura, M., 1994, Development of the Semi-active Suspension System Based on the Sky-hook Damper Theory, SAE Paper No. 940863.
44. Pinkos, A., Shtarkman, E., and Fitzgerald, T., 1994, An actively damped passenger car suspension system with low voltage electro-rheological magnetic fluid, *Proceedings International Symposium on Advanced Vehicle Control* (AVEC), Tsukuba, Japan, 311–317.
45. Sturk, M., Wu, M., and Wong, J. Y., 1995, Development and evaluation of a high voltage supply unit for electorheological fluid dampers, *Vehicle System Dynamics*, 24, 101–121.
46. Petek, N. K., Romstadt, D. L., Lizell, M. B., and Weyenberg, T. R., 1995, Demonstration of an automotive semi-active suspension using electro-rheological fluid, SAE Paper No. 950586.
47. Austin, S. A., 1993, The vibration damping effect of an electrorheological fluid, *ASME Journal of Vibration and Acoustics*, 115(1), 136–140.
48. Ginder, J. M. and Ceccio, S. L., 1995, The effect of electrical transients on the shear stresses in electrorheological fluids, *Journal of Rheology*, 39(1), 211–234.
49. Dimarogonas-Andrew, D. and Kollias, A., 1993, Smart electrorheological fluid dynamic vibration absorber, *Intelligent Structures, Materials, and Vibration, ASME Design Division,* 58, 7–15.
50. Lord Corporation, http://www.rheonetic.com.
51. Hubard, M. and Marolis, D., 1976, The semi-active spring: Is it a viable suspension concept?, *Proceedings 4th Intersociety Conference on Transportation*, 1–6.
52. Jalili, N. and Olgac, N., 2000, A sensitivity study of optimum delayed feedback vibration absorber, *ASME Journal of Dynamic Systems, Measurement, and Control,* 121, 314–321.
53. Liu, H. J., Yang, Z. C., and Zhao, L. C., 2000, Semi-active flutter control by structural asymmetry, *Journal of Sound and Vibration*, 229(1), 199–205.
54. Jalili, N., 2000, On adaptive-passive vibration suppression using distributed-parameter absorbers, *Proceedings of 2000 International Mechanical Engineering Congress and Exposition*, Orlando, FL.
55. Jalili, N., Fallahi, B., and Kusculuoglu, Z. K., 2001, A new approach to semi-active vibration suppression using adjustable inertia absorbers, *International Journal of Modelling and Simulation*, 21(2), 148–154.
56. Fujita, T., Katsu, M., Miyano, H., and Takanashi, S., 1991, Fundamental study of active-passive mass damper using XY-motion mechanism and hydraulic actuator for vibration control of tall building, *Transactions of Japan Society of Mechanical Engineers*, *Part C*, 57, 3532–3539.
57. Jalili N., 2000, A new perspective for semi-automated structural vibration control, *Journal of Sound and Vibration,* 238(3), 481–494.
58. ElBeheiry, E. M., Karnopp, D., ElAraby, M. E., and Abdelraaouf, A. M., 1995, Advanced ground vehicle suspension systems (a classified bibliography), *Vehicle System Dynamics*, 14(3), 231–258.
59. Karnopp, D. and Hess, G., 1991, Electronically controllable vehicle suspensions, *Vehicle System Dynamics*, 20(3-4), 207–217.
60. Hrovat, D., 1997, Survey of advanced suspension developments and related optimal control applications, *Automatica*, 33(10), 1781–1817.

© 2002 by CRC Press LLC

61. Hrovat, D., 1993, Applications of optimal control to advanced automotive suspension design, *ASME Journal of Dynamic Systems, Measurement, and Control*, 115, 328–342.
62. Jalili, N. and Esmailzadeh, E., 2001, Optimum active vehicle suspensions with actuator time delay, *ASME Journal of Dynamic Systems, Measurement, and Control*, 123, 54–61.
63. Anon., 1989, Nissan Active Hydraulic Suspension, Nissan Motor Co. Ltd., Tokyo, Japan.
64. Hrovat, D., 1990, Optimal active suspension structures for quarter-car vehicle models, *Automatica*, 26(5), 845–860.
65. Youn, I. and Hac, A., 1995, Semi-active suspensions with adaptive capability, *Journal of Sound and Vibration*, 180(3), 475–492.
66. Sharp, R. S., 1998, Variable geometry active suspension for cars, *IEE Computing and Control Engineering Journal*, October, 217–222.
67. Duclos, T. G., 1988, Design devices using electrorheological fluids, SAE Paper No. 881134.
68. Karnopp, D. and So, S. G., 1998, Energy flow in active attitude control suspensions: A bond graph analysis, *Vehicle System Dynamics*, 29, 69–91.
69. Margolis, D. L., Tylee, J. L., and Hrovat, D., 1975, Heave mode dynamics of a tracked air cushion vehicle with semi-active airbag secondary suspension, *ASME Journal of Dynamic Systems, Measurement, and Control*, 97(4), 399–407.
70. Hrovat, D., 1979, Optimal Passive Vehicle Suspension, Ph.D. thesis, University of California, Davis, CA.
71. Astrom, J. J. and Wittenmark, B., 1989, *Adaptive Control,* Addison-Wesley, Reading, MA.
72. Alleyne, A. and Hedrick, J. K., 1995, Nonlinear adaptive control of active suspensions, *IEEE Transactions on Control System Technology*, 3(1), 94–101.
73. Yoshimura, T., 1998, A semi-active suspension of passenger cars using fuzzy reasoning and the field testing, *International Journal of Vehicle Design*, 19(2), 150–166.
74. Venhovens, P. J., 1994, The development and implementation of adaptive semi-active suspension control, *Vehicle System Dynamics*, 23, 211–235.
75. Hedrick, J. K., Rajamani, R., and Yi, K., 1994, Observer design for electronic suspension applications, *Vehicle System Dynamics*, 23, 413–440.

© 2002 by CRC Press LLC

13

Semi-Active Suspension Systems II

13.1 Concepts of Semi-Active Suspension Systems
Karnopp's Original Concept • Sky-Hook for Comfort • Extended Ground-Hook for Road-Tire Forces • Semi-Active Actuators and Their Models

13.2 Control Design Methodology
General Design Methodology • Clipped Active Control • MOPO Approach • NQR Approach • Preview Control

13.3 Properties of Semi-Active Suspensions: Performance Indexes
Influence on Comfort • Influence on Road Friendliness

13.4 Examples of Practical Applications
Passenger Cars • Road-Friendly Trucks • Trains • Airplanes

Michael Valášek
Czech Technical University, Prague

Willi Kortüm
German Aerospace Research Establishment

The concept of suspension systems is treated for a broad array of vehicles including trains, airplanes (during ground motion), and off-road vehicles. The vehicle suspension has many important functions:

- Control of the attitude of the vehicle body with respect to the road surface
- Control of the attitude of the wheels with respect to both the road surface and the vehicle body
- Isolation of the vehicle body from forces generated by the roadway unevenness
- Control of the contact forces between wheels and the road surface
- Control of lateral and longitudinal motions

All these functions can be significantly improved via electronic control added into the system. This makes the suspension "active." In this section the treatment of suspension is limited just to the vibration control. We primarily focus on the isolation function and the contact force variation control (in particular, the normal component of it).

The lexicon commonly used in vehicle dynamics "semi-active" control implies that the control actuator requires very little power. Such control is where the actuator possesses many attributes of conventional (active) control but which requires very little control power. A semi-active actuator typically dissipates energy, thus it does not raise stability concerns.

13.1 Concepts of Semi-Active Suspension Systems

13.1.1 Karnopp's Original Concept

The fundamental concepts of semi-active suspension and semi-active vibration control go back to Karnopp's work.[1–3] For the suspension elements, which are electronically, controlled, a critical issue

© 2002 by CRC Press LLC

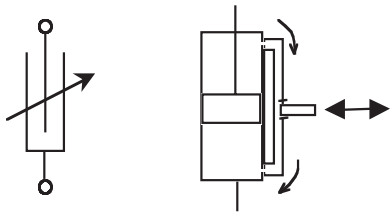


FIGURE 13.1 Variable damper and the working principle.

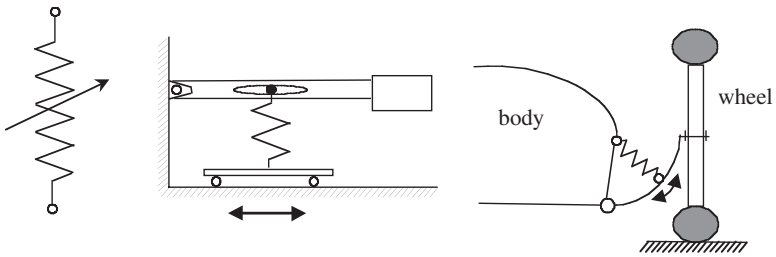


FIGURE 13.2 Variable force transfer and the working principle.

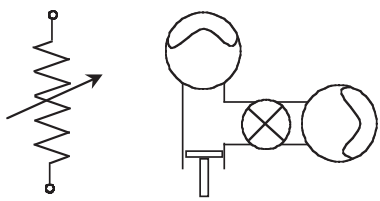


FIGURE 13.3 Variable spring stiffness and the working principle.

is the power consumed. The semi-active element must be either dissipative or conservative when it comes to their energy needs. There are a number of general classes of such devices.

The first class is the variable resistors, which dissipate energy. In a typical vehicle suspension it is the variable damper. Constitutive laws between the system variables of force and velocity characterize these elements. These relations can be rapidly altered using a control input, which consumes very little power (Figure 13.1). Practically, it is conceived as a variable orifice viscous damper. By closing or opening the orifice the damping characteristics change from soft to hard and vice versa. Recently, this flow control has been achieved using electro- and magneto-rheological fluids and is available as industrial products.[39]

The second class is variable force transformers, which conserve energy between suspension and spring storage. Within the vehicle's suspension it is the variable lever arm. These elements are characterized by controlled force variation, which consumes minimal power (Figure 13.2). The physical materialization is conceived as a variable lever on which the force acts. By moving the point of force application, the force transfer ratio change. If these points move orthogonally to the acting force, theoretically no mechanical work is involved in control.

The third class of semi-active components exhibits a variable stiffness feature, which again dissipates energy. These elements are characterized by a variable free length of a spring, which is changed deploying minimal control power (Figure 13.3). For this hydropneumatic spring, if the valve is shut, only one volume is connected and the spring is stiff. When the valve is open, both volumes are connected and the spring is soft. During switching of the valve opening, the pressure in the chambers is equalized and the accumulated energy is dissipated.

© 2002 by CRC Press LLC

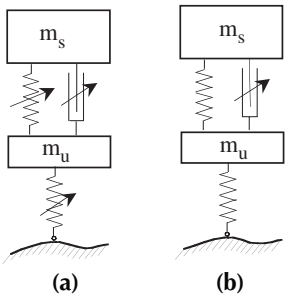


FIGURE 13.4 Semi-active vehicle suspension, (a) theoretical possibility (b) current practice.

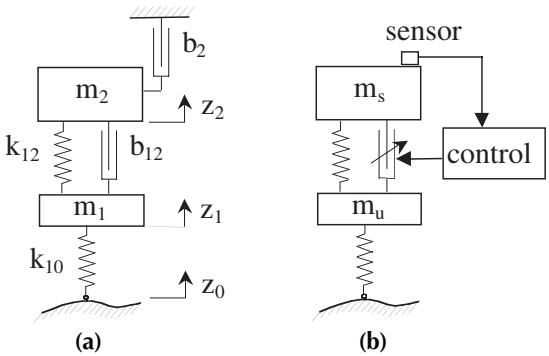


FIGURE 13.5 Sky-hook — ideal concept (a) and realization (b).

Using these semi-active devices the properties of vehicle suspension can be controlled according to the scheme on Figure 13.4 where m_s represents sprung mass and m_u unsprung mass. Theoretically, all elements of vehicle suspension can be adjusted (Figure 13.4a), but generally the semi-active damper feature for the shock absorber is controlled (Figure 13.4b).

13.1.2 Sky-Hook for Comfort

The initial concept the control of both semi-active and active suspensions originates again in Karnopp's work[1] and was developed by many other authors (bibliographic references are in Sharp and Crolla[4] and Elbeheiry[5]). The initial aim of controlled vehicle suspension is driver (passenger) comfort. This performance index is equivalent to the minimization of sprung mass acceleration (or its filtered form) with respect to the inertial space.[38] This proposition yields Karnopp's idea of a sky-hook. Sky-hook is a fictitious damper between the sprung mass and the inertial frame (fixed in the sky) (Figure 13.5). The damping force of this fictitious damper reduces the sprung mass vibration.

Further design considerations are based on the simple quarter-car model (Figure 13.6). Despite its simplicity, it covers the basic properties of suspension dynamics of a real vehicle. To introduce the control concept a linear quarter-car model is used. The nonlinearities of the structure are also taken into account in the text. The equations of motion of the quarter car model in Figure 13.6 are

$$\begin{aligned} m_1 \ddot{z}_1 + k_{10}(z_1 - z_0) + b_{10}(\dot{z}_1 - \dot{z}_0) - k_{12}(z_2 - z_1) + F_d = 0 \\ m_2 \ddot{z}_2 + k_{12}(z_2 - z_1) - F_d = 0 \end{aligned} \tag{13.1}$$

where m_1 is the unsprung mass, m_2 is the sprung mass, k_{12} is the stiffness of the main spring, k_{10} is the stiffness of the tire, b_{10} is the tire damping constant (usually negligible), and F_d is the force of the passive or semi-active damper or active element.

© 2002 by CRC Press LLC

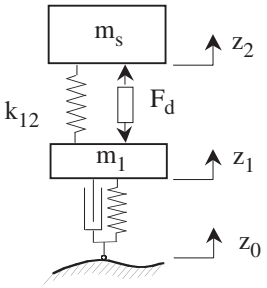


FIGURE 13.6 Quarter car model.

Figure 13.5a represents a fictitious case. Added parallel damper b_2 tries to complement b_{12} for the sprung mass. Figure 13.5b represents the realization of this concept. The fictitious force computed from the added sky-hook damper is applied by the actuator F_d. The actuator can be a fully active element (active force generator) or a semi-active element (variable shock absorber). The ideal active force F_d of this element according to the sky-hook control law is

$$F_d = -b_2 \dot{z}_2 \tag{13.2}$$

If this force is directly applied by an active force generator, then the process becomes the active sky-hook suspension. For semi-active suspension, this force is limited to the range of forces applicable by the semi-active devices. For a semi-active damper, this is done by the transformation from the required (active) force F_{act} to certain settings of the damping rate $b_{semi\text{-}active}$ such that the damping force is nearest to the desired value. For an ideal linear variable shock absorber, the damping rate $b_{semi\text{-}active}$ is set for the interval (b_{min}, b_{max}) as a linear saturation function:

$$b_{semi-active} = \begin{array}{lll} b_{max} & if & b_{max} < b_{act} \\ b_{act} & if & b_{min} < b_{act} < b_{max} \\ b_{min} & if & b_{act} < b_{min} \end{array} \tag{13.3}$$

$$b_{act} = \frac{F_{act}}{\left(\dot{z}_2 - \dot{z}_1\right)} \tag{13.4}$$

The semi-active sky-hook suspension is then realized by taking $F_{act} = F_d$ from (13.2) into (13.3) and (13.4) into (13.1). A real variable shock absorber is, however, nonlinear and time dependent with internal dynamics and the transformation (13.3)–(13.4) must reflect that.

The other problem is the usage of suitable sensors. The direct measurement of sprung mass velocity is usually not possible. Therefore, the acceleration sensor is used and the velocity is obtained by time integration after suitable filtering. Another new concept is the usage of acceleration feedback instead of velocity.[6]

The response of a nonlinear quarter car model with sky-hook control is given in Figure 13.7, for which a realistic model of the nonlinear damper is used and its response to chirp signal is observed. The responses of nonlinear semi-active damper and passive cases are compared.

13.1.3 Extended Ground-Hook for Road-Tire Forces

The road-tire forces are considered next as another performance index. Similar principle to the sky-hook called ground-hook is developed for this case.[7,8] The road-tire forces are proportional to

© 2002 by CRC Press LLC

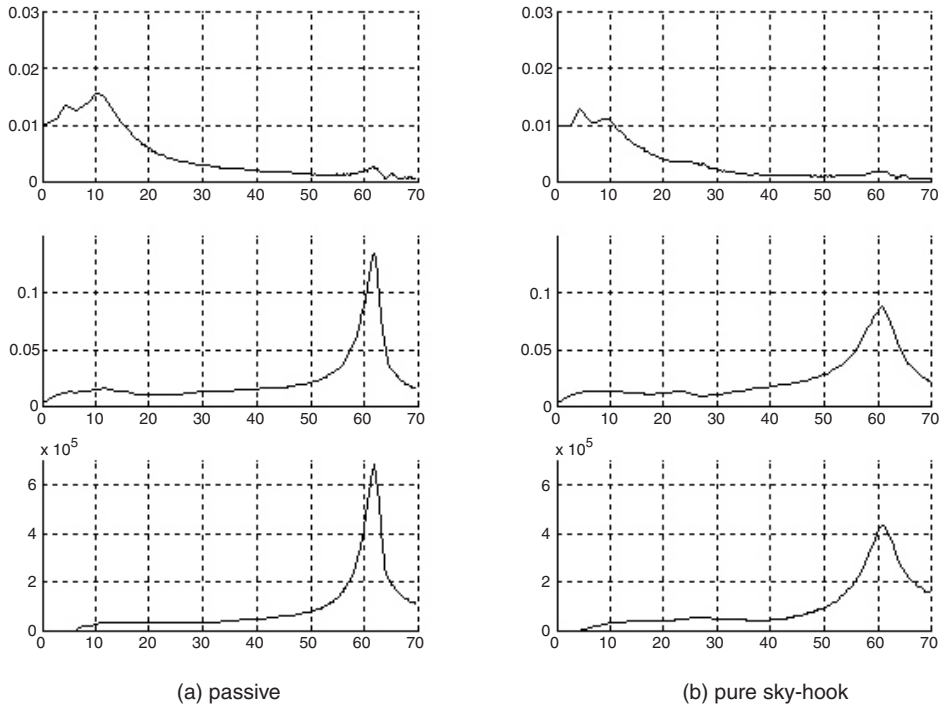


FIGURE 13.7 Comparison of response of passive suspension and pure sky-hook to the chirp signal in rad/s (figures of sprung mass, unsprung mass responses, road-tire forces).

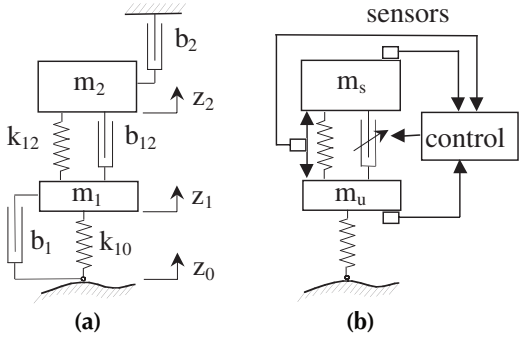


FIGURE 13.8 Extended ground-hook — ideal concept (a) and realization (b)

the tire deflection. Any reduction of this deflection by increased damping also reduces the road-tire forces. This leads to the principle of ground-hook, which is depicted in Figure 13.8.

Figure 13.8a represents a fictitious case. The real damping of the tire b_{10} has a very low value, and the parallel fictitious damper b_1 tries to add a higher damping value b_1 to it. Low accelerations of the sprung mass are obtained by the combination of a sky-hook and the ground-hook. Figure 13.8b represents the realization of this concept. The control force computed for this combination is applied by an actuator F_d. This actuator can be a fully active (active force generator) or a semi-active element (variable shock absorber). This concept was further developed into extended ground-hook. The ideal active force F_d of this element is

$$F_d = b_1\left(\dot{z}_1 - \dot{z}_0\right) - b_2\dot{z}_2 - b_{12}\left(\dot{z}_2 - \dot{z}_1\right) + \Delta k_{10}\left(z_1 - z_0\right) - \Delta k_{12}\left(z_2 - z_1\right) \tag{13.5}$$

© 2002 by CRC Press LLC

where Δk_{10} and Δk_{12} are additional terms for the fictitious stiffness cancellation, which is mainly important in the case of a fully active actuator. The direct application of the damping force (13.5) to the suspension in (13.1) gives the active extended ground-hook. The semi-active extended ground-hook suspension is then realized by combining $F_{act} = F_d$ from (13.5) into (13.3) and (13.4) and (13.1).

The advantage of the Equation (13.5) over standard [LQR] feedback is that all the terms are either directly measurable or reconstructable from other measurements. The sensors usually are the accelerometers on sprung and unsprung masses and the suspension displacement sensor. The value of $z_2 - z_1$ can be measured directly. The velocities $\dot{z}_1$ and $\dot{z}_2$ are observable from the measurements of accelerations, $\ddot{z}_1$ and $\ddot{z}_2$. By adding the Equations (13.1) the road-tire force is obtained as

$$F_{10} = m_1 \ddot{z}_1 + m_2 \ddot{z}_2 \approx k_{10}\left(z_1 - z_0\right) \tag{13.6}$$

If the tire damping b_1 is ignored, the tire deformation $z_1 - z_0$ can be solved from Equation (13.6).

By changing the parameters b_1, b_2, b_{12}, Δk_{10}, and Δk_{12} a variety of modified control laws for the suspension system can be obtained. For the systematic determination of these parameters, the multi-objective parameter optimization (MOPO) or nonlinear quadratic regulations (NQR) approaches are presented below.

The parameters of the extended ground-hook were originally considered to be constants for the entire velocity interval of the shock absorber. Because the characteristics of the shock absorber are nonlinear, the corresponding extended ground-hook control with state-dependent gains (gain scheduling) is used. The strong nonlinearity of the variable shock absorber (see Figure 13.9), especially its asymmetry, can be taken into account to determine control-law parameters. Therefore, the nonlinear extended ground-hook version, which enables the state-dependent coefficients (gains) of the control law (13.5), is developed resulting in a very desirable performance.[9,10] Their dependence on the relative velocity is determined by the optimization MOPO or NQR approaches.

The response of a nonlinear quarter-car model with the pure ground-hook control is on Figure 13.10. This is obtained as a response to chirp signal. A comparison between passive (Figure 13.10a) and active cases (Figure 13.10b) can be made. Flat response of the road-tire forces especially in low frequencies is noticeable.

13.2.4 Semi-Active Actuators and Their Models

Some conceptual models of dampers and semi-active dampers exist in the literature (for example, Duym[11] and Spencer et al.[39]) as well as the physical models (e.g., Besinger et al.[12] and Botelle et al.[37]). For a realistic investigation of semi-active suspension we consider the conceptual model of controllable dampers.

These models should entail the nonlinear characteristics of the damping force as a function of the relative velocity and control current $F_{act} = fct(v_{rel}, i_{act})$, for the special semi-active damper given in Figure 13.9. Then it must also consider the damper control's dynamic response. The dynamic behavior caused by the response time of the valve adaptation, hydraulics, and compliance of damper mounting is modeled as a low-pass filter of the steering current with two different time constants (τ_{HL}, τ_{LH}) (see Figure 13.11.) There is also the concept of real damper control. The control law determines a commanded damper force F_{des}. It is transformed on the basis of actual damper velocity in the control unit into the specific commanded control current value i_{des}, which is applied to the variable damper.

13.2 Control Design Methodology

13.2.1 General Design Methodology

Under the support of the Copernicus SADTS (Semi-Active Damping of Truck Suspension and Its Influence on Driver and Road Loads) project, a new advanced methodology for the design

© 2002 by CRC Press LLC

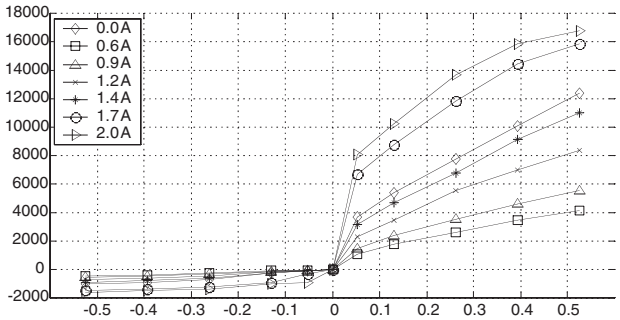


FIGURE 13.9 Semi-active damper characteristics.

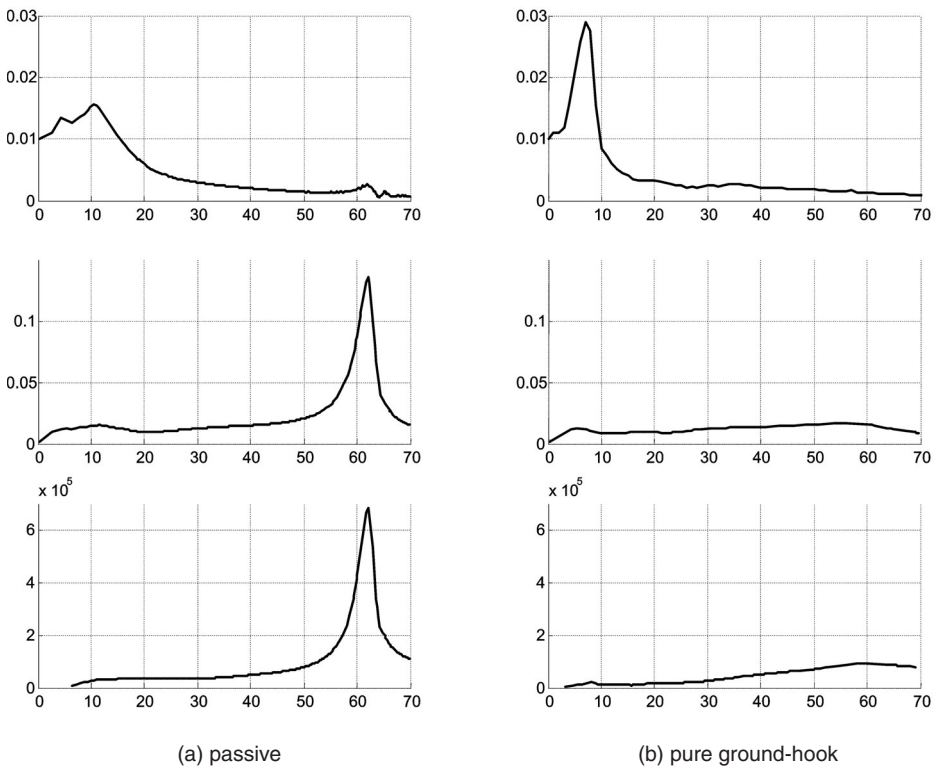


FIGURE 13.10 Comparison of response of passive suspension and pure ground-hook to the chirp signal in rad/s (figures of sprung mass, unsprung mass responses, road-tire forces).

of semi-active truck suspensions was developed.[9,13,14] It consists of a modification over earlier approaches. It also solves the problem of whether the control design should be done on a simple linear quarter-car model or on the complex fully nonlinear 3D-vehicle simulation model.

The state-of-the-art of design methodology of controlled vehicle suspensions is based on restricted design models (quarter-car or half-car models with few degrees-of-freedom, linear kinematics, linear force laws, mostly using the same models for control design and for evaluation), linear control laws (LQG design procedure and "clipped" optimal strategy), and limited experimental verification backed only by experimental parameter tuning.

© 2002 by CRC Press LLC

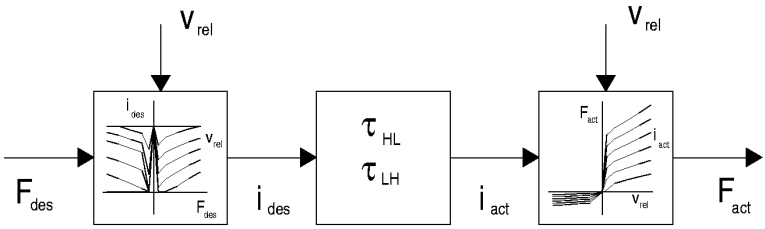


FIGURE 13.11 Control model of semi-active damper.

The design methodology developed and applied within SADTS can be described as follows:[14]

- **Models.** Multibody system modeling enables full 3D models to simplistic quarter-car models, all available in the same environment and with easy operability (as complex and with nonlinearities as required).
- **Control law design.** Without neglecting some degrees-of-freedom or essential nonlinearities, the complete nonlinear model is available within the control design environment usually by means of co-simulation interface between vehicle modeling and control design packages. Multi-objective parameter optimization (MOPO) or nonlinear quadratic regulator (NQR) approaches (described below) do not restrict the control design to oversimplified models. However, simplification (reduced-order models) are applied because:

  The design method may be restricted to linear or low-order plant models.
  The computational effort, physical insight etc. may suggest simpler models especially in the early design steps.

At any stage the use of more complex models is possible for evaluating the performance or more advanced design strategies. If the design methods allow, design-by-simulation can be performed, i.e., the use of the simulation model (in any desirable degree of complexity) within the design loop. As the simulation model is usually nonlinear, performance evaluation is possible only in the time domain.

- **Verification.** The SADTS program entails verification of the plant and also a final experimental demonstration on a controlled truck.

The main steps of the design methodology can be summarized:

1. Develop an appropriate multibody system model including all degrees-of-freedom and nonlinearities.
   a. Verify the model.
   b. Reduce the model for further design steps with respect to system order or system complexity including linearization.
2. Transfer model data into control design environment.
3. Develop the control design starting with simple models up to the advanced models. Complete the performance evaluation all along.
4. Perform multi-objective parameter optimization (MOPO): A way to achieve a fine tuning of the control system using the best (complex) model just as the engineers do with the hardware prototype.
5. Validate the dynamic structure via driving tests. Conduct trouble shooting for unexpected differences.

13.2.1.1 Design Tools

To apply the described design methodology a suitable design environment with particular software tools is necessary. There is usually a tool for modeling the vehicle as a multibody system including other components, a tool for modeling the control, optimization tool and suitable interface based

© 2002 by CRC Press LLC

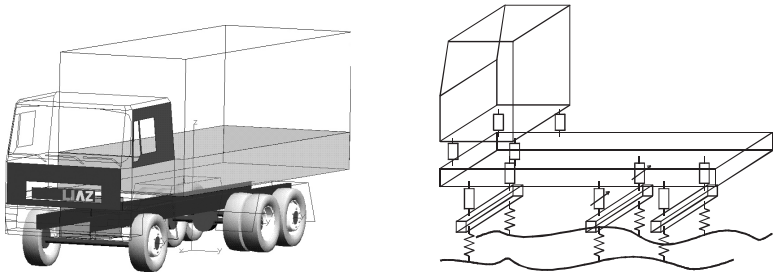


FIGURE 13.12 Truck simulation and multibody model.

on co-simulation (see Veitl et al.[15]). A brief list includes ADAMS, SIMPACK, MATLAB-SIMULINK MATRIXx-SystemBuild.

13.2.1.2 Design Models

For the control design, suitable models of the vehicle are necessary. An important result of new design methodology is that the models could be simplified (such as the quarter-car model) but they must display the main existing system's nonlinearities. Nevertheless, the final investigation and verification must be done on full 3D nonlinear simulation models.[10] Such a case is given in Figure 13.12 using a 3D model which is as close as we can recreate the real system.

13.2.2 Clipped Active Control

A systematic approach to semi-active control design is described here. It aims for the appropriate setting of $b_{semi\text{-}active}$ damping rate such that the damping force is nearest to the desired value (13.3) and (13.4). The ideal active force is computed according to the applied control design procedure (e.g., just sky-hook concept or LQR design). This force is then transformed (clipped) to the nearest realizable semi-active force. This is what we name "clipped active control."

This approach can accommodate any traditional (active) control design procedures. Probably, the most frequently used design methodology is the optimal LQR[3,16,17] for a linearized model (13.1), with a suitable cost function

$$J = \lim_{T\to\infty} \int_0^T (\mathbf{z}^T \mathbf{Q}\mathbf{z} + \mathbf{u}^T \mathbf{R}\mathbf{u}) dt \tag{13.7}$$

where $\mathbf{z} = [z_2 - z_1, \dot{z}_2, z_1 - z_0, \dot{z}_1]^T$ and $\mathbf{u} = F_d$. The semi-active control is then computed from the active force $F_{act} = F_d = \mathbf{u}$ as in Equations (13.3) and (13.4) and taking into account the limitation of the damper. Please note the selection of relative displacements as the state variables for practical reasons.

The clipped LQR control is investigated in Tseng and Hedrick.[18] It is really optimal control only for unconstrained semi-active control cases where $b_{min} = 0$ and $b_{max} = \infty$.

13.2.3 MOPO Approach

Due to the inherent nonlinearities of vehicle suspension structure, control synthesis has to be nonlinear. The traditional control methods based on linear design models cannot be used. The applicable approach for such a case is the MOPO.[19] The method is based on design-by-simulation. Control law is described in parametric form and its parameters are determined by the numerical optimization of the performance index evaluated by the simulation response of the plant to the excitations considered. Thus, by means of the MOPO approach, nonlinear models and models that

© 2002 by CRC Press LLC

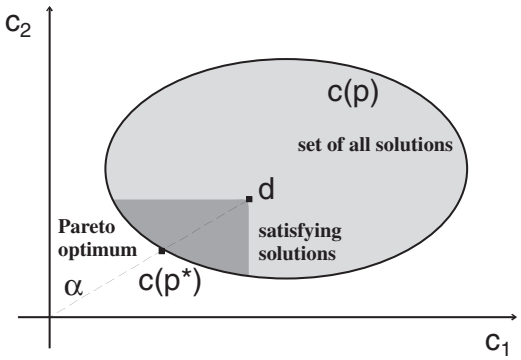


FIGURE 13.13 Pareto-optimum for two objective functions.

cannot be analytically expressed can also be treated. This approach enables not only finding parameters of nonlinear control of nonlinear plants, but also allows finding a satisfactory compromise among the performance criteria despite the possibility that they may conflict with each other. The MOPO approach is based on a search in the parameter space (Pareto optimality) by model simulation. Free system parameters and tuning parameters (e.g., control coefficients, mass properties, or installation positions) are varied within their limits until an optimal compromise is found. The parameter optimization is finished when the maximum of all weighted criteria cannot be decreased further. The result is a point on the Pareto-optimal boundary (see Figure 13.13).

For example, in the case of nonlinear extended ground-hook, the nonlinear suspension model of different complexity, the performance index of the time integral of square of the dynamic tire forces

$$I = \int_0^{T\text{large}} F_{10}^2 dt \tag{13.8}$$

and the excitation in the form of a cosine bump are used. By optimization the coefficients b_1, b_2, b_{12}, Δk_{10}, and Δk_{12} of extended ground-hook or even their dependence on relative damper velocity are determined.[9,10]

13.2.4 NQR Approach

The MOPO approach suffers from common problems of global numerical parametric optimization methods. As a remedy, a new direct control synthesis was developed.[20] It was based on recent results in nonlinear optimal control called NQR (nonlinear quadratic regulator)[21] or the SDRE (state-dependent Riccati equation).[22] The dynamics of the nonlinear system is generally described by the equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{f(x)} + \mathbf{g(x)}\,\mathbf{u} \tag{13.9}$$

where $\mathbf{x}(n \times 1)$ is the state and $\mathbf{u}(m \times 1)$ is the control and $\mathbf{f(0) = 0}$. If decomposition of the system dynamics exists

$$\mathbf{f(x)} = \mathbf{A(x)x} \tag{13.10}$$

which leads to the decomposed system

$$\frac{d\mathbf{x}}{dt} = \mathbf{A(x)x} + \mathbf{g(x)}\,\mathbf{u} \tag{13.11}$$

© 2002 by CRC Press LLC

with some properties like controllability of couple (**A(x), g(x)**) in each state position **x**, then for the quadratic performance index of the infinite horizon control problem

$$J = \int_0^\infty (\mathbf{x}^T \mathbf{Q}\, \mathbf{x} + \mathbf{u}^T \mathbf{R}\, \mathbf{u})\, dt \tag{13.12}$$

there exists the suboptimal nonlinear control

$$\mathbf{u} = -\mathbf{K}(\mathbf{x})\mathbf{x} \tag{13.13}$$

The state-dependent gain matrix **K(x)** is obtained as

$$\mathbf{K}(\mathbf{x}) = \mathbf{R}^{-1}\mathbf{g}^T(\mathbf{x})\mathbf{P}(\mathbf{x}) \tag{13.14}$$

where **P(x)** is the solution of the Riccatti equations

$$\mathbf{A}^T(\mathbf{x})\mathbf{P}(\mathbf{x}) + \mathbf{P}(\mathbf{x})\mathbf{A}(\mathbf{x}) + \mathbf{Q} - \mathbf{P}(\mathbf{x})\mathbf{g}(\mathbf{x})\mathbf{R}^{-1}\mathbf{g}^T(\mathbf{x})\mathbf{P}(\mathbf{x}) = \mathbf{0} \tag{13.15}$$

solved in each state position **x**. Recent investigations have shown that this is the only suboptimal solution for the minimization of (13.12). Nevertheless, all simulation investigations have demonstrated excellent results.

The key problem is an efficient computation of the decomposition (13.9) which is not unique. An efficient solution that needs only the evaluation of the right-hand sides of (13.9) without the necessity to manipulate symbolically the system dynamics (13.10) (e.g., differentiation or integration) is described in Valášek and Steinbauer.[21] This solution also takes into account the nonuniqueness of the decomposition. This procedure is capable of computing the decomposition of any complexity of dynamics **f**.

This theory can be extended for semi-active systems. In the case of semi-active actuation, the system Equations (13.12) are more complicated because the control is limited.

$$\mathbf{u}_{\min} \le \mathbf{u} \le \mathbf{u}_{\max} \tag{13.16}$$

The semi-active constraint (13.16) of control actuation is a highly nonlinear operation. However, the contol of such systems can be transformed into the standard NQR approach by adding artificial dynamics on the control variables $d\mathbf{u}/dt = (\mathbf{h}(\mathbf{x},\mathbf{u}) + \mathbf{w}/T$ with new artificial input variables **w**, small time constant T, and special dynamics **h** which produces the output **u** in the limited interval (13.16) for any unlimited input **w**. The reader is referred to Valášek and Kejval,[20] and Valášek and Steinbauer[21] for a detailed treatment of the topic.

A comparison of optimization results of MOPO and NQR approaches with passive suspension for the performance index of comfort and road-tire forces is in Figures 13.14 and 13.15.

13.2.5 Preview Control

Control of vehicle suspension by previewing is an attractive natural idea of usage of driver experience with proximity sensors. The principle is to use knowledge of the road profile before passing the profile by the vehicle itself. The preview principle has been investigated since the idea of active or semi-active suspension emerged.

However, the theoretical derivation of preview control or of its optimal control is substantially more difficult than control without preview.[23,24] The only investigated objective of preview control is comfort.

The problem is the introduction and reconstruction of the preview signal. In a majority of studies it is supposed that the preview signal is really the road profile in advance and the reconstruction

© 2002 by CRC Press LLC

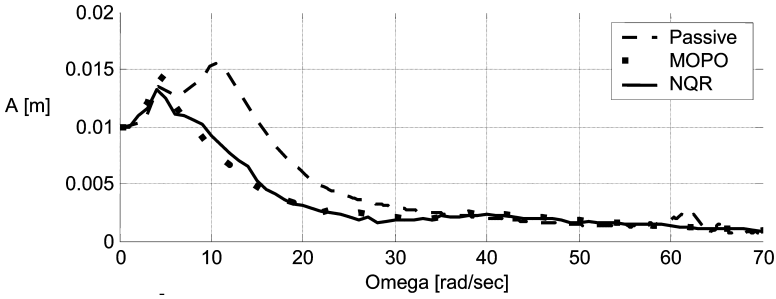


FIGURE 13.14 Comparison of optimization for comfort.

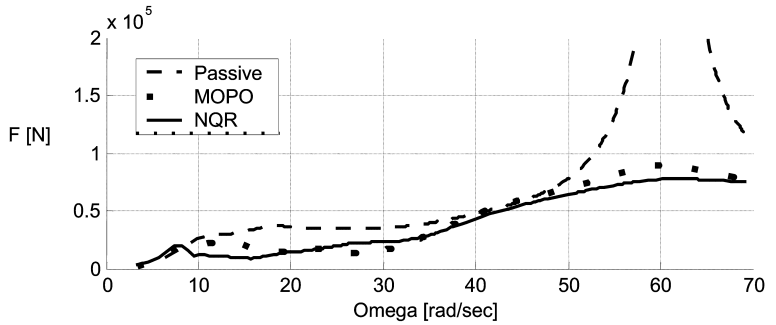


FIGURE 13.15 Comparison of optimization approaches for road-tire forces.

is not investigated at all. However, reconstruction of the road profile is rather difficult, and the treatment of feedback signal as an absolute coordinate of road profile can cause control problems, for example, because of road altitude change.

It is also supposed that if the front axle is treated as quarter-car model it could be used for measurement and reconstruction of road profile. This process could be very sensitive. Another possibility is to apply directly the signal of unsprung mass acceleration on the front axle or tire deflection on the front axle. Thus, control of all axles is connected.

13.3 Properties of Semi-Active Suspensions: Performance Indexes

There can be different performance indexes for the optimization of semi-active suspension than those already considered. Performance indexes originate from the different interactions between the vehicle and the environment. Two kinds of interaction are with the human driver and with the payload. Other interactions are with roads, with bridges, and soil for off-road vehicles. These performance indexes can be further combined with the ultimate objectives to compromise conflicting requirements. Because these optimized performance indexes improve interaction of the vehicle with the environment it speaks to friendliness toward the considered property. There are the driver-load-friendly, road-friendly, bridge-friendly, and soil-friendly suspensions.

13.3.1 Influence on Comfort

Comfort or the interaction of the driver (passenger) or payload has been investigated since the first suspension research.[1,17,25,27,36] The performance index is the root mean square (RMS) value of the weighted acceleration value according to ISO 2631.[38] The result of semi-active suspension on comfort is evident.

© 2002 by CRC Press LLC

13.3.2 Influence on Road Friendliness

Road–tire forces influenced by the semi-active suspension were investigated later.[26,27,28] Its importance recently increased with the growing demands on road transportation and the costs of road maintenance.[14,29,30] Therefore, road-friendly suspension is often spoken about.

Performance indexes besides the time integral of the square of the dynamic tire forces (13.8) can be divided into stochastic and simulated. The stochastic ones consider road damage on average, and the simulated ones try to consider the fact that road damage and maintenance are decided by one crack or hole on the road rather than average wear. The most widely used stochastic performance index is the dynamic load coefficient (DLC)[26]

$$DLC = \frac{RMS\ dynamic\ tire\ force}{static\ tire\ force} \tag{13.17}$$

and the dynamic load stress factor (DLSF)

$$DLSF = 1 + 6DLC^2 + 3DLC^4 \tag{13.18}$$

The simulated force is the aggregate force criterion. The dynamic tire forces of each axle raised to a power n and applied to each location along the road are added. The n-th power aggregate force measured at location k is

$$A_k^n = \sum_{j=1}^{N} P_{jk}^n, \qquad j = 1,2,...,N_r \tag{13.19}$$

where P_{jk} is the force from tire j to road location k, A is the number of vehicle axles and N_r is the number of locations of irregularities on the road. The power n is usually $n = 4$ representing the proportion of increased road damage due to the dynamic forces with respect to static ones. It is the well-known fourth-power law.

13.4 Examples of Practical Applications

13.4.1 Passenger Cars

Semi-active suspension for passenger car has been investigated since the first concepts. It is mostly based on the sky-hook concept and represents traditional semi-active vehicle suspension. An implementation on the passenger vehicle Nissan Cefiro in 1994 (Figure 13.16) is reported in Higashiyama.[31] The suspension is equipped with continuous shock absorbers (semi-active damper) that control all four wheels independently. The sensors are three vertical accelerometers together with the steering angle, vehicle speed, and brake sensors. Riding comfort improved by about 10% in almost the whole frequency range. Another implementation to a test passenger vehicle (Volvo) was reported in Venhovens.[25] It is based on adaptive sky-hook algorithm which takes into account the dynamic road–tire forces.

13.4.2 Road-Friendly Trucks

The first investigation of road-friendly truck suspension was performed by Yi and Hedrick[26] and followed by Valášek et al.,[9] Kortüm and Valášek,[14] and Besinger et al.[28] The recent investigation verified by prototype implementation and experimentation was done within the EU project Copernicus SADTS[9,14] and the IKA-DLR-CTU Workshop in Aachen.[32] In the first case, the prototype

© 2002 by CRC Press LLC

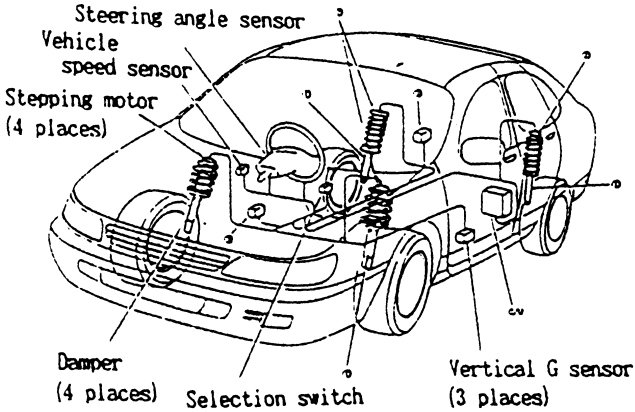


FIGURE 13.16 Semi-active suspension of passenger car. (From K. Higashiyama, T. Hirai, S. Kakizaki, and M. Hiramoto, *Proc. of AVEC,* 331–336, 1994. With permission.)

(a) **(b)**

FIGURE 13.17 Truck prototypes of SADTS Project (a) and IKA-DLR-CTU Workshop (b).

truck was a platform truck SKODA-LIAZ (Figure 13.17a), and in the second the MAN tractor with a Kögel semi-trailer (Figure 13.17b). In both cases, semi-active dampers from Mannesmann-Sachs were used on the rear-driven axle of the truck (Figure 13.9).

Each side of the rear-driven axle was equipped for control with two accelerometers (axle and chassis) and inductive displacement sensors. In addition, the road–tire forces were measured by strain gauges. Field experiments included passing the prototype truck over a good quality road such as an airport runway and then over known obstacles. The comparison between conventional, passive soft, passive hard, and controlled (semi-active) suspensions was also provided. The experimental results from the (airport) runway are shown in Figure 13.18a (evaluated as the DLSF factor) and from sinus-shaped obstacle in Figure 13.18b.

The table in Figure 13.19 summarizes the comparison of road-friendliness evaluations (advantages +/disadvantages –) for different damper settings and road excitations. Nonlinear extended ground-hook control of the damper always has advantages.

The overall evaluation of experimental results from both implementations is summarized as follows: semi-active truck suspensions to improve road friendliness demonstrate the capability of such control systems to reduce road–tire forces by about 10 to 20%, without adverse affects on riding comfort. This corresponds to a reduction of road damage up to 70% or to a possible payload increase by 1 ton for 10 tons of the design specifications. This simultaneously means a 50% reduction of truck loads fatigue.

© 2002 by CRC Press LLC

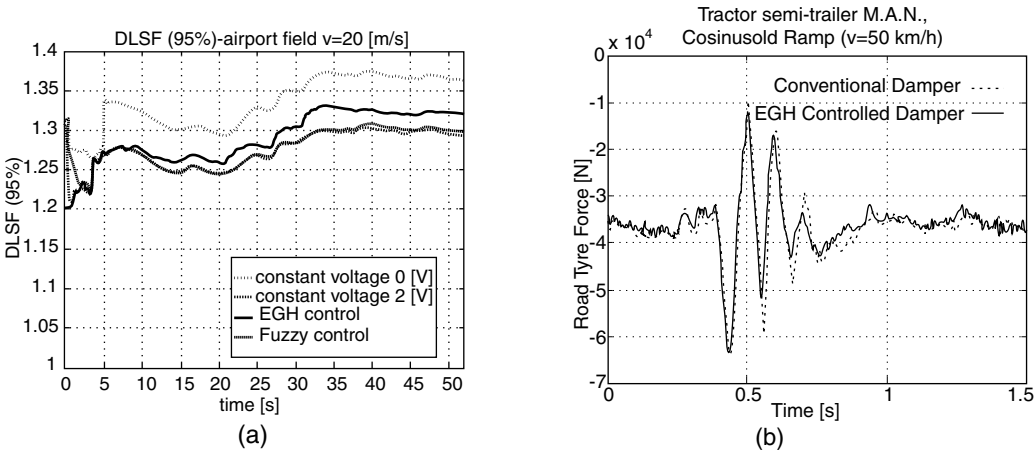


FIGURE 13.18 Experimental results (a) DLSF on stochastic road, (b) road-tire forces on sinus bump.

Damping	Big Obstacle	Small Obstacle	Stochastic Road
Conventional	+ -	+	+ -
Soft	- -	-	+ +
Hard	+ +	+	- -
Controlled	+ +	+ +	+ +

FIGURE 13.19 Road-friendliness for different damper settings.

13.4.3 Trains

Semi-active suspensions for railway vehicles have been studied theoretically and experimentally for about 25 years. Increasing operational speeds and comfort demands require focusing on both the vertical and lateral riding quality of railway vehicles. Furthermore, it seems that lighter vehicles with one or two axles per vehicle will replace classical bogie vehicles. The vehicles without classical two-axle bogies (primary train suspension) will no longer profit from so-called mechanical pre-filtering of rail irregularities, and moreover, vehicles will be lighter. Both these factors have a negative influence on ride comfort, and emphasize the necessity of more advanced suspensions.

Railway vehicles are typically equipped with two levels of suspension. The primary suspension connects the wheel sets with the bogie. Its main function is to maintain running stability and to offer curving performance as well as reducing unsprung masses. The secondary suspension is placed between the bogie and the car body. This suspension isolates the car body from rail irregularities.

Semi-active suspension is usually applied in the secondary suspension of railway vehicles, as secondary vertical, horizontal, or yaw suspensions.[33] Implementation of the primary suspension is not common; however, it is experimentally applied to wheel set yaw control. Secondary suspension can be performed by coil springs or air springs. Air springs offer the possibility of using controllable orifice damping instead of classical hydraulic semi-active dampers as performed by the pipe between the main air spring bellow and the additional volumes (semi-active air spring).

Semi-active suspensions were implemented in several vehicles in Europe and Japan for preliminary tests.[33] One of the advanced semi-active studies was performed by Siemens SGP.[34] Among other mechatronic systems, their tilting bogie SF 600 prototype was equipped with vertical and lateral semi-active suspensions in parallel to the air springs (see Figure 13.20). Limited-state feedback control is applied for the hydraulic semi-active dampers. Semi-active suspension was also implemented on an experimental vehicle. The field experiments indicated that an improvement of up to 15% in ride quality measured by RMS acceleration can be reached, which is typical for semi-active railway vehicles.

© 2002 by CRC Press LLC

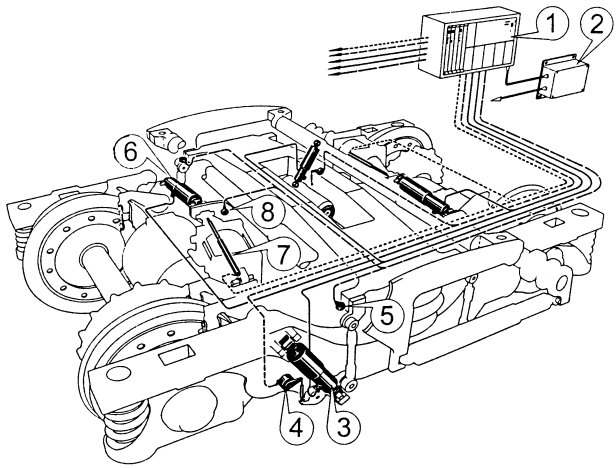


FIGURE 13.20 Prototype bogie SF 600 with semi-active damping system. 1 = control unit, 2 = inertial sensor system, 3 = semi-active vertical damper, 4 = angle-of-rotation sensor, 5 = vertical stand-alone accelerometers, 6 = semi-active lateral damper, 7 = lateral stroke sensor, and 8 = lateral stand alone accelerometer. (From A. Stribersky, A. Kienberger, G. Wagner, and H. Müller, *Vehicle System Dynamics Suppl.,* 28, 669–681, 1998. With permission.)

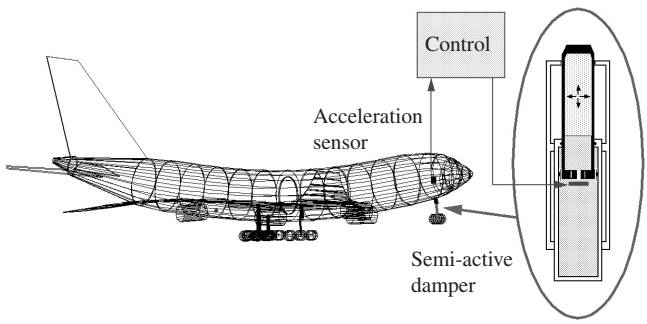


FIGURE 13.21 Semi-active landing gear.

13.4.4 Airplanes

Airplanes during ground motion in airports face many problems similar to traditional road vehicles. The shock absorber of an airplane's landing gear is designed primarily for the landing impact. The resulting dynamic characteristics can be unsatisfactory for airplane ground motion. This is especially true for modern large transport airplanes. Often, the fuselage of such an aircraft is so flexible that during ground motion the first natural modes of the structure are excited to such an extent that the resulting vibrations degrade the passengers' ride comfort. Accelerations can become so strong that the pilots can no longer read the cockpit instruments. To resolve this conflict between the landing gear design for touchdown and the design for taxiing a solution based on the semi-active suspension was developed. After the landing, the airplane's shock absorber is switched from a purely passive to a semi-active mode which is controlled by the sky-hook law from an acceleration sensor in the cockpit (Figure 13.21). The simulated results suggest significant improvements in comfort.[35]

© 2002 by CRC Press LLC

References

1. D. Karnopp, M.J. Crosby, and R.A. Harwood, Vibration control using semi-active force generators, *Journal of Engineering for Industry,* 96, 619–626, 1974.
2. D. Karnopp and G. Heess, Electronically controllable vehicle suspensions, *Vehicle System Dynamics,* 20, 207–217, 1991.
3. D. Karnopp: Active damping in road vehicle suspension systems, *Vehicle System Dynamics,* 12, 291–316, 1983.
4. R.S. Sharp and D.A. Crolla, Road vehicle suspension system design — a review, *Vehicle System Dynamics,* 16, 167–192, 1987.
5. E.M. Elbeheiry, D.C. Karnopp, M.E. Elaraby, and A.M. Abdelraaouf, Advanced ground vehicle suspension systems — a classified bibliography, *Vehicle System Dynamics,* 24, 231–258, 1995.
6. R. Holscher and Z. Huang, Das komfortorientierte semiaktive Damfungssystem, Forschritte der Fahrzeugtechnik 10, Aktive Fahrwerkstechnik, Vieweg Verlag, 1993.
7. M. Novak and M. Valášek, A new concept of semi-active control of truck's suspension, in *Proceedings of AVEC'96, Aachen,* 141–151, 1996.
8. M. Valášek, M. Novak, Z. Sika, and O. Vaculin, Extended ground-hook — new concept of semi-active control of truck's suspension, *Vehicle System Dynamics,* 27, 289–303, 1997.
9. M. Valášek, W. Kortüm, Z. Sika, L. Magdolen, and O. Vaculin.: Development of semi-active road-friendly truck suspensions, *Control Engineering Practice,* 6, 735–744, 1998.
10. M. Valášek and W. Kortüm, Nonlinear control of semi-active road-friendly truck suspension, in *Proceedings of AVEC 98, Nagoya,* 275–280, 1998.
11. S.W.R. Duym, Simulation tools, modeling and identification for an automotive shock absorber in the context of vehicle dynamics, *Vehicle System Dynamics,* 261–285, 33, 2000.
12. F.H. Besinger, D. Cebon, and D.J. Cole, Damper models for heavy vehicle ride dynamics, *Vehicle System Dynamics,* 24, 35–64, 1995.
13. M. Valášek and W. Kortüm, Development of semi-active road-friendly truck suspension, in *Proceedings of Nutzfahrzeuge mit tragenden Loesungen ins naechste Jahrhundert,* VDI, Duesseldorf, 95–104, 1999.
14. W. Kortüm and M. Valášek, (Eds.), SADTS Final Technical Report, DLR, Wessling, 1998.
15. A. Veitl, T. Gordon, A. van de Sand, M. Howell, M. Valášek, O. Vaculin, and P. Steinbauer, Methodologies for coupling simulation models and codes in mechatronic system analysis and design, *Vehicle System Dynamics Supplement,* 33, 231–243, 2000.
16. D. Margolis, The response of active and semi-active suspensions to realistic feedback signals, *Vehicle System Dynamics,* 12, 317–330, 1983.
17. D. Hrovat, D.L. Margolis, and M. Hubbard, An approach toward the optimal semi-active suspension, *Journal of Dynamic Systems, Measurement and Control,* 110, 288–296, 1988.
18. H.E.Tseng and J.K. Hedrick, Semi-active control laws — optimal and sub-optimal, *Vehicle System Dynamics,* 23, 545–569, 1994.
19. H.D. Joos, Informationstechnische Behandlung des mehrzieligen optimierungsgestutzten regelungstechnischen Entwurfs, Ph.D. thesis, Univ. Stuttgart, 1992.
20. M. Valášek and J. Kejval, New direct synthesis of nonlinear optimal control of semi-active suspensions, in *Proceedings of AVEC 2000,* Ann Arbor, 691–697, 2000.
21. M. Valášek and P. Steinbauer, Nonlinear control of multibody systems, in J. Ambrosio and W. Schiehlen (Eds.), *Proceedings of Euromech. Colloquium 404, Advances in Computational Multibody Dynamics*, Lisbon, 437–444, 1999.
22. J.R. Cloutier, State-dependent Riccati equation techniques: An overview, *Proceedings of 1997 American Control Conference,* Albuquerque, 932–936, 1997.
23. D. Hrovat, Optimal suspension performance for 2D vehicle models, *Journal of Sound and Vibration,* 146, 93–110, 1991.
24. A. Hac and I. Youn, Optimal semi-active suspension with preview based on quarter car model, *Journal of Vibration and Acoustic,* 114, 85–92, 1992.

© 2002 by CRC Press LLC

25. P.J. Venhovens, Optimal Control of Vehicle Suspensions, Ph.D. thesis, Delft University of Technology, 1993.
26. K. Yi and J.K. Hedrick, Active and Semi-Active Heavy Truck Suspensions to Reduce Pavement Damage, SAE Technical Paper 892486, 1989.
27. R. Rajamani and J.K. Hedrick, Semi-active suspensions – a comparison between theory and experiments, in G. Sauvage (Ed.), *Proceedings of 12th IAVSD Symposium,* Lyon, 504–518, 1991.
28. F.H. Besinger, D. Cebon, and D.J. Cole, An experimental investigation into the use of semi-active dampers on heavy lorries, in G. Sauvage (Ed.), *Proceedings of 12th IAVSD Symposium on the Dynamics of Vehicles on Roads and Tracks,* Lyon*,* 26–30, 1991.
29. D. Cebon, *Handbook of Vehicle-Road Interaction,* Swets and Zeitlinger, Lisse, 1999.
30. *Proceedings of DIVINE Concluding Conference*, OECD IRG, Rotterdam, 1997.
31. K. Higashiyama, T. Hirai, S. Kakizaki, and M. Hiramoto, Development of an active damper suspension, *Proceedings of AVEC,* 331–336, 1994.
32. *Proceedings of Workshop: Moderne Nutzfahrzeug-Fahrwerke, Potential zur Ladegut- und Fahrbahnshonung*, IKA RWTH Aachen, March 2000.
33. R. Goodall, Active railway suspensions: Implementation status and technological trends, *Vehicle System Dynamics,* 28, 87–117, 1997.
34. A. Stribersky, A. Kienberger, G. Wagner, and H. Müller, Design and evaluation of a semi-active damping system for rail vehicles, *Vehicle System Dynamics Suppl.,* 28, 669–681, 1998.
35. H. Wentscher, Design and Analysis of Semi-Active Landing Gears for Transport Aircraft, Ph.D. thesis, Univ. of Munich, 1995.
36. S. Ohsaku, T. Nakayama, I. Kamimura, and Y. Motozono, Nonlinear Hinf state feedback controller for semi-active controlled suspension, *Proceedings of AVEC 98, Nagoya,* 63–68, 1998.
37. E. Botelle, M. Alirand, and J. Saus, Modeling a force control actuator for semi-active car damper: Flow valve analysis, *Proceedings of AVEC 2000,* Ann Arbor, 677–684, 2000.
38. Guide for the evaluation of human exposure to whole-body vibration, ISO, 2631, 1974.
39. B.F. Spencer, S.J. Dyke, M.K. Sain, and J.D. Carlson, Phenomenological model of a magnetorheological damper, *ASCE Journal of Engineering Mechanics,* March 1996.

© 2002 by CRC Press LLC

14

Active Vibration Absorption and Delayed Feedback Tuning

14.1 Introduction

14.2 Delayed Resonator Dynamic Absorbers

The Delayed Resonator Dynamic Absorber with Acceleration Feedback • Automatic Tuning Algorithm for the Delayed Resonator Absorber • The Centrifugal Delayed Resonator Torsional Vibration Absorber

14.3 Multiple Frequency ATVA and Its Stability

Synopsis • Stability Analysis; Directional Stability Chart Method • Optimum ATVA for Wide-Band Applications

Nejat Olgac
University of Connecticut

Martin Hosek
University of Connecticut

14.1 Introduction

Vibration absorption has been a very attractive way of removing oscillations from structures under steady harmonic excitations. There are many common engineering applications yielding such undesired oscillations. Helicopter rotor vibration, unbalanced rotating power shafts, bridges under constant speed traffic can be counted as examples. We encounter numerous vibration absorption studies starting as early as the beginning of the 20th century to attenuate these vibrations (Frahm, 1911; den Hartog et al., 1928, 1930, 1938).

The fundamental premise in all of these works is to attach an additional substructure (the absorber) to the primary system in order to suppress its oscillations while it is subject to harmonic excitation with a time varying frequency. A simple answer to this effort appears as "*passive vibration absorber*" as described in most vibration textbooks (Rao, 1995; Thomson, 1988; Inman, 1994.) Figure 14.1a depicts one such configuration. The absorber section is designed such that it reacts to the excitation frequency above much more aggressively than the primary does. This makes the bigger part of the vibratory energy flow into the absorber instead of the primary system. This process complies with the literary meaning of the word 'absorption' of the excitation energy.

Based on the underlying premise there has been strong pursuit of new directions in the field of vibration absorption. A good survey paper to read in this area is (Sun et al., 1995). It covers the highlight topics with detailed discussions and the references on these topics. In this document we wish to overview the current trends in the active vibration absorption research and focus on a few highlight themes with some in-depth discussions.

© 2002 by CRC Press LLC

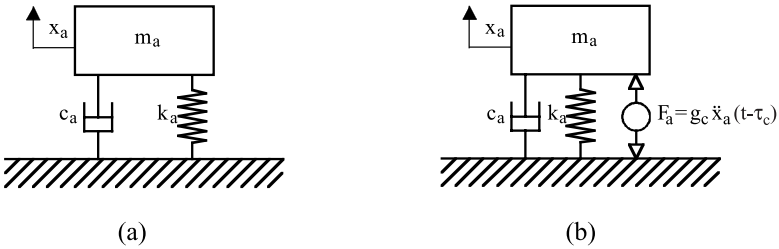


FIGURE 14.1 (a) Mass-spring-damper trio; (b) delayed resonator.

TABLE 14.1 Active Vibration Absorption Research Topics

a. Absorber tuning
1. Active
2. Semi-active
b. Mass ratio minimization
c. Spill-over phenomenon
d. Single and multiple frequency cases, wide-band absorption
e. Stability of controlled systems
f. Novel actuation means

Current research trends in vibration absorption (as displayed in Table 14.1):

- The first and most widely treated topic is the **absorber tuning**. A passive vibration absorber is known to suppress oscillations best in the vicinity of its natural frequencies. This range of effectiveness depends on the specific structural features of the absorber, and it is fixed for a given mechanical structure. Typically, the absorber is harmful, not helpful, outside the mentioned frequency range. That is, the undesired residual oscillations of the system with the absorber are larger in amplitude than those without.
- Can the tuning feature of such passive vibration absorber be improved by adding an active control to the dynamics? This question leads to the main topic of this section: **actively tuned vibration absorbers (ATVA)**. There are numerous methods for achieving active tuning. The format and the particularities of some of these active absorber-tuning methodologies will be covered in this document.

A sub-category of research under "absorber tuning" is semi-active tuning methodology, which is touched upon in two companion sections in this handbook (i.e., Jalili and Valášek). This text focuses on the active tuning methods, only.

- Mass ratio minimization. Most vibration sensitive operations are also weight conscience. Therefore, the application specialists look for minimum weight ratios between the absorber and the primary structure. (Puksand, 1975; Esmailzadeh et al., 1998; Bapat et al., 1979).
- Spill over effect constitutes another critical problem. As the TVA is tuned to suppress oscillations in a frequency interval it should not invoke some undesirable response in the neighboring frequencies. This phenomenon, known as 'spill over effect' needs to be avoided as much as possible (Ezure et al., 1994).
- Single frequency, multiple frequency, and wide-band suppression.

© 2002 by CRC Press LLC

- Stability of the active system.
- New actuators and smart materials. Primarily novel materials (such as piezoelectric and magnetostrictive) are driving the momentum in this field. (See the companion section by Wang.)

Out of these current research topics we focus on (d) and (e) (Table 14.1) in this chapter. In Section 14.2 an ATVA, the delayed resonator (DR) concept is revisited. Both the linear DR and the torsional counterpart, centrifugal delayed resonator (CDR), are considered. The latter also brings about nonlinear dynamics in the analysis. The focus of 14.3 is the multiple frequency DR (MFDR) and the wide-band vibration absorption, also the related optimization work and the stability analysis.

14.2 Delayed Resonator Dynamic Absorbers

The delayed resonator (DR) dynamic absorber is an unconventional vibration control approach which utilizes partial state feedback with time delay as a means of converting a passive mass-spring-damper system into an undamped real-time tunable dynamic absorber.

The core idea of the DR vibration control method is to reconfigure a passive single-degree-of-freedom system (mass-spring-damper trio) so that it behaves like an undamped absorber with a tunable natural frequency. A control force based on proportional partial state feedback with time delay is used to achieve this objective. The use of time delay is what makes this method unique. In contrast to the common tendency to *eliminate* delays in control systems due to their destabilizing effects (Rodellar et al., 1989; Abdel-Mooty and Roorda, 1991), the concept of the DR absorber *introduces* time delay as a tool for pole placement. Despite the vast number of studies on time delay systems available in the literature (Thowsen 1981a, 1981b and 1982; Zitek 1984), its usage for control advantage is rare and limited to stability- and robustness-related issues (Youcef-Toumi et al. 1990, 1991; Yang, 1991).

The delayed control feedback can be implemented using *position, velocity, or acceleration* measurements, depending on the type of sensor selected for a particular vibration control application at hand. In this chapter, acceleration feedback is presented as the core approach, mainly because of exceptional compactness, ruggedness, high sensitivity, and broad frequency range of piezoelectric accelerometers. All these features are essential for high-performance vibration control.

The concept of the tunable DR with absolute position feedback was introduced in Olgac and Holm-Hansen (1994) and Olgac (1995). A single-mass dual-frequency DR absorber was reported in Olgac et al., (1995, 1996) and Olgac (1996). Sacrificing the tuning capability, the single-mass dual-frequency DR absorber can eliminate oscillations at two frequencies simultaneously. As a practical modification of the DR concept, the absolute position feedback was replaced with relative position measurements (relative to the point of attachment of the absorber arrangement) in Olgac and Hosek (1997) and Olgac and Hosek (1995). Delayed acceleration feedback was proposed for high-frequency low-amplitude application in Olgac et al. (1997) and Hosek (1998). The issue of robustness against uncertainties and variations in the parameters of the absorber arrangement was addressed by automatic tuning algorithms presented in Renzulli (1996), Renzulli et al. (1999), and Hosek and Olgac (1999). The DR concept was extended to torsional vibration applications in Filipovic and Olgac (1998), where delayed *velocity* feedback was analyzed, and in Hosek (1997), Hosek et al. (1997a) and (1999a), where synthesis of the delayed control approach with a *centrifugal pendulum absorber* was presented. The concept of the DR absorber was demonstrated experimentally both for the linear and torsional cases in Olgac et al. (1995), Hosek et al. (1997b) and Filipovic and Olgac (1998).

The major contribution of the DR absorber is its ability to eliminate undesired harmonic oscillations with time-varying frequency. Other practical features include small number of operations executed in the control loop (delay and gain), simplicity of implementation (only one or at the most two variables need to be measured), complete decoupling of the control algorithm from the

© 2002 by CRC Press LLC

structural and dynamic properties of the primary system (uncertainties in the model of the primary structure do not affect the performance of the absorber provided that the combined system is stable), and fail tolerant operation (i.e., the feedback control is removed if it introduces instability and passive absorber remains).

In this section, the theoretical fundamentals of DR dynamic absorber are provided, an automatic and robust tuning algorithm is presented against uncertain variations in the mechanical properties. A topic of slightly different flavor, vibration control of rotating mechanical structures via a *centrifugal version of the DR* is also addressed.

The following terminology is used throughout the text: the *primary structure* is the original vibrating machinery alone; the *combined system* is the primary structure equipped with a dynamic absorber arrangement.

14.2.1 The Delayed Resonator Dynamic Absorber with Acceleration Feedback

The delayed feedback for the DR can be implemented in various forms: position (Olgac and Holm-Hansen 1994, Olgac and Hosek 1997), velocity (Filipovic and Olgac 1998) or acceleration (Olgac et al. 1997; Hosek 1998) measurements. The selection is based on the type of sensor that is appropriate for the practical application. In this section, the primary focus is delayed acceleration feedback especially for accelerometer's compactness, wide frequency range, and high sensitivity.

14.2.1.1 Real-Time Tunable Delayed Resonator

The basic mechanical arrangement under consideration is depicted schematically in Figure 14.1. Departing from a passive structure (mass-spring-damper) of Figure 14.1a, a control force F_a between the mass and the grounded base is added for Figure 14.1b. An acceleration feedback control with time delay is utilized in order to modify the dynamics of the passive arrangement:

$$F_a = g\ddot{x}_a(t-\tau) \tag{14.1}$$

where g and τ are the feedback gain and delay, respectively. The equation of motion for the new system and the corresponding (transcendental) characteristic equation are

$$m_a\ddot{x}_a(t) + c_a\dot{x}_a(t) + k_a x_a(t) - g\ddot{x}_a(t-\tau) = 0 \tag{14.2}$$

$$C(s) = m_a s^2 + c_a s + k_a - g s^2 e^{-\tau s} = 0 \tag{14.3}$$

Equation (14.3) possesses infinitely many characteristic roots. When the feedback gain varies from zero to infinity and the time delay is kept constant, these roots move in the complex plane along infinitely many *branches of root loci* (Olgac and Holm-Hansen 1994; Olgac et al. 1997; Hosek 1998).

To achieve pure resonance behavior, two dominant roots of the characteristic Equation (14.3) should be placed on the imaginary axis at the desired resonance frequency ω_c. Introducing this proposition, i.e., $s = \pm\omega_c i$, into Equation (14.3), the following expressions for feedback parameters are obtained*:

$$g_c = \frac{1}{\omega_c^2}\sqrt{\left(c_a\omega_c\right)^2 + \left(k_a - m_a\omega_c{}^2\right)^2} \tag{14.4}$$

*In Equation (14.5) atan2(y,x) is four quadrant arctangent of y and x, $-\pi \le \text{atan2}(y,x) \le +\pi$.

© 2002 by CRC Press LLC

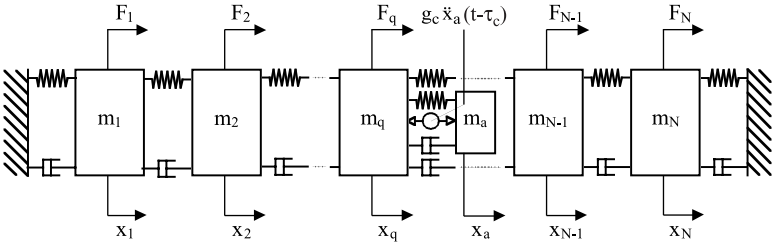


FIGURE 14.2 Schematic of MDOF structure with DR absorber.

$$\tau_c = \frac{\text{atan2}(c_a\omega_c, m_a\omega_c^2 - k_a) + 2(j_c - 1)\pi}{\omega_c}, \quad j_c = 1,2,3,\ldots * \tag{14.5}$$

By this selection of the feedback gain and delay, i.e., $g = g_c$ and $\tau = \tau_c$, the DR can be tuned to the desired frequency ω_c in real time. A complementary set of solutions which gives a negative feedback gain g_c also exists (Filipovic and Olgac 1998). However, for the sake of brevity, it is kept outside the treatment in this text.

The parameter j_c in expression (14.5) refers to the branch of the root loci which is selected to carry the resonant pair of the characteristic roots. While the control gain for a given ω_c remains the same for all branches (Equation 14.4), the values of the feedback delay (Equation 14.5) needed for operation on two consecutive branches of the root loci are related through the following expression:

$$\tau_{c,j_{c+1}} = \tau_{c,j_c} + 2\pi / \omega_c \tag{14.6}$$

The freedom in selecting higher values of j_c becomes a convenient design tool when the DR is coupled to a mechanical structure and employed as a vibration absorber. It allows the designer to relax restrictions on frequencies of operation which typically arise from stability-related issues and due to the presence of an inherent delay in the control loop (Olgac et al. 1997; Filipovic and Olgac 1998; Hosek 1998).

14.2.1.2 Vibration Control of Distributed Parameter Structures

The DR can be coupled to a mechanical structure and employed as a tuned dynamic absorber to suppress the dynamic response at the location of attachment, as depicted schematically in Figure 14.2. When the mechanical structure is subject to a harmonic force disturbance, the DR constitutes an ideal vibration absorber, provided that the control parameters are selected such that the resonance frequency of the DR and the frequency of the external disturbance coincide. The fundamental effect of the absorber is to reduce the amplitude of oscillation of the vibrating system to zero at the location where it is mounted (in this case, m_q).

It is a common engineering practice to represent distributed-parameter systems in a simplified reduced-order form, i.e., using a MDOF model. A typical representation of such a lumped-mass system is shown schematically in Figure 14.2. It consists of N discrete masses m_i which are coupled through spring and damping members and are acted on by harmonic disturbance forces $F_i = A_i \sin(\omega t + \varphi_i)$, $i = 1,2,\ldots,N$. A DR absorber is attached to the q-th mass in order to control oscillations resulting from the disturbance.

*In Equation (14.5) atan2(y,x) is a four-quadrant arctangent of y and x, $-\pi \le \text{atan2}(y,x) \le \pi$.

© 2002 by CRC Press LLC

The dynamic behavior of the primary structure is described by a linear differential equation of motion in conventional form:

$$[M]\{\ddot{x}(t)\}+[C]\{\dot{x}(t)\}+[K]\{x(t)\}=\{F(t)\} \tag{14.7}$$

where [M], [C], and [K] are $N \times N$ mass, damping and stiffness matrices, respectively, {F} is an $N \times 1$ vector of disturbance forces and {x(t)} denotes an $N \times 1$ vector of displacements. Equation (14.7) is represented in the Laplace domain as:

$$[A(s)]\{x(s)\}=\{F(s)\} \tag{14.8}$$

where:

$$[A(s)]=[M]s^2+[C]s+[K] \tag{14.9}$$

With the DR absorber on the q-th mass of the primary structure, Equation (14.9) takes the following form:

$$\left[\tilde{A}(s)\right]\{\tilde{x}(s)\}=\{\tilde{F}(s)\} \tag{14.10}$$

where:

$$\tilde{A}_{i,j}=A_{i,j},\ i,j=1,2,...,N\ \ except\ if\ \ i=j=q \tag{14.11}$$

$$\tilde{A}_{i,N+1}=0,\quad i=1,2,...,q-1\ and\ i=q+1,q+2,...,N \tag{14.12}$$

$$\tilde{A}_{N+1,i}=0,\quad i=1,2,...,q-1\ and\ i=q+1,q+2,...,N \tag{14.13}$$

$$\tilde{A}_{q,q}=A_{q,q}+c_a s+k_a \tag{14.14}$$

$$\tilde{A}_{q,N+1}=-c_a s-k_a+gs^2e^{-\tau s} \tag{14.15}$$

$$\tilde{A}_{N+1,q}=-c_a s-k_a \tag{14.16}$$

$$\tilde{A}_{N+1,N+1}=m_a s^2+c_a s+k_a-gs^2e^{-\tau s} \tag{14.17}$$

$$\tilde{F}_i=F_i,\ i=1,2,...,N \tag{14.18}$$

$$\tilde{F}_{N+1}=0 \tag{14.19}$$

$$\tilde{x}_i=x_i,\ i=1,2,...,N \tag{14.20}$$

$$\tilde{x}_{N+1}=x_a \tag{14.21}$$

© 2002 by CRC Press LLC

The coefficients $A_{i,j}$ are the corresponding elements of the matrix [A] defined in Equation (14.9). Applying Cramer's rule, the displacement of the q-th mass of the primary structure (i.e., the mass where the absorber is located) is obtained as:*

$$x_q(s) = \frac{(m_a s^2 + c_a s + k_a - g s^2 e^{-\tau s}) \det[Q(s)]}{\det[\tilde{A}(s)]} = \frac{C(s) \det[Q(s)]}{\det[\tilde{A}(s)]} \tag{14.22}$$

where:

$$Q_{i,j} = \tilde{A}_{i,j} = A_{i,j}, \; i = 1,2,...,N, \quad j = 1,2,...,N \;\; \textit{except if} \;\; j = q \tag{14.23}$$

$$Q_{i,q} = \tilde{F}_i = F_i, \quad i = 1,2,...,N \tag{14.24}$$

The factor $C(s)$ in the numerator is identical to the characteristic expression of Equation (14.3). Therefore, as long as the absorber is tuned to the frequency of disturbance, i.e., $\omega = \omega_c$, $g = g_c$, $\tau = \tau_c$, the expression for $x_q(\omega i)$ is zero. That is, provided that the denominator of Equation (14.22) possesses stable roots, the primary structure exhibits no oscillatory motion in the steady state:

$$\lim_{t \to \infty} x_q(t) = 0 \tag{14.25}$$

The frequency of disturbance, which is essential information for proper tuning of the DR absorber (see Equations 14.4 and 14.5), can be extracted from the acceleration of the absorber mass. Note that the frequency can be traced in this signal even when the primary structure has been quieted substantially by the DR absorber.

In summary, for the frequency of disturbance ω which agrees with the resonant frequency ω_c, the point of attachment of the absorber comes to quiescence. If the disturbance contains more than one frequency component, such as in the case of a square wave excitation, the delayed absorber is capable of eliminating any single frequency component selected (typically the fundamental frequency), as demonstrated in 14.2.1.6.

14.2.1.3 Stability Analysis of the Combined System

The DR absorber can track changes in the frequency of oscillation as explained above. In the meantime, the stability of the combined system should be assured for all the operating frequencies. We will see that this constraint plays a very critical role in the deployment of DR absorbers.

Stability is a critical issue in any feedback control. A system is said to have bounded-input-bounded-output (BIBO) stability if every bounded input results in a bounded output. A linear time-invariant system is BIBO stable if and only if all of the characteristic roots have negative real parts (e.g., Franklin et al. 1994).

In the following study, the objective is to explore stability properties of the combined system which comprises a multi-degree-of-freedom (MDOF) primary structure with the DR absorber, as depicted diagrammatically in Figure 14.2. It is stressed that the dynamics of the combined system is not related directly to the stability properties of the DR alone. That is, a substantially stable combined system can be achieved despite the fact that the absorber itself operates in a marginally stable mode.

*Abusing the notation slightly, $x_q(s)$ is written for the Laplace transform of $x_q(t)$.

© 2002 by CRC Press LLC

14.2.1.3.1 Characteristic Equation

As explained in 14.2.1.2, the combined system including a reduced-order (MDOF) model of the primary structure and a DR absorber (Figure 14.2) can be represented in the Laplace domain by the following system of equations:

$$\left[\tilde{A}(s)\right]\left\{\tilde{x}(s)\right\} = \left\{\tilde{F}(s)\right\} \tag{14.26}$$

The characteristic equation of the system of Equation (14.26) is identified as $\det[\tilde{A}(s)] = 0$. This determinant can be written out as:

$$[p(s) - gs^2 e^{-\tau s}]\det[P(s)] - [r(s) - gs^2 e^{-\tau s}]\det[R(s)] = 0 \tag{14.27}$$

where:

$$p(s) = m_a s^2 + c_a s + k_a \tag{14.28}$$

$$r(s) = c_a s + k_a \tag{14.29}$$

$$P_{i,j} = \tilde{A}_{i,j}, \quad i,j = 1,2,...,N \tag{14.30}$$

$$R_{i,j} = (-1)^{q+N+1}\tilde{A}_{i,j}, \; i = 1,2,...,q-1, \; j = 1,2,...,N \tag{14.31}$$

$$R_{i,j} = (-1)^{q+N+1}\tilde{A}_{i+1,j}, \; i = q,q+1,...,N, \; j = 1,2,...,N \tag{14.32}$$

For the sake of simplicity in formulation, the characteristic Equation (14.27) is manipulated into the following form:

$$CE(s) = A(s) - B(s)gs^2 e^{-\tau s} = 0 \tag{14.33}$$

where:

$$A(s) = p(s)\det[P(s)] - r(s)\det[R(s)] \tag{14.34}$$

$$B(s) = \det[P(s)] - \det[R(s)] \tag{14.35}$$

The characteristic Equation (14.33) is transcendental and possesses an infinite number of roots, all of which must have negative real parts (i.e., must stay in the left half of the complex plane) for stable behavior of the combined system. Since the number of the roots is not finite, their location must be explored without actually solving the characteristic equation. The well-known argument principle (e.g., Franklin et al. 1994) can be used for this purpose. However, this method requires repeated contour evaluations of the left hand side of the characteristic Equation (14.33) for every frequency of operation, which proves to be computationally demanding and inefficient. In the following section, an alternative method capable of revealing stability zones directly with less computational effort is explained.

14.2.1.3.2 Stability Chart Method

It can be shown that increasing control gain for a given feedback delay leads to instability of the combined system (Olgac and Holm-Hansen 1995a; Olgac et al. 1997). As a direct consequence, the following condition for stable operation of the DR absorber can be formulated: *the gain for*

© 2002 by CRC Press LLC

the absorber control should always remain smaller than the gain for which the combined system becomes unstable. The feedback gain and delay which lead to marginal stability of the combined system are to be determined from the characteristic Equation (14.33).

At the point where the root loci cross from the stable left half plane to the unstable right half plane, there are at least two characteristic roots on the imaginary axis, i.e., $s = \pm\omega_{cs} i$. Imposing this condition in Equation (14.33) yields:

$$g_{cs} = \frac{1}{\omega_{cs}^2} \left| \frac{A(\omega_{cs} i)}{B(\omega_{cs} i)} \right| \tag{14.36}$$

$$\tau_{cs} = \frac{1}{\omega_{cs}} \left[(2j_{cs} - 1)\pi + \angle \frac{A(\omega_{cs} i)}{B(\omega_{cs} i)} \right], j_{cs} = 1, 2, 3, \ldots \tag{14.37}$$

For a given $\tau_c = \tau_{cs}$ the inequality of $g_c < g_{cs}$ should be satisfied for stable operation. In order to visualize this condition, it is convenient to construct superposed parametric plots of $g_c(\omega_c)$ vs. $\tau_c(\omega_c)$ and $g_{cs}(\omega_{cs})$ vs. $\tau_{cs}(\omega_{cs})$ for the DR alone and the combined system, respectively. An example plot is shown and discussed in 14.2.1.5.

14.2.1.4 Transient Time Analysis

Once the stability of the combined system is assured, the transient behavior becomes another question of interest. It determines the time it takes the primary structure to reach a new steady state, i.e., the time needed for the absorption to take effect when any frequency change in the external disturbance occurs. The transient behavior also plays an important role in determination of the shortest allowable time between two consecutive updates of the feedback gain and delay when the absorber tunes to a different frequency. In general, the combined system must be allowed to settle before a new set of the control parameters is applied.

The settling time of the combined system is dictated by the dominant roots (i.e., the roots closest to the imaginary axis) of the characteristic Equation (14.33). Recalling that this equation has infinitely many solutions, a method is needed which determines the distance of the dominant roots from the imaginary axis, $|\alpha|$, without actually solving the equation. The argument principle can be utilized for this purpose (Olgac and Holm-Hansen 1995b; Olgac and Hosek 1997). The corresponding time constant is then obtained as the reciprocal value of $|\alpha|$, and the settling time for the combined system is estimated as four time constants:

$$t_s = 4 / |\alpha| \tag{14.38}$$

Based on the settling time analysis, the time interval is determined between two consecutive modifications of the control parameters. These modifications can take place periodically to track changes in the frequency of operation ω. The time period should always be longer than the corresponding transient response in order to allow the system to settle after the previous update of the control parameters.

14.2.1.5 Vibration Control of a 3DOF System

A three-degree-of-freedom (3DOF) primary structure with a DR absorber in the configuration of Figure 14.2 is selected as an example case. The primary structure consists of a trio of lumped masses m_i (0.6 kg each), which are connected through linear springs k_i (1.7×10^7 N/m each), damping members c_i (4.5×10^2 kg/s each) and acted on by disturbance forces F_i, i = 1, 2, 3. A DR absorber with acceleration feedback is implemented on the mass located in the middle of the system. The structural parameters of the absorber arrangement are defined as $m_a = 0.183$ kg, $k_a = 1.013 \times 10^7$ N/m, and $c_a = 62.25$ kg/s.

© 2002 by CRC Press LLC

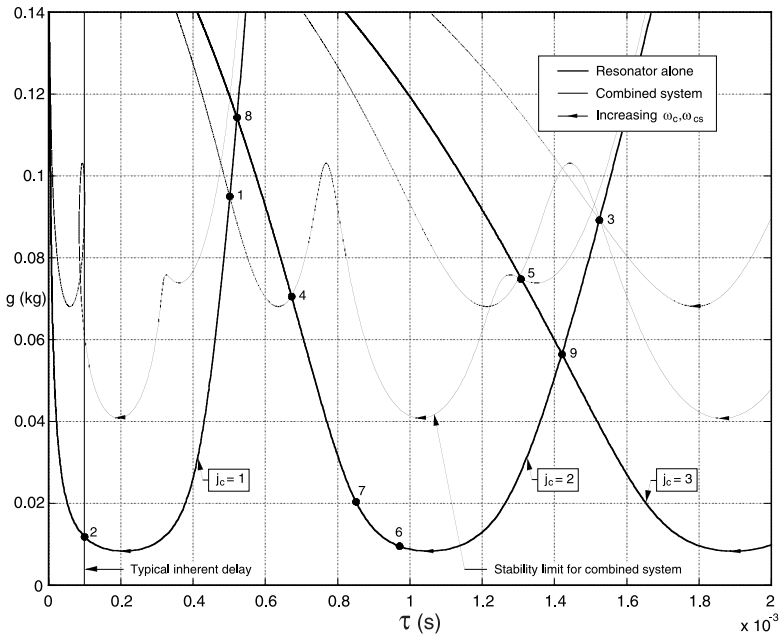


FIGURE 14.3 Plots of $g_c(\omega_c)$ vs. $\tau_c(\omega_c)$ and $g_{cs}(\omega_{cs})$ vs. $\tau_{cs}(\omega_{cs})$.

A stability chart for the example system is shown in Figure 14.3. It consists of superposed parametric plots of $g_c(\omega_c)$ vs. $\tau_c(\omega_c)$ and $g_{cs}(\omega_{cs})$ vs. $\tau_{cs}(\omega_{cs})$ constructed according to Equations (14.4), (14.5) and Equations (14.36), (14.37), respectively. As explained in 14.2.1.3, for a given $\tau_c = \tau_{cs}$ the inequality of $g_c < g_{cs}$ should be satisfied for stable operation. For operation on the first branch of the root loci, this condition is satisfied to the left of point 1. The corresponding operable range is $\tau < \tau_{cr}$ with the critical time delay $\tau_{cr} = 0.502 \times 10^{-3}$ s. In terms of frequency, the stable zone is defined as $\omega_c > \omega_{cr}$ with the lower bound at $\omega_{cr} = 962$ Hz. The upper frequency bound at point 2 results from the presence of an inherent delay in the control loop. For instance, a loop delay of 1×10^{-4} s limits the range of operation on the first branch to 1212 Hz. For the second branch of the root loci, the inequality of $g_c < g_{cs}$ is satisfied between points 3 and 4 in Figure 14.3, that is, for $0.672 \times 10^{-3}\,\text{s} < \tau < 1.524 \times 10^{-3}\,\text{s}$. The corresponding frequency range is found as $972\ \text{Hz} < \omega_c < 1{,}510$ Hz. The upper limit of operation on the third branch is represented by point 5 and corresponds to the frequency of 1530 Hz.

It is observed that operation on higher branches of the root loci introduces design flexibility which can increase operating range of the absorber and improve stability of the combined system. The stability limits can be built into the control algorithm to assure operation only in the stable range. As a preferred alternative, this scheme can be utilized to design the DR absorber with the stability limits desirably relaxed, so that the expected frequencies of disturbance remain operable.

Points 8 and 9 in Figure 14.3 indicate that there are two pairs of characteristic roots of the DR on the imaginary axis simultaneously. Therefore, the DR exhibits two distinct natural frequencies, and can suppress vibration at two frequencies at the same time. This situation is referred to as the **dual frequency fixed delayed resonator** (DFFDR) in the literature (Olgac et al. 1996; Olgac and Hosek 1995; Olgac et al. 1997). Point 8 corresponds to simultaneous operation of the absorber on the 1st and 2nd branches of the root loci. This point is unstable according to the stability chart. Point 9, on the other hand, represents a stable dual-frequency absorber created on the second and third branches of the root loci.

In order to illustrate the real-time tuning ability of the DR absorber, a simulated response of the example system to a step change in the frequency of disturbance is presented in Figure 14.4. Initially,

© 2002 by CRC Press LLC

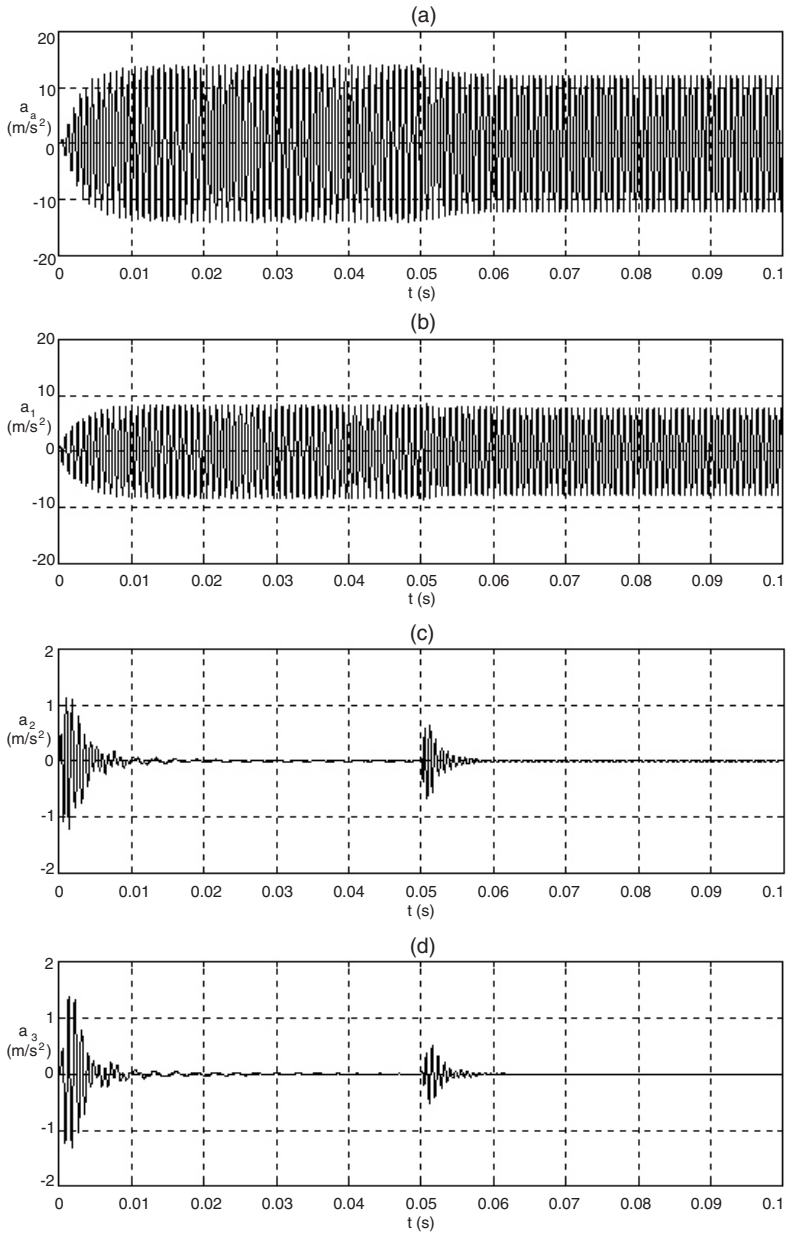


FIGURE 14.4 Simulated response to frequency change from 1200 Hz to 1250 Hz. (a) Absorber, (b) mass 1, (c) mass 2, and (d) mass 3.

a disturbance force in the form of $F_1 = A_1 \sin \omega t$ is applied to mass 1. The amplitude and frequency of disturbance are selected as $A_1 = 1$ N and $\omega = 1200$ Hz, respectively. The corresponding control parameters for the second branch of the root loci are determined as $g_c = 9.55 \times 10^{-3}$ kg and $\tau_c = 0.972 \times 10^{-3}$ s (see point 6 in Figure 14.3). After a short transient period, all undesired oscillations are substantially removed from elements 2 and 3 while mass 1, which is acted on by the disturbance force, keeps vibrating. In other words, the DR absorber creates an artificial node at mass 2, and isolates mass 3 from oscillations at mass 1. At the time t = 0.05 s, a step change in the frequency

© 2002 by CRC Press LLC

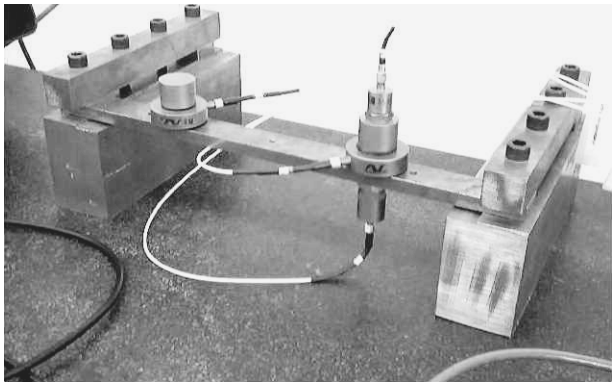


FIGURE 14.5a Experimental set-up.

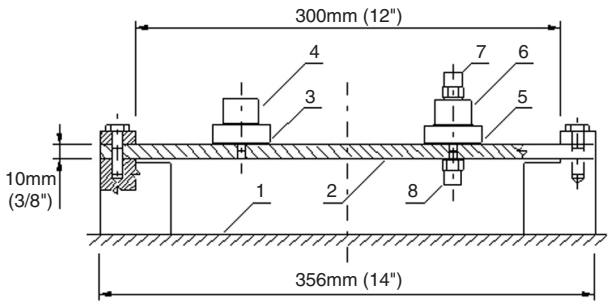


FIGURE 14.5b Side view of the test structure.

of disturbance takes place from 1200 Hz to 1250 Hz. The absorber is retuned accordingly by setting the feedback parameters to $g_c = 2.04 \times 10^{-2}$ kg and $\tau_c = 0.851 \times 10^{-3}$ s (see point 7 in Figure 14.3). After another transient period of approximately the same duration, the vibration suppression comes again into effect and elements 2 and 3 are quieted completely. In short, the DR absorber is capable of eliminating harmonic oscillations at different frequencies at the location where it is attached to the primary structure.

14.2.1.6 Vibration Control of a Flexible Beam

Implementation of the DR dynamic absorber for distributed parameter structures is illustrated by vibration control of a clamped-clamped flexible beam. The test structure is depicted in Figure 14.5a. A side view is detailed in Figure 14.5b. The setup is built on a heavy granite bed (1) which represents the ground. The primary system is selected as a steel beam (2) clamped at both ends. The dimensions of the beam are as follows (height × width × effective length): 10 mm × 25 mm × 300 mm or 3/8" × 1" × 12". A piezoelectric actuator (3) with a reaction mass (4) is mounted on the beam to generate excitation forces. The absorber arrangement comprises another piezoelectric actuator (5) with a reaction mass (6). In this particular case, the structural parameters of the absorber section are identified as $m_a = 0.183$ kg, $k_a = 9.691 \times 10^6$ N/m, and $c_a = 1.032 \times 10^2$ kg/s. The exciter and absorber actuators are located symmetrically at one quarter of the length of the beam from the center. A piezoelectric accelerometer (7) is mounted on the absorber mass (6) to provide signal for the feedback control. Another piezoelectric accelerometer (8) is attached to the beam at the base of the absorber to provide measurements for the automatic tuning algorithm (as described in 14.2) and to monitor vibration of the beam for evaluation purposes. A reduced-order lumped-parameter model of the test structure and the corresponding theoretical and experimental stability charts can

© 2002 by CRC Press LLC

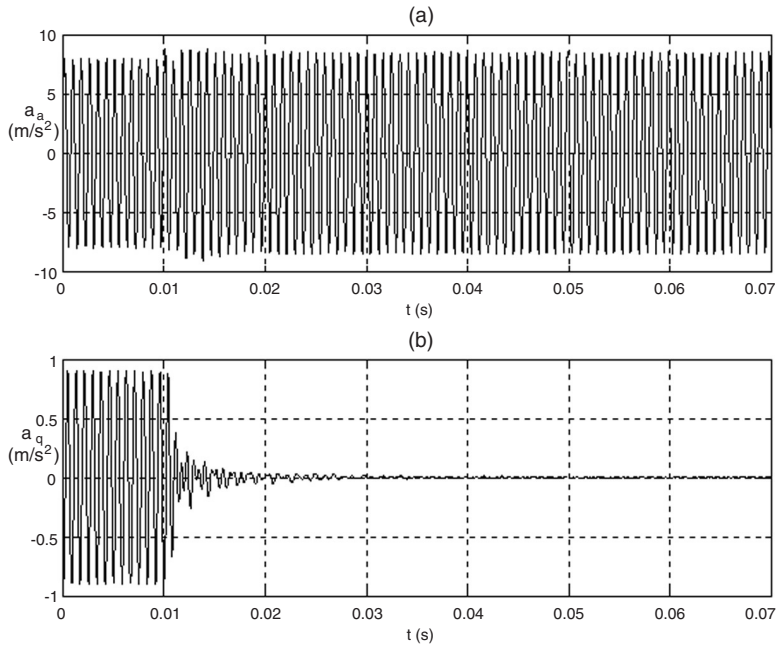


FIGURE 14.6 Time response to a harmonic disturbance at 1200 Hz, branch 2. (a) Acceleration of the absorber, (b) acceleration of the beam.

be found in Olgac et al. (1997). An alternative modal analysis approach is presented in Olgac and Jalili (1998).

In order to demonstrate the DR vibration control concept, a harmonic disturbance at 1200 Hz is applied. The corresponding feedback parameters for operation on the second branch of the root loci are found as $g_c = 1.92 \times 10^{-2}$ kg and $\tau_c = 0.939 \times 10^{-3}$ s. The corresponding time response is shown in Figure 14.6. The diagrams (a) and (b) represent plots of acceleration of the absorber mass (a_a) and acceleration of the beam at the absorber base (a_q), respectively. The control feedback is disconnected for the first 1×10^{-2} s of the test. After its activation, the amplitude of oscillation of the beam is reduced to the level of noise in the signal. The degree of vibration suppression is visualized in the DFT (discrete Fourier transformation) of the steady-state response, as depicted in Figure 14.7. The scale on the vertical axis is normalized with respect to the maximum magnitude of $a_q(\omega i)$, i.e., the ratio of $|a_q(\omega i)| / \max|a_q(\omega i)|$ expressed in percents is shown in the figure. The light line represents the DFT of the steady-state response of the beam with the control feedback disconnected. The bold line depicts the DFT when the control is active. It is observed that the oscillations of the primary structure at the point of attachment of the absorber are reduced by more than 99%.

The test is repeated with a square wave disturbance of the same fundamental frequency, i.e., 1200 Hz. The DFT of the steady-state response of the beam is depicted in Figure 14.8. Again, the ratio of $|a_q(\omega i)| / \max|a_q(\omega i)|$ expressed in percents is used on the vertical axis of the plot. The light line represents the response of the beam with the control feedback disconnected. The bold line is the response with the control active. It is observed that the dominant frequency component of 1200 Hz is suppressed by more than 99% again, while the rest of the frequency spectrum remains practically unchanged. That means no noticeable spill over effect is observed during the absorption.

The real-time tuning capability of the DR dynamic absorber is demonstrated in 14.2, where the beam is subject to a swept-frequency harmonic signal excitation from 650 to 750 Hz at the rate of 2.4 and 10 Hz/s.

© 2002 by CRC Press LLC

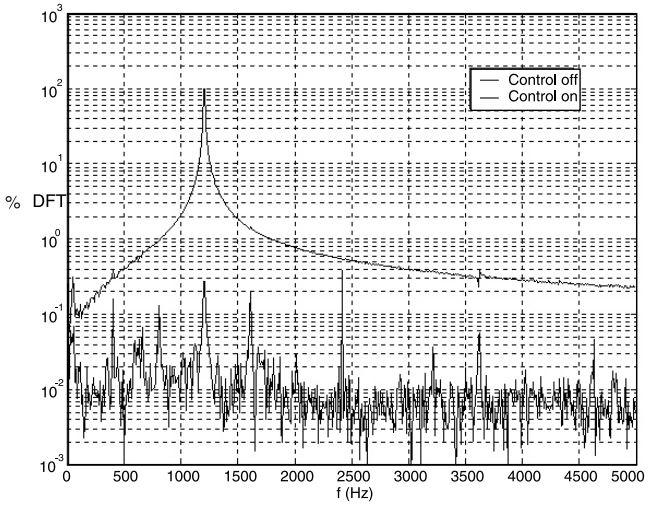


FIGURE 14.7 DFT of the beam response to a harmonic disturbance at 1200 Hz.

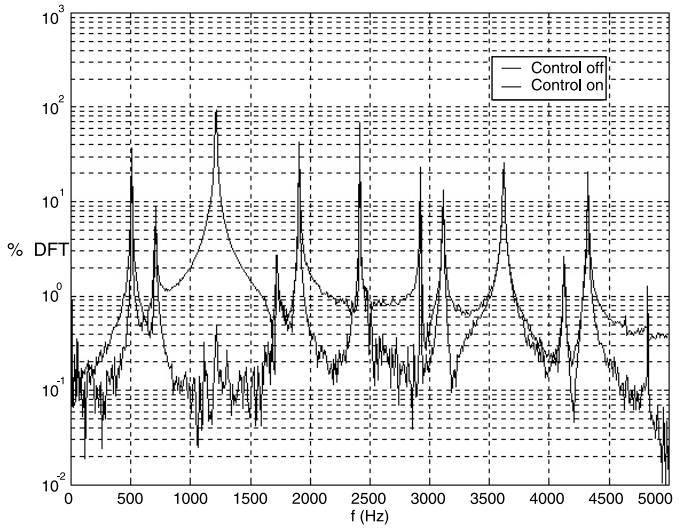


FIGURE 14.8 DFT of the beam response to a square-wave disturbance at 1200 Hz.

14.2.1.7 Summary

The delayed resonator (DR) is an active vibration control approach which utilizes partial state feedback with time delay as a means of converting a passive mass-spring-damper system into an ideal undamped real-time tunable dynamic absorber. The real-time tuning capability and complete suppression of harmonic oscillations at the point of attachment on the primary structure are not the only advantages of the DR absorber. Other practical features that can be found attractive in industrial applications are summarized below.

The frequency of disturbance can be detected conveniently from acceleration of the absorber mass. The feedback gain and delay are functions of the absorber parameters and the operating frequency only (see Equations 14.4 and 14.5). Therefore, the control is entirely decoupled from

© 2002 by CRC Press LLC

the dynamic and structural properties of the primary system. As such, it is insensitive to uncertainties and variations in the primary structure parameters, provided that the combined system remains stable.

Recalling Equation (14.5) and Figure 14.6, higher branches of the root loci can be used to tune the absorber to a given frequency ω_c. This freedom can be considered as a convenient design tool. If the feedback loop contains an inherent time delay, the designer is free to select a higher branch number and increase the required value of τ_c above the inherent delay in the loop. Proper selection of the branch of operation can also improve stability margin and transient response of the combined system (Hosek 1998).

Other practical features of the DR absorber include computational simplicity and fail-safe operation. Due to the simple structure of the feedback, a relatively small number of operations are performed within the control loop. This is particularly important in high-frequency applications where short sampling intervals are required. When the control system fails to operate and/or the feedback is disconnected, the device turns itself into a passive absorber with partial effectiveness, which is considered as a fail-safe feature.

14.2.2 Automatic Tuning Algorithm for the Delayed Resonator Absorber

Real mechanical structures tend to vary their physical properties with time. In particular, the damping and stiffness characteristics involved in their mathematical models often differ from the nominal values. As a natural consequence, insensitivity of the DR absorber performance to parameter variations and uncertainties is an essential requirement in practical applications.

Consider the combined system of a MDOF primary structure with the DR absorber as depicted in Figure 14.2. The Laplace transform of the displacement at the point of attachment of the absorber is in the form:*

$$x_q(s) = \frac{(m_a s^2 + c_a s + k_a - g s^2 e^{-\tau s}) \det[Q(s)]}{\det[\tilde{A}(s)]} \tag{14.39}$$

where the matrices $[Q(s)]$ and $[\tilde{A}(s)]$ are defined in Section 14.2.1.2. Assuming that the roots of the denominator assure stable dynamics for the combined system, the expression in the numerator must vanish for $s = \omega_c i$ in order to achieve zero steady-state response of the q-th element of the primary structure at the frequency $\omega = \omega_c$. Based on this proposition, the control parameters g and τ should be set as:

$$g_c = \frac{1}{\omega_c^2} \sqrt{(c_a \omega_c)^2 + (m_a \omega_c^2 - k_a)^2} \tag{14.40}$$

$$\tau_c = \frac{\operatorname{atan2}(c_a \omega_c, m_a \omega_c^2 - k_a) + 2(j_c - 1)\pi}{\omega_c}, \quad j_c = 1, 2, 3, \ldots \tag{14.41}$$

Equations (14.40) and (14.41) indicate that the control parameters depend on the mechanical properties of the absorber substructure and the frequency of disturbance only. That is, the performance of the DR absorber is insensitive to uncertainties in the parameters of the primary structure, as long as the combined system is stable (stability of the combined system is addressed separately in Section 14.2.1.3).

*Abusing the notation slightly, $x_q(s)$ is written for the Laplace transform of $x_q(t)$.

© 2002 by CRC Press LLC

Insensitivity to uncertainties in the parameters of the primary structure, however, does not guarantee sufficient robustness of the control algorithm. As mentioned earlier, some of the absorber parameters involved in Equations (14.40) and (14.41) are also likely to be contaminated by uncertainties. While the mass m_a can be determined quite accurately and typically does not change its value in time, the other parameters often exhibit undesirable fluctuations. The effective value of the stiffness k_a may depend, for instance, on the amplitude of oscillation of the absorber, and the damping coefficient c_a may be a function of the frequency of operation ω_c. Both of the parameters may also vary with other external factors, such as the temperature of the environment. Due to these uncertainties the actual values of the variables c_a and k_a are not available, and the control parameters g and τ can be set only according to estimated values of c_a and k_a in practice.

Two methods to improve robustness of the control algorithm against such parameter variations and uncertainties have been developed: a single-step automatic tuning algorithm based on on-line parameter identification of the absorber structural properties (Hosek 1998; Hosek and Olgac 1999), and a more general iterative approach which utilizes a gradient method for a direct search for satisfactory values of the control parameters (Renzulli 1996; Renzulli et al. 1999).

The key idea in the single-step approach is to apply control parameters based on the best estimates of the absorber properties available, evaluate the performance achieved, identify the actual mechanical properties of the absorber, calculate the corresponding control parameters, and utilize them in the feedback law. The parameter identification is achieved using the acceleration measurements taken at the absorber's mass and base. The process results in the estimates of two uncertain parameters, c_a and k_a. Details of the single-step automatic tuning algorithm can be found in Hosek (1998) and Hosek and Olgac (1999).

The more universal iterative approach (Renzulli 1996; Renzulli et al. 1998) is selected for presentation in this section. The procedure requires the initial g and τ to be in the vicinity of their actual values. Such a close starting point may be obtained by using the nominal, albeit imperfect, model of the absorber. The tuning process is accomplished through a gradient search method which iteratively converges to the desired values. The analytical formulation of the strategy is discussed first, and is then illustrated by vibration control of a flexible beam subject to swept-frequency excitation.

14.2.2.1 Iterative Automatic Tuning Algorithm

The dynamics of the DR section of the combined system in Laplace domain is given as:

$$(m_a s^2 + c_a s + k_a)x_a(s) + gs^2 e^{-\tau s} x_a(s) = (c_a s + k_a)x_q(s) \tag{14.42}$$

where $x_q(s)$ corresponds to the motion of the base of the absorber and $x_a(s)$ to the motion of the absorber proof mass. This equation can be rewritten as a transfer function between $x_q(s)$ and $x_a(s)$ as:

$$TF = \frac{x_q(s)}{x_a(s)} = \frac{m_a s^2 + c_a s + k_a + gs^2 e^{-\tau s}}{c_a s + k_a}. \tag{14.43}$$

Per Equation (14.39), $x_q(\omega i)$ should be zero if all the structural parameters are perfectly known, and g and τ are calculated as per Equations (14.40) and (14.41). When the parameters k_a and c_a vary, these control parameters must be readjusted for tuning the DR. Otherwise the point of attachment exhibits undesirable oscillations at ω. As a remedy, an adaptation law for the two control parameters, g and τ, is developed.

Before presenting the strategy, two points should be highlighted. First, the fundamental frequency, ω, is observed from the time trace of $\ddot{x}_a(t)$. Second, the ratio

$$\frac{x_q(\omega i)}{x_a(\omega i)} = \frac{\ddot{x}_q(\omega i)}{\ddot{x}_a(\omega i)} = TF(\omega i) \tag{14.44}$$

© 2002 by CRC Press LLC

can be evaluated in real time using the knowledge of ω. The resulting value of $TF(\omega i)$ is the frequency response of the system evaluated at the frequency ω. This is obtained by monitoring the accelerometer readings of $\ddot{x}_a$ and $\ddot{x}_q$, and convolving the time series of these two signals. An extended explanation of the steps involved is given in Renzulli (1996) and Renzulli et al. (1999). As demonstrated there, the convolution imposes minimal computational load when it is done progressively once at each sampling instant.

Assuming that the complex value of the transfer function TF(ωi) is known at ω, a tuning process for g and τ is presented next. Equation (14.43) can be rewritten for $s = \omega i$ as

$$TF(\omega i) = \frac{c_1(\omega i) - g\omega^2 e^{-\tau\omega i}}{c_2(\omega i)} \tag{14.45}$$

where $c_1(\omega i)$ and $c_2(\omega i)$ are complex numbers the nominal values of which are known only.

It is assumed that g and τ can be updated much faster than the speed of variations in c_a and k_a. This is a realistic assumption in most practical applications since the stiffness and damping values typically change gradually. Another assumption is that the absorber structure is capable of tuning itself to the changes in the excitation frequency ω much faster than they occur. These assumptions can be summarized as follows: rate of variations in c_a and k_a << rate of change in ω << sampling speed of g and τ. Consequently, during the robust tuning transition, c_a, k_a, and ω can be considered as constants, though unknown. A variational form of Equation (14.45) then can be written as:

$$\Delta TF(\omega i) = \frac{\partial TF}{\partial g}\Delta g + \frac{\partial TF}{\partial \tau}\Delta\tau + higher\,order\,terms\Big|_{s=\omega i} \tag{14.46}$$

where $\Delta TF(\omega i)$ is the complex variation of $TF(\omega i)$ due to the changes in g and τ (from their respective nominal values). These changes should preferably result in

$$\Delta TF(\omega i) = -TF(\omega i) \tag{14.47}$$

so that the new $TF(\omega i) + \Delta TF(\omega i) = 0$. Assuming that $TF(\omega i) \in C^{\infty}$, and Δg and $\Delta\tau$ are small, the higher order terms in Equation (14.46) can be ignored. This is a reasonable assumption since Δg and $\Delta\tau$ represent differences between the control parameters associated with the nominal values and the true values of k_a and c_a, which are expected to be close numbers. Evaluating the nominal values of the partial derivatives using Equation (14.43),

$$\left.\frac{\partial TF}{\partial g}\right|_{s=\omega i} = -\frac{\omega^2 e^{-\tau\omega i}}{c_a\omega i + k_a} \tag{14.48}$$

$$\left.\frac{\partial TF}{\partial \tau}\right|_{s=\omega i} = -\omega g i \left.\frac{\partial TF}{\partial g}\right|_{s=\omega i} = \frac{\omega^3 i g e^{-\tau\omega i}}{c_a\omega i + k_a}, \tag{14.49}$$

and substituting them in Equation (14.46), the following expression is obtained:

$$-TF(\omega i) = -\frac{\omega^2 e^{-\tau\omega i}}{c_a\omega i + k_a}\Delta g + \frac{\omega^3 i g e^{-\tau\omega i}}{c_a\omega i + k_a}\Delta\tau \tag{14.50}$$

In this equation, g and τ are known from the current control situation, and ω is detected from the zero-crossings of the $\ddot{x}_a$ signal. Though c_a and k_a are unknown, their nominal values are used

© 2002 by CRC Press LLC

(per the above discussion), and $TF(\omega i)$ is known from the complex convolution result (Renzulli 1996; Renzulli et al. 1999). The only unknowns in Equation (14.50) are Δg and $\Delta\tau$, which are solved from two algebraic equations that arise from the complex linear Equation (14.50):

$$\Delta g = \mathrm{Re}\left[TF(\omega i)\frac{c_a \omega i + k_a}{\omega^2 e^{-\tau\omega i}} \right] \tag{14.51}$$

$$\Delta\tau = -\frac{1}{g\omega}\,\mathrm{Im}\left[TF(\omega i)\frac{c_a \omega i + k_a}{\omega^2 - \tau\omega i} \right] \tag{14.52}$$

These are the increments necessary for reducing $|TF(\omega i)| = |TF(\omega i)_{old} + \Delta TF(\omega i)|$, i.e., for improving the DR absorption performance. In the next control step, $(g+\Delta g)$ and $(\tau+\Delta\tau)$ are used in place of g and τ, and the process described in Equations (14.44) to (14.52) is repeated. This leads to further reduction of $|TF(\omega i)|$ as the robust tuning evolves. The process is stopped when $|TF(\omega i)|$ falls within a desirably small value. The convergence of this process is assured if the assumptions regarding C^{∞} and the structural variation speeds hold.

Notice that this strategy requires nothing more than the two acceleration signals, i.e., acceleration of the mass and of the base of the DR. Therefore, the DR vibration absorption scheme remains free-standing. That is, the control logic (both for frequency tracking and robust tuning steps) does not require any external measurements, except those within the DR structure.

14.2.2.2 Tuning to Swept-Frequency Disturbance

The automatic tuning procedure is illustrated on vibration control of the flexible beam of Section 14.2.1.6 subject to disturbance with time-varying frequency. The test setup is shown in Figures 14.5a and 14.5b. In this particular case, the experimentally determined nominal absorber parameters are $m_a = 0.177$ kg, $c_a = 81.8$ kg/s, and $k_a = 3.49 \times 10^6$ N/m. The disturbance frequency is varied between 650 and 750 Hz at a constant rate, maintaining the amplitude fixed. The tests are carried out with sweep rates of 2.4 Hz/s and 10 Hz/s. It is logical to expect that the suppression for the swept-frequency disturbance is worse than in the fixed frequency case of Section 14.2.1.6. When the frequency sweeps, it changes before the DR attains the steady state, necessitating new values of gain and delay for perfect absorption. This settling delay of DR has a computational part (which is due to the iterations of DR autotuning) and an inertial part (due to the dynamic transients of the combined system). Therefore, it is natural that the tuning algorithm will always lag behind. Consequently, the higher the sweep rate, the worse the performance. The results of the two swept-frequency tests are shown in Figure 14.9 for a passive mode of operation, i.e., with the control feedback disconnected, and for the DR absorber with autotuning. The active vibration suppression level is 16 dB minimum for the 10 Hz/s sweep, and 32 dB minimum for the 2.4 Hz/s sweep.

14.2.3 The Centrifugal Delayed Resonator Torsional Vibration Absorber

The centrifugal delayed resonator (CDR) represents a synthesis of the delayed-feedback control strategy and a passive centrifugal pendulum absorber for vibration control of rotating mechanical structures (Hosek 1997; Hosek et al. 1997a and 1999b). The centrifugal pendulum absorber (Carter, 1929; Den Hartog, 1938; Wilson, 1968; Thomson, 1988) is an auxiliary vibratory arrangement in which the motion of the supplementary mass is controlled by a centrifugal force (Figure 14.10a). Considering its linear range of operation, the natural frequency of the centrifugal pendulum absorber is directly proportional to the angular velocity of the primary structure. Therefore, the absorber is effective when the ratio of the frequency of disturbance and the angular velocity of the primary

© 2002 by CRC Press LLC

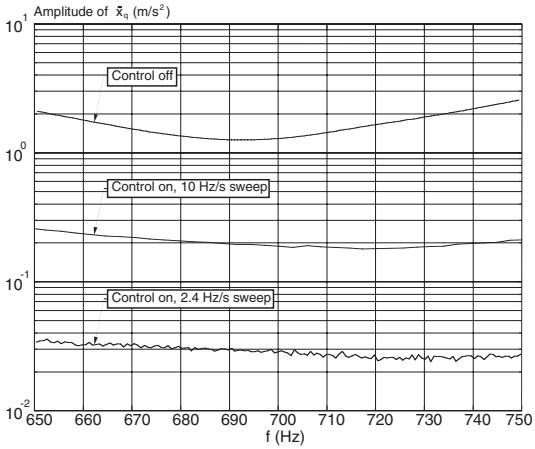


FIGURE 14.9 Beam response to swept-frequency excitation.

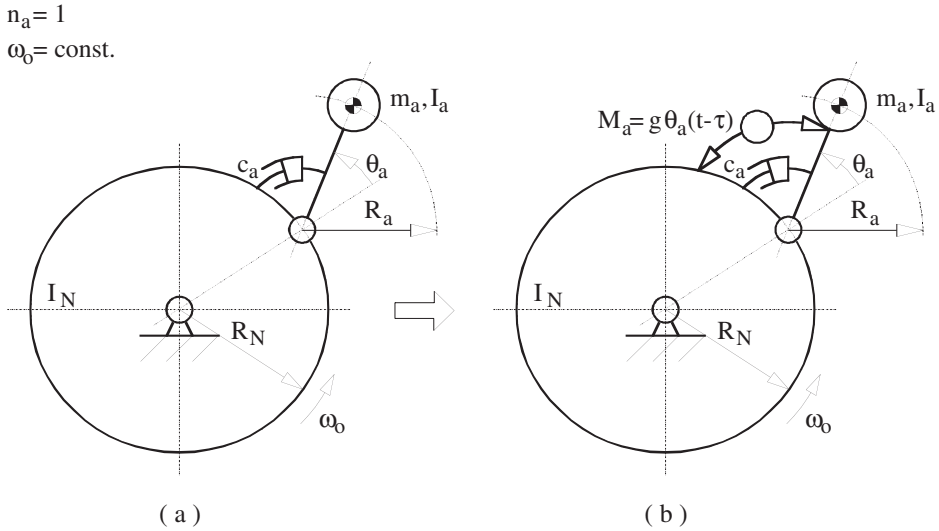


FIGURE 14.10 (a) Damped centrifugal pendulum, (b) centrifugal delayed resonator.

structure remains constant. This is the case in many applications. For instance, the fundamental frequency of the combustion-induced torques acting on the crankshaft in an internal combustion engine is a fixed multiple of the rotational velocity of the crankshaft.

In order to relax the constraint of a constant ratio of the frequency of disturbance and the angular velocity of the primary structure and/or to improve robustness against wear and tear, the CDR vibration suppression technique can be utilized. Similar to the DR vibration absorber, delayed partial state feedback is introduced to convert a damped centrifugal pendulum into an ideal frequency-tunable dynamic absorber. Introducing the real-time tuning ability feature, the CDR can improve performance of passive centrifugal pendulum absorbers in a variety of vibration problems. Typical examples can be seen in crankshaft and transmission systems of aero, automobile, and marine propulsion engines.

© 2002 by CRC Press LLC

14.2.3.1 Concept of the Centrifugal Delayed Resonator

A damped centrifugal pendulum attached to a rotating carrier is depicted schematically in Figure 14.10a. Considering small displacements θ_a and a constant angular velocity ω_o, the linearized differential equation of motion of the system of Figure 14.10a takes the following form (Hosek et al. 1997a, 1999b):

$$(I_a + m_a R_a^2)\ddot{\theta}_a + c_a\dot{\theta}_a + m_a R_N R_a \omega_0^2 \theta_a = 0 \tag{14.53}$$

The natural frequency, damping ratio and resonant (peaking) frequency of the centrifugal pendulum are found as:

$$\omega_a = \sqrt{\frac{R_N}{R_a + I_a/(m_a R_a)}}\,\omega_0 \tag{14.54}$$

$$\zeta_a = \frac{c_a}{2\omega_0\sqrt{m_a R_N R_a (I_a + m_a R_a^2)}} \tag{14.55}$$

$$\omega_p = \omega_a\sqrt{1-2\zeta_a^2} \approx \omega_a \text{ for light damping} \tag{14.56}$$

Equations (14.54) and (14.56) show that the (undamped) natural or resonant frequency of a lightly damped centrifugal pendulum is *directly proportional* to the rotational velocity ω_0. The proportionality constant $n = \omega_a/\omega_0$ is called the *order of resonance* of the passive centrifugal pendulum.

The proportionality between the natural frequency ω_a and the rotational velocity ω_0 has the following physical interpretation. The centrifugal field provides a restoring torque due to which the pendulum tends to return to a radially stretched position, i.e., it acts as a spring with an equivalent stiffness proportional to ω_0^2. Since the natural frequency ω_a is proportional to the square root of the equivalent spring stiffness, it is proportional to the angular velocity ω_0 as well.

Following the DR control philosophy (Secton 14.2.1), the core idea of the CDR concept is to reconfigure the dynamics of the damped centrifugal pendulum arrangement so that it behaves like an ideal tunable resonator. Departing from the passive arrangement in Figure 14.10a, a control torque M_a between the centrifugal pendulum and its carrier is applied in order to convert the system into a tunable resonator, as shown in Figure 14.10b. For this torque, a proportional position feedback with time delay is proposed, i.e., $M_a = g\theta_a(t-\tau)$. The new system dynamics is described by the linearized differential equation of motion (Hosek et al. 1997a, 1999b):

$$(I_a + m_a R_a^2)\ddot{\theta}_a(t) + c_a\dot{\theta}_a(t) + m_a R_N R_a \omega_0^2 \theta_a(t) + g\theta_a(t-\tau) = 0 \tag{14.57}$$

The corresponding Laplace domain representation leads to the following transcendental characteristic equation:

$$C(s) = (I_a + m_a R_a^2)s^2 + c_a s + m_a R_N R_a \omega_0^2 + ge^{-\tau s} = 0 \tag{14.58}$$

To achieve pure resonance, two dominant roots of the characteristic Equation (14.58) should be placed on the imaginary axis at the desired resonant frequency. This proposition results in the following control parameters:

$$g_c = \sqrt{(c_a\omega_c)^2 + [(I_a + m_a R_a^2)\omega_c^2 - m_a R_N R_a \omega_0^2]^2} \tag{14.59}$$

© 2002 by CRC Press LLC

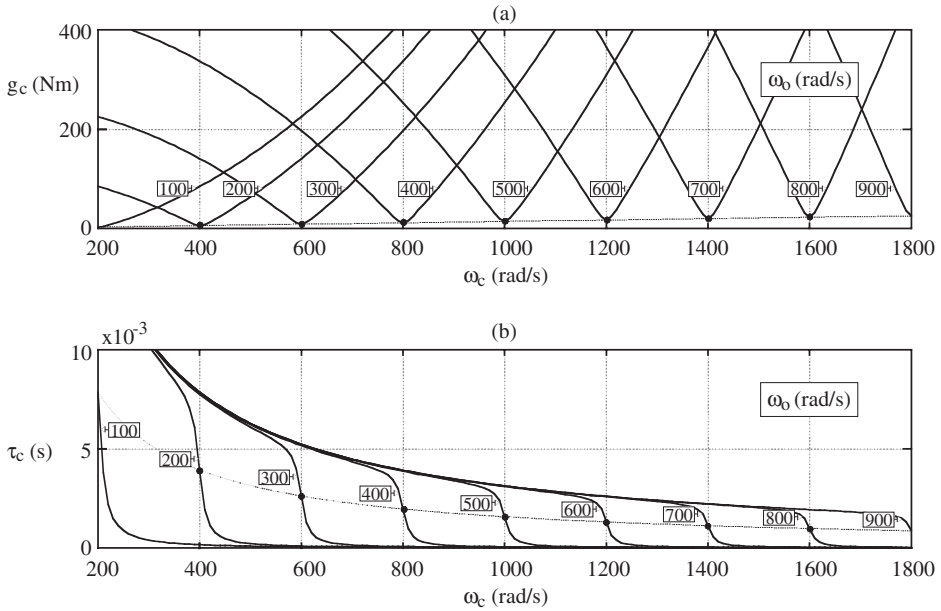


FIGURE 14.11 Feedback gain (a) and delay (b) for the CDR.

$$\tau_c = \frac{\text{atan2}[c_a\omega_c, (I_a + m_a R_a^2)\omega_c^2 - m_a R_N R_a \omega_0^2] + 2(\ell_c - 1)\pi}{\omega_c}, \quad \ell_c = 1,2,3,... \tag{14.60}$$

Similar to j_c in Equation (14.5), the parameter ℓ_c in expression (14.60) indicates the branch of the root loci which is selected to carry the resonant pair of the characteristic roots.

Example plots of the control parameters g_c and τ_c vs. the resonant frequency ω_c, as defined in Equations (14.59) and (14.60), are shown in Figure 14.11. In this particular case, the following parameters are used: $R_N = 0.15$ m, $R_a = 3.749 \times 10^{-2}$ m, $I_a = 2 \times 10^{-7}$ kgm^2, $m_a = 0.5$ kg, $c_a = 2.812 \times 10^{-5}$ kgm^2/s, $\omega_o = 500$ rad/s, and $\tau = 1.571 \times 10^{-3}$ s. The structural parameters R_a, I_a, m_a, and c_a are selected in such a way that the natural frequency (and thus, approximately, the frequency of the resonant peak) of the lightly damped centrifugal pendulum arrangement is twice the angular velocity of the carrier, i.e., $\omega_a = 2\omega_o$, see Equation (14.54). Indeed, the example structure given above possesses this property. The solid curves represent graphs of $g_c(\omega_c)$ and $\tau_c(\omega_c)$ for different values of the angular velocity ω_o in rad/s. The dashed curves correspond to the operating points where the ratio of ω_c and ω_o, i.e., the *order of resonance* for the CDR, remains fixed at $n = 2$.

Figure 14.11 shows that if the frequency ω_c fluctuates around the order of resonance $n = 2$, the CDR always operates near the minimum feedback gain g_c and the maximum sensitivity of the delay τ_c with respect to ω_c. This mode of operation is notable for low energy consumption and excellent tuning ability (Hosek 1997), both of which are desired features when the CDR is used as a tuned vibration absorber.

14.2.3.2 Vibration Control of MDOF Systems Using the CDR

When the CDR is implemented on a rotating multi-degree-of-freedom (MDOF) structure under harmonic torque disturbance, it constitutes an ideal torsional vibration absorber, provided that the control parameters are selected such that the resonant frequency of the CDR and the frequency of the external disturbance coincide.

© 2002 by CRC Press LLC

$n_a = 2$
ω_o = const.

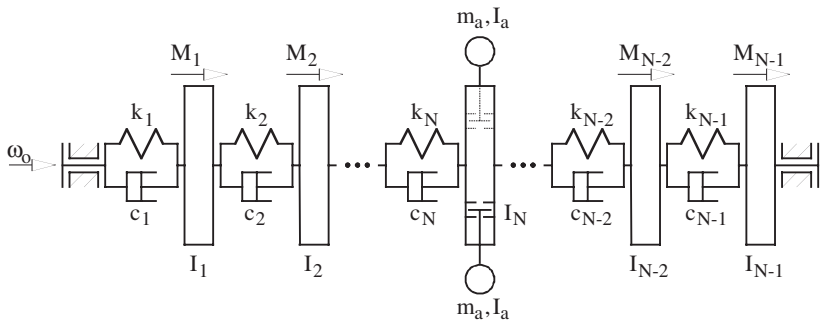


FIGURE 14.12 MDOF structure with the CDR absorber.

The combined system under consideration is depicted in Figure 14.12. A base turning at a constant angular velocity ω_o carries a MDOF primary structure that consists of N lumped disks of inertia moments I_i connected through torsional springs k_i and damping members c_i. The disks are acted on by harmonic disturbance torques $M_i = A_i \sin(\omega t + \varphi_i)$, i = 1, 2, … , N. A CDR absorber is employed at the N-th disk in order to control oscillations resulting from the external disturbance. Although the numbering scheme in this notation is selected so that the CDR is always attached to the disk number N, its implementation on any disk is practically possible, provided that the primary structure is renumbered accordingly.

The dynamic properties of the primary structure alone are represented by a linear differential equation of motion of the conventional form:

$$[I]\{\ddot{\Delta}\} + [C]\{\dot{\Delta}\} + [K]\{\Delta\} = \{M\} \tag{14.61}$$

where [I], [C], and [K] are NxN inertia, damping and stiffness matrices, respectively, $\{M\}$ is an N×1 vector of disturbance torques, and $\{\Delta\}$ represents an N×1 vector of angular differences defined as $\Delta_i = \theta_i - \theta_0$, i = 1, 2, … , N. The linear differential Equation (14.61) is represented in the Laplace domain as:

$$[A(s)]\{\Delta(s)\} = \{M(s)\} \tag{14.62}$$

where:

$$[A(s)] = [M]s^2 + [C]s + [K] \tag{14.63}$$

Considering small angular displacements of the centrifugal pendulum, Equation (14.63) can be expanded for the combined system of the primary structure with the CDR absorber as (Hosek et al. 1999b):

$$\left[\tilde{A}(s)\right]\left\{\tilde{\Delta}(s)\right\} = \left\{\tilde{M}(s)\right\} \tag{14.64}$$

where the matrix $[\tilde{A}(s)]$, augmented vectors of angular differences $\{\tilde{\Delta}(s)\}$ and disturbing torques $\{\tilde{M}(s)\}$ are defined as follows:

© 2002 by CRC Press LLC

$$\tilde{A}_{i,j} = A_{i,j},\ i,j = 1,2,...,N \ \ except \ \ i = j = N \tag{14.65}$$

$$\tilde{A}_{i,N+1} = 0, \quad i = 1,2,...N-1 \tag{14.66}$$

$$\tilde{A}_{N+1,j} = 0,\ j = 1,2,...,N-1 \tag{14.67}$$

$$\tilde{A}_{N,N} = A_{N,N} + n_a[I_a + m_a(R_a + R_N)^2]s^2 \tag{14.68}$$

$$\tilde{A}_{N,N+1} = n_a[(I_a + m_a R_a^2 + m_a R_a R_N)s^2 - c_a s - ge^{-\tau s}] \tag{14.69}$$

$$\tilde{A}_{N+1,N} = I_a + m_a(R_a^2 + R_a R_N)s^2 \tag{14.70}$$

$$\tilde{A}_{N+1,N+1} = (I_a + m_a R_a^2)s^2 + c_a s + m_a R_a R_N \omega_0^2 + ge^{-\tau s} \tag{14.71}$$

$$\tilde{M}_i = M_i,\ i = 1,2,...,N \tag{14.72}$$

$$\tilde{M}_{N+1} = 0 \tag{14.73}$$

$$\tilde{\Delta}_i = \Delta_i,\ i = 1,2,...,N \tag{14.74}$$

$$\tilde{\Delta}_{N+1} = \theta_a \tag{14.75}$$

Applying Cramer's rule, Equation (14.64) is solved for the angular displacement of the *N*-th disk of the primary structure, i.e., the disk to which the CDR is attached:

$$\Delta_N(s) = \frac{[(I_a + m_a R_a^2)s^2 + c_a s + m_a R_a R_N \omega_0^2 + ge^{-\tau s}]\det[Q(s)]}{\det\left[\tilde{A}(s)\right]} = \frac{C(s)\det[Q(s)]}{\det\left[\tilde{A}(s)\right]} \tag{14.76}$$

where:

$$Q_{i,j} = \tilde{A}_{i,j} = A_{i,j},\ i = 1,2,...,N, \quad j = 1,2,...,N-1 \tag{14.77}$$

$$Q_{i,N} = \tilde{M}_i = M_i, \quad i = 1,2,...,N \tag{14.78}$$

Similar to the conventional DR absorber (Section 14.2.1), the factor C(s) in the numerator is found to be identical to the left-hand side of Equation (14.58). Therefore, as long as the denominator possesses stable roots and the CDR is tuned to the frequency of disturbance, i.e., $\omega = \omega_c$, $g = g_c$, $\tau = \tau_c$, the expression $\Delta_N(\omega i)$ is zero and the *N*-th disk of the primary structure exhibits no oscillatory motion in the steady state:

$$\lim_{t\to\infty} \Delta_N(t) = 0 \tag{14.79}$$

In summary, for the frequency of disturbance ω which agrees with the resonant frequency ω_c, the disk of the CDR attachment is quieted completely.

© 2002 by CRC Press LLC

14.2.3.3 Stability of the Combined System

Similarly, to the conventional DR (Section 14.2.1), the range of frequencies of operation of the CDR absorber is restricted due to limitations which arise from stability related issues. Considering small angular displacements, the stability chart method of Section 14.2.1.3 can be used to assess stability of the combined system at a given angular velocity ω_o. Repeating the same stability analysis for angular velocity ω_o varying in a given range of interest, a set of stability limits can be obtained and built into the control algorithm to assure operation of the CDR in the stable zone (Hosek et al. 1997a).

In contrast to the conventional DR, two variables influence stability of the CDR absorber: the angular velocity ω_o and the frequency of disturbance ω. Any change in the angular velocity ω_o has direct influence on the stability properties of the combined system. In reality, however, the changes are smooth and relatively slow due to the inertias involved in the rotating structure. Since ω_o is monitored continuously for the CDR tuning, the stability limits can be updated periodically based on these measurements. The frequency of disturbance, on the other hand, is a property of the external disturbance. Therefore, it has no influence on the system stability until the control parameters g_c and τ_c are modified to correspond to the detected value of ω. Naturally, the controller should implement these modifications only if stable operation is expected, otherwise a passive mode is introduced by setting $g = 0$. Preferably, the stability analysis can be utilized to design the CDR absorber with desirably relaxed stability limitations so that the expected frequencies of disturbance fall into the stable zone.

Since the natural frequency of the centrifugal pendulum arrangement varies with the angular velocity of the primary structure, the overall range of operating frequencies is wider than that of the conventional DR absorber. However, full frequency range is not available at all rotational speeds (Hosek et al. 1997a).

14.2.3.4 Example Implementation

The concept of the CDR is illustrated by a simple prototype absorber. A photograph of the test structure is provided in Figure 14.13a, and a side view of the mechanical design is shown in Figure 14.13b. The main supporting component of the structure is a steel space frame (1). The primary system is represented by an aluminum disk (2) mounted on the shaft of an electric motor (3). The motor (3) is equipped with an integral tachometer to monitor the angular velocity of the shaft. The CDR absorber arrangement comprises the centrifugal pendulum (4), which is coupled pivotably to the disk (2) through an electric motor (5). A linear variable differential transformer (LVDT) (6) is mounted on the disk (2) to measure the relative displacements of the centrifugal pendulum (4). A rotating connector (7) is used to transmit control power for the electric motor (5) and to route low level signals associated with operation of the LVDT (6).

The control system for the test prototype performs two major tasks: nominal velocity control of the primary structure and delayed feedback control of the CDR absorber. The objective of the nominal velocity control is to track a desired overall profile of the angular velocity of the primary structure. This task corresponds, e.g., to a cruise control in an automobile engine application. The CDR control, on the other hand, eliminates undesired oscillations of the primary structure around its nominal velocity. These oscillations can originate, e.g., from periodic forces acting on pistons of an automobile engine. A simple harmonic signal generator is incorporated into the control system to emulate such an external disturbance.

As an example, a harmonic disturbance torque at the frequency of 12 Hz is applied to the primary structure while its nominal angular velocity is kept around 200 rpm. The degree of vibration suppression is visualized in the discrete Fourier transformation (DFT) of the steady-state response, as depicted in Figure 14.14. The scale on the vertical axis is normalized with respect to the maximum magnitude of $\Omega_1(\omega i)$, i.e., the ratio of $|\Omega_1(\omega i)| / \max|\Omega_1(\omega i)|$ is shown in the figure. The light line represents the DFT of the steady-state response of the primary structure with the control feedback disconnected. The bold line depicts the DFT when the CDR control is active. It is observed that the oscillations of the primary structure are reduced by 96%.

© 2002 by CRC Press LLC

FIGURE 14.13a Experimental set-up of CDR.

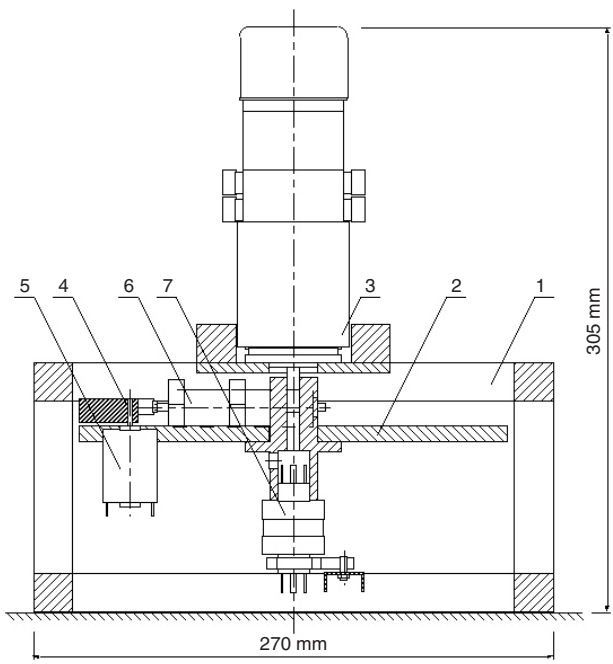


FIGURE 14.13b Side view of CDR test prototype.

© 2002 by CRC Press LLC

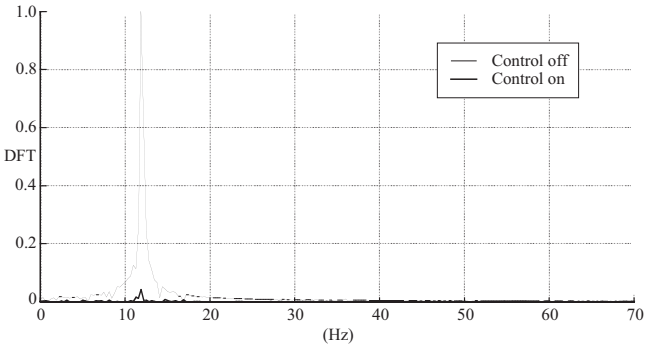


FIGURE 14.14 DFT of the steady-state response of the primary structure to a harmonic disturbance of 12 Hz at a nominal rotational velocity of 200 rpm.

14.2.3.5 Summary

The centrifugal delayed resonator (CDR) is a real-time tunable device for active suppression of torsional vibrations in rotating mechanical structures. It can be viewed as a fine tuning strategy beyond a passive centrifugal pendulum absorber. The preferred frequencies of operation are around the natural frequency of the passive centrifugal pendulum, which is *directly proportional* to the nominal angular velocity of the primary structure. In this mode of operation the CDR consumes relatively low energy and it is easy to tune. This is particularly favorable in applications where the frequency of disturbance is velocity dependent, such as in internal combustion engines, and it constitutes a key contribution of the CDR technique beyond the conventional DR (14.2.1). Similarly, to the conventional DR absorber, the frequency of disturbance can be detected by observing the displacements of the absorber relative to its carrier. Since the feedback gain and delay are functions of the absorber structural parameters and the angular velocity of the rotating base only, see Equations (14.59) and (14.60), the CDR control scheme is *entirely decoupled* from the mechanical and dynamic properties of the primary structure.

14.3 Multiple Frequency ATVA and Its Stability

14.3.1 Synopsis

The actively tuned vibration absorber contains a control, which sensitizes it at multiple frequencies. The feedback law to achieve this can be selected in variety of ways (full state feedback, linear quadratic regulator (LQR) etc.) Here we follow the unconventional feedback structure of Section 14.2. We adopt, however, a position feedback slightly differently, using multiple and unrelated time delays. This control sensitizes the absorber at a number of time-varying frequencies concurrently. It converts the absorber into a multi-frequency resonator (so-called multiple frequency delayed resonator or MFDR). The resonator, in turn, acts as a perfect absorber of vibration at these tuning frequencies when it is attached to a harmonically excited primary system.

The tuning scheme offers the same practical benefits as mentioned earlier (such as simplicity, on-line tuning ability, decoupled nature of the control from the primary). The emphasis of this section, however, is on the stability analysis of such systems with *multiple, unrelated time delays*. Such dynamics and its stability assessment are rarely treated in the literature (Thowsend, 1981b; Marshall, 1979). Thus the problem becomes mathematically very challenging.

An absorber is a dynamic structure, which offers minimum impedance between the excitation force on the primary system and the absorber mass (Sun et al., 1995). This yields relatively high amplitude of oscillations at the absorber for the excitation frequencies of concern (Figure 14.15).

© 2002 by CRC Press LLC

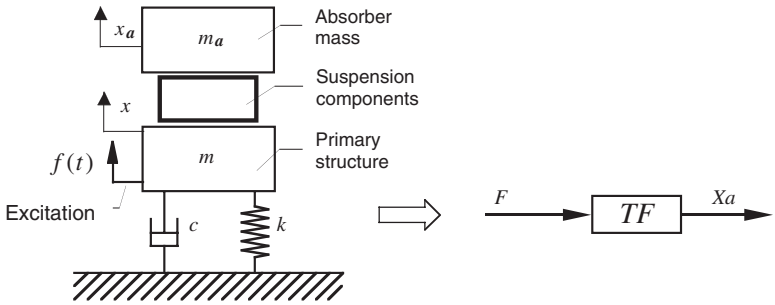


FIGURE 14.15 Typical vibration absorber set-up.

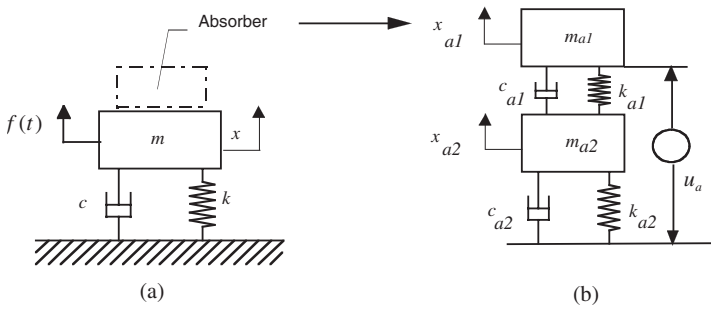


FIGURE 14.16 Primary (a) and absorber (b) structures.

The transfer function x_a/f (inverse of the impedance) should exhibit frequency response characteristics with maxima at the tuning frequencies. The aim in tuning (or 'sensitizing') is to match these frequencies with the dominant frequency contents of f. If this sensitization were assured, the amplitude x_a would be large, implying high energy flow into the absorber.

The most effective absorber is achieved by creating "a resonator," as discussed earlier, at the frequency of the excitation (Olgac et al., 1994; Olgac, 1995; Valášek, 1999). For this, the dominant poles of the absorber substructure are placed on the imaginary axis. Many researchers achieved this objective utilizing various feedback control laws from LQR (linear quadratic regulator) (Seto et al., 1991) to H_∞ (Nishimura et al., 1993) and even neural networks (DiDomenico, 1994) and the DR procedure as described earlier. Here we will expand on the latter concept, which contains a proportional position feedback with a time delay.

The theme of this section is to sensitize the absorber at multiple frequencies. Without loss of generality, we take the case of dual-frequency vibration absorption into account, although the treatment is expandable to general multiple frequency cases. The dynamic model is analyzed from the perspective of transmissibility; along with the stability of the controlled structure.

Consider the primary structure in Figure 14.16 which is subjected to time varying dual frequency excitation:

$$f(t) = f_1 \sin \omega_1 t + f_2 \sin \omega_2 t\,. \tag{14.80}$$

where ω_1 and ω_2 vary in steps as time progresses. In order to suppress undesired oscillations of the primary mass, m, a lightly damped passive absorber is suggested. Ideally, it should have two pairs of poles at $\pm i\omega_1$ and $\pm i\omega_2$. Even a small damping (i.e., c_{a1}, c_{a2}) could remove the poles from these ideal settings. Consequently, the vibration caused by the dual harmonic forcing will not be

© 2002 by CRC Press LLC

fully suppressed. A plausible two DOF absorber structure to achieve this is given in Figure 14.16b. Fine-tuning of the absorber poles requires a feedback control, which aims to

1. Make the absorber resonant at ω_1 and ω_2 by enforcing the absorber poles at $\pm i\omega_1$ and $\pm i\omega_2$. This pole placement operation is indeed the tuning of the absorber (i.e., to place the poles of the absorber structure at ω_1 and ω_2 precisely) so that the absorption is perfected at these frequencies. This process can also be considered as equivalent to eliminating the damping component at these poles.
2. Achieve tuning for time varying frequencies.

Typically, the excitation frequencies ω_1 and ω_2 are assumed to vary within a narrow range around the passive absorber's natural frequencies. Otherwise the above described pole placement procedure requires a large control authority, which is often not practicable.

Considering the system depicted in Figure 14.16, the governing equations are written as follows:

for the absorber alone (Figure 14.16b),

$$\begin{aligned} &m_{a1}\ddot{x}_{a1} + c_{a1}(\dot{x}_{a1} - \dot{x}_{a2}) + k_{a1}(x_{a1} - x_{a2}) + u_a = 0, \\ &m_{a2}\ddot{x}_{a2} + c_{a2}\dot{x}_{a2} + k_{a2}x_{a2} - c_{a1}(\dot{x}_{a1} - \dot{x}_{a2}) - k_{a1}(x_{a1} - x_{a2}) = 0; \end{aligned} \tag{14.81}$$

and for the combined system (primary structure and the absorber together),

$$\begin{aligned} &m_{a1}\ddot{x}_{a1} + c_{a1}(\dot{x}_{a1} - \dot{x}_{a2}) + k_{a1}(x_{a1} - x_{a2}) + u_a = 0, \\ &m_{a2}\ddot{x}_{a2} + c_{a2}(\dot{x}_{a2} - \dot{x}) + k_{a2}(x_{a2} - x) - c_{a1}(\dot{x}_{a1} - \dot{x}_{a2}) - k_{a1}(x_{a1} - x_{a2}) = 0, \\ &m\ddot{x} + c\dot{x} + kx - c_{a2}(\dot{x}_{a2} - \dot{x}) - k_{a2}(x_{a2} - x) - u_a = 0. \end{aligned} \tag{14.82}$$

Notice that the control feedback force, u_a, is implemented between the base of the absorber and the mass m_{a1}. This is a design choice that is not unique and does not affect the generalization of the work presented here.

Extending an earlier proposition (Valášek et al., 1999) the control for the above pole placement is chosen in the form of superimposed delayed position feedback:

$$u_a = \sum_{i=1}^{4} g_i(x_{a1}(t - \tau_i) - x(t - \tau_i)) \tag{14.83}$$

This is identical to having four retarded springs with stiffness g_i and retardation τ_i between the absorber mass m_{a1} and the absorber base. τ_i delays are preselected and the gains are periodically updated for tuning, as explained below. An important point to note is the nature of the feedback control. It is formed by the delayed relative displacements between the absorber and the primary structure. Therefore, the process is decoupled from the primary side. That is, the absorber can be tuned independently from the dynamic features of the primary structure, i.e., its mass, stiffness etc.

Laplace domain representation of Equations (14.81 and 14.82) is

for the **absorber**

$$\begin{bmatrix} m_{a1}s^2 + c_{a1}s + k_{a1} + \sum_{i=1}^{4} g_i e^{-\tau_i s} & -c_{a1}s - k_{a1} \\ -c_{a1}s - k_{a1} & m_{a2}s^2 + (c_{a2} + c_{a1})s + (k_{a2} + k_{a1}) \end{bmatrix} \cdot \begin{bmatrix} X_{a1} \\ X_{a2} \end{bmatrix} = \mathbf{A}_{\mathbf{abs}} \cdot \begin{bmatrix} X_{a1} \\ X_{a2} \end{bmatrix} = \mathbf{0} \tag{14.84}$$

© 2002 by CRC Press LLC

and for the **combined system**

$$\begin{bmatrix} m_{a1}s^2 + c_{a1}s + k_{a1} + \sum_{i=1}^{4} g_i e^{-\tau_i s} & -c_{a1}s - k_{a1} & -\sum_{i=1}^{4} g_i e^{-\tau_i s} \\ -c_{a1}s - k_{a1} & m_{a2}s^2 + (c_{a2} + c_{a1})s + (k_{a2} + k_{a1}) & -c_{a2}s - k_{a2} \\ -\sum_{i=1}^{4} g_i e^{-\tau_i s} & -c_{a2}s - k_{a2} & ms^2 + (c + c_{a2})s + (k + k_{a2}) + \sum_{i=1}^{4} g_i e^{-\tau_i s} \end{bmatrix} \cdot \begin{bmatrix} X_{a1} \\ X_{a2} \\ X \end{bmatrix} = \mathbf{A}_{\mathbf{cs}} \cdot \begin{bmatrix} X_{a1} \\ X_{a2} \\ X \end{bmatrix} = \mathbf{0} \qquad (14.85)$$

In these equations capital letters correspond to the Laplace transformed quantities of the respective lower case expressions; $\mathbf{A}_{\mathbf{cs}}$ is the system matrix for the combined system and $\mathbf{A}_{\mathbf{abs}}$ is that of the absorber.

The objective is to make this absorber resonant at two distinct frequencies at the same time. This tuned absorber is named dual frequency delayed resonator (DFDR). The two pairs of characteristic roots of the absorber substructure are placed at $\mp\omega_1 i$ and $\mp\omega_2 i$, by imposing these as roots to the respective characteristic equations:

$$CE_{abs} = \det \mathbf{A}_{\mathbf{abs}}\big|_{s=\omega i} = 0 \qquad \omega = \omega_1 \text{ and } \omega = \omega_2 \qquad (14.86)$$

Equation (14.86) is complex-valued and forms a set of four simultaneous linear equations in g_i's:

$$\begin{aligned} &\mathrm{Re}[CE_{\mathbf{abs}}] = 0 \text{ for } s = \omega_1 i \text{ and } s = \omega_2 i \\ &\mathrm{Im}[CE_{\mathbf{abs}}] = 0 \text{ for } s = \omega_1 i \text{ and } s = \omega_2 i \end{aligned} \qquad (14.87)$$

which can be solved readily. The solutions $\bar{g} = \{g_i\}$ and the selected τ_is are the feedback parameters. Notice that Equation (14.86) is transcendental, thus it invites many other finite roots as well as the four of interest. Desirably, but not necessarily, these additional roots should be in the stable left half plane. The stability of the combined system, therefore, needs to be studied for the effective suppression of vibration. Interestingly the stability of the absorber subsection is not required as long as the combined system is guaranteed to be stable. Details of this argument can be found in (Olgac et al., 1997) for single frequency DR cases.

When the absorber is tuned, the transfer function between the primary and absorber masses should exhibit an impedance picture with two zeros, such as Figure 4.17. In this figure $|x/x_a|$ refers to the frequency ratio of the two amplitudes, i.e., $|x(s=\omega i)/x_a(s=\omega i)|$. In order to suppress the oscillation of the primary at a given frequency, $|x/x_a|_{s=\omega i}$ should be very close to zero. This ratio, indeed, is the ratio of the two respective impedances, i.e., the impedance at the primary side, F/x, should be large, while

$$\frac{F/x_a}{F/x} = \frac{\text{impedance for } F \to x_a \text{ transition}}{\text{impedance for } F \to x \text{ transition}}$$

© 2002 by CRC Press LLC

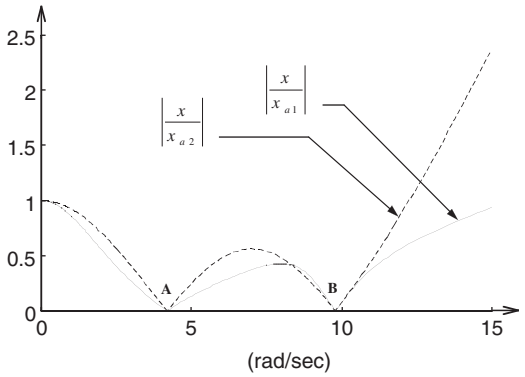


FIGURE 14.17 Primary/absorber amplitude ratios for a typical case of tuned absorber with dual resonance frequencies (at A and B).

is comparatively much smaller. This implies that the primary substructure passes the vibratory energy in full to the absorber section at the particular frequency of resonance. This is the core phenomenon yielding the absorber sensitization and effective vibration absorption.

Obviously, the absorber is sensitized to suppress oscillations at the two frequencies ω_1 and ω_2. This sensitization can be readjusted in real time for a different set of frequencies. The process requires simply recalculating the g_i's as described earlier. The two frequencies, ω_1 and ω_2, need to be detected as they vary in time. One way of achieving this is to check the time traces of the absorber motion and extract its frequency contents. This calls for a "shifting window" type FFT, and it is also possible to do in real-time (Orfanidis, 1996). In order to allow the system, however, to complete the transient regime prior to the adjustment in $\{g_i\}$, this process is repeated at a slower pace than the settling period of the combined system.

An outstanding operational issue in this active vibration suppression process is the on-line assessment of the stability. Notice that the system at hand (Equations 14.84 and 14.85) has transcendental characteristic equations with four unrelated time delays. There is very limited knowledge available in the dynamic systems literature on such systems containing unrelated delays (Huang and Olgac, 2000; Stepan, 1989). An interesting treatment can be found in (Hertz et al., 1984) for commensurate delays, i.e., delays that are integer multiples of a core delay. Recently, a systematic methodology was suggested by the group of the authors (Huang and Olgac, 2000) to address this issue. The method forms the highlight of the present work. Before going into the stability analysis we wish to touch upon the selection of τ_i's. A rudimentary selection method is used for the time delays τ_i. We take a relatively even distribution of $\{\tau_i\}$, such that even the largest τ_i would not cause aliasing in the feedback. That is, the largest time delay is much less than the smallest period of the excitation on hand. The case study example follows this logic.

14.3.1.2 Stability Analysis; Directional Stability Chart Method

The combined system as given in Equation (14.85), should be asymptotically stable. Notice that in this equation, only ω_i's (excitation frequencies) are the variables yielding the values of $\{g_i\}$ and in turn the stability of the operation. For an arbitrary set of (ω_1, ω_2), and the corresponding $\{g_i\}$ the characteristic equation

$$CE_{cs} = \det \mathbf{A}_{cs} = 0 \tag{14.88}$$

is either stable, marginally stable, or unstable. It has a general form of

$$CE_{cs}[\bar{g}, \bar{\tau}, s] = 0 \tag{14.89}$$

© 2002 by CRC Press LLC

where $\bar{\tau}$ is the preselected time delay vector with the components of τ_i, and $\bar{g} = \{g_i\}$. It is clear that for $\bar{g} = 0$ the system represents the passive structure, and it is inherently stable. From this point we follow a recently introduced stability assessment strategy which is called the directional stability chart method (DSCM) (Huang and Olgac, 2000) in the structured steps given below:

1. Substitute $\{g_i\}$ with $\{\lambda g_i\}$, where $\lambda \in [0,1]$ is the interpolation constant.
2. Suppressing the known terms in the argument the reduced form of Equation (14.88) becomes

$$CE_{cs}[\lambda, s] = 0 \tag{14.90}$$

3. Substitute $s = i\omega$ and find the values of $\lambda \in [0,1]$ which yield imaginary roots to Equation (14.90). That is, we solve λ and ω from the complex valued equation

$$CE_{cs}[\lambda, \omega] = 0 \tag{14.91}$$

which yields two real equations for the unknowns $\lambda \in [0,1]$ and $\omega \in [0, \omega_{high}]$. These two simultaneous nonlinear equations can be solved using commercial packages (such as MATLAB or MAPLE). Here ω_{high} is an upper bound for the root finding routine beyond which such an imaginary root does not exist. This value is case specific and can be determined mathematically or observed graphically.
Let's assume m such solution pairs (λ_i, ω_i), i = 1…m are found for which there is a pair of characteristic roots on the imaginary axis.
4. Determine the root sensitivities at each of these root crossing points with respect to λ by differentiating Equation 14.88 and using the imaginary root corresponding to this λ_i, i.e., s = ω_i i. That is,

$$\frac{\partial}{\partial s} CE_{cs}(\lambda, s)\ ds + \frac{\partial}{\partial \lambda} CE_{cs}(\lambda, s)\ d\lambda = 0 \tag{14.92}$$

The root sensitivity is defined as:

$$S_s^{\lambda} = \frac{ds}{d\lambda}\Big|_{\lambda_i, \varpi_i} \tag{14.93}$$

If the real part of the root sensitivity, Re (S_s^{λ}), is positive then increasing λ_i at that point would give rise to two new unstable poles. On the contrary, if the root sensitivity is negative, then increasing λ_i would decrease the number of unstable poles by two, implying the two unstable poles migrate to stable zone.
5. Scan the m solutions of (4) from the smallest to the largest value of $\lambda \in [0,1]$ and create a table of stability outlook which shows the number of unstable roots at each interval of λ. Note that the starting point $\lambda = 0$ (i.e., no feedback control) for most physical systems (certainly for the one chosen in Figure 14.16) is asymptotically stable, i.e., no unstable right half roots exists.
6. The last interval $\lambda_m < \lambda < 1$ dictates the stability properties of the system in Equation (14.83) which corresponds to $\lambda = 1$. Notice that, $\lambda = 1$ implies that the full values of the gains $\{g_i\}$ are used as determined from Equation (14.87).
7. If $\lambda = 1$ represents stable operation, we further continue for $\lambda > 1$ until such λ_c is reached that yields marginal stability. The value of λ_c indicates the proximity of the operating point to the marginal stability. The larger the λ_c the better the stability.

© 2002 by CRC Press LLC

The systematic steps of (1–7) represent a D-subdivision method (Kolmanovskii and Nosov, 1986) application in one dimension (λ). The method states that as parameter λ varies from 0 to 1 the number of unstable roots changes only at the points of (λ_i, ω_i), i = 1…m and the change is in the direction indicated by step (4). As this search is performed along the vector direction of $\{g_i\}$ the new methodology is called the directional stability chart method (DSCM).

DSCM procedure enables the user to assess the stability of the absorber at a given set of (ω_1, ω_2). This method can be used to establish operable (i.e., stable) vs. inoperable (i.e., unstable) sectioning in the (ω_1, ω_2) plane. These sections are determined off-line (a priori to the controls). In real-time applications the controller should first detect the present excitation frequencies (ω_1, ω_2), then check whether this point falls in the operable or inoperable zone, and ultimately activate the tuning control if the point (ω_1, ω_2) is in the operable zone.

14.3.1.3 Example Case

As a case study, an absorber with two masses is taken into account (as shown in Figure 14.16b), which is attached to an SDOF primary structure. The numerical values used for the simulations are:

$$m = 50 \text{ kg}; \quad c = 49.6 \text{ kg/s}; \quad k = 1922 \text{ N/m};$$
$$m_{a1} = 1 \text{ kg}; \quad c_{a1} = 1.32 \text{ kg/s}; \quad k_{a1} = 30.25 \text{ N/m};$$
$$m_{a2} = 1 \text{ kg}; \quad c_{a2} = 0.39 \text{ kg/s}; \quad k_{a2} = 42.25 \text{ N/m};$$

The feedback structure of Equation (14.83) is used with $\tau_1 = 0.05$ sec, $\tau_2 = 0.1$ sec, $\tau_3 = 0.2$ sec, and $\tau_4 = 0.25$ sec delays. As described earlier the gains are calculated for the resonance frequencies: $\overline{\omega} = [\omega_1, \omega_2] = [4.2,\ 9.8]$ rad/sec, as:

$$g_0 = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix} = \begin{bmatrix} -389.8606 \\ 667.5552 \\ -418.0914 \\ 194.7341 \end{bmatrix} \text{ (N/m)}$$

Notice that the τ_i's are evenly distributed between 0 and 0.25 sec. This maximum time delay is selected such that the aliasing will not appear even for the system response to 4.2 rad/sec excitation.

For this set of (ω_1, ω_2) we proceed with the directional stability chart method (DSCM) to check if Equation (14.91) has a solution in the interval of $\lambda \in [0,1]$. If it does not the combined system *together with the tuned absorber* is stable. As explained in Section 14.3.1.2(f) the range of the search is extended beyond $\lambda \in [0,1]$. The first root encountered is

$$\lambda_c = 1.0428; \quad \omega_{crossing} = 9.6648 \text{ rad/s};$$

and the corresponding $\text{Re}[ds/d\lambda] = 1.2178 > 0$ implying the passage of two stable roots to the right hand unstable plane. Notice the directional stability margin, the value of λ_c is only 4% larger than 1. That means keeping everything else fixed if we increase the operating gains $\overline{g}_0$ by 4% the combined system would become resonant (i.e., marginally stable).

Figure 14.18 shows the primary mass response $x(t)$, as well as the absorber displacements, $x_{a1}(t)$ and $x_{a2}(t)$ shown as insets when the primary structure experiences a dual harmonic excitation force

$$f(t) = \sin \omega_1 t + \sin \omega_2 t .$$

The oscillations are suppressed as the absorbers react with large amplitudes. The structure settles in at 50 seconds and then the two excitation frequencies are changed to

© 2002 by CRC Press LLC

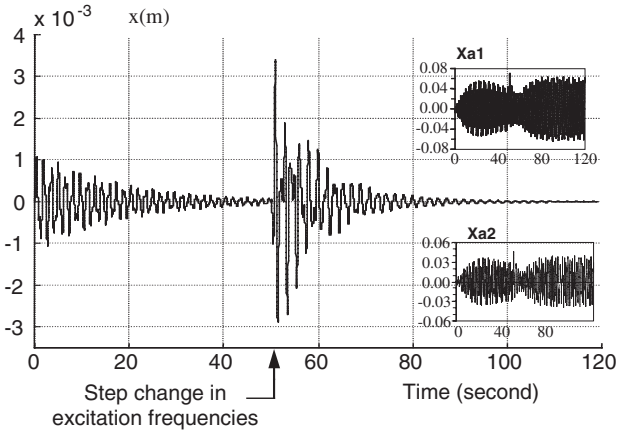


FIGURE 14.18 Time histories of x, x_{a1}, x_{a2} as the excitation frequencies ω_1 and ω_2 vary.

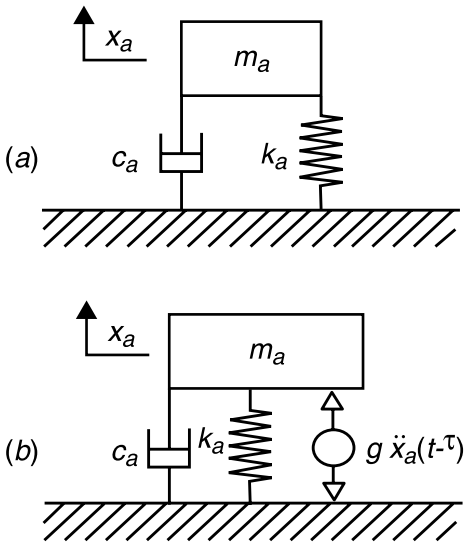


FIGURE 14.19 (a) Passive absorber, (b) delayed absorber with acceleration feedback.

$$\omega = [3.5,\ 9.2]\ \text{rad/sec}$$

by a step function. This sudden variation is observed in Figure 14.18. The absorber section is retuned to these two frequencies by the selection of new gains:

$$\begin{aligned} g_1 &= -494.6554 \\ g_2 &= 808.3025 \\ g_3 &= -514.6770 \\ g_4 &= 185.0157 \end{aligned} \qquad \text{(N/m)}$$

while keeping the delays (τ_is) the same.

© 2002 by CRC Press LLC

We follow the DSCM again for this new pair of (ω_1, ω_2) and show that there is no root for $\lambda \in [0,1]$. Therefore, the system still remains stable (operable). The suppression of the oscillation at the new frequencies, takes effect within about 50 seconds again. Notice, in this time interval the absorbers settle in their steady state response modes. This exercise shows the on-line tuning ability of the absorber section.

One critical issue in the above procedure is the determination of the excitation frequencies imposed on the primary system. It can only be achieved implementing FFT on the absorber displacement with a shifting window. The delay, which is caused by this process, is an issue in our ongoing research. It forces the control gain updates to be as frequent as the availability of the new ω_i's. For the example case above, it was assumed that the new values of ω_i's are instantly available to the controller.

14.3.2 Optimum ATVA for Wide-Band Applications

We next present a summary from an optimization effort reported in (Jalili and Olgac, 2000). The objective here is not a precise tuning of the ATVA to a number of frequencies; a frequency band of suppression is aimed for instead. This feature is achieved through a compromise of imperfect suppression at a given frequency.

14.3.2.1 Synopsis

The active tuning of the absorber is again achieved using partial state feedback with a controlled time delay. The final structure, which is named delayed feedback vibration absorber (DFVA), is optimized to yield minimum M_{peak} of the primary system involved for a given wide band of excitation frequencies. The optimization is performed over the absorber's structural parameters and the features of the feedback control. The assurance of the stability of the time-delayed system, which forms a critical constraint on the optimization, is also considered.

ATVA is typically formed by feedback which places the poles of the absorber section at some desired locations in order to create the spectral effects mentioned earlier (Jalili and Olgac, 2000). For instance, the DR strategy creates two dominant poles on the imaginary axis, but infinitely many finite poles at other locations (due to the time delay). As mentioned earlier, this procedure yields some undesirable M_{peak} at other frequencies. In this work we reiterate the DR feedback formation to place the dominant poles away from the imaginary axis, to an optimum location such that the primary system exhibits $min\{M_{peak}\}$ possible. Notice that the feedback strategy remains as simple as the case in DR, but the new flexibility, the placement of the dominant poles, offers a vehicle for further improvement such as performance optimization.

14.3.2.2 Delayed Feedback Vibration Absorber (DFVA)

A conventional passive absorber (Figure 14.19a) is reconfigured using a delayed *acceleration feedback* (Figure 14.19b). This forms the delayed feedback vibration absorber (DFVA). The corresponding new system dynamics and the respective characteristic equation are

$$m_a \ddot{x}_a(t) + c_a \dot{x}_a(t) + k_a x_a(t) - g\ddot{x}_a(t-\tau) = 0 \tag{14.94}$$

$$m_a s^2 + c_a s + k_a - g s^2 e^{-\tau s} = 0 \tag{14.95}$$

By selecting the control parameters g and τ properly, the dominant roots can be moved anywhere off the imaginary axis. This general DFVA lends itself to an optimization process for the most desirable location of these dominant roots. The effort in this work is to re-tune the DR feedback gain, g, and time delay, τ, such that the peak frequency response of the primary structure attains its global minimum, over a desired "*wide band*" frequency range. In the following sections, the

© 2002 by CRC Press LLC

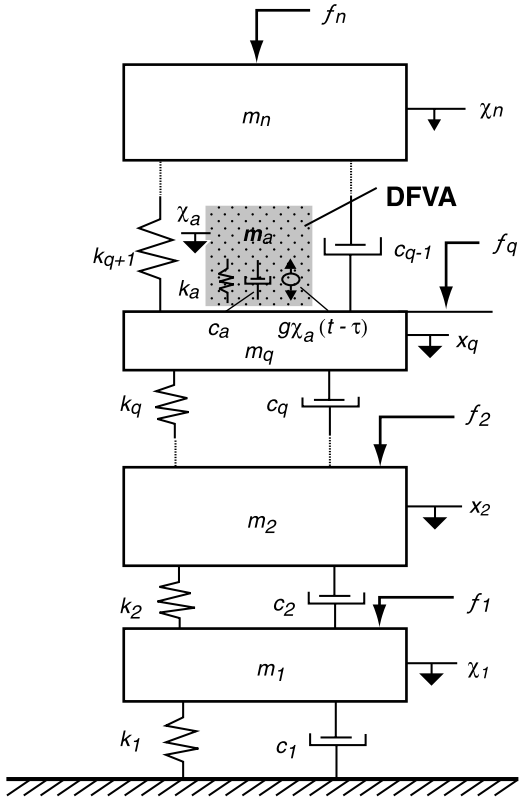


FIGURE 14.20 The DFVA implementation on an n-DOF structure.

governing equations of the dynamics and the selection procedure for the control parameters are presented.

The DR vibration absorber can be considered as a special case of DFVA. For an ideal DR, the dominant poles of the characteristics equation are placed on the imaginary axis $s = \pm j\omega_c$, at a designated frequency ω_c. As stated earlier, the DR offers perfect absorption at this frequency, but to the cost of worsening frequency response at the adjacent frequencies. This can be improved by different formation of feedback such that the dominant poles are positioned away from the imaginary axis, at some optimum locations. In this case, the primary system's frequency response would not be zero anywhere, but it will be flat over a desired frequency range, say $[\omega_{low}, \omega_{up}]$, around a specified center frequency ω_c. The strategy for the control parameters' selection and the end effects are discussed in what follows.

14.3.2.3 The Governing Equations

A general *n* degree-of-freedom (DOF) mechanical system is considered with a DFVA for suppressing undesired oscillations caused by wide band force excitations. The resulting combined system, the DFVA appended on the primary structure, is depicted in Figure 14.20. The DFVA is attached on the *q*-th mass (or DOF) of the primary structure. The objective of this work is twofold: to tune the control parameters g and τ, and to select the passive absorber characteristics m_a, c_a, and k_a, such that the frequency response of the *q*-th mass, on the primary system, is minimized in a given frequency interval.

The state-space representation of this system is written in the form of linear simultaneous differential equations, see (Jalili and Olgac, 2000) for details.

© 2002 by CRC Press LLC

$$\dot{\mathbf{y}}(t) = \mathbf{A}_0\mathbf{y}(t) + g\,\mathbf{A}_\tau\,\dot{\mathbf{y}}(t-\tau) + \mathbf{f}(t) \qquad t \geq t_0 \tag{14.96}$$

where

$$\mathbf{y}(t) = \{\ \underbrace{x_1, \dot{x}_1, ..., x_n, \dot{x}_n}_{2n}, x_{a1}, \dot{x}_{a1}\ \}^T,\ \mathbf{f}(t) = \{\ \underbrace{0, f_1, ..., 0, f_n}_{2n}, 0, 0\ \}^T \in \Re^{2(n+1)\times 1}$$

are the state variable and excitation vectors. The displacements of the primary structure and the absorber are denoted by $x_i(t)$, $i = 1, 2, ..., n$ and $x_a(t)$, respectively. $\mathbf{A}_0$ and $\mathbf{A}_\tau \in \Re^{2(n+1)\times 2(n+1)}$ are the constant system matrices, τ and g are the time delay and feedback gain.

The Laplace domain representation of the system is

$$(s\mathbf{I} - \mathbf{A}_0 - g\,s\,e^{-\tau s}\mathbf{A}_\tau)\,\mathbf{Y}(s) = \mathbf{F}(s) \;\Rightarrow\; \mathbf{H}(s)\,\mathbf{Y}(s) = \mathbf{F}(s) \tag{14.97}$$

and the corresponding characteristic equation is

$$|\mathbf{H}(s)| = \det[\ s(\mathbf{I} - g\,e^{-\tau s}\,\mathbf{A}_\tau) - \mathbf{A}_0\ ] = 0 \tag{14.98}$$

For each $g \neq 0$ and $\tau \neq 0$, Equation (14.98) has infinite number of roots, which are called the *spectrum* of the time-delayed system.

The Equation (14.97) yields a solution for $x_q(t)$. The objective of the ATVA is to bring $x_q(t)$ to zero or as close to it as possible.

14.3.2.4 Optimum DFVA

As stated before, the objective of this study is to minimize the maximum system response to a wide band excitation. By performing a number of simulations, it is shown that both flattening the frequency response and minimizing the peak frequency yield the same result. In fact, this is analytically proved for the passive absorber attached to an undamped primary structure (Bapat et al., 1979; Warburton et al., 1980). To the best of our knowledge, there is no companion analytical proof in cases where the primary is damped.

The numerical problem encountered here is a min-max problem: to find the absorber parameters m_a, c_a, k_a, g, and τ which minimize the supremum of the frequency response $X_q(\omega)$. As such, we seek the optimal absorber parameters $\mathbf{X}_{abs} = [m_a\ k_a\ c_a\ g\ \tau]^T$ to

$$\min_{\mathbf{X}_{abs}} \left\{ \max_{\omega_{low} \leq \omega \leq \omega_{up}} [X_q(\mathbf{X}_{abs}, \omega)] \right\} \tag{14.99}$$

subject to the physical bounds

$$m_a,\ k_a,\ c_a,\ g, \text{ and } \tau\ >\ 0 \tag{14.100}$$

The numerical procedure for this optimization is discussed in what follows. Notice that the passive absorber, i.e., $g = 0$, is always stable. The stability issue arises when active control is used. Especially, the delay element in the feedback can drive the system to instability. As the control parameters g and τ are used for optimization, their influence on the system stability should be studied in parallel. While the dominant characteristic poles of the absorber are placed at some optimum locations, away from the imaginary axis, the poles of the combined system should remain in the left-half of complex plane.

© 2002 by CRC Press LLC

The stability assurance imposes some limitations on the control parameters g and τ. This introduces an additional constraint to the inequalities (14.100) for the proposed optimization problem, described by Equation (14.99). This problem is treated next.

14.3.2.5 Stability of the Combined System

For a time-delayed system, such as the one described in Section 14.3.2.2, the stability analysis is relatively difficult since the characteristic equation is transcendental rather than algebraic. The stability properties of such a complicated configuration are addressed in some earlier works (Olgac et al. 1997; 1996). The sufficient and necessary condition for asymptotic stability is that the roots of the characteristics Equation (14.98), all have negative real parts. This equation is transcendental and the verification of the root locations is not trivial.

It is typical that increasing feedback gain g causes instability as the roots move from left to right of the complex plane as shown in Olgac and Hosek (1997) and briefly described in Section 14.2.1.3). In summary, Olgac and Hosek (1997) conclude the following: for a given delay τ, the operating gain g should be smaller than the gain for which the global system becomes marginally stable using the same delay, τ. The ratio of these two gains $g_{cs}(\tau)/g(\tau)$ can be defined as the stability margin of the control system, for that delay value τ. The comparison of the parametric plots of $g_{cs}(\omega_{cs})$ vs. $\tau_{cs}(\omega_{cs})$ for the combined system with the $g(\omega_c)$ vs. $\tau(\omega_c)$ of the DFVA reveals this stability picture for a range of τ's. See (Olgac et al. 1997) for details.

14.3.2.6 Optimization Scheme

We seek the optimum absorber characteristics vector $\mathbf{X}_{abs} = [m_a\ k_a\ c_a\ g\ \tau]^T$, such that the peak value of the frequency response amplitude of the q-th mass, as given in Equation (14.99), is minimized. The process would also flatten the frequency response of the system. This proposition forms a min-max problem: we wish to minimize the maximum frequency response over a desired frequency range. The optimization process has to comply with some constraints, as described next.

Following earlier explanations, increasing feedback gain g while time delay τ is kept constant, drives the combined system through the stable, marginally stable, and ultimately to the unstable behavior. For a particular delay $\tau = \tau_0$, the combined system crossings, $\omega_{csl}(\tau_0)$ are determined from Equation (14.97) along with the corresponding gains. Notice that, the multiplicity of the gains appears due to the transcendental nature of the characteristic equation for one fixed time delay. To ensure stability of the system, the feedback gain g should be smaller than the infemum of these $g_{csl}(\omega_{csl})$ values. That is,

$$g < g_{\min}(\omega_{csl}(\tau)), \tag{14.101}$$

where

$$g_{\min} = \text{infemum}\begin{Bmatrix} g_{cs}(\omega_{cs1}) \\ g_{cs}(\omega_{cs2}) \\ g_{cs}(\omega_{cs3}) \\ \vdots \end{Bmatrix}, \quad \text{for} \quad \tau = \tau_0 \tag{14.102}$$

Thus, the plot of $g_{min}(\omega_{cs})$ vs. $\tau_{cs}(\omega_{cs})$ is the lower envelope of the parameterized stability plot of the combined system, $g_{cs}(\omega_{cs})$ vs. $\tau_{cs}(\omega_{cs})$. This envelope is numerically obtained for each "m_a, c_a, k_a" set of absorber parameters, yielding the constraint in Equation (14.101).

It is assumed that primary system parameters are all kept fixed during the optimization procedure. For the passive absorber, typically the optimized values of the stiffness and damping are between

© 2002 by CRC Press LLC

their physical bounds, whereas the optimum value of the absorber mass is always found at the upper bound (Jacquot, 1978; Esmailzadeh et al., 1998; Ozguven et al., 1986). Thus, we select an absorber mass and keep it fixed for simplicity. The optimization is performed only over the four dimensional vector; the spring stiffness, damping coefficient, feedback gain, and time delay, which defines the absorber and control characteristics, $\mathbf{X}_{\text{abs}} = [k_a\ c_a\ g\ \tau]^{\text{T}}$.

The proposed constrained optimization problem is now recast in the form:

$$J = \underset{\mathbf{X}_{abs}}{Min}\ \left\{\ \underset{\omega_{low}\le\omega\le\omega_{up}}{Max}\ [X_q(\mathbf{X}_{abs},\omega)]\ \right\},\quad \text{for } \mathbf{X}_{abs} = [\,k_a,\ \ c_a,\ \ g\ \ \tau]^T \tag{14.103}$$

subject to physical bounds

$$\begin{aligned} h_1 &\equiv -k_a < 0,\\ h_2 &\equiv -c_a < 0,\\ h_3 &\equiv -g < 0,\\ h_4 &\equiv -\tau < 0, \end{aligned} \tag{14.104}$$

and the stability constraint

$$h_5 \equiv g - g_{\min} < 0 \tag{14.105}$$

The upper bounds k_a^{up} and c_a^{up} are chosen considering some practical limitations. A set of initial values of $\mathbf{X}_{\text{abs}}$ starts the process. These k_a and c_a parameters are taken from a preliminary study: the optimal passive absorber (obviously for $g = 0$), which is the result of the same optimization problem as in Equation (14.103) except for $g = 0$. The initial g and τ are selected based on the DR which uses the optimal passive absorber and operates at the resonance frequency of the primary structure, ω_{peak}. The optimum value of $\mathbf{X}_{\text{abs}}$ is numerically determined, in such a way that the cost function described in Equation (14.103) subject to constraints, Equations (14.104 and 14.105), is minimized.

It should be noted that the stability constraint, expression Equation (14.105), changes simultaneously in each iteration. This is due to the dependence of this constraint on the absorber spring stiffness, k_a, and damping coefficient, c_a, which happen to be a part of the parameter vector sought.

Since the computation of the higher derivatives of the objective function $G(\mathbf{X}_{\text{abs}})$ is very complicated, we deploy optimization techniques which use inferior information: the direct update methods. These methods require the computation of only first derivatives of the cost function. Using the information obtained from the previous iterations, convergence towards the minimum is accelerated. In this study, we follow BFGS (Broyden-Fletcher-Goldfarb-Shanno) method which has been proved to be most effective in similar applications. We refer the interested reader to (Gill et al., 1981) for detailed derivations of this method.

14.3.2.7 A Case Study

The primary system is taken as a single-degree-of-freedom structure which is subjected to a wide band frequency load in the interval of $\omega \in [400,\ 1500]$ Hz. The DFVA is appended to it, and the resulting combined structure is shown in Figure 14.21. The primary system parameters are taken as $m_1 = 5.77$ kg, $k_1 = 251.132 \times 10^6$ N/m, and $c_1 = 1142.0$ kg/s. This system has a peak frequency at $\omega_{\text{peak}} = 1050$ Hz. The absorber mass to primary mass ratio is taken to be 3.9% ($m_a = 0.227$ kg).

The following strategy is used to determine the initial guess $\mathbf{X}_{abs}^0$. Considering the given physical parameters, the optimal passive absorber (i.e., $g = \tau = 0$) is found first: $\bar{k}_a = 9.5471 \times 10^6$ N/m and $\bar{c}_a = 359.20$ kg/s with the same objective function in mind (Equation (14.103)). Based on this

© 2002 by CRC Press LLC

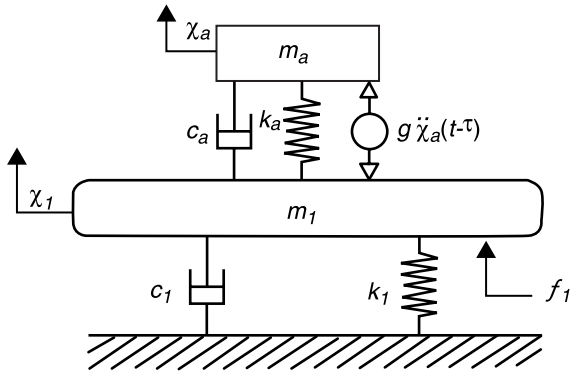


FIGURE 14.21 Implementation of a DFVA on a SDOF structure.

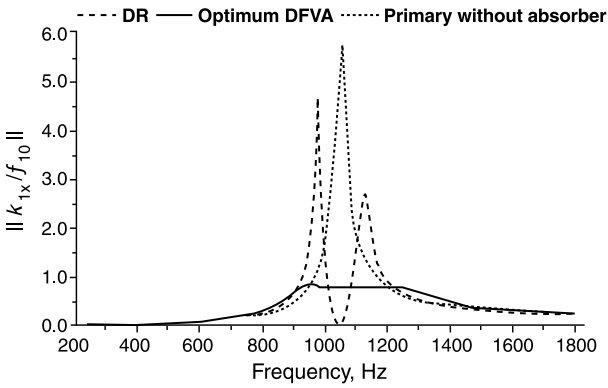


FIGURE 14.22 The frequency responses for different settings**.**

optimum passive absorber, the pure DR feedback gain and time delay are calculated at $\omega_c = 1050$ Hz: $g_c = 0.0550$ kg and $\tau_c = 1.1874 \times 10^{-3}$ sec. This setting of $[\bar{k}_a, \bar{c}_a, g_c, \tau_c]^T$ is taken as the starting vector: $\mathbf{X}^0_{abs} = [9.5471 \times 10^6, 359.20, 0.0550, 1.1874 \times 10^{-3}]^T$, for the following steps. The full scale optimization problem over $[k_a, c_a, g, \tau]^T \in \Re^4$ is handled to arrive at the *optimum DFVA* solution as:

$$\mathbf{X}^{opt}_{abs} = [k_a = 9.8014 \times 10^6,\ c_a = 35.4787,\ g = 0.0424,\ \tau = 0.7305 \times 10^{-3}]^T.$$

The frequency responses of the primary system alone, with DR (i.e., $\mathbf{X}^0_{abs}$), and with optimum DFVA are shown in Figure 14.22. All of these are obtained using the excitation force $f_1 = f_{10} \sin \omega t$. The optimum DFVA delivers better than 80% improvement over the DR in the vibration suppression at the first side frequency where the DR causes a substantial peak.

On the other hand, setting the absorber stiffness and damping (k_a, c_a) free during the optimization iterations offers a significant improvement as opposed to keeping them fixed. For open loop (passive) system ($g = 0$, $\tau = 0$) the characteristic roots are

$$s_{1,2} = -72.17 \pm j5921, s_{3,4} = -109.40 \pm j7348,\ f_1 = 942.4 \text{ Hz},\ f_2 = 1169.5 \text{ Hz} \quad (14.106)$$

The dominant roots for the corresponding optimum delayed feedback parameters are at

$$s^*_{1,2} = -406 \pm j5750,\ \text{for optimum DFVA case},\ g = 0.0424 \text{ kg},\ \tau = 0.7305 \text{ ms} \quad (14.107)$$

© 2002 by CRC Press LLC

The respective damped natural frequencies at the dominant roots are calculated as

$$f_1^* = 915 \text{ Hz}, f_2^* = 1079 \text{ Hz}, f_3^* = 1294 \text{ Hz} \quad \text{for optimum DFVA case} \tag{14.108}$$

From Figure 14.22, it is clear that the peak frequencies are at about 930 Hz and 1090 Hz for optimum DFVA. The root placement is simpler to do than verifying the frequency response outlook for design purposes.

Acknowledgments

The presented study is supported in part by Connecticut Innovations Inc. (Grant No. 00Y14). Authors also wish to acknowledge the contributions of Mr. Chang Huang and Dr. Nader Jalili during their graduate work. Some parts of the discussions are reflective of their research work and referenced articles.

References

Abdel-Mooty M. and Roorda J., 1991, Time delay compensation in active damping of structures, *Journal of Engineering Mechanics,* 117, 2549.

Bapat, V. A. and Kumaraswamy, H. V., 1979, Effect of primary system damping on the optimum design of an untuned viscous dynamic vibration absorber, *Journal of Sound and Vibration*, 63, 469–474.

Carter, B. C., 1929, Improvements in or relating to damping or oscillation-checking devices, British Patent 337,466.

Den Hartog, J. P., 1938, Tuned pendulums as torsional vibration eliminators, *Stephen Timoshenko 60th Anniversary Volume*, Macmillan, London.

Den Hartog, J. P. and Ormondroyd J., 1928, Theory of the dynamic vibration absorber, *Transactions of the ASME*, APM-50-7, 11–22.

Den Hartog, J. P. and Ormondroyd, J., 1930, Torsional vibration dampers, *Transactions of ASME,* 52, 133–152.

DiDomenico, E., 1994. Passive vibration tuning with neural networks, *Proceedings of Smart Structures and Materials*, SPIE, 152–162.

Esmailzadeh, E. and Jalili, N., 1998, Optimal design of vibration absorbers for structurally damped Timoshenko beams, *ASME Journal of Vibration and Acoustics*, 120(4), 833–841.

Ezure, K. and Seto, K., 1994, Vibration control of two-degrees-of-freedom system using active dynamic absorber, *Transactions of the Japan Society of Mechanical Engineers, Part C*, 60, 788–795.

Filipovic, D. and Olgac, N., 1998, Torsional delayed resonator with speed feedback, *IEEE/ASME Transactions on Mechatronics*, 3(1), 67–72.

Frahm, H., 1911, Device for damping vibrations of bodies, United States Patent 989,958.

Franklin G. F., Powell J. D., and Emami-Naeini A., 1994, Feedback control of dynamic systems, Addison-Wesley, Reading, Massachusetts.

Gill, P. E., Murray, W., and Wright, M. H., 1981, *Practical Optimization*, Academic Press, New York.

Hertz, D., Jury, E.I., and Zeheb, E., 1984, Simplified analytic stability test for systems with commensurate time delays, *IEEE Proceedings, Part D*, 131(1), 52–56.

Hosek, M., 1997, Tunable torsional vibration absorber: The centrifugal delayed resonator, Ph.D. Dissertation, University of Connecticut, Storrs, Connecticut.

Hosek, M., 1998, Robust delay-controlled dynamic absorber, Doctoral dissertation, Czech Technical University in Prague.

Hosek, M., Elmali, H., and Olgac N., 1997a, A tunable torsional vibration absorber: The centrifugal delayed resonator, *Journal of Sound and Vibration*, 205(2), 151.

© 2002 by CRC Press LLC

Hosek, M., Elmali, H., and Olgac N., 1997b, Centrifugal delayed resonator: Theory and experiments, *Proceedings of ASME Design Engineering Technical Conferences*, 16th Biennial Conference on Mechanical Vibration and Noise, Paper No. DETC97/VIB-3829, September 14–17, 1997, Sacramento, California.

Hosek, M., Elmali, H., and Olgac N., 1999, Centrifugal delayed resonator pendulum absorber, United States Patent No. 5934424.

Hosek, M. and Olgac, N., 1999, A single-step automatic tuning algorithm for the delayed resonator vibration absorber, *Proceedings of 1999 ASME International Mechanical Engineering Congress and Exposition*, Dynamic Systems and Control Division, 67, 157–164, November 14–19, Nashville, Tennessee.

Hosek, M., Olgac, N., and Elmali, H., 1999, The centrifugal delayed resonator as a tunable torsional vibration absorber for MDOF systems, *Journal of Vibration and Control*, 5, 2, 299–322.

Huang, C. and Olgac, N., *Proceedings, 2000 American Control Conference*, Chicago.

Inman, D. J., 1994, *Engineering Vibration*, Prentice-Hall, Englewood Cliffs, New Jersey.

Jacquot, R. G., 1978, Optimal dynamic vibration absorbers for general beams, *Journal of Sound and Vibration*, 60 (4), 535–542.

Jalili, N. and Olgac, N., 1999, Optimum delayed feedback vibration absorber for flexible beams, *Smart Structures*, NATO Science Series, Kluwer Academic, Amsterdam, 65, 237–246.

Jalili, N. and Olgac, N., 2000, A sensitivity study on optimum delayed feedback vibration absorber, *ASME Journal of Dynamic Systems, Measurement and Control*, 122, 2, 314–321.

Kolmanovskii, V.B. and Nosov, V.R., 1986, *Stability of Functional Differential Equations*, Academic Press, London.

Marshall, J. E., 1979, *Control of Time Delay Systems,* Peter Peregrinus Ltd., New York.

MATLAB/SIMULINK, version 5.3, Release R11, 1999, The MathWorks Inc., Natick, Massachusetts.

Nishimura, H., Nonami, K., Cui, W., and Shiba, A., 1993, H_∞ control of multi-degree-of-freedom structures by hybrid dynamic vibration absorber (experimental consideration of robustness and control performance), *Transactions of the Japan Society of Mechanical Engineers. Part C*., 59, 714–720.

Olgac, N., 1995, Delayed resonators as active dynamic absorbers, United States Patent 5,431,261.

Olgac, N., 1996, Single mass dual frequency fixed delayed resonator, United States Patent 5,505,282.

Olgac, N., Elmali, H., and Vijayan, S., 1996, Introduction to dual frequency fixed delayed resonator (DFFDR), *Journal of Sound and Vibration*, 189, 355–367.

Olgac, N., Elmali, H., Hosek, M., and Renzulli, M., 1995, High-frequency implementation of delayed resonator concept using piezoelectric actuators, *Proceedings of ACTIVE 95 — 1995 International Symposium on Active Control of Sound and Vibration*, 57–66.

Olgac, N., Elmali, H., Hosek, M., and Renzulli, M., 1997, Active vibration control of distributed systems using delayed resonator with acceleration feedback*, Journal of Dynamic Systems, Measurement and Control*, 119, 380–389.

Olgac, N. and Holm-Hansen, B., 1994, A novel active vibration absorption technique: Delayed resonator, *Journal of Sound and Vibration*, 176, 93–104.

Olgac, N. and Holm-Hansen, B., 1995a, Tunable active vibration absorber: The delayed resonator, *ASME Journal of Dynamic Systems, Measurement and Control*, 117, 513–519.

Olgac, N. and Holm-Hansen, B., 1995b, Design considerations for delayed-resonator vibration absorbers, *Journal of Engineering Mechanics*, 121, 80–89.

Olgac, N. and Hosek, M., 1995, Dual-frequency vibration absorption using delayed resonator with relative position measurement. *Proceedings of ASME Dynamic Systems and Control Division, ASME International Mechanical Engineering Congress and Exposition*, 2, 791, November 12–17, San Francisco, California.

Olgac, N. and Hosek, M., 1997, Active vibration absorption using delayed resonator with relative position measurement, *Journal of Vibration and Acoustics*, 119, 131–136.

Olgac, N. and Jalili, N., 1998, Modal analysis of flexible beams with delayed resonator vibration absorber: Theory and experiments, *Journal of Sound and Vibration*, 218(2), 307–331.

© 2002 by CRC Press LLC

Orfanidis S. J., 1996, *Introduction to Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey.
Ormondroyd, J. and Den Hartog, J. P., 1928, The theory of the dynamic vibration absorber, *Transactions of ASME*, 50, 9–22.
Ozguven, H. N. and Candir, B., 1986, Suppressing the first and second resonance of beams by dynamic vibration absorbers, *Journal of Sound and Vibration*, 111(3), 377–390.
Puksand, H., 1975, Optimum conditions for dynamic vibration absorbers for variable speed systems with rotating and reciprocating unbalance, *International Journal of Mechanical Engineering Education*, 3, 145–152.
Rao, S. S., 1995, *Mechanical Vibrations*, 3rd ed., Addison-Wesley, New York.
Renzulli, M., 1996, An algorithm for automatic tuning of the delayed resonator vibration absorber, M.S. Thesis, University of Connecticut, Storrs.
Renzulli M. E., Ghosh-Roy R., and Olgac N., 1999, Robust control of the delayed resonator vibration absorber, *IEEE Transactions on Control Systems Technology*, 7(6), 683.
Rodellar J., Chung L. L., Soong T. T., and Reinhorn A. M., 1989, Experimental digital control of structures, *Journal of Engineering Mechanics*, 115, 1245.
Seto, K. and Fumishi, Y., 1991, A Study on Active Dynamic Absorber, ASME paper DE - Vol. 38.
Stepan, G., 1989, *Retarded Dynamical Systems Stability and Characteristic Functions*, Longman, London.
Sun, J. Q., Jolly, M. R., and Norris, M. A., 1995, Passive, adaptive, and active tuned vibration absorbers — a survey, *ASME Transactions, Special 50th Anniversary, Design Issue*, 117, 234–242.
Thomson, W. T., 1988, *Theory of Vibration with Applications*, Prentice Hall, Englewood Cliffs, New Jersey.
Thowsen A., 1981a, An analytic stability test for a class of time-delay systems, *IEEE Transactions on Automatic Control*, AC-26, 735.
Thowsen A., 1981b, The Routh-Hurwitz method for stability determination of linear differential-difference systems, *International Journal of Control*, 33, 991.
Thowsen A., 1982, Delay-independent asymptotic stability of linear systems, *IEE Proceedings*, 129, Part D, 73, 1982.
Valášek, M. and Olgac, N., 1999, New concept of active multiple frequency vibration suppression technique, *Smart Structures*, NATO Science Series, Kluwer Academic, 65, 373–382.
Warburton, G. B. and Ayorinde, E. O., 1980, Optimum absorber parameters for simple systems, *Earthquake Engineering and Structural Dynamics*, 8, 197–217.
Wilson, W. K., 1968, *Practical Solution of Torsional Vibration Problems,* Chapman and Hall, London.
Yang, B., 1991, Noncollocated control of damped string using time delay, *Proceedings, 1991 American Control Conference*, Boston.
Youcef-Toumi K. and Bobbett J., 1991, Stability of uncertain linear systems with time delay, *Journal of Dynamic Systems, Measurements, and Control*, 113, 558.
Youcef-Toumi K. and Ito O., 1990, A time delay controller for systems with unknown dynamics, *Journal of Dynamic Systems, Measurements, and Control*, 112, 133.
Zitek P., 1984, Stability criterion for anisochronic dynamic systems, *Acta Technica CSAV*, 4, 399.

© 2002 by CRC Press LLC

15

Vibration Suppression Utilizing Piezoelectric Networks

15.1 Introduction
15.2 Passive and Semi-Active Piezoelectric Networks for Vibration Absorption and Damping
15.3 Active-Passive Hybrid Piezoelectric Network Treatments for General Modal Damping and Control
15.4 Active-Passive Hybrid Piezoelectric Network Treatments for Narrowband Vibration Suppression
15.5 Nonlinear Issues Related to Active-Passive Hybrid Piezoelectric Networks
15.6 Summary and Conclusions

Kon-Well Wang
Pennsylvania State University

15.1 Introduction

Because of their electromechanical coupling characteristics, piezoelectric materials have been explored extensively for structural vibration control applications. Some of the advantages of piezoelectric actuators include high bandwidth, high precision, compactness, and easy integration with existing host structures to form the so-called *smart* structures. In a purely active arrangement, an electric field is applied to the piezoelectric materials (which can be surface bonded or embedded in the host structure) based on sensor feedback and control commands. In response to the applied field, stress/strain will be induced in the piezoelectric material and active control force or moments can thus be created on the host structure to suppress vibration.

In recent years, a considerable amount of work has been performed to further utilize piezoelectric materials for structural control by integrating them with external electrical circuits to form piezoelectric networks. Such networks can be utilized for passive, semi-active, and active-passive hybrid vibration suppressions (Lesieutre, 1998; Tang, Liu, and Wang, 2000). Many interesting phenomena have been explored and promising results have been illustrated. The objective of this chapter is to review these efforts and assess the state-of-the-art of vibration control treatments utilizing piezoelectric networks. The basic concepts and development of passive and semi-active networks are discussed in Section 15.2. With the introduction of active actions, various issues, and recent advances regarding active-passive hybrid networks are presented in Sections 15.3 through 15.5.

© 2002 by CRC Press LLC

15.2 Passive and Semi-Active Piezoelectric Networks for Vibration Absorption and Damping

In a purely passive situation, piezoelectric materials are usually integrated with an external shunt circuit (Hagood and von Flotow, 1991; Lesieutre, 1998). As the host structure vibrates, the piezoelectric layer will be deformed. Because of the electromechanical coupling characteristic, electrical field/current will then be generated in the shunt circuit. With proper design of the shunt components (inductor, resistor, or capacitor), one can achieve the so-called electrical damper or electrical absorber effects.

Soon after Hagood and von Flotow provided the first quantitative analysis of piezoelectric shunt networks, Hagood and Crawley (1991) applied the resonant shunt piezoelectric (RSP) network to space truss structures. An important feature of that work is the usage of a synthetic inductor, which is essentially a circuit with an operational amplifier feeding back current rate, thus simulating the effect of an inductor. For small piezoelectric capacitance and low structural modes, the optimum RSP requires a large inductance with low electrical resistance, which could be difficult to realize. The introduction of the synthetic inductor can effectively circumvent this problem and, more importantly, ease the tuning of the circuit because the inductance can be changed by varying the gain of the feedback current rate. Following along the same line, Edberg et al. (1992) developed a simulated inductor composed of operational amplifiers and passive circuitry connected as a gyrator, which can produce hundreds or thousands of henries with just a few simple electronic components. Because the value of simulated inductance may be easily changed by a variable resistor, it may be possible to have passive damping circuits monitor the frequencies to which they are subjected and alter their own characteristics in order to optimize the behavior.

From the power-flow point of view, the effect of inductance in the RSP is to cancel the inherent capacitive reactance of the piezoelectric material. As proposed by Bondoux (1996) the same effect can be expected by introducing a negative capacitance. Although this negative capacitance is impossible to achieve passively, it can be realized by using a small operational amplifier circuit similar to the synthetic inductor. Bondoux compared the negative capacitance shunting and the RSP and found that the use of a negative capacitance provides a broadband efficiency allowing multiple-mode damping. A similar conclusion was also drawn by Spangler and Hall (1994) and Bruneau et al. (1999). In general, the negative capacitance can increase the electromechanical coupling coefficient and enhance the efficiency of piezoelectric damping in both the resistive shunt and RSP network. The disadvantages are that the negative capacitance can generate electrical instabilities (Bondoux, 1996), and the high ratio of capacitance compensation is difficult to achieve in practice without adding a sensor to the circuit to account for the thermal changes of the piezoelectric capacitance (Bruneau et al. 1999).

A common thread of the aforementioned studies is the usage of an electronic circuit with operational amplifiers. Although they are not true semi-active approaches, these studies laid down a foundation for semi-active (adaptive/variable) absorption and damping research that continues today. An immediate application of the tunable nature of the synthetic inductor is a self-tuning piezoelectric vibration absorber developed by Hollkamp and Starchville (1994) (see Figure 15.1, case a). An RSP network is formed as an electromechanical vibration absorber and the shunt inductance are controlled through varying the resistance of a motorized potentiometer in the synthetic inductor, which enables on-line adjustment of the RSP tuning to maximize the performance function. In their approach, an ad hoc performance function was selected as the ratio of the RMS voltage across the shunt and the RMS structure response. If the ratio increases, the change in the inductance is in the proper direction and the inductance is again changed in that direction. If the ratio decreases, the direction is reversed. Although one deficiency of this simple control scheme is that the absorber will never settle on a single tuning value, it is effective for slow time-varying systems which can tolerate the tuning fluctuations and the time it takes to initially tune the absorber.

© 2002 by CRC Press LLC

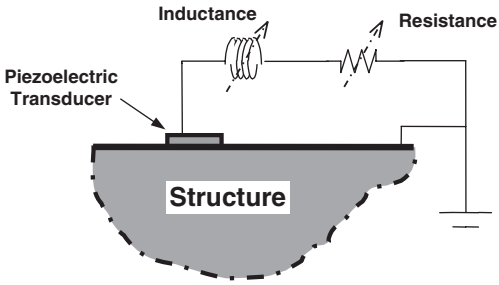


FIGURE 15.1 Schematics of some semi-active RSP damper/absorbers. Case (a): R = inherent resistance in the circuit; L on-line adjusted. Case (b): R and L on-line adjusted.

Wang et al. (1996) proposed a semi-active RSP scheme with variable inductance and resistance (see Figure 15.1, Case b). Their focus was on an improved control law that can handle not only quasi-steady-state scenarios but also structures with more general disturbances such as nonperiodic and transient loadings. They found that in such a semi-active configuration, the rates of the total system energy (the main structure mechanical energy plus the electrical and mechanical energies of the RSP) and the main structure energy are dependent on the circuit resistance, inductance, and inductance rate. It was recognized that an effective approach would be to reduce the total system energy while constraining the energy flowing into the main structure. Because two objectives were to be accomplished and they could contradict each other, an algorithm using variable resistance and changing rate of inductance as control inputs was developed to balance the energies. By selecting the total system energy as a Lyapunov functional, one can guarantee system stability through ensuring a negative rate of the system energy, while at the same time maximizing energy dissipation of the vibrating host structure.

Davis et al. (1997) and Davis and Lesieutre (1998) studied the possibility of tuning a mechanical absorber using shunted piezoelectric materials. The idea was initiated from the inertial piezoelectric actuator concept developed for structural vibration control (Dosch et al., 1995) where the forcing element in a proof mass actuator was replaced by a piezoelectric element with dual-unimorph displacement amplification effect. An important finding is that in such a configuration, the absorber stiffness is dependent on the ratio of the electrical impedance of the open circuit piezoelectric capacitance to the electrical impedance of the external shunt circuit. Therefore, by varying the impedance of an external shunt circuit, the natural frequency and, in some cases, the modal model damping of the vibration absorber will vary (Davis et al. (1997). Based upon this, Davis and Lesieutre (1998) developed an actively tuned solid-state piezoelectric vibration absorber. Because their goal was to maintain minimum structural response at a certain (may be varying) frequency, they adopted a capacitive shunting scheme without a resistive element, as damping is not needed in such applications. It should be noted that depending on different performance requirements, different shunting schemes could be optimally designed. To obtain variable capacitance, a "ladder" circuit of discrete capacitors wired in parallel was used. At a given time, the controller switches on some or all of the capacitors in parallel with the piezoelectric element, thereby changing the absorber stiffness and tuning the absorber frequency to the favorable value. The range of the adjustable stiffness is nevertheless limited by the piezoelectric electromechanical coupling coefficient. On a benchmark experimental setup, Davis and Lesieutre (1998) achieved a $\pm$ 3.7% tunable frequency band relative to the center frequency. Within the tuning band, increases in performance (vibration amplitude reduction) beyond passive performance were as great as 20dB. In addition, the averaged increase in performance across the tunable frequency band was over 10dB.

Piezoelectric materials realize a significant change in mechanical stiffness between their open-circuit and short-circuit states. This property was exploited by Larson et al. (1998) to develop a high-stroke acoustic source over a wide frequency range. By switching between the open-circuit

© 2002 by CRC Press LLC

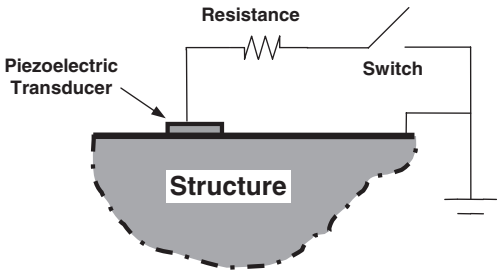


FIGURE 15.2 Schematics of some semi-active piezoelectric switching dampers. Case (a) Switching between open and short circuit states, R = 0. Case (b) switching between open circuit and resistive shunting, R = optimal passive value.

and short-circuit states, the acoustic driver's stiffness (and, therefore, its natural frequency) can be changed, allowing it to track a changing frequency with high amplitude. While Larson et al. (1998) proposed a practical realization of such a state-switched source for applications in active sonar systems, underwater research, and communication systems, Clark (1999a) found it is also useful in forming a semi-active piezoelectric damper. Using a typical energy-based control logic (Leitmann, 1994), Clark (1999a) illustrated how a piezoelectric actuator can be switched between the high and low stiffness states to achieve vibration suppression (see Figure 15.2, Case a). When the system is moving away from equilibrium, the circuit is switched to the high-stiffness state (open circuit), and the circuit is switched to the low-stiffness state (short circuit) when the system is moving toward equilibrium. This has the effect of suppressing deflection away from equilibrium, and then at the end of the deflection quarter-cycle, dissipating some of the stored energy so that it is not returned to the structure. In the open-circuit case, deflection stores energy by way of mechanical stiffness and the piezoelectric capacitance effect. When the system is switched to the short-circuit state, the charge stored across the capacitor is shunted to ground, effectively dissipating that portion of the energy. Clark (1999b) further studied the case that used a resistive shunt instead of a pure short circuit at low-stiffness state (see Figure 15.2, Case b), and compared the state-switching control with an optimally tuned passive resistive shunt. It was shown that for the example used in the study the optimal resistive shunt performed better for suppressing transient vibrations. The state-switching approach, however, provided better performance for off-resonance (particularly low-frequency) excitations, and was very robust to changes in system parameters.

Richard et al. (1999) also developed a piezoelectric damper using the switching concept (see Figure 15.2, Case a). The switch itself consisted simply of a pair of MOSFET transistors and little power was needed. The main difference between their approach and that proposed by Clark (1999a, 1999b) is in the switching law. Instead of switching between open and short circuits at different quarter-cycles of vibration, Richard et al. (1999) proposed to maintain the open circuit as the nominal state, and briefly switch to the short-circuit state to dump the electrical energy only when the structure displacement reaches a threshold value. Although no analytical results were available, they found that the best vibration suppression was achieved for a threshold corresponding to a maximum and a minimum of the displacement or output voltage in one vibration period. The time interval corresponding to the short-circuit time is also important and can be tuned. It was experimentally shown that the shortest time led to the best damping efficiency. They demonstrated enhanced damping performance of the proposed device over the passive resistive shunt.

Warkentin and Hagood (1997) studied a nonlinear piezoelectric shunting scheme with a four-diode full-wave rectifier and a DC voltage source. If the vibration amplitude is small, the voltage produced by the accumulation of charge on the piezoelectric capacitance is less than the DC voltage. Under this condition, all the diodes are reverse biased and no current will flow through the shunt, and the system is at the open-circuit condition. For larger motions, the diodes are turned on, current flows through the shunt, and the piezoelectric voltage is clipped at positive and negative DC voltage

© 2002 by CRC Press LLC

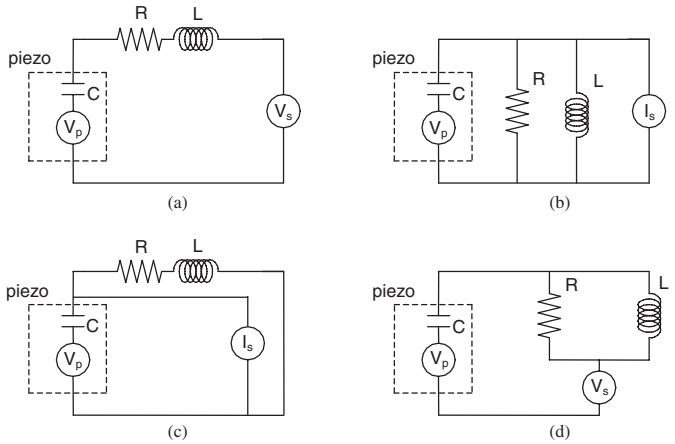


FIGURE 15.3 Schematics of active-passive hybrid piezoelectric networks. V_p: equivalent voltage generator attributed to the piezoelectric effect; V_s: voltage source; I_s: current or charge source; C: piezo capacitance; R: resistance; L: inductance. (From Tang, J., Liu, Y., and Wang, K. W., *Shock and Vibration Digest*, 32(3), 189–200, ©2000, Sage Publication, Inc.)

by the rectifier and the voltage source. The arrangement of the diodes ensures that the current always flows into the positive terminal of the DC source. If the DC source is implemented as a rechargeable battery or a regulated switching power circuit, the vibration energy removed from the structure may thus be recovered in a usable electrical form. The different stiffness exhibited at the open-circuit and short-circuit phases, combined with the voltage offset from the shunt voltage source, will produce a mechanical hysteresis. Although its performance was not as good when compared with the loss factor achieved by a conventional resistive shunt operating at optimum frequency, the rectified DC shunt is a frequency-independent device and its potential energy recovery ability remains an attractive feature. Warkentin and Hagood (1997) also studied resistive shunting with variable circuit resistance. An optimization approach was used to determine the ideal periodic resistance time history. The effective loss factors obtained in the simulations assuming sinusoidal deformation exceeded twice the values achieved by the fixed resistive shunt.

15.3 Active-Passive Hybrid Piezoelectric Network Treatments for General Modal Damping and Control

While the earlier investigations in RSP networks mostly focused on passive applications, it is clear that shunting the piezoelectric does not preclude the use of a coupled piezoelectric materials–shunt circuit as active actuators. That is, by integrating an active current or voltage control source with the passive shunt, one can achieve an active-passive hybrid piezoelectric network (APPN) configuration (Figure 15.3). The passive damping can be useful in stabilizing controlled structures in the manner analogous to proof mass actuators (Miller and Crawley, 1988; Zimmerman and Inman, 1990; Garcia et al., 1995). Hagood et al. (1990) developed a general modeling strategy for systems with dynamic coupling through the piezoelectric effect between a structure and an electrical network. Special attention was paid to the case where the piezoelectric electrodes are connected to an arbitrary electrical circuit with embedded voltage and current sources. They obtained good agreement between the analytical and experimental results, and concluded that the inclusion of electrical circuitry between the source and the structure gives the designer greater ability to model actual effects and to modify the system dynamics for closed-loop controls.

Niezrecki and Cudney (1994) addressed the power consumption characteristics of the piezoelectric actuators. The electrical property of a piezoelectric actuator is similar to a capacitor, which

© 2002 by CRC Press LLC

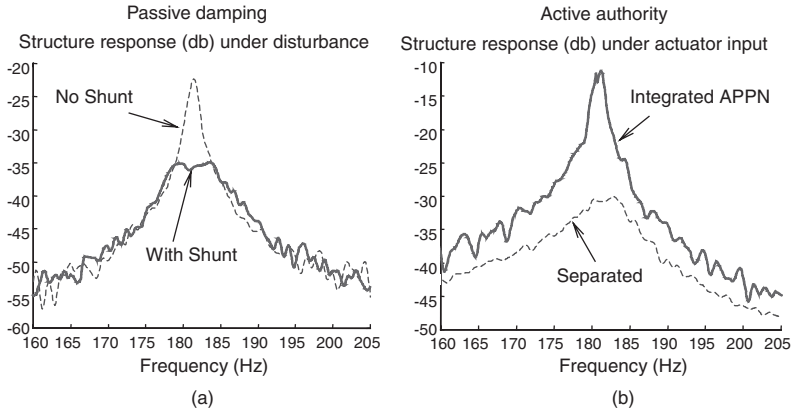


FIGURE 15.4 Experimental results on system passive damping and active authority of APPN. (From Tang, J., Liu, Y., and Wang, K. W., *Shock and Vibration Digest*, 32(3), 189–200, ©2000, Sage Publication, Inc.)

leads to a reactive current that provides only an electromagnetic field and does not perform work or result in useful power being delivered to the load. Therefore, the power factor of a piezoelectric actuator is approximately zero. Niezrecki and Cudney (1994) proposed to add an appropriate inductance to correct the power factor to unity within a small but useful frequency range. They studied two cases: adding inductors in parallel and in series with the piezoelectric actuator. In both cases, a resonant LC circuit was formed, and around the resonant frequency the reactive elements cancelled and the phase between current and voltage became zero, resulting in a unity power factor. They incorporated the internal resistance of the piezoelectric actuators and inductors in their analysis. Implementing the parallel LC circuit reduced the current consumption of the piezoelectric actuator by 75% when compared to the current consumption of the actuator used without an inductor. Implementing the series LC circuit produced a 300% increase in the voltage applied to the actuator compared to the case when no inductor was used. In both cases, the apparent power was reduced by 12dB.

From the above work, one may realize that the RSP network not only will increase the system's passive damping, but also will greatly increase the active control authority around the shunt resonant frequency. Agnes (1994, 1995) examined the simultaneous passive and active control actions of an RSP network through open-loop analyses. A modal model was developed to evaluate the hybrid vibration suppression effect, and open-loop experiments were performed for validation. Using Hagood and von Flotow's optimal RSP tuning results (1991) to determine the shunt circuit parameters, it was observed that not only the passive damping effect was significant, the modal response of the structure to the input voltage or current signal is also increased greatly. Using voltage as the driving source (Figure 15.3a), the shunted system frequency response was similar to the nonshunted response below the tuned (shunted mode) frequency, but exhibited greater roll-off above the tuned frequency. For broadband control, this would help prevent spillover because the magnitude of the response is, in general, lower for higher modes. When current source was used (Figure 15.3c), the shunted system's active action was less effective below the tuned frequency when compared to the nonshunted case, but no roll-off was observed in the high-frequency region. Tsai (1998) and Tsai and Wang (1999) also performed experimental investigations to illustrate the shunt circuit's passive damping ability (Figure 15.4a), as well as its active authority enhancement ability (Figure 15.4b) in APPN. Through exciting the structure with the actuator, they compared the open-loop structural response of the integrated APPN and the configuration with separated RSP and a piezoelectric actuator. While the two configurations have the same passive damping ability, the APPN configuration can drive the host structure much more effectively than the separated treatment does (Figure 15.4b), which clearly demonstrated the merit (high active authority) of the integrated APPN design.

© 2002 by CRC Press LLC

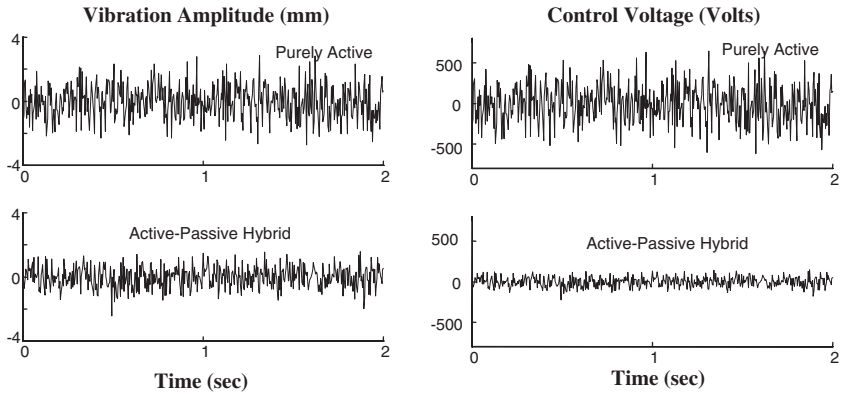


FIGURE 15.5 Comparisons of purely active and active-passive hybrid systems: performance and required voltage for vibration control. (From Tsai, M. S. and Wang, K. W., *Smart Materials and Structures*, 5(5), 695–703, ©1996, IOP Publishing, Inc.)

While Hagood and von Flotow's tuning results (1991) can minimize the maximum frequency response for a passive system, they are not necessarily good choices for an active-passive hybrid system. That is, the question of how to determine the system's active and passive parameters to achieve efficient hybrid vibration control still remains. From the driving voltage (control input) standpoint, the circuit inductance value will determine the electrical resonant frequency around which the active control authority will be amplified, and although appropriate resistance is required to achieve broadband passive damping, resistance in general reduces the active authority amplification effect (Tsai and Wang, 1999). To balance between active and passive requirement conflicts and performance tradeoffs and achieve an optimal configuration, a scheme was synthesized to concurrently design the passive elements and the active control law (Kahn and Wang, 1994, 1995; Tsai and Wang, 1996, 1999). This approach is to ensure that active and passive actions are configured in a systematic and integrated manner. The strategy developed is to combine the optimal control theory with an optimization process and to determine the active control gains together with the values of the passive system's parameters (the shunt circuit resistance and inductance). The procedure contains two major steps: (1) for a given set of passive parameters (resistance R and inductance L), form the system equations into a regulator control problem and derive the active gains to minimize a cost function representing vibration amplitude and control effort via the optimal control theory (Kwakernaak and Sivan, 1972); (2) for each set of the passive control parameters R and L, an optimal control exists with the corresponding minimized cost function, J, and control gains. That is, J is a function of R and L. Therefore, utilizing a nonlinear programming algorithm (Arora, 1989), one can determine the resistance and inductance that further reduce J. Note that as the R and L values are varied during the optimization process, step (1) is repeated to update the active gains simultaneously. In other words, by concurrently modifying the values of the active gains and passive parameters, an "optimized" optimal control system can be obtained.

The APPN system and the control/design scheme have been evaluated on various types of structures. In a multiple APPN ring vibration control problem (Tsai and Wang, 1996), a random sequence was generated to compare the structure displacements and control efforts (voltages) of the uncontrolled, the active, and the active-passive systems. From the results, it is clear that the active-passive action resulted in significant vibration reduction compared to the uncontrolled case (a 25dB reduction in standard deviation). In addition, the hybrid approach also outperformed the purely active system (Figure 15.5). Figure 15.5 also shows that the active-passive hybrid controller requires much less voltage than the active controller does.

Based on this simultaneous optimal-control/optimization strategy, Tsai (1998) and Tsai and Wang (1999) performed a detailed parametric analysis for the APPN design, showing that the optimal

© 2002 by CRC Press LLC

resistance and inductance values for the hybrid system could be quite different from those of the passive system, especially when demand on performance is high and/or when the number of actuators is much smaller than the number of controlled modes. For the APPN configuration, when the weighting on control effort increases, the optimal resistance (R) and inductance (L) values using the concurrent design will approach those derived from the passive optimization procedure. In general, when demand on control performance increases, the resistance value becomes smaller to enhance the active authority amplification effect, and inductance reduces to cover a wider frequency bandwidth. The excitation bandwidth also plays an important role, as it determines to which mode the RL values will be tuned.

Tsai and Wang (1998) addressed the robustness issue in systems controlled by APPN. They developed an algorithm with coupled μ synthesis (Zhou et al., 1996) and an optimization process to design a robust hybrid controller. In their example, they found that the structural uncertainty level that the hybrid controller can tolerate (the maximum uncertainty level at which the μ synthesis approach can find a solution) is much higher than what a purely active controller can tolerate, and thus the hybrid controller is much more robust than a purely active system.

Tang and Wang (1999a) applied the active-passive hybrid piezoelectric networks to rotationally periodic structures. Consisting of identical substructures, a rotationally periodic structure is essentially a multi-degrees-of-freedom system. The coupling between the substructures will split the otherwise repeated substructure frequency to a group of frequencies, which creates the problem of how to tune the shunt. By utilizing the unique property of rotationally periodic structures, Tang and Wang (1999a) developed an analytical method to determine the passive and active parameters for the control design, where the active control was used to compensate for the mistuning effect due to substructure coupling. The overall effect of the active and passive actions minimizes the maximum frequency response for all modes. Identical shunting circuit and control gains were applied to each substructure, which could bring convenience in implementations.

As mentioned earlier, while the resistor in the hybrid control system provides passive damping, it also tends to reduce the active control authority by dissipating a portion of the control power (Tsai and Wang, 1999). To further improve the efficiency of the active-passive hybrid piezoelectric network, Morgan and Wang (1998) proposed using a variable resistor in the circuit. The key feature in this control design was the introduction of a parametric control law to adjust the variable resistor. When electrical energy is flowing into the actuator/structure from the voltage source, the circuit is shorted to reduce the loss of control power. When the energy is flowing out of the actuator/structure, a positive value of resistance is selected for passive energy dissipation. They suggested using a digital potentiometer connected to the parametric controller to achieve the hardware realization. Their analysis showed that the parametric control law can significantly increase the efficiency of the active-passive hybrid control system, especially for narrowband and/or low to moderate gain applications. The reduced control effort could make it an attractive option for applications when minimizing the power consumption is critical.

Tsai and Wang (1999) concluded that the APPN will become less effective when the excitation bandwidth increases, because its passive damping and active authority amplification effects are narrowbanded. To circumvent this, they proposed to integrate the APPN with broadband damping treatments (Tsai and Wang, 1997). Specifically, they studied the integration with the enhanced active constrained layer (EACL) configuration (Liao and Wang, 1996, 1998a, 1998b; Liu and Wang, 1999), to which edge elements are added to the active constraining layer (ACL) (Park and Baz, 1999) to increase the transmissibility and active action authority. They found that adding the hybrid network to a traditional active constrained layer (ACL) treatment will not lead to much extra damping because of low transmissibility between the host structure strain and the piezoelectric coversheet deformation. However, the integration of APPN with EACL can achieve high damping. A comparison of the APPN, EACL, and combined APPN-EACL damping treatments was performed. An objective function was defined to reflect the vibration amplitude and control effort. In general, smaller objective function means better overall performance and thus better hybrid damping

© 2002 by CRC Press LLC

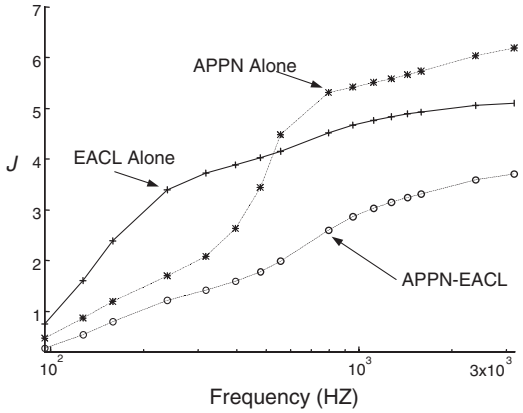


FIGURE 15.6 Objective function (J) comparison between different configurations.

ability. The minimized objective function, J, for different configurations vs. excitation bandwidth was obtained (Figure 15.6). As shown in the figure, APPN outperforms EACL when the bandwidth is small, but becomes less effective than EACL as bandwidth increases. On the other hand, the combined APPN-EACL system can outperform the individual APPN and EACL cases, under both narrowband and broadband excitations.

So far, in most active-passive hybrid piezoelectric network studies, only one of the series configurations has been considered. That is, the resistor, the inductor, and the power source (voltage source) were all connected in series with the piezoelectric actuator (Figure 15.3a). Wu (1996) found that by connecting the resistor and inductor in parallel with the piezoelectric material, one can achieve a similar passive vibration absorbing/damping effect as that of the series configuration proposed by Hagood and von Flotow (1991). Combining parallel and series passive configurations with the parallel and series active driving, one can envision a few different active-passive hybrid piezoelectric network configurations, some of which are shown in Figures 15.3b–d. From the viewpoint of linear system superposition, the structure response is a summation of that caused by external disturbance and that caused by control input. Therefore, for the passive effect to function normally in the absence of the active control input, we should use charge or current control when the power source is in parallel with the shunting elements, such as those shown in Figures 15.3b and c. Although one has to resort to complicated circuit design to obtain a charge source, it has the potential benefit of avoiding the piezoelectric hysteresis (Main et al., 1995). However, it should be noted that different configurations yield roughly the same passive and hybrid damping abilities (Tang and Wang, 2001).

15.4 Active-Passive Hybrid Piezoelectric Network Treatments for Narrowband Vibration Suppression

The focus of Section 15.3 is systems utilizing APPN for general modal damping and control. It has also been found that the APPN configuration could be very effective for narrowband vibration rejection. The active-passive hybrid approach is especially attractive for narrowband disturbances with varying frequencies (an example of this type of excitation is a machine with a rotating unbalance — the frequency variation could be a slow drift due to changes in operating conditions or a rapid spin-up when the machine is turned on), as discussed in this section.

While a passive piezoelectric vibration absorber (piezoelectric materials with passive resonant shunt) is effective for harmonic disturbance rejection (Hagood and von Flotow, 1991), it could be sensitive to frequency variations and system uncertainties. As stated in Section 15.2, semi-active piezoelectric absorber concepts have been proposed to suppress harmonic excitations with time-varying

© 2002 by CRC Press LLC

frequencies. The implementation of these semi-active absorbers requires either a variable inductor or a variable capacitor element. While they are conceptually valid, both of these methods have some inherent limitations. For instance, the variable capacitor method (Davis et al., 1997) limits tuning of the piezoelectric absorber to a relatively small frequency range. The variable inductor approach (Hollkamp and Starchville, 1994), which is usually accomplished using a synthetic inductance circuit, can add a significant parasitic resistance to the circuit that is generally undesirable for narrowband applications. In either case, the variable passive elements can be difficult to tune rapidly with high accuracy.

With the above arguments, Morgan et al. (2000) and Morgan and Wang (2000) developed a high-performance active-passive hybrid alternative to the semi-active absorbers, utilizing the APPN configuration. Throughout this study, the system being considered was a generic mechanical system with a single piezoelectric actuator attached. The piezoelectric was shunted with an *RL* circuit as well as an active voltage source (Figure 15.3a). The passive inductance value was tuned to a nominal excitation frequency. Because the interest here is to use the APPN absorber characteristic to suppress vibrations at distinct frequencies, low damping (resistance) is required in the absorber. Therefore, other than the inherent resistance in the circuit, no extra passive resistor was added.

The active control law consists of three modules. The first part of the control law is designed to imitate a variable inductor so that the absorber is always tuned to the correct frequency. In addition, an active negative resistance action is used to reduce the absorber damping (inherent resistance in the circuit) and increase the absorber narrowband performance. To further enhance the robustness of the piezoelectric absorber, the system's apparent electromechanical coupling is increased using the third active action. The advantages of the active inductor include fast and accurate adjustment, no parasitic resistance, and easier implementation compared to a semi-active inductor. To ensure that the active inductance is properly tuned, an expression for optimal tuning on a general multiple-degrees-of-freedom (MDOF) structure was derived. The closed-loop inductance was achieved using this optimal tuning law in conjunction with an algorithm that estimates the fundamental frequency of the measured excitation. Details of the mathematical formulation and derivation can be found in Morgan et al., 2000 and Morgan and Wang, 2000.

The APPN adaptive absorber concept was implemented and experimentally verified on a lab fixture. Details of the test procedure and setup are described in Morgan and Wang (2000). Two test cases were considered: the first case is for an off-resonant excitation, and the second is for an excitation near a resonant frequency of the structure. The baseline system for the resonant excitation case is an optimally damped passive piezoelectric absorber. That is, the absorber is tuned to the resonant frequency and sufficient damping (resistance) is added to give a flat frequency response around the resonant frequency. In the off-resonant case, a passive absorber would be a poor choice for an excitation of varying frequency because of its small effective bandwidth. Therefore, the baseline for the off-resonant case is selected to be the response of the structure with the piezoelectric actuator shorted (no shunt circuit). The inputs to the controller are the structure response signal, the voltage across the passive inductor, and the excitation signal. The controller also contained a frequency estimation algorithm, which uses the measured excitation signal to continually estimate the excitation frequency.

The purpose of this experiment was to study the performance of the system when subjected to a harmonic excitation with varying frequency. The simplest such excitation is a linear chirp signal, which is a sinusoid of linearly increasing frequency. The three parameters that characterize the chirp signal are the nominal frequency f_o, the bandwidth of the frequency variation Δ, and the frequency rate of change $\dot{f}$ (Hz/s). For the linear chirp used here the frequency starts at $(1-\Delta)f_o$ at time t_s and increases at a rate of $\dot{f}$ until it reaches a maximum frequency of $(1+\Delta)f_o$ at time t_f. In this experiment, the nominal excitation frequency and bandwidth were constant in each case and the frequency rate of change was varied. Four tests were carried out for both the near-resonant and off-resonant cases, with the frequency rate of change varying from 2 to 8 Hz per second. The excitation was applied at time $t = 0$, but the data acquisition system was set to have a trigger delay

© 2002 by CRC Press LLC

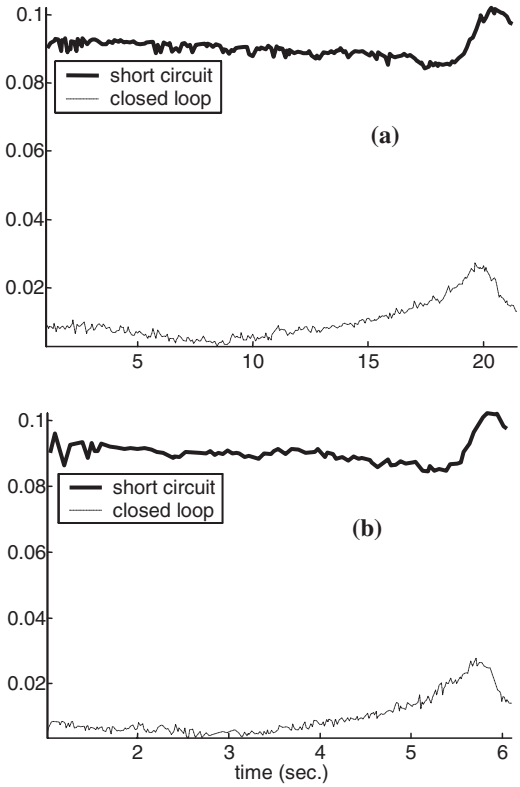


FIGURE 15.7 Experimental response (sensor voltage readings) envelopes (off-resonant case), $f_o = 205$ Hz. (a) $\dot{f}$ = 2Hz/s, (b) $\dot{f}$ = 8 Hz/s.

of $t_s = 1$ second. The purpose of this delay is to discard the large transient response caused by initially applying the excitation, which would give a better estimate of the performance of the system under extended operating conditions.

Experimental results for the off-resonant case are shown in Figure 15.7. The excitation bandwidth used for the off-resonant case is ±10% of the nominal frequency, which corresponds to approximately 40 Hz. These plots show the response envelopes for the cases $\dot{f}$ = 2 Hz/s and $\dot{f}$ = 8 Hz/s. From these results it appears that performance of the active-passive absorber is relatively unaffected by the frequency rate of change. This is somewhat surprising because the optimal tuning is determined using a quasi-steady-state assumption, which is only valid for excitations with very slowly changing frequency. The conclusion is that the combination of the quasi-steady-state tuning law and the active coupling enhancement allows the adaptive absorber to achieve good performance even for rapidly varying excitations. The combination of a rapidly changing excitation frequency and a very wide frequency bandwidth is a difficult problem for a semi-active device. However, the active-passive piezoelectric absorber presented here could have the performance and robustness necessary for these applications.

Experimental results for two of the near-resonant cases are shown in Figure 15.8. Once again we see that performance of the active-passive absorber is relatively unaffected by the frequency rate of change. Although the performance of the optimal passive absorber baseline is already much better than the original system (no absorber), the adaptive active-passive absorber still can outperform the baseline system significantly.

Through extensive parametric studies (Morgan and Wang, 2000), the proposed design was also compared with two active and active-passive vibration control methods: the Filter-X algorithm (Fuller et al., 1996) for off-resonant excitation and the concurrent APPN optimal control-optimization process

© 2002 by CRC Press LLC

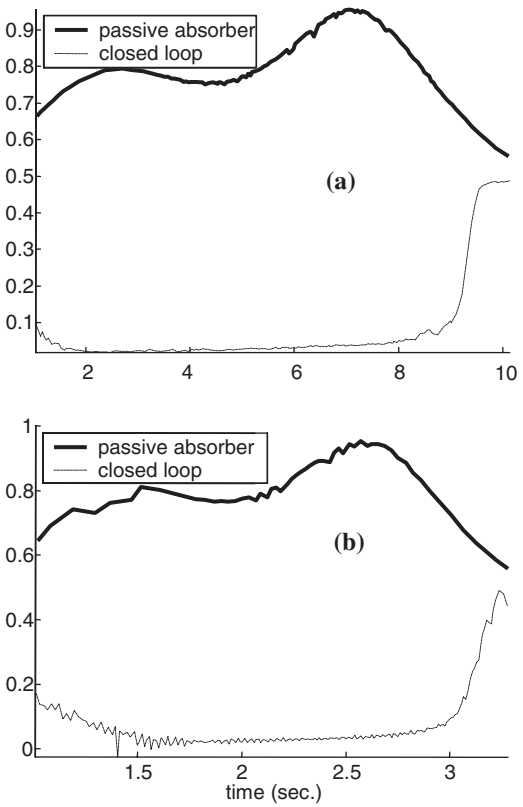


FIGURE 15.8 Experimental response (sensor voltage readings) envelopes (near-resonant case), $f_o = 92$ Hz. (a) $\dot{f} = 2$ Hz/s, (b) $\dot{f} = 8$ Hz/s.

(Tsai and Wang, 1999) for near-resonant situation. It was shown that the adaptive active-passive absorber can outperform the two systems significantly while requiring less control effort. These parametric studies also illustrated the effects of the absorber parameters and excitation characteristics on the performance of the adaptive active-passive piezoelectric absorber design.

While promising, the system proposed by Morgan et al. (2000) and Morgan and Wang (2000) is for suppression of excitations with only a single dominant frequency. To further enhance and expand the ability of such a device, a multi-frequency adaptive piezoelectric vibration absorber design was developed (Morgan and Wang, 2001). Building upon the single-frequency disturbance rejection network configuration, multiple circuit branches and an additional active law were added. The active control law effectively decouples the dynamics of the individual circuit branches. This decoupling action allows the tunings of the multi-frequency absorber to be calculated using an analytical optimal tuning law. The proposed design was shown to be effective for simultaneously suppressing two harmonic excitations with time-varying frequencies, and it can achieve better performance while requiring less control power than the Filtered-X algorithm. The design and analysis presented can be extended in a straightforward manner to cases with more excitation frequencies.

15.5 Nonlinear Issues Related to Active-Passive Hybrid Piezoelectric Networks

As mentioned, the piezoelectric network can result in high-performance vibration control through the dual effects (passive damping and active authority enhancements) of the shunt circuit. On the

© 2002 by CRC Press LLC

other hand, high performance corresponds to a high electrical field across the piezoelectric material, especially under high loading conditions. When the electrical field level is high, the linear assumption often made in most piezoelectric actuator-based systems (the linear constitutive relation between the stress, strain, electrical field, and electrical displacement of the piezoelectric material) may no longer be valid. This is because the material hysteresis and the high-order nonlinear relationship between the mechanical response and electrical field could become very significant when a high electric field occurs on the piezoelectric actuators.

In recent investigations performed by Tang et al. (1999) and Tang and Wang (1999b, 2000), the nonlinear behavior of the piezoelectric material was investigated experimentally and analytically. Through lab tests, one can clearly see the complexity of the material property, especially at high field levels. This fact suggests that one way to utilize the high field (high authority) regime is to consider the various nonlinear phenomena as uncertainties and develop robust controls to compensate for such uncertainties. By treating the nonlinearity (or part of it) as bounded uncertainties, a constitutive relation is proposed. For example, if the linear constitutive relation is used as the basic model, the actual actuation strain at a certain field will be the linear deterministic value plus some bounded uncertainty (Tang et al., 1999). For one-dimensional structures, the modified constitutive equations can be expressed in the following form,

$$\tau = E_p \varepsilon - h_{31} D$$

$$E = -h_{31}(\varepsilon - \varepsilon_0) + \beta_{33} D$$

where τ, ε, D, and E represent the stress, strain, electrical displacement (charge/area), and electrical field (voltage/length along the transverse direction) within the piezoelectric patch, respectively, and E_p, h_{31}, and β_{33} are the Young's modulus, piezoelectric constant, and dielectric constant of the material. Here, ε_0 represents the uncertainty in the strain-field relation, which is bounded as $|\varepsilon_0| < \varepsilon^*$. The bounds can be selected according to the maximum field level that the actuator will undergo and identified from experimental data.

Given the new constitutive equation and uncertainty bounds, a robust control algorithm based on the of sliding mode theory (Slotine and Li, 1991; Utkin, 1993) was then developed to compensate for the piezoelectric nonlinearities (Tang et al., 1999; Tang and Wang, 1999b, 2000). In general, the dynamics of a system so controlled consist of a reaching mode and a sliding mode. The strategy for designing a sliding mode controller involves: (1) the design of a switching manifold (sliding surface) on which the system will be asymptotically stable (the so-called sliding mode, where fast convergence is desired); and (2) designing a controller which can force the state trajectory to reach the switching manifold in finite time (the so-called reaching mode, where a brief reaching time is desired). When all the nonlinearities were considered as uncertainties, a linear-quadratic regulator (LQR) optimal control formulation (Kwakernaak and Sivan, 1972) was used to set up the sliding surface and ensure stability on the surface (Tang et al., 1999; Tang and Wang, 2000). When the high-order nonlinearity was included in the model and the other nonlinearities were treated as uncertainties, the Lyapunov stability approach was utilized to select the sliding surface (Tang and Wang, 1999b).

The effectiveness of the proposed approach was demonstrated through experimental studies and numerical analyses on vibration control of an APPN-treated cantilever beam structure. For the purpose of comparison, the simulation results of the beam tip displacement for the linear optimal controller are shown in Figure 15.9 (upper plot). The dashed line represents the ideal situation where there are no piezoelectric nonlinearities. However, since the voltage across the piezoelectric material is high, the actual result is given by a solid line, where the piezoelectric nonlinearities are simulated as bounded uncertainty. The performance of the linear controller is obviously degraded by the piezoelectric nonlinearities. The sliding mode control result considering all the piezoelectric nonlinearities as uncertainties (Tang and Wang, 2000) is then illustrated in Figure 15.9 (lower plot),

© 2002 by CRC Press LLC

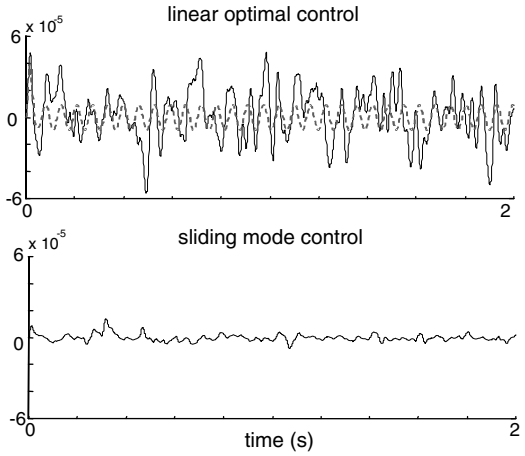


FIGURE 15.9 Beam tip vibration amplitude. Upper plot: linear control (dashed line, ideal case without piezoelectric nonlinearity; solid line, realistic case with piezoelectric nonlinearity). Lower plot: Sliding mode control compensating for piezo nonlinearity.

which clearly shows much better performance. Here, for a fair comparison, the two controllers are designed so that they utilize the same RMS value of control power input. These results illustrate that with such a nonlinear robust controller, one can fully utilize the high authority characteristics of the APPN system.

15.6 Summary and Conclusions

The major findings and achievements to date in vibration control utilizing piezoelectric networks can be summarized as follows:

- Passive and semi-active tuning of piezoelectric circuit elements can be effective in various damping and vibration absorption applications, especially for systems with no variations or under slow/small changes. Most of the control algorithms are based on energy or power analysis, and through adjustable resistance, inductance, and capacitance, as well as open- to short-circuit state switching.
- To further enhance the performance of piezoelectric networks, active voltage or current sources have been added to form an active-passive hybrid piezoelectric network (APPN). Circuit elements not only can provide passive damping, they also can increase the treatment's active authority. To tune the system properly for general modal damping and control applications, one approach is to employ a concurrent optimization scheme and simultaneously synthesize the active gains and passive parameters. Such an APPN approach could outperform a purely active system with less control effort.
- The active actions in an active-passive hybrid piezoelectric network can also be used to tune passive component parameters. Such an approach merely adds a dynamic compensator, with gains emulating the circuit variables and the electromechanical coupling parameter. This feature could be especially effective for rejecting narrowband excitations with variable frequencies, where the APPN adaptive absorber effect is utilized.
- The shunt circuit could significantly increase the APPN active authority through increasing the voltage across the piezoelectric material. That is, high performance corresponds to high electrical field, especially under high loading conditions. When the electrical field level is high, piezoelectric nonlinear characteristics should be considered in designing and controlling the system. One effective approach to utilize the nonlinear high authority features of the

© 2002 by CRC Press LLC

APPN is to synthesize a nonlinear robust controller (e.g., the sliding mode controller discussed in this chapter) to include and compensate for the actuator nonlinearities in the design process.

- Most investigations to date have concluded that a well-designed, self-contained APPN system could have the advantages of both purely active and passive systems and could outperform both approaches.

Acknowledgments

The author would like to thank several of his former and current graduate students (Steven P. Kahn, Jing-Shiun Lai, Ronald Morgan, Michael Philen, Jiong Tang, Meng-Shiun Tsai, and Wei-Kuei Yu) for contributing to the research presented in this chapter.

References

Agnes, G. S., 1994, Active/passive piezoelectric vibration suppression, *Proceedings of SPIE, Smart Structures and Materials*, 2193, 24–34.

Agnes, G. S., 1995, Development of a modal model for simultaneous active and passive piezoelectric vibration suppression, *Journal of Intelligent Material Systems and Structures*, 6(4), 482–487.

Arora, J. S., 1989, *Introduction to Optimum Design*, McGraw Hill, New York.

Bondoux, D., 1996, Piezo-damping: A low power consumption technique for semi-active damping of light structures, *Proceedings of SPIE, Smart Structures and Materials*, 2779, 694–699.

Bruneau, H., Le Letty, R., Claeyssen, F., Barillot, F., Lhermet, N., and Bouchilloux, P., 1999, Semi-passive and semi-active vibration control using new amplified piezoelectric actuators, *SPIE Smart Structures and Materials*, 3668, 814–821.

Clark, W. W., 1999a, Semi-active vibration control with piezoelectric materials as variable stiffness actuators, *Proceedings of SPIE, Smart Structures and Materials*, 3672, 123–130.

Clark, W. W., 1999b, State-switched piezoelectric systems for vibration control, *Proceedings of the 1999 AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit*, Part 4, 2623–2629.

Crawley, E. F. and Lazarus, K. B., 1990, Induced strain actuation of isotropic and anisotropic plates, *AIAA Journal*, 29(6), 944–951.

Davis, C. L. and Lesieutre, G. A., 1998, An actively tuned solid-state piezoelectric vibration absorber, *Proceedings of SPIE, Smart Structures and Materials*, 3327, 169–182.

Davis, C. L., Lesieutre, G. A., and Dosch, J. J., 1997, Tunable electrically shunted piezoelectric vibration absorber, *Proceedings of SPIE, Smart Structures and Materials*, 3045, 51–59.

Dosch, J. J., Lesieutre, G. A., Koopmann, G. H., and Davis, C. L., 1995, Inertial piezoelectric actuators for smart structures, *Proceedings of SPIE, Smart Structures and Materials*, 2447, 14–25.

Edberg, D. L., Bicos, A. S., Fuller, C. M., Tracy, J. J., and Fechter, J. S., 1992, Theoretical and experimental studies of a truss incorporating active members, *Journal of Intelligent Material Systems and Structures*, 3(2), 333–347.

Fuller, C.R., Elliott, S.J., and Nelson, P.A., 1996, *Active Control of Vibration*, Academic Press, London, 91–113.

Garcia, E., Webb, S., and Duke, J., 1995, Passive and active control of a complex flexible structure using reaction mass actuators, *ASME Journal of Vibration and Acoustics*, 117(1), 116–122.

Hagood, N. W., Chung, W. H., and von Flotow, A., 1990, Modeling of piezoelectric actuator dynamics for active structural control, *Proceedings of AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Part 4, 2242–2256.

Hagood, N. W. and Crawley, E. F., 1991, Experimental investigation of passive enhancement of damping for space structures, *Journal of Guidance, Control, and Dynamics*, 14(6), 1100–1109.

Hagood, N. W. and von Flotow, A., 1991, Damping of structural vibrations with piezoelectric materials and passive electrical networks, *Journal of Sound and Vibration*, 146(2), 243–268.

© 2002 by CRC Press LLC

Hollkamp, J. J. and Starchville, T. F., Jr., 1994, A self-tuning piezoelectric vibration absorber, *Journal of Intelligent Material Systems and Structures*, 5(4), 559–566.

Kahn, S. P. and Wang, K. W., 1994, Structural vibration controls via piezoelectric materials with active-passive hybrid networks, *Proceedings of ASME IMECE DE75*, 187–194.

Kahn, S. P. and Wang, K. W., 1995, On the simultaneous design of active-passive hybrid control actions for structures with piezoelectrical networks, ASME Paper # 95-WA/AD5.

Kwakernaak, H. and Sivan, R., 1972, *Linear Optimal Control Systems*, John Wiley and Sons, New York.

Larson, G. D., Rogers, P. H., and Munk, W., 1998, State switched transducers: A new approach to high-power, low frequency, under water projectors, *Journal of Acoustical Society of American*, 103(3), 1428–1441.

Leitmann, G., 1994, Semi-active control for vibration attenuation, *Journal of Intelligent Material Systems and Structures*, 5(6), 841–846.

Lesieutre, G. A., 1998, Vibration damping and control using shunted piezoelectric materials, *The Shock and Vibration Digest*, 30(3), 187–195.

Liao, W. H. and Wang, K. W., 1996, A new active constrained layer configuration with enhanced boundary actions, *Smart Materials and Structures*, 5, 638–648.

Liao, W. H. and Wang, K. W., 1998a, Characteristics of enhanced active constrained layer damping treatments with edge elements, Part 1: Finite element model and experimental validation, *ASME Journal of Vibration and Acoustics*, 120(4), 886–893.

Liao, W. H. and Wang, K. W., 1998b, Characteristics of enhanced active constrained layer damping treatments with edge elements, Part 2: System analysis, *ASME Journal of Vibration and Acoustics*, 120(4), 894–900.

Liu, Y. and Wang, K. W., 1999, A non-dimensional parametric study of enhanced active constrained layer damping treatments, *Journal of Sound and Vibration*, 223(4), 611–644.

Main, J. A., Garcia, E., and Newton, D. V., 1995, Precision position control of piezoelectric actuators using charge feedback, *Journal of Guidance, Control, and Dynamics*, 18(5), 1068–1073.

Miller, D. W. and Crawley, E. F., 1988, Theoretical and experimental investigation of space-realizable inertial actuation for passive and active control, *Journal of Guidance, Control, and Dynamics*, 11(5), 449–458.

Morgan, R. and Wang, K. W., 1998, An integrated active-parametric control approach for active-passive hybrid piezoelectric network with variable resistance, *Journal of Intelligent Material Systems and Structures*, 9(7), 564–573.

Morgan, R. and Wang, K. W., 2000, An active-passive piezoelectric vibration absorber for structural control under harmonic excitations with time-varying frequency, *Proceedings of ASME IMECE,* AD-60, 285–298.

Morgan, R. and Wang, K. W., 2001, A multi-frequency piezoelectric vibration absorber for variable frequency harmonic excitations, *Proceedings of SPIE, Smart Structures and Materials*.

Morgan, R., Wang, K. W., and Tang, J., 2000, Active tuning and coupling enhancement of piezoelectric vibration absorbers for variable-frequency harmonic excitations in multiple degrees of freedom mechanical systems, *Proceedings of SPIE, Smart Structures and Materials*, 3985, 497–509.

Niezrecki, C. and Cudney, H. H., 1994, Improving the power consumption characteristics of piezoelectric actuators, *Journal of Intelligent Material Systems and Structures*, 5(4), 522–529.

Park, C. H. and Baz, A., 1999, Vibration damping and control using active constrained layer damping: A survey, *The Shock and Vibration Digest*, 31, 355–364.

Richard, C., Guyomar, D., Audigier, D., and Ching, G., 1999, Semi-passive damping using continuous switching of a piezoelectric device, *Proceedings of SPIE, Smart Structures and Materials*, 3672, 104–111.

Sirohi, J. and Chopra, I., 1998, Fundamental behavior of piezoceramics sheet actuators, *Proceedings of SPIE, Smart Structures and Materials*, 3329, 626–646.

Slotine, J-J. E. and Li, W., 1991, *Applied Nonlinear Control*, Prentice-Hall, Englewood Cliffs, New Jersey.

Spangler, R. L. and Hall, S. R., 1994, Broadband active structural damping using positive real compensation and piezoelectric simultaneous sensing and actuation, *Smart Materials and Structures*, 3(4), 448–458.

© 2002 by CRC Press LLC

Tang, J., Liu, Y., and Wang, K. W., 2000, Semi-active and active-passive hybrid structural damping treatments via piezoelectric materials, *Shock and Vibration Digest*, 32(3), 189–200.

Tang, J. and Wang, K. W., 1999a, Vibration control of rotationally periodic structures using passive piezoelectric shunt networks and active compensation, *ASME Journal of Vibration and Acoustics*, 121(3), 379–390.

Tang, J. and Wang, K. W., 1999b, Vibration control using piezoelectric material with high order nonlinearity, *Proceedings of ASME IMECE,* AD-59, 149–160.

Tang, J. and Wang, K. W., 2000, High authority and nonlinearity issues in active-passive hybrid piezoelectric networks for structural damping, *Journal of Intelligent Material Systems and Structures,*11(3), 581–591.

Tang, J. and Wang, K. W., 2001, Active-passive hybrid piezoelectric networks for vibration control — comparisons and improvement, *Journal of Smart Materials and Structures,* 10(4), 1–12.

Tang, J., Wang, K. W., and Philen, M., 1999, Sliding mode control of structural vibrations via active-passive hybrid piezoelectric network, *Proceedings of SPIE, Smart Structures and Materials*, 3668, 543–554.

Tsai, M. S., 1998, Active-passive hybrid piezoelectric network-based smart structures for vibration controls, Ph.D. dissertation, The Pennsylvania State University.

Tsai, M. S. and Wang, K. W., 1996, Control of a ring structure with multiple active-passive hybrid piezoelectric networks, *Smart Materials and Structures*, 5(5), 695–703.

Tsai, M. S. and Wang, K. W., 1997, Integrating active-passive hybrid piezoelectric networks with active constrained layer treatments for structural damping, *Proceedings of ASME IMECE DE95*, 13–24.

Tsai, M. S. and Wang, K. W., 1998, A coupled robust control optimization approach for active-passive hybrid adaptive structures, *Proceedings of the 4th European Conference on Smart Structures and Materials*, 57–64.

Tsai, M. S. and Wang, K. W., 1999, On the structural damping characteristics of active piezoelectric actuators with passive shunt, *Journal of Sound and Vibration*, 221(1), 1–22.

Utkin, V. I., 1993, Sliding mode control design principles and applications to electric drives, *IEEE Transactions on Industrial Electronics*, 40(1), 23–36.

Wang, K. W., Lai, J. S., and Yu, W. K., 1996, An energy-based parametric control approach for structural vibration suppression via semi-active piezoelectric networks, *ASME Journal of Vibration and Acoustics*, 118(3), 505–509.

Warkentin, D. J. and Hagood, N. W., 1997, Nonlinear piezoelectric shunting for structural damping, *Proceedings of SPIE, Smart Structures and Materials*, 3041, 747–757.

Wu, S., 1996, Piezoelectric shunts with a parallel R-L circuit for structural damping and vibration control, *Proceedings of SPIE, Smart Structures and Materials*, 2720, 259–269.

Zhou, K., Doyle, J. C., and Glover, K., 1996, *Robust and Optimal Control*, Prentice-Hall, Upper Saddle River, New Jersey.

Zimmerman, D. C. and Inman, D. J., 1990, On the nature of the interaction between structures and proof-mass actuators, *AIAA Journal of Guidance, Control, and Dynamics,* 13(1), 82–88.

© 2002 by CRC Press LLC

16

Vibration Reduction via the Boundary Control Method

16.1 Introduction
16.2 Cantilevered Beam
System Model • Model-Based Boundary Control Law • Experimental Trials
16.3 Axially Moving Web
System Model • Model-Based Boundary Control Law • Experimental Trials
16.4 Flexible Link Robot Arm
System Model • Model-Based Boundary Control Law • Experimental Trials
16.5 Summary

Siddharth P. Nagarkatti
Lucent Technologies

Darren M. Dawson
Clemson University

16.1 Introduction

The dynamics of flexible mechanical systems that require vibration reduction are usually mathematically represented by partial differential equations (PDEs). Specifically, flexible systems are modeled by a PDE that is satisfied over all points within a domain and a set of boundary conditions. These static or dynamic boundary conditions must be satisfied at the points bounding the domain. Traditionally, PDE-based models for flexible systems have been discretized via modal analysis in order to facilitate the control design process. One of the disadvantages of using a discretized model for control design is that the controller could potentially excite the unmodeled, high-order vibration modes neglected during the discretization process (i.e., spillover effects), and thereby, destabilize the closed-loop system. In recent years, distributed control techniques using smart sensors and actuators (e.g., smart structures) have become popular; however, distributed sensing/actuation is often either too expensive to implement or impractical. More recently, boundary controllers have been proposed for use in vibration control applications. In contrast to using the discretized model for the control design, boundary controllers are derived from a PDE-based model and thereby, avoid the harmful spillover effects. In contrast to distributed sensing/actuation control techniques, boundary controllers are applied at the boundaries of the flexible system, and as a result, require fewer sensors/actuators.

In this chapter, we introduce the reader to the concept of applying boundary controllers to mechanical systems. Specifically, we first provide a motivating example to illustrate in a heuristic manner how a boundary controller is derived via the use of a Lyapunov-like approach. To this end, we now examine the following simple flexible mechanical system* described by the PDE

*This PDE model is the so-called wave equation which is often used to model flexible systems such as cables or strings.

© 2002 by CRC Press LLC

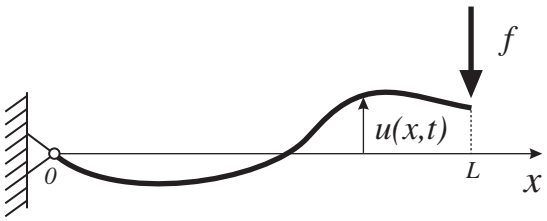


FIGURE 16.1 Schematic diagram of the string system.

$$u_{tt}(x,t) - u_{xx}(x,t) = 0 \tag{16.1}$$

along with the boundary conditions

$$u(0,t) = 0 \quad u_x(L,t) = f(t) \tag{16.2}$$

where $x \in [0, L]$ denotes the independent position variable, t denotes the independent time variable, $u(x,t)$ denotes the displacement at position x for time t, the subscripts x, t represent partial derivatives with respect to x, t, respectively, and $f(t)$ is a control input applied at the boundary position $x = L$. The flexible system described by Equations (16.1) and (16.2) can be schematically represented as shown in Figure 16.1.

The control objective involves designing the control force $f(t)$ to eliminate vibrations throughout the entire system domain using only boundary measurements. Specifically, the aim is to drive $u(x,t) \to 0 \forall x \in [0, L]$ as $t \to \infty$. The underlying philosophy of this control problem is that $f(t)$ should behave as an active *virtual damper* that sucks the energy out of the system. It should be noted that the degree of complexity of this damper-like force is often directly related to the system model. For the linear PDE model of (16.1) and (16.2), only a simple damper in the form of a negative boundary-velocity feedback term at $x = L$ is sufficient to eliminate vibrations throughout the entire system. However, as will be seen in later examples, a more sophisticated boundary control law is often required for more complicated flexible, mechanical system models.

To illustrate the boundary control design procedure, let us consider the following boundary control law for the system described by (16.1) and (16.2):

$$f(t) = -ku_t(L,t) \tag{16.3}$$

where k is a positive, scalar control gain. Note that the above controller is only dependent on measurement of the velocity $u_t(x,t)$ at the boundary position $x = L$. The structure of (16.3) is based on the concept that negative velocity feedback increases the damping in the system. A Lyapunov-like analysis method may be used to illustrate displacement regulation in the system. To this end, the following differentiable, scalar function, composed of the kinetic and potential energy, is defined as follows:

$$V(t) = \tfrac{1}{2}\int_0^L u_t^2(\sigma,t)d\sigma + \tfrac{1}{2}\int_0^L u_\sigma^2(\sigma,t)d\sigma + 2\beta\int_0^L \sigma u_t(\sigma,t)u_\sigma(\sigma,t)d\sigma \tag{16.4}$$

where β is a small, positive weighting constant that is used to ensure that $V(t)$ is non-negative. It should be noted that while the weighting constant β is used in the analysis, it does not appear in

© 2002 by CRC Press LLC

the control law of (16.3). After some algebraic manipulation and integration by parts,* the evaluation of the time derivative of (16.4) along (16.1), (16.2), and (16.3) produces

$$\dot{V}(t) \le -\beta \int_0^L \left(u_t^2(\sigma,t) + u_\sigma^2(\sigma,t) \right) d\sigma \tag{16.5}$$

for a sufficiently small β. Upon application of some standard integral inequalities[1] to (16.4) and (16.5), it can be shown that $u(x,t) \to 0 \forall x \in [0,L]$ as $t \to \infty$; hence, the vibration along the entire domain is driven to zero. We note that the third term of (16.4) is crucial in obtaining the structure of the time derivative of the Lyapunov function given by (16.5); however, the physical interpretation for this term in the Lyapunov function is difficult to explain.

With the above simple example serving to lay the groundwork, we will now focus our attention on the discussion of more complex PDE models often used to describe specific engineering applications. That is, we first present a model-based boundary controller that regulates the out-of-plane vibration of a cantilevered flexible beam with a payload mass attached to the beam free-end. This beam application is then followed by a discussion of a tension and speed setpoint regulating boundary controller for an axially moving web system. Finally, we present a model-based boundary controller that regulates the angular position of a flexible-link robot arm while simultaneously regulating the link vibrations.

16.2 Cantilevered Beam

In many flexible mechanical systems such as flexible link robots, helicopter rotor/blades, space structures, and turbine blades, the flexible element can be modeled as a beam-type structure. The most commonly used beam model that provides a good mathematical representation of the dynamic behavior of the beam is based on the Euler-Bernoulli theory, which is valid when the cross-sectional dimensions of the beam are small in comparison to its length. When deformation owing to shear forces is not inconsequential, a more accurate beam model is provided by the Timoshenko theory, which also incorporates rotary inertial energy. However, owing to its lower order, the Euler-Bernoulli model is often utilized for boundary control design purposes. This section focuses on the problem of stabilizing the displacement of a cantilevered Euler-Bernoulli beam wherein the actuator dynamics at the free-end of the beam have been incorporated into the model. The control law requires shear, shear-rate, and velocity measurements at the free-end of the beam.

16.2.1 System Model

The cantilevered Euler-Bernoulli beam system shown in Figure 16.2 is described by the following PDE:

$$\rho u_{tt}(x,t) + EIu_{xxxx}(x,t) = 0 \tag{16.6}$$

with the following boundary conditions:**

* The detailed mathematical analysis involved in obtaining the final result can be found in Reference 1.

**Given the clamped-end boundary conditions of (16.7), we also know that $u_t(0,t) = u_{xt}(0,t) = 0$.

© 2002 by CRC Press LLC

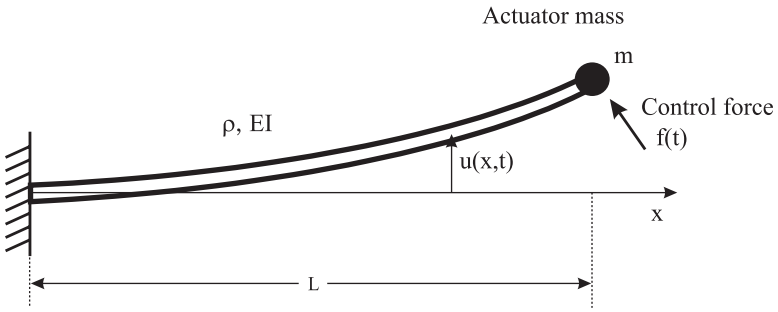


FIGURE 16.2 Schematic diagram of a cantilevered Euler-Bernoulli beam with a free-end payload mass.

$$u(0,t) = u_x(0,t) = u_{xx}(L,t) = 0 \tag{16.7}$$

and

$$mu_{tt}(L,t) - EIu_{xxx}(L,t) = f(t) \tag{16.8}$$

where x, t represent the independent spatial and time variables, the subscripts x, t denote the partial derivatives with respect to x, t, $u(x,t)$ denotes the beam displacement at the position x for time t, ρ is the mass/length of the beam, EI is the bending stiffness of the beam, L is the length of the beam, m represents the payload/actuator mass attached to the free end-point of the beam, and $f(t)$ denotes the boundary control input force.

16.2.2 Model-Based Boundary Control Law

The control objective is to design the boundary control force $f(t)$ that drives the beam displacement $u(x,t)$ to zero with time. Based on the system model, control objectives, and the stability analysis (see Reference 1 or 2 for details), the control force is designed as follows:

$$f(t) = \alpha mu_{xxxt}(L,t) - EIu_{xxx}(L,t) - k_s\eta(t) \tag{16.9}$$

where k_s is a positive control gain and the auxiliary signal $\eta(t)$ is defined as

$$\eta(t) = u_t(L,t) - \alpha u_{xxx}(L,t) \tag{16.10}$$

with α being a positive control gain. A Lyapunov-like analysis,[1] similar to the one given in the motivating example, can be used to show that the system energy (the sum total of the kinetic and potential energy) goes to zero exponentially fast. Standard inequalities can then be invoked to show that $u(x,t) \forall x \in [0, L]$ is bounded by an exponentially decaying envelope; thus, it can easily be established that the beam displacement $u(x,t)$ exponentially decays to zero.

16.2.3 Experimental Trials

A schematic of the experimental setup used in the real-time implementation of the controller is shown in Figure 16.3. A flexible beam 72 cm in length was attached to the top of a support structure with a small metal cylinder weighing 0.3 kg attached to the free end via a strain-gauge shear sensor.

© 2002 by CRC Press LLC

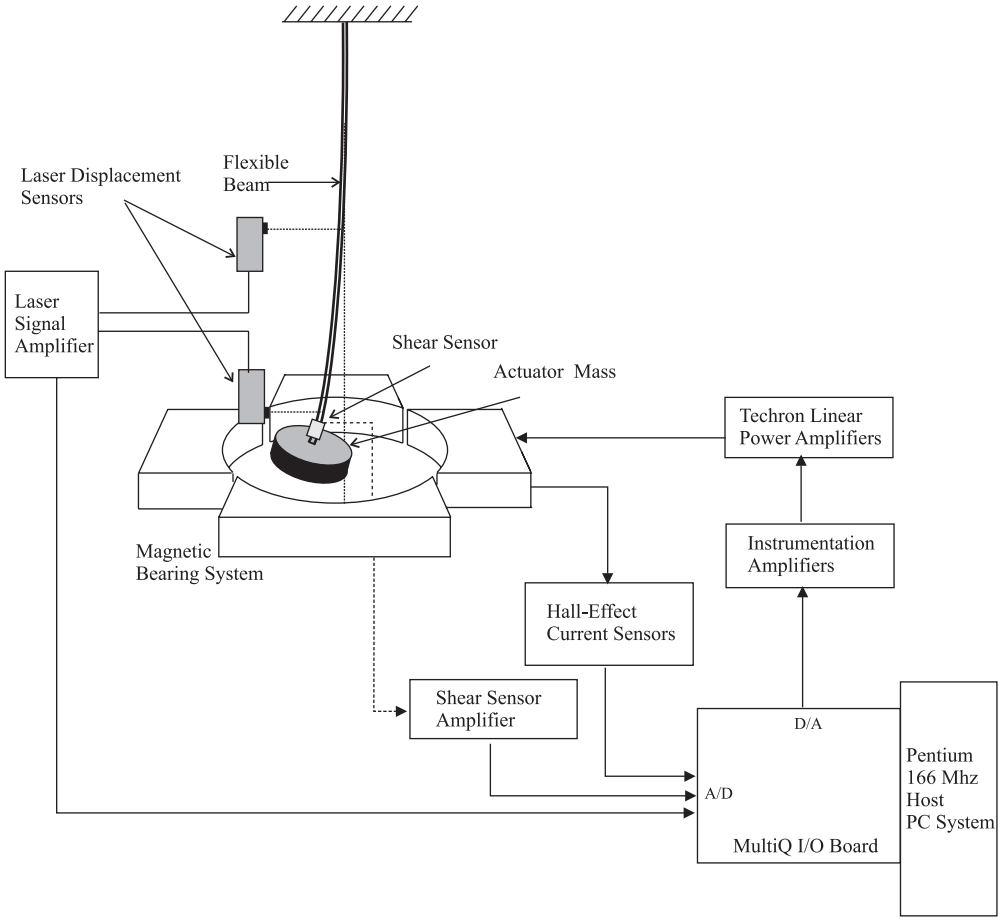


FIGURE 16.3 Schematic diagram of the cantilevered Euler-Bernoulli beam experimental setup.

The beam end-point displacement, $u(L,t)$, was sensed by a laser displacement sensor while another laser displacement sensor was used to monitor the beam mid-point displacement (note that this signal is not used in the control). A pair of electromagnets placed perpendicular to the beam free-end applied the boundary control input force to the payload mass and a custom designed software commutation strategy ensured that the desired input force commanded by the control law was applied to the mass. All time derivatives were calculated using a backwards difference algorithm and a second-order digital filter. The control algorithm was implemented at a 2 kHz sampling frequency on a Pentium 166 MHz PC running QNX (a real-time, micro-kernel-based operating system) under the *Qmotor*[3] graphical user environment.

For this experiment, we imparted an impulse excitation to an arbitrary point on the beam. To ensure a consistent excitation, an impulse hammer was released from a latched position and allowed to strike the beam only once and at the same point each time. The uncontrolled response of the beam's end-point and mid-point displacements when struck by the impulse hammer were recorded.

The response of the model-based boundary controller defined in (16.9) and implemented with three sets of control gains: (i) $k_s = 2.5, \alpha = 1.1$, (ii) $k_s = 5, \alpha = 0.55$, and (iii) $k_s = 7.5, \alpha = 0.38$ is shown in Figure 16.4. It can easily be observed that the model-based controller damps out both the low and high frequency oscillations. For a discussion and comparison of other experiments performed on this system, the reader is referred to Reference 2.

© 2002 by CRC Press LLC

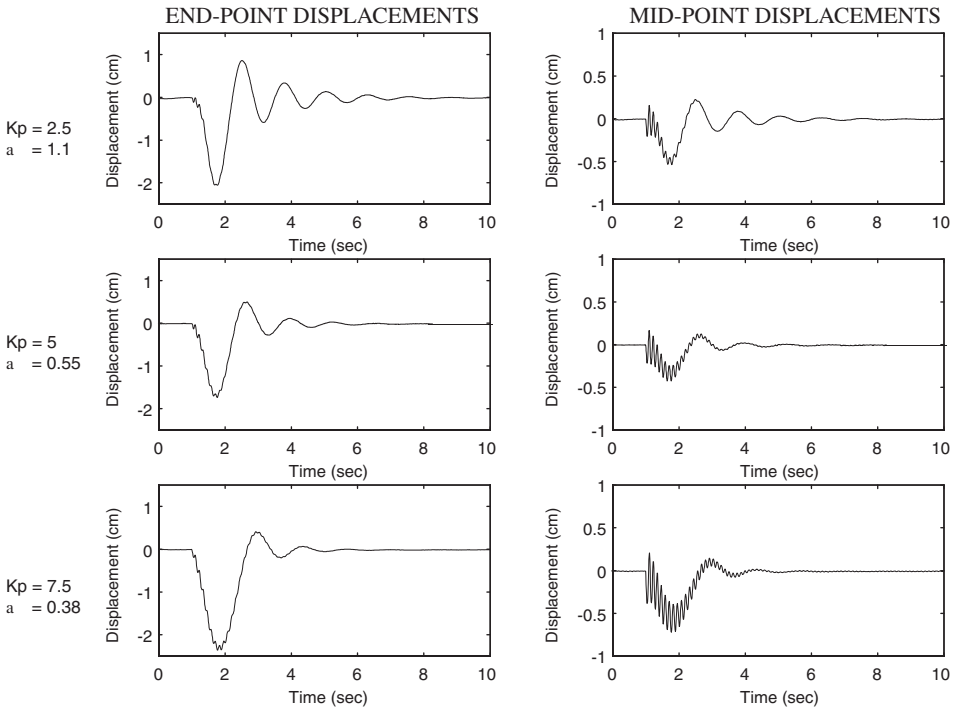


FIGURE 16.4 Cantilevered Euler-Bernoulli beam boundary control response to an impulse excitation.

16.3 Axially Moving Web

In high-speed manufacturing of continuous materials such as optical fibers, textile yarn, paper products, and plastic film, it is imperative to deploy accurate speed and tension control. Typically, rollers are driven to transport these materials through successive operations at varying speeds inherently increasing the risk of controller performance degradation due to tension-varying disturbances. Moreover, tension nonuniformities often lead to product degradation or even failure; hence, precise tension control is essential. Motivated by the need to increase throughput, many manufacturing processes such as those for textile yarn and fibers specify aggressive speed trajectories. Other processes such as label printing demand an aggressive start/stop motion; hence, precise control of such operations relies heavily on coordinated tension and speed control.

16.3.1 System Model

The axially moving web system, depicted in Figure 16.5, consists of a continuous material of length L, axial stiffness EA, and linear density ρ moving between two controlled rollers. Control torques are applied to each roller to regulate the speed of the moving web at a desired setpoint, maintain a constant desired web tension, and damp axial vibration. Based on standard linear web modeling assumptions,[4] the transformed field equation for the axial displacement of the web $u(x,t)$ is given by the following PDE:

$$\rho v_{tt}(x,t) - EAv_{xx}(x,t) = \rho \ddot{y}(t) \tag{16.11}$$

and the boundary conditions

$$v(L,t) = 0 \tag{16.12}$$

© 2002 by CRC Press LLC

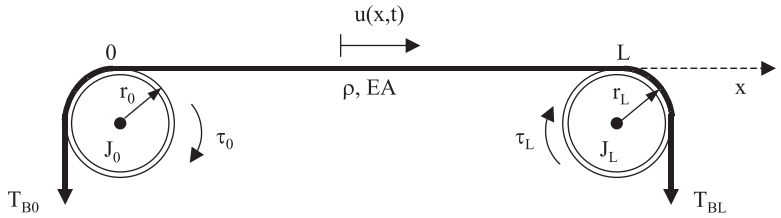


FIGURE 16.5 Schematic diagram of an axially moving web system.

$$mv_{tt}(0,t) - EAv_x(0,t) + P_D = m\ddot{y}(t) - f_0(t) \tag{16.13}$$

where the subscripts denote partial differentiation, the dots over variables denote time differentiation, m is the equivalent mass of the rollers, P_D is the constant desired web tension, and the following transformation* has been used:

$$y(t) = u(L,t) \quad v(x,t) = y(t) - u(x,t) + \frac{P_D}{EA}(x - L). \tag{16.14}$$

While the left roller (at $x = 0$) dynamics are incorporated into (16.13), the right roller dynamics (at $x = L$) are explicitly defined as follows:

$$m\ddot{y}(t) - EAv_x(L,t) + P_D = f_L(t). \tag{16.15}$$

The equivalent force control inputs $f_0(t)$ and $f_L(t)$ in (16.13) and (16.15) are related to the control torques $\tau_0(t)$ and $\tau_L(t)$ as follows:

$$f_0(t) = \frac{\tau_0(t)}{r_0} - T_{B0} \quad f_L(t) = \frac{\tau_L(t)}{r_L} + T_{BL} \tag{16.16}$$

where T_{B0}, T_{BL} denote the web tension in the respective adjacent span and r_0, r_L denote the radii of the rollers.

16.3.2 Model-Based Boundary Control Law

The primary control objective is to design roller torques $\tau_0(t)$ and $\tau_L(t)$ such that the web tension $P(x,t) \forall x \in [0, L]$ is regulated to a constant desired tension setpoint, denoted by P_D, and the web speed $u_t(x,t) \forall x \in [0, L]$ is regulated to a constant desired speed setpoint, denoted by v_d. Based on the system model, control objectives, and the stability analysis,[5] the speed setpoint control law is defined as follows:

$$f_L(t) = P(L,t) + k_p \eta_1(t) + k_i \int_0^t \eta_1(\tau) d\tau \tag{16.17}$$

*The definition of the position/stretch error distribution, denoted by $v(x,t)$, is motivated by the control objective and the stability analysis.

© 2002 by CRC Press LLC

where $P(L,t)$ denotes the web tension at $x = L$, the axial speed setpoint error $\eta_1(t)$ is defined as follows:

$$\eta_1(t) = v_d - \dot{y}(t) = v_d - u_t(L,t)\,, \tag{16.18}$$

v_d denotes the desired web speed setpoint, and k_p, k_i denote constant, positive, scalar gains. The tension setpoint control law is given by

$$f_0(t) = \frac{m\kappa}{EA} P_t(0,t) - P(0,t) + k_{s2}\eta_2(t) + k\eta_1(t) + k_i \int_0^t \eta_1(\tau)d\tau \tag{16.19}$$

where the tension setpoint error $\eta_2(t)$ is defined as

$$\eta_2(t) = v_t(0,t) - \kappa v_x(0,t) = u_t(L,t) - u_t(0,t) - \frac{\kappa}{EA}\left(P_D - P(0,t)\right), \tag{16.20}$$

$P(0,t)$ denotes the tension at $x = 0$, $P_t(0,t)$ denotes the rate change in tension at $x = 0$, and κ, k_{s2} denote positive scalar control gains. After using a Lyapunov-like analysis,[5] similar to one given in the motivating example, it can be shown that $v_t(x,t)$ and $v_x(x,t)$ exponentially decay to zero. Thus, the time derivative of (16.14) yields velocity setpoint regulation (i.e., $u_t(x,t)$ is exponentially driven to v_d). Furthermore, given that the web tension is related to the axial strain as follows: $P(x,t) = EAu_x(x,t)$, the spatial derivative of (16.14) yields tension setpoint regulation (i.e., $EAu_x(x,t)$ is exponentially driven to P_D). For more details, the reader is referred to Ulsoy.[5]

16.3.3 Experimental Trials

The experimental test stand consisted of an elastic rubber belt moving axially over two pulleys actuated by brushed DC motors (see Figure 16.6). Four tension sensors and roller assemblies laterally positioned the moving web and provided measurements of the forward boundary tensions $P(0,t)$ and $P(L,t)$ and the back boundary tensions T_{B0} and T_{BL} used by the controller. The encoders mounted on the motors measured the angular displacements of the rotors. The control algorithm was implemented with a sampling period of 0.5 msec on a Pentium 266 MHz PC running QNX OS under the *Qmotor* graphical user environment.[3]

The objective of the experiment was to regulate the material tension at 8.0 N and move the material according to a smooth, exponentially stepped, desired axial speed setpoint trajectory. In order to mimic real-world industrial processes (such as high-speed label printing), the desired speed of the material was aggressively driven to 0 m/s and back to 0.75 m/s within a time duration of 0.5 sec and was repeated every 10 sec. Process-line disturbances leading to a sudden change in material tension were also simulated by applying a constant reverse torque on the motor at $x = 0$ for a duration of 0.5 sec at 10 sec intervals. Figure 16.7 shows the boundary controller performance.

From the experimental results,[5] it was observed that the maximum speed error at $x = 0$ with the boundary controller was three times smaller than industry standard controllers. With a start-stop speed disturbance, the boundary controller improved tension setpoint regulation by a factor of three over a PI speed controller without tension feedback and a factor of two over a PI speed controller with tension feedback.

© 2002 by CRC Press LLC

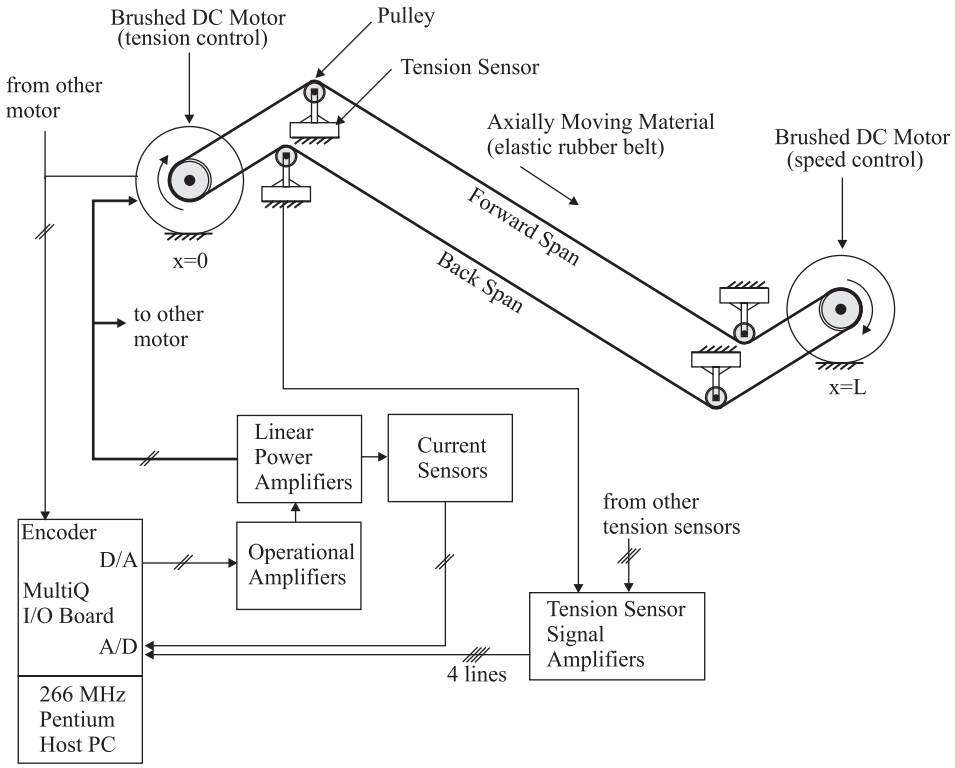


FIGURE 16.6 Schematic diagram of the axially moving web experimental setup.

16.4 Flexible Link Robot Arm

Owing to the prohibitive cost of placing heavy equipment in outer space, most structural designers prefer to utilize lightweight materials in the construction of space-based vehicles, satellites, etc. Indeed, space-based robot manipulators are more likely to be comprised of long links manufactured from lightweight metals or composites. Unfortunately, a major drawback in using lightweight links is the significant presence of deflection and/or vibration problems during position control applications. In this section, we focus our attention on regulating the angular displacement of a flexible link robot manipulator arm described by a nonlinear PDE model while simultaneously reducing the distributed vibration of the link itself.

16.4.1 System Model

The robot system, illustrated in Figure 16.8, is composed of an Euler-Bernoulli beam clamped to a rotating, rigid actuator hub with a payload/actuator mass attached to the free end of the beam. A torque input applied to the hub controls the angular position while a force input that is applied to the free-end mass regulates the beam displacement. The equations of motion of this single flexible-link robot are given by[6]

$$\rho w_{tt}(x,t) + EIw_{xxxx}(x,t) = \rho u(x,t)\dot{q}^2(t) \tag{16.21}$$

and

$$D(t)\ddot{q}(t) + \tfrac{1}{2}\dot{D}(t)\dot{q}(t) + V_m(t)\dot{q}(t) + mu(L,t)w_t(L,t)\dot{q}(t) - EIw_{xx}(0,t) = \tau(t) \tag{16.22}$$

© 2002 by CRC Press LLC

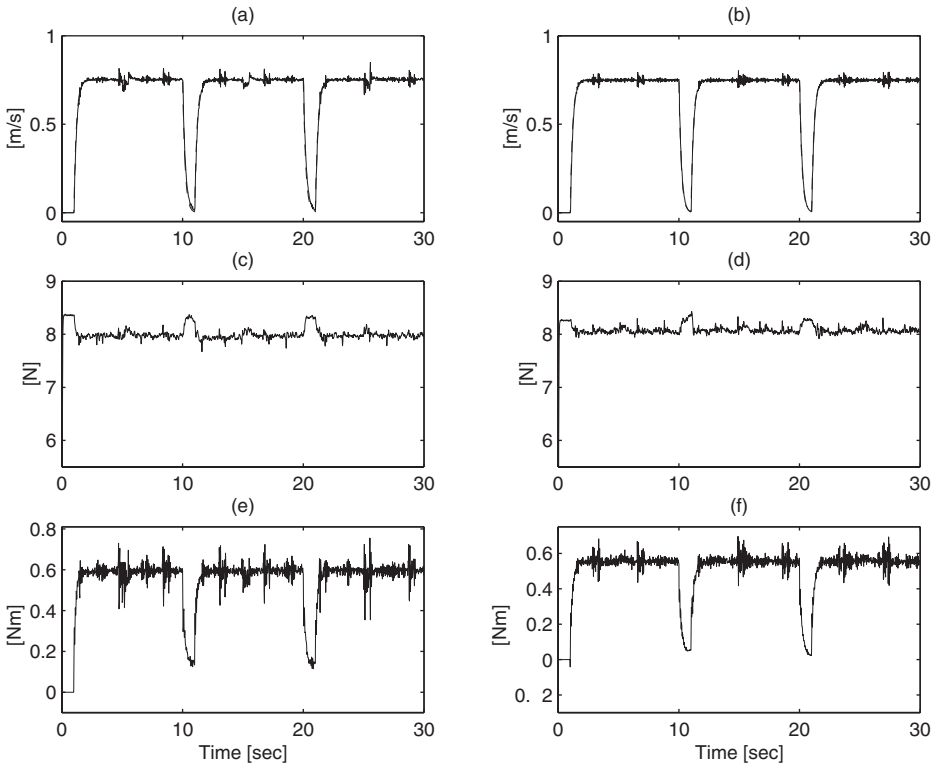


FIGURE 16.7 Axially moving web system control response.

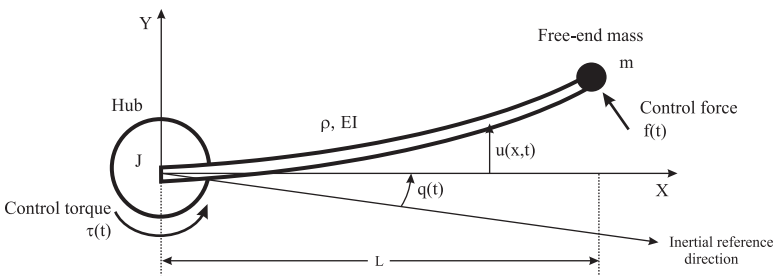


FIGURE 16.8 Schematic diagram of the flexible-link robot arm.

with the following boundary conditions:*

$$u(0,t) = u_x(0,t) = u_{xx}(L,t) = 0 \tag{16.23}$$

and

$$mw_{tt}(L,t) - mu(L,t)\dot{q}^2(t) - EIw_{xxx}(L,t) = f(t) \tag{16.24}$$

*Given the clamped boundary conditions of (16.23), we also know that $u_t(0,t) = u_{xt}(0,t) = 0$.

© 2002 by CRC Press LLC

where the distributed displacement $w(x,t)$ is defined as

$$w(x,t) = u(x,t) + xq(t) \tag{16.25}$$

and $u(x,t)$ denotes the link displacement at position x for time t with respect to the (X, Y) coordinate system that rotates with the hub (see Figure 16.8), $q(t), \dot{q}(t), \ddot{q}(t)$ represent the angular position, velocity, and acceleration of the hub, respectively, with respect to the inertial reference direction, ρ is the mass/length of the link, EI is the bending stiffness of the link, L is the length of the link, m represents a payload/actuator mass attached to the free end of the link, $\tau(t)$ is the control torque input applied to the hub, $f(t)$ denotes the boundary control force input applied to the mass, and the auxiliary functions $D(t), D(t)$, and $V_m(t)$ are defined as follows:

$$D(t) = J + mu^2(L,t) + \rho \int_0^L u^2(\sigma,t)d\sigma \geq J > 0 \tag{16.26}$$

$$\dot{D}(t) = \tfrac{d}{dt} D(t) = 2mu(L,t)u_t(L,t) + 2\rho \int_0^L u(\sigma,t)u_t(\sigma,t)d\sigma \tag{16.27}$$

and

$$V_m(t) = \rho \int_0^L u(\sigma,t)w_t(\sigma,t)d\sigma \tag{16.28}$$

with J denoting the hub's inertia.

16.4.2 Model-Based Boundary Control Law

The control objective is to ensure that: (i) $u(x,t) \to 0 \forall x \in [0,L]$ as $t \to \infty$ with respect to the rotating coordinate system (X, Y) attached to the hub, and (ii) $q(t) \to q_d$ as $t \to \infty$ with respect to the inertial reference direction, where q_d is a desired, constant angular position. To aid the analysis of the link displacement regulation objective, an auxiliary signal $\eta(t)$, is defined as follows:

$$\eta(t) = w_t(L,t) - w_{xxx}(L,t) \tag{16.29}$$

where $w(x,t)$ was defined in (16.25). The angular position regulation objective is quantified via the angular position setpoint error $e(t)$ as follows:

$$e(t) = q(t) - q_d \,. \tag{16.30}$$

Based on the form of (16.29), the stability analysis, and the control objective, the boundary control force is designed as follows:

$$f(t) = mw_{xxxt}(L,t) - EIw_{xxx}(L,t) - k_s\eta(t) \tag{16.31}$$

where k_s is a positive control gain. Similarly, the hub control torque is designed as follows:

$$\tau(t) = -k_v\dot{e}(t) - k_p e(t) + mu(L,t)\dot{q}(t)w_{xxx}(L,t) - \beta\rho L\dot{q}(t)u^2(L,t) \tag{16.32}$$

© 2002 by CRC Press LLC

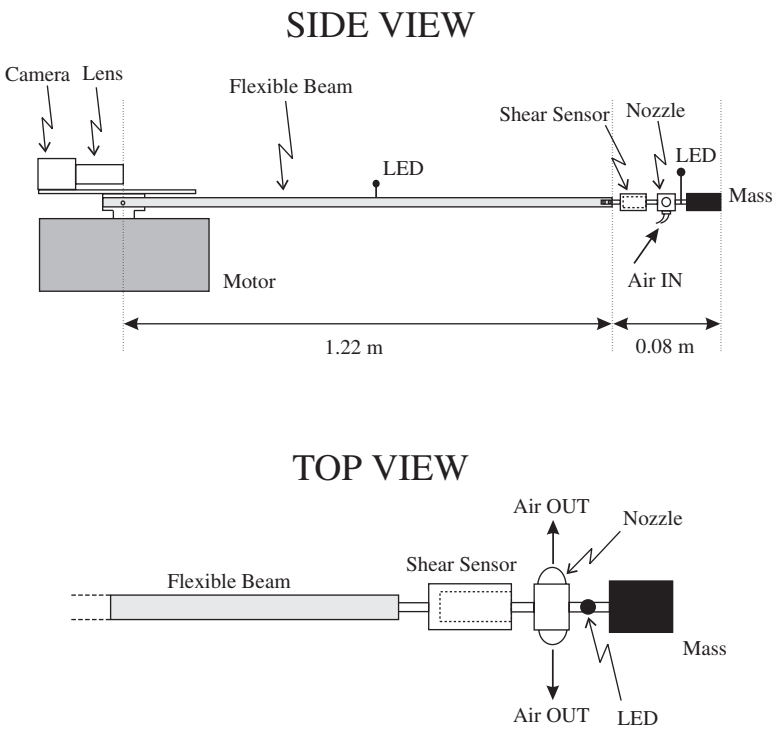


FIGURE 16.9 Schematic diagram of the flexible-link robot arm experimental setup.

where k_v, k_p, β are positive control gains. A Lyapunov-like analysis,[1] similar to one given in the motivating example, can be used to show that the system energy (the sum total of the kinetic and potential energy) asymptotically goes to zero. Standard inequalities can then be invoked to show that the beam displacement $w(x,t) \forall x \in [0, L]$ and the angular position setpoint error $e(t)$ are both asymptotically driven to zero. It is also interesting to note that the boundary control force of (16.31) contains a noncollocated term in the feedback loop (i.e., $\dot{q}(t)$ appears in the definition of $\eta(t)$ through $w_t(L,t)$) while the control torque of (16.32) contains noncollocated feedforward and feedback terms (i.e., the last two terms in (16.32)).

16.4.3 Experimental Trials

A schematic diagram of the experimental setup shown in Figure 16.9 consisted of a flexible aluminum beam attached to the shaft of a switched reluctance motor (SRM) that was used to apply the hub control torque. A lightweight plastic assembly supporting two air nozzles located at the end-point of the beam was used to apply the boundary control force with a 90 psi compressed air supply to high-speed proportional air valves. In addition, a modular line scan camera mounted on the motor shaft and a high luminescence LED mounted at the beam's end-point were used to measure the beam's end-point displacement, $u(L,t)$. A second monitoring LED was placed at $u(L/2,t)$. The signal $u_{xxx}(L,t)$ was measured via the shear force sensor attached to the beam free-end, while an incremental encoder mounted on the motor shaft was utilized to measure the hub angular position, $q(t)$. All time derivatives were calculated using a backwards difference algorithm and a second-order digital filter. The controller was implemented via the *Qmotor* real-time control environment[3] on a QNX platform using a sampling period of 0.5 msec.

The objective of the experiment was to regulate the hub angular position to a desired position of 20° (i.e., $q_d = 0.35$ rad) while driving the link displacement to zero. For comparison purposes,

© 2002 by CRC Press LLC

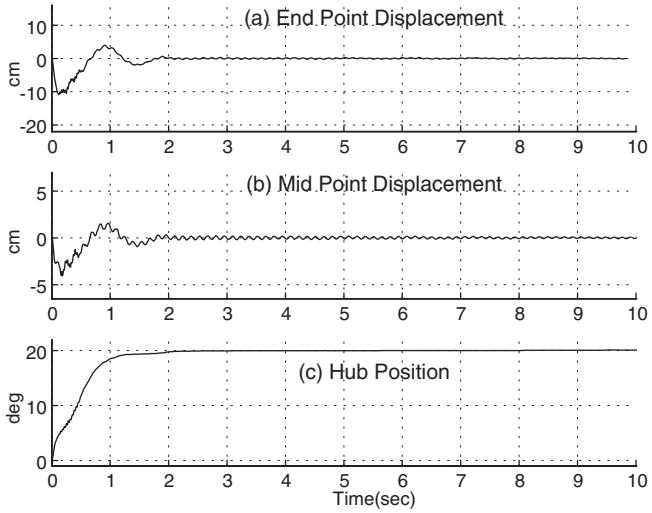


FIGURE 16.10 Flexible-link robot arm boundary control response.

a standard linear control algorithm along the hub angular setpoint error and with the air valves deactivated was implemented. The model-based boundary controller was implemented with $k_s = 3.5, k_v = 20, k_p = 70$, and $\beta = 0.35$ (see Figure 16.10). From the experimental results,[7] it was observed that while the standard linear control law did not damp out beam vibrations even after 10 sec, the model-based boundary controller was able to regulate the distributed beam vibrations within approximately 3 sec as can be seen in Figure 16.10.

16.5 Summary

A number of researchers have investigated vibration control for flexible mechanical systems. For example, Ulsoy[8] demonstrated how control and observation spillover can destabilize[9] the vibration of an axially moving string under state feedback control based on a reduced-order, discretized version of the infinite dimensional model. While Yang and Mote[10] used a transfer function approach to develop asymptotically stabilizing controllers to avoid spillover instabilities, Lee and Mote[11] developed Lyapunov-based, boundary control laws that asymptotically stabilized the vibration of an axially moving string. Other researchers such as Morgul,[12] Shahruz,[13] Joshi and Rahn,[14] and Baicu et al.[15] have designed boundary controllers for strings, overhead gantry crane systems, and flexible cable systems. More recently, Zhang et al.[16] and Nagarkatti et al.[17] designed boundary controllers for nonlinear string models and axially accelerating web systems, respectively.

For cantilevered beams, boundary controllers have been proposed by Morgul,[18,19] Chen et al.[20] Canbolat et al.,[2] and Rahn and Mote,[21] whereas boundary controllers for single flexible link robots were developed by Luo et al.[22-25] and Morgul.[26] Recently, Queiroz et al.[7] designed a boundary control strategy for a nonlinear hybrid model of flexible link robots that asymptotically regulated the link displacement and hub position.

In this chapter, we focused our attention on boundary controllers that regulated vibration for a class of mechanical systems described by second-order (in time) PDE models. It should be noted that the boundary control philosophy has also been applied to other kinds of systems. For example, Byrnes,[27] Krstic,[28] and van Ly[29] designed boundary controllers for the Burgers' equation, which serves as a model for a number of physical problems and is representative of many convection-dominated flow systems. Boundary controllers (e.g., Reference 30) have also been designed for the Kuramoto-Sivashinsky equation that is used to describe a variety of systems such as a plane flame front, flow of thin liquid films on inclined planes, and Alfven drift wave plasmas.

© 2002 by CRC Press LLC

Acknowledgments

The authors wish to thank Bret Costic, Marcio de Queiroz, and Erkan Zergeroglu for their assistance in the formatting of this chapter.

References

1. M. S. de Queiroz, D. M. Dawson, S. P. Nagarkatti, and F. Zhang, *Lyapunov-Based Control of Mechanical Systems*, Birkhauser, Basel, Switzerland, 1999.
2. H. Canbolat, D. Dawson, C. Rahn, and P. Vedagarbha, Boundary control of a cantilevered flexible beam with point-mass dynamics at the free end, *Mechatronics — An International Journal,* 8, 3, 163–186.
3. N. Costescu, D. M. Dawson, and M. Loffler, Qmotor 2.0 — a PC based real-time multitasking graphical control environment, *IEEE Control Systems Magazine Applications*, 18, 1, 13–22, 1999.
4. J. Wickert and C. D. Mote, Jr., Current research on the vibration and stability of axially-moving materials, *Shock and Vibration Digest*, 20, 5, 3–13, 1988.
5. S. P. Nagarkatti, F. Zhang, C. D. Rahn, and D. M. Dawson, Tension and speed setpoint regulation for axially moving materials, *ASME Journal of Dynamic Systems, Measurement, and Control,* 122, 3, 445–453, 2000.
6. J. L. Junkins and Y. Kim, *Introduction to Dynamics and Control of Flexible Structures*, AIAA Education Series, Washington, D.C., 1993.
7. M. S. de Queiroz, D. M. Dawson, M. Agarwal, and F. Zhang, Adaptive nonlinear boundary control of a flexible link robot arm, *IEEE Transactions on Robotics and Automation,* 15, 4, 779–787, 1999.
8. A. G. Ulsoy, Vibration control in rotating or translating elastic systems, *ASME Journal of Dynamic Systems, Measurement, and Control*, 106, 1, 6–14, 1984.
9. M. Balas, Active control of flexible systems, *Journal of Optimization Theory and Applications*, 25, 3, 415–436, 1978.
10. B. Yang and C. D. Mote, Jr., Active vibration control of the axially moving string in the S domain, *ASME Journal of Applied Mechanics*, 58, 189–196, 1991.
11. S. Y. Lee and C. D. Mote, Jr., Vibration control of an axially moving string by boundary control, *ASME Journal of Dynamic Systems, Measurement, and Control*, 118, 66–74, 1996.
12. O. Morgul, A dynamic control law for the wave equation, *Automatica*, 30, 11, 1785–1792, 1994.
13. S. M. Shahruz and L. G. Krishna, Boundary control of a nonlinear string, *Proceedings of the ASME Dynamics Systems and Control Division*, DSC-58, 831–835, 1996.
14. S. Joshi and C. D. Rahn, Position control of a flexible cable gantry crane: theory and experiment, *Proceedings of the American Control Conference*, Seattle, 2820–2824, 1995.
15. C. F. Baicu, C. D. Rahn, and B. D. Nibali, Active boundary control of elastic cables: Theory and experiment, *Journal of Sound and Vibration*, 198, 1, 17–26, 1996.
16. F. Zhang, D. M. Dawson, S. P. Nagarkatti, and C. D. Rahn, Boundary control for a general class of nonlinear actuator-string systems, *Journal of Sound and Vibration*, 229, 1, 113–132, 2000.
17. S. P. Nagarkatti, F. Zhang, B. T. Costic, D. M. Dawson, and C. D. Rahn, Velocity tracking control of an axially accelerating web and actuator system, *Mechanical Systems and Signal Processing*, accepted 2001 (to appear).
18. O. Morgul, Orientation and stabilization of a flexible beam attached to a rigid body: Planar motion, *IEEE Transactions on Automatic Control*, 36, 8, 953–962, 1991.
19. O. Morgul, Dynamic boundary control of a Euler-Bernoulli beam, *IEEE Transactions on Automatic Control*, 37, 5, 639–642, 1992.
20. G. Chen, M. Delfour, A. Krall, and G. Payre, Modeling, stabilization, and control of serially connected beams, *SIAM Journal of Control Optimizations*, 25, 3, 526–546, 1987.
21. C. Rahn and C. Mote, Axial force stabilization of transverse beam vibration, *Proceedings of the ASME Conference on Mechanical Vibration and Noise*, Albuquerque, DE-61, 29–34, 1993.
22. Z. Luo, Direct strain feedback control of flexible robot arms: New theoretical and experimental results, *IEEE Transactions on Automatic Control*, 38, 11, 1610–1622, 1993.

© 2002 by CRC Press LLC

23. Z. Luo and B. Guo, Further theoretical results on direct strain feedback control of flexible robot arms, *IEEE Transactions on Automatic Control,* 40, 4, 747–751, 1995.
24. Z. Luo, N. Kitamura, and B. Guo, Shear force feedback control of flexible robot arms, *IEEE Transactions on Robotics and Automation,* 11, 5, 760–765, 1995.
25. Z. H. Luo and B. Z. Guo, Shear force feedback control of a single-link flexible robot with a revolute joint, *IEEE Transactions on Automatic Control*, 42, 1, 53–65, 1997.
26. O. Morgul, On boundary control of single link flexible robot arms, *IFAC World Congress*, San Francisco, A, 127–132, 1996.
27. C. I. Byrnes, D. S. Gilliam, and V. I. Shubov, Boundary control for a viscous Burgers' equation, *Identification and Control for Systems Governed by Partial Differential Equations*, SIMA, 171–185, 1993.
28. M. Krstic, On global stabilization of Burgers' equation by boundary control, *Systems and Control Letters*, 37, 123–142, 1999.
29. H. van Ly, K. D. Mease, and E. S. Titi, Distributed and boundary control of the viscous Burgers' equation, *Numerical Function Analysis and Optimization*, 18, 143–188, 1997.
30. W. J. Liu and M. Krstic, Stability enhancement by boundary control in the Kuramoto-Sivashinsky equation, *Nonlinear Analysis*, 1999.

© 2002 by CRC Press LLC

III

Dynamics and Control of Aerospace Systems

Robert E. Skelton

© 2002 by CRC Press LLC

17

An Introduction to the Mechanics of Tensegrity Structures

Robert E. Skelton
University of California, San Diego

J. William Helton
University of California, San Diego

Rajesh Adhikari
University of California, San Diego

Jean-Paul Pinaud
University of California, San Diego

Waileung Chan
University of California, San Diego

17.1 Introduction
The Benefits of Tensegrity • Definitions and Examples • The Analyzed Structures • Main Results on Tensegrity Stiffness • Mass vs. Strength

17.2 Planar Tensegrity Structures Efficient in Bending
Bending Rigidity of a Single Tensegrity Unit • Mass Efficiency of the $C2T4$ Class 1 Tensegrity in Bending • Global Bending of a Beam Made from $C2T4$ Units • A Class 1 $C2T4$ Planar Tensegrity in Compression • Summary

17.3 Planar Class K Tensegrity Structures Efficient in Compression
Compressive Properties of the $C4T2$ Class 2 Tensegrity • $C4T2$ Planar Tensegrity in Compression • Self-Similar Structures of the $C4T1$ Type • Stiffness of the $C4T1^i$ Structure • $C4T1^i$ Structure with Elastic Bars and Constant Stiffness • Summary

17.4 Statics of a 3-Bar Tensegrity
Classes of Tensegrity • Existence Conditions for 3-Bar SVD Tensegrity • Load-Deflection Curves and Axial Stiffness as a Function of the Geometrical Parameters • Load-Deflection Curves and Bending Stiffness as a Function of the Geometrical Parameters • Summary of 3-Bar SVD Tensegrity Properties

17.5 Concluding Remarks
Pretension vs. Stiffness Principle • Small Control Energy Principle • Mass vs. Strength • A Challenge for the Future

Appendix 17.A Nonlinear Analysis of Planar Tensegrity

Appendix 17.B Linear Analysis of Planar Tensegrity

Appendix 17.C Derivation of Stiffness of the $C4T1^i$ Structure

© 2002 by CRC Press LLC

Abstract

Tensegrity structures consist of strings (in tension) and bars (in compression). Strings are strong, light, and foldable, so tensegrity structures have the potential to be light but strong and deployable. Pulleys, NiTi wire, or other actuators to selectively tighten some strings on a tensegrity structure can be used to control its shape. This chapter describes some principles we have found to be true in a detailed study of mathematical models of several tensegrity structures. We describe properties of these structures which appear to have a good chance of holding quite generally. We describe how pretensing all strings of a tensegrity makes its shape robust to various loading forces. Another property (proven analytically) asserts that the shape of a tensegrity structure can be changed substantially with little change in the potential energy of the structure. Thus, shape control should be inexpensive. This is in contrast to control of classical structures which require substantial energy to change their shapes. A different aspect of the chapter is the presentation of several tensegrities that are light but extremely strong. The concept of self-similar structures is used to find minimal mass subject to a specified buckling constraint. The stiffness and strength of these structures are determined.

17.1 Introduction

Tensegrity structures are built of bars and strings attached to the ends of the bars. The bars can resist compressive force and the strings cannot. Most bar–string configurations which one might conceive are not in equilibrium, and if actually constructed will collapse to a different shape. Only bar–string configurations in a stable equilibrium will be called **tensegrity structures**.

If well designed, the application of forces to a tensegrity structure will deform it into a slightly different shape in a way that supports the applied forces. Tensegrity structures are very special cases of trusses, where members are assigned special functions. Some members are always in tension and others are always in compression. We will adopt the words "strings" for the tensile members, and "bars" for compressive members. (The different choices of words to describe the tensile members as "strings," "tendons," or "cables" are motivated only by the scale of applications.) A tensegrity structure's bars cannot be attached to each other through joints that impart torques. The end of a bar can be attached to strings or ball jointed to other bars.

The artist Kenneth Snelson[1] (Figure 17.1) built the first tensegrity structure and his artwork was the inspiration for the first author's interest in tensegrity. Buckminster Fuller[2] coined the word "tensegrity" from two words: "tension" and "integrity."

17.1.1 The Benefits of Tensegrity

A large amount of literature on the geometry, artform, and architectural appeal of tensegrity structures exists, but there is little on the dynamics and mechanics of these structures.[2-19] Form-finding results for simple symmetric structures appear[10,20-24] and show an array of stable tensegrity units is connected to yield a large stable system, which can be deployable.[14] Tensegrity structures for civil engineering purposes have been built and described.[25-27] Several reasons are given below why tensegrity structures should receive new attention from mathematicians and engineers, even though the concepts are 50 years old.

17.1.1.1 Tension Stabilizes

A compressive member loses stiffness as it is loaded, whereas a tensile member gains stiffness as it is loaded. Stiffness is lost in two ways in a compressive member. In the absence of any bending moments in the axially loaded members, the forces act exactly through the mass center, the material spreads, increasing the diameter of the center cross section; whereas the tensile member reduces its cross-section under load. In the presence of bending moments due to offsets in the line of force application and the center of mass, the bar becomes softer due to the bending motion. For most materials, the tensile strength of a longitudinal member is larger than its buckling (compressive)

© 2002 by CRC Press LLC

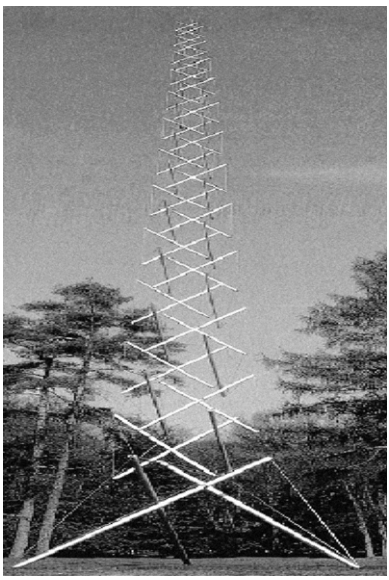


FIGURE 17.1 Snelson's tensegrity structure. (From Connelly, R. and Beck, A., *American Scientist,* 86(2), 143, 1998. Kenneth Snelson, Needle Tower 11, 1969, Kröller Müller Museum. With permission.)

strength. (Obviously, sand, masonary, and unreinforced concrete are exceptions to this rule.) Hence, a large stiffness-to-mass ratio can be achieved by increasing the use of tensile members.

17.1.1.2 Tensegrity Structures are Efficient

It has been known since the middle of the 20th century that continua cannot explain the strength of materials. The geometry of material layout is critical to strength at all scales, from nanoscale biological systems to megascale civil structures. Traditionally, humans have conceived and built structures in rectilinear fashion. Civil structures tend to be made with orthogonal beams, plates, and columns. Orthogonal members are also used in aircraft wings with longerons and spars. However, evidence suggests that this "orthogonal" architecture does not usually yield the minimal mass design for a given set of stiffness properties.[28] Bendsoe and Kikuchi,[29] Jarre,[30] and others have shown that the optimal distribution of mass for specific stiffness objectives tends to be neither a solid mass of material with a fixed external geometry, nor material laid out in orthogonal components. Material is needed only in the essential load paths, not the orthogonal paths of traditional manmade structures. *Tensegrity structures* use longitudinal members arranged in very unusual (and nonorthogonal) patterns to achieve strength with small mass. Another way in which tensegrity systems become mass efficient is with self-similar constructions replacing one tensegrity member by yet another tensegrity structure.

17.1.1.3 Tensegrity Structures are Deployable

Materials of high strength tend to have a very limited displacement capability. Such piezoelectric materials are capable of only a small displacement and "smart" structures using sensors and actuators have only a small displacement capability. Because the compressive members of tensegrity structures are either disjoint or connected with ball joints, large displacement, deployability, and stowage in a compact volume will be immediate virtues of tensegrity structures.[8,11] This feature offers operational and portability advantages. A portable bridge, or a power transmission tower made as a tensegrity structure could be manufactured in the factory, stowed on a truck or helicopter in a small volume, transported to the construction site, and deployed using only winches for erection through cable tension. Erectable temporary shelters could be manufactured, transported, and deployed in a similar manner. Deployable structures in space (complex mechanical structures combined with active control technology) can save launch costs by reducing the mass required, or by eliminating the requirement for assembly by humans.

© 2002 by CRC Press LLC

17.1.1.4 Tensegrity Structures are Easily Tunable

The same deployment technique can also make small adjustments for fine tuning of the loaded structures, or adjustment of a damaged structure. Structures that are designed to allow tuning will be an important feature of next generation mechanical structures, including civil engineering structures.

17.1.1.5 Tensegrity Structures Can be More Reliably Modeled

All members of a tensegrity structure are axially loaded. Perhaps the most promising scientific feature of tensegrity structures is that while the *global* structure bends with external static loads, none of the *individual* members of the tensegrity structure experience bending moments. (In this chapter, we design all compressive members to experience loads well below their Euler buckling loads.) Generally, members that experience deformation in two or three dimensions are much harder to model than members that experience deformation in only one dimension. The Euler buckling load of a compressive member is from a bending instability calculation, and it is known in practice to be very unreliable. That is, the actual buckling load measured from the test data has a larger variation and is not as predictable as the tensile strength. Hence, increased use of tensile members is expected to yield more robust models and more efficient structures. More reliable models can be expected for axially loaded members compared to models for members in bending.[31]

17.1.1.6 Tensegrity Structures Facilitate High Precision Control

Structures that can be more precisely modeled can be more precisely controlled. Hence, tensegrity structures might open the door to quantum leaps in the precision of controlled structures. The architecture (geometry) dictates the mathematical properties and, hence, these mathematical results easily scale from the nanoscale to the megascale, from applications in microsurgery to antennas, to aircraft wings, and to robotic manipulators.

17.1.1.7 Tensegrity is a Paradigm that Promotes the Integration of Structure and Control Disciplines

A given tensile or compressive member of a tensegrity structure can serve multiple functions. It can simultaneously be a load-carrying member of the structure, a sensor (measuring tension or length), an actuator (such as nickel-titanium wire), a thermal insulator, or an electrical conductor. In other words, by proper choice of materials and geometry, a grand challenge awaits the tensegrity designer: How to control the electrical, thermal, and mechanical energy in a material or structure? For example, smart tensegrity wings could use shape control to maneuver the aircraft or to optimize the air foil as a function of flight condition, without the use of hinged surfaces. Tensegrity structures provide a promising paradigm for integrating structure and control design.

17.1.1.8 Tensegrity Structures are Motivated from Biology

Figure 17.2 shows a rendition of a spider fiber, where amino acids of two types have formed hard β–pleated sheets that can take compression, and thin strands that take tension.[32,33] The β–pleated sheets are discontinuous and the tension members form a continuous network. Hence, the nano-structure of the spider fiber is a tensegrity structure. Nature's endorsement of tensegrity structures warrants our attention because per unit mass, spider fiber is the strongest natural fiber.

Articles by Ingber[7,34,35] argue that tensegrity is the fundamental building architecture of life. His observations come from experiments in cell biology, where prestressed truss structures of the tensegrity type have been observed in cells. It is encouraging to see the similarities in structural building blocks over a wide range of scales. If tensegrity is nature's preferred building architecture, modern analytical and computational capabilities of tensegrity could make the same incredible efficiency possessed by natural systems transferrable to manmade systems, from the nano- to the megascale. This is a grand design challenge, to develop scientific procedures to create smart tensegrity structures that can regulate the flow of thermal, mechanical, and electrical energy in a material system by proper choice of materials, geometry, and controls. This chapter contributes to this cause by exploring the mechanical properties of simple tensegrity structures.

© 2002 by CRC Press LLC

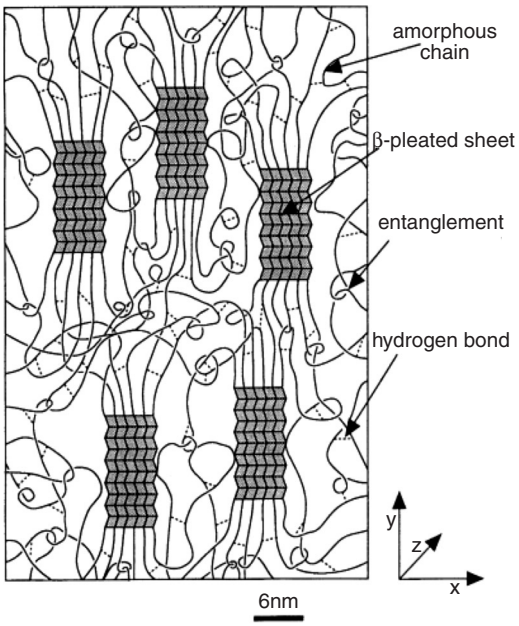


FIGURE 17.2 Structure of the Spider Fiber. (From Termonia, Y., *Macromolecules*, 27, 7378–7381, 1994. Reprinted with permission from the American Chemical Society.)

The remainder of the introduction describes the main results of this chapter. We start with formal definitions and then turn to results.

17.1.2 Definitions and Examples

This is an introduction to the mechanics of a class of prestressed structural systems that are composed only of axially loaded members. We need a couple of definitions to describe tensegrity scientifically.

Definition 17.1 *We say that the geometry of a material system is in a stable equilibrium if all particles in the material system return to this geometry, as time goes to infinity, starting from any initial position arbitrarily close to this geometry.*

In general, a variety of boundary conditions may be imposed, to distinguish, for example, between bridges and space structures. But, for the purposes of this chapter we characterize only the material system with free–free boundary conditions, as for a space structure. We will herein characterize the bars as rigid bodies and the strings as one-dimensional elastic bodies. Hence, a material system is in equilibrium if the nodal points of the bars in the system are in equilibrium.

Definition 17.2 *A* Class k tensegrity structure *is a stable equilibrium of axially loaded elements, with a maximum of k compressive members connected at the node(s).*

Fact 17.1 Class k tensegrity structures *must have tension members*.

Fact 17.1 follows from the requirement to have a stable equilibrium.

Fact 17.2 *Kenneth Snelson's structures of which* Figure *17.1 is an example are all* Class 1 tensegrity structures, *using Definition 17.1. Buckminster Fuller coined the word tensegrity to imply a connected set of tension members and a disconnected set of compression members. This fits our "Class 1" definition.*

A Class 1 tensegrity structure has a connected network of members in tension, while the network of compressive members is disconnected. To illustrate these various definitions, Figure 17.3(a)

© 2002 by CRC Press LLC

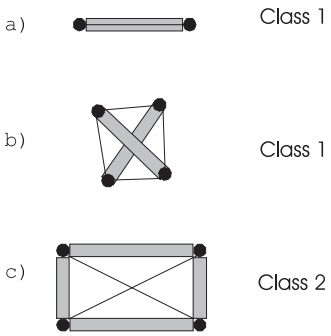


FIGURE 17.3 Tensegrity structures.

illustrates the simplest tensegrity structure, composed of one bar and one string in tension. Thin lines are strings and shaded bars are compressive members. Figure 17.3(b) describes the next simplest arrangement, with two bars. Figure 17.3(c) is a Class 2 tensegrity structure because two bars are connected at the nodes. Figure 17.3(c) represents a Class 2 tensegrity in the plane. However, as a three-dimensional structure, it is not a tensegrity structure because the equilibrium is unstable (the tensegrity definition requires a stable equilibrium).

From these definitions, the existence of a tensegrity structure having a specified geometry reduces to the question of whether there exist finite tensions that can be applied to the tensile members to hold the system in that geometry, in a stable equilibrium.

We have illustrated that the geometry of the nodal points and the connections cannot be arbitrarily specified. The role that geometry plays in the mechanical properties of tensegrity structures is the focus of this chapter.

The planar tensegrity examples shown follow a naming convention that describes the number of compressive members and tension members. The number of compressive members is associated with the letter C, while the number of tensile members is associated with T. For example, a structure that contains two compressive members and four tension members is called a *C*2*T*4 tensegrity.

17.1.3 The Analyzed Structures

The basic examples we analyzed are the structures shown in Figure 17.4, where thin lines are the strings and the thick lines are bars. Also, we analyzed various structures built from these basic structural units. Each structure was analyzed under several types of loading. In particular, the top and bottom loads indicated on the *C*2*T*4 structure point in opposite directions, thereby resulting in bending. We also analyzed a *C*2*T*4 structure with top and bottom loads pointing in the same direction, that is, a compressive situation. The *C*4*T*2 structure of Figure 17.4(b) reduces to a *C*4*T*1 structure when the horizontal string is absent. The mass and stiffness properties of such structures will be of interest under compressive loads, *F*, as shown. The 3-bar SVD (defined in 17.4.1) was studied under two types of loading: axial and lateral. Axial loading is compressive while lateral loading results in bending.

17.1.4 Main Results on Tensegrity Stiffness

A reasonable test of any tensegrity structure is to apply several forces each of magnitude F at several places and plot how some measure of its shape changes. We call the plot of $dF/dshape$ vs. F a stiffness profile of the structure. The chapter analyzes stiffness profiles of a variety of tensegrity structures. We paid special attention to the role of pretension set in the strings of the tensegrity. While we have not done an exhaustive study, there are properties common to these examples which we now describe. How well these properties extend to all tensegrity structures remains to be seen.

© 2002 by CRC Press LLC

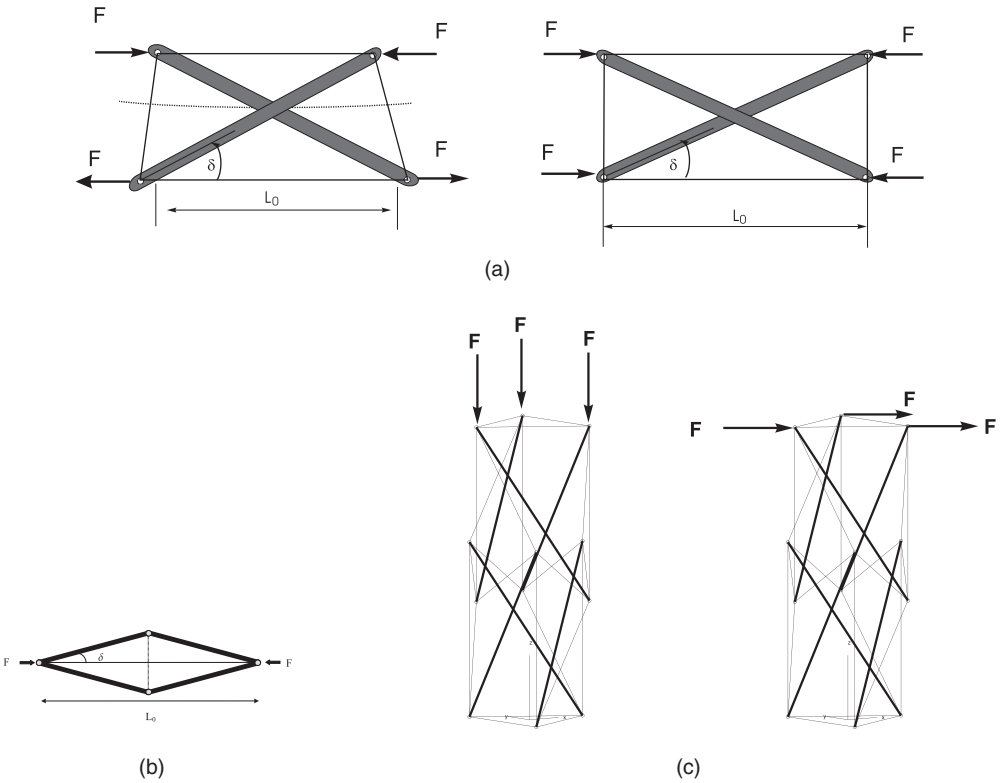


FIGURE 17.4 Tensegrities studied in this chapter (not to scale), (a) *C2T*4 bending loads (left) and compressive loads (right), (b) *C*4*T*2, and (c) 3-bar SVD axial loads (left) and lateral loads (right).

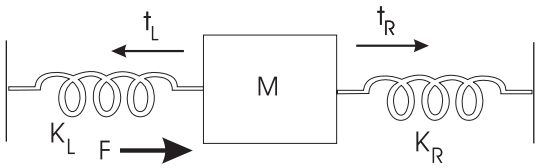


FIGURE 17.5 Mass–spring system.

However, laying out the principles here is an essential first step to discovering those universal properties that do exist.

The following example with masses and springs prepares us for two basic principles which we have observed in the tensegrity paradigm.

17.1.4.1 Basic Principle 1: Robustness from Pretension

As a parable to illustrate this phenomenon, we resort to the simple example of a mass attached to two bungy cords. (See Figure 17.5.)

Here K_L, K_R are the spring constants, F is an external force pushing right on the mass, and t_L, t_R are tensions in the bungy cords when $F = 0$. The bungy cords have the property that when they are shorter than their rest length they become inactive. If we set any positive pretensions t_L, t_R, there is a corresponding equilibrium configuration, and we shall be concerned with how the shape of this configuration changes as force F is applied. Shape is a peculiar word to use here when we mean position of the mass, but it forshadows discussions about very general tensegrity structures. The effect of the stiffness of the structure is seen in Figure 17.6.

© 2002 by CRC Press LLC

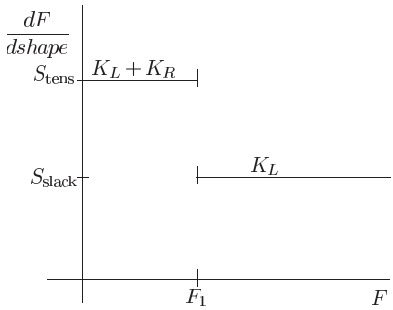


FIGURE 17.6 Mass–spring system stiffness profile.

There are two key quantities in this graph which we see repeatedly in tensegrity structures. The first is the critical value F_1 where the stiffness drops. It is easy to see that F_1 equals the value of F at which the right cord goes slack. Thus, F_1 increases with the pretension in the right cord. The second key parameter in this figure is the size of the jump as measured by the ratio

$$r := \frac{S_{tens}}{S_{slack}}$$

When $r = 1$, the stiffness plot is a straight horizontal line with no discontinuity. Therefore, the amount of pretension affects the value of F_1, but has no influence on the stiffness. One can also notice that increasing the value of r increases the size of the jump. What determines the size of r is just the ratio κ of the spring constants $\kappa := K_R/K_L$, since $r = 1 + \kappa$, indeed r is an increasing function of κ

$$r \cong \infty \quad \text{if} \quad \kappa \cong \infty.$$

Of course, pretension is impossible if $K_R = 0$. Pretension increases F_1 and, hence, allows us to stay in the high stiffness regime given by S_{tens}, over a larger range of applied external force F.

17.1.4.2 Robustness from Pretension Principle for Tensegrity Structures

Pretension is known in the structures community as a method of increasing the load-bearing capacity of a structure through the use of strings that are stretched to a desired tension. This allows the structure to support greater loads without as much deflection as compared to a structure without any pretension.

For a tensegrity structure, the role of pretension is monumental. For example, in the analysis of the planar tensegrity structure, the slackening of a string results in dramatic nonlinear changes in the bending rigidity. Increasing the pretension allows for greater bending loads to be carried by the structure while still exhibiting near constant bending rigidity. In other words, the slackening of a string occurs for a larger external load. We can loosely describe this as a robustness property, in that the structure can be designed with a certain pretension to accomodate uncertainties in the loading (bending) environment. Not only does pretension have a consequence for these mechanical properties, but also for the so-called prestressable problem, which is left for the statics problem. The prestressable problem involves finding a geometry which can sustain its shape without external forces being applied and with all strings in tension.[12,20]

17.1.4.2.1 Tensegrity Structures in Bending

What we find is that bending stiffness profiles for all examples we study have levels S_{tens} when all strings are in tension, S_{slack1} when one string is slack, and then other levels as other strings go slack or as strong forces push the structure into radically different shapes (see Figure 17.7). These very high force regimes can be very complicated and so we do not analyze them. Loose motivation for

© 2002 by CRC Press LLC

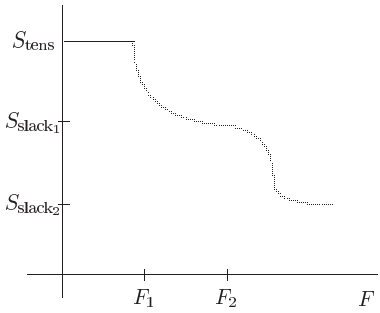


FIGURE 17.7 Gedanken stiffness profile.

the form of a bending stiffness profile curve was given in the mass and two bungy cord example, in which case we had two stiffness levels.

One can imagine a more complicated tensegrity geometry that will possibly yield many stiffness levels. This intuition arises from the possibility that multiple strings can become slack depending on the directions and magnitudes of the loading environment. One hypothetical situation is shown in Figure 17.7 where three levels are obtained. All tensegrity examples in the chapter have bending stiffness profiles of this form, at least until the force F radically distorts the figure. The specific profile is heavily influenced by the geometry of the tensegrity structure as well as of the stiffness of the strings, K_{string}, and bars, K_{bar}. In particular, the ratio

$$K := \frac{K_{string}}{K_{bar}}$$

is an informative parameter.

General properties common to our bending examples are

1. When no string is slack, the geometry of a tensegrity and the materials used have much more effect on its stiffness than the amount of pretension in its strings.
2. As long as all strings are in tension (that is, $F < F_1$), stiffness has little dependence on F or on the amount of pretension in the strings.
3. A larger pretension in the strings produces a larger F_1.
4. As F exceeds F_1 the stiffness quickly drops.
5. The ratio

$$r_1 := \frac{S_{tens}}{S_{slack}}$$

is an increasing function of K. Moreover, $r_1 \to \infty$ as $K \to \infty$ (if the bars are flabby, the structure is flabby once a string goes slack). Similar parameters, r_2, can be defined for each change of stiffness.

Examples in this chapter that substantiate these principles are the stiffness profile of *C2T*4 under bending loads as shown in Figure 17.12. Also, the laterally loaded 3-bar SVD tensegrity shows the same behavior with respect to the above principles, Figure 17.54 and Figure 17.55.

17.1.4.2.2 Tensegrity Structures in Compression

For compressive loads, the relationships between stiffness, pretension, and force do not always obey the simple principles listed above. In fact, we see three qualitatively different stiffness profiles in our compression loading studies. We now summarize these three behavior patterns.

© 2002 by CRC Press LLC

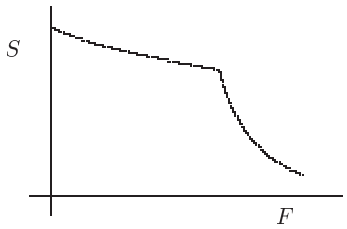


FIGURE 17.8 Stiffness profile for *C*4*T*2 in compression.

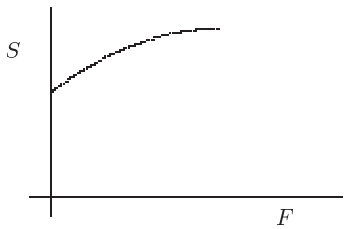


FIGURE 17.9 Stiffness profile of 3-bar SVD in compression.

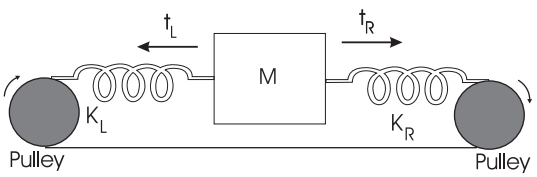


FIGURE 17.10 Mass–spring control system.

The *C*2*T*4 planar tensegrity exhibits the pretension robustness properties of Principles I, II, III, as shown in Figure 17.6. The pretension tends to prevent slack strings.

The *C*4*T*2 structure has a stiffness profile of the form in Figure 17.8. Only in the *C*4*T*1 and *C*4*T*2 examples does stiffness immediately start to fall as we begin to apply a load.

The axially loaded 3-bar-SVD, the stiffness profile even for small forces, is seriously affected by the amount of pretension in the structure. Rather than stiffness being constant for $F < F_1$ as is the case with bending, we see in Figure 17.9 that stiffness increases with F for small and moderate forces. The qualitative form of the stiffness profile is shown in Figure 17.9. We have not systematically analyzed the role of the stiffness ratio K in compression situations.

17.1.4.2.3 Summary

Except for the *C*4*T*2 compression situation, when a load is applied to a tensegrity structure the stiffness is essentially constant as the loading force increases unless a string goes slack.

17.1.4.3 Basic Principle 2: Changing Shape with Small Control Energy

We begin our discussion not with a tensegrity structure, but with an analogy. Imagine, as in Figure 17.10, that the rigid boundary conditions of Figure 17.5 become frictionless pulleys. Suppose we are able to actuate the pulleys and we wish to move the mass to the right, we can turn each pulley clockwise. The pretension can be large and yet very small control torques are needed to change the position of the nodal mass.

© 2002 by CRC Press LLC

Tensegrity structures, even very complicated ones, can be actuated by placing pulleys at the nodes (ends of bars) and running the end of each string through a pulley. Thus, we think of two pulleys being associated with each string and the rotation of the pulleys can be used to shorten or loosen the string. The mass–spring example foreshadows the fact that even in tensegrity structures, shape changes (moving nodes changes the shape) can be achieved with little change in the potential energy of the system.

17.1.5 Mass vs. Strength

The chapter also considers the issue of the strength vs. mass of tensegrity structures. We find our planar examples to be very informative. We shall consider two types of strength. They are the size of the bending forces and the size of compressive forces required to break the object.

First, in 17.2 we study the ratio of bending strength to mass. We compare this for our $C2T4$ unit to a solid rectangular beam of the same mass. As expected, reasonably constructed $C2T4$ units will be stronger. We do this comparison to a rectangular beam by way of illustrating the mass vs. strength question, because a thorough study would compare tensegrity structures to various kinds of trusses and would require a very long chapter.

We analyze compression stiffness of the $C2T4$ tensegrity. The $C2T4$ has worse strength under compression than a solid rectangular bar. We analyze the compression stiffness of $C4T2$ and $C4T1$ structures and use self-similar concepts to reduce mass, while constraining stiffness to a desired value. The $C4T1$ structure has a better compression strength-to-mass ratio than a solid bar when $\delta < 29°$. The $C4T1$, while strong (not easily broken), may not have an extremely high stiffness.

17.1.5.1 A 2D Beam Composed of Tensegrity Units

After analyzing one $C2T4$ tensegrity unit, we lay n of them side by side to form a beam. We derive in 17.2.3 that the Euler buckling formula for a beam adapts directly to this case. From this we conclude that the strength of the beam under compression is determined primarily by the bending rigidity $(EI)_n$ of each of its units. In principle, one can build beams with arbitrarily great bending strength. In practice this requires more study. Thus, the favorable bending properties found for $C2T4$ bode well for beams made with tensegrity units.

17.1.5.2 A 2D Tensegrity Column

In 17.3 we take the $C4T2$ structure in Figure 17.4(b) and replace each bar with a smaller $C4T2$ structure, then we replace each bar of this new structure with a yet smaller $C4T2$ structure. In principle, such a self-similar construction can be repeated to any level. Assuming that the strings do not fail and have significantly less mass than the bars, we find that the compression strength increases without bound if we keep the mass of the total bars constant. This completely ignores the geometrical fact that as we go to finer and finer levels in the fractal construction, the bars increasingly overlap. Thus, at least in theory, we have a class of tensegrity structures with unlimited compression strength to mass ratio. Further issues of robustness to lateral and bending forces would have to be investigated to insure practicality of such structures. However, our dramatic findings based on a pure compression analysis are intriguing. The self-similar concept can be extended to the third dimension in order to design a realistic structure that could be implemented in a column.

The chapter is arranged as follows: Section 17.2 analyzes a very simple planar tensegrity structure to show an efficient structure in bending; Section 17.3 analyzes a planar tensegrity structure efficient in compression; Section 17.4 defines a shell class of tensegrity structures and examines several members of this class; Section 17.5 offers conclusions and future work. The appendices explain nonlinear and linear analysis of planar tensegrity.

© 2002 by CRC Press LLC

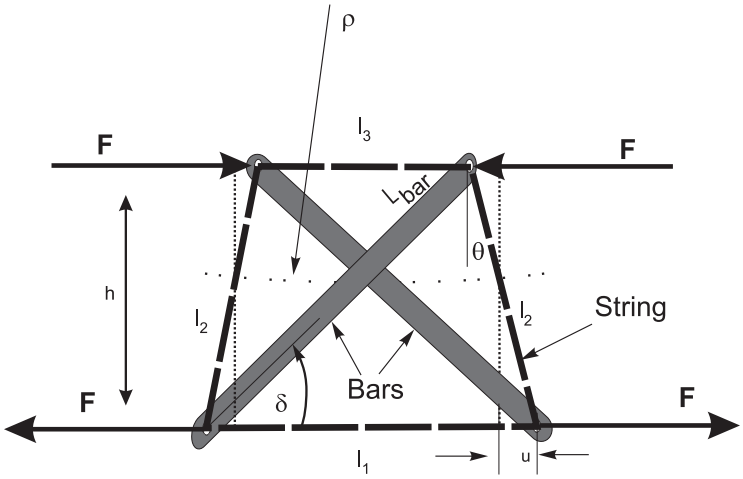


FIGURE 17.11 Planar one-stage tensegrity unit under pure bending.

17.2 Planar Tensegrity Structures Efficient in Bending

In this section, we examine the bending rigidity of a single tensegrity unit, a planar tensegrity model under pure bending as shown in Figure 17.11, where thin lines are the four strings and the two thick lines are bars. Because the structure in Figure 17.11 has two compressive and four tensile members, we refer to it as a *C2T*4 structure.

17.2.1 Bending Rigidity of a Single Tensegrity Unit

To arrive at a definition of bending stiffness suitable to *C2T*4, note that the moment M acting on the section is given by

$$M = FL_{bar} \sin \delta, \tag{17.1}$$

where F is the magnitude of the external force, L_{bar} is the length of the bar, and δ is the angle that the bars make with strings in the deformed state, as shown in Figure 17.11.

In Figure 17.11, ρ is the radius of curvature of the tensegrity unit under bending deformation. It can be shown from Figure 17.11 that

$$\rho = \left(\frac{L_{bar}}{2}\right)^2 \cos\delta \sin\delta \frac{1}{u}, \quad u = \frac{1}{2} L_{bar} \sin\delta \tan\theta. \tag{17.2}$$

The *bending rigidity* is defined by $EI = M\rho$. Hence,

$$EI = FL_{bar} \sin\delta\rho = FL_{bar} \sin^2\delta \left(\frac{L_{bar}}{2}\right)^2 \cos\delta \frac{1}{u}. \tag{17.3}$$

where EI is the equivalent bending rigidity of the planar one-stage tensegrity unit and u is the nodal displacement. The evaluation of the bending rigidity of the planar unit requires the evaluation of u, which will follow under various hypotheses. The bending rigidity will later be obtained by substituting u in (17.3).

© 2002 by CRC Press LLC

17.2.1.1 Effective Bending Rigidity with Pretension

In the absence of external forces **f**, let $\mathbf{A}_0$ be the matrix defined in Appendix 17.A in terms of the initial prestressed geometry, and let $\mathbf{t}_0$ be the initial pretension applied on the members of the tensegrity. Then,

$$\mathbf{A_0 t_0} = \mathbf{0}, \quad \mathbf{t}_0^T = [\mathbf{t}_{0_bars}\ \mathbf{t}_{0_strings}], \quad \mathbf{t}_{0_strings} \geq 0. \tag{17.4}$$

For a nontrivial solution of Equation (17.4), $\mathbf{A}_0$ must have a right null space. Furthermore, the elements of $\mathbf{t}_0$ obtained by solving Equation (17.4) must be such that the strings are always in tension, where $\mathbf{t}_{0\text{-}strings} \geq 0$ will be used to denote that each element of the vector is nonnegative. For this particular example of planar tensegrity, the null space of $\mathbf{A}_0$ is only one dimensional. $\mathbf{t}_0$ always exists, satisfying (17.4), and $\mathbf{t}_0$ can be scaled by any arbitrary positive scalar multiplier. However, the requirement of a stable equilibrium in the tensegrity definition places one additional constraint to the conditions (17.4); the geometry from which $\mathbf{A_0}$ is constructed must be a stable equilibrium.

In the following discussions, E_s, $(EA)_s$, and A_s denote the Young's modulus of elasticity, the axial rigidity and the cross-sectional area of the strings, respectively, whereas E_b, $(EA)_b$, and A_b, denote those of the bars, respectively. $(EI)_b$ denotes the bending rigidity of the bars.

The equations of the static equilibrium and the bending rigidity of the tensegrity unit are nonlinear functions of the geometry δ, the pretension $\mathbf{t}_0$, the external force **F**, and the stiffnesses of the strings and bars. In this case, the nodal displacement u is obtained by solving nonlinear equations of the static equilibrium (see Appendix 17.A for the underlying assumptions and for a detailed derivation)

$$\mathbf{A\ (u)\ KA\ (u)}^T\ \mathbf{u} = \mathbf{F} - \mathbf{A\ (u) t_0} \tag{17.5}$$

Also, $\mathbf{t}_0$ is the pretension applied in the strings, **K** is a diagonal matrix containing axial stiffness of each member, i.e., $K_{ii} = (EA)_i/L_i$, where L_i is the length of the i-th member; **u** represents small nodal displacements in the neighborhood of equilibrium caused by small increments in the external forces. The standard Newton–Raphson method is applied to solve (17.5) at each incremental load step $\mathbf{F}_k = \mathbf{F}_{k-1} + \Delta\mathbf{F}$. Matrix $\mathbf{A}(\mathbf{u}_k)$ is updated at each iteration until a convergent solution for $\mathbf{u}_k$ is found.

Figure 17.12 depicts EI as a function of the angle δ, pretension of the top string, and the rigidity ratio K which is defined as the ratio of the axial rigidity of the strings to the axial rigidity of the bars, i.e., $K = (EA)_s/(EA)_b$. The pretension is measured as a function of the prestrain in the top string Σ_0. In obtaining Figure 17.12, the bars were assumed to be equal in diameter and the strings were also assumed to be of equal diameter. Both the bars as well as the strings were assumed to be made of steel for which Young's modulus of elasticity E was taken to be $2.06 \times 10^{11} N/m^2$, and the yield strength of the steel σ_y was taken to be $6.90 \times 10^8 N/m^2$. In Figure 17.12, EI is plotted against the ratio of the external load F to the yield force of the string. The yield force of the string is defined as the force that causes the strings to reach the elastic limit. The yield force for the strings is computed as

$$\text{Yield force of string} = \sigma_y A_s,$$

where σ_y is the yield strength and A_s is the cross-sectional area of the string. The external force F was gradually increased until at least one of the strings yielded.

The following conclusions can be drawn from Figure 17.12:

1. Figure 17.12(a) suggests that the bending rigidity EI of a tensegrity unit with all taut strings increases with an increase in the angle δ, up to a maximum at $\delta = 90°$.
2. Maximum bending rigidity EI is obtained when none of the strings is slack, and the EI is approximately constant for any external force until one of the strings go slack.

© 2002 by CRC Press LLC

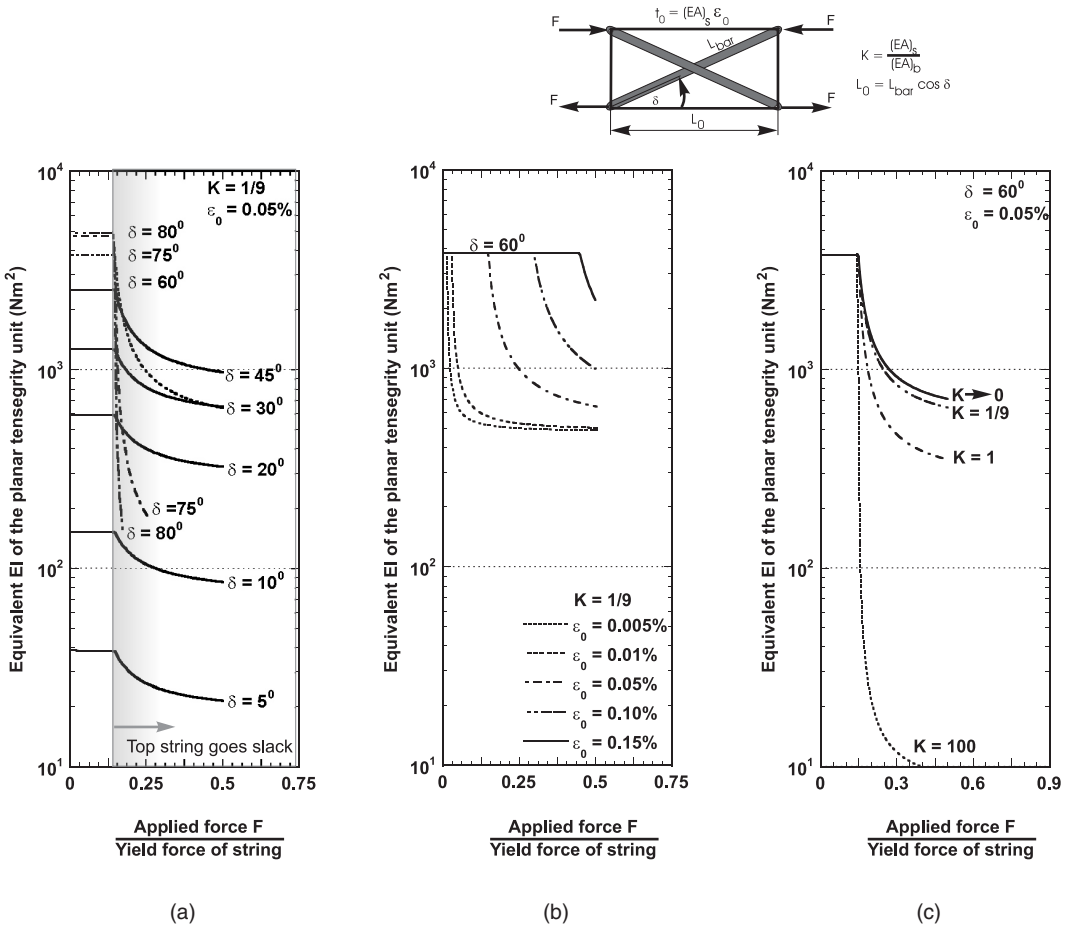


FIGURE 17.12 Bending rigidity *EI* of the planar tensegrity unit for (a) different initial angle δ with rigidity ratio $K = 1/9$ and prestrain in the top string $\varepsilon_0 = 0.05\%$, (b) different ε_0 with $K = 1/9$, (c) different K with $\delta = 60°$ and $\varepsilon_0 = 0.05\%$. L_{bar} for all cases is 0.25 m.

3. Figure 17.12(b) shows that the pretension does not have much effect on the magnitude of *EI* of a planar tensegrity unit. However, pretension does play a remarkable role in preventing the string from going slack which, in turn, increases the range of the constant *EI* against external loading. This provides robustness of *EI* predictions against uncertain external forces. This feature provides robustness against uncertainties in external forces.
4. In Figure 17.12(c) we chose structures having the same geometry and the same total stiffness, but different *K*, where *K* is the ratio of the axial rigidity of the bars to the axial rigidity of the strings. We then see that *K* has little influence on *EI* as long as none of the strings are slack. However, the bending rigidity of the tensegrity unit with slack string influenced *K*, with maximum *EI* occurring at $K = 0$ (rigid bars).

It was also observed that as the angle δ is increased or as the stiffness of the bar is decreased, the force-sharing mechanism of the members of the tensegrity unit changes quite noticeably. This phenomena is seen only in the case when the top string is slack. For example, for $K = 1/9$ and $\varepsilon_0 = 0.05\%$, for small values of δ, the major portion of the external force is carried by the bottom string, whereas after some value of δ (greater than 45°), the major portion of the external force is carried by the vertical side strings rather than the bottom string. In such cases, the vertical side strings

© 2002 by CRC Press LLC

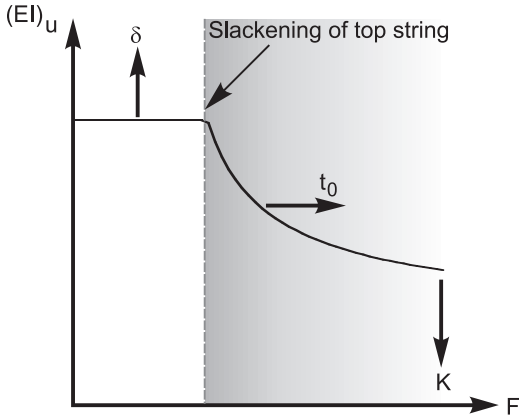


FIGURE 17.13 Trends relating geometry δ, prestress t_0, and material K.

could reach their elastic limit prior to the bottom string. Similar phenomena were also observed for a case of $K = 100$, $\delta = 60°$, and $\varepsilon_0 = 0.05\%$. In such cases, as shown in Figure 17.12(a) for $\delta = 70°$ and $\delta = 75°$, the EI drops drastically once the top string goes slack. Figure 17.13 summarizes the conclusions on bending rigidity, where the arrows indicate increasing directions of δ, $\mathbf{t}_0$, or K.

Note that when t_0 is the pretension applied to the top string, the pretension in the vertical side strings is equal to $t_0/\tan\delta$. The cases of $\delta > 80°$ were not computed, but it is clear that the bending rigidity is a step function as δ approaches 90°, with EI constant until the top string becomes slack, then the EI goes to zero as the external load increases further.

17.2.1.2 Bending Rigidity of the Planar Tensegrity for the Rigid Bar Case ($K = 0$)

The previous section briefly described the basis of the calculations for Figure 17.11. The following sections consider the special case $K = 0$ to show more analytical insight. The nonslack case describes the structure when all strings exert force. The slack case describes the structure when string 3 exerts zero force, due to the deformation of the structure. Therefore, the force in string 3 must be computed to determine when to switch between the slack and nonslack equations.

17.2.1.2.1 Some Relations from Geometry and Statics

Nonslack Case: Summing forces at each node we obtain the equilibrium conditions

$$f_c \cos\delta = F + t_3 - t_2 \sin\theta \tag{17.6}$$

$$f_c \cos\delta = t_1 + t_2 \sin\theta - F \tag{17.7}$$

$$f_c \sin\delta = t_2 \cos\theta, \tag{17.8}$$

where f_c is the compressive load in a bar, F is the external load applied to the structure, and t_i is the force exerted by string i defined as

$$t_i = k_i(l_i - l_{i0}). \tag{17.9}$$

The following relations are defined from the geometry of Figure 17.11:

$$l_1 = L_{bar} \cos\delta + L_{bar} \tan\theta \sin\delta$$

$$l_2 = L_{bar} \sin\delta \sec\theta$$

$$l_3 = L_{bar} \cos\delta - L_{bar} \sin\delta \tan\theta$$

$$h = L_{bar} \sin\delta, \tag{17.10}$$

© 2002 by CRC Press LLC

where l_i denote the geometric length of the strings. We will find the relation between δ and θ by eliminating f_c and F from (17.6)–(17.8)

$$\cos\theta = \frac{t_1 + t_3}{2t_2}\tan\delta. \tag{17.11}$$

Substitution of relations (17.10) and (17.9) into (17.11) yields

$$\cos\ \theta = \frac{k_1(L_{bar}\ \cos\delta\ +\ L_{bar}\tan\theta\sin\delta - l_{10}) + k_3(L_{bar}\cos\delta - L_{bar}\tan\theta\sin\delta - l_{30})}{2k_2(L_{bar}\sin\delta\sec\theta - l_{20})}\tan\delta. \tag{17.12}$$

If $k_i = k$, then (17.12) simplifies to

$$\tan\delta = \frac{2l_{20}}{l_{10} + l_{30}}\cos\theta\ =\ \beta\cos\theta. \tag{17.13}$$

Slack Case: In order to find a relation between δ and θ for the slack case when t_3 has zero tension, we use (17.12) and set k_3 to zero. With the simplification that we use the same material properties, we obtain

$$0 = L_{bar}\tan\theta\sin\delta\tan\delta + 2l_{20}\cos\theta - l_{10}\tan\delta - L_{bar}\sin\delta. \tag{17.14}$$

This relationship between δ and θ will be used in (17.22) to describe bending rigidity.

17.2.1.2.2 Bending Rigidity Equations

The bending rigidity is defined in (17.3) in terms of ρ and F. Now we will solve the geometric and static equations for ρ and F in terms of the parameters θ, δ of the structure. For the nonslack case, we will use (17.13) to get an analytical formula for the EI. For the slack case, we do not have an analytical formula. Hence, this must be done numerically.

From geometry, we can obtain ρ,

$$\tan\theta = \frac{l_1}{2(\rho + \frac{h}{2})}.$$

Solving for ρ we obtain

$$\begin{aligned}\rho &= \frac{l_1}{2\tan\theta} - \frac{h}{2}\\ &= \frac{L_{bar}\cos\delta + L_{bar}\tan\theta\sin\delta}{2\tan\theta} - \frac{L_{bar}\sin\delta}{2}\\ &= \frac{L_{bar}}{2}\frac{\cos\delta}{\tan\theta}.\end{aligned} \tag{17.15}$$

Nonslack Case: In the nonslack case, we now apply the relation in (17.13) to simplify (17.15)

$$\rho = \frac{L_{bar}}{2}\frac{1}{\tan\theta\sqrt{1+\beta^2\cos^2\theta}}. \tag{17.16}$$

© 2002 by CRC Press LLC

From (17.6)–(17.8) we can solve for the equilibrium external F

$$\begin{aligned} F &= \frac{1}{2}(t_1 + 2t_2 \sin\theta - t_3) \\ &= \frac{1}{2}(k_1 L_{bar} \cos\delta + k_1 L_{bar} \tan\theta \sin\delta - k_1 l_{10} \\ &\quad + 2k_2 L_{bar} \sin\delta \tan\theta - 2k_2 l_{20} \sin\theta \\ &\quad - k_3 L_{bar} \cos\delta + k_3 L_{bar} \sin\delta \tan\theta + k_3 l_{30}). \end{aligned} \tag{17.17}$$

Again, using (17.13) and $k_i = k$, Equation (17.17) simplifies to

$$F = \frac{2kL_{bar}\beta \sin\theta}{\sqrt{1+\beta^2 \cos^2\theta}} - \frac{k}{2}(l_{10} - l_{30} + 2l_{20} \sin\theta). \tag{17.18}$$

We can substitute (17.18) and (17.16) into (17.3)

$$El = \frac{L_{bar}^2 \beta \cos^2\theta}{2(1+\beta^2 \cos^2\theta)(\sin\theta)} \left(\frac{2kL_{bar}\beta \sin\theta}{\sqrt{1+\beta^2 \cos^2\theta}} - \frac{k}{2}(l_{10} - l_{30} + 2l_{20} \sin\theta) \right), \tag{17.19}$$

and we obtain the bending rigidity of the planar structure with no slack strings present. The expression for string length l_3 in the nonslack case reduces to

$$l_3 = L_{bar} \frac{1 - \beta \sin\theta}{\sqrt{1 + \beta^2 \cos^2\theta}}. \tag{17.20}$$

This expression can be used to determine the angle which causes l_3 to become slack.

Slack Case: Similarly, for the case when string 3 goes slack, we set $k_3 = 0$ and $k_i = k$ in (17.17), which yield simply

$$\begin{aligned} F_{slack} &= \frac{1}{2}(t_1 + 2t_2 \sin\theta) \\ &= \frac{1}{2}(kL_{bar} \cos\delta + 3kL_{bar} \tan\theta \sin\delta - kl_{10} - 2kl_{20} \sin\theta) \end{aligned} \tag{17.21}$$

and

$$EI_{slack} = \frac{L_{bar}^2 \sin\delta \cos\delta}{4 \tan\theta}(kL_{bar} \cos\delta + 3kL_{bar} \tan\theta \sin\delta - kl_{10} - 2kl_{20} \sin\theta). \tag{17.22}$$

See Figure 17.12(c) for a plot of EI for the $K = 0$ (rigid bar) case.

17.2.1.2.3 Constants and Conversions

All plots shown are generated with the following data which can then be converted as follows if necessary.

© 2002 by CRC Press LLC

$$\textit{Young's Modulus, } E = 2.06 \times 10^{11}\ N/m^2$$

$$\textit{Yield Stress, } \sigma = 6.9 \times 10^{8}\ N/m^2$$

$$\textit{Diameter of Tendons} = 1\ \text{mm}$$

$$\textit{Cross-Sectional Area of Tendon} = 7.8540 \times 10^{-7}\ m^2$$

$$\textit{Length of Bar, } L_{bar} = .25\ m$$

$$\textit{Prestress} = e_0$$

$$\textit{Initial Angle} = \delta_0$$

The spring constant of a string is

$$k = \frac{EA}{L_{bar}\cos(\delta_0)}. \tag{17.23}$$

The following equation can be used to compute the equivalent rest length given some measure of prestress t_0

$$t_0 = (\mathrm{EA})_s\ e_0 = k(l - l_0)$$

$$l_0 = L_{bar}\cos(\delta_0) - \frac{EAe_0}{k}. \tag{17.24}$$

17.2.1.3 Effective Bending Rigidity with Slack String ($K > 0$)

As noted earlier, the tensegrity unit is a statically indeterminate structure (meaning that matrix **A** is not full column rank) as long as the strings remain taut during the application of the external load. However, as soon as one of the strings goes slack, the tensegrity unit becomes statically determinate. In the following, an expression for bending rigidity of the tensegrity unit with an initially slack top string is derived. Even in the case of a statically determinate tensegrity unit with slack string, the problem is still a large displacement and nonlinear problem. However, a linear solution, valid for small displacements only, resulting in a quite simple and analytical form can be found. Based on the assumptions of small displacements, an analytical expression for *EI* of the tensegrity unit with slack top string has been derived in Appendix 17.B and is given below.

$$EI \approx \frac{1}{2}\frac{{L_{bar}}^2(EA)_s\sin^2\delta\cos^3\delta}{(\sin^3\delta + 2\cos^3\delta + K)}. \tag{17.25}$$

The *EI* obtained from nonlinear analysis, i.e., from (17.3) together with (17.5), is compared with the *EI* obtained from linear analysis, i.e., from (17.25), and is shown in Figure 17.14. Figure 17.14 shows that the linear analysis provides a lower bound to the actual bending rigidity. The linear estimation of *EI*, i.e., (17.25), is plotted in Figure 17.15 as a function of the initial angle δ for different values of the stiffness ratio *K*. Both bars and the strings are assumed to be made of steel, as before. It is seen in Figure 17.15 that the *EI* of the tensegrity unit with slack top string attains a maximum value for some value of δ. The decrease of *EI* (after the maximum) is due to the change in the force sharing mechanism of the members of the tensegrity unit, as discussed earlier. For small values of δ, the major portion of the external force is carried by the bottom string, whereas for larger values of δ, the vertical side strings start to share the external force. As δ is further

© 2002 by CRC Press LLC

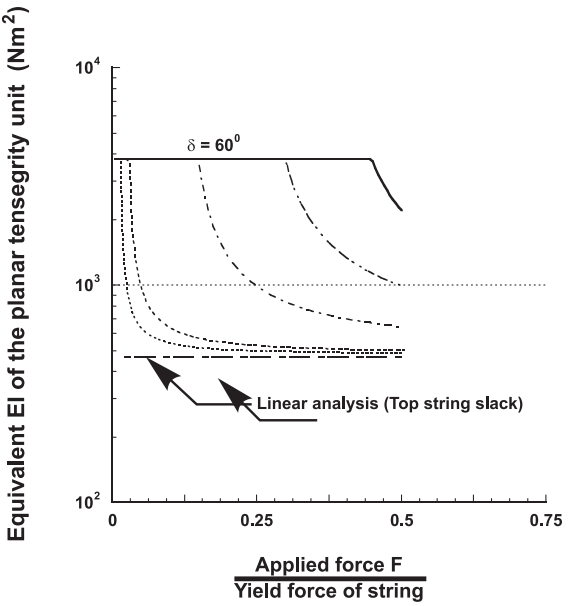


FIGURE 17.14 Comparison of *EI* from nonlinear analysis with the *EI* from linear analysis with slack top string (L_{bar} = 0.25 m, δ = 60° and K = 1/9).

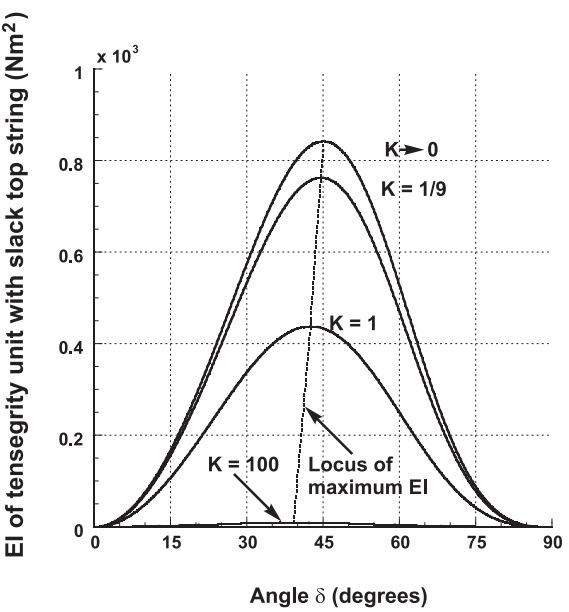


FIGURE 17.15 *EI* with slack top string with respect to the angle δ for L_{bar} = 0.25m.

increased, the major portion of the external force is carried by the vertical side strings rather than the bottom string. This explains the decrease in *EI* with the increase in δ after some values of δ for which *EI* is maximum.

The locus of the maximum *EI* is also shown in Figure 17.15. The maximum value of *EI* and the δ for which *EI* is maximum depend on the relative stiffness of the string and the bars, i.e., they depend on *K*. From Figure 17.15 note that the maximum *EI* is obtained when the bars are much

© 2002 by CRC Press LLC

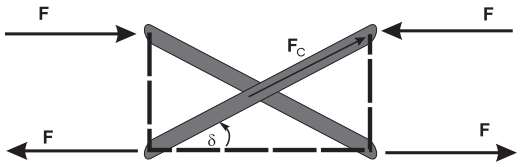


FIGURE 17.16 *C2T4* tensegrity with slack top string.

stiffer than the strings. *EI* is maximum when the bars are perfectly rigid, i.e., $K \to 0$. It is seen in Figure 17.15 and can also be shown analytically from (17.25) that for the case of bars much stiffer than the strings, $K \to 0$, the maximum *EI* of the tensegrity unit with slack top string is obtained when $\delta = 45°$. In constrast, note from Figure 17.12(a) that when no strings are slack, the maximum bending rigidity occurs with $\delta = 90°$.

17.2.2 Mass Efficiency of the *C2T4* Class 1 Tensegrity in Bending

This section demonstrates that beams composed of tensegrity units can be more efficient than continua beams. We make this point with a very specific example of a single-unit *C2T4* structure. In a later section we allow the number of unit cells to approach infinity to describe a long beam. Let Figure 17.16 describe the configuration of interest. Note that the top string is slack (because the analysis is easier), even though the stiffness will be greater before the string is slack. The compressive load in the bar, F_c.

$$F_c = F/\cos\,\delta$$

Designing the bar to buckle at this force yields

$$F_c = \frac{\pi^3 E_1 r_{bar}^4}{4L_{bar}^2},\ (L_{bar},\ r_{bar})\ =\ \text{length, radius of bar.}$$

where the mass of the two bars is (ρ_1 = bar mass density)

$$m_b = \ 2m_1 = 2\pi\,\rho_1\,L_{bar}\,r_{bar}^2 \Rightarrow r_{bar}^2 = \frac{m_b}{2\pi\rho_1 L_{bar}}.$$

Hence, eliminating r_{bar} gives for the force

$$F_c = \frac{\pi^3\,E_1}{4L_{bar}^2}\left(\frac{m_b^2}{4\pi^2\rho_1^2 L_{bar}^2}\right) = \left(\frac{\pi E_1}{16\rho_1^2 L_{bar}^4}\right) m_b^2$$

The moment applied to the unit is

$$M = F\,L_{bar}\sin\delta = \frac{\pi E_1}{16\rho_1^2 L_{bar}^3} m_b^2\cos\delta\sin\delta. \tag{17.30}$$

To compare this structure with a simple classical structure, suppose the same moment is applied to a single bar of a rectangular cross section with *b* units high and *a* units wide and yield strength σ_y such that

$$M = \frac{\sigma_y I}{C}, I = \frac{1}{12}ab^3, C = \frac{b}{2}, m_0 = \rho_0\,L_0\,ab, \tag{17.31}$$

© 2002 by CRC Press LLC

then, for the rectangular bar

$$M = \frac{\sigma_y \, m_0^2}{6a \, \rho_0^2 \, L_0^2}. \tag{17.32}$$

Equating (17.30) and (17.32), using $L_0 = L_{bar} \cos \delta$, yields the material/geometry conservation law ($\bar{\sigma}$ is a material property and g is a property of the geometry)

$$\mu^2 \triangleq \left(\frac{m_b}{m_0} \right)^2 = \frac{\delta}{3\pi} \bar{\sigma} g, \quad \bar{\sigma} \triangleq \frac{\sigma_y}{E_1}, \quad g \triangleq \frac{L_0}{a \cos^4 \delta \sin \delta} \tag{17.33}$$

The mass ratio μ is infinity if $\delta = 0°, 90°$, and the lower bound on the mass ratio is achieved when $\delta = \tan^{-1}(1/2) = 26.565°$.

Lemma 17.1 Let σ_y denote the yield stress of a bar with modulus of elasticity E_1 and dimension $a \times b \times L_0$. Let M denote the bending moment about an axis perpendicular to the b dimension. M is the moment at which the bar fails in bending. Then, the *C2T*4 tensegrity fails at the same M but has less mass if $\bar{\sigma} g < (3\pi/\delta)$, and minimal mass is achieved at $\delta = \tan^{-1}(1/2)$.

Proof: From (17.33),

$$\mu^2 = \frac{\delta}{3\pi} \bar{\sigma} g, \geq \frac{\delta}{3\pi} \bar{\sigma} \bar{g}, \bar{g} = 3.493856 \frac{L_0}{a} \tag{17.34}$$

where the lower bound $(\delta/3\pi) \bar{\sigma} \bar{g}$ is achieved at $\delta = \tan^{-1}(1/2)$ by setting $\partial g / \partial \delta = 0$ and solving $\cos^2 \delta = 4 \sin^2 \delta$, or $\tan \delta = 1/2$. ❑

For steel with $(\sigma_y, E_1) = (6.9 \times 10^8, 2 \times 10^{11})$ (N/m^2),

$$\mu^2 = \frac{\delta}{3\pi} \bar{\sigma} g \geq 0.008035869 \frac{L_0}{a} \tag{17.35}$$

where the lower bound is achieved for $\tan^{-1}(1/2) = 26.565°$. Hence, for geometry of the steel comparison bar given by $\left\{ L_0/a = 50, \text{and } \delta = \tan^{-1}(1/2) = 26.565^o \right\}$, then $m_b = 0.51 \, m_0$, showing 49% improvement in mass for a given yield moment. For the geometry $\left\{ L_0/a = 20 \text{ and } \delta = 26.565^o \right\}$, $m_b = 0.2 m_0$, showing 80% improvement in mass for a given yield moment, M. The main point here is that strength and mass efficiency are achieved by geometry ($\delta = 26.565°$), not materials.

It can be shown that the compressive force in a bar when the system *C2T*4 is under a pure bending load exhibits a similar robustness property that was shown with the bending rigidity. The force in a bar is constant until a string becomes slack, which is shown in Figure 17.17.

17.2.3 Global Bending of a Beam Made from *C2T*4 Units

The question naturally arises "what is the bending rigidity of a beam made from many tensegrity cells?" 17.2.3.2 answers that question. First, in Section 17.2.3.1 we review the standard beam theory.

17.2.3.1 Bucklings Load

For a beam loaded as shown in Figure 17.18, we have

$$EI \frac{d^2 v}{dz^2} = -Fv - Fe \tag{17.36}$$

© 2002 by CRC Press LLC

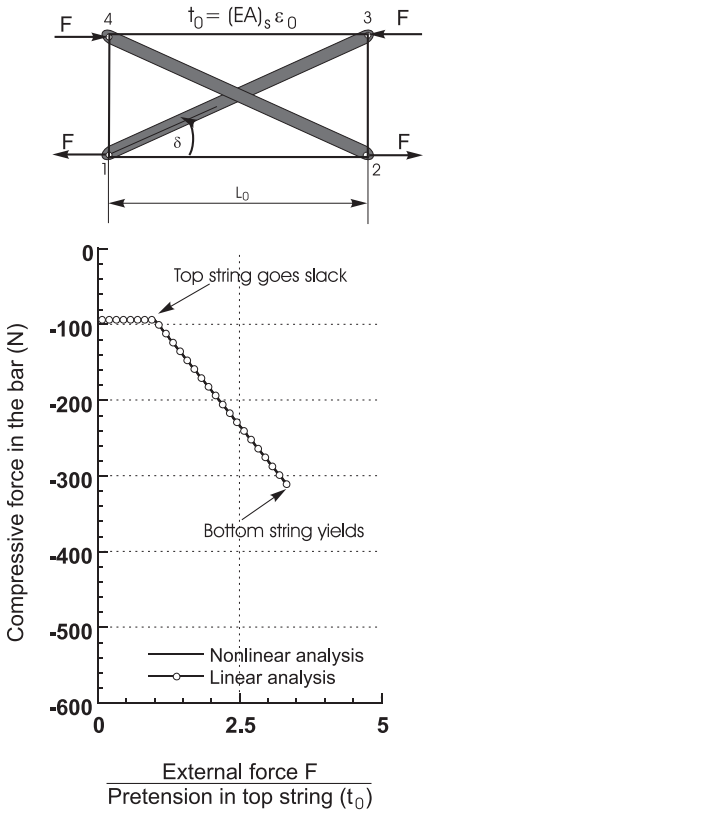


FIGURE 17.17 Comparison of force in the bar obtained from linear and nonlinear analysis for pure bending loading. (Strings and bars are made of steel, Young's modulus $E = 2.06 \times 10^{11}$ N/m^2, yield stress $\sigma_y = 6.90 \times 10^8$ N/m^2, diameter of string = 1 mm, diameter of bar = 3 mm, $K = 1/9$, $\delta = 30°$, $\varepsilon_0 = 0.05\%$ and $L_0 = 1.0$ m.)

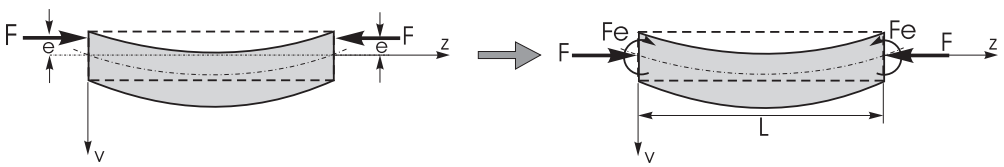


FIGURE 17.18 Bending of a beam with eccentric load at the ends.

equivalently,

$$\frac{d^2v}{dz^2} + p^2 v = -p^2 e \tag{17.37}$$

where

$$p^2 = F/EI, \tag{17.38}$$

where EI is the bending rigidity of the beam, v is the transverse displacement measured from the neutral axis (denoted by the dotted line in Figure 17.18), z represents the longitudinal axis, L is

© 2002 by CRC Press LLC

the length of the beam, e is the eccentricity of the external load F. The eccentricity of the external load is defined as the distance between the point of action of the force and the neutral axis of the beam.

The solution of the above equation is

$$v = A \sin pz + B \cos pz - e \tag{17.39}$$

where constants A and B depend on the boundary conditions. For a pin–pin boundary condition, A and B are evaluated to be

$$A = e \tan \frac{pL}{2}, \quad \text{and} \quad B = e \tag{17.40}$$

Therefore, the deflection is given by

$$v = e\left[\tan \frac{pL}{2} \sin pz + \cos pz - 1\right] \tag{17.41}$$

17.2.3.2 Buckling of Beam with Many *C2T4* Tensegrity Cells

Assume that the beam as shown in Figure 17.18 is made of n small tensegrity units similar to the one shown in Figure 17.11, such that $L = nL_0$, and the bending rigidity EI appearing in (17.36) and (17.38) is replaced by EI given by (17.25). Also, since we are analyzing a case when the beam breaks, we shall assume that the applied force is large compared to the pretension. The beam buckles at the unit receiving the greatest moment. Because the moment varies linearly with the bending and the bending is greatest at the center of the beam, the tensegrity unit at the center buckles. The maximum moment $M_{\max}$ leading to the worst case scenario is related to the maximum deflection at the center $v_{\max}$. From (17.41),

$$v_{\max} = e\left[\tan \frac{pL}{2} \sin \frac{pL}{2} + \cos \frac{pL}{2} - 1\right]. \tag{17.42}$$

Simple algebra converts this to

$$v_{\max} = e\left(\frac{1}{\cos \dfrac{pL}{2}} - 1\right) \tag{17.43}$$

The worst case $M_{\max}$ is equal to $Fv_{max} + Fe$ and is given by

$$M_{\max} = \frac{F}{\cos\left(\dfrac{nL_0}{2}\sqrt{\dfrac{F}{EI}}\right)} e \tag{17.44}$$

Now we combine this with the buckling formula for one tensegrity unit to get its breaking moment

$$M_{break} = eF_B = e\frac{\pi^2 (EI)_b}{L_0^2} \cos^3 \delta \tag{17.45}$$

© 2002 by CRC Press LLC

Thus, from Equations (17.44) and (17.45), if F exceeds F_{gB} given by

$$\frac{F_{gB}}{\cos\left(\frac{nL_0}{2}\sqrt{\frac{F_{gB}}{EI}}\right)} = \frac{\pi^2 (EI)_b}{L_0^2}\cos^3\delta \tag{17.46}$$

the central unit buckles, and F_{gB} is called the global buckling load.

Multiplying both sides of (17.46) by $(nL_0)^2$ and introducing three new variables,

$$\mathcal{F} = F_{gB}(nL_0)^2, \quad \mathcal{P} = \frac{\pi^2 (EI)_b}{L_0^2}\cos^3\delta, \qquad \mathcal{K} = \frac{1}{2}\sqrt{\frac{1}{EI}}, \tag{17.47}$$

we rewrite (17.46) as

$$\frac{\mathcal{F}}{\cos\left(\mathcal{K}\sqrt{\mathcal{F}}\right)} = \mathcal{P}n^2 L_0^2 \tag{17.48}$$

Equivalently,

$$\eta(\mathcal{F}) = \mathcal{P}n^2 L_0^2, \tag{17.49}$$

where η is a function defined as

$$\eta(\mathcal{F}) = \frac{\mathcal{F}}{\cos\left(\mathcal{K}\sqrt{\mathcal{F}}\right)}. \tag{17.50}$$

η is a monotonically increasing function in

$$0 \le \mathcal{F} \le \left(\frac{\pi}{2}\right)^2 \frac{1}{\mathcal{K}^2}, \tag{17.51}$$

satisfying

$$\eta \ge \mathcal{F} \tag{17.52}$$

It is interesting to know the buckling properties of the beam as the number of the tensegrity elements become large. As $n \to \infty$, $(nL_0)^2 \to \infty$, and from (17.49) and (17.51)

$$\mathcal{F} = \eta^{-1}\left[\mathcal{P}n^2 L_0^2\right] \to \left(\frac{\pi}{2}\right)^2 \frac{2}{\mathcal{K}^2} \tag{17.53}$$

and $\mathcal{F}$ approaches the limit from below. From Equations (17.47) and (17.49),

$$F_{gB} = \frac{1}{n^2}\frac{1}{L_0^2}\left[\eta^{-1}\left(\mathcal{P}n^2 L_0^2\right)\right] \tag{17.54}$$

Thus, for large n, using (17.53), we get

© 2002 by CRC Press LLC

$$F_{gB} \approx \frac{1}{n^2}\frac{1}{L_0^2}\left(\frac{\pi}{2}\right)^2 \frac{1}{\mathcal{K}^2}$$
$$\approx \frac{1}{n^2}\frac{1}{L_0^2}\left(\frac{\pi}{2}\right)^2 \frac{1}{\frac{1}{4}\frac{1}{EI}} \tag{17.55}$$
$$\approx \frac{1}{n^2}\frac{\pi^2 EI}{L_0^2}$$

The global buckling load as given by (17.55) is exactly the same as the classical Euler's buckling equation evaluated for the bending rigidity EI of the tensegrity unit. Therefore, asymptotically the buckling performance of the beam depends only on the characteristics of EI and L_0^2 just as a classical beam.

Note, for each n

$$F_{gB} \le \frac{1}{n^2}\frac{\pi^2 EI}{L_0^2}.$$

The implication here is that the standard Euler buckling formula applies where EI is a function of the geometrical properties of the tensegrity unit. Figure 17.12(a) shows that EI can be assigned any finite value. Hence, the beam can be arbitrarily stiff if the tensegrity unit has horizontal length arbitrarily small. This is achieved by using an arbitrarily large number of tensegrity units with large δ (arbitrarily close to 90°). More work is needed to define practical limits on stiffness.

17.2.4 A Class 1 *C*2*T*4 Planar Tensegrity in Compression

In this section we derive equations that describe the stiffness of the Class 1 *C*2*T*4 planar tensegrity under compressive loads. The nonslack case describes the structure when all strings exert force. The slack case describes the structure when string 3 and string 1 exert zero force, due to the deformation of the structure. Therefore, the force in string 3 and string 1 must be computed in order to determine when to switch between the slack and nonslack equations. We make the assumption that bars are rigid, that is, $K = 0$.

17.2.4.1 Compressive Stiffness Derivation

Nonslack Case: Summing forces at each node we obtain the equilibrium conditions

$$f_c \cos\delta = F + t_3 \tag{17.56}$$

$$f_c \cos\delta = F + t_1 \tag{17.57}$$

$$f_c \sin\delta = t_2\,, \tag{17.58}$$

where f_c is the compressive load in a bar, F is the external load applied to the structure, and t_i is the force exerted by string i defined as

$$t_i = k_i\left(l_i - l_{i0}\right).$$

The following relations are defined from the geometry of Figure 17.19:

© 2002 by CRC Press LLC

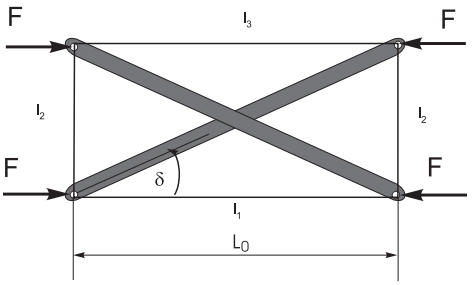


FIGURE 17.19 *C2T4* in compression.

$$l_1 = L_{bar}\cos\delta$$

$$l_2 = L_{bar}\sin\delta$$

$$l_3 = L_{bar}\cos\delta \tag{17.59}$$

Solving for F we obtain

$$F = k(l_{10} - \frac{l_{20}}{\tan\delta}). \tag{17.60}$$

Using the relation $L_0 = L_{\text{bar}}\cos\delta$ and $\tan\delta = \dfrac{\sqrt{L_{bar}^2 - L_0^2}}{L_0}$ results in

$$F = kl_{10} - \frac{kl_{20}L_0}{\sqrt{L_{bar}^2 - L_0^2}} \tag{17.61}$$

We will also make the assumption now that all strings have the same material properties, specifically, $l_{i0} = l_0$. Now, the stiffness can be computed as

$$K \triangleq -\frac{dF}{dL_0} = \frac{kl_0}{\sqrt{L_{bar}^2 - L_0^2}} + \frac{kl_0L_0^2}{\left(L_{bar}^2 - L_0^2\right)^{\frac{3}{2}}} = \frac{kl_0L_{bar}^2}{\left(L_{bar}^2 - L_0^2\right)^{\frac{3}{2}}}. \tag{17.62}$$

Similarly, for the slack case, when t_1 and t_3 are slack, we follow the same derivation setting $t_1 = t_3 = 0$ in (17.56)–(17.58)

$$F_{slack} = kL_{bar}\cos\delta - \frac{kl_0}{\tan\delta}. \tag{17.63}$$

Substitution of $L_0 = L_{bar}\cos\delta$ yields

$$F_{slack} = kL_0 - \frac{kl_0L_0}{\sqrt{L_{bar}^2 - L_0^2}}. \tag{17.64}$$

Taking the derivative with respect to L_0 gives

© 2002 by CRC Press LLC

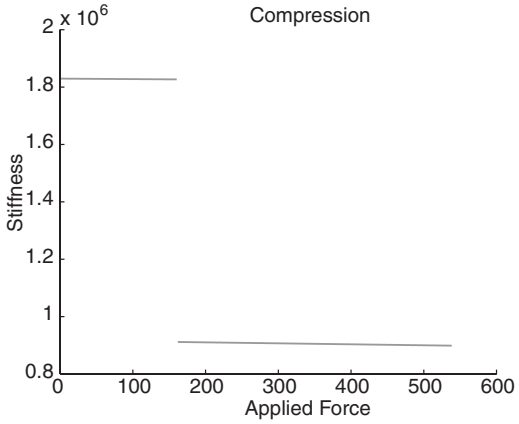


FIGURE 17.20 Stiffness of *C2T4* vs. applied load, plotted until strings yield.

$$K_{slack} = -\frac{dF_{slack}}{dL_0} = -k + \frac{kl_0}{\sqrt{L_{bar}^2 - L_0^2}} + \frac{kl_0 L_0^2}{\left(L_{bar}^2 - L_0^2\right)^{\frac{3}{2}}}$$

$$= k\left(\frac{l_0 L_{bar}^2}{\left(L_{bar}^2 - L_0^2\right)^{\frac{3}{2}}} - 1\right). \tag{17.65}$$

A plot of stiffness for the nonslack and slack case vs. applied force is given in Figure 17.20, where $k = 9.1523 \times 10^5$ *N/m*, $\delta = 45°$, $L_{bar} = 0.25$ *m*, and the force, *F*, ranges between 0 and 600 *N*.

17.2.5 Summary

Tensegrity structures have geometric structure that can be designed to achieve desirable mechanical properties. First, this chapter demonstrates how bending rigidity varies with the geometrical parameters. The bending rigidity is reduced when a string goes slack, and pretension delays the onset of slack strings. The important conclusions made in this section are

- Beams made from tensegrity units can be stiffer than their continuous beam counterparts.
- Pretension can be used to maintain a constant bending rigidity over a wider range of external loads. This can be important to robustness, when the range of external loads can be uncertain.
- For larger loads the bending stiffness is dominated by geometry, not pretension. This explains the mass efficiency of tensegrity structures since one can achieve high stiffness by choosing the right geometry.
- The ratio of mass to bending rigidity of the *C2T4* tensegrity is shown to be smaller than for a rectangular cross-section bar, provided the geometry is chosen properly (angle between bars must be less than 53°). Comparisons to a conventional truss would be instructive. There are many possibilities.

17.3 Planar Class K Tensegrity Structures Efficient in Compression

It is not hard to show that the Class 1 *C2T4* tensegrity of Figure 17.19 is not as mass efficient as a single rigid bar. That is, the mass of the structure in Figure 17.19 is greater than the mass of a

© 2002 by CRC Press LLC

FIGURE 17.21 A bar under compression.

single bar which buckles at the same load 2*F*. This motivates the examination of Class 2 tensegrity structures which have the potential of greater strength and stiffness due to ball joints that can efficiently transfer loads from one bar to another. Compressive members are disconnected in the traditional definition[2] of tensegrity structures, which we call Class 1 tensegrity. However, if stiff tendons connecting two nodes are very short, then for all practical purposes, the nodes behave as though they are connected. Hence, Class 1 tensegrity generates Class k tensegrity structures as special cases when certain tendons become relatively short. Class k tensegrity describes a network of axially loaded members in which the ends of not more than k compressive members are connected (by ball joints, of course, because torques are not permitted) at nodes of the network.

In this section, we examine one basic structure that is efficient under compressive loads. In order to design a structure that can carry a compressive load with small mass we employ Class k tensegrity together with the concept of self-similarity. Self-similar structures involve replacing a compressive member with a more efficient compressive system. This algorithm, or fractal, can be repeated for each member in the structure. The basic principle responsible for the compression efficiency of this structure is geometrical advantage, combined with the use of tensile members that have been shown to exhibit large load to mass ratios. We begin the derivation by starting with a single bar and its Euler buckling conditions. Then this bar is replaced by four smaller bars and one tensile member. This process can be generalized and the formulae are given in the following sections. The objective is to characterize the mass of the structure in terms of strength and stiffness. This allows one to design for minimal mass while bounding stiffness. In designing this structure there are trade-offs; for example, geometrical complexity poses manufacturing difficulties.

The materials of the bars and strings used for all calculations in this section are steel, which has the mass density $\rho = 7.862\ \ g/cm^3$, Young's modulus $E = 2.06^{11}\ \ N/m^2$ and yield strength $\sigma = 6.9^8\ \ N/m^2$. Except when specified, we will normalize the length of the structures $L_0 = 1$ in numerical calculations.

17.3.1 Compressive Properties of the *C4T2* Class 2 Tensegrity

Suppose a bar of radius r_0 and length L_0, as shown in Figure 17.21 buckles at load *F*. Then,

$$F = \frac{E_0 \pi^3 r_0^4}{L_0^2}, \tag{17.66}$$

where E_0 is the Young's modulus of the bar material.

The mass of the bar is

$$m_0 = \rho_0 \pi r_0^2 L_0, \tag{17.67}$$

where ρ_0 is the mass density of the bar.

Equations (17.66) and (17.67) yield the force–mass relationship

$$F = \frac{E_0 \pi m_0^2}{4\rho_0^2 L_0^4}. \tag{17.68}$$

© 2002 by CRC Press LLC

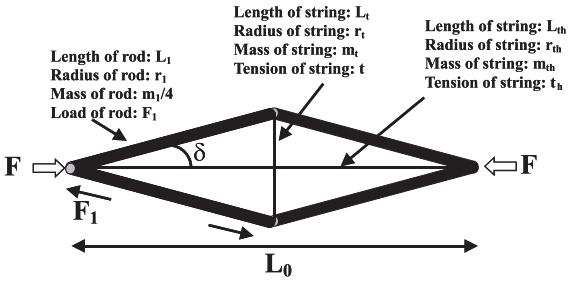


FIGURE 17.22 A *C4T2* planar Class 2 tensegrity structure.

Now consider the four-bar pinned configuration in Figure 17.22, which is designed to buckle at the same load *F*. Notice that the Class 2 tensegrity of Figure 17.22 is in the dual (where bars are replaced by strings and vice versa) of the Class 1 tensegrity of Figure 17.3(b), and is of the same type as the Class 2 tensegrity in Figure 17.3(c).

We first examine the case when tendon t_h is slack. The four identical bars buckle at the bar compressive load F_1 and the mass of each of the four bars is $1/4\, m_1$. Hence,

$$F_1 = \frac{E_1 \pi^2 r_1^4}{4L_1^2}, \quad m_1 = 4\rho_1 \pi r_1^2 L_1, \quad F_1 = \frac{E_1 \pi m_1^2}{64\rho_1^2 L_1^4}, \tag{17.69}$$

where (r_1, L_1, E_1, ρ_1) is respectively, the radius, length, Young's modulus, and mass density of each bar, and the mass of the system *C4T1* in Figure 17.22 is

$$m_1 = 4\rho_1 \pi r_1^2 L_1.$$

Since from the Figure 17.22, the length of each bar is L_1 and the compressive load in each bar is F_1 given by,

$$L_1 = \frac{L_0}{2\cos\delta}, \quad F_1 = \frac{F + t_h}{2\cos\delta}, \tag{17.70}$$

then, from (17.68)–(17.70)

$$F_1 = \frac{E_1 \pi m_1^2}{64\rho_1^2 L_1^4} = \frac{F + t_h}{2\cos\delta}. \tag{17.71}$$

Note from (17.70) that the *C4T2* structure with no external force *F* and tension $t_h = F_x$ in the horizontal string, places every member of the structure under the same load as a *C4T1* structure (which has no horizontal string) with an external load $F = F_x$. In both cases, $F_1 = F_x/2\cos\delta$.

Solving for the mass ratio, from (17.71)

$$\mu_1 \triangleq \left(\frac{m_1}{m_0}\right) = \frac{\rho_1}{\rho_0}\sqrt{\frac{E_0}{E_1}}\left(\frac{1 + \frac{t_h}{F}}{2\cos^5\delta}\right)^{\frac{1}{2}} \tag{17.72}$$

For slack tendon $t_h = 0$, note that $\mu_1 < 1$ if $\delta < \cos^{-1}\left(\frac{1}{2}\right)^{\frac{1}{5}} = 29.477°$. Of course, in the slack case (when $t_h = 0$), one might refer to Figure 17.22 as a *C4T1* structure, and we will use this designation to describe the system of Figure 17.22 when t_h is slack. Increasing pretension in t_h to generate the nonslack case can be examined later. The results are summarized as follows:

© 2002 by CRC Press LLC

Proposition 17.1 *With slack horizontal string $t_h = 0$, assume that strings are massless, and that the C4T1 system in Figure 17.22 is designed to buckle at the same load F as the original bar of mass m_0 in Figure 17.21. Then, the total mass m_1 of the C4T1 system is $m_1 = m_0(2\cos^5\delta)^{-\frac{1}{2}}$, which is less than m_0 whenever $\delta < 29.477$ degrees.*

Proof: This follows by setting $\mu_1 = 1$ in (17.72).

TABLE 17.1 Properties of the *C4T*1 Structure[a]

	$\delta = 10°$	$\delta = 20°$
r_1	$.602r_0$	$.623r_0$
m_1	$.735m_0$	$.826m_0$
L_1	$.508L_0$	$.532L_0$
$\frac{L_1}{r_1}$	$.844\frac{L_0}{r_0}$	$.854\frac{L_0}{r_0}$

[a] Strings are assumed massless.

Some illustrative data that reflect the geometrical properties of the *C4T*1 in comparison with a bar which buckles with the same force *F* are shown in Table 17.1. For example, when $\delta = 10^\circ$, the *C4T*1 requires only 73.5% of the mass of the bar to resist the same compressive force. The data in Table 17.1 are computed from the following relationships for the *C4T*1 structure. The radius of each bar in the *C4T*1 system is r_1

$$r_1^2 = \frac{m_1}{4\rho_1\pi L_1},$$

and

$$r_0^2 = \frac{m_0}{\rho_0\pi L_0}.$$

From this point forward we will assume the same material for all bars. Hence,

$$\left(\frac{r_1}{r_0}\right)^4 = \frac{1}{8\cos^3\delta}.$$

Likewise,

$$\left(\frac{L_1}{L_0}\right) = \frac{1}{2\cos\delta},$$

and

$$\left(\frac{r_1}{r_0}\right)^4 = \left(\frac{L_1}{L_0}\right)^3.$$

Also,

$$\frac{L_1}{r_1} = \frac{L_0\left(8\cos^3\delta\right)^{\frac{1}{4}}}{2r_0\cos\delta} = \frac{L_0}{r_0}\left(\frac{1}{2\cos\delta}\right)^{\frac{1}{4}}.$$

© 2002 by CRC Press LLC

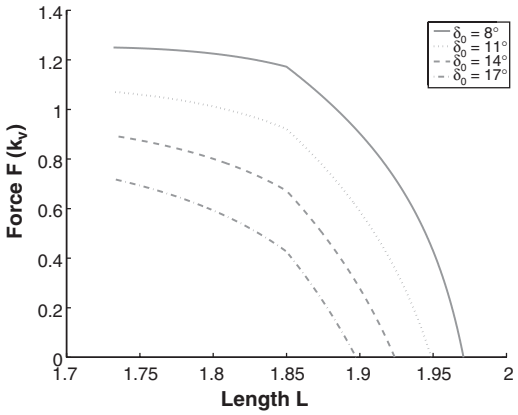


FIGURE 17.23 Load-deflection curve of *C4T*2 structure with different δ (K_e = 1, $K_h = 3K_t = 3$, L_l = 1).

17.3.2 *C4T*2 Planar Tensegrity in Compression

In this section we derive equations that describe the stiffness of the *C4T*2 planar tensegrity under compressive loads. Pretension would serve to increase the restoring force in the string, allowing greater loads to be applied with smaller deformations. This is clearly shown in the force balance Equation (17.70), where pretension can be applied through the use of the rest length L_{h0} of the string, and $t_h = k_h(L_0 - L_{h0})$, where k_h is the stiffness of the horizontal string.

17.3.2.1 Compressive Stiffness Derivation

From Figure 17.22, the equilibrium configuration can be expressed as

$$F = t\cot\delta - t_h = k_t(L_t - L_{t0})\frac{L_0}{L_t} - t_h = k_t\left(1 - \frac{L_{t0}}{L_t}\right)L_0 - t_h, \tag{17.73}$$

where t, L_0, L_t, and L_{t0} are the tension, length of the structure, length of the string, and the rest length of the vertical string, respectively. The length of the string can be written as

$$L_t^2 = 4L_1^2 - L_0^2,$$

where L_1 denotes the length of one bar. This relation simplifies the force balance equation to

$$F = k_t\left(1 - \frac{l_0}{\sqrt{4L_1^2 - L_0^2}}\right)L_0 - k_h(L_0 - l_{h0}). \tag{17.74}$$

Figure 17.23 shows the plot of the load deflection curve of a *C4T*2 structure with different δ. The compressive stiffness can be calculated by taking the derivative of (17.74) with respect to L_0 as follows,

$$\frac{dF}{dL_0} = k_t\left(1 - \frac{L_{t0}}{\sqrt{4L_1^2 - L_0^2}}\right) - k_t\frac{L_{t0}L_0^2}{\left(4L_1^2 - L_0^2\right)^{\frac{3}{2}}} - k_h$$

$$= k_t\left(1 - \frac{4L_{t0}L_1^2}{\left(4L_1^2 - L_0^2\right)^{\frac{3}{2}}}\right) - k_h. \tag{17.75}$$

© 2002 by CRC Press LLC

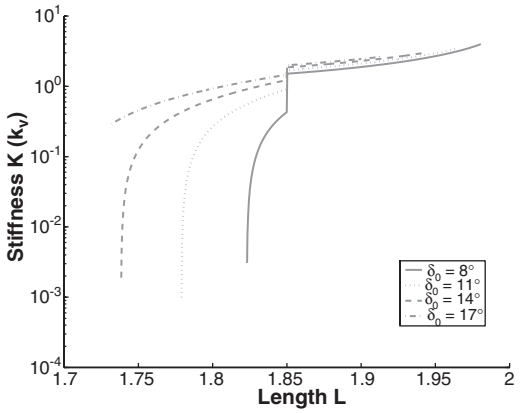


FIGURE 17.24 Stiffness vs. length of $C4T1^1$ structure with different δ_0 ($K_e = 1$, $K_h = 3K_e = 3$, $L_1 = 1$).

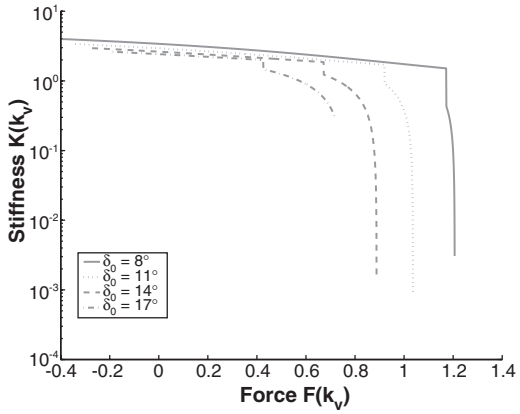


FIGURE 17.25 Stiffness vs. force of $C4T1$ structure with different δ ($K_e = 1$, $K_h = 3K_e = 3$, $L_1 = 1$).

Therefore, the stiffness is defined as

$$K \triangleq -\frac{dF}{dL_0} = k_t\left(\frac{4L_{t0}L_1^2}{\left(4L_1^2 - L_0^2\right)^{\frac{3}{2}}} - 1\right) + k_h = k_t\left(\frac{L_{t0}\cos\delta}{L_0\sin^3\delta} - 1\right) + k_h. \tag{17.76}$$

Figure 17.24 shows the plot of stiffness vs. the length of the structure and Figure 17.25 shows the plot of stiffness vs. the applied load F on the structure. Figures 17.23–17.25 demonstrate a step change in stiffness when tendon t_h goes slack. Note also that the $C4T1$ structure (t_h slack) demonstrates the property described in Figure 17.8. Structures which demonstrate robustness to external forces (that is, they maintain stiffness until strings go slack) do not preserve strength very well, whereas structures which demonstrate strength robustness have poor stiffness properties.

© 2002 by CRC Press LLC

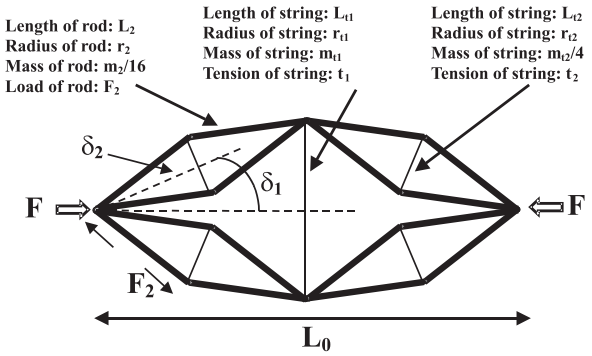


FIGURE 17.26 A $C4T1^2$ planar tensegrity structure. Points *A* are the same, and points *B* are the same, to illustrate that two identical bars overlap.

17.3.3 Self-Similar Structures of the *C4T*1 Type

Now let each of the four bars of the *C4T*1 system (that is, the *C4T*2 system with $t_h = 0$) in Figure 17.22 be replaced by another *C4T*1 system. The new 16 bar structure of Figure 17.26 is called $C4T1^2$, and is designed to buckle at the same load *F*. Hence, if F_2 represents the force in each of the 16 bars, with length L_2 and radius r_2, and mass $m_2/16$, then, for $\delta_1 = \delta_2 = \delta$ the relations below are obtained.

$$F_2 = \frac{E_0 \pi^3 r_2^4}{4L_2^2}, \quad m_2 = 16\rho_0 \pi r_2^2 L_2, \quad F_2 = \frac{E_0 \pi m_2^2}{64\rho_0^2 L_2^4} \tag{17.77}$$

$$L_2 = \frac{L_1}{2\cos\delta}, \quad F_2 = \frac{F_1}{2\cos\delta} \tag{17.78}$$

$$m_2^2 = \frac{m_1^2}{2\cos^5\delta} = \frac{m_0^2}{\left(2\cos^5\delta\right)^2} \tag{17.79}$$

$$\left(\frac{r_2}{r_1}\right)^4 = (2\cos\delta)^{-3} \tag{17.80}$$

$$\left(\frac{r_2}{r_0}\right)^4 = (2\cos\delta)^{-6} \tag{17.81}$$

$$\frac{L_2}{L_0} = \left(2\cos\delta\right)^{-1} \tag{17.82}$$

$$\frac{L_2}{L_0} = (2\cos\delta)^{-2} \tag{17.83}$$

$$\left(\frac{r_2}{r_1}\right)^4 = \left(\frac{L_2}{L_1}\right)^3 \tag{17.84}$$

© 2002 by CRC Press LLC

$$\left(\frac{r_2}{r_0}\right)^4 = \left(\frac{L_2}{L_0}\right)^3 \tag{17.85}$$

$$\frac{L_2}{r_2} = \frac{L_1}{r_1}(2\cos\delta)^{-\frac{1}{4}} = \frac{L_0}{r_0}(2\cos\delta)^{-\frac{1}{2}} \tag{17.86}$$

Now let us replace each bar in the structure of Figure 17.26 by yet another *C4T*1 structure and continue this process indefinitely. To simplify the language for these instructions, we coin some names that will simplify the description of the process we consider later.

Definition 17.3 *Let the operation which replaces the bar of length L_0 with the design of Figure 17.22 be called the "C4T1 operator." This replaces one compressive member with four compressive members plus one tension member, where the bar radii obey (17.88). Let δ be the same for any i. Let the operation which replaces the design of the bar Figure 17.21 with the design of Figure 17.26 be called the "C4T1² operator." If this C4T1 operation is repeated i times, then call it the C4T1^i operator, yielding the C4T1^i system.*

Lemma 17.2 *Let the C4T1^i operator be applied to the initial bar, always using the same material and preserving buckling strength. Then, $\delta_1 = \delta$ the mass m_i, bar radius r_i, bar length L_i of the C4T1^i system satisfy:*

$$\frac{m_i}{m_0} = (2\cos^5\delta)^{-\frac{i}{2}} \tag{17.87}$$

$$\frac{r_i}{r_0} = (2\cos\delta)^{-\frac{3i}{4}} = \left(\frac{L_i}{L_0}\right)^{\frac{3}{4}} \tag{17.88}$$

$$\frac{L_i}{L_0} = (2\cos\delta)^{-i} \tag{17.89}$$

$$\frac{L_i}{r_i} = \frac{L_0}{r_0}(2\cos\delta)^{-\frac{i}{4}} \tag{17.90}$$

$$F_i = \frac{E_0\pi r_i^4}{L_i^2} = \frac{E_0 m_i^2}{4^{2i}\rho_0^2 L_i^4} \tag{17.91}$$

$$m_i = 4^i\rho_0\pi r_i^2 L_i\,. \tag{17.92}$$

Note from (17.90) that the length-to-diameter ratio of the bars decreases with i if $\delta < 60^\circ$.

Figure 17.27 illustrates *C4T*1^i structures for $i = 3, 4, 5, 6$. Taking the limit of (17.87) as $i \to \infty$ proves the following:

Theorem 17.1 *Suppose the compressive force which buckles a C4T1^i system is a specified value, F. Then if $\delta < 29.477^\circ$, the total mass of the bars in the C4T1^i system approaches zero as $i \to \infty$.*

Proof: Take i toward infinity in (17.87). ❑

© 2002 by CRC Press LLC

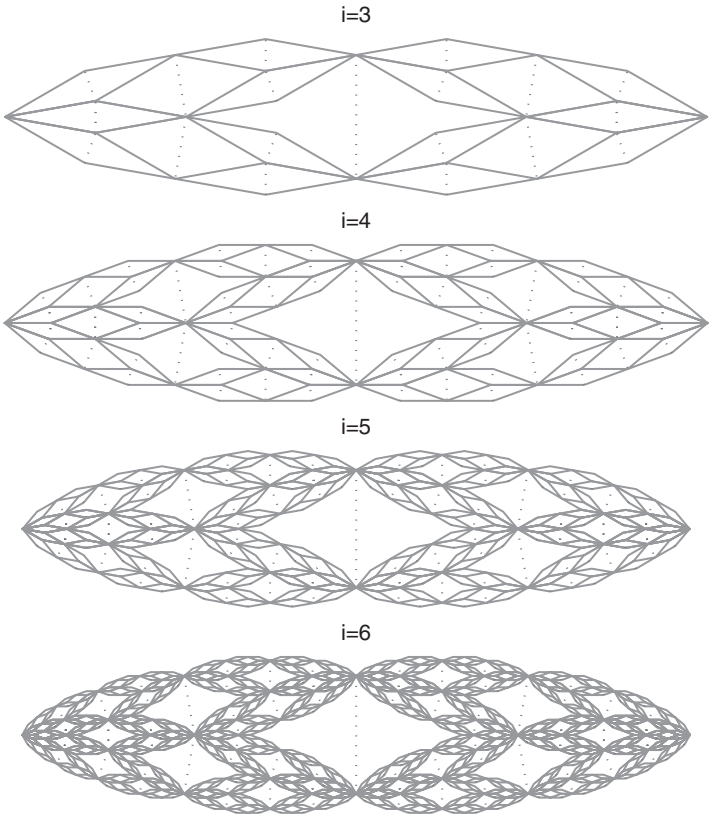


FIGURE 17.27 A $C4T1^i$ planar tensegrity structure for $i = 3, 4, 5, 6$.

Now suppose the number of self-similar iterations continue until the lengths of the bars are not longer than their diameters. Then, buckling cannot occur, and the structure is theoretically infinitely strong against buckling of the bars, but of course, the strings can still break. Therefore, ignoring the obvious overlapping of material as the iterations become large, we cite this result which is more intriguing than practical.

Proposition 17.2 *The $C4T1^i$ structure is infinitely strong against buckling if i, δ satisfy*

$$\left(2\cos\delta\right)^i \geq \left(\frac{L_0}{2r_0}\right)^4 \tag{17.93}$$

Proof: The Euler buckling formula $F_B = \pi^2 EI/L^2$ applies to beams whose diameter is smaller than the length. Otherwise, buckling cannot occur. From Lemma 17.2, the diameter equals the length of the bar when $L_i/2r_i = L_0/2r_0\left(2\cos\delta\right)^{-i/4} = 1$. From (17.90) the i such that $L_i/2r_i = 1$, satisfies

$$L_i/2r_i = L_0/2r_0\left(2\cos\delta\right)^{-i/4} = 1 \quad \text{❑} \tag{17.94}$$

As an example of (17.93) and (17.94), compared to a bar of length L_0 and radius r_0, the $C4T1^{18}$ structure with $\alpha = 10°$ buckles at the same load as the original bar, has .39% of the mass of the

© 2002 by CRC Press LLC

original bar, and is infinitely stronger than the bar. For a given specified strength, this example suggests that solid materials are quite wasteful of mass. Of course, the above result has ignored the fact that the material overlaps, if one tried to place all elements in the same plane. However, multiple planar layers of elements can be pinned to give the desired planar effect mathematically described herein. A more important omission of the above analysis is the calculation of string mass. The string mass increases with self-similar iterations (increases with i) because strings are added in the process. The mass of the bars decrease with *i,* so obviously minimal mass of the system (bars plus strings) occurs at finite *i.* This calculation will be shown momentarily.

17.3.3.1 Robustness of the *C*4*T*1

In this section we discuss briefly the issue of stability under a lateral force $F_L = 0$ in Figure 17.22. We begin by mentioning two disastrous circumstances. First, if the applied force F is small and F_L is big, then the *C*4*T*1 will collapse. Second, if the angle δ is very small and F is big, then a modest lateral force F_L will collapse the structure. Of course, a larger pretension in t_h will protect against larger F_L. Three important points on more general structures of this type: first, big F always helps lateral robustness; second, larger δ helps lateral robustness; third, increasing t_h helps robustness.

17.3.3.2 Mass and Tension of String in a *C*4*T*1[1] Structure

The mass m_{t1}, length L_{t1} and tension t_1 of string in the *C*4*T*1[1] structure are expressed as

$$m_{t1} = \rho_{t1} L_{t1} \pi r_{t1}^2, \tag{17.95}$$

$$L_{t1} = L_0 \tan\delta, \tag{17.96}$$

and

$$t_1 = F \tan\delta = \sigma_{t1} \pi r_{t1}^2, \tag{17.97}$$

respectively.

With (17.95) and (17.97),

$$t_1 = F \tan\delta = \sigma_{t1} \pi r_{t1}^2 = \sigma_{t1} \frac{m_{t1}}{\rho_{t1} L_{t1}}.$$

Hence,

$$m_{t1} = \frac{\rho_{t1}}{\sigma_{t1}} L_{t1} F \tan\delta$$

$$= \frac{\rho t1}{\sigma t1} (L_0 \tan\delta) \left(\frac{E_0 \pi m_0^2}{4\rho_0^2 L_0^4} \right) \tan\delta$$

$$= \left(\frac{E_0 \pi \tan^2\delta}{4\sigma_{t1} L_0^3} \right) \left(\frac{\rho_{t1}}{\rho_0^2} \right) m_0^2$$

$$= \left(\frac{E_0 \pi \tan^2\delta}{4\sigma_{t1} L_0^3} \right) \left(\frac{\rho_{t1}}{\rho_0^2} \right) \rho_0 \pi r_0^2 L_0 m_0.$$

So, the mass of string m_{t1} is

$$m_{t1} = \left(\frac{E_0 \pi^2 \tan^2\delta}{16\sigma_{t1} l_0^2} \right) \left(\frac{\rho_{t1}}{\rho_0} \right) m_0. \tag{17.98}$$

© 2002 by CRC Press LLC

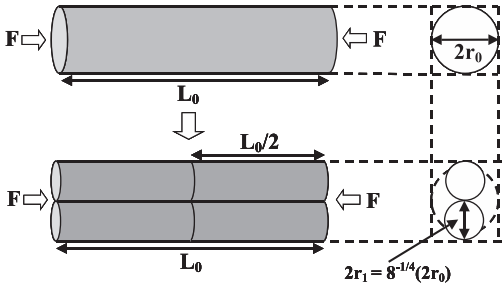


FIGURE 17.28 The minimal mass of $C4T1^1$ structure (bottom) that replaces the $C4T1^0$ structure (top) with a cross-section area comparison (right).

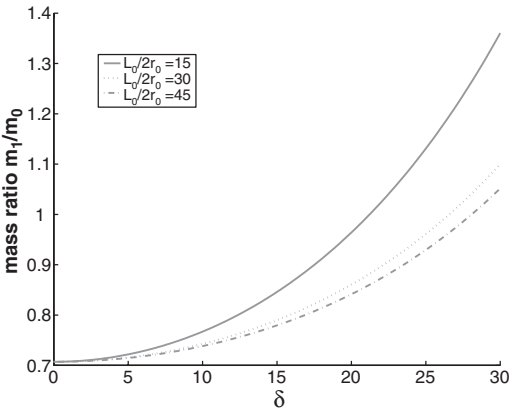


FIGURE 17.29 Mass ratio $\frac{m_1}{m_0}$ vs. δ for different length-to-diameter ratio $l_0 = \frac{L_0}{2r_o}$.

17.3.3.3 Total Mass of a $C4T1^1$ Structure

From (17.87) and (17.98), the total mass m_1 of the $C4T1^1$ structure is

$$m_1 = m_{t1} + m_{b1}$$

$$m_1 = \left[\left(\frac{E_0 \pi^2 \tan^2 \delta}{16 \sigma_{t1} l_0^2} \right) \left(\frac{\rho_{t1}}{\rho_o} \right) + \left(\frac{E_0}{E_1} \right)^{\frac{1}{2}} \left(\frac{\rho_1}{\rho_0} \right) \left(\frac{1}{2 \cos^5 \delta} \right)^{\frac{1}{2}} \right] m_0. \tag{17.99}$$

For the same material of bar and string as that of the original structure, (17.99) is reduced to

$$m_1 = \left[\left(\frac{E_0 \pi^2 \tan^2 \delta}{16 \sigma_{t1} l_0^2} \right) + \left(\frac{1}{2 \cos^5 \delta} \right)^{\frac{1}{2}} \right] m_0. \tag{17.100}$$

So, the minimal mass occurs at $\delta = 0^\circ$, yielding $m_1 = \frac{1}{\sqrt{2}} m_o$. This configuration is shown in Figure 17.28.

Figure 17.29 shows the plot of mass ratio m_1/m_0 vs. δ for different l_0. It can be seen that the upper bound of δ is less than 29.477318° for mass reduction and also depends on the length-to-diameter ratio l_0.

© 2002 by CRC Press LLC

17.3.3.4 *C4T1*^{i} Structures

For bars, the Young's modulus, density, and length-to-diameter ratio in *C4T1*^{i} will be denoted as E_i, ρ_i, and l_i, respectively. For strings, the Young's modulus, density, and yield strength in every stage j of *C4T1*^{i} (where $j = 0, 1, 2 \ldots i\text{-}1, i$) will be denoted as E_{tj}, ρ_{tj}, and σ_{tj}, respectively. The extra subscript "t" is used to distinguish the string from the bar. Applying the *C4T1*^{i} operator to the original bar allows one to proceed from the *C4T1*^{0} to *C4T1*^{i} system. Similar to the analysis before, the total mass of bars is

$$m_{bi} = 4^i \rho_i L_i \pi r_i^2. \tag{17.101}$$

The buckling load of each bar is

$$F_i = \frac{E_i \pi^3 r_i^4}{4L_i^2} = \frac{E_i \pi m_{bi}^2}{4^{2i+1} \rho_i^2 L_i^4}. \tag{17.102}$$

From the geometry of the structure, the length and load of each bar are

$$L_i = \frac{L_{e-1}}{2\cos\delta_i} = \frac{L_0}{\Pi_{j=1}^i 2\cos\delta_j}, \tag{17.103}$$

and

$$F_i = \frac{F_{i-1}}{2\cos\delta_i} = \frac{F}{\Pi_{j=1}^i 2\cos\delta j}, \tag{17.104}$$

respectively, where δ_j is the angle described in the same way as in Figure 17.26 and all δ_j might be equal, or might be different.

17.3.3.5 Mass of Bars in a *C4T1*^{i} Structure

From (17.101), (17.102), (17.103), and (17.106), the total mass of the bars in *C4T1*^{i} can be related to the mass of *C4T1*^{0} through

$$F_i = \frac{E_i \pi m_{bi}^2}{4^{2i+1} \rho_i^2 L_i^4} = \frac{F}{\Pi_{j=1}^i 2\cos\delta_j} = \frac{E_0 m_0^2}{4\rho_0^2 L_0^4 \Pi_{j=1}^i 2\cos\delta_j}$$

$$m_{bi}^2 = \left(\frac{E_0}{E_i}\right)\left(\frac{\rho_i}{\rho_0}\right)^2 \frac{1}{\Pi_{j=1}^i 2\cos^5\delta_j} m_0^2. \tag{17.105}$$

17.3.3.6 Length to Diameter Ratio of Bar in a *C4T1*^{i} Structure

The length-to-diameter ratio of the bars in *C4T1*^{i} will be l_i given by

$$l_i = \frac{L_i}{2r_i} = \left(\frac{E_i}{E_0}\right)^{\frac{1}{4}} \left(\frac{1}{\Pi_{j=1}^i 2\cos\delta_j}\right)^{\frac{1}{4}} l_0. \tag{17.106}$$

© 2002 by CRC Press LLC

17.3.3.7 Mass and Tension of Strings in a $C4T1^i$ Structure

Generalizing the concept from the previous section, the mass, length, and tension of the strings in the j-th iteration ($j = 1,2,3,\ldots i-1, i$) of $C4T1^i$ will be

$$m_{tj} = 4^{j-1}\rho_{tj}L_{tj}\pi r_{tj}^2, \tag{17.107}$$

where (r_{tj}, L_{tj}) is the radius and length of the strings,

$$L_{tj} = 2L_j \sin\delta_j = \frac{2L_0 \sin\delta_j}{\Pi_{r=1}^{j} 2\cos\delta_r}. \tag{17.108}$$

and

$$t_j = 2F_j \sin\delta_j = \frac{2F \sin\delta_j}{\Pi_{r=1}^{j} 2\cos\delta_j} = \sigma_{tj}\pi r_{tj}^2, \tag{17.109}$$

where σ_{tj} is the yield stress of the string.

With (17.107), (17.108), and (17.109), the mass of the each string in the i-th iteration can be related to the mass of $C4T1^0$

$$t_j = \frac{2\sin\delta_j}{\Pi_{r=1}^{j}\cos\delta_r}\left(\frac{E_0\pi m_0^2}{4\rho_0^2 L_0^4}\right) = \sigma_{tj}\left(\frac{4}{4i}\right)\left(\frac{m_{tj}\Pi_{s=1}^{j}2\cos\delta_s}{2\rho_{tj}L_0\sin\delta_j}\right)$$

$$m_{tj} = \frac{E_0\pi^2\sin^2\delta_j}{16\sigma_{tj}l_0^2\Pi_{r=1}^{j}\cos^2\delta_r}\left(\frac{\rho_{tj}}{\rho_0}\right)m_0. \tag{17.110}$$

17.3.3.8 Total Mass of $C4T1^i$ Structure

The total mass m_i of the $C4T1^i$ structure will be

$$m_i = m_{bi} + \sum_{j=1}^{i} m_{tj}.$$

With (17.105) and (17.110)

$$m_i = \left[\left(\frac{E_0}{E_i}\right)^{\frac{1}{2}}\left(\frac{\rho_i}{\rho_0}\right)\left(\frac{1}{\Pi_{j=1}^{i}2\cos^5\delta_j}\right)^{\frac{1}{2}} + \sum_{j=1}^{i}\frac{E_0\pi^2\sin^2\delta_j}{16\sigma_{tj}l_0^2\Pi_{r=1}^{j}\cos^2\delta_j}\left(\frac{\rho_{tj}}{\rho_0}\right)\right]m_0. \tag{17.111}$$

For the same angle $\delta_j = \delta$ and same material of bars and strings in every j-th stage, the total mass can be simplified to

$$m_i = \left[\left(\frac{1}{2\cos^5\delta}\right)^{\frac{i}{2}} + \frac{E_0\pi^2}{16\sigma_t l_0^2}\left(\frac{1}{\cos^{2i}\delta} - 1\right)\right]m_0. \tag{17.112}$$

© 2002 by CRC Press LLC

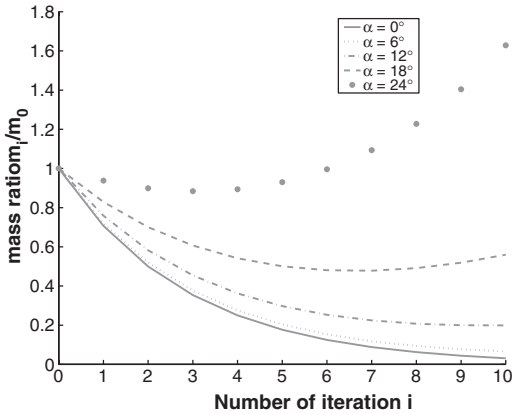


FIGURE 17.30 Mass ratio $\frac{m_i}{m_0}$ vs. number of iterations for length-to-diameter ratio $l_0 = 30$.

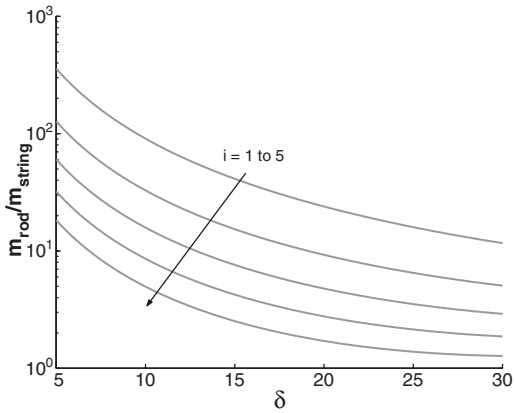


FIGURE 17.31 Mass ratio $\frac{m_{bar}}{m_{string}}$ vs. δ of $C4T1^i$ with $l_0 = \frac{L_0}{2r_0} = 30$.

Figure 17.30 shows the plot of mass ratio (m_i/m_0) vs. the number of iterations for different δ and $l_0 = 30$. From the figure, smaller angles δ will lead to larger mass reduction, and a larger length-to-diameter ratio l_0 also enhances the mass reduction effect.

Figure 17.31 shows the plot of the mass ratio of bars to strings vs. the angle δ for $l_0 = 30$. Bars will dominate the mass of structure at small δ and at a small number of iterations *i*.

Figure 17.32 shows the plot of δ vs. the number of iterations for different length-to-diameter ratios such that $m_i = m_0$. Regions below each curve are the allowed regions for mass reduction. Note that if $\delta \le 29.477318°$ and the use of materials are the same for every iteration, from (17.105), the mass of bars decreases as the number of iterations increases. However, the mass reduction will be offset by the increase of string mass as can be seen from (17.112). Therefore, maximum mass reduction can be achieved in some finite number of iterations that depends on the angle δ and length-to-diameter ratio l_0. In fact, from (17.112), the mass reduction will be maximum when the number of iterations *i* is given by the following theorem:

Theorem 17.2 *Assume all bars and strings are composed of the same material. Let the $C4T1^i$ operator be applied to the original bar to get the $C4T1^i$ system, where the iterations are designed*

© 2002 by CRC Press LLC

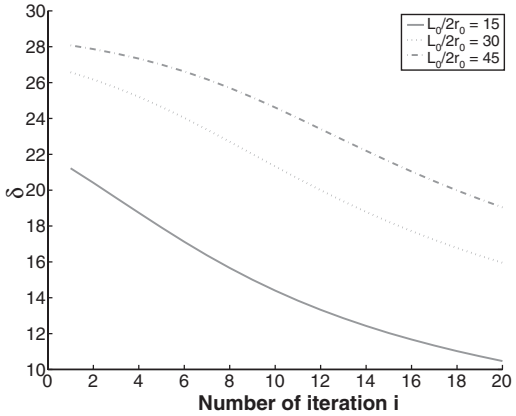


FIGURE 17.32 δ vs. number of iterations *i* for different for mass ratio $\frac{m_i}{m_0}$ = 1 of iteration i.

to preserve the buckling strength of the original bar. Minimum mass is achieved at a finite number of iterations and this number is given by either i_1 or i_2, where

$$i_1 = \left\lceil \frac{2}{In\left(\frac{1}{2\cos\delta}\right)} In\left(\frac{E_0\pi^2 In\cos\delta}{\sigma_t l_0^2 In\frac{1}{2\cos^5\delta}}\right)\right\rceil, \quad i_2 = \left\lfloor \frac{2}{In\left(\frac{1}{2\cos\delta}\right)} In\left(\frac{E_0\pi^2 In\cos\delta}{\sigma_t l_0^2 In\frac{1}{2\cos^5\delta}}\right)\right\rfloor, \tag{17.113}$$

where ⌈ ⌉ (⌊ ⌋) implies rounding up (down) to the closest integer. One must check the mass at both i_1 and i_2 to choose the smallest mass.

Proof: Let $\mu = m_i/m_0$, then, from (17.112),

$$\mu = \left(\frac{1}{2\cos^5\delta}\right)^{\frac{i}{2}} + \frac{E_0\pi^2}{16\sigma_t l_0^2}\left(\frac{\rho_t}{\rho_0}\right)\left(\frac{1}{\cos^{2i}\delta} - 1\right).$$

Take the derivative of μ w.r.t *i* by using the rule $da^x/dx = a^x In\, a$, and set it equal to zero to obtain

$$\frac{d\mu}{di} = \frac{1}{2}\left(\frac{1}{2\cos^5\delta}\right)^{\frac{i}{2}} In\left(\frac{1}{2\cos^5\delta}\right) + \frac{E_0\pi^2}{16\sigma_t l_0^2}\left(\frac{\rho_t}{\rho_0}\right)\frac{2}{\cos^{2i}\delta} In\left(\frac{1}{\cos\delta}\right) = 0.$$

Rearranging the equation gives

$$\left(\frac{1}{2\cos\delta}\right)^{\frac{1}{2}} = \frac{E_0\pi^2}{4\sigma_t l_0^2}\left(\frac{\rho_t}{\rho_0}\right)\frac{In\cos\delta}{In\left(\frac{1}{2\cos^5\delta}\right)}.$$

Solving for *i* yields (17.113). ❑

Figure 17.33 shows the plot of the optimal iteration in (17.113) vs. angle δ for maximum mass reduction.

Figure 17.34 shows the plot of the ratio of bar mass to string mass vs. δ at the optimal iteration given by (17.113). Note that at about δ= 17° and l_0= 30, the total bar mass and the total string mass are equal.

Figure 17.35 shows the corresponding plot of total mass ratio.

© 2002 by CRC Press LLC

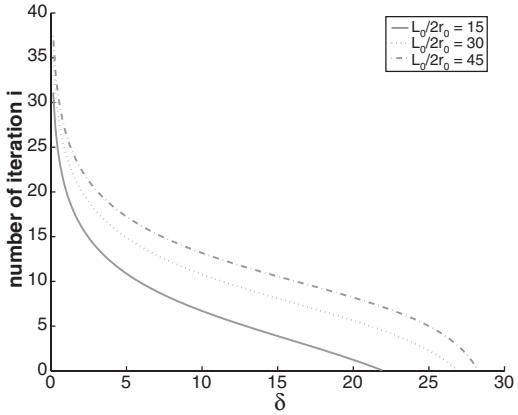


FIGURE 17.33 Optimal number of iterations i vs. δ for the maximum mass reduction.

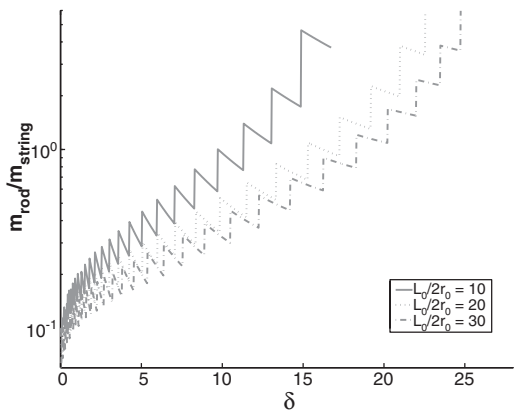


FIGURE 17.34 Mass ratio of bars to strings $\frac{m_{bar}}{m_{string}}$ vs. δ at the optimal iteration.

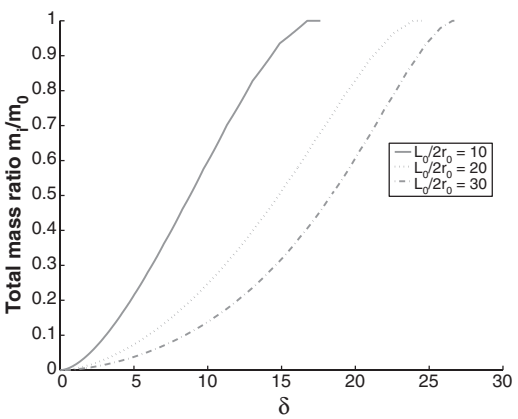


FIGURE 17.35 Total mass ratio $\frac{m_i}{m_0}$ vs. δ at the optimal iteration.

© 2002 by CRC Press LLC

17.3.4 Stiffness of the $C4T1^i$ Structure

17.3.4.1 Stiffness Definition

For the $C4T1^i$, the structure will change its length (measured from two nodal points where external load is applied) in the same direction as the applied force. Therefore, the stiffness calculation is a one-dimensional problem. For an external load F applied to the structure of length L, the stiffness K of the structure is defined as

$$K = -\frac{dF}{dL}, \tag{17.114}$$

where the negative sign means the length of the structure decreases as the applied load increases. Since the external load can be related to the potential energy of the structure U by

$$F = -\frac{dU}{dL},$$

the stiffness can also be calculated from the potential energy by

$$K = \frac{d^2U}{dL^2}. \tag{17.115}$$

17.3.4.2 The Stiffness Equation of a $C4T1^i$ Structure

In the calculations of stiffness (see Appendix 17.C), it is assumed that the stiffness of bars k_{bi} and strings k_{tj}, where $1 \le j \le i$, is constant under deformation. This is not always a good assumption, but other string stiffness models, such as $k = EA/L$ can be analyzed in a straightforward manner. If L_{i0} and L_{tj0} are the rest lengths of bars in the i-th iteration and strings in the j-th iteration, respectively, the stiffness of $C4T1^i$ is given by

$$K_i = k_{t1}\left\{\left(\frac{L_{t10}}{L_{t1}} - 1\right) + \left[4^i \frac{k_{t1}}{k_{bi}} \frac{L_i}{L_{i0}} \left(\frac{L_i}{L_0}\right)^2 + \sum_{j=1}^{i} 4^{j-1} \frac{k_{t1}}{k_{tj}} \frac{L_{tj}}{L_{tjo}} \left(\frac{L_{tj}}{L_0}\right)^2\right]^{-1}\right\}, \tag{17.116}$$

where, in the buckling design (see Appendix 17.B)

$$L_0 = L_i \Pi_{s=1}^{i} 2\cos\delta_s$$

$$L_{tj} = 2L_i \sin\delta_j \Pi_{s=j+1}^{i} 2\cos\delta_s$$

$$\frac{k_{t1}}{k_{tj}} = \frac{E_{t1}\sigma_{tj}}{E_{tj}\sigma_{t1}}$$

$$\frac{k_{t1}}{k_{bi}} = \frac{E_{t1}\pi}{16\sigma_{t1}l_0^2}\sqrt{\frac{E_0}{E_i}}\left(\Pi_{s=1}^{i} 2\cos\delta_s\right)^{\frac{1}{2}}$$

$$\frac{L_{tjo}}{L_{tj}} = 1 - \frac{\sigma_{tj}}{E_{tj}}$$

© 2002 by CRC Press LLC

$$k_{t1} = \frac{E_0 E_{t1} \pi^3 L_0}{16 \sigma_{t1} l_0^4}$$

$$\frac{L_{io}}{L_i} = 1 + \frac{\pi}{16 l_0^2} \left(\Pi_{s=1}^{i} 2 \cos \delta_s \right)^{\frac{1}{2}}.$$

In particular, if the materials of the bars and strings used are the same as those of original structure ($C4T1^0$) and $\delta_j = \delta$, the stiffness equation will be simplified to

$$\begin{aligned} K_i &= k_{t1} \left\{ -\frac{\sigma_t}{E_t} + \left[4^i \frac{k_{t1}}{k_{bi}} \frac{L_i}{L_{i0}} \left(\frac{L_i}{L_0} \right)^2 + \left(1 - \frac{\sigma_t}{E_t} \right)^{-1} \sum_{j=1}^{i} 4^{j-1} \left(\frac{L_{tj}}{L_0} \right)^2 \right]^{-1} \right\} \\ &= k_{t1} \left\{ -\frac{\sigma_t}{E_t} + \left[4^i \frac{k_{t1}}{k_{bi}} \frac{L_i}{L_{i0}} \left(\frac{L_i}{L_0} \right)^2 + \left(1 - \frac{\sigma_t}{E_t} \right)^{-1} \left(\frac{1}{\cos^{2i} \delta} - 1 \right) \right]^{-1} \right\} \end{aligned} \tag{17.117}$$

where

$$L_0 = L_i (2 \cos \delta)^i$$

$$\frac{k_{t1}}{k_{bi}} = \frac{E_t \pi^2}{16 \sigma_t l_0^2} \left(2 \cos \delta \right)^{\frac{i}{2}}$$

$$k_{t1} = \frac{E_0 E_t \pi^3 L_0}{64 \sigma_t l_0^4}$$

$$\frac{L_{i0}}{L_i} = 1 + \frac{\pi}{16 l_0^2} \left(2 \cos \delta \right)^{\frac{i}{2}}.$$

17.3.4.3 The Rigid Bar Case

If the bar has infinite rigidity (large compared to the stiffness of strings), this means

$$\frac{k_{ti}}{k_{bi}} \to 0,$$

then, the stiffness equation becomes

$$K_i = k_{t1} \left(\frac{\cos^{2i} \delta - \frac{\sigma_t}{E_t}}{1 - \cos^{2i} \delta} \right), \tag{17.118}$$

where

$$k_{t1} = \frac{E_0 E_t \pi^3 L_0}{64 \sigma_t l_0^4}.$$

Figure 17.36 shows the plot of the stiffness K_i vs. δ for length-to-diameter ratio $l_0 = 30$.

© 2002 by CRC Press LLC

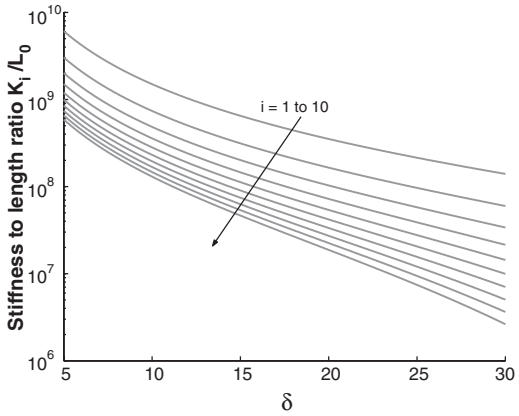


FIGURE 17.36 Stiffness K_i of $C4T1^i$ structure with rigid bars vs. δ for $l_0 = 30$.

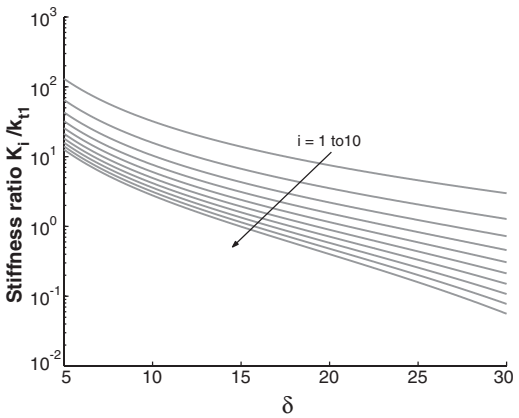


FIGURE 17.37 Stiffness ratio $\frac{K_i}{k_{t_1}}$ of $C4T1^i$ structure with rigid bars vs. δ.

Figure 17.37 shows the plot of the stiffness ratio K_i/k_{i1} vs. δ for the infinite rigid bar $C4T1^i$ structure. Note that the ratio is independent of l_0.

Because the stiffness reduces with each iteration i, it is of interest to know how many iterations may be taken before the stiffness violates a desired lower bound $\underline{K}$.

Proposition 17.3 *Given δ and a desired lower bound stiffness $\underline{K}$ of $C4T1^i$, that is, $\underline{K} \le K_i$, the number of iterations i which achieves this stiffness requirement is bounded by*

$$i \le \frac{\log \frac{\frac{\underline{K}}{k_{t1}} + \frac{\sigma_t}{E_t}}{1 + \frac{\underline{K}}{k_{t1}}}}{2 \log \cos\delta}. \tag{17.119}$$

Proof: From (17.118),

$$\underline{K} \le K_i$$

$$\le k_{t1}\left(\frac{\cos^{2i}\delta - \frac{\sigma_t}{E_t}}{1 - \cos^{2i}\delta}\right).$$

Rearrange the inequality to expose $\cos^{2i}\delta$ on one side and then take the log of both sides to obtain (17.119) ❑

© 2002 by CRC Press LLC

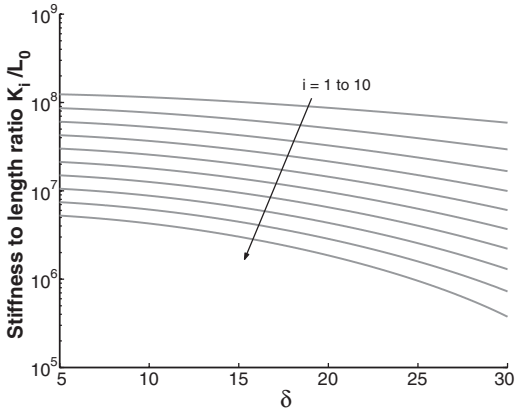


FIGURE 17.38 Stiffness-to-length ratio $\frac{K_i}{L_0}$ vs. δ for length-to-diameter ratio $l_0 = 30$, for elastic bars.

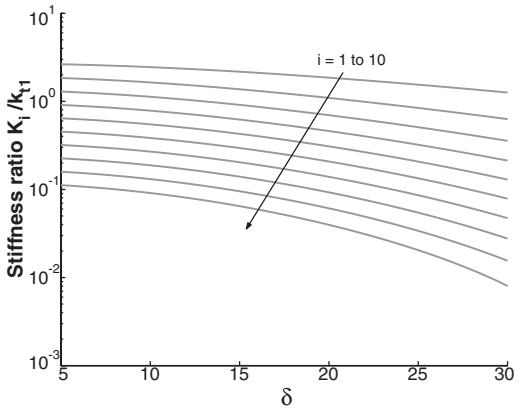


FIGURE 17.39 Stiffness ratio $\frac{K_i}{k_{t_1}}$ vs. δ for length-to-diameter ratio $l_0 = 30$, for elastic bars.

17.3.4.4 The Elastic Bar Case

If the bars in the $C4T1^i$ structure are elastic,

$$\frac{k_{ti}}{k_{bi}} \neq 0\,,$$

and hence (17.117) will be used for numerical calculations.

Figure 17.38 shows the plot of the stiffness K_i vs. δ for length-to-diameter ratio $l_0 = 30$.

Figure 17.39 shows the corresponding plot of the stiffness ratio K_i/K_{t1} vs. δ for Figure 17.38. A comparison of Figures 17.36 to 17.39 reveals that with elastic bars, system stiffness is much less than for a system made with rigid bars.

17.3.4.4.1 Stiffness Ratio K_i/K_0

The stiffness of $C4T1^0$ (a single bar) is

© 2002 by CRC Press LLC

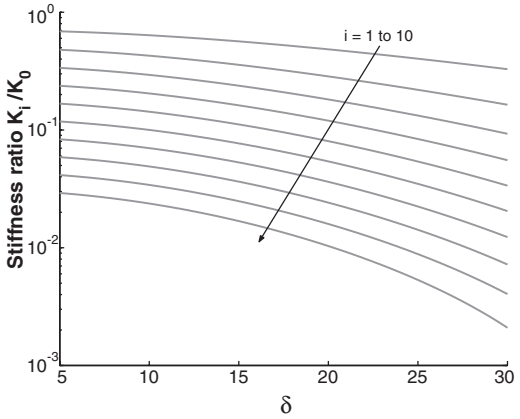


FIGURE 17.40 Stiffness ratio $\frac{K_i}{K_0}$ vs. δ for $l_0 = 30$, for elastic bars.

$$K_0 = \frac{E_0 \pi r_0^2}{L_0}. \tag{17.120}$$

With (17.117), the stiffness ratio K_i/K_0 is given by

$$\frac{K_i}{K_0} = \left(\frac{E_t \pi^2}{16\sigma_t l_0^2}\right)\left\{-\frac{\sigma_t}{E_t} + \left[4^i \frac{k_{ti}}{k_{bi}} \frac{L_i}{L_{i0}}\left(\frac{L_i}{L_{i0}}\right)^2 + \left(1-\frac{\sigma_t}{E_t}\right)^{-1}\left(\frac{1}{\cos^{2i}\delta} - 1\right)\right]^{-1}\right\}, \tag{17.121}$$

where other physical quantities are the same as those given in (17.117).

Figure 17.40 shows the plot of the stiffness ratio K_i/K_0 vs. δ for length-to-diameter ratio $l_0 = 30$.

Figures 17.37, 17.39, and 17.40 demonstrate that stiffness is much less sensitive to geometry (choice of δ) when bars are elastic than when the bars are rigid.

17.3.4.4.2 Stiffness to Mass Ratio

From (17.112) and (17.117), the stiffness-to-mass ratio is given by

$$\frac{K_i}{m_i} = \left(\frac{E_0 E_t \pi^2}{16\sigma_t \rho_0 l_0^2 L_0^2}\right)\left\{\frac{-\frac{\sigma_t}{E_t} + \left[4^i \frac{k_{ti}}{k_{bi}} \frac{L_i}{L_{i0}}\left(\frac{L_i}{L_0}\right)^2 + \left(1-\frac{\sigma_t}{E_t}\right)^{-1}\left(\frac{1}{\cos^{2i}\delta} - 1\right)^{-1}\right]}{\left(\frac{1}{2\cos^5\delta}\right)^{\frac{1}{2}} + \frac{E_0\pi}{16\sigma_t l_0^2}\left(\frac{1}{\cos^{2i}\delta} - 1\right)}\right\}, \tag{17.122}$$

where other physical quantities are the same as those given in (17.117).

Figures 17.41 plots the stiffness-to-mass ratio, K_i/m_i vs. δ, length-to-diameter ratio $l_0 = 30$. The stiffness-to-mass ratio remains constant with self-similar iteration i at $\delta = 0°$ and always decreases with the increase of iterations i and δ.

17.3.5 *C4T1*^{i} Structure with Elastic Bars and Constant Stiffness

Given a bar of radius r_0 and length L_0 under applied force F (not buckling load), this section shows how to design the *C4T1*^{i} structure to minimize the use of mass by replacing the bar with the *C4T1*^{i} structure such that the stiffness of the structure is the same as that of a bar under the same applied load F.

It is assumed that the stiffness of member bars and strings is constant and it is given by

© 2002 by CRC Press LLC

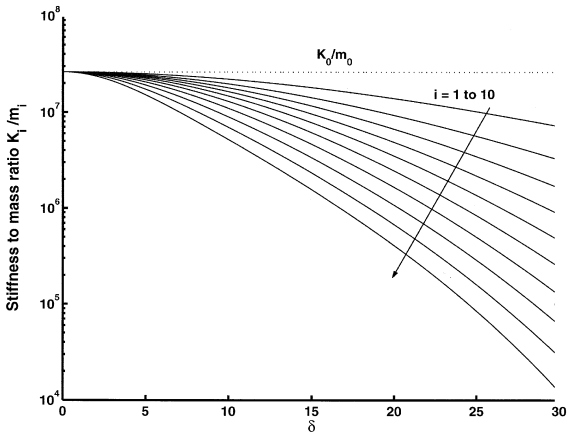


FIGURE 17.41 Stiffness-to-mass ratio $\frac{K_i}{m_i}$ vs. δ for $l_0 = 30$.

$$k = \frac{E\pi r^2}{L}, \tag{17.123}$$

where r and L are the cross-section radius and length of bars or strings when the $C4T1^i$ structure is under external load F.

17.3.5.1 $C4T1^1$ at δ = 0°

At δ = 0°, it is known from the previous section that the use of mass is minimum while the stiffness is maximum. Therefore, a simple analysis of $C4T1^1$ at δ = 0 will give an idea of whether it is possible to reduce the mass while preserving stiffness.

For the $C4T1^0$ structure, the stiffness is given by

$$K_0 = \frac{E\pi r_0^2}{L_0}. \tag{17.124}$$

For a $C4T1^1$ structure at δ = 0°, i.e., two pairs of parallel bars in series with each other, the length of each bar is $L_0/2$ and its stiffness is

$$k_b = \frac{E\pi r_1^2}{L_1} = \frac{2E\pi r_1^2}{L_0}. \tag{17.125}$$

For this four-bar arrangement, the equivalent stiffness is same as the stiffness of each bar, i.e.,

$$K_1 = \frac{2E\pi r_1^2}{L_0}. \tag{17.126}$$

To preserve stiffness, it is required that

$$K_1 = K_0$$

$$\frac{2E\pi r_1^2}{L_0} = \frac{E\pi r_0^2}{L_0}.$$

© 2002 by CRC Press LLC

So,

$$r_0^2 = 2r_1^2. \tag{17.127}$$

Then, the mass of $C4T1^1$ at $\delta = 0°$ for stiffness preserving design is

$$m_1 = 4\rho\pi r_1^2 L_1 = 4\rho\pi \frac{r_0^2}{2}\frac{L_0}{2} = \rho\pi_0^2 L_0 = m_0, \tag{17.128}$$

which indicates, at $\delta = 0°$, that the mass of $C4T1^1$ is equal to that of $C4T1^0$ in a stiffness-preserving design. Therefore, the mass reduction of $C4T1^i$ structure in a stiffness-preserving design is unlikely to happen. However, if the horizontal string t_h is added in the $C4T1^1$ element to make it a $C4T2$ element, then stiffness can be improved, as shown in (17.76).

17.3.6 Summary

The concept of self-similar tensegrity structures of Class k has been illustrated. For the example of massless strings and rigid bars replacing a bar with a Class 2 tensegrity structure $C4T1$ with specially chosen geometry, $\delta < 29°$, the mass of the new system is less than the mass of the bar, the strength of the bar is matched, and a stiffness bound can be satisfied. Continuing this process for a finite member of iterations yields a system mass that is minimal for these stated constraints. This optimization problem is analytically solved and does not require complex numerical codes. For elastic bars, analytical expressions are derived for the stiffness, and choosing the parameters to achieve a specified stiffness is straightforward numerical work. The stiffness and stiffness-to-mass ratio always decrease with self-similar iteration, and with increasing angle δ, improved with the number of self-similar iterations, whereas the stiffness always decreases.

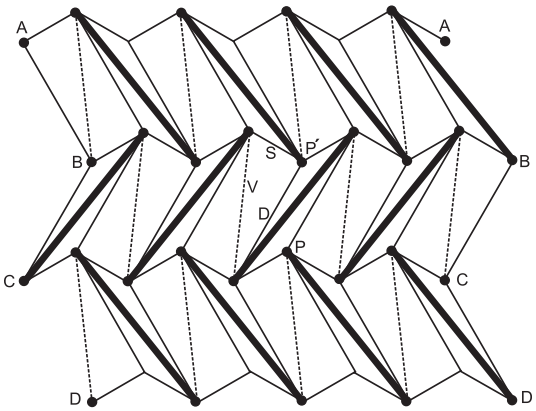
17.4 Statics of a 3-Bar Tensegrity

17.4.1 Classes of Tensegrity

The tensegrity unit studied here is the simplest three-dimensional tensegrity unit which is comprised of three bars held together in space by strings to form a tensegrity unit. A tensegrity unit comprising three bars will be called a 3-bar tensegrity. A 3-bar tensegrity is constructed by using three bars in each stage which are twisted either in clockwise or in counter-clockwise direction. The top strings connecting the top of each bar support the next stage in which the bars are twisted in a direction opposite to the bars in the previous stage. In this way any number of stages can be constructed which will have an alternating clockwise and counter-clockwise rotation of the bars in each successive stage. This is the type of structure in Snelson's Needle Tower, Figure 17.1. The strings that support the next stage are known as the "saddle strings (S)." The strings that connect the top of bars of one stage to the top of bars of the adjacent stages or the bottom of bars of one stage to the bottom of bars of the adjacent stages are known as the "diagonal strings (D)," whereas the strings that connect the top of the bars of one stage to the bottom of the bars of the same stage are known as the "vertical strings (V)."

Figure 17.42 illustrates an unfolded tensegrity architecture where the dotted lines denote the vertical strings in Figure 17.43 and thick lines denote bars. Closure of the structure by joining points A, B, C, and D yields a tensegrity beam with four bars per stage as opposed to the example in Figure 17.43 which employs only three bars per stage. Any number of bars per stage may be employed by increasing the number of bars laid in the lateral direction and any number of stages can be formed by increasing the rows in the vertical direction as in Figure 17.42.

© 2002 by CRC Press LLC



lllFIGURE 17.42 Unfolded tensegrity architecture.

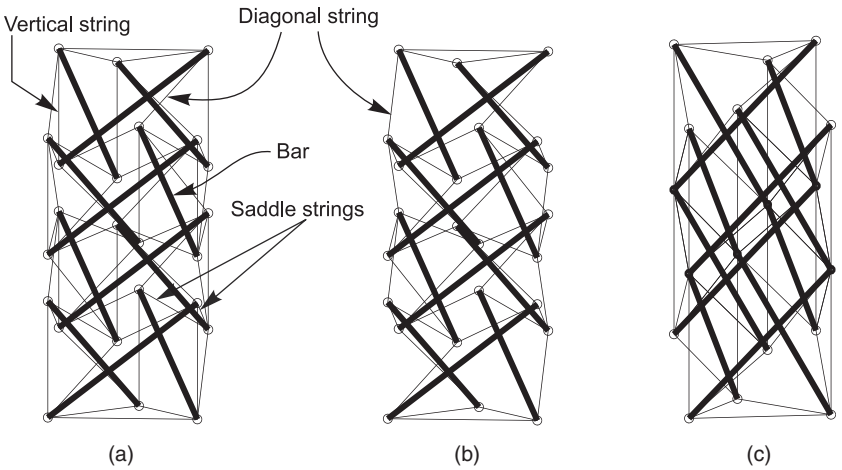


FIGURE 17.43 Types of structures with three bars in one stage. (a) 3-Bar SVD tensegrity; (b) 3-bar SD tensegrity, (c) 3-bar SS tensegrity.

Even with only three bars in one stage, which represents the simplest form of a three-dimensional tensegrity unit, various types of tensegrities can be constructed depending on how these bars have been held in space to form a beam that satisfies the definition of tensegrity. Three variations of a 3-bar per stage structure are described below.

17.4.1.1 3-Bar SVD Class 1 Tensegrity

A typical two-stage 3-bar SVD tensegrity is shown in Figure 17.43(a) in which the bars of the bottom stage are twisted in the counter-clockwise direction. As is seen in Figure 17.42 and Figure 17.43(a), these tensegrities are constructed by using all three types of strings, saddle strings (S), vertical strings (V), and the diagonal strings (D), hence the name SVD tensegrity.

17.4.1.2 3-Bar SD Class 1 Tensegrity

These types of tensegrities are constructed by eliminating the vertical strings to obtain a stable equilibrium with the minimal number of strings. Thus, a SD-type tensegrity only has saddle (S) and the diagonal strings (D), as shown in Figure 17.42 and Figure 17.43(b).

© 2002 by CRC Press LLC

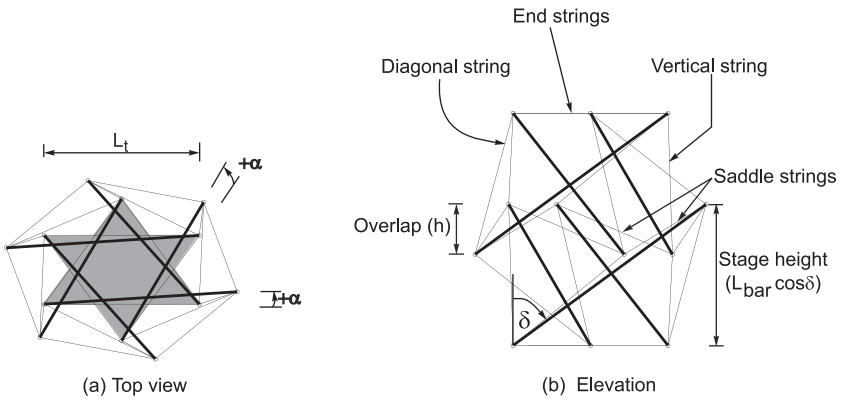


FIGURE 17.44 Top view and elevation of a two-stage 3-bar SVD tensegrity.

17.4.1.3 3-Bar SS Class 2 Tensegrity

It is natural to examine the case when the bars are connected with a ball joint. If one connects points P and P′ in Figure 17.42, the resulting structure is shown in Figure 17.43(c). The analysis of this class of structures is postponed for a later publication.

The static properties of a 3-bar SVD-type tensegrity is studied in this chapter. A typical two-stage 3-bar SVD-type tensegrity is shown in Figure 17.44 in which the bars of the bottom stage are twisted in the counter-clockwise direction. The coordinate system used is also shown in the same figure. The same configuration will be used for all subsequent studies on the statics of the tensegrity. The notations and symbols, along with the definitions of angles α and δ, and overlap between the stages, used in the following discussions are also shown in Figure 17.44.

The assumptions related to the geometrical configuration of the tensegrity structure are listed below:

1. The projection of the top and the bottom triangles (vertices) on the horizontal plane makes a regular hexagon.
2. The projection of bars on the horizontal plane makes an angle α with the sides of the base triangle. The angle α is taken to be positive (+) if the projection of the bar lies inside the base triangle, otherwise α is considered as negative (–).
3. All of the bars are assumed to have the same declination angle δ.
4. All bars are of equal length, L.

17.4.2 Existence Conditions for 3-Bar SVD Tensegrity

The existence of a tensegrity structure requires that all bars be in compression and all strings be in tension in the absence of the external loads. Mathematically, the existence of a tensegrity system must satisfy the following set of equations:

$$\mathbf{A}(\bar{q})\mathbf{t}_0 = 0, \qquad \mathbf{t}_{0_strings} > 0, \qquad \bar{q}: stable \quad equilibrium. \tag{17.129}$$

For our use, we shall define the conditions stated in (17.129) as the "tensegrity condition."

Note that $\mathbf{A}$ of (17.129) is now a function of α, δ, and h, the generalized coordinates, labeled q generically. For a given q, the null space of $\mathbf{A}$ is computed from the singular value decomposition of $\mathbf{A}$.[36,37] Any singular value of $\mathbf{A}$ smaller than 1.0×10^{-10} was assumed to be zero and the null vector $\mathbf{t}_0$ belonging to the null space of $\mathbf{A}$ was then computed. The null vector was then checked against the requirement of all strings in tension. The values of α, δ, and h that satisfy (17.129)

© 2002 by CRC Press LLC

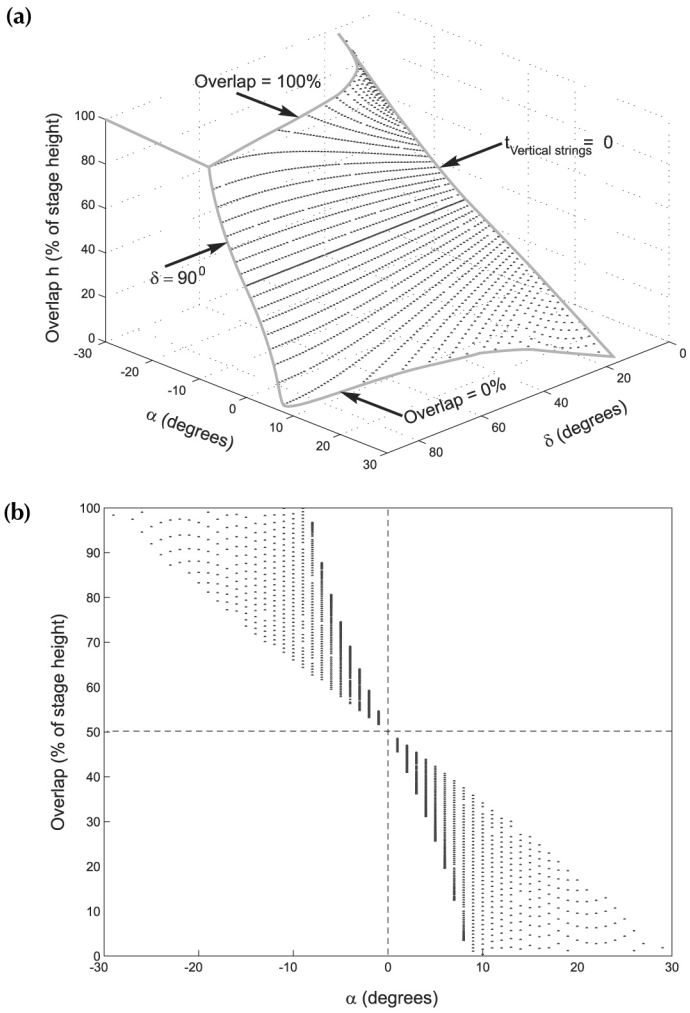


FIGURE 17.45 Existence conditions for a two-stage tensegrity. Relations between (a) α, δ, and the overlap, (b) α and overlap, (c) δ and overlap, and (d) δ and α giving static equilibria.

yield a tensegrity structure. In this section, the existence conditions are explored for a two-stage 3-bar SVD-type tensegrity, as shown in Figure 17.44, and are discussed below.

All of the possible configurations resulting in the self-stressed equilibrium conditions for a two-stage 3-bar SVD-type tensegrity are shown in Figure 17.45. While obtaining Figure 17.45, the length of the bars was assumed to be 0.40 m and L_t, as shown in Figure 17.44, was taken to be 0.20 m.

Figure 17.45 shows that out of various possible combinations of α–δ–*h*, there exists only a small domain of α–δ–*h* satisfying the existence condition for the two-stage 3-bar SVD-type tensegrity studied here. It is interesting to explore the factors defining the boundaries of the domain of α–δ–*h*. For this, the relation between α and *h*, δ and *h*, and also the range of α and δ satisfying the existence condition for the two-stage 3-bar SVD-type tensegrity are shown in Figures 17.45(b), (c), and (d). Figure 17.45(b) shows that when $\alpha = 30°$, there exists a unique value of overlap equal to 50% of the stage height. Note that $\alpha = 0°$ results in a perfect hexagonal cylinder. For any value of α other than 0°, multiple values of overlap exist that satisfies the existence condition. These overlap values for any given value of α depend on δ, as shown in Figure 17.45(c). It is also observed in Figure 17.45(b) and (c) that a larger value of negative α results in a large value of overlap and a

© 2002 by CRC Press LLC

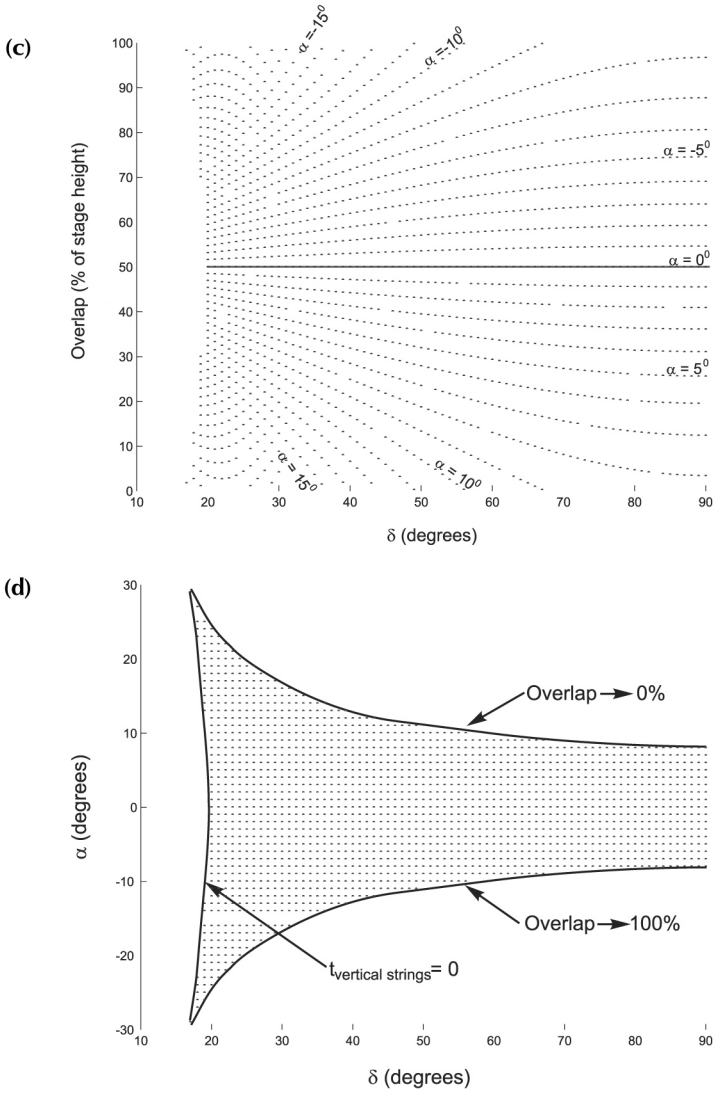


FIGURE 17.45 (Continued)

larger value of positive α results in a smaller value of overlap. Note that a large value of negative α means a "fat" or "beer-barrel" type structure, whereas larger values of positive α give an "hourglass" type of structure. It can be shown that a fat or beer-barrel type structure has greater compressive stiffness than an hourglass type structure. Therefore, a tensegrity beam made of larger values of negative α can be expected to have greater compressive strength.

Figure 17.45(d) shows that for any value of δ, the maximum values of positive or negative α are governed by overlap. The maximum value of positive α is limited by the overlap becoming 0% of the stage height, whereas the maximum value of negative α is limited by the overlap becoming 100% of the stage height. A larger value of negative α is expected to give greater vertical stiffness. Figure 17.45(d) shows that large negative α is possible when δ is small. However, as seen in Figure 17.45(d), there is a limit to the maximum value of negative α and to the minimum δ that would satisfy the existence conditions of the two-stage 3-bar SVD-type tensegrity. To understand this limit of the values of α and δ, the distribution of the internal pretensioning forces in each of the members is plotted as a function of α and δ, and shown in Figures 17.46 and 17.47.

© 2002 by CRC Press LLC

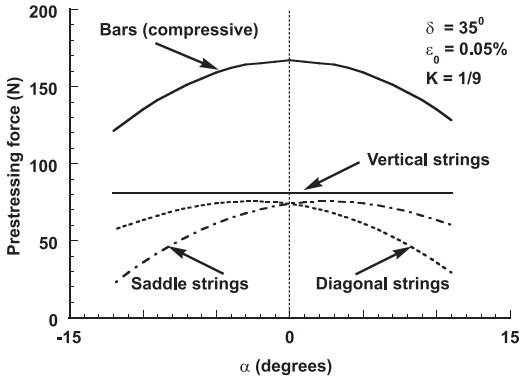


FIGURE 17.46 Prestressing force in the members as a function of α.

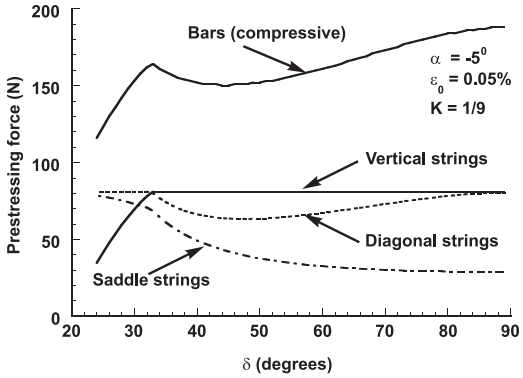


FIGURE 17.47 Prestressing force in the members as a function of δ.

Figure 17.46 shows the member forces as a function of α with δ= 35°, whereas Figure 17.47 shows the member forces as a function of δ with α= –5°. Both of the figures are obtained for K = 1/9, and the prestressing force in the strings is equal to the force due to a maximum prestrain in the strings ε_0 = 0.05% applied to the string which experiences maximum prestressing force. It is seen in both of the figures that for large negative α, the prestressing force in the saddle strings and the diagonal strings decreases with an increase in the negative α. Finally, for α below certain values, the prestressing forces in the saddle and diagonal strings become small enough to violate the definition of existence of tensegrity (i.e., all strings in tension and all bars in compression).

A similar trend is noted in the case of the vertical strings also. As seen in Figure 17.47, the force in the vertical strings decreases with a decrease in δ for small δ. Finally, for δ below certain values, the prestressing forces in the vertical strings become small enough to violate the definition of the existence of tensegrity. This explains the lower limits of the angles α and δ satisfying the tensegrity conditions.

Figures 17.46 and 17.47 show very remarkable changes in the load-sharing mechanism between the members with an increase in positive α and with an increase in δ. It is seen in Figure 17.46 that as α is gradually changed from a negative value toward a positive one, the prestressing force in the saddle strings increases, whereas the prestressing force in the vertical strings decreases. These trends continue up to α= 0°, when the prestressing force in both the diagonal strings and the saddle strings is equal and that in the vertical strings is small. For α < 0°, the force in the diagonal strings is always greater than that in the saddle strings. However, for α > 0°, the force in the diagonal strings decreases and is always less than the force in the saddle strings. The force in the vertical strings is the greatest of all strings.

© 2002 by CRC Press LLC

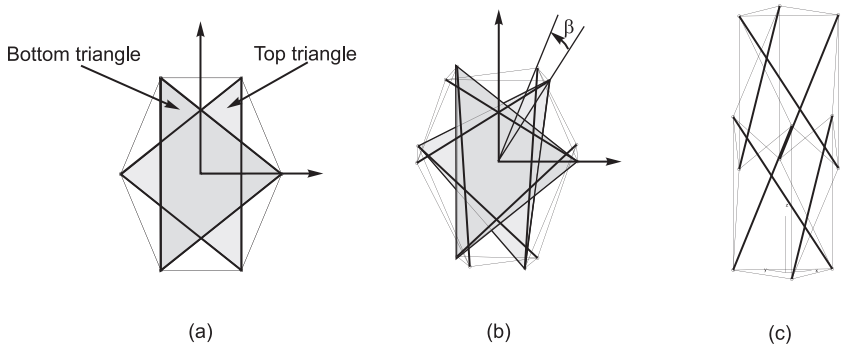


FIGURE 17.48 Rotation of the top triangle with respect to the bottom triangle for a two-stage cylindrical hexagonal 3-bar SVD tensegrity. (a) Top view when $\beta = 0$, (b) top view with β, and (c) elevation.

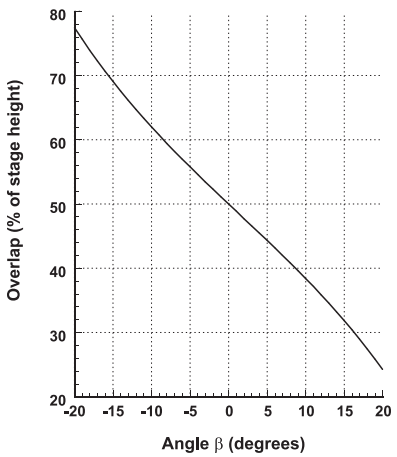


FIGURE 17.49 Existence conditions for a cylindrical two-stage 3-bars SVD tensegrity with respect to the rotation angle of the top triangle (anticlockwise β is positive).

Figure 17.45 showing all the possible configurations of a two-stage tensegrity can be quite useful in designing a deployable tensegrity beam made of many stages. The deployment of a beam with many stages can be achieved by deploying two stages at a time.

The existence conditions for a regular hexagonal cylinder (beam) made of two stages for which one of the end triangles is assumed to be rotated by an angle β about its mean position, as shown in Figure 17.48, is studied next. The mean position of the triangle is defined as the configuration when $\beta = 0$ and all of the nodal points of the bars line up in a straight line to form a regular hexagon, as shown in Figure 17.48. As is seen in Figure 17.49, it is possible to rotate the top triangle merely by satisfying the equilibrium conditions for the two-stage tensegrity. It is also seen that the top triangle can be rotated merely by changing the overlap between the two stages. This information can be quite useful in designing a Stewart platform-type structure.

17.4.3 Load-Deflection Curves and Axial Stiffness as a Function of the Geometrical Parameters

The load deflection characteristics of a two-stage 3-bar SVD-type tensegrity are studied next and the corresponding stiffness properties are investigated.

© 2002 by CRC Press LLC

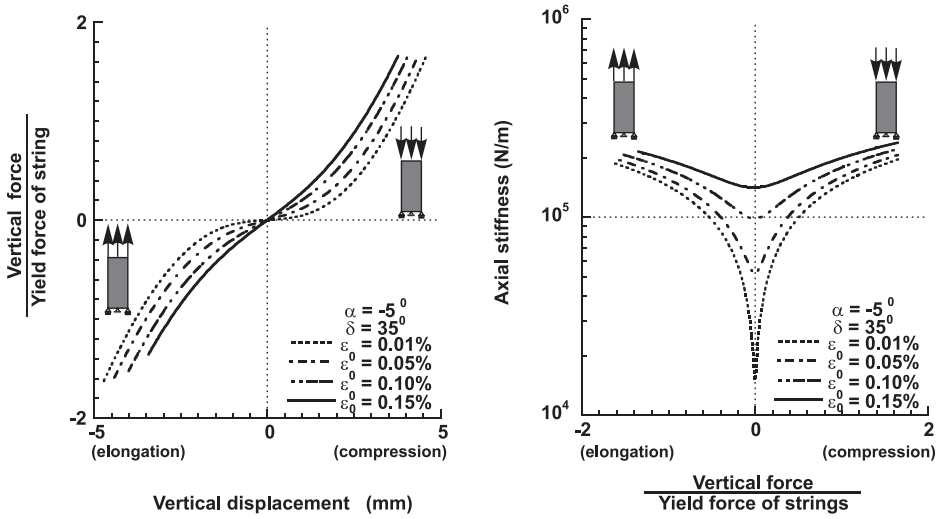


FIGURE 17.50 Load deflection curve and axial stiffness of a two-stage 3-bar SVD tensegrity subjected to axial loading.

Figure 17.50 depicts the load-deflection curves and the axial stiffness as functions of prestress, drawn for the case of a two-stage 3-bar SVD-type tensegrity subjected to axial loading. The axial stiffness is defined as the external force acting on the structure divided by the axial deformation of the structure. In another words, the stiffness considered here is the "secant stiffness."

Figure 17.50 shows that the tensegrity under axial loading behaves like a nonlinear spring and the nonlinear properties depend much on the prestress. The nonlinearity is more prominent when prestress is low and when the displacements are small. It is seen that the axial stiffnesses computed for both compressive and tensile loadings almost equal to each other for this particular case of a two-stage 3-bar SVD-type tensegrity. It is also seen that the axial stiffness is affected greatly by the prestress when the external forces are small (i.e., when the displacements are small), and prestress has an important role in increasing the stiffness of the tensegrity in the region of a small external load. However, as the external forces increase, the effect of the prestress becomes negligible.

The characteristics of the axial stiffness of the tensegrity as a function of the geometrical parameters (i.e., α, δ) are next plotted in Figure 17.51. The effect of the prestress on the axial stiffness is also shown in Figure 17.51. In obtaining the Figure 17.51, vertical loads were applied at the top nodes of the two-stage tensegrity. The load was gradually increased until at least one of the strings exceeded its elastic limit. As the compressive stiffness and the tensile stiffness were observed to be nearly equal to each other in the present example, only the compressive stiffness as a function of the geometrical parameters is plotted in Figure 17.51. The change in the shape of the tensegrity structure from a fat profile to an hourglass-like profile with the change in α is also shown in Figure 17.51(b).

The following conclusions can be drawn from Figure 17.51:

1. Figure 17.52(a) suggests that the axial stiffness increases with a decrease in the angle of declination δ (measured from the vertical axis).
2. Figure 17.51(b) suggests that the axial stiffness increases with an increase in the negative angle α. Negative α means a fat or beer-barrel-type structure whereas a positive α means an hourglass-type structure, as shown in Figure 17.51(b). Thus, a fat tensegrity performs better than an hourglass-type tensegrity subjected to compressive loading.
3. Figure 17.51(c) suggests that prestress has an important role in increasing the stiffness of the tensegrity in the region of small external loading. However, as the external forces are increased, the effect of the prestress becomes almost negligible.

© 2002 by CRC Press LLC

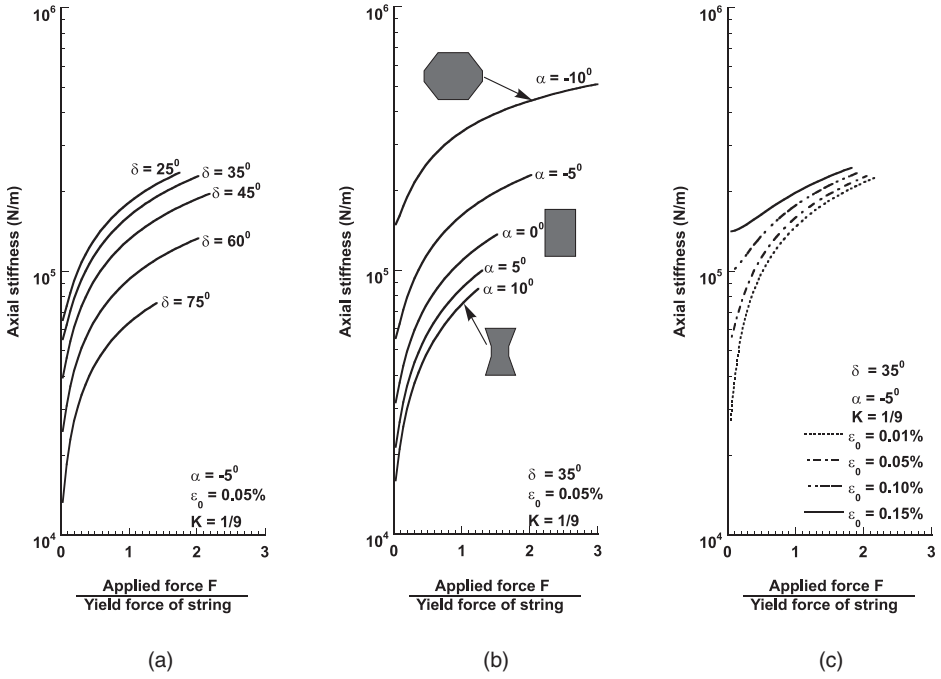


FIGURE 17.51 Axial stiffness of a two-stage 3-bar SVD tensegrity for different α, δ, and pretension.

17.4.4 Load-Deflection Curves and Bending Stiffness as a Function of the Geometrical Parameters

The bending characteristics of the two-stage 3-bar SVD tensegrity are presented in this section. The force is applied along the x-direction and then along the y-direction, as shown in Figure 17.52. The force is gradually applied until at least one of the strings exceeds its elastic limit.

The load deflection curves for the load applied in the lateral are plotted in Figure 17.52 as a function of the prestress. It was observed that as the load is gradually increased, one of the vertical strings goes slack and takes no load. Therefore, two distinct regions can be clearly identified in Figure 17.52. The first region is the one where none of the strings is slack, whereas the second region, marked by the sudden change in the slope of the load deflection curves, is the one in which at least one string is slack. It is seen in Figure 17.52 that in contrast to the response of the tensegrity subjected to the vertical axial loading, the bending response of the tensegrity is almost linear in the region of tensegrity without slack strings, whereas it is slightly nonlinear in the region of tensegrity with slack strings. The nonlinearity depends on the prestressing force. It is observed that the prestress plays an important role in delaying the onset of the slack strings.

The characteristics of the bending stiffness of the tensegrity as a function of the geometrical parameters (i.e., α, δ) are plotted next in Figures 17.53 and 17.54. Figure 17.53 is plotted for lateral force applied in the x-direction, whereas Figure 17.54 is plotted for lateral force applied in the y-direction. The effect of the prestress on the bending stiffness is also shown in Figures 17.53 and 17.54. The following conclusions about the bending characteristics of the two-stage 3-bar tensegrity could be drawn from Figures 17.53 and 17.54:

1. It is seen that the bending stiffness of the tensegrity with no slack strings is almost equal in both the x- and y-directions. However, the bending stiffness of the tensegrity with slack string is greater along the y-direction than along the x-direction.
2. The bending stiffness of a tensegrity is constant and is maximum for any given values of α, δ, and prestress when none of the strings are slack. However, as soon as at least one string goes

© 2002 by CRC Press LLC

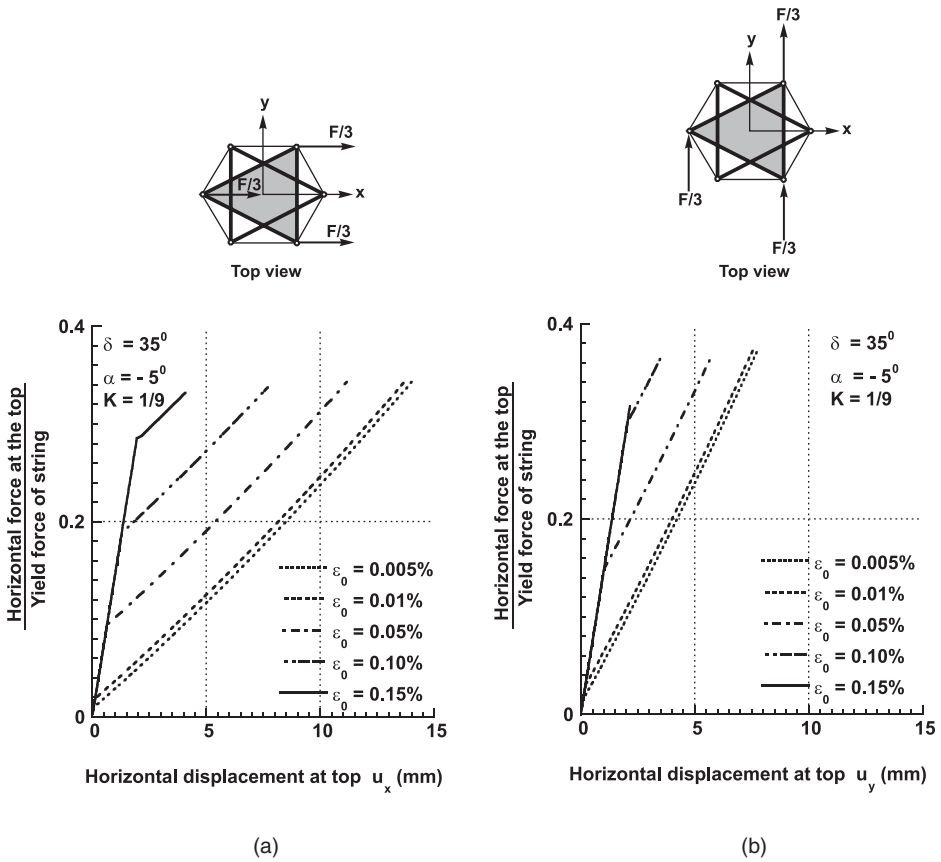


FIGURE 17.52 Load deflection curve of a two-stage 3-bar SVD tensegrity subjected to lateral loading, (a) loading along x-direction, and (b) loading along y-direction.

slack (marked by sudden drop in the stiffness curves in Figures 17.53 and 17.54), the stiffness becomes a nonlinear function of the external loading and decreases monotonically with the increase in the external loading. As seen in Figures 17.53 and 17.54, the onset of strings becoming slack, and hence the range of constant bending stiffness, is a function of α, δ, and prestress.

3. Figures 17.53(a) and 17.54(a) suggest that the bending stiffness of a tensegrity with no slack strings increases with the increase in the angle of declination δ (measured from the vertical axis). The bending stiffness of a tensegrity with a slack string, in general, increases with increase in δ. However, as seen in Figure 17.53(a), a certain δ exists beyond which the bending stiffness of a tensegrity with slack string decreases with an increase in δ. Hence, tensegrity structures have an optimal internal geometry with respect to the bending stiffness and other mechanical properties.
4. Figures 17.53(b) and 17.54(b) suggest that the bending stiffness increases with the increase in the negative angle α. As negative α means a fat or beer-barrel-type structure whereas a positive α means an hourglass-type structure, a fat tensegrity performs better than an hourglass-type tensegrity subjected to lateral loading.
5. Figures 17.53(a,b) and 17.54(a,b) indicate that both α and δ play a very interesting and important role in not only affecting the magnitude of stiffness, but also the onset of slackening of the strings (robustness to external disturbances). A large value of negative α and a large value of δ (in general) delay the onset of slackening of the strings, thereby increasing the range of constant bending stiffness. However, a certain δ exists for which the onset of the slack strings is maximum.

© 2002 by CRC Press LLC

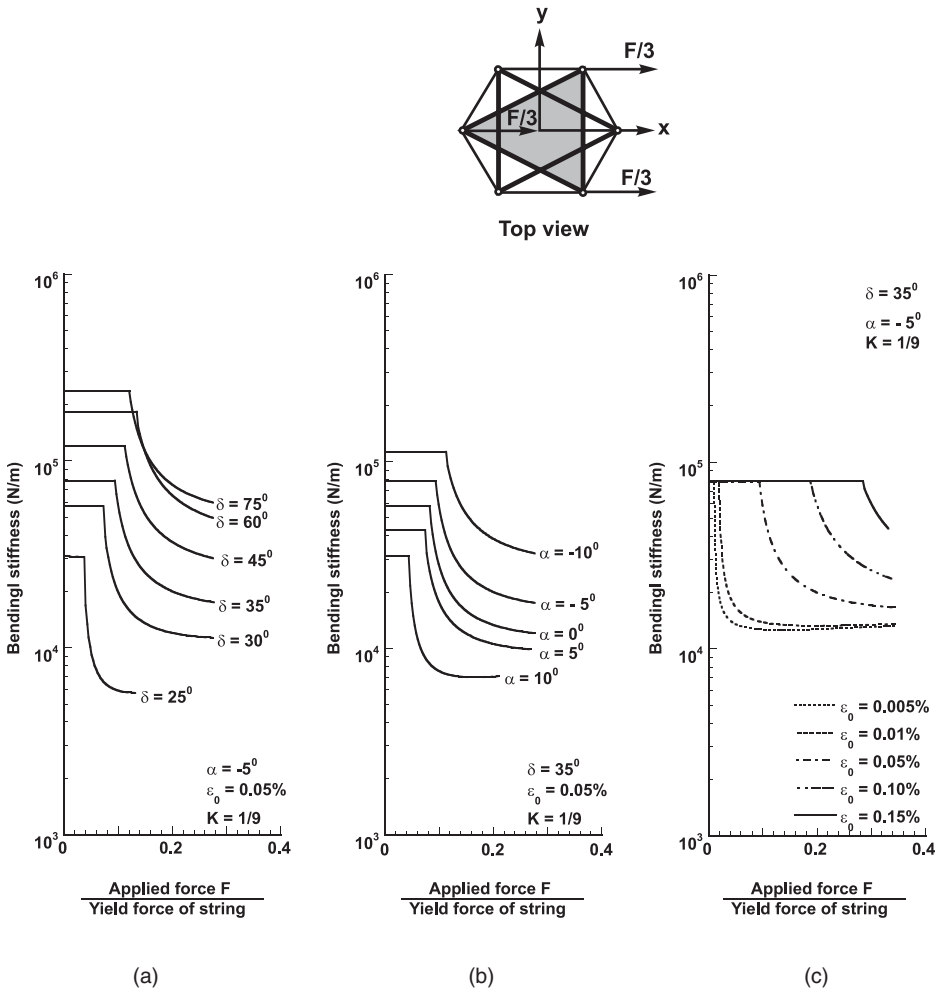


FIGURE 17.53 Bending stiffness of a two-stage 3-bar SVD tensegrity for different α, δ, and pretension. L-bar for all cases = 0.4 in.

6. Figures 17.53(c) and 17.54(c) suggest that prestress does not affect the bending stiffness of a tensegrity with no slack strings. However, prestress has an important role in delaying the onset of slack strings and thus increasing the range of constant bending stiffness.

17.4.5 Summary of 3-Bar SVD Tensegrity Properties

The following conclusions could be drawn from the present study on the statics of a two-stage 3-bar SVD-type tensegrities.

1. The tensegrity structure exhibits unique equilibrium characteristics. The self-stressed equilibrium condition exists only on a small subset of geometrical parameter values. This condition guarantees that the tensegrity is prestressable and that none of the strings is slack.
2. The stiffness (the axial and the bending) is a function of the geometrical parameters, the prestress, and the externally applied load. However, the effect of the geometrical parameters on the stiffness is greater than the effect of the prestress. The external force, on the other hand, does not affect the bending stiffness of a tensegrity with no slack strings, whereas it does affect the axial stiffness. The axial stiffness shows a greater nonlinear behavior even

© 2002 by CRC Press LLC

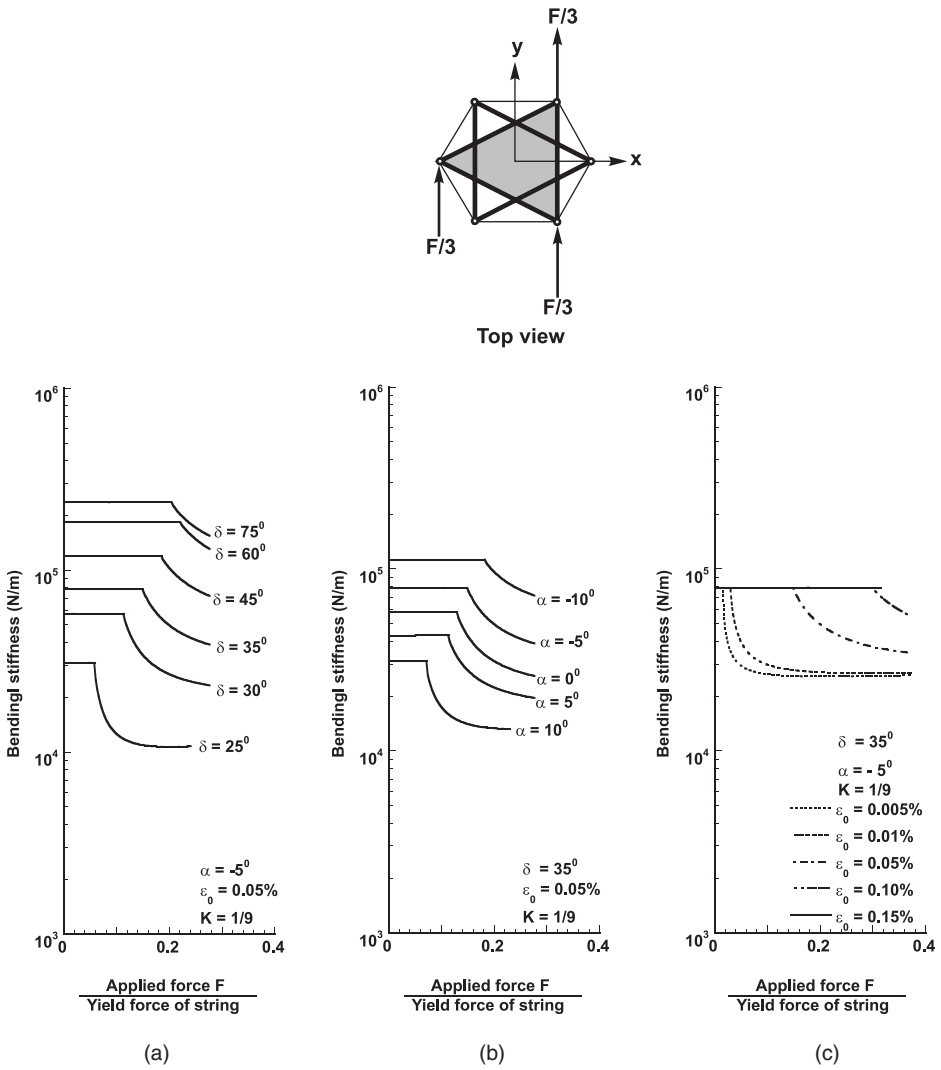


FIGURE 17.54 Bending stiffness of a two-stage 3-bar SVD tensegrity for different α, δ, and pretension.

up to the point when none of the strings are slack. The axial stiffness increases with an increase in the external loading, whereas the bending stiffness remains constant until at least one of the strings go slack, after which the bending stiffness decreases with an increase in the external loading.

3. Both the axial and the bending stiffness increase by making α more negative. That is, both the axial and the bending stiffness are higher for a beer-barrel-type tensegrity. The stiffness is small for an hourglass-type tensegrity.
4. The axial stiffness increases with a decrease in the vertical angle, whereas the bending stiffness increases with an increase in the vertical angle. This implies that the less the angle that the bars make with the line of action of the external force, the stiffer is the tensegrity.
5. Both the geometrical parameters α and δ, and prestress play an important role in delaying the onset of slack strings. A more negative α, a more positive δ, and prestress, all delay the onset of slack strings, as more external forces are applied. Thus, both α and δ also work as a hidden prestress. However, there lies a δ beyond which an increase in δ hastens the onset of slack strings, as more external force is applied.

© 2002 by CRC Press LLC

17.5 Concluding Remarks

Tensegrity structures present a remarkable blend of geometry and mechanics. Out of various available combinations of geometrical parameters, only a small subset exists that guarantees the existence of the tensegrity. The choice of these parameters dictates the mechanical properties of the structure. The choice of the geometrical parameters has a great influence on the stiffness. Pretension serves the important role of maintaining stiffness until a string goes slack. The geometrical parameters not only affect the magnitude of the stiffness either with or without slack strings, but also affect the onset of slack strings. We now list the major findings of this chapter.

17.5.1 Pretension vs. Stiffness Principle

This principle states that increased pretension increases robustness to uncertain disturbances. More precisely, for all situations we have seen (except for the $C4T2$):

When a load is applied to a tensegrity structure, the stiffness does not decrease as the loading force increases unless a string goes slack.

The effect of the pretension on the stiffness of a tensegrity without slack strings is almost negligible. The bending stiffness of a tensegrity without slack strings is not affected appreciably by prestress.

17.5.2 Small Control Energy Principle

The second principle is that the shape of the structure can be changed with very little control energy. This is because shape changes are achieved by changing the equilibrium of the structure. In this case, control energy is not required to hold the new shape. This is in contrast to the control of classical structures, where shape changes required control energy to work against the old equilibrium.

17.5.3 Mass vs. Strength

This chapter also considered the issue of strength vs. mass of tensegrity structures. We found planar examples to be very informative. We considered two types of strength: the size of bending forces and the size of compressive forces required to break the object. We studied the ratio of bending strength to mass and compression strength to mass. We compared this for two planar structures, one the $C2T4$ unit and the other a $C4T1$ unit, to a solid rectangular bar of the same mass.

We find:

- Reasonably constructed $C2T4$ units are stronger in bending than a rectangular bar, but they are weaker under compression.
- The $C2T4$ has worse strength under compression than a solid rectangular bar.
- The simple analysis we did indicates that $C4T2$ and $C4T1$ structures with reasonably chosen proportions have larger compression strength-to-mass ratios than a solid bar.
- On the other hand, a $C4T1$, while strong (not easily broken), need not be an extremely stiff structure.
- $C4T2$ and $C4T1$ structures can be designed for minimal mass subjected to a constraint on both strength and stiffness.

It is possible to amplify the effects stated above by the use of self-similar constructions:

- **A 2D Tensegrity Beam.** After analyzing a $C2T4$ tensegrity unit, we lay n of them side by side to form a beam. In principle, we find that one can build beams with arbitrarily great bending strength. In practice, this requires more study. However, the favorable bending properties found for $C2T4$ bode well for tensegrity beams.

© 2002 by CRC Press LLC

- **A 2D Tensegrity Column.** We take the *C*4*T*2 structure and replace each bar with a smaller *C*4*T*1 structure, then we replace each bar of this new structure by a yet smaller *C*4*T*1 structure. In principle, such a self-similar construction can be repeated to any level. Assuming that the strings do not fail and have significantly less mass than the bars, we find that we have a class of tensegrity structures with unlimited compression strength-to-mass ratio. Further issues of robustness to lateral and bending forces have to be investigated to ensure practicality of such structures.

The total mass including string and bars (while preserving strength) can be minimized by a finite numer of self-similar iterations, and the number of iterations to achieve minimal mass is usually quite small (less that 10). This provides an optimization of tensegrity structures that is analytically resolved and is much easier and less complex than optimization of classical structures. We emphasize that the implications of overlapping of the bars were not seriously studied.

For a special range of geometry, the stiffness-to-mass ratio increases with self-similar iterations. For the remaining range of geometry the stiffness-to-mass ratio decreases with self-similar iterations. For a very specific choice of geometry, the stiffness-to-mass ratio remains constant with self-similar iterations.

Self-similar steps can preserve strength while reducing mass, but cannot preserve stiffness while reducing mass. Hence, a desired stiffness bound and reconciliation of overlapping bars will dictate the optimal number of iterations.

17.5.4 A Challenge for the Future

In the future, the grand challenge with tensegrity structures is to find ways to choose material and geometry so that the thermal, electrical, and mechanical properties are specified. The tensegrity structure paradigm is very promising for the integration of these disciplines with control, where either strings or bars can be controlled.

Acknowledgment

This work received major support from a DARPA grant monitored by Leo Christodoulou. We are also grateful for support from DARPA, AFOSR, NSF, ONR, and the Ford Motor Company.

Appendix 17.A Nonlinear Analysis of Planar Tensegrity

17.A.1 Equations of Static Equilibrium

17.A.1.1 Static Equilibrium under External Forces

A planar tensegrity under external forces is shown in Figure 17.A.1, where F_i are the external forces and t_i represent the internal forces in the members of the tensegrity units. Note that **t** represents the net force in the members which includes the pretension and the force induced by the external forces. The sign convention adopted herein is also shown in Figure 17.A.1, where t_{ki} represents the member force t acting at the i-th node of the member k. We assume that $i < j$ and $t_{ki} = -t_{kj}$. With this convention, we write the force equilibrium equations for the planar tensegrity.

The equilibrium of forces in the x-direction acting on the joints yields the following equations

$$\begin{aligned} t_{1i}\cos\delta_{x1} + t_{4i}\cos\delta_{x4} + t_{6i}\cos\delta_{x6} &= F_1, \\ t_{1j}\cos\delta_{x1} + t_{2i}\cos\delta_{x2} + t_{5i}\cos\delta_{x5} &= -F_2, \\ t_{2i}\cos\delta_{x2} + t_{3i}\cos\delta_{x3} + t_{6j}\cos\delta_{x6} &= -F_3, \\ t_{3j}\cos\delta_{x3} + t_{4j}\cos\delta_{x4} + t_{5j}\cos\delta_{x5} &= F_4. \end{aligned} \tag{17.A.1}$$

© 2002 by CRC Press LLC

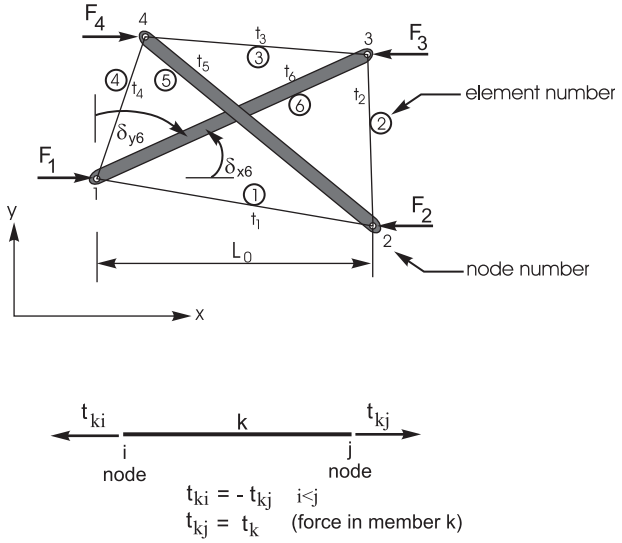


FIGURE 17.A.1 Forces acting on a planar tensegrity and the sign convention used.

Similarly, the equilibrium of forces in the y-direction acting on the joints yields the following equations

$$\begin{aligned} t_{1i}\cos\delta_{y1} + t_{4i}\cos\delta_{y4} + t_{6i}\cos\delta_{y6} &= 0,\\ t_{1j}\cos\delta_{y1} + t_{2i}\cos\delta_{y2} + t_{5i}\cos\delta_{y5} &= 0,\\ t_{2i}\cos\delta_{y2} + t_{3i}\cos\delta_{y3} + t_{6j}\cos\delta_{y6} &= 0,\\ t_{3j}\cos\delta_{y3} + t_{4j}\cos\delta_{y4} + t_{5j}\cos\delta_{y5} &= 0. \end{aligned} \tag{17.A.2}$$

In the above equations, $\cos\delta_{xk}$ represents the direction cosine of member k taken from the x-axis, whereas $\cos\delta_{yk}$ represents the direction cosine of member k taken from the y-axis.

The above equations can be rearranged in the following matrix form:

$$\mathbf{At} = \mathbf{f}, \tag{17.A.3}$$

where $\mathbf{t}$ is a vector of forces in the members and is given by $\mathbf{t}^T = [t_1\ t_2\ t_3\ t_4\ t_5\ t_6]$, matrix $\mathbf{A}$ (of size 8×6) is the equilibrium matrix, and $\mathbf{f}$ is a vector of nodal forces. For convenience, we arrange $\mathbf{t}$ such that the forces in the bars appear at the top of the vector, i.e.,

$$\mathbf{t}^T = [\mathbf{t}_{bars}\ \mathbf{t}_{strings}] = [t_5\ t_6\ t_1\ t_2\ t_3\ t_4]. \tag{17.A.4}$$

Matrix $\mathbf{A}$ and vector $\mathbf{f}$ are given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{C}^T & \vdots & \mathbf{0} \\ \cdots & \cdots & \cdots \\ \mathbf{0} & \vdots & \mathbf{C}^T \end{bmatrix} \begin{bmatrix} \mathbf{H}_x \\ \cdots \\ \mathbf{H}_y \end{bmatrix}, \quad \mathbf{f} = \begin{Bmatrix} \mathbf{f}_x \\ \cdots \\ \mathbf{f}_y \end{Bmatrix}. \tag{17.A.5}$$

In the above equation, matrices $\mathbf{H}_x$ and $\mathbf{H}_y$ are diagonal matrices containing the direction cosines of each member taken from the x-axis or y-axis, respectively, i.e., $H_{xii} = \cos\delta_{xi}$ and $H_{yii} = \cos\delta_{yi}$.

© 2002 by CRC Press LLC

Similar to the arrangement of $\mathbf{t}$, $\mathbf{H}_x$ and $\mathbf{H}_y$ are also arranged such that the direction cosines of bars appear at the top of $\mathbf{H}_x$ and $\mathbf{H}_y$ whereas the direction cosines of strings appear at the bottom of $\mathbf{H}_x$ and $\mathbf{H}_y$. Vectors $\mathbf{f}_x$ and $\mathbf{f}_y$ are the nodal forces acting on the nodes along x- and y-axes, respectively. Matrix $\mathbf{C}$ is a 6×4 (number of members $\times$ number of nodes) matrix. The k-th row of matrix $\mathbf{C}$ contains -1 (for i-th node of the k-th member), $+1$ (for j-th node of the k-th member) and 0. Matrix $\mathbf{C}$ for the present case is given as

$$\mathbf{C} = \begin{bmatrix} 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}0\\0\end{matrix}}\right\} \text{Bars} \\ \left.\vphantom{\begin{matrix}0\\0\\0\\0\end{matrix}}\right\} \text{Strings} \end{matrix} \tag{17.A.6}$$

It should be noted here that matrix $\mathbf{A}$ is a nonlinear function of the geometry of the tensegrity unit, the nonlinearity being induced by the matrices $\mathbf{H}_x$ and $\mathbf{H}_y$ containing the direction cosines of the members.

17.A.2 Solution of the Nonlinear Equation of Static Equilibrium

Because the equilibrium equation given in (17.A.3) is nonlinear and also $\mathbf{A}$ (of size 8×6) is not a square matrix, we solve the problem in the following way.

Let $\tilde{\mathbf{t}}$ be the member forces induced by the external force $\mathbf{f}$, then from (17.A.3)

$$\begin{aligned} &\mathbf{A}\mathbf{t} = \mathbf{f} \\ \Rightarrow\ &\mathbf{A}(\mathbf{t}_0 + \tilde{\mathbf{t}}) = \mathbf{f} \\ \Rightarrow\ &\mathbf{A}\tilde{\mathbf{t}} = \mathbf{f} - \mathbf{A}\mathbf{t}_0 \\ \Rightarrow\ &\mathbf{A}\mathbf{K}\mathbf{e} = \mathbf{f} - \mathbf{A}\mathbf{t}_0 \end{aligned} \tag{17.A.7}$$

where $\mathbf{e}$ is the deformation from the initial prestressed condition of each member, and from Hooke's law $\tilde{\mathbf{t}} = \mathbf{K}\mathbf{e}$, where $\mathbf{K}$ is a diagonal matrix of size 6×6, with $K_{ii} = (EA)_i/L_i$. $(EA)_i$ and L_i are the axial rigidity and the length of the i-th member. Note that $\mathbf{A}$ expressed above is composed of both the original $\mathbf{A}_0$ and the change in $\mathbf{A}_0$ caused by the external forces $\mathbf{f}$.

$$\mathbf{A} = \mathbf{A}_0 + \tilde{\mathbf{A}} \tag{17.A.8}$$

where $\tilde{\mathbf{A}}$ is the change in $\mathbf{A}_0$ caused by the external forces $\mathbf{f}$.

The nonlinear equation given above can be linearized in the neighborhood of an equilibrium. In the neighborhood of the equilibrium, we have the linearized relationship,

$$\mathbf{e}_k = \mathbf{A}_k^T \mathbf{u}_k \tag{17.A.9}$$

Let the external force $\mathbf{f}$ be gradually increased in small increments ($\mathbf{f}_k = \mathbf{f}_{k-1} + \Delta\mathbf{f}$ at the k-th step), and the equilibrium of the planar tensegrity be satisfied for each incremental force, then (17.A.7) can be written as

$$\mathbf{A}(\mathbf{u}_k)\mathbf{K}\mathbf{A}\ (\mathbf{u}_k)^T\ \mathbf{u}_k = \mathbf{f}_k - \mathbf{A}(\mathbf{u}_k)\mathbf{t}_0 \tag{17.A.10}$$

The standard Newton–Raphson method can now be used to evaluate $\mathbf{u}_k$ of (17.A.10) for each incremental load step $\Delta\mathbf{f}_k$. The external force is gradually applied until it reaches its specific value

© 2002 by CRC Press LLC

and $\mathbf{u}_k$ is evaluated at every load step. Matrix $\mathbf{A}$, which is now a nonlinear function of $\mathbf{u}$, is updated during each load step.

To compute the external force that would be required to buckle the bars in the tensegrity unit, we must estimate the force being transferred to the bars. The estimation of the compressive force in the bars following full nonlinear analysis can be done numerically. However, in the following we seek to find an analytical expression for the compressive force in the bars. For this we adopt a linear and small displacement theory. Thus, the results that follow are valid only for small displacement and small deformation.

© 2002 by CRC Press LLC

Appendix 17.B Linear Analysis of Planar Tensegrity

17.B.1 EI of the Tensegrity Unit with Slack Top String

17.B.1.1 Forces in the Members

A tensegrity with a slack top string does not have prestress. As mentioned earlier, we adopt the small displacement assumptions, which imply that the change in the angle δ due to the external forces is negligible. Therefore, in the following, we assume that δ remains constant. The member forces in this case are obtained as

$$\begin{aligned} t_1 &= 2F, \\ t_2 &= F\tan\delta, \\ t_3 &= 0, \\ t_4 &= F\tan\delta, \\ t_5 &= -\frac{F}{\cos\delta}, \\ t_6 &= -\frac{F}{\cos\delta}. \end{aligned} \tag{17.B.1}$$

The strain energy in each of the members is computed as

$$\begin{aligned} U_1 &= \frac{1}{2}\frac{4L_0F^2}{(EA)_s}, \\ U_2 &= \frac{1}{2}\frac{L_0F^2\tan^3\delta}{(EA)_s}, \\ U_3 &= 0, \\ U_4 &= \frac{1}{2}\frac{L_0F^2\tan^3\delta}{(EA)_s}, \\ U_5 &= \frac{1}{2}\frac{L_0F^2}{(EA)_b\cos^3\delta}, \\ U_6 &= \frac{1}{2}\frac{L_0F^2}{(EA)_b\cos^3\delta}. \end{aligned} \tag{17.B.2}$$

The total strain energy is then obtained as

$$U = \sum_i U_i = \frac{1}{2}\frac{L_0}{(EA)_s}\frac{2F^2}{\cos^3\delta}[\sin^3\delta + 2\cos^3\delta + K], \tag{17.B.3}$$

where K is defined as

$$K = \frac{(EA)_s}{(EA)_b}. \tag{17.B.4}$$

© 2002 by CRC Press LLC

Thus, large values of K mean that the strings are stiffer than the bars, whereas small values of K mean that the bars are stiffer than the strings. $K \to 0$ means bars are rigid.

17.B.1.2 External Work and Displacement

External work W is given by

$$W = 4\frac{1}{2}Fu = 2Fu, \tag{17.B.5}$$

where u is the displacement as shown in Figure 17.11.

Equating the total strain energy given by (17.B.3) to the work done by the external forces given by (17.B.5), and then solving for u yields

$$u = \frac{FL_0}{2(EA)_s \cos^3\delta}[\sin^3\delta + 2\cos^3\delta + K]. \tag{17.B.6}$$

17.B.1.3 Effective EI

Because $EI = M\rho$, we have

$$EI = FL_0 \tan\delta\left(\frac{L_0}{2}\right)^2 \tan\delta\frac{1}{u}. \tag{17.B.7}$$

Substitution of $\tilde{u}$ from (17.B.6) into (17.B.7) yields

$$EI = \frac{1}{2}\frac{L_0^2(EA)_s \sin^2\delta\cos\delta}{(\sin^3\delta + 2\cos^3\delta + K)}. \tag{17.B.8}$$

Substituting $L_0 = L_{bar}\cos\delta$ in (17.B.7) and (17.B.8) yields the following expressions for the equivalent bending rigidity of the planar section in terms of the length of the bars L_{bar},

$$EI = FL_{bar}\sin\delta\left(\frac{L_{bar}}{2}\right)^2 \cos\delta\sin\delta\frac{1}{u}. \tag{17.B.9}$$

or equivalently,

$$EI = \frac{1}{2}\frac{L_{bar}{}^2(EA)_s \sin^2\delta\cos^3\delta}{(\sin^3\delta + 2\cos^3\delta + K)}. \tag{17.B.10}$$

© 2002 by CRC Press LLC

Appendix 17.C Derivation of Stiffness of the $C4T1^i$ Structure

17.C.1 Derivation of Stiffness Equation

For a $C4T1^i$ structure under the buckling load F, the compressive load of bar in the i-th iteration is

$$F_i = \frac{F}{\prod_{j=1}^{i} 2\cos\delta_j}. \tag{17.C.1}$$

Similarly, the tension of strings in the i-th iteration is

$$t_j = \frac{2F\sin\delta_j}{\prod_{s=1}^{j} 2\cos\delta_s} \text{ for } j = 1,2,3,...,i-1,i. \tag{17.C.2}$$

So, the buckling load F can be written in terms of any one of the compressive bar loads or tension of strings in i-th iteration

$$F = F_i \prod_{s=1}^{i} 2\cos\delta_s = \frac{t_j}{2\sin\delta_j} \prod_{p=1}^{j} 2\cos\delta_p \quad \text{for } j = 1,2,3,...,i-1,i. \tag{17.C.3}$$

From the geometry of the structure,

$$\begin{aligned} \sin\delta_j &= \frac{L_{tj}}{2L_j}, \\ \cos\delta_j &= \frac{L_{j-1}}{2L_j}. \end{aligned} \tag{17.C.4}$$

Equation (17.C.3) can be simplified to

$$F = F_i \frac{L_0}{L_i} = t_j \frac{L_0}{L_{tj}}. \tag{17.C.5}$$

From this,

$$\frac{F}{L_0} = \frac{F_i}{L_i} = \frac{t_j}{L_{tj}} = \frac{t_1}{L_{t1}}. \tag{17.C.6}$$

This means the force-to-length ratio of every compressive or tensile members is the same. It is assumed that all the bars and strings have constant stiffness and, hence, are linear. With this assumption,

$$\begin{aligned} F_i &= k_{bi}(L_{i0} - L_i), \\ t_j &= k_{tj}(L_{tj} - L_{tj0}). \end{aligned} \tag{17.C.7}$$

So (17.C.6) becomes

$$k_{bi}\left(\frac{L_{i0}}{L_i} - 1\right) = k_{tj}\left(1 - \frac{L_{tj0}}{L_{tj}}\right) = k_{t1}\left(1 - \frac{L_{t10}}{L_{t1}}\right). \tag{17.C.8}$$

© 2002 by CRC Press LLC

Taking the infinitesimal of all the length quantities yields

$$-k_{bi}\frac{L_{i0}}{L_i^2}dL_i = k_{tj}\frac{L_{tj0}}{L_{tj}^2}dL_{tj} = k_{t1}\frac{L_{t10}}{L_{t1}^2}dL_{t1},$$

and hence,

$$\begin{aligned} dL_i &= -\frac{k_{t1}L_{t10}}{k_{bi}L_{i0}}\frac{L_i^2}{L_{t1}^2}dL_{t1}, \\ dL_{tj} &= \frac{k_{t1}L_{t10}}{k_{tj}L_{tj0}}\frac{L_{tj}^2}{L_{t1}^2}dL_{t1}. \end{aligned} \tag{17.C.9}$$

From the geometry of the structure,

$$\begin{aligned} L_0^2 &= 4L_1^2 - L_{t1}^2 \\ &= 4(4L_2^2 - L_{t2}^2) - L_{t1}^2 \\ &= \ldots \\ &= 4^i L_i^2 - \sum_{j=1}^{i} 4^{j-1} L_{tj}^2. \end{aligned} \tag{17.C.10}$$

Taking the infinitesimal of (17.C.10), noting that L_i is length of bars, yields

$$dL_0 = 4^i \frac{L_i}{L_0}dL_i - \frac{1}{L_0}\sum_{j=1}^{i} 4^{j-1} L_{tj}dL_{tj}. \tag{17.C.11}$$

Combining the (17.C.11) with (17.C.9) yields

$$\frac{dL_0}{dL_{t1}} = \frac{k_{t1}L_{t10}}{L_0 L_{t1}^2}\left(4^i \frac{L_i^3}{k_{bi}L_{i0}} + \sum_{j=1}^{i} 4^{j-1}\frac{L_{tj}^3}{k_{tj}L_{tj0}}\right). \tag{17.C.12}$$

From (17.C.5), it is natural to choose F in terms of the tension in the first iteration, i.e.,

$$F = t_1 \frac{L_0}{L_{t1}} = k_{t1}\left(1 - \frac{L_{t10}}{L_{t1}}\right)L_0.$$

The derivative of F w.r.t. L_0 yields

$$\frac{dF}{dL_0} = k_{t1}\left(1 - \frac{L_{t10}}{L_{t1}}\right) + k_{t1}L_0\frac{L_{t10}}{L_{t1}^2}\frac{dL_{t1}}{dL_0}. \tag{17.C.13}$$

With (17.C.12), the stiffness of $C4T1^i$ will be

$$\begin{aligned} K_i &= -\frac{dF}{dL_0} = k_{t1}\left(\frac{L_{t10}}{L_{t1}} - 1\right) + L_0^2\left(4^i \frac{L_i^3}{k_{bi}L_{i0}} + \sum_{j=1}^{i} 4^{j-1}\frac{L_{tj}^3}{k_{tj}L_{tj0}}\right)^{-1} \\ &= k_{t1}\left\{\left(\frac{L_{t10}}{L_{t1}} - 1\right) + \left[4^i \frac{k_{t1}}{k_{bi}}\frac{L_i}{L_{i0}}\left(\frac{L_i}{L_0}\right)^2 + \sum_{j=1}^{i} 4^{j-1}\frac{k_{t1}}{k_{tj}}\frac{L_{tj}}{L_{tj0}}\left(\frac{L_{tj}}{L_0}\right)^2\right]^{-1}\right\}. \end{aligned} \tag{17.C.14}$$

© 2002 by CRC Press LLC

17.C.2 Some Mathematical Relations in Buckling Design

In the strength-preserving design, the $C4T1^i$ system is designed to buckle at the same load as the original bar $C4T1^0$. The angles δ_j, where $j = 1, 2, \ldots, i–1, i$ are free variables to be specified to fix the geometry. Therefore, it is important to find out all the lengths and ratio quantities in terms of these angles.

17.C.2.1 Length of Structure and Strings

From the geometry of the structure, it can be shown that

$$\begin{aligned} L_0 &= L_i \prod_{s=1}^{i} 2\cos\delta_s, \\ L_{tj} &= 2L_i \sin\delta_j \prod_{s=j+1}^{i} 2\cos\delta_s. \end{aligned} \tag{17.C.15}$$

17.C.2.2 Computing the Stiffness Ratio of Strings, $\frac{k_{ts}}{k_{tj}}$ *for* $s, j = 1,2,3,..., i-1, i$

Consider the ratio

$$\begin{aligned} \frac{k_{t(j+1)}}{k_{tj}} &= \frac{E_{t(j+1)}A_{t(j+1)}}{L_{t(j+1)}} \frac{L_{tj}}{E_{tj}A_{tj}} \\ &= \frac{E_{t(j+1)}}{E_{tj}} \left(\frac{\pi r_{t(j+1)}^2}{\pi r_{tj}^2} \right) \left(\frac{L_{tj}}{L_{t(j+1)}} \right). \end{aligned}$$

From (17.109)

$$\frac{k_{t(j+1)}}{k_{tj}} = \frac{E_{t(j+1)}}{E_{tj}} \left(\frac{\sigma_{tj} t_{j+1}}{\sigma_{t(j+1)} t_j} \right) \left(\frac{L_{tj}}{L_{t(j+1)}} \right).$$

With (17.C.2) and (17.C.15), the ratio can be simplified to

$$\frac{k_{t(j+1)}}{k_{tj}} = \frac{E_{t(j+1)}}{E_{tj}} \left(\frac{\sigma_{tj}}{\sigma_{t(j+1)}} \right).$$

From this,

$$\frac{k_{ts}}{k_{tj}} = \frac{E_{ts}\sigma_{tj}}{E_{tj}\sigma_{ts}}. \tag{17.C.16}$$

In particular, if $E_{tj} = E_t$ and $\sigma_{tj} = \sigma_t$, then

$$\frac{k_{ts}}{k_{tj}} = 1. \tag{17.C.17}$$

© 2002 by CRC Press LLC

So, in the strength-preserving design, if the same material is used, then all strings have the same stiffness.

17.C.2.3 Computing the Stiffness Ratio of String to Bar

$$\frac{k_{tj}}{k_{bi}} \text{ where } j = 1, 2, 3, ..., i-1, i$$

The stiffness of bar and strings are defined by

$$k = \frac{EA}{L},$$

where E is the Young's modulus, A is the cross-section area, and L is the length of bar or strings at the buckling load. With this definition and (17.C.16),

$$\frac{k_{tj}}{k_{bi}} = \frac{k_{tj}}{k_{ti}} \frac{k_{ti}}{k_{bi}} = \frac{E_{tj}\sigma_{ti}}{E_{ti}\sigma_{tj}} \frac{E_{ti}\pi r_{ti}^2}{L_{ti}} \frac{L_i}{E_i \pi r_i^2}$$

$$= \frac{E_{tj}\sigma_{ti}}{E_i \sigma_{tj}} \left(\frac{1}{L_{ti} L_i} \right) \frac{L_i^2}{r_i^2} r_{ti}^2 .$$

From (17.106), (17.109), and (17.C.15),

$$\frac{k_{tj}}{k_{bi}} = \frac{E_{tj}\sigma_{ti}}{E_i \sigma_{tj}} \left(\frac{t_i}{\pi \sigma_{ti}} \right) \left(\frac{4 l_i^2}{2 L_i^2 \sin \delta_i} \right)$$

$$= \frac{E_{tj}\sigma_{ti}}{E_i \sigma_{tj}} \left(\frac{2F \sin \delta_i}{\pi \sigma_{ti} \Pi_{s=1}^{i} 2 \cos \delta_s} \right) \left(\frac{2}{L_i^2 \sin \delta_i} \right) \left(\frac{E_i}{E_0} \right)^{\frac{1}{2}} \left(\frac{1}{\Pi_{p=1}^{i} 2 \cos \delta_p} \right)^{\frac{1}{2}} l_0^2 .$$

Substitute F from (17.66) into the above equation to obtain

$$\frac{k_{tj}}{k_{bi}} = \frac{E_{tj}\pi^2}{16 \sigma_{tj} l_0^2} \sqrt{\frac{E_0}{E_i}} \left(\prod_{s=1}^{i} 2 \cos \delta_s \right)^{\frac{1}{2}} . \tag{17.C.18}$$

For some materials of bars and strings, (17.C.18) reduces to

$$\frac{k_{tj}}{k_{bi}} = \frac{E_{tj}\pi^2}{16 \sigma_{tj} l_0^2} \left(\prod_{s=1}^{i} 2 \cos \delta_s \right)^{\frac{1}{2}} . \tag{17.C.19}$$

17.C.2.4 Computing the Rest Length-to-Length Ratio of Strings, L_{tj0}/L_{tj}

The tension in the strings is given by

© 2002 by CRC Press LLC

$$t_j = k_{tj}\,(L_{tj} - L_{tj0}),$$

$$t_j = \frac{E_{tj}\pi r_{tj}^2}{L_{tj}}(L_{tj} - L_{tj0}).$$

From (17.109),

$$t_j = \frac{E_{tj}t_j}{L_{tj}\sigma_{tj}}(L_{tj} - L_{tj0}).$$

So,

$$\frac{L_{tj0}}{L_{tj}} = 1 - \frac{\sigma_{tj}}{E_{tj}}. \tag{17.C.20}$$

17.C.2.5 Computing the Rest Length to Length Ratio of Bars, L_{i0}/L_i

From (17.C.6),

$$\frac{F_i}{L_i} = \frac{t_1}{L_{t1}}$$

Hence,

$$k_{bi}\left(\frac{L_{i0}}{L_i} - 1\right) = k_{t1}\left(1 - \frac{L_{t10}}{L_{t1}}\right)$$

leading to

$$\frac{L_{io}}{L_i} = 1 + \frac{k_{t1}}{k_{bi}}\left(1 - \frac{L_{t10}}{L_{t1}}\right). \tag{17.C.21}$$

Using (17.C.20) and (17.C.21) reduces to

$$\frac{L_{i0}}{L_i} = 1 + \frac{k_{t1}}{k_{bi}}\frac{\sigma_{t1}}{E_{t1}}. \tag{17.C.22}$$

17.C.2.6 Computing the String Stiffness, k_{t1}

Recall that the string stiffness is given by

$$k_{t1} = \frac{E_{t1}\pi r_{t1}^2}{L_{t1}}.$$

$$= 1 + \frac{t1^2}{161_0^2}$$

Using (17.66), (17.108), and (17.109) yields

© 2002 by CRC Press LLC

$$
\begin{aligned}
k_{t1} &= \frac{E_{t1}}{L_{t1}} \frac{t_1}{\sigma_{t1}} \\
&= \frac{E_{t1}}{\sigma_{t1}} \frac{F \tan \delta_1}{L_0 \tan \delta_1} \\
&= \frac{E_{t1}}{\sigma_{t1}} \frac{E_0 \pi^3 r_0^4}{4 L_0^3} \\
&= \frac{E_0 E_{t1} \pi^3 L_0}{64 \sigma_{t1} l_0^4}.
\end{aligned}
\tag{17.C.23}
$$

References

1. K. Snelson, Continuous tension, discontinous compression structures, US Patent 3, 169, 611, 1965.
2. B. Fuller. *Tensile-integrity structures,* US Patent, 3, 063, 521, 1962.
3. R. Adhikari, R. E. Skelton, and W. J. Helton, *Mechanics of Tensegrity Beams,* UCSD, Structural System and Control Lab., Report No. 1998-1, 1998.
4. D. M. Campbell, D. Chen, P. A. Gossen, and K. P. Hamilton, *Effects of spatial triangulation on the behavior of "tensegrity" domes,* 652–663, IASS-ASCE International Symposium, April 24–28, Atlanta, GA, 1994.
5. R. Chu, Tensegrity, *Journal of Synergetics,* 2(l), 1988.
6. M. F. Coughlin and D. Stamenovic, A tensegrity structure with buckling compression elements: Application to cell mechanics, *Transactions of ASME, Journal of Applied Mechanics,* 64, 480–486, 1997.
7. D. E. Ingber, Tensegrity: The architectural basis of cellular mechanotransduction, *Annual Review of Physiology,* 59, 575–599, 1997.
8. I. J. Oppenheim and W. J. Williams, Tensegrity prisms as adaptive structures, ASME Annual meeting, Dallas, November 1997.
9. A. Pugh, *An Introduction to Tensegrity,* University of California Press, Berkeley, 1976.
10. R. Motro et al., Form finding numerical methods for tensegrity systems*,* IASS-ASCE International Symposium, Atlanta, GA, April 24–28, 706–713, 1994.
11. H. Faruya, Concept of deployable tensegrity structures in space applications, *International Journal of Space Structures*, 7(2), 143–151, 1992.
12. S. Pellegrino, Analysis of prestressed mechanisms, *International Journal of Solids and Structures*, 26(12), 1329–1350, 1989.
13. S. Pellegrino, A class of tensegrity domes, *International Journal of Space Structures*, 7(2), 1992.
14. A. Hanaor, Double-layer tensegrity grids as deployable structures, *International Journal of Space Structures*, 8, 1992.
15. S. Pellegrino, Foldable bar structures, *International Journal of Solids and Structures*, 34(15), 1825–1847, 1996.
16. D. Williamson and R. E. Skelton, A general class of tensegrity systems: Equilibrium analysis, Engineering Mechanics for the 21st Century, ASCE Conference, La Jolla, 1998.
17. R. E. Skelton, J. P. Pinaud, and D. L. Mingori, Dynamics of the shell class of tensegrity structures, *Journal of the Franklin Institute*, 338(2–3), 255–320, 2001.
18. H. Murakami and Y. Nishimura, Static and dynamic characterization of regular truncated icosahedral and dodecahedral tensegrity modules, *International Journal of Solids and Structures*, 2001 (to appear).
19. B. de Jager and R. E. Skelton, Optimizing stiffness properties of tensegrity structures, *Proceedings of International Mechanical Engineering Congress and Exposition,* 3330, New York, 2001.
20. C. Sultan, Modeling, design, and control of tensegrity structures with applications*,* Ph.D. dissertation, Purdue University, Lafayette, Indiana, 1999.

© 2002 by CRC Press LLC

21. C. Sultan, M. Corless, and R. E. Skelton, The prestressability problem of tensegrity structures: Some analytical solutions, *International Journal of Solids and Structures,* 2001 (to appear).
22. C. Sultan, M. Corless, and R. E. Skelton, Tensegrity flight simulator, *Journal of Guidance, Control, and Dynamics,* 23(6), 1055–1064, 2000.
23. C. Sultan, M. Corless, and R. E. Skelton. Linear dynamics of tensegrity structures, *Engineering Structures*, 2001 (to appear).
24. C. Sultan, M. Corless, and R. E. Skelton, Symmetrical reconfiguration of tensegrity structures, *International Journal of Solids and Structures,* 2001 (to appear).
25. L. Caviglione, Modellazione statica e dinamica di una tensegrity nell'ambito di una strategia di controllo attivo, Laurea thesis, University of Genoa*,* 1999.
26. M. Pedretti, Smart tensegrity structures for the Swiss expo, *Smart Structures and Materials*, 3330, San Diego, 1998.
27. L. Caviglione, A. Del Grosso, and M. Pedretti, Stabilisation and control of large tensegrity structures, private communication.
28. A. G. M. Michell, The limits of economy in frame structures, *Philosophical Magazine,* 8, 589–597, 1904.
29. M. P. Bendsoc and N. Kikuchi, Generating optimal topologies in structural design using a homozenization method, *Computer Methods in Applied Mechanics and Engineering*, 71, 197–224, 1988.
30. F. Jarre, M. Koevara, and J. Zowe, Optimal truss design by interior-point methods, *SIAM Journal on Optimization*, 8(4), 1084–1107, 1998.
31. J. Lu and R. E. Skelton, Optimal hybrid control for structures, *Computer-Aided Civil and Infrastructure Engineering*, 13, 405–414, 1998.
32. A. H. Simmons, C. A. Michal, and L. W. Jelinski, Molecular orientation and two-component nature of the crystalline fraction of dragline silk, *Science*, 271, 84–87, 1996.
33. Y. Termonia, Molecular modeling of spider silk elasticity, *Macromolecules,* 27, 7378–7381, 1994.
34. D. E. Ingber, Architecture of life, *Scientific American*, 48–57, January 1998.
35. D. E. Ingber, Cellular tensegrity: Defining new rules of biological design that govern the cytoskeleton, *Journal of Cell Science,* 104(3), 613–627, 1993.
36. R.E. Skelton, *Dynamic Systems Control — Linear Systems Analysis and Synthesis,* John Wiley & Sons, New York, 1988.
37. S. Pellegrino and C. R. Calladine, Matrix analysis of statistically and kinematically indeterminate frameworks, *International Journal of Solids and Stru*ctures, 22(4), 409–428, 1985.

© 2002 by CRC Press LLC

18

The Dynamics of the Class 1 Shell Tensegrity Structure

18.1 Introduction
18.2 Tensegrity Definitions
A Typical Element • Rules of Closure for the Shell Class
18.3 Dynamics of a Two-Rod Element
18.4 Choice of Independent Variables and Coordinate Transformations
18.5 Tendon Forces
18.6 Conclusion
Appendix 18.A Proof of Theorem 18.1
Appendix 18.B Algebraic Inversion of the Q Matrix
Appendix 18.C General Case for (n, m) = (i, 1)
Appendix 18.D Example Case (n,m) = (3,1)
Appendix 18.E Nodal Forces

Robert E. Skelton
University of California, San Diego

Jean-Paul Pinaud
University of California, San Diego

D. L. Mingori
University of California, Los Angeles

Abstract

A tensegrity structure is a special truss structure in a stable equilibrium with selected members designated for only tension loading, and the members in tension forming a continuous network of cables separated by a set of compressive members. This chapter develops an explicit analytical model of the nonlinear dynamics of a large class of tensegrity structures constructed of rigid rods connected by a continuous network of elastic cables. The kinematics are described by positions and velocities of the ends of the rigid rods; hence, the use of angular velocities of each rod is avoided.

The model yields an analytical expression for accelerations of all rods, making the model efficient for simulation, because the update and inversion of a nonlinear mass matrix are not required. The model is intended for shape control and design of deployable structures. Indeed, the explicit analytical expressions are provided herein for the study of stable equilibria and controllability, but control issues are not treated.

18.1 Introduction

The history of structural design can be divided into four eras classified by design objectives. In the prehistoric era, which produced such structures as Stonehenge, the objective was simply to oppose gravity, to take static loads. The classical era, considered the dynamic response and placed design constraints on the eigenvectors as well as eigenvalues. In the modern era, design constraints could be so demanding that the dynamic response objectives require feedback control. In this era, the

© 2002 by CRC Press LLC

control discipline followed the classical structure design, where the structure and control disciplines were ingredients in a multidisciplinary system design, but no interdisciplinary tools were developed to integrate the design of the structure and the control. Hence, in this modern era, the dynamics of the structure and control were not cooperating to the fullest extent possible. The post-modern era of structural systems is identified by attempts to unify the structure and control design for a common objective.

The ultimate performance capability of many new products and systems cannot be achieved until mathematical tools exist that can extract the full measure of cooperation possible between the dynamics of all components (structural components, controls, sensors, actuators, etc.). This requires new research. Control theory describes how the design of one component (the controller) should be influenced by the (given) dynamics of all other components. However, in systems design, where more than one component remains to be designed, there is inadequate theory to suggest how the dynamics of two or more components should influence each other at the design stage. In the future, controlled structures will not be conceived merely as multidisciplinary design steps, where a plate, beam, or shell is first designed, followed by the addition of control actuation. Rather, controlled structures will be conceived as an interdisciplinary process in which both material architecture and feedback information architecture will be jointly determined. New paradigms for material and structure design might be found to help unify the disciplines. Such a search motivates this work. Preliminary work on the integration of structure and control design appears in Skelton[1,2] and Grigoriadis et al.[3]

Bendsoe and others[4-7] optimize structures by beginning with a solid brick and deleting finite elements until minimal mass or other objective functions are extremized. But, a very important factor in determining performance is the paradigm used for structure design. This chapter describes the dynamics of a structural system composed of axially loaded compression members and tendon members that easily allow the unification of structure and control functions. Sensing and actuating functions can sense or control the tension or the length of tension members. Under the assumption that the axial loads are much smaller than the buckling loads, we treat the rods as rigid bodies. Because all members experience only axial loads, the mathematical model is more accurate than models of systems with members in bending. This unidirectional loading of members is a distinct advantage of our paradigm, since it eliminates many nonlinearities that plague other controlled structural concepts: hysteresis, friction, deadzones, and backlash.

It has been known since the middle of the 20th century that continua cannot explain the strength of materials. While science can now observe at the nanoscale to witness the architecture of materials preferred by nature, we cannot yet design or manufacture manmade materials that duplicate the incredible structural efficiencies of natural systems. Nature's strongest fiber, the spider fiber, arranges simple nontoxic materials (amino acids) into a microstructure that contains a continuous network of members in tension (amorphous strains) and a discontinuous set of members in compression (the β-pleated sheets in Figure 18.1).[8,9]

This class of structure, with a continuous network of tension members and a discontinuous network of compression members, will be called a Class 1 tensegrity structure. The important lessons learned from the tensegrity structure of the spider fiber are that

1. Structural members never reverse their role. The compressive members never take tension and, of course, tension members never take compression.
2. Compressive members do not touch (there are no joints in the structure).
3. Tensile strength is largely determined by the local topology of tension and compressive members.

Another example from nature, with important lessons for our new paradigms is the carbon nanotube often called the Fullerene (or Buckytube), which is a derivative of the Buckyball. Imagine

© 2002 by CRC Press LLC

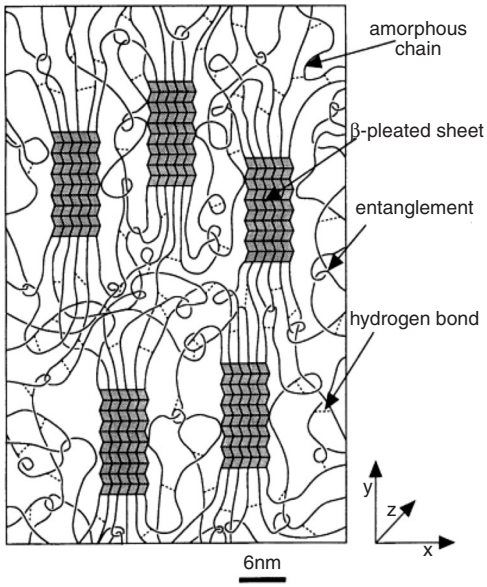


FIGURE 18.1 Nature's strongest fiber: the Spider Fiber. (From Termonia, Y., *Macromolecules*, 27, 7378–7381, 1994. Reprinted with permission from the American Chemical Society.)

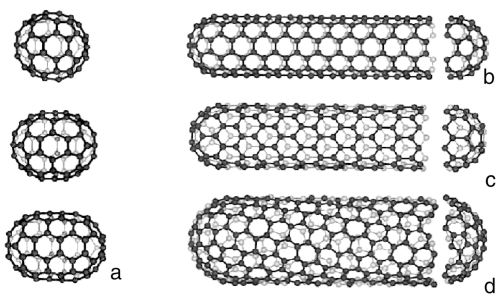


FIGURE 18.2 Buckytubes.

a 1-atom thick sheet of a graphene, which has hexagonal holes due to the arrangements of material at the atomic level (see Figure 18.2). Now imagine that the flat sheet is closed into a tube by choosing an axis about which the sheet is closed to form a tube. A specific set of rules must define this closure which takes the sheet to a tube, and the electrical and mechanical properties of the resulting tube depend on the rules of closure (axis of wrap, relative to the local hexagonal topology).[10] Smalley won the Nobel Prize in 1996 for these insights into the Fullerenes. The spider fiber and the Fullerene provide the motivation to construct manmade materials whose overall mechanical, thermal, and electrical properties can be predetermined by choosing the local topology and the rules of closure which generate the three-dimensional structure from a given local topology. By combining these motivations from Fullerenes with the tensegrity architecture of the spider fiber, this chapter derives the static and dynamic models of a shell class of tensegrity structures. Future papers will exploit the control advantages of such structures. The existing literature on tensegrity deals mainly[11-23] with some elementary work on dynamics in Skelton and Sultan,[24] Skelton and He,[25] and Murakami et al.[26]

© 2002 by CRC Press LLC

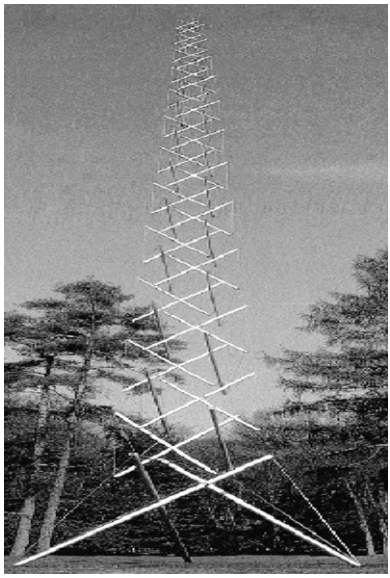


FIGURE 18.3 Needle Tower of Kenneth Snelson, Class 1 tensegrity. Kröller Müller Museum, The Netherlands. (From Connelly, R. and Beck, A., *American Scientist*, 86(2), 143, 1998. With permissions.)

18.2 Tensegrity Definitions

Kenneth Snelson built the first tensegrity structure in 1948 (Figure 18.3) and Buckminster Fuller coined the word "tensegrity." For 50 years tensegrity has existed as an art form with some architectural appeal, but engineering use has been hampered by the lack of models for the dynamics. In fact, engineering use of tensegrity was doubted by the inventor himself. Kenneth Snelson in a letter to R. Motro said, "As I see it, this type of structure, at least in its purest form, is not likely to prove highly efficient or utilitarian." This statement might partially explain why no one bothered to develop math models to convert the art form into engineering practice. We seek to use science to prove the artist wrong, that his invention is indeed more valuable than the artistic scope that he ascribed to it. Mathematical models are essential design tools to make engineered products. This chapter provides a dynamical model of a class of tensegrity structures that is appropriate for space structures.

We derive the nonlinear equations of motion for space structures that can be deployed or held to a precise shape by feedback control, although control is beyond the scope of this chapter. For engineering purposes, more precise definitions of tensegrity are needed.

One can imagine a truss as a structure whose compressive members are all connected with ball joints so that no torques can be transmitted. Of course, tension members connected to compressive members do not transmit torques, so that our truss is composed of members experiencing no moments. The following definitions are useful.

Definition 18.1 A given configuration of a structure is in a *stable equilibrium* if, in the absence of external forces, an arbitrarily small initial deformation returns to the given configuration.

Definition 18.2 A tensegrity structure is a stable system of axially loaded members.

Definition 18.3 A stable structure is said to be a "Class 1" tensegrity structure if the members in tension form a continuous network, and the members in compression form a discontinuous set of members.

© 2002 by CRC Press LLC

FIGURE 18.4 Class 1 and Class 2 tensegrity structures.

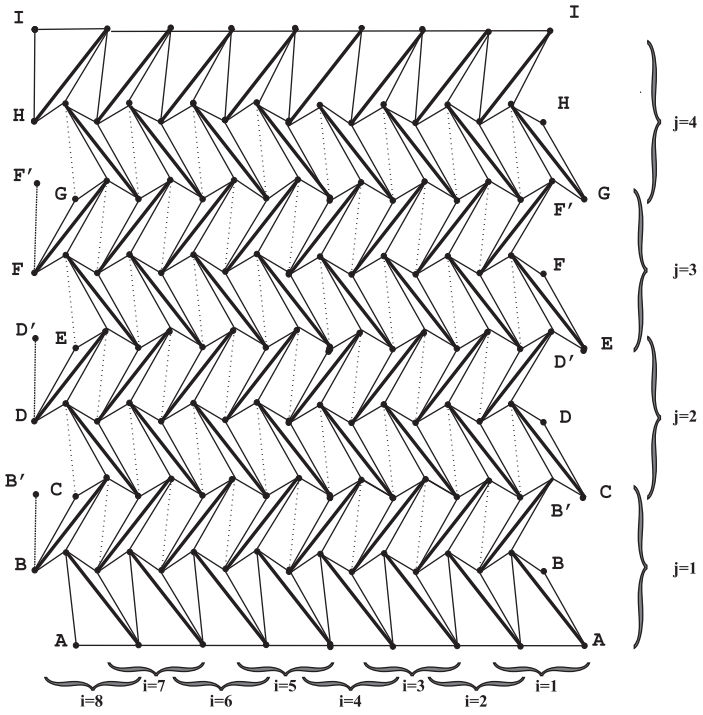


FIGURE 18.5 Topology of an (8,4) Class 1 tensegrity shell.

Definition 18.4 A stable structure is said to be a "Class 2" tensegrity structure if the members in tension form a continuous set of members, and there are at most two members in compression connected to each node.

Figure 18.4 illustrates Class 1 and Class 2 tensegrity structures.

Consider the topology of structural members given in Figure 18.5, where thick lines indicate rigid rods which take compressive loads and the thin lines represent tendons. This is a Class 1 tensegrity structure.

Definition 18.5 Let the topology of Figure 18.5 describe a three-dimensional structure by connecting points A to A, B to B, C to C, I to I. This constitutes a "Class 1 tensegrity shell" if there exists a set of tensions in all tendons $t_{\alpha\beta\gamma}$, $\alpha = 1(\rightarrow 10, \beta = 1 \rightarrow n, \gamma = 1 \rightarrow m)$ such that the structure is in a stable equilibrium.

© 2002 by CRC Press LLC

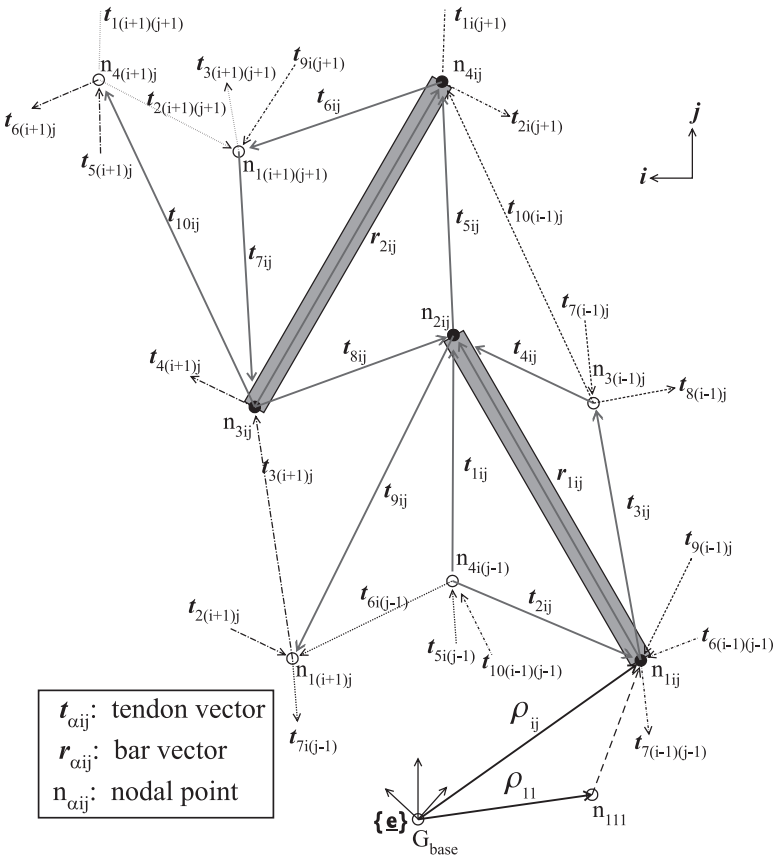


FIGURE 18.6 A typical *ij* element.

18.2.1 A Typical Element

The axial members in Figure 18.5 illustrate only the pattern of member connections and not the actual loaded configuration. The purpose of this section is two-fold: (i) to define a typical "element" which can be repeated to generate all elements, and (ii) to define rules of closure that will generate a "shell" type of structure.

Consider the members that make the typical *ij* element where $i = 1, 2, \ldots,$ n indexes the element to the left, and $j = 1, 2, \ldots,$ m indexes the element up the page in Figure 18.5. We describe the axial elements by vectors. That is, the vectors describing the *ij* element, are $\mathbf{t}_{1ij}$, $\mathbf{t}_{2ij}$, … $\mathbf{t}_{10ij}$ and $\mathbf{r}_{1ij}$, $\mathbf{r}_{2ij}$, where, within the *ij* element, $\mathbf{t}_{\alpha ij}$ is a vector whose tail is fixed at the specified end of tendon number α, and the head of the vector is fixed at the other end of tendon number α as shown in Figure 18.6 where $\alpha = 1, 2, \ldots, 10$. The *ij* element has two compressive members we call "rods," shaded in Figure 18.6. Within the *ij* element the vector $\mathbf{r}_{1ij}$ lies along the rod r_{1ij} and the vector $\mathbf{r}_{2ij}$ lies along the rod r_{2ij}. The first goal of this chapter is to derive the equations of motion for the dynamics of the two rods in the *ij* element. The second goal is to write the dynamics for the entire system composed of *nm* elements. Figures 18.5 and 18.7 illustrate these closure rules for the case $(n, m) = (8,4)$ and $(n, m) = (3,1)$.

Lemma 18.1 Consider the structure of Figure 18.5 with elements defined by Figure 18.6. A Class 2 tensegrity shell is formed by adding constraints such that for all $i = 1, 2, \ldots,$ n, and for $m > j > 1,$

© 2002 by CRC Press LLC

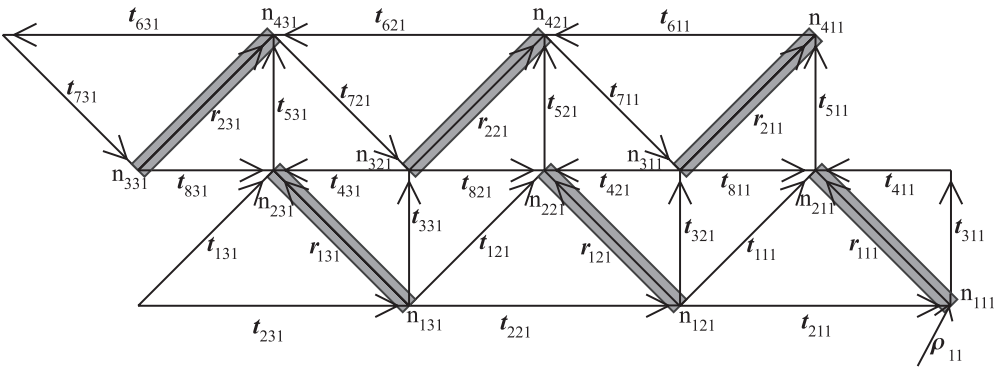


FIGURE 18.7 Class 1 shell: (n,m) = (3,1).

$$\begin{aligned} -\mathbf{t}_{1ij} + \mathbf{t}_{4ij} &= \mathbf{0}, \\ \mathbf{t}_{2ij} + \mathbf{t}_{3ij} &= \mathbf{0}, \\ \mathbf{t}_{5ij} + \mathbf{t}_{6ij} &= \mathbf{0}, \\ \mathbf{t}_{7ij} + \mathbf{t}_{8ij} &= \mathbf{0}. \end{aligned} \tag{18.1}$$

This closes nodes n_{2ij} and $n_{1(i+1)(j+1)}$ to a single node, and closes nodes $n_{3(i-1)j}$ and $n_{4i(j-1)}$ to a single node (with ball joints). The nodes are closed outside the rod, so that all tension elements are on the exterior of the tensegrity structure and the rods are in the interior.

The point here is that a Class 2 shell can be obtained as a special case of the Class 1 shell, by imposing constraints (18.1). To create a tensegrity structure not all tendons in Figure 18.5 are necessary. The following definition eliminates tendons $\mathbf{t}_{9ij}$ and $\mathbf{t}_{10ij}$, (i = 1 → n, j = 1 → m).

Definition 18.6 Consider the shell of Figures 18.4. and 18.5, which may be Class 1 or Class 2 depending on whether constraints (18.1) are applied. In the absence of dotted tendons (labeled t_9 and t_{10}), this is called a primal tensegrity shell. When all tendons t_9, t_{10} are present in Figure 18.5, it is called simply a Class 1 or Class 2 tensegrity shell.

The remainder of this chapter focuses on the general Class 1 shell of Figures 18.5 and 18.6.

18.2.2 Rules of Closure for the Shell Class

Each tendon exerts a positive force away from a node and $\mathbf{f}_{\alpha\beta\gamma}$ is the force exerted by tendon $\mathbf{t}_{\alpha\beta\gamma}$ and $\hat{\mathbf{f}}_{\alpha ij}$ denotes the force vector acting on the node $n_{\alpha ij}$. All $\mathbf{f}_{\alpha ij}$ forces are postive in the direction of the arrows in Figure 18.6, where $\mathbf{w}_{\alpha ij}$ is the external applied force at node $n_{\alpha ij}$, α = 1, 2, 3, 4. At the base, the rules of closure, from Figures 18.5 and 18.6, are

$$\mathbf{t}_{9i1} = -\mathbf{t}_{1i1},\ i = 1, 2, \ldots, n \tag{18.2}$$

$$\mathbf{t}_{6i0} = \mathbf{0} \tag{18.3}$$

$$\mathbf{t}_{600} = -\mathbf{t}_{2n1} \tag{18.4}$$

$$\mathbf{t}_{901} = \mathbf{t}_{9n1} = -\mathbf{t}_{1n1} \tag{18.5}$$

$$\mathbf{0} = \mathbf{t}_{10(i-1)0} = \mathbf{t}_{5i0} = \mathbf{t}_{7i0} = \mathbf{t}_{7(i-1)0},\ i = 1, 2, \ldots, n. \tag{18.6}$$

© 2002 by CRC Press LLC

At the top, the closure rules are

$$\mathbf{t}_{10im} = -\mathbf{t}_{7im} \tag{18.7}$$

$$\mathbf{t}_{100m} = -\mathbf{t}_{70m} = -\mathbf{t}_{7nm} \tag{18.8}$$

$$\mathbf{t}_{2i(m+1)} = \mathbf{0} \tag{18.9}$$

$$\begin{aligned} \mathbf{0} &= \mathbf{t}_{1i(m+1)} = \mathbf{t}_{9i(m+1)} = \mathbf{t}_{3(i+1)(m+1)} \\ &= \mathbf{t}_{1(i+1)(m+1)} = \mathbf{t}_{2(i+1)(m+1)}. \end{aligned} \tag{18.10}$$

At the closure of the circumference (where $i = 1$):

$$\mathbf{t}_{90j} = \mathbf{t}_{9nj}, \qquad \mathbf{t}_{60(j-1)} = \mathbf{t}_{6n(j-1)}, \qquad \mathbf{t}_{70(j-1)} = \mathbf{t}_{7n(j-1)} \tag{18.11}$$

$$\mathbf{t}_{80j} = \mathbf{t}_{8nj}, \qquad \mathbf{t}_{70j} = \mathbf{t}_{7nj}, \qquad \mathbf{t}_{100(j-1)} = \mathbf{t}_{10n(j-1)}. \tag{18.12}$$

From Figures 18.5 and 18.6, when $j = 1$, then

$$\mathbf{0} = \mathbf{f}_{7i(j-1)} = \mathbf{f}_{7(i-1)(j-1)} = \mathbf{f}_{5i(j-1)} = \mathbf{f}_{10(i-1)(j-1)}, \tag{18.13}$$

and for $j = m$ where,

$$\mathbf{0} = \mathbf{f}_{1i(m+1)} = \mathbf{f}_{9i(m+1)} = \mathbf{f}_{3(i+1)(m+1)} = \mathbf{f}_{1(i+1)(m+1)}. \tag{18.14}$$

Nodes n_{11j}, n_{3nj}, n_{41j} for $j = 1, 2, \ldots, m$ are involved in the longitudinal "zipper" that closes the structure in circumference. The forces at these nodes are written explicitly to illustrate the closure rules.

In 18.4, rod dynamics will be expressed in terms of sums and differences of the nodal forces, so the forces acting on each node are presented in the following form, convenient for later use. The definitions of the matrices $\mathbf{B}_i$ are found in Appendix 18.E.

The forces acting on the nodes can be written in vector form:

$$\mathbf{f} = \mathbf{B}^d \mathbf{f}^d + \mathbf{B}^o \mathbf{f}^o + \mathbf{W}^o \mathbf{w} \tag{18.15}$$

where

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_m \end{bmatrix}, \mathbf{f}^d = \begin{bmatrix} \bar{\mathbf{f}}_1^d \\ \mathbf{f}_2^d \\ \vdots \\ \mathbf{f}_m^d \end{bmatrix}, \mathbf{f}^o = \begin{bmatrix} \mathbf{f}_1^o \\ \vdots \\ \mathbf{f}_m^o \end{bmatrix}, \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix},$$

$$\mathbf{W}^o = BlockDiag\ [\cdots, \mathbf{W}_1, \mathbf{W}_1, \cdots],$$

$$\mathbf{B}^d = \begin{bmatrix} \mathbf{B}_3 & \mathbf{B}_4 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{B}_5 & \mathbf{B}_6 & \ddots & \ddots & \vdots \\ \mathbf{0} & \overline{\mathbf{B}}_5 & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{B}_6 & \mathbf{B}_4 \\ \mathbf{0} & \cdots & \mathbf{0} & \overline{\mathbf{B}}_5 & \mathbf{B}_8 \end{bmatrix}, \mathbf{B}^o = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_7 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \mathbf{B}_2 \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{B}_7 \end{bmatrix}$$

© 2002 by CRC Press LLC

and

$$\mathbf{f}_{ij}^{o} = \begin{bmatrix} \mathbf{f}_5 \\ \mathbf{f}_1 \end{bmatrix}_{ij}, \quad \mathbf{f}_{ij}^{d} = \begin{bmatrix} \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \\ \mathbf{f}_6 \\ \mathbf{f}_7 \\ \mathbf{f}_8 \\ \mathbf{f}_9 \\ \mathbf{f}_{10} \end{bmatrix}_{ij}, \quad \mathbf{w}_{ij} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \\ \mathbf{w}_4 \end{bmatrix}_{ij} \tag{18.16}$$

Now that we have an expression for the forces, let us write the dynamics.

18.3 Dynamics of a Two-Rod Element

Any discussion of rigid body dynamics should properly begin with some decision on how the motion of each body is to be described. A common way to describe rigid body orientation is to use three successive angular rotations to define the orientation of three mutually orthogonal axes fixed in the body. The measure numbers of the angular velocity of the body may then be expressed in terms of these angles and their time derivatives.

This approach must be reconsidered when the body of interest is idealized as a rod. The reason is that the concept of "body fixed axes" becomes ambiguous. Two different sets of axes with a common axis along the rod can be considered equally "body fixed" in the sense that all mass particles of the rod have zero velocity in both sets. This remains true even if relative rotation is allowed along the common axis. The angular velocity of the rod is also ill defined because the component of angular velocity along the rod axis is arbitrary. For these reasons, we are motivated to seek a kinematical description which avoids introducing "body-fixed" reference frames and angular velocity. This objective may be accomplished by describing the configuration of the system in terms of vectors located only the end points of the rods. In this case, no angles are used.

We will use the following notational conventions. Lower case, bold-faced symbols with an underline indicate vector quantities with magnitude and direction in three-dimensional space. These are the usual vector quantities we are familiar with from elementary dynamics. The same bold-faced symbols without an underline indicate a matrix whose elements are scalars. Sometimes we also need to introduce matrices whose elements are vectors. These quantities are indicated with an upper case symbol that is both bold faced and underlined.

As an example of this notation, a position vector can be expressed as

$$\underline{\mathbf{p}}_i = [\underline{\mathbf{e}}_1 \quad \underline{\mathbf{e}}_2 \quad \underline{\mathbf{e}}_3] \begin{bmatrix} p_{i1} \\ p_{i2} \\ p_{i3} \end{bmatrix} = \underline{\mathbf{E}}\mathbf{p}_i.$$

In this expression, $\mathbf{p}_i$ is a column matrix whose elements are the measure numbers of $\underline{\mathbf{p}}_i$ for the mutually orthogonal inertial unit vectors $\underline{\mathbf{e}}_1$, $\underline{\mathbf{e}}_2$, and $\underline{\mathbf{e}}_3$. Similarly, we may represent a force vector $\hat{\mathbf{f}}_i$ as

$$\hat{\underline{\mathbf{f}}}_i = \underline{\mathbf{E}}\hat{\mathbf{f}}_i.$$

Matrix notation will be used in most of the development to follow.

© 2002 by CRC Press LLC

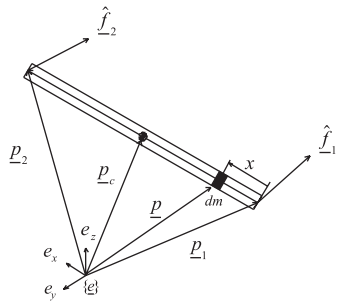


FIGURE 18.8 A single rigid rod.

We now consider a single rod as shown in Figure 18.8 with nodal forces $\hat{\underline{\mathbf{f}}}_1$ and $\hat{\underline{\mathbf{f}}}_2$ applied to the ends of the rod.

The following theorem will be fundamental to our development.

Theorem 18.1 Given a rigid rod of constant mass m and constant length L, the governing equations may be described as:

$$\ddot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{H}\tilde{\mathbf{f}} \tag{18.17}$$

where

$$\mathbf{q} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{p}_1 + \mathbf{p}_2 \\ \mathbf{p}_2 - \mathbf{p}_1 \end{bmatrix}$$

$$\tilde{\mathbf{f}} = \begin{bmatrix} \hat{\mathbf{f}}_1 + \hat{\mathbf{f}}_2 \\ \hat{\mathbf{f}}_1 - \hat{\mathbf{f}}_2 \end{bmatrix}, \quad \mathbf{H} = \frac{2}{m}\begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \frac{3}{L^2}\tilde{\mathbf{q}}_2^2 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & L^{-2}\dot{\mathbf{q}}_2^T\dot{\mathbf{q}}_2\mathbf{I}_3 \end{bmatrix}.$$

The notation $\tilde{\mathbf{r}}$ denotes the skew symmetric matrix formed from the elements of $\underline{\mathbf{r}}$:

$$\tilde{\mathbf{r}} = \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

and the square of this matrix is

$$\tilde{\mathbf{r}}^2 = \begin{bmatrix} -r_2^2 - r_3^2 & r_1 r_2 & r_1 r_3 \\ r_2 r_1 & -r_1^2 - r_3^2 & r_2 r_3 \\ r_3 r_1 & r_3 r_2 & -r_1^2 - r_2^2 \end{bmatrix}.$$

The matrix elements $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, etc. are to be interpreted as the measure numbers of the corresponding vectors for an orthogonal set of inertially fixed unit vectors $\underline{\mathbf{e}}_1$, $\underline{\mathbf{e}}_2$, and $\underline{\mathbf{e}}_3$. Thus, using the convention introduced earlier,

$$\underline{\mathbf{r}} = \underline{\mathbf{E}}\mathbf{r}, \quad \underline{\mathbf{q}} = \underline{\mathbf{E}}\mathbf{q}, \quad \text{etc.}$$

© 2002 by CRC Press LLC

The proof of Theorem 18.1 is given in Appendix 18.A. This theorem provides the basis of our dynamic model for the shell class of tensegrity structures.

Now consider the dynamics of the two-rod element of the Class 1 tensegrity shell in Figure 18.5. Here, we assume the lengths of the rods are constant. From Theorem 18.1 and Appendix 18.A, the motion equations for the *ij* unit can be described as

$$\begin{cases} \frac{m_{1ij}}{2}\ddot{\mathbf{q}}_{1ij} = \hat{\mathbf{f}}_{1ij} + \hat{\mathbf{f}}_{2ij} \\ \frac{m_{1ij}}{6}(\mathbf{q}_{2ij} \times \ddot{\mathbf{q}}_{2ij}) = \mathbf{q}_{2ij} \times (\hat{\mathbf{f}}_{2ij} - \hat{\mathbf{f}}_{1ij}) \\ \dot{\mathbf{q}}_{2ij} \cdot \dot{\mathbf{q}}_{2ij} + \mathbf{q}_{2ij} \cdot \ddot{\mathbf{q}}_{2ij} = \mathbf{0} \\ \mathbf{q}_{2ij} \cdot \mathbf{q}_{2ij} = L_{1ij}^2 \end{cases}, \tag{18.18}$$

$$\begin{cases} \frac{m_{2ij}}{2}\ddot{\mathbf{q}}_{3ij} = \hat{\mathbf{f}}_{3ij} + \hat{\mathbf{f}}_{4ij} \\ \frac{m_{2ij}}{6}(\mathbf{q}_{4ij} \times \ddot{\mathbf{q}}_{4ij}) = \mathbf{q}_{4ij} \times (\hat{\mathbf{f}}_{4ij} - \hat{\mathbf{f}}_{3ij}) \\ \dot{\mathbf{q}}_{4ij} \cdot \dot{\mathbf{q}}_{4ij} + \mathbf{q}_{4ij} \cdot \ddot{\mathbf{q}}_{4ij} = \mathbf{0} \\ \mathbf{q}_{4ij} \cdot \mathbf{q}_{4ij} = L_{2ij}^2 \end{cases}, \tag{18.19}$$

where the mass of the rod αij is $m_{\alpha ij}$ and $\|\mathbf{r}_{\alpha ij}\| = L_{\alpha ij}$. As before, we refer everything to a common inertial reference frame ($\underline{\mathbf{E}}$). Hence,

$$\mathbf{q}_{1ij} \triangleq \begin{bmatrix} q_{11ij} \\ q_{12ij} \\ q_{13ij} \end{bmatrix}, \ \mathbf{q}_{2ij} \triangleq \begin{bmatrix} q_{21ij} \\ q_{22ij} \\ q_{23ij} \end{bmatrix}, \ \mathbf{q}_{3ij} \triangleq \begin{bmatrix} q_{31ij} \\ q_{32ij} \\ q_{33ij} \end{bmatrix}, \ \mathbf{q}_{4ij} \triangleq \begin{bmatrix} q_{41ij} \\ q_{42ij} \\ q_{43ij} \end{bmatrix},$$

$$\mathbf{q}_{1ij} \triangleq \left[\mathbf{q}_{1ij,}^T \quad \mathbf{q}_{2ij,}^T \quad \mathbf{q}_{3ij}^T, \quad \mathbf{q}_{4ij}^T\right]^T,$$

and the force vectors appear in the form

$$\mathbf{H}_{1ij} = \frac{2}{m_{1ij}}\begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \frac{3}{L_{1ij}^2}\tilde{\mathbf{q}}_{2ij}^2 \end{bmatrix}, \ \mathbf{H}_{2ij} = \frac{2}{m_{2ij}}\begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \frac{3}{L_{2ij}^2}\tilde{\mathbf{q}}_{4ij}^2 \end{bmatrix},$$

$$\mathbf{H}_{ij} = \begin{bmatrix} \mathbf{H}_{1ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{2ij} \end{bmatrix}, \ \mathbf{f}_{ij} \triangleq \begin{bmatrix} \hat{\mathbf{f}}_{1ij} + \hat{\mathbf{f}}_{2ij} \\ \hat{\mathbf{f}}_{1ij} - \hat{\mathbf{f}}_{2ij} \\ \hat{\mathbf{f}}_{3ij} + \hat{\mathbf{f}}_{4ij} \\ \hat{\mathbf{f}}_{3ij} - \hat{\mathbf{f}}_{4ij} \end{bmatrix}.$$

© 2002 by CRC Press LLC

Using Theorem 18.1, the dynamics for the *ij* unit can be expressed as follows:

$$\ddot{\mathbf{q}}_{ij} + \boldsymbol{\Omega}_{ij}\mathbf{q}_{ij} = \mathbf{H}_{ij}\mathbf{f}_{ij},$$

where

$$\boldsymbol{\Omega}_{1ij} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & L_{1ij}^{-2}\dot{\mathbf{q}}_{2ij}^T\dot{\mathbf{q}}_{2ij}\mathbf{I}_3 \end{bmatrix}, \quad \boldsymbol{\Omega}_{2ij} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & L_{2ij}^{-2}\dot{\mathbf{q}}_{4ij}^T\dot{\mathbf{q}}_{4ij}\mathbf{I}_3 \end{bmatrix},$$

$$\boldsymbol{\Omega}_{ij} = \begin{bmatrix} \boldsymbol{\Omega}_{1ij} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{2ij} \end{bmatrix},$$

$$\mathbf{q} = \left[\mathbf{q}_{11}^T, \ldots, \mathbf{q}_{n1}^T, \mathbf{q}_{12}^T, \ldots, \mathbf{q}_{n2}^T, \ldots, \mathbf{q}_{1m}^T, \ldots, \mathbf{q}_{nm}^T\right]^T.$$

The shell system dynamics are given by

$$\ddot{\mathbf{q}} + \mathbf{K}_r\mathbf{q} = \mathbf{H}\mathbf{f}, \tag{18.20}$$

where **f** is defined in (18.15) and

$$\mathbf{q} = \left[\mathbf{q}_{11}^T, \ldots, \mathbf{q}_{n1}^T, \mathbf{q}_{12}^T, \ldots, \mathbf{q}_{n2}^T, \ldots, \mathbf{q}_{1m}^T, \ldots, \mathbf{q}_{nm}^T\right]^T,$$

$$\mathbf{K}_r = BlockDiag\left[\boldsymbol{\Omega}_{11}, \ldots, \boldsymbol{\Omega}_{n1}, \boldsymbol{\Omega}_{12}, \ldots, \boldsymbol{\Omega}_{n2}, \ldots, \boldsymbol{\Omega}_{1m}, \ldots, \boldsymbol{\Omega}_{nm}\right],$$

$$\mathbf{H} = BlockDiag\left[\mathbf{H}_{11}, \ldots, \mathbf{H}_{n1}, \mathbf{H}_{12}, \ldots, \mathbf{H}_{n2}, \ldots, \mathbf{H}_{1m}, \ldots, \mathbf{H}_{nm}\right].$$

18.4 Choice of Independent Variables and Coordinate Transformations

Tendon vectors $t_{\alpha\beta\gamma}$ are needed to express the forces. Hence, the dynamical model will be completed by expressing the tendon forces, **f**, in terms of variables **q**. From Figures 18.6 and 18.9, it follows that vectors $\hat{\boldsymbol{\rho}}_{ij}$ and $\boldsymbol{\rho}_{ij}$ can be described by

$$\boldsymbol{\rho}_{ij} = \boldsymbol{\rho}_{11} + \sum_{k=1}^{i}\mathbf{r}_{1k1} - \sum_{k=1}^{i-1}\mathbf{t}_{1k1} + \sum_{k=2}^{j}\mathbf{t}_{1ik} + \sum_{k=1}^{j-1}\mathbf{t}_{5ik} - \mathbf{r}_{1ij} \tag{18.21}$$

$$\hat{\boldsymbol{\rho}}_{ij} = \boldsymbol{\rho}_{ij} + \mathbf{r}_{1ij} + \mathbf{t}_{5ij} - \mathbf{r}_{2ij}. \tag{18.22}$$

To describe the geometry, we choose the independent vectors $\{\mathbf{r}_{1ij}, \mathbf{r}_{2ij}, \mathbf{t}_{5ij}$, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m\}$ and $\{\boldsymbol{\rho}_{11}, \mathbf{t}_{1ij}$, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, and $i < n$ when $j = 1\}$.

This section discusses the relationship between the **q** variables and the string and rod vectors $\mathbf{t}_{\alpha\beta\gamma}$ and $\mathbf{r}_{\beta ij}$. From Figures 18.5 and 18.6, the position vectors from the origin of the reference frame, E, to the nodal points, $\mathbf{p}_{1ij}$, $\mathbf{p}_{2ij}$, $\mathbf{p}_{3ij}$, and $\mathbf{p}_{4ij}$, can be described as follows:

© 2002 by CRC Press LLC

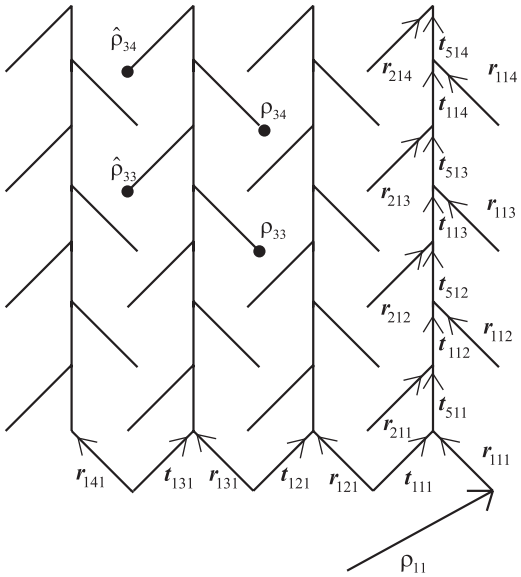


FIGURE 18.9 Choice of independent variables.

$$\begin{cases} \mathbf{p}_{1ij} = \boldsymbol{\rho}_{ij} \\ \mathbf{p}_{2ij} = \boldsymbol{\rho}_{ij} + \mathbf{r}_{1ij} \\ \mathbf{p}_{3ij} = \hat{\boldsymbol{\rho}}_{ij} \\ \mathbf{p}_{4ij} = \hat{\boldsymbol{\rho}}_{ij} + \mathbf{r}_{2ij} \end{cases} \tag{18.23}$$

We define

$$\begin{cases} \mathbf{q}_{1ij} \triangleq \mathbf{p}_{2ij} + \mathbf{p}_{1ij} = 2\boldsymbol{\rho}_{ij} + \mathbf{r}_{1ij} \\ \mathbf{q}_{2ij} \triangleq \mathbf{p}_{2ij} - \mathbf{p}_{1ij} = \mathbf{r}_{1ij} \\ \mathbf{q}_{3ij} \triangleq \mathbf{p}_{4ij} + \mathbf{p}_{3ij} = 2\hat{\boldsymbol{\rho}}_{1ij} + \mathbf{r}_{2ij} \\ \mathbf{q}_{4ij} \triangleq \mathbf{p}_{4ij} - \mathbf{p}_{3ij} = \mathbf{r}_{2ij} \end{cases} \tag{18.24}$$

Then,

$$\begin{aligned} \mathbf{q}_{ij} \triangleq \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \end{bmatrix}_{ij} &= \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ \mathbf{p}_4 \end{bmatrix}_{ij} \\ &= \begin{bmatrix} 2\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2\mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix}_{ij} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{ij} \end{aligned} \tag{18.25}$$

© 2002 by CRC Press LLC

In shape control, we will later be interested in the **p** vector to describe all nodal points of the structure. This relation is

$$\mathbf{p} = \mathbf{P}\mathbf{q} \quad \mathbf{P} = \mathit{BlockDiag}\ [\ldots,\mathbf{P}_1,\ \ldots,\mathbf{P}_1,\ldots] \tag{18.26}$$

$$\mathbf{P}_1 \triangleq \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix}$$

The equations of motion will be written in the **q** coordinates. Substitution of (18.21) and (18.22) into (18.24) yields the relationship between **q** and the independent variables $\mathbf{t}_5$, $\mathbf{t}_1$, $\mathbf{r}_1$, $\mathbf{r}_2$ as follows:

$$\begin{aligned}
\mathbf{q}_{1ij} &= 2\left[\boldsymbol{\rho}_{11} + \sum_{k=1}^{i}\mathbf{r}_{1k1} - \sum_{k=1}^{i-1}\mathbf{t}_{1k1} + \sum_{k=2}^{j}\mathbf{t}_{1ik} + \sum_{k=1}^{j-1}\mathbf{t}_{5ik}\right] - \mathbf{r}_{1ij} \\
\mathbf{q}_{2ij} &= \mathbf{r}_{1ij} \\
\mathbf{q}_{3ij} &= 2\left[\boldsymbol{\rho}_{11} + \sum_{k=1}^{i}\mathbf{r}_{1k1} - \sum_{k=1}^{i-1}\mathbf{t}_{1k1} + \sum_{k=2}^{j}\mathbf{t}_{1ik} + \sum_{k=1}^{j}\mathbf{t}_{5ik}\right] - \mathbf{r}_{2ij} \\
\mathbf{q}_{4ij} &= \mathbf{r}_{2ij}
\end{aligned} \tag{18.27}$$

To put (18.27) in a matrix form, define the matrices:

$$\mathbf{l}_{ij} = \begin{bmatrix} \mathbf{r}_{1ij} \\ \mathbf{r}_{2ij} \\ \mathbf{t}_{5ij} \\ \mathbf{t}_{1ij} \end{bmatrix} \text{for } j = 2,\ 3,\ldots,\ m,$$

$$\mathbf{l}_{11} = \begin{bmatrix} \boldsymbol{\rho}_{11} \\ \mathbf{r}_{111} \\ \mathbf{r}_{211} \\ \mathbf{t}_{511} \end{bmatrix},\ \mathbf{l}_{i1} = \begin{bmatrix} \mathbf{t}_{1(i\text{-}1)1} \\ \mathbf{r}_{1i1} \\ \mathbf{r}_{2i1} \\ \mathbf{t}_{5i1} \end{bmatrix} \text{for } i = 2,\ldots,\ n,$$

and

$$\mathbf{l} = \left[\mathbf{l}_{11}^T,\ \mathbf{l}_{21}^T,\ \ldots,\ \mathbf{l}_{n1}^T,\ \mathbf{l}_{12}^T,\ \ldots,\ \mathbf{l}_{n2}^T,\ \ldots,\ \mathbf{l}_{1m}^T,\ \ldots,\ \mathbf{l}_{nm}^T\right]^T,$$

$$\mathbf{A} = \begin{bmatrix} 2\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ 2\mathbf{I}_3 & 2\mathbf{I}_3 & -\mathbf{I}_3 & 2\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} -2\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -2\mathbf{I}_3 & 2\mathbf{I}_3 & -\mathbf{I}_3 & 2\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \end{bmatrix}$$

© 2002 by CRC Press LLC

$$\mathbf{C}=\begin{bmatrix}-\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & 2\mathbf{I}_3\\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I}_3 & 2\mathbf{I}_3 & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0}\end{bmatrix},\quad \mathbf{D}=\begin{bmatrix}2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ 2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}$$

$$\mathbf{E}=\begin{bmatrix}-2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ -2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix},\quad \mathbf{F}=\begin{bmatrix}2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ 2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}$$

$$\mathbf{J}=\begin{bmatrix}\mathbf{0} & \mathbf{0} & 2\mathbf{I}_3 & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & 2\mathbf{I}_3 & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix},\quad \mathbf{G}=\begin{bmatrix}-2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ 2\mathbf{I}_3 & 2\mathbf{I}_3 & \mathbf{0} & 2\mathbf{I}_3\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}$$

Then (18.27) can be written simply

$$\mathbf{q} = \mathbf{Q}\mathbf{I}, \tag{18.28}$$

where the $12nm \times 12nm$ matrix $\mathbf{Q}$ is composed of the 12×12 matrices A–H as follows:

$$\mathbf{Q}=\begin{bmatrix}\mathbf{Q}_{11} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \ddots & & \vdots & \\ \mathbf{Q}_{21} & \mathbf{Q}_{32} & \mathbf{Q}_{22} & \ddots & \vdots & \\ \mathbf{Q}_{21} & \mathbf{Q}_{32} & \mathbf{Q}_{32} & \mathbf{Q}_{22} & \mathbf{0} & \\ \mathbf{Q}_{21} & \mathbf{Q}_{32} & \mathbf{Q}_{32} & \mathbf{Q}_{32} & \mathbf{Q}_{22} & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots\end{bmatrix}, \tag{18.29}$$

$$\mathbf{Q}_{11}=\left.\begin{bmatrix}\mathbf{A} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0}\\ \mathbf{D} & \mathbf{B} & \ddots & & & \vdots\\ \mathbf{D} & \mathbf{E} & \mathbf{B} & \ddots & & \vdots\\ \mathbf{D} & \mathbf{E} & \mathbf{E} & \mathbf{B} & \ddots & \vdots\\ \vdots & \vdots & \vdots & \ddots & \ddots & 0\\ \mathbf{D} & \mathbf{E} & \mathbf{E} & \cdots & \mathbf{E} & \mathbf{B}\end{bmatrix}\right\} n \times n \text{ blocks of } 12 \times 12 \text{ matrices},$$

$$\mathbf{Q}_{21}=\left.\begin{bmatrix}\mathbf{F} & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0}\\ \mathbf{D} & \mathbf{G} & \ddots & & & \vdots\\ \mathbf{D} & \mathbf{E} & \mathbf{G} & \ddots & & \vdots\\ \mathbf{D} & \mathbf{E} & \mathbf{E} & \ddots & \ddots & \vdots\\ \vdots & \vdots & \vdots & \ddots & \ddots & \mathbf{0}\\ \mathbf{D} & \mathbf{E} & \mathbf{E} & \mathbf{E} & \mathbf{E} & \mathbf{G}\end{bmatrix}\right\} 12n \times 12n \text{ matrix},$$

$\mathbf{Q}_{22} = \mathit{BlockDiag}\ [\ldots,\mathbf{C}, \ldots,\mathbf{C}],$

$\mathbf{Q}_{32} = \mathit{BlockDiag}\ [\ldots,\mathbf{J}, \ldots,\mathbf{J}],$

© 2002 by CRC Press LLC

where each $\mathbf{Q}_{ij}$ is $12n \times 12n$ and there are m row blocks and m column blocks in $\mathbf{Q}$. Appendix 18.B provides an explicit expression for the inverse matrix $\mathbf{Q}$, which will be needed later to express the tendon forces in terms of $\mathbf{q}$.

Equation (18.28) provides the relationship between the selected generalized coordinates and an independent set of the tendon and rod vectors forming $\mathbf{l}$. All remaining tendon vectors may be written as a linear combination of $\mathbf{l}$. This relation will now be established. The following equations are written by inspection of Figures 18.5, 18.6, and 18.7 where

$$\mathbf{t}_{1n1} = \boldsymbol{\rho}_{n1} + \mathbf{r}_{1n1} - \boldsymbol{\rho}_{11} \tag{18.30}$$

and for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ we have

$$\begin{aligned}
\mathbf{t}_{2ij} &= \boldsymbol{\rho}_{ij} - (\hat{\boldsymbol{\rho}}_{i(j-1)} + \mathbf{r}_{2i(j-1)}), \quad (j > 1)\\
\mathbf{t}_{3ij} &= \hat{\boldsymbol{\rho}}_{(i-1)j} - \boldsymbol{\rho}_{ij}\\
\mathbf{t}_{4ij} &= -\mathbf{t}_{3ij} + \mathbf{r}_{1ij} = \boldsymbol{\rho}_{ij} + \mathbf{r}_{1ij} - \hat{\boldsymbol{\rho}}_{(i-1)j}\\
\mathbf{t}_{6ij} &= \boldsymbol{\rho}_{(i+1)(j+1)} - (\hat{\boldsymbol{\rho}}_{ij} + \mathbf{r}_{2ij}) \quad , \quad (j < m)\\
\mathbf{t}_{7ij} &= \hat{\boldsymbol{\rho}}_{ij} - \boldsymbol{\rho}_{(i+1)(j+1)} \quad , \quad (j < m)\\
\mathbf{t}_{8ij} &= \boldsymbol{\rho}_{ij} + \mathbf{r}_{1ij} - \hat{\boldsymbol{\rho}}_{ij} = -\mathbf{r}_{1ij} - \mathbf{t}_{5ij} + \mathbf{r}_{2ij}\\
\mathbf{t}_{9ij} &= \boldsymbol{\rho}_{(i+1)j} - (\boldsymbol{\rho}_{ij} + \mathbf{r}_{1ij})\\
\mathbf{t}_{10ij} &= \hat{\boldsymbol{\rho}}_{(i+1)j} + \mathbf{r}_{2(i+1)j} - \hat{\boldsymbol{\rho}}_{ij}.
\end{aligned} \tag{18.31}$$

For $j = 1$ we replace $\mathbf{t}_{2ij}$ with

$$\mathbf{t}_{2i1} = \boldsymbol{\rho}_{i1} - \boldsymbol{\rho}_{(i+1)1}.$$

For $j = m$ we replace $\mathbf{t}_{6ij}$ and $\mathbf{t}_{7ij}$ with

$$\begin{aligned}
\mathbf{t}_{6im} &= \hat{\boldsymbol{\rho}}_{(i+1)m} + \mathbf{r}_{2(i+1)m} - (\hat{\boldsymbol{\rho}}_{im} + \mathbf{r}_{2im})\\
\mathbf{t}_{7im} &= \hat{\boldsymbol{\rho}}_{im} - (\hat{\boldsymbol{\rho}}_{(i+1)m} + \mathbf{r}_{2(i+1)m}).
\end{aligned}$$

where $\boldsymbol{\rho}_{0j} \triangleq \boldsymbol{\rho}_{nj}$, $\hat{\boldsymbol{\rho}}_{oj} \triangleq \hat{\boldsymbol{\rho}}_{nj}$, and $i + n = i$. Equation (18.31) has the matrix form,

$$\mathbf{t}_{ij}^{d} \triangleq \begin{bmatrix} \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \\ \mathbf{t}_6 \\ \mathbf{t}_7 \\ \mathbf{t}_8 \\ \mathbf{t}_9 \\ \mathbf{t}_{10} \end{bmatrix}_{ij} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{i(j-1)} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i-1)j}$$

© 2002 by CRC Press LLC

$$+\begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{ij} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i+1)j}$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i+1)(j+1)},$$

$$\mathbf{t}_{i1}^{d} \triangleq \begin{bmatrix} \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \\ \mathbf{t}_6 \\ \mathbf{t}_7 \\ \mathbf{t}_8 \\ \mathbf{t}_9 \\ \mathbf{t}_{10} \end{bmatrix}_{i1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i-1)1} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{i1} \tag{18.32}$$

$$+\begin{bmatrix} -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i+1)1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i+1)2},$$

© 2002 by CRC Press LLC

$$\mathbf{t}_{im}^{d} \triangleq \begin{bmatrix} \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \\ \mathbf{t}_6 \\ \mathbf{t}_7 \\ \mathbf{t}_8 \\ \mathbf{t}_9 \\ \mathbf{t}_{10} \end{bmatrix}_{im} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{i(m-1)} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i-1)m}$$

$$+ \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{im} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{(i+1)m} .$$

Equation (18.25) yields

$$\begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{ij} = \begin{bmatrix} \frac{1}{2}\mathbf{I}_3 & -\frac{1}{2}\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{I}_3 & -\frac{1}{2}\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{ij} \tag{18.33}$$

Hence, (18.32) and (18.33) yield

$$\mathbf{t}_{ij}^{d} \triangleq \begin{bmatrix} \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \\ \mathbf{t}_6 \\ \mathbf{t}_7 \\ \mathbf{t}_8 \\ \mathbf{t}_9 \\ \mathbf{t}_{10} \end{bmatrix}_{ij} = \frac{1}{2} \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{i(j-1)} + \frac{1}{2} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{(i-1)j}$$

© 2002 by CRC Press LLC

$$+\frac{1}{2}\begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{ij} + \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{(i+1)j}$$

$$+\frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{(i+1)(j+1),}$$

$$\mathbf{t}_{i1}^{d} \triangleq \begin{bmatrix} \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \\ \mathbf{t}_6 \\ \mathbf{t}_7 \\ \mathbf{t}_8 \\ \mathbf{t}_9 \\ \mathbf{t}_{10} \end{bmatrix}_{i1} = \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{(i-1)1} + \frac{1}{2}\begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{i1}$$

$$+\frac{1}{2}\begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{(i+1)1} + \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{(i+1)2,} \tag{18.34}$$

© 2002 by CRC Press LLC

$$\mathbf{t}_{im}^{d} \triangleq \begin{bmatrix} \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \\ \mathbf{t}_6 \\ \mathbf{t}_7 \\ \mathbf{t}_8 \\ \mathbf{t}_9 \\ \mathbf{t}_{10} \end{bmatrix}_{im} = \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{i(m-1)} + \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{q}_{(i-1)m}$$

$$+ \frac{1}{2}\begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{im} + \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{q}_{(i+1)m}.$$

Also, from (18.30) and (18.32)

$$\mathbf{t}_{1n1} = \begin{bmatrix} -\mathbf{I}_{3,}\ \mathbf{0}, \mathbf{0}, \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{11} + \begin{bmatrix} \mathbf{I}_{3,}\ \mathbf{I}_{3,}\ \mathbf{0}, \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \hat{\boldsymbol{\rho}} \\ \mathbf{r}_2 \end{bmatrix}_{n1}$$

$$\mathbf{t}_{1n1} = \begin{bmatrix} -\tfrac{1}{2}\mathbf{I}_3, \tfrac{1}{2}\mathbf{I}_{3,}\ \mathbf{0}, \mathbf{0} \end{bmatrix} \mathbf{q}_{11} + \begin{bmatrix} \tfrac{1}{2}\mathbf{I}_3, \tfrac{1}{2}\mathbf{I}_3, \mathbf{0}, \mathbf{0} \end{bmatrix} \mathbf{q}_{n1},$$

$$= \mathbf{E}_6\mathbf{q}_{11} + \mathbf{E}_7\mathbf{q}_{n1}, \tag{18.35}$$

$$= \begin{bmatrix} \mathbf{E}_6, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{E}_7 \end{bmatrix} \begin{bmatrix} \mathbf{q}_{11} \\ \mathbf{q}_{21} \\ \vdots \\ \mathbf{q}_{n1} \end{bmatrix}, \mathbf{E}_6 \in^{3\times 12}, \mathbf{E}_7 \in \mathbf{R}^{3\times 12},$$

$$\mathbf{t}_{1n1} = \mathbf{R}_0\mathbf{q}_1 = \begin{bmatrix} \mathbf{R}_{0,}\ \mathbf{0} \end{bmatrix} \mathbf{q}, \mathbf{R}_0 \in \mathbf{R}^{3\times 12n}.$$

With the obvious definitions of the 24×12 matrices $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, $\mathbf{E}_4$, $\hat{\mathbf{E}}_4$, $\overline{\mathbf{E}}_4$, $\mathbf{E}_5$, equations in (18.34) are written in the form, where $\mathbf{q}_{01} = \mathbf{q}_{n1}$, $\mathbf{q}_{(n+1)j} = \mathbf{q}_{ij}$,

$$\begin{aligned} \mathbf{t}_{il}^{d} &= \mathbf{E}_2\mathbf{q}_{(i-1)1} + \mathbf{E}_3\mathbf{q}_{i1} + \hat{\mathbf{E}}_4\mathbf{q}_{(i+1)1} + \mathbf{E}_5\mathbf{q}_{(i+1)2}, \\ \mathbf{t}_{ij}^{d} &= \mathbf{E}_1\mathbf{q}_{i(j-1)} + \mathbf{E}_2\mathbf{q}_{(i-1)j} + \mathbf{E}_3\mathbf{q}_{ij} + \mathbf{E}_4\mathbf{q}_{(i+1)j} + \mathbf{E}_5\mathbf{q}_{(i+1)(j+1)}, \\ \mathbf{t}_{im}^{d} &= \mathbf{E}_1\mathbf{q}_{i(m-1)} + \mathbf{E}_2\mathbf{q}_{(i-1)m} + \mathbf{E}_3\mathbf{q}_{im} + \overline{\mathbf{E}}_4\mathbf{q}_{(i+1)m}. \end{aligned} \tag{18.36}$$

© 2002 by CRC Press LLC

Now from (18.34) and (18.35), define

$$\mathbf{l}^d = \left[\mathbf{t}_{1n1}^{dT}, \mathbf{t}_{11}^{dT}, \mathbf{t}_{21}^{dT}, \ldots, \mathbf{t}_{n1}^{dT} \middle| \mathbf{t}_{12}^{dT}, \ldots, \mathbf{t}_{n2}^{dT} \middle| \ldots, \mathbf{t}_{nm}^{dT}\right]^T$$

$$= \left[\mathbf{t}_{1n1}^{dT}, \mathbf{t}_1^{dT}, \mathbf{t}_2^{dT}, \ldots, \mathbf{t}_n^{dT}\right]^T,$$

to get

$$\mathbf{l}^d = \mathbf{R}\mathbf{q}, \mathbf{R} \in \mathbf{R}^{(24nm+3)\times 12nm}, \mathbf{q} \in \mathbf{R}^{12nm}, \mathbf{l}^d \in \mathbf{R}^{(24nm+3)}, \tag{18.37}$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \hat{\mathbf{R}}_{11} & \mathbf{R}_{12} & \ddots & & & \vdots \\ \mathbf{R}_{21} & \mathbf{R}_{11} & \mathbf{R}_{12} & \ddots & & \vdots \\ \mathbf{0} & \mathbf{R}_{21} & \mathbf{R}_{11} & \mathbf{R}_{12} & \ddots & \vdots \\ \vdots & \ddots & \mathbf{R}_{21} & \mathbf{R}_{11} & \ddots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & \mathbf{R}_{12} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{R}_{21} & \overline{\mathbf{R}}_{11} \end{bmatrix}, \mathbf{R}_{ij} \in \mathbf{R}^{24n\times 12n}, \mathbf{R}_0 \in \mathbf{R}^{3\times 12n},$$

$$\mathbf{R}_{11} = \begin{bmatrix} \mathbf{E}_3 & \mathbf{E}_4 & \mathbf{0} & \cdots & \cdots & \mathbf{E}_2 \\ \mathbf{E}_2 & \mathbf{E}_3 & \mathbf{E}_4 & \ddots & & \vdots \\ \mathbf{0} & \mathbf{E}_2 & \mathbf{E}_3 & \mathbf{E}_4 & \ddots & \vdots \\ \vdots & \ddots & \mathbf{E}_2 & \mathbf{E}_3 & \mathbf{E}_4 & \mathbf{0} \\ \mathbf{0} & & \ddots & \ddots & \ddots & \mathbf{E}_4 \\ \mathbf{E}_4 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{E}_2 & \mathbf{E}_3 \end{bmatrix}, \mathbf{R}_{12} = \begin{bmatrix} \mathbf{0} & \mathbf{E}_5 & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}_5 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & & & \ddots & \ddots & \mathbf{E}_5 \\ \mathbf{E}_5 & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{R}_{i(i+k)} = \mathbf{0} \text{ if } k > 1, \mathbf{R}_{(i+k)i} = \mathbf{0} \text{ if } k > 1$$

$$\mathbf{R}_{21} = BlockDiag\left[\cdots, \mathbf{E}_1\ \mathbf{E}_1, \cdots\right],\ \mathbf{E}_i \in \mathbf{R}^{24\times 12}\ i = 1 \to 5$$

$$\mathbf{R}_0 = \left[\mathbf{E}_6, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{E}_7\right], \mathbf{E}_6 = \frac{1}{2}\left[-\mathbf{I}_3, \mathbf{I}_3, \mathbf{0}, \mathbf{0}\right],$$

$$\mathbf{E}_7 = \frac{1}{2}\left[\mathbf{I}_3,\ \mathbf{I}_3,\ \mathbf{0},\ \mathbf{0}\right].$$

$\hat{\mathbf{R}}_{11}$ and $\overline{\mathbf{R}}_{11}$ have the same structure as $\mathbf{R}_{11}$ except $\mathbf{E}_4$ is replaced by $\hat{\mathbf{E}}_4$, and $\overline{\mathbf{E}}_4$, respectively. Equation (18.37) will be needed to express the tendon forces in terms of $\mathbf{q}$. Equations (18.28) and (18.37) yield the dependent vectors ($\mathbf{t}_{1n1}$, $\mathbf{t}_2$, $\mathbf{t}_3$, $\mathbf{t}_4$, $\mathbf{t}_6$, $\mathbf{t}_7$, $\mathbf{t}_9$, $\mathbf{t}_{10}$) in terms of the independent vectors ($\mathbf{t}_5$, $\mathbf{t}_1$, $\mathbf{r}_1$, $\mathbf{r}_2$). Therefore,

$$\mathbf{l}^d = \mathbf{R}\mathbf{Q}\mathbf{l}. \tag{18.38}$$

18.5 Tendon Forces

Let the tendon forces be described by

$$\mathbf{f}_{\alpha ij} = F_{\alpha ij} \frac{\mathbf{t}_{\alpha ij}}{\left\|\mathbf{t}_{\alpha ij}\right\|} \tag{18.39}$$

© 2002 by CRC Press LLC

For tensegrity structures with some slack strings, the magnitude of the force $F_{\alpha ij}$ can be zero, for taut strings $F_{\alpha ij} > 0$. Because tendons cannnot compress, $F_{\alpha ij}$ cannot be negative. Hence, the magnitude of the force is

$$F_{\alpha ij} = k_{\alpha ij}\left(\left\|\mathbf{t}_{\alpha ij}\right\| - \bar{L}_{\alpha ij}\right) \tag{18.40}$$

where

$$k_{\alpha ij} \triangleq \begin{cases} 0 & , \; if \; \bar{L}_{\alpha ij} > \left\|\mathbf{t}_{\alpha ij}\right\| \\ \bar{k}_{\alpha ij} > 0 & , \; if \; \bar{L}_{\alpha ij} \le \left\|\mathbf{t}_{\alpha ij}\right\| \end{cases}$$

$$\bar{L}_{\alpha ij} - u_{\alpha ij} + L^{o}_{\alpha ij} \ge 0 \tag{18.41}$$

where $L^{o}_{\alpha ij} > 0$ is the rest length of tendon $t_{\alpha ij}$ before any control is applied, and the control is $u_{\alpha ij}$, the change ın the rest length. The control shortens or lengthens the tendon, so $u_{\alpha ij}$ can be positive or negative, but $L^{o}_{\alpha ij} > 0$. So $u_{\alpha ij}$ must obey the constraint (18.41), and

$$u_{\alpha ij} \le L^{o}_{\alpha ij} > 0. \tag{18.42}$$

Note that for t_{1n1} and for $\alpha = 2, 3, 4, 6, 7, 8, 9, 10$ the vectors $\mathbf{t}_{\alpha ij}$ appear in the vector $\mathbf{l}^{d}$ related to $\mathbf{q}$ from (4.7) by $\mathbf{l}^{d} = \mathbf{R}\mathbf{q}$, and for $\alpha = 5, 1$ the vectors $\mathbf{t}_{\alpha ij}$ appear in the vector $\mathbf{l}$ related to $\mathbf{q}$ from (18.28), by $\mathbf{l} = \mathbf{Q}^{-1}\,\mathbf{q}$. Let $P_{\alpha ij}$ denote the selected row of $\mathbf{R}$ associated with $\mathbf{t}_{\alpha ij}$ for $\alpha ij = 1n1$ and for $\alpha = 2, 3, 4, 6, 7, 8, 9, 10$. Let $P_{\alpha ij}$ also denote the selected row of $\mathbf{Q}^{-1}$ when $\alpha = 5, 1$. Then,

$$\mathbf{t}_{\alpha ij} = \mathcal{R}_{\alpha ij}\mathbf{q}, \; \mathcal{R}_{\alpha ij} \in^{3\times 12nm} \tag{18.43}$$

$$\left\|\mathbf{t}_{\alpha ij}\right\|^{2} = \mathbf{q}^{T}\mathcal{R}^{T}_{\alpha ij}\mathcal{R}_{\alpha_{ij}}\mathbf{q} \tag{18.44}$$

From (18.39) and (18.40),

$$\mathbf{f}_{\alpha ij} = -\,\mathbf{K}_{\alpha ij}\,(\mathbf{q})\mathbf{q} + \mathbf{b}_{\alpha ij}\,(\mathbf{q})u_{\alpha ij}$$

where

$$\mathbf{K}_{\alpha ij}(\mathbf{q}) \triangleq k_{\alpha ij}\left(L^{o}_{\alpha ij}\left(\mathbf{q}^{T}\mathcal{R}^{T}_{\alpha ij}\mathcal{R}_{\alpha ij}\mathbf{q}\right)^{-\frac{1}{2}} - 1\right)\mathcal{R}_{\alpha ij}, \; \mathbf{K}_{\alpha ij} \in \boldsymbol{R}^{3\times 12nm} \tag{18.45}$$

$$\mathbf{b}_{\alpha ij}(\mathbf{q}) \triangleq k_{\alpha ij}(\mathbf{q}^{T}\mathbf{R}^{T}_{\alpha ij}\mathbf{R}_{\alpha ij}\mathbf{q})^{-\frac{1}{2}}\mathbf{R}_{\alpha ij}\mathbf{q}, \; \mathbf{b}_{\alpha ij} \in \mathbf{R}^{3\times 1} \tag{18.46}$$

Hence,

$$\mathbf{f}^{d}_{ij} = \begin{bmatrix} \mathbf{f}_{2ij} \\ \mathbf{f}_{3ij} \\ \mathbf{f}_{4ij} \\ \mathbf{f}_{6ij} \\ \mathbf{f}_{7ij} \\ \mathbf{f}_{8ij} \\ \mathbf{f}_{9ij} \\ \mathbf{f}_{10ij} \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_{2ij} \\ \mathbf{K}_{3ij} \\ \mathbf{K}_{4ij} \\ \mathbf{K}_{6ij} \\ \mathbf{K}_{7ij} \\ \mathbf{K}_{8ij} \\ \mathbf{K}_{9ij} \\ \mathbf{K}_{10ij} \end{bmatrix}\mathbf{q}$$

© 2002 by CRC Press LLC

$$+\begin{bmatrix} \mathbf{b}_{2ij} & & & & & & & & & \\ & \mathbf{b}_{3ij} & & & & & & & & \\ & & \mathbf{b}_{4ij} & & & & & & & \\ & & & \mathbf{b}_{6ij} & & & & & & \\ & & & & \mathbf{b}_{7ij} & & & & & \\ & & & & & \mathbf{b}_{8ij} & & & & \\ & & & & & & \mathbf{b}_{9ij} & & & \\ & & & & & & & \mathbf{b}_{10ij} & & \end{bmatrix} \begin{bmatrix} u_{2ij} \\ u_{3ij} \\ u_{4ij} \\ u_{6ij} \\ u_{7ij} \\ u_{8ij} \\ u_{9ij} \\ u_{10ij} \end{bmatrix}$$

or

$$\mathbf{f}_{ij}^{d} = -\mathbf{K}_{ij}^{d}\mathbf{q} + \mathbf{P}_{ij}^{d}\mathbf{u}_{ij}^{d}, \tag{18.47}$$

and

$$\mathbf{f}_{ij}^{o} = \begin{bmatrix} \mathbf{f}_{5ij} \\ \mathbf{f}_{1ij} \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_{5ij} \\ \mathbf{K}_{1ij} \end{bmatrix}\mathbf{q} + \begin{bmatrix} \mathbf{b}_{5ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_{1ij} \end{bmatrix} \begin{bmatrix} u_{5ij} \\ u_{1ij} \end{bmatrix}$$

or

$$\mathbf{f}_{ij}^{o} = -\mathbf{K}_{ij}^{o}\mathbf{q} + \mathbf{P}_{ij}^{o}\mathbf{u}_{ij}^{o}. \tag{18.48}$$

Now substitute (18.47) and (18.48) into

$$\bar{\mathbf{f}}_1^{d} = \begin{bmatrix} \mathbf{f}_{1n1} \\ \mathbf{f}_{11}^{d} \\ \mathbf{f}_{21}^{d} \\ \vdots \\ \mathbf{f}_{n1}^{d} \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_{1n1} \\ \mathbf{K}_{11}^{d} \\ \mathbf{K}_{21}^{d} \\ \vdots \\ \mathbf{K}_{n1}^{d} \end{bmatrix}\mathbf{q} + \begin{bmatrix} \mathbf{P}_{1n1} & & & & \\ & \mathbf{P}_{11}^{d} & & & \\ & & \mathbf{P}_{21}^{d} & & \\ & & & \ddots & \\ & & & & \mathbf{P}_{n1}^{d} \end{bmatrix} \begin{bmatrix} u_{1n1} \\ \mathbf{u}_{11}^{d} \\ \mathbf{u}_{21}^{d} \\ \vdots \\ \mathbf{u}_{n1}^{d} \end{bmatrix} = -\bar{\mathbf{K}}_1^{d}\mathbf{q} + \bar{\mathbf{P}}_1^{d}\bar{\mathbf{u}}_1^{d}$$

$$\mathbf{f}_2^{d} = \begin{bmatrix} \mathbf{f}_{12}^{d} \\ \mathbf{f}_{22}^{d} \\ \vdots \\ \mathbf{f}_{n2}^{d} \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_{12}^{d} \\ \mathbf{K}_{22}^{d} \\ \vdots \\ \mathbf{K}_{n2}^{d} \end{bmatrix}\mathbf{q} + \begin{bmatrix} \mathbf{P}_{12}^{d} & & & \\ & \mathbf{P}_{22}^{d} & & \\ & & \ddots & \\ & & & \mathbf{P}_{n2}^{d} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{12}^{d} \\ \mathbf{u}_{22}^{d} \\ \vdots \\ \mathbf{u}_{n2}^{d} \end{bmatrix} = -\mathbf{K}_2^{d}\mathbf{q}_2 + \mathbf{P}_2^{d}\mathbf{u}_2^{d}.$$

Hence, in general,

$$\mathbf{f}_j^{d} = -\mathbf{K}_j^{d}\mathbf{q} + \mathbf{P}_j^{d}\mathbf{u}_j^{d}$$

or by defining

$$\mathbf{K}^{d} = \begin{bmatrix} \bar{\mathbf{K}}_1^{d} \\ \mathbf{K}_2^{d} \\ \vdots \\ \mathbf{K}_m^{d} \end{bmatrix}, \quad \mathbf{P}^{d} = \begin{bmatrix} \bar{\mathbf{P}}_1^{d} \\ \mathbf{P}_2^{d} \\ \vdots \\ \mathbf{P}_m^{d} \end{bmatrix} \tag{18.49}$$

© 2002 by CRC Press LLC

$$\mathbf{f}^d = -\mathbf{K}^d\mathbf{q} + \mathbf{P}^d\mathbf{u}^d.$$

Likewise, for $\mathbf{f}_{11}^o$ forces (18.48),

$$\mathbf{f}_1^o = \begin{bmatrix} \mathbf{f}_{11}^o \\ \mathbf{f}_{21}^o \\ \vdots \\ \mathbf{f}_{n1}^o \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_{11}^o \\ \mathbf{K}_{21}^o \\ \vdots \\ \mathbf{K}_{n1}^o \end{bmatrix}\mathbf{q} + \begin{bmatrix} \mathbf{P}_{11}^o & & & \\ & \mathbf{P}_{21}^o & & \\ & & \ddots & \\ & & & \mathbf{P}_{n1}^o \end{bmatrix}\begin{bmatrix} \mathbf{u}_{11}^o \\ \mathbf{u}_{21}^o \\ \vdots \\ \mathbf{u}_{n1}^o \end{bmatrix}$$

$$\mathbf{f}_j^o = \begin{bmatrix} \mathbf{f}_{1j}^o \\ \mathbf{f}_{2j}^o \\ \vdots \\ \mathbf{f}_{nj}^o \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_{1j}^o \\ \mathbf{K}_{2j}^o \\ \vdots \\ \mathbf{K}_{nj}^o \end{bmatrix}\mathbf{q} + \begin{bmatrix} \mathbf{P}_{1j}^o & & & \\ & \mathbf{P}_{2j}^o & & \\ & & \ddots & \\ & & & \mathbf{P}_{nj}^o \end{bmatrix}\begin{bmatrix} \mathbf{u}_{1j}^o \\ \mathbf{u}_{2j}^o \\ \vdots \\ \mathbf{u}_{nj}^o \end{bmatrix}$$

$$\mathbf{f}_j^o = -\mathbf{K}_j^o\mathbf{q} + \mathbf{P}_j^o\mathbf{u}_j^o \tag{18.50}$$

$$\mathbf{f}^o = -\mathbf{K}^o\mathbf{q} + \mathbf{P}^o\mathbf{u}^o.$$

Substituting (18.49) and (18.50) into (18.E.21) yields

$$\mathbf{f} = -(\mathbf{B}^d\mathbf{K}^d + \mathbf{B}^o\mathbf{K}^o)\mathbf{q} + \mathbf{B}^d\mathbf{P}^d\mathbf{u}^d + \mathbf{B}^o\mathbf{P}^o\mathbf{u}^o + \mathbf{W}^o\mathbf{w}, \tag{18.51}$$

which is written simply as

$$\mathbf{f} = -\tilde{\mathbf{K}}\mathbf{q} + \tilde{\mathbf{B}}\mathbf{u} + \mathbf{W}^o\mathbf{w}, \tag{18.52}$$

by defining,

$$\tilde{\mathbf{K}} \triangleq \mathbf{B}^d\mathbf{K}^d + \mathbf{B}^o\mathbf{K}^o,$$

$$\tilde{\mathbf{B}} \triangleq \left[\mathbf{B}^d\mathbf{P}^d,\ \mathbf{B}^o\mathbf{P}^o\right],$$

$$\mathbf{B}^d\mathbf{P}^d = \begin{bmatrix} \mathbf{B}_3\overline{\mathbf{P}}_1^d & \mathbf{B}_4\mathbf{P}_2^d & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{B}_5\overline{\mathbf{P}}_1^d & \mathbf{B}_6\mathbf{P}_2^d & \mathbf{B}_4\mathbf{P}_3^d & \ddots & & \vdots \\ \mathbf{0} & \overline{\mathbf{B}}_5\mathbf{P}_2^d & \mathbf{B}_6\mathbf{P}_3^d & \ddots & \ddots & \vdots \\ \vdots & \ddots & \overline{\mathbf{B}}_5\mathbf{P}_3^d & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & \mathbf{B}_4\mathbf{P}_m^d \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \overline{\mathbf{B}}_5\mathbf{P}_{m-1}^d & \mathbf{B}_8\mathbf{P}_m^d \end{bmatrix},$$

© 2002 by CRC Press LLC

$$\mathbf{B}^o\mathbf{P}^o = \begin{bmatrix} \mathbf{B}_1\mathbf{P}_1^o & \mathbf{B}_2\mathbf{P}_2^o & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_7\mathbf{P}_2^o & \mathbf{B}_2\mathbf{P}_3^o & \ddots & \vdots \\ \vdots & \ddots & \mathbf{B}_7\mathbf{P}_3^o & \ddots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \mathbf{B}_2\mathbf{P}_m^o \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{B}_7\mathbf{P}_m^o \end{bmatrix}$$

$$\tilde{\mathbf{K}} = \mathbf{B}^d\mathbf{K}^d + \mathbf{B}^o\mathbf{K}^o = \begin{bmatrix} \mathbf{B}_3\overline{\mathbf{K}}_1^d + \mathbf{B}_1\mathbf{K}_1^o + \mathbf{B}_4\mathbf{K}_2^d + \mathbf{B}_2\mathbf{K}_2^o \\ \mathbf{B}_5\overline{\mathbf{K}}_1^d + \mathbf{B}_6\mathbf{K}_2^d + \mathbf{B}_4\mathbf{K}_3^d + \mathbf{B}_7\mathbf{K}_2^o + \mathbf{B}_2\mathbf{K}_3^o \\ \overline{\mathbf{B}}_5\mathbf{K}_2^d + \mathbf{B}_6\mathbf{K}_3^d + \mathbf{B}_4\mathbf{K}_4^d + \mathbf{B}_7\mathbf{K}_3^o + \mathbf{B}_2\mathbf{K}_4^o \\ \overline{\mathbf{B}}_5\mathbf{K}_3^d + \mathbf{B}_5\mathbf{K}_4^d + \mathbf{B}_4\mathbf{K}_5^d + \mathbf{B}_7\mathbf{K}_4^o + \mathbf{B}_2\mathbf{K}_5^o \\ \vdots \\ \overline{\mathbf{B}}_5\mathbf{K}_{m-2}^d + \mathbf{B}_6\mathbf{K}_{m-1}^d + \mathbf{B}_4\mathbf{K}_m^d + \mathbf{B}_7\mathbf{K}_{m-1}^o + \mathbf{B}_2\mathbf{K}_m^o \\ \overline{\mathbf{B}}_5\mathbf{K}_{m-1}^d + \mathbf{B}_6\mathbf{K}_m^d + \mathbf{B}_7\mathbf{K}_m^o \end{bmatrix} \tag{18.53}$$

$$\tilde{\mathbf{u}} = \begin{bmatrix} \overline{\mathbf{u}}_1^d \\ \mathbf{u}_2^d \\ \mathbf{u}_3^d \\ \mathbf{u}_4^d \\ \vdots \\ \mathbf{u}_m^d \\ \mathbf{u}_1^o \\ \mathbf{u}_2^o \\ \mathbf{u}_3^o \\ \mathbf{u}_4^o \\ \vdots \\ \mathbf{u}_m^o \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \hat{\mathbf{u}}_1^d \\ \mathbf{u}_2^d \\ \mathbf{u}_3^d \\ \mathbf{u}_4^d \\ \vdots \\ \hat{\mathbf{u}}_m^d \\ \mathbf{u}_1^o \\ \mathbf{u}_2^o \\ \mathbf{u}_3^o \\ \mathbf{u}_4^o \\ \vdots \\ \mathbf{u}_m^o \end{bmatrix} \tag{18.54}$$

In vector $\tilde{\mathbf{u}}$ in (18.54), u_{1n1} appears twice (for notational convenience u_{1n1} appears in $\overline{\mathbf{u}}_1^d$ and in $\mathbf{u}_1^o$). From the rules of closure, $\mathbf{t}_{9i1} = -\mathbf{t}_{1i1}$ and $\mathbf{t}_{7im} = -\mathbf{t}_{10im}$, $i = 1, 2, \ldots, n$, but $\mathbf{t}_{1i1}$, $\mathbf{t}_{7im}$, $\mathbf{t}_{9i1}$, $\mathbf{t}_{10im}$ all appear in (18.54). Hence, the rules of closure leave only $n(10m - 2)$ tendons in the structure, but (18.54) contains $10nm + 1$ tendons. To eliminate the redundant variables in (18.54) define $\tilde{\mathbf{u}} = \mathbf{T}\mathbf{u}$, where $\mathbf{u}$ is the independent set $\mathbf{u} \in^{n(10m-2)}$, and $\tilde{\mathbf{u}} \in \boldsymbol{R}^{10nm+1}$ is given by (18.54). We choose to keep $\mathbf{t}_{7im}$ in u and delete $\mathbf{t}_{10im}$ by setting $\mathbf{t}_{10im} = -\mathbf{t}_{7im}$. We choose to keep $\mathbf{t}_{1i1}$ and delete $\mathbf{t}_{9i1}$ by setting $\mathbf{t}_{9i1} = -\mathbf{t}_{1i1}$, $i = 1, 2, \ldots,$ n. This requires new definitions of certain subvectors as follows in (18.57) and (18.58). The vector $\tilde{\mathbf{u}}$ is now defined in (18.54). We have reduced the $\tilde{\mathbf{u}}$ vector by $2n + 1$ scalars to $\mathbf{u}$. The $\mathbf{T}$ matrix is formed by the following blocks,

© 2002 by CRC Press LLC

$$\mathbf{T} = \left(\begin{array}{c|c|c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots\ \mathbf{0}1 & \mathbf{0} \\ \hline \mathbf{T}_1 \;\ddots\; \mathbf{T}_1 & & & \mathbf{S} \;\ddots\; \mathbf{S} & \\ \hline & \mathbf{I}_8 \;\ddots\; \mathbf{I}_8 & & & \\ \hline & & \mathbf{T}_2 \;\ddots\; \mathbf{T}_2 & & \\ \hline & & & \mathbf{I}_2 \;\ddots\; \mathbf{I}_2 & \\ \hline & & & & \mathbf{I}_2 \;\ddots\; \mathbf{I}_2 \end{array}\right) \in \boldsymbol{R}^{(10nm+1)\times(n(10m-2))} \tag{18.55}$$

where

$$\mathbf{T}_1 = \left(\begin{array}{c|c} \mathbf{I}_6 & \mathbf{0}_{6\times 1} \\ \hline \mathbf{0}_{1\times 6} & 0 \\ \mathbf{0}_{1\times 6} & 1 \end{array}\right) \in \boldsymbol{R}^{8\times 7}$$

$$\mathbf{T}_2 = \left(\begin{array}{c} \mathbf{I}_7 \\ \hline 0\ 0\ 0\ 0\ -1\ 0\ 0 \end{array}\right) \in \boldsymbol{R}^{8\times 7}$$

$$\mathbf{S} = \left(\begin{array}{c|c} \mathbf{0}_{6\times 2} & \\ \hline 0 & -1 \\ 0 & \mathbf{0} \end{array}\right) \in \boldsymbol{R}^{8\times 2}. \tag{18.56}$$

There are n blocks labeled $\mathbf{T}_1$, $n(m-2)$ blocks labeled $\mathbf{I}_8$ (for $m \le 2$ no $\mathbf{I}_8$ blocks needed, see appendix 18.D), n blocks labeled $\mathbf{T}_2$, nm blocks labeled $\mathbf{I}_2$ blocks, and n blocks labeled $\mathbf{S}$.

The $\mathbf{u}_1^d$ block becomes

$$\hat{\mathbf{u}}_1^d \triangleq \begin{bmatrix} \hat{\mathbf{u}}_{11}^d \\ \hat{\mathbf{u}}_{21}^d \\ \hat{\mathbf{u}}_{31}^d \\ \vdots \\ \hat{\mathbf{u}}_{n1}^d \end{bmatrix}, \quad \hat{\mathbf{u}}_{i1}^d = \begin{bmatrix} u_{2n1} \\ u_{3n1} \\ u_{4n1} \\ u_{6n1} \\ u_{7n1} \\ u_{8n1} \\ u_{10n1} \end{bmatrix} \in \boldsymbol{R}^{7\times 1}, \quad i = 1,\ 2,\ 3,\ \ldots n,\ \ j = 1 \tag{18.57}$$

© 2002 by CRC Press LLC

The $\mathbf{u}_m^d$ block becomes

$$\hat{\mathbf{u}}_m^d \triangleq \begin{bmatrix} \hat{\mathbf{u}}_{1m}^d \\ \hat{\mathbf{u}}_{2m}^d \\ \hat{\mathbf{u}}_{3m}^d \\ \vdots \\ \hat{\mathbf{u}}_{nm}^d \end{bmatrix}, \quad \hat{\mathbf{u}}_{im}^d = \begin{bmatrix} u_{2nm} \\ u_{3nm} \\ u_{4nm} \\ u_{6nm} \\ u_{7nm} \\ u_{8nm} \\ u_{9nm} \end{bmatrix} \in \boldsymbol{R}^{7\times 1}, \quad i = 1,\ 2,\ 3,\ \ldots n, \quad j = m. \tag{18.58}$$

The $\mathbf{u}_1^d$ block is the $\overline{\mathbf{u}}_1^d$ block with the first element u_{1n1} removed, because it is included in $\mathbf{u}_{n1}^o$.

From (18.17) and (18.52),

$$\ddot{\mathbf{q}} + (\mathbf{K}_r(\dot{\mathbf{q}}) + \mathbf{K}_p(\mathbf{q}))\mathbf{q} = \mathbf{B}(\mathbf{q})\mathbf{u} + \mathbf{D}(\mathbf{q})\mathbf{w}, \tag{18.59}$$

where,

$$\mathbf{K}p = \mathbf{H}(\mathbf{q})\ \tilde{\mathbf{K}}(\mathbf{q}),$$

$$\mathbf{B} = \mathbf{H}(\mathbf{q})\ \tilde{\mathbf{B}}(\mathbf{q})\mathbf{T},$$

$$\mathbf{D} = \mathbf{H}(\mathbf{q})\mathbf{W}^o.$$

The nodal points of the structure are located by the vector $\mathbf{p}$. Suppose that a selected set of nodal points are chosen as outputs of interest. Then

$$\mathbf{y}p = \mathbf{C}\mathbf{p} = \mathbf{C}\mathbf{P}\mathbf{q} \tag{18.60}$$

where $\mathbf{P}$ is defined by (18.26). The length of tendon vector $\mathbf{t}_{\alpha ij} = \mathcal{R}_{\alpha ij}\mathbf{q}$ is given from (18.44). Therefore, the output vector $\mathbf{y}_l$ describing all tendon lengths, is

$$\mathbf{y}_l = \begin{bmatrix} \vdots \\ \mathbf{y}_{\alpha ij} \\ \vdots \end{bmatrix}, \quad \mathbf{y}_{\alpha ij} = \left(\mathbf{q}^T \mathcal{R}_{\alpha ij}^T \mathcal{R}_{\alpha ij}\mathbf{q}\right)^{\frac{1}{2}}.$$

Another ouput of interest might be tension, so from (18.40) and (18.44)

$$\mathbf{y}_f = \begin{bmatrix} \vdots \\ F_{\alpha ij} \\ \vdots \end{bmatrix}, \quad F_{\alpha ij} = k_{\alpha ij}(\mathbf{y}_{\alpha ij} - \overline{L}_{\alpha ij}).$$

The static equilibria can be studied from the equations

$$\mathbf{K}_p(\mathbf{q})\mathbf{q} = \mathbf{B}(\mathbf{q})\mathbf{u} + \mathbf{D}(\mathbf{q})\mathbf{w},\ \mathbf{y}_p = \mathbf{C}\mathbf{P}\mathbf{q}. \tag{18.61}$$

© 2002 by CRC Press LLC

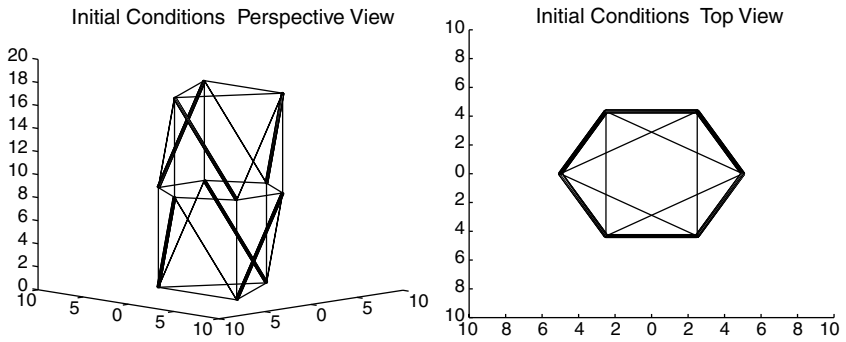


FIGURE 18.10 Initial conditions with nodal points on cylinder surface.

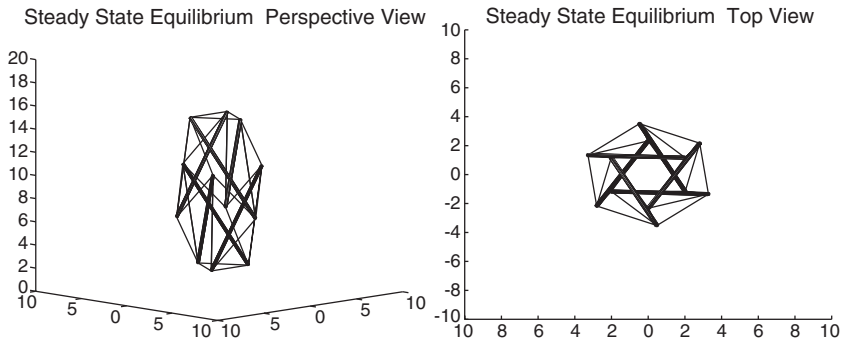


FIGURE 18.11 Steady-state equilibrium.

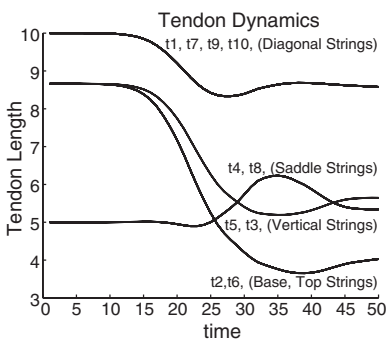


FIGURE 18.12 Tendon dynamics.

Of course, one way to generate equilibria is by simulation from arbitrary initial conditions and record the steady-state value of **q**. The exhaustive definitive study of the stable equilibria is in a separate paper.[27]

Damping strategies for controlled tensegrity structures are a subject of further research. The example case given in Appendix 18.D was coded in Matlab and simulated. Artificial critical damping was included in the simulation below. The simulation does not include external disturbances or control inputs. All nodes of the structure were placed symmetrically around the surface of a cylinder, as seen in Figure 18.10. Spring constants and natural rest lengths were specified equally for all tendons in the structure. One would expect the structure to collapse in on itself with this given initial condition. A plot of steady-state equilibrium is given in Figure 18.11 and string lengths in Figure 18.12.

© 2002 by CRC Press LLC

18.6 Conclusion

This chapter developed the exact nonlinear equations for a Class 1 tensegrity shell, having nm rigid rods and $n(10m - 2)$ tendons, subject to the assumption that the tendons are linear-elastic, and the rods are rigid rods of constant length. The equations are described in terms of $6nm$ degrees of freedom, and the accelerations are given explicitly. Hence, no inversion of the mass matrix is required. For large systems this greatly improves the accuracy of simulations.

Tensegrity systems of four classes are characterized by these models. Class 2 includes rods that are in contact at nodal points, with a ball joint, transmitting no torques. In Class 1 the rods do not touch and a stable equilibrium must be achieved by pretension in the tendons. The primal shell class contains the minimum number of tendons ($8nm$) for which stability is possible.

Tensegrity structures offer some potential advantages over classical structural systems composed of continua (such as columns, beams, plates, and shells). The overall structure can bend but all elements of the structure experience only axial loads, so no member bending. The absence of bending in the members promises more precise models (and hopefully more precise control). Prestress allows members to be uni-directionally loaded, meaning that no member experiences reversal in the direction of the load carried by the member. This eliminates a host of nonlinear problems known to create difficulties in control (hysteresis, dead zones, and friction).

Acknowledgment

The authors recognize the valuable efforts of T. Yamashita in the first draft of this chapter.

© 2002 by CRC Press LLC

Appendix 18.A Proof of Theorem 18.1

Refer to Figure 18.8 and define

$$\underline{\mathbf{q}}_1 = \underline{\mathbf{p}}_2 + \underline{\mathbf{p}}_1,\ \underline{\mathbf{q}}_2 = \underline{\mathbf{p}}_2 - \underline{\mathbf{p}}_1,$$

using the vectors $\underline{\mathbf{p}}_1$ and $\underline{\mathbf{p}}_2$ which locate the end points of the rod. The rod mass center is located by the vector,

$$\underline{\mathbf{p}}_c = \frac{1}{2}\underline{\mathbf{q}}_1. \tag{18.A.1}$$

Hence, the translation equation of motion for the mass center of the rod is

$$m\ddot{\underline{\mathbf{p}}}_c = \frac{m}{2}\ddot{\underline{\mathbf{q}}}_1 = (\hat{\underline{\mathbf{f}}}_1 + \hat{\underline{\mathbf{f}}}_2), \tag{18.A.2}$$

where a dot over a vector is a time derivative with respect to the inertial reference frame. A vector $\underline{\mathbf{p}}$ locating a mass element, *dm*, along the centerline of the rod is

$$\underline{\mathbf{p}} = \underline{\mathbf{p}}_1 + \rho(\underline{\mathbf{p}}_2 - \underline{\mathbf{p}}_1) = \frac{1}{2}\underline{\mathbf{q}}_1 + (\rho - \frac{1}{2})\underline{\mathbf{q}}_2,\ 0 \le \rho \le 1,\ \rho = \frac{x}{L}, \tag{18.A.3}$$

and the velocity of the mass *dm*, $\underline{\mathbf{v}}$, is

$$\underline{\mathbf{v}} = \dot{\underline{\mathbf{p}}} = \frac{1}{2}\dot{\underline{\mathbf{q}}}_1 + (\rho - \frac{1}{2})\dot{\underline{\mathbf{q}}}_2. \tag{18.A.4}$$

The angular momentum for the rod about the mass center, $\underline{\mathbf{h}}_c$, is

$$\underline{\mathbf{h}}_c = \int_m (-\underline{\mathbf{p}} - \underline{\mathbf{p}}_c) \times \dot{\underline{\mathbf{p}}}\, dm \tag{18.A.5}$$

where the mass *dm* can be described using $dx = L\, d\rho$ as

$$dm = \left(\frac{m}{L}\right)(L\, d\rho) = m\, d\rho. \tag{18.A.6}$$

Hence, (18.A.2) can be rewritten as follows:

$$\underline{\mathbf{h}}_c = \int_0^1 (\underline{\mathbf{p}} - \underline{\mathbf{p}}_c) \times \dot{\underline{\mathbf{p}}}(m\, d\rho) \tag{18.A.7}$$

where (18.A.1) and (18.A.3) yield

$$\underline{\mathbf{p}} - \underline{\mathbf{p}}_c = (\rho - \frac{1}{2})\underline{\mathbf{q}}_2. \tag{18.A.8}$$

(18.A.4)–(18.A.8) yield

© 2002 by CRC Press LLC

$$
\begin{aligned}
\underline{\mathbf{h}}_c &= m\int_0^1 (\rho - \frac{1}{2})\underline{\mathbf{q}}_2 \times \left\{\frac{1}{2}\dot{\underline{\mathbf{q}}}_1 + (\rho - \frac{1}{2})\dot{\underline{\mathbf{q}}}_2\right\} d\rho \\
&= m\underline{\mathbf{q}}_2 \times \left\{\dot{\underline{\mathbf{q}}}_1 \int_0^1 \frac{1}{2}(\rho - \frac{1}{2}) d\rho + \dot{\underline{\mathbf{q}}}_2 \int_0^1 (\rho - \frac{1}{2})^2 d\rho\right\} \\
&= m\underline{\mathbf{q}}_2 \times \left(\frac{1}{2}\left[\frac{1}{2}\rho^2 - \frac{1}{2}\rho\right]_0^1 \dot{\underline{\mathbf{q}}}_1 + \left[\frac{1}{3}(\rho - \frac{1}{2})^3\right]_0^1 \dot{\underline{\mathbf{q}}}_2\right) \\
&= \frac{m}{12}\underline{\mathbf{q}}_2 \times \dot{\underline{\mathbf{q}}}_2
\end{aligned}
\tag{18.A.9}
$$

The applied torque about the mass center, $\underline{\boldsymbol{\tau}}_c$, is

$$
\underline{\boldsymbol{\tau}}_c = \frac{1}{2}\underline{\mathbf{q}}_2 \times (\hat{\underline{\mathbf{f}}}_2 - \hat{\underline{\mathbf{f}}}_1)
$$

Then, substituting $\underline{\mathbf{h}}_c$ and $\underline{\boldsymbol{\tau}}_c$ from (18.A.9) into Euler's equations, we obtain

$$
\dot{\underline{\mathbf{h}}}_c = \underline{\boldsymbol{\tau}}_c
$$

or

$$
\begin{aligned}
\dot{\underline{\mathbf{h}}}_c &= \frac{m}{12}(\dot{\underline{\mathbf{q}}}_2 \times \dot{\underline{\mathbf{q}}}_2 + \underline{\mathbf{q}}_2 \times \ddot{\underline{\mathbf{q}}}_2) \\
&= \frac{m}{12}\underline{\mathbf{q}}_2 \times \ddot{\underline{\mathbf{q}}}_2 = \frac{1}{2}\underline{\mathbf{q}}_2 \times (\hat{\underline{\mathbf{f}}}_2 - \hat{\underline{\mathbf{f}}}_1)
\end{aligned}
\tag{18.A.10}
$$

Hence, (18.A.2) and (18.A.10) yield the motion equations for the rod:

$$
\begin{cases}
\frac{m}{2}\ddot{\underline{\mathbf{q}}}_1 = \hat{\underline{\mathbf{f}}}_1 + \hat{\underline{\mathbf{f}}}_2 \\
\frac{m}{6}(\underline{\mathbf{q}}_2 \times \ddot{\underline{\mathbf{q}}}_2) = \underline{\mathbf{q}}_2 \times (\hat{\underline{\mathbf{f}}}_2 - \hat{\underline{\mathbf{f}}}_1)
\end{cases}
\tag{18.A.11}
$$

We have assumed that the rod length L is constant. Hence, the following constraints for $\underline{\mathbf{q}}_2$ hold:

$$
\underline{\mathbf{q}}_2 \cdot \underline{\mathbf{q}}_2 = L^2
$$

$$
\frac{d}{dt}(\underline{\mathbf{q}}_2 \cdot \underline{\mathbf{q}}_2) = \dot{\underline{\mathbf{q}}}_2 \cdot \underline{\mathbf{q}}_2 + \underline{\mathbf{q}}_2 \cdot \dot{\underline{\mathbf{q}}}_2 = 2\underline{\mathbf{q}}_2 \cdot \dot{\underline{\mathbf{q}}}_2 = 0
$$

$$
\frac{d}{dt}(\underline{\mathbf{q}}_2 \cdot \dot{\underline{\mathbf{q}}}_2) = \dot{\underline{\mathbf{q}}}_2 \cdot \dot{\underline{\mathbf{q}}}_2 + \underline{\mathbf{q}}_2 \cdot \ddot{\underline{\mathbf{q}}}_2 = 0
$$

Collecting (18.A.11) and the constraint equations we have

$$
\begin{cases}
\frac{m}{2}\ddot{\underline{\mathbf{q}}}_1 = \hat{\underline{\mathbf{f}}}_1 + \hat{\underline{\mathbf{f}}}_2 \\
\frac{m}{6}(\underline{\mathbf{q}}_2 \times \ddot{\underline{\mathbf{q}}}_2) = \underline{\mathbf{q}}_2 \times (\hat{\underline{\mathbf{f}}}_2 - \hat{\underline{\mathbf{f}}}_1) \\
\dot{\underline{\mathbf{q}}}_2 \cdot \dot{\underline{\mathbf{q}}}_2 + \underline{\mathbf{q}}_2 \cdot \ddot{\underline{\mathbf{q}}}_2 = 0 \\
\underline{\mathbf{q}}_2 \cdot \underline{\mathbf{q}}_2 = L^2
\end{cases}
\tag{18.A.12}
$$

© 2002 by CRC Press LLC

We now develop the matrix version of (18.A.12). Recall that

$$\underline{\mathbf{q}}_i = \underline{\mathbf{E}}_{\mathbf{q}i} \quad ; \quad \hat{\underline{\mathbf{f}}}_i = \underline{\mathbf{E}}\hat{\underline{\mathbf{f}}}_i$$

Also note that $\underline{\mathbf{E}}^T \cdot \underline{\mathbf{E}} = 3 \times 3$ identity. After some manipulation, (18.A.12) can be written as:

$$\begin{aligned}
\tfrac{m}{2}\ddot{\mathbf{q}}_1 &= \hat{\mathbf{f}}_1 + \hat{\mathbf{f}}_2 \\
\tfrac{m}{6}\tilde{\mathbf{q}}_2\ddot{\mathbf{q}}_2 &= \tilde{\mathbf{q}}_2(\hat{\mathbf{f}}_2 - \hat{\mathbf{f}}_1) \\
\mathbf{q}_2^T\ddot{\mathbf{q}}_2 &= -\dot{\mathbf{q}}_2^T\dot{\mathbf{q}}_2 \\
\mathbf{q}_2^T\mathbf{q}_2 &= L^2.
\end{aligned} \tag{18.A.13}$$

Introduce scaled force vectors by dividing the applied forces by m and mL

$$\mathbf{g}_1 \triangleq (\hat{\mathbf{f}}_1 + \hat{\mathbf{f}}_2)\frac{2}{m}, \quad \mathbf{g}_2 \triangleq (\hat{\mathbf{f}}_1 - \hat{\mathbf{f}}_2)\frac{6}{mL^2}.$$

Then, (18.A.13) can be rewritten as

$$\begin{aligned}
\ddot{\mathbf{q}}_1 &= \mathbf{g}_1 \\
\tilde{\mathbf{q}}_2\ddot{\mathbf{q}}_2 &= \tilde{\mathbf{q}}_2(-\mathbf{g}_2 L^2) \\
\mathbf{q}_2^T\ddot{\mathbf{q}}_2 &= -\dot{\mathbf{q}}_2^T\dot{\mathbf{q}}_2 \\
\mathbf{q}_2^T\mathbf{q}_2 &= L^2.
\end{aligned} \tag{18.A.14}$$

Solving for $\ddot{\mathbf{q}}_2$ requires,

$$\begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}\ddot{\mathbf{q}}_2 = \begin{bmatrix} \mathbf{0} \\ -\dot{\mathbf{q}}_2^T\dot{\mathbf{q}}_2 \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{0} \end{bmatrix}\mathbf{g}_2 L^2 \tag{18.A.15}$$

Lemma *For any vector* $\mathbf{q}$, *such that* $\mathbf{q}^T\mathbf{q} = L^2$,

$$\begin{bmatrix} \tilde{\mathbf{q}} \\ \mathbf{q}^T \end{bmatrix}^T \begin{bmatrix} \tilde{\mathbf{q}} \\ \mathbf{q}^T \end{bmatrix} = L_2\mathbf{I}_3$$

Proof:

$$\begin{bmatrix} 0 & q_3 & -q_2 & q_1 \\ -q_3 & 0 & q_1 & q_2 \\ q_2 & -q_1 & 0 & q_3 \end{bmatrix} \begin{bmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \\ q_1 & q_2 & q_3 \end{bmatrix} = L^2\mathbf{I}_3 \qquad \Delta.$$

Since the coefficient of $\ddot{\mathbf{q}}_2$ in (18.A.15) has linearly independent columns by virtue of the Lemma, the unique solution for $\ddot{\mathbf{q}}_2$ is

© 2002 by CRC Press LLC

$$\ddot{\mathbf{q}}_2 = \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}^+ \left(\begin{bmatrix} \mathbf{0} \\ -\dot{\mathbf{q}}_2^T \dot{\mathbf{q}}_2 \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{0} \end{bmatrix} \mathbf{g}_2 L^2 \right), \tag{18.A.16}$$

where the pseudo inverse is uniquely given by

$$\begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}^+ = \left(\begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}^T \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix} \right)^{-1} \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}^T = L^{-2} \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}^T$$

It is easily verified that the existence condition for $\ddot{\mathbf{q}}_2$ in (18.A.15) is satisfied since

$$\left(\mathbf{I} - \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{q}_2^T \end{bmatrix}^+ \right) \left(\begin{bmatrix} \mathbf{0} \\ -\dot{\mathbf{q}}_2^T \dot{\mathbf{q}}_2 \end{bmatrix} - \begin{bmatrix} \tilde{\mathbf{q}}_2 \\ \mathbf{0} \end{bmatrix} \mathbf{g}_2 L^2 \right) = 0,$$

Hence, (18.A.16) yields

$$\ddot{\mathbf{q}}_2 = -\frac{\dot{\mathbf{q}}_2^T \dot{\mathbf{q}}_2}{L^2} \mathbf{q}_2 + \tilde{\mathbf{q}}_2^2 \mathbf{g}_2 \tag{18.A.17}$$

Bringing the first equation of (18.A.14) together with (18.A.17) leads to

$$\begin{bmatrix} \ddot{\mathbf{q}}_1 \\ \ddot{\mathbf{q}}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\dot{\mathbf{q}}_2^T \dot{\mathbf{q}}_2}{L^2} \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{q}}_2^2 \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \tag{18.A.18}$$

Recalling the definition of $\mathbf{g}_1$ and $\mathbf{g}_2$ we obtain

$$\begin{bmatrix} \ddot{\mathbf{q}}_1 \\ \ddot{\mathbf{q}}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\dot{\mathbf{q}}_2^T \dot{\mathbf{q}}_2}{L^2} \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} = \frac{2}{m} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \frac{3}{L^2} \tilde{\mathbf{q}}_2^2 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} \tag{18.A.19}$$

where we clarify

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{f}}_1 + \hat{\mathbf{f}}_2 \\ \hat{\mathbf{f}}_1 - \hat{\mathbf{f}}_2 \end{bmatrix}.$$

Equation (18.A.19) is identical to (18.17), so this completes the proof of Theorem 18.1.

Example 1

$$\mathbf{q}_1 = \mathbf{p}_1 + \mathbf{p}_2 = \begin{bmatrix} p_{11} + p_{21} \\ p_{12} + p_{22} \end{bmatrix} = \begin{bmatrix} q_{11} \\ q_{12} \end{bmatrix},$$

$$\mathbf{q}_2 = \mathbf{p}_2 - \mathbf{p}_1 = \begin{bmatrix} p_{21} - p_{11} \\ p_{22} - p_{12} \end{bmatrix} = \begin{bmatrix} q_{21} \\ q_{22} \end{bmatrix},$$

The generalized forces are now defined as

© 2002 by CRC Press LLC

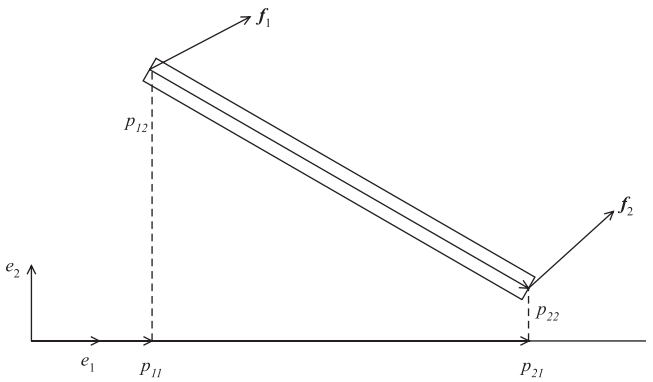


FIGURE 18.A.1 A rigid bar of the length L and mass m.

$$\mathbf{g}_1 = \frac{2}{m}(\hat{\mathbf{f}}_1 + \hat{\mathbf{f}}_2) = \frac{2}{m}\begin{bmatrix} f_{11} + f_{21} \\ f_{12} + f_{22} \end{bmatrix},\ \mathbf{g}_2 = \frac{6}{mL^2}(\hat{\mathbf{f}}_2 - \hat{\mathbf{f}}_1) = \frac{6}{mL^2}\begin{bmatrix} f_{21} - f_{11} \\ f_{22} - f_{12} \end{bmatrix}$$

From (18.A.16),

$$\begin{bmatrix} \ddot{q}_{11} \\ \ddot{q}_{12} \\ \ddot{q}_{21} \\ \ddot{q}_{22} \end{bmatrix} = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -\frac{\dot{q}_{21}^2 + \dot{q}_{22}^2}{L^2} & 0 \\ 0 & 0 & 0 & -\frac{\dot{q}_{21}^2 + \dot{q}_{22}^2}{L^2} \end{array}\right] \begin{bmatrix} q_{11} \\ q_{12} \\ q_{21} \\ q_{22} \end{bmatrix} + \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & q_{22}^2 & -q_{21}q_{22} \\ 0 & 0 & -q_{21}q_{22} & q_{21}^2 \end{array}\right] \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$$

Example 2

Using the formulation developed in 18.A.3, 18.A.4, and 18.A.5 we derive the dynamics of a planar tensegrity. The rules of closure become:

$$\mathbf{t}_5 = -\mathbf{t}_4$$

$$\mathbf{t}_8 = \mathbf{t}_1$$

$$\mathbf{t}_7 = -\mathbf{t}_2$$

$$\mathbf{t}_6 = -\mathbf{t}_3$$

We define the independent vectors l^o and l^d:

$$\mathbf{l}^o = \begin{bmatrix} \rho \\ \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{t}_5 \end{bmatrix},\ \mathbf{l}^d = \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \end{bmatrix}$$

The nodal forces are

$$\bar{\mathbf{f}} = \begin{bmatrix} \hat{\mathbf{f}}_1 + \hat{\mathbf{f}}_2 \\ \hat{\mathbf{f}}_1 - \hat{\mathbf{f}}_2 \\ \hat{\mathbf{f}}_3 + \hat{\mathbf{f}}_4 \\ \hat{\mathbf{f}}_3 - \hat{\mathbf{f}}_4 \end{bmatrix} = \begin{bmatrix} (\mathbf{f}_3 - \mathbf{f}_2 + \mathbf{w}_1) + (\mathbf{f}_5 - \mathbf{f}_1 + \mathbf{w}_2) \\ (\mathbf{f}_3 - \mathbf{f}_2 + \mathbf{w}_1) - (\mathbf{f}_5 - \mathbf{f}_1 + \mathbf{w}_2) \\ (\mathbf{f}_1 + \mathbf{f}_2 + \mathbf{w}_3) + (-\mathbf{f}_3 - \mathbf{f}_5 + \mathbf{w}_4) \\ (\mathbf{f}_1 + \mathbf{f}_2 + \mathbf{w}_3) - (-\mathbf{f}_3 - \mathbf{f}_5 + \mathbf{w}_4) \end{bmatrix}$$

© 2002 by CRC Press LLC

We can write

$$\bar{\mathbf{f}} = \begin{bmatrix} \mathbf{I}_2 \\ -\mathbf{I}_2 \\ -\mathbf{I}_2 \\ \mathbf{I}_2 \end{bmatrix} \mathbf{f}^o + \begin{bmatrix} -\mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{I}_2 \\ \mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{I}_2 \\ \mathbf{I}_2 & \mathbf{I}_2 & -\mathbf{I}_2 \\ \mathbf{I}_2 & \mathbf{I}_2 & \mathbf{I}_2 \end{bmatrix} \mathbf{f}^d + \begin{bmatrix} \mathbf{I}_2 & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & -\mathbf{I}_2 \end{bmatrix} \mathbf{w}$$

where

$$\mathbf{f}^o = \begin{bmatrix} \mathbf{f}_5 \end{bmatrix}, \ \mathbf{f}^d = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \\ \mathbf{w}_4 \end{bmatrix}.$$

Or, with the obvious definitions for $\mathbf{B}^o$, $\mathbf{B}^d$, and $\mathbf{W}_1$, in matrix notation:

$$\bar{\mathbf{f}} = \mathbf{B}^o \ \mathbf{f}^o + \mathbf{B}^d \ \mathbf{f}^d + \mathbf{W}_1 \mathbf{w}. \tag{18.A.20}$$

The nodal vectors are defined as follows:

$$\begin{cases} \mathbf{p}_1 = \boldsymbol{\rho} \\ \mathbf{p}_2 = \boldsymbol{\rho} + \mathbf{r}_1 \\ \mathbf{p}_3 = \hat{\boldsymbol{\rho}} \\ \mathbf{p}_4 = \hat{\boldsymbol{\rho}} + \mathbf{r}_2 \end{cases},$$

and

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\rho} + \mathbf{r}_1 + \mathbf{t}_5 - \mathbf{r}_2.$$

We define

$$\begin{cases} \mathbf{q}_1 \triangleq \mathbf{p}_2 + \mathbf{p}_1 = 2\boldsymbol{\rho} + \mathbf{r}_1 \\ \mathbf{q}_2 \triangleq \mathbf{p}_2 - \mathbf{p}_1 = \mathbf{r}_1 \\ \mathbf{q}_3 \triangleq \mathbf{p}_4 + \mathbf{p}_3 = 2\hat{\boldsymbol{\rho}} + \mathbf{r}_2 = 2(\boldsymbol{\rho} + \mathbf{r}_1 + \mathbf{t}_5) - \mathbf{r}_2 \\ \mathbf{q}_4 \triangleq \mathbf{p}_4 - \mathbf{p}_3 = \mathbf{r}_2 \end{cases}$$

The relation between $\mathbf{q}$ and $\mathbf{p}$ can be written as follows:

$$\mathbf{q} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_2 & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_2 & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ \mathbf{p}_4 \end{bmatrix},$$

© 2002 by CRC Press LLC

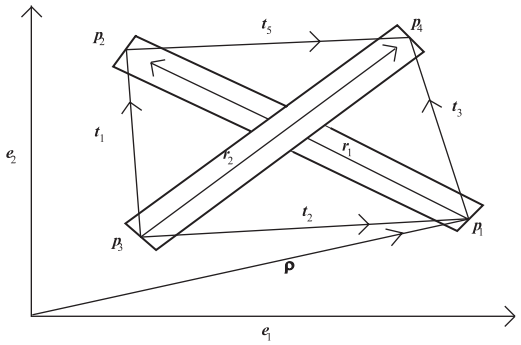


FIGURE 18.A.2 A planar tensegrity.

and

$$\mathbf{q} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \end{bmatrix} = \begin{bmatrix} 2\mathbf{I}_2 & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ 2\mathbf{I}_2 & 2\mathbf{I}_2 & -\mathbf{I}_2 & 2\mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{t}_5 \end{bmatrix} = \mathbf{Q}\mathbf{l}^o. \tag{18.A.21}$$

We can now write the dependent variables $\mathbf{l}^d$ in terms of independent variables $\mathbf{l}^o$. From (18.30) and (18.31):

$$\mathbf{t}_1 = \boldsymbol{\rho} + \mathbf{r}_1 - (\boldsymbol{\rho} + \mathbf{r}_1 + \mathbf{t}_5 - \mathbf{r}_2),$$

$$\mathbf{t}_2 = \boldsymbol{\rho} - (\boldsymbol{\rho} + \mathbf{r}_1 + \mathbf{t}_5 - \mathbf{r}_2).$$

By inspection of Figure 18.A.2, (18.30) and (18.31) reduce to:

$$\begin{aligned} \mathbf{t}_1 &= \mathbf{r}_2 - \mathbf{t}_5 \\ \mathbf{t}_2 &= -\mathbf{r}_1 + \mathbf{r}_2 - \mathbf{t}_5 \\ \mathbf{t}_3 &= \mathbf{r}_1 + \mathbf{t}_5, \end{aligned} \tag{18.A.22}$$

or,

$$\mathbf{I}^d = \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & -\mathbf{I}_2 \\ \mathbf{0} & -\mathbf{I}_2 & \mathbf{I}_2 & -\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{t}_5 \end{bmatrix}$$

Equation (18.A.21) yields

© 2002 by CRC Press LLC

$$\mathbf{l}^o = \begin{bmatrix} \boldsymbol{\rho} \\ \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{t}_5 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 2\mathbf{I}_2 \\ -\mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{I}_2 & \mathbf{I}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \end{bmatrix} = \mathbf{Q}^{-1}\mathbf{q}.$$

Hence,

$$\mathbf{l}^d = \frac{1}{2} \begin{bmatrix} \mathbf{I}_2 & \mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{I}_2 \\ \mathbf{I}_2 & -\mathbf{I}_2 & -\mathbf{I}_2 & \mathbf{I}_2 \\ -\mathbf{I}_2 & \mathbf{I}_2 & \mathbf{I}_2 & \mathbf{I}_2 \end{bmatrix} \mathbf{q} = \mathbf{R}\mathbf{q}$$

We can now write out the tendon forces as follows:

$$\mathbf{f}^d = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix} = -\begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \\ \mathbf{K}_3 \end{bmatrix} \mathbf{q} + \begin{bmatrix} \mathbf{b}_1 & & \\ & \mathbf{b}_2 & \\ & & \mathbf{b}_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix},$$

or

$$\mathbf{f}^d = -\mathbf{K}^d\mathbf{q} + \mathbf{P}^d\mathbf{u}^d$$

and

$$\mathbf{f}^o = [\mathbf{f}_5] = -[\mathbf{K}_5]\ \mathbf{q} + [\mathbf{b}_5]\ [u_5],$$

or

$$\mathbf{f}^o = -\mathbf{K}^o\mathbf{q} + \mathbf{P}^o\mathbf{u}^o,$$

using the same definitions for **K** and **b** as found in (18.45) and (18.46), simply by removing the *ij* element indices. Substitution into (18.A.20) yields:

$$\bar{\mathbf{f}} = \mathbf{B}^o(-\mathbf{K}^o\mathbf{q} + \mathbf{P}^o\mathbf{u}^o) + \mathbf{B}^d(-\mathbf{K}^d\mathbf{q} + \mathbf{P}^d\mathbf{u}^d)$$

$$= -(\mathbf{B}^o\mathbf{K}^o + \mathbf{B}^d\mathbf{K}^d)\mathbf{q} + (\mathbf{B}^o\mathbf{P}^o\mathbf{u}^o + \mathbf{B}^d\mathbf{P}^d\mathbf{u}^d).$$

With the matrices derived in this section, we can express the dynamics in the form of (18.20):

$$\ddot{\mathbf{q}} + (\mathbf{K}_r + \mathbf{K}_p)\mathbf{q} = \mathbf{B}\mathbf{u} + \mathbf{D}\mathbf{w},$$

$$\mathbf{K}_r = \boldsymbol{\Omega}_1,$$

$$\mathbf{K}_p = \mathbf{H}\tilde{\mathbf{K}},$$

$$\mathbf{B} = \mathbf{H}\tilde{\mathbf{B}},$$

© 2002 by CRC Press LLC

$$\mathbf{D} = \mathbf{H}\mathbf{W}^o = \mathbf{H}_1\mathbf{W}_1,$$

where

$$\mathbf{H} = \mathbf{H}_1 = \frac{2}{m}\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{3}{L_1^2}\tilde{\mathbf{q}}_2^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{3}{L_2^2}\tilde{\mathbf{q}}_2^2 \end{bmatrix},$$

$$\mathbf{K}_r = \boldsymbol{\Omega}_1\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & L_1^{-2}\dot{\mathbf{q}}_2^T\dot{\mathbf{q}}_2\mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & L_2^{-2}\dot{\mathbf{q}}_4^T\dot{\mathbf{q}}_4\mathbf{I}_4 \end{bmatrix},$$

$$\tilde{\mathbf{K}} \underline{\underline{\Delta}} \mathbf{B}^d\mathbf{K}^d + \mathbf{B}^o\mathbf{K}^o,$$

$$\tilde{\mathbf{B}} \underline{\underline{\Delta}} [\mathbf{B}^d\mathbf{P}^d, \mathbf{B}^o\mathbf{P}^o],$$

and

$$\tilde{\mathbf{q}}_2^2 = \begin{bmatrix} -q_{22}^2 & q_{21}q_{22} \\ q_{21}q_{22} & -q_{21}^2 \end{bmatrix},\ \tilde{\mathbf{q}}_4^2 = \begin{bmatrix} -q_{42}^2 & q_{41}q_{42} \\ q_{41}q_{42} & -q_{41}^2 \end{bmatrix},$$

$$\dot{\mathbf{q}}_2^T\dot{\mathbf{q}}_2 = \dot{q}_{21}^2 + \dot{q}_{22}^2,\quad \dot{\mathbf{q}}_4^T\dot{\mathbf{q}}_4 = \dot{q}_{41}^2 + \dot{q}_{42}^2.$$

© 2002 by CRC Press LLC

Appendix 18.B Algebraic Inversion of the Q Matrix

This appendix will algebraically invert a 5×5 block **Q** matrix. Given **Q** in the form:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{21} & \mathbf{Q}_{32} & \mathbf{Q}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{21} & \mathbf{Q}_{32} & \mathbf{Q}_{32} & \mathbf{Q}_{22} & \mathbf{0} \\ \mathbf{Q}_{21} & \mathbf{Q}_{32} & \mathbf{Q}_{32} & \mathbf{Q}_{32} & \mathbf{Q}_{22} \end{bmatrix} \tag{18.B.1}$$

we define **x** and **y** matrices so that

$$\mathbf{Qx} = \mathbf{y} \tag{18.B.2}$$

where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \end{bmatrix} \tag{18.B.3}$$

Solving (18.B.2) for **x** we obtain

$$\mathbf{x} = \mathbf{Q}^{-1}\mathbf{y} \tag{18.B.4}$$

Substituting (18.B.1) and (18.B.3) into (18.B.2) and carrying out the matrix operations we obtain

$$\begin{cases} \mathbf{Q}_{11}\mathbf{x}_1 & = \mathbf{y}_1 \\ \mathbf{Q}_{21}\mathbf{x}_1 + \mathbf{Q}_{22}\mathbf{x}_2 & = \mathbf{y}_2 \\ \mathbf{Q}_{21}\mathbf{x}_1 + \mathbf{Q}_{32}\mathbf{x}_2 + \mathbf{Q}_{22}\mathbf{x}_3 & = \mathbf{y}_3 \\ \mathbf{Q}_{21}\mathbf{x}_1 + \mathbf{Q}_{32}\mathbf{x}_2 + \mathbf{Q}_{32}\mathbf{x}_3 + \mathbf{Q}_{22}\mathbf{x}_4 & = \mathbf{y}_4 \\ \mathbf{Q}_{21}\mathbf{x}_1 + \mathbf{Q}_{32}\mathbf{x}_2 + \mathbf{Q}_{32}\mathbf{x}_3 + \mathbf{Q}_{32}\mathbf{x}_4 + \mathbf{Q}_{22}\mathbf{x}_5 & = \mathbf{y}_5 \end{cases} \tag{18.B.5}$$

Solving this system of equations for **x** will give us the desired $\mathbf{Q}^{-1}$ matrix. Solving each equation for **x** we have

$$\begin{cases} \mathbf{x}_1 = \mathbf{Q}_{11}^{-1}\mathbf{y}_1 \\ \mathbf{x}_2 = \mathbf{Q}_{22}^{-1}(-\mathbf{Q}_{21}\mathbf{x}_1 + \mathbf{y}_2) \\ \mathbf{x}_3 = \mathbf{Q}_{22}^{-1}(-\mathbf{Q}_{21}\mathbf{x}_1 - \mathbf{Q}_{32}\mathbf{x}_2 + \mathbf{y}_3) \\ \mathbf{x}_4 = \mathbf{Q}_{22}^{-1}(-\mathbf{Q}_{21}\mathbf{x}_1 - \mathbf{Q}_{32}\mathbf{x}_2 - \mathbf{Q}_{32}\mathbf{x}_3 + \mathbf{y}_4) \\ \mathbf{x}_5 = \mathbf{Q}_{22}^{-1}(-\mathbf{Q}_{21}\mathbf{x}_1 - \mathbf{Q}_{32}\mathbf{x}_2 - \mathbf{Q}_{32}\mathbf{x}_3 - \mathbf{Q}_{32}\mathbf{x}_4 + \mathbf{y}_5) \end{cases} \tag{18.B.6}$$

Elimination of **x** on the right side of (18.B.6) by substitution yields

© 2002 by CRC Press LLC

$$\begin{cases} \mathbf{x}_1 &= \mathbf{\Lambda}_{11}\mathbf{y}_1 \\ \mathbf{x}_2 &= \mathbf{\Lambda}_{21}\mathbf{y}_1 + \mathbf{\Lambda}_{22}\mathbf{y}_2 \\ \mathbf{x}_3 &= \mathbf{\Lambda}_{31}\mathbf{y}_1 + \mathbf{\Lambda}_{32}\mathbf{y}_2 + \mathbf{\Lambda}_{33}\mathbf{y}_3 \\ \mathbf{x}_4 &= \mathbf{\Lambda}_{41}\mathbf{y}_1 + \mathbf{\Lambda}_{42}\mathbf{y}_2 + \mathbf{\Lambda}_{43}\mathbf{y}_3 + \mathbf{\Lambda}_{44}\mathbf{y}_4 \\ \mathbf{x}_5 &= \mathbf{\Lambda}_{51}\mathbf{y}_1 + \mathbf{\Lambda}_{52}\mathbf{y}_2 + \mathbf{\Lambda}_{53}\mathbf{y}_3 + \mathbf{\Lambda}_{54}\mathbf{y}_4 + \mathbf{\Lambda}_{55}\mathbf{y}_5 \end{cases} \tag{18.B.7}$$

Or, in matrix form,

$$\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{\Lambda}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_{31} & \mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_{41} & \mathbf{\Lambda}_{42} & \mathbf{\Lambda}_{43} & \mathbf{\Lambda}_{44} & \mathbf{0} \\ \mathbf{\Lambda}_{51} & \mathbf{\Lambda}_{52} & \mathbf{\Lambda}_{53} & \mathbf{\Lambda}_{54} & \mathbf{\Lambda}_{55} \end{bmatrix} \tag{18.B.8}$$

where we have used the notation $\Lambda_{row,\,column}$ with the following definitions

$$\mathbf{\Lambda}_{11} = \mathbf{Q}_{11}^{-1}$$

$$\mathbf{\Lambda}_{21} = -\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} = -\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{\Lambda}_{11} = -\mathbf{\Lambda}_{22}\mathbf{Q}_{21}\mathbf{\Lambda}_{11}$$

$$\mathbf{\Lambda}_{31} = \mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} - \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}$$

$$\mathbf{\Lambda}_{41} = -\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} + 2\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} - \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}$$

$$\begin{aligned} \mathbf{\Lambda}_{51} = &-3\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} + 3\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} \\ &+ \mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} - \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{11}^{-1} \end{aligned}$$

$$\mathbf{\Lambda}_{22} = \mathbf{\Lambda}_{33} = \mathbf{\Lambda}_{44} = \mathbf{\Lambda}_{55} = \mathbf{Q}_{22}^{-1}$$

$$\mathbf{\Lambda}_{32} = \mathbf{\Lambda}_{43} = \mathbf{\Lambda}_{54} = -\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}$$

$$\mathbf{\Lambda}_{42} = \mathbf{\Lambda}_{53} = \mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1} - \mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}$$

$$\mathbf{\Lambda}_{52} = -\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1} + 2\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1} - \mathbf{Q}_{22}^{-1}\mathbf{Q}_{32}\mathbf{Q}_{22}^{-1}$$

Using repeated patterns, the inverse may be computed by

$$\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{\Lambda}_{11} & & & & & & \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} & & & & & \\ \mathbf{\Theta}\mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{22} & & & & \\ \mathbf{\Theta}^2\mathbf{\Lambda}_{21} & \mathbf{\Theta}\mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{22} & & & \\ \mathbf{\Theta}^3\mathbf{\Lambda}_{21} & \mathbf{\Theta}^2\mathbf{\Lambda}_{31} & \mathbf{\Theta}\mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{22} & & \\ \mathbf{\Theta}^4\mathbf{\Lambda}_{21} & \mathbf{\Theta}^3\mathbf{\Lambda}_{31} & \mathbf{\Theta}^2\mathbf{\Lambda}_{31} & \mathbf{\Theta}\mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{32} & \mathbf{\Lambda}_{22} & \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{18.B.9}$$

© 2002 by CRC Press LLC

using

$$\boldsymbol{\Theta} = (\mathbf{I} - \boldsymbol{\Lambda}_{22}\mathbf{Q}_{32})$$

$$\boldsymbol{\Lambda}_{11} = \mathbf{Q}_{11}^{-1}$$

$$\boldsymbol{\Lambda}_{21} = -\boldsymbol{\Lambda}_{22}\mathbf{Q}_{21}\boldsymbol{\Lambda}_{11}$$

$$\boldsymbol{\Lambda}_{22} = \mathbf{Q}_{22}^{-1}$$

$$\boldsymbol{\Lambda}_{32} = -\boldsymbol{\Lambda}_{22}\mathbf{Q}_{32}\boldsymbol{\Lambda}_{22}.$$

Only $\boldsymbol{\Lambda}_{11}, \boldsymbol{\Lambda}_{22}, \boldsymbol{\Lambda}_{21}, \boldsymbol{\Lambda}_{32},$ and powers of Θ need to be calculated to obtain $\mathbf{Q}^{-1}$ for any (n, m). The only matrix inversion that needs to be computed to obtain $\mathbf{Q}^{-1}$ is $\mathbf{Q}_{11}^{-1}$ and $\mathbf{Q}_{22}^{-1}$, substantially reducing computer processing time for computer simulations.

© 2002 by CRC Press LLC

Appendix 18.C General Case for (n, m) = (i, 1)

In Appendix E the forces acting on each node are presented, making special exceptions for the case when ($j = m = 1$). The exceptions arise for ($j = m = 1$) because one stage now contains both closure rules for the base and the top of the structure. In the following synthesis we use $\hat{\mathbf{f}}_{111}$ and $\hat{\mathbf{f}}_{211}$ from (18.E.1) and $\hat{\mathbf{f}}_{311}$ and $\hat{\mathbf{f}}_{411}$ from (18.E.4). At the right, where ($i = 1$, $j = 1$):

$$\bar{\mathbf{f}}_{11} = \begin{bmatrix} \hat{\mathbf{f}}_{111} + \hat{\mathbf{f}}_{211} \\ \hat{\mathbf{f}}_{111} - \hat{\mathbf{f}}_{211} \\ \hat{\mathbf{f}}_{311} + \hat{\mathbf{f}}_{411} \\ \hat{\mathbf{f}}_{311} - \hat{\mathbf{f}}_{411} \end{bmatrix}$$

$$= \begin{bmatrix} (-\mathbf{f}_{211} + \mathbf{f}_{311} + \mathbf{f}_{1n1} + \mathbf{f}_{2n1} + \mathbf{w}_{111}) + (\mathbf{f}_{511} - \mathbf{f}_{111} - \mathbf{f}_{411} - \mathbf{f}_{811} + \mathbf{w}_{211}) \\ (-\mathbf{f}_{211} + \mathbf{f}_{311} + \mathbf{f}_{1n1} + \mathbf{f}_{2n1} + \mathbf{w}_{111}) - (\mathbf{f}_{511} - \mathbf{f}_{111} - \mathbf{f}_{411} - \mathbf{f}_{811} + \mathbf{w}_{211}) \\ (\mathbf{f}_{811} - \mathbf{f}_{711} - \mathbf{f}_{321} + \mathbf{f}_{421} + \mathbf{w}_{311}) + (\mathbf{f}_{611} - \mathbf{f}_{511} - \mathbf{f}_{6n1} + \mathbf{f}_{7n1} + \mathbf{w}_{411}) \\ (\mathbf{f}_{811} - \mathbf{f}_{711} - \mathbf{f}_{321} + \mathbf{f}_{421} + \mathbf{w}_{311}) - (\mathbf{f}_{611} - \mathbf{f}_{511} - \mathbf{f}_{6n1} + \mathbf{f}_{7n1} + \mathbf{w}_{411}) \end{bmatrix}.$$

At the center, where ($1 < i < n$, $j = 1$):

$$\bar{\mathbf{f}}_{i1} = \begin{bmatrix} \hat{\mathbf{f}}_{1i1} + \hat{\mathbf{f}}_{2i1} \\ \hat{\mathbf{f}}_{1i1} - \hat{\mathbf{f}}_{2i1} \\ \hat{\mathbf{f}}_{3i1} + \hat{\mathbf{f}}_{4i1} \\ \hat{\mathbf{f}}_{3i1} - \hat{\mathbf{f}}_{4i1} \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{f}_{1(i-1)1} + \mathbf{f}_{2(i-1)1} - \mathbf{f}_{2i1} + \mathbf{f}_{3i1} + \mathbf{w}_{1i1}) + (-\mathbf{f}_{1i1} - \mathbf{f}_{4i1} + \mathbf{f}_{5i1} - \mathbf{f}_{8i1} + \mathbf{w}_{2i1}) \\ (\mathbf{f}_{1(i-1)1} + \mathbf{f}_{2(i-1)1} - \mathbf{f}_{2i1} + \mathbf{f}_{3i1} + \mathbf{w}_{1i1}) - (-\mathbf{f}_{1i1} - \mathbf{f}_{4i1} + \mathbf{f}_{5i1} - \mathbf{f}_{8i1} + \mathbf{w}_{2i1}) \\ (\mathbf{f}_{8i1} - \mathbf{f}_{7i1} - \mathbf{f}_{3(i+1)1} + \mathbf{f}_{4(i+1)1} + \mathbf{w}_{3i1}) + (-\mathbf{f}_{5i1} + \mathbf{f}_{6i1} - \mathbf{f}_{6(i-1)} + \mathbf{f}_{7(i-1)1} + \mathbf{w}_{4i1}) \\ (\mathbf{f}_{8i1} - \mathbf{f}_{7i1} - \mathbf{f}_{3(i+1)1} + \mathbf{f}_{4(i+1)1} + \mathbf{w}_{3i1}) - (-\mathbf{f}_{5i1} + \mathbf{f}_{6i1} - \mathbf{f}_{6(i-1)} + \mathbf{f}_{7(i-1)1} + \mathbf{w}_{4i1}) \end{bmatrix}.$$

At the left end of the base in Figure 18.5, where ($i = n$, $j = 1$):

$$\bar{\mathbf{f}}_{n1} = \begin{bmatrix} \hat{\mathbf{f}}_{1n1} + \hat{\mathbf{f}}_{2n1} \\ \hat{\mathbf{f}}_{1n1} - \hat{\mathbf{f}}_{2n1} \\ \hat{\mathbf{f}}_{3n1} + \hat{\mathbf{f}}_{4n1} \\ \hat{\mathbf{f}}_{3n1} - \hat{\mathbf{f}}_{4n1} \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{f}_{1(n-1)1} + \mathbf{f}_{2(n-1)1} + \mathbf{f}_{3n1} - \mathbf{f}_{2n1} + \mathbf{w}_{1n1}) + (-\mathbf{f}_{1n1} - \mathbf{f}_{4n1} + \mathbf{f}_{5n1} - \mathbf{f}_{8n1} + \mathbf{w}_{2n1}) \\ (\mathbf{f}_{1(n-1)1} + \mathbf{f}_{2(n-1)1} + \mathbf{f}_{3n1} - \mathbf{f}_{2n1} + \mathbf{w}_{1n1}) - (-\mathbf{f}_{1n1} - \mathbf{f}_{4n1} + \mathbf{f}_{5n1} - \mathbf{f}_{8n1} + \mathbf{w}_{2n1}) \\ (-\mathbf{f}_{7n1} + \mathbf{f}_{8n1} - \mathbf{f}_{311} + \mathbf{f}_{411} + \mathbf{w}_{3n1}) + (-\mathbf{f}_{6(n-1)1} - \mathbf{f}_{5n1} + \mathbf{f}_{6n1} + \mathbf{f}_{7(n-1)1} + \mathbf{w}_{4n1}) \\ (-\mathbf{f}_{7n1} + \mathbf{f}_{8n1} - \mathbf{f}_{311} + \mathbf{f}_{411} + \mathbf{w}_{3n1}) - (-\mathbf{f}_{6(n-1)1} - \mathbf{f}_{5n1} + \mathbf{f}_{6n1} + \mathbf{f}_{7(n-1)1} + \mathbf{w}_{4n1}) \end{bmatrix}. \quad (18.C.1)$$

Using,

© 2002 by CRC Press LLC

$$\mathbf{f}_{ij}^{o}=\begin{bmatrix}\mathbf{f}_5\\ \mathbf{f}_1\end{bmatrix}_{ij},\mathbf{f}_{ij}^{d}=\begin{bmatrix}\mathbf{f}_2\\ \mathbf{f}_3\\ \mathbf{f}_4\\ \mathbf{f}_6\\ \mathbf{f}_7\\ \mathbf{f}_8\\ \mathbf{f}_9\\ \mathbf{f}_{10}\end{bmatrix}_{ij},\mathbf{w}_{ij}=\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\\ \mathbf{w}_3\\ \mathbf{w}_4\end{bmatrix}_{ij},$$

$$\bar{\mathbf{f}}_{11}=\begin{bmatrix}\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_3&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{n1}^{d}+\begin{bmatrix}-\mathbf{I}_3&\mathbf{I}_3&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ -\mathbf{I}_3&\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_3&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&-\mathbf{I}_3&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{11}^{d}$$

$$+\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{21}^{d}+\begin{bmatrix}\mathbf{I}_3&-\mathbf{I}_3\\ -\mathbf{I}_3&\mathbf{I}_3\\ -\mathbf{I}_3&\mathbf{0}\\ \mathbf{I}_3&\mathbf{0}\end{bmatrix}\mathbf{f}_{11}^{o}+\begin{bmatrix}\mathbf{I}_3\\ \mathbf{I}_3\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}\mathbf{f}_{1n1}$$

$$+\begin{bmatrix}\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{I}_3&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{I}_3&\mathbf{I}_3\\ \mathbf{0}&\mathbf{0}&\mathbf{I}_3&-\mathbf{I}_3\end{bmatrix}\mathbf{w}_{11}.$$

Or, in matrix notation, with the implied matrix definitions,

$$\bar{\mathbf{f}}_{11}=\hat{\mathbf{B}}_{n1}^{d}\mathbf{f}_{n1}^{d}+\mathbf{B}_{11}^{d}\mathbf{f}_{11}^{d}+\mathbf{B}_{21}^{d}\mathbf{f}_{21}^{d}+\mathbf{B}_{11}^{o}\mathbf{f}_{11}^{o}+\mathbf{B}_{1n1}\mathbf{f}_{1n1}+\mathbf{W}\mathbf{w}_{11}. \tag{18.C.2}$$

$$\bar{\mathbf{f}}_{i1}=\begin{bmatrix}\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_3&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{(i-1)1}^{d}+\begin{bmatrix}-\mathbf{I}_3&\mathbf{I}_3&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ -\mathbf{I}_3&\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_3&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&-\mathbf{I}_3&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{i1}^{d}$$

$$+\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&-\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{(i+1)1}^{d}+\begin{bmatrix}\mathbf{0}&\mathbf{I}_3\\ \mathbf{0}&\mathbf{I}_3\\ \mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}\end{bmatrix}\mathbf{f}_{(i-1)1}^{o}$$

$$+\begin{bmatrix}\mathbf{I}_3&-\mathbf{I}_3\\ -\mathbf{I}_3&\mathbf{I}_3\\ -\mathbf{I}_3&\mathbf{0}\\ \mathbf{I}_3&\mathbf{0}\end{bmatrix}\mathbf{f}_{i1}^{o}+\begin{bmatrix}\mathbf{I}_3&\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{I}_3&-\mathbf{I}_3&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{I}_3&\mathbf{I}_3\\ \mathbf{0}&\mathbf{0}&\mathbf{I}_3&-\mathbf{I}_3\end{bmatrix}\mathbf{w}_{i1}.$$

© 2002 by CRC Press LLC

Or, in matrix form,

$$\bar{\mathbf{f}}_{i1} = \hat{\mathbf{B}}^d_{01}\mathbf{f}^d_{(i-1)1} + \mathbf{B}^d_{11}\mathbf{f}^d_{i1} + \mathbf{B}^d_{21}\mathbf{f}^d_{(i+1)1} + \mathbf{B}^o_{01}\mathbf{f}^o_{(i-1)1} + \mathbf{B}^o_{11}\mathbf{f}^o_{i1} + \mathbf{W}\mathbf{w}_{i1}. \tag{18.C.3}$$

$$\begin{aligned}
\bar{\mathbf{f}}_{n1} = {} & \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(n-1)1} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n1} \\
& + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{11} + \begin{bmatrix} \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{(n-1)1} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{nl} \\
& + \begin{bmatrix} -\mathbf{I}_3 \\ \mathbf{I}_3 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{f}_{1n1} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{n1}.
\end{aligned}$$

Or, in matrix form,

$$\bar{\mathbf{f}}_{n1} = \hat{\mathbf{B}}^d_{01}\mathbf{f}^d_{(n-1)1} + \mathbf{B}^d_{11}\mathbf{f}^d_{n1} + \mathbf{B}^d_{21}\mathbf{f}^d_{11} + \mathbf{B}^o_{01}\mathbf{f}^o_{(n-1)1} + \mathbf{B}^o_{n1}\mathbf{f}^o_{n1} + \overline{\mathbf{B}}_{1n1}\mathbf{f}_{1n1} + \mathbf{W}\mathbf{w}_{n1}. \tag{18.C.4}$$

Now assemble (18.C.2)–(18.C.4) into the form

$$\mathbf{f}_1 = \mathbf{B}_3\bar{\mathbf{f}}^d_1 + \mathbf{B}_1\mathbf{f}^o_1 + \mathbf{W}_1\mathbf{w}_1, \tag{18.C.5}$$

where

$$\mathbf{f}_j = \begin{bmatrix} \bar{\mathbf{f}}_{1j} \\ \bar{\mathbf{f}}_{2j} \\ \vdots \\ \bar{\mathbf{f}}_{nj} \end{bmatrix}, \ \bar{\mathbf{f}}^d_1 = \begin{bmatrix} \mathbf{f}_{1n1} \\ \mathbf{f}^d_1 \end{bmatrix}, \ \mathbf{f}^o_j = \begin{bmatrix} \mathbf{f}^o_{1j} \\ \mathbf{f}^o_{2j} \\ \vdots \\ \mathbf{f}^o_{nj} \end{bmatrix}, \ \mathbf{w}_j = \begin{bmatrix} \mathbf{w}_{1j} \\ \mathbf{w}_{2j} \\ \vdots \\ \mathbf{w}_{nj} \end{bmatrix}$$

$$\mathbf{B}_3 = \begin{bmatrix} \mathbf{B}_{1n1} & \mathbf{B}^d_{11} & \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \hat{\mathbf{B}}^d_{n1} \\ \mathbf{0} & \hat{\mathbf{B}}^d_{01} & \ddots & \ddots & \ddots & & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & & \ddots & \ddots & \ddots & \mathbf{B}^d_{21} \\ \overline{\mathbf{B}}_{1n1} & \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \hat{\mathbf{B}}^d_{01} & \mathbf{B}^d_{11} \end{bmatrix}, \ \mathbf{B}_1 = \begin{bmatrix} \mathbf{B}^o_{11} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{B}^o_{01} & \ddots & \ddots & & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{B}^o_{11} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^o_{01} & \mathbf{B}^o_{n1} \end{bmatrix},$$

$$\mathbf{W}_1 = \mathit{BlockDiag}[\ldots, \mathbf{W}, \mathbf{W}, \ldots].$$

Or, simply, (18.C.5) has the form (18.E.21), where

© 2002 by CRC Press LLC

$$\mathbf{f} = \mathbf{f}_1, \mathbf{f}^d = \bar{\mathbf{f}}_1^d, \mathbf{f}^o = \mathbf{f}_1^o, \mathbf{w} = \mathbf{w}_1, \mathbf{W}^o = \mathbf{W}_1,$$

$$\mathbf{B}^d = \mathbf{B}_3, \mathbf{B}^o = \mathbf{B}_1.$$

The next set of necessary exceptions that apply to the model (i,1) arises in the form of the **R** matrix that relates the dependent tendons set to the generalized coordinates ($\mathbf{l}^d = \mathbf{Rq}$). For any (i,1) case **R** takes the form following the same procedure as in (18.32) and (18.33).

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 \\ \tilde{\mathbf{R}}_{11} \end{bmatrix}, \tag{18.C.6}$$

where

$$\tilde{\mathbf{R}}_{11} = \begin{bmatrix} \mathbf{E}_3 & \tilde{\mathbf{E}}_4 & \mathbf{0} & \cdots & \cdots & \mathbf{E}_2 \\ \mathbf{E}_2 & \mathbf{E}_3 & \tilde{\mathbf{E}}_4 & \ddots & & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_2 & \mathbf{E}_3 & \tilde{\mathbf{E}}_4 & \ddots & \vdots \\ \vdots & \ddots & \mathbf{E}_2 & \mathbf{E}_3 & \tilde{\mathbf{E}}_4 & \mathbf{0} \\ \mathbf{0} & & \ddots & \ddots & \ddots & \tilde{\mathbf{E}}_4 \\ \tilde{\mathbf{E}}_4 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{E}_2 & \mathbf{E}_3 \end{bmatrix}$$

$$\tilde{\mathbf{E}}_4 = \frac{1}{2}\begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix}.$$

The transformation matrix **T** that is applied to the control inputs takes the following form. The only exception to (18.55) is that there are no $\mathbf{I}_8$ blocks due to the fact that there are no stages between the boundary conditions at the base and the top of the structure. The second set of $\mathbf{I}_2$ blocks is also not needed since $m = 1$. Hence, the appropriate **T** matrix for $\tilde{u} = \mathbf{Tu}$ is

$$\mathbf{T} = \left[\begin{array}{ccc|ccc|c} 0 & \cdots & & & & & 01 \\ \hline \mathbf{T}_1' & & & \mathbf{S} & & & \\ & \ddots & & & \ddots & & \\ & & \mathbf{T}_1' & & & \mathbf{S} & \\ & & & \mathbf{I}_2 & & & \\ & & & & \ddots & & \\ & & & & & \mathbf{I}_2 & \end{array}\right] \in \boldsymbol{R}^{(10n+1)\times 8n},$$

where **S** is defined by (18.56), and

© 2002 by CRC Press LLC

$$\mathbf{T}_1' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \in \boldsymbol{R}^{8\times 6}.$$

There are n $\mathbf{T}_1'$ blocks, n $\mathbf{S}$ blocks, and n $\mathbf{I}_2$ blocks. The control inputs are now defined as

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1^d \\ \mathbf{u}_1^o \end{bmatrix}.$$

The $\mathbf{u}_1^d$ block becomes

$$\mathbf{u}_1'^d = \begin{bmatrix} \mathbf{u}_{11}'^d \\ \mathbf{u}_{21}'^d \\ \mathbf{u}_{31}'^d \\ \vdots \\ \mathbf{u}_{n1}'^d \end{bmatrix}, \mathbf{u}_{i1}'^d = \begin{bmatrix} u_{2n1} \\ u_{3n1} \\ u_{4n1} \\ u_{6n1} \\ u_{7n1} \\ u_{8n1} \end{bmatrix} \in \boldsymbol{R}^{6\times 1},\ i = 1,\ 2,\ 3, \cdots n,\ j = m = 1$$

Appendix 18.D will explicitly show all matrix forms for the specific example $(n,m) = (3,1)$.

© 2002 by CRC Press LLC

Appendix 18.D Example Case (n,m) = (3,1)

Given the equation for the dynamics of the shell class of tensegrity structures:

$$\ddot{\mathbf{q}} + (\mathbf{K}_r(\dot{\mathbf{q}}) + \mathbf{K}_p(\mathbf{q}))\mathbf{q} = \mathbf{B}(\mathbf{q})\mathbf{u} + \mathbf{D}(\mathbf{q})\mathbf{w}.$$

We explicitly write out the matrices that define the problem:

$$\mathbf{q} = \mathbf{q}_1 = \begin{bmatrix} \mathbf{q}_{11} \\ \mathbf{q}_{21} \\ \mathbf{q}_{31} \end{bmatrix} = \begin{bmatrix} \mathbf{q}_{111} \\ \mathbf{q}_{211} \\ \mathbf{q}_{311} \\ \underline{\mathbf{q}_{411}} \\ \mathbf{q}_{121} \\ \mathbf{q}_{221} \\ \mathbf{q}_{321} \\ \underline{\mathbf{q}_{421}} \\ \mathbf{q}_{131} \\ \mathbf{q}_{231} \\ \mathbf{q}_{331} \\ \mathbf{q}_{431} \end{bmatrix} = \mathbf{Q}_{11}\mathbf{l}_1 = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{D} & \mathbf{B} & \mathbf{0} \\ \mathbf{D} & \mathbf{E} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{l}_{11} \\ \mathbf{l}_{21} \\ \mathbf{l}_{31} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{D} & \mathbf{B} & \mathbf{0} \\ \mathbf{D} & \mathbf{E} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}_{11} \\ \mathbf{r}_{111} \\ \mathbf{r}_{211} \\ \mathbf{r}_{511} \\ \underline{\mathbf{r}_{111}} \\ \mathbf{r}_{121} \\ \mathbf{r}_{221} \\ \underline{\mathbf{r}_{521}} \\ \mathbf{r}_{121} \\ \mathbf{r}_{131} \\ \mathbf{r}_{321} \\ \mathbf{r}_{531} \end{bmatrix},$$

where Q_{11} is (36×36). Furthermore,

$$\mathbf{K}_r(\dot{\mathbf{q}}) = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Omega}_{31} \end{bmatrix}$$

which is also a (36×36) matrix.

$$\mathbf{K}_p(\mathbf{q}) = \mathbf{H}\tilde{\mathbf{K}} = \mathbf{H}(\mathbf{B}^d\mathbf{K}^d + \mathbf{B}^o\mathbf{K}^o) = \mathbf{H}_1(\mathbf{B}_3\overline{\mathbf{K}}_1^d + \mathbf{B}_1^o\mathbf{K}_1^0)$$

$$\mathbf{K}_p = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{31} \end{bmatrix} \left(\begin{bmatrix} \mathbf{B}_{1n1} & \mathbf{B}_{11}^d & \mathbf{B}_{21}^d & \hat{\mathbf{B}}_{n1}^d \\ \mathbf{0} & \hat{\mathbf{B}}_{01}^d & \mathbf{B}_{11}^d & \mathbf{B}_{21}^d \\ \overline{\mathbf{B}}_{1n1} & \mathbf{B}_{21}^d & \hat{\mathbf{B}}_{01}^d & \mathbf{B}_{11}^d \end{bmatrix} \begin{bmatrix} \mathbf{K}_{131} \\ \mathbf{K}_{11}^d \\ \mathbf{K}_{21}^d \\ \mathbf{K}_{31}^d \end{bmatrix} \right.$$

$$\left. + \begin{bmatrix} \mathbf{B}_{11}^o & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{01}^o & \mathbf{B}_{11}^o & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{01}^o & \mathbf{B}_{n1}^o \end{bmatrix} \begin{bmatrix} \mathbf{K}_{11}^o \\ \mathbf{K}_{21}^o \\ \mathbf{K}_{31}^o \end{bmatrix} \right)$$

$$(36 \times 36) = (36 \times 36)\ [(36 \times 75) * (75 \times 36) + (36 \times 18) * (18 \times 36)]$$

In order to form $\overline{\mathbf{K}}_1^{d1}$, $\mathbf{R}$ is needed. In order to form $\mathbf{K}_1^o$, $\mathbf{Q}^{-1}$ is needed. Therefore, we obtain $\mathbf{R}$ as follows:

© 2002 by CRC Press LLC

$$\mathbf{l}^d = \mathbf{l}_1^d = \begin{bmatrix} \mathbf{t}_{131}^d \\ \mathbf{t}_1^d \end{bmatrix} = \mathbf{R}\mathbf{q} = \begin{bmatrix} \mathbf{R}_0 \\ \tilde{\mathbf{R}}_{11} \end{bmatrix} [\mathbf{q}_1],$$

$$\begin{bmatrix} \mathbf{t}_{131}^d \\ \hline \mathbf{t}_{11}^d \\ \mathbf{t}_{21}^d \\ \mathbf{t}_{31}^d \end{bmatrix} = \left[\begin{array}{ccc} \mathbf{E}_6 & \mathbf{0} & \mathbf{E}_7 \\ \hline \mathbf{E}_3 & \tilde{\mathbf{E}}_4 & \mathbf{E}_2 \\ \mathbf{E}_2 & \mathbf{E}_3 & \tilde{\mathbf{E}}_4 \\ \tilde{\mathbf{E}}_4 & \mathbf{E}_2 & \mathbf{E}_3 \end{array}\right] \begin{bmatrix} \mathbf{q}_{11} \\ \mathbf{q}_{21} \\ \mathbf{q}_{31} \end{bmatrix},$$

where the matrix is dimension (75 × 36).

$$\mathbf{B}(\mathbf{q}) = \mathbf{H}\tilde{\mathbf{B}}\mathbf{T} = \mathbf{H}[\mathbf{B}^d\mathbf{P}^d, \mathbf{B}^o\mathbf{P}^o]\mathbf{T}$$

$$\left[\mathbf{B}^d\mathbf{P}^d\right] = \left[\mathbf{B}_3\bar{\mathbf{P}}_1^d\right] = \begin{bmatrix} \mathbf{B}_{1n1} & \mathbf{B}_{11}^d & \mathbf{B}_{21}^d & \hat{\mathbf{B}}_{n1}^d \\ \mathbf{0} & \hat{\mathbf{B}}_{01}^d & \mathbf{B}_{11}^d & \mathbf{B}_{21}^d \\ \bar{\mathbf{B}}_{1n1} & \mathbf{B}_{21}^d & \hat{\mathbf{B}}_{01}^d & \mathbf{B}_{11}^d \end{bmatrix} \begin{bmatrix} \mathbf{P}_{1n1}^d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{11}^d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{21}^d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{31}^d \end{bmatrix}$$

where the dimensions are (36 × 25) = (36 × 75) ∗ (75 × 25).

$$\left[\mathbf{B}^o\mathbf{P}^o\right] = \left[\mathbf{B}_1\mathbf{P}_1^o\right] = \begin{bmatrix} \mathbf{B}_{11}^o & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{01}^o & \mathbf{B}_{11}^o & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{01}^o & \mathbf{B}_{n1}^o \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11}^o & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{21}^o & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{31}^o \end{bmatrix}$$

and the dimensions are (36 × 6) = (36 × 18) ∗ (18 × 6).

$$\mathbf{B} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{31} \end{bmatrix} \left[\mathbf{B}_3\bar{\mathbf{P}}_1^d, \mathbf{B}_1\mathbf{P}_1^o\right] \left[\begin{array}{ccccc|c} 0 & \cdots & & \cdots & 0 & 01 \\ \hline \mathbf{T}_1' & \mathbf{0} & \mathbf{0} & \mathbf{S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_1' & \mathbf{0} & \mathbf{0} & \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_1' & \mathbf{0} & \mathbf{0} & \mathbf{S} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_2 \end{array}\right]$$

and the dimensions are (36 × 24) = (36 × 36) ∗ (36 × 31) ∗ (31 × 24).

The control inputs $u_{\alpha ij}$ are defined as

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}^d \\ \mathbf{u}^o \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1'^d \\ \mathbf{u}_1^o \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{11}'^d \\ \mathbf{u}_{21}'^d \\ \mathbf{u}_{31}'^d \\ \mathbf{u}_{11}^o \\ \mathbf{u}_{21}^o \\ \mathbf{u}_{31}^o \end{bmatrix} \quad (24 \times 1)$$

© 2002 by CRC Press LLC

where

$$\mathbf{u}_{i1}'^{d} = \begin{bmatrix} u_{2n1} \\ u_{3n1} \\ u_{4n1} \\ u_{6n1} \\ u_{7n1} \\ u_{8n1} \end{bmatrix} \in \boldsymbol{R}^{6\times 1},\ i = 1,\ 2,\ 3\ \ j = 1.$$

$$\mathbf{Bu} \in \boldsymbol{R}^{(36\times 1)}.$$

The external forces applied to the nodes arise in the product **Dw**, where

$$\mathbf{D(q)} = \mathbf{HW}^{o} = \mathbf{H}\left[\mathbf{W}_1\right] = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{31} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W} \end{bmatrix}$$

with dimensions: $(36 \times 36) = (36 \times 36) * (36 \times 36)$.

$$\mathbf{w} = \mathbf{w}_1 = \begin{bmatrix} \mathbf{w}_{11} \\ \mathbf{w}_{21} \\ \mathbf{w}_{31} \end{bmatrix} (36\times 1)$$

so

$$\mathbf{Dw} \in \boldsymbol{R}^{(36\times 1)}.$$

© 2002 by CRC Press LLC

Appendix 18.E Nodal Forces

At the base, right end of Figure 18.5, where ($i = 1, j = 1$):

$$\bar{\mathbf{f}}_{11} = \begin{bmatrix} \hat{\mathbf{f}}_{111} + \hat{\mathbf{f}}_{211} \\ \hat{\mathbf{f}}_{111} - \hat{\mathbf{f}}_{211} \\ \hat{\mathbf{f}}_{311} + \hat{\mathbf{f}}_{411} \\ \hat{\mathbf{f}}_{311} - \hat{\mathbf{f}}_{411} \end{bmatrix}$$

$$= \begin{bmatrix} (-\mathbf{f}_{211} + \mathbf{f}_{311} + \mathbf{f}_{1n1} + \mathbf{f}_{2n1} + \mathbf{w}_{111}) + (\mathbf{f}_{511} - \mathbf{f}_{111} - \mathbf{f}_{411} - \mathbf{f}_{811} + \mathbf{w}_{211}) \\ (-\mathbf{f}_{211} + \mathbf{f}_{311} + \mathbf{f}_{1n1} + \mathbf{f}_{2n1} + \mathbf{w}_{111}) - (\mathbf{f}_{511} - \mathbf{f}_{111} - \mathbf{f}_{411} - \mathbf{f}_{811} + \mathbf{w}_{211}) \\ (\mathbf{f}_{811} - \mathbf{f}_{711} + \mathbf{f}_{1011} - \mathbf{f}_{321} + \mathbf{f}_{421} + \mathbf{w}_{311}) + (\mathbf{f}_{611} - \mathbf{f}_{511} + \mathbf{f}_{212} + \mathbf{f}_{112} + \mathbf{w}_{411}) \\ (\mathbf{f}_{811} - \mathbf{f}_{711} + \mathbf{f}_{1011} - \mathbf{f}_{321} + \mathbf{f}_{421} + \mathbf{w}_{311}) - (\mathbf{f}_{611} - \mathbf{f}_{511} + \mathbf{f}_{212} + \mathbf{f}_{112} + \mathbf{w}_{411}) \end{bmatrix}$$

At the center of the base, where ($1 < i < n, j = 1$):

$$\bar{\mathbf{f}}_{i1} = \begin{bmatrix} \hat{\mathbf{f}}_{1i1} + \hat{\mathbf{f}}_{2i1} \\ \hat{\mathbf{f}}_{1i1} - \hat{\mathbf{f}}_{2i1} \\ \hat{\mathbf{f}}_{3i1} + \hat{\mathbf{f}}_{4i1} \\ \hat{\mathbf{f}}_{3i1} - \hat{\mathbf{f}}_{4i1} \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{f}_{1(i-1)1} + \mathbf{f}_{2(i-1)1} - \mathbf{f}_{2i1} + \mathbf{f}_{3i1} + \mathbf{w}_{1i1}) + (-\mathbf{f}_{1i1} - \mathbf{f}_{4i1} + \mathbf{f}_{5i1} - \mathbf{f}_{8i1} + \mathbf{w}_{2i1}) \\ (\mathbf{f}_{1(i-1)1} + \mathbf{f}_{2(i-1)1} - \mathbf{f}_{2i1} + \mathbf{f}_{3i1} + \mathbf{w}_{1i1}) - (-\mathbf{f}_{1i1} - \mathbf{f}_{4i1} - \mathbf{f}_{5i1} - \mathbf{f}_{8i1} + \mathbf{w}_{2i1}) \\ (\mathbf{f}_{8i1} - \mathbf{f}_{7i1} + \mathbf{f}_{10i1} - \mathbf{f}_{3(i+1)1} + \mathbf{f}_{4(i+1)1} + \mathbf{w}_{3i1}) + (-\mathbf{f}_{5i1} + \mathbf{f}_{6i1} - \mathbf{f}_{10(i-1)} + \mathbf{f}_{1i2} + \mathbf{f}_{2i2} + \mathbf{w}_{4i1}) \\ (\mathbf{f}_{8i1} - \mathbf{f}_{7i1} + \mathbf{f}_{10i1} - \mathbf{f}_{3(i+1)1} + \mathbf{f}_{4(i+1)1} + \mathbf{w}_{3i1}) - (-\mathbf{f}_{5i1} + \mathbf{f}_{6i1} - \mathbf{f}_{10(i-1)} + \mathbf{f}_{1i2} + \mathbf{f}_{2i2} + \mathbf{w}_{4i1}) \end{bmatrix}$$

At the left end of the base in Figure 18.5, where ($i = n, j = 1$):

$$\bar{\mathbf{f}}_{n1} = \begin{bmatrix} \hat{\mathbf{f}}_{1n1} + \hat{\mathbf{f}}_{2n1} \\ \hat{\mathbf{f}}_{1n1} - \hat{\mathbf{f}}_{2n1} \\ \hat{\mathbf{f}}_{3n1} + \hat{\mathbf{f}}_{4n1} \\ \hat{\mathbf{f}}_{3n1} - \hat{\mathbf{f}}_{4n1} \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{f}_{1(n-1)1} + \mathbf{f}_{2(n-1)1} + \mathbf{f}_{3n1} - \mathbf{f}_{2n1} + \mathbf{w}_{1n1}) + (-\mathbf{f}_{1n1} - \mathbf{f}_{4n1} + \mathbf{f}_{5n1} - \mathbf{f}_{8n1} + \mathbf{w}_{2n1}) \\ (\mathbf{f}_{1(n-1)1} + \mathbf{f}_{2(n-1)1} + \mathbf{f}_{3n1} - \mathbf{f}_{2n1} + \mathbf{w}_{1n1}) - (-\mathbf{f}_{1n1} - \mathbf{f}_{4n1} + \mathbf{f}_{5n1} - \mathbf{f}_{8n1} + \mathbf{w}_{2n1}) \\ (-\mathbf{f}_{7n1} + \mathbf{f}_{8n1} + \mathbf{f}_{10n1} - \mathbf{f}_{311} + \mathbf{f}_{411} + \mathbf{w}_{3n1}) + (-\mathbf{f}_{10(n-1)1} - \mathbf{f}_{5n1} + \mathbf{f}_{6n1} + \mathbf{f}_{1n2} + \mathbf{f}_{2n2} + \mathbf{w}_{4n1}) \\ (-\mathbf{f}_{7n1} + \mathbf{f}_{8n1} + \mathbf{f}_{10n1} - \mathbf{f}_{311} + \mathbf{f}_{411} + \mathbf{w}_{3n1}) - (-\mathbf{f}_{10(n-1)1} - \mathbf{f}_{5n1} + \mathbf{f}_{6n1} + \mathbf{f}_{1n2} + \mathbf{f}_{2n2} + \mathbf{w}_{4n1}) \end{bmatrix} \quad (18.E.1)$$

At the second stage, where ($1 \le i \le n, j = 2$):

© 2002 by CRC Press LLC

$$\bar{\mathbf{f}}_{12} = \begin{bmatrix} \hat{\mathbf{f}}_{112} + \hat{\mathbf{f}}_{212} \\ \hat{\mathbf{f}}_{112} - \hat{\mathbf{f}}_{212} \\ \hat{\mathbf{f}}_{312} + \hat{\mathbf{f}}_{412} \\ \hat{\mathbf{f}}_{312} - \hat{\mathbf{f}}_{412} \end{bmatrix}$$

$$= \begin{bmatrix} \left(-\mathbf{f}_{6n1} + \mathbf{f}_{7n1} - \mathbf{f}_{9n2} - \mathbf{f}_{212} + \mathbf{f}_{312} + \mathbf{w}_{112}\right) + \left(-\mathbf{f}_{112} - \mathbf{f}_{412} + \mathbf{f}_{512} - \mathbf{f}_{812} + \mathbf{f}_{912} + \mathbf{w}_{212}\right) \\ \left(-\mathbf{f}_{6n1} + \mathbf{f}_{7n1} - \mathbf{f}_{9n2} - \mathbf{f}_{212} + \mathbf{f}_{312} + \mathbf{w}_{112}\right) - \left(-\mathbf{f}_{112} - \mathbf{f}_{412} + \mathbf{f}_{512} - \mathbf{f}_{812} + \mathbf{f}_{912} + \mathbf{w}_{212}\right) \\ \left(-\mathbf{f}_{712} + \mathbf{f}_{812} + \mathbf{f}_{1012} - \mathbf{f}_{322} + \mathbf{f}_{422} + \mathbf{w}_{312}\right) + \left(-\mathbf{f}_{512} + \mathbf{f}_{612} - \mathbf{f}_{10n2} + \mathbf{f}_{113} + \mathbf{f}_{213} + \mathbf{w}_{412}\right) \\ \left(-\mathbf{f}_{712} + \mathbf{f}_{812} + \mathbf{f}_{1012} - \mathbf{f}_{322} + \mathbf{f}_{422} + \mathbf{w}_{312}\right) - \left(-\mathbf{f}_{512} + \mathbf{f}_{612} - \mathbf{f}_{10n2} + \mathbf{f}_{113} + \mathbf{f}_{213} + \mathbf{w}_{412}\right) \end{bmatrix}$$

$$\bar{\mathbf{f}}_{i2} = \begin{bmatrix} \hat{\mathbf{f}}_{1i2} + \hat{\mathbf{f}}_{2i2} \\ \hat{\mathbf{f}}_{1i2} - \hat{\mathbf{f}}_{2i2} \\ \hat{\mathbf{f}}_{3i2} + \hat{\mathbf{f}}_{4i2} \\ \hat{\mathbf{f}}_{3i2} - \hat{\mathbf{f}}_{4i2} \end{bmatrix} \tag{18.E.2}$$

$$= \begin{bmatrix} \left(-\mathbf{f}_{6(i-1)1} + \mathbf{f}_{7(i-1)1} - \mathbf{f}_{9(i-1)2} - \mathbf{f}_{2i2} + \mathbf{f}_{3i2} + \mathbf{w}_{1i2}\right) + \left(-\mathbf{f}_{1i2} - \mathbf{f}_{4i2} + \mathbf{f}_{5i2} - \mathbf{f}_{8i2} + \mathbf{f}_{9i2} + \mathbf{w}_{2i2}\right) \\ \left(-\mathbf{f}_{6(i-1)1} + \mathbf{f}_{7(i-1)1} - \mathbf{f}_{9(i-1)2} - \mathbf{f}_{2i2} + \mathbf{f}_{3i2} + \mathbf{w}_{1i2}\right) - \left(-\mathbf{f}_{1i2} - \mathbf{f}_{4i2} + \mathbf{f}_{5i2} - \mathbf{f}_{8i2} + \mathbf{f}_{9i2} + \mathbf{w}_{2i2}\right) \\ \left(-\mathbf{f}_{7i2} + \mathbf{f}_{8i2} + \mathbf{f}_{10i2} - \mathbf{f}_{3(i+1)2} + \mathbf{f}_{4(i+1)2} + \mathbf{w}_{3i2}\right) + \left(-\mathbf{f}_{5i2} + \mathbf{f}_{6i2} - \mathbf{f}_{10(i-1)2} + \mathbf{f}_{1i3} + \mathbf{f}_{2i3} + \mathbf{w}_{4i2}\right) \\ \left(-\mathbf{f}_{7i2} + \mathbf{f}_{8i2} + \mathbf{f}_{10i2} - \mathbf{f}_{3(i+1)2} + \mathbf{f}_{4(i+1)2} + \mathbf{w}_{3i2}\right) - \left(-\mathbf{f}_{5i2} + \mathbf{f}_{6i2} - \mathbf{f}_{10(i-1)2} + \mathbf{f}_{1i3} + \mathbf{f}_{2i3} + \mathbf{w}_{4i2}\right) \end{bmatrix}$$

$$\bar{\mathbf{f}}_{n2} = \begin{bmatrix} \hat{\mathbf{f}}_{1n2} + \hat{\mathbf{f}}_{2n2} \\ \hat{\mathbf{f}}_{1n2} - \hat{\mathbf{f}}_{2n2} \\ \hat{\mathbf{f}}_{3n2} + \hat{\mathbf{f}}_{4n2} \\ \hat{\mathbf{f}}_{3n2} - \hat{\mathbf{f}}_{4n2} \end{bmatrix}$$

$$= \begin{bmatrix} \left(-\mathbf{f}_{6(n-1)1} + \mathbf{f}_{7(n-1)1} - \mathbf{f}_{9(n-1)2} - \mathbf{f}_{2n2} + \mathbf{f}_{3n2} + \mathbf{w}_{1n2}\right) + \left(-\mathbf{f}_{1n2} - \mathbf{f}_{4n2} + \mathbf{f}_{5n2} - \mathbf{f}_{8n2} + \mathbf{f}_{9n2} + \mathbf{w}_{2n2}\right) \\ \left(-\mathbf{f}_{6(n-1)1} + \mathbf{f}_{7(n-1)1} - \mathbf{f}_{9(n-1)2} - \mathbf{f}_{2n2} + \mathbf{f}_{3n2} + \mathbf{w}_{1n2}\right) - \left(-\mathbf{f}_{1n2} - \mathbf{f}_{4n2} + \mathbf{f}_{5n2} - \mathbf{f}_{8n2} + \mathbf{f}_{9n2} + \mathbf{w}_{2n2}\right) \\ \left(-\mathbf{f}_{7n2} + \mathbf{f}_{8n2} + \mathbf{f}_{10n2} - \mathbf{f}_{312} + \mathbf{f}_{412} + \mathbf{w}_{3n2}\right) + \left(-\mathbf{f}_{5n2} + \mathbf{f}_{6n2} - \mathbf{f}_{10(n-1)2} + \mathbf{f}_{1n3} + \mathbf{f}_{2n3} + \mathbf{w}_{4n2}\right) \\ \left(-\mathbf{f}_{7n2} + \mathbf{f}_{8n2} + \mathbf{f}_{10n2} - \mathbf{f}_{312} + \mathbf{f}_{412} + \mathbf{w}_{3n2}\right) - \left(-\mathbf{f}_{5n2} + \mathbf{f}_{6n2} - \mathbf{f}_{10(n-1)2} + \mathbf{f}_{1n3} + \mathbf{f}_{2n3} + \mathbf{w}_{4n2}\right) \end{bmatrix}$$

At the typical stage ($1 \le j < m$, $1 \le i \le n$):

$$\bar{\mathbf{f}}_{1j} = \begin{bmatrix} \hat{\mathbf{f}}_{11j} + \hat{\mathbf{f}}_{21j} \\ \hat{\mathbf{f}}_{11j} - \hat{\mathbf{f}}_{21j} \\ \hat{\mathbf{f}}_{31j} + \hat{\mathbf{f}}_{41j} \\ \hat{\mathbf{f}}_{31j} - \hat{\mathbf{f}}_{41j} \end{bmatrix}$$

$$= \begin{bmatrix} \left(-\mathbf{f}_{6n(j-1)} + \mathbf{f}_{7n(j-1)} - \mathbf{f}_{9nj} - \mathbf{f}_{21j} + \mathbf{f}_{31j} + \mathbf{w}_{11j}\right) + \left(-\mathbf{f}_{11j} - \mathbf{f}_{41j} + \mathbf{f}_{51} - \mathbf{f}_{81j} + \mathbf{f}_{91j} + \mathbf{w}_{21j}\right) \\ \left(-\mathbf{f}_{6n(j-1)} + \mathbf{f}_{7n(j-1)} - \mathbf{f}_{9nj} - \mathbf{f}_{21j} + \mathbf{f}_{31j} + \mathbf{w}_{11j}\right) - \left(-\mathbf{f}_{11j} - \mathbf{f}_{41j} + \mathbf{f}_{51j} - \mathbf{f}_{81j} + \mathbf{f}_{91j} + \mathbf{w}_{21j}\right) \\ \left(-\mathbf{f}_{71j} + \mathbf{f}_{81j} + \mathbf{f}_{101j} - \mathbf{f}_{32j} + \mathbf{f}_{42j} + \mathbf{w}_{31j}\right) + \left(-\mathbf{f}_{51j} + \mathbf{f}_{61j} - \mathbf{f}_{10nj} + \mathbf{f}_{11(j+1)} + \mathbf{f}_{21(j+1)} + \mathbf{w}_{41j}\right) \\ \left(-\mathbf{f}_{71j} + \mathbf{f}_{81j} + \mathbf{f}_{101j} - \mathbf{f}_{32j} + \mathbf{f}_{42j} + \mathbf{w}_{31j}\right) - \left(-\mathbf{f}_{51j} + \mathbf{f}_{61j} - \mathbf{f}_{10nj} + \mathbf{f}_{11(j+1)} + \mathbf{f}_{21(j+1)} + \mathbf{w}_{41j}\right) \end{bmatrix}$$

© 2002 by CRC Press LLC

$$\bar{\mathbf{f}}_{ij} = \begin{bmatrix} \hat{\mathbf{f}}_{1ij} + \hat{\mathbf{f}}_{2ij} \\ \hat{\mathbf{f}}_{1ij} - \hat{\mathbf{f}}_{2ij} \\ \hat{\mathbf{f}}_{3ij} + \hat{\mathbf{f}}_{4ij} \\ \hat{\mathbf{f}}_{3ij} - \hat{\mathbf{f}}_{4ij} \end{bmatrix}$$

$$= \begin{bmatrix} (-\mathbf{f}_{6(i-1)(j-1)} + \mathbf{f}_{7(i-1)(j-1)} - \mathbf{f}_{9(i-1)j} - \mathbf{f}_{2ij} + \mathbf{f}_{3ij} + \mathbf{w}_{1ij}) + (-\mathbf{f}_{1ij} - \mathbf{f}_{4ij} + \mathbf{f}_{5ij} - \mathbf{f}_{8ij} + \mathbf{f}_{9ij} + \mathbf{w}_{2ij}) \\ (-\mathbf{f}_{6(i-1)(j-1)} + \mathbf{f}_{7(i-1)(j-1)} - \mathbf{f}_{9(i-1)j} - \mathbf{f}_{2ij} + \mathbf{f}_{3ij} + \mathbf{w}_{1ij}) - (-\mathbf{f}_{1ij} - \mathbf{f}_{4ij} + \mathbf{f}_{5ij} - \mathbf{f}_{8ij} + \mathbf{f}_{9ij} + \mathbf{w}_{2ij}) \\ (-\mathbf{f}_{7ij} + \mathbf{f}_{8ij} + \mathbf{f}_{10ij} - \mathbf{f}_{3(i+1)j} + \mathbf{f}_{4(i+1)j} + \mathbf{w}_{3ij}) + (-\mathbf{f}_{5ij} + \mathbf{f}_{6ij} - \mathbf{f}_{10(i-1)j} + \mathbf{f}_{1i(j+1)} + \mathbf{f}_{2i(j+1)} + \mathbf{w}_{4ij}) \\ (-\mathbf{f}_{7ij} + \mathbf{f}_{8ij} + \mathbf{f}_{10ij} - \mathbf{f}_{3(i+1)j} + \mathbf{f}_{4(i+1)j} + \mathbf{w}_{3ij}) - (-\mathbf{f}_{5ij} + \mathbf{f}_{6ij} - \mathbf{f}_{10(i-1)j} + \mathbf{f}_{1i(j+1)} + \mathbf{f}_{2i(j+1)} + \mathbf{w}_{4ij}) \end{bmatrix} \quad (18.E.3)$$

$$\bar{\mathbf{f}}_{nj} = \begin{bmatrix} \hat{\mathbf{f}}_{1nj} + \hat{\mathbf{f}}_{2nj} \\ \hat{\mathbf{f}}_{1nj} - \hat{\mathbf{f}}_{2nj} \\ \hat{\mathbf{f}}_{3nj} + \hat{\mathbf{f}}_{4nj} \\ \hat{\mathbf{f}}_{3nj} - \hat{\mathbf{f}}_{4nj} \end{bmatrix}$$

$$= \begin{bmatrix} (-\mathbf{f}_{6(n-1)(j-1)} + \mathbf{f}_{7(n-1)(j-1)} - \mathbf{f}_{9(n-1)j} - \mathbf{f}_{2nj} + \mathbf{f}_{3nj} + \mathbf{w}_{1nj}) + (-\mathbf{f}_{1nj} - \mathbf{f}_{4nj} + \mathbf{f}_{5nj} - \mathbf{f}_{8nj} + \mathbf{f}_{9nj} + \mathbf{w}_{2nj}) \\ (-\mathbf{f}_{6(n-1)(j-1)} + \mathbf{f}_{7(n-1)(j-1)} - \mathbf{f}_{9(n-1)j} - \mathbf{f}_{2nj} + \mathbf{f}_{3nj} + \mathbf{w}_{1nj}) - (-\mathbf{f}_{1nj} - \mathbf{f}_{4nj} + \mathbf{f}_{5nj} - \mathbf{f}_{8nj} + \mathbf{f}_{9nj} + \mathbf{w}_{2nj}) \\ (-\mathbf{f}_{7nj} + \mathbf{f}_{8nj} + \mathbf{f}_{10nj} - \mathbf{f}_{31j} + \mathbf{f}_{41j} + \mathbf{w}_{3nj}) + (-\mathbf{f}_{5nj} + \mathbf{f}_{6nj} - \mathbf{f}_{10(n-1)j} + \mathbf{f}_{1n(j+1)} + \mathbf{f}_{2n(j+1)} + \mathbf{w}_{4nj}) \\ (-\mathbf{f}_{7nj} + \mathbf{f}_{8nj} + \mathbf{f}_{10nj} - \mathbf{f}_{31j} + \mathbf{f}_{41j} + \mathbf{w}_{3nj}) - (-\mathbf{f}_{5nj} + \mathbf{f}_{6nj} - \mathbf{f}_{10(n-1)j} + \mathbf{f}_{1n(j+1)} + \mathbf{f}_{2n(j+1)} + \mathbf{w}_{4nj}) \end{bmatrix}$$

$(1 \le i \le n,\ j = m)$

$$\bar{\mathbf{f}}_{1m} = \begin{bmatrix} \hat{\mathbf{f}}_{11m} + \hat{\mathbf{f}}_{21m} \\ \hat{\mathbf{f}}_{11m} - \hat{\mathbf{f}}_{21m} \\ \hat{\mathbf{f}}_{31m} + \hat{\mathbf{f}}_{41m} \\ \hat{\mathbf{f}}_{31m} - \hat{\mathbf{f}}_{41m} \end{bmatrix}$$

$$= \begin{bmatrix} (-\mathbf{f}_{6n(m-1)} + \mathbf{f}_{7n(m-1)} - \mathbf{f}_{21m} + \mathbf{f}_{31m} - \mathbf{f}_{9nm} + \mathbf{w}_{1nm}) + (-\mathbf{f}_{41m} + \mathbf{f}_{51m} - \mathbf{f}_{81m} - \mathbf{f}_{11m} + \mathbf{f}_{91m} + \mathbf{w}_{2nm}) \\ (-\mathbf{f}_{6n(m-1)} + \mathbf{f}_{7n(m-1)} - \mathbf{f}_{21m} + \mathbf{f}_{31m} - \mathbf{f}_{9nm} + \mathbf{w}_{1nm}) - (-\mathbf{f}_{41m} + \mathbf{f}_{51m} - \mathbf{f}_{81m} - \mathbf{f}_{11m} + \mathbf{f}_{91m} + \mathbf{w}_{2nm}) \\ (-\mathbf{f}_{71m} + \mathbf{f}_{81m} - \mathbf{f}_{32m} + \mathbf{f}_{42m} + \mathbf{w}_{31m}) + (-\mathbf{f}_{51m} + \mathbf{f}_{61m} - \mathbf{f}_{6nm} + \mathbf{f}_{7nm} + \mathbf{w}_{41m}) \\ (-\mathbf{f}_{71m} + \mathbf{f}_{81m} - \mathbf{f}_{32m} + \mathbf{f}_{42m} + \mathbf{w}_{31m}) - (-\mathbf{f}_{51m} + \mathbf{f}_{61m} - \mathbf{f}_{6nm} + \mathbf{f}_{7nm} + \mathbf{w}_{41m}) \end{bmatrix}$$

$$\bar{\mathbf{f}}_{im} = \begin{bmatrix} \hat{\mathbf{f}}_{1im} + \hat{\mathbf{f}}_{2im} \\ \hat{\mathbf{f}}_{1im} - \hat{\mathbf{f}}_{2im} \\ \hat{\mathbf{f}}_{3im} + \hat{\mathbf{f}}_{4im} \\ \hat{\mathbf{f}}_{3im} - \hat{\mathbf{f}}_{4im} \end{bmatrix} \quad (18.E.4)$$

$$= \begin{bmatrix} (-\mathbf{f}_{6(i-1)(m-1)} + \mathbf{f}_{7(i-1)(m-1)} - \mathbf{f}_{9(i-1)m} - \mathbf{f}_{2im} + \mathbf{f}_{3im} + \mathbf{w}_{1im}) + (-\mathbf{f}_{1im} - \mathbf{f}_{4im} + \mathbf{f}_{5im} - \mathbf{f}_{8im} + \mathbf{f}_{9im} + \mathbf{w}_{2im}) \\ (-\mathbf{f}_{6(i-1)(m-1)} + \mathbf{f}_{7(i-1)(m-1)} - \mathbf{f}_{9(i-1)m} - \mathbf{f}_{2im} + \mathbf{f}_{3im} + \mathbf{w}_{1im}) - (-\mathbf{f}_{1im} - \mathbf{f}_{4im} + \mathbf{f}_{5im} - \mathbf{f}_{8im} + \mathbf{f}_{9im} + \mathbf{w}_{2im}) \\ (-\mathbf{f}_{7im} + \mathbf{f}_{8im} - \mathbf{f}_{3(i+1)m} + \mathbf{f}_{4(i+1)m} + \mathbf{w}_{3im}) + (-\mathbf{f}_{5im} - \mathbf{f}_{6(i-1)m} + \mathbf{f}_{6im} + \mathbf{f}_{7(i-1)m} + \mathbf{w}_{4im}) \\ (-\mathbf{f}_{7im} + \mathbf{f}_{8im} - \mathbf{f}_{3(i+1)m} + \mathbf{f}_{4(i+1)m} + \mathbf{w}_{3im}) - (-\mathbf{f}_{5im} - \mathbf{f}_{6(i-1)m} + \mathbf{f}_{6im} + \mathbf{f}_{7(i-1)m} + \mathbf{w}_{4im}) \end{bmatrix}$$

$$\bar{\mathbf{f}}_{nm} = \begin{bmatrix} \hat{\mathbf{f}}_{1nm} + \hat{\mathbf{f}}_{2nm} \\ \hat{\mathbf{f}}_{1nm} - \hat{\mathbf{f}}_{2nm} \\ \hat{\mathbf{f}}_{3nm} + \hat{\mathbf{f}}_{4nm} \\ \hat{\mathbf{f}}_{3nm} - \hat{\mathbf{f}}_{4nm} \end{bmatrix}$$

$$= \begin{bmatrix} (-\mathbf{f}_{6(n-1)(m-1)} + \mathbf{f}_{7(n-1)(m-1)} - \mathbf{f}_{9(n-1)m} - \mathbf{f}_{2nm} + \mathbf{f}_{3nm} + \mathbf{w}_{1nm}) + (-\mathbf{f}_{1nm} - \mathbf{f}_{4nm} + \mathbf{f}_{5nm} - \mathbf{f}_{8nm} + \mathbf{f}_{9nm} + \mathbf{w}_{2nm}) \\ (-\mathbf{f}_{6(n-1)(m-1)} + \mathbf{f}_{7(n-1)(m-1)} - \mathbf{f}_{9(n-1)m} - \mathbf{f}_{2nm} + \mathbf{f}_{3nm} + \mathbf{w}_{1nm}) - (-\mathbf{f}_{1nm} - \mathbf{f}_{4nm} + \mathbf{f}_{5nm} - \mathbf{f}_{8nm} + \mathbf{f}_{9nm} + \mathbf{w}_{2nm}) \\ (-\mathbf{f}_{7nm} + \mathbf{f}_{8nm} - \mathbf{f}_{31m} + \mathbf{f}_{41m} + \mathbf{w}_{3nm}) + (-\mathbf{f}_{5nm} + \mathbf{f}_{6nm} - \mathbf{f}_{6(n-1)m} + \mathbf{f}_{7(n-1)m} + \mathbf{w}_{4nm}) \\ (-\mathbf{f}_{7nm} + \mathbf{f}_{8nm} - \mathbf{f}_{31m} + \mathbf{f}_{41m} + \mathbf{w}_{3nm}) - (-\mathbf{f}_{5nm} + \mathbf{f}_{6nm} - \mathbf{f}_{6(n-1)m} + \mathbf{f}_{7(n-1)m} + \mathbf{w}_{4nm}) \end{bmatrix}$$

© 2002 by CRC Press LLC

$$\mathbf{f}_{ij}^{o} = \begin{bmatrix} \mathbf{f}_5 \\ \mathbf{f}_1 \end{bmatrix}_{ij}, \mathbf{f}_{ij}^{d} = \begin{bmatrix} \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \\ \mathbf{f}_6 \\ \mathbf{f}_7 \\ \mathbf{f}_8 \\ \mathbf{f}_9 \\ \mathbf{f}_{10} \end{bmatrix}_{ij}, \mathbf{w}_{ij} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \\ \mathbf{w}_4 \end{bmatrix}_{ij}$$

$$\bar{\mathbf{f}}_{11} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{n1}^{d} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}_{11}^{d}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{12}^{d} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{21}^{d}$$

$$+ \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}_{11}^{o} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}_{12}^{o} + \begin{bmatrix} \mathbf{I}_3 \\ \mathbf{I}_3 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{f}_{1n1} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{11}.$$

Or, in matrix notation, with the implied matrix definitions,

$$\bar{\mathbf{f}}_{11} = \mathbf{B}_{n1}^{d}\mathbf{f}_{n1}^{d} + \mathbf{B}_{11}^{d}\mathbf{f}_{11}^{d} + \mathbf{B}_{12}^{d}\mathbf{f}_{12}^{d} + \mathbf{B}_{21}^{d}\mathbf{f}_{21}^{d} + \mathbf{B}_{11}^{o}\mathbf{f}_{11}^{o} + \mathbf{B}_{12}^{o}\mathbf{f}_{12}^{o} + \mathbf{B}_{1n1}\mathbf{f}_{1n1} + \mathbf{W}\mathbf{w}_{11}. \tag{18.E.5}$$

$$\bar{\mathbf{f}}_{i1} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}_{(i-1)1}^{d} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}_{i1}^{d}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{i2}^{d} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{(i+1)1}^{d}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{(i-1)1}^{o} + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}_{i1}^{o} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}_{i2}^{o} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{i1}$$

© 2002 by CRC Press LLC

Or, in matrix form,

$$\begin{aligned}\bar{\mathbf{f}}_{i1} = {} & \mathbf{B}^d_{01}\mathbf{f}^d_{(i-1)1} + \mathbf{B}^d_{11}\mathbf{f}^d_{i1} + \mathbf{B}^d_{12}\mathbf{f}^d_{i2} + \mathbf{B}^d_{21}\mathbf{f}^d_{(i+1)1} + \mathbf{B}^o_{01}\mathbf{f}^o_{(i-1)1} \\ & + \mathbf{B}^o_{11}\mathbf{f}^o_{i1} + \mathbf{B}^o_{12}\mathbf{f}^o_{i2} + \mathbf{W}\mathbf{w}_{i1}.\end{aligned} \tag{18.E.6}$$

$$\begin{aligned}\bar{\mathbf{f}}_{n1} = {} & \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{(n-1)1} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{n1} \\ & + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n2} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^{\,d}_{11} \\ & + \begin{bmatrix} \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{(n-1)1} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{n1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}^o_{n2} + \begin{bmatrix} -\mathbf{I}_3 \\ \mathbf{I}_3 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{f}_{1n1} \\ & + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{n1}.\end{aligned}$$

Or, in matrix form,

$$\begin{aligned}\bar{\mathbf{f}}_{n1} = {} & \mathbf{B}^d_{01}\mathbf{f}^d_{(n-1)1} + \mathbf{B}^d_{11}\mathbf{f}^d_{n1} + \mathbf{B}^d_{12}\mathbf{f}^d_{n2} + \mathbf{B}^d_{21}\mathbf{f}^d_{11} + \mathbf{B}^o_{01}\mathbf{f}^o_{(n-1)1} \\ & + \mathbf{B}^o_{n1}\mathbf{f}^o_{n1} + \mathbf{B}^o_{12}\mathbf{f}^o_{n2} + \overline{\mathbf{B}}_{1n1}\mathbf{f}_{1n1} + \mathbf{W}\mathbf{w}_{n1}.\end{aligned} \tag{18.E.7}$$

$$\begin{aligned}\bar{\mathbf{f}}_{12} = {} & \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{12} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{22} \\ & + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{13} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n1}\end{aligned}$$

© 2002 by CRC Press LLC

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}_{n2}^d + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}_{12}^o + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}_{13}^o$$

$$+\begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{12}.$$

Or, in matrix form,

$$\begin{aligned} \bar{\mathbf{f}}_{12} = {} & \overline{\mathbf{B}}_{12}^d \mathbf{f}_{12}^d + \mathbf{B}_{21}^d \mathbf{f}_{22}^d + \mathbf{B}_{12}^d \mathbf{f}_{13}^d + \overline{\mathbf{B}}_{n1}^d \mathbf{f}_{n1}^d + \mathbf{B}_{n2}^d \mathbf{f}_{n2}^d \\ & + \mathbf{B}_{11}^o \mathbf{f}_{12}^o + \mathbf{B}_{12}^o \mathbf{f}_{13}^o + \mathbf{W}\mathbf{w}_{12}. \end{aligned} \tag{18.E.8}$$

$$\bar{\mathbf{f}}_{i2} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{(i-1)1}^d + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}_{i2}^d$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{i3}^d + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}_{(i-1)2}^d$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{(i+1)2}^d + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}_{i2}^o$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}_{i3}^o + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{i2}.$$

Or, in matrix form,

$$\begin{aligned} \bar{\mathbf{f}}_{i2} = {} & \overline{\mathbf{B}}_{n1}^d \mathbf{f}_{(i-1)1}^d + \overline{\mathbf{B}}_{12}^d \mathbf{f}_{i2}^d + \mathbf{B}_{12}^d \mathbf{f}_{i3}^d + \mathbf{B}_{n2}^d \mathbf{f}_{(i-1)2}^d + \mathbf{B}_{21}^d \mathbf{f}_{(i+1)2}^d \\ & + \mathbf{B}_{11}^o \mathbf{f}_{i2}^o + \mathbf{B}_{12}^o \mathbf{f}_{i3}^o + \mathbf{W}\mathbf{w}_{i2}. \end{aligned} \tag{18.E.9}$$

$$\bar{\mathbf{f}}_{n2} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{12}^d + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}_{(n-1)1}^d$$

© 2002 by CRC Press LLC

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{(n-1)2} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{n2}$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n3} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}^o_{n3} + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{n2}$$

$$+\begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{n2}.$$

Or, in matrix form,

$$\begin{aligned} \bar{\mathbf{f}}_{n2} = {} & \mathbf{B}^d_{21}\mathbf{f}^d_{12} + \overline{\mathbf{B}}^d_{n1}\mathbf{f}^d_{(n-1)1} + \mathbf{B}^d_{n2}\mathbf{f}^d_{(n-1)2} + \overline{\mathbf{B}}^d_{12}\mathbf{f}^d_{n2} + \mathbf{B}^d_{12}\mathbf{f}^d_{n3} + \mathbf{B}^o_{12}\mathbf{f}^o_{n3} \\ & + \mathbf{B}^o_{11}\mathbf{f}^o_{n2} + \mathbf{W}\mathbf{w}_{n2}. \end{aligned} \tag{18.E.10}$$

$$\bar{\mathbf{f}}_{1j} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n(j-1)} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{1j}$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{nj} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{2j}$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{1(j+1)} + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{1j}$$

$$+\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}^o_{1(j+1)} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{1j}.$$

Or, in matrix form,

$$\begin{aligned} \bar{\mathbf{f}}_{1j} = {} & \overline{\mathbf{B}}^d_{n1}\mathbf{f}^d_{n(j-1)} + \overline{\mathbf{B}}^d_{12}\mathbf{f}^d_{1j} + \mathbf{B}^d_{n2}\mathbf{f}^d_{nj} + \mathbf{B}^d_{21}\mathbf{f}^d_{2j} + \mathbf{B}^d_{12}\mathbf{f}^d_{1(j+1)} \\ & + \mathbf{B}^o_{11}\mathbf{f}^o_{1j} + \mathbf{B}^o_{12}\mathbf{f}^o_{1(j+1)} + \mathbf{W}\mathbf{w}_{1j}. \end{aligned} \tag{18.E.11}$$

© 2002 by CRC Press LLC

$$\bar{\mathbf{f}}_{ij} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(i-1)(j-1)} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{ij}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{(i-1)j} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(i+1)j}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{i(j+1)} + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{ij}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}^o_{i(j+1)} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{ij}.$$

Or, in matrix form,

$$\begin{aligned} \bar{\mathbf{f}}_{ij} = {} & \overline{\mathbf{B}}^d_{n1} \mathbf{f}^d_{(i-1)(j-1)} + \overline{\mathbf{B}}^d_{12} \mathbf{f}^d_{ij} + \mathbf{B}^d_{n2} \mathbf{f}^d_{(i-1)j} + \mathbf{B}^d_{21} \mathbf{f}^d_{(i+1)j} + \mathbf{B}^d_{12} \mathbf{f}^d_{i(j+1)} \\ & + \mathbf{B}^o_{11} \mathbf{f}^o_{ij} + \mathbf{B}^o_{12} \mathbf{f}^o_{i(j+1)} + \mathbf{W} \mathbf{w}_{ij}. \end{aligned} \tag{18.E.12}$$

$$\bar{\mathbf{f}}_{nj} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(n-1)(j-1)} + \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{nj}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \mathbf{f}^d_{(n-1)j} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{1j}$$

$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n(j+1)} + \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{nj} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix} \mathbf{f}^o_{n(j+1)}$$

$$+ \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{nj}.$$

© 2002 by CRC Press LLC

Or, in matrix form,

$$\bar{\mathbf{f}}_{nj} = \overline{\mathbf{B}}^d_{n1}\mathbf{f}^d_{(n-1)(j-1)} + \overline{\mathbf{B}}^d_{12}\mathbf{f}^d_{nj} + \mathbf{B}^d_{n2}\mathbf{f}^d_{(n-1)j} + \mathbf{B}^d_{21}\mathbf{f}^d_{1j} + \mathbf{B}^d_{12}\mathbf{f}^d_{n(j+1)} + \mathbf{B}^o_{11}\mathbf{f}^o_{nj} + \mathbf{B}^o_{12}\mathbf{f}^o_{n(j+1)} + \mathbf{W}\mathbf{w}_{nj}. \tag{18.E.13}$$

$$\bar{\mathbf{f}}_{1m} = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{1m} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{2m}$$
$$+ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{nm} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{n(m-1)}$$
$$+ \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{1m} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{W}_{1m}.$$

Or, in matrix form,

$$\bar{\mathbf{f}}_{1m} = \mathbf{B}^d_{1m}\mathbf{f}^d_{1m} + \mathbf{B}^d_{21}\mathbf{f}^d_{2m} + \mathbf{B}^d_{nm}\mathbf{f}^d_{nm} + \overline{\mathbf{B}}^d_{n1}\mathbf{f}^d_{n(m-1)} + \mathbf{B}^o_{11}\mathbf{f}^o_{1m} + \mathbf{W}\mathbf{w}_{1m}. \tag{18.E.14}$$

$$\bar{\mathbf{f}}_{im} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(i-1)(m-1)} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(i-1)m}$$
$$+ \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{1m} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(i+1)m}$$
$$+ \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{im} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{im}.$$

Or, in matrix notation,

$$\bar{\mathbf{f}}_{im} = \overline{\mathbf{B}}^d_{n1}\mathbf{f}^d_{(i-1)(m-1)} + \mathbf{B}^d_{nm}\mathbf{f}^d_{(i-1)m} + \mathbf{B}^d_{1m}\mathbf{f}^d_{im} + \mathbf{B}^d_{21}\mathbf{f}^d_{(i+1)m} + \mathbf{B}^o_{11}\mathbf{f}^o_{im} + \mathbf{W}\mathbf{w}_{im}. \tag{18.E.15}$$

© 2002 by CRC Press LLC

$$\bar{\mathbf{f}}_{nm} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(n-1)(m-1)} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{(n-1)m}$$

$$+ \begin{bmatrix} -\mathbf{I}_3 & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{nm} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{f}^d_{1m}$$

$$+ \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix} \mathbf{f}^o_{nm} + \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_3 & -\mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix} \mathbf{w}_{nm}.$$

Or, in matrix notation,

$$\begin{aligned} \bar{\mathbf{f}}_{nm} &= \bar{\mathbf{B}}^d_{n1}\mathbf{f}^d_{(n-1)(m-1)} + \mathbf{B}^d_{nm1}\mathbf{f}^d_{(n-1)m} + \mathbf{B}^d_{1m}\mathbf{f}^d_{nm} + \mathbf{B}^d_{21}\mathbf{f}^d_{1m} \\ &\quad + \mathbf{B}^o_{11}\mathbf{f}^o_{nm} + \mathbf{W}\mathbf{w}_{nm}. \end{aligned} \tag{18.E.16}$$

Now assemble (18.E.5)–(18.E.16) into the form

$$\mathbf{f}_1 = \mathbf{B}_3\bar{\mathbf{f}}^d_1 + \mathbf{B}_4\mathbf{f}^d_2 + \mathbf{B}_1\mathbf{f}^o_1 + \mathbf{B}_2\mathbf{f}^o_2 + \mathbf{W}_1\mathbf{w}_1 \tag{18.E.17}$$

where

$$\bar{\mathbf{f}}^d_1 = \begin{bmatrix} \mathbf{f}_{1n1} \\ \mathbf{f}^d_1 \end{bmatrix}, \mathbf{f}^d_j = \begin{bmatrix} \mathbf{f}^d_{1j} \\ \mathbf{f}^d_{2j} \\ \vdots \\ \mathbf{f}^d_{nj} \end{bmatrix}, \mathbf{f}^o_j = \begin{bmatrix} \mathbf{f}^o_{1j} \\ \mathbf{f}^o_{2j} \\ \vdots \\ \mathbf{f}^o_{nj} \end{bmatrix}, \mathbf{w}_j = \begin{bmatrix} \mathbf{w}_{1j} \\ \mathbf{w}_{2j} \\ \vdots \\ \mathbf{w}_{nj} \end{bmatrix}, \mathbf{f}_j = \begin{bmatrix} \bar{\mathbf{f}}_{1j} \\ \bar{\mathbf{f}}_{2j} \\ \vdots \\ \bar{\mathbf{f}}_{nj} \end{bmatrix}$$

$$\mathbf{B}_3 = \begin{bmatrix} \mathbf{B}_{1n1} & \mathbf{B}^d_{11} & \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^d_{n1} \\ \mathbf{0} & \mathbf{B}^d_{01} & \ddots & \ddots & \ddots & & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & & \ddots & \ddots & \ddots & \mathbf{B}^d_{21} \\ \bar{\mathbf{B}}_{1n1} & \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^d_{01} & \mathbf{B}^d_{11} \end{bmatrix}$$

$$\mathbf{B}_4 = BlockDiag[\ldots, \mathbf{B}^d_{12}, \mathbf{B}^d_{12}, \ldots],$$

$$\mathbf{B}_2 = BlockDiag[\ldots, \mathbf{B}^o_{12}, \mathbf{B}^d_{12}, \ldots],$$

© 2002 by CRC Press LLC

$$\mathbf{B}_1 = \begin{bmatrix} \mathbf{B}^o_{11} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{B}^o_{01} & \ddots & \ddots & & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{B}^o_{11} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^o_{01} & \mathbf{B}^o_{n1} \end{bmatrix},$$

$$\mathbf{W}_1 = \textit{BlockDiag}\ [\ldots, \mathbf{W}, \mathbf{W}, \ldots].$$

Also, from (18.E.5)–(18.E.16)

$$\mathbf{f}_2 = \mathbf{B}_5\bar{\mathbf{f}}^d_2 + \mathbf{B}_6\mathbf{f}^d_2 + \mathbf{B}_4\mathbf{f}^d_3 + \mathbf{B}_7\mathbf{f}^o_2 + \mathbf{B}_2\mathbf{f}^o_3 + \mathbf{W}_1\mathbf{w}_2, \tag{18.E.18}$$

where

$$\mathbf{B}_5 = \left[\mathbf{0}, \overline{\mathbf{B}}_5\right], \mathbf{B}_7 = \textit{BlockDiag}\left[\cdots, \mathbf{B}^o_{11}, \mathbf{B}^o_{11}, \cdots\right],$$

$$\overline{\mathbf{B}}_5 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \overline{\mathbf{B}}^d_{n1} \\ \overline{\mathbf{B}}^d_{n1} & \ddots & \ddots & & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \overline{\mathbf{B}}^d_{n1} & \mathbf{0} \end{bmatrix}, \mathbf{B}_6 = \begin{bmatrix} \overline{\mathbf{B}}^d_{12} & \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^d_{n2} \\ \mathbf{B}^d_{n2} & \ddots & \ddots & \ddots & & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & & \ddots & \ddots & \ddots & \mathbf{B}^d_{21} \\ \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^d_{n2} & \overline{\mathbf{B}}^d_{12} \end{bmatrix}.$$

Also from (18.E.5)–(18.E.16)

$$\mathbf{f}_3 = \overline{\mathbf{B}}_5\mathbf{f}^d_2 + \mathbf{B}_6\mathbf{f}^d_3 + \mathbf{B}_4\mathbf{f}^d_4 + \mathbf{B}_7\mathbf{f}^o_3 + \mathbf{B}_2\mathbf{f}^o_4 + \mathbf{W}_1\mathbf{w}_3, \tag{18.E.19}$$

$$\mathbf{f}_m = \overline{\mathbf{B}}_5\mathbf{f}^d_{(m-1)} + \mathbf{B}_8\mathbf{f}^d_m + \mathbf{B}_7\mathbf{f}^o_m + \mathbf{W}_1\mathbf{w}_m, \tag{18.E.20}$$

$$\mathbf{B}_8 = \begin{bmatrix} \mathbf{B}^d_{1m} & \mathbf{B}^d_{21} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}^d_{nm} \\ \mathbf{B}_{nm} & \ddots & \ddots & \ddots & & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & & \ddots & \ddots & \ddots & \mathbf{B}^d_{21} \\ \mathbf{B}^d_{21} & \mathbf{0} & & \mathbf{0} & \mathbf{B}^d_{nm} & \mathbf{B}^d_{1m} \end{bmatrix}$$

Or, simply, the vector form of (18.E.17)–(18.E.19) is

$$\mathbf{f} = \mathbf{B}^d\mathbf{f}^d + \mathbf{B}^o\mathbf{f}^o + \mathbf{W}^o\mathbf{w}, \tag{18.E.21}$$

where

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_m \end{bmatrix}, \mathbf{f}^d = \begin{bmatrix} \bar{\mathbf{f}}^d_1 \\ \mathbf{f}^d_2 \\ \vdots \\ \mathbf{f}^d_m \end{bmatrix}, \mathbf{f}^o = \begin{bmatrix} \mathbf{f}^o_1 \\ \vdots \\ \mathbf{f}^o_m \end{bmatrix}, \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix},$$

© 2002 by CRC Press LLC

$$\mathbf{W}^o = BlockDiag\,[\cdots,\ \mathbf{W}_1,\ \mathbf{W}_1, \cdots],$$

$$\mathbf{B}^d = \begin{bmatrix} \mathbf{B}_3 & \mathbf{B}_4 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{B}_5 & \mathbf{B}_6 & \ddots & \ddots & \vdots \\ \mathbf{0} & \overline{\mathbf{B}}_5 & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{B}_6 & \mathbf{B}_4 \\ \mathbf{0} & \cdots & \mathbf{0} & \overline{\mathbf{B}}_5 & \mathbf{B}_8 \end{bmatrix}, \mathbf{B}^o = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_7 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \mathbf{B}_2 \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{B}_7 \end{bmatrix}.$$

References

1. R. E. Skelton, Systems design, the absentee in system theory, *American Automatic Control Conference,* San Diego, CA, June 1999.
2. R. E. Skelton, A systems approach to integrate structure and control design, *SPIE Conference on Mathematics for Smart Structures,* Newport Beach, CA, March 2000.
3. K. M. Grigoriadis, G. Zhu, and R. E. Skelton, Optimal redesign of linear systems, *Journal of Dynamic Systems, Measurement and Control*, 118(3), 598–605, September 1996.
4. M. P. Bendsoe, Optimal shape design as a material distribution problem, *Structural Optimization*, 1, 193–202, 1989.
5. M. P. Bendsoe, *Optimization of Structural Topology, Shape, and Material,* Springer, Berlin, 1995.
6. M. P. Bendsoe and N. Kikuchi, Generating optimal topologies in structural design using a homozenization method, *Computer Methods in Applied Mechanics and Engineering*, 71, 197–224, 1988.
7. A. R. Diaz and M. P. Bendsoe, Shape optimization of structures for multiple loading conditions using a homogenization method, *Structural Optimization*, 4, 17–22, 1992.
8. A. H. Simmons, C. A. Michal, and L. W. Jelinski, Molecular orientation and two-component nature of the crystalline fraction of dragline silk, *Science*, 271, 84–87, 1996.
9. Y. Termonia, Molecular modeling of spider silk elasticity, *Macromolecules,* 27, 7378–7381, 1994.
10. B. I. Yakobson and R.E. Srnalley, Fullerene nanotubes beyond, *American Scientist,* 85, 324–337, 1997.
11. R. Chu, Tensegrity, *Journal of Synergetics,* 2(l), 1988.
12. R. B. Fuller and R. Marks, *The Dymaxion World of Buckminister Fuller*, Anchor Books Edition, Garden City, NY, 1973.
13. R. Buckminister Fuller, Tensile-integrity structures, US Patent 3,063,521, 1962.
14. R. Buckminister Fuller, *Synergetics Explorations in the Geometry of Thinking,* Collier Macmillan Publishers, London, 1975.
15. H. Furuya, Concept of deplorable tensegrity structures in space applications, *International Journal of Space Structures*, 7(2),143–151, 1992.
16. D. E. Ingber, Cellular tensegrity: Defining new rules of biological design that govern the cytoskeleton, *Journal of Cell Science*, 104(3), 613–627, 1993.
17. D. E. Ingber, Tensegrity: The architectural basis of cellular mechanotransduction, *Annual Review of Physiology,* 59, 575–599, 1997.
18. D. E. Ingber, Architecture of life, *Scientific American*, 48–57, January 1998.
19. H. Kenner, *Geodesic Math and How to Use It*, University of California Press, Berkeley, 1976.
20. R. Motro, Tensegrity systems: The state of the art, *International Journal of Space Structures*, 7(2), 75–83, 1992.
21. I. J. Oppenheim and W. J. Williams, Tensegrity prisms as adaptive structures, ASME Annual meeting, Dallas, November 1997.
22. A. Pugh, *An Introduction to Tensegrity*, University of California Press, Berkeley, 1976.
23. R. Adhikari, R. E. Skelton, and W. J. Helton, Mechanics of Tensegrity Beams, UCSD, Structural System and Control Lab., Report No. 1998-1, 1998.

© 2002 by CRC Press LLC

24. R. E. Skelton and C. Sultan, Controllable tensegrity, a new class of smart structures, SPIE *Conference, Mathematics and Control in Smart Structures,* San Diego, March 1997.
25. R. E. Skelton and M. He., Smart tensegrity structure for nestor, *SPIE Conference, Smart Structures and Integrated Systems,* San Diego, CA, March 1997.
26. H. Murakami, Y. Nishimura, T. J. Impesullo, and R. E. Skelton. A virtual reality environment for tensegrity structures, In *Proceedings of the 12th ASCE Engineering Mechanics Conference,* 1998.
27. D. Williamson and R. E. Skelton. A general class of tensegrity systems: Geometric definition, *ASCE Conference Engineering Mechanics for the 21st Century*, pp. 164–169, La Jolla, CA, May 1998.

© 2002 by CRC Press LLC

IV

Robotics

Miomir Vukobratović

© 2002 by CRC Press LLC

19

Robot Kinematics

19.1 Introduction
19.2 Description of Orientation
Rotation Matrix • Unit Quaternion • Euler Angles
19.3 Direct Kinematics
Homogeneous Transformation • Denavit-Hartenberg Convention • Joint Space and Task Space
19.4 Inverse Kinematics
Closed-Form Solutions
19.5 Differential Kinematics
Geometric Jacobian • Analytical Jacobian • Singularities
19.6 Differential Kinematics Inversion
Pseudoinverse • Redundancy • Damped Least-Squares Inverse • User-Defined Accuracy
19.7 Inverse Kinematics Algorithms
Jacobian Pseudoinverse • Jacobian Transpose • Use of Redundancy • Orientation Errors
19.8 Further Reading

Bruno Siciliano
Università degli Studi di Napoli Federico II

19.1 Introduction

From a mechanical viewpoint, a robotic system generally consists of a locomotion apparatus (legs, wheels) to move in the environment and a manipulation apparatus to operate on the objects present. It is then important to distinguish between mobile robots and robot manipulators.

The mechanical structure of a robot manipulator consists of a sequence of links connected by means of joints. Links and joints are usually made as rigid as possible to achieve high precision in robot positioning. The presence of elasticity at the joint transmissions or the use of lightweight materials for the links poses a number of interesting issues that lead to separating the study of flexible robot manipulators from that of rigid robot manipulators. The latter are implicitly meant by the term "robots" throughout this chapter.

This chapter surveys the fundamentals of robot kinematics. Basic mathematical tools such as the rotation matrix, the unit quaternion, and the Euler angles are briefly recalled. They serve to describe the orientation of the robot's end effector that, together with the position can be expressed as a function of the joint variables. This is the direct kinematics equation that is derived through a systematic procedure based on the use of homogeneous transformations and the so-called Denavit-Hartenberg convention. The inverse kinematics problem is considered and closed-form solutions are found for simple geometries. Further, a treatment of differential kinematics based on the robot's Jacobian matrix, hereafter simply called the Jacobian (geometric or analytical) is provided. Specific attention is paid to the occurrence of singularities or redundancy in the context of the differential kinematics inversion. The material ends with the presentation of inverse kinematics algorithms with special emphasis on the definition of the end-effector orientation error; both a pseudoinverse and a transpose of the Jacobian are considered.

© 2002 by CRC Press LLC

19.2 Description of Orientation

Robot manipulation tasks are typically specified in terms of the position and orientation of an end-effector frame with respect to a base frame. Position is uniquely described by the Cartesian coordinates of the origin of the end-effector frame, whereas various representations of orientation exist. Therefore, as a natural prelude to deriving the direct kinematics equation of a robot, some basic concepts about the orientation of a rigid body in space are briefly recalled in the following.

19.2.1 Rotation Matrix

The location of a rigid body in space is typically described in terms of the (3×1) *position vector* $\boldsymbol{p}$ and the (3×3) *rotation matrix* $\boldsymbol{R}$ describing the origin and the orientation of a frame attached to the body with respect to a fixed reference frame, i.e.,

$$\boldsymbol{R} = [\boldsymbol{x} \;\; \boldsymbol{y} \;\; \boldsymbol{z}] \tag{19.1}$$

where $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ are the unit vectors expressing the direction cosines of the axes of the body frame with respect to the reference frame. It is straightforward to verify that the matrix $\boldsymbol{R}$ is orthogonal, meaning that

$$\boldsymbol{R}^{\mathrm{T}} \boldsymbol{R} = \boldsymbol{I} \tag{19.2}$$

thus implying the useful result that the transpose of a rotation matrix is equal to its inverse, i.e., $\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{R}^{-1}$. Frame orientation is conventionally taken to be left-handed.

A rotation matrix possesses three equivalent geometrical meanings:

- It describes the mutual orientation between two coordinate frames (as above).
- It represents the coordinate transformation between the coordinates of a point expressed in two different frames (with common origin).
- It is the operator that allows rotating a vector in the same coordinate frame.

Elementary rotations are those made about one of the coordinate axes,

$$\boldsymbol{R}_X(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{19.3}$$

$$\boldsymbol{R}_Y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \tag{19.4}$$

$$\boldsymbol{R}_Z(\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix} \tag{19.5}$$

which denote the *elementary rotation matrices* with respect to the *X, Y, Z* axes. These are useful to describe rotations about an arbitrary axis in space, as shown below.

Rotation matrices between multiple frames — say frames 0, 1, 2 — can be nicely composed according to the simple rule

$${}^0\boldsymbol{R}_2 = {}^0\boldsymbol{R}_1 {}^1\boldsymbol{R}_2 \tag{19.6}$$

© 2002 by CRC Press LLC

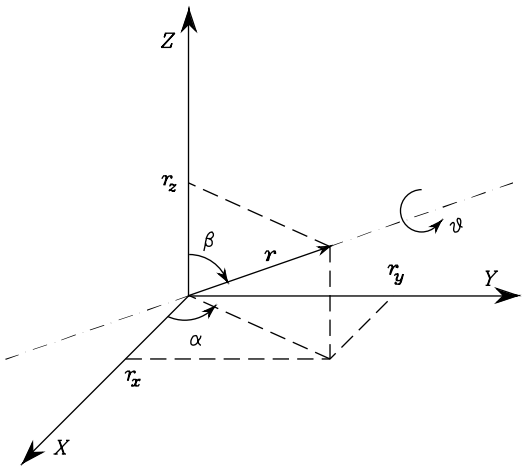


FIGURE 19.1 Rotation of a given angle about an arbitrary axis.

where the notation ${}^{j}\boldsymbol{R}_{i}$ denotes the rotation matrix of frame *i* with respect to frame *j*, and successive rotations are composed with respect to the axes of the current frame. Note also that ${}^{i}\boldsymbol{R}_{j} = ({}^{j}\boldsymbol{R}_{i})^{\mathrm{T}}$.

Expressing a rotation of a given angle about an arbitrary axis in space is often desired. Let $\boldsymbol{r} = [r_x \; r_y \; r_z]^{\mathrm{T}}$ be the unit vector of a rotation axis with respect to the reference frame. To derive the rotation matrix $\boldsymbol{R}(\vartheta, \boldsymbol{r})$ expressing the rotation of an angle ϑ about axis $\boldsymbol{r}$, it is convenient to compose elementary rotations about the coordinate axes of the reference frame. The angle is positive if the rotation is made counter-clockwise about axis $\boldsymbol{r}$.

As shown in Figure 19.1, a possible solution is obtained through the following sequence of rotations:

- Align *Z* with $\boldsymbol{r}$, which is obtained as the sequence of a rotation by α about *Z* and a rotation by β about *Y*;
- Rotate by ϑ about *Z*;
- Realign with the initial direction of *Z*, which is obtained as the sequence of a rotation by $-\beta$ about *Y* and a rotation by $-\alpha$ about *Z*.

The resulting rotation matrix is

$$\boldsymbol{R}(\vartheta, \boldsymbol{r}) = \boldsymbol{R}_Z(\alpha)\, \boldsymbol{R}_Y(\beta)\, \boldsymbol{R}_Z(\vartheta)\, \boldsymbol{R}_Y(-\beta)\, \boldsymbol{R}_Z(-\alpha). \tag{19.7}$$

By using the following relations:

$$\sin\alpha = \frac{r_y}{\sqrt{r_x^2 + r_y^2}} \qquad \cos\alpha = \frac{r_x}{\sqrt{r_x^2 + r_y^2}}$$
$$\sin\beta = \sqrt{r_x^2 + r_y^2} \qquad \cos\beta = r_z,$$

the rotation matrix of the *angle/axis description* in Equation (19.7) can be expressed as

$$\boldsymbol{R}(\vartheta, \boldsymbol{r}) = \begin{bmatrix} r_x^2(1-c_\vartheta) + c_\vartheta & r_x r_y(1-c_\vartheta) - r_z s_\vartheta & r_x r_z(1-c_\vartheta) + r_y s_\vartheta \\ r_x r_y(1-c_\vartheta) + r_z s_\vartheta & r_y^2(1-c_\vartheta) - c_\vartheta & r_y r_z(1-c_\vartheta) - r_x s_\vartheta \\ r_x r_z(1-c_\vartheta) - r_y s_\vartheta & r_y r_z(1-c_\vartheta) - r_x s_\vartheta & r_z^2(1-c_\vartheta) + c_\vartheta \end{bmatrix} \tag{19.8}$$

© 2002 by CRC Press LLC

where the standard abbreviations for $\cos\vartheta$ and $\sin\vartheta$ have been used. Equation (19.8) can be cast in the more compact form

$$\boldsymbol{R}(\vartheta,\boldsymbol{r}) = c_\vartheta \boldsymbol{I} + (1 - c_\vartheta)\,\boldsymbol{r}\boldsymbol{r}^{\mathrm{T}} - s_\vartheta \boldsymbol{S}(\boldsymbol{r}) \tag{19.9}$$

where $\boldsymbol{I}$ is the (3×3) identity matrix and $\boldsymbol{S}(\cdot)$ is the matrix operator performing the cross product between two (3×1) vectors, i.e., $\boldsymbol{S}(\boldsymbol{a})\boldsymbol{b} = \boldsymbol{a} \times \boldsymbol{b}$.

Although the axis can be arbitrary, the three components of $\boldsymbol{r}$ are constrained by the unit norm condition

$$\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} = 1. \tag{19.10}$$

Also, it is clear that $\boldsymbol{R}(-\vartheta, -\boldsymbol{r}) = \boldsymbol{R}(\vartheta, \boldsymbol{r})$, i.e., a rotation by $-\vartheta$ about $-\boldsymbol{r}$ cannot be distinguished from a rotation by ϑ about $\boldsymbol{r}$; hence, for $\vartheta = \pi$ the representation is not unique.

The angle and axis corresponding to a given rotation matrix

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{19.11}$$

are

$$\vartheta = \cos^{-1}\left(\frac{r_{11} + r_{22} + r_{33} - 1}{2}\right)$$

$$\boldsymbol{r} = \frac{1}{2\sin\vartheta}\begin{bmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{bmatrix} \tag{19.12}$$

for $\sin\vartheta \neq 0$. Instead, if $\sin\vartheta = 0$, then it is necessary to refer directly to the particular expressions attained by $\boldsymbol{R}$ and find the solving formulæ in the two cases: if $\vartheta = 0$ the unit vector is arbitrary (no rotation has occurred), while if $\vartheta = \pi$, the above nonuniqueness problem is encountered. This drawback can be overcome by adopting a different four-parameter description, namely, the unit quaternion introduced next.

19.2.2 Unit Quaternion

With reference to the above angle/axis description of orientation, the *unit quaternion* (viz. Euler parameters) is defined as

$$\boldsymbol{Q} = \{\eta, \boldsymbol{\varepsilon}\} \tag{19.13}$$

where

$$\eta = \cos\frac{\vartheta}{2}$$

$$\boldsymbol{\varepsilon} = \sin\frac{\vartheta}{2}\boldsymbol{r}, \tag{19.14}$$

© 2002 by CRC Press LLC

with $\eta \geq 0$ for $\vartheta \in [-\pi, \pi]$; η is called the scalar part of the quaternion while $\boldsymbol{\varepsilon}$ is called the vector part of the quaternion.

The constraint Equation (19.10) transforms into

$$\eta^2 + \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} = 1. \tag{19.15}$$

It is worth remarking that, different than the angle/axis description, a rotation by $-\vartheta$ about $-\boldsymbol{r}$ gives a vector part of the quaternion of the opposite sign from the one associated with a rotation by ϑ about $\boldsymbol{r}$, while the scalar part does not change. This solves the above nonuniqueness problem. The rotation matrix corresponding to a given quaternion is

$$\boldsymbol{R}(\eta, \boldsymbol{\varepsilon}) = (\eta^2 - \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon})\boldsymbol{I} + 2\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}} - 2\eta\boldsymbol{S}(\boldsymbol{\varepsilon}). \tag{19.16}$$

On the other hand, the unit quaternion corresponding to a given rotation matrix Equation (19.11) is

$$\begin{aligned} \eta &= \frac{1}{2}\sqrt{r_{11} + r_{22} + r_{33} + 1} \\ \boldsymbol{\varepsilon} &= \begin{bmatrix} \frac{1}{2}\,\mathrm{sgn}(r_{32} - r_{23})\sqrt{r_{11} - r_{22} - r_{33} + 1} \\ \frac{1}{2}\,\mathrm{sgn}(r_{13} - r_{31})\sqrt{r_{22} - r_{33} - r_{11} + 1} \\ \frac{1}{2}\,\mathrm{sgn}(r_{21} - r_{12})\sqrt{r_{33} - r_{11} - r_{22} + 1} \end{bmatrix}. \end{aligned} \tag{19.17}$$

19.2.3 Euler Angles

Rotation matrices in general give a redundant description of frame orientation; in fact, they are characterized by nine elements that are not independent but are related by six constraints due to the orthogonality conditions in Equation (19.2). Even in the case of describing orientation in terms of rotation about an arbitrary axis or a unit quaternion, a representation in terms of four parameters is obtained. These components are not independent but are constrained by either condition (19.10) or condition (19.15). This implies that there are actually three free parameters to describe orientation.

A minimal representation of orientation can be obtained by using a set of three *Euler angles* $\boldsymbol{\varphi} = [\alpha \;\; \beta \;\; \gamma]^{\mathrm{T}}$. Among the 12 possible definitions of Euler angles, without loss of generality, the *XYZ* representation is considered to lead to the rotation matrix

$$\begin{aligned} \boldsymbol{R}(\boldsymbol{\varphi}) &= \boldsymbol{R}_X(\alpha)\,\boldsymbol{R}_Y(\beta)\,\boldsymbol{R}_Z(\gamma) \\ &= \begin{bmatrix} c_\beta c_\gamma & -c_\beta s_\gamma & s_\beta \\ s_\alpha s_\beta c_\gamma + c_\alpha s_\gamma & -s_\alpha s_\beta s_\gamma + c_\alpha c_\gamma & -s_\alpha c_\beta \\ -c_\alpha s_\beta c_\gamma + s_\alpha s_\gamma & c_\alpha s_\beta s_\gamma + s_\alpha c_\gamma & c_\alpha c_\beta \end{bmatrix} \end{aligned} \tag{19.18}$$

The set of the Euler angles corresponding to a given rotation matrix (19.11) is

$$\begin{aligned} \alpha &= \mathrm{Atan2}(-r_{23}, r_{33}) \\ \beta &= \mathrm{Atan2}\left(r_{13}, \sqrt{r_{11}^2 + r_{12}^2}\right) \\ \gamma &= \mathrm{Atan2}(-r_{12}, r_{11}) \end{aligned} \tag{19.19}$$

© 2002 by CRC Press LLC

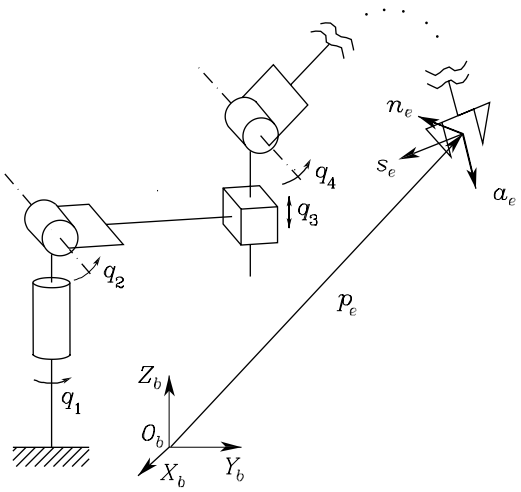


FIGURE 19.2 Schematic of an open-chain robot manipulator with a base frame and end-effector frame.

with $\beta \in (-\pi/2, \pi/2)$, whereas the solution is

$$\alpha = \text{Atan2}(r_{23}, -r_{33})$$
$$\beta = \text{Atan2}\left(r_{13}, -\sqrt{r_{11}^2 + r_{12}^2}\right) \tag{19.20}$$
$$\gamma = \text{Atan2}(r_{12}, -r_{11})$$

with $\beta \in (\pi/2, 3\pi/2)$; the function Atan2($y$, x) computes the arctangent of the ratio y/x but utilizes the sign of each argument to determine to which quadrant the resulting angle belongs.

Solutions (19.19) and (19.20) degenerate when $\beta = \pm\pi/2$; in this case, it is possible to determine only the sum or difference of α and γ, i.e.,

$$\alpha \pm \gamma = \text{Atan2}(r_{21},\ r_{22}) \tag{19.21}$$

where the plus sign applies for $\beta = +\pi/2$ and the minus sign applies for $\beta = -\pi/2$.

19.3 Direct Kinematics

A robot manipulator consists of a kinematic chain of $n + 1$ links connected by means of n joints. Joints can essentially be of two types: *revolute* and *prismatic*; complex joints can be decomposed into these simple joints. Revolute joints are usually preferred because of their compactness and reliability. One end of the chain is connected to the base link to which a suitable base frame is attached, whereas an *end-effector* is connected to the other end and a suitable end-effector frame is attached. The basic structure of a robot is the open kinematic chain that occurs when only one sequence of links connects the two ends of the chain. Alternatively, a robot contains a closed kinematic chain when a sequence of links forms a loop. In Figure 19.2, an open-chain robot manipulator is illustrated with conventional representation of revolute and prismatic joints.

Direct kinematics of a robot consist of determining the mapping between the joint variables and the position and orientation of the end-effector frame with respect to the base frame.

© 2002 by CRC Press LLC

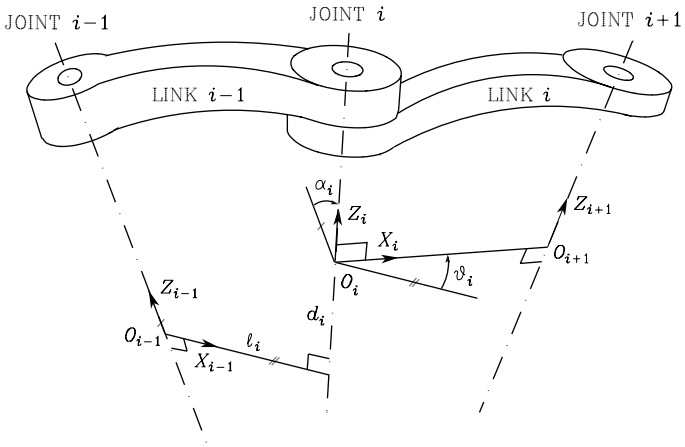


FIGURE 19.3 Kinematic parameters with modified Denavit-Hartenberg convention.

19.3.1 Homogeneous Transformation

As discussed above, the position of a rigid body in space is expressed in terms of the position of a suitable point on the body with respect to a reference frame (translation), while its orientation is expressed in terms of the components of the unit vectors of a frame attached to the body (with origin in the above point) with respect to the same reference frame (rotation).

The complete *coordinate transformation* between two frames (say frames 0, 1) is given by composing the translation ${}^0\boldsymbol{p}_1$ between the origins of the frames and the rotation ${}^0\boldsymbol{R}_1$ between the axes of the frames into a (4×4) homogeneous transformation matrix

$$ {}^0\boldsymbol{T}_1 = \begin{bmatrix} & {}^0\boldsymbol{R}_1 & & {}^0\boldsymbol{p}_1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{19.22} $$

Similar to the composition of rotations expressed by (19.6), a sequence of coordinate transformations from frame 0 to frame n can be composed as in the product

$$ {}^0\boldsymbol{T}_n = {}^0\boldsymbol{T}_1 \, {}^1\boldsymbol{T}_2 \ldots {}^{n-1}\boldsymbol{T}_n \tag{19.23} $$

where ${}^{i-1}\boldsymbol{T}_i$ denotes the homogeneous transformation expressing the position and orientation of frame i with respect to frame $i-1$. The relationship (19.23) is the basic tool for deriving the direct kinematics equation of a robot.

19.3.2 Denavit-Hartenberg Convention

An effective procedure for computing the direct kinematics function for a general robot is based on the so-called modified *Denavit-Hartenberg* convention. According to this convention, a coordinate frame is attached to each link of the chain and the overall transformation matrix from link 0 to link n is derived by composition of transformations between consecutive frames. With reference to Figure 19.3, let joint i connect link $i-1$ to link i, where the links are assumed to be rigid; frame i is attached to link i and can be defined as follows:

- Choose axis Z_i aligned with the axis of joint i.
- Choose axis X_i along the common normal to axes Z_i and Z_{i+1} with direction from joint i to joint $i+1$.
- Choose axis Y_i to complete a right-handed frame.

© 2002 by CRC Press LLC

Once the link frames have been established, the position and orientation of frame i with respect to frame $i-1$ are completely specified by the following *kinematic parameters*.

α_i Angle between Z_{i-1} and Z_i about X_{i-1} measured counter-clockwise
ℓ_i Distance between Z_{i-1} and Z_i along X_{i-1}
ϑ_i Angle between X_{i-1} and X_i about Z_i measured counter-clockwise
d_i Distance between X_{i-1} and X_i along Z_i

Let $\mathbf{Rot}(K,\delta)$ $\left(\mathbf{Trans}(K,\delta)\right)$ denote the homogeneous transformation matrix expressing the rotation (translation) about (along) axis K by an angle (distance) δ. Then, the coordinate transformation of frame i with respect to frame $i-1$ can be expressed in terms of the above four parameters by the matrix

$$
\begin{aligned}
{}^{i-1}\boldsymbol{T}_i &= \mathbf{Rot}(X,\alpha_i)\mathbf{Trans}(X,\ell_i)\mathbf{Rot}(Z,\vartheta_i)\mathbf{Trans}(Z,d_i) \\
&= \begin{bmatrix}
\cos\vartheta_i & -\sin\vartheta_i & 0 & \ell_i \\
\cos\alpha_i\sin\vartheta_i & \cos\alpha_i\cos\vartheta_i & -\sin\alpha_i & -d_i\sin\alpha_i \\
\sin\alpha_i\sin\vartheta_i & \sin\alpha_i\cos\vartheta_i & \cos\alpha_i & d_i\cos\alpha_i \\
0 & 0 & 0 & 1
\end{bmatrix} \\
&= \begin{bmatrix}
 & {}^{i-1}\boldsymbol{R}_i & & {}^{i-1}\boldsymbol{p}_i \\
0 & 0 & 0 & 1
\end{bmatrix}
\end{aligned}
\tag{19.24}
$$

where ${}^{i-1}\boldsymbol{R}_i$ is the (3×3) matrix defining the orientation of frame i with respect to frame $i-1$, and ${}^{i-1}\boldsymbol{p}_i$ is the (3×1) vector defining the origin of frame i with respect to frame $i-1$.

Dually, the transformation matrix defining frame $i-1$ with respect to frame i is given by

$$
\begin{aligned}
{}^{i}\boldsymbol{T}_{i-1} &= \mathbf{Trans}(Z,-d_i)\mathbf{Rot}(Z,-\vartheta_i)\mathbf{Trans}(X,-\ell_i)\mathbf{Rot}(X,-\alpha_i) \\
&= \begin{bmatrix}
 & & & -\ell_i\cos\vartheta_i \\
 & {}^{i-1}\boldsymbol{R}_i^{\mathrm{T}} & & \ell_i\sin\vartheta_i \\
 & & & -d_i \\
0 & 0 & 0 & 1
\end{bmatrix}
\end{aligned}
\tag{19.25}
$$

Two of the four parameters (ℓ_i and α_i) are always constant and depend only on the size and shape of link i. Of the remaining two parameters, only one is variable (degree of freedom) depending on the type of joint that connects link $i-1$ to link i. If q_i denotes the joint i variable, then it is

$$
q_i = \bar{\xi}_i\vartheta_i + \xi_i d_i \tag{19.26}
$$

where $\bar{\xi}_i = 1-\xi_i$, i.e.,

- $\xi_i = 0$ if joint i is revolute ($q_i = \vartheta_i$),
- $\xi_i = 1$ if joint i is prismatic ($q_i = d_i$).

In view of (19.26), the equation

$$
\bar{q}_i = \xi_i\vartheta_i + \bar{\xi}_i d_i \tag{19.27}
$$

© 2002 by CRC Press LLC

gives the constant parameter at each joint to add to α_i and ℓ_i.

The above procedure does not yield a unique definition of frames 0 and n that can be chosen arbitrarily. Also, in all cases of nonuniqueness in the definition of the frames, it is convenient to make as many link parameters zero as possible, because this will simplify kinematics computation. A number of remarks are in order.

- A simple choice to define frame 0 is to take it coincident with frame 1 when $q_1 = 0$; this makes $\alpha_1 = 0$ and $\ell_1 = 0$, and $\bar{q}_1 = 0$.
- A similar choice for frame n is to take X_n along X_{n-1} when $q_n = 0$; this makes $\bar{q}_n = 0$.
- If joint i is prismatic, the direction of Z_i is fixed while its location is arbitrary; it is convenient to locate Z_i either at the origin of frame $i - 1$ ($\ell_i = 0$) or at the origin of frame $i + 1$ ($\ell_{i+1} = 0$).
- When the joint axes i and $i + 1$ are parallel, it is convenient to locate X_i to achieve either $d_i = 0$ or $d_{i+1} = 0$ if either joint is revolute.

In view of (19.23), through the composition of the individual link transformations, the coordinate transformation describing the position and orientation of frame n with respect to frame 0 is given by

$$ {}^0\boldsymbol{T}_n(\boldsymbol{q}) = {}^0\boldsymbol{T}_1(q_1)\,{}^1\boldsymbol{T}_2(q_2)\cdots{}^{n-1}\boldsymbol{T}_n(q_n), \tag{19.28} $$

where $\boldsymbol{q}$ denotes the $(n \times 1)$ vector of joint variables. To derive the direct kinematics, two further *constant* transformations have to be introduced; namely, the transformation from the base frame b to frame 0 ($^bT_0$) and the transformation from frame n to the end-effector frame e (^nT_e), i.e.,

$$ \begin{aligned} {}^b\boldsymbol{T}_e(\boldsymbol{q}) &= {}^b\boldsymbol{T}_0\,{}^0\boldsymbol{T}_n(\boldsymbol{q})\,{}^n\boldsymbol{T}_e \\ &= \begin{bmatrix} {}^b\boldsymbol{n}_e(\boldsymbol{q}) & {}^b\boldsymbol{s}_e(\boldsymbol{q}) & {}^b\boldsymbol{a}_e(\boldsymbol{q}) & {}^b\boldsymbol{p}_e(\boldsymbol{q}) \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{19.29} $$

where the normal, sliding, and approach unit vectors $\boldsymbol{n}$, $\boldsymbol{s}$, $\boldsymbol{a}$ have been formally introduced (Figure 19.2). Subscripts and superscripts can be omitted when the relevant frames are clear from the context.

The "modified" Denavit-Hartenberg convention stems from the fact that, in the "classical" convention, axis Z_i is aligned with the axis of joint $i + 1$ and the kinematic parameters differ accordingly.

An example of an open-chain robot is the anthropomorphic robot.

With reference to the frames illustrated in Figure 19.4, the Denavit-Hartenberg parameters are specified in Table 19.1.

Computing the transformation matrices in (19.24) and composing them as in (19.28) gives

$$ {}^0\boldsymbol{T}_6 = \begin{pmatrix} {}^0\boldsymbol{n}_6 & {}^0\boldsymbol{s}_6 & {}^0\boldsymbol{a}_6 & {}^0\boldsymbol{p}_6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{19.30} $$

where

$$ {}^0\boldsymbol{p}_6 = \begin{bmatrix} c_1\left(c_2\,\ell_3 - s_{23}d_4\right) \\ s_1\left(c_2\,\ell_3 - s_{23}d_4\right) \\ s_2\,\ell_3 + c_{23}d_4 \end{bmatrix} \tag{19.31} $$

© 2002 by CRC Press LLC

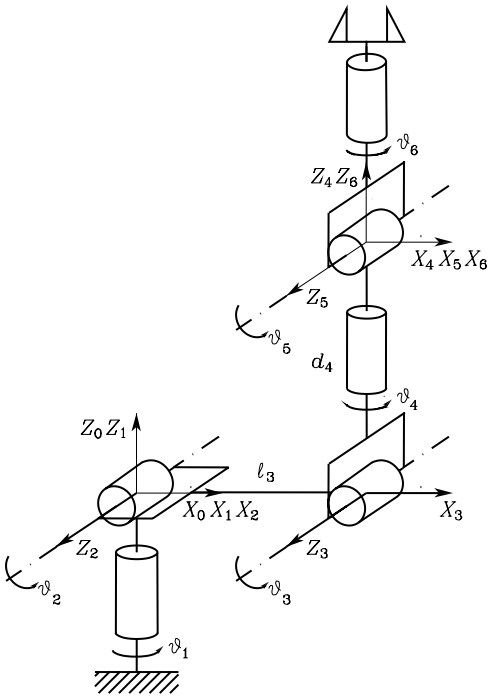


FIGURE 19.4 Anthropomorphic robot with frame assignment.

TABLE 19.1 Denavit-Hartenberg Parameters of the Anthropomorphic Robot

i	α_i	ℓ_i	ϑ_i	d_i
1	0	0	q_1	0
2	$\pi/2$	0	q_2	0
3	0	ℓ_3	q_3	0
4	$-\pi/2$	0	q_4	d_4
5	$\pi/2$	0	q_5	0
6	$-\pi/2$	0	q_6	0

for the position, and

$$
{}^0\boldsymbol{n}_6 = \begin{bmatrix} c_1\left(c_{23}\left(c_4c_5c_6 - s_4s_6\right) - s_{23}s_5c_6\right) - s_1\left(s_4c_5c_6 + c_4s_6\right) \\ s_1\left(c_{23}\left(c_4c_5c_6 - s_4s_6\right) - s_{23}s_5c_6\right) + c_1\left(s_4c_5c_6 + c_4s_6\right) \\ s_{23}\left(c_4c_5c_6 - s_4s_6\right) + c_{23}s_5c_6 \end{bmatrix} \tag{19.32}
$$

© 2002 by CRC Press LLC

$$
{}^0\boldsymbol{s}_6 = \begin{bmatrix} c_1\left(-c_{23}\left(c_4c_5s_6 + s_4c_6\right) + s_{23}s_5s_6\right) + s_1\left(s_4c_5s_6 - c_4c_6\right) \\ s_1\left(-c_{23}\left(c_4c_5s_6 + s_4c_6\right) + s_{23}s_5s_6\right) - c_1\left(s_4c_5s_6 - c_4c_6\right) \\ -s_{23}\left(c_4c_5s_6 + s_4s_6\right) - c_{23}s_5s_6 \end{bmatrix} \tag{19.33}
$$

$$
{}^0\boldsymbol{a}_6 = \begin{bmatrix} -c_1\left(c_{23}c_4s_5 + s_{23}c_5\right) + s_1s_4s_5 \\ -s_1\left(c_{23}c_4s_5 + s_{23}c_5\right) - c_1s_4s_5 \\ -s_{23}c_4s_5 + c_{23}c_5 \end{bmatrix} \tag{19.34}
$$

for the orientation, where $c_i = \cos\vartheta_i$, $s_i = \sin\vartheta_i$, $c_{23} = \cos\left(\vartheta_2 + \vartheta_3\right)$, and $s_{23} = \sin\left(\vartheta_2 + \vartheta_3\right)$.

19.3.3 Joint Space and Task Space

If a task has to be assigned to the end-effector, it is necessary to specify both the end-effector's position and orientation. This is easy for the position $\boldsymbol{p}_e$. However, specifying the orientation through the unit vector triple ($\boldsymbol{n}_e$, $\boldsymbol{s}_e$, $\boldsymbol{a}_e$) is difficult, because their nine components must be guaranteed to satisfy the orthonormality constraints imposed by (19.2). Even with a four-parameter description of the orientation, one constraint in the form of either (19.10) or (19.15) should be satisfied.

On the other hand, if a minimal representation is adopted in terms of the Euler angles describing the orientation of the end-effector frame with respect to the base frame, a suitable ($m \times 1$) vector can be considered as

$$
\boldsymbol{x} = \begin{bmatrix} \boldsymbol{p}_e \\ \boldsymbol{\varphi}_e \end{bmatrix}, \tag{19.35}
$$

where $\boldsymbol{p}_e$ describes the end-effector position and $\boldsymbol{\varphi}_e$ its orientation. This representation of position and orientation allows the description of the end-effector task in terms of a number of inherently independent parameters. The vector $\boldsymbol{x}$ is defined in the space in which the robot task is specified; hence, this space is typically called *task space* (operational space). The dimension of the task space is at most $m = 6$, because three coordinates specify position and three angles specify orientation. Nevertheless, depending on the geometry of the task, a reduced number of task space variables may be specified; for instance, for a planar robot it is $m = 3$, because two coordinates specify position and one angle specifies orientation.

On the other hand, the *joint space* (configuration space) denotes the space in which the ($n \times 1$) vector of joint variables $\boldsymbol{q}$ is defined. Taking into account the dependence of position and orientation from the joint variables, the direct kinematics equation can be written in a form other than (19.24), i.e.,

$$
\boldsymbol{x} = \boldsymbol{k}(\boldsymbol{q}). \tag{19.36}
$$

It is worth noticing that the explicit dependence of the function $\boldsymbol{k}(\boldsymbol{q})$ from the joint variables for the orientation components is not available except for simple cases. In fact, on the most general assumption of a six-dimensional task space ($m = 6$), the computation of the three components of the function $\boldsymbol{\varphi}_e(q)$ cannot be performed in closed form but goes through the computation of the elements of the rotation matrix.

© 2002 by CRC Press LLC

The notion of joint space and task space naturally allows introducing the concept of *kinematic redundancy.* This occurs when the dimension of the task space is smaller than the dimension of the joint space ($m < n$). Redundancy is a concept *relative* to the task assigned to the robot; a robot can be redundant with respect to a task and nonredundant with respect to another, depending on the number of task space variables of interest.

For instance, a three-degree-of-freedom planar robot becomes redundant if end-effector orientation is of no concern ($m = 2$, $n = 3$). Yet, the typical example of redundant robot is the human arm that has seven degrees of freedom: three in the shoulder, one in the elbow, and three in the wrist, without considering the degrees of freedom in the fingers ($m = 6$, $n = 7$).

19.4 Inverse Kinematics

The direct kinematics equation, either in the form (19.24) or in the form (19.36), establishes the functional relationship between the joint variables and the end-effector position and orientation. *Inverse kinematics* concerns the determination of the joint variables $\boldsymbol{q}$ corresponding to a given end-effector position $\boldsymbol{p}_e$ and orientation $\boldsymbol{R}_e$. The solution to this problem is of fundamental importance in order to translate the specified motion, naturally assigned in the task space, into the equivalent joint space motion that allows execution of the desired task.

With regard to the direct kinematics Equation (19.24), the end-effector position and rotation matrix are uniquely computed, once the joint variables are known. In general, this cannot be said for Equation (19.36), because the Euler angles are not uniquely defined. On the other hand, the inverse kinematics problem is much more complex for the following reasons.

- The equations to solve are general nonlinear equations for which it is not always possible to find closed-form solutions.
- Multiple solutions may exist.
- Infinite solutions may exist, e.g., in the case of a kinematically redundant robot.
- There might not be admissible solutions, in view of the robot kinematic structure.

Of course, the existence of solutions is guaranteed if the given end-effector position and orientation belong to the robot workspace.

On the other hand, the problem of multiple solutions depends not only on the number of degrees of freedom but also on the Denavit-Hartenberg parameters; in general, the greater the number of nonnull parameters, the greater the number of admissible solutions. For a six-degrees-of-freedom robot without mechanical joint limits, in general up to 16 admissible solutions exist. This occurrence demands some criteria to choose among admissible solutions.

The computation of closed-form solutions requires either algebraic intuition to find those significant equations containing the unknowns, or geometric intuition to discover those significant points on the structure for which it is convenient to express position and orientation. Or, in all those cases when there are no — or it is difficult to find — *closed-form solutions,* it might be appropriate to resort to numerical solution techniques. These clearly have the advantage of being applicable to any kinematic structure, but generally they do not allow computation of all admissible solutions.

19.4.1 Closed-Form Solutions

Most of the existing robots are kinematically simple, because they are typically formed by an arm (three or more degrees of freedom) which provides mobility and by a wrist which provides dexterity (three degrees of freedom). This choice is partially motivated by the difficulty of finding solutions to the inverse kinematics problem in the general case. In particular, a six-degrees-of-freedom robot has closed-form inverse kinematics solutions if three consecutive revolute joint axes intersect at a common point. This situation occurs when a robot has a so-called *spherical wrist* that is characterized by

© 2002 by CRC Press LLC

$$\ell_5 = d_5 = \ell_6 = 0 \qquad \xi_4 = \xi_5 = \xi_6 = 0, \tag{19.37}$$

with $\sin\alpha_5 \neq 0$ and $\sin\alpha_6 \neq 0$ so as to avoid parallel axes (degenerate robot). In that case, it is possible to divide the inverse kinematics problem into two subproblems, because the solution for the *position* is *decoupled* from that for the *orientation*.

In the case of a three-degrees-of-freedom arm, for given end-effector position ${}^0\boldsymbol{p}_e$ and orientation ${}^0\boldsymbol{R}_e$, the inverse kinematics can be solved according to the following steps:

- Compute the wrist position ${}^0\boldsymbol{p}_4$ from ${}^0\boldsymbol{p}_e$;
- Solve inverse kinematics for (q_1, q_2, q_3);
- Compute ${}^0\boldsymbol{R}_3\ (q_1, q_2, q_3)$;
- Compute ${}^3\boldsymbol{R}_6(q_4, q_5, q_6) = {}^3\boldsymbol{R}_0\ {}^0\boldsymbol{R}_e\ {}^e\boldsymbol{R}_6$;
- Solve inverse kinematics for (q_4, q_5, q_6).

Therefore, on the basis of this kinematic decoupling, it is possible to solve the inverse kinematics for the arm separately from the inverse kinematics for the spherical wrist.

Consider the anthropomorphic robot in Figure 19.4, whose direct kinematics was given in (19.30). Finding the vector of joint variables $\boldsymbol{q}$ corresponding to given end-effector position ${}^0\boldsymbol{p}_e$ and orientation ${}^0\boldsymbol{R}_e$ is desired; without loss of generality, assume that ${}^0\boldsymbol{p}_e = {}^0\boldsymbol{p}_6$ and ${}^6\boldsymbol{R}_e = \boldsymbol{I}$.

Observing that ${}^0\boldsymbol{p}_6 = {}^0\boldsymbol{p}_4$, the first three joint variables can be solved from (19.31) which can be rewritten as

$$\begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} c_1\left(c_2\,\ell_3 - s_{23}d_4\right) \\ s_1\left(c_2\,\ell_3 - s_{23}d_4\right) \\ s_2\,\ell_3 + c_{23}d_4 \end{bmatrix}. \tag{19.38}$$

From the first two components of (19.38), it is

$$q_1 = \text{Atan}\,2\left(p_{y,}p_x\right). \tag{19.39}$$

Notice that another solution is

$$q_1 = \pi + \text{Atan}\,2\left(p_{y,}p_z\right). \tag{19.40}$$

The second joint variable can be found by squaring and summing the first two components of (19.38), i.e.,

$$p_x^2 + p_y^2 = \left(c_2\,\ell_3 - s_{23}d_4\right)^2; \tag{19.41}$$

then, squaring the third component and summing it to (19.41) lead to the solution

$$q_3 = \text{Atan}\,2\left(s_3, c_3\right) \tag{19.42}$$

where

$$s_3 = \frac{\ell_3^2 + d_4^2 - p_x^2 - p_y^2 - p_z^2}{2\,\ell_3 d_4} \qquad c_3 = \pm\sqrt{1 - s_3^2}.$$

© 2002 by CRC Press LLC

Substituting q_3 in (19.41), taking the square root thereof and combining the result with the third component of (19.38) lead to a system of equations in the unknowns s_2 and c_2; its solution can be found as

$$s_2 = \frac{\left(\ell_3 - s_3 d_4\right) p_z - c_3 d_4 \sqrt{p_x^2 + p_y^2}}{p_x^2 + p_y^2 + p_z^2}$$

$$c_2 = \frac{\left(\ell_3 - s_3 d_4\right) \sqrt{p_x^2 + p_y^2} + c_3 d_4 p_z}{p_x^2 + p_y^2 + p_z^2},$$

and thus the second joint variable is

$$q_2 = \mathrm{Atan}\,2\left(s_2, c_2\right). \tag{19.43}$$

Notice that four admissible solutions are obtained according to the values of q_1, q_2, q_3, namely, shoulder-right/elbow-up, shoulder-left/elbow-up, shoulder-right/elbow-down, shoulder-left/elbow-down.

To solve for the three joint variables of the wrist, the following procedure can be used. Given the matrix

$${}^0\boldsymbol{R}_6 = \begin{pmatrix} n_x & s_x & a_x \\ n_y & s_y & a_y \\ n_z & s_z & a_z \end{pmatrix}, \tag{19.44}$$

the matrix ${}^0\boldsymbol{R}_3$ can be computed from the first three joint variables via (19.24), and thus the following equation is to be considered:

$$\begin{bmatrix} {}^3n_x & {}^3s_x & {}^3a_x \\ {}^3n_y & {}^3s_y & {}^3a_y \\ {}^3n_z & {}^3s_z & {}^3a_z \end{bmatrix} = \begin{bmatrix} c_4c_5c_6 - s_4s_6 & -c_4c_5s_6 - s_4c_6 & -c_4s_5 \\ s_5c_6 & -s_5s_6 & c_5 \\ -s_4c_5c_6 - c_4s_6 & s_4c_5s_6 - c_4c_6 & s_4s_5 \end{bmatrix}. \tag{19.45}$$

The elements of the matrix on the right-hand side of (19.45) have been obtained by computing ${}^3\boldsymbol{R}_6$ via (19.24), whereas the elements of the matrix on the left-hand side of (19.45) can be computed as ${}^3\boldsymbol{R}_0\ {}^0\boldsymbol{R}_6$ with ${}^0\boldsymbol{R}_6$ as in (19.44), i.e.,

$$\begin{aligned} {}^3n_x &= c_{23}\left(c_1n_x + s_1n_y\right) + s_{23}n_z \\ {}^3n_y &= -s_{23}\left(c_1n_x + s_1n_y\right) + s_{23}n_z \\ {}^3n_z &= s_1n_x - c_1n_y; \end{aligned} \tag{19.46}$$

the other elements $\left({}^3s_x, {}^3s_y, {}^3s_z\right)$ and $\left({}^3a_x, {}^3a_y, {}^3a_z\right)$ can be computed from (19.46) by replacing $\left(n_x, n_y, n_z\right)$ with $\left(s_x, s_y, s_z\right)$ and $\left(a_x, a_y, a_z\right)$, respectively.

At this point, inspecting (19.45) reveals that from the elements [1, 3] and [3, 3], q_4 can be computed as

© 2002 by CRC Press LLC

$$q_4 = \text{Atan2}\left({}^3a_z, -{}^3a_x\right). \tag{19.47}$$

Then, q_5 can be computed by squaring and summing the elements [1, 3] and [3, 3], and from the element [2, 3] as

$$q_5 = \text{Atan2}\left(\sqrt{\left({}^3a_x\right)^2 + \left({}^3a_z\right)^2}, {}^3a_y\right). \tag{19.48}$$

Finally, q_6 can be computed from the elements [2, 1] and [2, 2] as

$$q_6 = \text{Atan2}\left(-{}^3s_y, {}^3n_y\right). \tag{19.49}$$

It is worth noticing that another set of solutions is given by the triplet

$$q_4 = \text{Atan2}\left(-{}^3a_z, {}^3a_x\right) \tag{19.50}$$

$$q_5 = \text{Atan2}\left(-\sqrt{\left({}^3a_x\right)^2 + \left({}^3a_z\right)^2}, {}^3a_y\right) \tag{19.51}$$

$$q_6 = \text{Atan2}\left({}^3s_y, -{}^3n_y\right). \tag{19.52}$$

Notice that both sets of solutions degenerate when ${}^3a_x = {}^3a_z = 0$; in this case, q_4 is arbitrary and simpler expressions can be found for q_5 and q_6.

In conclusion, four admissible solutions have been found for the arm and two admissible solutions have been found for the wrist, resulting in a total of eight admissible inverse kinematics solutions for the anthropomorphic robot with a spherical wrist.

19.5 Differential Kinematics

The (3×1) vector $\dot{\boldsymbol{p}}$ of linear velocity of a rigid body in space is given by the time derivative of the position vector, while the (3×1) vector ω of angular velocity can be defined through the time derivative of the rotation matrix in the form

$$\dot{\boldsymbol{R}} = \boldsymbol{S}(\boldsymbol{\omega})\boldsymbol{R}. \tag{19.53}$$

With reference to the other descriptions of orientation, the relationship between the angular velocity and the time derivative of the unit quaternion is

$$\begin{bmatrix} \dot{\eta} \\ \dot{\boldsymbol{\varepsilon}} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & -\boldsymbol{\omega}^{\mathrm{T}} \\ \boldsymbol{\omega} & -\boldsymbol{S}(\boldsymbol{\omega}) \end{bmatrix}\begin{bmatrix} \eta \\ \boldsymbol{\varepsilon} \end{bmatrix} \tag{19.54}$$

which is known as the quaternion propagation rule, whereas that between the angular velocity and the time derivative of the Euler angles is

$$\boldsymbol{\omega} = \boldsymbol{T}(\boldsymbol{\varphi})\dot{\boldsymbol{\varphi}} \tag{19.55}$$

© 2002 by CRC Press LLC

where $\boldsymbol{T}(\boldsymbol{\varphi})$ depends on the particular choice of Euler angles.

The mapping between the $(n \times 1)$ vector of joint velocities $\dot{\boldsymbol{q}}$ and the (6×1) vector of end-effector (linear and angular) velocities $\boldsymbol{v}$ is established by the differential kinematics equation

$$\boldsymbol{v} = \begin{bmatrix} \dot{\boldsymbol{p}} \\ \boldsymbol{\omega} \end{bmatrix} = \boldsymbol{J}(\boldsymbol{q})\dot{\boldsymbol{q}}, \tag{19.56}$$

where $\boldsymbol{J}(\boldsymbol{q})$ is the $(6 \times n)$ Jacobian matrix. The computation of this matrix usually follows a geometric procedure that is based on computing the contributions of each joint velocity to the linear and angular end-effector velocities. Hence, $\boldsymbol{J}(\boldsymbol{q})$ can be termed the geometric Jacobian of the robot.

19.5.1 Geometric Jacobian

In view of simple geometry, the velocity contributions of each joint to the linear and angular velocities of link n give the following relationship:

$$\begin{bmatrix} \dot{p}_n \\ \boldsymbol{\omega}_n \end{bmatrix} = \begin{bmatrix} \xi_1 z_1 + \bar{\xi}_1 (z_1 \times p_{1n}) & \cdots & \xi_n z_n + \bar{\xi}_n (z_n \times p_{nn}) \\ \bar{\xi}_1 z_1 & \cdots & \bar{\xi}_n z_n \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \vdots \\ \dot{q}_n \end{bmatrix} = J_n(q)\dot{q} \tag{19.57}$$

where z_k is the unit vector of axis Z_k and $\boldsymbol{p}_{kn}$ denotes the vector from the origin of frame k to the origin of frame n. Notice that $\boldsymbol{J}_n$ is a function of $\boldsymbol{q}$ through the vectors z_k and $\boldsymbol{p}_{kn}$ that can be computed on the basis of direct kinematics.

The geometric Jacobian can be computed with respect to any frame i; in that case, the k-th column of ${}^{i}\boldsymbol{J}_n$ is given by

$${}^{i}J_{nk} + \begin{bmatrix} \xi_k {}^{i}z_k + \bar{\xi}_k {}^{i}R_k S({}^{k}z_k) {}^{k}p_n \\ \bar{\xi}_k {}^{i}z_k \end{bmatrix} \tag{19.58}$$

where ${}^{k}\boldsymbol{p}_n = {}^{k}\boldsymbol{p}_{kn}$. In view of the expression of ${}^{k}z_k = [0\ 0\ 1]$, Equation (19.58) can be rewritten as

$${}^{i}\boldsymbol{J}_{nk} = \begin{bmatrix} \xi_k {}^{i}\boldsymbol{z}_k + \bar{\xi}_k (-{}^{k}p_{ny} {}^{i}\boldsymbol{x}_k + {}^{k}p_{nx} {}^{i}\boldsymbol{y}_k) \\ \bar{\xi}_k {}^{i}\boldsymbol{z}_k \end{bmatrix} \tag{19.59}$$

where ${}^{k}p_{nx}$ and ${}^{k}p_{ny}$ are the x and y components of ${}^{k}\boldsymbol{p}_n$. A number of remarks are in order.

- The transformation of the Jacobian from frame i to a different frame l can be obtained as

$${}^{l}\boldsymbol{J}_n = \begin{bmatrix} {}^{l}\boldsymbol{R}_i & 0 \\ 0 & {}^{l}\boldsymbol{R}_i \end{bmatrix} {}^{i}\boldsymbol{J}_n. \tag{19.60}$$

- The Jacobian relating the end-effector velocity to the joint velocities can be computed either by using (19.57) and replacing $\boldsymbol{p}_{kn}$ with $\boldsymbol{p}_{ke}$, or by using the relationship

© 2002 by CRC Press LLC

$$ {}^{i}\boldsymbol{J}_{e} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{S}\left({}^{i}\boldsymbol{p}_{ne}\right) \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix} {}^{i}\boldsymbol{J}_{n}. \tag{19.61} $$

A Jacobian ${}^{i}\boldsymbol{J}_{n}$ can be decomposed as the product of three matrices, where the first two are full-rank, while the third one has the same rank as ${}^{i}\boldsymbol{J}_{n}$ but contains simpler elements to compute. To achieve this, the Jacobian of link n can be expressed as a function of a generic Jacobian

$$ J_{n,h} = \begin{bmatrix} \xi_1 z_1 + \bar{\xi}_1 \left(z_1 \times p_{1h}\right) & \cdots & \xi_n z_n + \bar{\xi}_n \left(z_n \times p_{nh}\right) \\ \bar{\xi}_1 z_1 & \cdots & \bar{\xi}_n z_n \end{bmatrix} \tag{19.62} $$

giving the velocity of a frame fixed to link n attached instantaneously to frame h. Then $\boldsymbol{J}_{\mathrm{n}}$ can be computed via (19.61) as

$$ \boldsymbol{J}_{n} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{S}\left(\boldsymbol{p}_{hn}\right) \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix} \boldsymbol{J}_{n,h} \tag{19.63} $$

which can be expressed with respect to frame i, giving

$$ {}^{i}\boldsymbol{J}_{n} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{S}\left({}^{i}\boldsymbol{R}_{h}\,{}^{h}\boldsymbol{p}_{n}\right) \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix} {}^{i}\boldsymbol{J}_{n,h}. \tag{19.64} $$

Combining (19.60) with (19.64) yields the result that the matrix ${}^{l}\boldsymbol{J}_{n}$ can be computed as the product of three matrices

$$ {}^{l}\boldsymbol{J}_{n} = \begin{bmatrix} {}^{l}\boldsymbol{R}_{i} & \boldsymbol{O} \\ \boldsymbol{O} & {}^{l}\boldsymbol{R}_{i} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{S}\left({}^{i}\boldsymbol{R}_{h}\,{}^{h}\boldsymbol{p}_{n}\right) \\ \boldsymbol{O} & \boldsymbol{I} \end{bmatrix} {}^{i}\boldsymbol{J}_{n,h}, \tag{19.65} $$

where remarkably the first two matrices are full-rank. In general, the values of h and i leading to the Jacobian ${}^{i}\boldsymbol{J}_{n,h}$ of simplest expression are given by

$$ i = \mathrm{int}(n/2) \qquad h = \mathrm{int}(n/2) + 1. $$

Hence, for a robot with six degrees of freedom, the matrix ${}^{3}\boldsymbol{J}_{6,4}$ is expected to have the simplest expression; if the wrist is spherical ($\boldsymbol{p}_{46} = 0$), then the second matrix in (19.65) is identity and ${}^{3}\boldsymbol{J}_{6,4} = {}^{3}\boldsymbol{J}_{6}$.

As an example, the geometric Jacobian for the anthropomorphic robot in Figure 19.4 can be computed on the basis of the matrix

$$ {}^{3}\boldsymbol{J}_{6} = \begin{bmatrix} 0 & \ell_3 s_3 - d_4 & -d_4 & 0 & 0 & 0 \\ 0 & \ell_3 c_3 & 0 & 0 & 0 & 0 \\ -\ell_3 c_2 + d_4 s_{23} & 0 & 0 & 0 & 0 & 0 \\ s_{23} & 0 & 0 & 0 & s_4 & -c_4 s_5 \\ c_{23} & 0 & 0 & 1 & 0 & c_5 \\ 0 & 1 & 1 & 0 & c_4 & s_4 s_5 \end{bmatrix}. \tag{19.66} $$

© 2002 by CRC Press LLC

19.5.2 Analytical Jacobian

If the end-effector position and orientation are specified in terms of a minimum number of parameters in the task space as in (19.36), it is possible to also compute the Jacobian matrix by direct differentiation of the direct kinematics equation, i.e.,

$$\dot{\boldsymbol{x}} = \begin{bmatrix} \dot{\boldsymbol{p}}_e \\ \dot{\boldsymbol{\varphi}}_e \end{bmatrix} = \boldsymbol{J}_a(\boldsymbol{q})\dot{\boldsymbol{q}}, \tag{19.67}$$

where the matrix $\boldsymbol{J}_a(\boldsymbol{q}) = \partial \boldsymbol{k}/\partial \boldsymbol{q}$ is termed *analytical Jacobian*.

The relationship between the analytical Jacobian and the geometric Jacobian is expressed as

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}(\boldsymbol{\varphi}_e) \end{bmatrix} = \boldsymbol{T}_a(\boldsymbol{\varphi}_e)\boldsymbol{J}_a, \tag{19.68}$$

where $\boldsymbol{T}(\boldsymbol{\varphi}_e)$ is the transformation matrix defined in (19.55) that depends on the particular set of Euler angles used to represent end-effector orientation.

It can be easily recognized that the two Jacobians are in general different; note, however, that the two coincide for the positioning part. Concerning their use, the geometric Jacobian is adopted when physical quantities are of interest, while the analytical Jacobian is adopted when task space quantities are the focus. It is always possible to pass from one Jacobian to the other, except when the transformation matrix is singular. The orientations at which the determinant of $\boldsymbol{T}(\boldsymbol{\varphi}_e)$ vanishes are called *representation singularities* of $\boldsymbol{\varphi}_e$. For instance, with reference to the *XYZ* representation in (19.18), the transformation matrix is

$$\boldsymbol{T}(\boldsymbol{\varphi}_e) = \begin{bmatrix} 1 & 0 & s_\beta \\ 0 & c_\alpha & -s_\alpha c_\beta \\ 0 & s_\alpha & c_\alpha c_\beta \end{bmatrix}. \tag{19.69}$$

$\boldsymbol{T}$ becomes singular at the representation singularities $\beta = \pm \pi/2;$ notice that, in these configurations, it impossible to describe an arbitrary angular velocity with a set of Euler angle time derivatives. It should be remarked that each of the other Euler angle descriptions suffers from the occurrence of two representation singularities.

19.5.3 Singularities

The differential kinematics Equation (19.56) defines a linear mapping between the vector of joint velocities $\dot{\boldsymbol{q}}$ and the vector of end-effector velocities $\boldsymbol{v}$. The Jacobian is in general a function of the robot configuration $\boldsymbol{q}$. Those configurations at which $\boldsymbol{J}$ is rank-deficient are called *kinematic singularities*.

The simplest means to find singularities is to compute the determinant of the Jacobian matrix. For instance, for the above Jacobian in (19.66) it is

$$\det\left({}^3\boldsymbol{J}_6\right) = \ell_3 d_4 c_3 s_5 \left(d_4 s_{23} - \ell_3 c_2\right) \tag{19.70}$$

leading to three types of singularities $\left(\ell_3, d_4 \neq 0\right)$: *elbow singularity*

$$c_3 = 0$$

© 2002 by CRC Press LLC

when links 2 and 3 are aligned; *shoulder singularity*

$$d_4 s_{23} - \ell_3 c_2 = 0$$

when the origin of frame 4 is along axis Z_0; and *wrist singularity*

$$s_5 = 0$$

when axes Z_4 and Z_6 are aligned. Notice that elbow singularity is not troublesome because it occurs at the boundary of the robot workspace $\left(q_3 = \pm\pi/2\right)$. Shoulder singularity is characterized in the task space and thus it can be avoided when planning an end-effector trajectory. Instead, wrist singularity is characterized in the joint space $\left(q_5 = 0,\ \pi\right)$, and thus it is difficult to predict when planning an end-effector trajectory.

An effective tool for analyzing the linear mapping from the joint velocity space into the task velocity space defined by (19.56) is offered by the singular value decomposition (SVD) of the Jacobian matrix is given by

$$\boldsymbol{J} = \boldsymbol{U}\,\boldsymbol{\Sigma}\boldsymbol{V}^{\boldsymbol{T}} = \sum_{i=1}^{m} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\boldsymbol{T}}, \tag{19.71}$$

where $\boldsymbol{U}$ is the $(m \times m)$ matrix of the output singular vectors $\boldsymbol{u}_i$, $\boldsymbol{V}$ is the $(n \times n)$ matrix of the input singular vectors $\boldsymbol{v}_i$, and $\Sigma = [\boldsymbol{S} \quad \boldsymbol{O}]$ is the $(m \times n)$ matrix whose $(m \times m)$ diagonal submatrix $\boldsymbol{S}$ contains the singular values σ_i of the matrix $\boldsymbol{J}$. If r denotes the rank of $\boldsymbol{J}$, the following properties hold:

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_m = 0$,
- $\boldsymbol{R}(\boldsymbol{J}) = \text{span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$,
- $\boldsymbol{N}(\boldsymbol{J}) = span\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$.

The null space $\boldsymbol{N}(\boldsymbol{J})$ is the set of joint velocities that yield null task velocities at the current configuration; these joint velocities are termed *null space joint velocities*. A base of $\boldsymbol{N}(\boldsymbol{J})$ is given by the $(n - r)$ last input singular vectors, which represent independent linear combinations of the joint velocities. Hence, one effect of a singularity is to increase the dimension of $\boldsymbol{N}(\boldsymbol{J})$ by introducing a linear combination of joint velocities that produce a null task velocity.

The range space $\boldsymbol{R}(\boldsymbol{J})$ is the set of task velocities that can be obtained as a result of all possible joint velocities; these task velocities are termed *feasible space task velocities*. A base of $\boldsymbol{R}(\boldsymbol{J})$ is given by the first r output singular vectors that represent independent linear combinations of the single components of task velocities. Accordingly, another effect of a singularity is to decrease the dimension of $\boldsymbol{R}(\boldsymbol{J})$ by eliminating a linear combination of task velocities from the space of feasible velocities.

The singular value decomposition (19.71) shows that the i-th singular value of $\boldsymbol{J}$ can be viewed as a gain factor relating the joint velocity along the $\boldsymbol{v}_i$ direction to the task velocity along the $\boldsymbol{u}_i$ direction. When a singularity is approached, the r-th singular value tends to zero and the task velocity produced by a fixed joint velocity along $\boldsymbol{v}_r$ is decreased proportionally to s_r. At the singular configuration, the joint velocity along $\boldsymbol{v}_r$ is in the null space and the task velocity along $\boldsymbol{u}_r$ becomes infeasible.

In the general case, the joint velocity has components in any $\boldsymbol{v}_i$ direction, and the resulting task velocity can be obtained as a combination of the single components along each output singular vector direction.

© 2002 by CRC Press LLC

19.6 Differential Kinematics Inversion

The differential kinematics equation, in terms of either the geometric or the analytical Jacobian, establishes a linear mapping between joint space velocities and task space velocities, even if the Jacobian is a function of joint configuration. This feature suggests the use of the differential kinematics Equation (19.56) to solve the inverse kinematics problem.

Assume that a task space trajectory is given ($\boldsymbol{x}(t)$, $\boldsymbol{v}(t)$). The goal is to find a feasible joint space trajectory $\left(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t)\right)$ that reproduces the given trajectory. Joint velocities can be obtained by solving the differential kinematics equation for $\dot{\boldsymbol{q}}$ at the current joint configuration; then, joint positions $\boldsymbol{q}(t)$ can be computed by integrating the velocity solution over time with known initial conditions. This approach is based on knowledge of the robot Jacobian and is applicable to any robot structure, on the condition that a suitable inverse for the matrix $\boldsymbol{J}$ can be found.

19.6.1 Pseudoinverse

With reference to the geometric Jacobian, the basic inverse solution to (19.56) is obtained by using the *pseudoinverse* $\boldsymbol{J}^{\dagger}$ of the matrix $\boldsymbol{J}$; this is a unique matrix satisfying the Moore-Penrose conditions

$$\boldsymbol{J}\boldsymbol{J}^{\dagger}\boldsymbol{J}=\boldsymbol{J} \qquad \boldsymbol{J}^{\dagger}\boldsymbol{J}\boldsymbol{J}^{\dagger}=\boldsymbol{J}^{\dagger}$$

$$\left(\boldsymbol{J}\boldsymbol{J}^{\dagger}\right)^{\mathbf{T}}=\boldsymbol{J}\boldsymbol{J}^{\dagger} \qquad \left(\boldsymbol{J}^{\dagger}\boldsymbol{J}\right)^{\mathbf{T}}=\boldsymbol{J}^{\dagger}\boldsymbol{J} \tag{19.72}$$

or, alternatively, the equivalent conditions

$$\begin{aligned} \boldsymbol{J}^{\dagger}\boldsymbol{a}&=\boldsymbol{a} && \forall \boldsymbol{a}\in\boldsymbol{N}^{\perp}(\boldsymbol{J}) \\ \boldsymbol{J}^{\dagger}\boldsymbol{b}&=0 && \forall \boldsymbol{b}\in\boldsymbol{R}^{\perp}(\boldsymbol{J}) \\ \boldsymbol{J}^{\dagger}(\boldsymbol{a}+\boldsymbol{b})&=\boldsymbol{J}^{\dagger}\boldsymbol{a}+\boldsymbol{J}^{\dagger}\boldsymbol{b} && \forall \boldsymbol{a}\in\boldsymbol{R}(\boldsymbol{J}),\forall \boldsymbol{b}\in\boldsymbol{R}^{\perp}(\boldsymbol{J}). \end{aligned} \tag{19.73}$$

The inverse solution can then be written as

$$\dot{\boldsymbol{q}}=\boldsymbol{J}^{\dagger}(\boldsymbol{q})\boldsymbol{v} \tag{19.74}$$

that provides a least-squares solution with minimum norm to Equation (19.56); in detail, solution (19.74) satisfies the condition

$$\min_{\dot{\boldsymbol{q}}}\|\dot{\boldsymbol{q}}\| \tag{19.75}$$

of all $\dot{\boldsymbol{q}}$ that fulfill

$$\min_{\dot{\boldsymbol{q}}}\|\boldsymbol{v}-\boldsymbol{J}\dot{\boldsymbol{q}}\|. \tag{19.76}$$

If the Jacobian matrix is full-rank, the right pseudoinverse of $\boldsymbol{J}$ can be computed as

$$\boldsymbol{J}^{\dagger}=\boldsymbol{J}^{\mathrm{T}}\left(\boldsymbol{J}\boldsymbol{J}^{\mathrm{T}}\right)^{-1}, \tag{19.77}$$

© 2002 by CRC Press LLC

and (19.74) provides an exact solution to (19.56). Further, if $\boldsymbol{J}$ square, the pseudoinverse (19.77) reduces to the standard inverse Jacobian matrix $\boldsymbol{J}^{-1}$.

To gain insight into the properties of the inverse mapping described by (19.74), it is useful to consider the singular value decomposition (19.71) of $\boldsymbol{J}$, and thus

$$\boldsymbol{J}^{\dagger} = \boldsymbol{V}\boldsymbol{\Sigma}^{\dagger}\boldsymbol{U}^{T} = \sum_{i=1}^{r} \frac{1}{\sigma_i} \boldsymbol{v}_i \boldsymbol{u}_i^{T} \tag{19.78}$$

where r denotes the rank of $\boldsymbol{J}$. The following properties hold:

- $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_m = 0$,
- $\boldsymbol{R}(\boldsymbol{J}^{\dagger}) = \boldsymbol{N}^{\perp}(\boldsymbol{J}) = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$,
- $\boldsymbol{N}(\boldsymbol{J}^{\dagger}) = \boldsymbol{R}^{\perp}(\boldsymbol{J}) = \text{span}\{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n\}$.

The null space $\boldsymbol{N}(\boldsymbol{J}^{+})$ is the set of task velocities that yields null joint space velocities at the current configuration; these task velocities belong to the orthogonal complement of the feasible space task velocities. Hence, one effect of the pseudoinverse solution (19.74) is to filter the infeasible components of the given task velocities while allowing exact tracking of the feasible components; this is due to the minimum norm property (19.75).

The range space $\boldsymbol{R}(\boldsymbol{J}^{\dagger})$ is the set of joint velocities that can be obtained as a result of all possible task velocities. Because these joint velocities belong to the orthogonal complement of the null space joint velocities, the pseudoinverse solution (19.74) satisfies the least-squares condition (19.76).

If a task velocity is assigned along $\boldsymbol{u}_i$, the corresponding joint velocity computed via (19.74) lies along $\boldsymbol{v}_i$ and is magnified by the factor $1/\sigma_i$. When a singularity is approached, the r-th singular value tends to zero and a fixed task velocity along $\boldsymbol{u}_r$ requires large joint velocities. At a singular configuration, the $\boldsymbol{u}_r$ direction becomes infeasible and $\boldsymbol{v}_r$ adds to the set of null space velocities of the robot.

19.6.2 Redundancy

For a kinematically redundant robot a nonempty null space $\boldsymbol{N}(\boldsymbol{J})$ exists which is available to set up systematic procedures for an effective handling of redundant degrees of freedom. The general inverse solution can be written as

$$\dot{\boldsymbol{q}} = \boldsymbol{J}^{\dagger}(\boldsymbol{q})\boldsymbol{v} + \left(\boldsymbol{I} - \boldsymbol{J}^{\dagger}(\boldsymbol{q})\boldsymbol{J}(\boldsymbol{q})\right)\dot{\boldsymbol{q}}_0 \tag{19.79}$$

which satisfies the least-squares condition (19.76) but loses the minimum norm property (19.75) by virtue of the addition of the homogeneous term $\left(\boldsymbol{I} - \boldsymbol{J}^{\dagger}\boldsymbol{J}\right)\dot{\boldsymbol{q}}_0$. The matrix $\left(\boldsymbol{I} - \boldsymbol{J}^{\dagger}\boldsymbol{J}\right)$ is a projector of the joint vector $\dot{\boldsymbol{q}}_0$ onto $\boldsymbol{N}(\boldsymbol{J})$.

In terms of the singular value decomposition, solution (19.79) can be written in the form

$$\dot{\boldsymbol{q}} = \sum_{i=1}^{r} \boldsymbol{v}_i \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{v} + \sum_{i=r+1}^{m} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} \dot{\boldsymbol{q}}_0 + \sum_{i=m+1}^{n} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} \dot{\boldsymbol{q}}_0. \tag{19.80}$$

Three contributions can be recognized in (19.80), namely, the least-squares joint velocities, the null space joint velocities due to singularities (if $r < m$), and the null space joint velocities due to redundant degrees of freedom (if $m < n$).

This result is of fundamental importance for redundancy resolution, because solution (19.79) evidences the possibility of choosing the vector $\dot{\boldsymbol{q}}_0$ to exploit the redundant degrees of freedom.

© 2002 by CRC Press LLC

In fact, the contribution of $\dot{\boldsymbol{q}}_0$ is to generate null space motions of the structure that do not alter the task space configuration but allow the robot to reach more dexterous postures for the execution of the given task.

A typical choice of the null space joint velocity vector is

$$\dot{\boldsymbol{q}}_0 = \alpha\left(\frac{\partial w(\boldsymbol{q})}{\partial \boldsymbol{q}}\right)^{\mathrm{T}} \tag{19.81}$$

with $\alpha > 0$; $w(\boldsymbol{q})$ is a scalar objective function of the joint variables, and $\left(\partial w(\boldsymbol{q}) / \partial \boldsymbol{q}\right)^{\mathrm{T}}$ is the vector function representing the gradient of w. In this way, locally optimizing w in accordance with the kinematic constraint expressed by (19.56) is sought. Usual objective functions are

- The *manipulation* measure defined as

$$w(\boldsymbol{q}) = \sqrt{\det\left(\boldsymbol{J}(\boldsymbol{q})J^{\mathrm{T}}(\boldsymbol{q})\right)}, \tag{19.82}$$

  which vanishes at a singular configuration, and thus redundancy may be exploited to escape singularities.
- The *distance from mechanical joint limits* defined as

$$w(\boldsymbol{q}) = -\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{q_i - \overline{q}_i}{q_{iM} + q_{im}}\right)^2, \tag{19.83}$$

  where q_{iM} (q_{im}) denotes the maximum (minimum) limit for q_i and $\overline{q}_i$ the middle of the joint range, and thus redundancy may be exploited to keep the robot from joint limits.
- The *distance from an obstacle* defined as

$$w(\boldsymbol{q}) = \min_{\boldsymbol{p},\boldsymbol{o}} \left\| \boldsymbol{p}(\boldsymbol{q}) - \boldsymbol{o} \right\|, \tag{19.84}$$

  where $\boldsymbol{o}$ is the position vector of an opportune point on the obstacle and $\boldsymbol{p}$ is the position vector of the closest robot point to the obstacle, and thus redundancy may be exploited to avoid collisions with obstacles.

19.6.3 Damped Least-Squares Inverse

In the neighborhood of singular configurations the use of a pseudoinverse is not adequate and a numerically robust solution is achieved by the *damped least-squares inverse* technique based on the solution to the modified differential kinematics equation

$$\boldsymbol{J}^{\mathrm{T}}\boldsymbol{v} = \left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} + \lambda^2\boldsymbol{I}\right)\dot{\boldsymbol{q}} \tag{19.85}$$

in place of Equation (19.56); in (19.85) the scalar λ is the so-called *damping factor*. Note that when $\lambda = 0$, Equation (19.85) reduces to (19.56).

The solution to (19.85) can be written in either of the equivalent forms

$$\dot{\boldsymbol{q}} = \boldsymbol{J}^{\mathrm{T}}\left(\boldsymbol{J}\boldsymbol{J}^{\mathrm{T}} + \lambda^2\boldsymbol{I}\right)^{-1}\boldsymbol{v} \tag{19.86}$$

© 2002 by CRC Press LLC

$$\dot{\boldsymbol{q}} = \left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} + \lambda^2\boldsymbol{I}\right)^{-1}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{\upsilon}. \tag{19.87}$$

The computational load of (19.86) is lower than that of (19.87), being usually $n \geq m$. Let

$$\dot{\boldsymbol{q}} = \boldsymbol{J}^{\#}(\boldsymbol{q})\boldsymbol{\upsilon} \tag{19.88}$$

indicate the damped least-squares inverse solution computed with either of the above forms. Solution (19.88) satisfies the condition

$$\min_{\dot{\boldsymbol{q}}} \left\| \boldsymbol{\upsilon} - \boldsymbol{J}\dot{\boldsymbol{q}} \right\|^2 + \lambda^2 \left\| \dot{\boldsymbol{q}} \right\|^2 \tag{19.89}$$

that gives a trade-off between the least-squares condition (19.76) and the minimum norm condition (19.75). In fact, condition (19.89) accounts for both accuracy and feasibility in choosing the joint space velocity $\dot{\boldsymbol{q}}$ required to produce the given task space velocity $\boldsymbol{\upsilon}$. In this regard, it is essential to select a suitable value for the damping factor; small values of λ give accurate solutions but low robustness in the neighborhood of singular configurations, while large values of λ result in low tracking accuracy even if feasible and accurate solutions are possible.

Resorting to the singular value decomposition, the damped least-squares inverse solution (19.88) can be written as

$$\dot{\boldsymbol{q}} = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \lambda^2}\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{\upsilon}. \tag{19.90}$$

Remarkably, it is

- $\boldsymbol{R}\left(\boldsymbol{J}^{\#}\right) = \boldsymbol{R}\left(\boldsymbol{J}^{\dagger}\right) = \boldsymbol{N}^{\perp}(\boldsymbol{J}) = \mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$,
- $\boldsymbol{N}\left(\boldsymbol{J}^{\#}\right) = \boldsymbol{N}\left(\boldsymbol{J}^{\dagger}\right) = \boldsymbol{R}^{\perp}(\boldsymbol{J}) = \mathrm{span}\{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n\}$,

that is, the structural properties of the damped least-squares inverse solution are analogous to those of the pseudoinverse solution.

It is clear that with respect to the pure least-squares solution (19.74) the components for which $\sigma_i \gg \lambda$ are not influenced by the damping factor, because in this case it is

$$\frac{\sigma_i}{\sigma_i^2 + \lambda^2} \approx \frac{1}{\sigma_i}. \tag{19.91}$$

On the other hand, when a singularity is approached, the smallest singular value tends to zero while the associated component of the solution is driven to zero by the factor σ_i/λ^2; this progressively reduces the joint velocity to achieve near-degenerate components of the commanded task velocity. At the singularity, solutions (19.88) and (19.74) behave identically as long as the remaining singular values are significantly larger than the damping factor. Note that an upper bound of $1/2\lambda$ is set on the magnification factor relating the task velocity component along $\boldsymbol{u}_i$ to the resulting joint velocity along $\boldsymbol{v}_i$; this bound is reached when $\sigma_i = \lambda$.

The damping factor λ determines the degree of approximation introduced with respect to the pure least-squares solution. Then, using a constant value for λ may turn out to be inadequate for obtaining good performance over the entire robot workspace. An effective choice is to adjust λ as a function of some measure of closeness to the singularity at the current configuration of the robot. To this purpose, manipulability measures or estimates of the smallest singular value can be adopted.

© 2002 by CRC Press LLC

Remarkably, currently available microprocessors even allow real-time computation of full singular-value decomposition.

A singular region can be defined on the basis of the smallest singular value estimate of $\boldsymbol{J}$. Outside the region the exact solution is used, while inside the region a configuration-varying damping factor is introduced to obtain the desired approximate solution. The factor must be chosen so that continuity of joint velocity $\dot{\boldsymbol{q}}$ is ensured in the transition at the border of the singular region.

Without loss of generality, for a six-degree-of-freedom robot, the damping factor can be selected according to the following law:

$$\lambda^2 = \begin{cases} 0 & \hat{\sigma}_6 \geq \varepsilon \\ \left(1 - \left(\frac{\hat{\sigma}_6}{\varepsilon}\right)^2\right)\lambda_{\max}^2 & \hat{\sigma}_6 < \varepsilon, \end{cases} \tag{19.92}$$

where $\hat{\sigma}_6$ is the smallest singular value estimate, and ε defines the size of the singular region; the value of $\lambda_{\max}$ is at the user's disposal to suitably shape the solution in the neighborhood of a singularity.

Equation (19.92) requires computation of the smallest singular value. To avoid a full singular-value decomposition, we can resort to a recursive algorithm to find an estimate of the smallest singular value. Suppose that an estimate $\hat{\boldsymbol{v}}_6$ of the last input singular vector is available, so that $\hat{\boldsymbol{v}}_6 \approx \boldsymbol{v}_6$ and $\left\|\hat{\boldsymbol{v}}_6\right\|$. This estimate is used to compute the vector $\hat{\boldsymbol{v}}_6'$ from

$$\left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} + \lambda^2\boldsymbol{I}\right)\hat{\boldsymbol{v}}_6' = \hat{\boldsymbol{v}}_6. \tag{19.93}$$

Then the square of the estimate $\hat{\sigma}_6$ of the smallest singular value can be found as

$$\hat{\sigma}_6^2 = \frac{1}{\left\|\hat{\boldsymbol{v}}_6'\right\|} - \lambda^2, \tag{19.94}$$

while the estimate of $\boldsymbol{v}_6$ is updated using

$$\hat{\boldsymbol{v}}_6 = \frac{\hat{\boldsymbol{v}}_6'}{\left\|\hat{\boldsymbol{v}}_6'\right\|}. \tag{19.95}$$

The above estimation scheme is based on the assumption that $\boldsymbol{v}_6$ is slowly rotating, which is normally the case. However, if the robot is close to a double singularity (e.g., a shoulder and a wrist singularity for the anthropomorphic robot), the vector $\boldsymbol{v}_6$ will instantaneously rotate if the two smallest singular values cross. The estimate of the smallest singular value will then track σ_5 initially, before $\hat{\boldsymbol{v}}_6$ converges again to $\boldsymbol{v}_6$. Therefore, it is worth extending the scheme by estimating not only the smallest but also the second smallest singular value. Assume that the estimates $\hat{\boldsymbol{v}}_6$ and $\hat{\sigma}_6$ are available and define the matrix

$$\boldsymbol{M} = \boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} + \lambda^2\boldsymbol{I} - \left(\hat{\sigma}_6^2 + \lambda^2\right)\hat{\boldsymbol{v}}_6\hat{\boldsymbol{v}}_6^{\mathrm{T}}. \tag{19.96}$$

With this choice, the second smallest singular value of $\boldsymbol{J}$ plays in

$$\boldsymbol{M}\hat{\boldsymbol{v}}_5' = \hat{\boldsymbol{v}}_5 \tag{19.97}$$

the same role as σ_6 in (19.93) and then will provide a convergent estimate of $\hat{\boldsymbol{v}}_5$ to $\boldsymbol{v}_5$ and $\hat{\sigma}_5$ to σ_5.

© 2002 by CRC Press LLC

At this point, suppose that $\hat{\boldsymbol{v}}_5$ is an estimate of $\boldsymbol{v}_5$ so that $\hat{\boldsymbol{v}}_5 \approx \boldsymbol{v}_5$ and $\|\hat{\boldsymbol{v}}_5\| = 1$. This estimate is used to compute $\hat{\boldsymbol{v}}_5'$ from (19.97). Then, an estimate of the square of the second smallest singular value of $\boldsymbol{J}$ is found from

$$\hat{\sigma}_5^2 = \frac{1}{\|\hat{\boldsymbol{v}}_5'\|} - \lambda^2, \tag{19.98}$$

and the estimate of $\boldsymbol{v}_5$ is updated using

$$\hat{\boldsymbol{v}}_5 = \frac{\hat{\boldsymbol{v}}_5'}{\|\hat{\boldsymbol{v}}_5'\|}. \tag{19.99}$$

On the basis of this modified estimation algorithm, crossing of singularities can be effectively detected; also, by switching the two singular values and the associated estimates $\hat{\boldsymbol{v}}_5$ and $\hat{\boldsymbol{v}}_6$, the estimation of the smallest singular value will be accurate even when the two smallest singular values cross.

19.6.4 User-Defined Accuracy

The above damped least-squares inverse method achieves a compromise between accuracy and robustness of the solution. This is performed without specific regard to the components of the particular task assigned to the robot's end-effector. The *user-defined accuracy* strategy based on the weighted, damped, least-squares inverse method allows discriminating between directions in the task space where higher accuracy is desired and directions where lower accuracy can be tolerated. This is the case, for instance, of spot welding or spray painting in which the tool angle about the approach direction is not essential to the fulfillment of the task.

Let a weighted end-effector velocity vector be defined as

$$\bar{\boldsymbol{v}} = \boldsymbol{W}\boldsymbol{v} \tag{19.100}$$

where $\boldsymbol{W}$ is the $(m \times m)$ task-dependent weighting matrix taking into account the anisotropy of the task requirements. Substituting (19.100) into (19.56) gives

$$\bar{\boldsymbol{v}} = \bar{\boldsymbol{J}}(\boldsymbol{q})\dot{\boldsymbol{q}} \tag{19.101}$$

where $\bar{\boldsymbol{J}} = \boldsymbol{W}\boldsymbol{J}$. It is worth noticing that if $\boldsymbol{W}$ is full-rank, solving (19.56) is equivalent to solving (19.101), but with different conditioning of the system of equations to solve. This suggests selecting only the strictly necessary weighting action to avoid undesired ill-conditioning of $\bar{\boldsymbol{J}}$.

Equation (19.101) can be solved by using the weighted, damped, least-squares inverse technique, i.e.,

$$\bar{\boldsymbol{J}}^{\mathrm{T}}(\boldsymbol{q})\bar{\boldsymbol{v}} = \left(\bar{\boldsymbol{J}}^{\mathrm{T}}(\boldsymbol{q})\bar{\boldsymbol{J}}(\boldsymbol{q}) + \lambda^2\boldsymbol{I}\right)\dot{\boldsymbol{q}}. \tag{19.102}$$

Again, the singular value decomposition of the matrix $\bar{\boldsymbol{J}}$ is helpful, i.e.,

$$\bar{\boldsymbol{J}} = \sum_{i=1}^{r} \bar{\sigma}_i \bar{\boldsymbol{u}}_i \bar{\boldsymbol{v}}_i^{\mathrm{T}} \tag{19.103}$$

and the solution to (19.102) can be written as

© 2002 by CRC Press LLC

$$\dot{\boldsymbol{q}} = \sum_{i=1}^{r} \frac{\bar{\sigma}_i}{\bar{\sigma}_i^2 + \lambda^2} \bar{\boldsymbol{v}}_i \bar{\boldsymbol{u}}_i^{\mathrm{T}} \boldsymbol{v}. \tag{19.104}$$

It is clear that the singular values $\bar{\sigma}_i$ and the singular vectors $\bar{\boldsymbol{u}}_i$ and $\bar{\boldsymbol{v}}_i$ depend on the choice of the weighting matrix $\boldsymbol{W}$. While this has no effect on the solution $\dot{\boldsymbol{q}}$ as long as $\bar{\sigma}_r \gg \lambda$, close to singularities where $\bar{\sigma}_r \ll \lambda$, for some $r < m$, the solution can be shaped by properly selecting the matrix $\boldsymbol{W}$.

For a six-degree-of-freedom robot with a spherical wrist, it is worthwhile to devise special handling of the wrist singularity, because such a singularity is difficult to predict at the planning level in the task space. It can be recognized that at the wrist singularity only two components of the angular velocity vector can be generated by the wrist itself. The remaining component might be generated by the inner joints, although at the expense of loss of accuracy along some other task space directions. For this reason, lower weight should be put on the angular velocity component that is infeasible to the wrist. For the anthropomorphic robot, this is easily expressed in the frame attached to link 4; let $\boldsymbol{R_4}$ denote the rotation matrix describing orientation of this frame with respect to the base frame so that the infeasible component is aligned with the X-axis. Then the weighting matrix can be chosen as

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{R}_{4\mathrm{diag}}\{w,1,1\}\boldsymbol{R}_4^T \end{bmatrix}. \tag{19.105}$$

Similar, to the choice of the damping factor as in (19.92), the weighting factor w is selected according to the following expression:

$$(1-w)^2 = \begin{cases} 0 & \hat{\sigma}_6 \geq \varepsilon \\ \left(1 - \left(\dfrac{\hat{\sigma}_6}{\varepsilon}\right)^2\right)\left(1 - w_{\min}\right)^2 & \hat{\sigma}_6 < \varepsilon, \end{cases} \tag{19.106}$$

where $w_{\min} > 0$ is a design parameter.

19.7 Inverse Kinematics Algorithms

The differential kinematics equation has been utilized above to solve for joint velocities. Open-loop reconstruction of joint variables through numerical integration unavoidably leads to solution drift and then to task space errors. This drawback can be overcome by devising a closed-loop *inverse kinematics algorithm* based on the task space error between the desired and actual end-effector locations $\boldsymbol{x}_d$ and $\boldsymbol{x}$, i.e., $\boldsymbol{e} = \boldsymbol{x}_d - \boldsymbol{x}(\boldsymbol{q})$. It is also worth considering the differential kinematics equation in the form (19.67) where the definition of the task error has required consideration of the analytical Jacobian $\boldsymbol{J_a}$ in lieu of the geometric Jacobian.

19.7.1 Jacobian Pseudoinverse

The joint velocity vector should be chosen so that the task error tends to zero. The simplest algorithm is obtained by using the *Jacobian pseudoinverse*

$$\dot{\boldsymbol{q}} = \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\left(\dot{\boldsymbol{x}}_d + \boldsymbol{K}\boldsymbol{e}\right), \tag{19.107}$$

which plugged into (19.67) gives

$$\dot{\boldsymbol{e}} + \boldsymbol{K}\boldsymbol{e} = \boldsymbol{0}. \tag{19.108}$$

© 2002 by CRC Press LLC

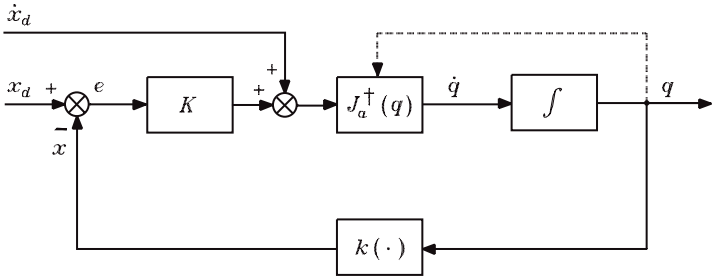


FIGURE 19.5 Block scheme of the inverse kinematics algorithm with the Jacobian pseudoinverse.

If $\boldsymbol{K}$ is a positive definite (diagonal) matrix, the linear system (19.108) is asymptotically stable; the tracking error along the given trajectory converges to zero with a rate depending on the eigenvalues of $\boldsymbol{K}$.

A block scheme of the inverse kinematics algorithm based on the Jacobian pseudoinverse is illustrated in Figure 19.5.

If it is desired to exploit redundant degrees of freedom, solution (19.107) can be generalized to

$$\dot{\boldsymbol{q}} = \boldsymbol{J}_a^\dagger(\boldsymbol{q})\left(\dot{\boldsymbol{x}}_d + \boldsymbol{K}\boldsymbol{e}\right) + \left(\boldsymbol{I} - \boldsymbol{J}_a^\dagger(\boldsymbol{q})\boldsymbol{J}_a(\boldsymbol{q})\right)\dot{\boldsymbol{q}}_0 \tag{19.109}$$

that logically corresponds to (19.79). In the case of numerical problems in the neighborhood of singularities, the pseudoinverse can be replaced with a suitable damped least-squares inverse.

19.7.2 Jacobian Transpose

A computationally efficient inverse kinematics algorithm can be derived by considering the *Jacobian transpose* in lieu of the pseudoinverse.

Consider the joint velocity vector

$$\dot{\boldsymbol{q}} = \boldsymbol{J}_a^{\mathrm{T}}(\boldsymbol{q})\boldsymbol{K}\boldsymbol{e} \tag{19.110}$$

where $\boldsymbol{K}$ is a symmetric positive definite matrix. A simple Lyapunov argument can be used to analyze the convergence of the algorithm. Consider the positive definite function candidate

$$\boldsymbol{V} = \frac{1}{2}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{e}; \tag{19.111}$$

its time derivative along the trajectories of the system (19.67) and (19.110) is

$$\dot{\boldsymbol{V}} = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{K}\dot{\boldsymbol{x}}_d - \boldsymbol{e}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{J}_a(\boldsymbol{q})\boldsymbol{J}_a^{\mathrm{T}}(\boldsymbol{q})\boldsymbol{K}\boldsymbol{e}. \tag{19.112}$$

If $\dot{\boldsymbol{x}}_d = \boldsymbol{0}$, it is easy to see that $\dot{\boldsymbol{V}}$ is negative definite as long as $\boldsymbol{J}_a$ is full-rank, and then it can be concluded that $\boldsymbol{e} = \boldsymbol{0}$ is an asymptotically stable equilibrium point for the system (19.67) and (19.110) as long as $\boldsymbol{J}_a$ is full-rank for all joint configurations $\boldsymbol{q}$. A number of remarks are in order.

- If $\dot{\boldsymbol{x}}_d \neq \boldsymbol{0}$, only boundedness of tracking errors can be established; an estimate of the bound is given by

$$\|\boldsymbol{e}\|_{\max} = \frac{\|\dot{\boldsymbol{x}}_d\|_{\max}}{k\sigma_r^2(\boldsymbol{J}_a)} \tag{19.113}$$

© 2002 by CRC Press LLC

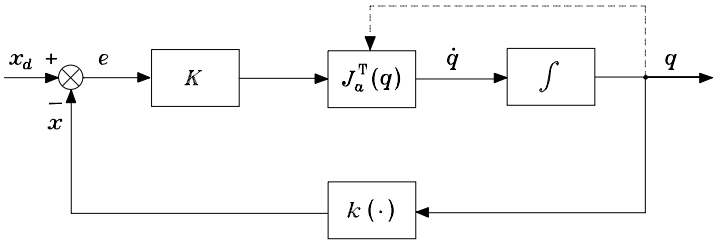


FIGURE 19.6 Block scheme of the inverse kinematics algorithm with Jacobian transpose.

where $\boldsymbol{K}$ has been conveniently chosen as a diagonal matrix $\boldsymbol{K} = k\boldsymbol{I}$. It is anticipated that k can be increased to diminish the errors, but, in practice, upper bounds exist due to discrete-time implementation of the algorithm.

- When a singularity is encountered, $N\left(\boldsymbol{J}_a^T\right)$ is nonempty and $\dot{V}$ is only semi-definite; $\dot{V} = 0$ for $\boldsymbol{e} \neq \boldsymbol{0}$ with $\boldsymbol{Ke} \in N\left(\boldsymbol{J}_a^T\right)$, and the algorithm may get stuck. It can be shown, however, that such an equilibrium point is unstable as long as $\dot{\boldsymbol{x}}_d$ drives $\boldsymbol{Ke}$ outside $N\left(\boldsymbol{J}_a^T\right)$. An enhancement of the algorithm can be achieved by rendering the matrix $\boldsymbol{J}_a^T\boldsymbol{K}$ less sensitive to variations of joint configurations along the task trajectory. This is accomplished by choosing a configuration-dependent $\boldsymbol{K}$ that compensates for variations of $\boldsymbol{J}_a$.

A block scheme of the inverse kinematics algorithm based on the Jacobian transpose is illustrated in Figure 19.6.

The most attractive feature of the Jacobian transpose algorithm is certainly the need of computing only direct kinematics functions $\boldsymbol{k}(\boldsymbol{q})$ and $\boldsymbol{J}_a(\boldsymbol{q})$. Further insight into the performance of solution (19.110) can be gained by considering the singular value decomposition of the Jacobian transpose, and thus

$$\boldsymbol{J}^T = \sum_{i=1}^{m} \sigma_i \boldsymbol{v}_i \boldsymbol{u}_i^T \tag{19.114}$$

which reveals continuous, smooth behavior of the solution close and through singular configurations. Note that in (19.114) the geometric Jacobian has been considered, and it has been assumed that no representation singularities are introduced.

19.7.3 Use of Redundancy

In case of redundant degrees of freedom, it is possible to combine the Jacobian pseudoinverse solution with the Jacobian transpose solution as illustrated below. This is carried out in the framework of the so-called *augmented task space* approach to exploit redundancy in robotic systems. The idea is to introduce an additional constraint task by specifying a $(p \times 1)$ vector $\boldsymbol{x}_c$ as a function of the robot joint variables, i.e.,

$$\boldsymbol{x}_c = \boldsymbol{k}_c(\boldsymbol{q}), \tag{19.115}$$

with $\boldsymbol{p} \leq n - m$ to constrain at most all the available redundant degrees of freedom. The constraint task vector $\boldsymbol{x}_c$ can be chosen by embedding scalar objective functions of the kind introduced in (19.82)–(19.84).

Differentiating (19.115) with respect to time gives

$$\dot{\boldsymbol{x}}_c = \boldsymbol{J}_c(\boldsymbol{q})\dot{\boldsymbol{q}} \tag{19.116}$$

© 2002 by CRC Press LLC

where $\boldsymbol{J}_c(\boldsymbol{q}) = \partial \boldsymbol{k}_c / \partial \boldsymbol{q}$ is the constraint Jacobian. The result is an augmented differential kinematics equation given by (19.67) and (19.116), based on a Jacobian matrix

$$\boldsymbol{J}' = \begin{bmatrix} \boldsymbol{J}_a \\ \boldsymbol{J}_c \end{bmatrix}. \tag{19.117}$$

When a constraint task is specified independently of the end-effector task, there is no guarantee that the matrix $\boldsymbol{J}'$ remains full-rank along the entire task path. Even if rank $(\boldsymbol{J}_a) = m$ and rank $(\boldsymbol{J}_c) = p$, then rank $(\boldsymbol{J}') = m + p$ if and only if $\boldsymbol{R}(\boldsymbol{J}_a^{\mathrm{T}}) \cap R(\boldsymbol{J}_c^{\mathrm{T}}) = \{\varnothing\}$. Singularities of $\boldsymbol{J}'$ are termed *artificial singularities,* and it can be shown that those are given by singularities of the matrix $\boldsymbol{J}_c\left(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}\boldsymbol{J}_a\right)$.

The above discussion suggests that, when solving for joint velocities, a *task priority strategy* is advisable to avoid conflicting situations between the end-effector task and the constraint task. Substituting (19.109) into (19.116) gives

$$\dot{\boldsymbol{x}}_c = \boldsymbol{J}_c(\boldsymbol{q})\boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\,(\dot{\boldsymbol{x}}_d + \boldsymbol{K}\boldsymbol{e}) + \boldsymbol{J}_c(\boldsymbol{q})\left(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\boldsymbol{J}_a(\boldsymbol{q})\right)\dot{\boldsymbol{q}}_0 \tag{19.118}$$

which could be solved for $\dot{\boldsymbol{q}}_0$ provided that artificial singularities — those of the matrix $\boldsymbol{J}_c(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}\boldsymbol{J}_a)$ — are avoided. Observing that equality (19.118) can be achieved only for the components of $\dot{\boldsymbol{x}}_c$ belonging to $\boldsymbol{R}(\boldsymbol{J}_c)$, it is sufficient to consider the equation

$$\boldsymbol{J}_c^{\dagger}(\boldsymbol{q})\dot{\boldsymbol{x}}_c = \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\,(\dot{\boldsymbol{x}}_d + \boldsymbol{K}\boldsymbol{e}) + \left(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\boldsymbol{J}_a(\boldsymbol{q})\right)\dot{\boldsymbol{q}}_0 \tag{19.119}$$

that can be solved for $\dot{\boldsymbol{q}}_0$ giving

$$\dot{\boldsymbol{q}}_0 = \left(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\boldsymbol{J}_a(\boldsymbol{q})\right)^{\dagger}\left(\boldsymbol{J}_c^{\dagger}(\boldsymbol{q})\dot{\boldsymbol{x}}_c - \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})(\dot{\boldsymbol{x}}_d + \boldsymbol{K}\boldsymbol{e})\right). \tag{19.120}$$

By recalling that $(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}\boldsymbol{J}_a)^{\dagger} = (\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}\boldsymbol{J}_a)$, solution (19.120) reduced to the simple form

$$\dot{\boldsymbol{q}}_0 = \left(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\boldsymbol{J}_a(\boldsymbol{q})\right)\boldsymbol{J}_c^{\dagger}(\boldsymbol{q})\dot{\boldsymbol{x}}_c. \tag{19.121}$$

Folding (19.121) back into (19.109) and exploiting the idempotence of $(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}\boldsymbol{J}_a)$ gives

$$\dot{\boldsymbol{q}} = \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})(\dot{\boldsymbol{x}}_d + \boldsymbol{K}\boldsymbol{e}) + \left(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}(\boldsymbol{q})\boldsymbol{J}_a(\boldsymbol{q})\right)\boldsymbol{J}_c^{\dagger}(\boldsymbol{q})(\dot{\boldsymbol{x}}_{cd} + \boldsymbol{K}_c\boldsymbol{e}_c) \tag{19.122}$$

where $\boldsymbol{e_c} = \boldsymbol{x}_{cd} - \boldsymbol{x}_c$, $\boldsymbol{x}_{cd}$ being the desired value of the constraint task, and $\boldsymbol{K}_c$ is a positive definite matrix. The operator $(\boldsymbol{I} - \boldsymbol{J}_a^{\dagger}\boldsymbol{J}_a)$ projects the secondary velocity contribution $\dot{\boldsymbol{q}}_0$ on the null space $N(\boldsymbol{J}_a)$, guaranteeing correct execution of the primary end-effector task while the secondary constraint task is correctly executed as long as it does not interfere with the end-effector task. Obviously, if desired, the order of priority can be switched, e.g., in an obstacle avoidance task when an obstacle is along the end-effector path.

In the case when $\boldsymbol{J}_c$ becomes singular, a damped least-squares inverse of $\boldsymbol{J}_c$ in lieu of the pseudoinverse in (19.121) can be used. Otherwise, by recalling the Jacobian transpose solution for the end-effector task (19.110), the null space joint velocity vector can be conveniently chosen as

$$\dot{\boldsymbol{q}}_0 = \boldsymbol{J}_c^{\mathrm{T}}(\boldsymbol{q})\boldsymbol{K}_c(\boldsymbol{x}_{cd} - \boldsymbol{x}_c), \tag{19.123}$$

© 2002 by CRC Press LLC

which allows the algorithm to work at a singularity of $\boldsymbol{J}_c$ and even at an artificial singularity. A tracking error arises for the constraint task but, observing that the desired constraint task is often constant over time ($\dot{\boldsymbol{x}}_{cd} = \boldsymbol{0}$), it can be concluded that the solution based on (19.123) performs equally well.

19.7.4 Orientation Errors

The above inverse kinematics algorithms make use of the analytical Jacobian since they operate on error variables (position and orientation) that are defined in the task space. More insight about the implications of different end-effector orientation descriptions can be gained by separating position from orientation components. With reference to the pseudoinverse algorithm based on (19.107), using the geometric Jacobian in lieu of the analytical Jacobian, the solution can be rewritten as

$$\dot{\boldsymbol{q}} = \boldsymbol{J}^{\dagger}(\boldsymbol{q}) \begin{bmatrix} \boldsymbol{\upsilon}_p \\ \boldsymbol{\upsilon}_o \end{bmatrix} \tag{19.124}$$

where $\boldsymbol{\upsilon}_p$, $\boldsymbol{\upsilon}_o$ represent two resolved velocities chosen to ensure tracking of the desired end-effector motion. Substituting (19.124) into (19.56) gives

$$\dot{\boldsymbol{p}}_e = \boldsymbol{\upsilon}_p \tag{19.125}$$

$$\boldsymbol{\omega}_e = \boldsymbol{\upsilon}_o \tag{19.126}$$

where the explicit end-effector linear and angular velocities have been separated.

For position, the choice is rather straightforward, i.e.,

$$\boldsymbol{\upsilon}_p = \dot{\boldsymbol{p}}_d + \boldsymbol{K}_p \boldsymbol{e}_p \tag{19.127}$$

where the position error

$$\boldsymbol{e}_p = \boldsymbol{p}_d - \boldsymbol{p}_e(\boldsymbol{q}) \tag{19.128}$$

between the desired and actual end-effector positions has been defined. Substituting (19.127) into (19.125) gives

$$\dot{\boldsymbol{e}}_p + \boldsymbol{K}_p \boldsymbol{e}_p = \boldsymbol{0} \tag{19.129}$$

and the choice of a positive definite matrix $\boldsymbol{K}_p$ guarantees asymptotic stability of the error system which in turn implies tracking of $\boldsymbol{p}_d$.

On the other hand, for the orientation error, some considerations are in order depending on the type of description adopted. If Euler angles are adopted, the resolved angular velocity in (19.124) is chosen as

$$\boldsymbol{\upsilon}_o = \boldsymbol{T}(\boldsymbol{\varphi}_e)(\dot{\boldsymbol{\varphi}}_d + \boldsymbol{K}_o \boldsymbol{e}_{o,\mathrm{Eul}}) \tag{19.130}$$

where

$$\boldsymbol{e}_{o,\mathrm{Eul}} = \boldsymbol{\varphi}_d - \varphi_e(\boldsymbol{q}) \tag{19.131}$$

© 2002 by CRC Press LLC

is the orientation error. Substituting (19.130) into (19.126) gives

$$\dot{\boldsymbol{e}}_{o,\mathrm{Eul}} + \boldsymbol{K}_o \boldsymbol{e}_{o,\mathrm{Eul}} = \boldsymbol{0} \tag{19.132}$$

provided that the matrix $\boldsymbol{T}(\boldsymbol{\varphi}_e)$ is nonsingular. The system (19.132) is asymptotically stable for a positive definite $\boldsymbol{K}_o$, which in turn implies tracking of $\boldsymbol{\varphi}_d$.

To overcome the drawback of representation singularities in (19.130), an algorithm based on an alternative Euler angles description can be conceived that makes use of the rotation matrix describing the mutual orientation between the desired and the actual end-effector frame, i.e.,

$${}^e\boldsymbol{R}_d = \boldsymbol{R}_e^T(\boldsymbol{q})\boldsymbol{R}_d. \tag{19.133}$$

Differentiating (19.133) with respect to time and accounting for (19.53) gives

$${}^e\dot{\boldsymbol{R}}_d = \boldsymbol{S}({}^e\boldsymbol{\omega}_{de})\,{}^e\boldsymbol{R}_d \tag{19.134}$$

where $\boldsymbol{\omega}_{de} = \boldsymbol{\omega}_d - \boldsymbol{\omega}_e(\boldsymbol{q})$ is the end-effector angular velocity error.

Let $\boldsymbol{\varphi}_{de}$ denote the set of Euler angles that can be extracted from ${}^e\boldsymbol{R}_d$. Then, in view of (19.55) and (19.53), the angular velocity ${}^e\boldsymbol{\omega}_{de}$ in (19.134) is related to the time derivative of $\boldsymbol{\varphi}_{de}$ as

$${}^e\boldsymbol{\omega}_{de} = \boldsymbol{T}(\boldsymbol{\varphi}_{de})\dot{\boldsymbol{\varphi}}_{de}. \tag{19.135}$$

At this point, the resolved angular velocity in (19.124) can be chosen as

$$\boldsymbol{v}_o = \boldsymbol{\omega}_d + \boldsymbol{R}_e\boldsymbol{T}(\boldsymbol{\varphi}_{de})\boldsymbol{K}_o\boldsymbol{e}_{o,\mathrm{EulAlt}} \tag{19.136}$$

where

$$\boldsymbol{e}_{o,\mathrm{EulAlt}} = \boldsymbol{\varphi}_{de}. \tag{19.137}$$

Substituting (19.136) into (19.126) gives

$$\dot{\boldsymbol{e}}_{o,\mathrm{EulAlt}} + \boldsymbol{K}_o\boldsymbol{e}_{o,\mathrm{EulAlt}} = \boldsymbol{0} \tag{19.138}$$

provided that the matrix $\boldsymbol{T}(\boldsymbol{\varphi}_{de})$ is nonsingular.

The clear advantage of the alternative over the classical Euler angles algorithm based on (19.130) is that by adopting a representation ϕ_{de} for which $\boldsymbol{T}(\boldsymbol{0})$ is nonsingular, representation singularities occur only for large orientation errors, e.g., when $\beta_{de} = \pm\pi/2$ for the *XYZ* representation. In other words, the ill-conditioning of matrix $\boldsymbol{T}$ is not influenced by the desired or actual end-effector orientation but only by the orientation error; hence, as long as the error parameter $|\beta_{de}| < \pi/2$, the behavior of system (19.138) is not affected by representation singularities. In this respect, the choice of a particular Euler angle description among the 12 possible should be carefully made, i.e., in the sense of avoiding a representation singularity for the second angle of the type $\beta = 0$.

To overcome the problem of representation singularities, an inverse kinematics algorithm based on the angle/axis description of orientation can be devised. From (19.133), the rotation ϑ_{de}, and the unit vector $\boldsymbol{r}_{de}$ can be extracted using the formulæ (19.12). Then, the orientation error can be defined as

$$\boldsymbol{e}_{o,\mathrm{AnAx}} = \sin\vartheta_{de}\boldsymbol{r}_{de}. \tag{19.139}$$

© 2002 by CRC Press LLC

Notice that (19.139) gives a unique solution for $-\pi/2 < \vartheta < \pi/2$, but this interval is not limiting for a convergent inverse kinematics algorithm. It can be shown that a computational expression of the orientation error in (19.139) is given by

$$\boldsymbol{e}_{o,\mathrm{AnAx}} = \frac{1}{2}\left(\boldsymbol{S}(\boldsymbol{n}_e(\boldsymbol{q}))\boldsymbol{n}_d + \boldsymbol{S}(\boldsymbol{s}_e(\boldsymbol{q}))\boldsymbol{s}_d + \boldsymbol{S}(\boldsymbol{a}_e(\boldsymbol{q}))\boldsymbol{a}_d\right), \tag{19.140}$$

where the triplet of unit vectors has been used for both the desired and the actual end-effector rotation matrix. Note that the above limitation on ϑ sets the conditions $\boldsymbol{n}_e^{\mathrm{T}}\boldsymbol{n}_d \geq 0, \boldsymbol{s}_e^{\mathrm{T}}\boldsymbol{s}_d \geq 0, \boldsymbol{a}_e^{\mathrm{T}}\boldsymbol{a}_d \geq 0$.

Differentiation of (19.140) with respect to time gives

$$\dot{\boldsymbol{e}}_{o,\mathrm{AnAx}} = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\omega}_d - \boldsymbol{L}\boldsymbol{\omega} \tag{19.141}$$

where

$$\boldsymbol{L} = -\frac{1}{2}\left(\boldsymbol{S}(\boldsymbol{n}_d)\boldsymbol{S}(\boldsymbol{n}_e) + \boldsymbol{S}(\boldsymbol{s}_d)\boldsymbol{S}(\boldsymbol{s}_e) + \boldsymbol{S}(\boldsymbol{a}_d)\boldsymbol{S}(\boldsymbol{a}_e)\right). \tag{19.142}$$

At this point, the resolved angular velocity in (19.124) can be chosen as

$$\boldsymbol{\upsilon}_o = \boldsymbol{L}^{-1}(\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\omega}_d + \boldsymbol{K}_o\boldsymbol{e}_{o,\mathrm{AnAx}}). \tag{19.143}$$

Substituting (19.143) into (19.126) gives

$$\dot{\boldsymbol{e}}_{o,\mathrm{AnAx}} + \boldsymbol{K}_o\boldsymbol{e}_{o,\mathrm{AnAx}} = \boldsymbol{0} \tag{19.144}$$

provided that the matrix $\boldsymbol{L}$ is nonsingular. In this respect, if the angle ϑ_{de} is extended to the interval $(-\pi, \pi)$, then a singularity occurs at $\vartheta_{de} = \pm\pi/2$ for the matrix $\boldsymbol{L}$ which does not allow the computation of $\boldsymbol{\upsilon}_o$ as in (19.143).

The final inverse kinematics algorithm is based on the unit quaternion description of orientation. Let $\boldsymbol{Q}_d = \{\eta_d, \boldsymbol{\varepsilon}_d\}$ and $\boldsymbol{Q}_e = \{\eta_e, \boldsymbol{\varepsilon}_e\}$ represent the unit quaternions associated with $\boldsymbol{R}_d$ and $\boldsymbol{R}_e$, respectively. The mutual orientation can be expressed in terms of the unit quaternion $\boldsymbol{Q}_{de} = \{\eta_{de}, \boldsymbol{\varepsilon}_{de}\}$ where

$$\eta_{de} = \eta_e(\boldsymbol{q})\eta_d + \boldsymbol{\varepsilon}_e^{\mathrm{T}}(\boldsymbol{q})\boldsymbol{\varepsilon}_d$$

$$\boldsymbol{\varepsilon}_{de} = \eta_e(\boldsymbol{q})\boldsymbol{\varepsilon}_d - \eta_d\boldsymbol{\varepsilon}_e(\boldsymbol{q}) - \boldsymbol{S}(\boldsymbol{\varepsilon}_d)\boldsymbol{\varepsilon}_e(\boldsymbol{q}). \tag{19.145}$$

It can be recognized that $\boldsymbol{Q}_{de} = \{1, \boldsymbol{0}\}$ if and only if $\boldsymbol{R}_e$ and $\boldsymbol{R}_d$ are aligned, and thus it is sufficient to consider $\boldsymbol{\varepsilon}_{de}$ to express an end-effector orientation error, i.e.,

$$\boldsymbol{e}_{o,\mathrm{Quat}} = \boldsymbol{\varepsilon}_{de}. \tag{19.146}$$

Note that the explicit computation of $\eta_e(\boldsymbol{q})$ and $\boldsymbol{\varepsilon}_e(\boldsymbol{q})$ is not possible, but it requires intermediate computation of the rotation matrix $\boldsymbol{R}_e(\boldsymbol{q})$ that is available from the robot direct kinematics; then, the unit quaternion can be extracted using the formulæ (19.17).

At this point, the resolved angular velocity in (19.124) can be chosen as

$$\boldsymbol{\upsilon}_o = \boldsymbol{\omega}_d + \boldsymbol{K}_o\boldsymbol{e}_{o,\mathrm{Quat}} \tag{19.147}$$

© 2002 by CRC Press LLC

Substituting (19.147) into (19.126) gives

$$\boldsymbol{\omega}_{de} + \boldsymbol{K}_o \boldsymbol{e}_{o,\mathrm{Quat}} = \mathbf{0}. \tag{19.148}$$

It should be observed that now the orientation error equation is not homogeneous in $\boldsymbol{e}_{o,\mathrm{Quat}}$ since it contains the end-effector angular velocity error instead of the time derivative of the orientation error. To study the stability of system (19.148), consider the positive definite Lyapunov function

$$\boldsymbol{V} = (\eta_d - \eta_e)^2 + (\boldsymbol{\varepsilon}_d - \boldsymbol{\varepsilon}_e)^{\mathrm{T}}(\boldsymbol{\varepsilon}_d - \boldsymbol{\varepsilon}_e). \tag{19.149}$$

In view of the quaternion propagation (19.54), the time derivative of $\boldsymbol{V}$ along the trajectories of system (19.148) is given by

$$\dot{\boldsymbol{V}} = -\,\mathbf{e}_{o,\mathrm{Quat}}^{\mathrm{T}} \boldsymbol{K}_o \boldsymbol{e}_{o,\mathrm{Quat}} \tag{19.150}$$

which is negative definite, implying that $\boldsymbol{e}_{o,\mathrm{Quat}}$ converges to zero.

19.8 Further Reading

Kinematic modelling of rigid robot manipulators can be found in any classical robotics textbook, e.g., Craig,[17] Dombre and Khalil,[19] Paul,[43] Sciavicco and Siciliano,[51] Spong and Vidyasagar,[54] Vukobratović.[56] Important reference sources on kinematics are also Angeles,[1] Bottema and Roth,[3] Hunt,[25] McCarthy,[38] Vukobratović and Kirćanski.[57] Symbolic software packages have been developed to derive robot kinematic models, Khalil.[26]

The Denavit-Hartenberg notation dates back to the original work of Denavit and Hartenberg,[18] which was recently modified in Craig[17] and Khalil and Kleinfinger.[28] One advantage of the so-called modified Denavit-Hartenberg notation over the classical one is that it can also be used for tree-structured and closed-chain robots.[28] The homogeneous transformation representation for direct kinematics of open-chain robots was first proposed in Pieper.[45]

Sufficient conditions for the inverse kinematics problem of closed-form solutions were given in Pieper.[45] These ensure the existence of solutions to six-degrees-of-freedom robots provided there are three revolute joints with intersecting axes or three prismatic joints; in the former case, at most eight admissible solutions exist, while the number reduces to two in the latter case. The kinematic decoupling resulting for spherical-wrist robots was developed in Featherstone,[22] Hollerbach,[24] Khalil and Bennis,[27] and Paul and Zhang.[44] An algebraic approach to the inverse kinematics problem for robots having closed-form solutions was presented in Paul,[43] and consists of successively post- (or pre-) multiplying both sides of the direct kinematics equation by partial transformation matrices to isolate the joint variables one after another; the types of equations obtained with this approach were formalized in Dombre and Khalil.[19] Recent methods[32,46] have found the inverse kinematics solution to general six-revolute-joint robots in the form of a polynomial equation of degree 16, i.e., the maximum number of admissible solutions is 16. On the other hand, numerical solution techniques based on iterative algorithms have been proposed, e.g., Goldenberg et al.[23] and Tsai and Morgan.[55]

The geometric Jacobian of the differential kinematics equation was originally proposed in Whitney.[59] The decomposition of the Jacobian into the product of three matrices is due to Renaud.[47] The problem of efficient Jacobian computation was addressed in Orin and Schrader.[42] The analytical Jacobian concept was introduced in Khatib[29] in connection with the operational space control problem. A treatment of differential kinematics mapping properties can be found in Sciavicco and Siciliano[51]; the reader is referred to Klema and Laub[31] for SVD decomposition.

© 2002 by CRC Press LLC

The inversion of differential kinematics dates back to Whitney[59] under the name of resolved motion rate control. The adoption of the pseudoinverse of the Jacobian is due to Klein and Huang.[30] More on the properties of the pseudoinverse can be found in Boullion and Odell.[4] The use of null-space joint velocities for redundancy resolution was proposed in Liégeois,[33] and further refined in Maciejewski and Klein[36] and Yoshikawa[60] concerning the choice of objective functions. The reader is referred to Nakamura[39] for a complete treatment of redundant robots.

The adoption of the damped least-squares inverse was independently presented in Nakamura and Hanafusa[40] and Wampler.[58] More about kinematic control in the neighborhood of kinematic singularities can be found in Chiaverini.[9] The technique for estimating the smallest singular value of the Jacobian is due to Maciejewski and Klein,[37] and its modification to include the second smallest singular value was achieved by Chiaverini.[10] The use of the damped least-squares inverse for redundant robots was presented in Egeland et al.[21] The user-defined accuracy strategy was proposed in Chiaverini et al.[12] and further refined in Chiaverini et al.[13] A review of the damped least-squares inverse kinematics with experiments on an industrial robot was recently presented.[16]

Closed-loop inverse kinematics algorithms are discussed in Sciavicco and Siciliano.[51] The original Jacobian transpose inverse kinematics algorithm was proposed in Sciavicco and Siciliano;[49] the choice of suitable gains for achieving robustness to singularities was discussed in Chiacchio and Siciliano.[7] Singular value decomposition of the Jacobian transpose is due to Chiaverini et al.[14] Combining the Jacobian transpose solution with the pseudoinverse solution was proposed in Chiacchio and Siciliano.[8] References on the augmented task space approach are Egeland,[20] Samson et al.,[48] Sciavicco and Siciliano,[50] and Seraji.[52] The occurrence of artificial singularities was pointed out in Baillieul,[2] and their properties were studied in Chiacchio et al.[6] The task priority strategy was originally proposed in Nakamura et al.[41] and has recently been refined in Chiaverini[11] concerning robustness to artificial singularities. The use of the Jacobian transpose for the constraint task was presented in Chiaverini et al.[15] and Siciliano.[53] The expression of the end-effector orientation error based on an angle/axis description of orientation is due to Luh et al.[35] and its properties were studied in Lin.[34] The use of a quaternion-based orientation error is due to Yuan.[61] More about the possible definitions of the orientation error can be found in Caccavale et al.[5]

References

1. Angeles, J., *Spatial Kinematic Chains: Analysis, Synthesis, Optimization,* Springer-Verlag, Berlin, 1982.
2. Baillieul, J., Kinematic programming alternatives for redundant manipulators, in *Proc. 1985 IEEE Int. Conf. Robotics and Automation,* St. Louis, MO, 1985, 722.
3. Bottema, O. and Roth, B., *Theoretical Kinematics,* North Holland, Amsterdam, 1979.
4. Boullion, T. L. and Odell, P. L., *Generalized Inverse Matrices,* Wiley, New York, 1971.
5. Caccavale, F., Natale, C., Siciliano, B., and Villani, L., Resolved-acceleration control of robot manipulators: A critical review with experiments, *Robotica,* 16, 565, 1998.
6. Chiacchio, P., Chiaverini, S., Sciavicco, L., and Siciliano, B., Closed-loop inverse kinematics schemes for constrained redundant manipulators with task space augmentation and task priority strategy, *Int. J. Robotics Res.,* 10, 410, 1991.
7. Chiacchio, P., and Siciliano, B., Achieving singularity robustness: An inverse kinematic solution algorithm for robot control, in *Robot Control: Theory and Applications,* IEE Control Engineering Series 36, Warwick, K. and Pugh, A., Eds., Peter Peregrinus, Herts, U.K., 149, 1988.
8. Chiacchio, P. and Siciliano, B., A closed-loop Jacobian transpose scheme for solving the inverse kinematics of nonredundant and redundant robot wrists, *J. Robotic Systems,* 6, 601, 1989.
9. Chiaverini, S., Inverse differential kinematics of robotic manipulators at singular and near-singular configurations, in *Prepr. 1992 IEEE Int. Conf. Robotics Automation — Tutorial on Redundancy: Performance Indices, Singularities Avoidance, and Algorithmic Implementations,* Nice, 1992.
10. Chiaverini, S., Estimate of the two smallest singular values of the Jacobian matrix: Application to damped least-squares inverse kinematics, *J. Robotic Systems,* 10, 991, 1993.

© 2002 by CRC Press LLC

11. Chiaverini, S., Singularity-robust task-priority redundancy resolution for real-time kinematic control of robot manipulators, *IEEE Trans. Robotics Automation,* 13, 398, 1997.
12. Chiaverini, S., Egeland, O., and Kanestrøm, R. K., Achieving user-defined accuracy with damped least-squares inverse kinematics, in *Proc. 5th Int. Conf. Advanced Robotics,* Pisa, I, 672, 1991.
13. Chiaverini, S., Egeland, O., Sagli, J. R., and Siciliano, B., User-defined accuracy in the augmented task space approach for redundant manipulators, *Lab. Robotics Automation,* 4, 59, 1992.
14. Chiaverini, S., Sciavicco, L., and Siciliano, B., Control of robotic systems through singularities, in *Advanced Robot Control,* Lecture Notes in Control and Information Science 162, Canudas de Wit, C., Ed., Springer-Verlag, Berlin, 285, 1991.
15. Chiaverini, S., Siciliano, B., and Egeland, O., Redundancy resolution for the human-arm-like manipulator, *Robotics Autonomous Systems,* 8, 239, 1991.
16. Chiaverini, S., Siciliano, B., and Egeland, O., Review of the damped least-squares inverse kinematics with experiments on an industrial robot manipulator, *IEEE Trans. Control Systems Technology,* 2, 123, 1994.
17. Craig, J. J., *Introduction to Robotics: Mechanics and Control,* 2nd ed., Addison-Wesley, Reading, MA, 1989.
18. Denavit, J. and Hartenberg, R. S., A kinematic notation for lower-pair mechanisms based on matrices, *ASME J. Appl. Mech.,* 22, 215, 1955.
19. Dombre, E. and Khalil, W., *Modélisation et Commande des Robots,* Hermès, Paris, 1988.
20. Egeland, O., Task-space tracking with redundant manipulators, *IEEE J. Robotics Automation,* 3, 471, 1987.
21. Egeland, O., Sagli, J. R., Spangelo, I., and Chiaverini, S., A damped least-squares solution to redundancy resolution, in *Proc. 1991 IEEE Int. Conf. Robotics Automation,* Sacramento, CA, 945, 1991.
22. Featherstone, R., Position and velocity transformations between robot end-effector coordinates and joint angles, *Int. J. Robotics Res.,* 2(2), 35, 1983.
23. Goldenberg, A. A., Benhabib, B., and Fenton, R. G., A complete generalized solution to the inverse kinematics of robots, *IEEE J. Robotics Automation,* 1, 14, 1985.
24. Hollerbach, J. M., Wrist-partitioned inverse kinematic accelerations and manipulator dynamics, *Int J. Robotics Res.,* 2(4), 61, 1983.
25. Hunt, K. H., *Kinematic Geometry of Mechanisms,* Clarendon, Oxford, U.K., 1978.
26. Khalil, W., A system for generating the symbolic models of robots, in *Postpr. 4th IFAC Symp. Robot Control,* Capri, 416, 1994.
27. Khalil, W. and Bennis, F., Automatic generation of the inverse geometric model of robots, *Robotics and Autonomous Systems,* 7, 1, 1991.
28. Khalil, W. and Kleinfinger, J. F., A new geometric notation for open and closed-loop robots, in *Proc. 1986 IEEE Int. Conf. Robotics Automation,* San Francisco, CA, 1174, 1986.
29. Khatib, O., A unified approach for motion and force control of robot manipulators: The operational space formulation, *IEEE J. Robotics Automation,* 3, 43, 1987.
30. Klein, C. A. and Huang, C. H., Review of pseudoinverse control for use with kinematically redundant manipulators, *IEEE Trans. Systems, Man, Cybernetics,* 13, 245, 1983.
31. Klema, V. C. and Laub, A. J., The singular value decomposition: Its computation and some applications, *IEEE Trans. Automatic Control,* 25, 164, 1980.
32. Lee, H. Y. and Liang, C. G., Displacement analysis of the general 7-link 7R mechanism, *Mechanism Machine Theory,* 23, 219, 1988.
33. Liégeois, A., Automatic supervisory control of the configuration and behavior of multibody mechanisms, *IEEE Trans. Systems, Man, Cybernetics,* 7, 868, 1977.
34. Lin, S. K., Singularity of a nonlinear feedback control scheme for robots, *IEEE Trans. Systems, Man, Cybernetics,* 19, 134, 1989.
35. Luh, J. Y. S., Walker, M. W., and Paul, R. P. C., Resolved-acceleration control of mechanical manipulators, *IEEE Trans. Automatic Control,* 25, 468, 1980.
36. Maciejewski, A. A. and Klein, C. A., Obstacle avoidance for kinematically redundant manipulators in dynamically varying environments, *Int. J. Robotics Res.,* 4(3), 109, 1985.

© 2002 by CRC Press LLC

37. Maciejewski, A. A. and Klein, C. A., Numerical filtering for the operation of robotic manipulators through kinematically singular configurations, *J. Robotic Systems,* 5, 527, 1988.
38. McCarthy, J. M., *An Introduction to Theoretical Kinematics,* MIT Press, Cambridge, MA, 1990.
39. Nakamura, Y., *Advanced Robotics: Redundancy and Optimization,* Addison-Wesley, Reading, MA, 1991.
40. Nakamura, Y. and Hanafusa, H., Inverse kinematic solutions with singularity robustness for robot manipulator control, *ASME J. Dynamic Systems, Measurement, Control,* 108, 163, 1986.
41. Nakamura, Y., Hanafusa, H., and Yoshikawa, T., Task-priority based redundancy control of robot manipulators, *Int. J. Robotics Res.,* 6(2), 3, 1987.
42. Orin, D. E. and Schrader, W. W., Efficient computation of the Jacobian for robot manipulators, *Int. J. Robotics Res.,* 3(4), 66, 1984.
43. Paul, R. P., *Robot Manipulators: Mathematics, Programming, and Control,* MIT Press, Cambridge, MA, 1981.
44. Paul, R. P. and Zhang, H., Computationally efficient kinematics for manipulators with spherical wrists based on the homogeneous transformation representation, *Int. J. Robotics Res.,* 5(2), 32, 1986.
45. Pieper, D. L., *The Kinematics of Manipulators under Computer Control,* memo. AIM 72, Stanford Artificial Intelligence Laboratory, 1968.
46. Raghavan, M. and Roth, B., Inverse kinematics of the general 6R manipulator and related linkages, *ASME J. Mechanical Design,* 115, 502, 1990.
47. Renaud, M., Calcul de la matrice jacobienne necessaire à la commande coordonnee d'un manipulateur, *Mechanism and Machine Theory,* 15, 81, 1980.
48. Samson, C., Le Borgne, M., and Espiau, B., *Robot Control: The Task Function Approach,* Oxford Engineering Science Series 22, Clarendon, Oxford, U.K., 1991.
49. Sciavicco, L. and Siciliano, B., Coordinate transformation: A solution algorithm for one class of robots, *IEEE Trans. Systems, Man, Cybernetics,* 16, 550, 1986.
50. Sciavicco, L. and Siciliano, B., A solution algorithm to the inverse kinematic problem for redundant manipulators, *IEEE J. Robotics Automation,* 4, 403, 1988.
51. Sciavicco, L. and Siciliano, B., *Modelling and Control of Robot Manipulators,* 2nd ed., Springer, London, 2000.
52. Seraji, H., Configuration control of redundant manipulators: Theory and implementation, *IEEE Trans. Robotics Automation,* 5, 472, 1989.
53. Siciliano, B., Solving manipulator redundancy with the augmented task space method using the constraint Jacobian transpose, in *Prepr. 1992 IEEE Int. Conf. Robotics and Automation — Tutorial on Redundancy: Performance Indices, Singularities Avoidance, and Algorithmic Implementations,* Nice, 1992.
54. Spong, M. W. and Vidyasagar, M., *Robot Dynamics and Control,* Wiley, New York, 1989.
55. Tsai, L. W. and Morgan, A. P., Solving the kinematics of the most general six- and five-degree-of-freedom manipulators by continuation methods, *ASME J. Mechanisms, Transmission, Automation Design,* 107, 189, 1985.
56. Vukobratović, M., *Introduction to Robotics,* Springer, Berlin, 1989.
57. Vukobratović, M. and Kirćanski, M., *Kinematics and Trajectory Synthesis of Manipulation Robots,* Scientific Fundamentals of Robotics 3, Springer, Berlin, 1986.
58. Wampler, C. W., Manipulator inverse kinematic solutions based on vector formulations and damped least-squares methods, *IEEE Trans. Systems, Man, Cybernetics,* 16, 93, 1986.
59. Whitney, D. E., Resolved motion rate control of manipulators and human prostheses, *IEEE Trans. Man–Machine Systems,* 10, 47, 1969.
60. Yoshikawa, T., Manipulability of robotic mechanisms, *Int. J. Robotics Res.,* 4(2), 3, 1985.
61. Yuan, J. S.-C., Closed-loop manipulator control using quaternion feedback, *IEEE J. Robotics Automation,* 4, 434, 1988.

© 2002 by CRC Press LLC

20

Robot Dynamics

20.1 Fundamentals of Robot Dynamic Modeling
Basic Ideas • Robot Geometry • Equations of Dynamics

20.2 Recursive Formulation of Robot Dynamics
Velocities and Accelerations of Robot Links • Elimination of Reactions — Minimization of Dynamic Model Form • Calculation of Direct and Inverse Dynamics

20.3 Complete Model of Robot Dynamics
Dynamic Model of a DC-Driven Robot • Generalized Form of the Dynamic Model

20.4 Some Applications of Computer-Aided Dynamics
Dynamics and Robot Design • Dynamics in On-Line Control

20.5 Extension of Dynamic Modeling — Some Additional Dynamic Effects
Robot Dynamics — Problems and Research • Dynamics of Robot in Constrained Motion • Robot in Contact with Dynamic Environment • Effects of Elastic Transmissions

Appendix: Calculation of Transformation Matrices

Miomir Vukobratović
Mihajlo Pupin Institute

Viljko Potkonjak
University of Belgrade

We start our discussion on robot dynamics from the standpoint that successful design and control of any system require appropriate knowledge of its behavior. This is certain, but we should discuss what is meant by "appropriate knowledge." Let us consider a robot as an example of a technical system. Appropriate knowledge of its behavior may, but need not, include the mathematical model of its dynamics. In the earlier phases of robotics development, design was not based on the exact calculation of robot dynamics but followed experience from machine design. Control did not take into account many dynamic effects. Large approximations were made to reduce the problem to the well-known theory of automatic control. The undeveloped robot theory could not support a more exact approach. For a long time, the practice of robotics (design, manufacture, and implementation) grew independently of the theory that was too academic. However, this did not stop manufacturers from producing many successful robots.

Presently, the need for complex, precise, and fast robots requires a close connection between theory and practice. Regarding the application of robot dynamics, the main breakthrough was made with the development of computer-aided methods for dynamic modeling.[1-3] Such methods allowed fast and user-friendly calculations of all relevant dynamic effects. In this way dynamic modeling and simulation became the essential tools in robot design. The other possibility for application of robot dynamics is the synthesis of the so-called dynamic control.

In this subsection we first discuss the principles of dynamic modeling, the approach to the description of dynamics, and the derivation of the mathematical model. Then, special attention is

© 2002 by CRC Press LLC

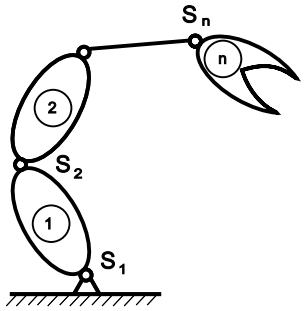


FIGURE 20.1 Robot as a simple and open chain.

paid to computer-based methods. To complete the information on robot dynamics, the mechanism model should be supplemented with the driving system model. After that, we briefly describe the application of the dynamic model. One of the promising directions is the development of CAD systems for robots. The other is dynamic control. Some extension of robot dynamics is made and discuss different effects that were not included in the initial model, trying to locate those of main importance (contact problems, elastic deformations, friction, impact, etc.).

20.1 Fundamentals of Robot Dynamic Modeling

20.1.1 Basic Ideas

From the notion *dynamic modeling* we understand the system of differential equations that describes robotic dynamic behavior. We expect the reader to possess the knowledge necessary to understand the derivation of the model. However, we will try to give enough information at an adequate level of presentation to allow readers to follow the text easily.

Here we consider a manipulation robot as an open and simple kinematic chain (as shown in Figure 20.1) consisting of n rigid bodies (robot links) interconnected by means of n one-degree-of-freedom (one-DOF) joints. A joint allows one relative rotation (revolute joint) or one relative translation (linear joint). Because the complete chain has n DOFs, its dynamics can be described by means of n differential equations of motion. They are second-order equations. This set is called the *dynamic model*.

Several approaches have been used to describe system dynamics: laws of linear momentum and angular momentum.[1-4] Lagrange's equations,[5,6] and Gauss' principle.[7,8] All approaches lead to the same dynamic model but the model formation procedure is different. Here, we use the laws of linear momentum and angular momentum. This approach is often called Newton–Euler equations. In the authors' opinion it is the most appropriate for the majority of readers.

Let us introduce one position coordinate for each joint, angle in the revolute joint, and longitudinal displacement in the linear joint. This set of coordinates uniquely describes the position of the chain. We usually call this set the *internal coordinates* (or joint coordinates, or generalized coordinates). If the coordinate for joint S_j is marked by q_j, then the complete position vector is

$$q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}^T \tag{20.1}$$

20.1.2 Robot Geometry

At this point we have to decide the mathematical presentation of robot geometry and kinematics. Up to now, two ways have been defined. One is based on the Rodrigues' formulae of finite rotation and the other uses the Denavit–Hartenberg parameters. The latter method is more widely accepted

© 2002 by CRC Press LLC

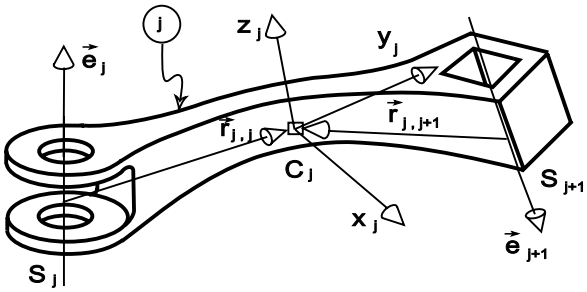


FIGURE 20.2 Geometry of a link.

because it allows simpler expression of transformation matrices and probably faster calculation of robot kinematics. However, if the intention is to discuss dynamics, it should be stressed that the first method is more appropriate. It is more general and follows the rigid body motion approach used in all standard textbooks on mechanics. For these reasons, we utilize the method based on Rodrigues' formulae.

Figure 20.2 shows one link of the robot chain, the j-th one. Joint S_j is shown as revolute and S_{j+1} as linear. To define the link geometry, it is necessary to describe the position and the orientation of the joints with respect to the mass center (MC).

The motion direction in each joint is defined by means of an axis, that is, by a unit vector. It can describe rotation or translation, depending on the type of joint. Thus $\vec{e}_j$ corresponds to joint S_j and $\vec{e}_{j+1}$ to S_{j+1}. The relative position of MC with respect to the joints is defined by means of vectors $\vec{r}_{j,j}$ and $\vec{r}_{j,j+1}$ as shown in the Figure 20.2. MC is marked by C_j.

During robot motion, positions of all links, and accordingly, geometry vectors expressed in the immobile external frame, change. However, if geometry vectors are considered relative to the corresponding link, they become constant and represent the property of the link itself. To express these constant values, we introduce a Cartesian system fixed to the link and with the origin in the MC (link-fixed frame). The axes are x_j, y_j, and z_j. The system may be oriented in an arbitrary way but is most suitable if its axes coincide with the so-called principal axes of inertia. Consider now vector $\vec{e}_j$. It can be expressed by means of three constant projections onto the axes of the frame fixed to the link j: $e_j^{x_j}$, $e_j^{y_j}$, and $e_j^{z_j}$. For this triple we introduce the notation

$$\tilde{\vec{e}}_j = \left(e_j^{x_j},\ e_j^{y_j},\ e_j^{z_j}\right) \tag{20.2}$$

The tilde "~" above the letter indicates that the vector is expressed in the link-fixed frame. Notation $\vec{e}_j$ (without tilde) denotes three projections onto the axes of an external immobile frame. If the same is applied to vectors $\vec{r}_{j,j}$ and, $\vec{r}_{j,j+1}$, two constant triples are obtained

$$\tilde{\vec{r}}_{j,j} = \left(r_{j,j}^{x_j},\ r_{j,j}^{y_j}, r_{j,j}^{z_j}\right) \tag{20.3}$$

$$\tilde{\vec{r}}_{j,j+1} = \left(r_{j,j+1}^{x_j},\ r_{j,j+1}^{y_j}, r_{j,j+1}^{z_j}\right) \tag{20.4}$$

Vector $\vec{e}_{j+1}$ is constant if expressed in the frame fixed to link j and a suitable notation is needed for these projections. Notation $\tilde{\vec{e}}_{j+1}$ indicates that the vector is considered relative to link $j + 1$ (analogously to relation (20.2)). Hence, a new notation is introduced to indicate the projections onto link j:

$$\tilde{\vec{e}}_{j+1} = \left(e_{j+1}^{x_j}, e_{j+1}^{y_j}, e_{j+1}^{z_j}\right) \tag{20.5}$$

© 2002 by CRC Press LLC

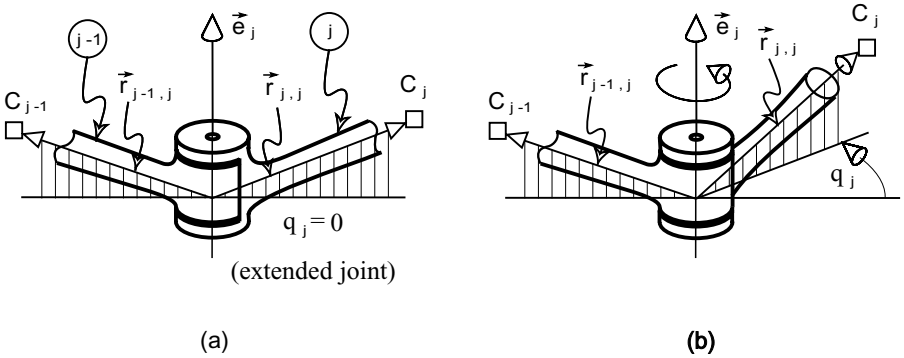


FIGURE 20.3 Definition of a coordinate in a revolute joint.

Generally, for any vector $\vec{a}_j$ having index j, the tilde "~" above the letter ($\tilde{\vec{a}}_j$) indicates the projections onto frame j, while the tilde under the letter ($\underset{\sim}{\vec{a}}_j$) indicates the projections onto the preceding frame, $j - 1$. Notation without the tilde ($\vec{a}_j$) indicates projections onto the external immobile frame.

Now, it should be stated that four vectors, $\tilde{\vec{e}}_j$, $\underset{\sim}{\vec{e}}_{j+1}$, $\tilde{\vec{r}}_{j,j}$, $\tilde{\vec{r}}_{j,j+1}$, define the geometry of link j. To define the geometry of the complete chain, one has to prescribe these four vectors for all links.

It is still necessary to distinguish between the revolute and linear joints. For this purpose we introduce the indicator s_j for each joint:

$$s_j = \begin{cases} 0, & \text{if } S_j \text{ is a revolute joint} \\ 1, & \text{if } S_j \text{ is a linear joint.} \end{cases} \tag{20.6}$$

Now, it is possible to define the joint coordinates more precisely. We consider the revolute joints first. If S_j is a revolute joint, then coordinate q_j represents the angle of rotation measured from the extended position. The exact definition is shown in Figure 20.3. The angle lies in a plane perpendicular to axis $\vec{e}_j$. The negative projection of $\vec{r}_{j-1,j}$ defines the extended position ($q_j = 0$) and the angle is measured to the projection of $\vec{r}_{j,j}$. Figure 20.3a shows the extended position and Figure 20.3b the rotated position.

Suppose now that joint S_j is linear. Coordinate q_j defines the length of translation along $\vec{e}_j$ and its precise definition requires previous introduction of the zero position. This zero-point can be adopted anywhere on the axis of translation. It is marked by S'_j in Figure 20.4. Once adopted, this point determines the vector $\vec{r}_{j-1,j}$. Coordinate q_j is defined as the displacement $\overrightarrow{S'_j S''_j}$ with the proper sign with respect to $\vec{e}_j$ (see Figure 20.4).

It is necessary to introduce an additional vector. $\overrightarrow{S''_j C_j} = \vec{r}'_{j,j}$. It follows from Figure 20.4. that

$$r'_{j,j} = \overrightarrow{S'_j S''_j} + \overrightarrow{S''_j C_j} = q_j \vec{e}_j + \vec{r}_{j,j} \tag{20.7}$$

or, more generally,

$$\vec{r}'_{j,j} = \vec{r}_{j,j} + s_j q_j \vec{e}_j. \tag{20.8}$$

For a linear joint ($s_j = 1$) relation (20.8) becomes (20.7) and for revolute joints ($s_j = 0$) it reduces to $\vec{r}'_{j,j} = \vec{r}_{j,j}$. Thus, expression (20.8), i.e., vector $\vec{r}'_{j,j}$, may replace $\vec{r}_{j,j}$ for any type of joint.

© 2002 by CRC Press LLC

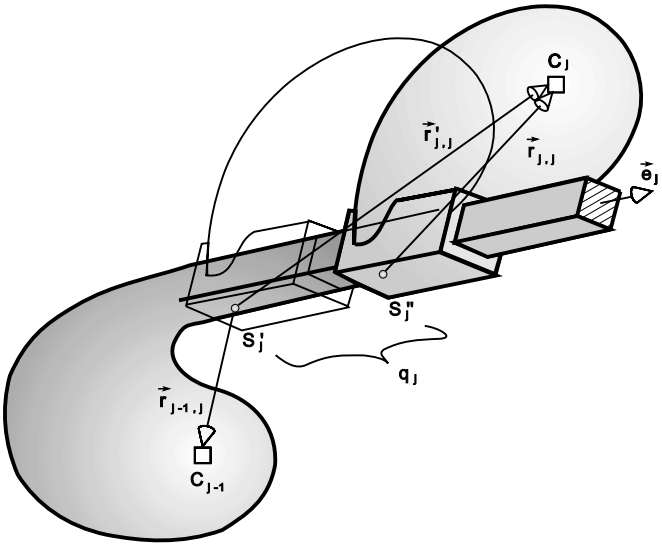


FIGURE 20.4 Definition of coordinate in a linear joint.

After introducing different frames (link-fixed and external immobile) the question arises about the possibility of transforming a vector from one frame to another. Let us consider a vector $\vec{a}_j$. For the transition between the frames, transformation matrices are applied:

From the *j*-th link-fixed frame to the external one

$$\vec{a}_j = A_j \tilde{\vec{a}}_j, \quad \dim A_j = 3 \times 3 \tag{20.9}$$

and in the opposite direction

$$\tilde{\vec{a}}_j = A_j^{-1} \vec{a}_j = A_j^T \vec{a}_j \tag{20.10}$$

From the *j*-th link-fixed frame to the $(j-1)$-th one

$$\underset{\sim}{\vec{a}}_j = A_{j-1,j} \tilde{\vec{a}}_j \tag{20.11}$$

and in the opposite direction

$$\tilde{\vec{a}}_j = A_{j,j-1} \underset{\sim}{\vec{a}}_j = A_{j-1,j}^{-1} \underset{\sim}{\vec{a}}_j = A_{j-1,j}^T \underset{\sim}{\vec{a}}_j . \tag{20.12}$$

It could be noted that the matrices are orthogonal and thus the inverse equals the transpose.

A detailed explanation of how to calculate the transformation matrices is given in this chapter's appendix.

20.1.3 Equations of Dynamics

We start by considering dynamics from one of the links, the *j*-th one. For this purpose we fictively interrupt the chain in joints S_j and S_{j+1}. The disconnection in joint S_j is shown in Figure 20.5a. The

© 2002 by CRC Press LLC

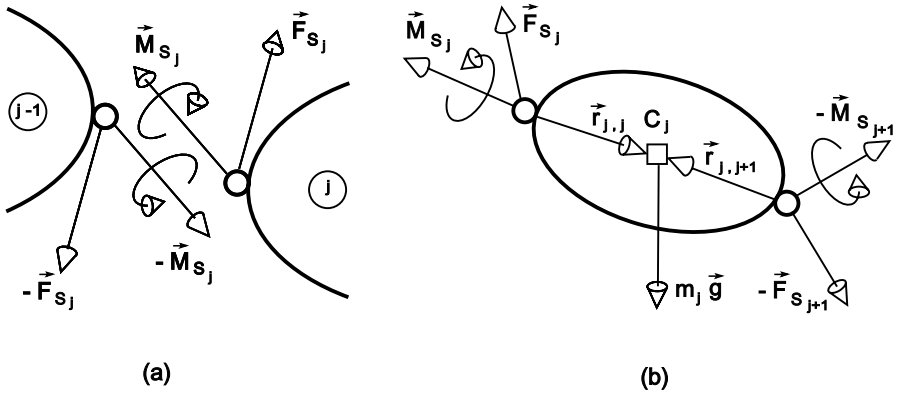


FIGURE 20.5 Extraction of one link from the chain.

mutual influence of the two links is expressed in terms of a force and a couple. $\vec{F}_{S_j}$ denotes the force and $\vec{M}_{S_j}$ denotes the moment of the couple. They act in the forward direction (from link $j-1$ to link j) while $-\vec{F}_{S_j}$ and $-\vec{M}_{S_j}$ act in the backward direction (from j to $j-1$).

Figure 20.5b shows the link j together with all forces and couples acting upon it. We use the law of linear momentum (Newton's law) to describe the motion of MC:

$$m_j \vec{w}_j = m_j \vec{g} + \vec{F}_{S_j} - \vec{F}_{S_{j+1}} \tag{20.13}$$

where m_j is the link mass, $\vec{w}_j$ is MC acceleration, and $\vec{g}$ is the acceleration due to gravity (9.81 m/s^2). The law expresses the equilibrium of the inertial and real forces.

Now, we discuss the rotation of the link about its MC. It can be described by the law of angular momentum (Euler's equations):

$$\tilde{J}_j \tilde{\vec{\varepsilon}}_j + \tilde{\vec{\omega}}_j \times \left(\tilde{J}_j \tilde{\vec{\omega}}_j \right) = \tilde{\vec{M}}_{S_j} - \tilde{\vec{M}}_{S_{j+1}} - \tilde{\vec{r}}'_{j,j} \times \tilde{\vec{F}}_{S_j} + \tilde{\vec{r}}_{j,j+1} \times \tilde{\vec{F}}_{S_{j+1}} \tag{20.14}$$

where $\tilde{\vec{\varepsilon}}_j$ and $\tilde{\vec{\omega}}_j$ are the angular acceleration and angular velocity. The tilde "~" indicates that the vectors are expressed in the frame fixed to the link j. $\tilde{J}_j$ is the tensor of inertia calculated for the axes of the link-fixed frame. In a general case, the tensor has the form

$$\tilde{J}_j = \begin{bmatrix} J_{xx_j} & J_{xy_j} & J_{xz_j} \\ J_{yx_j} & J_{yy_j} & J_{yz_j} \\ J_{zx_j} & J_{zy_j} & J_{zz_j} \end{bmatrix} \tag{20.15}$$

However, if the frame axes are placed to coincide with the principal inertial axes, then the tensor takes the diagonal form

$$\tilde{J}_j = \begin{bmatrix} J_{x_j} & & \\ & J_{y_j} & \\ & & J_{z_j} \end{bmatrix} \tag{20.16}$$

where J_{x_j} is the inertial moment with respect to axis x_j and analogously holds for J_{y_j} and J_{z_j}.

© 2002 by CRC Press LLC

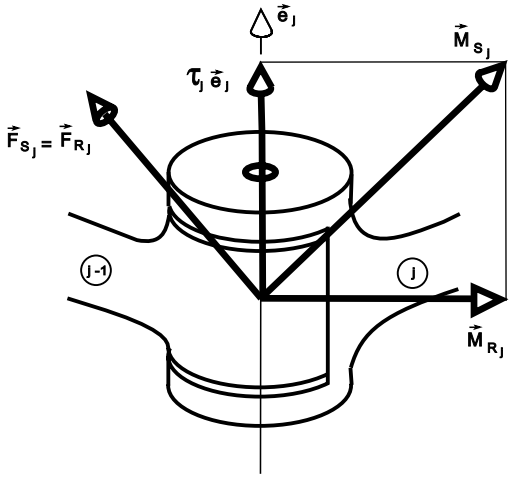


FIGURE 20.6 The total force $(\vec{F}_{S_j})$ and the total couple $(\vec{M}_{S_j})$ in a revolute joint.

It should be mentioned that Equation (20.13) is written in the external frame, while (20.14) is written in the link-fixed frame. Equation (20.13) could be simply rewritten and expressed in the link-fixed frame, while (20.14) cannot change the frame so easily. It would have to be multiplied by transformation matrix A_j.

We now discuss force $\vec{F}_{S_j}$ and couple $\vec{M}_{S_j}$ transmitted through the revolute joint S_j (Figure 20.6). First, we consider the "pure" reactions, force and couple, that follow from the mechanical connection of the two links. Because the joint permits one rotation (about $\vec{e}_j$), the reaction force $\vec{F}_{R_j}$ (from link $j-1$ to link j) may be of arbitrary direction, while the reaction couple $\vec{M}_{R_j}$ is perpendicular to axis $\vec{e}_j$ (that is, $\vec{M}_{R_j} \perp \vec{e}_j$). In the revolute joint a driving torque τ_j acting about axis $\vec{e}_j$ exists. In the vector form, the drive is $\tau_j \vec{e}_j$. Thus, the total force and couple transmitted through the joint are

$$\vec{F}_{S_j} = \vec{F}_{R_j}$$

$$\vec{M}_{S_j} = \vec{M}_{R_j} + \tau_j \vec{e}_j. \tag{20.17}$$

Consider now a linear joint S_j (Figure 20.7). The total force $\vec{F}_{S_j}$ transmitted through the joint has two components, the reaction force $\vec{F}_{R_j}$, and the driving force $\tau_j \vec{e}_j$. The reaction is perpendicular to the axis $\vec{e}_j$ (that is, $\vec{F}_{R_j} \perp \vec{e}_j$). The total couple $\vec{M}_{S_j}$ consists of reaction only and can be of arbitrary direction. Thus, it holds that

$$\vec{F}_{S_j} = \vec{F}_{R_j} + \tau_j \vec{e}_j$$

$$\vec{M}_{S_j} = \vec{M}_{R_j} \tag{20.18}$$

Equations (20.17) and (20.18) can be written in a unique way that fits both the revolute and the linear joints:

$$\vec{F}_{S_j} = \vec{F}_{R_j} + s_j \tau_j \vec{e}_j$$

$$\vec{M}_{S_j} = \vec{M}_{R_j} + (1 - s_j)\tau_j \vec{e}_j. \tag{20.19}$$

© 2002 by CRC Press LLC

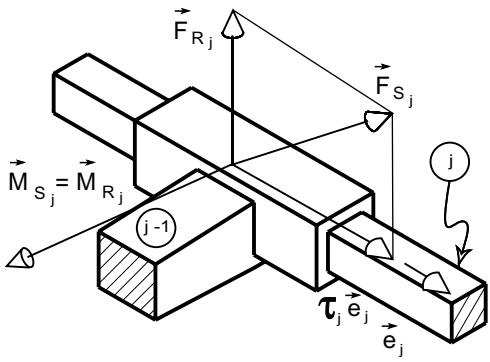


FIGURE 20.7 The total force $(\vec{F}_{S_j})$ and the total couple $(\vec{M}_{S_j})$ in a linear joint.

where s_j indicates the type of joint (see (20.6)).

Equations (20.13) and (20.14), together with (20.19), describe the dynamics of link j. If we apply these equations to all links, we arrive to the system of $2n$ vector equations describing the dynamics of the complete chain:

$$m_j \vec{w}_j = m_j \vec{g} + \vec{F}_{R_j} + s_j \tau_j \vec{e}_j - \vec{F}_{R_{j+1}} - s_{j+1} \tau_{j+1} \vec{e}_{j+1}$$

$$\tilde{J}_j \vec{\tilde{\varepsilon}}_j + \vec{\tilde{\omega}}_j \times \left(\tilde{J}_j \vec{\tilde{\omega}}_j \right) = \vec{\tilde{M}}_{R_j} + \left(1 - s_j\right) \tau_j \vec{\tilde{e}}_j - \vec{\tilde{M}}_{R_{j+1}} - \left(1 - s_{j+1}\right) \tau_{j+1} \vec{\tilde{e}}_{j+1}$$

$$- \vec{\tilde{r}}'_{j,j} \times \left(\vec{\tilde{F}}_{R_j} + s_j \tau_j \vec{\tilde{e}}_j \right) + \vec{\tilde{r}}_{j,j+1} \times \left(\vec{\tilde{F}}_{R_{j+1}} + s_{j+1} \tau_{j+1} \vec{\tilde{e}}_{j+1} \right)$$

$$j = 1, \ldots, n \tag{20.20}$$

Our intention is to find the dynamic model in its minimal form. It should directly connect the input (drives τ_j) and the output (motions q_j). First, it is necessary to express all kinematic variables (velocities and accelerations) in terms of the joint coordinates (q_j) and their derivatives ($\dot{q}_j$, $\ddot{q}_j$). After that we eliminate all the reactions, $\vec{F}_{R_j}$ and $\vec{M}_{R_j}$, $j = 1, \ldots, n$. Let us analyze the number of equations and the number of unknowns that should be eliminated. System (20.20) consists of $2n$ vector equations, that is, $6n$ scalar equations. The number of reaction vectors is also $2n$, two in each joint. However, the number of unknown scalar components for elimination is $5n$, that is, five in each joint. This is because the reaction vector with an arbitrary direction has three unknown scalar components, while the one perpendicular to the joint axis has two. Thus, after eliminating $5n$ unknowns from $6n$ equations, we obtain a system of n scalar equations that does not contain reactions, but only the drives $\tau_1, \ldots, \tau_n$. The methodology for elimination of joint reactions represents the characteristics of each method for dynamic modeling. One method will be explained in the next paragraph.

Regardless of the method chosen for dynamic modeling, a set of differential equations is obtained. Equations are linear with respect to the second derivatives:

$$\begin{aligned} H_{11}(q)\ddot{q}_1 \quad &+ \quad \cdots + H_{1n}(q)\ddot{q}_n + h_1(q,\dot{q}) = \tau_1 \\ &\vdots \\ H_{n1}(q)\ddot{q}_1 \quad &+ \quad \cdots + H_{nn}(q)\ddot{q}_n + h_n(q,\dot{q}) = \tau_n. \end{aligned} \tag{20.21}$$

© 2002 by CRC Press LLC

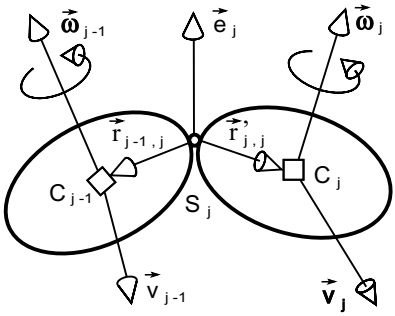


FIGURE 20.8 Velocities of robot links.

Notation $H_{ij}(q)$ indicates that the coefficient depends on all elements of the joint position vector $q = [q_1 \ldots q_n]^T$, and $h_j(q, \dot{q})$ indicates that the free member depends on the coordinates q and their derivatives $\dot{q}$. System (20.21) may be written in matrix form

$$H(q)\ddot{q} + h(q, \dot{q}) = \tau \tag{20.22}$$

where $H = [H_{ij}]_{\dim = n \times n}$ is called the inertial matrix, $h = [h_j]_{\dim = n \times 1}$ takes into account the gravity, centrifugal, and Coriolis effects, and $\tau = [\tau_1 \ldots \tau_n]^T$ is the vector of drives.

Let us discuss drives τ a little more. τ_j represents the driving torque about the joint shaft if the joint is revolute, or the force along the sliding axis if the joint is linear. In any case, the drive is produced by some actuation system (electric, hydraulic, or pneumatic) and then transmitted to the joint (by means of gears, chain, belt, etc.). Thus, τ_j is the output action of an actuator-plus-transmission assembly. This means that the previously derived model (Equation 20.22) describes only one part of robot, its mechanism, while the actuation system model is still missing.

20.2 Recursive Formulation of Robot Dynamics

In the previous paragraph we described robot dynamics by means of a set of $2n$ vector equations (model (20.20)). The model included the reaction forces and couples ($\vec{F}_{R_j}$ and $\vec{M}_{R_j}$) and their elimination was recognized as essential for the minimization of the model form. Here, we use a recursive approach to kinematics and dynamics to carry out this elimination. The method is specially suitable for the creation of a computer procedure for dynamic modeling.

20.2.1 Velocities and Accelerations of Robot Links

Position and speed define the state of each link of a kinematic chain and accordingly the state of the entire chain. The position was discussed previously and it was found that the joint coordinates $q = [q_1 \ldots q_n]^T$ describe the chain position in the most appropriate way. Now, we are going to discuss speed.

If we consider the spatial motion of the link j, we find that it is characterized by the velocity of its MC (e.g., of point C_j in Figure 20.8) and the angular velocity. Let $\vec{v}_j$ and $\vec{\omega}_j$ be these velocities. Our intention is to express these quantities in terms of joint coordinates q and joint velocities $\dot{q}$. The recursive approach will be applied, and hence, we consider two links, $j - 1$ and j, interconnected by means of joint S_j (Figure 20.8).

The angular velocity is obtained by the superposition of rotations. If S_j is a linear joint (indicator $s_j = 1$), it does not contribute to rotation and if it is revolute ($s_j = 0$), its contribution is $\dot{q}_j \vec{e}_j$. Thus, we may write a general expression

$$\vec{\omega}_j = \vec{\omega}_{j-1} + \dot{q}_j(1 - s_j)\vec{e}_j. \tag{20.23}$$

© 2002 by CRC Press LLC

showing the recursive character of angular velocity. Making the derivative of (20.23) one obtains the expression for acceleration:

$$\vec{\varepsilon}_j = \vec{\varepsilon}_{j-1} + (\ddot{q}_j \vec{e}_j + \dot{q}_j(\vec{\omega}_j \times \vec{e}_j))(1 - s_j) \tag{20.24}$$

For the MC position vector, $\vec{r}_{c_j}$, the recursive expression holds:

$$\vec{r}_{c_j} = \vec{r}_{cj-1} - \vec{r}_{j-1,j} + \vec{r}'_{j,j} \tag{20.25}$$

from which the first derivative gives MC velocities

$$\vec{v}_j = \vec{v}_{j-1} - \vec{\omega}_{j-1} \times \vec{r}_{j-1,j} + \vec{\omega}_j \times \vec{r}'_{j,j} + \dot{q}_j s_j \vec{e}_j \tag{20.26}$$

and the second derivative gives the MC accelerations

$$\begin{aligned} \vec{w}_j = \vec{w}_{j-1} - \vec{\varepsilon}_{j-1} \times \vec{r}_{j-1,j} - \vec{\omega}_{j-1} \times (\vec{\omega}_{j-1} \times \vec{r}_{j-1,j}) + \vec{\varepsilon}_j \times \vec{r}'_{j,j} \\ + \vec{\omega}_j \times (\vec{\omega}_j \times \vec{r}'_{j,j}) + (\ddot{q}_j \vec{e}_j + 2\dot{q}_j(\vec{\omega}_j \times \vec{e}_j)) s_j \end{aligned} \tag{20.27}$$

All the vectors in Equations (20.23) to (20.27) are expressed in the external immobile frame. Note that it is possible to transform the equations to the link-fixed frame j or $j - 1$.

We now express the velocities and accelerations of link j in terms of coordinates q and derivatives $\dot{q}$ and $\ddot{q}$. Velocities $\vec{\omega}_j$ and $\vec{v}_j$ represent linear forms with respect to joint velocities:

$$\vec{\omega}_j = \sum_{k=1}^{j} \vec{\lambda}_k^j \dot{q}_k \tag{20.28}$$

$$\vec{v}_j = \sum_{k=1}^{j} \vec{\xi}_k^j \dot{q}_k \tag{20.29}$$

where j in $\vec{\lambda}_k^j$ and $\vec{\xi}_k^j$ represents an upper index and not an exponent. Accelerations of the link are linear forms with respect to the joint accelerations:

$$\vec{\varepsilon}_j = \sum_{k=1}^{j} \vec{\lambda}_k^j \ddot{q}_k + \vec{\gamma}^j \tag{20.30}$$

$$\vec{w}_j = \sum_{k=1}^{j} \vec{\xi}_k^j \ddot{q}_k + \vec{\delta}^j \tag{20.31}$$

We now turn to matrix notation. For this reason we introduce a 3×1 matrix for each vector and use a proper notation. For instance, ω_j denotes the 3×1 matrix corresponding to vector $\vec{\omega}_j$, and analogously holds for all other vectors (we simply omit "$\rightarrow$"). Relations (20.28) to (20.31) can now be written in the form

$$\omega_j = \Lambda^j \dot{q} \tag{20.32}$$

© 2002 by CRC Press LLC

$$v_j = \Xi^j \dot{q} \tag{20.33}$$

$$\varepsilon_j = \Lambda^j \ddot{q} + \Gamma^j \tag{20.34}$$

$$w_j = \Xi^j \ddot{q} + \Delta^j \tag{20.35}$$

where Λ^j and Γ^j are $3 \times n$ and 3×1 matrices (respectively) containing the coefficients and the free member of the linear forms (20.28) and (20.30):

$$\Lambda^j = [\lambda_1^j \cdots \lambda_j^j 0 \cdots 0]_{\dim=3\times n} \tag{20.36}$$

$$\Gamma^j = [\gamma^j]_{\dim=3\times 1} \tag{20.37}$$

and Ξ^j and Δ^j are $3 \times n$ and 3×1 matrices that contain the coefficients and the free member of (20.29) and (20.31):

$$\Xi^j = [\xi_1^j \ldots \xi_j^j 0 \ldots 0]_{\dim=3\times n} \tag{20.38}$$

$$\Delta^j = [\delta^j]_{\dim=3\times 1} \tag{20.39}$$

Matrices Λ, Γ, Ξ, and Δ are very often written without the upper index j. This is because in software realization one variable is used for each matrix and it is modified when a new index j is considered. Thus, j is the index of iteration. Now, let us analyze how matrices Λ, Γ, Ξ and Δ change when j increases. Starting from (20.23) to (20.27), we find that when passing from $j - 1$ to j, the following modifications of the matrices are needed:

$$\begin{cases} \vec{\lambda}_k^j = \vec{\lambda}_k^{j-1}, k = 1, \ldots, j-1 \\ \vec{\lambda}_j^j = (1 - s_j)\vec{e}_j \end{cases} \tag{20.40}$$

$$\vec{\gamma}^j = \vec{\gamma}^{j-1} + \dot{q}_j(\vec{\omega}_j \times \vec{e}_j)(1 - s_j) \tag{20.41}$$

$$\begin{cases} \vec{\xi}_k^j = \vec{\xi}_k^{j-1} - \vec{\lambda}_k^{j-1} \times \vec{r}_{j-1,j} + \vec{\lambda}_k^j \times \vec{r}'_{j,j}, k = 1, \ldots, j-1 \\ \vec{\xi}_j^j = \vec{e}_j s_j + \vec{\lambda}_j^j \times \vec{r}'_{j,j} \end{cases} \tag{20.42}$$

$$\vec{\delta}^j = \vec{\delta}^{j-1} - \vec{\gamma}^{j-1} \times \vec{r}_{j-1,j} + \vec{\gamma}^j \times \vec{r}'_{j,j} - \vec{\omega}_{j-1} \times (\vec{\omega}_{j-1} \times \vec{r}_{j-1,j}) + \vec{\omega}_j \times (\vec{\omega}_j \times \vec{r}'_{j,j}) + 2\vec{\omega}_j \times \vec{e}_j s_j \dot{q}_j \tag{20.43}$$

Thus, it is possible to form matrices Λ, Γ, Ξ, and Δ in a recursive manner. In each iteration, a new link is added to the chain (e.g., link j). Transformation matrix A_j is calculated to allow expressing the geometry vectors in the external frame. Now, applying the recursive expressions (20.40) to (20.43), matrices Λ^j, Γ^j, Ξ^j and Δ^j are found starting from Λ^{j-1}, Γ^{j-1}, Ξ^{j-1}, and Δ^{j-1}.

© 2002 by CRC Press LLC

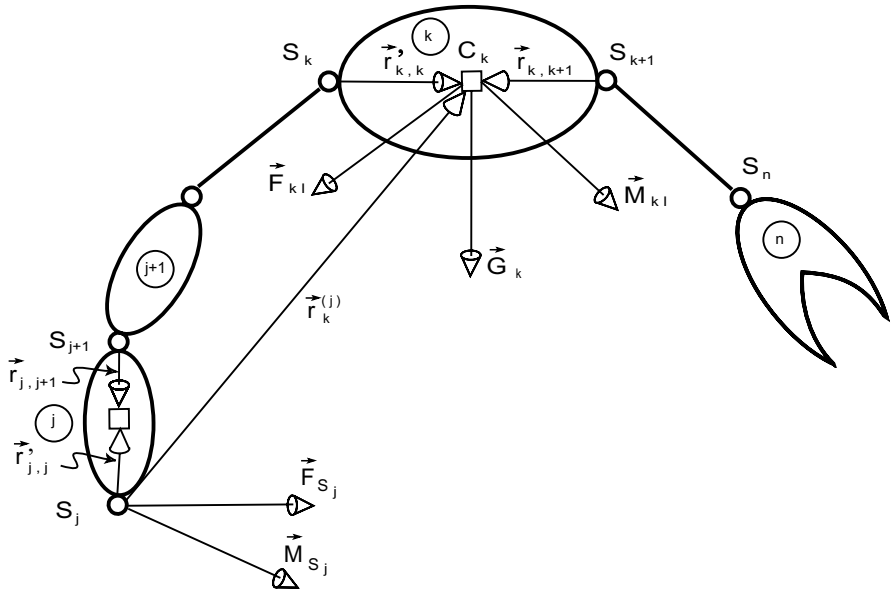


FIGURE 20.9 A part of the robot chain.

20.2.2 Elimination of Reactions — Minimization of Dynamic Model Form

Dynamics of the robot chain was described by means of $2n$ vector Equations (20.20). It was stressed that reactions in joints should be eliminated to reduce the dynamic model to its minimal form (20.21), that is, to n scalar equations. Here, we describe the recursive procedure for the elimination. The procedure begins by considering a part of the kinematic chain, from a selected joint to the free end (Figure 20.9). Let the selected joint be S_j. We now consider the subchain that consists of the links $j, j+1, \ldots, n$. The fictive break is made in joint S_j and the influence of the preceding link is expressed in terms of force $\vec{F}_{S_j}$ and couple $\vec{M}_{S_j}$. Now, we apply D'Alambert's principle and establish the equilibrium of the real and inertial forces acting on the considered part of the chain. The system of forces includes gravity forces, inertial forces, joint force $\vec{F}_{S_j}$, and joint couple $\vec{M}_{S_j}$.

The inertial load of a link (let it be link k) is distributed all over it, but can be reduced to a resultant inertial force acting at the MC, and a resultant inertial couple. The force is

$$\vec{F}_{kI} = -m_k \vec{w}_k \tag{20.44}$$

and acts in point C_k; m_k is the link mass. The moment of the couple can be expressed in Euler's form

$$\vec{\tilde{M}}_{kI} = -\left(\tilde{J}_k \vec{\tilde{\varepsilon}}_k + \vec{\tilde{\omega}}_k \times \left(\tilde{J}_k \vec{\tilde{\omega}}_k\right)\right) \tag{20.45}$$

where $\tilde{J}_k$ is the tensor of inertia. Couple (20.45) is expressed in the link-fixed frame and transformation matrix A_k could be used to transfer it to the external frame: $\vec{M}_{kI} = A_k \vec{\tilde{M}}_{kI}$. The gravity load of the link is

$$\vec{G}_k = m_k \vec{g} \ . \tag{20.46}$$

where $\vec{g}$ denotes the gravity acceleration.

D'Alambert's equilibrium concerns $\vec{F}_{kI}, \vec{M}_{kI}, \vec{G}_{k,}\ k = j, \ldots n, \vec{F}_{S_j}, \vec{M}_{S_j}$. Equilibrium of forces gives

© 2002 by CRC Press LLC

$$\vec{F}_{S_j} = -\sum_{k=j}^{n}\left(\vec{G}_k + \vec{F}_{kI}\right) \tag{20.47}$$

Equilibrium of moments of forces is found with respect to point S_j, yielding

$$\vec{M}_{S_j} = -\sum_{k=j}^{n}\left(\vec{M}_{kI} + \vec{r}_k^{\,(j)} \times \left(\vec{G}_k + \vec{F}_{kI}\right)\right) \tag{20.48}$$

where

$$\vec{r}_k^{\,(j)} = \overrightarrow{S_j C_k} = \sum_{p=j}^{k-1}\left(\vec{r}'_{p,p} - \vec{r}_{p,p+1}\right) + \vec{r}'_{k,k} \tag{20.49}$$

Suppose now that joint S_j is linear (indicator $s_j = 1$), as shown in Figure 20.7. According to (20.18), force $\vec{F}_{S_j}$ consists of the driving component $\tau_j \vec{e}_j$ acting along the joint axis and the reaction force $\vec{F}_{R_j}$ perpendicular to that axis. To eliminate the reaction, we multiply Equation (20.47) by the joint axis vector $\vec{e}_j$, obtaining the scalar expression for the driving force

$$\tau_j = \vec{F}_{S_j}\vec{e}_j = \sum_{k=j}^{n} m_k\left(\vec{w}_k - \vec{g}\right)\vec{e}_j. \tag{20.50}$$

where (20.44) and (20.46) are substituted.

In the other case we suppose that S_j is a revolute joint (indicator $s_j = 0$) as shown in Figure 20.6. According to (20.17), the moment of couple $\vec{M}_{S_j}$ consists of the driving torque $\tau_j \vec{e}_j$ acting about the joint axis, and the reaction couple $\vec{M}_{R_j}$ perpendicular to the axis. Multiplication of Equation (20.48) by axis vector $\vec{e}_j$ eliminates the reaction, thus giving the scalar expression for the driving torque

$$\tau_j = \vec{M}_{S_j}\vec{e}_j = \sum_{k=j}^{n}\left(A_k\left(\tilde{J}_k\vec{\tilde{\varepsilon}}_k + \vec{\tilde{\omega}}_k \times \left(\tilde{J}_k\vec{\tilde{\omega}}_k\right)\right) + \vec{r}_k^{\,(j)} \times m_k\left(\vec{w}_k - \vec{g}\right)\right)\vec{e}_j\,. \tag{20.51}$$

where expression (20.45) multiplied by A_k, and expressions (20.46) and (20.44) are substituted.

To summarize, in the case of a linear joint, the drive is expressed in form (20.50) and for a revolute joint in form (20.51). This can be applied to joints $j = 1, \ldots, n$. So, dynamics of the robot chain is described by means of n scalar equations expressing the drives. In relations (20.50) and (20.51), MC accelerations, $\vec{w}_k$, and angular accelerations, $\vec{\varepsilon}_k$, appear. These accelerations can be expressed in terms of joint accelerations by using linear forms (20.34) and (20.35). In this way (20.50) is transformed to

$$\tau_j = \sum_{k=j}^{n} e_j^T m_k\left(\Xi^k\ddot{q} + \Delta^k - g\right). \tag{20.52}$$

where vector notation is replaced by 3×1 matrices (omitting "→"). Further transformation yields

$$\tau_j = H'_j\ddot{q} + h'_j \tag{20.53}$$

© 2002 by CRC Press LLC

where

$$H'_j = \sum_{k=j}^{n} e_j^T m_k \Xi^k, \quad h'_j = \sum_{k=j}^{n} e_j^T m_k \left(\Delta^k - g\right) \tag{20.54}$$

are of dimensions $1 \times n$ and 1×1, respectively.

If (20.51) is transformed to depend on joint accelerations (introducing (20.34) and (20.35)), one obtains

$$\tau_j = \sum_{k=j}^{n} e_j^T A_k \left(\tilde{J}_k A_k^{-1}\left(\Lambda^k \ddot{q} + \Gamma^k\right) + \underline{\tilde{\omega}}_k \tilde{J}_k \tilde{\omega}_k\right) + e_j^T \underline{r}_k^{(j)} m_k \left(\Xi^k \ddot{q} + \Delta^k - g\right) \tag{20.55}$$

where for any vector, e.g., $\vec{a}$, notation $\underline{\underline{a}}$ understands the matrix

$$\underline{\underline{a}} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \tag{20.56}$$

that serves to perform the vector product in matrix form. Further transformation of (20.55) yields

$$\tau_j = H''_j \ddot{q} + h''_j \tag{20.57}$$

where

$$H''_j = \sum_{k=j}^{n} e_j^T \left(A_k \tilde{J}_k A_k^{-1} \Lambda^k + \underline{r}_k^{(j)} m_k \Xi^k\right),$$

$$h''_j = \sum_{k=j}^{n} e_j^T \left(A_k \tilde{J}_k A_k^{-1} \Gamma^k + A_k \underline{\tilde{\omega}}_k \tilde{J}_k \tilde{\omega}_k + \underline{r}_k^{(j)} m_k \left(\Delta^k - g\right)\right) \tag{20.58}$$

are of dimensions $1 \times n$ and 1×1, respectively.

In the described way we have found the expressions for the joint drives in terms of joint accelerations. For a linear joint, form (20.53), along with (20.54), holds while for a revolute joint form (20.57), along with (20.58), applies. For the complete chain (all joints), the following matrix relation can be written

$$\tau = H\ddot{q} + h \tag{20.59}$$

where

$$\tau = \begin{bmatrix} \tau_1 \\ \vdots \\ \tau_j \\ \vdots \\ \tau_n \end{bmatrix}, \quad H = \begin{bmatrix} H_1 \\ \vdots \\ H_j \\ \vdots \\ H_n \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ \vdots \\ h_j \\ \vdots \\ h_n \end{bmatrix}, \quad (H_j \text{ and } h_j) = \begin{cases} H'_j \text{ and } h'_j, & \text{for } s_j = 1 \\ H''_j \text{ and } h''_j, & \text{for } s_j = 0 \end{cases} \tag{20.60}$$

© 2002 by CRC Press LLC

are of dimensions: τ ($n \times 1$), H ($n \times n$), h ($n \times 1$). Thus, we have come to the previously stated form of the dynamic model, that is, Equation (20.22). Expressions (20.54) and (20.58) offer the possibility of recursive calculation of matrices H and h.

20.2.3 Calculation of Direct and Inverse Dynamics

The notions of *direct* and *inverse dynamics* are introduced in theoretical mechanics when considering the behavior of a mechanical system under the action of forces. The direct dynamic problem understands that system motion is known and the forces that cause this motion are to be calculated. The opposite problem, calculation of motion for the given forces, is called inverse dynamics. One should note that some authors interchange these terms. Although one might find some justification for this, we keep the definition as given above because "our" inverse dynamics needs the inversion of the inertial matrix, thus making a suitable association. For a robotic chain, direct dynamics means calculating drives τ for the prescribed robot motion $q(t)$, while inverse dynamics understands the calculation of motion q for the given drives τ. Model (20.22) offers the possibility of solving both problems. Here, we give a general approach to problem solution. Details cannot be discussed, because they depend on the used method. For instance, if the dynamic model was found in its symbolic form, the procedure for the solution would differ to some extent from the procedure used with numerical models.

Consider the direct dynamics first. When we say that the motion $q(t)$ is known, it means that velocity $\dot{q}(t)$ and acceleration $\ddot{q}(t)$ are known, too. Hence, direct dynamics generally involves derivation. Note that sometimes the motion is given by directly prescribing the acceleration. This is the case if a trapezoidal velocity profile is required (constant acceleration is followed by a uniform motion, and finally, constant deceleration stops the system). In such cases the solution includes integration to find the velocity and the position. In any case, relation (20.22) is used to calculate the drives τ. Calculation of direct dynamics is needed for several reasons. First, it is possible to obtain the preliminary information about the drive and power requirements for different robot tasks without an actual experiment. This is important in the process of robot design and enables the development of CAD systems for robots. Second, direct dynamics is used to create dynamic control. It leads to the introduction of feedforward to the robot control scheme.

Inverse dynamics understands the integration of the dynamic model (20.22). Initial state (position $q(0)$ and velocity $\dot{q}(0)$) are needed. Most standard procedures for numerical integration require the differential equations systems be written in a canonical form. Hence, model (20.22) is rewritten:

$$\frac{d}{dt}\begin{bmatrix} q \\ \dot{q} \end{bmatrix} = \begin{bmatrix} \dot{q} \\ H^{-1}(q)(\tau - h(q,\dot{q})) \end{bmatrix} \tag{20.61}$$

The calculation of inverse dynamics plays the central role in any simulation system. This fact sufficiently explains the great importance of this problem.

The discussion in Section 20.2 proves that the computer can successfully be used for forming and solving the robot dynamic model. This conclusion is important because any attempt to write the dynamic model of a multiple joint system "by hand" would probably lead to numerous errors. This is because of extreme complexity of the equations that describe the dynamics of spatial multijoint chains. The research efforts in the field of computer-aided dynamic modeling have resulted in several commercially available program packages.[9,10]

20.3 Complete Model of Robot Dynamics

In the previous paragraphs discussion on dynamics was restricted to robot mechanisms. Dynamics was understood as a relation between joint drives τ and joint motions q. Now, we note that with a

© 2002 by CRC Press LLC

real robotic system, joint torques (or forces) cannot be considered as control variables. Some actuators produce them and the actuators introduce their own dynamics. Each actuator has an input control variable and the drive it produces depends on the control as well as on the dynamic behavior of the entire system. Hence, we may say that the joint torques (or forces) τ describe the interaction between the two subsystems that form a robot: the chain and the actuators.

To describe the dynamics of the entire system, it is necessary to discuss the possible actuators. A completely general discussion that would cover all the possible types of actuators would be very extensive, because robots can be equipped with many types of different actuators: DC motors, synchronous AC motors, stepping motors, different electrohydraulic actuators, and pneumatic systems. Our intention is to show the modeling methodology for the dynamics of the entire robot. For this reason we first choose one type of actuator, a DC motor. After completing the model for this particular case, we make some generalizations to cover different types of actuators. More detailed discussion on robot driving system is given in Chapter 21.

20.3.1 Dynamic Model of a DC-Driven Robot

Permanent magnet DC motors are very common actuators with robotic systems. Their main advantage is simple control by varying the input voltage. The main disadvantage, however, follows from graphite brushes. Some other problems that are common characteristics of all electric drives should be mentioned. It is a fact that they rotate fast and produce relatively small torques. Thus, a gear-box is usually needed to reduce speed and increase torque. Further, motors are very often displaced from joints and moved toward the robot base to unload the arm statically. Hence, different transmissions (chains, belts, gears, shafts, etc.) are needed. This complicates robot construction and may influence the accuracy of motion. However, good controllability still makes DC motors very popular.

The dynamics of a DC motor that drives a robot joint S_j is described by the following relations expressing the mechanical and electrical equilibrium:

$$J_j\ddot{\theta}_j = C_{M_j} i_j - B_j\dot{\theta}_j - M_j \tag{20.62}$$

$$u_j = R_j i_j + L_j \frac{di_j}{dt} + C_{E_j}\dot{\theta}_j \tag{20.63}$$

where θ_j is the angle of the motor shaft rotation, i_j is the armature current, M_j is the output torque, and u_j is the input voltage. Motor parameters are J_j the rotor moment of inertia; C_{M_j} and C_{E_j} the constants of torque and counter electromotive force; B_j, the viscous friction coefficient; R_j, the armature resistance; and L_j, inductivity. The dynamic equations can be united to obtain a more compact canonical form:

$$\dot{x}_j = C_j x_j + f_j M_j + d_j u_j \tag{20.64}$$

where the state vector x_j has dimension three. The state vector and the system matrices are

$$x_j = \begin{bmatrix} \theta_j \\ \dot{\theta}_j \\ i_j \end{bmatrix}, \quad C_j = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -B/J & C_M/J \\ 0 & -C_E/L & -R/L \end{bmatrix}, \quad f_j = \begin{bmatrix} 0 \\ -1/J \\ 0 \end{bmatrix}, \quad d_j = \begin{bmatrix} 0 \\ 0 \\ 1/L \end{bmatrix}. \tag{20.65}$$

where, for simplicity, index j is omitted from the elements of the system matrices.

If inductivity L is small enough (it is a rather common case), the term Ldi/dt can be neglected. Equation (20.63) now becomes

© 2002 by CRC Press LLC

$$u_j = R_j i_j + C_{E_j}\dot{\theta}_j \tag{20.66}$$

and the number of state variables reduces to two. The state vector and the system matrices in Equation (20.64) are

$$x_j = \begin{bmatrix} \theta_j \\ \dot{\theta}_j \end{bmatrix}, \quad c_j = \begin{bmatrix} 0 & 0 \\ 0 & -C_M C_E / RJ - B/J \end{bmatrix}, \quad f_j = \begin{bmatrix} 0 \\ -1/J \end{bmatrix}, \quad d_j = \begin{bmatrix} 0 \\ C_M/RJ \end{bmatrix}. \tag{20.67}$$

It was stated that DC motors are usually followed by some transmission system. The transmission defines the relation between the joint coordinate q_j and the motor coordinate θ_j. If the transmission is considered ideal (no backlash, no elastic deformation), the transmission ratio is constant:

$$q_j = \theta_j / N_j \tag{20.68}$$

If the dynamics of the transmission system (inertial properties) and the loss due to friction are neglected, it holds that

$$\tau_j = M_j N_j \tag{20.69}$$

If friction has to be discussed, it is considered through power loss. The efficiency coefficient is introduced: $0 < \eta_j < 1$. Now, Equation (20.69) is modified. If the motion is in the direction of the drive, $N_j \eta_j'$ is used instead of N_j. However, if the motion is opposite to the action of the drive, then N_j / η_j'' applies. Note that η_j' and η_j'' are generally different. The efficiency of a gear-box in the reverse direction is smaller ($\eta_j'' < \eta_j'$).

The dynamics of the robot chain was discussed in Sections 20.1 and 20.2. It was described by means of n scalar Equations (20.21) or by means of the matrix relation (20.22). If the chain is considered as one subsystem of the robot and the actuators as the other, then the complete dynamics can be described by combining the two models: (20.21) for the chain and (20.62), (20.63) for the motors. To simplify the formulation and stress the main dynamic effects, we neglect the inductivity L and friction B (real numerical values justify this approximation). Equations (20.62) and (20.63) now yield

$$J_j \ddot{\theta}_j = \frac{C_{M_j}}{R_j} u_j - \frac{C_{M_j} C_{E_j}}{R_j} \dot{\theta}_j - M_j \tag{20.70}$$

Relations (20.68) and (20.69) that describe the transmission are needed to connect the motor variables and the joint ones. The motor variables in Equation (20.70), θ and M, are replaced by the joint variables, q and τ. The torque τ is then substituted from such a modified relation (20.70) into system (20.21) thus yielding the model

$$\begin{aligned} &H_{j1}\ddot{q}_1 + \ldots + \left(H_{jj} + J_j N_j^2\right)\ddot{q}_j + \ldots + H_{jn}\ddot{q}_n + h_j + \left(C_{M_j} C_{E_j} / R_j\right) N_j^2 \dot{q}_j \\ &= \left(N_j C_{M_j} / R_j\right) u_j, \; j = 1, \ldots, n \end{aligned} \tag{20.71}$$

or in the matrix form

$$H^*(q)\ddot{q} + h^*(q, \dot{q}) = Du \tag{20.72}$$

© 2002 by CRC Press LLC

where

$$H_{kj}^{*}=\begin{cases} H_{kj}, \text{ for } k\neq j, \\ H_{jj}+J_jN_j^2, \text{ for } k=j, \end{cases}, \quad h_j^{*}=h_j+\left(C_{M_j}C_{E_j}/R_j\right)N_j^2\dot{q}_j, \quad D=diag\left[N_jC_{M_j}/R_j\right] \tag{20.73}$$

By comparing models (20.21) and (20.71) one may conclude that the introduction of motor dynamics results in increased proper inertia of the joint (diagonal coefficient H_{jj} is augmented) while dynamic coupling between joints remains the same (nondiagonal coefficients H_{kj} do not change).

20.3.2 Generalized Form of the Dynamic Model

Here we present the procedure for obtaining the complete dynamic robot model for any kind of actuators.[11] The only restriction is that the actuator can be described by a linear model. Let the dynamics of the robot chain be described by model (20.22) and let the actuator for joint S_j have a linear model of the form

$$\dot{x}_j=C_jx_j+f_jM_j+d_ju_j \tag{20.74}$$

where x_j is a state vector of dimension n_j (e.g., $n_j=3$ for a DC motor as shown in Equation (20.65)), M_j is the output torque, and u_j is the control variable. Note that u_j is subject to the saturation-type constraint: $-u_{j_{max}}<u_j<u_{j_{max}}$.

To simplify the derivation we do not discuss the transmission, that is, we assume a direct connection between the actuators and the joints. In this case, the motion of a joint and that of the corresponding actuator are equal and both are defined by means of q_j, and the torques M_j and τ_j (motor and joint) coincide. This simplification does not compromise the generality of presentation, because the transmission ratio can easily be incorporated when needed.

Let the dynamic model of the chain (Equation (20.22)) be rewritten in a canonical form, according to (20.61):

$$\frac{d\zeta}{dt}=\begin{bmatrix} \zeta_2 \\ H^{-1}(\zeta_1)(\tau-(\zeta_1,\zeta_2)) \end{bmatrix} \tag{20.75}$$

where $\zeta_1=q$, $\zeta_2=\dot{q}$, and $\zeta=\left[\zeta_1^T\zeta_2^T\right]^T$. Thus, ζ is the column vector defining the state of the chain and has the dimension $2n$. This equation can further be rewritten:

$$\dot{\zeta}=K(\zeta)+V(\zeta)\tau \tag{20.76}$$

where

$$K(\zeta)=\begin{bmatrix} \zeta_2 \\ -H^{-1}(\zeta_1)h(\zeta_1,\zeta_2) \end{bmatrix}, \quad V(\zeta)=\begin{bmatrix} 0 \\ -H^{-1}(\zeta_1) \end{bmatrix} \tag{20.77}$$

Let k_j elements of vector x_j coincide with elements of ζ, that is, k_j state coordinates of the j-th actuator are already included in the state vector of the chain. Joint position q_j and joint velocity $\dot{q}_j$ are usually part of the state vector x_j. This means that $k_j=2$, and summation over all joint actuators, $\sum_{j=1}^{n}k_j=2n$, covers the complete vector ζ. Note that in a general case q_j and $\dot{q}_j$ need not directly

© 2002 by CRC Press LLC

be the elements of x_j. Instead, linear or nonlinear dependence may exist between q_j, $\dot{q}_j$ and some elements of x_j.

The dynamics of all actuators can be described if Equations (20.74), for $j = 1, \ldots, n$, written in the compact form:

$$\dot{x} = Cx + F\tau + Du \tag{20.78}$$

where vector $x = [x_1^T \ldots x_n^T]^T$, of dimension $N = \sum_{j=1}^{n} n_j$, defines the state of the entire system.

Vector $\tau = [\tau_1 \ldots \tau_n]^T$ contains the drives, and vector $u = [u_1 \ldots u_n]^T$ the control inputs. In addition, $C = diag[C_j]$, $F = diag[f_j]$, and $D = diag[d_j]$.

Now, we are going to unite the model of the chain, Equation (20.22), and the model of the actuators, Equation (20.78). Let us introduce the matrix T_j of dimension $1 \times n_j$ such that $\ddot{q}_j = T_j \dot{x}_j$. For instance, $T_j = [010]$ for the DC drive that has the state $x_j = [q_j \dot{q}_j i_j]^T$. Model (20.22) can now be rewritten:

$$\tau = HT\dot{x} + h \tag{20.79}$$

where $T = diag[T_j]$ is an $n \times N$ matrix. Substituting $\dot{x}$ from (20.78) into (20.79) one obtains

$$\tau = (E_n - HTF)^{-1} (HT(Cx + Du) + h) \tag{20.80}$$

where E_n is the n-dimensional unit matrix. After substituting τ from (20.80) into (20.78), the complete robot dynamics model is obtained in the form

$$\dot{x} = \hat{C}(x) + \hat{D}(x)u. \tag{20.81}$$

The system matrices are

$$\hat{C} = Cx + F(E_n - HTF)^{-1}(HTCx + h), \quad \hat{D} = D + (E_n - HTF)^{-1} HTD \tag{20.82}$$

with dimensions $N \times 1$ and $N \times n$, respectively.

It is clear that the new form of robot dynamics model requires reformulation of the direct and the inverse problem. The direct dynamics understands the calculation of the control u that is needed to produce the prescribed robot motion. Inverse dynamics means the solution of motion for the prescribed control inputs. This latter problem is called simulation.

20.4 Some Applications of Computer-Aided Dynamics

The formulation of computer procedures for modeling robot kinematics and dynamics made possible the implementation of robot theory for practical work in design and control problems. Before that, handwritten dynamics were limited to simple cases and thus could not be successfully applied. Computer-aided kinematics and dynamics can be used to derive more sophisticated control algorithms and, on the other hand, to assist in robot design. Chapter 21 is devoted to the problem of robot design. Control issues are elaborated in Chapter 22. A survey of advanced results in these fields is given there. Hence, we briefly present only some ideas and the principal references. For historical reasons, we discuss design issues first and then control.

20.4.1 Dynamics and Robot Design

Computer-aided kinematics enables the transformation of robot coordinates from internal (joints) to external (end-effector) and vice versa. Computer procedures to calculate dynamics solve the

© 2002 by CRC Press LLC

direct and inverse problems. The direct procedure starts from the prescribed motion and calculates torques and control inputs. These are primary results, but a lot of additional characteristics can be found, too. Let us discuss these characteristics briefly.

Based on the dynamic equations for robot links, it is possible to find the vectors of reactions in robot joints. Further, the distribution of forces along each link can be solved. This result enables calculation of the mechanical stresses and elastic deformation (bending and torsion of links). For such calculation, some supplementary dynamic blocks (approximate or exact) would be needed to accomplish the model explained here. These possibilities show that robot dynamics can be successfully applied in the design of robot mechanical structure (geometry, dimensions, cross sections, choice of materials, etc.).[12-14]

Calculation of robot dynamics offers a lot of results useful for choosing the appropriate drives and the design of control systems. In addition to the torques, we can compute the power and energy requirements, form the diagrams speed vs. torque, etc.[13,14] Supplementary software calculates motor heating.[15] Because the dynamic model relates motion and driving input, it can be used to synthesize the control parameters (e.g., the feedback gains).

The variables calculated for nonperturbed mode by means of the direct dynamic procedure can be recalculated by simulating the perturbed mode (inverse dynamic procedure). In this way, it is very easy to prescribe a robot configuration and a task and then examine robot behavior by computing the different dynamic characteristics. We call this *dynamic analysis*. The software realization of such a procedure represents a very useful tool in the robot design process. A designer can quickly check a large number of different configurations. He or she can vary parameters to see their influence on some dynamic characteristics. This is of considerable help in fast and successful design. The next step is made if the limits that we impose in the design (e.g., maximal elastic deformation) are given to the computer and the software package checks the calculated characteristics against these limits. If a test is negative, some expert system might suggest how to change the relevant constructive parameters. In Vukobratović and Potkonjak[14] this approach is called the *interactive design system*. The final step is formulation of the complete CAD system for robots. The algorithms and appropriate software that would automatically find the optimal robot parameters based on the required performances and the imposed limits should be derived. The criterion of optimality is needed as well as optimization techniques. Certain results in this direction are currently available.[14,16]

20.4.2 Dynamics in On-Line Control

During the early stages of robot theory, an idea to formulate the control algorithm that takes care of robot dynamics, the so-called *dynamic control* appeared.[11] For a long time this idea was just a theoretical possibility. On one hand, the calculation of dynamics was not possible in real time, and on the other, the relatively low speed and acceleration required did not justify dynamic control. However, with current very fast and precise robots, implementation of dynamics in the control algorithm becomes necessary. It is used to form the feedforward control signal. Figure 20.10 shows such control schemes. In case (a) the feedforward signal (nominal control u^*) is found based on referent (prescribed) motion. If the motion is completely defined in advance, it is possible to calculate the direct dynamics offline and store the nominal control in the computer memory. It is recalled when the robot starts to execute the task. If referent motion is defined online (e.g., guided by means of a joystick or a sensor), on-line computation of nominal control is necessary. In case (b), the feedforward signal is found on the basis of a real state (measured data). Such a scheme understands online calculation of robot dynamics. The quality of the dynamic model (effects that are included) is relevant for the quality of robot control.

© 2002 by CRC Press LLC

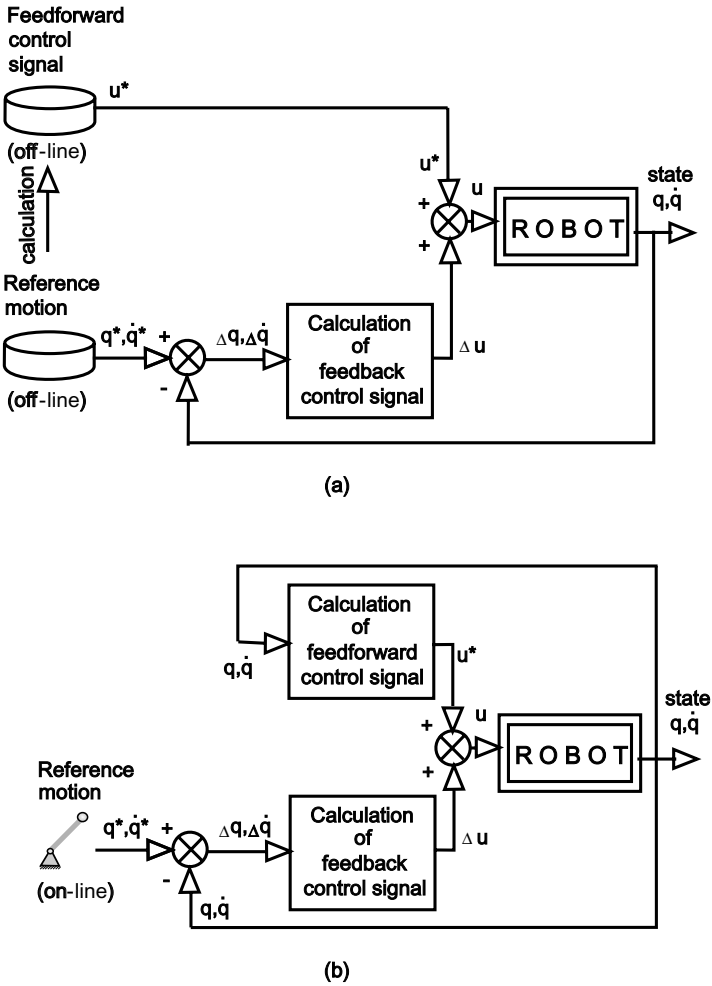


FIGURE 20.10 Scheme of dynamic control with feedforward signal.

20.5 Extension of Dynamic Modeling — Some Additional Dynamic Effects

Our purpose is to explain some dynamic effects that are not usually discussed in books on general robotics. However, these effects could be very important in some theoretical and practical work on control and design. We start with a review of the problems significant in robot dynamics and indicate the sources of particular effects. Next, some of the problems are selected and explained in more detail.

20.5.1 Robot Dynamics — Problems and Research

Beginning research in robot dynamics dealt with robots represented by an open kinematic chain. Robot links, as well as all transmission elements (shafts, gears, etc.), were considered nondeformable (Problem 1 in Figure 20.11). Many authors have worked in this field and we mention only few early results.[1-8] This approach covered many important dynamic effects and for a long time discussions on robot dynamics were restricted to such problems. This is still the case with most textbooks.

© 2002 by CRC Press LLC

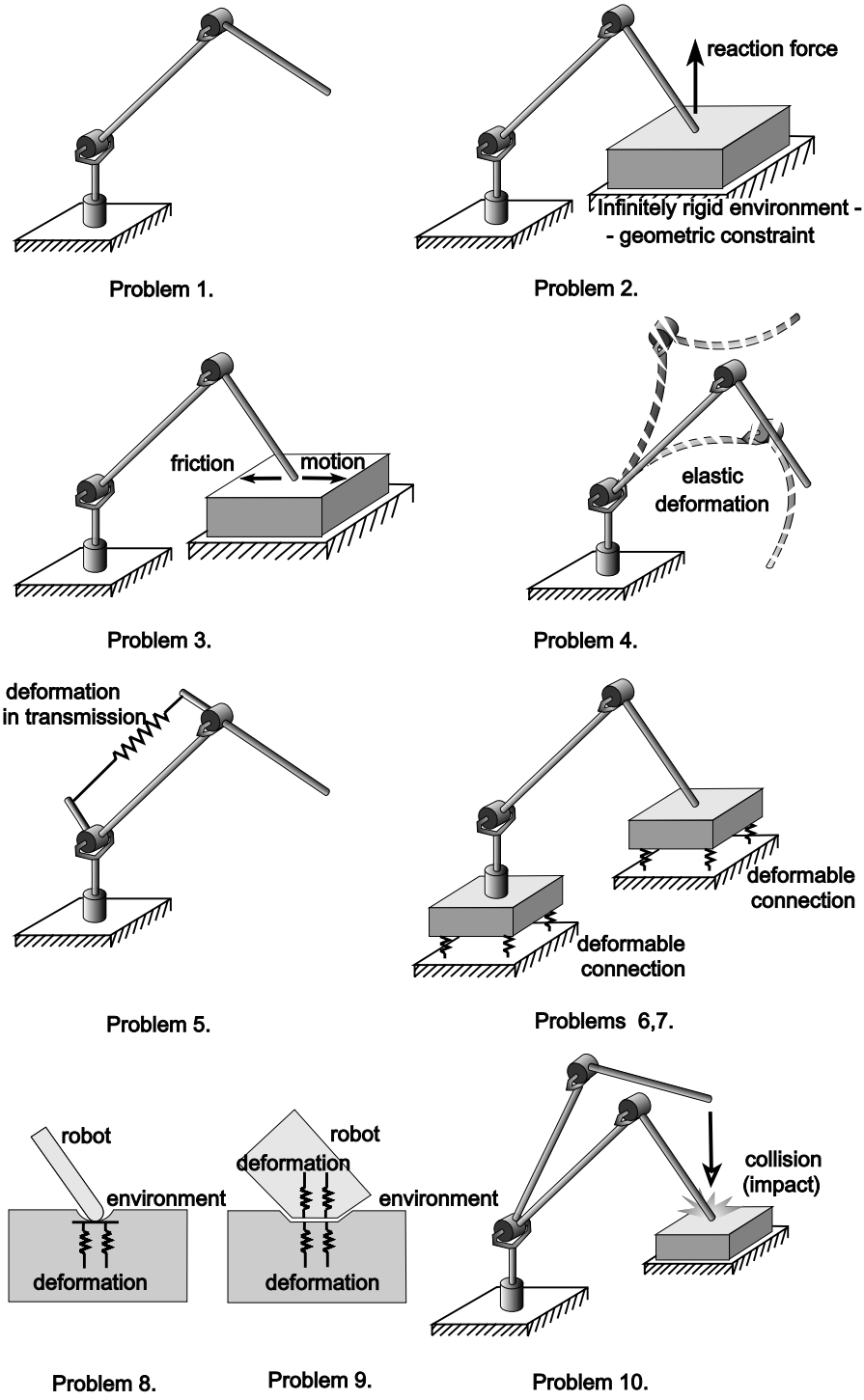


FIGURE 20.11 Different effects in robot dynamics.

The fact that many robot tasks include contact of the end-effector and the robot environment led to the first research on contact dynamics.[14] Robot environment was considered in the form of a geometric constraint (Problem 2 in Figure 20.11). The stationary and nonstationary cases were

© 2002 by CRC Press LLC

discussed. The constraints imposed to the end-effector resulted in reaction forces. Dynamic models were derived for arbitrary constraints (restricting one to six degrees of freedom). All robot elements were considered rigid. The derived models enabled solving motion along with computing contact forces. Friction between the robot and the environment was included (Problem 3). Collision problems and the solution to nonelastic impact were discussed. Practical examples considered were writing and assembly task.

After solving the dynamics of rigid robots, researchers attention turned to the elastic effects. The problem of flexible links was considered first (Problem 4). Some of the initial research looked for a simplified but fast solution of a flexible chain of the general structure,[13,14] while others intended to find a more exact model for some practical examples, i.e., single-link or two-link flexible arms.[17,18] Almost all existing dynamic models for flexible multi-link arms are found in some specific discretization methods, such as lumped mass,[13,14] assumed modes,[19] or finite elements.[20] A few researchers considered the computational efficiency of the proposed procedures.[21,22] Real-time calculation has become significant in advanced control algorithms for flexible robots.[23] Many researchers considered the linear deformation and neglected the effects of coupling between the components of deformation. Later research[24,25] takes care of these problems, thus producing more general models.

The next source of elastic effects is the transmission of torque. With electrical drives complex transmission between motor and joint shaft is usually needed. It is necessary to reduce speed and multiply torque. Thus, a kind of gear-box is present. Depending on their construction, gear-boxes introduce smaller or larger elastic deformations. The deformation is specially expressed with a harmonic-drive reducer, because elasticity is the essential property for its operation. If robot construction is such that gravitational load is reduced by placing all motors close to the robot base, then a system is needed to transmit torque and motion from the motor to the corresponding joint. This may be a chain, belt, shaft, etc. Any of these systems introduces its elastic deformation (Problem 5 in Figure 20.11). If the transmission is considered deformable, the joint shaft motion becomes independent of motor motion and only relatively high stiffness makes these motions close to each other. The number of DOFs is at least doubled. The initial results in this field were presented by Spong[26] and Potkonjak,[27] the foundations for further research. A mathematical model was derived to describe the dynamics of robots with elastic transmissions. The torque transmission included the harmonic-drive reducer, gears, and chains. Research[27] followed from practical work in robot design. Special attention was paid to some practical problems in forming the control loop: Should one measure the joint position or the motor angle? Generally speaking, the presence of unpowered DOFs represented the main problem with control of such robots. Further works elaborated this subject in more detail. One way to solve the tracking problem was presented in Kircanski, Timcenko, and Vukobratović[28] and included the measurement of torque for feedback formation. The next step in this research was to introduce constraints upon the motion of the end-effector, and thus, consider elastic joint robots in contact tasks. Several approaches to simultaneous force and position control in constrained robot systems with joint flexibility have been proposed in the literature.[29-32]

The elastic effects can be expressed with the robot support (Problem 6 in Figure 20.11). If connections of robot arm and its support were considered deformable, or if the robot was placed on a platform with pneumatic wheels, then oscillations would appear. However, these effects can be included in the existing models of rigid system by adding passive DOFs with stiffness and damping. A similar problem may appear on the environmental side. If the object to be grasped or processed in some other way is connected to its support by means of deformable connections, then oscillations arise (Problem 7). A dynamic environmental model is then needed.

With contact tasks, the most interesting deformation effects are expressed in the vicinity of contact points (Problems 8 and 9). Two bodies in contact produce a force upon each other and the force depends strongly on the elastic properties. For exact contact modeling, the elastodynamics in the contact zone has to be taken into account. In most research contact deformation was considered on the environment side only (Problem 8). The terminal link of the robot was assumed

© 2002 by CRC Press LLC

nondeformable. Such an approach can be justified in many industrial applications. It is due to the fact that tools are generally harder than the objects upon which they are acting. The general approach, however, would require the analysis of deformation on both sides of contact: environment and robot (Problem 9). Such discussion is not needed for pure theoretical reasons. With some applications, such as peg-in-hole assembly, it is a real situation. The same material is used for the peg and for the object with the hole. Hence, it is likely that both bodies in contact will be deformed. Most efforts in the field of contact deformation were made to create control strategies for contact tasks.[33-36] The main problem relates to the need to control position and force simultaneously. Different approaches to solving the problem could be distinguished. In the first approach, robot dynamics was considered in its rigid-body form and deformation in the contact zone was treated through stiffness and damping.[34] Because massless spring and damper were in question, no dynamics of the environment exists. From the standpoint of environment modeling, a more exact method has been proposed in Hogan.[35] The suggested control strategy was called impedance control. Environment was modeled by appropriate impedance. Thus, dynamics took place, but was restricted to the linear model. Complete dynamics of environment, including nonlinear effects, is the topic of position/force control of robot interacting with dynamic environment.[36,37]

Special discussion should be given about the collision (Problem 10). It is an always present effect because no contact can be precisely made to avoid impact. The first study of impact with robotic systems was given in Chumenko and Yuschenko.[38] The nonelastic impact between a robot and an object being grasped was solved. In Vukobratović and Potkonjak,[14] the collision of robot end-effector and a geometric constraint was elaborated. The impact was still considered plastic. The effect of friction was included. Both of these studies followed the classical approach based on the law of momentum. Another early result is Zheng and Hemami.[39] The influence of friction on body collision was discussed in detail in Keller[40] and Stronge.[41]

Hurmuzlu and Marghitu[42] considered a rigid-body collision of planar kinematic chains with multiple contact points. A successful algorithm for the numerical integration of a system subject to impacts was presented in Drenovac and Potknojak.[43] Brogliato and Orhant[44] formed the mathematical model of impulsive collision dynamics through the use of Schwartz's distributions, then studied the relationships between impulsive and continuous dynamic models, and analyzed the difficulties associated with transition phase control. Acaccia et al.[45] modeled the impact as a "black box," without a need to explicitly observe the compression and restitution phases. To achieve better insight, the collision could be modeled through elastodynamics. One way to do this was by means of the lumped mass approach.[46]

The final problem mentioned in this survey is redundancy. In early research in this field, redundancy was considered a problem of kinematics (avoiding obstacles, avoiding singular positions, etc.). Later research, however, saw redundancy as a possibility to improve robot dynamic performance.[47,48] The biomechanical approach to the solution of redundancy of a humanoid robot arm was proposed in Potknojak et al.[49] Special kind of redundancy appears in so-called systems with variable geometry.[50,51] The mechanism is designed to have an augmented number of DOFs (more than the kinematics of the task requires). However, this redundancy does not change end-effector maneuvering capabilities. It changes the inner structure of the robot. For this reason, it is called *internal redundancy*. The role of such redundancy is to avoid limitations imposed to actuators (torque and speed limits) and thus improve robot dynamic capabilities.

Among various problems in robot dynamics, in this chapter we emphasized the following: motion subject to geometric constraints, interaction with the dynamic environment, and effects of elastic transmissions.

20.5.2 Dynamics of Robot in Constrained Motion

Here we discuss contact tasks considering robot environment as a geometric constraint imposed to the motion of the end-effector. The discussion starts from the free motion of a rigid-body robot.

© 2002 by CRC Press LLC

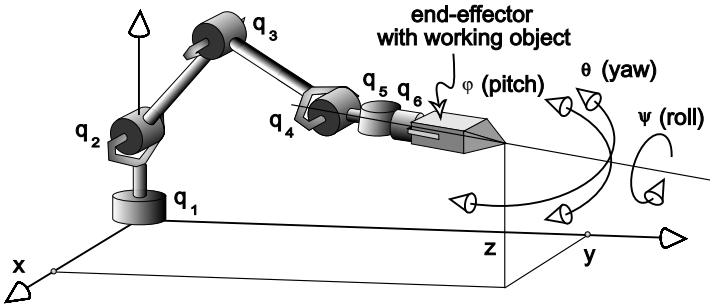


FIGURE 20.12 Internal position $q = (q_1, \ldots, q_6)$ and external position $X = (x, y, z, \theta, \phi, \psi)$.

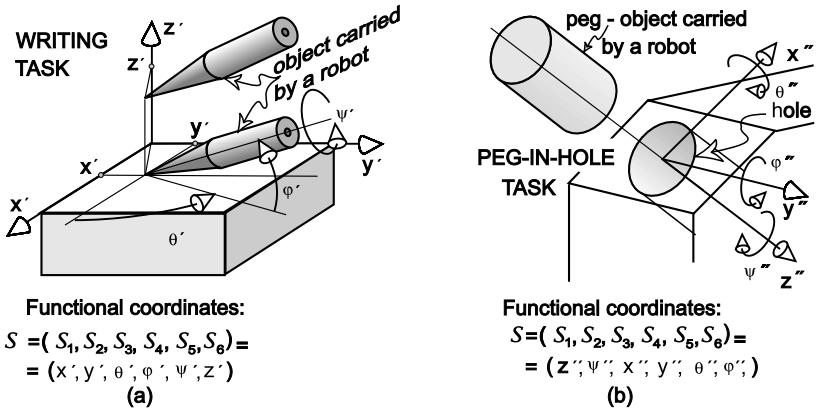


FIGURE 20.13 Two examples of functional coordinates.

The position of the chain, consisting of n links and n one-DOF joints (Figure 20.12), is defined by means of the joint coordinates: $q = [q_1 \ \ldots \ q_n]^T$. In practical operation, joint coordinates are not always suitable for use. Many tasks cannot be described in this way. For this reason, a new set of coordinates is defined, the *external coordinates*. By this term we usually understand position and orientation of the end-effector with respect to the immobile frame: Cartesian coordinates of the robot tip (x, y, z) and yaw, pitch, and roll angles (θ, ϕ, ψ), as shown in Figure 20.12. The complete external position vector is $X = [xyz\theta\phi\psi]^T$ and has a dimension of six. For simplicity, we assume that the robot has six joints and, thus, six joint coordinates ($n = 6$). In this way the elaboration of redundancy is avoided but the intention of our discussion is not compromised. In the majority of manipulation tasks, the external position is very suitable to apply. However, with contact tasks and some other process operations, it is more appropriate if the position of the end-effector is defined relative to the object being processed. Hence, we introduce a frame fixed to the object. Because the object may be mobile, the new frame would incorporate the law of motion. We assume that the object moves according to a given law that cannot be affected by robot action. This is necessary if one intends to describe the environment as a geometric constraint. With respect to the new frame, the position and the orientation of the end-effector are expressed by means of six coordinates: $s = [s_1 \ \ldots \ s_6]^T$. We call them *functional coordinates*. Two examples are shown in Figure 20.13. Case (a) represents the surface-type constraint appropriate for modeling the writing task. Case (b) represents the peg-in-hole assembly task. Note that the order of coordinates in vector s may be adopted arbitrarily, and we adopt the order suitable for the discussion that follows.

Each contact task consists of three phases: approaching, impact, and constrained motion. In the approaching phase the end-effector moves toward the constraint. The motion is usually planned to

© 2002 by CRC Press LLC

achieve zero-velocity contact and avoid impact. However, in a real situation, the always present perturbations cause the collision. The impact occurs, producing sudden change in velocities. After the impact phase, regular constrained motion starts. From the standpoint of modeling, the first phase represents free motion. Both internal coordinates (q) and functional coordinates (s) are free. All the discussion from the previous paragraphs holds and the dynamics is described by model (20.22). The kinematics should be described by means of s-coordinates. The relation between coordinates q and s can be expressed via a nonlinear function that can be nonstationary or stationary:

$$s = s(q, t) \text{ or } s = s(q) \tag{20.83}$$

With the nonstationary problems, the explicit appearance of time t is due to the mobility of the s-frame. For simplicity, we restrict our discussion to stationary problems (immobile frame s). The second derivative of the latter relation from (20.83) produces

$$\ddot{s} = J(q)\ddot{q} + A(q,\dot{q}) \tag{20.84}$$

where $J = \partial s/\partial q$ is the Jacobian matrix, and $A = (\partial^2 s / \partial q^2)\dot{q}^2$ is the adjoint vector.

Although all the coordinates from the set s are free and independent (in the approaching phase), it is useful to separate them into two subsets, s^f and s^c, of dimensions $6 - m$ and m, respectively. This separation follows from the nature of the constraint being approached, and will be explained later. With vector s separated into two subvectors, relation (20.84) becomes

$$\ddot{s}^f = J_f(q)\ddot{q} + A_f(q,\dot{q}) \tag{20.85}$$

$$\ddot{s}^c = J_c(q)\ddot{q} + A_c(q,\dot{q}) \tag{20.86}$$

where the dimensions are J_f $((6 - m) \times 6)$, A_f $(6 - m)$, J_c $(m \times 6)$, A_c (m).

The approaching phase ends when the end-effector touches the constraint, that is, when the corresponding coordinate becomes zero. For the example from Figure 20.13a, the contact is established when coordinate s_6 (that is, z') reduces to zero. The impact occurs in the instant of contact. We find it more convenient to discuss impact dynamics later and elaborate the constrained motion (the third phase) now. In this case, we assume that the contact has already been established and the effects characteristic for the transient process have been finished. The imposed constraint restricts some motion of the end effector. Let this restriction be expressed in terms of s-coordinates by means of the condition

$$s^c = 0. \tag{20.87}$$

For example, (a) from Figure 20.13, the separation of vector s should be: $s^f = [s_1 \ldots s_5]^T$, $s^c = [s_6]$, $m = 1$; while for example (b) separation $s^f = [s_1\ s_2]^T$, $s^c = [s_3\ s_4\ s_5\ s_6]^T$, $m = 4$, is applied. Thus, s^c represents the set of constrained coordinates (upper index "c" stands for "constrained"). In example (a) there is only one such coordinate and accordingly one DOF is lost. In example (b), four DOFs are lost. In a general case, the constrained robot has 6-m remaining DOFs and its motion is described by means of $6 - m$ coordinates forming the vector s^f (upper index "f" stands for "free").

The restriction of motion results in reaction forces. Reaction will appear in the directions of the constrained coordinates s^c. Thus, there will be m independent components of reaction. Let these components form the vector F. Reaction may be in the form of a force, if translation is constrained, or in the form of a torque, if rotation is constrained. Figure 20.14 shows the reactions for the examples defined in Figure 20.13.

© 2002 by CRC Press LLC

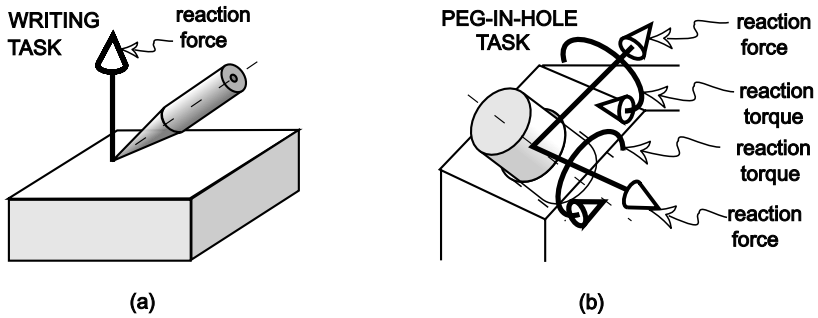


FIGURE 20.14 Two examples of reactions.

The dynamic model (20.22) used for free motion should now be reformulated to incorporate the reaction effects. Because F is the external force acting along directions s^c, the model can be written in the form:

$$H(q)\ddot{q} + h(q,\dot{q}) = \tau + J_c^T(q)F \tag{20.88}$$

where $J_c = \partial s^c/\partial q$ was introduced in (20.86). To solve the dynamics of the constrained motion, model (20.88) has to be considered along with the condition (20.87). It is more suitable if the constraint is expressed in differential form, and hence, (20.86) should be used, yielding

$$\ddot{s}^c = J_c(q)\ddot{q} + A_c(q,\dot{q}) = 0. \tag{20.89}$$

Equations (20.88) and (20.89) complete the dynamic model. They consist of 6 + m scalar equations and can be solved for 6 + m unknowns, accelerations $\ddot{q}$ and reactions F.

Now, we return to the second phase of the contact task, the problem of impact. At the beginning, we consider the simple surface-type constraint shown in Figures 20.13a and 20.14a. Impact occurs when s^c, that is, coordinate s_6, becomes zero and the impact action is directed along this coordinate. The impact may be more or less elastic and depending on that skipping could appear. To avoid such a complex discussion, we restrict consideration to the nonelastic problem. In this case, the end-effector will not leave the surface after the first contact, but move along it.

A slightly more complex problem would be constraint in the form of two surfaces. Although the intention is to move the end-effector along the line of intersection, in a real situation collision with one surface would happen first. Thus, two impacts would occur, one after the other. However, one might neglect this effect and consider the two contacts as simultaneous. In this case a complex impact has two components acting along the two constrained coordinates (dimension $s^c = 2$). Further generalization leads to the m-component constraint, that is, m restricted coordinates in vector s^c. The complex impact force F would have m components acting simultaneously. The example for $m = 4$ is shown in Figures 20.13b and 20.14b. Note that in real motion the complex impact represents a series of collisions with surfaces, but in the discussion that follows we neglect this effect. To solve the dynamics of the impact, we integrate the model of constrained motion over the impact interval (interval of transition). Let t' denote the beginning of transition, that is, the instant when s^c reduces to zero and the contact occurs. Let t'' be the instant when transition effects may be considered finished. The impact interval is then $\Delta t = [t',\ t'']$. For a geometric constraint and a nonelastic impact, it is justified to consider this interval infinitely short, that is, $\Delta t \to 0$ and $t'' \to t'$. We integrate model (20.88) over this interval to obtain:

$$H(q')\Delta\dot{q} = J_c^T(q')F\Delta t,\ \Delta\dot{q} = \dot{q}'' - \dot{q}' \tag{20.90}$$

© 2002 by CRC Press LLC

where $\dot{q}' = \dot{q}(t')$ and $\dot{q}'' = \dot{q}(t'')$. Equation (20.89) gives:

$$J_c(q')\Delta\dot{q} = -J_c(q')\dot{q}' \tag{20.91}$$

Relations (20.90) and (20.91) describe the dynamics of impact. They represent the set of $6 + m$ scalar equations and can be solved for $\Delta\dot{q}$ and $F\Delta t$. $\Delta\dot{q}$ is the change in joint velocity (6 components) and $F\Delta t$ is the impact impulse (m components). Note that $F\Delta t \neq 0$ although $\Delta t \to 0$. This means that $F \to \infty$.

Let us briefly explain how we use the dynamic models to solve the three phases of the contact task numerically. In the approaching phase, model (20.22) is integrated checking the value of s^c. When s^c becomes zero, we turn to the impact model (20.90), (20.91) and solve the change in velocity. With this new initial state we start solving the dynamics of the constrained motion by integrating model (20.88) and (20.89).

20.5.3 Robot in Contact with Dynamic Environment

In the preceding paragraphs we established the contact problem as an important part of most robot tasks in industry. Geometric constraints were one way to form the dynamic model for such systems. The geometric constraints understood rigid-body contact and could be called "rigid constraints." However, with some theoretical and practical problems it was shown that this concept was not justified. The dynamic behavior of the environment appeared to be important. This led first to the concept of *soft constraint* and later to the idea of *dynamic environment*.

We do not intend to solve the general case of a robot interacting with a dynamic environment, but only to illustrate the idea. We model the robot's dynamics as if it were a rigid-body system and the dynamics of the environment is reduced to deformation and elastodynamic effects in the contact zone.

Let us consider the surface-type constraint (e.g., a writing task) shown in Figures 20.13a and 20.14a. Writing along the base has to be performed by applying some pressure force upon it. With the rigid base, coordinate z', that is, s_6, was constrained. If the base is not considered as infinitely rigid, but deformable, the motion in the perpendicular direction (negative s_6) will be possible. Thus, from the standpoint of kinematics there is no restriction on motion and no degrees of freedom are lost. Some kind of restriction follows from the base dynamics. Its elastic properties will keep the perpendicular motion small. Now the question arises of how to model environment dynamics (base, in this example). To avoid too complex a discussion on elastodynamics, we adopt the lumped-mass approach. Figure 20.15 shows the steps in the introduction of environment dynamics. Here, we elaborate model (c). Model (d) will be mentioned only briefly.

The contact task, as explained before, consists of three phases. In the first, approaching the surface, the problem of dynamic environment does not differ from the problem of geometric constraint. Robot dynamics is integrated using q-coordinates, but at the same time, we use functional s-coordinates to check whether the surface is reached. When s_6 (that is, s^c in a general case) reduces to zero, contact is present. At that instant impact occurs, and we assume that it is completely nonelastic. Impact is the second phase of the task. However, it is more convenient to explain the dynamics of contact motion (the third phase) before discussing the impact.

Figure 20.16 shows contact of the robot end-effector and the elastic base. Robot dynamics is described by model (20.88) and Equation (20.86) is used to express the relation between coordinates q and s^c. For the constraints of the surface-type, vector s^c reduces to one component (coordinate s_6), and hence, we rewrite (20.88) and (20.86) as

$$H(q)\ddot{q} - h(q,\dot{q}) = \tau + J_6^T(q)F \tag{20.92}$$

© 2002 by CRC Press LLC

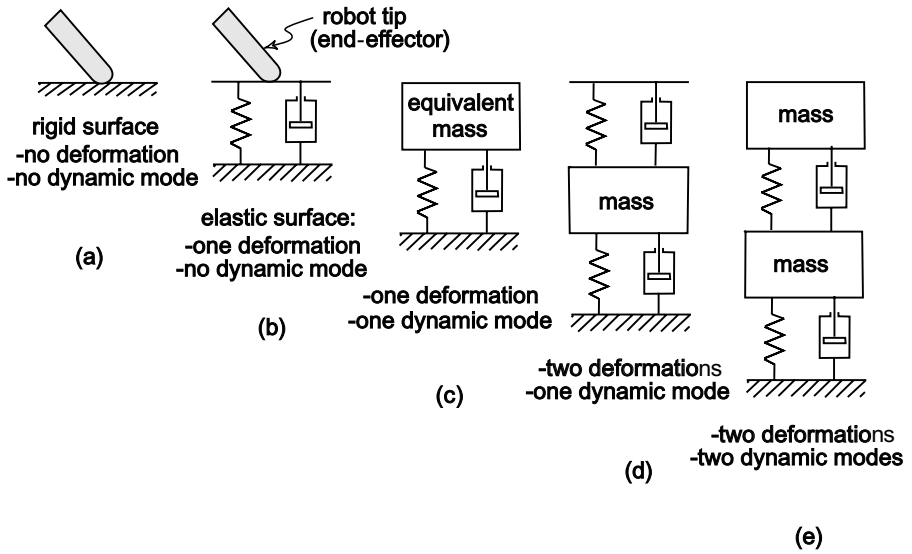


FIGURE 20.15 Modeling environment dynamics.

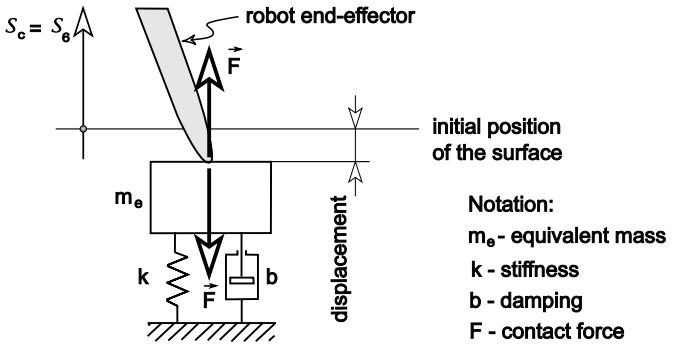


FIGURE 20.16 Elastodynamics of the base.

and

$$\ddot{s}_6 = J_6(q)\ddot{q} + A_6(q,\dot{q}) \tag{20.93}$$

where dimensions of Jacobian matrix J_6 are 1×6, and variables s_6, F, and A_6 are scalars. The dynamics of the base can be described by means of Newton's law applied to the equivalent mass m_e:

$$m_e\ddot{s}_6 = -F - ks_6 - b\dot{s}_6 \tag{20.94}$$

(Note that s_6 is negative all the time.) The set of Equations (20.92), (20.93), and (20.94) defines the dynamics of contact motion. The set consists of 6 + 1 + 1 = 8 scalar equations and could be solved for 8 scalar unknowns: 6- component $\ddot{q}$, and scalars $\ddot{s}_6$ and F.

We now return to impact. It has been assumed nonelastic and the integration of Equations (20.92)–(20.94) over the impact interval $\Delta t \to 0$ yields

© 2002 by CRC Press LLC

$$H(q')\Delta\dot{q} = J_6^T(q')F\Delta t \tag{20.95}$$

$$\Delta\dot{s}_6 = J_6(q')\Delta\dot{q} \tag{20.96}$$

$$m_e\Delta\dot{s}_6 = -F\Delta t \tag{20.97}$$

where q' refers to the instant of contact (t') and $\Delta\dot{q}$ and $\Delta\dot{s}_6$ represent the changes in velocities. Equations (20.95), (20.96), and (20.97) define the impact dynamics and could be solved for $\Delta\dot{q}$, $\Delta\dot{s}_6$, and the impact impulse $F\Delta t$ (since $\Delta t \to 0$, it has to be $F \to \infty$ in order to keep $F\Delta t \neq 0$).

If the environment is modeled as shown in Figure 20.15d, the entire deformation consists of the external and internal components. Thus, one additional DOF appears. The next specific of this model is that it does not require special impact treatment. The external deformation in the form of a massless spring eliminates the impact, and thus, after approaching, the system immediately enters the contact motion phase. One may say that the impact represents an initial period of contact motion while transient effects are exhibited.

20.5.4 Effects of Elastic Transmissions

Here, we consider an open-chain mechanism (unconstrained motion) and concentrate on the problem of transmitting torque from motors to joints. With robots driven by electrical motors, it is very common that motors are displaced from joints. They are moved toward the robot base to achieve better statics (that is, to unload the arm). In such cases, a different transmission could be applied between the motor and joint shafts. Some kind of a gear-box is usually present followed by chains, belts, or some other elements. Such transmission is a source of deviation, because each element may introduce its elastic deformation. Here, we assume that each joint has only one deformable element in the transmission. To make the discussion clearer, let this elastic element be the harmonic-drive reducer (HD). This will not compromise the generality because any other transmission element may be modeled in the same manner.

Figure 20.17a shows joint S_j with its motor and transmission. Motor dynamics can be described by mechanical Equation (20.62) and electrical Equation (20.63). If we neglect friction and inductivity, then these two equations could be combined to give the form (20.70). This model is expressed in terms of motor shaft angle θ_j. M_j that appears in the model represents motor output torque. The model holds for each motor, $j = 1, \ldots, n$. Dynamics of robot links are described by model (20.22) and expressed in terms of joint coordinates q_j, $j = 1, \ldots, n$. Vector $\tau = [\tau_1 \ \ldots \ \tau_n]^T$ that appears in the model contains the input torques for robot links (joint shaft). In Section 20.3, when nondeformable transmission was assumed, the relation between the motor variables (θ_j, M_j) and joint variables (q_j, τ_j) was defined as linear. It was expressed by means of Equations (20.68) and (20.69). The motion of the motor and that of the corresponding joint were dependent on each other. If flexibility of transmission is introduced, the situation becomes rather different. Motor motion (coordinate θ_j) and joint motion (coordinate q_j) become kinematically independent and only high stiffness keeps q_j close to θ_j/N_j. This means that the overall number of DOFs is doubled. We now concentrate on the concrete example of transmission shown in Figure 20.17. To generalize the discussion, in addition to the elasticity we take care of transmission inertia. In the schematic presentation (Figure 20.17b), the dashed line indicates the transmission stage. Its input torque is M_j and the output is τ_j. If the moments of inertia of the gears are I_j' and I_j'' and if other inertial effects are neglected, dynamics can be described by means of equation

$$(I_j' + I_j''/N_j^2)\ddot{\theta}_j = M_j - \tau_j/N_j. \tag{20.98}$$

© 2002 by CRC Press LLC

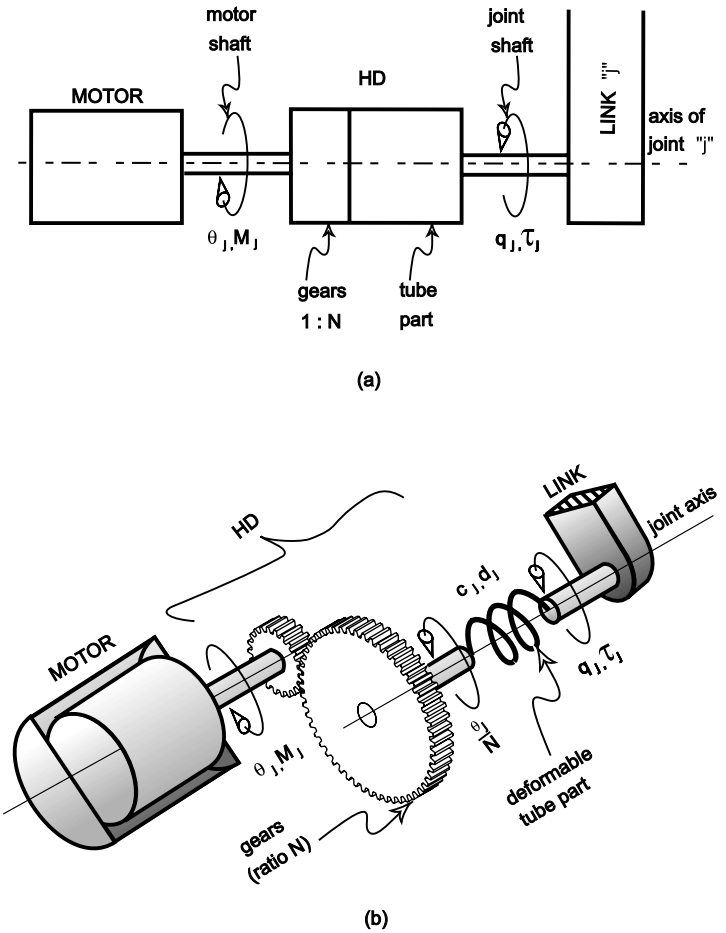


FIGURE 20.17 Deformable transmission.

It should be noted that any other transmission element (chain, belt, etc.) can be modeled in the same way, that is, as a combination of two rigid gears and a torsion spring element. In a generalized case, the transmission may consist of several stages, one after the other. Each stage would be treated as explained for the HD reducer in the present example. Inertia of gears may be expressed stronger or weaker, especially if compared with transmission friction that we neglected. Let us return to Figure 20.17. Joint torque τ_j follows from deformation of spring and hence

$$\tau_j = c_j(\theta_j/N_j - q_j) + d_j(\dot{\theta}_j/N_j - \dot{q}_j) \tag{20.99}$$

where c_j and d_j denote torsion stiffness and damping coefficients.

The complete dynamic model now includes Equations (20.22) for links, (20.70), $j = 1, \ldots, n$ for motors, and (20.98) and (20.99), $j = 1, \ldots, n$ for transmissions. The model can be integrated to give motions θ_j and q_j, $j = 1, \ldots, n$.

Appendix: Calculation of Transformation Matrices

We consider the mechanism as a kinematic chain consisting of n rigid links interconnected by one-DOF joint which can be either revolute or linear. The position of the chain is described by means

© 2002 by CRC Press LLC

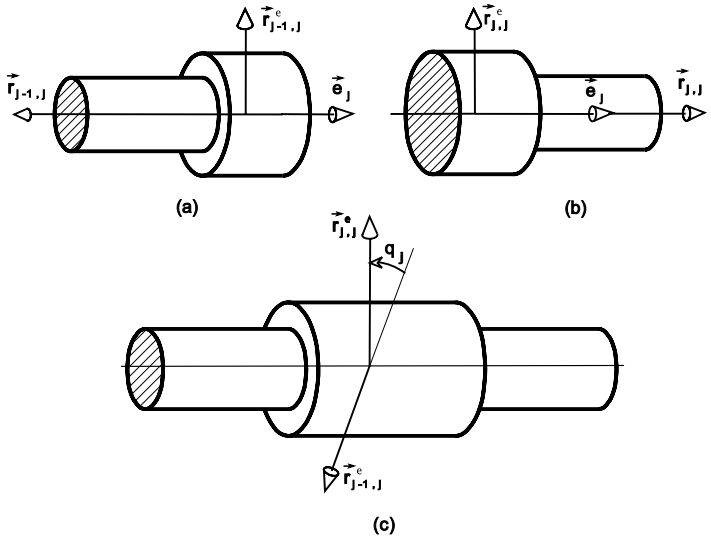


FIGURE 20.18 Definition of joint coordinate in the case of "specificity."

of coordinates $q_1, \ldots, q_n$. The definition of revolute coordinate requires some additional discussion. We start by referring to Figure 20.2 where the geometry of a link is defined and to Figure 20.3 that introduces the coordinate for a revolute joint. However, this definition of coordinate is not general. It requires that neither of vectors $\vec{r}_{j-1,j}$ or $\vec{r}_{j,j}$ is parallel with the axis $\vec{e}_j$. If it happens that one or both vectors are parallel with the joint axis, then modification of the definition is needed. Let us discuss this problem in detail.

If $\vec{r}_{j-1,j} \| \vec{e}_j$, then we say that there exists a specificity at the upper end of link $j-1$ (the upper end of a link is the end oriented to the next link; e.g., joint S_j represents the upper end of link $j-1$). In this case, it is necessary to introduce an additional vector, $\vec{r}^{\,*}_{j-1,j}$, which is not parallel with $\vec{e}_j$, and use it instead of $\vec{r}_{j-1,j}$ when defining the coordinate q_j (Figure 20.18a,c). The condition that $\vec{r}^{\,*}_{j-1,j}$ is not parallel with $\vec{e}_j$ is the only condition imposed in the choice of this additional vector. However, it is most convenient to define it as a unit vector perpendicular to $\vec{e}_j$.

If $\vec{r}_{j,j} \| \vec{e}_j$, then we say that there exists a "specificity" at the lower end of the j-th link (joint S_j represents the lower end of link j). In this case we introduce an additional vector, $\vec{r}^{\,*}_{j,j}$, which is not parallel with $\vec{e}_j$, and use it instead of $\vec{r}_{j,j}$ when defining the coordinate q_j (Figure 20.18b,c).

In this way, vectors $\vec{r}^{\,*}_{j-1,j}$ and $\vec{r}^{\,*}_{j,j}$ become the input data for the software that calculates the position or kinematics of the robot chain. Vector $\vec{r}^{\,*}_{j-1,j}$ is expressed in frame $j-1$, and vector $\vec{r}^{\,*}_{j,j}$ in frame j.

We are now going to explain the calculation of the transformation matrix between the frame fixed to link j and the frame fixed to link $j-1$. It is called the relative transformation matrix and is marked by $A_{j-1,j}$. The matrix is used to turn some vector from frame j to $j-1$. If some vector $\vec{\tilde{c}}_j$ is considered, then $\vec{\tilde{c}}_j$ can be calculated starting from $\vec{\tilde{c}}_j$:

$$\vec{\tilde{c}}_j = A_{j-1,j}\vec{\tilde{c}}_j \tag{20.100}$$

To calculate the transformation matrix we need three vectors linearly independent of each other. We have to know the projections of these vectors to both frames, $j-1$ and j. The calculation starts with solving the transformation matrix that corresponds to the extended joint. The first vector of the triple is $\vec{e}_j$. Its projections are known: $\vec{\tilde{e}}_j$ onto frame $j-1$ and $\vec{\tilde{e}}_j$ onto frame j. We now introduce vector $\vec{a}_j$ as a unit vector of axis "a" shown in Figure 20.19. Axis "a" defines the extended position of the joint. The projections of the vector $\vec{a}_j$ onto frames $j-1$ and j are

© 2002 by CRC Press LLC

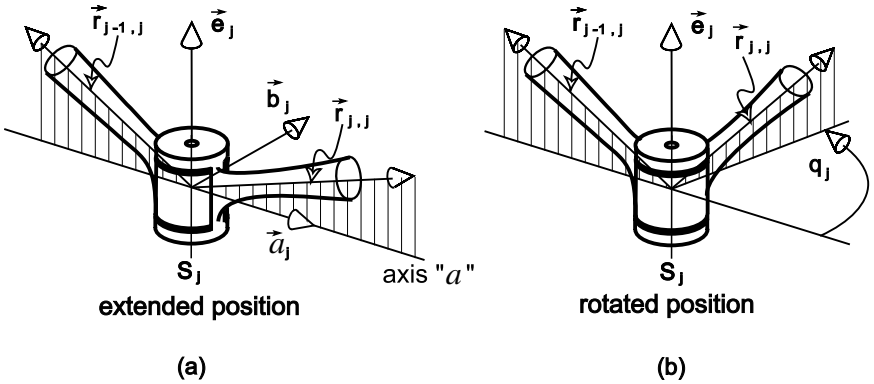


FIGURE 20.19 Revolute joint: extended and rotated position.

$$\underset{\sim}{\vec{a}}_j = \frac{-\underset{\sim}{\vec{e}}_j \times (\underset{\sim}{\vec{r}}_{j-1,j} \times \underset{\sim}{\vec{e}}_j)}{|\underset{\sim}{\vec{e}}_j \times (\underset{\sim}{\vec{r}}_{j-1,j} \times \underset{\sim}{\vec{e}}_j)|}, \quad \tilde{\vec{a}}_j = \frac{\tilde{\vec{e}}_j \times (\tilde{\vec{r}}_{j,j} \times \tilde{\vec{e}}_j)}{|\tilde{\vec{e}}_j \times (\tilde{\vec{r}}_{j,j} \times \tilde{\vec{e}}_j)|} \tag{20.101}$$

Vector $\vec{a}_j$ is the second vector of the triple. The third vector is obtained as

$$\vec{b}_j = \vec{e}_j \times \vec{a}_j \tag{20.102}$$

and has the projections

$$\underset{\sim}{\vec{b}}_j = \underset{\sim}{\vec{e}}_j \times \underset{\sim}{\vec{a}}_j, \quad \tilde{\vec{b}}_j = \tilde{\vec{e}}_j \times \tilde{\vec{a}}_j \tag{20.103}$$

Vectors $\vec{e}_j, \vec{a}_j$, and $\vec{b}_j$ are perpendicular to each other and thus linearly independent. The relation between projections onto two frames may be expressed (according to (20.100)) in the form:

$$\underset{\sim}{\vec{e}}_j = A^0_{j-1,j} \tilde{\vec{e}}_j, \quad \underset{\sim}{\vec{a}}_j = A^0_{j-1,j,} \tilde{\vec{a}}_j \quad \underset{\sim}{\vec{b}}_j = A^0_{j-1,j} \tilde{\vec{b}}_j \tag{20.104}$$

where the upper index "0" indicates the extended joint. Relations (20.104) can be united:

$$\left[\underset{\sim}{e}_j \, \underset{\sim}{a}_j \, \underset{\sim}{b}_j\right] = A^0_{j-1,j}\left[\tilde{e}_j \tilde{a}_j \tilde{b}_j\right] \tag{20.105}$$

where e_j is a column vector of dimension 3×1 that contains the projections of vector $\vec{e}_j$, and analogously holds for the other two vectors. From Equation (20.105), the transformation matrix of the extended joint is obtained:

$$A^0_{j-1,j} = \left[\underset{\sim}{e}_j \, \underset{\sim}{a}_j \, \underset{\sim}{b}_j\right]\left[\tilde{e}_j \tilde{a}_j \tilde{b}_j\right]^{-1} \tag{20.106}$$

Matrix $A^0_{j-1,j}$ $(\dim = 3 \times 3)$ has three columns that represent the unit vectors of frame j expressed in frame $j - 1$. Let us denote these columns by $\underset{\sim}{V}^0_{j1}$, $\underset{\sim}{V}^0_{j2}$ and $\underset{\sim}{V}^0_{j3}$. Thus,

$$A^0_{j-1,j} = \left[\underset{\sim}{V}^0_{j1} \, \underset{\sim}{V}^0_{j2} \, \underset{\sim}{V}^0_{j3}\right] \tag{20.107}$$

© 2002 by CRC Press LLC

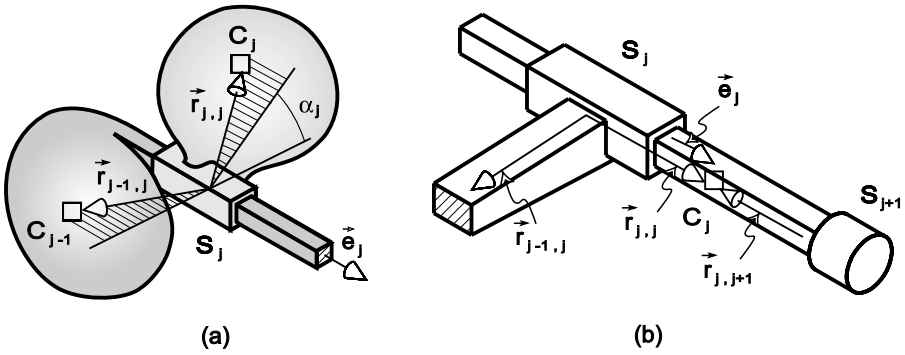


FIGURE 20.20 Linear joint.

Up to this point, the extended joint was considered, that is, the joint coordinate was assumed to be zero ($q_j = 0$). To find the transformation matrix $A_{j-1,\,j}$ that corresponds to some nonzero position ($q_j \neq 0$), it is necessary to rotate each of the unit vectors, $\vec{V}^{\,0}_{jk}, k = 1,\ 2,\ 3$, about the axis $\vec{e}_j$ by the angle q_j. For the rotation, we use the Rodrigues' formulae and mark the rotated vectors by $\vec{V}_{jk}$:

$$\vec{V}_{jk} = \vec{V}^{\,0}_{jk}\cos q_j + \left(1-\cos q_j\right)\left(\vec{e}_j \times \vec{V}^{\,0}_{jk}\right)\vec{e}_j + \vec{e}_j \times \vec{V}^{\,0}_{jk}\sin q_j, \quad k = 1,\ 2,\ 3 \tag{20.108}$$

After rotation, unit vectors $\vec{V}_{jk}$ define the new transformation matrix. Thus, matrix $A_{j-1,j}$ that corresponds to some given position q_j is

$$A_{j-1,\,j} = \left[\vec{V}_{j1}\ \vec{V}_{j2}\ \vec{V}_{j3}\right] \tag{20.109}$$

We now turn to the problem when two links ($j - 1$ and j) are connected by means of a linear joint. Figure 20.4 explains the definition of a coordinate in such a joint. Since the relative motion between two links is a translation, it is clear that the transformation matrix is constant. The matrix $A_{j-1,j}$ can be solved in a manner analogous to that used for revolute joints. If that procedure is applied, from Figure 20.20a we conclude that for a linear joint the extended position is only fictive. The rotation is performed by the constant angle α_j. This angle has to be prescribed among the geometrical parameters. Further, Figure 20.20b shows that with linear joints specificity is a very common feature. Additional vectors (like $\vec{r}^{\,*}_{j,j}$) are needed in such cases.

The discussion in this Appendix concerned the transformation between two adjacent frames. However, we are often interested in transforming a vector from a link-fixed frame to the external stationary one. Relation (20.9) introduced notation A_j for such transformation matrix:

$$\vec{c}_j = A_j \vec{\tilde{c}}_j \tag{20.110}$$

Moving the vector toward the robot support, from one frame to the other by using recursive expression (20.100), we finally reach the stationary frame. Thus, the transformation from the link-fixed frame to the external one is described by matrix

$$A_j = A_{0,1}A_{1,2}\ldots A_{j-1,\,j.} \tag{20.111}$$

© 2002 by CRC Press LLC

References

1. Stepanenko, Yu., Method of analysis of spatial lever mechanisms, (in Russian), *Mechanics of Machines,* 23, 2, 48–57, 1970.
2. Vukobratović, M. and Stepanenko, Yu., Mathematical models of general anthropomorphic systems, *Math. Biosci.,* 17, 191–242, 1973.
3. Stepanenko, Yu. and Vukobratović, M., Dynamics of articulated open-chain active mechanisms, *Math. Biosci.,* 28, 1/2, 137–170, 1976.
4. Luh, J.Y.S., Walker, M.W., and Raul, R.P.C., On-line computational scheme for mechanical manipulators, *J. Dynam. Syst., Meas. Control,* 102, 1980.
5. Vukobratović, M. and Potkonjak, V., Contribution to computer construction of active chain models via lagrangian form, *J. Appl. Mech. — Trans. ASME,* 46, 1, 181–185, 1979.
6. Hollerbach, J.M., A recursive formulation of Lagrangian manipulator dynamics, *IEEE Trans. on SMC,* 10, 11, 730–736, 1980.
7. Popov, E.P., Vereschagin, A.F., and Zenkevich, S.A., *Manipulation Robots: Dynamics and Algorithms,* (in Russian), Nauka, Moscow, 1978.
8. Potkonjak, V. and Vukobratović, M., Two new methods for computer forming of dynamic equations of active mechanisms, *J. Mechanisms Machine Theory,* 14, 3, 189–200, 1979.
9. Corke, P., A Robotics toolbox for MATLAB. *IEEE Robot. Automat.*, 3, 1, 24–32, 1996.
10. Nethery, J.F., Robotica: A structured environment for computer-aided design and analysis of robots, Master's thesis, Univ. of Illinois, Urbana, 1993.
11. Vukobratović, M. and Stokić, D., *Control of Manipulation Robots: Theory and Application,* Springer-Verlag, Berlin, 1982.
12. Vukobratović, M., Potkonjak, V., and Hristic, D., Dynamic method for the evaluation and choice of industrial manipulators, *Proc. 9th ISIR,* Washington, 1979.
13. Vukobratović, M. and Potkonjak, V., *Dynamics of Manipulation Robots: Theory and Application,* Springer-Verlag, Berlin, 1982.
14. Vukobratović, M. and Potkonjak, V., *Applied Dynamics and CAD of Manipulation Robots,* Springer-Verlag, Berlin, 1985
15. Potkonjak, V., Thermal criterion for the selection of d.c. drives for industrial robots, *Proc. 16th ISIR,* Brussels, 218–224, 1986.
16. Inoue, K., Shiina, K., Takano, M., and Sasaki, K., Study on total computer-aided design system for robot manipulators, *Proc. 24th ISIR,* 729–736, 1993.
17. Cannon, R.H. and Schmitz, E., Initial experiments on the end-point control of a flexible one-link robot, *J. Robotics Res.,* 3, 62–75, 1984.
18. Fukuda, T. and Arakawa, A., Modeling and control characteristics for a two-degrees-of-freedom coupling system of flexible robotic arm, *JSME,* 30, 1458–1464, 1987.
19. Truckenbrodt, A., Dynamics and control methods for moving flexible structures and their application to industrial robots, *Proc. 5th World Cong. Theory Machines Mechanisms,* ASME, 1981.
20. Sunada, H.W., Dynamic analysis of flexible spatial mechanisms and robotic manipulators, Ph.D. Thesis, Univ. of California, Los Angeles, 1981.
21. Book, W.J., Recursive Lagrangian dynamics of flexible manipulator arms, *J. Robotics Res.,* 3, 87–101, 1984.
22. King, J.O., Gourishankar, V.G., and Rink, R.E., Lagrangian dynamics of flexible manipulators using angular velocities instead of transformation matrices, *IEEE Trans. SMC,* 17, 1059–1068, 1987.
23. Uchiyama, M. and Conno, A., Computed acceleration control for the vibration suppression of flexible robotic manipulators, *Proc. 5th ICAR,* Pisa, Italy, 126–131, 1991.
24. Šurdilović, D. and Vukobratović, M., One method for efficient dynamic modeling of flexible manipulators, *Mechanisms Machine Theory,* 31, 3, 397–315, 1996.
25. Šurdilović, D. and Vukobratović, M., Deflection compensation for large flexible manipulators, *Mechanisms Machine Theory,* 31, 3, 317–329, 1996.
26. Spong, M.W., Modeling and control of elastic joint robots, *J. Dynam. Syst., Meas. Control,* 109, 309–319, 1987.
27. Potkonjak, V., Contribution to the dynamics and control of robots having elastic transmission, *Robotica,* 6, 63–69, 1988.

© 2002 by CRC Press LLC

28. Kircanski, N., Timcenko, A., and Vukobratović, M., Position control of robot manipulators with elastic joints using force feedback, *J. Rob. Syst.,* 7, 4, 535–554, 1990.
29. Ahmad, S., Constrained motion (force/position) control of flexible joint robots, *Proc. IEEE Conf. Decision Control,* Brighton, England, 1397–1402, 1991.
30. Mills, J.K., Stability and control of elastic-joint robotic manipulators during constrained motion tasks, *IEEE Trans. Robotics and Automation,* 8, 1, 119–125, 1992.
31. Brogliato, B. and Lozano-Leal, R., Adaptive force/position control of constrained manipulators with flexible joints, *Proc. Am. Control Conf.,* Boston, 968–969, 1992.
32. Krishnan, H., Robot force and position control including effects of joint flexibility and actuator dynamics, *Proc. 26th ISIR,* Singapore, 185–190, 1995.
33. Raibert, M.H. and Craig, J.J., Hybrid position/force control of manipulators, *J. Dynam. Syst., Meas. Control,* 103, 126–133, 1981.
34. Khatib, O., A unified approach for motion and force control of robot manipulators: the operational space formulation, *IEEE J. Robotics Automation,* 5, 3, 107–115, 1987.
35. Hogan, N., Impedance control: an approach to manipulation, Part I–III, *J. Dynam. Syst., Meas. Control,* 107, 1–24, 1985.
36. Vukobratović, M. and Ekalo, Y., New approach to control of robotic manipulators interacting with dynamic environment, *Robotica,* 14, 1, 31–39, 1996.
37. Vukobratović, M., How to control robots interacting with dynamic environment, *J. Intelligent Robotic Systems,* 19, 119–152, 1997.
38. Chumenko, V.N. and Yuschenko, A.S., Impact effects upon manipulation robot mechanism, (in Russian), *Technical Cybernetics,* 22, 4, 103–108, 1981.
39. Zheng, Y.F. and Hemami, H., Mathematical modeling of a robot collision with its environment, *J. Robotic Systems,* 2(3), 289–307, 1985.
40. Keller, J.B., Impact with friction, *J. Appl. Mech.,* 53(1–4), 1986.
41. Stronge, W.J., Rigid body collision with friction, *Proc. R. Soc. Lond. A.,* 431, 169–181, 1990.
42. Hurmuzlu, Y. and Marghitu, D.B., rigid body collision of planar kinematic chain with multiple contact points, *Int. J. Robotic Res.,* 13, 1, 82–92, 1994.
43. Drenovac, V. and Potkonjak, V., Contribution to the modeling of impact with robotic systems, *Robotica,* 11, Pt. 4, 445–452, 1993.
44. Brogliato, B. and Orhant, P., On the transition phase in robotics: impact models, dynamics and control, *Proc. IEEE Conf. Robotics Automation,* San Diego, 1994.
45. Acaccia, G.M., Gagetti, P.C., Callegari, M., Michelini, R.C., and Molfino, R.M., Modeling the impact dynamics of robotic manipulators, *Proc. 4th IFAC Symp. Robot Control,* Capri, Italy, 559–564.
46. Tornambe, A., Modeling and controlling two-degrees-of-freedom impacts, *Proc. 3rd IEEE Mediterranean Conf. Control Automation,* Lymassol, Cyprus, 1995.
47. Hogan, N., Sharon, A., and Hardt, E.D., High bandwidth force regulation and inertia reduction using a macro/micro manipulator system, *Proc. IEEE Conf. Robotics Automation,* 126–132, 1988.
48. Potkonjak, V. and Krstulovic, A., Contribution to the kinematics and dynamics of redundant robots via distributed positioning, *J. Intelligent Robotic Systems,* 5, 229–239, 1992.
49. Potkonjak, V., Popovic, M., Lazarevic, M., and Sinanovic, J., Redundancy problem in writing: from human to anthropomorphic robot arm, *IEEE Trans. SMC,* 28, 6, 790–806, 1998.
50. Vukobratović, M., Potkonjak, V., Systems with variable geometry: Concept and prospects, *ASME J. Dyn. Syst., Meas. Control,* 2, 121, 308–312, 1999.
51. Vukobratović, M., Potkonjak, V., and Matijevic, V., Internal redundancy — the way to improve robot dynamics and control performances, *J. Intelligent Robotic Systems,* 27, 31–66, 2000.

© 2002 by CRC Press LLC

21

Actuators and Computer-Aided Design of Robots

Miomir Vukobratović
Mihajlo Pupin Institute

Veljko Potkonjak
University of Belgrade

Kenji Inoue
Osaka University

Masaharu Takano
Kansai University

21.1 Robot Driving Systems
Present State and Prospects • DC Motors: Principles and Mathematics • How to Mount Motors to Robot Arms • Hydraulic Actuators: Principles and Mathematics • Pneumatic Actuators: Principles and Mathematics

21.2 Computer-Aided Design
Robot Manipulator Design Problem • Robot Design Procedure • Design Condition Input • Fundamental Mechanism Design • Inner Mechanism Design • Detailed Structure Design • Design Example

At the beginning of a discussion on robot design one should recall the history of robotics. During the early stage of robotics, no exact theory existed to assist engineers in designing robots. The designers followed the rich experience of machine building. In the 1970s, the theory of robotics started to grow fast. At the same time, industry manufactured and implemented rather complex robots capable of solving many industrial tasks. However, there was little connection between theory and industrial practice. The theory of robots was too academic. The problems considered were often too advanced for the industrial robotics of that time. Theoretical research dealt with mathematical modeling of robot dynamics, problems of control of nonlinear multivariable systems like robots, stability of control, even force feedback, etc. It seems that robot industry did not believe the need for some exact theory.

Experience in machine building and control represented sufficient background for design of many successful robots. Presently, the necessity for complex, precise, and high-speed robots requires a close connection between theory and practice. Regarding the application of robot dynamics, the main break-through was made when computer-aided methods for dynamic modeling were developed (see Chapter 20). Such methods allowed fast and user-friendly calculation of all relevant dynamic effects. It became possible to examine the a robot's behavior in advance, that is before it was actually built. A mathematical model replaced the real system. Such simulation was relevant depending on the quality of the model. In the beginning, the models were restricted to open-type linkages. Links were considered infinitely rigid and joints frictionless. In spite of these approximations, the dynamic model covered the main effects, inertial behavior of the spatial robotic system. Later, other relevant effects were included as explained in Section 20.5.

If a simulation system based on dynamic model is supplemented with appropriate testing of the dynamic characteristics and the user-friendly interface for changing robot parameters, one obtains

© 2002 by CRC Press LLC

a very useful design tool. A designer can examine the influence of certain parameters to robot performance and then change the parameters to improve the results. In this way, step by step, he or she approaches the optimal design. Finally, it is possible to create a software system that includes optimization procedures, thus automating the choice of robot parameters. This is a brief idea of something called computer-aided design (CAD).[1,16]

When selecting the topics for this chapter devoted to robot design we started from the fact that technology grows fast. Thus, some currently advanced constructive solutions might soon become obsolete. Hence, we decided to avoid presentation of specific constructive solutions and try to explain advanced principles of robot design. First, it was necessary to discuss robot-driving systems. It is important because the choice of actuator type (electric, hydraulic, or pneumatic) is one of the first decisions in the design process and many constructive solutions depend on this choice. Also, dynamic models of actuators are needed for knowledge of overall robot dynamics and to create the simulation system. Actuators and their impact to robot design are discussed in Section 21.1; Section 21.2 gives the principles of advanced design. A CAD system for industrial robots is described.

21.1 Robot Driving Systems

Discussion on robot-driving systems is important for several reasons. First, we address the problem of dynamic modeling. The actuators represent a subsystem of the entire robot. It is often said that a robot consists of a mechanical part (robot mechanism) and actuators. For mathematical modeling of robot dynamics it is necessary to take care of all dynamic effects, those introduced by the mechanism (e.g., link inertia) and those due to actuators (e.g., rotor inertia, counter electromotive forces, etc.).[1,2] Such a model of the complete dynamics is derived in Sections 21.3.1 and 21.3.2.

The problem of control is strongly influenced by the choice of actuators. For instance, DC motors, stepper motors, and hydraulic actuators require different hardware and software solutions. The problem of control is closely related to the driving characteristics of different types of actuators. Finally, constructive solutions of the robot's mechanical part depend on the choice of actuators. For instance, if a hydraulic cylinder drives a robot elbow, it is attached to the upperarm and the forearm, representing a kind of direct drive. On the other hand, if a DC motor is used, it is usually displaced from the elbow and located on the robot base. This concept understands a mechanism for transmitting motor power to the joint.

So, when elaborating robot actuators it is necessary to stress the following points: operation principles, mathematical modeling, driving characteristics, and mounting on the robot arm. Section 21.1.1 presents a review of actuators currently used in robots and automation. The main characteristics, advantages, and drawbacks are mentioned without detailed explanation. The idea is to stress those points that are important for implementation. In the paragraphs that follow we discuss the principles of operation and the mathematical description of most common types of actuators. Some constructive aspects of actuator implementation are also considered (especially the transmission). Presentation covers DC motors, hydraulic actuators, and pneumatic drive.

21.1.1 Present State and Prospects

In the early stage of robotics, pneumatic cylinders were often used to drive the manipulation mechanisms. Such devices had limited motion possibilities. This follows from the binary character of pneumatic actuators. The piston can extend to the final position or retract to the initial state and no control is achieved between these two positions. This is due to the compressibility of the air that flows through the cylinder. Thus, the manipulator can reach a set of points in space and programming of motion means only the definition of the sequence of working points. Although some special designs of pneumatic drives offer the possibility of achieving closed-loop control, such actuators are not widely used in advanced robotic systems. However, there is still a need for

© 2002 by CRC Press LLC

pick-and-place industrial systems positioned by mechanical stops. For such devices pneumatic actuation represents a fast, cheap, and reliable solution.

The hydraulic actuator is to some extent similar to the pneumatic one but avoids its main drawbacks. The uncompressible hydraulic oil flows through a cylinder and applies pressure to the piston. This pressure force causes motion of the robot joint. Control of motion is achieved by regulating the oil flow. The device used to regulate the flow is called a servovalve. Hydraulic systems can produce linear or rotary actuation. There are many advantages of the hydraulic drives. Its main benefit is the possibility of producing a very large force (or torque) without using geartrains. At the same time, the effector attached to the robot arm allows high concentration of power within small dimensions and weight. This is due to the fact that some massive parts of the actuator, like the pump and the oil reservoir, are placed beside the robot and do not load the arm. With hydraulic drives it is possible to achieve continuous motion control. The drawbacks one should mention are:

Hydraulic power supply is inefficient in terms of energy consumption
Leakage problem is present.
A fast-response servovalve is expensive.
If the complete hydraulic system is considered (reservoir, pump, cylinder and valve), the power supply becomes bulky.

Electric motors (electromagnetic actuators) are the most common type of actuators in robots today. They are used even for heavy robots for which some years ago hydraulics was exclusive. This can be justified by the general conclusion that electric drives are easy to control by means of a computer. This is especially the case with DC motors. However, it is necessary to mention some drawbacks of electromagnetic actuation. Today, motors still rotate at rather high speed. Rated speed is typically 3000 to 5000 r.p.m. At the same time, the output motor torque is small compared with the value needed to move a robot joint. For instance, rated torque for a 250W DC motor with rare-earth magnets may be 0.9 Nm. Hence, electric motors are in most cases followed by a reducer (gear-box), a transmission element that reduces speed and increases torque. It is not uncommon for a large reduction ratio to be needed (up to 300). The always present friction in gear-boxes produces loss of energy. The efficiency (output to input power ratio) of a typical reducer, the Harmonic Drive, is about 0.75. The next problem is backlash that has a negative influence on robot position accuracy. Similar problems may arise from the unsatisfactory stiffness of the transmission.

An important question concerns the allocation of the motor on the robot arm. To unload the arm and achieve better static balance, motors are usually displaced from the joints they drive. Motors are moved toward the robot base. In such cases, additional transmission is needed between the motor and the corresponding joint. Different types of shafts, chains, belts, ball screws, and linkage structures may be used. The questions of efficiency, backlash, and stiffness are posed again. Finally, the presence of transmission elements makes the entire structure more complex and expensive. This main disadvantage of electric motors can be eliminated if direct drive is applied. This understands motors powerful enough to operate without gearboxes or other types of transmission. Such motors are located directly in the robot joints. Direct drive motors are used in advanced robots, but not very often. Problems arise if high torques are needed. However, direct drive is a relatively new and very promising concept.[5]

The most widely used electromagnetic drive is the permanent magnet DC motor. Classical motor structure has a rotor with wire windings and a stator with permanent magnets and includes brush-commutation. There are several forms of rotors. A cylindrical rotor with iron has high inertia and slow dynamic response. An ironless rotor consists of a copper conductor enclosed in a epoxy glass cup or disk. A cup-shaped rotor retains the cylindrical-shaped motor while the disc-shaped rotor allows short overall motor length. This might be of importance when designing a robot arm. A disadvantage of ironless armature motors is that rotors have low thermal capacity. As a result, motors have rigid duty cycle limitations or require forced-air cooling when driven at high torque

© 2002 by CRC Press LLC

levels. Permanent magnets strongly influence the overall efficiency of motors. Low-cost motors use ceramic (ferrite) magnets. Advanced motors use rare-earth (samarium-cobalt and neodymium-boron) magnets. They can produce higher peak torques because they can accept large currents without demagnetization. Such motors are generally smaller in size (better power to weight ratio). However, large currents cause increased brush wear and rapid motor heating.

The main drawback of the classical structure comes from commutation. Graphite brushes and a copper bar commutator introduce friction, sparking, and the wear of commutating parts. Sparking is one of the factors that limits motor driving capability. It limits the current at high rotation speed and thus high torques are only possible at low speed. These disadvantages can be avoided if wire windings are placed on the stator and permanent magnets on the rotor. Electronic commutation replaces the brushes and copper bar commutator and supplies the commutated voltage (rectangular or trapezoidal shape of signal). Such motors are called brushless DC motors. Sometimes, the term synchronous AC motor is used although a difference exists (as will be explained later). In addition to avoiding commutation problems, increased reliability and improved thermal capacity are achieved. On the other hand, brushless motors require more complex and expensive control systems. Sensors and switching circuitry are needed for electronic commutation.

The synchronous AC motor differs from the brushless DC motor only in the supply. While the electronic commutator of a brushless DC motor supplies a trapezoidal AC signal, the control unit of an AC synchronous motor supplies a sinusoidal signal. For this reason, many books and catalogues do not differentiate between these two types of motors.

Inductive AC motors (cage motors) are not common in robots. They are cheap, robust, and reliable, and at the same time offer good torque characteristics. However, control of such motors is rather complicated. Advanced vector controllers are expensive and do not guarantee the same quality of servo-operation as DC motors. Still, it should be pointed out that these motors should be regarded as prospective driving systems. The price of controllers has a tendency to decrease and control precision is being improved constantly. Presently, cage AC motors are used for automated guided vehicles, and for different devices in manufacturing automation.

Stepper motors are often used in low-cost robots. Their main characteristic is discretized motion. Each move consists of a number of elementary steps. The magnitude of the elementary step (the smallest possible move) depends on the motor design solution. The hardware and software needed to control the motor are relatively simple. This is because these motors are typically run in an open-loop configuration. In this mode the position is not reliable if the motor works under high load — the motor may loose steps. This can be avoided by applying a closed-loop control scheme, but at a higher price.

Let us now discuss some ideas for robot drives that are still the topic of research. First, we notice that all the discussed actuators can be described as kinematic pairs of the fifth class, i.e., pairs that have one degree of freedom (DOF). Accordingly, such an actuator drives a robot joint that also has one DOF. This means that multi-DOF joints must not appear in robots, or they have to be passive. If a multi-DOF connection is needed, it is designed as a series of one-DOF joints. However, with advanced robots it would be very convenient if true multi-DOF joints could be utilized. As an example, one may consider humanoid robots that really need spherical joints (for shoulder and hip). To achieve the possibility of driving a true spherical joint one needs an actuation element that could be called an artificial muscle. It should be long, thin, and flexible. Its main feature would be the ability to control contraction. Although there have been many varying approaches to this problem (hydraulics, pneumatics, materials that change the length in a magnetic field or in contact with acids, etc.), the applicable solution is still missing.

21.1.2 DC Motors: Principles and Mathematics

DC motors are based on the well-known physical phenomenon that a force acting upon a conductor with the current flow appears if this conductor is placed in a magnetic field. Hence, a magnetic

© 2002 by CRC Press LLC

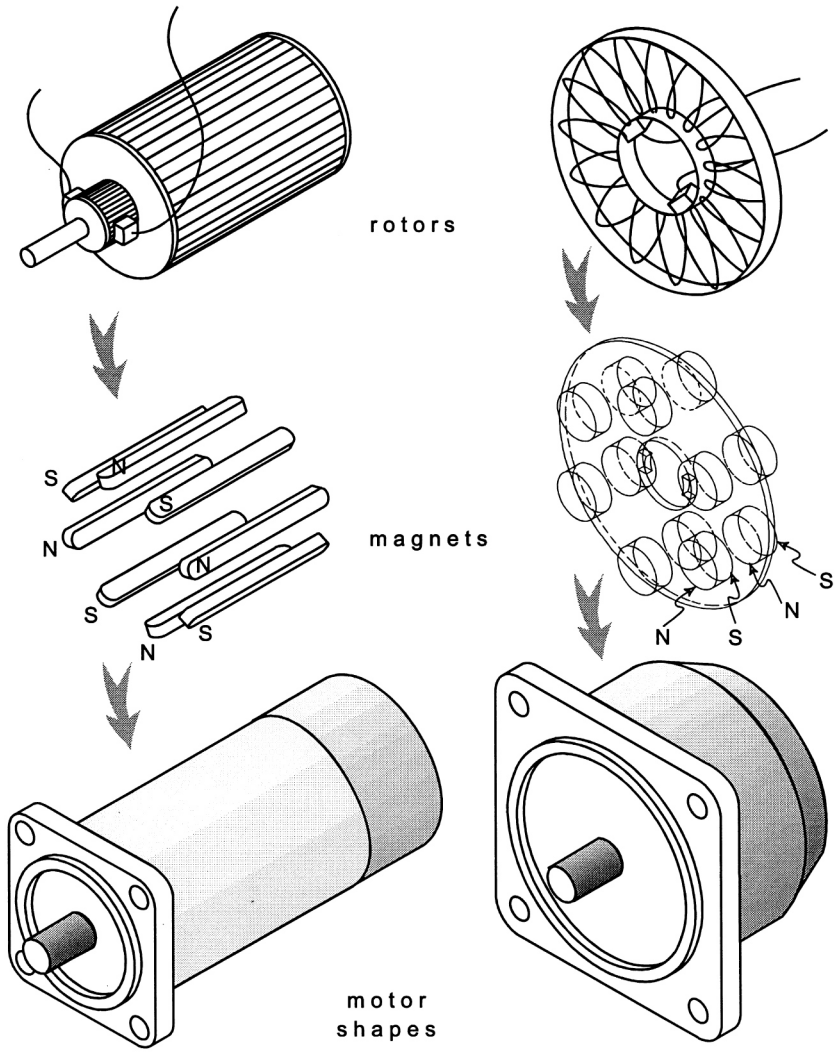


FIGURE 21.1 Different rotor shapes enable different overall shape of motors.

field and electrical circuit are needed. Accordingly, a motor has two parts, one carrying the magnets (we assume permanent magnets because they are most often used) and the other carrying the wire windings. The classical design means that magnets are placed on the static part of the motor (stator) while windings are on the rotary part (rotor). This concept understands brush-commutation. An advanced idea places magnets on the rotor and windings on the stator, and needs electronic commutation (brushless motors). The discussion starts with the classical design.

Permanent magnets create magnetic field inside the stator. If current flows through the windings (on rotor), force will appear producing a torque about the motor shaft. Figure 21.1 shows two rotor shapes, cylindrical and disc. Placement of magnets and finally the overall shape of the motor are also shown.

Let the angle of rotation be θ. This coordinate, together with the angular velocity $\dot{\theta}$, defines the rotor state. If rotor current is i, then the torque due to interaction with the magnetic field is $C_M i$. The constant C_M is known as the torque constant and can be found in catalogues. This torque has to solve several counter-torques. Torque due to inertia is $J\ddot{\theta}$, where J is the rotor's moment of

© 2002 by CRC Press LLC

inertia and $\ddot{\theta}$ is angular acceleration. Torque that follows from viscous friction is $B\dot{\theta}$, where B is the friction coefficient. Values for J and B can be found in catalogues. Finally, the torque produced by the load has to be solved. Let the moment of external forces (load) be denoted by M. Very often this moment is called the output torque. Now, equilibrium of torques gives

$$C_M i = J\ddot{\theta} + B\dot{\theta} + M \tag{21.1}$$

To solve the dynamics of the electrical circuit we apply the Ohm's law. The voltage u supplied by the electric source covers the voltage drop over the armature resistance and counter-electromotive forces (e.m.f.):

$$u = Ri + C_E\dot{\theta} + Ldi / dt \tag{21.2}$$

Ri is the voltage drop where R is the armature resistance. $C_E\dot{\theta}$ is counter e.m.f. due to motion in magnetic field and C_E is the constant. Finally, Ldi/dt is counter e.m.f. due to self-inductance, where L is inductivity of windings. Values R, C_E, and L can be found in catalogues. The dynamics of electrical circuit introduces one new state variable, current i.

Equations (21.1) and (21.2) define the dynamics of the entire motor. If one wishes to write the dynamic model in canonical form, the state vector $x = [\theta\ \dot{\theta}\ i]^T$ should be introduced. Equations (21.1) and (21.2) can now be united into the form

$$\dot{x} = Cx + fM + du \tag{21.3}$$

The system matrices are

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -B/J & C_M/J \\ 0 & -C_E/L & -R/L \end{bmatrix},\ f = \begin{bmatrix} 0 \\ -1/J \\ 0 \end{bmatrix},\ d = \begin{bmatrix} 0 \\ 0 \\ 1/L \end{bmatrix} \tag{21.4}$$

This is the third-order model of motor dynamics.

If inductivity L is small enough (it is a rather common case), the term Ldi/dt can be neglected. Equation (21.2) now becomes

$$u = Ri + C_E\dot{\theta} \tag{21.5}$$

and the number of state variables reduces to two. The state vector and the system matrices in eq. (21.3) are

$$x = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix},\ C = \begin{bmatrix} 0 & 0 \\ 0 & -C_M C_E / RJ - B/J \end{bmatrix},\ f = \begin{bmatrix} 0 \\ -1/J \end{bmatrix},\ d = \begin{bmatrix} 0 \\ C_M / R_J \end{bmatrix} \tag{21.6}$$

The motor control variable is u. By changing the voltage, one may control rotor speed or position.

If the motor drives a robot joint, for instance, joint j, we relate the motor with the joint by using index j with all variables and constants in the dynamic model (21.3). This was done in Section 20.3.1. when the motor model is integrated with the arm links model to obtain the dynamic model of the entire robot. There the second-order model in the form of Equations (21.1) and (21.5) was

© 2002 by CRC Press LLC

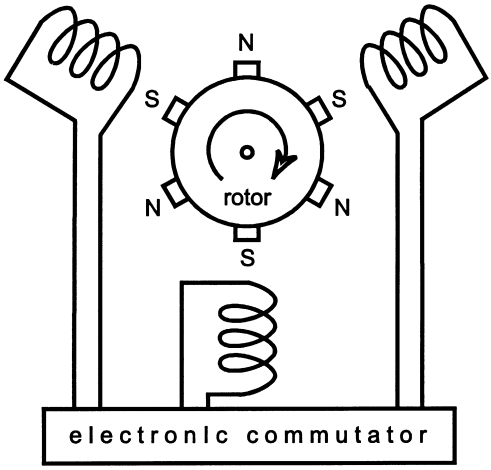


FIGURE 21.2 Scheme of brushless motor.

applied. If the third-order model is to be used, then the canonic form of motor dynamics, Equation (21.3), is combined with arm dynamics as explained in Section 20.3.2.

As already stated, the main disadvantage of the classical design of DC motors follows from brush-commutation. To avoid it, brushless motors place permanent magnets on the rotor and wire windings on the stator (Figure 21.2). The interaction between the magnetic field and the electrical circuit, which forces the rotor to move, still exists. Brushes are not needed because there is no current in the rotor. To synchronize switching in the electrical circuit and the angular velocity, Hall's sensors are used. They give the information for the device called an electronic commutator. In this way the electronic commutator imitates the brush commutation. We are not going to discuss the details of such a commutation system. Figure 21.2 shows the scheme of a brushless motor with three pairs of magnetic poles and three windings.

Let us briefly discuss the voltage supplied to the windings. It is a rectangular or trapezoidal signal switching between positive and negative values. Switching in a winding shifts with respect to the preceding winding. Because periods of constant voltage exist, we still deal with a DC motor. However, better performances can be achieved if a trapezoidal voltage profile is replaced with a sinusoidal one. In this case we have a three-phase AC supply, producing a rotating magnetic field of constant intensity. The magnetic force appears between the rotating field and the permanent magnets placed on the rotor, causing rotor motion. The rotating field pulls the rotor and they both rotate at the speed defined by the frequency of the AC signal. Changing the frequency, one may control the motor speed. This concept is called the synchronous AC motor. It is clear that the difference between a DC brushless motor and an AC synchronous motor is only in the supply.

21.1.3 How to Mount Motors to Robot Arms

When searching for the answer to the question posed in the heading, we face two criteria that conflict with each other. First, we prefer to use direct drive motors. They eliminate transmission and thus simplify arm construction and avoid backlash, friction, and deformation. Direct drive motors are used in robots, but not very often. Particularly, they are not appropriate for joints that are subject to a large gravitational load. The other criterion starts with the demand to unload the arm. With this aim, motors are displaced from the joints they drive. Motors are moved toward the robot base, creating better statics of the arm and reducing gravity in terms in joint torques. This concept introduces the need for a transmission mechanism that would connect a motor with the

© 2002 by CRC Press LLC

corresponding joint. The presence of a transmission complicates the arm design (thus increasing the price) and introduces backlash (leading to lower accuracy when positioning some object), friction (energy loss due to friction and problems in controlling the system with friction), and elastic deformation (undesired oscillations). Despite all these drawbacks, some type of transmission is present in the majority of robots. It should be noted that the role of transmission is threefold. First, power is transmitted at distance. Second, speed can be reduced and torque increased if needed. Finally, it is possible to change the character of motion from the input to the output of transmission system: rotation to translation (R/T) or translation to rotation (T/R). If such change is not needed, the original character is kept: rotation (R/R) and translation (T/T). Here, we review some typical transmission systems that appear in robots, paying attention to the three mentioned roles of transmission.[3]

Spur gearing is an R/R transmission that has low backlash and high stiffness to stand large moments. It is not used for transmitting at a distance, but for speed reduction. One pair of gears has a limited reduction ratio (up to 10), and thus, several stages might be needed; however, the system weight, friction, and backlash will increase. This transmission is often applied to the first rotary arm axis. *Helical gears* have some advantages over spur gears. In robots, a large reduction of speed is often required. The problem with spur gears may arise from lack of an adequate gear tooth contact ratio. Helical gears have higher contact ratios and hence produce smoother output. However, they produce undesired axial gear loads. The mentioned gearing (spur and helical) is applied if the input and output rotation have parallel axes. If the axes are not parallel, then *bevel gearing* may be applied. An example of bevel gearing in a robot wrist is shown in Figure 21.7.

Worm gear allows a high R/R reduction ratio using only one pair. The main drawbacks are increased weight and friction losses that cause heat problems (e.g., efficiency less than 0.5).

Planetary gear is an R/R transmission used for speed reduction. The reduction ratio may be high but very often several stages are needed. Disadvantages of this system are that it is heavy in weight and often introduces backlash. So-called zero-backlash models are rather expensive. Note that buying a motor and a gearbox already attached to it and considering this assembly as one unit are recommended.

Harmonic drive is among the most common speed reduction systems in robots. This R/R transmission allows a very high reduction ratio (up to 300 and even more) using only one pair. As a consequence, compact size is achieved. Another advantage is small backlash, even near zero if selective assembly is conducted in manufacturing the device. On the other hand, static friction in these drives is high. The main problem, however, follows from the stiffness that allows considerable elastic deformation. Such torsion in joints may sometimes compromise robot accuracy.

Cyclo reducer is a R/R transmission that may increase the speed ratio up to 120 at one stage. As advantages, we also mention high stiffness and efficiency (0.75 to 0.85). The main drawbacks are heaviness and high price.

Toothed rack-and-pinion transmission allows R/T and T/R transformation of motion. In robots, R/T operation appears when long linear motion has to be actuated by an electric motor. The rack is attached to the structure that should be moved and motor torque is applied to the pinion (Figure 21.3a). The same principle may be found in robot grippers. T/R transmission can be applied if the hydraulic cylinder has to move a revolute joint. One example, actuation of rotary robot base, is shown in Figure 21.3b. Rack-and-pinion transmission is precise and inexpensive.

Recirculating ball nut and screw represent a very efficient R/T transmission. It also provides very high precision (zero backlash and high stiffness) and reliability along with great reduction of speed. A quality ball screw is an expensive transmission. One example of a ball screw applied in robots is presented in Figure 21.4. It is used to drive the vertical translation in a cylindrical robot.

Linkages and linkage structures may be considered transmission elements, although they are often structural elements as well. They feature very high stiffness and efficiency and small backlash. In Figure 21.5 a ball screw is combined with a linkage to drive the forearm of the ASEA robot.

© 2002 by CRC Press LLC

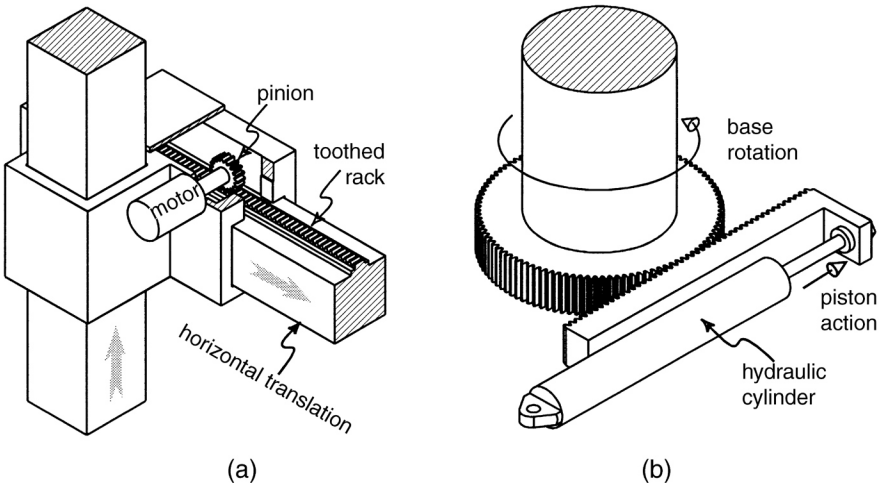


FIGURE 21.3 Toothed rack-and-pinion transmission.

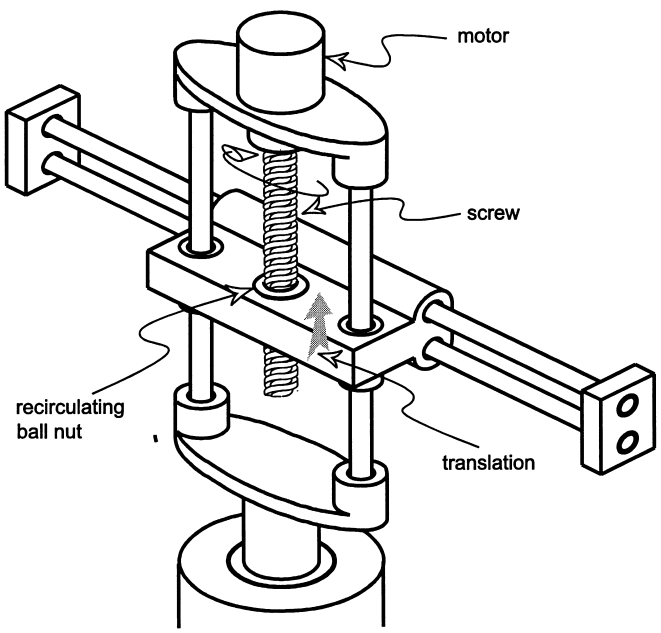


FIGURE 21.4 Application of ball screw transmission to vertical linear joint of a cylindrical robot.

Torsion shafts or torque tubes are R/R transmissions often used in robots to transmit power at a distance. They do not reduce speed. The problem of torsion deformation always exists with such systems. For this reason, it is recommended to transmit power at high speed (and low torque) because it allows smaller diameter and wall thickness, and lower weight. An example is shown in Figure 21.6. Wrist motors are located to create a counterbalance for the elbow. Motor power is transmitted to the wrist by means of three coaxial torque tubes.

Toothed belts can be found in low-cost robots. They are used to transmit rotary motion (R/R) at long distances. It is possible to reduce rotation speed, but it is not common. The usual speed ratio is 1:1. Toothed belt transmissions are very light in weight, simple, and cheap. The problems follow mainly from backlash and elastic deformation that cause vibrations. Figure 21.7 shows how the

© 2002 by CRC Press LLC

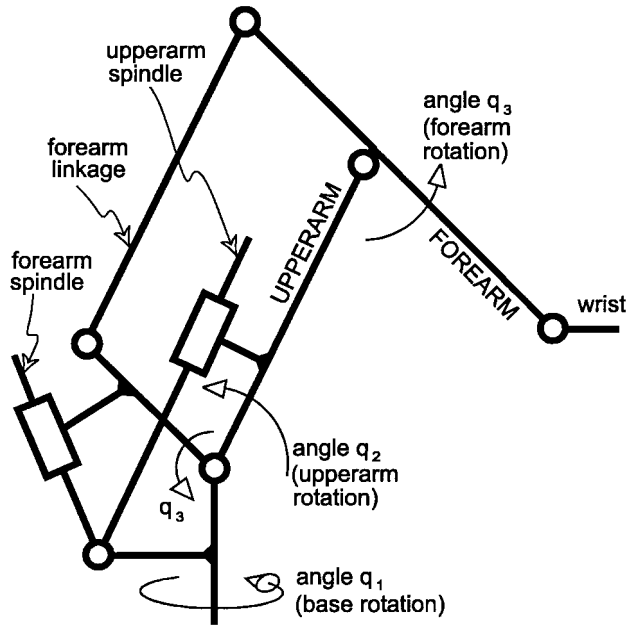


FIGURE 21.5 Ball screw combined with a linkage transmission.

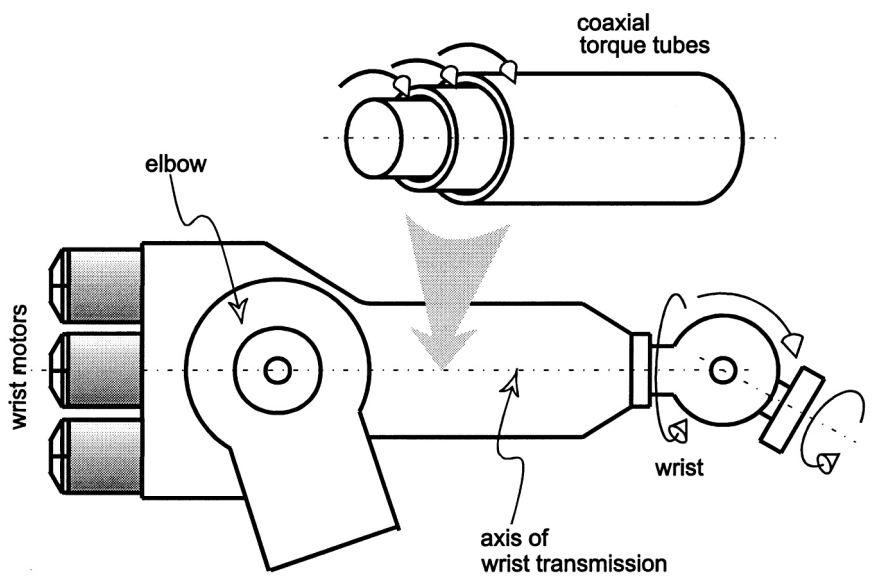


FIGURE 21.6 Wrist motors are used as a counterbalance and power is transmitted by means of coaxial torque tubes.

wrist can be driven by motors located at the robot base. Three belts are used for each motor to transmit power to the joint. In the wrist, bevel gearing is applied. The combined action of two motors can produce pitch and roll motion.

Chain drive can replace the toothed belt for transmitting rotary motion at a distance. It has no backlash and can be made to have stiffness that prevents vibrations. However, a chain transmission is heavy. Chain is primarily used as an R/R transmission, but sometimes it is applied for R/T and T/R operations.

© 2002 by CRC Press LLC

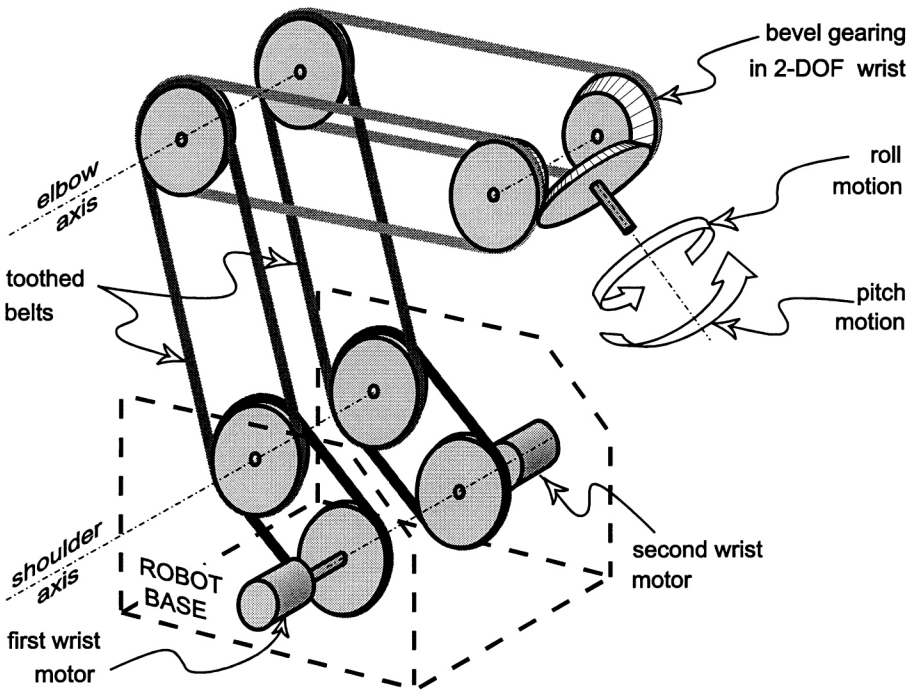


FIGURE 21.7 Motors driving the wrist are located at the robot base.

Mathematical model of transmission. Let us discuss the mathematical representation of transmission systems. If some actuator drives a robot joint, then motor motion θ and motor torque M represent the input for the transmission system. Joint motion q and joint torque τ are the output. An ideal transmission is characterized by the absence of backlash, friction, elastic deformation (infinite stiffness), and inertia. In modeling robot dynamics this is a rather common assumption. In such a case, there is a linear relation between the input and the output:

$$q = \theta/N, \tag{21.7}$$

$$\tau = MN \tag{21.8}$$

where N is the reduction ratio. This assumption allows simple integration of motor dynamics to the dynamic model of robot links.

However, transmission is never ideal. If backlash is present, relation (21.7) does not hold. Modeling of such a system is rather complicated, and hence, backlash is usually neglected. Friction is an always-present effect. Neglecting it would not be justified. It is well known that static friction introduces many problems in dynamic modeling. For this reason, friction is usually taken into account through power loss. We introduce the efficiency coefficient η as the output-to-input power ratio. Note that $0 < \eta < 1$. Now, Relation (21.8) is modified. If the motion is in the direction of the drive, then $N\eta'$ is used instead of N. However, if the motion is opposite to the action of the drive, then N/η'' is applied. Note that η' and η'' are generally different. The efficiency of a transmission in the reverse direction is usually smaller $\eta_j'' < \eta_j'$.

If transmission stiffness is not considered to be infinite, then the elastic deformation should be taken into account. Relation (21.7) does not hold since q and θ become independent coordinates. However, stiffness that is still high will keep the values q and θ/N close to each other. To solve the elastic deformation, one must know the values of stiffness and damping. The problem becomes even more complex if the inertia of transmission elements is not neglected. In that case, the

© 2002 by CRC Press LLC

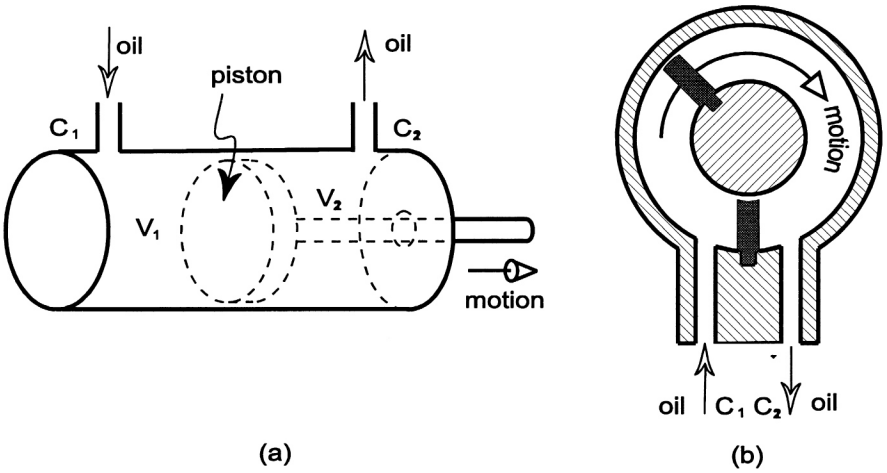


FIGURE 21.8 Hydraulic cylinder (a) and hydraulic vane motor (b).

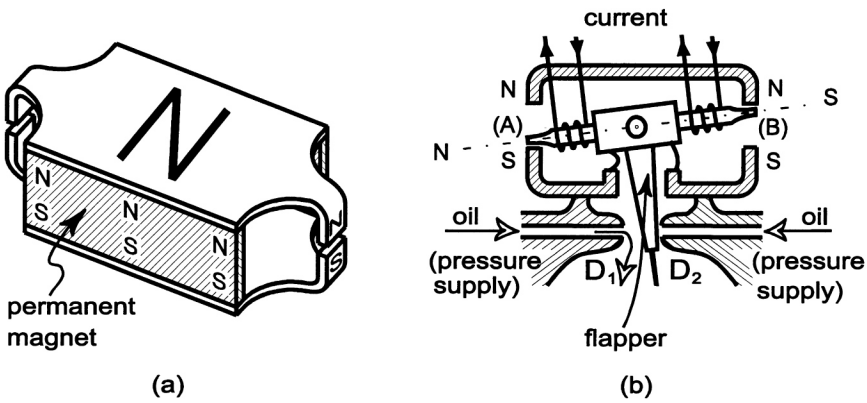


FIGURE 21.9 Torque motor: structure and operation.

transmission system requires dynamic modeling. One approach to this problem was presented in Section 20.5.4.

21.1.4 Hydraulic Actuators: Principles and Mathematics

Hydraulic servoactuator consists of a cylinder with a piston, a servovalve with a torque motor, an oil reservoir, and a pump. The term electrohydraulic actuator is also used. A reservoir and pump are necessary for the operation of the hydraulic system, but they are not essential for explaining operation principles. So, we restrict our consideration to the cylinder and the servovalve. The pump is seen simply as a pressure supply. A cylinder with a piston is shown in Figure 21.8a. If the pump forces the oil into port C_1, the piston will move to the right and volume V_1 will increase, V_2 will decrease and the oil will drain through port C_2. Oil flow and the difference in pressure on the two sides of the piston define the direction and speed of motion as well as the output actuator force. The same principle can be used to create a rotary actuator, a hydraulic vane motor (Figure 21.8b).

We explain the servovalve operation by starting with the torque motor (magnetic motor). The scheme of the motor is presented in Figure 21.9. If current flows through the armature windings as shown in Figure 21.9b, magnetic north will appear on side *A* and south on side *B*. Interaction

© 2002 by CRC Press LLC

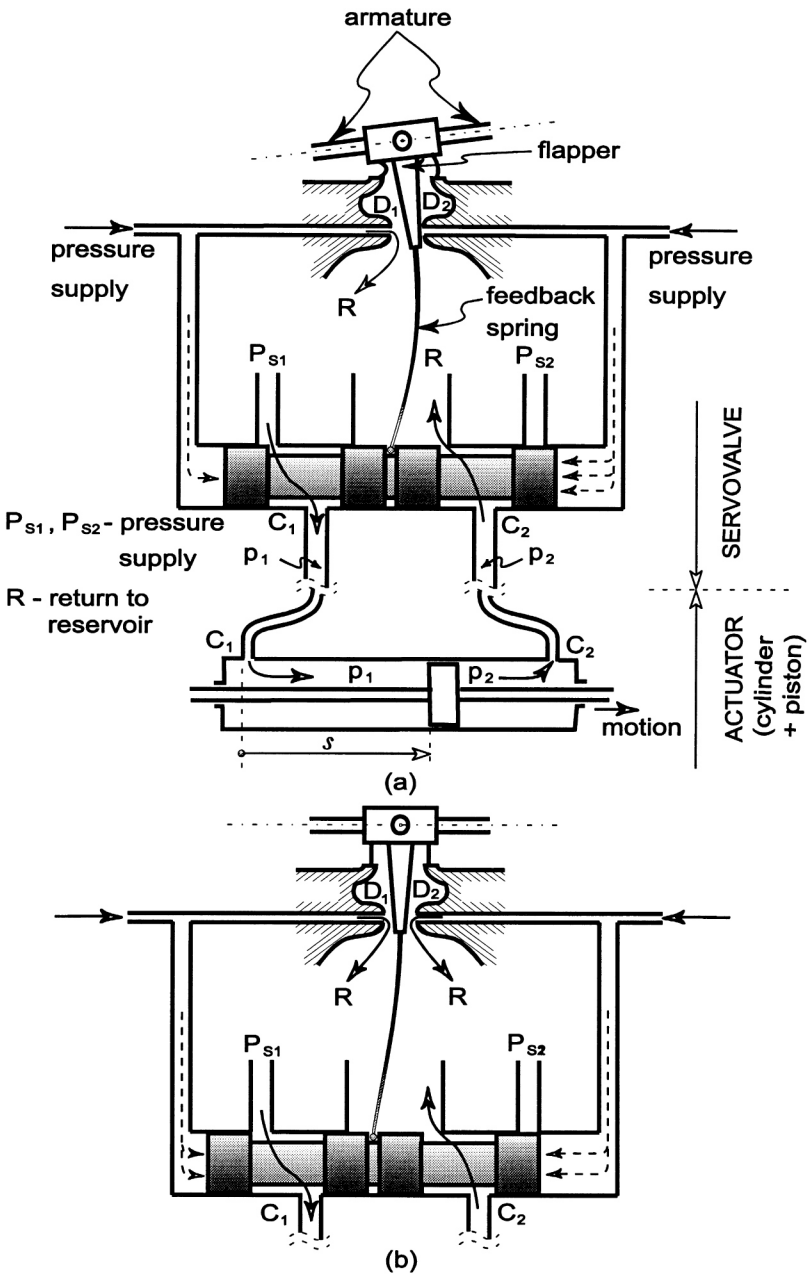


FIGURE 21.10 Operation of a servovalve.

with the permanent magnet will turn the armature to the left. Changing the current direction will turn the armature to the opposite side. When the armature moves, the flapper closes nozzle D_1 or D_2.

Figure 21.10 shows the complete servovalve. Let us explain how it works.[4] Suppose that current forces the armature to turn to the left (Figure 21.10a). The flapper moves to the right, thus closing nozzle D_2. The pressure supply line P_{s_2} is now closed and the oil from the left line, P_{s_1}, flows through pipe C_1 into the cylinder. The actuator piston moves to the right. Pipe C_2 allows the oil to flow out from the cylinder to the return line R (back to the reservoir). Since nozzle D_2 is closed,

© 2002 by CRC Press LLC

the oil in the right supply line exerts strong pressure upon the right-hand side of the servovalve piston forcing it to move to the left. This motion causes deformation of the feedback spring. At some deformation, the elastic torque of the deformed spring starts to turn the armature to the right and the flapper to the left, thus opening nozzle D_2. When the oil begins to flow through D_2, the pressure acting upon the right-hand side of the piston reduces, but it is still stronger than the pressure acting upon the left-hand side. Hence, the piston continues moving to the left. The pressure on both sides of the servovalve piston balances when the flows through D_1 and D_2 become equal. This means the vertical position of the flapper, that is, the horizontal position of the armature (Figure 21.10b). The motion of the piston stops. In this position the motor torque equals the spring deformation torque. Let coordinate z define the position of the servovalve piston. The equilibrium of torques may be expressed by the relation

$$C_M i = \gamma z \tag{21.9}$$

where C_M is the motor torque constant, i is the armature current, γ is the coefficient of elastic deformation torque, and z expresses the magnitude of deformation. The equilibrium position z of the servovalve corresponds to some value of oil flow and accordingly some velocity of the piston in the actuator cylinder. Since current i can change the motor torque, and thus position z (according to Equation (21.9)), the possibility of controlling the flow and the actuator speed is achieved. Current i represents the control variable. One should note that after the change of the current, a transient phase takes place before the new equilibrium is established. However, one may neglect dynamics of the servovalve and avoid analysis of the transient phase. In such case, Equation (21.9) is satisfied all the time and thus servovalve position z immediately follows the changes of the current. The nonlinear static characteristic of the servovalve (flow depending on the pressure and the piston position) has the form

$$Q = Dwz\sqrt{\frac{1}{\rho}\left(p_s - \mathrm{sgn}(z)p_d\right)} \tag{21.10}$$

where p_s is the pressure in the supply line, $p_d = p_1 - p_2$ is the differential pressure, sgn(z) is the sign of the position coordinate z, ρ is the oil density, w is the area gradient of rectangular port (the rate of change of orifice area with servovalve piston motion), and D is a dimensionless coefficient. Differential pressure means the difference in pressures in pipes C_1 and C_2, and at the same time, the difference in pressure on the two sides of the actuator piston. For this reason it is often called the load pressure.

When modeling the dynamics of an actuator we assume, for simplicity, symmetry of the piston (Figure 21.11). Let coordinate s define the position of the actuator piston. The pressures on the two sides of the piston are p_1 and p_2, and hence, the oil exerts the force to the piston: $p_1A - p_2A = p_dA$, where A is the piston area. Dynamic equilibrium of forces acting on piston gives

$$p_dA = m\ddot{s} + B\dot{s} + F \tag{21.11}$$

where m is the mass (total mass of the piston and load referred to the piston), B is the viscous friction coefficient, and F is the external load force on the piston (often called the output force).

Consider now oil flow through a cylinder (Figure 21.11) and denote it by Q. It consists of three components. The first component follows from the piston motion. It is a product of piston area and velocity, $A\dot{s}$. The second component is due to leakage. Since leakage depends on pressure, we introduce leakage coefficient c as leakage per unit pressure. There are two kinds of leakage, internal and external, as shown in Figure 21.11. If the coefficient of internal leakage is c_i and that of the external is c_e, and if the coefficient of total leakage is defined as $c = c_i + c_e/2$, then the flow due to leakage is cp_d. Finally, the third component follows from oil compression. Its value is $(V/4\beta)\,\dot{P}_d$,

© 2002 by CRC Press LLC

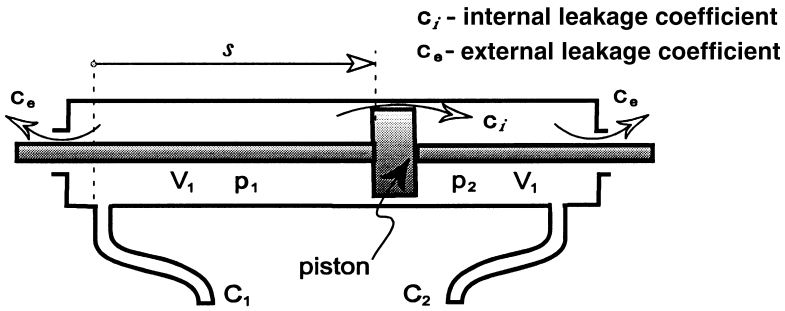


FIGURE 21.11 Oil flow through a hydraulic cylinder.

where V is the total volume ($V = V_1 + V_2$), and β is the compression coefficient. Total volume includes cylinder, pipes, and servovalve. Now, the flow is

$$Q = A\dot{s} + cp_d + \frac{V}{4\beta}\dot{p}_d \tag{21.12}$$

In this way we arrive at the mathematical model of the electrohydraulic actuator. The system dynamics is described by Equations (21.9) to (21.12). The model is nonlinear. The system state is defined by the three-dimensional vector $x = [s\ \dot{s}\ p_d]^T$. The control input is current i. The nonlinear model may be written in canonical form

$$\dot{x} = Cx + fF + d(x)i \tag{21.13}$$

where we tried to find analogy with the model (21.3) used for DC motors. Model matrices are

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -B/m & A/m \\ 0 & -(4\beta/V)A & -(4\beta/V)c \end{bmatrix}, \quad f = \begin{bmatrix} 0 \\ -1/m \\ 0 \end{bmatrix},$$
$$d(x) = \begin{bmatrix} 0 \\ 0 \\ (4\beta/V)Dw(C_M/\gamma)\sqrt{(1/\rho)(p_s - \mathrm{sgn}(z)p_d)} \end{bmatrix} \tag{21.14}$$

If a linear model is required, the expression (21.10) should be linearized by expansion into a Taylor series about a particular operating point $K(z_K, p_{dK}, Q_K)$:

$$Q - Q_K = \left.\frac{\partial Q}{\partial z}\right|_K (z - z_K) + \left.\frac{\partial Q}{\partial p_d}\right|_K (p_d - p_{dK}) \tag{21.15}$$

The most important operating point is the origin of the flow-pressure curve ($Q_K = p_{dK} = z_k = 0$). In such a case relation (21.15) becomes

$$Q = k_1 z + k_2 p_d \tag{21.16}$$

where: $k_1 = \partial Q/\partial z$ and $k_2 = \partial Q/\partial p_d$ are called the valve coefficients. They are extremely important in determining stability, frequency response, and other dynamic characteristics. The flow gain k_1

© 2002 by CRC Press LLC

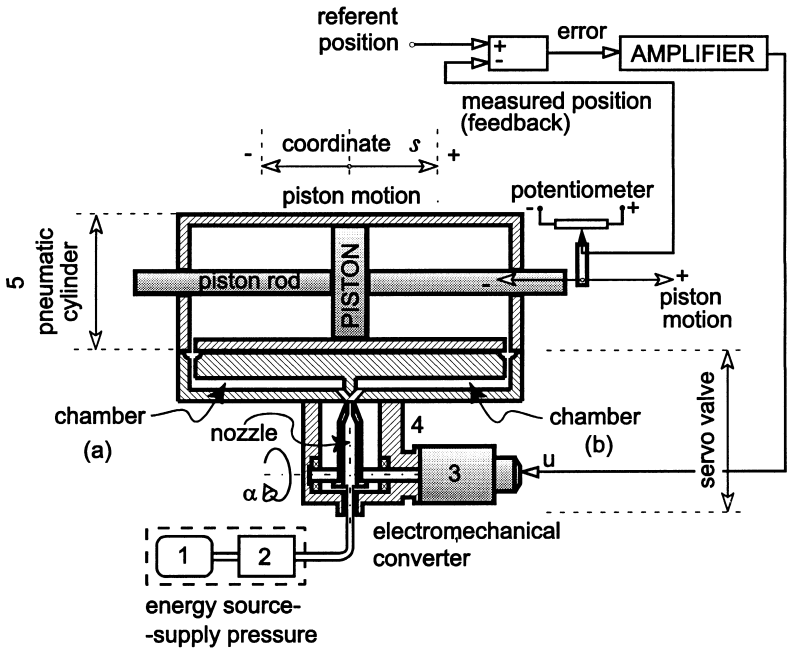


FIGURE 21.12 Scheme of a pneumatic servoactuator.

has a direct influence on system stability. The flow-pressure coefficient k_2 directly affects the damping ratio of valve–cylinder combination. Another useful quantity is the pressure sensitivity defined by $k_p = \partial p_d / \partial z = k_1 / k_2$. The pressure sensitivity of valves is quite high, which accounts for the ability of valve–cylinder combinations to break away large friction loads with little error. If (21.16) is used instead of (21.10), dynamic model (21.13) becomes linear with system matrices

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -B/m & A/m \\ 0 & -(4\beta/V)A & (4\beta/V)(k_2 - c) \end{bmatrix}, \quad f = \begin{bmatrix} 0 \\ -1/m \\ 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \\ (4\beta/V)k_1(C_M/\gamma) \end{bmatrix} \tag{21.17}$$

Model (21.13) in its linear or nonlinear form can be combined with arm dynamics, as explained in Section 20.3.2, to obtain the dynamic model of the complete robot system.

21.1.5 Pneumatic Actuators: Principles and Mathematics

A pneumatic servoactuator (often called a electropneumatic actuator) consists of an electropneumatic servovalve and a pneumatic cylinder with a piston. Figure 21.12 presents the scheme of the actuator. Let us explain how it operates.[5] Numbers 1 and 2 in the figure indicate an independent source of energy: (1) gas under pressure with (2) a valving and pressure reduction group. An electromechanical converter (3), a kind of torque motor, transforms the electrical signal (voltage u that comes from the amplifier) into an angle of its output shaft (angle α). The nozzle fixed to the shaft turns by the same angle. A mechanical-pneumatic converter (4) provides the difference in pressure and flow in chambers (a) and (b) proportional to the angle of the nozzle. The electromechanical converter and the mechanical-pneumatic converter together form the servovalve. The pneumatic cylinder (5) is supplied with differential pressure (p_d) and flow (Q_d), and hence, the piston moves. Thus, the voltage applied to the electromechanical converter represents the actuator-input variable that offers the possibility of controlling piston motion. Feedback is realized by using a sliding potentiometer (6). The potentiometer

© 2002 by CRC Press LLC

provides for voltage proportional to piston displacement. This is analog information describing the position of the piston. The information is used to form the error signal by subtracting this position from the referent position. The error signal is amplified and then applied to the electromechanical converter. In this way the closed-loop control scheme is obtained.

Let us describe the dynamics of the pneumatic servoactuator mathematically. We first find the relation between the input and the output of the electromechanical converter. If the inductivity of the coil is neglected, the input voltage u reduces to:

$$u = R_c i \tag{21.18}$$

where R_c is the resistance of the circuit and i is the current. If the dynamics of the rotating parts (rotor, shaft, nozzle) is neglected, the output angle α will be proportional to the current:

$$\alpha = K_i i \tag{21.19}$$

where K_i is the coefficient of proportionality.

The flow through the mechanical-pneumatic converter is

$$Q_d = K_\alpha \alpha + K_p p_d \tag{21.20}$$

where p_d is the differential pressure (in two chambers), K_α is the flow gain coefficient with respect to angle α, and K_p is the flow gain coefficient with respect to pressure.

Now we consider the cylinder. Let the coordinate s define the position of the piston. Flow through the pneumatic cylinder can be described by the relation

$$Q_d = \frac{M p_s \zeta A}{R T_s} \dot{s} + \frac{M V_0}{k R T_s} \dot{p}_d \tag{21.21}$$

where M is the molecular mass of gas, p_s is the supply pressure, ζ is the pressure-loss coefficient, R is the universal gas constant, T_s is the supply temperature, A is the active piston area, k is the polytropic exponent, V_0 is the total volume. Dynamic equilibrium of forces acting on piston gives

$$p_d A = m\ddot{s} + B\dot{s} + F \tag{21.22}$$

where m is the total piston mass (including rod and other load referred to the piston), B is the viscous friction coefficient, and F is the external load force on the piston (often called the output force). Note that there may exist other forces like dry friction (F_{fr} sgn $\dot{s}$) or linear force (cs). In such cases Equation (21.22) has to be augmented.

Equations (21.18) to (21.22) describe the dynamics of the electropneumatic actuator. If the equations are rearranged, canonical form of the dynamic model can be obtained. The system state is defined by the three-dimensional vector $x = [s\ \dot{s}\ p_d]^T$. The control variable is voltage u. Equations (21.18) to (21.22) can be united in the linear matrix model

$$\dot{x} = Cx + f F + du \tag{21.23}$$

where model matrices are

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -B/m & A/m \\ 0 & -p_s \zeta A k / V_0 & R T_s k K_p / M V_0 \end{bmatrix}, f = \begin{bmatrix} 0 \\ -1/m \\ 0 \end{bmatrix}, d = \begin{bmatrix} 0 \\ 0 \\ R T_s k K_\alpha K_i / M V_0 R_c \end{bmatrix} \tag{21.24}$$

© 2002 by CRC Press LLC

Model (21.23) can be combined with the arm dynamics, as explained in Section 20.3.2, to obtain the dynamic model of the complete robot system.

It should be said that electropneumatic servosystems cannot be applied practically for servodrives of robotic manipulators. This is because all gases to be applied as driving media are compressible, i.e., their specific volume is pressure dependent. In this way elasticity is introduced into the driving system. Under a load, especially in cases of longer strokes, large loads, and big pneumatic cylinders, this phenomenon leads to oscillations of loaded links of manipulator chain, thus rendering the electropneumatic drives practically unusable for robotic servodrives. This is a real situation present on the market and industry today. Pneumatic drives are applied in simple pick-and-place industrial systems positioned by mechanical stops.

The other variants of driving units need more extensive presentation.

21.2 Computer-Aided Design

As the number of industrial robots used in manufacturing systems increases and robots tend to be used in many nonindustrial fields, additional functions and performance improvements, such as high speed motion and high precision positioning, are desirable. It is, however, difficult to design robots by the conventional method of experimentation and trial manufacturing because robots involve many design parameters and evaluation functions. Accordingly, computer-aided design (CAD) is significant for designing suitable robots for objective tasks and saving manpower, time, and costs required for design.

21.2.1 Robot Manipulator Design Problem

Designing a robot manipulator (or robot) requires a determination of all design parameters of its mechanism.

- Fundamental mechanism:
  1. Degrees of freedom (D.O.F.)
  2. Joint types (rotational/sliding)
  3. Arm lengths and offsets
- Inner mechanism:
  1. Motor allocations
  2. Types of transmission mechanisms
  3. Motors
  4. Reduction gears and their reduction ratios
  5. Arm cross-sectional dimensions
  6. Machine elements

The designed robot should have suitable functions and abilities to perform certain tasks. The following design functions must be evaluated:

- Kinematic evaluation:
  1. Workspace
  2. Joint operating range
  3. Maximum workpiece velocity and acceleration
  4. Maximum joint velocity and acceleration
- Static/dynamic evaluation:
  1. Maximum motor driving torque
  2. Total motor power
  3. Total weight

© 2002 by CRC Press LLC

TABLE 21.1 Relationship between Design Parameters and Evaluation Functions

	Evaluation Function											
	Kinematics				Dynamics						Both	
Design Parameter	Workspace	Joint Disp. Limit	Max. Joint Vel./Acc.	Max. Workpiece Vel./Acc.	Max. Motor Torque	Total Motor Power	Total Weight	Weight Capacity	Deflection	Natural Frequency	Positioning Accuracy	Cost
Fundamental Mechanism												
D.O.F.	∀	Δ	Δ	Δ	Δ	○	○	Δ	○	○	○	∀
Joint type	∀	○	Δ	Δ	Δ	○	○	Δ	○	○	○	∀
Arm length	∀	Δ	∀	∀	∀	∀	∀	∀	∀	∀	∀	○
Offset	∀	∀	Δ	Δ	Δ	Δ	Δ	○	Δ	Δ	Δ	Δ
Inner Mechanism												
Motor allocation	Δ	Δ	Δ	Δ	∀	∀	∀	∀	Δ	Δ	Δ	Δ
Trans. mech.	Δ	∀	Δ	Δ	Δ	Δ	∀	Δ	∀	∀	∀	∀
Motor	Δ	Δ	∀	∀	∀	∀	∀	∀	Δ	Δ	Δ	∀
Reduction gear	Δ	Δ	∀	∀	∀	∀	∀	∀	∀	∀	∀	∀
Arm cross. dim.	Δ	Δ	Δ	Δ	∀	Δ	∀	∀	∀	∀	∀	○
Machine element	Δ	Δ	Δ	Δ	∀	Δ	∀	∀	∀	∀	∀	○

∀ = strong, ○ = medium, Δ = weak.

Source: Modified from Inoue, K., et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.

4. Weight capacity
5. Maximum deflection
6. Minimum natural frequency

The relationship between design parameters and evaluation functions is shown in Table 21.1. It shows that kinematics strongly depends on fundamental mechanism, while dynamics depends on inner mechanism.

Many robot CAD systems have been developed throughout the world.[6-18] Among them is TOCARD (total computer-aided robot design), which has the ability to design robots comprehensively and will be explained later.

21.2.2 Robot Design Procedure

Figure 21.13 shows the total robot design procedure in this CAD system. First, the operator (designer) inputs the design conditions which are prescribed by the objective tasks. Then, the procedure consists of three design systems — *fundamental mechanism design, inner mechanism design,* and *detailed structure design*, described as follows:

1. Fundamental mechanism design is based on kinematic evaluation — workspace, joint displacement, velocity and acceleration, and workpiece velocity and acceleration.
2. Inner mechanism design requires determination of motor allocations and the types of transmission mechanisms. The arm cross-sectional dimensions are calculated roughly, and the machine elements including the motors and the reduction gears are selected from their catalog data temporarily, based on rough evaluation of dynamics — motor driving torque, total motor power, total weight, weight capacity, and deflection.
3. Detailed structure design involves modification of the arm cross-sectional dimensions and reselection of the machine elements based on precise evaluation of dynamics — total weight, deflection, and natural frequency.

© 2002 by CRC Press LLC

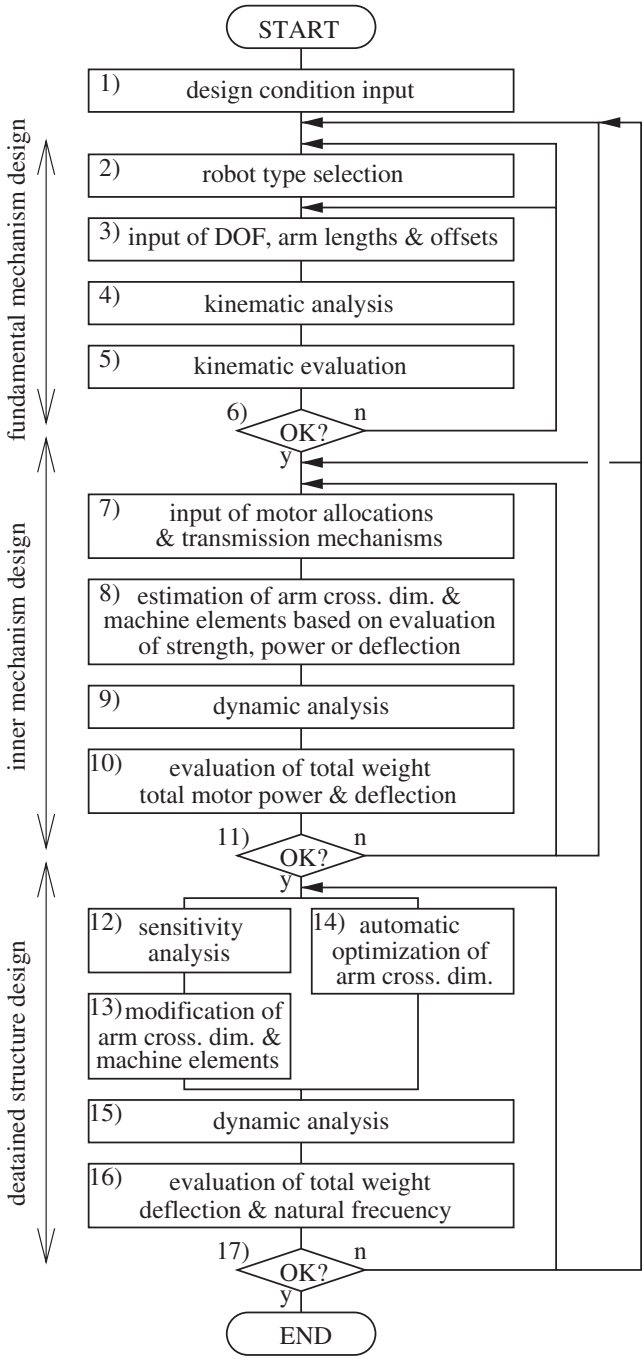


FIGURE 21.13 Total robot design procedure in TOCARD. (Modified from Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

Some of the design parameters are locally optimized in each system. However, if sufficient performance cannot be obtained in a system, the operator returns back to the previous system and tries the previous design again. The CAD system is an interactive design system; the operator can repeatedly alternate between design change and evaluation. The details of the above-mentioned design systems are described in the following sections.

© 2002 by CRC Press LLC

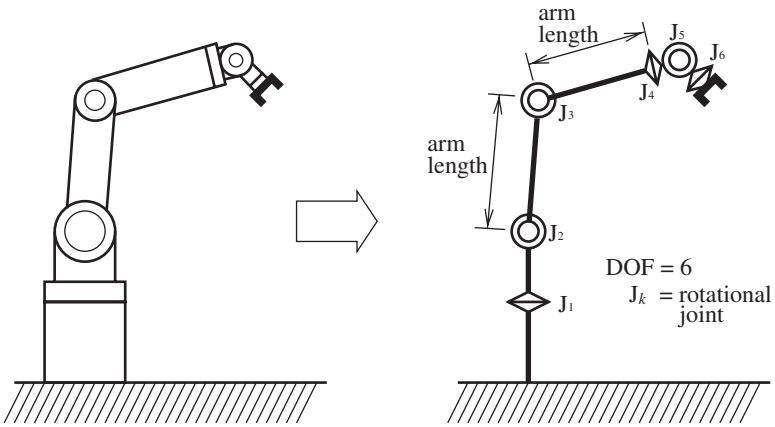


FIGURE 21.14 Fundamental mechanism of robot.

21.2.3 Design Condition Input

21.2.3.1 Step 1

The operator inputs the design conditions or constraints prescribed by the objective tasks:

1. Sizes and weights of workpieces (including end-effectors)
2. Reference trajectories of workpieces
3. Required working space
4. Allowable deflection and natural frequency

21.2.4 Fundamental Mechanism Design

The kinematic design parameters of a robot that is a serial link mechanism as shown in Figure 21.14, are called the "fundamental mechanism:"

1. Degrees of freedom
2. Joint types (rotational/sliding)
3. Arm lengths and offsets

The fundamental mechanism is determined based on kinematic evaluation.

21.2.4.1 Step 2

The type of robot mechanism — how rotational or sliding joints are serially arranged — is called robot type, and most industrial robots are classified into the following categories:

- Cartesian robot (or rectangular robot)
- Cylindrical robot
- Spherical robot (or polar robot)
- Articulated robot
- SCARA robot

Generally, a robot design expert selects a suitable robot type for the objective tasks from these categories, using empirical knowledge concerning the characteristics of the performance of each robot type. Here a new method is introduced for selecting the most suitable robot type for the tasks from the typical six-D.O.F. industrial robot types based on rough evaluation of the performances using fuzzy theory. In this method, the performances of robot types derived from the design expert's knowledge are roughly compared with the performances required for the tasks

© 2002 by CRC Press LLC

using fuzzy theory.[19] The method is outlined below, where the italics in the examples are expressed using fuzzy sets.

1. Five performances, workspace, dexterity, speed, accuracy, and weight capacity, are evaluated in robot type selection, in the same way as a robot design expert's method. These and the suitability for the objective tasks of a robot type are expressed by fuzzy sets.
   Example 21.1: Workspace is *large*.
   Example 21.2: Suitability for task is *very high* (*very suitable* for task).
2. The empirical knowledge of the design expert concerning performance of each robot type is expressed in the form of a fact. All such knowledge is stored in the system beforehand.
   Example 21.3: Workspace of articulated robot is *very large*.
3. The operator analyzes the tasks and obtains the performances required for them, which are expressed as a set of rules; these rules are input by the operator.
   Example 21.4: If workspace of robot type is *large*, it is *suitable* for painting task.
   Example 21.5: If workspace of robot type is *small*, it is *never suitable* for painting task.
4. The suitability of each robot type for the tasks is obtained from 2 and 3 above by fuzzy reasoning (Mamdani's method).
   Example 21.6: Articulated robot is *very suitable* for painting task.
5. After the suitabilities of all robot types are obtained, the operator selects the most suitable type.

21.2.4.2 Step 3

After the robot type is selected, the operator inputs and modifies arm lengths and offsets. He can also add new joints or can remove the joints that do not move when the robot moves along the reference trajectories given as the design condition, thus increasing/reducing degrees of freedom.

21.2.4.3 Step 4

Once the fundamental mechanism is determined using the above two steps, then kinematic analyses are applied to the designed mechanism.

Forward kinematics (Figure 21.15) — Forward kinematics calculates the workpiece position and orientation $\boldsymbol{R}$ from the joint displacement vector $\boldsymbol{q}$. Transformation matrix is often used for the forward kinematics of a serial link manipulator; this system uses the revised transformation matrix of the Denavit–Hartenberg method.[20]

Inverse kinematics (Figure 21.15) — An efficient algorithm of inverse kinematics problem calculating $\boldsymbol{q}$ from $\boldsymbol{R}$ was developed by Takano.[20] This algorithm is applicable to all types of a six-DOF robot with three rotational joints in the wrist and can obtain a maximum eight sets of solutions. Inverse kinematics is used for calculating the joint trajectories $\boldsymbol{q}[t]$ corresponding to the workpiece reference trajectories $\boldsymbol{R}[t]$ given as the design condition.

Workspace analysis (Figure 21.15) — Evaluating the workspace generated by three joints near the base is sufficient for the robot design. The method developed by Inoue can efficiently obtain the boundary surface of such workspace of any type of robot, considering the joint operating range.[21]

Velocity/acceleration analysis (Figure 21.16) — Luh's algorithm[22] includes the process calculating the workpiece velocity **v** and acceleration **a** from $\boldsymbol{q}$, $\dot{\boldsymbol{q}}$, and $\ddot{\boldsymbol{q}}$; it is used here.

21.2.4.4 Step 5

Kinematic performances of the designed fundamental mechanism are evaluated by:

- The workspace considering the joint operating range must cover the required working space for the objective tasks given as the design condition (Figure 21.15).
- The joint operating range is limited by the structure of the joint. While wide joint operating range makes workspace large, a long sliding joint makes the robot heavy, and using a

© 2002 by CRC Press LLC

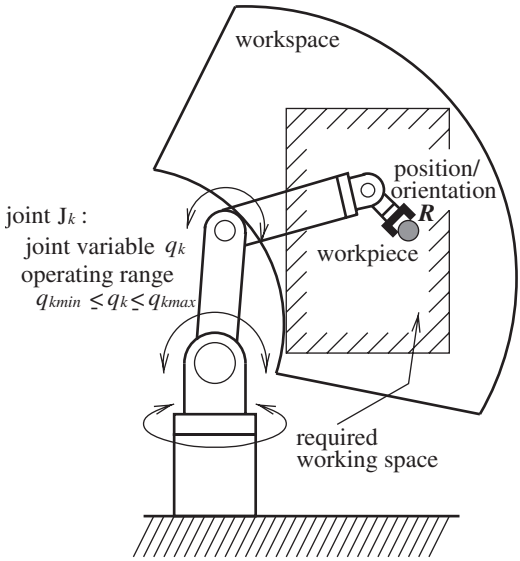


FIGURE 21.15 Position kinematics, workspace, and joint operating range.

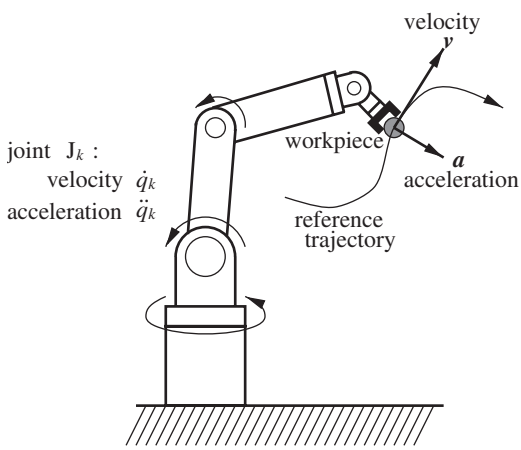


FIGURE 21.16 Workpiece velocity/acceleration and joint velocity/acceleration.

rotational joint with offset reduces the stiffness of the joint shaft to radial force. Thus, the joint operating range is evaluated as described above (Figure 21.15).

- The maximum workpiece velocity and acceleration required for the objective tasks are given indirectly as the design condition — the reference trajectories of workpieces (Figure 21.16).
- The maximum joint velocity and acceleration on the given trajectories should be as small as possible so that the robot can be moved by small and light motors (Figure 21.16).

21.2.4.5 Step 6

The operator repeats the design change and evaluation alternately in Steps 3 through 5. If the above interactive design fails, the operator goes back to Step 2 and selects another suitable robot type. This procedure is repeated until the suitable fundamental mechanism is obtained.

© 2002 by CRC Press LLC

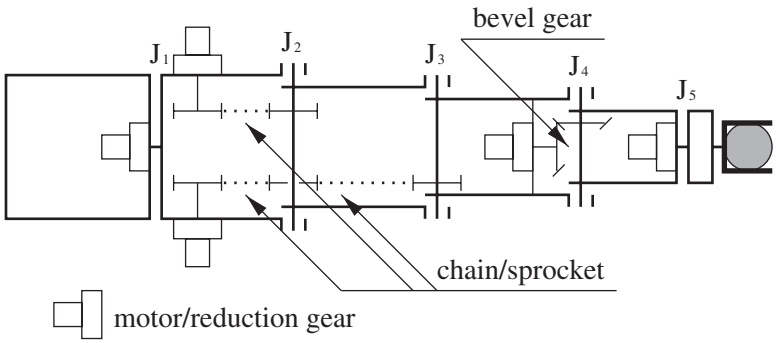


FIGURE 21.17 Joint driving systems of robot. (Modified from Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

21.2.5 Inner Mechanism Design

The following design parameters are called "inner mechanism:"

1. Motor allocations (where the motors are to be attached)
2. Types of transmission mechanisms
3. Motors
4. Reduction gears and their reduction ratio
5. Arm cross-sectional dimensions
6. Machine elements (bearings, chains, bevel gears, etc.)

In the inner mechanism design, (1) and (2) are determined, (5) is calculated roughly, and (3), (4), and (6) are selected from catalog data temporarily, based on rough evaluation of dynamics — motor driving torque, total motor power, total weight, weight capacity, and deflection.

21.2.5.1 Step 7

As shown in Figure 21.17, a joint driving system consists of an actuator, a reduction gear, and transmission mechanisms (if needed). Five types of driving elements used in this CAD system are

1. Motor/reduction gear element
2. Shaft element
3. Chain/sprocket element
4. Bevel gear element
5. Ball screw/nut element

We adopted motors and harmonic drives as actuators and reduction gears respectively, because these are used in many industrial robots in the present time. Direct drive motors can be modeled as motor/reduction gear elements without reduction gears. Ordinary belts and timing belts are dealt with as chain/sprocket elements, because they are the same as chains kinematically, and only have different stiffnesses and weights. In this step, the operator inputs motor allocations and types of transmission mechanisms, as illustrated in Figure 21.18.

21.2.5.2 Step 8

Five types of arm/joint elements are used here (Figure 21.19).

1. Cylindrical arm element
2. Prismatic arm element
3. Revolute joint element (type 1)
4. Revolute joint element (type 2)
5. Sliding joint element

© 2002 by CRC Press LLC

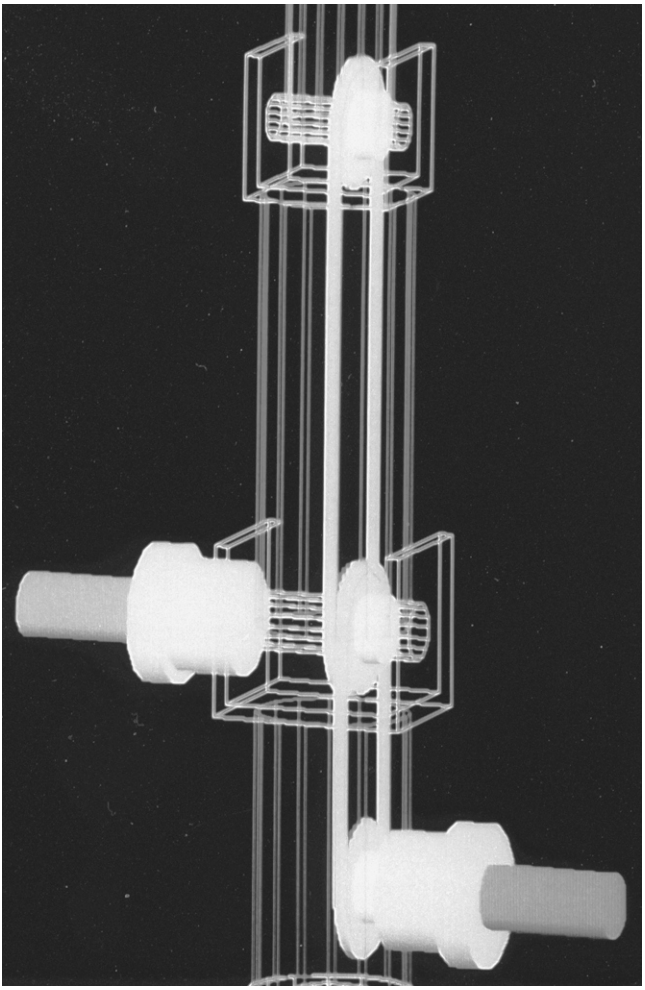
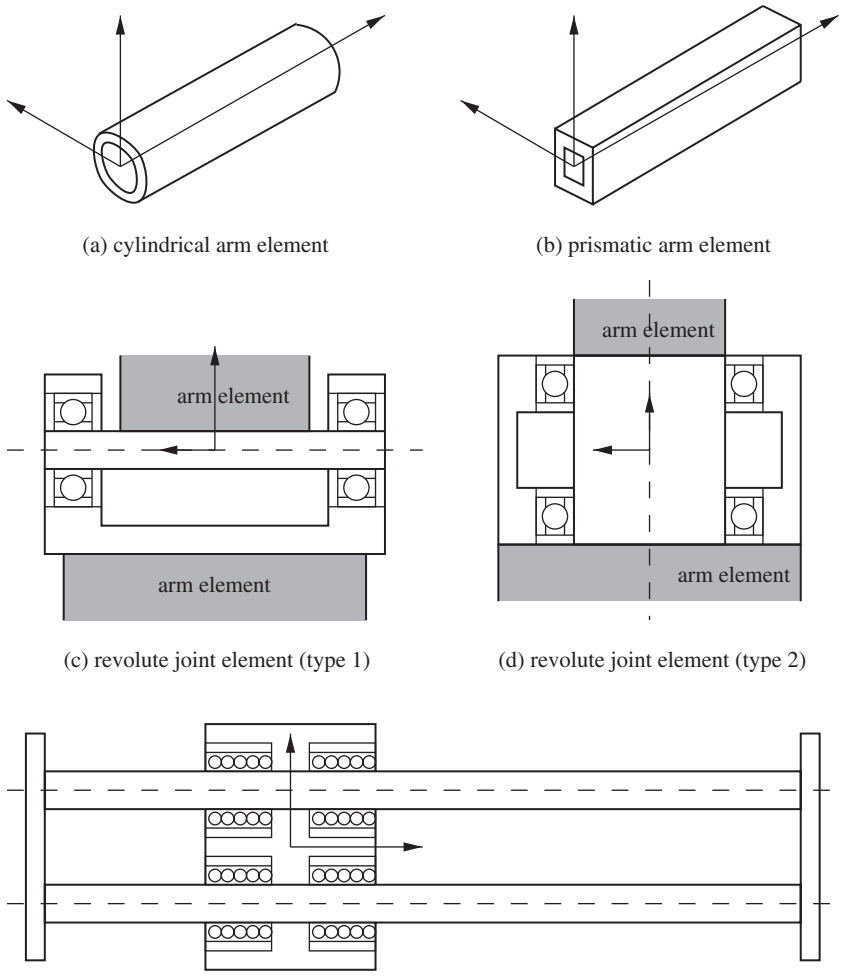


FIGURE 21.18 Example of designed transmission mechanisms. (From Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

Because the arm cross-sectional dimensions of arm elements, the bearings used in joint elements, and the machine elements used in driving elements are design parameters, the system roughly calculates arm cross-sectional dimensions and selects motors, reduction gears, and machine elements (bearings, chains, bevel gears, etc.) from catalog data temporarily. This is done so that each arm, joint, or driving element will have enough strength and stiffness against the internal force acting on it and each motor will have enough power and torque to move the robot. In Figure 21.20, B_i is the i-th arm element, and $\boldsymbol{f}_i$ is the force/moment acting on the lower arm element B_{i-1} from B_i. If the joint element J_i is rotational, the moment around the joint axis of $\boldsymbol{f}_i$ is the joint driving torque τ_i of J_i; if J_i is sliding, the force in the joint axis direction of $\boldsymbol{f}_i$ is the joint driving force, which is converted into τ_i with the ball screw/nut element.

1. The cross-sectional dimension of the arm element B_i is determined to minimize the weight of B_i under the constraint that its deflection to the maximum value of the force/moment $\boldsymbol{f}_{i+1}$ acting on B_i from the upper arm element B_{i+1} is less than its allowable deflection.
2. The weight, the position of center of gravity, and the inertia tensor of B_i are calculated from the determined dimension.

© 2002 by CRC Press LLC



(a) cylindrical arm element

(b) prismatic arm element

(c) revolute joint element (type 1)

(d) revolute joint element (type 2)

(e) sliding joint element

FIGURE 21.19 Arm/joint elements. (Modified from Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

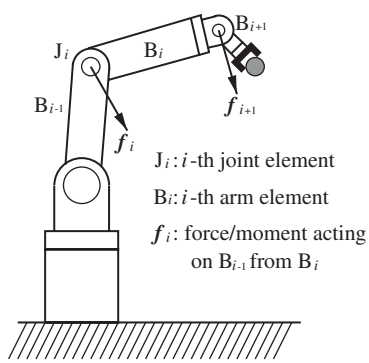


FIGURE 21.20 Automatic design of each element in inner mechanism design.

© 2002 by CRC Press LLC

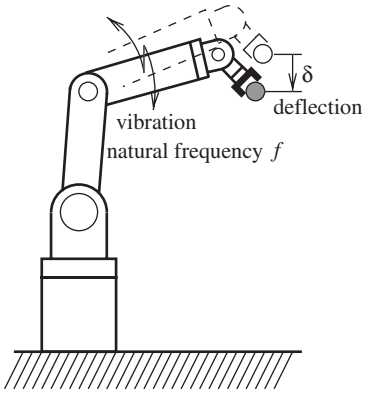


FIGURE 21.21 Deflection and natural frequency.

3. The force/moment $\boldsymbol{f}_i$ acting on the lower arm element B_{i-1} from B_i, thus the joint driving torque τ_i, can be obtained via the inverse dynamics of B_i along the trajectories given as the design condition.
4. The bearing used in the joint element J_i is selected from the catalog data so that the bearing may be lightest and have greater allowable radial/thrust load than the maximum value of $\boldsymbol{f}_i$.
5. Each machine element used in the transmission mechanism of J_i is selected from the catalog data so that it may be lightest and have greater allowable torque than the maximum value of τ_i.
6. The reduction ratio of the reduction gear for J_i is determined via mechanical impedance matching.*
7. The motor driving J_i is selected from the catalog data so that it may be lightest and have enough rated power and allowable torque to move J_i.
8. Repeating the above-mentioned procedure alternately from the tip arm element to the base arm element allows us to determine the design parameters of the elements temporarily.

21.2.5.3 Step 9

Determining all design parameters of the robot temporarily in this way permits the following dynamic analyses.

Inverse dynamics — We expanded Luh's algorithm so that it can be applied to robots with transmission mechanisms as shown in Figure 21.17; the revised method can calculate both the joint driving torque τ and the motor driving torque τ_m when the robot motion $\boldsymbol{q}$, $\dot{\boldsymbol{q}}$, and $\ddot{\boldsymbol{q}}$ are given. This method can also calculate the internal force/moment $\boldsymbol{f}_i$ acting on each arm element, which is used in design of each element as described above.

Deflection analysis (Figure 21.21) — Generally, the stiffness of bearings in joints, reduction gears, and transmission mechanisms is not negligible because it is less than the stiffness of arms. Thus we developed an elastic model of a robot by the finite element method (FEM), which is applicable to robots with transmission mechanisms and deals with the stiffness of arms as well as that of bearings, reduction gears, and transmission mechanisms.[23,24] Using this model, we calculate the deflection δ when the robot motions $\boldsymbol{q}$, $\dot{\boldsymbol{q}}$, and $\ddot{\boldsymbol{q}}$ are given.

21.2.5.4 Step 10

The operator evaluates the dynamic performances of the designed robot:

Maximum motor driving torque — The maximum motor driving torque on the trajectories given as the design condition should be as small as possible in order to use small and light motors.

*When the moment of inertia of motor and arm are I_m and I_a, reduction ratio $n = \sqrt{I_a / I_m}$ gives the maximum arm acceleration by the constant motor torque. It is called "mechanical impedance matching."

© 2002 by CRC Press LLC

Total motor power — Total motor power required for the robot to do the objective tasks should be as small as possible, leading to low cost.

Total weight — The total weight of the robot should also be as small as possible, to reduce cost. Lightening the robot enables high speed motion and makes the transportation and installation of the robot easy.

Weight capacity — The robot must have enough weight capacity to carry the workpieces dealt with in the objective tasks. This constraint is given as a design condition — the weight of workpieces.

Maximum deflection — Deflection strongly depends on the balance of the stiffness and weight of the robot, and the deflection affects the accuracy. The maximum deflection on the given trajectories should be nearly equal to the allowable deflection given as the design condition. The constraint on the deflection will be evaluated again in the detailed structure design.

21.2.5.5 Step 11

If sufficient performance cannot be obtained in Step 10, the operator returns to Step 7 and changes the motor allocations and the types of transmission mechanisms. Steps 7–10 are repeated until the suitable motor allocations and the suitable types of transmission mechanisms for the objective tasks are obtained.

21.2.6 Detailed Structure Design

The arm cross-sectional dimensions and/or machine elements, which have already been locally optimized on each element in the inner mechanism design, are modified/reselected to minimize the total weight under the constraints that the deflection is lower than the allowable deflection and that the natural frequency is higher than the allowable frequency:

$$\begin{cases} \text{objective function} & = & m \rightarrow \min \\ \text{constraints} & = & \delta \le \delta_0, \quad f \ge f_0 \end{cases} \tag{21.25}$$

where m = total weight, δ = maximum deflection of robot at the point of grasped workpiece, f = natural frequency of the first vibration mode of robot, δ_0 = allowable deflection given as design condition, and f_0 = allowable natural frequency given as design condition.

The above-mentioned global optimization is important to attain both high speed motion and high precision positioning. This optimization problem can be rewritten into the objective function Q to be minimized:

$$Q \equiv m + \frac{w_d}{\delta_0^2}\left\{\max(0,\ \delta - \delta_0)\right\}^2 + \frac{w_f}{f_0^2}\{\max(0, f_0 - f)\}^2 \rightarrow \min, \tag{21.26}$$

where $w_d(>0)$ = penalty for deflection constraint and $w_f(>0)$ = penalty for natural frequency constraint.

δ_0, f_0, w_d, and w_f are input by the operator as the design conditions. Steps 12, 13, and 14 help the operator determine the arm cross-sectional dimensions and the machine elements.

21.2.6.1 Steps 12 and 13

Letting the parameter p_i be one of the arm cross-sectional dimensions and machine elements, the sensitivity S_i of p_i is defined by:

$$s_i \equiv \delta Q / \delta p_i \tag{21.27}$$

Obviously, the most sensitive parameter (the parameter with negative, and the smallest, sensitivity) is most effective for minimizing Q. In Step 12, the system calculates all sensitivities of the

© 2002 by CRC Press LLC

TABLE 21.2 Workpiece Allowable Deflection, and Allowable Natural Frequency Given as Design Conditions

Length of workpiece		100.0 mm
Weight of workpiece		2.0 kg
Allowable deflection	δ_0	0.5 mm
Allowable natural frequency	f_0	25.0 Hz
Penalty for deflection	w_d	10000.0 kg
Penalty for frequency	w_f	10000.0 kg

Source: From Inoue, K., et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.

arm cross-sectional dimensions and the machine elements. Then, in Step 13, the operator selects and modifies the most sensitive parameter.

21.2.6.2 Step 14

The system calculates the optimal arm cross-sectional dimensions which minimize Q under the constraints on their allowable ranges by the gradient projection method.

21.2.6.3 Step 15

In addition to deflection analysis, we perform natural frequency analysis (Figure 21.21). The natural frequency f for the given joint displacement $\boldsymbol{q}$ (the given pose of robot) is calculated by FEM, using the elastic model used in deflection analysis.

21.2.6.4 Step 16

The operator evaluates the total weight, maximum deflection, and minimum natural frequency according to Equation (21.25).

- The total weight should be minimized.
- The maximum deflection on the given trajectories must be less than the allowable deflection.
- Natural frequency also depends on the balance of the stiffness and weight of the robot, and low natural frequency causes low accuracy because of residual vibration after positioning. The minimum natural frequency on the given trajectories must be higher than the allowable natural frequency.

21.2.6.5 Step 17

The above-mentioned design change and evaluation, Steps 12 through 16, are repeated alternately until the optimal arm cross-sectional dimensions and machine elements are obtained; then the robot design is terminated.

21.2.7 Design Example

The design conditions are summarized in Table 21.2 and Figures 21.22 and 21.23. Figure 21.24 shows the robot designed by the above-mentioned procedure, and Tables 21.3 through 21.5, summarize its main design parameters. Furthermore, Figures 21.22, 21.23, 21.25 and Table 21.6 show the various performances of the designed robot. Figure 21.26 illustrates the change in the total weight, deflection, and natural frequency with repetition of design change in the detailed structure design. As shown here, we could have reduced the total weight with a few repetitions, keeping the deflection and the natural frequency within the allowable ranges given as the design conditions. The robot finally obtained is about 1.1 kg lighter than that obtained temporarily with the inner mechanism design.

© 2002 by CRC Press LLC

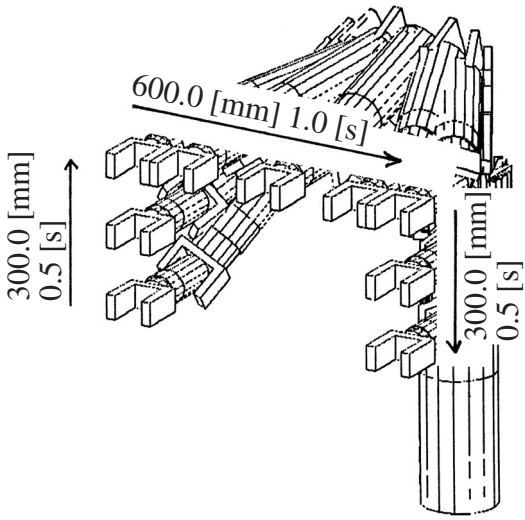


FIGURE 21.22 Trajectory given as design condition and animation of designed robot moving along it. (From Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

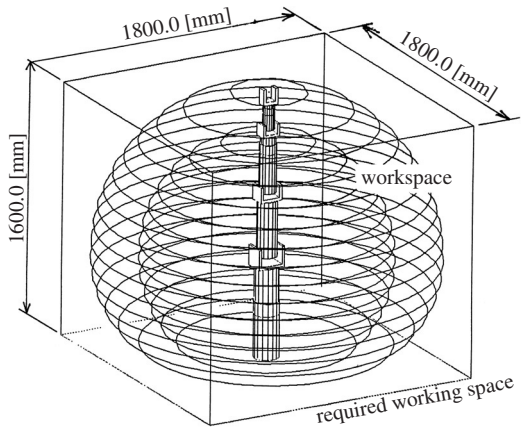


FIGURE 21.23 Required working space given as design condition and workspace of designed robot. (From Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

© 2002 by CRC Press LLC

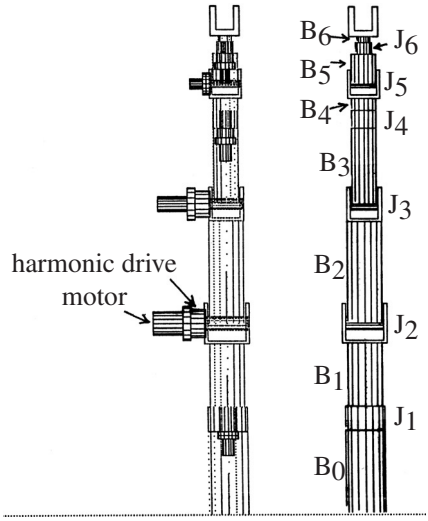


FIGURE 21.24 Designed robot using TOCARD. (From Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

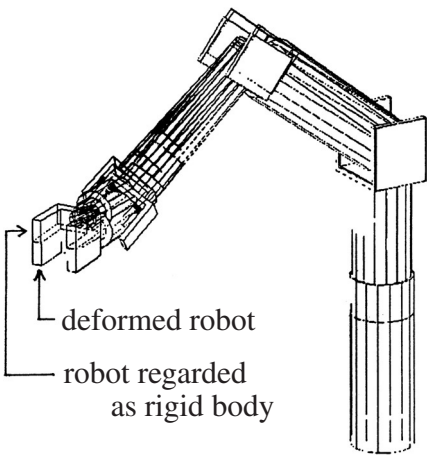


FIGURE 21.25 Maximum deflection (0.468 mm) on given trajectory of designed robot. (From Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

© 2002 by CRC Press LLC

TABLE 21.3 Arm Parameters of Designed Robot

Arm	Length (mm)	Diameter (mm)	Thickness (mm)
B_0	320.0	130.0	9.0
B_1	320.0	120.0	10.0
B_2	400.0	120.0	2.0
B_3	280.0	84.0	2.0
B_4	120.0	84.0	2.0
B_5	120.0	84.0	2.0
B_6	40.0	44.5	2.0
Total	1600.0		

Cylindrical arm element (Duralumin).

Source: From Inoue, K., et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.

TABLE 21.4 Joint Parameters of Designed Robot

Joint	Type	Distance between Bearings (mm)	Bearing Internal Diameter (mm)
J_1	(b)	80.0	6812 (60.0)
J_2	(a)	140.0	6809 (45.0)
J_3	(a)	106.0	6807 (35.0)
J_4	(b)	60.0	6808 (40.0)
J_5	(a)	100.0	6804 (20.0)
J_6	(b)	40.0	6805 (25.0)

(a) = Revolute (type 1) joint element; (b) = revolute (type 2) joint element.

Source: From Inoue, K., et al., *J. Robotics Soc. Jpn*, 14, 710, 1996. With permission.

TABLE 21.5 Joint Driving System Parameters of Designed Robot

Motor Joint	Rated Power (W)	Harmonic Drive (Reduction Ratio)
J_1	L402-011 (23.0)	FRS-20-160 (160)
J_2	L720-012 (200.0)	CSS-32-160 (160)
J_3	L511-012 (110.0)	CSS-32-160 (160)
J_4	L402-011 (23.0)	CSS-20-80 (80)
J_5	R301-011 (11.0)	FRS-20-160 (160)
J_6	R301-011 (11.0)	CSS-20-50 (50)

The motor and harmonic drive are directly attached to their joint axis.

Source: From Inoue, K., et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.

© 2002 by CRC Press LLC

TABLE 21.6 Performance Parameters of Designed Robot

Parameter	Performance
Total motor power	378.0 W
Total weight	24.0 kg
Maximum deflection on given trajectory	0.468 mm
Minimum natural frequency on given trajectory	25.0 Hz

Source: Modified from Inoue, K., et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.

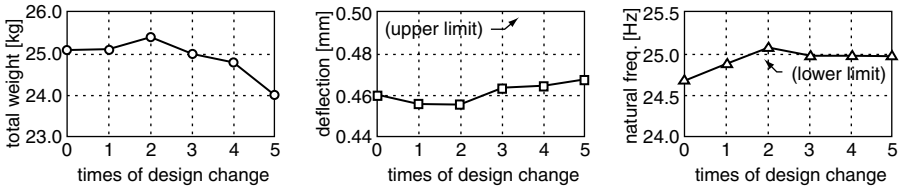


FIGURE 21.26 Change in total weight, deflection, and natural frequency with iteration of design change in detailed structure design. (From Inoue, K. et al., *J. Robotics Soc. Jpn.*, 14, 710, 1996. With permission.)

References

1. Vukobratović, M., and Potkonjak, V., *Dynamics of Manipulation Robots: Theory and Application*, *Sci. Fundam. Robotics,* Springer Verlag, Berlin, 1982.
2. Vukobratović, M. and Stokić, D., *Control of Manipulation Robots, Theory and Application*, Springer Verlag, Berlin, 1982.
3. Nof, S.Y. (Ed.), *Handbook of Industrial Robotics,* John Wiley & Sons, New York, 1985.
4. Merit, H.E., *Hydraulic Control Systems,* John Wiley & Sons, New York, 1967.
5. Vukobratović, M., *Applied Dynamics of Manipulation Robots,* Springer Verlag, Berlin, 1989.
6. Chedmail, P., Robot structures and actuators optimization, *Proc. 16th Int. Symp. Ind. Robots*, 185, 1986.
7. Erdman, A.G., Thompson, T., and Riley, D.R., Type selection of robot and gripper kinematic topology using expert system, *Int. J. Robotics Res.*, 5(2), 183, 1986.
8. Fenton, R.G. and Lipitkas, J., Optimum design of manipulators, *Comput.-Aided Eng. Robotics*, 68, 1984.
9. Gosselin, C.M. and Guillot, M., The synthesis of manipulators with prescribed workspace, *ASME J. Mech. Design*, 113, 451, 1991.
10. Kim, J. and Khosla, P.K., A formulation for task based design of robot manipulators, *Proc. 1993 IEEE/RSJ Int. Conf. Intelligent Robots Syst.*, 2310, 1993.
11. Lin, C.D., and Freudenstein, F., Optimization of the workspace of a three-link turning-pair connected robot arm, *Int. J. Robotics Res.*, 5(2), 104, 1986.
12. Paden, B., Optimal kinematic design of 6R manipulators, *Int. J. Robotics Res.*, 7(2), 43, 1988.
13. Paredis, C.J.J. and Khosla, P.K., Kinematic design of serial link manipulators from task specifications, *Int. J. Robotics Res.*, 12(3), 274, 1993.
14. Potkonjak, V., Vukobratović, M., and Hristic, D., Interactive procedure for computer-aided design of industrial robots mechanisms, *Proc. 13th Int. Symp. Ind. Robots*, 16, 1983.
15. Potkonjak, V. and Vukobratović, M., Computer-aided design of manipulation robots via multiparameter optimization, *Mechanism and Machine Theory*, 18(6), 431, 1983.

© 2002 by CRC Press LLC

16. Vukobratović, M., and Potkonjak, V., Applied Dynamics and CAD of Manipulation Robots, *Sci. Fundam. Robotics*, Springer Verlag, Berlin, chap. 6, 1985.
17. Vijaykumar, R., Waldron, K.J., and Tsai, M.J., Geometric optimization of serial chain manipulator structures for working volume and dexterity, *Int. J. Robotics Res.*, 5(2), 91, 1986.
18. Yang, D.C.H. and Lee, T.W., Heuristic combinatorial optimization in the design of manipulator workspace, *IEEE Trans. SMC*, 14(4), 571, 1984.
19. Inoue, K., Takano, M., and Sasaki, K., Type selection of robot manipulators using fuzzy reasoning in robot design system, *Proc. 1993 IEEE/RSJ Int. Conf. Intelligent Robots Sys.*, 926, 1993.
20. Takano, M., A new effective solution for inverse kinematics problem (synthesis) of a robot with any type of configuration, *J. Fac. Eng. Univ. Tokyo*(*B*), 38(2), 107, 1985.
21. Inoue, K., Takano, M., and Sasaki, K., Development of robot CAD system (5th report) calculation method of workspace using inverse kinematics, *Proc. 1989 JSPE* (*Autumn*), 279, 1989, (in Japanese).
22. Luh, J.Y.S., Walker, M.W., and Paul, R.P.C., On-line computational scheme for mechanical manipulators, *ASME J. Dynamic Syst., Meas., Control*, 102, 69, 1980.
23. Inoue, K., Takano, M., and Sasaki, K., Optimum design of robot manipulators using sensitivity analysis, *Proc. 20th Int. Symp. Ind. Robots*, 513, 1989.
24. Inoue, K., Takano, M., and Sasaki, K., Optimal design system of robot structures using sensitivity to dynamical characteristics, *J. Robotics Soc. Jpn.*, 9(1), 18, 1991 (in Japanese).

© 2002 by CRC Press LLC

22

Control of Robots

22.1 Introduction
22.2 Hierarchical Control of Robots
Mission Layer • Task Layer • Action Layer
22.3 Control of a Single Joint of the Robot
Model of Actuator and Joint Dynamics • Synthesis of Servosystem • Influence of Variable Moments of Inertia • Influence of Gravity Moment and Friction • Synthesis of the Servosystem for Trajectory Tracking
22.4 Control of Simultaneous Motion of Several Robot Joints
Analysis of the Influence of Dynamic Forces • Dynamic Control of Robots • Inverse Problem Technique • Effects of Payload Variation and the Notion of Adaptive Control

Miomir Vukobratović
Mihajlo Pupin Institute

Dragan Stokić
ATB Institute

22.1 Introduction

This chapter is dedicated to the synthesis of basic control of manipulation robots. Because the successful application of robots in industry and other domains often depends to a great extent upon the efficiency, reliability, and capabilities of a control system, it is obvious that the synthesis of adequate control systems is of the highest importance for further application of robotics in industrial practice.

Control systems of robots can be realized in different ways. Historically speaking, different open-loop control systems were applied to control the first manipulation robots. However, current robots include digital (microprocessor)-based control systems that enable flexible specification of the tasks, adaptation to environment changes, etc. A robot's joints are controlled by servo-systems (or servos) based on the feedback loops providing information on positions, velocity, and accelerations of the joints.

In this chapter we mainly focus upon the synthesis of servos for robots. In order to enable efficient specification of the tasks to be fulfilled by the robot, modern control systems include options to specify directly desired the position of the gripper (hand). To accomplish various tasks, the hand of the robot (or the payload, or the tool) has to be placed in the desired locations at the workplace and take the desired orientation (and, sometimes, to produce certain desired forces upon the other objects in the workspace). If an operator, when specifying the task for the robot to accomplish, intends to place the hand of the robot in a desired position by specifying the positions of the joints, he or she would have to determine the corresponding positions of the joints in an iterative way. For some robot structures this may be easy task (e.g., a robot with three linear joints, or a robot with a cylindrical structure, etc.). However, for the majority of robot structures, this can be a very tedious and time-consuming job. Therefore, it is necessary to enable the user to directly specify the desired positions of the robot hand, either by programming the robot, or by a teaching-box, or by some other means. In this case, the operator of the robot has to specify the desired

© 2002 by CRC Press LLC

position and orientation of the hand, and the control system has to compute automatically the corresponding positions of the joints. This means that the control system has to compute internal (joint) coordinates of the robot based on the desired position/orientation of the robot hand, i.e., based on the specified so-called external (or Cartesian) coordinates. This calculation can be performed by the control computer in various ways. Most modern robots are equipped with control systems that enable direct specification of the Cartesian coordinates.

Modern industry and other application domains are assigning more and more complex tasks to robots. Apart from the simplest task (such as pick-and-place, which can be reduced to a free motion of the robot and payload from one position to another), modern robots have to ensure movements along prespecified paths in the workspace (for example, arc-welding by robots, gluing by robots, moving a robot in a workspace with many obstacles, etc.). In these tasks, the operator has to specify the desired path of the robot's hand and the control system has to calculate the corresponding trajectories of the robot's joints and ensure their execution (i.e., the robot's joints are tracking these trajectories which, in turn, should ensure that the hand is following the desired path in the workspace).

Often robot tasks can be complex and the operator may need a very long time to specify the positions through which the robot has to move, or the paths along which the robot hand has to move to perform the desired tasks. For example, if the robot has to move very close to various machines and equipment in its workspace (i.e., if the robot has to move close to various obstacles), the operator has to plan all the intermediate positions through which the robot has to pass, or to plan paths along which the robot's hand has to move to avoid collisions of the hand or any of the robot's links with the obstacles. Obviously, such trajectory planning task can be very difficult, which is why it is desirable to have a control system capable of solving such problems automatically, and by this the operator is no longer responsible for path planning tasks. A number of modern robots include such control systems with automatic path planning. The user has to specify the task in relatively high abstract form (e.g., replace an object from one position to another), and the control system then automatically plans all movements of the robot (approaching the object, orientation of the hand, grasping the object, lifting the object, replacement of the object to another location with obstacle avoidance, putting the object into another location etc.). This automatic planning of the robot's paths and tasks represents the main prerequisites to introducing robots to flexible manufacturing systems. Obviously, it is also a prerequisite for further spreading of robots in various nonindustrial applications (e.g., service domains, space applications, etc.). Therefore, control systems of the current and future generations of robots required such capabilities.

22.2 Hierarchical Control of Robots

Control systems which can accommodate the requirements explained above, are obviously very complex. To simplify the synthesis and implementation of the control system, it has to be carefully structured. The usual approach to structure control system is to apply hierarchical architecture in which the robot's control system is organized in several levels, with each control level solving its specific task. One such (simple) hierarchical structure is presented in Figure 22.1. In this structure the control system includes three levels:[1]

1. The strategic control level has to plan the robot's paths. This level receives its tasks from the operator who is communicating with the control system by a programming language (normally each control system has special language enabling easy programming of the robot task). The strategic control level has to plan each motion of the robot. The operator specifies the tasks to be accomplished by the robot, and the strategic control level defines those paths of the robot's hand which have to be realized. If the workspace of the robot is predefined (i.e., if all obstacles are prespecified), the strategic control level can plan the paths in the space without additional information from external systems (e.g., sensors). However, if

© 2002 by CRC Press LLC

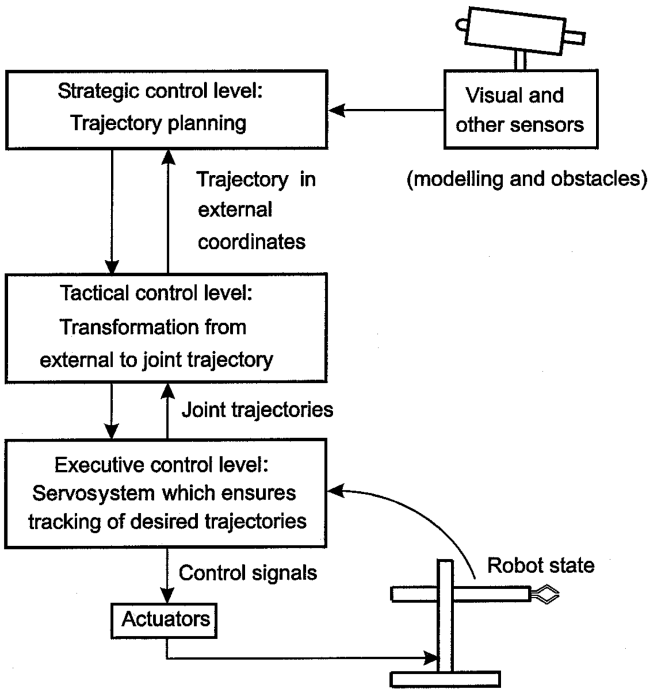


FIGURE 22.1 Simple hierarchical structure of a robot's control system.

locations of different obstacles are not (accurately) defined in advance or may change during the operation of the robot (e.g., movements of the parts not defined in advance), path planning must be performed based upon the sensor information (e.g., cameras, proximity sensor, etc. that provide information on the actual, current positions and shapes of different obstacles). In this case, the strategic level often must solve path-planning problems in real time, i.e., during process execution, which is a much more complex problem than if it can be done off-line (before task execution). In both cases, the strategic control level generates the trajectories of the robot's hand, i.e., it defines trajectories of the external coordinates of the robot.

2. The tactical control level has to map the trajectories from the external into internal (joint) coordinates of the robot. That is, the strategic control level provides the trajectories of the robot's hand coordinates and the tactical control level has to compute the corresponding trajectories of the robot's joints which have to be realized to execute the imposed hand trajectories. This problem is solved using the so-called "inverse kinematic model of the robot." Output of the tactical control level is joint trajectories. This control level can operate in either an off-line or on-line mode, depending on the conditions imposed in the specific tasks.
3. The executive control level has to realize the trajectories (or positions) of the robot's joints which are imposed by the higher, tactical control level. This control level must ensure realization of the imposed trajectories on the basis of information on the actual robot state (positions, speeds, and accelerations of the joints). By ensuring the tracking of the imposed joint trajectories, the trajectories of the robot hand are also accomplished, and the task imposed by the operator is accomplished.

It should be noted that some control systems do not include all three control levels; however, all control levels must include an executive control level to realize desired positions or trajectories of the robot's joints. As explained above, modern robots incorporate specifications of hand coordinates, which

© 2002 by CRC Press LLC

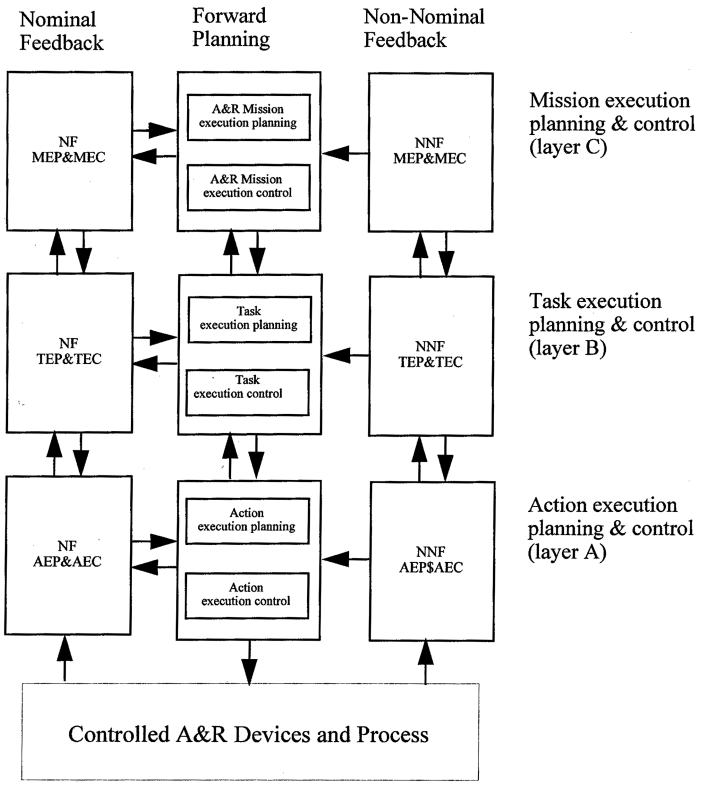


FIGURE 22.2 A&R (automation and robotics) FRM structure. (From Dornier GmbH, 1992.[4] With permission.)

means that they include a tactical control level. However, a number of modern robots still do not include a strategic control level, which means they are not capable of automatically planning hand trajectories. For such robots the operator has to impose the desired trajectories (or positions) of the robot hand, and plan the paths using the robot's programming languages and teaching boxes, etc. Some robots have the strategic control level in a very rudiment form. Given the tasks demanded of modern robots, robots in the near future must include more sophisticated and complex strategic control levels.

The presented control structure is relatively simple. In order to cover the various complex tasks requested by different applications, the control systems must have a much more complex architecture. Different control architectures have been developed for industrial, space, and service applications. Attempts were made to defined general standard structures.[2-4] For example, with the introduction of automation and robot (A&R) technology in space applications, the European Space Agency ESA has identified the need for generic approaches in the development of such systems and has defined the so-called functional reference model (FRM) which provides a unified representation of essential robot control functions.[4] This reference model offers an essential functional and information architecture (a logical model) of general robot control systems that is independent of particular applications, operational scenarios, and implementations.

FRM is presented in Figure 22.2. It includes three main levels (or layers) and three main paths: the forward path where control actions are planned and executed, the nominal feedback path (NF) which establishes the feedback loop from the sensors to correct planned actions based on the current state of the robot and its environment under nominal conditions, and the non-nominal feedback path (NNF) which ensures an appropriate reaction of the robot in non-nominal situations, i.e., when some exceptional, accidental, and unforeseen situations appear (e.g., an actuator or a sensor failure, etc.)

© 2002 by CRC Press LLC

22.2.1 Mission Layer

This layer covers the overall system planning functions. Its functions perform time-lining (planning), scheduling, and dispatching tasks to A&R devices.

The inputs for this layer are the missions from an external system. Missions are process-oriented instructions which do not explicitly specify the way the mission should be accomplished by the robot (e.g., "assemble the parts" etc.).

The output is a set of tasks which specifies how the input mission will be executed by an A&R device such as a manipulation robot or a mobile vehicle.

22.2.2 Task Layer

The basic function of the task layer is to transform the process-oriented input tasks into device-oriented actions of robot. To do this, the most important function for a controller is the planning of trajectories to reach the given locations. The task layer performs path-planning activities (and adjusts these on the basis of sensor readings) to produce executable path segments specified by mathematical curves (straight lines, polynomials, clothoids, etc.).

Typically, this layer contains the following modules: path planning, path control, object recognition, subactivity planning (scheduling), and subactivity control (dispatching).

Input is a set of tasks that describes locations to reach and activities to perform in each location. Output is the path to be followed by the robot and activities to be performed in different parts of the path.

22.2.3 Action Layer

The action layer serves to transform the device-oriented action instructions from the task layer into control commands to the actuator and sensor hardware. The transformation requires the transition from Cartesian space in which the input is specified, into the configuration space of the actuators and sensors. Controlling the robot at this level also includes reactions to obstacles and avoidance of collisions. At this layer this is performed locally, which means with respect to the configuration space of actuators and sensors.

Typical modules for this layer are trajectory interpolation, trajectory control, actuator path interpolation, actuator control (servo control), local position estimation, obstacle detection and avoidance (local), detection of failure to reach goal, elementary activity planning, and elementary activity control.

Input is the path to be followed by the robot's hand, and the actions to be performed in different parts of the path. Outputs are control output signals to the actuator and sensor hardware.

In this chapter we focus on the problems related to the synthesis of the executive control level (Figure 22.1), i.e., the actuator control module in FRM. This means that we consider control of the actuators that drive the joints of the robot to maintain positions and trajectories imposed either by a higher (tactical) control level, or directly by the operator. In doing this we observe both problems: if the robot moves point-to-point (from one position to another), and if it has to move along desired continuous trajectories. It should be mentioned that the synthesis of the executive control level considered is relevant for all generations of robots and for both remote and manual robot control. We present some of the simplest approaches for robot control synthesis, those most often applied in practice. More sophisticated methods may be found in the corresponding literature.

22.3 Control of a Single Joint of the Robot

First we consider a simple case when a single joint of the robot is moving while all other joints are fixed. Let us assume the i-th joint of the robot has to be moved. The joints of the robot are driven by the actuators, and therefore, we consider the synthesis of control of an actuator driving the i-th joint while all other joints are fixed.

© 2002 by CRC Press LLC

22.3.1 Model of Actuator and Joint Dynamics

The actuators driving the joints may be D.C. or A.C. electromotors, hydraulic or pneumatic actuators. Because a large number of robots are driven by D.C. electromotors, we consider synthesis of the control for such actuators. However, these considerations can be easily extended to other types of actuators.[5]

The model of the dynamics of a D.C. electromotor, with a permanent magnet driving the i-th joint can be written in the following form.[5,6] The equation of moments equilibrium around the motor axis can be written as:

$$N_M^i N_V^i J_M^i \ddot{q}_i + P_i = N_M^i C_M^i i_r^i - B_C^i \dot{q}_i \tag{22.1}$$

where J_M^i is the moment of inertia of the rotor of the motor, q_i is the angle of rotation of the motor (joint) P_i is the load acting around the motor axis, C_M^i is the mechanical constant of the motor (the coefficient of proportionality between the moments developed by the motor and the current of the rotor coil), i_M^i is the current in the rotor coil, B_C^i is the coefficient of the viscous friction of the motor N_M^i is the moment reduction ratio at the motor axis (the ratio between the moment behind and in front of the gear), N_V^i is the speed reduction ratio of the gear (the ratio between the speed of the input and output shafts of the gear). The equation describing the equilibrium in the electric circuit of the rotor coil can be written in the form (assuming that the inductivity of the coil can be ignored):

$$R_r^i i_r^i + C_e^i N_V^i \dot{q}_i = u_i \tag{22.2}$$

where u_i is the input voltage on the rotor circuit, C_e^i is the coefficient of proportionality between the contra-electromotor force of the rotor and the rotation speed of the motor (this force is the voltage developing due to rotation of the rotor coil in the magnetic field), and R_r^i is the rotor coil resistance. Based on Equations (22.1) and (22.2) we can write:

$$J_M^i N_M^i N_V^i \ddot{q}_i + P_i = C_M^{i'} u_i - (B_C^i + C_E^{i'}) \dot{q}_i \tag{22.3}$$

where the following notations are introduced:

$$C_M^{i'} = N_M^i C_M^i / R_r^i \text{ and } C_E^{i'} = N_V^i N_M^i C_M^i C_e^i / R_r^i$$

In order to write the model of actuators in the state space, let us introduce the state vector in the form:

$$x_i = (q_i, \dot{q}_i)^T \tag{22.4}$$

Now, instead of Equation (22.3) we may write:

$$\dot{x}_i = A_i x_i + b_i u_i + f_i P_i \tag{22.5}$$

where A_i is the matrix of dimensions 2×2, and f_i are vectors of dimensions 2×1 given by:

$$A_i = \begin{bmatrix} 0 & 1 \\ 0 & -(B_C^i + C_E^{i'}) / (J_M^i N_V^i N_M^i) \end{bmatrix}, \quad b_i = \begin{bmatrix} 0 \\ C_M^{i'} / (J_M^i N_V^i N_M^i) \end{bmatrix}, \quad f_i = \begin{bmatrix} 0 \\ -1 / (J_M^i N_V^i N_M^i) \end{bmatrix} \tag{22.6}$$

© 2002 by CRC Press LLC

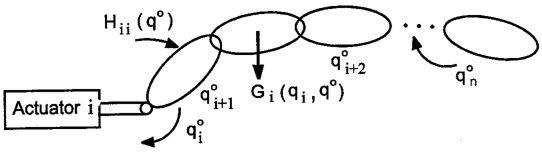


FIGURE 22.3 Actuator in the i-th joint of the robot (the remaining joints are fixed).

The actuator is driving the i-th joint while all other joints are in some fixed positions $q_j = q_j^0$, j = 1, 2, …, n, j ≠ i. The i-th actuator is driving the mechanical part of the robot (kinematic chain) around the i-th joint. In the given fixed positions of the joints q_j^0, j > i, the mechanical part of the robot has a constant moment of inertia around the i-th joint $H_{ii}(q_j^0)$ (see Figure 22.3). The actuator practically drives the set of links which together has a moment of inertia around the i-th joint $H_{ii}(q_j^0)$. These links also produce gravitational moment around the axis of the i-th joint.

This moment depends on the positions in which the joints are fixed and the current (variable) angle (linear displacement) of the i-th joint, i.e., $G_i(q_j^0, q_i)$. Thus, the moment produced by the mechanism around the i-th axis (i.e., around the shaft of the i-th motor) might be written as:

$$P_i = H_{ii}(q_j^0)\,\ddot{q}_i + G_i(q_j^0, q_i) \tag{22.7}$$

If we introduce the dynamic model of the rotation of the mechanism around the i-th axis in the model of the actuator (22.5), we obtain the model of the actuator's dynamics and the mechanism driven by the actuator in the following form (for simplicity, we shall write $H_{ii} = H_{ii}(q_j^0)$ and $G_i = G_i(q_j^0, q_i)$):

$$\dot{x}_i = \overline{A}_i\, x_i + \overline{b}_i\, u_i + \overline{f}_i\, P_i \tag{22.8}$$

where

$$\overline{A}_i = \begin{bmatrix} 0 & 1 \\ 0 & -(B_C^i + C_E^{i'})/(J_M^i N_V^i N_M^i + H_{ii}) \end{bmatrix}, \quad \overline{b}_i = \begin{bmatrix} 0 \\ C_M^{i'}/(J_M^i N_V^i N_M^i + H_{ii}) \end{bmatrix},$$

$$\overline{f}_i = \begin{bmatrix} 0 \\ -1/(J_M^i N_V^i N_M^i + H_{ii}) \end{bmatrix}$$

The model (22.8) represents the object of control (the actuator and the mechanism, the mechanical part of the robot which has to be controlled).

22.3.2 Synthesis of Servosystem

The task is to synthesize such a control law of the actuator and the joint (robot mechanical part) which should ensure that once the position of the joint is set at desired value q_i^0 the joint will be driven to this position in an adequate way. The control law accomplishing this task is usually a servosystem (servo), the scheme of which is presented in Figure 22.4.

A servo for control of the i-th actuator and joint consists of the following (basic) elements: a position sensor which provides information on the current (actual) position of the i-th joint and of the shaft of the actuator q_i (usually a potentiometer, or opto-encoder, etc.); a rotational (or displacement) velocity sensor of the joint and the motor $\dot{q}_i$ (usually tachogenerators are used, or numerical differentiation of the position/angle is applied); a differentiator which provides the difference

© 2002 by CRC Press LLC

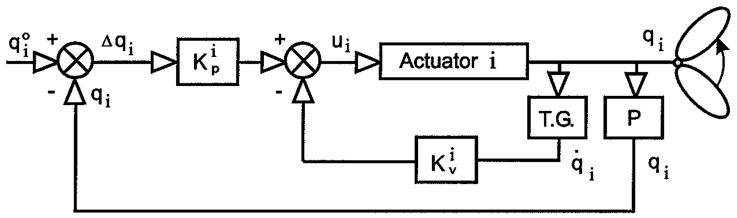


FIGURE 22.4 Positional servosystem (servo).

between the set (desired) position of the i-th joint, q_i^0 and actual position, q_i, obtained from the position sensor; an amplifier of the position error which amplifies the position error signal $\Delta q_i = q_i^0 - q_i$ by k_p^i times, where k_p^i represents the position gain; the velocity signal amplifier (i.e., information on current velocity) which amplifies the signal from the velocity sensor of the joint $\dot{q}_i$ by k_v^i times, where k_v^i represents the velocity feedback gain (in the following, we call it velocity gain).

The way a servosystem operates is obvious from Figure 22.4. The information on the actual joint position is returned as feedback and the difference between the desired and actual position is amplified by k_p^i times. This represents the input signal for the actuator. If $q_i^0 > q_i$ this produces a positive signal which drives the motor so increase q_i until it reaches q_i^0; if $q_i^0 < q_i$, negative signal appears which drives the actuator toward decrement of the angle q_i until it reaches q_i^0; when q_i reaches q_i^0 the error Δq_i reduces to zero, and the signal at the actuator input also falls to zero, which in turn means that the actuator is stopped. However, due to rotor of the motor's inertia and the inertia of the mechanism $J_M^i N_V^i N_M^i + H_{ii}(q_j^0)$, the motor cannot be stopped instantly, and it could incur over-shooting, i.e., the real position may overshoot the desired position, q_i^0, before the motor stops. To ensure an adequate positioning of the joint (without overshoot) we have introduced a velocity feedback loop: the information (signal) from the velocity sensor is amplified k_v^i times and brought to the actuator input to dampen too sharp changes in the actuator motion that may be caused by the position feedback loop.

Therefore, the servosystem generates the following signal at the actuator input:

$$u_i = -k_p^i (q_i - q_i^0) - k_v^i \dot{q}_i = k_p^i \Delta q_i - k_v^i \dot{q}_i \tag{22.9}$$

The synthesis of a servosystem means selection of the position and velocity gains to achieve a satisfactory positioning of the joint in the desired position q_i^0. This means that if, e.g., a signal of a step type which corresponds to a desired position of the joint (Figure 22.5) is fed at the servosystem input, the servosystem response, i.e., the resulting movement of the joint, depends upon the selection of the feedback gains. It can be shown, by solving the differential Equations (22.8) with the input defined by (22.9), that the response of the servo can appear in three various forms depending on the selection of k_p^i and k_v^i (see Figure 22.5):

1. The servo can be underdamped, in which case the joint rapidly moves from its initial position and reaches the desired position q_i^0 but then overshoots it, i.e., q_i gets values that are higher than q_i^0 and, then, it oscillates around the desired position before settling at the final desired value.
2. The servo can be critically damped. In this case, the joint reaches the desired position relatively quickly, but there are no overshoots or oscillations, and the joint quickly settles at the given q_i^0.
3. The servo can be overdamped, in which case as the joint slowly approaches the desired position, there are no overshoots or oscillations, but the settling period is considerably longer than in the case of a critically damped servo.[7]

© 2002 by CRC Press LLC

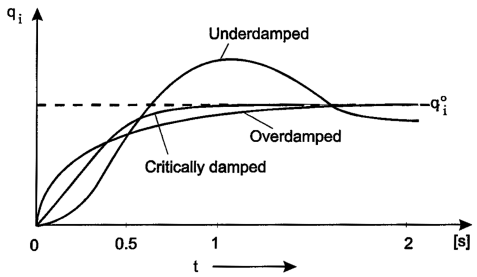


FIGURE 22.5 Responses of servo to step input.

These three types of servo responses can be described by the following functions (as the solutions of the differential Equations (22.8) and (22.9), depending on k_p^i and k_v^i):

1. The under-damped servo:

$$q_i - q_i^0 \approx C_1 e^{-\xi_i \omega_i t} \sin(\omega_i \sqrt{1-\xi_i^2}\, t) + C_2 e^{-\xi_i \omega_i t} \cos(\omega_i \sqrt{1-\xi_i^2}\, t) \tag{22.10}$$

2. The critically damped servo:

$$q_i - q_i^0 \approx C_3 e^{-\omega_i t} \tag{22.11}$$

3. The overdamped servo:

$$q_i - q_i^0 \approx C_4 e^{-(\xi_i \omega_i t + \omega_i \sqrt{1-\xi_i^2}\, t)} + C_5 e^{-(\xi_i \omega_i t - \omega_i \sqrt{1-\xi_i^2}\, t)} \tag{22.12}$$

In the functions (22.10)–(22.12) C_i represent constants (which depend on q_i^0) while ξ_i represents the damping factor and ω_i is the so-called characteristic frequency of the servo. The damping factor ξ_i defines whether the servo is critically damped, overdamped, or underdamped. If:

$$\begin{aligned} &\xi_i < 1, \text{ then the servo is underdamped} \\ &\xi_i = 1, \text{ then the servo is critically damped} \\ &\xi_i > 1, \text{ then the servo is overdamped.} \end{aligned} \tag{22.13}$$

The damping factor ξ_i and the characteristic frequency ω_i of the system are the features of the servo that are direct functions of the selection of the feedback gains k_p^i and k_v^i, as well of the parameters of the actuator and the mechanical part of the robot. It can be shown that[8]

$$\xi_i = (C_M^{i'} k_v^i + C_E^{i'} + B_C^i) / [2\sqrt{C_M^{i'} k_p^i (J_M^i N_V^i N_M^i + H_{ii})}] \tag{22.14}$$

$$\omega_i = [\sqrt{C_M^{i'} k_p^i} / \sqrt{J_M^i N_V^i N_M^i + H_{ii}}] \tag{22.15}$$

It is obvious by the selection of the gains k_p^i and k_v^i that it is possible to directly influence ξ_i and ω_i, and by this it is possible to directly change the character of the servo's response, and to directly influence the way in which the joint is driven to the imposed (desired) position.

© 2002 by CRC Press LLC

In the selection of k_p^i and k_v^i several requirements have to be satisfied:

1. The servo controlling the joint of a robot must not be underdamped under any circumstances. If a servo is underdamped, an overshoot of a desired joint position would occur and oscillations would appear. This is not acceptable with robots, because if the desired position of the link is close to some obstacle in a workspace, and an overshoot occurs, the robot could hit or collide with the obstacle. The servo, therefore, has to be overdamped ($\xi_i > 1$) or critically ($\xi_i = 1$) damped. As the servo's response is significantly slower if it is overdamped, to achieve a response as fast as possible (but without an overshoot and oscillations), it is most suitable that the servo is critically damped.
2. Up to now we have ignored the influence of the gravitational moment about the joint and actuator axis G_i. All the above considerations are valid assuming that the external moments are not acting upon the actuator (except the inertia moment, $H_{ii}\ddot{q}_i$). Let us consider the influence of the gravitational moment. When the joint comes close to the desired position q_i^0, the gravitational moment of the mechanism $G_i(q_j^0, q_i^0)$ is acting about the axis of the joint and the actuator. Because the error between the desired and actual position Δq_i would drop to zero and as the actuator is stopped the velocity, $\dot{q}_i$ also would fall to zero, and the signal at the actuator input would also have to drop to zero in accordance to Equation (22.9). This means that the driving torque produced by the actuator would also fall to zero. However, the actuator should produce the torque to compensate for the gravitational moment G_i (if not, the gravitational moment causes movement of the joint). To produce the actuator torque which would compensate for the external load G_i, some signal u_i must be generated at the actuator input. Looking at Equation (22.9) it is obvious that such a signal can be generated only if some error occurs between the actual and the desired positions, once the joint motion is terminated. The error in the positioning of the joint which appears in a steady state due to external load G_i is called the steady-state error. From Equations (22.8) and (22.9) it is easy to calculate this error as:

$$\Delta q(\infty) = [G_i(q_j^0, q_i^0) / C_M^{i'} k_p^i) \tag{22.16}$$

   i.e., the steady-state error is inversely proportional to the position gain. Because our aim is to reduce the error in robot positioning to the minimum, it is obviously necessary to increase the position gain as much as possible.
3. The structure of the robot itself has its own frequency at which the resonant oscillations of the entire robot structure appear. This frequency is called the structural frequency ω_o. According to requirement (1), the gains have to be selected in a way to ensure that the servos are always critically damped. However, because the damping factor ξ_i depends upon the different parameters of the actuators and the mechanism, it is possible that the oscillations of the servos with the frequency ω_i yet may appear. If the characteristic frequency of the servo ω_i is close (equal) to the structural frequency ω_o, the resonant oscillations of the whole structure may appear. Because these oscillations must not be allowed under any circumstances, the characteristic frequency of the servo must be sufficiently below the range of any possible structural frequency; that is, the characteristic frequency must satisfy:[7]

$$\omega_i \leq 0.5\omega_O \tag{22.17}$$

   If condition (22.17) is met, the characteristic frequency is sufficiently low so that the structural frequency cannot be excited and the undesired oscillations cannot appear. The problem lies in the fact that the structural frequency is often hard to determine theoretically and usually is identified experimentally. Because according to Equation (22.15) the characteristic frequency of the servo is directly proportional to the position gain, condition (22.17) means

© 2002 by CRC Press LLC

that the position gain has to be limited, it must not be too high to prevent the servo's characteristic frequency from becoming too high and reaching the range of the structural frequency of the robot mechanism.

4. The electrical signals in the servos in Figure 22.4 are never ideally "clean," but always include a certain "noise" superimposed upon the useful information. For example, apart from the useful information, signals from sensors (potentiometers, tachogenerators, etc.) may include noise which originates from various sources (voltage sources are never accurate, certain oscillatory modes always appear, etc.). The noise is usually an order-of-magnitude lower signal than the useful signal. These signals are amplified by the amplifiers k_p^i and k_v^i. If these gains are too high, they amplify not only the useful signals but also the noises; thus, the influence of these noises upon the servo's performance may become significant, which is why limited values of the gains have to be selected.

Based upon the above listed requirements, the gains k_p^i and k_v^i have to be selected. Requirements (3) and (4) are essentially the same, and both demand that the gains to be limited (i.e., the gains must not be too high). Usually if requirement (3) is satisfied, requirement (4) is also met. However, requirement (2) is opposite to these two, as it demands that the position gain should be as high as possible (to keep the steady-state error minimal). Because of this, the following procedure for selecting the gains is usually applied:

1. The maximum allowed position gain is selected to satisfy requirement (3). Based upon Equation (22.15) and (22.17) we get:

$$k_p^i = \frac{\omega_O^2}{4C_M^{i'}}(J_M^i N_V^i N_M^i + H_{ii}) \tag{22.18}$$

2. It is necessary to check whether or not the gain k_p^i calculated by Equation (22.18) also satisfies requirement (4). Because we have selected the maximum allowed k_p^i, we have also satisfied requirement (2) to the highest possible degree.
3. Because the servo has to be critically damped, $\xi_i = 1$, the velocity gain is defined by:

$$k_v^i = [2\sqrt{C_M^{i'} k_p^i}\sqrt{(J_M^i N_V^i N_M^i + H_{ii})} - C_E^{i'} - B_C^i] / C_M^{i'} \tag{22.19}$$

In this way we obtain the gains which satisfy all requirements to the maximum possible degree.

It should be noted that, because the linear servos are applied not only in robotics, but for the control of a number of other systems as well, it is possible to synthesize the feedback gains by applying various other methods developed in automatic control theory. These methods, such as methods in frequency domain, pole-placement methods, linear optimal regulator, etc. can be easily found in the relevant references.[8,9]

Example: For the first joint of the manipulator presented in Figure 22.6, a synthesis of the servo gains should be carried out. The joint is driven by a D.C. electromotor of the type IG2315-P20, the parameters of which are presented in Table 22.1. The data on masses, moments of inertia, lengths, and positions of the centers of masses of the robot links are provided in Table 22.2. It is rather easy to show that the moment of inertia of the mechanical part of the robot around the axis of the first joint is given by

$$H_{11}(q_j^0) = J_{z1} + J_{z2} + J_{z3} + m_3(l_3 + q_3)^2 \tag{22.20}$$

If the third link is fixed in the position $q_3^0 = 0$, the moment of inertia of the mechanism around the axis of the first joint is $H_{11} = 0.403$ kgm^2 ($l_3 = 0.035$ m). Using the values of the actuator

© 2002 by CRC Press LLC

TABLE 22.1 Data on Actuators for the Robot Presented in Figure 22.6

Actuator	1	2	3
C_e^i (V/rad/s)	0.0459	0.0459	0.0459
C_M^i (M/A)	0.0480	0.0480	0.0480
J_M^i (kgm^2)	0.00003	0.00003	0.00003
N_V^i (–)	31.17	2616.0	1570.0
N_M^i (–)	31.17	2616.0	1570.0
R_r^i (Ω)	1.6	1.6	1.6
B_C^i (Nm/rad/s)	0.0058	0.0154	0.00092
			3.0

TABLE 22.2 Data on Robot Presented in Figure 22.6

Link	1	2	3
Mass (kg)	10.0	7.0	4.15
Length (m)	0.213	0.026	0.035
J_x (kg m^2)	—	—	—
J_y (kg m^2)	—	—	—
J_z (kg m^2)	0.0294	0.055	0.318

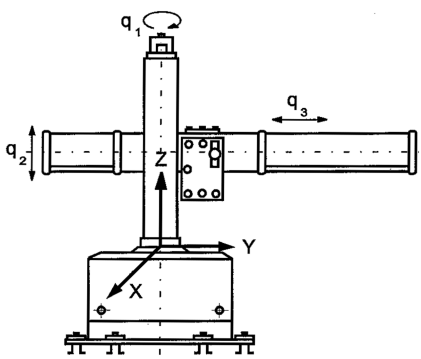


FIGURE 22.6 Robot with three joints.

parameters as given in Table 22.1, we can get the model of the actuator and the joint dynamics in the form (22.8) where the matrices are given by:

$$\bar{A}_1 = \begin{bmatrix} 0 & 1 \\ 0 & -3.117 \end{bmatrix}, \bar{b}_1 = \begin{bmatrix} 0 \\ 2.17 \end{bmatrix}, \bar{f}_1 = \begin{bmatrix} 0 \\ -2.13 \end{bmatrix} \tag{22.21}$$

The structure of the servo to be synthesized is given in Figure 22.4. The gains of the servo are selected according to the above presented approach. Let us assume that the structural frequency is identified (experimentally) to be ω_o = 24 Hz. Based on Equation (22.18) we obtain the position feedback gain as:

© 2002 by CRC Press LLC

$$k_p^i = 62.2 \text{ [V/rad]}$$

Assuming that the noises in the sensor that measures the position of the joint do not exceed 1% of the useful signal and assuming that the total angle of a rotation of this joint is ±180°, we can determine the signal at the amplifier output due to noises to be 0.3 V, which may be considered as negligible. The velocity feedback gain is obtained based on Expression (22.19):

$$k_v^i = 9.62 \text{ [V/rad/s]}$$

This gain is also relatively low so it will not cause significant influence of the noise.

22.3.3 Influence of Variable Moments of Inertia

The described synthesis of a servo is in essence the standard synthesis of a servosystem for mechanical systems. However, robotic systems have some essential differences to other mechanical systems. For example, robots have variable moments of inertia of the mechanisms about the joint axes. We have assumed that only i-th joint can move while all the other joints are fixed in the given positions q_j^0. The moment of inertia of the mechanism about the axis of the i-th joint $H_{ii}(q_j^0)$ depends on the angles (positions) q_j^0 at which the joints behind the i-th joint in the kinematic chain are fixed. If the position (angle) of any joint behind the i-th joint is changed, the moment of inertia of the mechanism about the axis of the i-th joint will change as well.

Let us briefly consider how the variations of the moment of inertia of the mechanism influence the performance of the servo in the i-th joint. Let us assume the gains k_p^i and k_v^i are calculated for such a position of the joints of the robot q_j^0 for which the moment of inertia around the axis of the i-th joint has the value of $\overline{H}_{ii}$. In this case the gains are given by:

$$\begin{aligned} k_p^i &= \frac{1}{4C_M^{i'}} \omega_o^2(\overline{H}_{ii})(J_M^i N_V^i N_M^i + \overline{H}_{ii}) \\ k_v^i &= [2\sqrt{C_M^i k_p^i}\sqrt{(J_M^i N_V^i N_M^i + \overline{H}_{ii})} - C_E^{i'} - B_C^i] / C_M^{i'} \end{aligned} \tag{22.22}$$

where by $\omega_o(\overline{H}_{ii})$ we have denoted the structural frequency of the robot for the moment of inertia $\overline{H}_{ii}$. It has been shown[7] that the structural frequency is inversely proportional to the square root of the moment of inertia of the mechanism, i.e.,

$$\omega_o(H_{ii}) = \frac{k}{\sqrt{(J_M^i N_V^i N_M^i + H_{ii})}} = \frac{\omega_o(\overline{H}_{ii})\sqrt{(J_M^i N_V^i N_M^i + \overline{H}_{ii})}}{\sqrt{(J_M^i N_V^i N_M^i + H_{ii})}} \tag{22.23}$$

where k is the proportionality factor.

If any of the joints in the kinematic chain of the robot (behind the i-th joint) change its position $q_j \neq q_j^0$, then the moment of inertia about the i-th joint axis will also change and become $H_{ii}(q_j) \neq \overline{H}_{ii}$. In this case the characteristic frequency of the i-th joint servo can be obtained in the following form (if we introduce the expression (22.22) for the position gain in Equation (22.15)):

$$\omega_i(H_{ii}) = \frac{\omega_o(\overline{H}_{ii})\sqrt{(J_M^i N_V^i N_M^i + \overline{H}_{ii})}}{2\sqrt{(J_M^i N_V^i N_M^i + H_{ii})}} \tag{22.24}$$

Obviously, the characteristic frequency of the i-th servo has to satisfy the following inequality:

© 2002 by CRC Press LLC

$$\omega_i(H_{ii}) = \frac{\omega_o(\overline{H}_{ii})\sqrt{(J_M^i N_V^i N_M^i + \overline{H}_{ii})}}{2\sqrt{(J_M^i N_V^i N_M^i + H_{ii})}} \leq \frac{1}{2}\omega_o(H_{ii}) \tag{22.25}$$

By introducing the expression (22.23) for the structural frequency into (22.25), it can be easily checked that this inequality is always satisfied. This means that regardless of the moment of inertia of the mechanism, requirement (3) (given by Equation (22.17)), which stipulates that the characteristic frequency of the servo has to be sufficiently beyond the structural frequency, is always satisfied (if the position gain is selected according to Relation (22.22)).

However, the damping factor of the servo in the i-th joint varies with the moment of inertia of the mechanism according to the following Equation (based on Equations (22.14) and (22.22)):

$$\xi_i = \frac{C_M^{i'} k_v^i + C_E^{i'} + B_C^i}{2\sqrt{C_M^{i'} k_p^i \sqrt{J_M^i N_V^i N_M^i + H_{ii}}}} = \frac{\sqrt{J_M^i N_V^i N_M^i + \overline{H}_{ii}}}{\sqrt{J_M^i N_V^i N_M^i + H_{ii}}} \tag{22.26}$$

If the j-th joint changes its position to the one in which the mechanism's moment of inertia around the i-th joint H_{ii} is less than $\overline{H}_{ii}$ for which the servo gains were computed, i.e., if $\overline{H}_{ii} > H_{ii}$, the servo is obviously overdamped in the new position of the mechanism, i.e., $\xi_i > 1$. However, if the mechanism comes into the position in which the robot's moment of inertia mechanism around the i-th joint is greater than the moment of inertia $\overline{H}_{ii}$ for which the gains were computed, i.e., if $\overline{H}_{ii} < H_{ii}$ it is obviously $\xi_i < 1$. This means the servo would be underdamped. As we have explained above (requirement 1), the servo for robots must not be underdamped under any circumstances. To ensure that the servo is always over-critically damped ($\xi_i > 1$), we must not allow the case $\overline{H}_{ii} < H_{ii}$. This leads to the following conclusion: to ensure that the servo is always over-critically damped, the gains have to be selected for the mechanism's position for which the moment of inertia of the mechanism around the i-th joint is maximal. As can be seen from Equation (22.26), the damping factor does not depend upon the selection of the position gain (if the velocity gain is selected according to Equation (22.22)). Thus, we have to select the velocity gain for the mechanism's position for which the mechanism's moment of inertia around the axis of the i-th joint $\overline{H}_{ii}$ is at the maximum possible.

The procedure is as follows. All possible positions of the mechanism should be examined (by varying the joints angles q_j) and the maximum moment of inertia of the mechanism $\overline{H}_{ii} = \max H_{ii}(q_j)$ should be determined. For the defined moment of inertia we have to compute the velocity gain k_v^i according to Equation (22.22). In all positions of the mechanism for which $H_{ii}(q_j) \neq \overline{H}_{ii}$ the servo must be overdamped (according to Equation (22.26) because $\overline{H}_{ii} > H_{ii}$). However, if the moment of inertia varies so much that in some positions of the mechanism $\overline{H}_{ii} >> H_{ii}$, the damping factor can become too high $\xi_i >> 1$, which in turn means that the servo is very over-critically damped, the positioning is very slow, and the performance of the servo then may become nonuniform depending on the mechanism position, which is unacceptable for any robot application. To ensure that robot performance is nearly uniform in all positions of the mechanism, we have to ensure that the damping factor is approximately constant. To achieve this we must introduce the variable velocity gain k_v^i (because the damping factor does not depend upon the selection of the position gain). For each position of the mechanism we have to compute the moment of inertia $H_{ii}(q_j)$ and determine the gains k_v^i so as to achieve $\xi_i = 1$. The implementation of a variable gain is significantly more complex than the implementation of fixed gains. Another way to compensate for the influence of the variable moment of inertia of the mechanism is by an introduction of global gain (see 22.4.2.).

However, if the variation of the mechanism's moment of inertia is not too high, quite satisfactory performance of the servo can be obtained even with constant velocity gains (computed for max $H_{ii}(q_j)$). If we consider Equation (22.26) for the damping factor, it is obvious that the moment of inertia of the motor rotor and the reduction ratio of the gears have an effect upon the variation of

© 2002 by CRC Press LLC

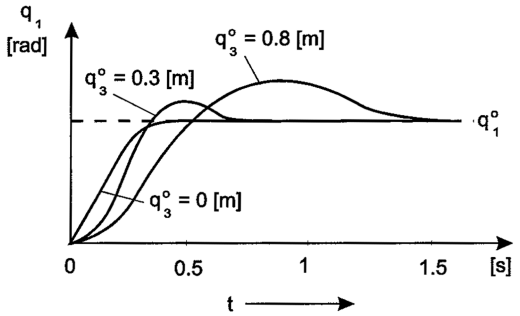


FIGURE 22.7 Responses of the servo in the first joint of the robot presented in Figure 22.6 for various positions of the third joint.

the damping factor with the variation of $H_{ii}(q_j)$. If $J_M^i N_V^i N_M^i >> (\overline{H}_{ii} - H_{ii}(q_j))$, it is obvious that the damping factor will not change significantly regardless of the moment of inertia's variation of the mechanism. In other words, if the equivalent moment of inertia of the motor's rotor is large with regard to the variation of the mechanism's moment of inertia, we may expect that the performance of the servo will be uniform (and approximately critically damped) for all positions of the mechanism, even if we keep the velocity gain fixed. Thus, by selecting a large (powerful) motor and gears we may eliminate the influence of the variable mechanism's moment of inertia. This approach is often applied in the design of robots. However, it is obvious that such a solution has certain drawbacks from the point of view of power consumption, unnecessary loading of joints, as well as the use of unnecessarily powerful actuators and large (heavy) gears.

The bigger gears may be especially inconvenient due to a large backlash and high dry friction coefficients which they may introduce in the system. The introduction of direct-drive actuators (i.e., motors without gears) effectively solves the problems regarding the backlash and friction, but on the other hand, the variation of the mechanism's moment of inertia may affect the servo's performance with such actuators and, therefore, a more complex control law (e.g., with variable velocity gain) has to be applied.

Example: For the servo in the first joint of the robot presented in Figure 22.6, in the previous example, we have computed the gains when the third joint is in the position $q^0 = 0$. Considering Equation (22.20) for the moment of inertia of the mechanism around the axis of the first joint, it is obvious that if the third joint is set in the position $q^0 > 0$ the moment of inertia of the mechanism H_{ii} will be higher and the damping factor will be less than 1. Using Equation (22.26), the damping factor for the position of the third joint, $q_3^0 = 0.3$ m, can be calculated as:

$$\xi_1 = \frac{\sqrt{0.435}}{\sqrt{0.895}} < 1$$

Thus, the gains selected in the previous example will not be satisfactory for all positions of the mechanism. In Figure 22.7 the servo's responses for the various positions of the third joint are presented. This is why the gains must be selected for the mechanism's position for which $\overline{H}_{ii} = \max H_{ii}(q_j)$. In this case, H_{ii} is at maximum if q_3^0 is at maximum, i.e., for $q_3^0 = 0.8$ m. We may calculate that $\overline{H}_{ii}$ ($q_3^0 = 0.8$ m) = 3.323 kg m^2, and the gains are obtained as:

$$k_p^i = 62.2 \text{ [V/rad]}, \quad k_v^i = 27.5 \text{ [V/rad/s]}$$

If we compute the gains in this way, the servo will be overdamped for all positions of the mechanism. According to Equation (22.26) the damping factor changes with the variation of q_3^0 as

© 2002 by CRC Press LLC

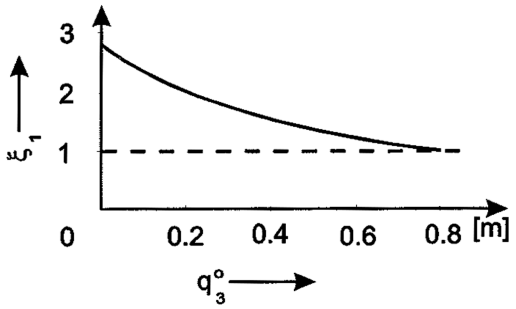


FIGURE 22.8 The variation of the damping factor of the servo in the first joint of the robot presented in Figure 22.6 for various positions of the third joint.

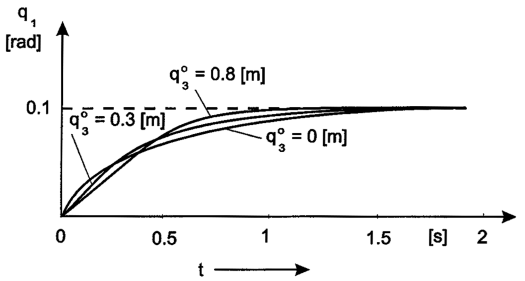


FIGURE 22.9 Responses of the servo in the first joint of the robot presented in Figure 22.6 for various positions of the third joint.

presented in Figure 22.8. It can be seen that for $q_3^0 = 0$, the servo is strongly overdamped, which causes slow positioning of the first joint. In Figure 22.9, the response of the first joint for various positions of the third joint is presented. To achieve a more uniform positioning of the joint it is necessary to introduce (a) variable gains, (b) global control loop, or (c) to apply a larger actuator and gears with a higher equivalent moment of inertia of the rotor.

22.3.4 Influence of Gravity Moment and Friction

We have already explained that the gravity moment of the mechanism causes a steady-state error in robot positioning. Because our aim is to minimize the errors in robot positioning, we have to consider various possibilities to compensate for the influence of gravity moments:

1. We have shown above that a steady-state error is directly proportional to the gravity moment and inversely proportional to the position gain and the moment coefficient of the actuator. We have shown as well that if we select higher position gain the steady-state error will be reduced. However, the position gain is limited by the resonant structural frequency and noises, so the steady-state error cannot be eliminated beyond a certain limit by purely increasing the position gain. Obviously, by the selection of a more powerful actuator (with a higher moment coefficient) and larger gears (with a higher moment reduction ratio), one may decrease the steady-state error, but this solution has some drawbacks, as already pointed out above.
2. Gravity moments can be compensated for by introducing an additional signal at the actuator input; this signal is proportional to the gravity moment (see Figure 22.10). In this case the control system has to compute the gravity moment of the i-th joint $G_i(q_j^0, q_i)$ as a function of the coordinates (positions) of the robot's joints, and generate at the actuator input an

© 2002 by CRC Press LLC

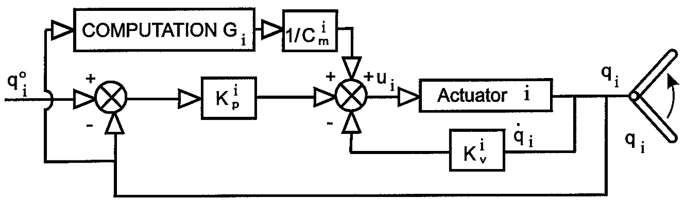


FIGURE 22.10 Positional servo with gravity moment compensation.

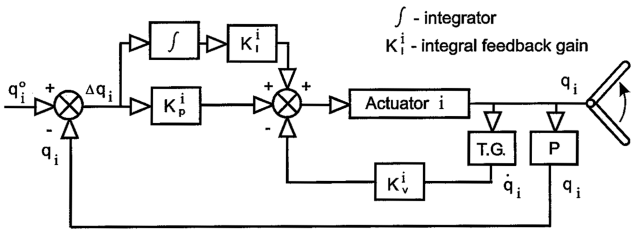


FIGURE 22.11 PID regulator in the i-th joint of the robot.

additional signal which will produce a compensating torque. Thus, the input signal for the actuator is defined by:

$$u_i = -k_p^i(q_i - q_i^0) - k_v^i \dot{q}_i + \frac{G_i}{C_M^i} \tag{22.27}$$

In this way we can eliminate the steady-state error due to the gravity moment. However, a drawback of this solution lies in the fact that it requires the control system to compute gravity moments which, in turn, ask for an accurate identification of the parameters of the mechanism (masses, centers of masses, lengths of links).

3. A steady-state error can be eliminated by introducing an integral feedback loop, i.e., a feedback loop with respect to the integral of the position error (see Figure 22.11). Thus, the PID regulator is obtained (P, proportional; I, integral action; D, differential) which is often applied in practice to a number of systems. In this case, the signal for the actuator input is generated as:

$$u_i = -k_p^i(q_i - q_i^0) - k_v^i \dot{q}_i + k_I^i \int_0^t (q_i^0(t) - q_i)dt \tag{22.28}$$

where k_I^i is the integral feedback gain. The integral feedback has the role of producing a signal proportional to the integral of the position error when the servo approaches the desired position. This signal obviously compensates for the external load and eliminates the steady-state error. There are obvious advantages of this solution over the previous one: the PID solution does not require knowledge of the robot parameters, and the PID regulator compensates other (time constant) perturbing moments' action about the joint axis (these perturbing moments need not be identified, but the PID regulator may compensate them). However, the synthesis of the gains for the PID regulator (which will not be considered here, see, e.g., Paul,[13] is not simple because with the PID regulator it is not possible to satisfy all

© 2002 by CRC Press LLC

the above defined requirements upon the servo (e.g., it is not possible to eliminate overshoots, etc.).

4. Finally, a steady-state error due to gravity moments can be reduced by introducing brakes in the joints, which should hold a joint in the desired position once the servo reaches it. This solution is rather simple regarding the control, but often it cannot technically be applied and is inconvenient for an elimination of errors due to the gravity moments if trajectory tracking has to be realized.

Besides the gravity moments, friction forces may also affect the performance of a servo. These forces about the joint axis also cause errors in servo positioning and operation. In this, special problems arise due to static friction forces that appear when the joint starts to move from the still position which differs from the dynamic friction forces during the motion. Compensation of these forces can be realized by one of the above listed methods for the compensation of gravity moments. However, the model and parameters of these forces are often very difficult to identify, and therefore, computation and introduction of additional compensation signals (analogous to the solution in Figure 22.10) cannot be easily implemented. The reliability of such a solution may not be appropriate. The compensation signal can be identified experimentally.

Backlash in the gears, elastic modes, and other nonlinear effects, the models of which are not simple to identify, also may affect performance of the servo. One must carefully consider these effects during synthesis and implementation of servos for robots.

Additionally, it should be mentioned that the amplitude of the input signal to the actuators is constrained, which limits the speed of the servo's positioning if the given (desired) position is far from the initial position of the joint.

22.3.5 Synthesis of the Servosystem for Trajectory Tracking

Up to now we have considered the problem of positioning of the joint in the set (desired) position q_i^0. At the input of the servo a desired position q_i^0 is fed and the joint is positioned following the above described process. However, as we have already underlined, modern industry and other applications of robots require robots which have to be not only precisely positioned in various positions in the workspace, but can also track continual trajectories. For example, with arc welding, the robot hand should move along a prescribed trajectory in the workspace with an accurately defined velocity. Often, a definition of the desired trajectory can be achieved by imposing a set of discrete points (positions) in the space through which the robot hand has to pass (the point-to-point motion). However, with the above-mentioned example of arc welding it is necessary to implement a motion of the hand (tool) along a continuous path in the workspace. In this case, all joints of the robot have to realize their desired trajectories as continual functions of time $q_i^0(t)$. This is why it is necessary to consider how the servo can ensure tracking of the continual trajectory of the joint coordinates (assuming the rest of the joints are fixed).

Let us assume that at the servo input (Figure 22.4) a signal $q_i^0(t)$ introduced which is a continuous function of time. This signal corresponds to the desired nominal trajectory of the i-th joint, i.e., to the desired variation of the joint angle along the time. This means that the joint angle has to track the trajectory $q_i^0(t)$. The servo must ensure that the actual joint position is as close as possible to $q_i^0(t)$ at each time instant. Even more important, it should ensure that the rotational speed of the joint $\dot{q}_i(t)$ is as close as possible to the desired trajectory of the speed $\dot{q}_i^0(t)$ at each time instant. However, if we just feed the desired trajectory $q_i^0(t)$ at the input of the servo (Figure 22.4), the servo output–joint angle will undoubtedly have a delay with respect to the given (desired) trajectory $q_i^0(t)$. This delay is due to dynamic characteristics of the actuator and the mechanism driven by the actuator (i.e., the inertia of the actuator rotor and of the mechanism, friction, and contra-electromotive force which is generated in the motor). Here, we will not analyze mathematically this phenomenon, but it is clear that it is necessary to compensate for this delay in order for the

© 2002 by CRC Press LLC

joint to implement accurately the desired trajectory. To compensate for this delay caused by servo dynamics, we may introduce the feedforward signal (precompensation signal).

The feedforward term has to compensate for a delay of the servo along the given nominal trajectory and can be synthesized in various ways. Here, we briefly present one simple procedure for the synthesis of the feedforward term for a robot servo. The model of the actuator of the i-th joint is given by Equation (22.8) in the state space. The nominal trajectory of the joint $q_i^0(t)$ has to be realized. Because the trajectory of the joint is given, by differentiating it we can obtain the desired variation (a trajectory) of the joint velocity $\dot{q}_i^0(t)$. The state vector of the servo and the actuator is given by (22.4). This means that the desired nominal trajectory of the state vector of the servo and the actuator is given as well, $x_i^0(t)=(q_i^0(t), \dot{q}_i^0(t))^T$. At each time instant t the difference between the actual state vector x_i (t) and the nominal trajectory $x_i^0(t)$ should be as small as possible. The feedforward term represents the signal at the actuator input $\bar{u}_i^0(t)$ which satisfies the following equation:[10]

$$\dot{x}_i^0(t)=\bar{A}_i\, x_i^0(t)+\bar{b}_i\, \bar{u}_i^0(t)+\bar{f}_i\, G_i(q_j^0, q_i^0) \tag{22.29}$$

i.e., the signal $\bar{u}_i^0(t)$ satisfies the model of the actuator and joint (Equation (22.8)) along the specified trajectory $x_i^0(t)$. The signal $\bar{u}_i^0(t)$ represents the programmed signal as a function of time and is called local nominal, programmed control. The name "local" originates from the fact that this signal is computed for one local actuator and one joint ignoring the other joints (i.e., they are assumed to be fixed). The name "programmed" originates from the fact that this control is a function exclusively of time, and not of the actual (temporary) state of the joint and the actuator (i.e., it is not dependent on the actual position and speed of the joint), and therefore, it represents the programmed input for the actuator corresponding to the programmed trajectory $x_i^0(t)$. Taking into account the form of the matrix and vectors in Equation (22.8) it can be easily shown that the signal $\bar{u}_i^0(t)$ satisfying Equation (22.29) can be computed according to the following equation:

$$\bar{u}_i^0(t)=[(J_M^i N_V^i N_M^i+\bar{H}_{ii})\,\ddot{q}_i^0(t)+(B_C^i+C_E^{i'})\,\dot{q}_i^0(t)+G_i(q_j^0, q_i^0(t))]\,/\,C_M^{i'} \tag{22.30}$$

where $\ddot{q}_i^0(t)$ represents the desired variation of joint acceleration along the specified trajectory $q_i^0(t)$, and it is obtained by the differentiation of the nominal trajectory of the velocity $\dot{q}_i^0(t)$. Based on (22.30) we obtain the local nominal control using the specified nominal trajectory of the joint. If the local nominal control is fed into the input of the actuator (as a programmed signal), and if no perturbation is acting upon the joint, the actuator and joint would move along the specified trajectory $q_i^0(t)$. However, it is obvious certain perturbations always act upon the system, and the model and parameters used for the computation of (22.30) are not ideally accurate. In addition, in the initial moment t = 0, the joint angle $q_i(0)$ need not correspond to the nominal angle $q_i^0(0)$. Because of this, the motion of the joint always deviates from the nominal trajectory when we feed the actuator with the programmed nominal control $\bar{u}_i^0(t)$ only. The behaviour of the actuator and the joint in this case is described by:

$$\dot{x}_i(t)=\bar{A}_i\, x_i(t)+\bar{b}_i\, \bar{u}_i^0(t)+\bar{f}_i\, G_i(q_j^0, q_i) \tag{22.31}$$

Obviously, if $x_i(0)\neq x_i^0(0)$, the actual state vector $x_i(t)$ will not coincide with the nominal trajectory $x_i^0(t)$. Due to this, an additional signal Δu_i must be fed into the input of the actuator to ensure that the state vector $x_i(t)$ is as close as possible to $x_i^0(t)$ when the perturbations are acting upon the system and when $x_i(0)\neq x_i^0(0)$. Let us introduce a vector of deviation of the system state from the nominal trajectory as a difference between the actual state and the nominal state $\Delta x_i(t)=x_i(t)-x_i^0(t)$. The model (22.31) can then be written in the following form:

$$\Delta\dot{x}_i(t)=\bar{A}_i\,\Delta x_i(t)+\bar{b}_i\,\Delta u_i(t)+\bar{f}_i[G_i(q_j^0, q_i)-G_i(q_j^0, q_i^0)] \tag{22.32}$$

© 2002 by CRC Press LLC

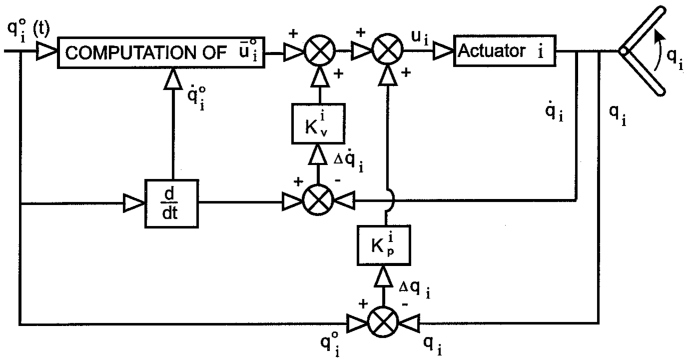


FIGURE 22.12 Servosystem with local nominal control.

Equation (22.32) is called the model of deviation of the system state from the nominal trajectory. We have to ensure that the deviation of the state $\Delta x_i(t) = x_i(t) - x_i^0(t)$ approaches zero, i.e., we have to synthesize the additional control signal Δu_i in such a way that it ensures that the deviation vector is close to zero. The model (22.32) is analogous to the basic model of the actuator and the joint (22.8). It is, therefore, obvious that the additional control signal Δu_i for the deviation model (22.32) can be generated analogously as for the actuator and joint positioning. The problem of a reduction of the deviation state vector of the deviation model (22.32) to zero is analogous to the problem of the positioning the actuator and joint (22.8) into the position $x_i = (0,0)^T$. Thus, the additional control signal can be generated as:

$$\Delta u_i = k_p^i \Delta q_i + k_v^i \Delta\dot{q}_i = -k_i^T \Delta x_i \tag{22.33}$$

where by $k_i = (k_p^i, k_v^i)^T$ is denoted a vector of feedback gains. The total signal which has to be fed to the actuator input is

$$u_i(t) = \bar{u}_i^0(t) + \Delta u_i(t) = \bar{u}_i^0(t) - k_i^T \Delta x_i \tag{22.34}$$

Figure 22.12 presents the control scheme which ensures tracking of the trajectory. The servo for trajectory tracking has a similar structure as the servo for positioning (Figure 22.4). The only difference is in the feedforward term which represents the computation of the local nominal control according to Equation (22.30), and in the fact that instead of feedback by the velocity we introduce the difference between the actual velocity and the nominal velocity $\Delta\dot{q}_i$. This difference (velocity error) is amplified by k_v^i. Because the deviation model (22.32) is analogous to the model (22.8), the synthesis of the gains k_p^i and k_v^i for the servo with the feedforward term is analogous to the synthesis of the gains for positioning of the servo.

It should be mentioned that for the computation of the local nominal control according to Equation (22.30), for the moment of inertia of the mechanism $H_{ii}(q_j)$ the least possible value this moment of inertia may have depending on the position of the rest of the joints q_j should be taken. The reason for this is to avoid overshoots. Therefore, the procedure for selecting the moment of inertia of the mechanism for the calculation of local nominal control is analogous to the one for the computation of the velocity feedback gain (but the minimum value is searched for).

Example: Let us assume that at the input of the servo for the first joint of the robot presented in Figure 22.6 (the servo was synthesized in Section 22.3.2), instead of the position, the signal corresponding to the joint trajectory $q_1^0(t)$ is fed. This trajectory is presented in Figure 22.13 and can be described by

© 2002 by CRC Press LLC

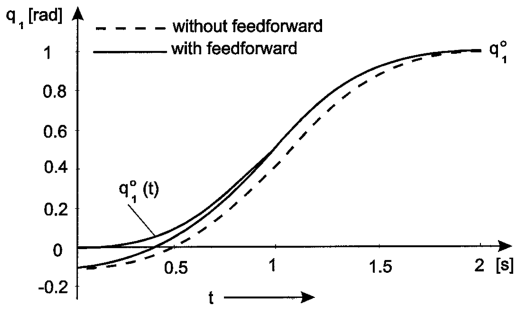


FIGURE 22.13 Trajectory tracking with and without a feedforward term (the first joint of the robot presented in Figure 22.6).

$$q_1^0(t) = a_1 t^2 / 2, \quad 0 < t \le \tau / 2$$

$$q_1^0(t) = a_1 (\tau t - t^2 / 2 - \tau^2 / 4), \quad \tau / 2 < t < \tau$$

where $a_1 = 1$ rad/sec^2 is the acceleration, and $\tau = 2$ s is the time duration of the movement. If the feedforward is not introduced, but we directly feed the desired trajectory at the input of the servo, the actual trajectory will be delayed to the nominal one, as can be seen from Figure 22.13 (for an initial error $\Delta q_1(0) = q_1^0(0) - q_1(0) = 0.1$ rad). This is why it is necessary to introduce a feedforward term to compensate for this delay.

22.4 Control of Simultaneous Motion of Several Robot Joints

Up to now we have analyzed the control of one single joint of the robot, assuming that all the remaining joints are fixed. However, to execute different control tasks by the robot, the hand of the robot has to be positioned in the workspace. To do this, it is necessary to drive all joints of the robot into certain positions (angles) which correspond to the desired position of the hand. Generally speaking, it would be possible to drive the joints to the certain positions successively, one by one, so that while each joint is moving the remaining joints are fixed. In this case, the control we have observed up to now would be satisfactory. However, it is obvious that such a positioning of the robot, joint by joint, is not efficient from the point of view of the time required to accomplish the task. Obviously, the time required for the positioning of the hand into the desired position if all joints are moving simultaneously is considerably less than if the joints are moving successively. Because one of the main goals in robot design is to achieve a quick as possible working speed, it is clear that for the modern robots simultaneous positioning of all joints must be ensured. Even more if the tracking of a given path of the hand is required, it is obvious that all joints of the robot must move simultaneously (i.e., they have to track their corresponding trajectories simultaneously). Thus, the simultaneous movement of all robot's joints is a must for modern robots.

If several joints of the robot are moving simultaneously, dynamic coupling between the joints must appear. This dynamic coupling must affect the performance of the servo in the robot's joints. We consider the influence of the dynamic forces, and the synthesis of the control which ensures a satisfactory performance of the robot when its joints are moving simultaneously.

22.4.1 Analysis of the Influence of Dynamic Forces

The model of robot dynamics was presented in previous chapters. The moments about the joint axes can be described as functions of the joint angles, velocities, and accelerations:

© 2002 by CRC Press LLC

$$P = H(q)\ddot{q} + h(q,\dot{q}) \tag{22.35}$$

where $H(q)$ is the inertia matrix n × n, $h(q,\dot{q})$ is the vector of the centrifugal, Coriolis, and gravity moments n × 1. If several joints are moving simultaneously, the dynamic moment P_i is acting around the i-th joint. This means that an external load is acting upon the i-th actuator and this load P_i is the function of the angles, velocities, and accelerations of all joints. If only the i-th joint is moving (and all the remaining joints are fixed) upon the actuator is acting the load given by (22.7) which we took into account within the synthesis of the servo in the i-th joint. The dynamic moment caused by simultaneous motion of several joints loads the i-th servo and affects servo performance. In the text to follow we consider qualitatively how certain dynamic forces (moments) in simultaneous motion of several joints affect the performance of the servos. We consider both the positioning problem (by simultaneous positioning of all robot's joints) and the problem of tracking the robot hand trajectory (by simultaneous tracking all the joint trajectories). Let us assume that the desired positions of the joints q_i^0 or the trajectories $q_i^0(t)$ are fed simultaneously at the inputs of all robot servos.

1. **Variable moment of inertia.** If several joints are moving simultaneously, the moment of inertia of the mechanism around the i-th joint varies during the motion, because H_{ii} depends on the positions of all the joints in the kinematic chain behind the i-th joint. In 22.3.3 we considered the influence of the variation of the moment of inertia of the mechanism upon the servo performance. We have shown that it is necessary to compute the feedback gains for the maximum possible value of the moment of inertia of the mechanism to prevent the system from being underdamped. However, we have also seen that if the moment of inertia is significantly varied, the performance of the robot (servo) can be uneven. This is especially inconvenient with simultaneous motion of several joints, as the moment of inertia about a joint varies during the motion, which may cause oscillatory tracking of trajectories or positioning. However, this problem can be solved in one of the ways previously mentioned.
2. **Cross-inertia members.** Accelerations of the j-th joint cause the moment in the i-th joint through cross-inertia moments, which themselves represent the elements of the inertia matrix H off the main diagonal $H_{ij}(i \neq j)$. Thus, due to acceleration in the j-th joint $\ddot{q}_j$, an external load $H_{ij}(q)\ddot{q}_j$ acts upon the i-th servo. As we explained above, the external load upon the shaft of the actuator causes errors in the positioning of the joint, or in the tracking of trajectories (because the servo must overcome this external load through a position error which will create a corresponding compensating signal). However, this moment is significant only if accelerations are relatively high. When the robot stops in the desired position, the accelerations drop to zero and, therefore, they do not influence the positioning of the joints, i.e., they do not cause steady-state errors. These moments can cause errors in tracking the trajectories if they are with high accelerations $\ddot{q}_j$. If in a certain application task it is not essential to ensure accurate tracking of fast trajectories, the effects of these moments can be ignored. However, if accurate tracking of the fast trajectories is essential (which means that in each moment the difference between the actual position of the joint and the nominal trajectory must be minimal), then the moments due to cross-inertia members must be compensated for.
3. **Gravity moments.** The effects of gravity moments have already been considered in 22.3.4. In the simultaneous motion of several joints the gravity moments vary during the movement, causing errors both in positioning and tracking of trajectories. The compensations for these moments can be reached through one of the previously described manners, but it should be kept in mind that the gravity moments vary during the tracking of trajectories and, therefore, they cannot be completely compensated for by PID regulators.
4. **Centrifugal and Coriolis moments.** These moments are produced by the velocities $\dot{q}_j$ in the robot's joints. They also act as external loads upon the servos. However, because the centrifugal

© 2002 by CRC Press LLC

and Coriolis forces are directly proportional to rotational (or linear) velocities of the joints $\dot{q}_j$, these forces are significant only if the joints are moving at relatively high speeds. When the robot starts to move or it stops, these forces are negligible, which means they do not affect the positioning of the robot in any desired positions and do not cause steady-state errors. These forces only cause errors in the tracking of fast trajectories. Similarly, as in the case of cross-inertia moments, here we can also conclude that if an accurate tracking of fast trajectories is not required, the effects of these forces can be ignored. However, if the tracking of fast trajectories is essential for the robot application, we must take into account the centrifugal and Coriolis forces within the synthesis of control.

As can be seen from these considerations, when several joints are moving simultaneously, dynamic coupling between the joints appears which affects the positioning and tracking of trajectories. The influence of these dynamic forces upon the positioning and tracking of slow trajectories is not significant and, therefore, the servos synthesized for isolated joints of the robot can easily overcome them. In the previous paragraphs, we presented how at the level of local servos we can compensate for the effects of variable moment of inertia and external loads. These compensations are often quite sufficient to ensure positioning and tracking of slow trajectories even if several joints are moving simultaneously. For such tasks it is quite acceptable to control the robot by local servos synthesized for isolated joints. However, if tracking of fast trajectories is required, the effects of dynamic forces cannot be ignored. Because these forces act as external loads, if the servo's feedback gains are high, the errors caused by these forces may be very negligible, so that even in the case of relatively fast trajectories we may accept the servo synthesized for the isolated joints. However, because the gains are normally limited, as explained before, if we have to ensure accurate tracking of fast trajectories we cannot apply only local servos, the dynamic forces must be compensated for.

22.4.2 Dynamic Control of Robots

A number of robots available in the market are not capable of ensuring accurate tracking of fast trajectories, because their applications were sufficient for local servos without introducing any compensation for dynamic forces. However, because demands upon the robots regarding the high speed and quality of operation (accuracy of tracking of the desired paths) in modern industry and other application domains are increasing, the control systems of the newest generations of robots have to take into account the dynamics of robots. The control law which takes into account all (or some of) dynamic forces in the robotic systems is called the dynamic control of robots.

The basic problem when applying a dynamic control lies in the fact that the dynamic forces acting within the robotic mechanism are generally very complex functions comprising coordinates, velocities, and accelerations of the joints. Thus, if we want the control system to compensate for the effects of these forces, this may lead to relatively complex control laws.

Various approaches have been developed for the synthesis of the dynamic control of robots. A survey of these approaches can be found in Vukobratović et al.[11] Here, we very briefly consider just two approaches: nominal programmed control and global control.[10–12]

In Section 22.3.5 we showed that by applying the nominal local programmed control delays in servos along the trajectory, we can be compensated for the tracking process. If several joints are moving simultaneously, local nominal control cannot compensate for dynamic moments which act about the i-th joint, which is why instead of the local nominal control we may apply the nominal programmed control computed on the basis of the complete dynamic model of the system. This control can be computed in the following way. Let the nominal trajectories of the robot $q_j^0(t)$, i = 1, 2, …, n be given. By differentiating, we get the nominal trajectories of the velocities and accelerations. Now, based upon the model of the mechanism dynamics (22.35), we can compute the nominal driving torques in the robot's joints:

© 2002 by CRC Press LLC

$$P^0(t)=H(q^0(t))\ddot{q}^0(t)+h(q^0(t),\dot{q}^0(t)) \tag{22.36}$$

where by $P^0(t)$ we denote the vector of nominal driving torques, while by $q^0(t)$ we denote the vector of nominal trajectories in all n joints. The nominal driving torques represent the moments which have to be implemented about the joints of the robot to ensure that it moves along the desired nominal trajectories $q_j^0(t)$.

The models of the actuators in the joints are given by Equation (22.5) where P_i represents the external load by which the mechanism acts upon the actuators, i.e., P_i is the driving torque realized by the actuator. The nominal control signal at the actuator input which has to ensure the nominal driving torque must satisfy the equations:

$$\dot{x}_i^0(t)=A_i\,x_i^0(t)+b_i\,u_i^0(t)+f_i\,P_i^0(t),\ \ i=1,2,...,n \tag{22.37}$$

The nominal programmed control $u_i^0(t)$ which satisfies Equation (22.37) can be computed by:

$$u_i^0(t)=[(J_M^i N_V^i N_M^i+\overline{H}_{ii})\ddot{q}_i^0(t)+(B_C^i+C_E^{i'})\dot{q}_i^0(t)+P_i(t))]\,/\,C_M^{i'} \tag{22.38}$$

Obviously, the nominal control $u_i^0(t)$ differs from the local nominal control $\overline{u}_i^0(t)$ computed by (22.30), since the former includes the total nominal driving torque which represents the dynamic moment due to the movements of all joints of the robot. Thus, the nominal programmed control compensates not only for the dynamics of the actuator and a single joint, but it also compensates for the dynamics of the complete mechanism (but, only along the nominal trajectory). If we feed the nominal programmed control $u_i^0(t)$ at the inputs of all actuators, and if the models of the actuators and of the mechanism were exact, and the robot in the initial moment is in such a position that $q_i(0)=q_i^0(0)$ for all joints, and if no perturbation is acting upon the robot, the joints of the robot would move along the imposed trajectories. However, all these assumptions do not hold often (or nearly never hold). This is the reason why deviations of the joint coordinates from the nominal trajectories appear whenever we apply the nominal programmed control only. The motion of the joints can be described by the model

$$\dot{x}_i(t)=A_i\,x_i(t)+b_i\,u_i(t)+f_i\,(P_i^0(t)+\Delta P_i),\ \ i=1,2,...,n \tag{22.39}$$

where ΔP_i represents a deviation of the actual load in the joint from the nominal value $P^0(t)$. If $x_i(0)\neq x_i^0(0)$, then $x_i(t)$ will not coincide with $x_i^0(t)$ so we have to introduce an additional control signal at the actuator input. This additional control has to ensure that the actual state $x_i(t)$ is as close as possible to $x_i^0(t)$. The model (22.39) can be written as (based upon Equation (22.37)):

$$\Delta\dot{x}_i(t)=A_i\,\Delta x_i(t)+b_i\,\Delta u_i(t)+f_i\,\Delta P_i,\ \ i=1,2,...,n \tag{22.40}$$

where $\Delta x_i(t)=x_i(t)-x_i^0(t)$. The model (22.40) represents the model of deviation of the actuator state from its nominal trajectory if the motions of the joints of the robot deviate from the nominal trajectory. This model is similar to the basic model of the actuator (22.5) save for the fact that as external moment instead the total moment P_i, here acts as a deviation of the dynamic moment from the nominal driving torque. The nominal programmed control compensates for the nominal driving torque (i.e., nominal moment $P^0(t)$) and by this reduces the effects of the dynamic forces upon the servo performance. If we apply the servo synthesized in Section 22.3.5, the reduced ΔP_i dynamic moment acts upon it. Thus, the servos can overcome this external load more efficiently if we apply the nominal programmed control (see Figure 22.14).

© 2002 by CRC Press LLC

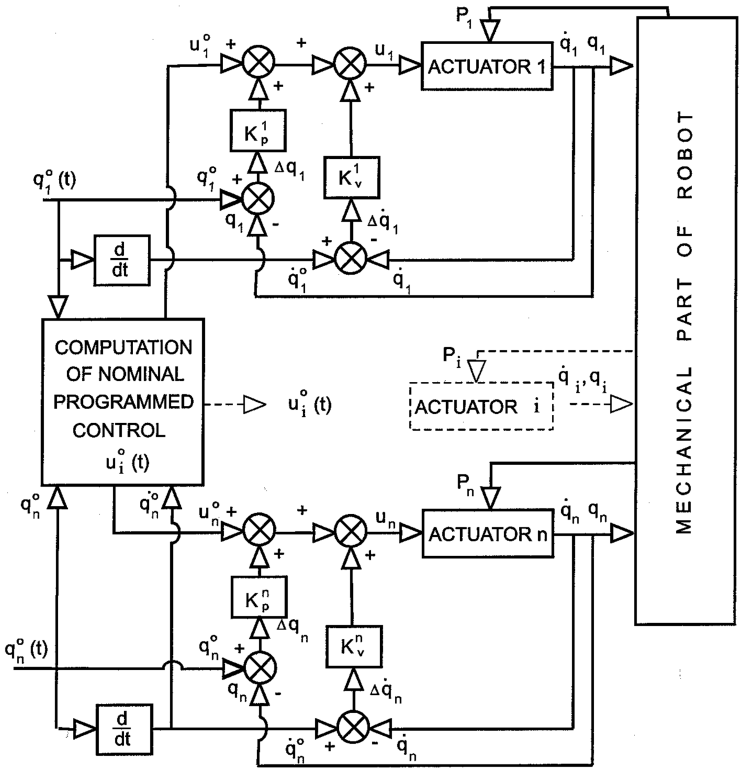


FIGURE 22.14 Control scheme of the robot including the nominal programmed control.

However, the application of the nominal programmed control involves numerous difficulties. To compute the nominal programmed control on the basis of Equation (22.38) we have to compute on-line the complete dynamic model of the robot (22.35) which may be rather complex and requires powerful computer. However, this can be easily achieved by applying standard microcomputers; the main problem is to accurately identify parameters of the actuators and of the mechanisms as well as model different dynamic effects not included in the assumed models (e.g., elastic effects, etc.).

The second possibility to compensate for the dynamic moments during the trajectory tracking is application of the global control. If we apply local nominal programmed control (22.3.5), the external load comes from the moments of the mechanism dynamics. To compensate for the effects of these dynamic moments we may introduce an additional compensating control signal at the actuator input. This additional signal has to be proportional to the dynamic moment P_i acting upon the i-th joint. This additional signal can be calculated in the following form:

$$\Delta u_i^G = -k_i^G \tilde{P}_i \tag{22.41}$$

where k_i^G represents the global gain, and $\tilde{P}_i$ represents the value which is proportional (or equal) to the dynamic moment P_i in the i-th joint. This additional control signal is called the global control because it represents the feedback between the joints. The local servo includes the feedback by local coordinates and velocities of a joint which is controlled by the servo, so it has just local information, compared with the global control which represents an exchange of information between the servos (cross-feedback global loops, see Figure 22.15).

Global control has to compensate for the effects of the dynamic moments by generating the signal at the actuator input which is proportional to this moment. Therefore, the basic problem

© 2002 by CRC Press LLC

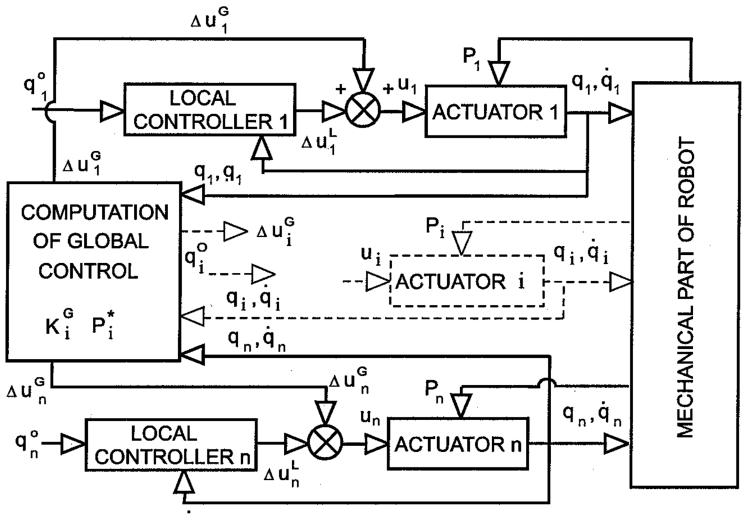


FIGURE 22.15 Control scheme of the robot including global control.

regarding global control is how to obtain information on the actual dynamic moment acting upon the servo P_i.

Two options for implementation of the global control are proposed:[10,11]

1. The dynamic force (moment) which acts upon the servo can be measured directly by a force transducer. By placing a force transducer in the shaft of an actuator, the dynamic force is directly measured. The signal from the force transducer is used as a feedback to the actuator input (multiplied with the global gain k_i^G). By this, the signal is generated at the actuator input which is proportional to the actual dynamic force and produces the additional driving torque to compensate for the dynamic moment. Global control is realized by force feedback. The problems of this option are related to the elastic effects introduced when the force tranducer is installed in the actuator shaft.
2. The second option is by on-line computation of the dynamic moment. Based on the information of the actual robot state (joint coordinates, velocities, and accelerations) obtained from the sensors, the control computer computes the dynamic moments P_i according to the model (22.35). Thus, the computer generates the signal at the actuator input which is proportional to the computed value of the dynamic moment. As we have already explained, the model of the robot dynamics may be very complex, which in turn requires fast computers capable of computing on-line the dynamic moments as a function of the actual state of the robot. This is easy achievable with standard computers. The main problem lies in the accuracy of the model of robot dynamics and in the identification of parameters. However, the effects of all dynamic forces need not be significant, and in many cases it is not obligatory to compute all the dynamic moments of the complete model. Instead, it may be sufficient to compute certain components of the dynamic moments (e.g., gravity moments, cross-inertia members, etc.). In other words, the dynamic moments can be computed by approximative models. The problem is determining which components of the dynamic moments must be compensated for by global control. Generally, this analysis requires application of the computer, i.e., the selection of the adequate dynamic control is most effectively performed by computer-aided control synthesis.

It should be mentioned that various combinations of the control laws are possible. It is possible to apply a nominal programmed control (which compensates for the nominal dynamic moments), and the local servosystems and global control (which in this case has to compensate for the deviation of the real moments from the nominal ones, ΔP_i).

© 2002 by CRC Press LLC

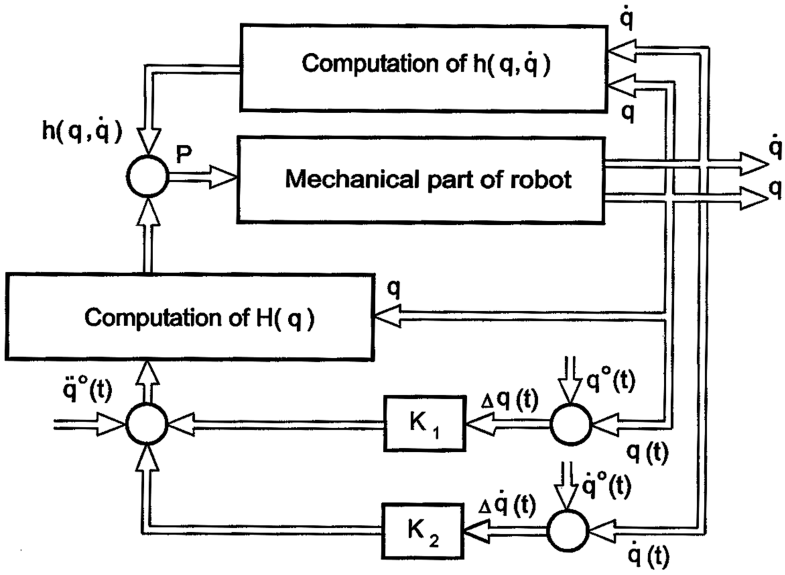


FIGURE 22.16 Control scheme of the inverse problem technique.

22.4.3 Inverse Problem Technique

One of the most investigated dynamic control law is the inverse problem technique. Paul[13] investigated the inverse problem technique, which was called the "computed torque" technique by Bejczy.[14] A similar approach has been taken by Pavlov and Timofeyev.[15] Their approaches include on-line computation of the complete model of robot dynamics, i.e., they involve computation of driving torques by Equation (22.35) using the measured values of internal coordinates q and velocities $\dot{q}$ of the robot and the computed values of internal accelerations $\ddot{q}^o(t)$. Namely, if the desired trajectory is computed, we can obtain $q^o(t)$, $\dot{q}^o(t)$, $\ddot{q}^o(t)$. It has been shown[15] that the robot is asymptotically stable around the nominal trajectory if the driving torques are computed by:

$$P(t) = H(q)\cdot[\ddot{q}^o(t) + K_1(q(t) - q^0(t)) + K_2(\dot{q}(t) - \dot{q}^o(t))] + h(q,\dot{q}) \tag{22.42}$$

where K_1 is an nxn matrix of position feedback gains and K_2 is an nxn matrix of velocity feedback gains; K_1 and K_2 must be chosen in such a way that a trivial solution of

$$\ddot{e} = K_1 e + K_2 \dot{e}$$

is asymptotically stable, where e is $nx1$ vector. However, only driving torques are computed in (22.42). It is necessary to also include the models of actuators (22.5), i.e., to calculate inputs u_i that correspond to computed driving torques (22.42).

The control scheme is presented in Figure 22.16. Obviously, this scheme combines a closed-loop controller with nominal control signals computed on the basis of the equations of motion. In this scheme compensation is provided for time-varying gravitational, centrifugal, and Coriolis forces; the feedback gains are adjusted according to the changes in matrix $H(q)$ of moments of inertia; an acceleration feedforward term is also included to compensate for changes along nominal trajectory.

However, these approaches suffer from several disadvantages. The main problem is that in (22.42) computation of the complete dynamic model of robot is required. For complex robot structures this

© 2002 by CRC Press LLC

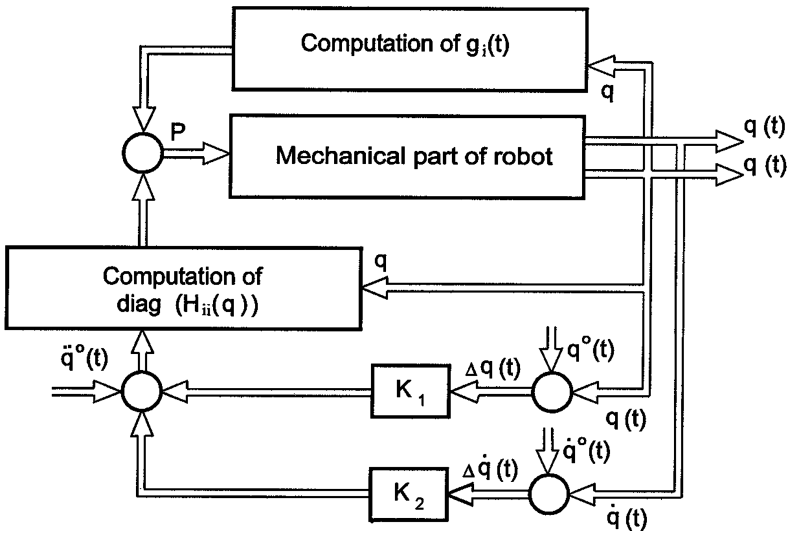


FIGURE 22.17 Control scheme of the inverse problem technique, simplified computation.

requirement may be difficult to satisfy. This is the reason why some authors have tried to implement an approximative model of robot dynamics. They have omitted cross-coupling terms of moment of inertia in the matrix $H(q)$ and centrifugal and Coriolis forces. This means that the model is reduced to diagonal terms in the inertia matrix $H(q)$ and to gravity terms, i.e., the control (22.42) is reduced to:

$$P_i(t) = H_{ii}(q) \cdot [\ddot{q}_i^{\,o}(t) + K_1(q(t) - q^0(t)) + K_2(\dot{q}(t) - \dot{q}^o(t))] + g_i(q(t)) \tag{22.43}$$

where g_i is the gravity force (moment) in the i-th joint. The control scheme is now reduced to the one presented in Figure 22.17. This computation is considerably simplified, but it is still cumbersome for some types of manipulators. On the other hand, it is questionable whether or not the control efficiency is lost by these simplifications. Paul[13] has found that in his experiments with the Stanford manipulator, the contribution by Coriolis and centrifugal terms is relatively insignificant.

Raibert and Horn[16] have used a partial table lookup approach to automatically simplify the computation on a digital computer. Rather than compute the coefficients in (22.42) each time they are needed (every sampling period), their approach (called the configuration space method) is to look them up in a predefined multidimensional memory organized by the positional variables $q(t)$ (the configuration space controller).

Thus, it is obvious that the main difficulty with the inverse problem technique is on-line computation of the dynamic model of the robot. Actually, the analysis of the complexity of the model that is needed for this control law to achieve sufficiently good tracking of nominal trajectory has not been given. However, several other problems with this approach exist. The implementation of control law (22.42) requires perfect knowledge of mechanism parameters. So, it is questionable whether the control (22.42) is robust enough to withstand all parameter variations.

Timofeyev has extended this approach to adaptive control in the case of unknown and variable parameters of the robot.[17]

© 2002 by CRC Press LLC

22.4.4 Effects of Payload Variation and the Notion of Adaptive Control

Up to now we have assumed that all the robot's parameters are constant and precisely known in advance. However, some parameters in rotobtic systems (such are the coefficients of viscous and dry friction in the actuators and joints, backlash, and some actuator cofficients) are not sufficiently defined precisely in advance and can vary during task execution. During the operation of the robot these parameters vary, but rather slowly. Often they do not affect robot performance if the values of the parameters stay with the limits allowed. Thus, it is often unnecessary to include any compensation for variations of these parameter in control law. However, some parameters of the robotic system vary significantly and relatively fast, and have considerable effect upon robot performance. Such parameters are masses, dimensions, and moments of inertia of the payload that is carried by the robot.

The presence of the payload causes changes to the moments of inertia of the mechanism around the axes of the robot joints. In Section 22.3.3 we considered the influence of the mechanism's moment of inertia variation upon servosystem performance. We have seen that when selecting the velocity feedback gain we have include the maximum value of the moment of inertia of the mechanism around the joint in order to prevent overshoots. When the gripper (hand) of the robot grasps the payload, the moments of inertia of the mechanism around the joint axis increase. If during synthesis of servosystems we have not included the mass and the moment of inertia of the payload, the servosystems can become underdamped which causes oscillations in the robotic system. Thus, it is necessary to synthesize the servosystem gains taking into account the maximum payload mass and the moments of inertia that can be carried by the robot. However, if a relatively large variation of the payload is assumed, this can cause uneven performance of the servosystems. For example, when the gripper is empty, or when it carries a payload that is lighter than the maximum assumed payload, the servosystem can become very overdamped ($\xi_i >> 1$) with a resulting slow response.

This problem can be solved by introducing the variable velocity feedback gain. Such adaptive control ensures that for each payload the performance of the servosystems will be nearly equal. Such control requires that the parameters of the payload be known. However, generally, the parameters of the payload are not known, so if the control has to adapt to payload parameter variation it is necessary to ensure identification of these parameters. Identification of the payload parameters can be realized in various ways. The most efficient approach is by direct measuring of the forces at the contact points between the gripper and the payload. In these contact points force transducers are implemented that give direct information on the dynamics of the payload.[11]

Adaptive control can be introduced in various ways. If payload variation with respect to the parameters of the robot links and actuators is relatively small, then it is not often necessary to introduce adaptive control. It can be assumed that servosystems synthesized with constant feedback gains are sufficiently robust to withstand payload parameter variations. However, new robots are appearing that can carry payloads with much larger masses than the masses of the links. Obviously, for such robots the influence of payload variations can cause uneven performance of the servosystems, making it necessary to apply adaptive control. With such robots, elastic effects of the links usually appear. In previous considerations, we have assumed that all the robot's links are rigid. If the masses of the links are small with respect to the mass of the payload, the elastic effects can significantly affect system performance. The control system in these cases must be concerned not only about payload parameter variations, but also about the elastodynamic effects that can complicate control laws and their implementations. However, it should be pointed out that the problems related to computer implementation of dynamic on-line control algorithms are constantly diminishing. Thus, the present-day conclusions about these problems, like those concerning many other technological issues, are only conditionally true.

© 2002 by CRC Press LLC

References

1. Popov, E.P, Vereschagin, F.A., and Zenkevich, S.L., Manipulation robots: *Dynamics and Algorithms* (in Russian), in Scientific Fundamentals of Robotics, Nauka, Moscow, 1978.
2. Albus, J.S., McCain, F.G., and Lumia, R., NASA/NBS Standard Reference Model for Telerobot Control System Architecture (NASREM), NIST Tech. Note 1235, 1989.
3. Albus, J.S., Concept for a reference model architecture for real time intelligent control, ARTICS. NIST/NBS. Tech. Note. TN-1227. U.S.A., 1990.
4. Dornier GmbH, A&R Control Development Methodology Definition Report, Doc. No. CT2/CDR/DO Issue 2.0, September 1992.
5. Vukobratović, K. M., *Applied Dynamics of Manipulation Robots: Modeling, Analysis and Examples*, Springer-Verlag, Berlin, 1989.
6. Vukobratović, K. M., *Introduction to Robotics*, Springer-Verlag, 1989.
7. Paul, P. *Robot Manipulators: Mathematics, Programming and Control*, The MIT Press, 1981.
8. Chestnut, H. and Mayer, R.W. Servomechanisms and Regulating System Design. John Wiley and Sons, New York, 1963.
9. Athans, M. and Falb, P.I., *Optimal Control: An Introduction to the Theory and its Application*, McGraw-Hill, New York, 1966.
10. Vukobratović, K. M. and Stokić, M.D., *Control of Manipulation Robots: Theory and Application*, Springer-Verlag, Berlin, 1982.
11. Vukobratović, K.M., Stokić, M.D., and Kircanski M.N., *Non-adaptive and Adaptive Control of Manipulation Robots*, Springer-Verlag, Berlin, 1985.
12. Vukobratović, K.M. and Stokić, M.D., *Applied Control of Manipulation Robots: Analysis, Synthesis and Exercises*, Springer-Verlag, Berlin, 1989.
13. Paul, R.C., Modeling, Trajectory Calculation, and Servoing of a Computer Controlled Arm, A.I. Memo 177, Stanford Artifical Intelligence Laboratory, Stanford University, September 1972, also in Russian, Nauka, Moscow, 1976.
14. Bejczy, K.A., Robot Arm Dynamic and Control, Technical Memorandum 33-669, Jet Propulsion Laboratory, February, 1974.
15. Pavlov, A.V. and Timofejev, V.A., Calculation and stabilization of programmed motion of a moving robot-manipulator, (in Russian), *Teknicheskaya Kibernetika,* 6, 91–101, 1976.
16. Raibert, H.M. and Horn, P.K.B., Manipulator control using the configuration space method, *The Industrial Robot,* 5, 2, 69–73, 1978.
17. Timofeyev, V.A. and Ekalo, V.Yu., Stability and stabilization of programmed motion of robots: manipulators, (in Russian), *Avtomatikaand Telemechanika,* 10, 148–156, 1976.

© 2002 by CRC Press LLC

23

Control of Robotic Systems in Contact Tasks

23.1 Introduction
23.2 Contact Tasks
23.3 Classification of Robotized Concepts for Constrained Motion Control
23.4 Model of Robot Performing Contact Tasks
23.5 Passive Compliance Methods
Nonadaptable Compliance Methods • Adaptable Compliance Methods
23.6 Active Compliant Motion Control Methods
Impedance Control • Hybrid Position/Force Control • Force/Impedance Control • Position/Force Control of Robots Interacting with Dynamic Environment
23.7 Contact Stability and Transition
23.8 Synthesis of Impedance Control at Higher Control Levels
Compliance C-Frame • Operating Modes • Change of Impedance Gains — Relax Function • Impedance Control Commands • Control Algorithms • Implicit Force Control Integration
23.9 Conclusion

Dragoljub Šurdilović
Fraunhofer Institute

Miomir Vukobratović
Mihajlo Pupin Institute

23.1 Introduction

This chapter reviews the state of the art of the control of compliant motion. It covers early ideas and later improvements, as well as new control concepts and recent trends. A comprehensive review of various compliant motion control methods proposed in the literature would certainly be voluminous, since the research in this area has grown rapidly in recent years. Therefore, for practical reasons, a limited number of the most relevant or representative investigations and methods are discussed. Before we review the results, we categorize compliant motion tasks and proposed control concepts based on various classifying criteria. Particular attention is paid to traditional indices of control performance and to the reliability and applicability of algorithms and control schemes in industrial robotic systems.

23.2 Contact Tasks

Robotic applications can be categorized in two classes based on the nature of interaction between a robot and its environment. The first one covers *noncontact*, e.g., unconstrained, motion in a free

© 2002 by CRC Press LLC

space, without environmental influence exerted on the robot. The robot's dynamics have a crucial influence upon its performance of noncontact tasks. A limited number of frequently performed simple robotic tasks such as pick-and-place, spray painting, gluing, and welding, belong to this group.

In contrast, many advanced robotic applications such as assembly and machining require the manipulator to be mechanically coupled to the other objects. In principle, two basic *contact task* subclasses can be distinguished. The first one includes *essential force tasks* whose nature requires the end effector to establish physical contact with the environment and exert a process-specific force. In general, these tasks require the positions of the end effector and the interaction force to be simultaneously controlled. Typical examples of such tasks are machining processes such as grinding, deburring, polishing, and bending. Force is an inherent part of the process and plays a decisive role in task fulfillment (e.g., metal cutting or plastic deformation). In order to prevent overloading or damage to the tool during operation, this force must be controlled in accordance with some definite task requirements.

The prime emphasis within the second subclass lies on the requirement for end effector motion near the constrained surfaces (*compliant motion*). A typical representative task is the part mating process. The problem of controlling the robot during these tasks is, in principle, the problem of accurate positioning. However, due to imperfections inherent in the process and the sensing and control system, these tasks are inevitably accompanied by contact with constrained surfaces, which produces reaction forces. The measurement of interaction force provides useful information for error detection and allows appropriate modification of the prescribed robot motion.

Future research will certainly develop more tasks for which interaction with the environment will be fundamental. Recent medical robot applications (e.g., spine surgery, neurosurgical and microsurgical operations, and knee and hip joint replacements) may also be considered *essential contact tasks*. Comprehensive research programs in automated construction, agriculture, and food industry focus on the robotization of other types of contact tasks such as underground excavation and meat deboning.

Common to all contact tasks is the presence of the constraints upon robot motion due to environmental objects. If all parameters of the environment and robot were known and robot positioning was precise, it might be possible to accomplish the majority of these tasks using the same control strategies and techniques developed for the control of robot motion in free space. However, none of these conditions can be fulfilled in reality. Hence, contact tasks are characterized by the dynamic interaction between robot and environment, which often cannot be predicted accurately. The magnitude of the mechanical work exchanged between the robot and the environment during contact may vary drastically and cause significant alteration of performance of the robotic control system. Therefore, for successful completion of contact tasks, the interaction forces have to be monitored and controlled, or control concepts ensuring the robot interacts compliantly with the environment must be applied.

Compliance, i.e., accommodation,[1] can be considered a measure of the ability of a manipulator to react to interaction forces. This term refers to a variety of control methods in which the end effector motion is modified by contact forces.

23.3 Classification of Robotized Concepts for Constrained Motion Control

The previous classification of elementary robotic tasks provides a framework for the further systematization of compliant motion control. Recently, the problems encountered in the control of compliant motion have been extensively investigated and several control strategies and schemes have been proposed. These methods can be systematized according to different criteria. The primary systematization requires considering the kind of compliance. According to this criterion, two basic groups of control concepts for compliant motion are distinguishable (Figure 23.1):

© 2002 by CRC Press LLC

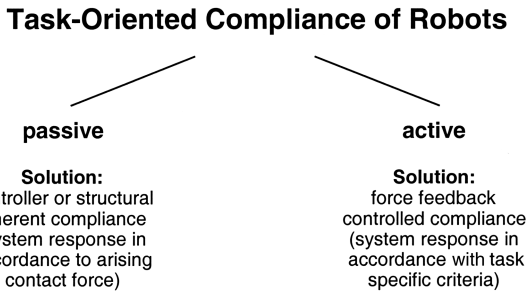


FIGURE 23.1 Basic classification of robot compliance.

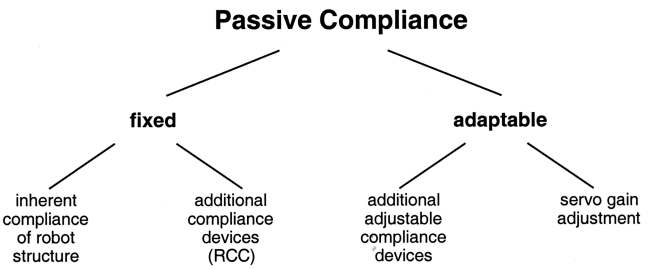


FIGURE 23.2 Passive compliance classification.

1. *Passive compliance,* whereby the position of the end effector is accommodated by the contact forces due to compliance inherent in the manipulator structure, servos, or special compliant devices.
2. *Active compliance,* whereby the compliance is provided by constructing a force feedback in order to achieve a programmable robot reaction, either by controlling interaction force* or by generating task-specific compliance at the robot end point.

Regarding the possibility of adjusting system compliance to specific process requirements, passive compliance methods can be categorized as *adaptable* and *nonadaptable*. Based on the dominant sources of compliance, two methods within these groups can be distinguished (Figure 23.2):

1. Fixed (or nonadaptable) passive compliance:
   a. Methods based on the inherent compliance of the robot's mechanical structure, such as elasticity of the arm, joints, and end effectors.[2]
   b. Methods that use specially constructed passive deformable structures attached near the end effectors and designed for particular problems. The best known is the remote center compliance (RCC) element.[3]
2. Adaptable passive compliance:
   a. Methods based on devices with tunable compliance.[4]
   b. Compliance achieved by the adjustment of joint servo-gains.[5]

The basic classification of *active compliance* control methods is based on the classifying tasks as *essential* or *potential*. Using the terminology of bond–graph formalisms, robot behavior that performs *essential contact tasks* can be generalized as a source of *effort* (force) that should raise

*By *force* we mean *force and torque* and, accordingly, *position* should be interpreted as *position and orientation.*

© 2002 by CRC Press LLC

a *flow* (motion) reaction by environmental objects. The robot behavior associated with the second or *potential* subclass corresponds to *impedance,* characterized by the reaction of robot's motion on external forces exerted by the environment. The active control force method can be classified into two groups:

1. *Force*, i.e., *position/force control* or *admittance control*, whereby both desired interaction force and robot position are controlled. A desired force trajectory is commanded and force measurements are required to realize feedback control.
2. *Impedance control,*[6] uses the different relationships between acting forces and manipulator position to adjust the mechanical impedance of the end effector to external forces. Impedance control can be defined as allowing interaction forces to govern the error between the nominal and actual positions of the end effector according to the target impedance law. Impedance control is based on position control and requires position commands and measurements to close the feedback loop. Force measurements are needed to effect the target impedance behavior.

Position/force control methods can be divided into two categories:

1. *Hybrid position/force control*, whereby position and force are controlled in a nonconflicting way in two orthogonal subspaces defined in a task-specific frame (*compliance* or *constraint frame*). For force-controlled end-effector degrees of freedom (DOF), the contact force is essential for performing the task. The motion is most important in position DOF. Force is commanded and controlled along directions constrained by the environment, while position is controlled in directions in which the manipulator is free to move (*unconstrained*). *Hybrid control* is usually referred to as the method of Raibert and Craig.[7] However, according to Mason's[1] definition, this term is used in a more general sense and is defined as any controller based on the division into force and position controlled directions.
2. *Unified position/force control,* which differs essentially from the above conventional hybrid control schemes.Vukobratović and Ekalo[8,33] have established a dynamic approach to simultaneously control both the position and force in an environment with completely dynamic reactions. The approach of dynamic interction control[8,33] defines two control subtasks responsible for stabilization of robot position and interaction force. Both control subtasks utilize a dynamic model of the robot and the environment in order to ensure the tracking of the nominal motion and the force.
3. *Parallel position-force control,*[9] is based on the appropriate tuning of the position and force controllers. The force loop is designed to dominate the position control loop along constrained task directions where a force interaction is expected. From this viewpoint, the parallel control can be considered as impedance/force control.

Taking into account the way in which the force information is included in the forward control path, the following position/force control schemes can further be distinguished:

1. *Explicit* or *force-based*[7,10,11] whereby force control signals (i.e., the difference between the desired and actual forces) are used to generate the torque inputs for the actuators in the joints.
2. *Implicit* or *position-based* algorithms[12,13] whereby the force control error is converted to an appropriate motion adjustment in force-controlled directions and then added to the positional control loop.

Impedance control methods can also be distinguished by the way the robotic mechanism is treated: either as an actuator (i.e., source) of position or as an actuator of a force. The aim in impedance control is to provide specific relationships between effort and motion rather than follow a prescribed force trajectory as in the case of force control. Considering the arrangement of position and force sensor and control signals within control loops (inner or outer), the following two common approaches to provide task-specific impedance via feedback control can be distinguished:[14]

© 2002 by CRC Press LLC

1. *Position mode* or *outer loop control,* whereby a target impedance control block relating the force exerted on the end effector and its relative position is added within an additional control loop around the position-controlled manipulator. An inner loop is closed based on the position sensor and an outer loop is closed around it based on the force sensor.[15,16]
2. *Force mode* or *inner loop* control, whereby position is measured and force command is computed to satisfy target impedance objectives.[14]

Regarding the force–motion relationship or the impedance order, impedance control schemes can be further categorized into: *stiffness control,*[17] *damping control*,[18] and *general impedance control,*[19,20] using zeroth, first, and second order impedance models respectively.

There are additional criteria that allow further classification of active compliant motion control concepts. For example, we can distinguish the methods with respect to the *source of force information* (with or without direct interaction force sensing), and the *allocation of force sensor* (wrist, torque sensor in joints, force-sensing pedestal, force sensor at the contact surface, sensors at robot links, fingers, etc.). To avoid the problems associated with noncollocation between measurement of contact forces and actuation in robot joints, which can cause instability,[21] the use of redundant force information combining joint force sensing with one of the above force sensing principles was proposed.

Regarding the space in which the active force control is performed in, one can distinguish between two methods:

1. *Operational space control* techniques where control takes place in the same frame in which actions are specified.[22,23] This approach requires the construction of a model describing the system dynamic behavior as perceived at the end effector where the task is specified (operational point, i.e., coordinate frame). The traditional approach for specifying compliant motion uses a *task* or *compliance frame* approach.[1] This geometrical approach introduces a Cartesian-compliant frame with orthogonal force and position (velocity)-controlled directions. To overcome the limitations of this approach, new methods were proposed.[24,25] These approaches, referred to as *explicit task specification of compliant motion*, are based on the model of the constraint topology for every contact configuration and utilize projective geometry metrics to define a hybrid contact task.
2. *Joint space control*, whereby control objectives and actions are mapped into joint space.[26] Associated with this control approach are transformations of action attributes, compliance, and contact forces from the task into the joint space.

Further, considering control issues, such as variations of control parameters (gains) during execution, one can distinguish:

1. *Nonadaptive* active compliance control algorithms that use fixed gains assuming small variations in the robot *and* environment parameters
2. *Adaptive control*, which can adapt the variation of process[27,28]
3. *Robust control* approaches, which maintain model imprecision and parametric uncertainties within specified bounds[29,30]

Depending on the extent to which system dynamics is involved in the applied control laws, it is possible to further distinguish:

1. *Nondynamic*, i.e., *kinematic model-based* algorithms, such as *hybrid control*,[7] *stiffness control,*[17] etc., which approximate the contact problem considering its static aspects only.
2. *Dynamic model-based* control schemes, such as *resolved acceleration control,*[31] *dynamic hybrid control,*[11] *constrained robot control,*[32] and *dynamic force-position control in contact with dynamic environment,*[8,33] based on complete dynamic models of the robot and the environment that take into account all dynamic interactions between position- and force-controlled directions.

© 2002 by CRC Press LLC

Although contact motion is characterized by relatively low velocities, high dynamic interaction (i.e., exchange of energy) between a robot and its environment affects the control system significantly and can jeopardize the stability of the control system.[34] Consequently, the role of both dynamics, namely dynamics of the robot[35] and dynamics of the environment,[8,33] in the control of compliant motion, becomes essential. Kinematic algorithms are mostly based on Jacobian matrix computation, while the complexity of the dynamic methods is much greater.[36]

The seminal hybrid control method proposed by Raibert and Craig[7] essentially provides a quasistatic approach to compliance control based on an idealized simple geometric model of a constrained motion task (Mason's *constraint frame formalism*). With hybrid control, the dynamics of both robot and environment (dynamic interaction) is neglected. The *dynamic hybrid control*[11] and *constrained motion control*[32] approaches consider the constraints upon robot motion in the form of algebraic equations defining a hyper surface. These methods take the robot dynamic model and the model of the environment into account in order to synthesize dynamic control laws to ensure admissible robot motion with the constraint and achieve desired interaction forces. Generalization of the constrained motion problem leads to introducing active dynamic contact forces (dynamic environment), also described by differential equations. In a dynamic environment, the interaction forces are not compensated by constraint reactions; they produce active work on the environment. Obviously, contact with a dynamic environment requires consideration of the complete system dynamics involving robot and interaction models to obtain admissible robot motion and interaction forces. The "pure dynamic" interaction without passive reaction was considered by Vukobratović and Ekalo in papers dedicated to the dynamic control of robots interacting with the dynamic environment.[8,33,89–91] A suitable model structure has been proposed by De Luca and Manes[37] that handles a most general case in which purely kinematic constraints on the robot end-effector live together with the dynamic interactions.

Although very inclusive, the above classification cannot encompass all of the proposed concepts to date. Some approaches combine two or more methods categorized in distinct groups, and attempt to use the benefits of both to offset disadvantages of single solution strategies. Such methods use compliant motion control approaches that combine force and impedance control.[12,38] Some methods integrate control mechanical system design.[39] This approach is based on *micro–macro* manipulator structures that provide inherently stable and well-suited subsystems for high bandwidth active force control.

The terminology used above represents, in some measure, a trade-off among different nomenclatures used in the literature. Mason[1] designates the control concepts by specifying the linear relation between effector force and position as *explicit feedback*, while Whitney[6] uses the phrase *explicit control* to refer to techniques having a desired force input other than position or velocity input. The classification and the terminology reflect, in our opinion, the essential aspects of appropriate control strategies. The above classification is summarized in Figures 23.1 through 23.3.

23.4 Model of Robot Performing Contact Tasks

During the execution of a contact task, the kinematic structure of the robot changes from an open to a closed chain. Contact with the environment imposes kinematic and dynamic constraints on the motion of the end effector. One of the most difficult aspects of dynamic modeling concerns the interactions of bodies in contact. We will briefly consider simplified models of constrained motion to be used for the analysis of contact motion control concepts.

In order to form a mathematical model that describes the dynamics of the closed configuration manipulator, let us consider an open robot structure whose last link (end effector) is subjected to a generalized external force (Figure 23.4). A dynamics model of rigid manipulation robot interacting with the environment is described by the vector differential equation in the form:

$$\mathbf{H}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{h}(\mathbf{q}, \dot{\mathbf{q}}) + \mathbf{g}(\mathbf{q}) = \boldsymbol{\tau}_a + \mathbf{J}^T(\mathbf{q})\mathbf{F} \tag{23.1}$$

© 2002 by CRC Press LLC

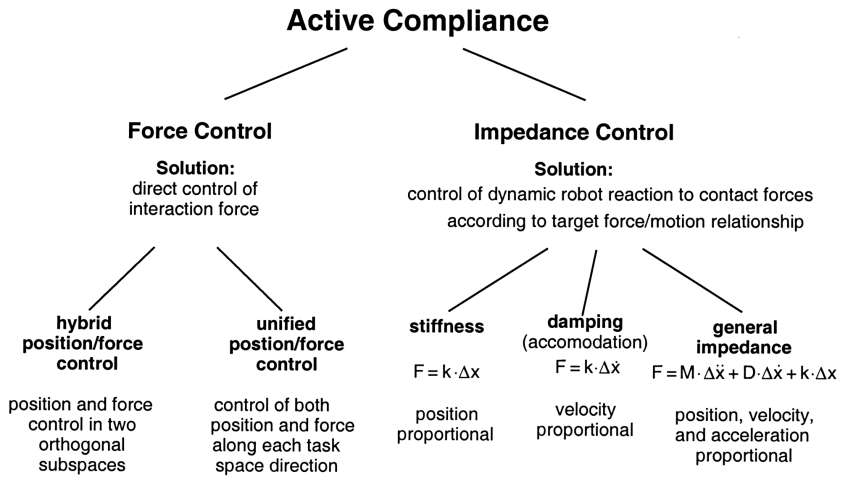


FIGURE 23.3 Active compliance control methods.

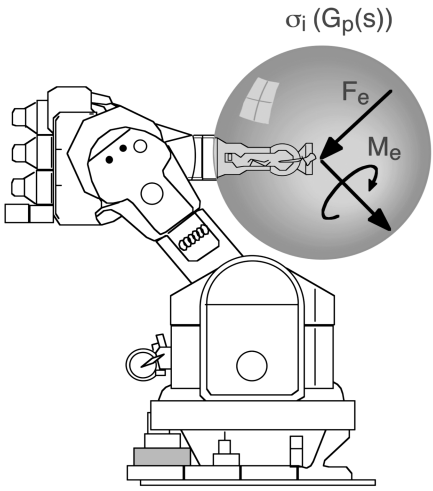


FIGURE 23.4 Open kinematic chain exposed to an external force action.

where $\mathbf{q} = \mathbf{q}(t)$ is an n-dimensional vector of robot generalized coordinates; $\mathbf{H}(\mathrm{q})$ is an $n \times n$ positive definite matrix of inertia moments of the manipulator mechanism; $\mathbf{h}(\mathrm{q}, \dot{\mathrm{q}})$ is an n-dimensional nonlinear function of centrifugal and Coriolis moments; $\mathbf{g}(\mathrm{q})$ is a vector of gravitational moments; $\boldsymbol{\tau}_a = \boldsymbol{\tau}_a(t)$ is an n-dimensional vector of generalized joint axes driving torques; $\mathbf{J}(\mathrm{q})$ is an $n \times m$ Jacobian matrix relating joint space velocity to task space velocity; and $\mathbf{F} = \mathbf{F}(t)$ is an m-dimensional vector of external forces and moments acting on the end effector.

The dynamic model of the actuator (we confine discussion to robot manipulators driven by DC motors) that drive the robot joints must be added to the above equations. It is convenient to adopt this model in linear form. Taking into account that electric time constants of DC motors driving almost all commercial robotic systems are very low, we shall adopt a second order model of actuators:

$$n_i^2 I_{mi} \ddot{q}_{mi} + n_i^2 b_{mi} \dot{q}_{mi} + \tau_{ai} = n_i \tau_{mi} \tag{23.2}$$

where q_{mi} is the output angle of the motor shaft after-reducer; n_i is the gear ratio; I_{mi} is the inertia of the motor actuator; b_{mi} is the viscous friction coefficient; τ_{mi} is the control input to the i-th

© 2002 by CRC Press LLC

actuator (i.e., motor torque); and where i denotes the local i-th subsystem. The torque produced by the motor is proportional to the armature current, that is:

$$\tau_{mi} = k_{mi} i_{mi} \tag{23.3}$$

where k_{mi} is the torque constant. If we assume the stiffness in the joints (gears) to be infinite, the relation between the coordinate of the mechanism q_i coincides with the actuator coordinate q_{mi}.

The dynamic models of the actuators and mechanical parts of the robot are related by joint torques (loads). If we substitute τ_{ai} from (23.2) into (23.1) we get the entire model of the robotic mechanism in joint coordinate space:

$$\overline{\mathbf{H}}(\mathrm{q})\ddot{\mathbf{q}} + \mathbf{B}_m \dot{\mathbf{q}} + \mathbf{h}(\mathrm{q}, \dot{\mathrm{q}}) + \mathbf{g}(\mathrm{q}) = \boldsymbol{\tau}_{\mathrm{q}} + \mathbf{J}(\mathrm{q})^T \mathbf{F} \tag{23.4}$$

where:

$$\overline{\mathbf{H}}(\mathrm{q}) = \mathbf{H}(\mathrm{q}) + \mathbf{I}_m = \mathbf{H}(\mathrm{q}) + diag(n_i^2 I_{m_i})$$

$$\mathbf{B}_m = diag(n_i^2 b_{mi}) \tag{23.5}$$

and $\boldsymbol{\tau}_q$ is 6×1 vector of input torques at the joint shaft (after-reducer):

$$\boldsymbol{\tau}_{\mathbf{q}} = \left[n_1 \tau_{m1} .... n_6 \tau_{m6} \right]^T.$$

The above dynamical model can be transformed into an equivalent form that is more convenient for analysis and synthesis of a robot controller for contact tasks. When the manipulator interacts with the environment, it is convenient to describe its dynamics in the space where manipulation task is described, rather than in joint coordinate space (also termed *configuration space*). The end effector position and orientation with respect to a reference coordinate system can be described by a six-dimensional vector **x**. The reference system is chosen to suit a particular robot application. Most frequently, a fixed coordinate frame attached to the manipulator base is considered as the reference system. Using the Jacobian matrix, we can transform the dynamic models (23.4) from the joint into the end effector coordinate system:

$$\boldsymbol{\Lambda}(x)\ddot{\mathbf{x}} + \mathbf{B}(x)\dot{\mathbf{x}} + \boldsymbol{\mu}(x, \dot{x}) + \mathbf{p}(x) = \boldsymbol{\tau} + \mathbf{F} \tag{23.6}$$

where relationships among corresponding matrices and vectors from Equations (23.1) and (23.6) are given by the following equations:

$$\begin{aligned}
\boldsymbol{\Lambda}(x) &= \mathbf{J}(q)^{-T} \mathbf{H}(\mathbf{q}) \mathbf{J}(q)^{-1} \\
\mathbf{B}(x) &= \mathbf{J}(\mathbf{q})^{-T} \mathbf{B}_{\mathbf{m}} \mathbf{J}(\mathbf{q})^{-1} \\
\boldsymbol{\mu}(x, \dot{x}) &= \mathbf{J}(q)^{-T} \mathbf{h}(q, \dot{q}) - \boldsymbol{\Lambda}(x) \dot{\mathbf{J}}(q, \dot{q}) \dot{\mathbf{q}} \\
\mathbf{p}(x) &= \mathbf{J}(q)^{-T} \mathbf{g}(\mathrm{q}) \\
\boldsymbol{\tau} &= \mathbf{J}(q)^{-T} \boldsymbol{\tau}_{\mathbf{q}}
\end{aligned} \tag{23.7}$$

© 2002 by CRC Press LLC

The description, analysis, and control of manipulator systems with respect to the dynamic characteristics of their end effectors are referred to as the *operational space formulation*.[22] Analogous to the joint space quantities, $\mathbf{\Lambda}(x)$ is the operational space inertia matrix, $\boldsymbol{\mu}(x,\dot{x})$ is the vector of Coriolis and centrifugal forces, $\mathbf{p}(x)$ is the vector of gravity terms, and $\boldsymbol{\tau}$ is the applied input control force in the operational space.

The interaction force is influenced by robot motion and also by the nature of the environment. Since mechanical interaction is generally very complex and difficult describe mathematically, we are compelled to introduce certain simplifications and thus partly idealize the problem. In practice, the interaction force **F** is commonly modeled as a function of the robot dynamics, i.e., end-effector motion (position, velocity, and acceleration) and control input:

$$\mathbf{F} = \mathbf{F}\left(\mathbf{x},\ \dot{\mathbf{x}},\ \ddot{\mathbf{x}}, \boldsymbol{\tau}, \mathbf{d}, \mathbf{d}_e\right) \tag{23.8}$$

where $\mathbf{d}$ and $\mathbf{d}_e$ denote sets of robot and environment model parameters, respectively. The following general work environment models have been mostly applied in the literature for describing constrained motion: *rigid hypersurface*, *dynamic environment*, and *compliant environment*.

In contact with a rigid hypersurface, robot motion (i.e., surface penetration) is prevented in the direction orthogonal to the surface. For maintaining the constraint, only an infinitesimal displacement in the tangential hyperplane is allowed. Different models describing robot constrained motion on a rigid hypersurface have been presented in Yoshikawa et al.[11] and McClamroch and Wang.[32] These models can be applied for simulation or control design, i.e., computation of control laws ensuring the robot remains on the constraint manifold. However, the complexity of these models is great. In the special case of a rigid plane, model decomposition is relatively simple and does not require that computations are repeated for every step. In general, however, computing and integrating these models involves extensive computations and solutions of numerical problems.

If the environment does not possess displacements (DOFs) independent of the robot motion, the mathematical model of the environment dynamics in the frame of robot coordinates can be described by nonlinear differential equations:[8]

$$\mathbf{M}(\mathrm{q})\ddot{\mathbf{q}} + \mathbf{L}(\mathrm{q},\dot{\mathrm{q}}) = -\mathbf{S}^{\mathbf{T}}(\mathrm{q})\mathbf{F} \tag{23.9}$$

where $\mathbf{M}(\mathrm{q})$ is a nonsingular $n \times n$ matrix; $\mathbf{L}(\mathrm{q},\dot{\mathrm{q}})$ is a nonlinear n-dimensional vector function; and $\mathbf{S}^{\mathbf{T}}(\mathrm{q})$ is an $n \times n$ matrix with rank equal to n. The system (23.4-23.9) then describes the dynamics of robot interaction with dynamic environment. We assume that all the mentioned matrices and vectors are continuous functions of the arguments for the contact cases.

In operational space, the model of a pure dynamic environment has the form:[40]

$$\boldsymbol{M}(\mathrm{x})\ddot{\mathbf{x}} + \mathbf{l}(\mathrm{x},\dot{\mathrm{x}}) = -\mathbf{F}$$

$$\boldsymbol{M}(\mathrm{x}) = \mathbf{J}(q)^{-T}\mathbf{M}(q)\mathbf{J}(q)^{-1}$$

$$\mathbf{l}(\mathrm{x},\dot{\mathrm{x}}) = \mathbf{J}(q)^{-T}\mathbf{L}(q,\dot{q}) - \boldsymbol{M}(\mathrm{x})\dot{\mathbf{J}}(q)\dot{\mathbf{q}}$$

In effect, a general environment model involves geometrical (kinematic) constraints plus dynamic constraints.[37] An example of such a dynamic environment is when a robot is turning a crank or sliding a drawer. Dynamics is relevant for the robot motion and cannot be neglected. However, the dynamic model of kinematic–dynamic constraints is rather complex and its computation involves several difficulties. The crucial problem is the decomposition of DOFs, i.e., force and independent coordinate parameterization, which is not unique from a mathematical viewpoint. Although in several elemental contact cases, the feasible model parameterization is

© 2002 by CRC Press LLC

obvious,[37] it is difficult to perform model parameterization in many practical contact tasks. Planning and computing tools supporting automatic minimal parameterization of a dynamic constrained motion problem based on task specification do not exist. Moreover, the differentiation of constraint equations can lead to unstable numerical solutions, causing constraint violation in real-time simulations. By introducing inaccuracy in the robot and environment (e.g., for robust control design purposes), the problem becomes even more complicated.

For control design purposes, it is customary to utilize a linearized model of manipulator and environment. The applicability of a linearized model in constrained motion control design, especially in industrial robotic systems, was demonstrated in Goldenberg[41] and Šurdilović.[42] Neglecting nonlinear Coriolis and centrifugal effects due to relatively low operating velocities (rate linearization) during contact, and assuming the gravitational effect to be ideally compensated for, we obtain a linearized model around a nominal trajectory in Cartesian space $\mathbf{x}_0$ in the form:

$$\mathbf{\Lambda}(x_0)\ddot{\mathbf{x}} + \mathbf{B}(x_0)\dot{\mathbf{x}} = \boldsymbol{\tau}(\mathbf{x_0}) + \mathbf{F} \tag{23.10}$$

In passive linear environments, it is convenient to adopt the relationship between forces and motion around the contact point in the form (*linear elastic environment*):

$$-\mathbf{F} = \mathbf{M}_e\ddot{\mathbf{p}} + \mathbf{B}_e\dot{\mathbf{p}} + \mathbf{K}_e\mathbf{p} \tag{23.11}$$

where $\mathbf{p}$ denotes the end effector *penetration* through the surface defined by $\mathbf{p} = \mathbf{x} - \mathbf{x}_e$, $\mathbf{x}_e$ represents contact point locations, and $\mathbf{M}_e$, $\mathbf{B}_e$, and $\mathbf{K}_e$ are inertial, damping, and stiffness matrices, respectively.

23.5 Passive Compliance Methods

According to the classifications presented above, we first review the compliant control methods based on passive accommodation (with no actuator involved). *Passive compliance* is a concept often used to overcome the problems arising from positional and angular misalignments between the manipulator and its working environment.

23.5.1 Nonadaptable Compliance Methods

The passive compliance method, which is based on *inherent robot structural elasticity,* is more interesting as a theoretical solution than a feasible approach. This method assumes that the compliance of the mechanical structure has a determining effect on the compliance of the entire system. However, this assumption is opposite to the real performance of commercial robotic systems which are designed to achieve high positioning accuracy. Elastic properties of the arms are insignificant. The dominant influence on a somewhat larger deflexion of the manipulator tip position is, in some cases, joint compliance, e.g., due to reducer elasticity (harmonic drive) or compressibility of the hydraulic actuator.[43] In practice, the mechanical compliance of the robotic structure can be utilized for contact tasks purposes under very restricted conditions. The endpoint compliance is often unknown and too complex to be modeled. Due to high stiffness levels, the accommodation range within an acceptable contact force level is usually extremely small and without any practical values. This method does not offer any possibility to adapt system compliance to the various task requirements. The idea of utilizing flexible manipulator arms as an instrumented compliant system[2] is relatively new and poses additional problems due to complex modeling and controlling of elastic robots.

The method based on *mechanical compliance devices*, in principle, also utilizes structural compliance. The most influential source of multi-axis compliance in this case, however, is a specially constructed device whose behavior is known and sufficiently repeatable. Relatively good performances have been achieved, especially in the robotic assembly. Different types of such devices

© 2002 by CRC Press LLC

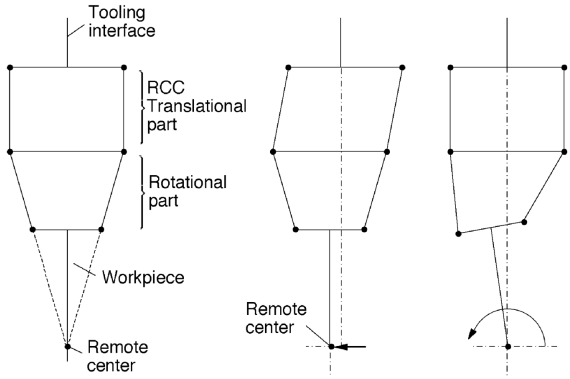


FIGURE 23.5 Remote center compliance (RCC).

have been developed; the best known is the RCC (remote center compliance)[3] developed in the Charles Stark Draper Laboratory. RCC is designed to make the workpiece rotate around a defined center of compliance. The compliance center is a point at which application of a force causes only translation, while a torque applied around an axis through this point will cause rotation of the workpiece (Figure 23.5). A crucial feature of the RCC is that it consists of translational and rotational parts; this combination allows lateral and angular errors to be accommodated independently.

RCC elements provide a simple and effective solution that permits fast and easy interfacing of mechanical parts in spite of initial positioning errors. The main advantage is that a simple positional controller can be applied, without any additional force sensor feedback or complex calculations. However, an RCC element cannot be applied to tasks involving parts of lengths and weights. A solution to this problem may be to design a set of compliance adapters that can be changed according to the needs of specific tasks.

Instrumented Remote Center Compliance (IRCC)[44] represents an improvement of RCC which provides the fast error absorption characteristic of RCC and the measurement characteristic of a multi-DOF sensor. Contact force and deformation data can be used for task monitoring, calibration, contour following, or positioning feedback.

23.5.2 Adaptable Compliance Methods

Further development of RCC has led to *adaptable compliant devices*[4] which enable the location of the center of compliance to be automatically controlled to a prescribed extent in accordance with parts of different lengths and weights. These devices are usually also instrumented to provide information about end point deflections for robot control.

The *controller gain adjustment method* is based on the compliance of the robotic controller and attempts to provide a universally programmable passive compliance at endpoints, by the relatively simple adjustment of servo gains. The basic principle is to tune the positional servo gains to make the robot behave as a linear six-dimensional spring in Cartesian space with programmable stiffness. Therefore, taking into account the relationship between forces exerted upon the robot and its reaction (stiffness-like behavior), the gain adjustment method was considered equivalent to the impedance (i.e., stiffness) control.

The choice of Cartesian stiffness matrix is strongly dependent on the task specification. In the case of part mating, for example, the elements of the stiffness matrix that relate force and motion in the direction of insertion should be estimated sufficiently high so that axial force does not cause the insertion to stop. Conversely, in lateral directions, the corresponding elements should be sufficiently low to enable the peg to move easily as it encounters the chamfer. A strategy for systematic setting of Cartesian stiffness in different phases of peg/hole assembly is proposed by Simons and Van Brussel.[5]

© 2002 by CRC Press LLC

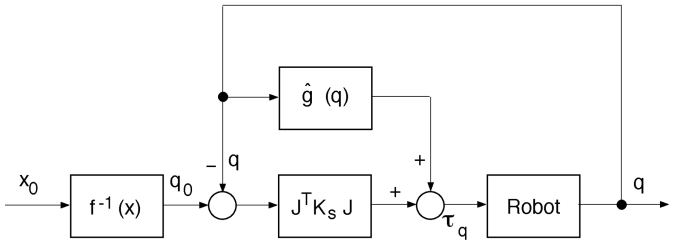


FIGURE 23.6 Passive gain adjustment scheme.

The basic gain adjustment control scheme is sketched in (Figure 23.6), where $\mathbf{x}^0$ and $\mathbf{q}^0$ are nominal Cartesian and joint position vectors, respectively; $\mathbf{f}^{-1}(\mathbf{x})$ denotes the inverse kinematic transformation; $\mathbf{q}$ is the actual joint position; and $\hat{\mathbf{g}}(q)$ is the computed gravitational torque. The control torque $\boldsymbol{\tau}_q$ is obtained according to:

$$\boldsymbol{\tau}_q = \mathbf{k}_p(\mathbf{q}_0 - \mathbf{q}) + \hat{\mathbf{g}}(q) \tag{23.12}$$

where $\mathbf{k}_p$ represents the joint stiffness matrix which should be tuned to ensure the arm will behave with the desired stiffness $\mathbf{K}_S$. The relationship between the joint and Cartesian stiffness matrices is given by:

$$\begin{aligned} \mathbf{k}_p &= \mathbf{J}^T \mathbf{K}_S \mathbf{J} \\ \mathbf{K}_S &= \mathbf{J}^{-T} \mathbf{k}_p \mathbf{J}^{-1} \end{aligned} \tag{23.13}$$

where $\mathbf{J}(q)$ represents Jacobian matrix-relating velocities (i.e., forces) between a Cartesian frame attached at the compliance center and the joint coordinate space. At the center of compliance, the Cartesian stiffness matrix is diagonal, but corresponding joint stiffness $\mathbf{k}_P$ is, according to Equation (23.13), a fully symmetric matrix. This means that the joint stiffness matrix is highly coupled and a position error in one joint will affect the commanded torque in all other joints.

Equation (23.12) represents the central formulation of active gain adjustment methods. Assuming the static (gravitational) forces are exactly compensated for and dynamic forces due to slow displacements are negligible, it is relatively easy to prove that the linearized robot-and-environment model is always stable. Control adjustment allows us to adopt the location of center of compliance (by the aid of Jacobian matrix $\mathbf{J}$) and Cartesian stiffness (choosing $\mathbf{K}_S$). However, although this stiffness-like behavior could be theoretically adjusted on-line while running a task, we have classified this method as passive compliance, because the compliant motion is performed in a purely passive way by the action of external forces, rather than by force feedback as with active stiffness control.

While the adaptable passive compliance method provides a simple and flexible solution for many compliant motion tasks (without requirements for force sensing and feedback), the aim of having the entire robot structure behave loosely in some directions is difficult to achieve. This concept is coupled with several problems. Most contemporary robotic systems cannot accurately achieve the desired spring-like behavior. Several nonlinearities such as friction and backlash in mechanical transmission and process frictional phenomena like jamming can destroy the stiffness force/position causality. Furthermore, by setting very low control gains in some directions, the entire system is made more sensitive to perturbations. Different disturbances and nonlinearities can affect performance, and that can be extremely dangerous in some environments. Since integral control action is not applied, all static effects such as gravitation must be completely compensated for.

All these factors make the performance of this control approach uncertain, thus imposing the need to introduce additional sensor information to monitor task execution. Relevant improvements

© 2002 by CRC Press LLC

can be achieved by including force sensor information in a rule-based assembly strategy,[5] or by introducing an internal force feedback loop.[45] However, the simplicity of passive gain adjustment is lost when these additional strategies are applied. An equivalent improvement in performance can be achieved by applying a simple active force control concept.

The principle of adaptable control gains is more suitable for direct drive, multifingered, or wrist hands. This method appears similar to those described above, which use special adaptable compliant devices.

If the passive gain adjustment concept is used in industrial practice, one should consider that conventional robotic systems are *nonbackdriveable* due to high gear ratios and Coulomb friction/stiction effects in joints. (The order of equivalent friction force in Cartesian space is about 10^2 N.) Hence, although compliant control is applied, a force exerted at the end effector will not cause a corresponding detectable displacement in joints. Therefore, the method can be applied only in manipulation tasks that permit large interaction forces. Due to relatively high costs and low robustness of force sensors, though, there is increased interest on the part of industrial robot manufacturers in appling this method in specific tasks such as handling of castings (e.g., the new *soft servo* or *soft float* industrial robot control functions).

23.6 Active Compliant Motion Control Methods

The active compliance control methods best utilize reprogrammability of manipulation robots. This is done by representing the manipulation robots' main characteristic, that is, their ability to switch from one production task to another.

23.6.1 Impedance Control

Whitney first reported use of force feedback control of a manipulator for impedance control.[6] Impedance control is a fundamental approach toward allowing a stiff industrial robot to interact with the environment. Impedance control is mainly directed to contact tasks for which the control of interaction force is not essential for successful task execution. These contact tasks, such as insert, require a specific motion of the workpiece that adheres to external constraints in the presence of possible contact with the environment (constrained or compliant robot motion).

These compliant motion tasks require solution of motion control problems. The objective of the impedance control is to reduce contact impedance or stiffness of the position-controlled robot. This is done by controlling dynamic reaction to the external contact forces (robot compliance) to compensate for uncertainties and tolerances in the robot–environment location, while maintaining acceptable force magnitudes. The interaction force between a robot and a fixed environment depends on motion and target impedance. Under certain circumstances, impedance control may also be applied to produce a desired force.

An impedance control task is specified in terms of desired motion trajectory and relationships between position error and interaction force exerted at the end effector. To ensure successful accomplishment of a constrained motion task, the commonly stiff robot position control behavior must be replaced with a compliant *target impedance model.*

The objective of impedance control differs from the conventional control goals in the sense that the main control issue is not to ensure tracking of a reference input signal (e.g., nominal position or force). The aim is to produce a reference target model (*target impedance*) specifying the interaction of robot and environment, i.e., the desired relationship between acting forces and robot motion reaction (position error). A conventional control system is usually analyzed for its ability to track standard input signals (e.g., step, ramp) within the allowed time. The main impedance control performance specification, however, addresses the capability of achieving the target model.

The impedance control problem can be defined as designing a controller so that interaction forces govern the error between desired and actual positions of the end effector. The control input

© 2002 by CRC Press LLC

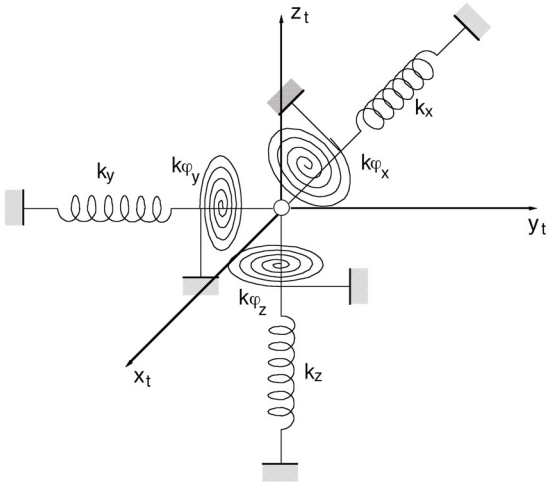


FIGURE 23.7 Target stiffness model in C frame.

describing a desired target impedance relation may, in principle, have an arbitrary functional form, but it is commonly adopted in the linear second order differential equation form describing the simple six-dimensional decoupled mass–spring–damper mechanical system. The reason is that the dynamics of a second order system is well understood. Lee and Lee[46] developed a control algorithm referred to as generalized impedance control by introducing a higher order impedance relation between position and force errors, which includes force derivatives.

In other words, impedance control is a general approach to contact task control in which the robot behaves as a mass–spring–dashpot system whose parameters can be specified arbitrarily. This can be achieved by feedback control using position and force sensing. The following control objective should be obtained:

$$\mathbf{F} = \mathbf{M}_t(\ddot{\mathbf{x}} - \ddot{\mathbf{x}}_0) + \mathbf{B}_t(\dot{\mathbf{x}} - \dot{\mathbf{x}}_0) + \mathbf{K}_t(\mathbf{x} - \mathbf{x}_0) = \mathbf{M}_t\ddot{\mathbf{e}} + \mathbf{B}_t\dot{\mathbf{e}} + \mathbf{K}_t\mathbf{e} \tag{23.14}$$

or in the s domain:

$$\mathbf{F}(s) = \mathbf{Z}_t(s)(\mathbf{x} - \mathbf{x}_0) = \mathbf{Z}_t(s)\mathbf{e} = (\mathbf{M}_t s^2 + \mathbf{B}_t s + \mathbf{K}_t)(\mathbf{x} - \mathbf{x}_0) \tag{23.15}$$

where $\mathbf{Z}_t(s) = \mathbf{M}_t s^2 + \mathbf{B}_t s + \mathbf{K}_t$ is the *target robot impedance* in Cartesian space, $\mathbf{x}_0$ describes the desired position trajectory, $\mathbf{x}$ is the actual position vector, $\mathbf{e}$ is the position control error, $\mathbf{F}$ is the external force exerted upon the robot, and $\mathbf{M}_t$, $\mathbf{B}_t$, and $\mathbf{K}_t$ are positive definite matrices that define target impedance, where $\mathbf{K}_t$ is the stiffness matrix, $\mathbf{B}_t$ is the damping matrix, and $\mathbf{M}_t$ is the inertia matrix. The diagonal elements of these target model matrices describe the desired robot mechanical behavior during contact.

One of the most common approaches for representation of robot and object positions is based on coordinate frames. It is convenient to describe the robot impedance reaction to external forces with respect to a frame, referred to as a *compliance* or *C frame*. Along each C frame direction, the target model describes a mechanical system presented in (Figure 23.7) with the programmable impedance (mechanical parameters); for simplicity, only spring elements are depicted. The model describes a virtual spatial system consisting of mutually independent spatial mass–damper–spring subsystems in six Cartesian directions. A corresponding decoupled physical system is difficult to

© 2002 by CRC Press LLC

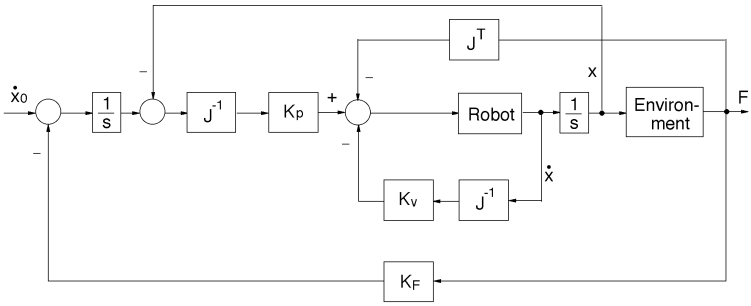


FIGURE 23.8 Damping control.

realize (for example, by combining Cartesian linear axes and Cardan frames). Appropriate selection of target impedance parameters along specific axes is required to achieve active impedance control.

The target impedance matrices can be selected to correspond to various objectives of the given manipulation task.[14] Obviously, high levels of stiffness are required in the directions where the environment is compliant and positioning accuracy is important. Low stiffness can be selected in directions where small interaction forces must be maintained. Large $\mathbf{B}_t$ values are specified when energy must be dissipated, and $\mathbf{M}_t$ is used to provide smooth transient system response during contact.

To assess how well a designed impedance controller meets the above control objective, it is customary to specify performance criteria. A reasonable measure to express the performance of the impedance control is the difference between the target model and real system behavior described by robot motion and interaction forces.[47] Depending on which of these physical values is used to characterize the system behavior (force or position), the *impedance control error* can be expressed by means of *force measure* (*force model error*):

$$\mathbf{e}_f = \mathbf{M}_t(\ddot{\mathbf{x}} - \ddot{\mathbf{x}}_0) + \mathbf{B}_t(\dot{\mathbf{x}} - \dot{\mathbf{x}}_0) + \mathbf{K}_t(\mathbf{x} - \mathbf{x}_0) - \mathbf{F} \tag{23.16}$$

or by *position measure* (*position model error*):

$$\mathbf{e}_p = \mathbf{x} - \mathbf{x}_0 - \delta\mathbf{x}_f \tag{23.17}$$

where the target position deviation is obtained as the solution of the target model differential equation:

$$\mathbf{F} = \mathbf{M}_t\delta\ddot{\mathbf{x}}_f + \mathbf{B}_t\delta\dot{\mathbf{x}}_f + \mathbf{K}_t\delta\mathbf{x}_f \tag{23.18}$$

for the initial conditions: $\mathbf{F}(t_0) = 0;\ \delta\mathbf{x}_f(t_0) = \delta\mathbf{x}_0$.

The computing of the model errors requires both force and the robot position to be measured.

The above defined control goal can be achieved using various control strategies. Impedance control represents a strategy for constrained motion rather than a concrete control scheme. Various control concepts and schemes were established for controlling the relation between robot motion and interaction force.

One of the first approaches to impedance control was proposed by Whitney[18] (Figure 23.8). In this approach known as *damping* or *accommodation control*, the force feedback is closed around the velocity control loop. The interaction force is converted into a velocity modification command by a constant damping coefficient K_F. Using a simplified example of discrete time force control, Whitney defined the condition for system stability during contact as:

$$0 < T\,K_f\,K_e < 1 \tag{23.19}$$

© 2002 by CRC Press LLC

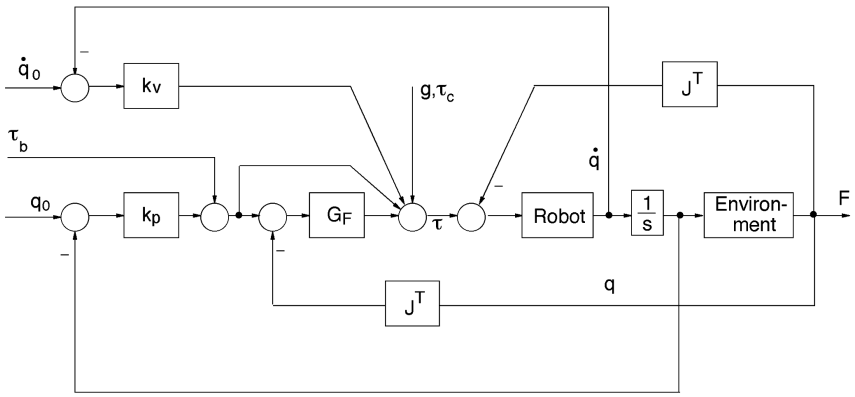


FIGURE 23.9 Stiffness control.

where T is the sampling period, K_f is the force control gain (damping coefficient), and K_e is the stiffness of the environment. This condition implies that if K_e is high, the product TK_f must be small. To avoid large contact forces, a very high sampling rate, i.e., small T is required. Alternatively, for contact with a very stiff object Whitney proposed introduction of a passive compliance in order to achieve the equivalent environmental stiffness K_e smaller (including the stiffness of the robot structure, environment, sensor, etc.).

Salisbury[17] proposed modification of the end effector position in accordance with the interaction force (Figure 23.9). This concept is based on a *generalized stiffness* formulation $\mathbf{F} = \mathbf{K}\delta\mathbf{x}$ where $\delta\mathbf{x}$ is a generalized displacement from a nominal commanded end effector position, and $\mathbf{K}$ is a six-dimensional stiffness matrix. Based on the difference between the desired and actual end positions, a nominal force is computed and converted into joint torques using the transpose of the Jacobian matrix. This force is then used to determine the torque error on each joint that is further used to correct applied torque so that the desired force (i.e., stiffness) is maintained at the robot hand. The requirements of the stiffness matrix elements and their designs for specific tasks are considered in Whitney.[6]

These impedance control schemes are simple and relatively easy to implement. However, the achieved closed loop impedance behavior in the Cartesian space depends on robot configuration. Obviously, to replace the nonlinear dynamic model with the linear time-invariant target system (e.g., mass–damper–spring system) generally requires the control law to compensate for relevant system nonlinearities (*model-based dynamic control*).

The most common impedance control concept was established by Hogan[19] who defined a unified theoretical framework for understanding the mechanical interactions between physical systems. This approach focuses on the characterization and control of dynamic interaction based on manipulator behavior modification. In this sense, impedance control is an augmentation of position control. The actions of the manipulator control and hardware and the interaction between a robot and its environment are described by network analysis. The important issue is that the command and control of a vector such as position or force is not enough to control the interaction between systems (dynamic networks). The controller must also be able to command and control a relationship between system variables. The proposed control design strategy is to adapt the robot behavior to become the inverse of the environment. This means that if the environment behaves like admittance, the impedance control should be applied and vice versa.

23.6.1.1 Force-Based Impedance Control

Most of the impedance control algorithms utilize the computed torque method to cancel nonlinearity in robot dynamics in order to achieve linear target impedance behavior. This popular approach requires computation of a complete dynamic model of constrained motion, which make its realization rather

© 2002 by CRC Press LLC

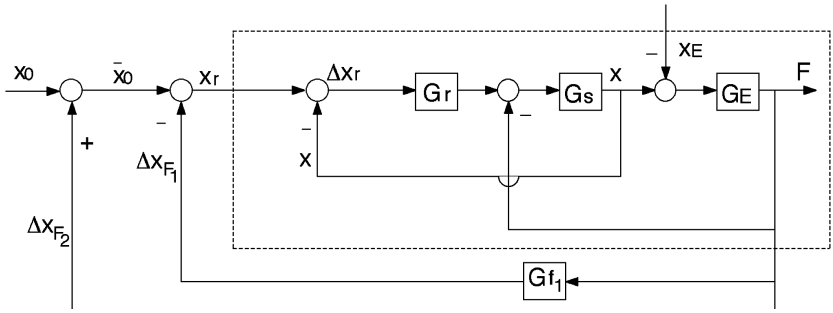


FIGURE 23.10 Force-based dynamic impedance control.

complex. An important drawback of this approach is sensitivity to model uncertainties and parameter variations. Performance improvements that can be achieved with algorithms in industrial robotics are not in proportion to implementation efforts.

Hogan[48] proposed several techniques with and without force feedback for modulating the end point impedance of a general nonlinear manipulator. Assuming the Cartesian dynamic model perfectly matches the real system, Hogan proposed the following nonlinear control law:

$$\boldsymbol{\tau} = \hat{\Lambda}\mathbf{M}_t^{-1}\left[\mathbf{K}_t(\mathbf{x}_0 - \mathbf{x}) - \mathbf{B}_t\dot{\mathbf{x}} + \mathbf{F}\right] + \hat{\boldsymbol{\mu}}(x,\dot{x}) + \hat{\mathbf{p}}(x) - \mathbf{F} \tag{23.20}$$

be applied to reach a reasonable target impedance behavior in the ideal case in the form:

$$\mathbf{F} = \mathbf{M}_t\ddot{\mathbf{x}} + \mathbf{B}_t\dot{\mathbf{x}} + \mathbf{K}_t(\mathbf{x} - \mathbf{x}_0). \tag{23.21}$$

The control scheme corresponding to the above control law is sketched in (Figure 23.10). A distinction is made in the figure between the active force exerted by the robot ($\overline{\mathbf{F}}$) and the reactive external force (F), which can be computed assuming a simple spring-like environmental model:

$$\overline{\mathbf{F}} = \mathbf{K}_e\left(\mathbf{x} - \mathbf{x}_e\right) = -\mathbf{F} \tag{23.22}$$

where $\mathbf{K}_e$ is the stiffness of the environment. This control law essentially represents a nonlinear control algorithm that combines the *inverse control technique*[49] (also known as *computed torque method* and *nonlinear decoupling*) and force-based (inner loop) impedance control. In force-based impedance control algorithms (Figure 23.10), an expected reference force is computed to satisfy the desired impedance specification based on position error and target impedance $\mathbf{F}_0(s) = \mathbf{Z}_t(s)\left(\mathbf{x}_0 - \mathbf{x}\right)$. The expected active force $\mathbf{F}_0$ is compared with the actual force sensed by the force sensor and a force error is computed. This error is further multiplied with inertia matrices $\hat{\Lambda}\mathbf{M}_t^{-1}$. Finally, the product is summed with dynamic compensation terms (Coriolis and gravitation vectors) and feed-forward force $\overline{\mathbf{F}}$ to obtain Cartesian control force, which is further transferred into the robot joint via the transposed Jacobian $\mathbf{J}^T$ to get the actuator torque control inputs. It is relatively easy to prove that the control law:

$$\boldsymbol{\tau} = \hat{\Lambda}\left\{\ddot{\mathbf{x}}_0 + \mathbf{M}_t^{-1}\left[\mathbf{K}_t(\mathbf{x}_0 - \mathbf{x}) + \mathbf{B}_t(\dot{\mathbf{x}}_0 - \dot{\mathbf{x}}) + \mathbf{F}\right]\right\} + \hat{\boldsymbol{\mu}}(x,\dot{x}) + \hat{\mathbf{p}}(x) - \mathbf{F} \tag{23.23}$$

realizes the impedance control behavior specified in Equation (23.15).

The reason impedance control methods based on force control input cannot be suitably applied in commercial robotic system lies in the fact that commercial robots are designed as positioning

© 2002 by CRC Press LLC

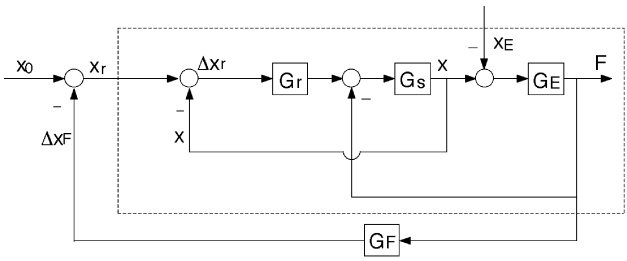


FIGURE 23.11 Position model-error impedance control.

devices. In the above methods, the driving torque vector ensuring the desired target impedance behavior has been computed and then multiplied by the transpose of the Jacobian matrix in order to be realized around the actuated robot joints. However, the realization of computed torque is not accurate in commercial robotic systems because the local servos are position controlled and there is no force feedback with respect to the torques around the joints. Consequently, the realization of desired torques is poor, since high friction and other nonlinearities in the transmission mechanisms contribute significantly to the inaccuracy of current/torque causality. Because of these difficulties, the implementation of force-based impedance control can be successfully performed only by a new generation of direct-drive robots[50] with accurate joint torque controls. Force-based impedance control requires a completely new control system.

23.6.1.2 Position Based Impedance Control

As mentioned above, force-based impedance control is mainly intended for robotic systems with relatively good causality between joint and end effector forces, such as direct-drive manipulators. In commercial robots, the effects of nonlinear friction in transmission systems with high gear ratios significantly destroy this causality. Therefore, in commercial robotic systems, it is feasible to implement only the position-mode impedance control by closing a force-sensing loop around position controller. Position-based impedance control is most reliable and suitable for implementation in industrial robot control systems since no modification of a conventional positional controller is required.

Two basic impedance control schemes with internal position controls can be distinguished.[51] The first scheme is sketched in Figure 23.11. An inner position control loop is closed based on position sensing; it is surrounded by a closed outer loop based on force sensing. The force loop is naturally closed when the end effector encounters the environment. The outer loop includes a force feedback compensator $\mathbf{G}_F$, basically representing admittance since its role is to shape the relation between contact force and corresponding nominal position modifications $\Delta\mathbf{x}_f$. This block is imposed on the system to regulate the force response to the commanded and actual motions according to the target admittance $\mathbf{Z}_t^{-1}$.

Other control blocks in Figure 23.11 represent a common industrial robot position control system involving the following transfer function matrices: $\mathbf{G}_r$, position control regulator; $\mathbf{G}_s$, robot plant; and $\mathbf{G}_e$, environment. The position correction $\Delta\mathbf{x}_F$ is subtracted from the nominal position $\mathbf{x}_0$ and the command input vector for the positional controller, referred to as *reference position* $\mathbf{x}_r$, is computed. A good tracking of the reference position must be achieved by the internal position controller. Assuming $\mathbf{G}_F = \mathbf{Z}_t^{-1}$, the position error input to the position controller $\Delta\mathbf{x}_r$ becomes:

$$\Delta\mathbf{x}_r = \mathbf{x}_r - \mathbf{x} = \mathbf{x}_0 - \Delta\mathbf{x}_F - \mathbf{x} = \mathbf{x}_0 - \mathbf{x} - \mathbf{Z}_t^{-1}\mathbf{F} = \mathbf{e}_p . \tag{23.24}$$

This means that the control system in Figure 23.11 utilizes the position-related impedance model error $\mathbf{e}_p$ (23.17) to achieve target impedance behavior. The impedance model error $\mathbf{e}_p$ is fed forward to the position controller $\mathbf{G}_r$ in order to be nullified within internal position control loop. Since the

© 2002 by CRC Press LLC

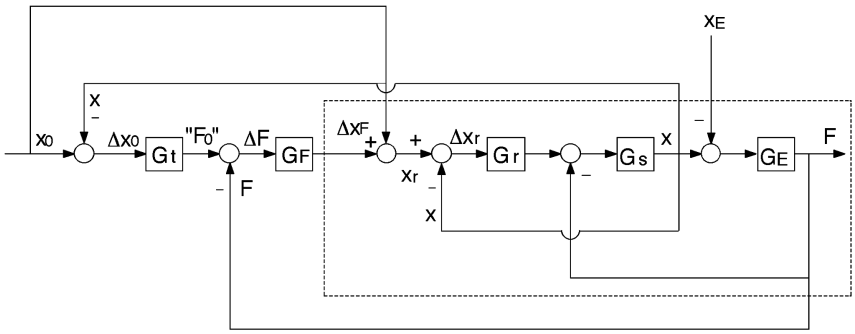


FIGURE 23.12 Force model error-based impedance control.

purpose of the system in (Figure 23.11) is to control position, it will be referred to as *position impedance model error control*.

The second position-based impedance control structure is depicted in (Figure 23.12). This scheme provides a generalization of the original scheme proposed by Maples and Becker[15] and is referred to as *outer/inner loop stiffness control*. The control scheme consists of two parallel feedback loops superimposed to the internal position control and closed using measurements from both the wrist force sensor and position sensors. Analyzing the control scheme, it can be seen that the position error $\mathbf{e} = \Delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}$ is multiplied by the task-specific target impedance $\mathbf{G}_t(s) = \mathbf{Z}_t(s)$ to provide a *nominal* (reference) *force* $\mathbf{F}_0$, which corresponds to the target impedance behavior on the output. The tracking of this force is realized by the next feedback loop closed on the sensed force $\mathbf{F}$. In the ideal case, we have $\mathbf{F}_0 = \mathbf{F}$, describing the target behavior. Thus Figure 23.12 basically represents a force control system with target impedance added to regulate the motion response to the interaction force. Following the control flow, we see that the force error in this control scheme corresponds to the previously defined force impedance model error (23.17):

$$\Delta\mathbf{F} = \mathbf{F}_0 - \mathbf{F} = \mathbf{G}_t\left(\mathbf{x}_0 - \mathbf{x}\right) - \mathbf{F} = \mathbf{e}_f \tag{23.25}$$

Therefore, we will refer to the control system in Figure 23.12 as *force model error impedance control*. Similarly, to the previous system (Figure 23.11), the model error Equation (23.25) is further relayed to the internal control part in order to regulate this error to zero as time increases. However, different from the position model error control in Figure 23.11, where the position model error is eliminated by the internal position control, in the control system in Figure 23.12, the regulation of the model error is realized by means of the compensator $\mathbf{G}_F$. In order to retain the internal position control loop, the implicit force control structure is implemented by passing the force error $\Delta\mathbf{F}$ through the admittance filter $\mathbf{G}_F$, providing nominal path modification $\Delta\mathbf{x}_F$. The position correction is further added to the Cartesian nominal position $\mathbf{x}_0$, and via reference position $\mathbf{x}_r$ feeds forward to the position servo. Obviously, to achieve $\Delta\mathbf{F} = \mathbf{e}_f \to 0$ as $t \to \infty$ which ensures a steady state position deviation $\left(\mathbf{x}_0 - \mathbf{x}\right)_\infty = \mathbf{e}_\infty$ corresponding to the target impedance (stiffness) model, the regulator $\mathbf{G}_F$ has to involve an integral control term.

This scheme was originally developed as a position-based realization of Salisbury's stiffness control algorithm.[17] In this seminal work,[15] $\mathbf{G}_t$ block was a diagonal stiffness matrix that allowed the user to specify compliance along Cartesian directions, while compensator $\mathbf{G}_F$ was realized as a pure integrator ensuring desired stiffness steady state.

Both control approaches utilize similar concepts to produce the target impedance model by reducing the impedance model errors $\mathbf{e}_p$ and $\mathbf{e}_f$ to zero. Each approach has specific advantages and disadvantages.[51] The $\mathbf{e}_p$-based scheme (Figure 23.11) is simpler and easier to implement. Under some circumstances, this scheme allows different target impedances to be realized by setting the

© 2002 by CRC Press LLC

compensator $\mathbf{G}_F$ to the target admittance, while the position controller undertakes feedback control. This is similar to an open loop target impedance control. Conversely, in the *force model error control* scheme (Figure 23.12), the target impedance is specified in the outer loop using $\mathbf{G}_t$ block, while the role of the internal loop compensator $\mathbf{G}_F$ is to ensure the tracking of the selected model using force feedback. The internal position control loop is retained to achieve robust position-based, i.e., implicit, force control and control robot motion in the free space. This scheme offers more possibilities to adjust the system contact behavior by choosing $\mathbf{G}_t$ and tuning $\mathbf{G}_F$. However, the opportunity to arbitrarily select the target model and dynamically maintain the force/motion relationship is limited by the complex structure of this scheme.

The main problem with the $\mathbf{e}_f$-based scheme lies in the transition to and from contact (constrained motion). The external impedance loop in this scheme is closed even in the free space when the contact force is zero, and thus affects position control performance. Although the magnitude of the position deviation can be insignificant, considering that the stiffness of the position control is essentially greater than the target one and the inner position loop is faster than the external impedance loops, this effect is not desirable in practice. The compensator $\mathbf{G}_F$ has to be tuned to achieve the required control goal in the presence of a stiff environment, e.g., a large amount of damping to ensure a stable transition. However, that is contrary to the position control performance needed in the free space. In the $\mathbf{e}_p$-based scheme (Figure 23.11), the force feedback loop is closed naturally by physical contact and interaction force sensing. In the free space, only the forward position control is active.

To avoid this shortcoming of the $\mathbf{e}_f$-based impedance control manifested by deviations of position control performance in the free space through impedance control blocks $\mathbf{G}_t$ and $\mathbf{G}_f$ (Figure 23.12), the outer part of the control scheme providing the position modification $\Delta\mathbf{x}_F$ can be deactivated in the free space and activated only on contact with the environment (control switching, variable structure control). The contact state can be observed using force sensor information and a force threshold, which should prevail over noise effects in the force sensing (e.g., offsets, high frequency oscillations, gripper inertial forces during robot motion, etc.). Generally, however, the switching algorithms are not easy to implement. This causes the force model error control scheme to be even more difficult to integrate into today's industrial controllers. Moreover, in conjunction with control delays, the change of control structure can cause undesirable chattering in the contact task, which will lead to contact and system instability. Thus, the design of a stable impedance controller becomes a complex undertaking with this scheme.

The $\mathbf{e}_p$*-based* control scheme (Figure 23.11) was recently implemented in the new SPARCO space control system[52] developed based on industrial robot standards. Its impedance control is completely integrated at several levels including servo control, virtual force sensor (data processing, filtering, calibration), motion planning, language supports, and monitoring functions. The SPARCO control servo scheme involves an improved position-based control law. The impedance control design problem is split into two subproblems: realization of target impedance model, and choice of target impedance parameters to achieve stable interaction with the environment and required performance. The compensator $\mathbf{G}_F$ that produces the target impedance $\mathbf{G}_t$ is obtained from the following relations (Figure 23.11):

$$\begin{aligned} \mathbf{G}_s^{-1}(s)\mathbf{x} &= \mathbf{G}_r(s)\left[\mathbf{x}_r - \mathbf{x}\right] - \mathbf{F} \\ \mathbf{x}_r &= \mathbf{x}_0 - \Delta\mathbf{x}_f \\ \Delta\mathbf{x}_F &= \mathbf{G}_F(s)\mathbf{F} \end{aligned} \tag{23.26}$$

Substituting:

$$\mathbf{x} - \mathbf{x}_0 = \mathbf{G}_t^{-1}(s)\mathbf{F} \tag{23.27}$$

© 2002 by CRC Press LLC

in (23.26), we derive the expression for the position modification $\Delta\mathbf{x}_f$ which ensures the realization of the target model in the form:

$$\Delta\mathbf{x}_F = \mathbf{G}_p^{-1}(s)\left[\left(\mathbf{G}_t^{-1}(s) - \mathbf{S}_p(s)\mathbf{G}_s(s)\right)\mathbf{F} - \mathbf{S}_p(s)\mathbf{x}_0\right] \tag{23.28}$$

where $\mathbf{S}_p(s)$ is the sensitivity transfer function matrix $\mathbf{S}_p(s) = \mathbf{I} - \mathbf{G}_p(s)$. This control law involves the impedance compensator:

$$\mathbf{G}_F(s) = \mathbf{G}_p^{-1}(s)\left(\mathbf{G}_t^{-1}(s) - \mathbf{S}_p(s)\mathbf{G}_s(s)\right) = \mathbf{G}_p^{-1}(s)\mathbf{G}_t^{-1}(s) - \mathbf{G}_r^{-1}(s) \tag{23.29}$$

and an additional nominal position feed forward term:

$$\mathbf{G}_p^{-1}(s)\mathbf{S}_p(s)\mathbf{x}_0 = \mathbf{G}_r^{-1}(s)\mathbf{G}_s^{-1}(s)\mathbf{x}_0 \tag{23.30}$$

In the linearized robot control system, this control law provides equivalent effect as the computed torque-based impedance control (Equation 23.23). Essentially, the main issue is to compensate for dynamic effects in the forward position control in order to achieve the given target model, which is similar to the nonlinear control (Equation 23.23) goal. The difference is that control law defined in Equation (23.29) is based on linearized compensation techniques, which are less complex than computation of nonlinear robot dynamics. However, the impedance compensator (Equation 23.29) includes the inverse of position controller $\mathbf{G}_r^{-1}(s)$ and the position control closed loop system matrix $\mathbf{G}_p^{-1}(s)$. Generally these matrices depend on robot configuration. Moreover, using the inverse compensators is not well suited in practice, since inverse systems produce large control signals, amplify high frequency noise, and may introduce unstable pole zero cancellations.

However, as demonstrated in Šurdilović,[53] these shortcomings do not appear in industrial robots. The performance of commercial industrial robotic systems allows significant simplification of impedance control design and implementation. The robustness of internal position control allows the disturbances due to interaction force and joint friction effects to be neglected. In other words, the term $\mathbf{G}_r^{-1}(s)$ from Equation 23.29 can be omitted, since the internal position controller (Figure 23.11) significantly reduces the interaction force disturbance effects. Furthermore, due to high gear ratios and accurate design of joint position controllers, the closed loop position control transfer matrix $\mathbf{G}_p(s)$ is *normal*, *diagonally dominant,* and *spatially rounded* with good approximation. In other words, it exhibits similar performance independent of Cartesian directions, and compliance frame selection achieves similar performance in a large workspace area (Figure 23.4).

Necessary conditions to ensure the *spatial roundness* and *diagonal dominance* of convenient position control systems of industrial robots are derived in Šurdilović.[53] In the majority of industrial robot systems, diagonal dominance is achieved by high transmission ratios in joints, causing constant rotor inertia to prevail over variable inertia of the robot arm. The spatial roundness in the joint and Cartesian space is achieved by uniform tuning of local axis position controllers. This characteristic is illustrated in Figure 23.4 by the spherical form of the principal gain space of the closed loop position control transfer matrix $\mathbf{G}_p(s)$. These characteristics are important in decentralized position control in order to ensure robust and uniform performance in Cartesian space. They allow impedance control to be implemented simply, using the constant compensator $\mathbf{G}_F$.

In spite of implementation of inverse compensators, we can require that $\mathbf{G}_p^{-1}(s)$ show inverse characteristics only over some finite frequency range. To obtain a proper compensator, we can employ a low pass filter (by inserting more poles), or utilize the low pass performance of the target admittance $\mathbf{G}_t^{-1}(s)$. Moreover, assuming that the nominal motion exhibits slow acceleration/deceleration in the vicinity of constraints and during contact, which is a reliable premise due to unknown

© 2002 by CRC Press LLC

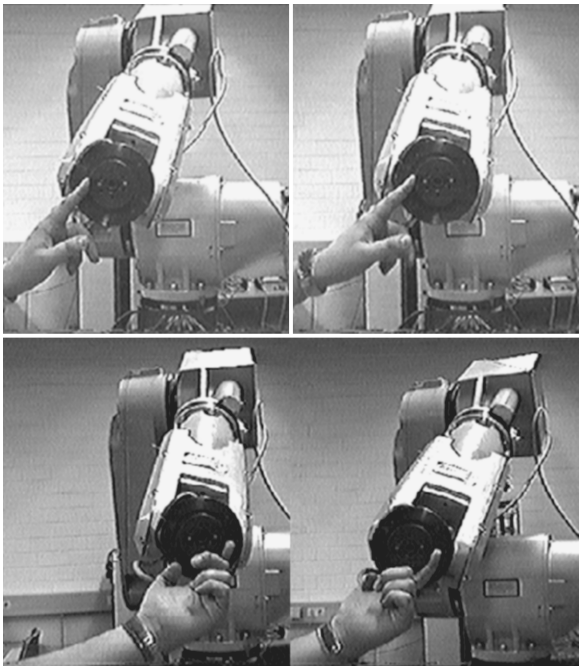


FIGURE 23.13 Target model realization experiment.

constraints, we can also neglect the feed forward term (Equation 23.30) and thus substantially simplify the control law:

$$\mathbf{G}_F(s) = \hat{\mathbf{G}}_p^{-1}(s)\mathbf{G}_t^{-1}(s) \tag{23.31}$$

where $\mathbf{G}_t$ is the diagonal target end effector impedance matrix specifying the target behavior in each compliance frame direction corresponding to Equation (23.14) and $\hat{\mathbf{G}}_p$ is the diagonal estimate of the closed loop position transfer matrix, i.e., the estimation of its dominant diagonal part. The controller (Equation 23.31) practically consists of a diagonal and, for a given task, constant compensator. The above control law provides the following nominal closed loop contact behavior:

$$\mathbf{x} = \mathbf{G}_p(s)\mathbf{x}_0 - \mathbf{G}_t^{-1}(s)\mathbf{F} \tag{23.32}$$

In other words, the controller (Equation 23.31) accurately realizes the desired target model in the industrial robot control system. It is obvious that the role of this controller is to shape the sensitivity transfer functions, i.e., the relationship between external interaction force disturbance and the position tracking error according to the desired target impedance model (Equation 23.14), without influencing the nominal position control performance in the free space. Only the sensitivity transfer function to the interaction force sensed by the force sensor and used in the external control loop is modified by the impedance control. The impedance controller does not influence the robust and good perturbation rejection properties of the position controller toward other disturbance effects, such as friction.

A typical result of a target model realization experiment (Figure 23.13) by the control law (Equation 23.31) with the industrial Manutec r3 robot is presented in Figure 23.14. Obviously, a very good match of model and experimental contact forces was achieved. The bandwidth of the position-based impedance controller is theoretically limited by the bandwidth of the internal position

© 2002 by CRC Press LLC

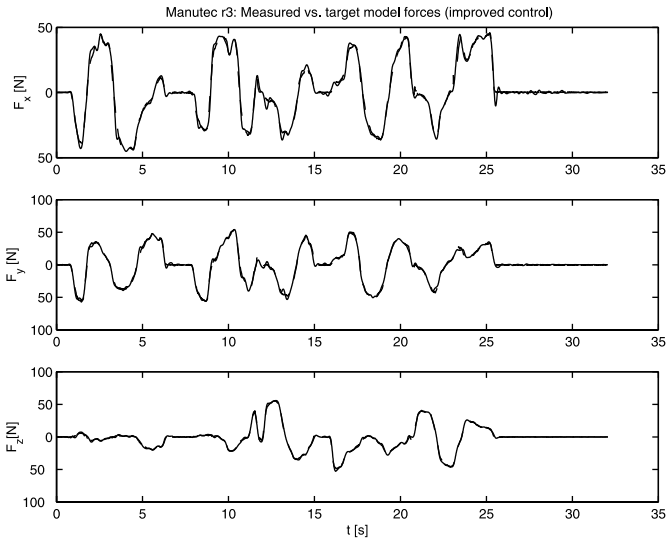


FIGURE 23.14 Target model (solid) and measured (dashed) forces (improved law).

controller (commonly about 10 Hz). However, in practice, impedance controller bandwidth up to 5 Hz is reliable.

The main advantage of the *position model error* scheme over the *force model* scheme, lies in its reliability and simpler design and implementation. The achieved system behavior is easy to understand. Furthermore, taking into account the reliable performance of the industrial robot position control, a sufficiently accurate and robust desired impedance behavior can be achieved with this scheme.

The position-based impedance approach in general suffers from its inability to provide soft impedance due to limits in the accuracy of the position control system and sensor resolution. This approach is mainly suitable for applications that require high position accuracy in some Cartesian directions, which is accomplished by stiff and robust joint control. Design and implementation of this scheme is simple and does not require complex computations.

The force (i.e., torque)-based approach is better suited to providing small impedance (stiffness and damping) while reducing the contact force. From a computational viewpoint, this approach is reasonable for applications where manipulator gravity is small and slow motion is required. In other cases, manipulator modeling details (i.e., complete dynamic models) are needed. Contrary to the position-based impedance control, the force-based control is mainly intended for robotic systems with relatively good causality between joint torques and end effector forces, such as direct drive manipulators.

23.6.1.3 Other Impedance Control Approaches

Considerable research efforts addressed the development of *adaptive impedance control algorithms*. Daneshmend et al.[27] proposed a model reference adaptive control scheme with Whitney's damping control loop. Several authors have pursued Craig's adaptive inverse dynamic control algorithms[54] and expanded its application to contact motion. Lu and Goldenberg[47] proposed a sliding mode-based control law for impedance control. The proposed controller consists of two parts: a nominal dynamic model to compensate for nonlinearities in robot dynamics, and a compensator ensuring the impedance error (i.e., the difference between nominal target model and the actual impedance) proceeds asymptotically to zero on the sliding surface. In order to cope with the chattering effects in the variable structure sliding mode control, a continuous switching algorithm in a small region around sliding surface is proposed. Al-Jarah and Zheng[55] proposed an interesting adaptive impedance control algorithm intended to minimize the interaction force between manipulator and environment.

© 2002 by CRC Press LLC

Dawson et al.[30] developed a robust position/force control algorithm based on the impedance approach. The control scheme consists of two blocks: a desired trajectory generator computing the modified command position based on the target impedance model and using the nominal position and force measurements, and a controller involving a PD regulator and robust control part. The purpose of the robust controller is to ensure that the control tracking error (i.e., the difference between target and actual robot impedance) proceeds asymptotically to zero in spite of model uncertainties within specified bounds. Robust control design is currently one of the most challenging topics in controlling contact tasks.

Under some circumstances, the impedance control can be applied to achieve desired contact forces. When an impedance-controlled manipulator is in contact with the environment, the interaction force is completely determined by the input position, target impedance, and the model (impedance) of the environment. It is then apparent from Equations (23.14-15) that the interaction forces can be precisely controlled using the impedance approach as long as an exact model of the environment and the robot is available. By using the force-based approach in this case, the desired force can be achieved in the open loop, and a force sensor is not needed. Such an approach is very similar to the passive gain adjustment.

In general, however, it is difficult to exactly know the location and impedance of the environment and robotic system. If the stiffness of the environment is much greater than the stiffness of the target impedance and the robot, the force can also be controlled in a desired accuracy range by using only the impedance model, rather than only knowledge about the environment.[51] When these conditions are not fulfilled, i.e., stiffness of the environment is not much greater than that of the target impedance, it is necessary to perform estimation experiments to obtain the model of the environment and control the contact force. However, the on-line estimation of the environment is complex and coupled with several practical problems: uncertain robot motion sensing at low velocities, noise, disturbances due to friction and vibrations, impact, etc., that can significantly influence the results. Using the robot to acquire the data for an off-line estimation is risky in principle, and in tasks with variable environment, virtually impossible.

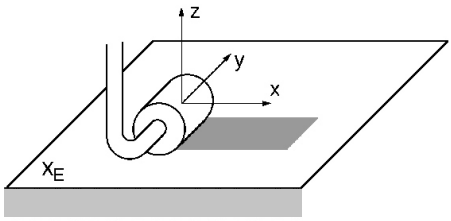
23.6.2 Hybrid Position/Force Control

This approach is based on a theory of compliant force and position control formalized by Mason[1] and concerns a large class of tasks involving partially constrained motion of the robot. Depending on the specific mechanical and geometrical characteristics of the contact problem, this approach makes a distinction between two sets of constraints upon robot motion and contact forces. The constraints that are natural consequences of the task configuration, i.e., of the nature of the desired contact between an end effector held by the robot and a constrained surface, are called *natural constraints*. Physical objects impose natural constraints. As already mentioned, a suitable frame in which the task to be performed is easily described, i.e., in which *constraints* are specified, is referred to as the *constraint frame* (or *task frame* or *compliance frame*).[56] For example, for a surface sliding contact task, it is customary to adopt the Cartesian constraint frame as sketched in Figure 23.15. Assuming an ideal rigid and frictionless contact between the end effector and the constraint surface, it is obvious that natural constraints restrict end effector motion in z direction and rotations about x and y axes. The frictionless contact prevents the forces in these directions and allows the torque around the z axis to be applied.

In order to specify the task of the robot with respect to the compliant frame, *artificial constraints* must be introduced. The artificial constraints must be imposed by the control system. These constraints essentially partition the possible DOFs of motion in those that must be position controlled and those that should be force controlled in order to perform the given task. The need to define an artificial constraint with respect to force when there is a natural constraint on the end-effector motion in this direction (i.e., DOF) and vice versa (Figure 23.15) is obvious.

To implement hybrid position/force control, a diagonal Boolean matrix $\mathbf{S}$, called the *compliance selection matrix*,[7] has been introduced in the feedback loops to filter out sensed end effector forces

© 2002 by CRC Press LLC



Natural Constraints		Artificial Constraints	
	$F_x = 0$		$v_x = V$
	$F_y = 0$		$v_y = 0$
$v_z = 0$		$F_z = 0$	
$\omega_x = 0$		$M_x = 0$	
$\omega_y = 0$		$M_y = 0$	
	$M_z = 0$		$\omega_z = 0$

FIGURE 23.15 Specification of surface sliding hybrid position/force control task.

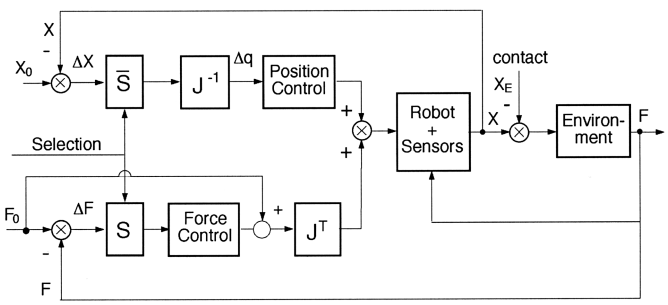


FIGURE 23.16 Explicit hybrid position/force control.

and displacements that are inconsistent with the contact task model. In accordance with the specified artificial constraints, the *i*-th diagonal element of this matrix has the value 1 if the *i*-th DOF with respect the task frame is to be force controlled and the value 0 if it is position controlled. To specify a hybrid contact task, according to Mason,[1] the following information sets must be defined:

1. Position and orientation of the task frame
2. Denotation of position and force controlled directions with respect to the task frame (selection matrix)
3. Desired position and force setpoints expressed in the task frame

Once the contact task is specified, the next step is to select the appropriate control algorithms. The relevant methods are discussed below.

23.6.2.1 Explicit Force Control

The most important method within this group is certainly the algorithm proposed by Raibert and Craig.[7] Figure 23.16 represents the control scheme that illustrates the main idea. The control consists of two parallel feedback loops, the upper one for the position, and the lower one for the force

© 2002 by CRC Press LLC

feedback loop. Each of these loops uses separate sensor systems. The positional loop utilizes the information obtained from the positional sensors at the robot joints, and the force loop is based on force-sensing data. Separate control laws are adopted for each loop. The central idea of this hybrid control method is to apply two outwardly independent control loops assigned to each DOF in the task frame. Both control loops cooperate simultaneously to control each of the manipulator joints. This concept, at first glance, appears to be ideal for solving hybrid position/force control problems. However, a deeper insight into the method reveals some essential difficulties and problems.

The first problem is related to the opposite requirements of the hybrid control concept concerning position and force control subtasks. Namely, the position control must be very stiff to keep the positioning errors in the selected directions as small as possible. The force control requires a relatively low stiffness of the robot (corresponding to the desired force) in the force controlled direction with respect to the task frame to ensure that the end effector behaves compliantly with the environment. As explained above, the explicit hybrid control attempts to solve this problem by control decoupling into two independent parts that are position and force controlled (Figure 23.16). In the force-controlled directions, the position errors decrease to zero by multiplication with the selection matrix orthogonal complement (position selection matrix) defined as $\bar{\mathbf{S}} = \mathbf{I} - \mathbf{S}$.* This implies that the position control part does not interfere with the force control loop, but that is not the case. The joint space nature of robot control realization results in a coupling between position and force control loops that are previously decoupled mathematically in the task frame. Assuming a proportional plus differential (PD) position control law, and assuming that the force control consists of a proportional plus integral controller (PI) with gain $\mathbf{K}_{Fp}$ and $\mathbf{K}_{Fi}$, respectively, and a force feed forward part, the control law according to the scheme in Figure 23.16 can be written in the Cartesian space as:

$$\boldsymbol{\tau} = \mathbf{K}_p\, \bar{\mathbf{S}}\, \Delta\mathbf{x} + \mathbf{K}_v\, \bar{\mathbf{S}}\, \Delta\dot{\mathbf{x}} + \mathbf{K}_{Fp}\mathbf{S}\Delta\mathbf{F} + \mathbf{K}_{Fi}\mathbf{S}\int \Delta\mathbf{F}dt + \mathbf{F}_0 \tag{23.33}$$

Based on relationships between Cartesian and joint space gains, Zhang and Paul[26] proposed an equivalent hybrid control law in the joint space:

$$\boldsymbol{\tau}_\mathbf{q} = \mathbf{J}^T\tau = \mathbf{k}_p\mathbf{J}^{-1}\bar{\mathbf{S}}\,\mathbf{J}\,\Delta\mathbf{q} + \mathbf{k}_v\,\mathbf{J}^{-1}\,\bar{\mathbf{S}}\,\mathbf{J}\,\Delta\dot{\mathbf{q}} + \mathbf{J}^T(\mathbf{K}_{fp}\mathbf{S}\Delta\mathbf{F} + \mathbf{K}_{fi}\mathbf{S}\int \Delta\mathbf{F}dt + \mathbf{F}_0) \tag{23.34}$$

Since each robot joint contributes to the control of both position and force, couplings in the manipulator's mechanical structure (implied in the Jacobian matrix) cause a control input to the actuator, corresponding to the force loop (e.g., force-controlled directions) to produce additional forces in position-controlled directions in the task frame, and vice versa. It is obvious from Equation (23.33) that by setting the position errors in the force controlled directions to zero (i.e., by filtering the position error through $\bar{\mathbf{S}}$), the position feedback gains in all directions are changed in comparison with the position control in free space. This causes the entire system to become more sensitive to perturbations. As a consequence, the performance of a robot with this scheme is not applicable for all robot configurations or all position/force-commanded directions. Moreover, one can find certain configurations with which, depending on selected force and position directions, the robot becomes unstable with the control law (Equation 23.33). This can be easily demonstrated on a simplified linearized robot model, derived from Equation (23.6) by neglecting the nonlinear Coriolis and centrifugal effects (due to small velocities in the contact task) and assuming that gravitational effects are ideally compensated for:

*For the sake of simplicity it is assumed that the task frame coincides with the Cartesian frame. Generally the selection matrix **S** is not diagonal in Cartesian space.[35]

© 2002 by CRC Press LLC

$$\mathbf{\Lambda}(x)\ddot{\mathbf{x}} = \boldsymbol{\tau} + \mathbf{F}. \tag{23.35}$$

Let us analyze the case where the manipulator is in free space and a noncontacting environment (e.g., in the transition phase when the force-controlled robot is approaching a contact surface after being switched from the position-control mode). Assume that some directions (e.g., orthogonal to the contact surface) have been selected for force control and remaining directions for position control. Taking into account that the force is zero, substituting Equation (23.33) in Equation (23.35) yields:

$$\mathbf{\Lambda}(x)\ddot{\mathbf{x}} + \mathbf{K}_v\bar{\mathbf{S}}\dot{\mathbf{x}} + \mathbf{K}_p\bar{\mathbf{S}}\mathbf{x} = \mathbf{K}_v\bar{\mathbf{S}}\dot{\mathbf{x}}_0 + \mathbf{K}_p\bar{\mathbf{S}}\mathbf{x}_0 \tag{23.36}$$

with a robot closed loop system matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \Lambda^{-1}\mathbf{K}_p\bar{\mathbf{S}} & \Lambda^{-1}\mathbf{K}_v\bar{\mathbf{S}} \end{bmatrix}. \tag{23.37}$$

To analyze the stability of this system, we determine the eigenvalues of **A**. As shown in Stokić and Šurdilović,[57] the closed-loop matrix becomes unstable in a number of configurations. Even if we introduce feedback loops with respect to the integrals of position errors in directions that are position controlled, it is always possible to find unstable configurations. These unstable configurations build working subspaces far away from singular positions where the system matrix **A** is intrinsically unstable due to the degeneration of the Jacobian matrix. Moreover, only alterations of the selection matrix can cause switching of robot behavior from stable to unstable and vice versa. The kinematic instability was experimentally tested and proven using the industrial robot control systems.[57]

Although the above stability analysis was based on a linearized model and therefore has some limitations, it provides a simple explanation of the nature of stability problems in hybrid position/force control. Since only the robot's position and the selection matrix influence the instability, this phenomenon is referred to as *kinematic instability.*[58] This phenomenon does not depend on whether the robot is in contact with the constraint surface. However, in contact situations, analysis of this problem is complicated by force/position relationship and the tests become very dangerous. It may be concluded that the kinematic instability problem encountered in the considered explicit hybrid position/force control represents a serious deficiency of this method and significantly reduces its applicability.

In order to overcome the difficulties related to kinematic instability, Zhang[59] proposed to introduce an additional selection of input forces. In other words, the input torques from position and force control parts (Figure 23.16) are decoupled in the task frame before they are applied to the joints. When the robot is in free space, the joint torque from the position control part (Equation 23.34) is initially transferred in the Cartesian-compliant frame, then multiplied with the selection matrix, and again transferred back using the static force transformation (i.e., Jacobian matrix) that provides the following control law for the position loop:

$$\boldsymbol{\tau}_q^{\,p} = \mathbf{J}^T\bar{\mathbf{S}}\mathbf{J}^{-T}\mathbf{k}_p\mathbf{J}^{-1}\bar{\mathbf{S}}\mathbf{J}\Delta\mathbf{q} + \mathbf{J}^T\bar{\mathbf{S}}\mathbf{J}^{-T}\mathbf{k}_v\,\mathrm{J}^{-1}\,\bar{\mathbf{S}}\mathbf{J}\Delta\dot{\mathbf{q}}. \tag{23.38}$$

It is relatively easy to prove that the linearized model (Equation 23.36) becomes kinematically stable with this control law. However, similar to the original control scheme, the eigenvalues of the system change with variation of the robot configuration and with the given task, i.e., selection matrix. This causes the robot performance to be strongly dependent on the configuration and selection of controlled directions.

© 2002 by CRC Press LLC

Fisher and Mujtaba[60] have shown that kinematic instability is not inherent to the explicit hybrid position/force control scheme; it is a result of an inappropriate mathematic formulation of position/force decomposition via selection matrix **S**. It was demonstrated that in the original hybrid control formulation (Equations 23.33 and 34), the position control loop is responsible for inducing the instability, namely the term $\mathbf{J}^{-1}\,\bar{\mathbf{S}}\,\mathbf{J}$ in Equation (23.34). The crucial error in the position control loop is, in the authors' opinion, made by the decomposition of the robot coordinate (DOF) to position- and force-controlled. Instead, to compute the selected position-controlled DOF and the corresponding selected joint errors, respectively, based on:

$$\mathbf{x}^p = \bar{\mathbf{S}}\mathbf{x} \tag{23.39}$$

and

$$\Delta\mathbf{q}^p = \mathbf{J}^{-1}\Delta\mathbf{x}^p = \mathbf{J}^{-1}\bar{\mathbf{S}}\Delta\mathbf{x} = \mathbf{J}^{-1}\bar{\mathbf{S}}\mathbf{J}\,\Delta\mathbf{q}. \tag{23.40}$$

the authors proposed to use the "correct" relationship between the selected Cartesian errors and the joint errors:

$$\Delta\mathbf{x}^p = \left(\bar{\mathbf{S}}\mathbf{J}\right)\Delta\mathbf{q} \tag{23.41}$$

Taking into account the selection matrix structure, it is obvious that $\left(\bar{\mathbf{S}}\mathbf{J}\right)$ is a singular matrix (with zero rows corresponding to the force DOF). Hence, the selected joint errors equivalent to the selected Cartesian position error are obtained as the minimal 2-norm solution:

$$\Delta\mathbf{q}^p = \left(\bar{\mathbf{S}}\mathbf{J}\right)^+\Delta\mathbf{x}^p = \left(\bar{\mathbf{S}}\mathbf{J}\right)^+\bar{\mathbf{S}}\Delta\mathbf{x} = \left(\bar{\mathbf{S}}\mathbf{J}\right)^+\Delta\mathbf{x} = \left(\bar{\mathbf{S}}\mathbf{J}\right)^+\mathbf{J}\,\Delta\mathbf{q} \tag{23.42}$$

or, when the robot is a singular position type, or has a redundant number of joints, with an additional term from the null space of the Jacobian **J**:

$$\Delta\mathbf{q}^p = \left(\bar{\mathbf{S}}\mathbf{J}\right)^+\Delta\mathbf{x} + \left[\mathbf{I} - \mathbf{J}^+\mathbf{J}\right]\mathbf{z}_q \tag{23.43}$$

where $\mathbf{z}_q$ is an arbitrary vector in the joint space and the plus sign denotes the Moor–Penrose pseudoinverse matrix. Thus, for the case in Equation (23.42), the control law of the position hybrid control part becomes:

$$\boldsymbol{\tau}_q{}^p = \mathbf{k}_p\left(\bar{\mathbf{S}}\,\mathbf{J}\right)^+\mathbf{J}\,\Delta\mathbf{q} + \mathbf{k}_v\left(\bar{\mathbf{S}}\,\mathbf{J}\right)^+\mathbf{J}\,\Delta\dot{\mathbf{q}}. \tag{23.44}$$

To determine how the above kinematic transformations can induce instability of the hybrid control, the authors defined a sufficient condition for kinematic stability. From the control viewpoint, this criterion prevents the second order system gain matrices (Equation 23.33) from becoming negative definite, which is a condition that produces system instability.[59] By testing the kinematic stability conditions for both original and correct selection and position error transformation solutions, the authors have proven that the instability can occur in the first case. The new hybrid control scheme, however, always satisfies the kinematic stability condition — it is always possible to find a vector $\mathbf{z}_q$ to ensure kinematic stability.

The second problem relates to *dynamic stability* issues in force control.[61] These effects concern high gain effect of force sensor feedback (caused by high environment stiffness), unmodeled high

© 2002 by CRC Press LLC

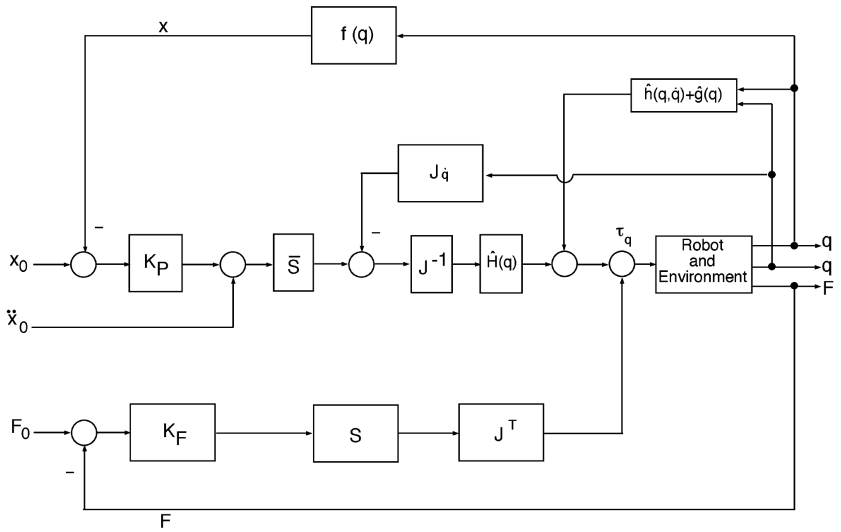


FIGURE 23.17 Resolved acceleration–motion force control.

frequency dynamic effects (due to arm and sensor elasticity), contact with a stiff environment, noncollocated sensing and control, and other factors.

To overcome dynamic problems of hybrid position/force control, several researchers pursued the idea to include the robot dynamic model in the control law. The resolved acceleration control originally formulated for the position control[62] belongs to the group of dynamic position control algorithms. Shin and Lee[31] extended this approach to the hybrid position/force control. The joint space implementation of the proposed control scheme is shown in Figure 23.17. The driving torque compensates for the gravitational, centrifugal, and Coriolis effects, and feedback gains are adjusted according to the changes in the inertial matrix. An acceleration feed-forward term is also included to compensate for changes of nominal motion in position directions. Finally, the control inputs are computed by:

$$\boldsymbol{\tau} = \hat{\Lambda}\ddot{\mathbf{x}}^{*} + \hat{\boldsymbol{\mu}}(x,\dot{x}) + \hat{\mathbf{p}}(x) + \mathbf{S}\mathbf{f}^{*} \tag{23.45}$$

where $\ddot{\mathbf{x}}^{*}$ is the commanded equivalent acceleration:

$$\ddot{\mathbf{x}}^{*} = \ddot{\mathbf{x}}_0 + \mathbf{K}_v\left(\dot{\mathbf{x}}_0 - \dot{\mathbf{x}}\right) + \mathbf{K}_p\left(\mathbf{x}_0 - \mathbf{x}\right) \tag{23.46}$$

and $\mathbf{f}^{*}$ is the command vector from the force control parts whose form depends on the applied control law. To minimize the force error, it is convenient to introduce the PI force regulator of the form:

$$\mathbf{f}^{*} = \mathbf{K}_{fp}\left(\mathbf{F}_0 - \mathbf{F}\right) + \mathbf{K}_{fi}\int\left(\mathbf{F}_0 - \mathbf{F}\right)dt\,. \tag{23.47}$$

Khatib[22] introduced an active damping term into the force control part to avoid bouncing and minimize force overshoots during transition (impact effects):

$$\boldsymbol{\tau}_f = \mathbf{S}\mathbf{f}^{*} - \hat{\boldsymbol{\Lambda}}\mathbf{S}\mathbf{K}_{vf}\dot{\mathbf{x}} \tag{23.48}$$

© 2002 by CRC Press LLC

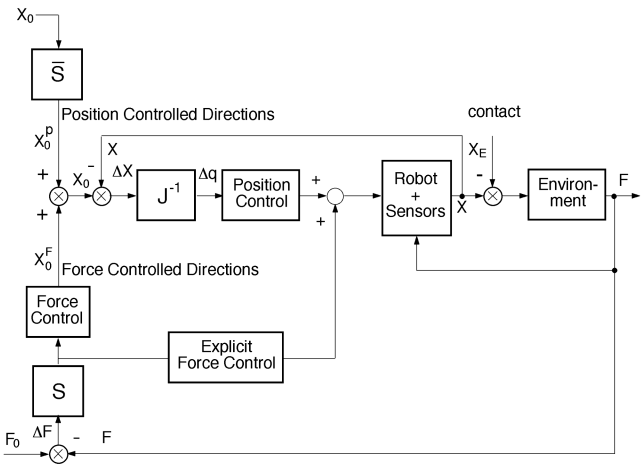


FIGURE 23.18 Implicit hybrid position/force control.

where $\mathbf{K}_{vf}$ is a diagonal Cartesian damping matrix. Bona and Indri[63] proposed further modifications of the control scheme. To compensate for the coupling between force and position control loops and for the disturbance of the position controller due to reaction force, the authors modified the position control law according to:

$$\boldsymbol{\tau}_p = \hat{\Lambda}\bar{\mathbf{S}}\left[\ddot{\mathbf{x}}^* - \hat{\Lambda}^{-1}\left(\mathbf{S}\mathbf{f}^* - \mathbf{F}\right)\right] + \hat{\boldsymbol{\mu}}(x,\dot{x}) + \hat{\mathbf{p}}(x) \tag{23.49}$$

If the dynamic modeling used for computation of the control law is exact, the above control law provides complete decoupling between position and force control in the task frame, i.e., the following closed loop behavior:

$$\ddot{\mathbf{x}} = \bar{\mathbf{S}}\ddot{\mathbf{x}}^* + \mathbf{S}\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{S}\mathbf{f}^* - \mathbf{S}\hat{\boldsymbol{\Lambda}}^{-1}\mathbf{F} - \mathbf{S}\mathbf{K}_{vf}\dot{\mathbf{x}}. \tag{23.50}$$

An experimental evaluation and comparison of explicit force control strategies was presented in Volpe and Khosla.[64]

23.6.2.2 Position Based (Implicit) Force Control

The reason explicit force control methods cannot be suitably applied in commercial robotic systems lies in the fact that commercial robots are designed as positioning devices. The feedback term, i.e., the signal proportional to the force errors, is multiplied by the transposition of the Jacobian matrix in order to calculate the driving torques that have to be realized around the joints to achieve the desired force action (Figure 23.16). These signals are directly fed to the inputs of the local servo parts. However, the computed torques may not be accurate for commercial robotic systems. Since there is no position feedback loop in the force-controlled direction, the robot will move due to various disturbances acting upon it, such as controller and sensor drifts, etc.[57] The implementation of explicit force control can be successfully performed only by a new generation of direct drive robots.

In commercial applied robotic systems, implementing *implicit* or *position-based* force control by closing a force-sensing loop around the position controller (Figure 23.18) appears promising. The input to the force controller is the difference between desired and actual contact force in the task frame. The output is an equivalent position in force-controlled directions which is used as reference input to the positional controller. According to the hybrid force/position control concept,

© 2002 by CRC Press LLC

the equivalent position in force direction $\mathbf{x}_0^F$ is superimposed to the orthogonal vector $\mathbf{x}_0^P$ in the compliance frame, which defines the nominal position in orthogonal position-controlled directions. The robot behavior in force direction is affected only by the acting force. The positional controller remains unchanged, except for the additional transformations between Cartesian and task frame which have to be introduced since these two frames are not coincident. Since a positional controller provides a basis for realization of force control, this concept is referred to as *implicit* or *position-based* force control,[15] or *external force* control.[13]

The role of force control block in this scheme is two-fold, first, to compensate for the effects of environment (contact process), and second, to achieve tracking of the desired force. Another important quality of a force-controlled manipulator is the ability to respond to positional variations of the contact surfaces. Commonly, a PI force controller has been applied. A more complex force controller including the compensation of the internal position control effects has been proposed in Stokić and Šurdilović.[65] In Figure 23.18 an explicit force control block is added. This scheme combines the implicit and explicit control with the aim of using benefits (robustness and reliability of *implicit* force control and fast reaction of the *explicit* one) and compensating specific disadvantages of single force control approaches.

The main features of the implicit force control scheme are its reliability and robustness. Implemented in commercial robotic systems, this scheme is neither configuration dependent nor sensitive to parameter variation. This control algorithm can be used for arbitrary processes. However, this scheme also exhibits some drawbacks. The accuracy of contact forces is mainly limited by the precision of robot positioning (sensor resolution). The precision can be disturbed when contact with a very stiff environment is requested. Fortunately, inherent compliance of the robot structure or force sensor is always present and reduces the equivalent system stiffness. The performance of implicit force control is significantly limited by the bandwidth of the position controller. A slightly higher bandwidth can be achieved by using a compensator of a higher order. However, due to coupling between position and force-controlled degrees of freedom, whether force control can become significantly faster is questionable.

23.6.2.3 Other Force Control Approaches

The next group of algorithms considers more complex constraints on robot motion. They are described as a set of rigid hypersurfaces in the spaces of end effector Cartesian coordinates,[11] or in the joint coordinate space.[32] The system model is described by a typical set of linearly implicit second order differential algebraic equations (mechanical differential algebraic equations). This model is used to compute the control law to linearize and decouple the system dynamics and divide the control problem into position- and force-controlled directions.

To improve reliability, the dynamic hybrid control is extended to unknown environments that consist of hypersurfaces.[66] The improved control schemes involve on-line identification algorithms based on force and position measurements and adaptive control mechanisms. However, the adaptive constrained motion control is theoretically attractive, but impractical in reality. Hence, the hybrid control algorithms become even more complex and difficult to implement in real time with the computational and sensing resources available for robotic manipulators today.

The hybrid position/force task specification has been a subject of several investigations. Lipkin and Duffy[24] demonstrated that Mason's position/force decomposition approach based on *geometrical orthogonality* is in fact erroneous. The resulting planning for hybrid control is not invariant with respect to translation of origin or change of unit length. The authors proposed a more general and mathematically consistent invariant hybrid task formulation based on screw algebra. The complementarity between motion (modeled by a twist) and force (represented by a wrench) is expressed via a *reciprocity relationship* independent of coordinate frame, scaling, or units. Two fundamental relations between twist and wrench, referred to as *freedom* and *constraint equations*, have been introduced to test task compatibility with the model of the environment. These relations correspond to analytical expressions for natural and artificial constraints in the noninvariant hybrid approach. For every constrained motion

© 2002 by CRC Press LLC

task, two screw subspaces that correspond to artificial constraints can be derived. These subspaces represent the sets of screws about which twists and wrenches can be controlled.

In certain simple tasks and reference frames, both conventional and reciprocity-based decomposition show the same results. However, the reciprocity-based approach provides a more general decomposition applicable when the freedom and constraint subspaces do not span a six-dimensional space or have nonzero intersections and also to manipulators that have fewer than six DOF.[67]

If the twist and wrench are consistent with the environment (i.e., the freedom and constraint equations are satisfied) the specified task is feasible for hybrid control. In the opposite case, the specified twist and wrench must be filtered to obtain a kinestatically realizable control action (so-called kinestatic filtering).

A procedure to apply the reciprocity-based task decomposition to manipulator dynamics to obtain equations of motion relevant for hybrid control was presented by Sinha and Goldenberg.[68] Several model-based tools for task specification using this approach were presented by Khatib.[22] The reciprocity concept is well suited for nominal specification of arbitrary motion constraints and also serves to define possible uncertainties and on-line identification and observation of real motion constraints. This strategy generally makes task execution against uncertainties very robust. This is particularly essential for contour-following tasks. An overall hybrid position force control scheme based on general decomposition formalism including identification of geometrical uncertainties was proposed by De Schutter and Bruyninckx.[25] Design of appropriate controllers is subject to further researches.

23.6.3 Force/Impedance Control

Several attempts have been made to combine impedance and force control with the aim of compensating for specific disadvantages of single control approaches. Although it is possible under some circumstances to demonstrate correspondence between force and impedance control laws,[69] there are essential differences between these main constraint motion control concepts.

The main advantage of impedance control over force control is easier task specification and programming. A contact task is specified in terms of motion sequences, so the impedance control does not require modifications of conventional free space planning control concepts and algorithms (the programmer can take advantage of existing off-line programming). Moreover, impedance control can be activated in free space during approach motion. Thus, it can be applied for the transition to and from the constraint motion, without specific control-switching algorithms.

Impedance control allows closed-loop position control in free space, while in contact with rigid environments, it offers force open-loop capabilities. Conversely, the force (admittance) control approach allows closed-loop force control capabilities in contact, but exhibits open-loop position control characteristics in free space. Therefore, the activation of force control in free space is only possible under specific circumstances. In general, however, a discontinuous control strategy is required for the transition from noncontact to contact motion phase or vice versa. The control structure change is done during the most critical phase when the manipulator is in contact with the environment. That represents a major drawback of force control. To cope with unexpected collisions, additional sensors (e.g., distance) have to be integrated into the control system. The fundamental superiority of force control is, however, that the interaction force is the result of the control action, rather than a result of deviation of the environment position and the chosen target impedance.

In Goldenberg's algorithm,[38] force control is closed around an internal impedance control loop. Desired force and force error are used to compute an equivalent desired relative motion of the end effector. Impedance control is included with the aim of achieving a suitable relationship between force and relative motion during contact. This is realized in the internal velocity loop by compensator gain adjustment to obtain target impedance. A similar reliable position-based force/impedance control scheme suitable for implementation in industrial robots has been proposed by Šurdilović and Kirchhof.[70] An external implicit force controller loop is closed around an internal position-based impedance controller (Figure 23.19). The main goal of the internal loop is to achieve target

© 2002 by CRC Press LLC

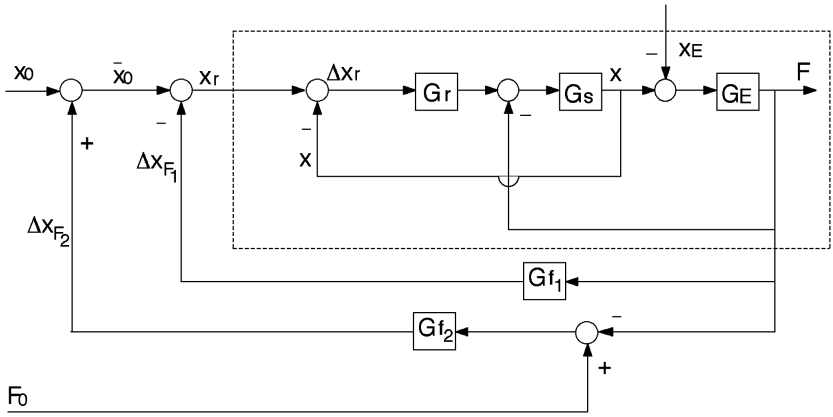


FIGURE 23.19 Position-based force/impedance control.

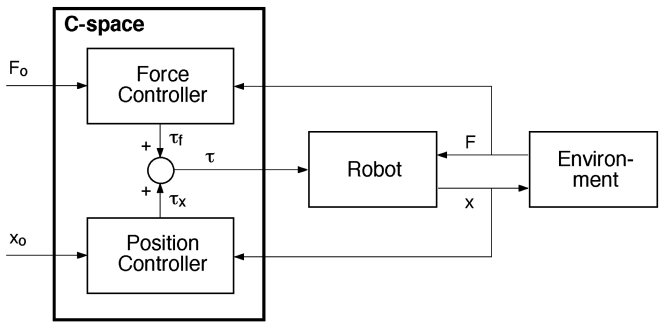


FIGURE 23.20 Parallel position/force control.

impedance while the external loop takes care of desired force realization. The selection between position (i.e., impedance) and force-controlled directions is not needed. Indeed impedance and force control affect all directions. A disproportion between motion and force planning is not critical in the control scheme (Figure 23.19), since the internal control loop behaves as a low stiffness target impedance system allowing relatively large differences between input position command and real robot position in the output. In the reverse, the internal loop in the implicit force control (Figure 23.18) is a very stiff position control, and the selection is inevitable.

Anderson and Spong[12] proposed an approach referred to as *hybrid impedance control* algorithm to control contact forces. The kernel part of the algorithm is Raibert and Craig's hybrid position/force control scheme, with the selection matrix applied to decompose position- and force-controlled subspaces. Both control parts use the feedback of contact force to realize desired system impedance (position-based and force-based impedance control) along each DOF.

A controller that combines an internal position control, a position-based impedance compensator, and a desired force feed forward was proposed by Mayeda et al.[71] The authors suggest that integral control actions be applied for both impedance (damping control) and force filters to ensure the compliance and the desired steady state force.

A different approach to position/force control, referred to as *parallel control* (Figure 23.20), has been proposed by Chiaverini and Sciavicco.[9] Contrary to the hybrid control, the key feature of the parallel approach is to have both force and position controls along the same task space direction without a selection mechanism. In general, both position and force cannot be effectively controlled in an uncertain environment. Therefore, the logical conflict between the position and force actions is managed by the dominance of the force control action over the position action along the constrained task direction

© 2002 by CRC Press LLC

where a force interaction is expected. The force control is designed to prevail over the position control in constrained motion directions. This means that force tracking is dominant in directions where an interaction with the environment is expected, while the position control loop allows compliance, i.e., a deviation from the nominal position in order to reach the desired forces. For this reason, the parallel control method can be considered a force/impedance control approach. The designed position tracking quality in constrained motion directions corresponds to target impedance behavior.

A set of sufficient local asymptotic stability conditions has been derived by Chiaverini et al.[72] for a parallel controller case consisting of a PD action on the position loop and a PI control in the force loop, together with gravity compensation and the desired force feed forward. Stability analysis and simulation results on an industrial robot are included. These conditions imply a relatively high damping (i.e., velocity gain) to ensure system stability.

23.6.4 Position/Force Control of Robots Interacting with Dynamic Environment

Vukobratović and Ekalo[8,33] established a unified approach to simultaneously control position and force in an environment with completely dynamic reactions. This fully dynamic approach to the control of robots interacting with dynamic environments will be presented in a condensed way. It will be assumed that $n = m$, where n is the number of robot DOFs and m is the number of contact force components. The general case in which $n > m$ has been considered by Vukobratović et al.[73]

When the environment does not possess displacements (DOFs) that are independent of robot motion, the environment dynamics in the robot coordinate space can be described by the model (23.9). Then the system (23.1 through 23.9) describes the dynamics of robot interaction with a dynamic environment. It is assumed in the contact case that all mentioned matrices and vectors are continuous functions and that the robot is in permanent unilateral contact with the environment.

In the case of contact with the environment, the robot control task can be described as motion along a programmed trajectory $\mathbf{q}_p(t)$ representing a twice continuous differentiable function, when a desired force of interaction $\mathbf{F}_p(t)$ acts between the robot and the environment. The nonlinear model programmed motion $\mathbf{q}_p(t)$ and desired force $\mathbf{F}_p(t)$ must satisfy the relation:

$$\begin{aligned} \mathbf{F_p}(t) &\equiv \mathbf{f}(q_p(t), \dot{q}_p(t), \ddot{q}_p(t)) \\ \mathbf{f}(q, \dot{q}, \ddot{q}) &= -(\mathbf{S}^{\mathbf{T}}(q))^{-1}[\mathbf{M}(q)\ddot{q} + \mathbf{L}(q, \dot{q})]. \end{aligned} \tag{23.51}$$

The control goal of robot interaction with a dynamic environment can be formulated by defining the control $\boldsymbol{\tau}(t)$ for $t \geq t_0$ that is to satisfy the target conditions:

$$\mathbf{q}(t) \to \mathbf{q_p}(t), \quad \mathbf{F}(t) \to \mathbf{F_p}(t), \quad \text{as} \quad t \to \infty \tag{23.52}$$

The two questions are addressed to the control design problem. Can we choose such a control law that, by satisfying preset robot motion quality, would enable the attainment of the control goals that satisfy the relation of Equation (23.52)? Is it possible to choose the control law in such a way as to ensure the preset quality of the robot interaction force and the attainment of the control goals? The answer to the first question is quite simple:[8,33] the inverse dynamics methods ensure that desired motion quality is achieved and at the same time guarantee that the interaction force is stable. The answer to the second question depends on the environment dynamics.

The task of stabilizing the programmed interaction force $(PFI)\mathbf{F}_p(t)$ can be posed by considering a family of transient responses with respect to force in the form $\boldsymbol{\mu} = \mathbf{F}(t) - \mathbf{F_p}(t)$ and by choosing a continuous vector function $\mathbf{Q}$ $(\mathbf{Q}(0) = 0)$ of dimension n, such that the asymptotic stability as a whole is ensured for the trivial solution of $\boldsymbol{\mu}(t) \equiv 0$.

© 2002 by CRC Press LLC

Let us consider pure force control according to the assumption that $m = n$, i.e., when the number of the contact force components is equal to the number of the powered DOFs of the robot. For convenience, when describing the quality of transient response perturbation force dynamics, $\dot{\boldsymbol{\mu}} = \mathbf{Q}(\mu)$, we shall use an equivalent relation of the form:

$$\dot{\boldsymbol{\mu}}(t) = \boldsymbol{\mu}_0 + \int_{t_0}^{t} \mathbf{Q}(\mu(\omega))d\omega. \tag{23.53}$$

With no loss in generality, we can adopt $\boldsymbol{\mu}_0 \equiv 0$, because the stabilization of $\boldsymbol{\mu}$ in the sense of preset quality Equation (23.53) directs stabilization according to the preset quality $\dot{\boldsymbol{\mu}} = \mathbf{Q}(\mu)$ independently from the value of $\boldsymbol{\mu}_0$.

Let us consider only one of the possible control laws with the feedback loops with respect to $\mathbf{q}$, $\dot{\mathbf{q}}$, and $\mathbf{F}$ of the form[8,33]

$$\boldsymbol{\tau} = \mathbf{H}(\mathrm{q})\mathbf{M}^{-1}(\mathrm{q})\left\{-\mathbf{L}(\mathrm{q},\dot{\mathrm{q}}) + \mathbf{S}^{T}(\mathrm{q})\left[\mathbf{F}_{\mathbf{p}} + \int_{t_0}^{t} \mathbf{Q}(\mu(\omega))\mathrm{d}\omega\right]\right\} + \mathbf{h}(\mathrm{q},\dot{\mathrm{q}}) + \mathbf{g}(\mathrm{q}) - \mathbf{J}^{\mathrm{T}}(\mathrm{q})\mathbf{F}. \tag{23.54}$$

By applying this control law to the robot dynamics model Equation (23.1) we obtain the following law of robot operating in contact with the environment:

$$\mathbf{M}(q)\ddot{\mathbf{q}} + \mathbf{L}(\mathrm{q},\dot{\mathrm{q}}) = \mathbf{S}^{\mathrm{T}}(q)\left(\mathbf{F}_{\mathbf{p}} + \int_{\mathbf{t_0}}^{\mathbf{t}} \mathbf{Q}(\mu(\omega))\mathrm{d}\omega\right)$$

Taking into account the environment dynamics model Equation (23.9), we obtain the following closed-form control system:

$$\mathbf{S}^{\mathrm{T}}(q)\left(\boldsymbol{\mu}(t) - \int_{\mathbf{t_0}}^{\mathbf{t}} \mathbf{Q}(\mu(\omega))\mathbf{d}\omega\right) = \mathbf{0} \tag{23.55}$$

and, because $rank(S) = n$, (23.55) is equivalent to: $\boldsymbol{\mu}(t) = \int_{t_0}^{t} \mathbf{Q}(\mu(\omega))d\omega$, from which $\dot{\boldsymbol{\mu}}(t) = \mathbf{Q}(\mu(\omega))$ follows directly. In this way, the control law (23.54) ensures the desired quality of stabilization of $(PFI)F_p(t)$.

The stability of the real motion (position) when asymptotic stability of the contact force is fulfilled has been considered.[8,33,89–91] Sufficient conditions for contrained motion stability based on the generalized Lyapunov's stability theorem in the first approximation of the system with perturbation have been derived. The theorem conditionally defines the internal stability properties of the environment because the fulfillment of stability conditions depends in general not only on environment dynamics but also on the nature of the programmed motion.

23.7 Contact Stability and Transition

The types of contact tasks may vary substantially in relation to specific requirements, but in all cases of performing a contact task the robot must perform three kinds of motions:

- *Gross motion*, related to movement in free space (free motion mode)
- *Compliant* or *fine* motion, related to movement constrained by environment
- *Transition motion*, representing all passing phases between free and compliant motion

© 2002 by CRC Press LLC

The contact transition can be considered stable if the contact is not lost after the manipulator meets the environment. A stable contact transition can be characterized by nonzero force (after contact is detected), positive penetration of manipulator end point into environment, nonappearance of bouncing, etc. The most critical issue in transition control is initial impact against a stiff environment. A stable controller should ensure the passage through the transition phase and maintain contact until all impact energy has been absorbed.

In most of the proposed control algorithms, instability occurs when the contact between end effector and environment is stiff. However, the investigations were primarily concerned with the question of *coupled stability* (i.e., will the robot remain stable when it is interconnected with the environment?) of robots and the environment under various control algorithms, while assuming the manipulator initially is and remains in contact with environment. Surprisingly, relatively little research has addressed the problem of *contact transition stability* (i.e., will the robot during transition from free to contact motion establish a continuous contact with the environment without multiple impacts?) which is most fundamental for performing contact tasks. The contact transition stability problem is important for both unilateral (force) and bilateral (geometric) constraints. A bilateral constraint is usually achieved by closing the gripper, due to position misalignment usually resulting from unilateral contact between gripper jaws and grasping object.

In *impedance control*, contact stability issues have mainly been considered based on simplified models of interaction between a target impedance system and the environment. Colgate and Hogan[74] defined necessary and sufficient conditions to ensure the stability of a linear robotic system coupled to a linear environment. The authors applied the network theory to describe the manipulator- and environment-interactive behavior at the equilibrium point. For the coupled interactive system described by the linear models, the equilibrium is defined by:

$$\begin{aligned}
\mathbf{p}^* &= \mathbf{p}(t; t \to \infty) = \mathbf{x}^* - \mathbf{x}_e \\
\mathbf{p}_0^* &= \mathbf{p}_0(t; t \to \infty) = \mathbf{x}_0^* - \mathbf{x}_e \\
\mathbf{e}^* &= \mathbf{e}(t; t \to \infty) = \mathbf{p}_0^* - \mathbf{p}^* = \mathbf{x}_0^* - \mathbf{x}^* \\
\mathbf{F}^* &= \mathbf{F}(t; t \to \infty)
\end{aligned} \tag{23.56}$$

where $\mathbf{p}_0 = \mathbf{x}_0 - \mathbf{x}_e$ and $\mathbf{p} = \mathbf{x} - \mathbf{x}_e$ denote *nominal penetration*, expressing a position planning failure due to tolerances, a desired entry into the environment, and *actual robot penetration*, respectively. For the adopted linear target impedance and environment models defined by Equations (23.14) and (23.11) respectively, these equilibriums can be expressed as:

$$\begin{aligned}
\mathbf{F}^* &= \left[\mathbf{I} + \mathbf{G}_e(0)\hat{\mathbf{G}}_t^{-1}(0)\right]^{-1}\mathbf{G}_e(0)\mathbf{p}_0^* = \left[\mathbf{I} + \mathbf{K}_e\mathbf{K}_t^{-1}\right]^{-1}\mathbf{K}_e\mathbf{p}_0^* \\
\mathbf{p}^* &= \left[\mathbf{I} + \mathbf{K}_e\mathbf{K}_t^{-1}\right]^{-1}\mathbf{p}_0^* \\
\mathbf{e}^* &= \left[\mathbf{I} + \mathbf{K}_e\mathbf{K}_t^{-1}\right]^{-1}\mathbf{K}_e\mathbf{K}_t^{-1}\mathbf{p}_0^*
\end{aligned} \tag{23.57}$$

Expressing the essential impedance control characteristics, interaction force **F**, penetration **p**, and position error **e,** in terms of nominal penetration are useful for the analysis of both coupled and contact transition stability.[53] During contact establishment, $\mathbf{p}_0(t)$ is a positive monotone-increasing function. In a passive stationary environment, two time-invariant networks coupled along interaction ports (Figure 23.21) can represent the interactive model around the equilibrium $\mathbf{p}_0(\infty) = \mathbf{p}_0^*$. The coupling makes the velocities of the robot and the environment at contact point equal, while the forces acting upon the robot and the environment have opposite activities (action

© 2002 by CRC Press LLC

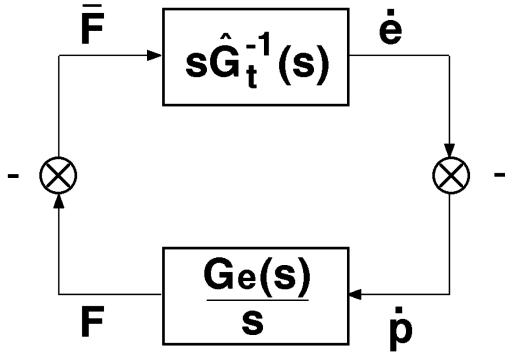


FIGURE 23.21 Robot/environment interaction model.

and reaction). If the environmental transfer matrix $\mathbf{G}_e(s)/s$ is positive real, representing a passive Hamiltonian environment, then a necessary and sufficient condition to ensure stability of a linearized robotic control system is that the *realized admittance* $s\mathbf{G}_t^{-1}(s)$ be *positive real*.[74] In other words, it should represent the driving point impedance of a passive network. In a SISO system, the *coupled stability* has been proven by using the Nyquist criterion and the positive real transfer function that has a limited phase of ± 90°.[74] It is then relatively easy to prove that the mapping of the Nyquist contour of a positive real environmental impedance $\mathbf{G}_e(s)/s$ through an also positive real admittance $s\hat{\mathbf{G}}_t^{-1}(s)$, altering the phase by ± 90° and changing the magnitude by a factor 0 to ∞, provides a stable system, i.e., a stable Nyquist plot of the open-loop coupled system transfer function.

The system *passivity concept* provides a relatively simple test for the assessment of coupled system stability. Only the passivity of the environment can be proven without accurate knowledge of parameters. Assuming that the ideal target impedance response Equation (23.15) is realized, the passivity of target admittance $s\mathbf{G}_t^{-1}(s)$ implies positive definite matrices $\mathbf{M}_t$, $\mathbf{B}_t$, and $\mathbf{K}_t$, and consequently, the closed-loop system should be stable in contact with any passive environment to which it is directly coupled. The explicit design of a positive-real robot control system, however, may become cumbersome.[75] Moreover, various practical control implementation effects, including computational time delay, sampling effects, and unmodeled dynamics (e.g., high order actuator and arm dynamic effects), may result in a nonpassive real impedance control response.[75]

The above stability results can be extended to nearly passive control systems. However, a passive environment can destabilize the coupled system. To simplify coupled stability analysis, Colgate and Hogan[74] used *worst* or *most destabilizing environment* to denote the most critical environment for coupled system stability. Such environmental impedance $\mathbf{G}_e(s)/s$ shapes the Nyquist contour of $s\hat{\mathbf{G}}_t^{-1}(s)$ by minimizing the distance from the critical point –1 to the nearest point on the Nyquist plot of the loop transfer function $\mathbf{G}_e(s)\mathbf{G}_t^{-1}(s)$. Since the driving point impedance of simple passive environmental models, such as mass or spring (M_e/s and sK_e), performs the maximum rotation in the Nyquist plane, the authors found that the *worst passive environment* for coupled stability consists of a set of *pure masses and springs*. If both the environment and the realized admittance are stable, the coupled stability of the interactive system in Figure 23.21 can also be assessed by means of the *small gain theorem* by which a feedback loop composed of stable operators will certainly remain stable if the product of all operator gains is smaller than unity:

$$\left\|\mathbf{G}_e(j\omega)\hat{\mathbf{G}}_t^{-1}(j\omega)\right\|_\infty < 1 \tag{23.58}$$

The small gain theorem provides a general law, valid for continuous- or discrete-time, SISO and MIMO, and linear and nonlinear systems. It is also the convergence criterion used in many iterative

© 2002 by CRC Press LLC

processes. Furthermore, the norm inequality criterion (Equation 23.58) can easily be extended to maintain the uncertainties in the target system and environment models.

However, the small gain theorem only provides sufficient stability conditions, which in many cases are too conservative to be of much use in practical contact tasks. For example, assuming that ideal second-order target impedance has been achieved, the condition (Equation 23.58) implies the admissible target stiffness to be $K_t \geq K_e$ to ensure a stable interaction. In stiff environments, this has no practical relevance. This result is similar to the stability analysis performed by Kazerooni et al.[16] The established interaction stability criterion practically implies that the gain of feedback compensator (i.e., the target admittance) should be limited by the magnitude of sum of environmental admittance and robot position control sensitivity. For a SISO system, this imposes, in the steady state:

$$K_t \geq \min\left(K_p, K_e\right) \tag{23.59}$$

In direct drive robotic systems with significantly less position control stiffness (due to elimination of the transmission) than in industrial robots, this condition might provide reliable target models for practical tasks. The sufficiency of stability condition (Equation 23.59) has experimentally been demonstrated on a lightweight direct drive University of Minnesota robot. However, from the viewpoint of industrial robot performance, this condition is conservative and practically useless. In industrial robots, with stiff servo gains (e.g., position control gains usually have the order 10^6 N/m), the value is $K_p > K_e$ and the above condition also requires the target stiffness to be higher than the environmental one. Moreover, no target model, i.e., the compliance feedback compensator G_f (Figure 23.11), can be found to enable interaction with an infinitely rigid environment ($K_e \to \infty$). Therefore, one of the main conclusions in Kazerooni et al.[16] pointed out the need for intrinsic compliance either in the robot or in the environment to maintain interactive stability.

The *coupled stability* analysis around the equilibrium point cannot be applied for the analysis of *contact transition stability*, and this represents a fundamental contact control task problem. Reliable criteria ensuring contact stability of a linearized robotic control system under impedance control during transition from the free space to a unilateral contact within a passive environment has been established by Šurdilović.[42] The contact transition stability conditions require interaction force, i.e., actual penetration to be nonnegative $\mathbf{p}(t) = \mathbf{x} - \mathbf{x}_e \geq 0$, or the position deviation to be less than the nominal penetration $\mathbf{e}(t) \leq \mathbf{p}_0(t)$ i.e.,

$$\frac{\left|\mathbf{e}(t)\right|}{\left|\mathbf{p}_0(t)\right|} = \frac{\left|\mathbf{p}_0(t) - \mathbf{p}(t)\right|}{\left|\mathbf{p}_0(t)\right|} = \frac{\left|\mathbf{x}_0(t) - \mathbf{x}(t)\right|}{\left|\mathbf{x}_0(t) - \mathbf{x}_e\right|} \leq 1 \tag{23.60}$$

This relation implies the actual end effector position during a stable contact transition will always be located between the position of environment and the nominal position. Since this contact stability condition is based on a simple geometric consideration, it is referred to as the *geometric criterion.*[42] This criterion theoretically can be applied in cases when the actual position overshoots the nominal one, i.e., when $\mathbf{p} > \mathbf{p}_0$, which provides a negative position error. However, in a contact with an industrial robot with a realistic stiff environment, the actual motion is nearly stopped by the resistant force and impedance control effects, so this case has no practical relevance. The advantage of the geometric criterion is that it compares two time signals. The norm comparison offers possibilities to apply relatively simple and efficient system theory formalisms for contact stability analysis. This criterion has been utilized[76] to derive the *robust contact stability* condition ensuring both coupled and contact transition stability based on:

$$\sup \frac{\left\|\mathbf{e}(t)\right\|_2}{\left\|\mathbf{p}_0(t)\right\|_2} < \left\|\mathbf{W}(s)\left(\mathbf{I} + \mathbf{G}_e^{-1}(s)\mathbf{G}_t(s)\right)^{-1}\right\|_\infty \leq 1 \tag{23.61}$$

© 2002 by CRC Press LLC

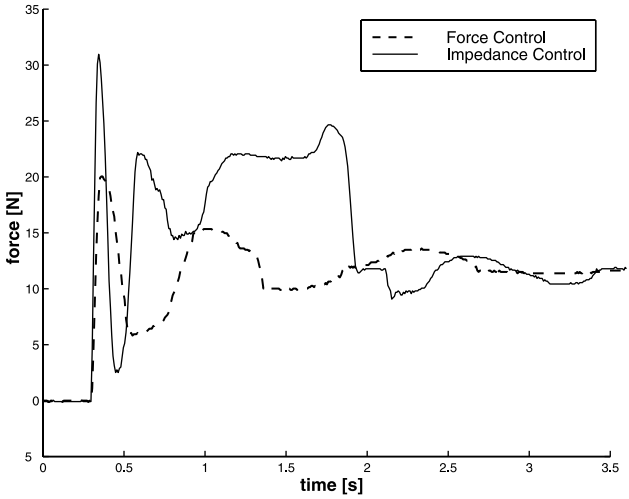


FIGURE 23.22 Performance comparison: impedance vs. implicit hybrid control.

where stable weighting transfer function matrix $\mathbf{W}(s)$ describes uncertainties of the environmental model and the target impedance realization. In a SISO system, this condition implies:

$$\xi_t \geq \frac{1}{2}\left(\sqrt{1+2\kappa}-1\right) \tag{23.62}$$

where $\xi_t = B_t / 2\sqrt{M_t K_t}$ and $\kappa = K_e / K_t$ are target damping and stiffness ratios, respectively. However, in spite of an effective and simple formulation, this criterion ensures *sufficient* contact stability conditions, but not the *necessary* ones. Consequently, the obtained contact stability indices might be conservative. It should be mentioned that the damping ratio bound Equation (23.62) is still less than the usually applied *dominant real pole* solution[77] imposing $\xi_t \geq \sqrt{1+\kappa}$. A very important advantage of the input/output criterion Equation (23.61) is that it can be applied for both continuous and discrete systems including the time lags. The control time lag has been identified by Šurdilović[76] as the critical destabilizing contact transition effect. In general, a retarded system requires a significantly higher amount of damping to stabilize the transition process with delayed force signals.

Typical transition experimental results in position-based impedance and force control during contact with a stationary environment are presented in Figure 23.22.[42] The force transition in hybrid control is characterized by lower overshoots. The reason is that force control represents an explicit aspect of hybrid control that is achieved by appropriate control structure and design. In the impedance control, however, the aim is to passively modify a preplanned motion in accordance with the interaction forces. Therefore, the force transition in the impedance control is greatly measured and influenced by selected target impedance parameters and nominal motion.[42]

Lawrence[14] analyzed the destabilizing influence of time delays on impedance control performance. Vossoughi and Donath[78] investigated the influence of nonlinear friction effects on the performance of an impedance-controlled hand. High position control gain (or integral gain) leads to limit cycles due to friction/stiction effects.

In contrast to the impedance control algorithms that provide the same control structure for the three motion phases, the transition to and from contact motion is usually based on discontinuous control in force control schemes. The change of the control strategy from position control to force control occurs in the free space, and the transition is realized in the force control mode after contact is established. Most force control algorithms execute the transition control in the *force mode*. The reason is that the impact force can be very large, especially when due to high approach velocities and delay in a stiff position controller. One method to reduce impact is to use a soft force sensor,[79]

© 2002 by CRC Press LLC

i.e., passive compliance, but this reduces the position accuracy during position control. The underlying idea of most methods concerned with the impact problem is to increase damping in the collision direction.[22] Assuming a simplified stiffness model of environment, the damping effect can be achieved either by force derivatives or approach velocity feedbacks. However, both methods have practical limits. The force signals are usually noisy and the derivation is inaccurate. Qian and De Schutter[80] proposed low-pass sensor filtering and nonlinear damping to cope with the transition problem. The velocity sensing is also not reliable at slow approach motion before contact.

Moreover, in a stiff environment, relatively fast oscillations in force and velocity can cause instability due to time (phase) lags between sensing and control action. These difficulties have been recently addressed in several works aimed at designing a stable force controller without velocity measurement,[81] and without end effector contact force sensing,[82] when the system dynamics is well known. However, these innovative schemes are complex to implement and require further tests.

Independent of the active damping method, the transition control based on the force mode generally requires some modification of control strategy or gains before and after impact. For example, in integral explicit force control, the force error integration in the free space causes the robot to accelerate in the force direction. Hence, the maximum impact velocity or integration wind-up should be limited in the free space. In implicit integral force control, a constant force error corresponds to a constant position correction velocity, but usually the different gains should be used in the free space and during contact to achieve desired system performance. The gains synthesized for stiff contact provide very slow free motion, while contact stability is jeopardized in the opposite case.

In the second transition control concept, the approach phase is realized in the *position control mode*; after contact is established, the control is switched to force control mode. Numerous discontinuous transition control algorithms have recently been tested. By treating the discontinuous controller as an entire generalized system, Mills and Lokhorst[83] proposed a discontinuous control scheme that guarantees global asymptotic stability of the closed-loop system, asymptotic trajectory tracking of position and force inputs, and reestablishment of contact after an inadvertent loss.

Wu et al.[84] proposed the addition of a positive acceleration feedback to the force control in the impact direction. In addition, a switching control strategy is introduced to eliminate unexpected bouncing. A similar control strategy for the transition problem in both force and impedance control has been developed by Volpe and Khosla.[85] They recommended use of positive force feedback during transition, and integral force control after stable contact is established. A force-regulated switch triggers the transition from position control to impact control. For the further switch to integral force control, several options were proposed. Based on the equivalency of force and impedance control, the authors established the transition stability condition for the impedance control-imposing ratio (robot inertia through target inertia) at less than one. Using a direct drive robot at very high impact velocity (0.7 m/s) with a relatively stiff environment (10^4 N/m), the authors demonstrated the reliability of established criteria. However, these results are not applicable to industrial robots, with high Cartesian inertia levels (> 500 kg), very stiff position controllers, and time lags that cause switching algorithms to be critical.

Gorinevski et al.[86] examined the transition problem of both impedance and general force control during contact with stationary and dynamic environments. They tested linear control and sliding mode control. The influence of several effects, such as time delay and elasticity of robot end effector, transmissions and mechanical structure, on the contact stability has been examined. The contact stability criteria for single and two DOF systems are derived in the explicit closed form in terms of control gains and limits on robot and environments velocities.

Several authors consider transition control a short-impulse dynamic problem. This model is valid for very fast systems (e.g., micro–macro manipulators), but is seldom used in practice. In industrial robotic systems, the transition problem can be accurately analyzed in a finite time period. Most industrial control systems still do not provide mechanisms to control short-impulse impact effects.

McClamroch and Wang[32] emphasized the importance of constraints in constrained dynamics. They presented global conditions for tracking based on a modified computed torque and local conditions for

© 2002 by CRC Press LLC

feedback stabilization using a linear controller. The closed-loop properties of force disturbances, dynamics in the force feedback loops, and uncertainties in constrained functions were also investigated. Eppinger and Seering[87] have studied the influence of unmodeled dynamics on contact task stability, introducing additional (elastic) degrees of freedom of both the robot and the environment.

A treatment of the contact stability, considering the environment as a nonlinear dynamic system is given by Hogan.[34] It is shown that if the impedance control is applied, enabling the robot to be asymptotically stable in free space, the robot interacting with the environment is a passive system and is stable in isolation. However, the conclusion is valid only if the robot in contact is at rest and for this reason the result cannot be considered complete. The stability issue, i.e., the establishment of the conditions under which a particular control law guarantees the stability of the robot in contact with the environment, is of great importance.

Vukobratović and Ekalo,[8,33] and Vukobratović[88] focused attention on control laws that simultaneously stabilize the motion of the robot and the forces of its interaction with a dynamic environment, ensuring the exponential stability of the closed-loop systems (based on the analysis of a complete dynamic model of the robot and the dynamic environment). The papers formulate conditions ensuring an asymptotically stable position of the system in the first approximation (local stability). The character of the position stability depends particularly on the nature of the programmed (desired) motion. In spite of sufficient conditions of the linearized system, asymptotic stability is conservative, and the dynamic character of the interaction of the environment with the robot can lead to positional instability. This problem deserves the full attention of researchers and designers of robot controllers dedicated to diverse contact tasks. This linear analysis provides very important criterion that must be fulfilled by any force-based law. However, the model uncertainties representing a crucial problem in control of robots interacting with a dynamic environment still have not been appropriately addressed. Therefore, it can be difficult to achieve the asymptotic (exponential) stability of the system (unless robust control laws including factors for compensationg these perturbations and uncertainties are used). Inaccuracies of robot and environment dynamic models and dynamic control robustness have been considered by Ekalo and Vukobratović.[89-91]

Problems arising from parameter uncertainties may also be resolved by applying knowledge-based techniques.[92] Taking into account external perturbations and model uncertainties, it may be difficult to achieve asymptotic (exponential) stability. Therefore, it is of practical interest to require less restrictive stability conditions, i.e., to consider the so-called practical stability of the system. An approach to analysis of the practical stability of manipulation robots interacting with a dynamic environment based on a centralized model of the system is presented by Stokić and Vukobratorić.[93] The test conditions for practical stability of the robot interacting with dynamic environment are recently derived.[93,94] The presented tests might be too conservative due to the number of linearizations (approximations) made. More refined approximations by, e.g., taking into account possible dependencies of the model elements (matrix of inertia, Coriolis forces, Jacobian matrix, etc.) on the parameters, may lead to less conservative tests.

23.8 Synthesis of Impedance Control at Higher Control Levels

Although several sophisticated control strategies have been proposed, the numbers of advanced robotic contact task applications remain insignificant. The reason is that most new concepts concern particular problems and the integration of the required algorithms and control concepts is tedious. Most of the studies on impedance and force control relate to servo control. Except for the seminal works on compliance control,[1,48] contact task planning and programming issues have been somewhat neglected in the research studies.

The next section briefly addresses solving high-level impedance control problems in industrial robotic systems. The problems with impedance control motion planning and programming layers were investigated during the development of the new space robot control system (SPARCO),[95] in order to develop a completely integrated reliable impedance control system including control,

© 2002 by CRC Press LLC

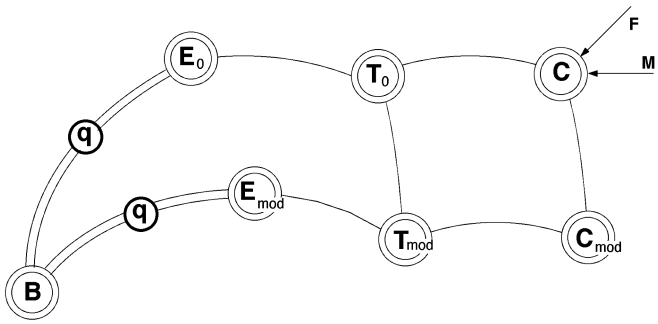


FIGURE 23.23 Compliance frame and position modification.

programming, and monitoring functions. Several control algorithms for the integration of a position-based impedance control scheme in conventional industrial robot control systems will be presented.

23.8.1 Compliance C-Frame

One of the most common approaches for word modeling (specification of robot and object positions) in robot programs is based on coordinate frames. Beside coordinate frames that are convenient for the programming of robot motion in the free space (e.g., robot base B, end point E, tool T frame, etc.), a compliance control includes two new frames, specific for compliant motion programming: force sensing S and compliance frame C. The S frame is a force sensor-specific frame in which the forces and torques are measured. This frame is commonly defined relative to the robot end point E. With respect to the C frame, the target impedance behavior (robot impedance reaction) is specified and controlled. Since the location of the C frame depends on the current task, we have chosen to specify the convenient C frame relative to task T frame (Figure 23.23) taking into account that the T frame is a variable selected to meet specific task motion requirements. Usually the desired robot position specifies the location of the T frame (specified relative to the E frame, e.g., tip of the tool) with respect to an object frame.

The basic specification of compliance control required for the implementation addresses both the definition of C frame location with respect to the tool frame and the selection of appropriate target impedance parameters in the C frame based on the geometric model (Figure 23.7) for the specific task. A set of associated robot programming commands is needed to handle this specification in the robot program.

A specific problem in the position-based impedance control is the computation of the position correction $\Delta\mathbf{x}_F$ (Figure 23.11) corresponding to the interaction force. An obvious approach is to compute this modification in the C frame where the compliance behavior is specified. While the computation of $\Delta\mathbf{x}_F$ for the translational DOFs is straightforward, there are several possibilities for managing the rotational difference dependent on the representation of orientation. The description of orientation affects the relationship between the position displacement $\mathbf{x}_0 - \mathbf{x} = \mathbf{p}_0 - \mathbf{p}$ and the forces/torques. The form of the target impedance matrices (Equation 23.14) in Cartesian space also depends on rotation representation. The SPARCO control system approach[95,96] utilizes the angle axis orientation description in the C frame. The compliance model (Figure 23.7) is based on ideally decoupled translational and rotational stiffness (impedance). This is indeed an idealized representation. As demonstrated by Lonari,[97] a point at which translational and rotational elasticity are completely decoupled does not always exist for a compliantly supported spatial rigid body. The point at which they are maximally decoupled is referred to as the *center of stiffness*. The simple SPARCO approach allows the rotational impedance parameters to be directly related to the task geometry described in the C frame, i.e., in the tool-frame T_0. The selection of the C frame location is based on geometric task analysis in T_0 and consideration of the force/displacement equilibrium.[96]

© 2002 by CRC Press LLC

Additional possible rotation description models for compliance control are presented by Caccavale et al.[98] Several authors considered achieving target stiffness with linear and rotational springs,[99] spring systems,[100] and serial elastic mechanisms.[101] Besides the synthesis of passive compliance devices, these works are relevant for the better understanding of spatial compliant behavior in robot contact interaction tasks. A more general compliance planning (impedance selection) approach has recently been proposed by Fasse and Broenink.[102] This approach is based on *spatial affine* presentation of compliance described by sets of geometrical and nongeometrical parameters independent of robot/object configuration. A spatially affine family of compliances provides invariant wrenches with respect to the rigid body transformation. This simplifies the selection of the target impedance.

23.8.2 Operating Modes

Additional control functions and commands are required to manage dynamic communication between the basic position control system and the impedance control module. This section describes impedance control functions that support programming compliance control applications.

Although the impedance control feedback is activated and deactivated by contact with the environment (Figure 23.11), the mechanisms must functionally manage the activation of the control modules and functions. It is convenient to introduce the following impedance control operating modes:

- *Stopped mode*, with no impedance control functions performed. This mode corresponds to conventional position control systems.
- *Monitoring mode*, performing monitoring functions in real control time such as monitoring sensor and process force/torque limits, checking position correction bounds, contact check, collision detection, end effector monitoring functions, etc.
- *Running mode*, in which monitoring and exteroceptive control functions (computation of impedance control loop) are executed at each sampling interval.

In the initial stopped status, the impedance control can be initialized by selecting the sensor type and transferring the control and machine data to the local control functions. Before activating the impedance control functions, the status is changed to monitoring mode. This mode allows a change to task-dependent control parameters such as tool and compliance frame locations, impedance gains, contact force limits, and desired force still in contact with the environment, without deactivating the external control. An algorithm, referred to as the *relax* control function,[95] has been used to meet the conditions for a continuous change of impedance control parameters.

Once the running mode has been started, all subsequent robot motions are automatically modified by the corrections corresponding to the force and selected impedance target gains. Any transition to the monitoring or stopped modes automatically resets position correction offset (position synchronization) by replacing the nominal robot position in the interpolator by the actual current robot position. This allows the robot motion to continue in the position control mode starting from the actual position.

23.8.3 Change of Impedance Gains — Relax Function

Since the impedance control parameters define a desired target mechanical system, they should often be changed, depending on the current action (some actions, such as *insert,* require target impedance to be changed several times during execution) to meet action-specific requirements. The location of the compliant C frame in which the impedance is controlled must also be varied. The main problem with the parameter changes in a control system is achieving a continuous, transient switching (bumpless parameter change).

When the robot is in contact with the environment, any discontinuous parameter change can cause control chattering due to interruption of the impedance force/motion relationship, for example,

© 2002 by CRC Press LLC

by alteration of gains (i.e., variation of stiffness) or compliance frame location (change of force and torque components). The control gain switching can be especially critical in directions in which high stiffness level is replaced by low stiffness and vice versa. In free space, however, impedance parameter switching is not critical since the impedance feedback loop is inactive due to zero contact force.

The way to obtain bumpless parameter changes is to achieve similar contact conditions as in free space by reducing the interaction force to a minimum level while maintaining contact with the environment. This is realized by the relax built-in program language functions consisting of the following steps:

Step 1: Switch the impedance control to the monitoring mode and reset position correction offset.
Step 2: Settle the damping control gains (target stiffness is zero) in all C frame directions and desired force to zero.
Step 3: Switch the impedance control to the running mode.
Step 4: Due to contact force, the robot moves until the given small force threshold is reached in all directions during a selected time period.
Step 5: Switch the impedance control again to the monitoring and reset offset and initial impedance gains.

The main issue of the *apply-force* function is to realize a specified steady state force. In impedance control, the force can generally be regulated only in an open loop by proprietary generation of robot motion in accordance with the selected target system. However, this approach requires accurate knowledge of contact point location, and is sensitive to disturbances (e.g., friction forces). In the SPARCO-implemented approach, a constant force is achieved in the closed loop based on the damping control that provides a correction velocity proportional to the error between actual and desired forces (integral force control). Thus, the robot corrects its position in corresponding directions until the force error becomes less than the given threshold.

23.8.4 Impedance Control Commands

Program instructions are required to allow the programmer to manage impedance control parameters and monitor and handle various contact exceptions. These instructions select impedance control gains, read the contact force, check the force limits and contact with the environment, and indicate when the desired force is achieved. In conjunction with the standard robot programming language commands, exception handling, and motion synchronization mechanisms, these devices provide a powerful framework for programming complex impedance control algorithms. These new commands are specified by Šurdilović.[95]

Setting impedance control parameters is done in an implicit manner by using understandable linguistic descriptions using high, medium, or low attributes in conjunction with target impedance or damping models. To each description case a set of gains is designed for SPARCO applications in a CAT environment and put in a look-up table initialized during system setup. The user can select an individual set of control gains that should be read from a specified file. A built-in function is provided to facilitate the selection of user gains (set/user/gf). The control gains for the impedance controller are put in the system memory (look-up table).

23.8.5 Control Algorithms

A major question in using sensor-based robot control is how to apply sensory information to perform a given task in the presence of uncertainty and errors. This requires new algorithms to be developed to predict and detect various events and generate corresponding reliable robot actions. Facilitating the use of sensory information, robot-programming language should provide the mechanisms to

© 2002 by CRC Press LLC

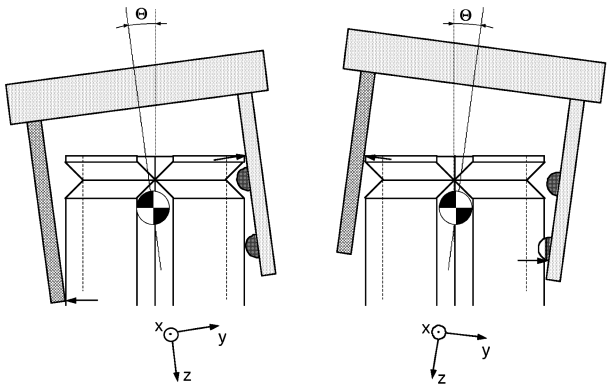


FIGURE 23.24 Grasping using impedance effect.

easily access sensory information and use it to control the robot. Two examples are presented to illustrate the practical contact control task using new impedance control functions and commands.

23.8.5.1 Grasping

Nominal relative configuration (position) between end effector and object before gripper closing with relative high accuracy must be achieved to grasp a fixed object properly. The grasping algorithm based on impedance control allows the gripper to self-align to the fixed objects to be grasped, and thus compensates for inaccuracies in environment and robot control.
The following basic specification of the impedance control parameters is required to meet the grasping requirement:

- The compliance C Frame is located approximately at the gripper middle point between the hemispheres (Figure 23.24) used to support the gripper self-alignment along grooves, based on the impedance control effect.
- Low impedance is selected in each direction.

To achieve a good centralized grasp (the compliance center is located near the grapple fixture axis), internal grasp forces are approximately balanced, and the resulting torque causes a rotation about the center of compliance. In an opposite case, lateral misplacements are dominant, causing a one-sided contact. The lateral force and corresponding torque around the compliance center cause the robot to correct initial position and orientation and both jaws grasp the object (Figure 23.24). The robot moves until the desired internally stable grasp is achieved. (The jaw hemispheres and grapple fixture notches in conjunction with contact friction prevent further motion.

23.8.5.2 Insertion

Industrial programmers have experience with the insertion function using RCC passive elements. The impedance control provides a similar approach. Moreover, the control system capabilities that change impedance control gains or compliance frames in various task phases also provide a programmable compliance device.

The insertion (Figure 23.25) control algorithm requires three procedures: engagement, insertion, and termination. The selection of impedance control gains in these phases is shown in (Figure 23.26). Engagement requires that part chamfers meet and slide past one another. The following impedance control specifications apply to the engagement phase:

- The C frame should be located near interacting force directions (on the peg top).
- The insertion (i.e., engagement) motion consists of a linear displacement in the positive z direction along the hole axis. The target position is chosen below the nominal front surfaces (below the ends of the chamfers).

© 2002 by CRC Press LLC

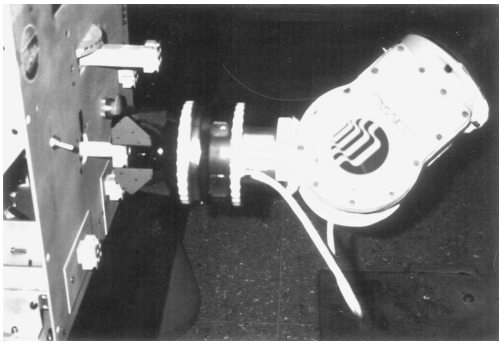


FIGURE 23.25 Insert task.

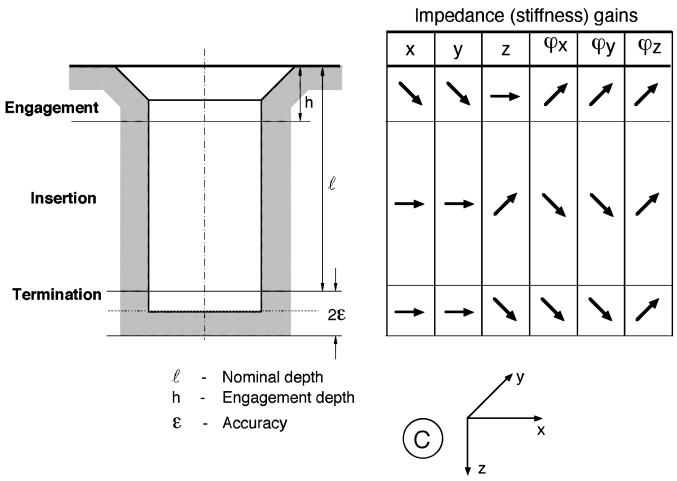


FIGURE 23.26 Impedance gains selection.

Determining a medium stiffness (Figure 23.26) in the z direction during engagement should take into account that the robot motion is unconstrained in this direction. To slide along the chamfers, the stiffness in lateral directions (x, y) must be less than the axial stiffness, causing a faster misalignment than encountering the surface (the lateral force/stiffness behavior serves as a cue to the desired corrective motion). This also reduces contact forces (Figure 23.26), allowing the peg to cope easily with the friction. A high rotational stiffness is required, taking into account the goal to compensate for lateral misalignment only, without introducing unwanted rotation.

After engagement depth is reached, the insertion procedure is started. The lateral rotation impedance (stiffness) around the x and y C frame directions is switched (after relaxing the contact) to low to compensate for the rotation error. This impedance in the insertion direction (z) along the hole is set to medium since the robot motion is unconstrained in this direction and is required to compensate for the disturbing friction forces between the peg and hole during insertion. The insertion phase ends when a termination pose in front of the peg bottom is reached. It is selected dependent on estimated position tolerances to avoid contact with the bottom of the hole. The C frame should be located near the middle point of the hole to relax the peg (similar to Figure 23.24) and change gains for the next phase. The termination phase algorithms are dependent on the termination strategy (governed by position, external signal, or force). In each case the low impedance is selected in the insertion direction to avoid large contact forces due to inaccuracies.

© 2002 by CRC Press LLC

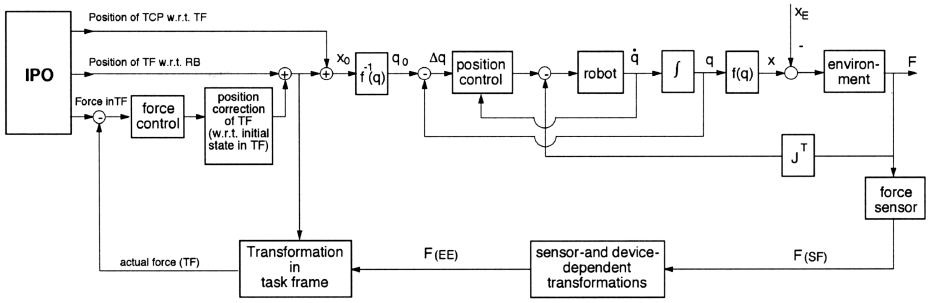


FIGURE 23.27 Integration of implicit hybrid control.

23.8.6 Implicit Force Control Integration

A practical method to integrate implicit force control and orthogonal decomposition in a position control system for robot manufacturing applications (e.g., deburring) was proposed by Lonari.[97] The hybrid control integration is realized at motion interpolation level (Figure 23.27). In order to decompose position and force control DOF, two virtual kinematic structures in Cartesian space are assumed: the first driven by the position interpolator and the second obtaining the position commands (i.e., relative modifications) from the force control in the compliance i.e., task frame TF. The interpolator (IPO) generates the nominal position of the tool center point (TCP) in the position control DOF in the task frame. The integral implicit force control modifies the position of TF in force-controlled directions relative to its initial location at the robot reference frame (e.g., robot base frame, object frame) according to the force error. Superimposition of both positions produces the command for the internal position controller $\mathbf{x}_0$. Independent of modifications of the tool position and/or orientation in force direction (e.g., tool feed direction) based on interaction forces, it is possible to realize the desired motion in position-controlled directions, such as changing the tool clearance in toll directions etc.

23.9 Conclusion

During the past several years, compliant motion control has emerged as one of the most attractive and useful research areas in robotics. The control of constrained motion of robots is challenging and a successful solution will affect considerably use of robots in industry and increase their efficiency and productivity.

It may be of interest to indicate possible future investigation areas. Clear formulation and specification of hybrid control are required. Further simulation and experimental tests of recently proposed compliant motion control algorithms, such as parallel position force control, adaptive and variable structure algorithms, and, particularly, dynamic control of a robot interaction with a dynamic environment, are also of interest. Effort should also be devoted to solving the problem of contact task control where the uncertainties and nonlinear effects in the environment and the robotic system, such as friction, multipoint contact, elasticity, etc. must be considered. In impedance control, further advances are expected in adaptation of target impedance to complex task requirements. The compliant motion capability analysis of industrial robots and requirements for the next generation are of interest to designers. Robust control continues to be the focus of control design. Comparison of available algorithms, definition of benchmark tests, investigation of compliant control in uncertain and dynamic environments, examination of nonlinear effects in robot and environment dynamics, and solving control problems at higher control levels certainly deserve further computational/experimental studies.

© 2002 by CRC Press LLC

References

1. Mason, M.T., Compliance and force control for computer controlled manipulators, *IEEE Trans. Syst., Man, Cybernetics,* SMC 11(4), 1981.
2. Pfeifer, F., Richter, K., and Wapenhans H., Elastic robot trajectory planning with force control, *Proc. IFIP Int. Symp. Theory Robots*, Rome, 412–418, 1990.
3. Whitney, D.E. and Nevins, J.L., What is remote centre compliance (rcc) and what can it do, in *Robot Sensors, Volume 2: Tactile and Non-Vision*, Springer-Verlag, Berlin, 1986, 3–17.
4. Cutkosky, M.R. and Wright P.K, Active control of a compliant wrist in manufacturing tasks, in *Robot Sensors, Volume 2: Tactile and Non-Vision*, Springer-Verlag, Berlin, 1986, 17.
5. Simons, J. and Van Brussel, H., Force control schemes for robot assembly, international trends, in *Manufacturing Technology: Robotic Assembly,* Springer-Verlag, Berlin, 1985, 253.
6. Whitney, D.E., Historical perspective and state of the art in robot force control, *Int. J. Robotic Res.,* 6(1), 3–7, 1987.
7. Raibert, M.H. and Craig, J.J., Hybrid position/force control of manipulators, *ASME J. Dynamic Syst., Meas. Control,* 102, 126–133, 1981.
8. Vukobratović, M. and Ekalo, Y., Unified approach to control laws: synthesis for robotic manipulators in contact with dynamic environment, Tutorial SS, *Proc. IEEE Int. Conf. Robotics and Automation,* 213–229, 1993.
9. Chiaverini, S. and Sciavicco, L., Force/position control of manipulators in task space with dominance in force, *Proc. 2nd IFAC Symp. Robot Control,* Karlsruhe, 51.1–51.7, 1988.
10. Wedel, D.L. and Saridis, G.N., An experiment in hybrid position/force control of a six DOF revolute manipulator, *Proc. IEEE Int. Conf. Robotics Automation,* 1638–1642, 1988.
11. Yoshikawa, T., Sugie, T., and Tanaka, M., Dynamic hybrid position/force control of robot manipulators: controller design and experiment, *IEEE J. Robotics and Automation,* 4(6), 1988.
12. Anderson, R. and Spong, M., Hybrid impedance control of robotic manipulators, *IEEE J. Robotics and Automation,* (4)5, 1988.
13. De Schutter, J. and Van Brussel, H., Compliant robot motion II: a control approach based on external control loops, *Int. J. Robotic Res.,* 7(4), 18–33, 1988.
14. Lawrence, D.A., Impedance control stability properties in common implementations, *Proc. IEEE Int. Conf. Robotics and Automation,* 1185–1190, 1988.
15. Maples, J.A. and Becker, J.J., Experiments in force control of robotic manipulators, *Proc. IEEE Int. Conf. Robotics Automation,* 695–703, 1986.
16. Kazerooni, H., Waibel, B.J., and Kim, S., On the stability of robot compliant motion control: theory and experiments, *J. Dynamics Syst., Meas. Control,* 112, 417–426, 1990.
17. Salisbury, J.K., Active stiffness control of a manipulator in cartesian coordinates, *Proc. 19th IEEE Conf. Decision Control,* 95–100, 1980.
18. Whitney, D.E., Force feedback control of manipulator fine motions, *ASME J. Dynamics, Syst., Meas. Control,* 91–97, 1977.
19. Hogan, N., Impedance control: an approach to manipulation, *J. Dynamics Syst., Meas. Control,* Vol. 107, 1–24, 1985.
20. Kazerooni, K.H., Waibel, B.J., and Kim, S., On the stability of robot compliant motion control: Theory and experiments, *J. Dynam. Syst., Meas. Control,* 112, 417–426, 1990.
21. Eppinger, S.D. and Seering, W.P., Understanding bandwidths limitation in robot force control, *Proc. IEEE Int. Conf. Robotics Automation,* Raleigh, NC, 904–909, 1987.
22. Khatib, O., A unified approach for motion and force control of robot manipulators: the operational space formulation, *IEEE J. Robotics Automation,* 3(1), 1987.
23. De Luca, A., Manes, C., and Nicolo, F., A task space decoupled approach to hybrid control of manipulators, *Proc. 2nd IFAC Symp. Robot Control,* Karlsruhe, 1988, 270–277.
24. Lipkin, H. and Duffy, J., Hybrid twist and wrench control for a robotic manipulator, *Trans. ASME J. Mech., Trans. Automation Design,* 110, 138–144, 1988.
25. De Shutter, J. and Bruyninckx, H., Model-based specification and execution of compliant motion, *Proc. IEEE Conf. Robotics Automation,* Tutorial M6, Nice, 1992.

© 2002 by CRC Press LLC

26. Zhang, H. and Paul, R., Hybrid control of robot manipulator, *Proc. IEEE Int. Conf. Robotics,* 602–607, 1985.
27. Daneshmend, L., Hayward, V., and Pelletier, M., *Adaptation to Environment Stiffness in the Control of Manipulators, Experimental Robotics I,* (Hayward, K. and Khatib, O., Eds.,), Springer-Verlag, Berlin, 1990, 150–156.
28. Stokić, D., Vukobratović, M., and Šurdilović, D., An adaptive hybrid control scheme for manipulation robots with implicit force control, *Proc. V ICAR,* Pisa, 1505–1508, 1991.
29. Kazerooni, K.H., Sheridan, T., and Houpt, P., Robust compliant motion for manipulators, part i: the fundamental concepts of compliant motion; part ii design method, *IEEE J. Robotics Automation,* RA2(2), 83–106, 1986.
30. Dawson, D.M., Lewis, F.L., and Dorsey, J.F., Robust force control of a robot manipulator, *Int. J. Robotic Res.,* 11(4), 1992.
31. Shin, K.G. and Lee, C.P., Compliant control of robotic manipulators with resolved acceleration, *Proc. 24th IEEE Conf. Decision Control,* 350–357, 1985.
32. Mc Clamroch, H. and Wang, D., Feedback stabilization and tracking of constrained robots, *IEEE Trans. Automatic Control,* 33(5), 1988.
33. Vukobratović, M. and Ekalo, Y., New approach to control of robotic manipulators interacting with dynamic environment, *Robotica,* 14, 31–39, 1996.
34. Hogan, N., On the stability of manipulator performing contact tasks, *IEEE J. Robotics Automation,* 4(6), 1990.
35. An, C.H. and Hollerbach, J.M., The role of dynamic models in Cartesian force control, *Int. J. Robotics Res.,* 8(4), 1989.
36. Kircanski, N., Computational aspects of force/impedance control algorithms, Vukobratović, M. and Nakamura Y., Eds., *IEEE Int. Conf. Robotics Automation,* Tutorial S5, 33–48, 1993.
37. De Luca, A. and Manes, C., Modeling of robots in contact with a dynamic environment, *IEEE Trans. Robotics Automation,* 10(4), 542, 1994.
38. Goldenberg, A.A., Force and impedance control of robot manipulators, *Proc. Winter Annu. Meet. ASME,* 161–168, 1987.
39. Sharon, A., Hogan, N., and Hardt, D., More analysis and experimentation of a micro/macro manipulator system, *Proc. ASME Winter Annu. Meet.,* Boston, 417–422, 1987.
40. Vukobratović, M. and Stojić, R., On position/force control of robot interacting with dynamic environment in cartesian space, *J. Dynamic Syst., Meas. Control,* 118, 187, 1996.
41. Goldenberg, A.A., Analysis of force control based on linear models, *Proc. IEEE Int. Conf. Robotics Automation,* Nice, 1348–1353, 1992.
42. Šurdilović, D., Contact stability issues in position based impedance control: theory and experiments, *Proc. IEEE Int. Conf. Robotics Automation,* Minneapolis, 1675, 1996.
43. Chernousko, F.L., Bolotnik, N.N., and Gradetsky, V.G., *Manipulation Robots: Dynamics, Control, and Optimization,* CRC Press, Boca Raton, FL, 1994.
44. De Fazio, T.L., Seltzer, D.S., and Whitney, D.E., The IRCC instrumented remote center compliance, in *International Trends in Manufacturing Technology: Robotic Assembly,* Springer-Verlag, 1985, 33.
45. Mason, M.T and Salisbury, J.K., *Robot Hands and the Mechanics of Manipulation,* MIT Press, Cambridge, 1985.
46. Lee, S. and Lee, H.S., Generalized impedance of manipulators: its application to force and position control, *Proc. Fifth Int. Conf. Adv. Robotics ICAR '91,* 1477–1480, 1991.
47. Lu, Z. and Goldenberg, A.A., Robust impedance control and force regulation: theory and experiments, *Int. J. Robotics Res.,* 14(3), 1995.
48. Hogan, N., Stable execution of contact tasks using impedance control, *Proc. IEEE Int. Conf. Robotics Automation,* Raleigh, NC, 1047, 1987.
49. Vukobratović, M., Stokić, D., and Kircanski, N., *Non-Adaptive and Adaptive Control of Manipulation Robots,* Scientific Fundamentals of Robotics 5, Springer-Verlag, Berlin, 1985.
50. Visher, D. and Khatib, O., Design and development of torque-controlled joints, *Proc. First Int. Symp. Exp. Robotics,* 271–276, 1989.

© 2002 by CRC Press LLC

51. Šurdilović, D., Anton, S., and Al-Keshmery, A., *Compliant Motion Control Concept for Space Robotics for Upgrading of the ESTEC Robotic Laboratory Controller (IMCO),* ESA/ESTEC Contract 103929, Study Report, IPK-Berlin, 1993.
52. Šurdilović, D., Bernhardt, R., Columbina, G., and Grassini, F., A space robot control with advanced sensor-based capabilities: terrestrial spin-off, *Proc. 27th ISIR,* Milan, 243, 1996.
53. Šurdilović, D., *Compliance Control Design in Industrial Robotic Systems,* Ph.D. Thesis, University of Niš, Yugoslavia, 2001.
54. Craig, J.J., Hsu, P., and Sastry, S.S., Adaptive control of mechanical manipulators, Int. *J. Robotic Res.,* 6(2), 1987.
55. Al-Jarrah, O.M. and Zheng, Y.F., Intelligent compliant motion control, *Proc. IEEE Int. Conf. Robotics Automation,* Minnesota, 2610–2615, 1996.
56. Paul, R.P. and Shimano, B., Compliance and control, *Proc. Joint Autom. Control Conf.,* San Francisco, 694–699, 1976.
57. Stokić, D.M. and Šurdilović, D.T., *Position/Force Control of Industrial Robots: Implementation Problems,* IPK-Berlin, 1990.
58. An, C.H. and Hollerbach, J., Kinematic stability issues in force control of manipulators, *Proc. IEEE Int. Conf. Robotics Automation,* Raleigh, 897–903, 1987.
59. Zhang, H., A kinematically stable hybrid control formulation, *Proc. 3rd Int. Symp. ISRAM '90,* British Columbia, 287–292, 1990.
60. Fisher, W.D. and Mujtaba, M.S., Hybrid position/force control: A correct formulation, *Int. J. Robotics Res.,* 11(4), 1992.
61. An, C.H. and Hollerbach, J., Dynamic stability issues in force control of manipulators, *Proc. IEEE Int. Conf. Robotics Automation,* Raleigh, 890–896, 1987.
62. Luh, J.Y.S., Fisher, W.D., and Paul, R.P.C., Joint torque control by a direct feedback for industrial robots, *IEEE Trans. Autom. Control,* AC-28, 153–161, 1983.
63. Bona, B. and Idri, M., Exact decoupling of the force-position control using the operational space formulation, *Proc. IEEE Int. Conf. Robotics Automation,* Nice, 1394–1398, 1992.
64. Volpe, R. and Khosla, P., An experimental evaluation and comparison of explicit control strategies for robotic manipulators, *Proc. IEEE Int. Conf. Robotics Automation,* Nice, 1382–1386, 1992.
65. Stokić, D.M. and Šurdilović, D.T., Simulation and control of robotic deburring, *Int. J. Robotics Automation,* 5(3), 107, 1990.
66. Yoshikawa, T. and Sudou, A., Dynamic hybrid position/force control of robot manipulators: estimation of unknown constraint, *Proc. IEEE Int. Conf. Robotics Automation,* 1231, 1990.
67. Bruyninckx, H., De Schutter, J., and Allota, B., Model-based constrained motion: introducing on-line identification and observations of motion constraints, *Proc. 5th Int. Conf. Adv. Robotics, ICAR '91,* Pisa, 976–981, 1991.
68. Sinha, P. and Goldenberg, A.A., A unified theory for hybrid control of manipulators, *Proc. IEEE Int. Conf. Robotics Automation,* Atlanta, 343–348, 1993.
69. Volpe, R. and Khosla, P., The equivalence of second-order impedance control and proportional gain explicit force control, *Int. J. Robotics Res.,* 14(6), 574, 1995.
70. Šurdilović, D., and Kirchhof, J., A new position based force/impedance control for industrial robots, *Proc. IEEE Int. Conf. Robotics Automation,* Minneapolis, 629–634, 1996.
71. Mayeda, H., Ikeda, N., and Miyaji, K., Position/force/impedance control for robot tasks, *Proc. 2nd ISMCR '92,* Tsukuba, Japan, 551–558, 1992.
72. Chiaverini, S., Sicilliano, B., and Villani, L., Force/position regulation of compliant robot manipulators, *IEEE Trans. Aut. Control,* 39(3), 1994.
73. Vukobratović, M., Stojić, R., and Ekalo, Yu, Contribution to the problem solution of position/force control of manipulation robots in contact with dynamic environment – a generalization, *IFAC Automatica,* 34(10), 1998.
74. Colgate, J.E. and Hogan, M., Robust control of dynamically interacting systems, *Int. J. Control,* 48(1), 1988.
75. Chapel, J.D. and Su R., Coupled stability characteristics of nearly passive robots, *Proc. IEEE Int. Conf. Robotics Automation,* Nice, 1342–1346, 1992.

© 2002 by CRC Press LLC

76. Šurdilović, D., Contact transition stability in the impedance control, theory and experiments, *Proc. IEEE Int. Conf. Robotics Automation,* Albuquerque, 847, 1997.
77. Benady, M., Flash, T., and Gershon, D., Robot learning of contact tasks, in *Robotic Systems and AMT*, Halevi, G., Yudilevich, I., and Weill, I., Eds., Elsevier Science Publishers, Amsterdam, 147, 1990.
78. Vossoughi, R. and Donath, M., Robot hand impedance control in the presence of mechanical non-linearities, *Proc. Winter Annu. Meet. ASME,* 67, 1985.
79. Xu, Y., Paul, R.P., and Corke, P.I., Hybrid position force control of robot manipulator with an instrumented compliant wrist, *Proc. 1st Int. Symp. Exp. Robotics,* Springer-Verlag, 244, 1990.
80. Quian, H.P. and De Schutter, J., Introducing active linear and nonlinear damping to enable stable high gain force control in case of stiff contact, *Proc. IEEE Int. Conf. Robotics Automation,* Nice, 1374, 1992.
81. Siciliano, B. and Villani, L., A force/position regulator for robot manipulators without velocity measurements, *Proc. IEEE Int. Conf. Robotics Automation,* 2567, 1996.
82. De Quieroz, M., Dawson, D., and Burg, T., Position/force control of robot manipulators without velocity/force measurements, *Proc. IEEE Int. Conf. Robotics Automation,* 2561, 1996.
83. Mills, J. and Lokhorst, D., Control of robotic manipulators during general task execution: a discontinuous control approach, *Int. J. Robotic Research,* 12(2), 146, 1993.
84. Wu, Y., Tarn, T.J., Xi, N., and Isidori, A., On robust impact control via positive acceleration feedback for robot manipulators, *Proc. IEEE Int. Conf. Robotics Automation,* Minneapolis, 1891, 1996.
85. Volpe, R. and Khosla, P., A theoretical and experimental investigation of impact control for manipulators, *Int. J. Robotics Res.,* 12(4), 351, 1993.
86. Gorinevski, D.M., Formalski, A.M., and Šneider, A.I., *Control of Manipulation Robots Based on Force Information* Physical–Mathematical Literature, Moscow, 1994 (in Russian).
87. Eppinger, S.D. and Seering, W.P., Introduction to dynamic robot models for robot force control, *IEEE Control Syst. Mag.,* 2(2), 1987.
88. Vukobratović, M., The role of environment dynamics in contact force control of manipulation robots, *ASME J. Dynamic Syst., Meas. Control,* 119, 86, 1997.
89. Ekalo Y. and Vukobratović, M., Robust and adaptive position/force stabilization of robotic manipulators in contact tasks, *Robotica,* 11(4), 1993.
90. Ekalo Y. and Vukobratović, M., Adaptive stabilization of motion and forces in contact tasks for robotic manipulators with non-stationary dynamics, *Int. J. Robotics Automation,* 9(3), 1994.
91. Ekalo Y. and Vukobratović, M., Quality of stabilization of robot interacting with dynamic environment, *J. Intelligent Robotic Syst.,* 14(2), 1995.
92. Vukobratović, M., and Katić, D., Robust stabilizing position/force control of robots interacting with dynamic environment by learning connectionist structure, *IFAC Automatica,* 32(12), 1733, 1996.
93. Stokić, D. and Vukobratović, M., An improved method for analysis of practical stability of robots interacting with dynamic environment, *Int. J. Russ. Acad. Sci.: Theory and Control Systems,* 4, 186–292, 2000.
94. Stokić, D. and Vukobratović, M., Practical stabilization of robots interacting with dynamic environment, *Int. J. Robotics and Automation,* 13, 4, 1998.
95. Šurdilović, D., Synthesis of impedance control laws at higher control levels: algorithms and experiments, *Proc. IEEE Int. Conf. Robotics and Automation,* Leuven, 213, 1998.
96. Šurdilović, D., *Synthesis of SPARCO Control Algorithms,* ESA/ESTEC Report, Berlin, 1996.
97. Lončarić, J., Normal form of stiffness and compliance matrices, *IEEE Trans. Robotics Automation,* 3, 567, 1987.
98. Caccavale, F., Natale, C., Siciliano, B., and Villani, L., Six-DOF impedance control based on angle/axis representations, *IEEE Trans. Robotics and Automation,* 13(2), 289, 1997.
99. Ciblak, N. and Lipkin, H., Synthesis of cartesian stiffness for robotic application, *Proc. IEEE Int. Conf. Robotics Automation,* Detroit, 2147, 1999.

© 2002 by CRC Press LLC

100. Roberts, R., Minimal realization of arbitrary spatial stiffness matrix with a parallel connection of simple and complex springs, *Proc. IEEE Int. Conf. Robotics and Automation,* San Francisco, 3302, 2000.
101. Huang, S. and Schimmels, J.M., Synthesis of spatial compliances with simple serial elastic mechanisms, *Proc. IEEE Int. Conf. Robotics and Automation,* San-Francisco, 3328, 2000.
102. Fasse, E.D. and Broenink, J.F., A spatial impedance controller for robotic manipulation, *IEEE Trans. Robotics and Automation,* 13, 546, 1997.

© 2002 by CRC Press LLC

24

Intelligent Soft-Computing Techniques in Robotics

24.1 Introduction
24.2 Connectionist Approach in Robotics
Basic Concepts • Connectionist Models with Applications in Robotics • Learning Principles and Rules
24.3 Neural Network Issues in Robotics
Kinematic Robot Learning by Neural Networks • Dynamic Robot Learning at the Executive Control Level • Sensor-Based Robot Learning
24.4 Fuzzy Logic Approach
Introduction • Mathematical Foundations • Fuzzy Controller • Direct Applications • Hybridization with Model-Based Control
24.5 Neuro-Fuzzy Approach in Robotics
24.6 Genetic Approach in Robotics
24.7 Conclusion

Dustic M. Katić
Mihajlo Pupin Institute

Branko Karan
Mihajlo Pupin Institute

24.1 Introduction

Robots and machines that perform various tasks in an intelligent and autonomous manner are required in many contemporary technical systems. Autonomous robots have to perform various anthropomorphic tasks in both unfamiliar or familiar working environments by themselves much like humans. They have to be able to determine all possible actions in unpredictable dynamic environments using information from various sensors. In advance, human operators can transfer to robots the knowledge, experience, and skill to solve complex tasks. In the case of a robot performing tasks in an unknown enviroment, the knowledge may not be sufficient. Hence, robots have to adapt and be capable of acquiring new knowledge through learning. The basic components of robot intelligence are actuation, perception, and control. Significant effort has been attempted to make robots more intelligent by integrating advanced sensor systems as vision, tactile sensing, etc. But, one of the ultimate and primary goals of contemporary robotics is development of intelligent algorithms that can further improve the performance of robotic systems, using the above-mentioned human intelligent functions.

Intelligent control is a new discipline that has emerged from the classical control disciplines with primary research interest in specific kinds of technological systems (systems with recognition

© 2002 by CRC Press LLC

in the loop, systems with elements of learning and self-organization, systems that sometimes do not allow for representation in a conventional form of differential and integral calculus). Intelligent control studies high-level control in which control strategies are generated using human intelligent functions such as perception, simultaneous utilization of a memory, association, reasoning, learning, or multi-level decision making in response to fuzzy or qualitative commands. Also, one of the main objectives of intelligent control is to design a system with acceptable performance characteristics over a very wide range of structured and unstructured uncertainties.

The conditions for development of intelligent control techniques in robotics are different. It is well known that classic model-based control algorithms for manipulation robots cannot provide desirable solutions, because traditional control laws are, in most cases, based on a model with incomplete information and partially known or inaccurately defined parameters. Classic algorithms are extremely sensitive to the lack of sensor information, unplanned events, and unfamiliar situations in robots' working environment. Robot performance is not able to capture and utilize past experience and available human expertise. The previously mentioned facts and examples provide motivation for robotic intelligent control capable of ensuring that manipulation robots can sense the environment, process the information necessary for uncertainty reduction, and plan, generate, and execute high-quality control action. Also, efficient robotic intelligent control systems must be based on the following features:

1. Robustness and great adaptability to system uncertainties and environment changes
2. Learning and self-organizing capabilities with generalization of acquired knowledge
3. Real-time implementation on robot controllers using fast processing architectures

The fundamental aim of intelligent control in robotics represents the problem of uncertainties and their active compensation. Our knowledge of robotic systems is in most cases incomplete, because it is impossible to describe their behavior in a rigorous mathematical manner. Hence, it is very important to include learning capabilities in control algorithms, i.e., the ability to acquire autonomous knowledge about robot systems and their environment. In this way, using learning active compensation of uncertainties is realized, which results in the continous improvement of robotic performances. Another important characteristic that must be included is knowledge generalization, i.e., the application of acquired knowledge to the general domain of problems and work tasks.

Few intelligent paradigms are capable of solving intelligent control problems in robotics. In addition, symbolic knowledge-based systems (expert systems), connectionist theory, fuzzy logic, and evolutionary computation theory (genetic algorithms) are very important in the development of intelligent robot control algorithms. Also, important in the development of efficient algorithms are hybrid techniques based on integration of particular techniques such as neuro-fuzzy networks, neuro-genetic, and fuzzy-genetic algorithms.

Connectionist systems (neural networks) represent massively parallel distributed networks with the ability to serve in advanced robot control loops as learning and compensation elements using nonlinear mapping, learning, parallel processing, self-organizing, and generalization. Usually, learning and control in neurocontrollers are performed simultaneously, and learning continues as long as perturbations are present in the robot under control and/or its environment.

Fuzzy control systems based on mathematical formulation of fuzzy logic have the ability to represent human knowledge or experience as a set of fuzzy rules. Fuzzy robot controllers use human knowhow or heuristic rules in the form of linguistic if–then rules, while a fuzzy inference engine computes efficient control action for a given purpose.

The theory of evolutionary computation with genetic algorithms represents a global optimization search approach that is based on the mechanics of natural selection and natural genetics. It combines survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with expected ever-improving perfomance.

© 2002 by CRC Press LLC

The purpose of this chapter is to present intelligent techniques as new paradigms and tools in robotics. Basic principles and concepts are given, with an outline of a number of algorithms that have been shown to simulate or use a diversity of intelligent concepts for sophisticated robot control systems.

24.2 Connectionist Approach in Robotics

24.2.1 Basic Concepts

Connectionism is the study of massively parallel networks of simple neuron-like computing units.[9.19] The computational capabilities of systems with neural networks are in fact amazing and very promising; they include not only so-called "intelligent functions" like logical reasoning, learning, pattern recognition, formation of associations, or abstraction from examples, but also the ability to acquire the most skillful performance for control of complex dynamic systems. They also evaluate a large number of sensors with different modalities providing noisy and sometimes inconsistent information. Among the useful attributes of neural networks are

- *Learning*. During the training process, input patterns and corresponding desired responses are presented to the network, and an adaptation algorithm is used to automatically adjust the network so that it responds correctly to as many patterns as possible in a training set.
- *Generalization*. Generalization takes place if the trained network responds correctly with a high probability of inputting patterns that were not included in the training set.
- *Massive parallelism*. Neural networks can perform massive parallel processing.
- *Fault tolerance*. In principle, damage to a few links need not significantly impair overall performance. Network behavior gradually decays as the number of errors in cell weights or activations increases.
- *Suitability for system integration*. Networks provide uniform representation of inputs from diverse resources.
- *Suitability for realization in hardware*. Realization of neural networks using VLSI circuit technology is attractive, because identical structures of neurons make fabrication of neural networks cost-effective. However, the massive interconnection may result in some technical difficulties, such as power consumption and circuitry layout design.

Neural networks consist of many interconnected simple nonlinear systems that are typically modeled by appropriate activation functions. These simple nonlinear elements, called nodes or neurons, are interconnected, and the strengths of the interconnections are denoted by parameters called weights. A basic building block of nearly all artificial neural networks, and most other adaptive systems, is the adaptive linear combinier, cascaded by a nonlinearity which provides saturation for decision making. Sometimes, a fixed preprocessing network is applied to the linear combinier to yield nonlinear decision boundaries. In multi-element networks, adaptive elements are combined to yield different network topologies. At input, an adaptive linear combinier receives analog or digital input vector $x = [x_0, x_1, \ldots, x_n]^T$ (input signal, input pattern), and using a set of coefficients, the weight vector, $w = [w_0, w_1, \ldots w_n]^T$, produces the sum s of weighted inputs on its output together with the bias member b:

$$s = x^T w + b \tag{24.1}$$

The weighted inputs to a neuron accumulate and then pass to an activation function that determines the neuron output:

$$o = f(s) \tag{24.2}$$

© 2002 by CRC Press LLC

The activation function of a single unit is commonly a simple nondecreasing function like threshold, identity, sigmoid, or some other complex mathematical function. A neural network is a collection of interconnected neurons. Neural networks may be distinguished according to the type of interconnection between the input and output of network. Basically, there are two types of networks: feedforward and recurrent. In a feedforward network, there are no loops, and the signals propagate in only one direction from an input stage through intermediate neurons to an output stage. With the use of a continuous nonlinear activation function, this network is a static nonlinear map that can be used efficiently as a parallel computational model of a continuous mapping. If the network possesses some cycle or loop, i.e., signals may propagate from the output of any neuron to the input of any neuron, then it is a feedback or recurrent neural network. In a recurrent network the system has an internal state, and thereby the output will also depend on the internal state of the system. Hence, the study of recurrent neural networks is connected to analysis of dynamic systems.

Neural networks are able to store experiential knowledge through learning from examples. They can also be classified in terms of the amount of guidance that the learning process receives from an outside agent. An *unsupervised learning* network learns to classify input into sets without being told anything. A *supervised learning* network adjusts weights on the basis of the difference between the values of the output units and the desired values given by the teacher using an input pattern. Neural networks can be further characterized by their network topology, i.e., by the number of interconnections, the node characteristics that are classified by the type of nonlinear elements used (activation rule), and the kind of learning rules implemented.

The application of neural networks in technical problems consists of two phases:

1. "Phase of learning/adaptation/design" is the special phase of learning, modifying, and designing the internal structure of the network when it acquires knowledge about the real system as a result of interaction with system and real environment using a trial-error method, as well as the result of the appropriate meta rules inherent to global network context.
2. "Pattern associator phase or associative memory mode" is a special phase when, using the stored associations, the network converges toward the stable attractor or a desired solution.

24.2.2 Connectionist Models with Applications in Robotics

In contemporary neural network research, more than 20 neural network models have been developed. Because our attention is focused on the application of neural networks in robotics, we briefly introduce some important types of network models that are commonly used in robotics applications. There are multilayer perceptrons (MP), radial basis function networks (RBF), recurrent version of multilayer perceptron (RMP), Hopfield networks (HN), CMAC networks, and ART networks.

For the study and application of feedforward networks it is convenient to use in addition to single-layer neural networks, more structured ones known as multilayer networks or *multilayer perceptrons*. These networks with an appropriate number of hidden levels have received considerable attention because of better representation capabilities and the possibility of learning highly nonlinear mappings. The typical network topology that represents a multilayer perceptron (Figure 24.1) consists of an input layer, a sufficient number of hidden layers, and the output layer. The following recursive relations define the network with $k + 1$ layers:

$$y_0 = u \tag{24.3}$$

$$y_l = f_l(W_l \bar{y}_{l-1}), \quad l = 1, \ldots, k \tag{24.4}$$

where y_l is vector of neuron inputs in the l-layer ($y_k = y$ – output of $k + 1$ is the network layer, u is network input, f_l is the activation function for the l layer, W_l is the weighting matrix between layers $l-1$ i $l, \bar{y}_j = [y_j, 1]$ is the adjoint vector y_j. In the previous equation, bias vector is absorbed by the weighting matrix.

© 2002 by CRC Press LLC

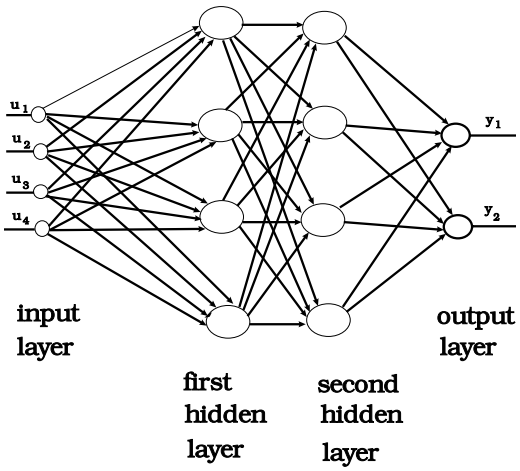


FIGURE 24.1 Multilayer perceptron.

Each layer has an appropriate number of neural units, where each neural unit has some specific activation function (usually a logistic sigmoid function). The weights of the networks are incrementally adjusted according to appropriate learning rules, depending on the task, to improve the system performance. They can be assigned new values in two ways: either via some prescribed offline algorithm that remains fixed during the operation, or adjusted by a learning process. Several powerful learning algorithms exist for feedforward networks, but the most commonly used algorithm is the *backpropagation algorithm*.[9] The backpropagation algorithm as a typical supervised learning procedure that adjusts weights in the local direction of greatest error reduction (steepest descent gradient algorithm) using the square criterion between the real network output and desired network output.

An RBF network approximates an input–output mapping by employing a linear combination of radially symmetric functions. The $k - th$ output y_k is given by:

$$y_k(u) = \sum_{i=1}^{m} w_{ki}\phi_i(u) \tag{24.5}$$

where:

$$\phi(u) = \phi\left(\left\|u - c_i\right\|\right) = \phi(r_i) = \exp\left(\frac{-\eta^2}{2\sigma_i^2}\right),\ r_i \geq 0, \sigma_i \geq 0 \tag{24.6}$$

The RBF network always has one hidden layer of computational modes with a nonmonotonic activation function $\phi(.)$. Theoretical studies have shown that the choice of activation function $\phi(.)$ is not very crucial to the effectiveness of the network. In most cases, the Gaussian RBF given by (24.6) is used, where c_i and σ_i are selected centers and widths, respectively.

One of the earliest sensory connectionist methods capable of serving as an alternative to the well-known backpropagation algorithm is the CMAC (cerebellar model arithmetic computer)[20] (Figure 24.2). The CMAC topology consists of a three-layer network, one layer being the sensory or command input, the second the association layer, and the third the output layer. The association layer is conceptual memory with high dimensionality. On the other hand, the output layer is the actual memory with low dimensionality. The connections between these two layers are chosen in a random way. The adjustable weights exist only between the association layer and the output layer. Using supervised learning, the training set of patterns is presented and, accordingly, the weights are adjusted. CMAC uses the Widrow-Hoff LMS algorithm[6] as a learning rule.

© 2002 by CRC Press LLC

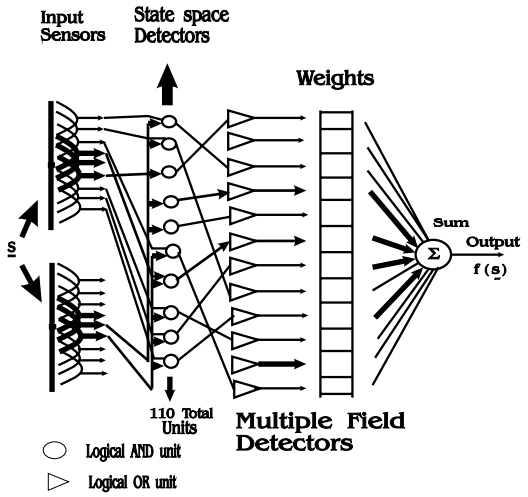


FIGURE 24.2 Structure of CMAC network.

CMAC is an associative neural network using the feature that only a small part of the network influences any instantaneous output. The associative property built into CMAC enables local generalization; similar inputs produce similar outputs while distant inputs produce nearly independent outputs. As a result, we have fast convergence properties. It is very important that practical hardware realization using logical cell arrays exists today.

If the network possesses some cycle or loop, then it is a feedback or recurrent neural network. In a recurrent network the system has an internal state, and the output will also depend on the internal state of the system. These networks are essentially nonlinear dynamic systems with stability problems. There are many different versions of inner and outer recurrent neural networks (recurrent versions of multilayer perceptrons) for which efficient learning and stabilization algorithms must be synthesized. One of the most commonly used recurrent networks is the Hopfield[23] type neural network that is very suitable for optimization problems. Hopfield introduced a network that employed a continuous nonlinear function to describe the output behavior of the neurons. The neurons are an approximation to biological neurons in which a simplified set of important computational properties is retained. This neural network model, which consists of nonlinear graded-response model neurons organized into networks with effectively symmetric synaptic connections, can be easily implemented with electronic devices. The dynamics of this network is defined by the following equation:

$$\dot{y}_i = -\alpha y_i + \beta f_i\left(\sum_j w_{ij} y_j\right) + I_i \qquad i = 1,..,n, \tag{24.7}$$

where α, β are positive constants and I_i is the array of desired network inputs.

A Hopfield network can be characterized by its energy function:

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1} w_{ij} y_i y_j - \sum_{i=1}^{n} I_i y_i \tag{24.8}$$

The network will seek to minimize the energy function as it evolves into an equilibrium state. Therefore, one may design a neural network for function minimization by associating variables in an optimization problem with variables in the energy function.

© 2002 by CRC Press LLC

ART networks are neural networks based on the Adaptive Resonance Theory of Carpenter and Grossberg.[17] An ART network selects its first input as the exemplar for the first cluster. The next input is compared to the first cluster exemplar. It is clustered with the first if the distance to the first cluster is less than a threshold. Otherwise it is the exemplar for a new cluster. This procedure is repeated for all the following inputs. If an input is clustered with the *j*th cluster, the weights of the network are updated according to the following formulae

$$w_{ij}(t+1) = \frac{v_{ij}(t)u_i}{0.5 + \sigma_{i=1}^{n} v_{ij}(t)u_i} \tag{24.9}$$

$$\nu_{ij}(t+1) = u_i v_{ij}(t) \tag{24.10}$$

where $i = 1, 2, \ldots, M$. ART networks belong to the class of unsupervised learning networks. They are stable because new input patterns do not erase previously learned information. They are also adaptive because new information can be incorporated until full capacity of the architecture is utilized.

Proposed neural networks can be classified according to their ability to generalize. CMAC is a local generalizing neural network, while MLPs and recurrent MLPs are suitable for global generalization. RBF networks are placed between them. The choice for either one of the networks depends on the requirement for local generalization. When a strong local generalization is needed, a CMAC is most suitable. For global generalization, MLPs and recurrent MLPs provide a good alternative, combined with an improved weight adjustment algorithm.

24.2.3 Learning Principles and Rules

Adaptation (or machine learning) deals with finding weights (and sometimes a network topology) that will produce the desired behavior. Usually, the learning algorithm works from training examples, where each example incorporates correct input–output pairs (*supervised learning*). This learning form is based on the acquisition of mapping by the presentation of training exemplars (input–output data). Different than supervised learning, *reinforcement learning* considers the improvement of system performances by evaluating some realized control action that is included in the learning rules. Unsupervised learning in connectionist learning is when processing units respond only to interesting patterns on their inputs that are based on internal learning function.

The topology of the network during the training process can be fixed or variable based on evolution and regeneration principles.

The different iterative adaptation algorithms proposed so far are essentially designed in accordance with the *minimal disturbance principle:* Adapt to reduce output error for the current training pattern, with minimal disturbance to responses already learned. Two principal classes of algorithms can be distinguished:

Error-correction rules, alter the weights of a network to correct the error in the output response to the present input pattern.

Gradient-based rules, alter the weights of a network during each pattern presentation by a gradient descent with the objective of reducing mean-square error, averaged over training patterns.

The error-correction rules for networks often tend to be ad hoc. They are most often used when training objectives are not easily quantified, or when a problem does not lend itself to tractable analysis (for instance, networks that contain discontinuous functions, e.g., signum networks).

Gradient adaptation techniques are intended for minimization of the mean-square error associated with an entire network of adaptive elements:

© 2002 by CRC Press LLC

$$e^2 = \sum_{t=1}^{T} \sum_{i=1}^{N_y} [e_i(t)]^2 \tag{24.11}$$

where $e_i^2(t)$ is the square error for particulary patterns.

The most practical and efficient algorithms typically work with one pattern presentation at a time. This approach is referred to as *pattern learning*, as opposite to *batch learning*, in which weights are adapted after presentation of all the training patterns (true *real-time learning* is similar to pattern learning, but it is performed with only one pass through the data). Similar, to the single-element case, in place of the true MSE function, the instantaneous sum squared error $e^2(t)$ is considered, which is the sum of the square errors at each of the N_y outputs of the network:

$$e^2(t) = \sum_{i=1}^{N_y} [e_i(t)]^2 \tag{24.12}$$

The corresponding instantaneous gradient is

$$E = \hat{\nabla}(t) = \frac{\partial e^2(t)}{\partial w(t)} \tag{24.13}$$

where $w(t)$ denotes a vector of all weights in the network. The steepest descent with the instantaneous gradient is a process presented by

$$w(t + 1) = w(t) + \Delta w(t)$$

$$\Delta w(t) = \mu(-\hat{\nabla}(t)) \tag{24.14}$$

The most popular method for estimating the gradient $\hat{\nabla}(t)$ is the backpropagation algorithm.

The backpropagation algorithm or generalized delta rule is the basic training algorithm for multilayer perceptrons. The basic analysis of an algorithm application will be shown using a three-layer perceptron (one hidden layer with a sigmoid function in the hidden and output layers). The main relations in the training process for one input–output pair $p = p(t)$ are given by the following relations:

$$s_2^p = W_{12}^{pT} u_1^p \quad s_2^p \varepsilon R^{L_1} \tag{24.15}$$

$$o_{2a}^p = 1/(1 + \exp(-s_{2a}^p)) \quad a = 1, \ldots, L_1 \quad o_{20}^p = 1 \tag{24.16}$$

$$s_3^p = W_{23}^{pT} o_2^p \quad s_3^p \varepsilon R^{N_y} \tag{24.17}$$

$$o_{3b}^p = 1/(1 + \exp(-s_{3b}^p)) \qquad b = 1, \ldots, N_y \tag{24.18}$$

$$y_c^p = o_{3c}^p \qquad c = 1, \ldots, N_y \tag{24.19}$$

where s_2^p, s_3^p are input vectors of the hidden and output layers of the network; o_2^p, o_3^p are output vectors of the hidden and output layers; $W_{12}^p = [w_{12ij\,Nu+1\times L1}(t)]$, $W_{23}^p = [w_{23ij_{L1+1\times N_y}}(t)]$ are weighting factors; w_{tuij} is the weighting factor that connects neuron j in layer t with neuron i in output layer

© 2002 by CRC Press LLC

u; u_1^p is the input vector ($u_{10}^p = 1$; N_u-number of inputs; y^p is the output vector (N_y_ number of outputs; L_1 = number of neurons in a hidden layer).

The square error criterion can be defined as:

$$E = \sum_{p \in P} E^p = 0.5 \sum_{p \in P} \left| \hat{y}^p - y^p \right|^2 \tag{24.20}$$

where $\hat{y}^p$ is the desired value of the network output; y^p je output value of the networks; E^p is the value of the square criterion for one pair of input–output data; P is the set of input–output pairs.

The corresponding gradient component for the output layer is

$$\frac{\partial E}{\partial w_{23ij}} = \sum_{p \in P} \frac{\partial E^p}{\partial w_{23ij}} = \sum_{p \in P} \frac{\partial E_p}{s_{3i}^p} \frac{\partial s_{3i}^p}{\partial w_{23ij}} = -\sum_{p \in P} \delta_{3i}^p o_{2j}^p \tag{24.21}$$

$$\delta_{3i}^p = (\hat{y}_i^p - y_i^p) df_{3i}^p / ds_{3i}^p = (\hat{y}_i^p - y_i^p) f'_{3i}(s_{3i}^p) \tag{24.22}$$

where f_{gi} is the activation function for neuron i in layer g.

For the hidden layer, the gradient component is defined by:

$$\begin{aligned} \frac{\partial E}{\partial w_{12ij}} &= \sum_{p \in P} \frac{\partial E^p}{\partial w_{12ij}} = \sum_{p \in P} \frac{\partial E_p}{s_{2i}^p} \frac{\partial s_{2i}^p}{\partial w_{12ij}} \\ &= \sum_{p \in P} \sum_r \frac{\partial E^p}{\partial s_{3r}^p} \frac{\partial s_{3r}^p}{\partial o_{2i}^p} \frac{\partial o_{2j}^p}{\partial s_{2i}^p} \frac{\partial s_{2i}^p}{\partial w_{12ij}} \\ &= -\sum_{p \in P} \sum_r \delta_{3r}^p w_{23ri} f'_{2i}(s_{2i}^p) u_{1j}^p \\ &= -\sum_{p \in P} \delta_{2i}^p u_{1j}^p \end{aligned} \tag{24.23}$$

$$\delta_{2i}^p = \sum_r \delta_{3\tau}^p w_{23ri} f'_{2i}(s_{2i}^p) \tag{24.24}$$

Based on previous equations, starting from the output layer and going back, the error backpropagation algorithm is synthesized. The final version of the algorithm modified by weighting factors is defined by the following relations:

$$\delta_{3i}(t) = (\hat{y}_i(t) - y_i(t)) f'_{3i}(s_{3i}(t)) \tag{24.25}$$

$$\Delta w_{23ij}(t) = -\eta \frac{\partial E}{\partial w_{23ij}} = \eta \delta_{3i}(t) o_{2j}(t) \tag{24.26}$$

$$\delta_{2i}(t) = \sum_r \delta_{3r}(t) w_{23ri}(t) f'_{2i}(s_{2i}(t)) \tag{24.27}$$

© 2002 by CRC Press LLC

$$\Delta w_{12ij}(t) = -\eta \frac{\partial E}{\partial w_{12ij}} = \eta \delta_{2i}(t) u_{1j}(t) \tag{24.28}$$

$$w_{23ij}(t+1) = w_{23ij}(t) + \Delta w_{23ij}(t) \tag{24.29}$$

$$w_{12ij}(t+1) = w_{12ij}(t) + \Delta w_{12ij}(t) \tag{24.30}$$

where η is the learning rate.

Also, numerous variants are used to speed up the learning process in the backpropagation algorithm. The one important extension is the *momentum technique* which involves a term proportional to the weight change from the previous iteration:

$$w(t+1) = w(t) + \Delta w(t)$$

$$\Delta w(t) = (1-\eta) \cdot \mu(-\hat{\nabla}(t)) + \eta \cdot \Delta w(t-1) \tag{24.31}$$

The momentum technique serves as a low-pass filter for gradient noise and is useful in situations when a clean gradient estimate is required, for example, when a relatively flat local region in the mean square error surface is encountered. All gradient-based methods are subject to convergence on local optima. The most common remedy for this is the sporadic addition of noise to the weights or gradients, as in simulated annealing methods. Another technique is to retrain the network several times using different random initial weights until a satisfactory solution is found. Backpropagation adapts the weights to seek the extremum of the objective function whose domain of attraction contains the initial weights. Therefore, both choice of the initial weights and the form of the objective function are critical to the network performance. The initial weights are normally set to small random values. Experimental evidence suggests choosing the initial weights in each hidden layer in a quasi-random manner, which ensures that at each position in a layer's input space the outputs of all but a few of its elements will be saturated, while ensuring that each element in the layer is unsaturated in some region of its input space.

There are more different learning rules for speeding up the convergence process of the backpropagation algorithm. One interesting method is using recursive least square algorithms and the extended Kalman approach instead of gradient techniques.[12]

The training procedure for the RBF networks involves a few important steps:

Step 1: Group the training patterns in M subsets using some clustering algorithm (k-means clustering algorithm) and select their centers c_i.
Step 2: Compute the widths, σ_i, ($i = 1, \ldots, m$), using some heuristic method (p-nearest neighbor algorithm).
Step 3: Compute the RBF activation functions $\phi_i(u)$, for the training inputs.
Step 4: Compute the weight vectors by least squares.

24.3 Neural Network Issues in Robotics

Possible applications of neural networks in robotics include various purposes suh as vision systems, appendage controllers for manufacturing, tactile sensing, tactile feedback gripper control, motion control systems, situation analysis, navigation of mobile robots, solution to the inverse kinematic problem, sensory-motor coordination, generation of limb trajectories, learning visuomotor coordination of a robot arm in 3D, etc.[5,11,16,38,39,43] All these robotic tasks can be categorized according to the type of hierarchical control level of the robotic system, i.e., neural networks can be applied at a strategic control level (task planning), at a tactic control level (path planning), and at an executive

© 2002 by CRC Press LLC

control level (path control). All these control problems at different hierarchical levels can be formulated in terms of optimization or pattern association problems. For example, autonomous robot path planning and stereovision for task planning can be formulated as optimization problems, while on the other hand, sensor/motor control, voluntary movement control, and cerebellar model articulation control can be formulated as pattern association tasks. For pattern association tasks, neural networks in robotics can have the role of function approximation (modeling of input/output kinematic and dynamic relations) or the role of pattern classification necessary for control purposes.

24.3.1 Kinematic Robot Learning by Neural Networks

It is well known in robotics that control is applied at the level of the robot joints, while the desired trajectory is specified through the movement of the end-effector. Hence, a control algorithm requires the solution of the inverse kinematic problem for a complex nonlinear system (connection between internal and external coordinates) in real time. However, in general, the path in Cartesian space is often very complex and the end-effector location of the arm cannot be efficiently determined before the movement is actually made. Also, the solution of the inverse kinematic problem is not unique, because in the case of redundant robots there may be an infinite number of solutions. The conventional methods of solution in this case consist of closed-form and iterative methods. These are either limited only to a class of simple non-redundant robots or are time-consuming and the solution may diverge because of a bad initial guess. We refer to this method as the *position-based inverse kinematic control*. The *velocity-based inverse kinematic control* directly controls the joint velocity (determined by the external and internal velocities of the Jacobian matrix). Velocity-based inverse kinematic control is also called inverse Jacobian control.

The goal of kinematic learning methods is to find or approximate two previously defined mappings: one between the external coordinate target specified by the user and internal values of robot coordinates (position-based inverse kinematic control) and a second mapping connected to the inverse Jacobian of the robotic system (velocity-based inverse kinematic control).

In the area of position-based inverse kinematic control problems various methods have been proposed to solve them. The basic idea common to all these algorithms is the use of the same topology of the neural network (multilayer perceptron) and the same learning rule: the backpropagation algorithm. Although the backpropagation algorithms work for robots with a small number of degrees of freedom, they may not perform in the same way for robots with six degrees of freedom. In fact, the problem is that these methods are naive, i.e., in the design of neural network topology some knowledge about kinematic robot model has not been incorporated. One solution is to use a hybrid approach, i.e., a combination of the neural network approach with the classic iterative procedure. The iterative method gives the final solution in joint coordinates within the specified tolerance.

In the velocity-based kinematic approaches, the neural network has to map the external velocity into joint velocity. A very interesting approach has been proposed using the context-sensitive networks. It is an alternative approach to the reduction of complexity, as it proposes partition of the network input variables into two sets. One set (context input) acts as the input to a context network. The output of the context network is used to set up the weights of the function network. The function network maps the second set of input variables (function input) to the output. The original function to be learned is decomposed into a parameterized family of functions, each of which is simpler than the original one and is thus easier to learn.

Generally, the main problem in all kinematic approaches is accurately tracking a predetermined robot trajectory. As is known, in most kinematic connectionist approaches, the kinematic input/output mapping is learned offline and then control is attempted. However, it is necessary to examine the proposed solutions by learning control of manipulation robots in real-time, because the robots are complex dynamic systems.

© 2002 by CRC Press LLC

24.3.2 Dynamic Robot Learning at the Executive Control Level

As a solution in the context of robot dynamic learning, neural network approaches provide the implementation tools for complex input/output relations of robot dynamics without analytic modeling. Perhaps the most powerful property of neural networks in robotics is their ability to model the whole controlled system itself. In this way the connectionist controller can compensate for a wide range of robot uncertainties. It is important to note that the application of the connectionist solution for robot dynamic learning is not limited only to noncontact tasks. It is also applicable to essential contact tasks, where inverse dynamic mapping is more complex, because dependence on contact forces is included.

The application of the connectionist approach in robot control can be divided according to the type of learning into two main classes: neurocontrol by supervised and neurocontrol by unsupervised learning.

For the first class of neurocontrol a teacher is assumed to be available, capable of teaching the required control. This is a good approach in the case of a human-trained controller, because it can be used to automate a previously human-controlled system. However, in the case of automated linear and nonlinear teachers, the teacher's design requires *a priori* knowledge of the dynamics of the robot under control. The structure of the supervised neurocontrol involves three main components, namely, a teacher, the trainable controller, and the robot under control.[1] The teacher can be either a human controller or another automated controller (algorithm, knowledge-based process, etc.). The trainable controller is a neural network appropriate for supervised learning prior to training. Robot states are measured by specialized sensors and are sent to both the teacher and the trainable controller. During control of the robot by the teacher, the control signals and the state variables of the robot are sampled and stored for neural controller training. At the end of successful training the neural network has learned the right control action and replaces the teacher in controlling the robot.

In unsupervised neural learning control, no external teacher is available and the dynamics of the robot under control is unknown and/or involves severe uncertainties. There are different principal architectures for unsupervised robot learning.

In the *specialized learning architecture* (Figure 24.3), the neural network is tuned by the error between the desired response and actual response of the system. Another solution, *generalized learning architecture* (Figure 24.4), is proposed in which the network is first trained offline based on control error, until good convergence properties are achieved, and then put in a real-time feedforward controller where the network continues its adaptation to system changes according to specialized learning procedures.

The most appropriate learning architectures for robot control are *feedback-error learning architecture* and *adaptive learning architecture*. The feedback-error learning architecture (Figure 24.5) is an exclusively online achitecture for robot control that enables simultaneous processing of learning and control. The primary interest is learning an inverse dynamic model of robot mechanism for the tasks with holonomic constraints, where exact robot dynamics is generally unknown. The neural network as part of feedforward control generates necessary driving torques in robot joints as a nonlinear mapping of robot desired internal coordinates, velocities, and accelerations:

$$P_i = g(w_{jk}^{ab}, q_d, \dot{q}_d, \ddot{q}_d) \quad i = 1, \ldots, n. \tag{24.32}$$

where $P_i \varepsilon R^n$ is a joint-driving torque generated by a neural network; w_{jk}^{ab} are adaptive weighting factors between neuron j in $a - th$ layer and neuron k in $b - th$ layer; g is nonlinear mapping.

According to the integral model of robotic systems, the decentralized control algorithm with learning has the form

© 2002 by CRC Press LLC

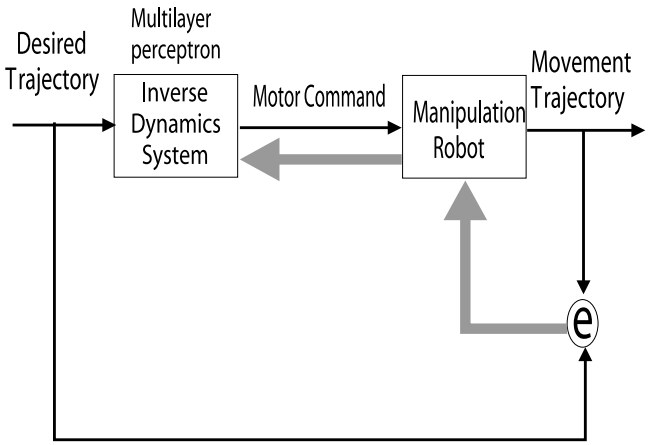


FIGURE 24.3 Specialized learning architecture.

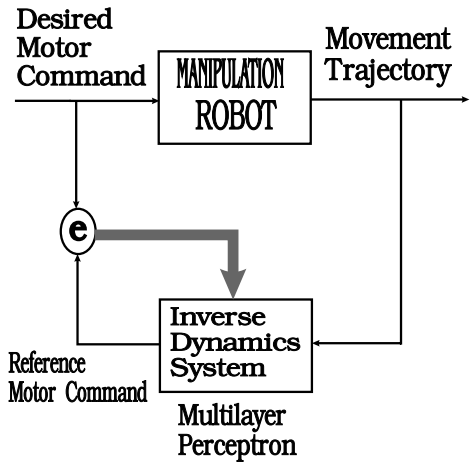


FIGURE 24.4 Generalized learning architecture.

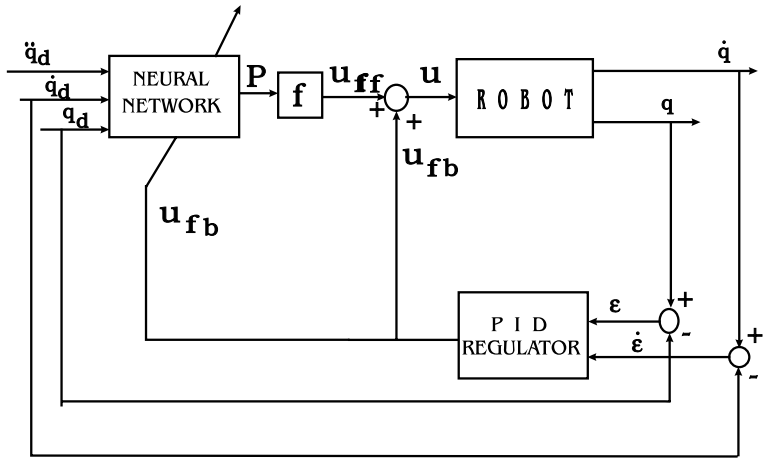


FIGURE 24.5 Feedback-error learning architecture.

© 2002 by CRC Press LLC

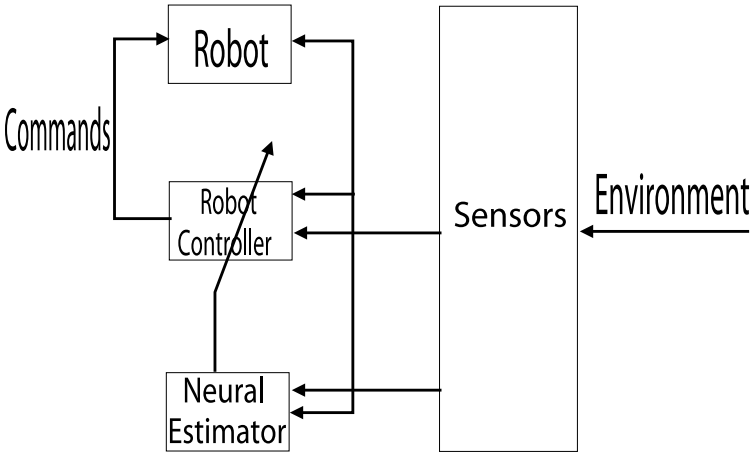


FIGURE 24.6 Sensor-based learning architecture.

$$u_i = u_{ii}^{ff} + u_{ii}^{fb} \qquad i = 1, \ldots, n. \tag{24.33}$$

$$u_i = f_i(q_d, \dot{q}_d, \ddot{q}_d, P) - KP_{ii}\varepsilon_i - KD_{ii}\dot{\varepsilon}_i - KI_{ii}\int \varepsilon_i dt \quad i = 1, \ldots, n. \tag{24.34}$$

where f_i is the nonlinear mapping which describes the nature of the robot actuator model; $KP, KF, KI \varepsilon R^{n \times n}$ are position, velocity, and integral local feedback gains, respectively; $\varepsilon \varepsilon R^n$ is the feedback error. Training and learning the proposed connectionist structure can be accomplished using the well-known backpropagation algorithm.[9] In the process of training we can use the feedback control signal:

$$e_i^{bp} = u_i^{fb} \quad i = 1, \ldots, n \tag{24.35}$$

where $e_i^{bp} \varepsilon R^n$ is the output error for the backpropagation algorithm.

A more recent and sophisticated learning architecture (adaptive learning architecture) involves the neural estimator that identifies some robot parameters using available information from robot sensors (Figure 24.6). Based on information from the neural estimator, the robot controller modifies its parameters and then generates a control signal for robot actuators. The robot sensors observe the status of the system and make available information and parameters to the estimator and robot controller. Based on this input, the neural estimator changes its state, moving in the state space of its variables. The state variables of the neural estimator correspond exactly to the parameters of robot controller. Hence, the stable-state topology of this space can be designed so that the local minima correspond to an optimal law.

The special reactive control strategy applied to robotic dynamic control[51] can be characterized as reinforcement learning architecture. In contrast to the supervised learning paradigm, the role of the teacher in reinforcement learning is more evaluative than instructional. The teacher provides the learning system with an evaluation of the system performance of the robot task according to a certain criterion. The aim of this learning system is to improve its performance by generating appropriate outputs. In Gullapalli[51] a stochastic reinforcement learning approach with application in robotics for learning functions with continuous outputs is presented. The learning system computes real-valued output as some function of a random activation generated using normal distribution. The parameters of normal distribution are the mean and the standard deviation that

© 2002 by CRC Press LLC

depend on current input patterns. The environment evaluates the unit output in the context of input patterns and sends a reinforcement signal to the learning system. The aim of learning is to adjust the mean and the standard deviation to increase the probability of producing the optimal real value for each input pattern.

A special group of dynamic connectionist approaches is the methods that use the "black-box" approach in the design of neural network algorithms for robot dynamic control. The "black box" approach does not use any *a priori* experience or knowledge about the inverse dynamic robot model. In this case it is a multilayer neural network with a sufficient number of hidden layers. All we need to do is feed the multilayer neural network the necessary information (desired positions, velocities, and accelerations at the network input and desired driving torque at the network output) and let it learn by test trajectory. In Ozaki et al.[48] a nonlinear neural compensator that incorporates the idea of computed torque method is presented. Although the pure neural network approach without knowledge about robot dynamics may be promising, it is important to note that this approach will not be very practical because of the high dimensionality of input–output spaces. Bassi and Bekey[10] use the principle of functional decomposition to simplify robot dynamics learning. This method includes *a priori* knowledge about robot dynamics which, instead of being specific knowledge corresponding to a certain type of robot models, incorporates common invormation about robot dynamics. In this way, the unknown input–output mapping is decomposed into simpler functions that are easier to learn because of smaller domains. In Katić and Vukobratović,[12] similar ideas in the development of the fast learning algorithm were used with decomposition at the level of internal robot coordinates, velocities, and accelerations.

The connectionist approach is very efficient in the case of robots with flexible links or for a flexible materials handling system by a robotic manipulators where the parameters are not exactly known and the learning capability is important to deal with such problems. Because of the complex nonlinear dynamical model, the recurrent neural network is very suitable for compensating flexible effects.

With recent extensive research in the area of robot position/force control, a few connectionist learning algorithms for constrained manipulation have been proposed. We can distinguish two essential different approaches: one, whose aim is the transfer of human manipulation skills to robot controllers, and the other, in which the manipulation robot is examined as an independent dynamic system in which learning is achieved through repetition of the work task.

The principle of transferring human manipulation skill (Figure 24.7) has been developed in the papers of Asada and co-workers.[18] The approach is based on the acquisition of manipulation skills and strategies from human experts and subsequent transfer of these skills to robot controllers. It is essentially a playback approach, where the robot tries to accomplish the working task in the same way as an experienced worker. Various methods and techniques have been evaluated for acquisition and transfer of human skills to robot controllers.

This approach is very interesting and important, although there are some critical issues related to the explicit mathematical description of human manipulation skill because of the presence of subconscious knowledge and inconsistent, contradictory, and insufficient data. These data may cause system instability and wrong behavior by the robotic system. As is known, dynamics of the human arm and a robot arm are essentially different, and therefore it is not possible to apply human skill to robot controllers in the same way. The sensor system for data acquisition of human skill can be insufficient for extracting a complete set of information necessary for transfer to robot controllers. Also, this method is inherently an offline learning method, whereas for robot contact tasks online learning is a very important process because of the high level of robot interaction with the environment and unpredictable situations that were not captured in the skill acquisition process.

The second group of learning methods, based on autonomous online learning procedures with working task repetition, have also been evaluated through several algorithms. The primary aim is to build internal robot models with compensation of the system uncertainties or direct adjustment of control signals or parameters (reinforcement learning). Using a combination of different intelligent paradigms (fuzzy + neuro) Kiguchi and Fukuda[25] proposed a special algorithm for approach,

© 2002 by CRC Press LLC

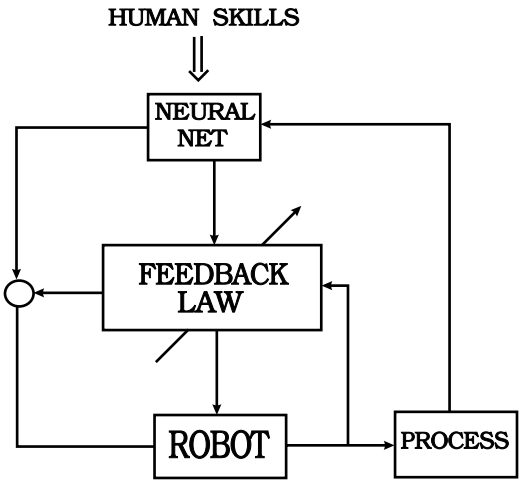


FIGURE 24.7 Transfer of human skills to robot controllers by the neural network approach.

contact, and force control of robot manipulators in an unknown environment. In this case, the robot manipulator controller, which approaches, contacts, and applies force to the environment, is designed using fuzzy logic to realize human-like control and then modeled as a neural network to adjust membership functions and rules to achieve the desired contact force control.

As another exposed problem in control robotic contact tasks, the connectionist approach is used for dynamic environment identification. A new learning control concept based on neural network classification of unknown dynamic environment models and neural network learning of robot dynamic model has been proposed.[13] The method classifies characteristics of environments by using multilayer perceptrons based on the first neural network, and then determines the control parameters for compliance control using the estimated characteristics. Simultaneously, using the second neural network, compensation of robot dynamic model uncertainties is accomplished. The classification capability of the neural classifier is realized by an efficient offline training process. It is important that the pattern classification process can work in an online manner as a part of selected compliance control algorithm.

The first objective is the application of connectionist structures to fast online learning of robotic system uncertainties as a part of the stabilizing control algorithm mentioned previously. The role of the connectionist structure has a broader sense, because its aim is to compensate possible uncertainties and differences between real robot dynamics and assumed dynamics defined by the user in the process of control synthesis. Hence, to achieve good tracking performance in the presence of model uncertainties, a fixed non-recurrent multilayer perceptron is integrated into the non-learning control law with the desired quality of transient processing for interaction force.

In this case, compensation by neural network is connected to the uncertainties of robot dynamic model. But, the proposed learning control algorithm does not work in a satisfactory way if there is no sufficiently accurate information about the type and parameters of the robot environment model. Hence, to enhance connectionist learning of the general robot-environment model, a new method is proposed whose main idea is using a neural network approach through an offline learning process and online sufficiently exact classification of robot dynamic environment. The neural network classifier based on a four-layer perceptron is chosen due to good generalization properties. Its objective is to classify the model profile and parameters of environment in an online manner. In the acquisition process, based on real-time realization of proposed contact control algorithms and using previously chosen sets of different working environments and model profiles of working environments, some force data from force sensors are measured, calculated, and stored as special input patterns for training the neural network. On the other side, the acquisition process must be

© 2002 by CRC Press LLC

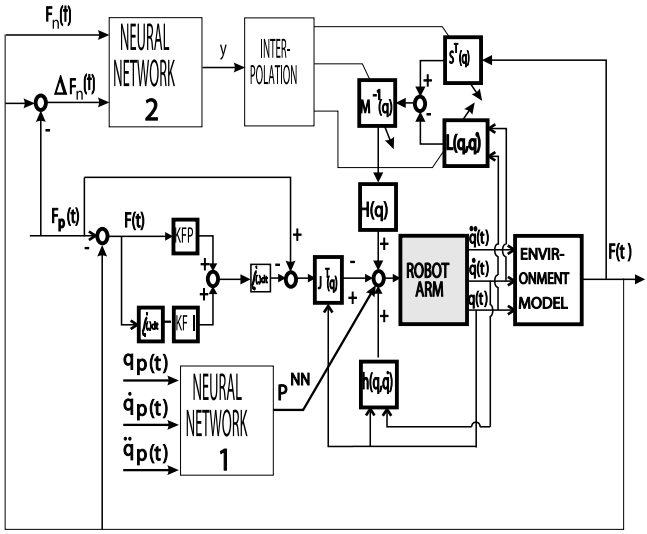


FIGURE 24.8 Scheme of the connectionist control law stabilizing interaction force.

accomplished using various robot environments, starting with the environment with a low level of system characteristics (for example, with a low level of environment stiffness) and ending with an environment with a high level of system characteristics (with high level of environment stiffness). As another important characteristic in the acquisition process, different model profiles of the environment are used based on additional damping and stiffness members that are added to the basic general impedance model.

After that, during the extensive offline training process, the neural network receives a set of input–output patterns, where the input variables form a previously collected set of force data. As a desired output, the neural network has a value between 0 and a value defined by the environment profile model (the whole range between 0 and 1) that exactly defines the type of training robot environment and environment model. The aim of connectionist training is for the real output of the neural network for given inputs to be exact or very close to the desired output value determined for an appropriate training robot environment model.

After the offline training process with different working environments and different environment model profiles, the neural classifier is included in the online version of the control algorithm to produce some value at the network's output between 0 and 1. In the case of an unknown environment, information from the neural classifier output can be utilized efficiently for calculating the necessary environment parameters by linear interpolation procedures. Figure 24.8 shows the overall structure of the proposed algorithm.

24.3.3 Sensor-Based Robot Learning

A completely different approach of connectionist learning uses sensory information for robot neural control. Sensor-based control is a very efficient method in overcoming problems with robot model and environment uncertainties, because sensor capabilities help in the adaptation proces without explicit control intervention. It is adaptive sensor-motor coordination that uses various mappings given by the robot sensor system. Particular attention has been paid to the problem of visuo-motor coordination, in particular for eye–head and arm–eye systems. In general, in visuo-motor coordination by neural networks, visual images of the mechanical parts of the systems can be directly related to posture signals. However, tactile-motor coordination differs significantly from visuo-motor because the intrinsic dependency on the contacted surface. The direct association of tactile sensations with positioning of the robot end-effector is not feasible in many cases, hence it is very

© 2002 by CRC Press LLC

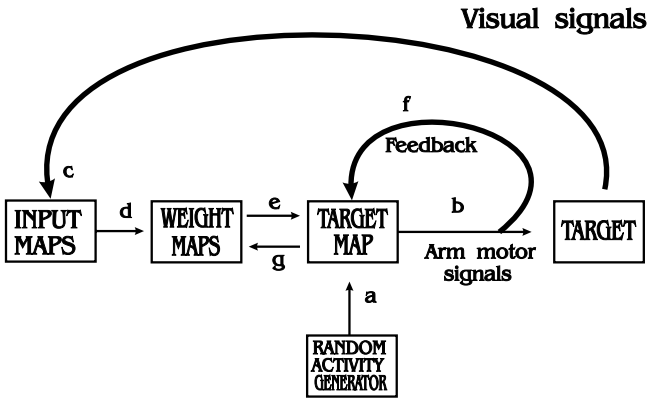


FIGURE 24.9 Sensory-motor circular reaction.

important to understand how a given contact condition will be modified by motor actions. The task of the neural network in these cases is to estimate the direction of a feature-enhancing motor action on the basis of modifications in the sensed tactile perception.

After many years of being thought impractical in robot control, it was demonstrated that CMAC could be very useful in learning state-space dependent control responses.[56] A typical demonstration of CMAC application in robot control involves controlling an industrial robot using a video camera. The robot's task is to grasp an arbitrary object lying on a conveyor belt with a fixed orientation or to avoid various obstacles in the workspace. In the learning phase, visual input signals about the objects are processed and combined into a target map through modifiable weights that generate the control signals for the robot's motors. The errors between the actual motor signals and the motor signals computed from the camera input are used to incrementally change the weights. Kuperstain[33] has presented a similar approach using the principle of sensory-motor circular reaction (Figure 24.9). This method relies on consistency between sensory and motor signals to achieve unsupervised learning. This learning scheme requires only availability of the manipulator, but no formal knowledge of robotic kinematics. Opposite to previously mentioned approaches for visuo-motor coordination, Rucci and Dario[34] experimentally verified autonomous learning of tactile-motor coordination by a Gaussian network for a simple robotic system composed of a single finger mounted on a robotic arm.

24.4 Fuzzy Logic Approach

24.4.1 Introduction

The basic idea of fuzzy control was conceived by L. Zadeh in his papers from 1968, 1972, and 1973.[59,61,62] The heart of his idea is describing control strategy in linguistic terms. For instance, one possible control strategy of a single-input, single-output system can be described by a set of control rules:

If (error is positive and error change is positive), then
  control change = negative
Else if (error is positive and error change is negative), then
  control change = zero
Else if (error is negative and error change is positive), then
  control change = zero
Else if (error is negative and error change is negative), then
  control change = positive

© 2002 by CRC Press LLC

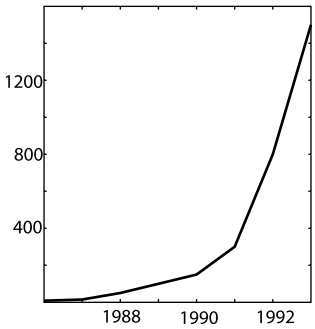


FIGURE 24.10 Estimated number of commercial applications of fuzzy systems.

Further refining of the strategy might take into account cases when, e.g., the error and error change are small or big. Such a procedure could make it possible to describe the control strategy used, e.g., by trained operators when controlling a system manually.

Statements in natural language are intrinsically imprecise due to the imprecise manner of human reasoning. Development of techniques for modeling imprecise statements is one of the main issues in implementation of automatic control systems based on using linguistic control rules. With fuzzy controllers, modeling of linguistic control rules (as well as derivation of control action on the basis of given set of rules and known state of the controlled system) is based on the theory of fuzzy sets introduced by Zadeh in 1965.[58]

In 1974, Mamdani described the first application of fuzzy set theory to automatic control.[30] However, almost 10 years passed before broader interest was reestablished for fuzzy logic and its applications in automatic control. The number of reported fuzzy applications has been increasing exponentially (Figure 24.10). Current applications based on fuzzy control appear in such diverse areas as the automatic control of trains, road cars, cranes, lifts, nuclear plants, home appliances, etc. Commercial applications in robotics still do not exist; however, numerous research efforts promise that fuzzy robot control systems will be developed, notably in the fields of robotized part processing, assembly, mobile robots, and robot vision systems.

Thanks to its ability to manipulate imprecise and incomplete data, fuzzy logic offers the possibility of incorporating expertise into automatic control systems. Fuzzy logic already has proven itself useful in cases where the process is too complex to be analyzed by conventional quantitative techniques, or where the available information is qualitative, imprecise, or unreliable. Considering that it is based on precise mathematical theory, fuzzy logic additionally offers the possibility of integrating heuristic methods with conventional techniques for analysis and synthesis of automatic control systems, thus facilitating further refinement of fuzzy control-based systems.

24.4.2 Mathematical Foundations

24.4.2.1 Fuzzy Sets

At the heart of fuzzy set theory is the notion of fuzzy sets that are used to model statements in natural (or artificial) language. Fuzzy set is a generalization of classical (crisp) sets. The classical set concept assumes that it is possible to divide particles of some universe into two parts: those that are members of the given set, and those that are not. This partitioning process can be described by means of a characteristic *membership function*. For a given universe of discourse X and a given set A, membership function $\mu_A(\cdot)$ assigns a value to each particle $x \in X$ so that

$$\mu_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

© 2002 by CRC Press LLC

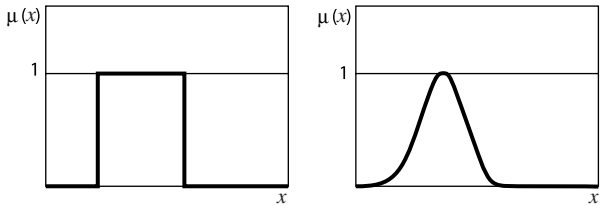


FIGURE 24.11 Membership functions of conventional and fuzzy sets.

With fuzzy sets, the set's boundary is not strict between the members and nonmembers. This softening of the boundary is defined mathematically using the *membership degree function,* which assigns each particle a value that indicates the degree of membership in the given set (see Figure 24.11). Accordingly, fuzzy set $\tilde{A}$ in the universe of discourse X is defined by its degree of membership function $\mu_{\tilde{A}}(\cdot)$ of the form:

$$\mu_{\tilde{A}}: X \mapsto [0, 1].$$

For each fuzzy set, its support can be defined. The support of fuzzy set $\tilde{A}$ is an ordinary set A that contains all elements from the universe X with nonzero membership degrees in $\tilde{A}$:

$$\text{supp}(\tilde{A}) = \{x \in X: \mu_{\tilde{A}}(x) > 0\}.$$

The notion of support allows a formal definition of empty fuzzy sets. An *empty fuzzy set* is a fuzzy set with empty support.

It is customary to represent fuzzy sets by fuzzy singletons. A *fuzzy singleton* is a fuzzy set for which its support is a single particle x from the universe X. If fuzzy set $\tilde{A}$ has a finite support $\text{supp}(\tilde{A}) = \{x_1, x_2, \ldots, x_n\}$ with degrees of membership $\mu_{\tilde{A}}(x_i)$, $i = 1, 2, \ldots, n,$ such a fuzzy set is conveniently written as:

$$\tilde{A} = \mu_{\tilde{A}}(x_1)/x_1 + \mu_{\tilde{A}}(x_2)/x_2 + \cdots + \mu_{\tilde{A}}(x_n)/x_n = \sum_{i=1}^{n} \mu_{\tilde{A}}(x_i)/x_i$$

Here, the plus sign indicates that pairs $\mu_{\tilde{A}}(x_i)/x_i$ collectively form the definition of fuzzy set $\tilde{A}$. If universe X is an interval of real numbers, then the following notation for fuzzy set $\tilde{A}$ in X is customary:

$$\tilde{A} = \int_X \mu_{\tilde{A}}(x)/x$$

The notions of fuzzy subsets and equality between fuzzy sets are also defined in terms of membership degree functions. Fuzzy set $\tilde{A}$ is said to be a *subset* of $\tilde{B}$ if all particles $x \in X$ have degrees of membership to $\tilde{A}$ lower or equal to their degrees of membership to $\tilde{B}$:

$$\tilde{A} = \tilde{B} \text{ iff } \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(x) \text{ for all } x \in X$$

Fuzzy sets are *equal* if their membership functions are equal for all elements in the universe of discourse:

© 2002 by CRC Press LLC

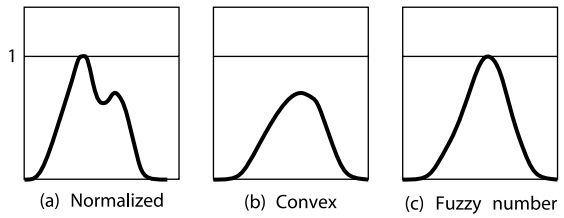


FIGURE 24.12 Examples of fuzzy sets.

$$\tilde{A} = \tilde{B} \text{ iff } \mu_{\tilde{A}}(x) = \mu_{\tilde{B}}(x) \text{ for all } x \in X$$

An important class of fuzzy sets is normalized fuzzy sets. A fuzzy set $\tilde{A}$ is said to be normalized if its height $h(\tilde{A})$, defined as the largest degree of membership attained by elements in its support, is equal to 1:

$$h(\tilde{A}) \equiv \max_X \mu_{\tilde{A}}(x) = 1$$

The value $m \in X$ for which $\mu_{\tilde{A}}(m) = h(\tilde{A})$ is called the *modal value* of the fuzzy set.

Fuzzy set $\tilde{A}$ in Euclidean space R^n is *convex* if, for any vectors $\mathbf{x}, \mathbf{y} \in R^n$, the following is valid:

$$\mu_{\tilde{A}}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \min[\mu_{\tilde{A}}(\mathbf{x}), \mu_{\tilde{A}}(\mathbf{y})]$$

Fuzzy sets that are normalized, convex, and, additionally, have a piecewise continuous membership degree function, are denoted as *fuzzy intervals*. A special class of fuzzy intervals is fuzzy numbers. A *fuzzy number* is a fuzzy interval with an unique modal value. The concept of fuzzy numbers is based on fuzzy artihmetic that may be considered a generalization of classical arithmetic. Examples of membership functions of normalized, convex fuzzy sets, and fuzzy numbers are shown in Figure 24.12.

24.4.2.2 Operations on Fuzzy Sets

The basic principle for generalization of classical mathematical concepts to the field of fuzzy sets is known as the principle of extension.[63] Formally, given a function $f: X \to Y$, mapping elements of ordinal set X into elements of set Y, and an arbitrary fuzzy set $\tilde{A} \in \tilde{P}(X)$, e.g.,

$$\tilde{A} = \mu_1/x_1 + \mu_2/x_2 + \cdots + \mu_n/x_n$$

the principle of extension states that the following relation has to be preserved:

$$\begin{aligned} f(\tilde{A}) &\triangleq f(\mu_1/x_1 + \mu_2/x_2 + \cdots + \mu_n/x_n) \\ &= \mu_1/f(x_1) + \mu_2/f(x_2) + \cdots + \mu_n/f(x_n) \end{aligned}$$

In other words, operations on fuzzy sets should preserve important properties of operations on classical sets. Unfortunately, it turns out that it is not possible to define of basic fuzzy set operations that would preserve all the important properties of the corresponding operations on classical sets. For example, it is shown that arbitrary fuzzy complement, union, and intersection operations satisfying the law of contradiction and law of excluded middle are not distributive. Therefore, the

© 2002 by CRC Press LLC

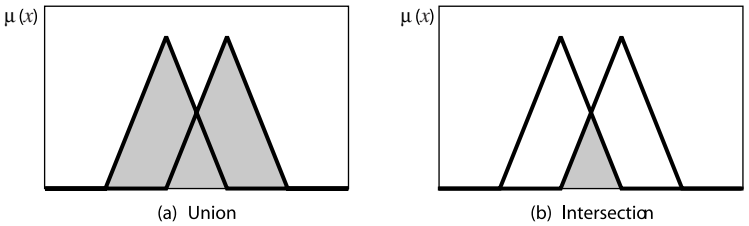


FIGURE 24.13 Standard operations on fuzzy sets.

choice of basic fuzzy set operations has to be made by considering the context in which these operations will be carried out. The most often used set of basic standard operations of fuzzy set theory is (see Figure 24.13):

$$\textit{Complement:} \quad \mu_{\bar{\tilde{A}}}(x) = 1 - \mu_{\tilde{A}}(x)$$

$$\textit{Union:} \quad \mu_{\tilde{A} \cup \tilde{B}}(x) = \max[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$$

$$\textit{Intersection:} \quad \mu_{\tilde{A} \cap \tilde{B}}(x) = \min[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$$

Fuzzy set theory based on such defined operators is usually referred to as *possibility theory.* However, in some situations, different definitions of basic fuzzy set operators are preferable. For example, a union $\tilde{A} \cup \tilde{B}$ intuitively is a disjunction of the concepts represented by $\tilde{A}$ and $\tilde{B}$. Additionally, the notion of union normally implies a certain level of interchangeability between the concepts represented by its arguments. On the other hand, a standard union max operator is rigid in the sense that it does not assume such an interchangeability. If the union were specified by the function

$$f_u: [0, 1] \times [0, 1] \rightarrow [0, 1]$$

that assigns a value $f_u[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$ to given pair of membership degrees $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$, then the intuitive meaning of the union implies the following relation:

$$f_u[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)] \geq \max[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$$

It is evident that standard union operation, defined as $\max[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$, yields the lowest possible degree of membership. For this reason, in some cases, alternative formulations are used in place of the max operator. All potential formulations $f_u(\cdot)$ are required to satisfy the minimum axiomatic conditions:

U1. *Boundary conditions:* $f_u(0, 0) = 0$ and $f_u(0, 1) = f_u(1, 0) = f_u(1, 1) = 1$
U2. *Commutativity:* $f_u(x, y) = f_u(y, x)$
U3. *Monotony:* if $x \leq x'$ and $y \leq y'$, then $f_u(x, y) \leq f_u(x', y')$
U4. *Associativity:* $f_u(f_u(x, y), z) = f_u(x, f_u(y, z))$

The functions satisfying these axioms are called *triangular conorms* (*t-conorms*). Evidently, the standard union operation is a t-conorm. Other t-conorms are proposed as well, such as algebraic sum, bounded sum, etc.

Fuzzy intersection $\tilde{A} \cap \tilde{B}$ intuitively denotes a conjunction of concepts represented by $\tilde{A}$ and $\tilde{B}$. As in the case of union, the intersection operation can be specified using the function:

$$f_i: [0, 1] \times [0, 1] \rightarrow [0, 1]$$

© 2002 by CRC Press LLC

The minimum axiomatic skeleton that functions $f_i(\cdot)$ have to satisfy to qualify as candidates for defining fuzzy intersection consists of conditions:

I1. *Boundary conditions:* $f_i(1, 1) = 1$ and $f_i(0, 0) = f_i(0, 1) = f_i(1, 0) = 0$
I2. *Commutativity:* $f_i(x, y) = f_i(y, x)$
I3. *Monotony:* if $x \le x'$ and $y \le y'$, then $f_i(x, y) \le f_i(x', y')$
I4. *Associativity:* $f_i(f_i(x, y), z) = f_i(x, f_i(y, z))$

The functions satisfying axioms I1–I3 are called *triangular norms* (*t-norms*). Obviously, the standard min operation is a t-norm.

Analogous to the case of the union, intersection of fuzzy sets normally implies a certain requirement level for the simultaneous satisfaction of concepts represented by its arguments. On the other hand, the standard min operation is rigid in the sense that it does not account for the benefits of simultaneous memberships. Hence, alternative t-norms are proposed in which different intensities of intersections are achieved: algebraic product, bounded product, etc. Standard min operation is the upper bound of the possible intersection operations (the weakest intersection).

24.4.2.3 Fuzzy Relations

Fuzzy relations are generalizations of the classical concept of relations among elements of two or more sets. Additionally, fuzzy relations allow the specification of different levels of strength of association among individual elements. The levels of association are represented by degrees of membership to the fuzzy relations, in the same manner as the degree of membership to a fuzzy set is represented.

Formally, a fuzzy relation among elements of ordinary sets $X_1, X_2, \ldots, X_n$ is a fuzzy subset $\tilde{R} = \tilde{R}(X_1, X_2, \ldots, X_n)$ of Cartesian product $X_1 \times X_2 \times \ldots \times X_n$ and it is defined by the membership degree function:

$$\mu_{\tilde{R}}: X_1 \times X_2 \times \ldots \times X_n \to [0,1]$$

Thus, tuples $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in X_1 \times X_2 \times \ldots \times X_n$ may have different degrees of membership $\mu_{\tilde{R}}(x_1, x_2, \ldots, x_n) \in [0,1]$ to the fuzzy relation.

When the sets $X_1, X_2, \ldots, X_n$ are finite, fuzzy relation $\tilde{R}(X_1, X_2, \ldots, X_n)$ is suitably represented by an *n*-dimensional *membership matrix,* whose elements show the degree to which the individual tuples belong to a given fuzzy relation. For instance, binary fuzzy relation $\tilde{R}(X, Y)$ between sets $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ is conveniently represented by the matrix:

$$\tilde{\mathbf{R}} = \begin{bmatrix} \mu_{x_1,y_1} & \cdots & \mu_{x_1,y_m} \\ \vdots & \ddots & \vdots \\ \mu_{x_n,y_1} & \cdots & \mu_{x_n,y_m} \end{bmatrix}$$

For a given family of sets $\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_n$, defined in the universes $X_1, X_2, \cdots, X_n$, the Cartesian product of fuzzy sets:

$$\tilde{A}_1 \times \tilde{A}_2 \times \cdots \times \tilde{A}_n$$

is a fuzzy set in the universe of discourse $X_1 \times X_2 \times \ldots \times X_n$. Consequently, the Cartesian product is an *n*-ary fuzzy relation with the degree of membership function defined by:

$$\mu_{\tilde{A}_1 \times \tilde{A}_2 \times \cdots \times \tilde{A}_n}(x_1, x_2, \ldots, x_n) = \mu_{\tilde{A}_1}(x_1) * \mu_{\tilde{A}_2}(x_2) * \cdots * \mu_{\tilde{A}_n}(x_n)$$

© 2002 by CRC Press LLC

for all $x_1 \in X_1$, $x_2 \in X_2$, …, $x_n \in X_n$, where the sign $*$ denotes one of the triangular norms (i.e., the intersection operation).

Among the operations over fuzzy relations, compositions of binary relations are of special significance. For ordinary binary relations $P(X, Y)$ and $Q(Y, Z)$, defined in the common set Y, the *composition* of P and Q:

$$R(X, Z) = P(X, Y) \circ Q(Y, Z)$$

is defined as a subset $R \subseteq X \times Z$ such that

$$(x, z) \in R \text{ iff there exists } y \in Y \text{ for which } (x, y) \in P \text{ and } (x, z) \in Q$$

The concept of composition is extended to fuzzy relations in a number of ways aimed at preserving important properties of corresponding compositions of classical relations. The most important types of compositions of binary fuzzy relations are

- *Max–min composition.* Denoted by $\tilde{P}(X,Y) \circ \tilde{Q}(Y,Z)$, this operation is defined by

$$\mu_{\tilde{P}\circ\tilde{Q}}(x,z) = \max_{y\in Y} \min[\mu_{\tilde{P}}(x,y), \quad \mu_{\tilde{Q}}(y,z)]$$

  Thus, the strength of the relation between elements x and z is equal to the strength of the strongest chain between these elements, whereas the strength of each chain x–y–z is equal to the strength of its weakest link.
- *Max–product composition.* The composition is denoted by $\tilde{P}(X,Y) \circ \tilde{Q}(Y,Z)$ and defined by:

$$\mu_{\tilde{P}\circ\tilde{Q}}(x,z) = \max_{y\in Y}[\mu_{\tilde{P}}(x,y) \cdot \mu_{\tilde{Q}}(y,z)]$$

The max–min and max–product compositions may be regarded as specializations of the more general *sup-star composition,* denoted by $\tilde{P}(X,Y) \circ \tilde{Q}(Y,Z)$ and defined by

$$\mu_{\tilde{P}\circ\tilde{Q}}(x,z) = \sup_{y\in Y}[\mu_{\tilde{P}}(x,y) \ * \ \mu_{\tilde{Q}}(y,z)]$$

where the sign $*$ represents any triangular norm, and the sup operator denotes supremum (the lowest upper bound).

Compositions of binary relations in finite sets may be efficiently realized using membership matrices. For example, the composition $\tilde{P} \circ \tilde{Q}$ can be calculated as a matrix product:

$$\tilde{\mathbf{P}} \cdot \tilde{\mathbf{Q}}$$

where multiplication is replaced by the min, and addition by the max operator.

24.4.2.4 Fuzzy Logic

Fuzzy logic is a discipline comprising formal principles of approximate reasoning.[64] Its main issue is modeling of imprecise modes of human reasoning in conditions characterized by unreliability and imprecision, whereby the theory of fuzzy sets is used as a basic methodology.

Fuzzy logic is an extension to classical logic, in which the basic objects are logical propositions that may take one of the possible values of truth: true or false, i.e., 1 or 0. Contrary to classical formal systems, fuzzy logic allows evaluation of the truth of a proposition as, e.g., a real number in interval [0,1]. The basis of fuzzy logic is the theory of fuzzy sets. For example, the characterization

© 2002 by CRC Press LLC

of fuzzy set $\tilde{A}$ with membership function $\mu_{\tilde{A}}(x)$, $x \in X$, can be interpreted as the truth value of the proposition:

$$x \text{ is element of } \tilde{A}$$

To enable work with imprecise propositions, fuzzy logic permits use of:

- *Fuzzy predicates.* Truth values of an imprecise predicate $P(x)$ (e.g., x is *small, big,* etc.) can be described for any $x \in X$ by the fuzzy set with membership function $\mu_P(x)$, determined by the predicate $P(\cdot)$.
- *Fuzzy truth values.* Fuzzy sets, defined on the interval [0, 1], can be used to describe different levels of truth (e.g., *fairly true, completely false,* etc.).
- *Fuzzy quantifiers.* In addition to the usual quantifiers from classical logic ($\forall$, $\exists$), imprecise statements may contain imprecise quantifiers (e.g., *sometimes, almost always*) represented by fuzzy numbers.
- *Fuzzy modifiers.* Different forms of fuzzy modifiers (*probably, fairly,* etc.) can be described by utilizing special operations on fuzzy sets representing the modified propositions.

The central problem of *quantitative fuzzy semantics* is calculating the meaning of *linguistic variables,* i.e., the variables whose values are sentences in a specific (natural or artificial) language.[60] The linguistic variable can be regarded as a variable whose value is a fuzzy number (the *meaning* of the variable) or as a variable whose values are linguistically defined (the *label* of the variable).[63] Generally, the label of a linguistic variable is obtained by concatenating the terms of the language according to some rules. In simple cases, these terms can be divided into four categories:

1. *Primary terms* that represent labels of specific fuzzy sets
2. Negation *not* and connectives *or* and *and*
3. *Modifiers* that modify the basic concept to which they are applied (e.g., *very, extremely,* etc.)
4. *Markers,* such as parentheses

Negation *not* and connectives *or* and *and* may be considered labels of the corresponding operations on fuzzy sets:

- *Complement* $\overline{\tilde{A}}$ that represents the fuzzy concept "*not* $\tilde{A}$"
- *Union* $\tilde{A} \cup \tilde{B}$ that represents the fuzzy concept "$\tilde{A}$ *or* $\tilde{B}$"
- *Intersection* $\tilde{A} \cap \tilde{B}$ that represents the fuzzy concept "$\tilde{A}$ *and* $\tilde{B}$"

Linguistic modifiers can be expressed by specific operations on the fuzzy set $\tilde{A}$ describing the basic concept, e.g.,

- *Exponent* $\tilde{A}^{\alpha}$, defined as

$$\mu_{\tilde{A}^{\alpha}}(x) = [\mu_{\tilde{A}}(x)]^{\alpha}$$

- *Concentration,* defined as

$$\text{con}\,(\tilde{A}) = \tilde{A}^2$$

The operation of concentration may be interpreted as "*very* $\tilde{A}$" and its effect is a large reduction of the degrees of membership of those values of x that already have a small degree of membership $\mu_{\tilde{A}}(x)$ to the basic concepts $\tilde{A}$, with an additional small reduction for those x with high membership $\mu_{\tilde{A}}(x)$.

© 2002 by CRC Press LLC

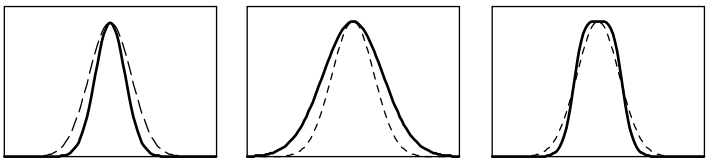


FIGURE 24.14 Examples of linguistic modifiers.

- *Dilatation,* defined by

$$\text{dil}(\tilde{A}) = \tilde{A}^{0.5}$$

The operation of dilatation can be described as "*more or less* $\tilde{A}$" and its effect is opposite to that of concentration.

- *Contrast intensification* int($\tilde{A}$), defined as

$$\mu_{\text{int}(\tilde{A})}(x) = \begin{cases} 2[\mu_{\tilde{A}}(x)]^2 & \text{for } \mu_{\tilde{A}}(x) \le 0.5 \\ 1 - 2[1 - \mu_{\tilde{A}}(x)]^2 & \text{for } \mu_{\tilde{A}}(x) > 0.5 \end{cases}$$

This operation has the consequence of increasing the values $\mu_{\tilde{A}}(x)$ that are above the *crossover point* 0.5 and the reduction of values below this point.

Examples of linguistic modifiers are illustrated in Figure 24.14.

Among various forms of fuzzy propositions, fuzzy implications are of special importance. *Fuzzy implication* is a statement of the form:

$$\tilde{A}(x) \Rightarrow \tilde{B}(y)$$

or, equivalently,

$$\text{if } \tilde{A}(x) \text{ than } \tilde{B}(y)$$

where $x \in X$, $y \in Y$ are linguistic variables and $\tilde{A}(\cdot)$ $\tilde{B}(\cdot)$ are fuzzy predicates in universes of discourse *X, Y,* respectively. Essentially, such a statement describes the fuzzy relation:

$$\tilde{R}_{\Rightarrow}(X,Y) \subseteq X \times Y$$

between the two fuzzy sets, i.e., between the equivalent fuzzy propositions $\tilde{A}(x)$ and $\tilde{B}(y)$.

Fuzzy implication is important because of its role in automatic inferencing. The two basic *fuzzy inference rules* that are based on fuzzy implication are

- *Generalized modus ponens*: $(\tilde{A}'(x) \wedge (\tilde{A}(x) \Rightarrow \tilde{B}(y))) \Rightarrow \tilde{B}'(y)$
- *Generalized modus tollens*: $(\tilde{B}'(y) \wedge (\tilde{A}(x) \Rightarrow \tilde{B}(y))) \Rightarrow \tilde{A}'(x)$

The generalized modus ponens is closely related to the mechanism of forward inferencing (data-driven inference) and it reduces to the classical modus ponens when $\tilde{A}' = \tilde{A}$ and $\tilde{B}' = \tilde{B}$. Analogously, the generalized modus tollens is closely related to the mechanism of backward inferencing (goal-driven inference) and it reduces to the classical modus tollens when $\tilde{A}' = \overline{\tilde{A}}$ and $\tilde{B}' = \overline{\tilde{B}}$.

© 2002 by CRC Press LLC

A basic technique that lies at the heart of most implementations of automatic fuzzy inference is the *compositional rule of inference* proposed by Zadeh.[62] According to this rule, binary fuzzy relation $\tilde{R}$ from X to Y and fuzzy set $\tilde{x} \subseteq X$ induce the fuzzy set $\tilde{y} \subseteq Y$ determined by the sup-star composition

$$\tilde{y} = \tilde{x} \circ R$$

in which $\tilde{x}$ plays the role of unary fuzzy relation. When setting $\tilde{R} = \left(\tilde{A} \Rightarrow \tilde{B}\right) \tilde{x} = \tilde{A}'$, and $\tilde{y} = \tilde{B}'$ in the compositional rule, the rule becomes an implementation of generalized modus ponens:

$$\tilde{B}' = \tilde{A}' \circ (\tilde{A} \Rightarrow \tilde{B}).$$

If $\tilde{A}, \tilde{A}', \tilde{B}, \tilde{B}'$ are nonfuzzy and $\tilde{A}' = \tilde{A}$, the compositional rule of inference becomes

$$\tilde{B}' = \tilde{A} \circ (\tilde{A} \Rightarrow \tilde{B}) = \tilde{B}.$$

Thus, the compositional rule can be regarded as an approximate extension, i.e., a fuzzy generalization of modus ponens: The more different $\tilde{A}'$ is from $\tilde{A}$, the less sharply defined is $\tilde{B}'$.

Because of the significance of fuzzy implication, a number of distinct fuzzy implication functions have been proposed for its implementation. The proposed functions can be divided into five families:[14]

1. Material implication: $\tilde{A} \Rightarrow \tilde{B} \stackrel{\Delta}{=} \overline{\tilde{A}} \cup \tilde{B}$
2. Implication in propositional calculus: $\tilde{A} \Rightarrow \tilde{B} \stackrel{\Delta}{=} \overline{\tilde{A}} \cup (\tilde{A} \cap \tilde{B})$
3. Extended implication in propositional calculus: $\tilde{A} \Rightarrow \tilde{B} \stackrel{\Delta}{=} (\overline{\tilde{A}} \cap \overline{\tilde{B}}) \cup \tilde{B}$
4. Generalization of modus ponens: $\tilde{A} \Rightarrow \tilde{B} \stackrel{\Delta}{=} \sup_{\tilde{C}} \{\tilde{C} : \tilde{C} \cap \tilde{A} \subseteq \tilde{B}\}$
5. Generalization of modus tollens: $\tilde{A} \Rightarrow \tilde{B} \stackrel{\Delta}{=} \inf_{\tilde{C}} \{\tilde{C} : \tilde{C} \cup \tilde{B} \subseteq \tilde{A}\}$

Several authors have analyzed axiomatic requirements and criteria for selection of appropriate functions for implementation of fuzzy implication.[4,15,28] One of the widely accepted definitions is the *standard fuzzy implication*, an implementation of generalized modus ponens in which the standard union and intersection operators are used:

$$\mu_{\tilde{A} \Rightarrow \tilde{B}}(x, y) = \begin{cases} 1 & \text{for } \mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(y) \\ \mu_{\tilde{B}}(y) & \text{for } \mu_{\tilde{A}}(x) > \mu_{\tilde{B}}(y) \end{cases}$$

24.4.3 Fuzzy Controller

Fuzzy control approaches the control problem in a radically different way compared to the traditional model-based techniques. Instead of precise mathematical models, fuzzy control uses an imprecise and incomplete description of the process and/or the way the system is controlled by human operators, where the theory of fuzzy sets is used as a principle tool.

A fuzzy controller consists of four basic components (see Figure 24.15): condition (fuzzification) interface, knowledge base, inference mechanism, and action interface.

The block denoted as the condition interface performs measurement of input (state) variables:

$$x = [x_1, x_2, \ldots, x_n]^T$$

© 2002 by CRC Press LLC

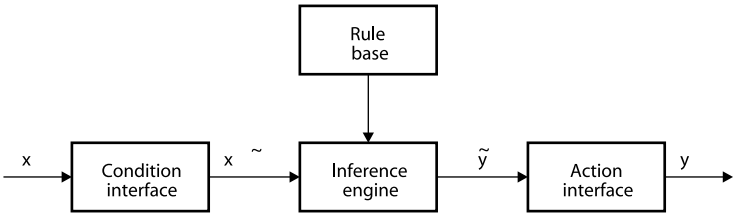


FIGURE 24.15 Components of fuzzy controller.

of the controlled process and translates them into fuzzy linguistic terms X_1, X_2, …, X_n that are represented by fuzzy sets $\tilde{x}_1, \tilde{x}_2,\ldots,\tilde{x}_n$ in appropriate universes of discourse U_1,U_2, …, U_n, respectively. The obtained fuzzy values constitute the *fuzzy state* of the process:

$$\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2,\ldots,\tilde{x}_n]^T$$

in the state space $U = U_1 \times U_2 \times \ldots \times U_n$. The fuzzy state variables are further used in evaluation of fuzzy control rules.

The knowledge base consists of control rules and a fuzzy set definition base. The definition base provides the definitions necessary to characterize fuzzy control rules and manipulation of fuzzy data. The rule base consists of heuristic fuzzy control rules that describe control goals and policy. A *fuzzy control rule* is a fuzzy conditional statement (fuzzy implication) in which the antecedent is a condition and the consequent is a control action. Thus, the rule base can be represented as:

R_1: if X_1 is A_{11} and … and X_n is A_{1n}, then $Y_1 = B_{11}$ and … and $Y_m = B_{1m}$

R_2: if X_1 is A_{21} and … and X_n is A_{2n}, then $Y_1 = B_{21}$ and … $Y_m = B_{2m}$

⋮

R_r: if X_1 is A_{r1} and … and X_n is A_{rn}, then $Y_1 = B_{r1}$ and … and $Y_m = B_{rm}$

where:

$X_1, X_2, \ldots, X_n$ = labels of fuzzy state variables $\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_n$ in universes $U_1, U_2, \ldots,U_n$

$Y_1, Y_2, \ldots, Y_m$ = labels of fuzzy actions $\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_n$ in universes $V_1, V_2, \ldots, V_m$

A_{ki} and B_{kj} = labels of fixed linguistic values represented by fuzzy sets $\tilde{a}_{ki} \subseteq U_i, \tilde{b}_{kj} \subseteq V_j$

Rules R_k, $k = 1, 2, \ldots, r$ are also mutually interconnected via implicit connectives. Each control rule is implemented by the fuzzy relation $\tilde{\mathbf{R}}_k$ in $\mathbf{U} \times \mathbf{V}$, where $\mathbf{U} = U_1 \times U_2 \times \ldots \times U_m$ and $\mathbf{V} = V_1 \times V_2 \times \ldots \times \boldsymbol{V_m}$. The rule base is an aggregate of individual rules. By integration of particular relations $\tilde{\mathbf{R}}_k, k = 1, 2, \ldots, r$, the aggregate relation of the whole rule base is obtained as:

$$\tilde{\mathbf{R}} \subseteq \mathbf{U} \times \mathbf{V}$$

The block designated as the inference mechanism is responsible for evaluation of control rules. Evaluation is commonly carried out using the sup-star compositional rule:[62]

$$\tilde{\mathbf{y}} = \tilde{\mathbf{x}} \circ \tilde{\mathbf{R}}$$

The result is the *fuzzy control action* $\tilde{\boldsymbol{y}} = [\tilde{y}_1\ \tilde{y}_2 \ldots \tilde{y}_m]^T$ in universe $\mathbf{V}$ of possible control actions. Within the action interface, the fuzzy action $\tilde{\mathbf{y}}$ is converted into defuzzified action $\mathbf{y} = (y_1, y_2, \ldots, y_m)$.

© 2002 by CRC Press LLC

24.4.3.1 Condition Interface

The task of the condition interface is (1) to perform *scale mapping,* which transfers the range of values of input variables into corresponding universes of discourse, and (2) to perform *fuzzification,* which converts crisp inputs into fuzzy sets.

The most frequent fuzzification strategy consists of transforming the measured value x into a fuzzy singleton $\tilde{x}$. Thus, input x is interpreted as a fuzzy set $\tilde{x}$ with the membership function equals zero in all points $u \in U$ except for the point u_0, where $\mu_{\tilde{x}}(u_0) = 1$.

24.4.3.2 Fuzzy Set Definition Base

The fuzzy set definition base contains definitions of fuzzy sets $\tilde{a}_{ki}$ and $\tilde{b}_{kj}$ ($i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. $k = 1, 2, \ldots, r$) that correspond to linguistic labels A_{ki} and B_{kj} appearing in the control rules. These fuzzy sets are frequently designated as *primary fuzzy sets.*

The universes of discourse for input and output control signals can be discrete or continuous. To attain a more efficient manipulation with fuzzy sets, two basic transformations are commonly applied to the input/output spaces:

- *Normalization*, by which the universe of discourse U is transformed into the normalized closed interval $U_N = [-1, +1]$. The transformation function $f_N(\cdot)$ may be linear or nonlinear and its synthesis assumes *a priori* knowledge on the possible range $U = [u_{\min}, u_{\max}]$ of the signal. For the case of linear mapping,

$$f_N(u) = [(u - u_{\max}) + (u - u_{\min})]/(u_{\max} - u_{\min})$$

  By choosing appropriate nonlinear transformation, a uniform distribution of symmetric and mutually equal primary sets may be achieved.
- *Discretization* (*quantization*), by which the continuous universe U or U_N is partitioned into a finite number of segments:

$$\overline{u}_1 = [\hat{u}_0, \hat{u}_1], \overline{u}_2 = (\hat{u}_1, \hat{u}_2], \ldots, \overline{u}_q = (\hat{u}_{q-1}, \hat{u}_q],$$

- specified by quantization levels $u_{\min} = \hat{u}_0 < \hat{u}_1 < \cdots < \hat{u}_q = \hat{u}_{\max}$. Each segment $\overline{u}_i$, $i = 1, 2, \ldots, q$ is treated as a generic element representing all elements $u \in \overline{u}_i$. In this manner, fuzzy sets can now be defined by assigning degrees of membership to each generic element of the universe $\overline{U} = \{\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_q\}$.

  Quantization may also be linear or nonlinear. The number of quantization levels should be sufficiently large to ensure adequate approximation and yet be small enough to save memory space. In the majority of applications, the number of quantization levels is 16 to 32.

Primary fuzzy sets are usually represented by linquistic labels such as: NB, negative big; NM, negative medium; NS, negative small; ZE, zero; PS, positive small; PM, positive medium; and PB, positive big. The set of different labels:

$$A_i = \bigcup_{k=1}^{r} \{A_{ki}\}$$

is called *fuzzy input space* of the i-th input variable, $i = 1, 2, \ldots, n$. Analogously, the set of different labels:

$$B_j = \bigcup_{k=1}^{r} \{B_{kj}\}$$

© 2002 by CRC Press LLC

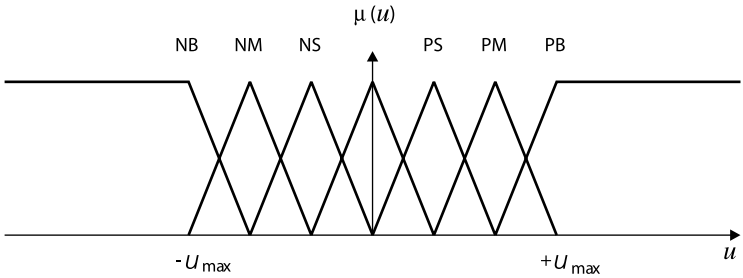


FIGURE 24.16 Primary fuzzy sets.

is called *fuzzy control space* of *j*-th control variable, $j = 1, 2, \ldots, m$. The number of different labels in fuzzy input space determines the number of possible control rules. Finding the optimal fuzzy partition of the input space is a difficult task and is usually performed in a heuristic way.

Depending on whether the underlying universe of discourse is continuous or discrete, primary fuzzy sets are specified using a functional or numerical definition. For the case of continuous universe, commonly applied functional forms of membership functions are

- *Triangular functions*: $\mu_f(x) = \max\left(1 - \left|1 + \frac{x - x_f}{\sigma_f}\right|, 0\right)$
- *Bell-shaped functions*: $\mu_f(x) = e^{-\frac{1}{2}\left(\frac{x - x_f}{\sigma_f}\right)^2}$

An example of triangular primary fuzzy sets is given in Figure 24.16. If the universe of discourse is discrete, a fuzzy set is represented as a vector whose elements are values of the membership degree.

24.4.3.3 Control Rules

A rule R_k in the rule base typically takes the form of a *state evaluation fuzzy control rule*:

R_k: if (X_1 is A_{k1} and … and X_n is A_{kn}), then $Y = B_k$

Linguistic variables appearing on the left side of the implication are typically the process state error (i.e., deviation from desired state), and error change (i.e., time derivative of the error). The variable on the right side is usually the control output or a change of the control output.

A more general form is the *functional control rule,* where premises and consequences are specified as (logical) functions:

$$R_k\text{: if } f_k\ (X_1 \text{ is } A_{k1}, \ldots, X_n \text{ is } A_{kn}), \text{ then } Y = g_k\ (X_1, X_2, \ldots X_n)$$

A popular form of functional rules is the *Sugeno-type rules*,[44] where the antecendent propositions are connected by fuzzy conjunction, and the consequent is a linear function of input variables:

R_k: if X_1 is A_{k1} and … and X_n is A_{kn}, then $y_k = g_{k0} + g_{k1} \times_1 + \cdots + g_{kn}\, x_n$

The principal questions about implementation of fuzzy controllers are connected to the derivation and validation of control rules. Keeping in mind that fuzzy control is primarily efficient in cases when only qualitative and incomplete information is available, the rules are often derived in a heuristic way. Use of expert knowledge and imitation of procedures employed by trained operators are commonly used. Adjustments of control parameters to improve system performances are often made using *ad hoc* procedures that usually reduce to trial-and-error.

Intensive investigations were conducted on development of systematic methods of deriving fuzzy control rules. Most of them use notion of the *fuzzy process model,* i.e., the linguistic description of dynamic characteristics of controlled process.[45,57] The fuzzy model approaches identification of

© 2002 by CRC Press LLC

linguistic structure and model parameters in a systematic manner. Based on the known fuzzy model, control rules can be generated for attaining optimal system behavior.

An alternative solution lies in adding fuzzy controller learning capabilities, i.e., facilities to adapt general control rules to an actual situation. In principle, a fuzzy controller with learning capabilities has a hierarchical structure consisting of two rule bases: a general rule base and base of meta-rules. The meta-rules exhibit human-like learning ability to create and modify the general rules based on the observed and desired performance characteristics of the system. The first such system with learning capabilities, a *self-organizing controller,* was described by Procyk and Mamdani.[37]

24.4.3.4 Inference Mechanism

Consider the rule base:

$R_{1:}$ if X_1 is A_{11} and … and X_n is A_{1n}, then $Y = B_1$

$R_{2:}$ if X_1 is A_{21} and … and X_n is A_{2n}, then $Y = B_2$

⋮

$R_{r:}$ if X_1 is A_{r1} and … and X_n is A_{rn}, then $Y = B_r$

The antecedent of each rule R_k, $k = 1, 2, \ldots, r$ is interpreted as a fuzzy set:

$$\tilde{\boldsymbol{a}}_k = \tilde{a}_{k1} \times \tilde{a}_{k2} \times \cdots \times \tilde{a}_{kn}$$

in the product space $\mathbf{U} = U_1 \times U_2 \times \cdots \times U_n$, with the membership function $\boldsymbol{\mu}_{\tilde{\mathbf{a}}_k(\cdot)}$ given for all $\mathbf{u} \in \mathbf{U}$ by:

$$\mu_{\tilde{\mathbf{a}}_k}(u) = \mu_{\tilde{\mathbf{a}}_k}(u_1, u_2, \ldots, u_n) = f_i(\mu_{\tilde{\mathbf{a}}_{k1}}(u_1), u_{\tilde{\mathbf{a}}_{k2}}(u_2), \ldots, u_{\tilde{\mathbf{a}}_{kn}}(u_n))$$

where $f_i(\cdot)$ denotes any t-norm (intersection) function, such as min or algebraic product. Thus, the rule base can be represented in the form:

R_1: if $\tilde{\mathbf{a}}_1$, then $\tilde{b}_1$

R_2: if $\tilde{\mathbf{a}}_2$, then $\tilde{b}_2$

⋮

R_r: if $\tilde{\mathbf{a}}_r$, then $\tilde{b}_r$

where the antecendents are fuzzy sets in the universe **U**, and the consequents are fuzzy sets in the universe *V.*

If the rule base is *complete* (i.e., it contains all possible fuzzy conditions and, additionally, for every input **u** there exists a *dominant rule* R_k with applicability degree $\mu_{\tilde{\mathbf{a}}_k}(\mathbf{u})$ higher than some level, say, 0.5), then such a base can be interpreted as a sequence of *fuzzy conditional statements:*

if $\tilde{\mathbf{a}}_1$, then $\tilde{b}_1$

else if $\tilde{\mathbf{a}}_2$, then $\tilde{b}_2$

⋮

else if $\tilde{\mathbf{a}}_r$, then $\tilde{b}_r$

© 2002 by CRC Press LLC

It is natural to interpret the fuzzy conditional statements as Cartesian products, and connectives between the conditional statements as union. Thus, the relation represented by the rule base is naturally implemented as:

$$\tilde{\mathbf{R}} = \bigcup_{k=1}^{r} \tilde{\mathbf{R}}_k = \bigcup_{k=1}^{r} \tilde{\mathbf{a}}_k \times \tilde{b}_k$$

Taking into account computational aspects, Cartesian products $\tilde{\mathbf{R}}_k, k = 1, 2, \ldots, r$ are frequently implemented by using min or algebraic product functions, yielding two commonly used *operation rules:*

- *Mamdani's mini-operation rule*: $\mu_{\tilde{\mathbf{R}}_k}(\mathbf{u}, v) = \min(\mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}), \mu_{\tilde{b}_k}(v))$;
- *Larsen's product operation rule*: $\mu_{\tilde{\mathbf{R}}_k}(\mathbf{u}, v) = \mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}).\mu_{\tilde{b}_k}(v)$.

Detailed analysis of the influence of different fuzzy implication functions and union and intersection operators on control quality can be found in, e.g., Lee,[28] Mizumoto,[32] and Stachowicz and Koshanska.[42]

The inference mechanism is based on the sup-star compositional rule of inference:

$$\tilde{y} = \tilde{\mathbf{x}} \circ \tilde{\mathbf{R}}$$

The rule is usually implemented using sup-min or sup-product compositional operator. If this is the case, and if the union is implemented using the max function, fuzzy control action $\tilde{y}$ can be expressed as:

$$\tilde{y} = \bigcup_{k=1}^{r} \tilde{\mathbf{x}} \circ \tilde{\mathbf{R}}_k = \bigcup_{k=1}^{r} \tilde{y}_k$$

or, in terms of the degree of membership function, as:

$$\mu_{\tilde{y}}(v) = \max_{k \in 1,\ldots,r} \tilde{y}_k$$

where $\tilde{y}_k = \tilde{\mathbf{x}} \circ \tilde{\mathbf{R}}_k$ is a local fuzzy control action inferred from the *k*th rule. In terms of the degree of membership function, the local fuzzy control action is determined by:

$$\mu_{\tilde{y}_k}(v) = \begin{cases} \sup_{\mathbf{u}} \min[\mu_{\tilde{\mathbf{x}}}(\mathbf{u}), \mu_{\tilde{\mathbf{R}}_k}(\mathbf{u}, \mathbf{v})] & \text{in case of sup-min composition} \\ \sup_{\mathbf{u}} [\mu_{\tilde{\mathbf{x}}}(\mathbf{u}) \cdot \mu_{\tilde{\mathbf{R}}_k}(\mathbf{u}, v)] & \text{in case of sup-product composition} \end{cases}$$

If input $\tilde{x}$ is a fuzzy singleton with the membership function equaling zero at all points except at the point $\mathbf{u}_0$ at which $\mu_{\tilde{\mathbf{x}}}(\mathbf{u}_0) = 1$, then both versions of the compositional rule of inference reduce to:

$$\mu_{\tilde{y}_k}(v) = \mu_{\tilde{\mathbf{R}}_k}(\mathbf{u}_0, v)$$

In this way, local fuzzy action is determined by the membership function:

$$\mu_{\tilde{y}_k}(v) = \begin{cases} \min[\mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0), \mu_{\tilde{b}_k}(v)] & \text{in case of Mamdani's rule} \\ \mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0) \cdot \mu_{\tilde{b}_k}(v) & \text{in case of Larsen's rule} \end{cases}$$

© 2002 by CRC Press LLC

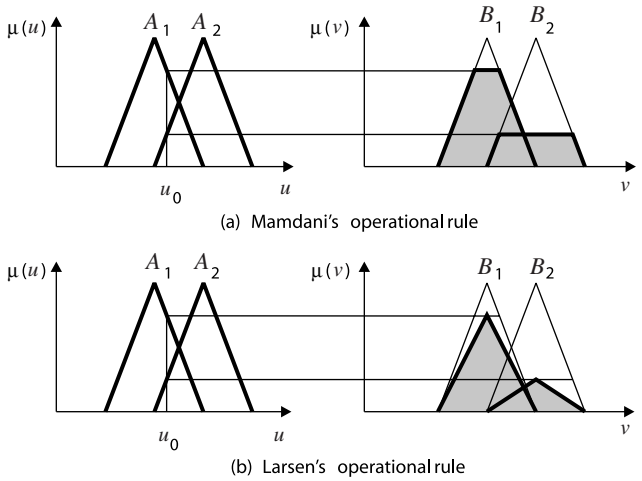


FIGURE 24.17 Operational rules.

The quantity $\mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0)$ is referred to as the *firing strength* of the kth rule and represents a measure of the contribution of the kth rule to integral fuzzy control action. It is computed by applying the corresponding operation rule:

$$\mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0) = \begin{cases} \min[\mu_{\tilde{\mathbf{a}}_{k1}}(u_{10}), \mu_{\tilde{\mathbf{a}}_{k2}}(u_{20}), \ldots, \mu_{\tilde{a}_{kn}}(u_{n0})] & \text{for Mamdani's rule} \\ \mu_{\tilde{\mathbf{a}}_{k1}}(u_{10}) . \mu_{\tilde{\mathbf{a}}_{k2}}(u_{20}), \ldots, \mu_{\tilde{\mathbf{a}}_{kn}}(u_{n0}) & \text{for Larsen's rule} \end{cases}$$

The mechanism of inference in these two methods is illustrated in Figure 24.17. The first method has the advantage of enhancing the contribution of the dominant rule, so that it is widely used in fuzzy applications. On the other hand, the second method has an advantage of preserving the contribution of all rules to the control action.

The obtained fuzzy control action $\tilde{y}$ (or a set of local control actions y_k, $k = 1, 2, \ldots, r$) is transferred to the action, i.e., defuzzification interface, where the actual crisp control signal is generated.

24.4.3.5 Action Interface

Degree of membership function of fuzzy control action can be interpreted as a distribution of possibility $\mu_{\tilde{y}}(v)$ to achieve a control goal by the signal v. Action interface's purpose is to generate a control signal that will best represent the possibility distribution of inferred fuzzy action. Frequently used strategies employed by the action interface are

- *The mean-of-maximum method.* With this strategy, control action is derived as a mean value of all points v at which the membership function of the fuzzy control reaches the global maximum $M = \max_v[\mu_{\tilde{y}}(v)]$. In case of a discrete universe $V = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_q\}$ the control y is computed as:

$$y = \frac{1}{|L|} \sum_{l \in L} \bar{v}_l$$

where L is a set of all indices for which the grade membership function reaches maximum, i.e., $L = \{l : \mu_{\tilde{y}}(\bar{y}_l) = M\}$.

© 2002 by CRC Press LLC

- *The center-of-mass method.* With this method, control is generated as the center of mass of the possibility distribution of fuzzy control. In case of the discrete universe,

$$y = \frac{\sum_{l=1}^{q} \mu_{\tilde{y}}(\bar{v}_l) \cdot v_l}{\sum_{l=1}^{q} \mu_{\tilde{y}}(\bar{v}_l)}.$$

Analytical indices that could serve as guides for selecting the preferred method are not known. However, experiments conducted by several authors, e.g., Mandič et al.[31] have shown that the mean-of-maximum method yields a better transient performance while the center-of-mass method yields a better steady-state performance.

A slightly different scheme is employed with Sugeno-type rules, where fuzzy logic is employed only to describe conditions for the application of the rule whereas the control actions are fuzzy singletons, i.e., the control signals in classical sense:

$$R_k\text{: if } X_1 \text{ is } A_{k1} \text{ and } \ldots \text{ and } X_n \text{ is } A_{kn} \text{ then } y_k = g_{k0} + \sum_{i=1}^{n} g_{ki} x_i$$

Here, it is natural to employ the firing strengths $\mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0)$ of the rules to directly determine the crisp control signal y. A standard technique is to generate the aggregate signal as a weighted average of local controls, where the firing strengths are used as weighting factors:

$$y = \frac{\sum_{k=1}^{r} \mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0) \cdot \mathbf{y}_k}{\sum_{k=1}^{r} \mu_{\tilde{\mathbf{a}}_k}(\mathbf{u}_0)}$$

24.4.4 Direct Applications

Frequently used control rules in fuzzy controllers are of the type:

$$R_k\text{: if } (E \text{ is } A_k \text{ and } \Delta E \text{ is } B_k) \text{ then } U = C_k$$

where E represents value of error e, ΔE represents error change Δe between successive operation cycles of the controller, and U represents fuzzy control action that is transferred to the action interface which in turn generates control signal u. For the operation of such a controller of special importance are *normalizing gains* that are effectively applied within normalization that takes place during conversion of actual signal values and their fuzzy representations. Normalized values of the signals can be represented as:

$$e' = e/E_{\max} = G_E \cdot e$$

$$\Delta e' = \Delta e/\Delta E_{\max} = G_{\Delta E} \cdot \Delta e$$

$$u' = u/U_{\max} = u/G_U$$

© 2002 by CRC Press LLC

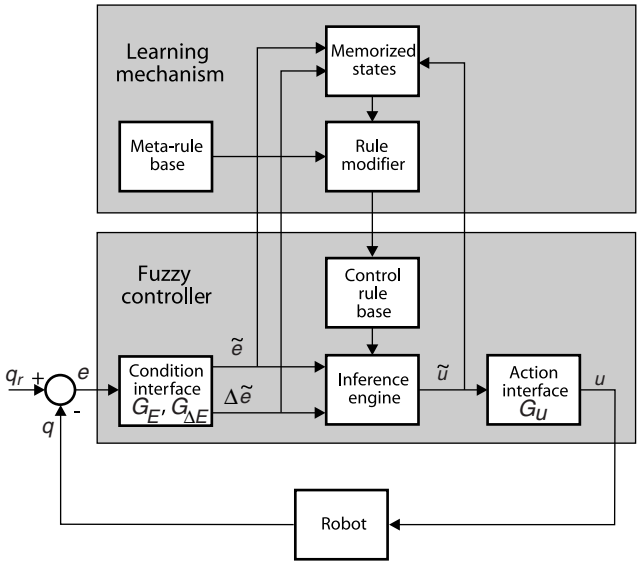


FIGURE 24.18 Self-organizing controller.

where E_{max}, ΔE_{max}, U_{max} are maximum values of the signals, e', $\Delta e'$, u' are corresponding normalized values, and G_E, $G_{\Delta E}$, G_U denote normalizing gains. If the operation of the fuzzy controller is regarded as nonlinear mapping from e and Δe to u, e.g.,

$$u = G_U \cdot f\,(G_E \cdot e,\ G_{\Delta E} \cdot \Delta e)$$

then it is obvious that fuzzy controller is actually a nonlinear PD controller. Conversely, the conventional PD controller can be considered as a special case of the fuzzy PD controller for which the consequent of the rule is

$$u = K_p e + K_d \Delta e$$

Analogously, a fuzzy controller with control rules of the type:

$$R_k\text{: if } (E \text{ is } A_k \text{ and } \Delta E \text{ is } B_k) \text{ then } \Delta U = C_k$$

where ΔU represents change in control output, is actually a nonlinear PI controller. This analogy with conventional controllers is often employed in synthesis and adjustments of fuzzy control.

The first application of fuzzy control to robotic manipulators was described by Mandič, Scharf, and Mamdani.[31] In their 1985 paper, these authors described a series of experiments with a two-DOF robot controlled by two independent self-organizing controllers. Both controllers are of the same structure (see Figure 24.18) that consists of two levels. The lower level is a usual fuzzy controller with control rules, whereas the upper level is a system that realizes the mechanism of automatic learning, i.e., the generation and modification of the rules at the lower level. Control rules employed at the lower level were of the type:

$$R_k\text{: if } (E \text{ is } A_k \text{ and } \Delta E \text{ is } B_k) \text{ then } U = C_k$$

where E represents joint position error $e = q - q_0$, ΔE represents error change, and U represents control action. The upper level is responsible for evaluating controller performance and modifying

© 2002 by CRC Press LLC

the control rule base. Evaluation of performance is achieved using a production system whose structure is identical to the basic fuzzy controller. Performance is evaluated using a local criterion that roughly expresses deviation P between the actual and desired system response. Evaluation criteria are expressed using a set of meta-rules of the form:

$$\Pi_k\text{: if } (E \text{ is } V_{Ek} \text{ and } \Delta E \text{ is } V_{\Delta Ek}), \text{ then } P = V_{Pk}$$

In these rules, parameters of the primary fuzzy set V_{Pk} = zero define a tolerance range for system response, whereas the values different from zero imply the desired degree of correction. Remarkably, such defined rules depend to a very small extent on the controlled process and really express the tolerable errors and degree of acceptability of the errors. If the base of meta-rules is represented by fuzzy relation Π, then the output of the evaluator is a nonlinear function:

$$p = \pi(e, \Delta e)$$

If the precise model of the controlled system were available, then a needed correction in control Δu in principle could be calculated from the known index p. Because the use of the model is always accompanied by inaccuracies, the self-organizing controller instead performs a modification of the control rules that is based on simplified assumptions that (1) the current system performance index $p(t)$ is a consequence of control $u(t - nT)$ generated n operation cycles prior the current time instant t, and (2) the necessary correction in control $\Delta u(t - nT)$ is proportional to $p(t)$:

$$r(t) = \Delta u(t - nT) = \lambda \cdot p(t)$$

In other words, it is accepted that corrections in control are not 100% accurate and that the learning process is slower.

Modifications in the rule base are achieved using fuzzy set operations. Namely, the rule base at the time instant t can be represented as union:

$$\tilde{R}(t) = \bigcup_{k=1}^{r(t)} \tilde{e}_k \times \Delta\tilde{e}_k \times \tilde{u}_k$$

If the function transforming value x into fuzzy singleton $\tilde{x} = F\{x\}$ is denoted by $f\{\cdot\}$, then the control, generated at the time instant $t - nT$ may be regarded as a value that corresponds to the fuzzy implication:

$$\tilde{R}'(t) = f_E\{e(t - nT)\} \times f_{\Delta E}\{\Delta e(t - nT)\} \times f_U\{u(t - nT)\}$$

whereas the desired control at the current time instant is regarded as a value corresponding to the implication:

$$\tilde{R}''(t) = f_E\{e(t - nT)\} \times f_{\Delta E}\{\Delta e(t - nT)\} \times f_U\{u(t - nT) + \lambda \cdot p(t)\}$$

Now, the problem of modifying control rules can be expressed as a problem of substituting implication $\tilde{R}'(t)$ with the implication $\tilde{R}''(t)$. One of the ways to achieve it is to describe the substitution with the expression:

$$\tilde{R}(t + T) = [\tilde{R}(t) \text{ and not } \tilde{R}'(t)] \text{ or } \tilde{R}''(t)$$

© 2002 by CRC Press LLC

or, equivalently:

$$\tilde{R}(t+T) = \left[\tilde{R}(t) \cap \overline{\tilde{R}'(t)}\right] \cup \tilde{R}''(t)$$

Direct application of this formula would lead to exponential growth of the number of rules. Therefore, an approximate method is used where only a single rule is modified at a time. The modified rule is the dominant rule, i.e., the rule $\tilde{R}_k = \tilde{e}_k \times \Delta\tilde{e}_k \times \tilde{u}_k$ for which the heights of intersections $\tilde{e}_k \cap f_E\{e(t-nT)\}$ and $\Delta\tilde{e}_k \cap f_{\Delta E}\{\Delta e(t-nT)\}$ are at least equal to 0.5. Once the dominant rule is identified, its old fuzzy action $\tilde{u}_k$ is replaced by the action $f_U\{u(t - nT) + \lambda \cdot p(t)\}$.

In experiments by Mandič, Scharf, and Mamdani,[31] it was demonstrated on a real robot that a self-organizing controller was able, after only a few adaptation cycles, to attain steady performance that was completely comparable to that of a conventional PID controller. Tanscheit and Scharf,[46] who described several experiments with self-organizing controllers applied to control of a second-order linear system that represented the transfer function of a DC motor with variable load (the variable load was represented by the variable moment of inertia) arrived at similar results.

These and other works, in which direct control of manipulation robots by fuzzy controllers was tried, emphasized two main problems. The first is manifested by the lack of analytical tools for control synthesis, i.e., the selection of parameters of fuzzy controllers (or initial values of the parameters in cases of self-organizing controllers). Second, ordinary fuzzy controllers have attained performances similar to, or slightly better than simple PID schemes. Therefore, it may be expected that direct application of fuzzy controllers will not yield satisfactory performance in more complex robotic tasks, such as tracking fast trajectories. The appearance of these problems can be partially explained by the fact that the early works primarily concentrated on demonstrating the ability of fuzzy logic-based methods to effectively master the nonlinear control problems without need for exact mathematical modeling of the controlled system. For this reason, the role of *a priori* available mathematical knowledge (in situations where the system dynamics is deterministic) as well as the established model-based control techniques was somewhat overshadowed.

24.4.5 Hybridization with Model-Based Control

The problem of merging fuzzy logic-based control with analytic methodologies to exploit advantages of both approaches in real-time robot control was addressed by several authors. Lim and Hiyama[29] have proposed a decentralized control strategy that incorporates a PI controller and a simple fuzzy logic controller. In their approach, the PI controller was used to enhance transient response and steady-state accuracy, whereas fuzzy control was to enhance damping of the overall system. A tighter connection between fuzzy and standard control methods was proposed by Tzafestas and Papanikolopoulos,[50] who suggested employing a two-level hierarchy in which a fuzzy logic-based expert system is used for fine tuning low-level PID control. A similar approach was applied to robot control by Popović and Shekhawat.[36] However, the two-level control hierarchy by itself does not actually solve the problem of weak performance in situations characterized by quickly varying robot dynamics. In such cases, knowledge of readily available mathematical models of robot dynamics cannot be ignored. Therefore, fuzzy logic-based control should not be viewed as a pure alternative to model-based robot control. Instead, a combined approach is preferred, and it may yield superior control schemes over both simple model-based or fuzzy logic-based approaches.

The general idea behind the hybrid approach is utilization of a satisfactory approximation of the model of robot dynamics to decrease dynamic coupling between robot joints and then engage the fuzzy logic-based heuristics as a effective tool for creating a nonlinear performance-driven PID control to handle the effects uncovered by the approximate model. A similar concept was formulated by de Silva and MacFarlane,[8] who proposed a three-level hierarchy for robot control. The proposed hierarchy consists of:

© 2002 by CRC Press LLC

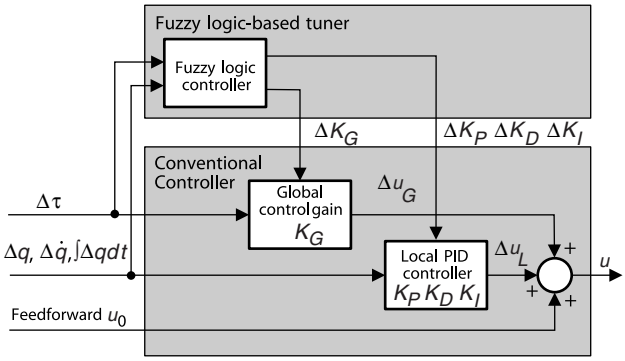


FIGURE 24.19 Hybrid control scheme.

1. Conventional robot controller, i.e., a set of PID controllers closed around a fast decoupling controller. The important role of this controller is to assure the decoupling and linearization needed to efficiently apply expert knowledge for tuning the PID controllers (in an idealized case, every joint subsystem would behave as an independent oscillator with PID control).
2. Intelligent pre-processor, i.e., a set of knowledge-based observers. Each observer is implemented as a fuzzy system and its outputs are the attributes of response at a corresponding joint (e.g., accuracy, oscillations, error convergence, divergence, and steady-state error).
3. Fuzzy tuner, i.e., a fuzzy controller that is used for tuning the gains of the PID controllers at the lowest level.

de Silva and MacFarlane[8] have tested their approach by simulation of a two-link manipulator with an assumption of idealized effectiveness of the low-level global nonlinear feedback. Therefore, robot dynamics was approximated in their work by a set of joint subsystems modelled as second-order systems with unknown acceleration-type disturbances.

The idea of a hybrid approach to robot control was elaborated in detail by Vukobratović and Karan,[52,53] who have employed fuzzy logic to express control policy and have determined analytically conditions on values of fuzzy control parameters that assure stability of a closed-loop robot control system. The authors have analyzed a hybrid design that is an extension to decentralized control strategy. The proposed controller consists of a set of subsystems closed around individual robot joints where each of the subsystems comprises two components: conventional model-based controller and fuzzy logic-based tuner (see Figure 24.19). Inputs to ith joint subsystem, $i = 1, \ldots, n$, where n is the number of actuated joints, are nominal control signal u_{0i}, joint position error Δq_i, joint velocity error $\Delta\dot{q}_i$, and integral error $\int \Delta q_i dt$. In cases where a highly precise tracking of fast trajectory is necessary, an optional global feedback loop (full dynamic compensation) can be added. Global feedback is generated on the basis of computed or measured deviation of dynamic torque $\Delta\tau_i$ acting at the joint and is synthesized to assure practical system stability.[54] A further refinement introduces the upper level that tunes the gains of the PID controllers. The tuner is designed as a fuzzy controller that monitors joint response characteristics and modifies the gains to provide better responses for large deviations of monitored quantities. Although its general structure permits construction of sophisticated control rules for tuning the gains, Vukobratović and Karan have considered a simple decentralized scheme consisting of independent joint servo tuners operating on the basis of observed joint position error Δq, velocity error $\Delta\dot{q}$, and integral error $\int \Delta q dt$. A rather simple heuristics for synthesizing gain-tuning rules was used:

1. If the observed errors are large and do not show a significant tendency to decrease, the proportional gain is enlarged to speed-up error convergence.
2. When the errors are small, the proportional gain is decreased to prevent resonance oscillations and attenuate undesired noise effects.

© 2002 by CRC Press LLC

3. If errors are large, but the error convergence is satisfactory, the proportional gain is gradually decreased to the appropriate value for small-error conditions.
4. The values of derivative and integral gains are changed simultaneously with the changes in proportional gain so that the stability of the whole system is preserved. The actual values are derived from stability analysis of the closed-loop system. Importantly, the readily available stability conditions for a fixed-gain controller were reused to determine the conditions on parameters of nonlinear fuzzy tuners that are sufficient for overall system stability.

In spite of their simplicity, the rules resulted in significant improvements compared to those of the fixed-gain model-based controller. Simulation experiments on a real-scale six-DOF industrial robot have shown that the resulting variable-gain controllers in many respects outperform constant-gain schemes. The most obvious advantage was the improvement in accuracy demonstrated in both positioning and trajectory tracking tasks. An important feature is that the accuracy improvement was not accompanied by degradation in other performance characteristics, such as energy consumption and maximum developed torques. The second considerable aspect is the possibility of reducing the computational complexity of the nominal robot model (by employing approximate robot models) without the significant loss in control quality that was notable with fixed-gain control. Although the issues related to sensitivity to parameter variations were not explicitly investigated, an improved robustness of the variable-gain controller is implied by the results obtained from experiments with approximate robot models.

24.5 Neuro-Fuzzy Approach in Robotics

Although fuzzy logic can directly and easily encode expert knowledge using rules with linguistic labels, it often takes a lot of time to design and tune the membership functions that quantitatively define these linguistic labels. Wrong membership functions can lead to poor controller performance and possible instability. An execellent solution is to apply learning techniques by neural networks that can be used to design membership functions automatically, thus reducing development time and cost while improving performance. These combined neuro-fuzzy networks can learn faster than neural networks. Also, they provide a connectionist architecture that is easy for very large scale integrated (VLSI) implementation of the functions of a traditional fuzzy logic controller with distributed learning abilities.

The most proposed neuro-fuzzy networks are, in fact Takagi-Sugeno controllers,[49] where the consequent parts of linguistic rules are constant values. Figure 24.20 shows the commonly used connectionist fuzzy system. The system has a total of five layers. Nodes at layer one are input nodes (linguistic nodes) which represent input linguistic variables, Nodes at layer two act as membership functions (in Figure 24.20 they are the Gaussain functions) to represent the terms of the respective linguistic variable. Each node at layer three is a rule node that represents one fuzzy rule. Thus, all layer-three nodes form a fuzzy rule base. Layer five is the output layer. Links at layers three and four function as a connectionist inference engine.

There are many different supervising learning methods for neuro-fuzzy networks.[27,40] Many learning methods are application of the backpropagation method to neuro-fuzzy systems. In addition to gradient-descent techniques, reinforcement learning[7] and some hybrid learning techniques[21] are proposed. One of the most important is ANFIS (adaptive-network-based fuzzy inference system).[21] The learning rule is a hybrid method that combines the gradient descent and the least square estimate to identify the parameters of ANFIS. Usually, neuro-fuzzy networks are trained by applying hybrid techniques where the consequent parts of the rules are adapted with a supervised method and the parameters of the antecedent parts are updated with an unsupervised technique (vector quantization). The idea comes from the field of radial basis functions neural networks.

The one of most important applications of fuzzy-neural networks in robotics is in the field of mobile robotics.[22] A mobile robot is a nonlinear plant that is difficult to model. The state variables

© 2002 by CRC Press LLC

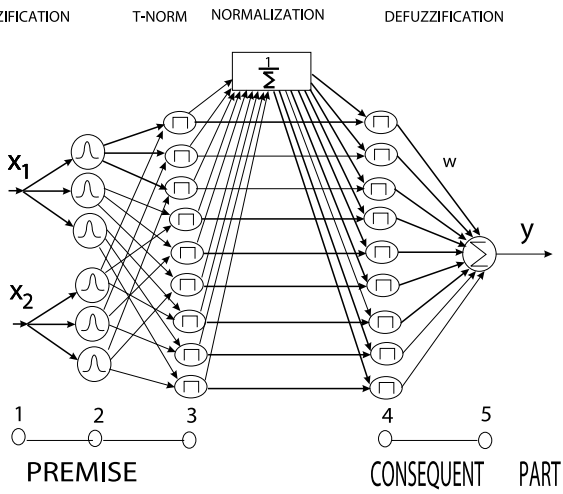


FIGURE 24.20 The structure of fuzzy-neural network.

of a mobile robot are easy to visualize: they have an intuitive relation to the robot's behavior. Therefore, the linguistic if–then rules could be defined in an intuitive way. The problem occurs when a robot has many sensors and actuators. The complexity of the controller increases and the construction of the rule base is more complicated, especially if complex behavior is required. Hence, fuzzy-neural networks are specially applied for the complex task of mobile robot navigation and obstacle avoidance in real time. The network's input data are based on direct or indirect data from many laser, infra-red, and ultrasonic sensors and some other robot velocity sensors that measure the robot's distance from obstacles in the environment, the heading angle between the robot and the specified target, and the velocity of the robot. The network's output values are the control signals for the robot's wheels to determine the appropriate direction of the motion and velocity of the robot. Learning appropriate behavior in the training process defined by the fuzzy-neural controller can be accomplished by the supervisor or by using reinforcement learning for unsupervised learning. In this way, by learning the designer can extract fuzzy rules. Experimental results[55] show that proposed neuro-fuzzy systems can improve navigation performance in complex and unknown environments. This architecture is suitable for robot navigation by multisensor fusion and integration.

Fuzzy-neural networks can be efficiently applied to learning dynamic control and position/force control.[25] It is especially effective in the case when control is applied to an unknown environment. The input data for this type of problem are the appropriate position, velocity, or force errors, while the output of fuzzy-neural network is the control signal.

24.6 Genetic Approach in Robotics

Genetic algorithms (GA) are the global search algorithms for solving optimization problems based on the mechanism of natural selection and natural genetics. It is not a gradient search technique, because the algorithms combine survival of the fittest among string structures (binary or nonbinary type) with a structured yet randomized infomation exchange. Furthermore, GA is not considered a mathematically guided algorithm. It is merely a stochastic, discrete event and a nonlinear process that gives the optima containing the best elements of previous generations. GA is inspired by the biological process in which stronger individuals are likely be the winners in a competing environment. It presumes that the potential solution of a problem is an individual and can be represented by a set of parameters. These parameters are regarded as the genes of a chromosome and can be structured by a string of values in binary or nonbinary form. The fitness value is used to reflect the degree of "goodness" of the chromosome for

© 2002 by CRC Press LLC

solving the problem, and this value is closely related to its objective value. Through genetic evolution, a fitter chromosome tends to yield good-quality offspring, which means a better solution to the optimization problem. In each cycle of genetic operation, an evolving process, a subsequent generation is created from the chromosomes in the current population. This process is achieved through a specific selection routine. The genes of the parents are mixed and recombined to produce offspring. The cycle of evolution is repeated until a desired termination criterion is reached. This criterion can also be set by the number of evolution cycles, the amount of variation of individuals between different generations, or a predefined value of fitness.

To facilitate the GA evolution cycle, two fundamental operators, *crossover* and *mutation,* are required. The crossover process is a reform operation for the survival candidates and is performed by exchanging pieces of string using information from old strings. The pieces are crossed in pairs of strings selected randomly. However, mutation is applied to each offsping individually with random alteration of each bit with a small probability with a typical value of less than 0.1. The choice of crossover and mutation probability can be a complex, nonlinear otimization problem.

The general structure of GA is shown by the following algorithm:

Genetic Algorithm

```
{
*** initial time
t:=0;
*** initialize a random population of individuals
initpopulation P(t);
*** evaluate fitness of all individuals in population
evaluate P(t);
*** test for termination criterion
while not done do
   *** increase the time counter
   t:=t+1;
   *** select a sub-population of offspring
   P':=selectparents P(t);
   *** recombine the genes of selected parents
   recombine P'(t);
   *** mutation of each offspring
   mutate P' (t);
   *** evaluate the new fitness
   evaluate P' (t);
   *** select the survivors from actual fitness
   P:=survive P,P' (t);
od
}
```

GA can be efficiently applied in the various research areas of mobile, industrial, and locomotion robotics. The one of dominant application of GA is the kinematic domain for trajectory optimization and navigation in mobile robotics. Michalewicz[65] adopted the order-based coding in his evolutionary navigator for mobile robot. A chromosome in an evolutionary navigator (EN) is an ordered list of path nodes. Each of path nodes, apart from the pointer to the next node, consists of x and y coordinates of an intermediate knot point along the path, and a Boolean variable b indicating whether the given node is feasible or not. EN unifies off-line and on-line planning with a simple map of high-fidelity and efficient planning algorithms. The off-line planner searches for the optimal global path from the start to the desired destination, whereas the on-line planner is responsible for handling possible collisions of previous objects by replacing a part of the original global path with the optimal subtour.

© 2002 by CRC Press LLC

An interesting approach for GA optimization in robotics is tuning control parameters for some specific robot applications. For many robot controllers there are currently no systematic approaches to choose controller parameters to obtain desired performance. Controller parameters are usually determined by trial-and-error through simulations and experimental tests. In such cases, the paradigm of GA appears to offer an effective way for automatically and efficiently searching for a set of controller parameters for better performance. The effectiveness of this approach is demonstrated by applying a simple and efficient decimal GA optimization procedure for tuning and optimizing robust controllers for position/force control and control of flexible link robots.[41] The robust controller is developed based on stability theory and uses special fitness functions. It is a special GA algorithm with decimal real-number type representation (instead of binary type). The specially used fitness functions are integral time-multiplied absolute value of errors (ITAE) and the normally used integral of squared errors (ISE).

For locomotion robots, GA can be efficiently applied to hierarchical trajectory generation of the natural motion of bipeds using energy optimization.[24,35] The hierarchical trajectory generation method consists of two layers, one is the GA that minimizes the total energy of all actuators, and the other is the evolutionary programming layer that optimizes interpolated configuration of biped locomotion robots. The second example is application of GA to PD local gain tuning and determination of nominal trajectory for dynamic biped walking. Designs to achieve different goals, such as being able to walk on an inclined surface, walk at a high speed, or walk with specified step size, have evolved with the use of GA.

GA are particularly efficient as hybrid techniques with other intelligent soft-computing methods. Together with neural networks, GA can be efficiently applied to determine optimal weighting factors, the topology of networks (number of neurons, number of layers, types of activation functions) and parameters of learning rules. Together with fuzzy rules, GA can be efficently applied to optimization of membership functions, the number of rules, and the parameters of consequent part of rules. On the other hand, fuzzy logic and neural networks can be evaluation functions for GA in the case of complex optimization problems.

In fuzzy-genetic algorithms, it is necessary to solve some problems connected with transformation between the domain of fuzzy knowledge and the GA-coded domain together with using initial expert knowledge for better further searching. In one special example (hierachical fuzzy controller for control of flexible link robots),[3] GA performs optimization of two fuzzy systems: the fuzzy extractor of features at the higher control layer and the fuzzy controller at the lower control layer. For this problem there is very interesting hardware solution where the fuzzy controller works on a DSP board in direct connection with a GA that is executed on a Pentium 133 MHz board. Another typical application of the GA–fuzzy approach in robotics is tuning of local fuzzy-PID controller gains.[2]

GA is applied with the connectionist approach to control visually guided swing motions of a two-armed bipedal robot.[26] The goal is that the neural network learns from the GA swing motion based on visual informatiion from the virtual environment. The goal of visual swing motion is increasing swing amplitude by changing the gravity center in direction of the swing radius acquired through the process of environment recognition using cameras. The inputs in the network are optical signals from cameras, while the outputs are knee joint angles. GA optimizes the three sets of weighting factors of the proposed multilayer perceptron (Figure 24.21). The initial population is 200. The GA–connectionist approach includes determining the weighting factor for recurrent networks for generation of stable biped motion.[47]

24.7 Conclusion

The challenge to future intelligent control system researchers and designers in robotics is to take advantage of the desirable properties of different composite soft-computing control paradigms. It is important to combine the experience and dependability of classic and traditional adaptive control

© 2002 by CRC Press LLC

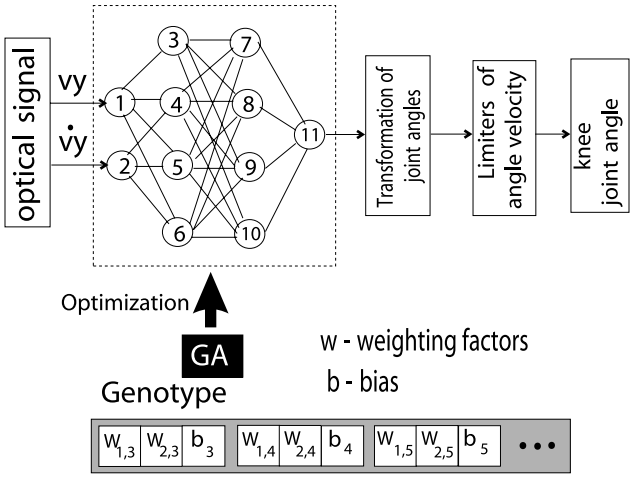


FIGURE 24.21 GA–connectionist approach for robot swing motion optimization.

methods with the potential and promise of soft-computing techniques. One promising idea is the integration of various artificial intelligence paradigms (expert systems, connectionist systems, fuzzy logic, evolutionary algorithms) for the purpose of robot control. The goal of the hybrid approach for robot control is to overcome the weaknesses of each individual intelligent technique by combining it with another complementary intelligent paradigm. This approach is the basis for the development of a generation of intelligent, highly adaptive robotic systems.

References

1. A. Guez and J. Selinsky. Neurocontroller design via supervised and unsupervised leaning. *Journal of Intelligent and Robotic Systems,* 2(2–3):307–335, 1989.
2. A. Homaifar, M. Bikdash, and V. Gopalan. Design using genetic algorithms of hierarchical hybrid fuzzy-PID controllers of two-link robotic arms. *Journal of Robotic Systems,* 14(5):449–463, June 1997.
3. M.-R. Akbarzadeh and M. Jamshidi. Evolutionary fuzzy control of a flexible-link. *Intelligent Automation and Soft Computing,* 3(1):77–88, 1997.
4. J. F. Baldwin and B.W. Pitsworth. Axiomatic approach to implication of approximate reasoning with fuzzy logic. *Fuzzy Sets and Systems,* 3:193–219, 1980.
5. B. Horne, M. Jamshidi, and N. Vadiee. Neural networks in robotics: A survey. *Journal of Intelligent and Robotic Systems,* 3:51–66, 1990.
6. B. Widrow. Generalization and information storage in networks of Adaline neurons. In *Self-Organizing Systems*, pp. 435–461, Spartan Books, New York, 1962.
7. C.-T. Lin and C.S.G. Lee. Reinforcement structure/parameter learning for neural network-based fuzzy logic control system. *IEEE Transactions on Fuzzy Systems,* 2(1):46–63, 1994.
8. C.W. de Silva and A.G.J. MacFarlane. *Knowledge-Based Control with Application to Robots.* Springer, Berlin, 1989.
9. D.E. Rumelhart and J.L. McClelland. *Parallel Distributed Processing* (*PDP*)*: Exploration in the Microstructure of Cognition,* Vol. 1–2. MIT Press, Cambridge, 1986.
10. D.F. Bassi and G.A. Bekey. Decomposition of neural networks models of robot dynamics: A feasibility study. In W. Webster, Ed., *Simulation and AI*, pp. 8–13. The Society for Computer Simulation International, 1989.
11. D. Katić and M. Vukobratović. Connectionist approaches to the control of manipulation robots at the executive hierarchical level: An overview. *Journal of Intelligent and Robotic Systems,* 10:1–36, 1994.

© 2002 by CRC Press LLC

12. D. Katić and M. Vukobratović. Highly efficient robot dynamics learning by decomposed connectionist feedforward control structure. *IEEE Transactions on Systems, Man and Cybernetics*, 25(1):145–158, 1995.
13. D. Katić and M. Vukobratović. Robot compliance algorithm based on neural network classification and learning of robot-environment dynamic models. In *Proceedings of the IEEE International Conference on Robotics and Automation,* pp. 2632–2637, Albuquerque, NM, April 1997.
14. D. Dubois and H. Prade. Fuzzy logics and the generalized modus ponens revisited. *Cybern. Syst.,* 15:3–4, 1984.
15. S. Fukami, M. Mizumoto, and K. Tanaka. Some considerations of fuzzy conditional inference. *Fuzzy Sets and Systems,* 4:243–273, 1980.
16. G.A. Bekey and K.Y. Goldberg. *Neural Networks in Robotics.* Kluwer, Norwell, MA, 1993.
17. G.A. Carpenter and S. Grossberg. The art of adaptive pattern recognition by a self-organizing neural network. *Computer*, pp. 77–88, March 1988.
18. H. Asada and S. Liu. Transfer of human skills to neural network robot controllers. In *Proceedings of the IEEE International Conference on Robotics and Automation,* pp. 2442–2448, Sacramento, April 1991.
19. S. Haykin. *Neural Networks — A Comprehensive Foundation,* Macmillan, Englewood Cliffs, 1994.
20. J. Albus. A new approach to manipulator control: The cerebellar model articulation controller. *Journal of Dynamic Systems, Measurement and Control,* 97:220–227, September 1975.
21. J.-S.R. Jang. ANFIS, adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man and Cybernetics,* 23(3):665–685, 1992.
22. J. Godjevac. *Neuro-Fuzzy Controllers*. Presses Polytechniques et Universitaires Romandes, Lausanne, 1997.
23. J. Hopfield. Neural networks and physical systems with emergent collection computational abilities. In *Proceedings of the National Academy of Science,* pp. 2554–2558, April 1988.
24. J. Xiao, Z. Michalewicz, L. Zhang, and K. Trojanowski. Adaptive evolutionary planner/navigator for mobile robots. *IEEE Transactions on Evolutionary Computation,* 1(1):18–28, April 1997.
25. K. Kiguchi and T. Fukuda. Robot manipulator contact force control application of fuzzy-neural network. In *Proceedings of the IEEE International Conference on Robotics and Automation,* pp. 875–880, Nagoya, May 1995.
26. K. Nagasaka, A. Konno, M. Inaba, and H. Inoue. Acquisition of visually guided swing motion based on genetic algorithms and neural networks in two-armed bipedal robot. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation,* pp. 2944–2949, Albuquerque, NW, April 1997.
27. L.-X. Wang. *Adaptive Fuzzy Systems and Control: Design and Stability Analysis.* Prentice Hall, 1994.
28. C.C. Lee. Fuzzy logic in control systems: Fuzzy logic controller. *IEEE Transactions on Systems, Man and Cybernetics,* 20(2):404–435, April 1990.
29. C.M. Lim and T. Hiyama. Application of fuzzy logic control to a manipulator. *IEEE Transactions on Robotics and Automation,* 7(5):688–691, October 1991.
30. E.H. Mamdani. Application of fuzzy algorithms for control of simple dynamic plant. *Proc. IEEE Control and Science,* 121(12):1585–1588, December 1974.
31. N.J. Mandič, E.M. Scharf, and E.H. Mamdani. Practical application of a heuristic fuzzy rule-based controller to the dynamic control of a robot arm. *IEEE Proceedings on Control Theory and Applications,* 132(4):190–203, July 1985.
32. M. Mizumoto. Fuzzy controls under various approximate reasoning methods. In *Proceedings of the 2nd IFSA Congress,* pp. 143–146, Tokyo, 1987.
33. M. Kuperstain. Adaptive visual-motor coordination in multijoint robots using parallel architecture. In *Proceedings of the IEEE International Conference on Robotics and Automation,* pp. 1595–1601, Raleigh, March 1987.
34. M. Rucci and P. Dario. Autonomous learning of tactile-motor coordination in robotics. In *Proceedings of the IEEE International Conference on Robotics and Automation,* pp. 3230–3236, San Diego, May 1994.

© 2002 by CRC Press LLC

35. M. Zhao, N. Ansari, and E. Hou. Mobile manipulator path planning by a genetic algorithm. In *Proceedings of the 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems,* pp. 681–688, Raleigh, NC, July 1992.
36. D. Popović and R.S. Shekhawat. A fuzzy expert tuner for robot controller. In *Proceedings of the 1991 IFAC/IFIP/IMACS Symposium on Robot Control,* pp. 229–233, Vienna, 1991.
37. T.J. Procyk and E.H. Mamdani. A linguistic self-organizing process controller. *Automatica*, 15(1):15–30, January 1979.
38. J-N. Hwang and S-Y. Kung. Neural network architectures for robotic applications. *IEEE Transactions on Robotics and Automation,* 5(5):641–657, October 1989.
39. S.H. Huang and H.-C. Zhang. Artificial neural networks in manufacturing: Concepts, applications, and perspectives. *IEEE Transactions on Components, Packaging, and Manufacturing Technology — Part A*, 17(2):212–228, June 1994.
40. S. Horikawa, T. Furuhashi, and Y. Uchikawa. On identification of structures in premises of a fuzzy model using a fuzzy neural network. In *Proceedings of the 2nd IEEE International Conference on Fuzzy Systems,* pp. 661–666, 1993.
41. S.S. Ge, T.H. Lee, and G. Zhu. Genetic algorithm tuning of Lyapunov-based controllers: An application to a single-link flexible robot system. *IEEE Transactions on Industrial Electronics,* 43(5):567–574, October 1996.
42. M. S. Stachowicz and M.E. Kochanska. Fuzzy modeling of the process. In *Proceedings of the 2nd IFSA Congress,* pp. 86–89, Tokyo, 1987.
43. S. Tzafestas. Fuzzy and neural approaches to robot control. In *Proceedings of the 1st ECPD International Conference on Advanced Robotics and Intelligent Automation,* pages 34–55, Athens, September 1995.
44. M. Sugeno and M. Nishida. Fuzzy control of a model car. *Fuzzy Sets and Systems,* 16:103–113, 1985.
45. T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, January 1985.
46. R. Tanscheit and E.M. Scharf. Experiments with the use of a rule-based self-organizing controller for robotics applications. *Fuzzy Sets and Systems,* 26(1):195–214, 1988.
47. T. Fukuda, Y. Komata, and T. Arakawa. Stabilization control of biped locomotion robot based learning with GAs having self-adaptive mutation and recurrent neural network. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation,* pp. 217–222, Albuquerque, NM, April 1997.
48. T. Ozaki, T. Suzuki, T. Furuhashi, S. Okuma, and Y. Uchikawa. Trajectory control of robotic manipulators using neural networks. *IEEE Transactions on Industrial Electronics,* 38(3):195–202, June 1991.
49. T. Takagi and M. Sugeno. Derivation of fuzzy control rules from human operator's control actions. In *Proceedings of the IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis,* pp. 55–60, 1983.
50. S. Tzafestas and N.P. Papanikolopoulos. Incremental fuzzy expert PID control. *IEEE Transactions on Industrial Electronics,* 37(5):365–371, October 1990.
51. V. Gullapalli. A stochastic reinforcement learning algorithm for learning real-valued functions. *Neural Networks,* 3:671–692, 1990.
52. M. Vukobratović and B. Karan. Experiments with fuzzy logic robot control with model-based dynamic compensation. In *Proceedings of the 25th International Symposium on Industrial Robots,* pp. 215–222, Hanover, 1994.
53. M. Vukobratović and B. Karan. Experiments with fuzzy logic robot control with model-based dynamic compensation in nonadaptive decentralized control scheme. *International Journal on Robotics and Automation,* 11(3):118–131, 1996.
54. M. Vukobratović, D. Stokić, and N. Kirćanski. *Non-Adaptive and Adaptive Control of Manipulation Robots.* Springer, Berlin, 1985.
55. W. Li, C. Ma, and F.M. Wahl. A neuro-fuzzy system architecture for behaviour-based control of a mobile robot in an unknown environment. *Fuzzy Sets and Systems,* 87(2):133–140, 1997.

© 2002 by CRC Press LLC

56. W.T. Miller. Sensor-based control of manipulation robots using a general learning algorithm. *IEEE Journal on Robotics and Automation,* 3:157–165, April 1987.
57. C.W. Xu and Y.Z. Lu. Fuzzy model identification and self-learning for dynamic systems. In *IEEE Transactions on Systems, Man and Cybernetics,* 17(4):683–689, August 1987.
58. L.A. Zadeh. Fuzzy sets. *Information and Control,* 8:338–353, 1965.
59. L.A. Zadeh. Fuzzy algorithms. *Information and Control,* 12:94–102, 1968.
60. L.A. Zadeh. Quantitative fuzzy semantics. *Information Sciences,* 3:159–176, 1971.
61. L.A. Zadeh. A rationale for fuzzy control. *Journal of Dynamic Systems, Measurement, and Control,* 94:3–4, 1972.
62. L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics*, 3(1):28–44, January 1973.
63. L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences,* 8:199–249, 301–357, 1975.
64. L.A. Zadeh. Fuzzy logic. *IEEE Computer,* 21(4):83–93, April 1988.
65. Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Program.* Springer Verlag, Berlin, 1994.

© 2002 by CRC Press LLC

25

Teleoperation and Telerobotics

25.1 Introduction
25.2 Hand Controllers
Control Handles • Control Input Devices • Universal Force-Reflecting Hand Controller (FRHC)
25.3 FRHC Control System
25.4 ATOP Computer Graphics
25.5 ATOP Control Experiments
25.6 Anthropomorphic Telerobotics
25.7 New Trends in Applications

Antal K. Bejczy
California Institute of Technology

25.1 Introduction

In a general sense, teleoperator devices enable human operators to remotely perform mechanical actions usually performed by the human arm and hand. Thus, teleoperators or the activities of teleoperation extend the manipulative capabilities of the human arm and hand to remote, physically hostile, or dangerous environments. In this sense, teleoperation conquers space barriers by performing manipulative mechanical actions at remote sites, as telecommunication conquers space barriers by transmitting information to distant places.

Teleoperator systems were developed in the mid-1940s to create capabilities for handling highly radioactive material. Such systems allowed a human operator to handle radioactive material in its radioactive environment from a workroom separated by a 1-m thick, radiation-absorbing concrete wall. The operator could observe the task scene through radiation resistant viewing ports in the wall. The development of teleoperators for the nuclear industry culminated in the introduction of bilateral force-reflecting master–slave manipulator systems. In these very successful systems, the slave arm at the remote site is mechanically or electrically coupled to the geometrically identical or similar master arm handled by the operator and follows the motion of the master arm. The coupling between the master and slave arms is two-way; inertia or work forces exerted on the slave arm can back-drive the master arm, enabling the operator to feel the forces that act on the slave arm. Force information available to the operator is an essential requirement for dexterous control of remote manipulators, since general purpose manipulation consists of a series of well-controlled contacts between handling device and objects and also implies the transfer of forces and torques from the handling device to objects.

Teleoperators in this age of modern information technology are classified as specialized robots, called telerobots, performing manipulative mechanical work remotely where humans cannot or do not want to go. Teleoperator robots serve to extend, through mechanical, sensing, and computational techniques, the human manipulative, perceptive, and cognitive abilities into an environment that is hostile to or remote from the human operator. Teleoperator robots, or telerobots, typically perform

© 2002 by CRC Press LLC

nonrepetitive or singular servicing, maintenance, or repair work under a variety of structured and unstructured environmental conditions. Telerobot control is characterized by the direct involvement of the human operator in the control since, by definition of task requirements, teleoperator systems extend human manipulative, perceptual, and cognitive skills to remote places.

Continuous human operator control in teleoperation has both advantages and disadvantages. The main advantage is that overall task control can rely on human perception, judgment, decision, dexterity, and training. The main disadvantage is that the human operator must cope with a sense of remoteness, be alert to and integrate many information and control variables, and coordinate the control of one or two mechanical arms each having many (typically six) degrees of freedom (DOFs) — and handling all these tasks with limited resources. Furthermore, in cases like space and deep sea applications, communication time delay interferes with continuous human operator control.

Modern development trends in teleoperator technology are aimed at amplifying the advantages and alleviating the disadvantages of the human element in teleoperator control. This is being done through the development and the use of advanced sensing and graphics displays, intelligent computer controls, and new computer-based human–machine interface devices and techniques in the information and control channels. The use of model and sensor data-driven automation in teleoperation offers significant new possibilities to enhance overall task performance by providing efficient means for task level controls and displays.

Later in this section, we will focus on mechanical, control, and display topics that are specific to the human–machine system aspect of teleoperation and telerobotics: hand controllers, task level manual and automatic controls, and overlaid, calibrated graphics displays aimed to overcome telecommunication time delay problems in teleoperation. Experimental results will be briefly summarized. The section will conclude with specific issues in anthropomorphic telerobotics and a brief outline of emerging application areas.

25.2 Hand Controllers

The human arm and hand are powerful mechanical tools and delicate sensory organs through which information is received and transmitted to and from the outside world. Therefore, the human arm–hand system (from now on simply called the hand) is a key communication medium in teleoperator control. Complex position, rate, or force commands can be formulated to control a remote robot arm–hand system in all workspace directions with hand actions. The human hand also can receive contact force, torque, and touch information from the remote robot hand or end effector. The human fingers provide capabilities to convey new commands to a remote robot system from a suitable hand controller. Hand controller technology is, therefore, an important component in the development of advanced teleoperators. Its importance is particularly stressed when one considers the computer control that connects the hand controller to a remote robot arm system.

We will review teleoperator system design issues and performance capabilities from the viewpoint of the operator's hand and hand controllers through which the operator exercises manual control communication with remote manipulators. Through a hand controller, the operator can write commands to and also read information from a remote manipulator in real time. It is conceptually appropriate and illuminating to view the operator's manual control actions as a control language and, subsequently, to consider the hand controller as a translator of that control language to machine-understandable control actions.

A particular property of manual control as compared to computer keyboard control in teleoperation is that the operator's hand motion, as translated by the hand controller, directly describes a full trajectory to the remote robot arm in the time continuum. In the case of a position control device, the operator's manual motion contains direct position, velocity, acceleration, and even higher order derivative motion command information. In the case of a rate input device, the position information is indirect since it is the integral of the commanded rate, but velocity, acceleration, and even higher order derivative motion command information is direct in the time continuum. All

© 2002 by CRC Press LLC

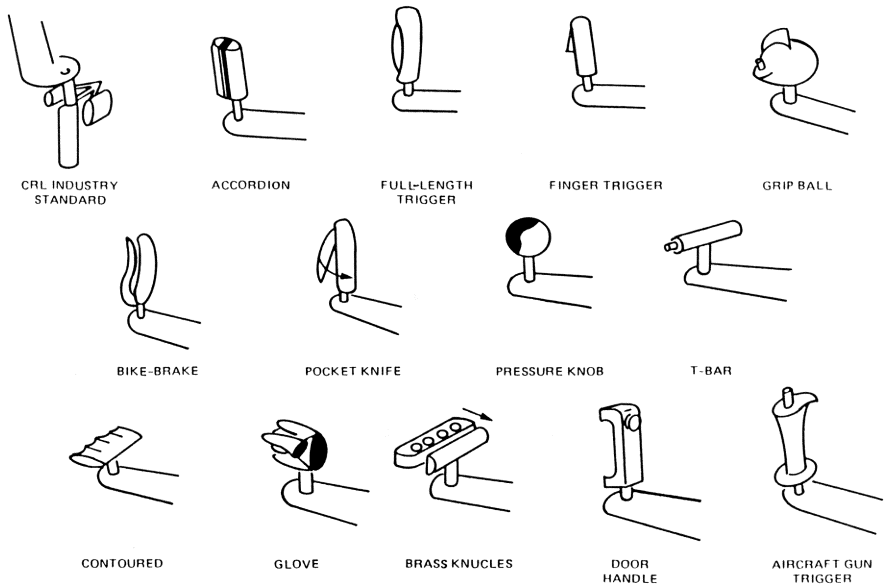


FIGURE 25.1 Basic grip and trigger concepts.

this direct operator hand motion relation to the remote robot arm's motion behavior in real time through the hand controller is in sharp contrast to computer keyboard commands which, by their very nature, are symbolic and abstract, and require the specification of some set of parameters within the context of a desired motion.

First, a brief survey of teleoperator hand controller technology will discuss both hand grips and complete motion control input devices, as well as the related control modes or strategies. Then a specific example, a general purpose force-reflecting position hand controller will be briefly discussed, implemented, and evaluated at the Jet Propulsion Laboratory (JPL), including a novel switch module attached to the hand grip.

25.2.1 Control Handles

The control handles are hand grips through which the operator's hand is physically connected to the complete hand controller device. Fourteen basic handle concepts (Figure 25.1) have been considered and evaluated by Brooks and Bejczy[1]:

1. Nuclear industry standard handle — a squeeze-grasp gripper control device that exactly simulates the slave end effector squeeze-type grasp motion.
2. Hydraulic accordion handle — a finger-heel grasp device using a linear motion trigger driven by hydraulic pressure.
3. Full-length trigger — a finger-heel type, linear motion gripper control device driven by a mechanism.
4. Finger trigger — a linear or pivoted gripper control device that only requires one or two fingers for grasp actuation.
5. Grip ball — a ball-shaped handle with a vane-like protuberance that prevents slippage of the ball when sandwiched between two fingers.
6. Bike brake — a finger-heel-type grasp control device in which the trigger mechanism is pivoted at the base of the handle.
7. Pocket knife — similar to the bike brake, but the trigger mechanism is pivoted at the top of the handle.

© 2002 by CRC Press LLC

8. Pressure knob — a uni-body ball-shaped handle consisting of a rigid main body and a semi-rigid rubber balloon gripper control driven by hydraulic pressure.
9. T-bar — a one-piece T-shaped handle with a thumb button for gripper control.
10. Contoured bar — a one-piece contoured T-type handle with gripper control surface located on the underside.
11. Glove — a mechanical device that encases the operator's hand.
12. Brass knuckles — a two-piece T-type handle, the operator's fingers slip into recesses or holes in the gripper control.
13. Door handle — a C-shaped handle with a thumb button gripper control.
14. Aircraft gun trigger — a vertical implementation handle using a lateral grasp for trigger control combined with wrap-around grasp for firm spatial control.

The 14 handle concepts have been evaluated based on 10 selection criteria and grouped into four major categories:

1. *Engineering development:* This category considers the handle's developmental requirements in terms of (a) design simplicity, (b) difficulty of implementation, (c) extent to which a technological base has been established, and (d) cost.
2. *Controllability:* This category considers the operator's ability to control the motion of the slave manipulator through the handle. Two major categories were used as selection criteria: (a) stimulus-response compatibility, and (b) cross-coupling between the desired arm motion/forces and the grasp action. Stimulus-response compatibility considers the extent to which the handle design approaches the stimulus-response compatibility of the industry standard. This category only considers the desirability of a stimulus-response compatibility from a motion-in/motion-out standpoint; it does not take into account fatigue (fatigue is considered in category 4). Cross-coupling, considers the extent of cross-coupling between the motion or force applied to the arm and the desired motion or force of the gripper.
3. *Human-handle interaction:* This category considers the effects of the interface and the interaction between the human and the handle. Four major categories were used as selection criteria: (a) secondary function control, (b) force-feedback ratio, (c) kinesthetic feedback, and (d) accidental activation potential. Secondary function control considers the appropriateness of secondary switch placement from the standpoint of the operator's ability to activate a given function. Force feedback considers the extent to which the remote forces must be scaled for a given handle configuration. The third category rates the degree of kinesthetic feedback, particularly with regard to the range of trigger motion with respect to an assumed 3-in. open/close motion of the end effector. The fourth category deals with the potential for accidental switch activation for a given design. The lower the rating, the more potential exists for accidental activation.
4. *Human limitations:* This category considers the limitations of the operator as a function of each design (assuming a normalized operator). Two areas were of concern in the handle selection: (a) endurance capacity, and (b) operator accommodation. The first category deals with the relative duration with respect to the other handle configurations during which an operator can use a given design without becoming fatigued or stressed. The second category considers the extent to which a given design can accommodate a wide range of operators.

Details of subjective ratings for each of the 14 handle concepts based on the four categories of criteria can be found in Brooks and Bejczy.[1] The value analysis is summarized in Table 25.l. As shown in this table, the finger-trigger design stands out as the most promising handle candidate. From a simple analysis, it also appears that the most viable technique for controlling trigger DOFs while simultaneously controlling six spatial DOFs through handle holding should obey the following guidelines:

© 2002 by CRC Press LLC

TABLE 25.1 Tradeoff and Value Analysis of Handle Designs

	Engineering Development				Controllability		Human-Handle Interaction				Human Limitations		
	Design Simplicity	Difficulty of Implementation	Technology Base	Cost	Stimulus-Response Compatibility	Cross Coupling	Secondary-Function Control	Force Feedback	Kinesthetic Feedback	Accidental Activation	Endurance Capacity	Operator Accommodation	Total Figure of Merit Σ Value × Score
Value	2	1	5	4	3	5	5	4	4	4	3	2	
Industry Standard	2	2	3	2	3	3	1	3	3	2	1	2	97
Accordion	3	3	1	3	2	1	3	3	3	3	2	2	98
Full-Length Trigger	2	2	3	2	2	1	3	3	3	3	2	2	101
Finger Trigger	3	3	3	3	2	3	3	2	3	3	3	2	117
Grip Ball	3	3	2	2	2	3	1	2	2	1	2	3	85
Bike-Brake	3	3	3	3	2	1	3	3	3	3	2	2	108
Pocket Knife	3	3	3	3	2	1	3	3	3	3	2	2	108
Pressure Nub	3	3	1	3	1	1	1	1	1	1	1	3	60
T-Bar	3	3	3	3	2	3	1	2	2	1	2	3	94
Contoured	2	2	1	2	1	1	3	1	1	2	1	3	67
Glove	1	1	1	1	3	3	1	3	3	2	2	1	81
Brass Knuckle	2	2	3	2	2	1	3	3	3	2	2	3	99
Door Handle	3	3	3	3	2	3	2	2	2	2	2	3	103
Aircraft Gun Trigger	3	3	3	3	2	3	1	2	2	1	2	3	94

Ratings: 1 = lowest; 3 = highest.

- The handle must be held firmly with at least two fingers and the heel of the hand at all times to adequately control the six spatial DOFs.
- At least one of the stronger digits of the hand (i.e., thumb or index finger) must be dedicated to the function of trigger actuation and force feedback; that is, it must be independent of spatial control functions.
- The index finger, having restricted lateral mobility, makes a good candidate for single-function dedication since it cannot move as freely as the thumb from one switch to another.
- The thumb makes a better candidate for multiple switch activation.

25.2.2 Control Input Devices

Twelve hand controllers have been evaluated for manual control of six DOF manipulators in Brooks and Bejczy.[1] Some descriptive details of their designs and their detailed evaluation are included. We will only summarize their basic characteristics.

1. *Switch controls* generally consist of simple spring-centered, three-position (–, off, +) discrete action switches. Each switch is assigned to a particular manipulator joint or to end effector control.
2. *Potentiometer controls* or potentiometers are used for proportional control inputs for either position or rate commands. They can be either force-operated (e.g., spring centered) or displacement-operated. Typically, each pot is assigned to one manipulator joint and to end effector control.
3. *The isotonic joystick controller* is a position-operated fixed-force (isotonic) device used to control two or more DOFs with one hand within a limited control volume. A trackball is a well-known example.

© 2002 by CRC Press LLC

4. *The isometric joystick controller* is a force-operated minimal-displacement (isometric) device used to control two or more DOFs with one hand from a fixed base. Its command output directly corresponds to the forces applied by the operator and drops to zero unless manual force is maintained.
5. *Proportional joystick controller* is a single-handed, two or more DOF device with a limited operational volume in which the displacement is a function of the force applied by the operator (F = kx), and the command output directly corresponds to the displacement of the device.
6. *The hybrid joystick controller* is composed of isotonic, isometric, and proportional elements (that are mutually exclusive for a given DOF), used to control two or more DOFs within a limited control volume with a single hand. It has two basic implementation philosophies: concurrent and sequential. In the concurrent implementation, some DOFs are position-operated and some are force-operated (either isometrically or proportionally). In the sequential implementation, position and force inputs are switched for any DOF. For details of these two implementations, see Brooks and Bejczy.[1]
7. *The replica controller* has the same geometric configuration as the controlled manipulator but built on a different scale. Hence, there is a one-to-one correspondence between replica controller and remote manipulator joint movements without actual one-to-one spatial correspondence between control handle and end effector motion.
8. *The master–slave controller* has the same geometric configuration and physical dimensions as the controlled manipulator. There is a one-to-one correspondence between master and slave arm motion. These and the replica devices can be unilateral (no force feedback) or bilateral (with force feedback) in the implementation.
9. *The anthropomorphic controller* derives the manipulator control signals from the configuration motion of the human arm. It may or may not have a geometric correspondence with the remotely controlled manipulator.
10. *The nongeometric analogic controller* does not have the same geometric configuration as the controlled manipulator, but it maintains joint-to-joint or spatial correspondence between the controller and the remote manipulator.
11. *The universal force-reflecting hand controller* is a six DOF position control device which, through computational transformations, is capable of controlling the end effector motion of any geometrically dissimilar manipulator and can be backdriven by forces sensed at the base of the remote manipulator's end effector (i.e., it provides force feedback to the operator). For more details of this device, see section 25.2.3.
12. *The universal floating-handle controller* is a nongeometric six DOF control device, without joints and linkages, which is used for controlling the slave arm end effector motion in hand-referenced control. It can be either unilateral or bilateral in the control mode. An example of unilateral version is the data glove.

25.2.3 Universal Force-Reflecting Hand Controller (FRHC)

In contrast to the standard force-reflecting master–slave systems, a new form of bilateral, force-reflecting manual control of robot arms has been implemented at JPL. It is used for a dual-arm control setting in a laboratory work cell to carry out performance experiments.

The feasibility and ramifications of generalizing the bilateral force-reflecting control of master–slave manipulators has been under investigation at JPL for more than 10 years. Generalization means that the master arm function is performed by a universal force-reflecting hand controller that is dissimilar to the slave arm both kinematically and dynamically. The hand controller under investigation is a backdrivable six DOF isotonic joystick. It controls a six DOF mechanical arm equipped with a six-dimensional force-torque sensor at the base of the mechanical hand. The hand controller provides position and orientation control for the mechanical hand. Forces and torques

© 2002 by CRC Press LLC

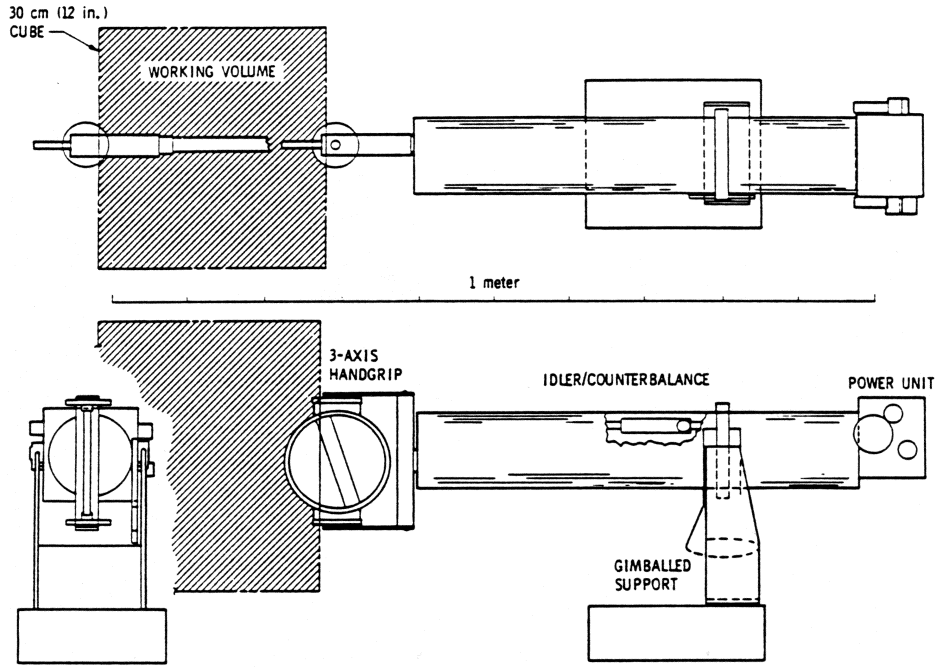


FIGURE 25.2 Overall schematic of six-axis force-reflecting hand controller.

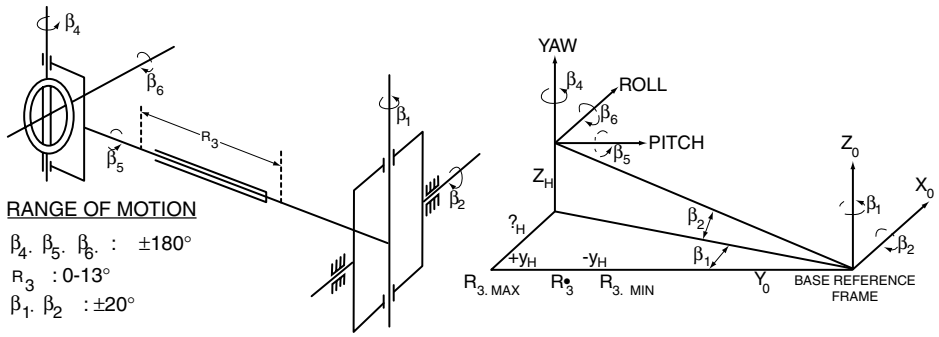


FIGURE 25.3 Hand controller kinematics and command axes.

sensed at the base of the mechanical hand back drive the hand controller so that the operator feels the forces and torques acting at the mechanical hand while he controls the position and orientation of the mechanical hand.

The overall schematic of the six DOF force-reflecting hand controller employed in the study is shown in Figure 25.2. (The mechanism of the hand controller was designed by J.K. Salisbury, Jr., now at MIT, Cambridge, MA.) The kinematics and the command axes of the hand controller are shown in Figure 25.3.

The hand grip is supported by a gimbal with three intersecting axes of rotation (β_4, β_5, β_6). A translation axis (R_3) connects the hand gimbal to the shoulder gimbal which has two more intersecting axes (β_1, β_2). The motors for the three hand gimbal and translation axes are mounted on a stationary drive unit at the end of the hand controller's main tube. This stationary drive unit forms a part of the shoulder gimbal's counterbalance system. The moving part of the counterbalance system is connected to the R_3, translation axis through an idler mechanism that moves at one half

© 2002 by CRC Press LLC

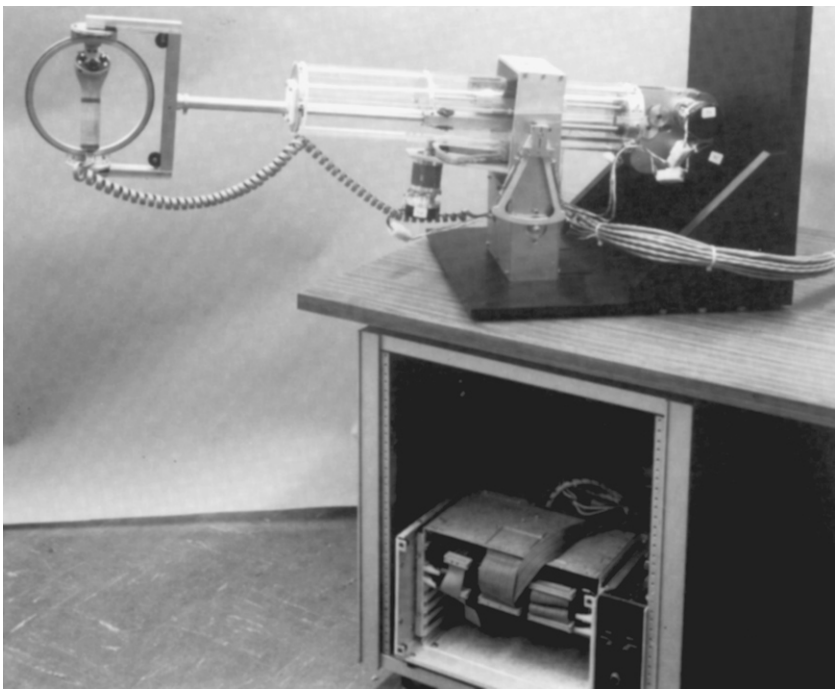


FIGURE 25.4 Universal force-reflecting hand controller with basic computer control system.

the rate of R_3. It serves (1) to maintain the hand controller's center of gravity at a fixed point, and (2) to maintain the tension in the hand gimbal's drive cables as the hand gimbal changes its distance from the stationary drive unit. The actuator motors for the two shoulder joints are mounted to the shoulder gimbal frame and to the base frame of the hand controller, respectively. A self-balance system renders the hand controller neutral against gravity. Thus, the hand controller can be mounted both horizontally or vertically, and the calculation of motor torques to backdrive the hand controller does not require gravity compensation.

In general, the mechanical design of the hand controller provides a dynamically transparent input/output device for the operator. This is accomplished by low backlash, low friction, and low effective inertia at the hand grip. More details of the mechanical design of the hand controller can be found in Bejczy and Salisbury.[2]

The main functions of the hand controller are: (1) to read the position and orientation of the operator's hand, and (2) to apply forces and torques to his hand. It can read the position and orientation of the hand grip within a 30-cm cube in all orientations, and can apply arbitrary force and torque vectors up to 20 N and 1.0 Nm, respectively, at the hand grip.

A computer-based control system establishes the appropriate kinematic and dynamic control relations between the FRHC and the robot arm controlled by the FRHC. Figure 25.4 shows the FRHC and its basic control system. The computer-based control system supports four modes of control. Through an on-screen menu, the operator can designate the control mode for each task space (Cartesian space) axis independently. Each axis can be controlled for position, rate, force-reflecting, and compliant control modes.

Position control mode servos the slave position and orientation to match the master's. Force/torque information from the six-axis sensor in the smart hand generates feedback to the operator of environmental interaction forces via the FRHC. The indexing function allows slave excursions beyond the 1-cubic foot workvolume of the FRHC, and allows the operator to work at any task site from his most comfortable position. This mode is used for local manipulation.

© 2002 by CRC Press LLC

Rate control sets slave endpoint velocity in task space based on the displacement of the FRHC. The master control unit delivers force commands to the FRHC to enforce a software spring by which the operator has a better sensation of command, and provides a zero referenced restoring force. Rate mode is useful for tasks requiring large translations.

Position, force-reflecting, and rate modes exist solely on the master side. The slave receives the same incremental position commands in either case. In contrast, variable compliance control resides at the slave side. It is implemented through a low-pass software filter in the hybrid position force control loop. This permits the operator to control a springy or less stiff robot. Active compliance with damping can be varied by changing the filter parameters in the software menu. Setting the spring parameter to zero in the low-pass filter will reduce it to a pure damper which results a high stiffness in the hybrid position force control loop.

The present FRHC has a simple hand grip equipped with a deadman switch and three function switches. To better utilize the operator's finger input capabilities, an exploratory project evaluated a design concept that would place computer keyboard features attached to the hand grip of the FRHC. To accomplish this, three DATAHAND™[3] switch modules were integrated into the hand grip as shown in Figure 25.5. Each switch module at a finger tip contains five switches as indicated in Figure 25.6. Thus, the three switch modules at the FRHC hand grip can contain 15 function keys that can directly communicate with a computer terminal. This eliminates the need for the operator to move his hand from the FRHC hand grip to a separate keyboard to input messages and commands to the computer. A test and evaluation using a mock-up system and ten test subjects indicated the viability of the finger-tip switch modules as part of a new hand grip unit for the FRHC as a practical step toward a more integrated operator interface device. More on this concept and evaluation can be found in Knight.[4]

25.3 FRHC Control System

An advanced teleoperator (ATOP) dual–arm laboratory breadboard system was set up at JPL using two FRHC units in the control station to experimentally explore the active role of computers in system operation.

The overall ATOP control organization permits a spectrum of operations including full manual, shared manual, automatic, and full automatic (called traded) control, and the control can be operated with variable active compliance referenced to force moment sensor data. More on the overall ATOP control system can be found in Bejczy et al.[5,17] and Bejczy and Szakaly.[6,8] Only the salient features of the ATOP control system are summarized here. The overall control/information data flow diagram (for a single arm) is shown in Figure 25.7. The computing architecture of this original ATOP system is a fully synchronized pipeline, where the local servo loops at the control station and the remote manipulator nodes can operate at a 1000-Hz rate. The end-to-end bilateral (i.e., force-reflecting) control loop can operate at a 200-Hz rate. More on the computational system critical path functions and performance can be found in Bejczy and Szakaly.[9]

The actual data flow depends on the control mode chosen. The different selectable control modes are: freeze mode, neutral mode, current mode, joint mode, and task mode. In the freeze mode, the brakes of joints are locked, the motors are turned off, and some joints are servoed to maintain their last positions. This mode is primarily used when the robot is not needed for a short time and turning it off is not desired. In the neutral mode, all position gains are set to zero, and gravity compensation is active to prevent the robot from falling. In this mode, the user can manually move the robot to any position, and it will stay there. In the current mode, the six motor currents are directly commanded by the data coming in from the communication link. This mode exists for debugging only. In the joint mode, the hand controller axes control individual motors of the robot. In the task mode, the inverse kinematic transformation is performed on the incoming data, and the hand controller controls the end effector tip along the three Cartesian and pitch, yaw, and roll axes. This mode is the most frequently used for task execution or experiments, and is shown in Figure 25.7.

© 2002 by CRC Press LLC

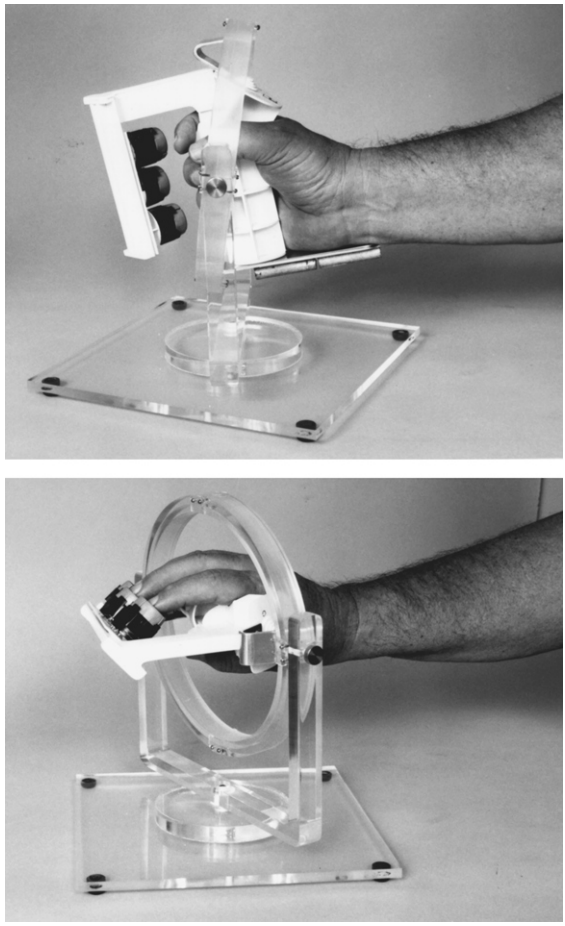


FIGURE 25.5 DATAHAND™ switch modules integrated with FRHC hand grip.

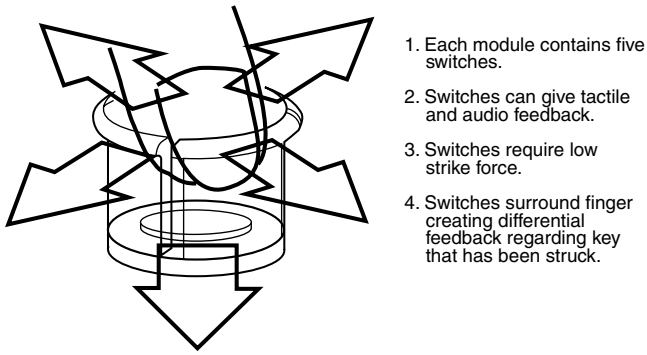


FIGURE 25.6 Five key-equivalent switches at a DATAHAND™ fingertip switch module.

The control system on the remote site is designed to prevent sudden robot motions. The motion commands received are incremental and are added to the current parameter under control. Sudden large motions are also prevented in case of mode changes. This necessitates proper initialization of the inverse kinematics software at the time of the mode transition. This is done by inputting the current Cartesian coordinates from the forward kinematics into the inverse kinematics. The data

© 2002 by CRC Press LLC

flow diagram shown in Figure 25.7 illustrates the organization of several servo loops in the system. The innermost loop is the position control servo at the robot site. This servo uses a PD control algorithm, where the damping is purely a function of the robot joint velocities. The incoming data to this servo is the desired robot trajectory described as a sequence of points at 1 ms intervals. This joint servo is augmented by a gravity compensation routine to prevent the weight of the robot from causing a joint positioning error. Because this is a first order servo, there will be a constant position error that is proportional to the joint velocity.

In the basic Cartesian control mode, the data from the hand controller are added to the previous desired Cartesian position. From this the inverse kinematics generate the desired joint positions. The joint servo moves the robot to this position. The forward kinematics compute the actual Cartesian positions from the actual joint position.. The force-torque sensor data and the actual positions are fed back to the hand controller side to provide force feedback.

This basic mode can be augmented by the addition of compliance control, Cartesian servo, and stiction/friction compensation. Figure 25.8 shows the compliance control and the Cartesian servo augmentations. The two forms of compliance are an integrating type and a spring type. With integrating compliance, the velocity of the robot end effector is proportional to the force felt in the corresponding direction. To eliminate drift, a deadband is used. The zero velocity band does not have to be a zero force; a force offset may be used. Such a force offset is used if, for example, we want to push against the task board at some given force while moving along other axes. Any form of compliance can be selected along any axis independently. In the case of the spring-type compliance, the robot position is proportional to the sensed force. This is similar to a spring centering action. The velocity of the robot motion is limited in both the integrating and spring cases.

As is shown in Figure 25.8, the Cartesian servo acts on task space (X, Y, Z, pitch, yaw, roll) errors directly. These errors are the difference between desired and actual task space values. The actual task space values are computed from the forward kinematic transformation of the actual joint positions. This error is then added to the new desired task space values before the inverse kinematic transformation determines the new joint position commands from the new task space commands.

A trajectory generator algorithm was formulated based on observations of profiles of task space trajectories generated by the operators manually through the FRHC. Based on these observations, we formulated a harmonic motion generator (HMG) with a sinusoidal velocity-position phase function profile as shown in Figure 25.9. The motion is parameterized by the total distance traveled, the maximum velocity, and the distance used for acceleration and deceleration. Both the accelerating and decelerating segments are quarter sine waves connected by a constant velocity segment. This scheme still has a problem: the velocity is zero before the motion starts. This problem is corrected by adding a small constant to the velocity function.

The HMG discussed here is quite different from the typical trajectory generator algorithms employed in robotics which use polynomial position–time functions. The HMG algorithm generates motion as a trigonometric (harmonic) velocity vs. position function. More on performance results generated by HMG, Cartesian servo, and force-torque sensor data filtering in compliance control can be found in Bejczy and Szakaly.[6,10]

25.4 ATOP Computer Graphics

Task visualization is a key problem in teleoperation, since most of the operator's control decisions are based on visual or visually conveyed information. For this reason, computer graphics plays an increasingly important role in advanced teleoperation. This role includes: (1) *planning* actions, (2) *previewing* motions, (3) *predicting* motions in real time under communication time delay, (4) helping operator *training*, (5) *enabling visual perception of nonvisible events* like forces and moments, and (6) serving as a *flexible operator interface* to the computerized control system.

The capability of task planning aided by computer graphics offers flexibility, visual quality, and a quantitative design base to the planning process. The ability to graphically preview motions

© 2002 by CRC Press LLC

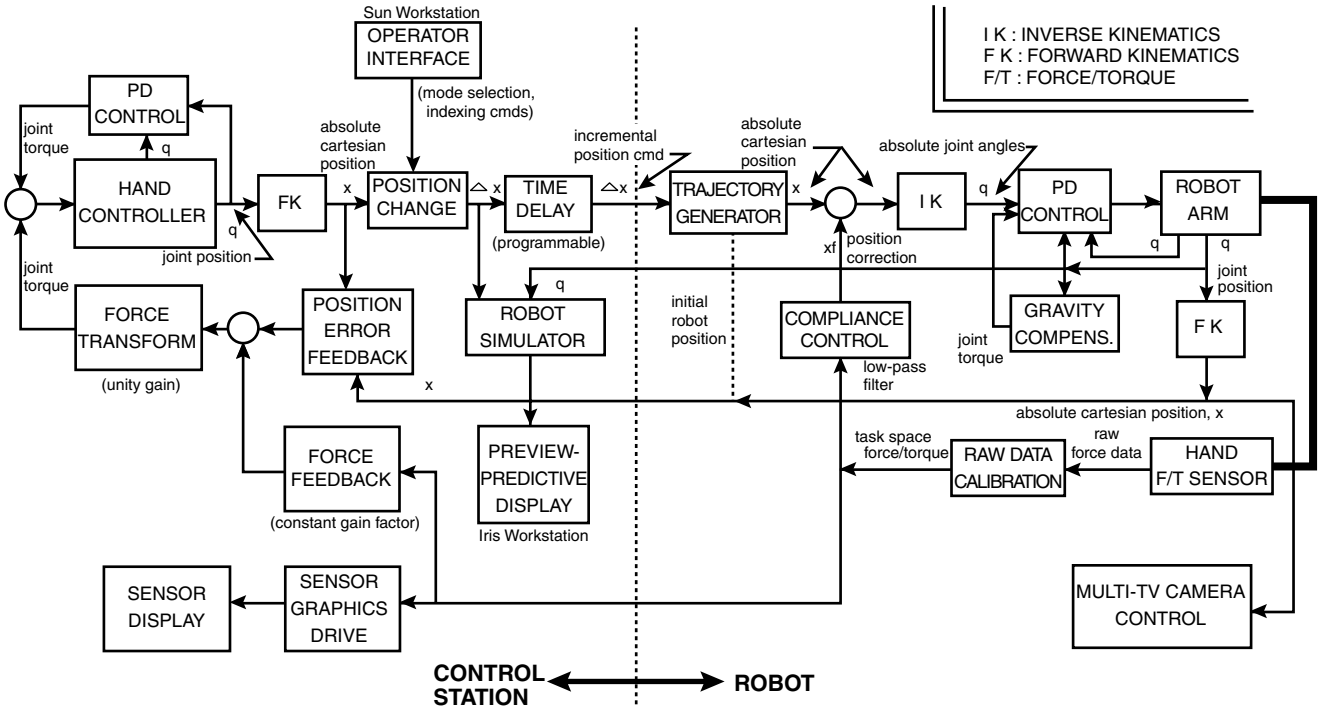


FIGURE 25.7 Control system flow diagram.

© 2002 by CRC Press LLC

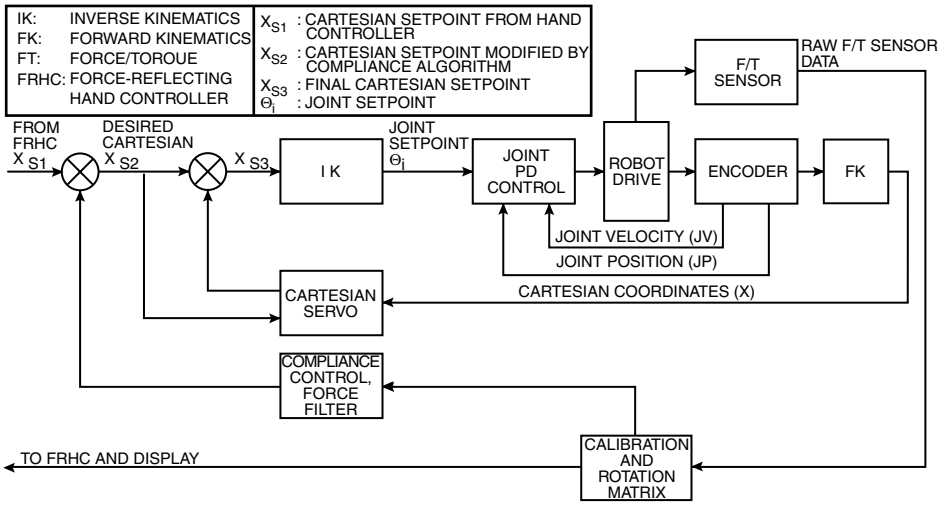


FIGURE 25.8 Control schemes: joint servo, Cartesian servo, and compliance control.

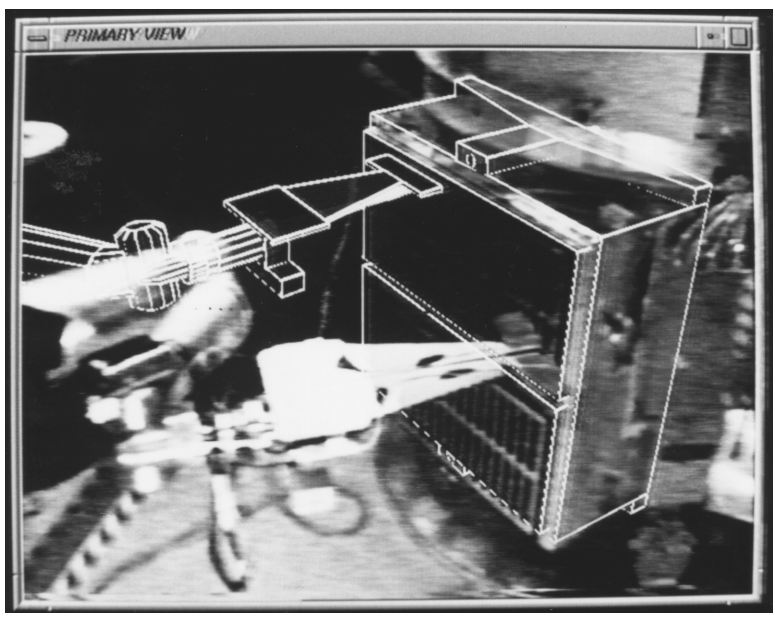


FIGURE 25.9A Predictive/preview display of end point motion.

enhances the quality of teleoperation by reducing trial-and-error strategies in the hardware control and increases the operator's confidence in decision making during task execution. Predicting consequences of motion commands in real time under communication time delay permits longer action segmentations as opposed to the move-and-wait control strategy normally employed when no predictive display is available, increases operation safety, and reduces total operation time. Operator training through a computer graphics display system is a convenient tool for familiarizing the operator with the teleoperated system without turning the hardware system on. Visualization of nonvisible effects (like control forces) enables visual perception of different nonvisual sensor data, and helps manage system redundancy by providing a suitable geometric image of a multidimensional

© 2002 by CRC Press LLC

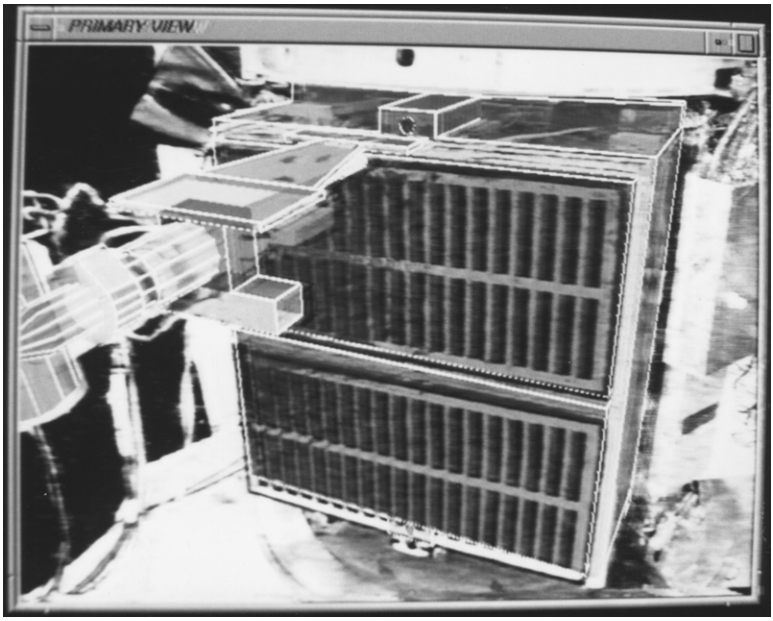


FIGURE 25.9B Status of predicted end point after motion execution from a tv camera view different from the view shown in Figure 25.9a.

system state. Computer graphics, as a flexible operator interface to the control systems, replace complex switchboard and analog display hardware in a control station.

The utility of computer graphics in teleoperation depends on the fidelity of graphics models that represent the teleoperated system, the task, and the task environment. The JPL ATOP effort focused on the development of *high-fidelity calibration of graphics* images into TV images of task scenes. This development has four major ingredients: (1) creation of high-fidelity 3-D graphics models of robot arms and objects of interest for robot arm tasks; (2) high-fidelity calibration of the 3-D graphics models relative to TV camera 2-D image frames that cover both the robot arm and the objects of interest; (3) high-fidelity overlay of the calibrated graphics models over the actual robot arm and object images in a TV camera image frame on a monitor screen; and (4) high-fidelity motion control of the robot arm graphics image by using the same control software that drives the robot.

The high-fidelity fused virtual and actual reality image displays are very useful tools for planning, previewing, and predicting robot arm motions without commanding and moving the robot hardware. The operator can generate visual effects of robot motion by commanding and controlling the motion of the robot's graphics image superimposed over TV pictures of the live scene. Thus, the operator can see the consequences of motion commands in real time, before sending the commands to the remotely located robot. The calibrated virtual reality display system can also provide high-fidelity *synthetic* or *artificial* TV camera views to the operator. These synthetic views can make critical motion events visible that are otherwise hidden from the operator in a TV camera view or for which no TV camera view is available. More on the graphics system in the ATOP control station can be found in Bejczy et al.,[11] Bejczy and Kim,[12] Kim and Bejczy,[13,16] Kim,[14,17] Fiorini et al.,[15] and Lee et al.[18]

25.5 ATOP Control Experiments

To evaluate computer-augmented and sensor-aided advanced teleoperation capabilities, two types of experiments were designed and conducted: experiments with *generic* tasks and experiments with *application* tasks. Generic tasks are idealized, simplified tasks that serve the purpose of evaluating

© 2002 by CRC Press LLC

some specific advanced teleoperation features. Application tasks simulate real-world uses of teleoperation.

In the *generic task experiments*, described in detail by Hannaford et al.,[19] four tasks were used: attach and detach velcro; peg insertion and extraction; manipulating three electrical connectors; and manipulating a bayonet connector. Each task was broken down into subtasks. The test operators were chosen from a population with technical background but without an in-depth knowledge of robotics and teleoperation. Each test subject received 2 to 4 hours of training on the control station equipment. The practice consisted of four to eight 30-minute sessions.

The generic task experiments focused on the evaluation of kinesthetic force feedback vs. no force feedback, using the specific force feedback implementation techniques of the JPL ATOP project. The evaluation of the experimental data supports the idea that multiple measures of performance must be used to characterize human performance in sensing and computer-aided teleoperation. For instance, in most cases, kinesthetic force feedback significantly reduced task completion time. In some cases, however, it did not, but it sharply reduced extraneous forces. More information on the results can be found in Hannaford et al.[19,20]

Two major *application task experiments* were performed: *one without* communication time delay and *one with* communication time delay. The experiments without communication time delay were grouped around a simulated satellite repair task that duplicated the solar maximum satellite repair (SMSR) mission performed by two astronauts orbiting Earth in the Space Shuttle in 1984. Thus, it offered a realistic performance reference database. This repair was a challenging task, because the satellite was not designed for repair. Very specific auxiliary subtasks had to be performed (e.g., a hinge attachment) to accomplish the basic repair which, in our simulation, was the replacement of the main electric box (MEB) of the satellite. The total repair performed by two astronauts in Earth orbit took about 3 hours, and comprised the following subtasks: thermal blanket removal, hinge attachment for MEB opening, opening of the MEB, removal of electrical connectors, replacement of MEB, securing parts and cables, replug of electrical connectors, closing of MEB, and reinstating thermal blanket. The two astronauts trained for this repair on the ground for about one year.

Several important observations were made during the performance experiments. The two most important observations are: (1) the remote control problem in any teleoperation mode and using any advanced component or technique is at least 50% a visual perception problem to the operator, influenced greatly by view angle, illumination, and contrasts in color or in shading, and (2) the training or, more specifically, the training cycle has a dramatic effect upon operator performance. The practical purpose of training is, in essence, to help the operator develop a mental model of the system and the task. During task execution, the operator acts through the aid of this mental model. It is, therefore, critical that the operator completely understands the response characteristics of the sensing and computer-aided ATOP system which has a variety of selectable control modes, adjustable control gains, and scale factors. More on application experiments results can be found in Hannaford et al.[20] and Das et al.[21,22]

The performance experiments with communication time delay, conducted on a large laboratory scale in early 1993, utilized a simulated life-size satellite servicing task set up at the Goddard Space Flight Center (GSFC) and controlled 4000 km away from the JPL ATOP control station. Three fixed TV camera settings were used at the GSFC worksite, and TV images were sent to the JPL control station over the NASA-select satellite TV channel at video rate. Command and control data from JPL to GSFC and status and sensor data from GSFC to JPL were sent through the Internet computer communication network. The roundtrip command/information time delay varied from 4 to 8 sec between the GSFC worksite and the JPL control station, dependent on the data communication protocol.

The task involved the exchange of a satellite module. This required inserting a 45-cm long power screwdriver attached to the robot arm through a 45-cm long hole to reach the latching mechanism

© 2002 by CRC Press LLC

at the module backplane, unlatching the module from the satellite, connecting the module rigidly to the robot arm, and removing the module from the satellite. The placement of a new module back to the satellite frame followed the reverse sequence of actions.

Four camera views were calibrated for this experiment, entering 15 to 20 correspondence points in total from three to four arm poses for each view. The calibration and object localization errors at the critical tool insertion task amounted to about 0.2 cm each, well within the allowed insertion error tolerance. This 0.2-cm error is referenced to the zoom-in view (fov = 8°) from the overhead (front view) camera which was about 1 m away from the tool tip. For this zoom-in view, the average error on the image plane was typically 1.2 to 1.6% (3.2 to 3.4% maximum error); a 1.4% average error is equivalent to a 0.2-cm displacement error on the plane 1 m in front of the camera. These successful experiments showed the practical utility of high-fidelity predictive-preview display techniques, combined with sensor referenced automatic compliance control to complete a demanding telerobotic servicing task under communication time delay. More on these experiments and on the related error analysis can be found in Kim and Bejczy[16] and Kim.[17] Figures 25.9a and 25.9b illustrate a few typical overlay views.

A few notes are in order regarding the use of calibrated graphics overlays for time-delayed remote control:

1. The operator must exercise a number of computation activities and needs an easy and user-friendly interface to the computation system.
2. The selection of the matching graphics and TV image points by the operator has an impact on the calibration results. The operator must select significant points. This requires some rule-based knowledge about what point is significant in a given view. The operator must also use good visual acuity to click the selected significant points with a mouse.

The following general lessons were learned from the development and experimental evaluation of the JPL ATOP:

1. The sensing, computer- and graphics-aided advanced teleoperation system truly provides new and improved technical features. To transform these features into new and improved task performance capabilities, the operators of the system must be transformed from naive to skilled operators through education and training.
2. Carrying out a task requires that the operator follow a clear procedure or protocol that has to be worked out off-line, tested, modified, and finalized. This procedure- or protocol-following habit will help develop the experience and skill of the operator.
3. The final skill of an operator can be tested and graded by his or her ability to successfully improve recovery from unexpected errors and complete a task.
4. The variety of I/O activities in the ATOP control station requires workload distribution between two operators. The primary operator controls the sensing and computer-aided robot arm system, while the secondary operator controls the TV camera and monitor system and assures protocol compliance. Thus, the coordinated training of two cooperating operators is essential to successful use of the ATOP system for performing realistic tasks. It is not yet known what a single operator could do or how. To configure and integrate the current ATOP control station for successful use by a single operator is challenging research and development work.
5. ATOP system development require us to find ways to improve technical components and create new subsystems. The final challenge is to integrate the improved or new technical features with the natural capabilities of the operator through appropriate human–machine interface devices and techniques to produce improved overall system performance. Figure 25.10 illustrates in a summary view the machine environment of the JPL ATOP control station.

© 2002 by CRC Press LLC

FIGURE 25.10 JPL ATOP control station.

25.6 Anthropomorphic Telerobotics

The use of an industrial type robot arm with industrial type parallel claw end effectors sets definite limits for the task performance capabilities of the arm as dexterity in manipulation resides in the mechanical and sensing capabilities of the hands (or end effectors). The use of industrial arms and end effectors in space would essentially require the design of space manipulation tasks to match the capabilities of industrial arms and end effectors. Existing space manipulation tasks (except the handling of large space cargos) are designed for astronauts and their tools. Well over 300 tools are available today and certified for use by extra-vehicular activity (EVA) astronauts in space. Motivated by these facts, an effort parallel to the ATOP project was initiated at JPL to develop and evaluate human-equivalent or human-rated dexterous telemanipulation capabilities for potential applications in space because all manipulation-related tools used by EVA astronauts are human rated.

The actual design and laboratory prototype development included the following technical features: (1) the system is fully electrically driven; (2) the hand and glove have four fingers (the little finger is omitted) and each finger has four DOFs; (3) the base of the slave fingers follow the curvature of the human fingers base; (4) the slave hand and wrist form a mechanically integrated closed subsystem, that is, the hand cannot be used without its wrist; (5) the lower slave arm that connects to the wrist houses the full electromechanical drive system for the hand and wrist (total of 19 DOFs), including control electronics and microprocessors; and (6) the slave drive system electromechanically emulates the dual functions of human muscles, namely, position and force control. This implies a novel and unique implementation of active compliance. All of the specific technical features taken together make this exoskeleton unique among the few similar systems. No other previous or ongoing developments have all the aforementioned technical features in one integrated system, and some of the specific technical features are not represented in similar systems. More on this system can be found in Jau[23] and Jau et al.[24]

© 2002 by CRC Press LLC

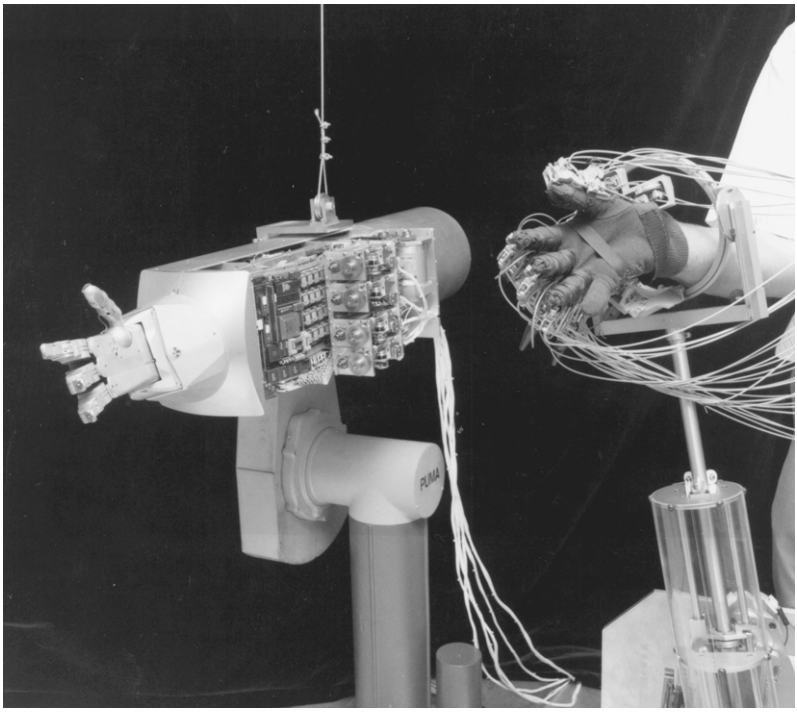


FIGURE 25.11 Master glove controller and anthropomorphic hand.

The JPL anthropomorphic telemanipulation system was assembled and tested in a *terminus control configuration*. The master glove was integrated with a previously developed nonanthropomorphic six DOF force-reflecting hand controller (FRHC), and the mechanical hand and forearm were mounted to an industrial robot (PUMA 560), replacing its standard forearm. The terminus control mode refers to the fact that only the terminus devices (glove and robot hand) are of anthropomorphic nature, and the master and slave arms are nonanthropomorphic. The system is controlled by a high-performance distributed computer controller. Control electronics and computing architecture were custom developed for this telemanipulation system.

The anthropmorphic telemanipulation system in terminus control configuration is shown in Figure 25.11. The master arm/glove and the slave arm/hand have 22 active joints each. The manipulator lower arm has five additional drives to control finger and wrist compliance. This active electromechanical compliance (AEC) system provides the equivalent muscle dual functions of position and stiffness control. A cable links the forearm to an overhead gravity balance suspension system, relieving the PUMA upper arm of this additional weight. The forearm has two sections, one rectangular and one cylindrical. The cylindrical section, extending beyond the elbow joint, contains the wrist actuation system. The rectangular cross-section houses the finger drive actuators, all sensors, and the local control and computational electronics. The wrist has three DOFs with angular displacements similar to the human wrist. The wrist is linked to an AEC system that controls wrist stiffness. The slave hand, wrist, and forearm form a mechanically closed system, that is, the hand cannot be used without its wrist. A glove-type device is worn by the operator. Its force sensors enable hybrid position/force control and compliance control of the mechanical hand. Four fingers are instrumented, each having four DOFs. Position feedback from the mechanical hand provides position control for each of the 16 glove joints. The glove's feedback actuators are remotely located and linked to the glove through flex cables. One-to-one kinematic mapping exists between the master glove and slave hand joints, thus reducing the computational efforts and control complexity

© 2002 by CRC Press LLC

of the terminus subsystem. The exceptions to the direct mapping are the two thumb base joints that need kinematic transformations.

The system was successfully tested on 18 astronaut-equivalent tool handling tasks. It became clear during the tests, however, that many EVA tool handling tasks require a dual-arm fingered-hand system with four fingers and with seven DOF compliant robot arms. The tests also demonstrated the distinct advantages of the terminus control configuration in anthropomorphic telemanipulation as compared to a fully exoskeleton master arm configuration.

25.7 New Trends in Applications

Applications of teleoperators or telerobots are numerous, in particular in the nuclear and munitions industries, maintenance and reclaiming industries operating in hostile environments, and industries that support space and underwater operations and explorations. Robotics and teleoperation technology recently started breaking ground in the medical field. Diagnostic and treatment surgeries, including microsurgery and telesurgery, seem to be receptive fields for potential use of robotic and teleoperator tools and techniques.

An interesting robot-assisted microsurgery (RAMS) telerobotic workstation was developed at JPL recently in collaboration with Steve Charles, M.D., a vitreo-retinal surgeon. RAMS is a prototype system that will be completely under the manual control of a surgeon. The system, shown in Figures 25.12a and 25.12b, has a slave robot that holds surgical instruments. The slave robot motions replicate in six DOFs the motions of the surgeon's hand measured using a master input device with a surgical instrument-shaped handle. The surgeon commands motions for the instrument by moving the handle in the desired trajectories. The trajectories are measured, filtered, scaled down, and then used to drive the slave robot.

The RAMS workstation is a six DOF master–slave telemanipulator with programmable controls. The primary control mode is telemanipulation, which includes task frame-referenced manual force feedback and textural feedback. The operator is able to interactively designate or share automated control of robot trajectories. RAMS refines the physical scale of state-of-the-art microsurgical procedures and enables more positive outcomes for average surgeons during typical procedures. The RAMS workstation controls include features to enhance manual positioning and tracking in the face of myoclonic jerk and tremor that limit most surgeons' fine-motion skills. More on RAMS can be found in Schenker et al.[25] and Charles.[26]

Acknowledgment

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

© 2002 by CRC Press LLC

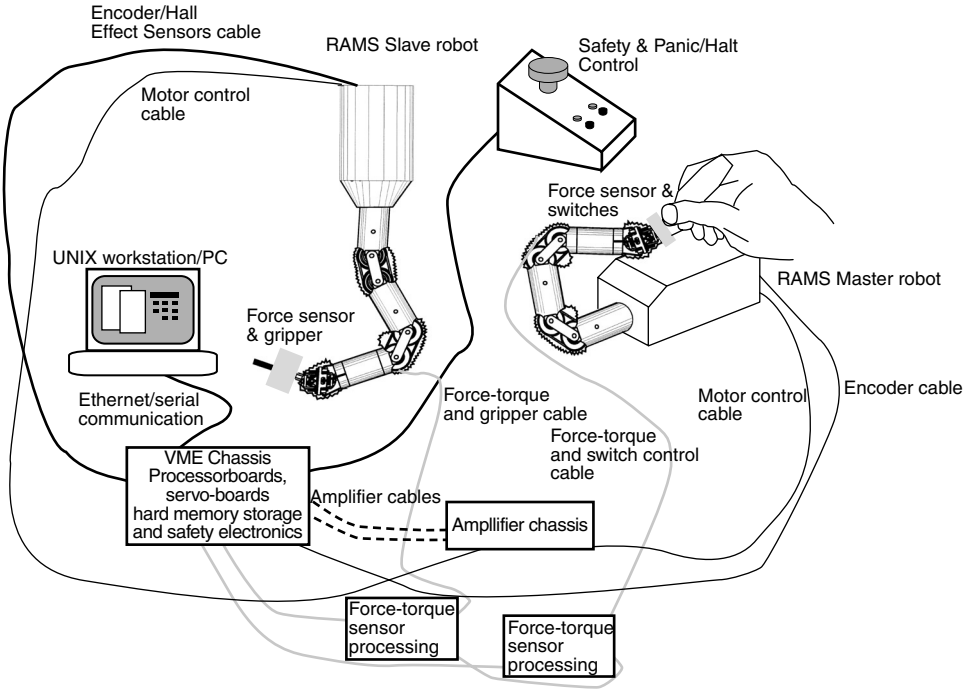


FIGURE 25.12A Schematic of RAMS master–slave system.

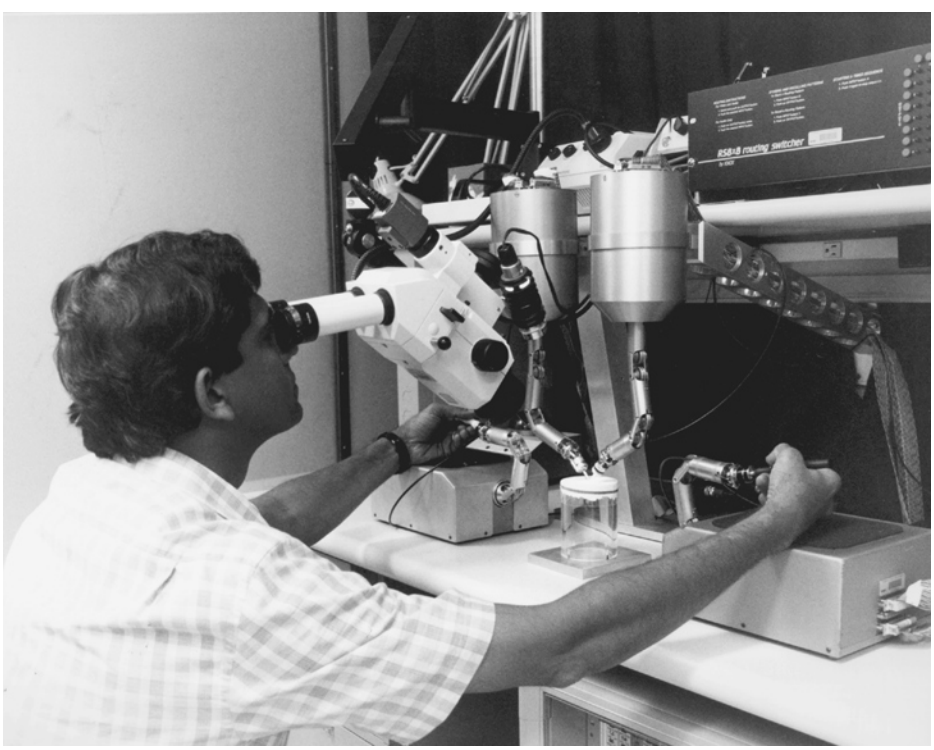


FIGURE 25.12B Fine suturing test with two-handed RAMS system.

© 2002 by CRC Press LLC

References

1. Brooks, T.L. and Bejczy, A.K., *Hand Controllers for Teleoperation: A State-of-the-Art Technology Survey and Evaluation,* JPL Publication 85-11, March 1, 1985.
2. Bejczy, A.K. and Salisbury, J.K., Jr., Kinesthetic coupling between operator and remote manipulator, *Proc. ASME Int. Comput. Technol. Conf.,* San Francisco, 1, 197, August 12-15, 1980; also, controlling remote manipulators through kinesthetic coupling, *Comput. Mech. Eng.,* 1(1), 46, July 1983.
3. Knight, L.W., and Retter, D., Datahand™: Design, potential performance, and improvements in the computer keyboard and mouse, *Natl. Hum. Factors Soc. Conf.,* Denver, November 1989.
4. Knight, L.W., *Single Operator Environment: Experimental Hand-Grip Controller for ATOP,* Jet Propulsion Laboratory, Pasadena, CA, July 31, 1992, and July 30, 1993.
5. Bejczy, A.K., Szakaly, Z., and Kim, W.S., A laboratory breadboard system for dual arm teleoperation, *Third Annu. Workshop on Space Operations, Automation and Robotics,* NASA Conf. Pub. 3059, Johnson Space Center, Houston, TX, 649, July 1989.
6. Bejczy, A.K., and Szakaly, Z., Performance capabilities of a JPL dual-arm advanced teleoperation system, *Space Operations, Appl. Res. Symp. Proc.,* Albuquerque, NM, 168, June 26, 1990.
7. Bejczy, A.K., Szakaly, Z., and Ohm, T., Impact of end effector technology on telemanipulation performance, *Third Annu. Workshop on Space Operations, Automation and Robotics,* NASA Conf. Pub. 3059, Johnson Space Center, Houston, TX, 429, 1989.
8. Bejczy, A.K., and Szakaly, Z., An 8-DOF dual arm system for advanced teleoperation performance experiments, *Space Operations, Appli. Res. Symp.* NASA Conf. Pub. 3127, Johnson Space Center, Houston, TX, 282, 1991; also Lee, S. and Bejczy, A.K., Redundant arm kinematic control based on parametrization, *Proc. IEEE Int. Conf. Robotics Automation,* Sacramento, CA, 458, April 1991.
9. Bejczy, A.K., and Szakaly, Z.F., Universal computer control system (UCCS) for space telerobots, *Proc. IEEE Int. Conf. Robotics Automation,* Raleigh, NC, March 30-April 3, 1987.
10. Bejczy, A.K., and Szakaly, Z., A harmonic motion generator for telerobotic applications, *Proc. IEEE Int. Conf. Robotics Automation,* Sacramento, CA, 2032, 1991.
11. Bejczy, A.K., Kim, W.S., and Venema, S., The phantom robot: predictive display for teleoperation with time delay, *Proc. IEEE Int. Conf. Robotics Automation,* Cincinnati, OH, 546, May 1990.
12. Bejczy, A.K., and Kim, W.S., Predictive displays and shared compliance control for time delayed telemanipulation, *Proc. IEEE Int. Workshop Intelligent Robots Systems,* Tsuchiura, Japan, 407, July 1990.
13. Kim, W.S., and Bejczy, A.K., Graphics displays for operator aid in telemanipulation, *Proc. IEEE Int. Conf. Syst., Man Cybernetics,* Charlottesville, VA, 1059, Oct. 1991.
14. Kim, W.S., Graphical operator interface for space telerobotics, *Proc. IEEE Int. Conf. Robotics Automation,* Atlanta, GA, 95, May 1993.
15. Fiorini, P., Bejczy, A.K., and Schenker, P., Integrated interface for advanced teleoperation, *IEEE Control Syst. Mag.,* 13(5), 15, October 1993.
16. Kim, W.S., and Bejczy, A.K., Demonstration of a high-fidelity predictive/preview display technique for telerobotics servicing in space, *IEEE Trans. Robotics and Automation,* Special Issue on Space Telerobotics, 698, October 1993; also, Kim, W.S., Schenker, R S., Bejczy, A.K., Leake, S., and Ollendorf, S., An advanced operator interface design with preview/predictive displays for ground-controlled space telerobotic servicing, *SPIE Conf. Pub. 2057, Telemanipulator Technology and Space Telerobotics,* Boston, September 1993.
17. Kim, W.S., Virtual reality calibration for telerobotic servicing, *Proc. IEEE Int. Conf. Robotics Automation,* San Diego, 2769, May 1994.
18. Lee, P., Hannaford, B., and Wood, L., Telerobotic configuration editor, *Proc. IEEE Int Conf. Syst., Man Cybernetics,* Los Angeles, 121, 1990.
19. Hannaford, B., Wood, L., Guggisberg, B., McAffee, D., and Zak, H., *Performance Evaluation of a Six-Axis Generalized Force-Reflecting Teleoperator,* Jet Propulsion Laboratory, Pub. 89-18, Pasadena, CA, June 15, 1989.

© 2002 by CRC Press LLC

20. Hannaford, B., Wood, L., Guggisberg, B., McAffee, D., and Zak, H., Performance evaluation of a six-axis force-reflecting teleoperation, *IEEE Trans. Syst., Man Cybernetics,* 21(3), 1991.
21. Das, H., Zak, H., Kim, W.S., Bejczy, A.K., and Schenker; P.S., Performance experiments with alternative advanced teleoperator control modes for a simulated solar max satellite repair, Proc. Space Operations, Automation Robotics Symp. NASA Conf. Pub. 3127, Johnson Space Center, Houston, TX, 294, July 9-11, 1991.
22. Das, H., Zak, H., Kim, W.S., Bejczy, A.K., and Schenker, P.S., Performance with alternative control modes in teleoperation, in *Teleoperators and Virtual Environments,* MIT Press, Cambridge, MA, 1(2), 219, 1993.
23. Jau, B.M., Man-equivalent teleopresence through four fingered human-like hand system, *Proc. IEEE Int. Conf. Robotics Automation* Nice, France, IEEE Press, Los Alamitos, CA, 843, 1992.
24. Jau, B.M., Lewis, M.A., and Bejczy, A.K., Anthropomorphic telemanipulation system in terminus control mode, Proc. ROMANSY '94, Springer-Verlag, Berlin, 1994.
25. Schenker, P., Das, H., and Ohm, T., A new robot for high dexterity microsurgery, *Proc. First Int. Conf., CVRMed* Nice, April 1995; also in Computer vision, virtual reality and robotics in medicine, *Lecture Notes in Computer Science,* Ayache, Nicholas, Ed., Springer-Verlag, Berlin, 1995.
26. Charles, S., Das, H. et al., JPL, *Proc. 8th Int. Conf. Advanced Robotics,* Monterey, CA, 5, July 7-9, 1997.

© 2002 by CRC Press LLC

26

Mobile Robotic Systems

26.1 Introduction
26.2 Fundamental Issues
Definition of a Mobile Robot • Stanford Cart • Intelligent Vehicle for Lunar/Martian Robotic Missions • Mobile Robots — Nonholonomic Systems
26.3 Dynamics of Mobile Robots
26.4 Control of Mobile Robots

Nenad M. Kircanski
University of Toronto

26.1 Introduction

This subsection is devoted to modeling and control of mobile robotic systems. Because a mobile robot can be used for exploration of unknown environments due to its partial or complete autonomy is of fundamental importance. It can be equipped with one or more manipulators for performing mission-specific operations. Thus, mobile robots are very attractive engineering systems, not only because of many interesting theoretical aspects concerning intelligent behavior and autonomy, but also because of applicability in many human activities. Attractiveness from the theoretical point of view is evident because no firm fundamental theory covering intelligent control independent of human assistance exists. Also, because wheeled or tracked mobile robots are nonholonomic mechanical systems, they are attractive for nonlinear control and modeling research. In Section 26.2 of this chapter, fundamental issues are explained regarding nonholonomic systems and how they differ from holonomic ones. Although we will focus attention mostly on wheeled mobile robots, those equipped with tracks and those that rely on legged locomotion systems are addressed as well. The term "mobility" is addressed from the standpoint of wheeled and tracked platform geometry. Examples provided are also showing how different platforms have been built in practice.

Section 26.3 covers dynamics of mobile robots. Models range from very complex ones that include dynamics of deformable bodies to relatively simple models mostly used to facilitate development of control algorithms. The discussion concludes with some model transformations that help obtain relatively simple models.

The next section is devoted to control issues from the standpoint of both linear and nonlinear control theories. We explain the difference between controllability in the linear system theory and controllability of mobile robots, having in mind that a mobile robot is a nonlinear system.

26.2 Fundamental Issues

26.2.1 Definition of a Mobile Robot

A definition of a mobile robotic system does not exist. The International Standards Organization (ISO), has defined an industrial robot as:

© 2002 by CRC Press LLC

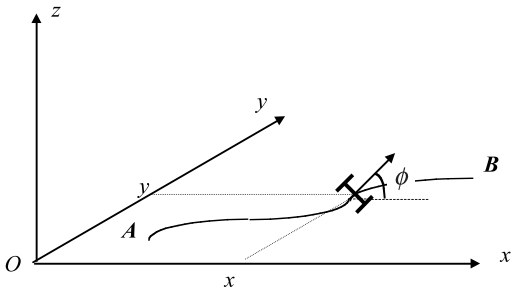


FIGURE 26.1 Definition of coordinates.

Definition 1: An *industrial robot* is an automatic, servo-controlled, freely programmable, multipurpose *manipulator*, with several axes, for the handling of workpieces, tools, or special devices. Variably programmed operation make possible the execution of a multiplicity of tasks.

This definition clearly indicates that the term "industrial robot" is linked to a "manipulator," meaning that such a mechanical system is attached to a base. Typically, the base is fixed with respect to a ground or single-degree-of-freedom platform mounted on rails. We also observe that an industrial robot must have a programmable control system so that the same robot can be used for different tasks.

A mobile robot has two essential features that are not covered by this definition. The first is obviously mobility, and the second is autonomy. A minimum requirement for a mobile robot is to be capable of traversing over flat horizontal surfaces. Given a point $\boldsymbol{A}$ on such a surface S, where the robot is positioned at a time instant t, the mobile robot must be capable of reaching any other point $\boldsymbol{B}$ at a certain distance $d < \infty$ from $\boldsymbol{A}$, in a finite time T. Here, we should clarify the meaning of the term "position." Let us assume that a coordinate frame $Oxyz$ is attached to the surface so that O belongs to the surface, while the z-axis is normal to the surface (Figure 26.1). Clearly, the position of any point on the surface is defined by coordinates (x, y). But, the position of the robot is actually defined by the coordinates (x, y, ϕ), where ϕ defines the orientation of the chassis with respect to the x axis. Of course the orientation can be defined in many different ways (for example, with respect to y or any other axis in the plane surface). So, mobility means that the robot is capable of traversing from the position (x_A, y_A, ϕ_A) to (x_B, y_B, ϕ_B) in a finite time interval T.

Mobility, as discussed above, is limited in terms of the system's ability to traverse different surfaces. The simplest case is a flat horizontal surface z = constant. Most 4-wheel mobile robots are designed for such terrains. In the case of a smooth surface $z = f(x, y)$, where f is an arbitrary continuous function of x and y, the ability of a wheeled robot to reach any desired point $\boldsymbol{B}$ from a point $\boldsymbol{A}$ on the surface depends upon (1) the ability of the robot to produce enough driving force to compensate for gravity force while moving toward the goal point; and (2) the presence of sufficient friction forces between the wheels and the ground to prevent continuous slippage. Notice also that there is no uniquely defined path between the points $\boldsymbol{A}$ and $\boldsymbol{B}$, and the robot may be incapable of traversing some trajectories, but still capable of reaching point $\boldsymbol{B}$ provided the trajectory is conveniently selected.

In discussions related to mobility a fundamental question concerns climbing and descending stairs, over-crossing channels, etc. Previously we have implicitly assumed that the function f is differentiable with respect to x and y. If this does not hold, as is true for a staircase, mobility can be achieved with tracks or legs. Robots with legs are called "legged-locomotion robots." Such robots are rarely used in practice due to the complexity (and thus reliability and cost) of the locomotion subsystem. Tracked robots are usually six-wheel robots with a set of two tracks mounted on three wheels on the left and three wheels on the right side of the chassis. Each track has a tread

© 2002 by CRC Press LLC

that engages with the edge of the first stair of a staircase. Such engagement allows for lifting the front side of the chassis while the back side remains on the ground. In this phase of climbing, the vehicle tilts backward while moving forward and finally reaches an inclination angle equal to that of the staircase. The tread on the tracks engages with several stairs simultaneously, allowing the robot to move forward.

In the second part of this section, the mobility will be highlighted from the standpoint of nonholonomic constraints. We explain why a manipulator is a holonomic and why a mobile robot is a nonholonomic system. Prior to that, though, we attempt to define a mobile robot. As mentioned at the beginning of this section, the second essential feature of a mobile robot is autonomy. We know that vehicles have been used as a means of transportation for centuries, but vehicles were never referred to as "mobile robots" before, because the fundamental feature of a robot is to perform a task without human assistance. In an industrial, well-structured environment it is not difficult to program a robot manipulator to perform a task. On the other hand, the term "mobile robot" does not necessarily correlate to an industrial environment, but a natural or urban environment. Industrial mobile robots are called automatic guided vehicles (AGVs). AGVs are mobile platforms typically guided by an electromagnetic source (a set of wires) placed permanently under the floor cover. Tracking of the guidelines is realized through a simple feedback/feedforward control. Thus, an AGV is not referred to as a mobile robot because it is not an autonomous system.

An autonomous system must be capable of performing a task without human assistance and without relying on an electronic guidance system. It must have sensors to identify environmental changes, and it must incorporate planning and navigation features to accomplish a task. More details about these features are given in later sections, but now we provide an example of a simple mobile robot currently used in urban environments: a vacuum-cleaning mobile robot. Such commercially available robots have an ultrasonic-based range-finder mounted on a pan-and-tilt unit. This unit is located on the front end of the chassis and constantly rotates left-and-right and up-and-down independently of the speed of the vehicle. The range-finder is an ultrasonic transceiver/receiver sensor mounted on the unit end-point. The echoes are processed by an on-board computer to identify obstacles around the robot. A planner is a software module that describes the "desired path" so that cleaning is performed uniformly all over the floor surface. The navigator is the software module that provides changes in desired trajectories in accordance with the obstacles/walls located by the sensorial system. Such a robot will automatically slow down and avoid a collision should another vehicle or a human traverse its trajectory. Clearly, autonomy is not necessarily correlated to artificial intelligence. "Intelligent control" can be a feature of a mobile robot, but it is not a must in practice.

Based on the previous discussion we can define a mobile robot as follows:

Definition 2: A *mobile robot* is an autonomous system capable of traversing a terrain with natural or artificial obstacles. Its chassis is equipped with wheels/tracks or legs, and, possibly, a manipulator setup mounted on the chassis for handling of work pieces, tools, or special devices. Various preplanned operations are executed based on a preprogrammed navigation strategy taking into account the current status of the environment.

Although this is not an official definition proposed by ISO, it contains all the essential features of a mobile robot. According to this definition, an AGV is not a mobile robot because it lacks autonomy and the freedom to traverse a terrain (it is basically a single-degree-of-freedom moving platform along a built-in guide path). Similarly, "teleoperators," used in the nuclear industry for decades, are not mobile robots for the same reason: a human operator has full control over the vehicle. A teleoperator looks like a mobile robot because it has a chassis and a manipulator arm on top of it, but its on-board computer is programmed to follow the remote operator's commands. An example of a real mobile robot is the four-wheel Stanford Cart built in the late seventies.[1] This relatively simple robot, as well as some advanced ones, including an intelligent robotic vehicle recently developed for Lunar/Martian robotic missions, are described in the text to follow.

© 2002 by CRC Press LLC

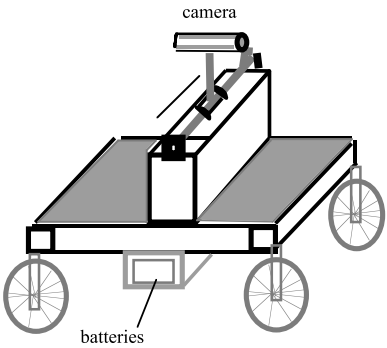


FIGURE 26.2 The Stanford Cart.

26.2.2 Stanford Cart

The cart was developed at the Stanford Artificial Intelligence Laboratory as a research setup for Ph.D. students (Figure 26.2). The robot was equipped with an on-board TV system and a computer dedicated to image processing and driving the vehicle through obstacle-cluttered spaces. The system gained its knowledge entirely from images. Objects were located in three dimensions, and a model of the environment was built with information gained while the vehicle was traversing a terrain. The system was unreliable for long runs and very slow (1 m in 10 to 15 min).

The operation would start at a certain point on a flat horizontal surface (flat floor) cluttered with obstacles. The camera was mounted on a sliding unit (50-cm track on top of the chassis) so that it was able to move sideways while keeping the line-of-sight forward. Such sidewise movements allowed the collection of several images of the same scene with a fixed lateral offset. By correlating those images the control system was able to identify locations of obstacles in the camera's field of vision. Control was simplified because these images were collected while the cart was inactive. After identifying the location of obstacles as simple fuzzy ellipsoids projected on the floor surface, the vehicle itself was modeled as a fuzzy ellipsoid projected on the same surface.

Based on the environment model a Path Planner was used to determine the shortest possible path to the goal-point. This program was capable of finding the path that was either a straight segment between the end and initial points, or a set of tangential segments and arcs along the ellipses (Figure 26.3). To simplify the algorithm, the ellipses were actually approximated by circles.

The navigation module was very primitive because the chart motor control lacked feedback. Thus, the vehicle was moved roughly in a certain direction by activating, driving, and steering motors for a brief time. After moving the vehicle for about 1 m, the whole procedure was repeated.

Although the whole process was extremely slow (roughly 4 to 6 m/h), and vehicle control very primitive, this was one of the first platforms that had all features needed for a robot to be regarded as a real mobile robot. It was autonomous and adaptable to environmental variations.

26.2.3 Intelligent Vehicle for Lunar/Martian Robotic Missions

In contrast to the Stanford Cart built as a students' experimental setup in the late 1970s, the intelligent robotic vehicle system (IRVS) was developed by the UA/NASA Space Engineering Research Center in the early 1990s.[2] This robot was developed to facilitate *in situ* exploration missions on the lunar/Martian surface. The system was designed to determine (1) site topography using two high-resolution CCD cameras and stereo-photogrammetry techniques; (2) surface mineral composition using two spectrometers, an oven soil heater, and a gas analyzer; and (3) regolith depths using sonar sounders. The primary goal of such missions was to provide accurate information that incorporates *in situ* resource utilization on the suitability of a site to become a lunar/Martian outpost. Such a lunar base would be built using locally available construction materials (rocks and

© 2002 by CRC Press LLC

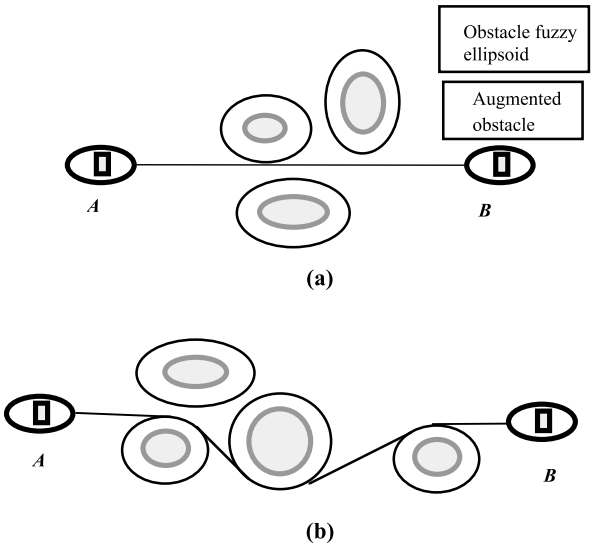


FIGURE 26.3 Path Planning results for two distinct scenarios: (a) a straight line segment exists between the initial and final point, ***A*** and ***B***; and (b) a path consists of a set of straight segments tangential to augmented obstacles, and arcs along the obstacle boundaries that are optimal in terms of its length.

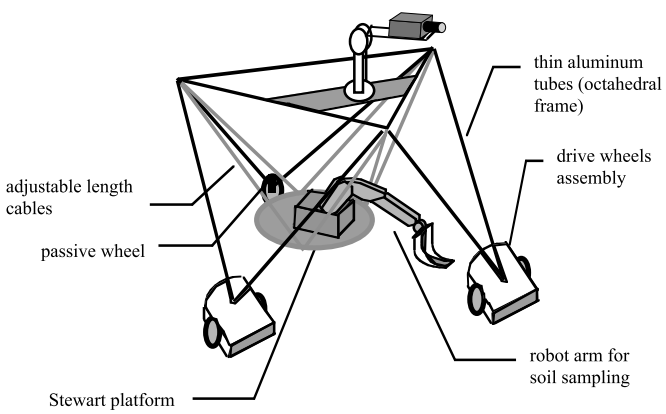


FIGURE 26.4 IRVS mobile robot.

minerals). The base would allow building plants for the production of oxygen and hydrogen for rocket fuel, helium for nuclear energy, and some metals. These materials would be used for building space stations with a cost far lower than the cost of transporting them from Earth.

The IRVS consists of a mobile platform, a manipulator arm, and a set of mission sensors. The most important requirement for the platform is exceptionally high payload-to-mass ratio. This was achieved by using a Stewart platform system developed by the U.S. National Institute of Standards and Technology (Figure 26.4). The structure consists of (1) an octahedral frame constructed of thin walled aluminum tubing, (2) three wheel assemblies (two of them have speed/skid steering control, while the third is a single free wheel), and (3) a work platform suspended by six cables arranged as a Stewart platform. The system is equipped with two high-resolution cameras with power zoom, auto iris, and focus capabilities mounted on a pan/tilt unit at the top of the octahedral frame. Ultrasonic ranging sensors were added for detects objects within a range of 0.2 to 12 m with a field of view of 6°. The system is also equipped with roll-and-pitch sensors that are used for controlling the six cables so that the work platform is always horizontal.

© 2002 by CRC Press LLC

The IRVS control system is nontraditional, i.e., it is not based on sensing, planning, and executing control levels. It consists of a number of behavior programs organized in control levels: *organization*, *coordination,* and *execution*. The organization level consists of four behavior programs: (1) *site-navigator*, (2) *alternative sample collection point* (SCP) *selector*, (3) SCP *recorder*, and (4) SCP *organizer.*

The site-navigator uses a potential field method to calculate the vehicle's trajectory to the next SCP based on vision and range measurements. The alternative SCP selector picks an alternative SCP when a scheduled SCP cannot be reached due to obstacles/craters. The SCP recorder marks the points already visited so that the vehicle cannot sample a SCP twice. The SCP organizer generates a sequence of manipulator and instrument deployment tasks when the robot arrives at a SCP.

The coordination level contains *task-dispatcher* and *behavior arbitrator* programs. The task-dispatcher program analyzes the tasks submitted from the organization level, and activates the behaviors (tasks at the execution level) needed for successful completion of the task's requirements. It also implements a set of failure procedures when a given task cannot be executed because of possible failure (unstable vehicle, etc.). The behavior arbitrator assigns priorities to behaviors so that only the highest-priority behavior will be executed when two or more are simultaneously activated.

Execution level behaviors include the following tasks: obstacle-avoider, open-terrain explorer, etc. Obstacle avoiders are activated when an ultrasonic sensor measurement indicates the presence of an obstacle. Then, the site-navigator behavior is immediately suppressed due to its lower priority than that of the obstacle avoider. The purpose of the open-terrain explorer is to monitor obstacles in an open terrain situation and prevent the vehicle from becoming trapped among obstacles.

Clearly, IRVS control architecture is similar to that of a multitasking real-time kernel. Control is divided over a large number of tasks (called behaviors). The tasks are activated from a control kernel so that the highest-priority one will run first. The control algorithm implemented within a task (behavior) is usually simple and easy to test. The interdependencies among the control laws are implemented within the task's intercommunication network. Message envelopes, circular buffers, semaphores, sockets, and other communication means are used for this purpose. Such control architecture has "fine-granularity" so that elementary control tasks are simple. Still, the overall control architecture is very complex and difficult for theoretical analysis.

26.2.4 Mobile Robots — Nonholonomic Systems

The Stanford Cart and IRVS are just two examples of mobile robots. From these examples we see that mobile platforms can differ in many aspects including geometry, number of wheels, frame structure, etc. From a mechanical point of view there is a common feature to all systems: they are nonholonomic systems. In this section we explain exactly what that means.

Recall that the dynamic model of a manipulator with n degrees of freedom is described by

$$H(q)\ddot{q} + h(q,\dot{q}) = \tau - J^T(q)f \tag{26.1}$$

where $H(q)$ is the $n \times n$ inertia matrix; $h(q,\dot{q})$ is the n-vector due to gravity, centrifugal, and Coriolis forces; τ is the k-dimensional input vector (note that not all joints are necessarily equipped with actuators); $J(q)$ is a $m \times n$ Jacobian matrix; and f is the m vector of constraint forces. The constraint equation generally has the form

$$C(q,\dot{q}) = 0 \tag{26.2}$$

© 2002 by CRC Press LLC

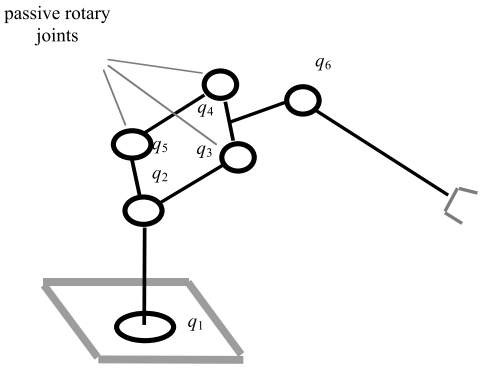


FIGURE 26.5 A manipulator with a closed-loop chain within its structure.

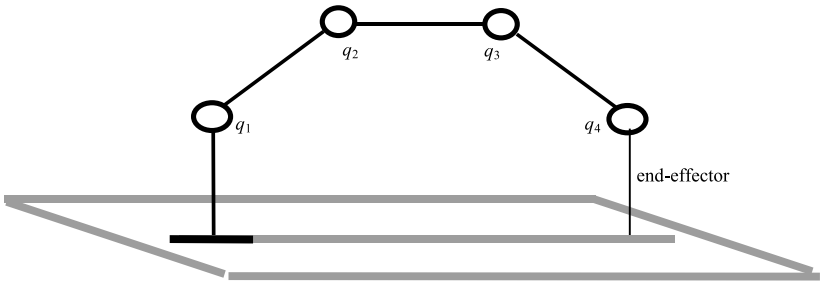


FIGURE 26.6 A manipulator in contact with the environment.

where C is an m vector. Note that the constraint Equation (26.2) involves both the generalized coordinates and its derivatives. In other words, the constraints may have their origins in the system's geometry and/or kinematics.

A typical system with geometric constraints is the robot shown in Figure 26.5. It has six joints (generalized coordinates), but only three degrees of freedom. Assuming that the closed loop chain ABCD is a parallelogram (Figure 26.5), the constraint equations are

$$q_3 + q_2 - \pi = 0$$

$$q_4 - q_2 = 0$$

$$q_5 + q_2 - \pi = 0$$

In this case the constraint equations have form $C_i(q) = 0$. Such constraints, or those that can be integrated into this form, are called *holonomic constraints*.

Another example is a four-degrees-of-freedom manipulator in contact with the bottom surface with an end-effector normal to the surface (Figure 26.6). Assuming that the link lengths are equal we obtain the following constraints:

$$q_1 - q_4 = 0$$

$$q_2 - q_3 = 0$$

$$q_1 + q_2 + q_3 + q_4 - 3\pi = 0$$

© 2002 by CRC Press LLC

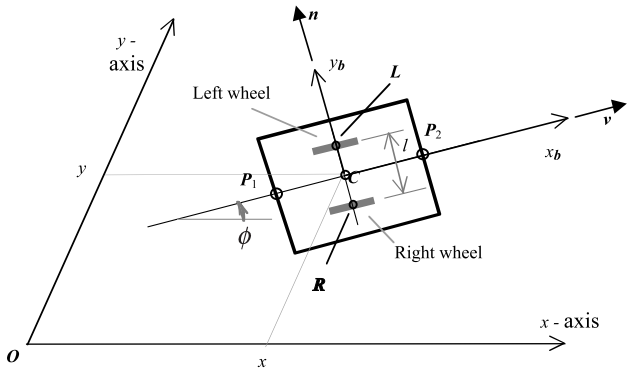


FIGURE 26.7 A simple mobile platform.

A typical system with both holonomic and nonholonomic constraints is a two-wheel platform supported by two additional free wheels in points $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ (Figure 26.7). Because the vehicle consists of three rigid bodies (a chassis and two wheels), we can select the following five generalized coordinates: x and y coordinates of the central point $\boldsymbol{C}$; an angle ϕ between the longitudinal axis of the chassis (x_b) and the x-axis of the reference frame; and θ_L and θ_R, the angular displacements of the left and right wheel, respectively. We assume that the wheels are independently driven and parallel to each other. The distance between the wheels is l.

The constraint equations can be derived from the fact that the vehicle velocity vector $\boldsymbol{v}$ is always along the axis x_b. In other words, the lateral component of the velocity vector (the one that is normal to the wheels) is zero. From Figure 26.7 we observe that the unit vector along x_b is $\mathbf{x}_b = (\cos\phi \sin\phi)^T$, while the vector normal to direction of motion is

$$\mathbf{x}_b = (\cos\phi \;\; \sin\phi)^T$$

Because $\mathbf{v} = (\dot{x}, \;\; \dot{y})^T$, and $\mathbf{v}\cdot\mathbf{n} = 0$, we obtain the first constraint equation:

$$\dot{y}\cos\phi - \dot{x}\sin\phi = 0 \tag{26.3}$$

The other two constraint equations are obtained from the condition that the wheels roll, but do not slip, over the ground surface:

$$v_R = r\,\dot{\theta}_R$$

$$v_L = r\,\dot{\theta}_L$$

where v_R (v_L) is the velocity of the platform at the points $\boldsymbol{R}$ ($\boldsymbol{L}$) in Figure 26.7. Velocities $\dot{\theta}_R$ and $\dot{\theta}_L$ are angular velocities of the right- and left-hand side wheel. The velocity vector in either of these two points has two components: one due to the linear velocity of the chassis, and another due to the rotation of the chassis. The first component is easily obtained as

$$v = \dot{x}\cos\phi + \dot{y}\sin\phi \tag{26.4}$$

from

$$\dot{x} = v\cos\phi \tag{26.5}$$

$$\dot{y} = v\sin\phi \tag{26.6}$$

© 2002 by CRC Press LLC

By multiplying Equation (26.5) by $\cos\phi$ and Equation (26.6) by $\sin\phi$, we easily get Equation (26.4). The second component of the velocity of the platform at the point $\boldsymbol{R}$ is

$$v_R^{'} = \frac{l}{2}\dot{\phi}$$

At the point $\boldsymbol{L}$ the velocity has the same magnitude, but the opposite sign

$$v_L^{'} = \frac{l}{2}\dot{\phi}$$

Now, we get the constraint equations for the wheels:

$$\dot{x}\cos\phi + \dot{y}\sin\phi + \frac{l}{2}\dot{\phi} = r\dot{\theta}_R \tag{26.7}$$

$$\dot{x}\cos\phi + \dot{y}\sin\phi - \frac{l}{2}\dot{\phi} = r\dot{\theta}_L \tag{26.8}$$

The obtained set of constraint equations can be easily converted into the matrix form Equation (26.2). Because the generalized coordinate vector has the form

$$\mathbf{q} = (x, y, \phi, \theta_R, \theta_L)^T$$

the constraint Equations (26.3), (26.7), and (26.8) can be presented in the matrix form

$$\begin{bmatrix} \sin\phi & -\cos\phi & 0 & 0 & 0 \\ -\cos\phi & -\sin\phi & l/2 & 0 & r \\ -\cos\phi & -\sin\phi & -l/2 & r & 0 \end{bmatrix} \dot{\mathbf{q}} = \mathbf{0} \tag{26.9}$$

This is a very characteristic form for nonholonomic constraints: $C(q, \dot{q}) = R(q)\dot{q} = 0$, with

$$R(q) = \begin{bmatrix} \sin\phi & -\cos\phi & 0 & 0 & 0 \\ -\cos\phi & -\sin\phi & l/2 & 0 & r \\ -\cos\phi & -\sin\phi & -l/2 & r & 0 \end{bmatrix}$$

In our case the matrix $R(q)$ is a 3×5 matrix. We also note that there are five generalized coordinates, and three constraint equations. This means that there are two dynamic equations to be written to complete the system (that is, to derive the full dynamic model of the system). Let us now return to the constraint equations. The question is, How many constraint equations are nonholonomic out of the three listed above? The general solution is based on the properties of matrix R, but such a solution is rather complicated (readers who are interested in this topic can find more information in Campion et al.[3]). Instead, we can come to the same conclusion by observing that there is a holonomic equation (constraint) hidden among the three constraint equations given above. To obtain this equation, we subtract Equation (26.8) from Equation (26.7):

$$l\dot{\phi} = r(\dot{\theta}_R - \dot{\theta}_L) \tag{26.10}$$

© 2002 by CRC Press LLC

We can now integrate this equation over time and obtain

$$l\phi = r(\theta_R - \theta_L) + const. \tag{26.11}$$

where *const.* is a constant that depends on initial conditions (angles). This equation can be easily derived straight from the geometry of the system. Because there is no velocity-dependent term in this constraint equation, it is a holonomic one. The set of constraints now becomes

$$\begin{aligned} \dot{x}\cos\phi + \dot{y}\sin\phi + (l/2)\dot{\phi} &= r\dot{\theta}_R \\ \dot{y}\cos\phi - \dot{x}\sin\phi &= 0 \\ l\phi - r(\theta_R - \theta_L) &= 0 \end{aligned} \tag{26.12}$$

In conclusion, the mobile platform shown in Figure 26.7 has one holonomic and two nonholonomic constraints.

26.3 Dynamics of Mobile Robots

Although there has been a vast amount of research effort on modeling open and closed kinematic chains (manipulators), study of systems that include both the mobile platforms and manipulators mounted on top of them is very limited. The dynamic equations of such systems are far more complicated than those of simple manipulators. The first noticeable difference is in the state vector. It is common with manipulators to select joint coordinates and velocities as components of a state vector, but with mobile platforms there is no such simple clear rule. We recall (Figure 26.7) that the coordinates describing the platform position and orientation are x, y, ϕ, θ_R, θ_L. These coordinates are often referred to as "generalized coordinates." The state vector contains these five coordinates and their time-derivatives. The total number of state coordinates is thus ten. On the other hand, the vehicle in Figure 26.7 has only two degrees of freedom (from any position it can only advance for a vector $\Delta\boldsymbol{r}$ along its longitudinal axis, and rotate by an angle $\Delta\phi$ about its vertical rotation). Thus, only two equations are sufficient to describe the system dynamics. These two equations plus the three constraint equations derived in the previous paragraph constitute the mathematical model of the system.

The dynamic equations can be derived from Newton–Euler's formalism, or Lagrange equations, etc. Let us illustrate the derivation of the equations for the vehicle shown in Figure 26.7 using Newton's equations. This method relies on the system's forces and geometry. The forces that act on the chassis are imposed by the torques about the wheel axes (Figure 26.8).

The relationship between the force and the torque is

$$F_R r + I_W \ddot{\theta}_R = \tau_R \tag{26.13}$$

for the right-hand side wheel, and

$$F_L r + I_W \ddot{\theta}_L = \tau_L \tag{26.14}$$

for the left wheel. Here, I_W is the inertia of the wheel, while $\ddot{\theta}_L$ and $\ddot{\theta}_R$ are angular accelerations of the corresponding wheels. The forces $\boldsymbol{F}_R$ and $\boldsymbol{F}_L$ act on the vehicle at points R and L along the longitudinal axis $\boldsymbol{x}_b$. In general, they have different magnitudes, but their vectors are always parallel to each other. They may also have opposite signs, thus turning the chassis about the vertical axis.

© 2002 by CRC Press LLC

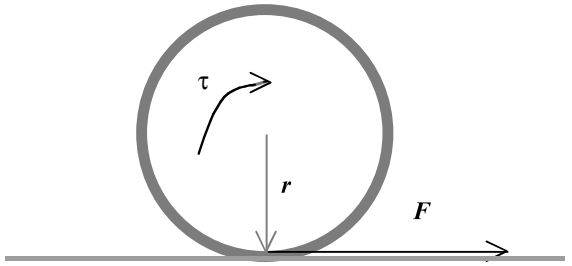


FIGURE 26.8 Wheel force and torque.

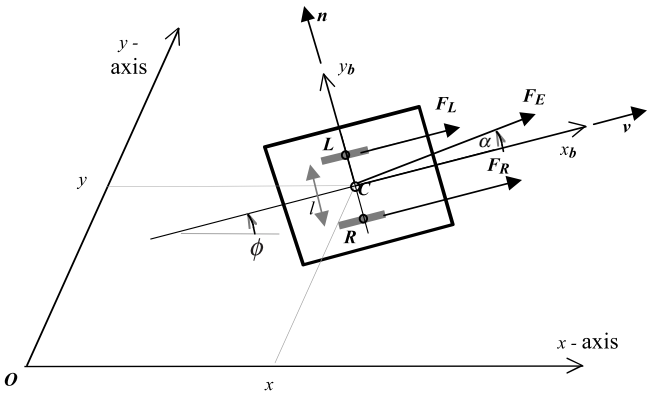


FIGURE 26.9 Forces acting on the mobile platform.

For the sake of generality, let us assume that an external force $\boldsymbol{F}_{\mathrm{E}}$ acts on the chassis at point C in addition to the forces $\boldsymbol{F}_{\mathrm{R}}$ and $\boldsymbol{F}_{\mathrm{L}}$.

Figure 26.9 shows that the total force along the axis x_b is

$$F_{x_b} = F_R + F_L + F_E \cos\alpha$$

while the force along the y_{b} axis is

$$F_{y_b} = F_E \sin\alpha$$

This yields the components along the x and y axes of the reference frame:

$$\begin{aligned} F_x &= (F_R + F_L + F_E \cos\alpha)\cos\phi - (F_E \sin\alpha)\sin\phi \\ F_y &= (F_R + F_L + F_E \cos\alpha)\sin\phi + (F_E \sin\alpha)\cos\phi \end{aligned} \tag{26.15}$$

Finally, we obtain the equations of motion

$$\begin{aligned} F_x &= m\ddot{x} \\ F_y &= m\ddot{y} \end{aligned} \tag{26.16}$$

© 2002 by CRC Press LLC

where m is the mass of the vehicle (chassis + wheels). The moment equation is rather simple:

$$\frac{l}{2}(F_R - F_L) = I\ddot{\phi} \tag{26.17}$$

By combining Equations (26.15) and (26.16), and by grouping these two equations with Equations (26.13), (26.14), and (26.17), we obtain the dynamic model of the vehicle:

$$\begin{aligned} m\ddot{x} - (F_R + F_L + F_E\cos\alpha)\cos\phi + (F_E\sin\alpha)\sin\phi = 0 \\ m\ddot{y} - (F_R + F_L + F_E\cos\alpha)\sin\phi - (F_E\sin\alpha)\cos\phi = 0 \\ I\ddot{\phi} - \frac{l}{2}(F_R - F_L) = 0 \\ I_W\ddot{\theta}_R + F_R r = \tau_R \\ I_W\ddot{\theta}_L + F_L r = \tau_L \end{aligned} \tag{26.18}$$

The forces F_R and F_L are generated between the ground and the wheels. The external force has two components: one along the chassis, and another perpendicular to the chassis. Obviously, the latter does not contribute to motion and can have any value below a limit that would cause lateral sliding of the chassis. The external force can be generated by mechanical means (cable-pulling system), electromagnetic means (attraction or repulsion force in a magnetic or electrostatic field), chemical reaction force (by the action of jets), etc. For the sake of simplicity, let us assume that the external force component acting along the vehicle is zero so that only a lateral component exists. In this case, the angle α equals 90°, and we have the following equations:

$$\begin{aligned} m\ddot{x} - (F_R + F_L)\cos\phi + F_E\sin\phi = 0 \\ m\ddot{y} - (F_R + F_L)\sin\phi - F_E\cos\phi = 0 \\ I\ddot{\phi} - \frac{l}{2}(F_R - F_L) = 0 \\ I_W\ddot{\theta}_R + F_R r = \tau_R \\ I_W\ddot{\theta}_L + F_L r = \tau_L \end{aligned} \tag{26.19}$$

The dynamic model Equation (26.19) includes constrained forces F_R and F_L, and F_E. The model Equation (26.19) and the constraint Equations (26.12) describe the dynamic behavior of the system. The matrix form of the constraint equations derived in the previous section is

$$C(q,\dot{q}) = R(q)\dot{q} = 0 \tag{26.20}$$

Note that the matrix form of Equation (26.19) has the form equal to the one that holds for the manipulators in Equation (26.1). It can be easily shown that Equation (26.19) can be presented in the form

© 2002 by CRC Press LLC

$$H(q)\ddot{q} + h(q,\dot{q}) = T(q)\boldsymbol{\tau} - R^T(q)\boldsymbol{f} \tag{26.21}$$

where q is the vector of generalized coordinates:

$$q = [x\ y\ \phi\ q_R\ q_L]^T$$

$\boldsymbol{f}$ is a vector of constrained forces $\boldsymbol{f} = (F_R, F_L, F_E)^T$, and $\boldsymbol{\tau} = (\tau_R, \tau_L)^T$ is the driving torque vector. The matrix T is called the "input transformation matrix." Comparing Equations (26.20) and (26.21) we see that $H(q) = \text{diag}(m, m, I, I_W, I_W)$. This means that the inertia matrix is a 5×5 diagonal matrix. This matrix is always symmetric, but not necessarily diagonal. Any offset in the center of mass would actually bring nondiagonal elements into existence. The matrix h is zero, but in general it has velocity-dependent terms. These terms have only quadratic forms (either square or the product of two velocities). Matrix T is a 5×2 matrix with the elements:

$$T(q) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Finally, matrix R has the form

$$R(q) = \begin{bmatrix} \sin\phi & -\cos\phi & 0 & 0 & 0 \\ -\cos\phi & -\sin\phi & l/2 & 0 & r \\ -\cos\phi & -\sin\phi & -l/2 & r & 0 \end{bmatrix}$$

which is equal to that derived in the previously (see Equation 26.9).

$R(q)\dot{q} = 0$ shows that a certain relationship exists between the components of the vector $\dot{q}$. Indeed, each component of $\dot{q}$ can be expressed in terms of $\dot{\theta}_R$ and $\dot{\theta}_L$. For example, $\dot{x} = v\cos\varphi$, where v is the velocity of the chassis equal to

$$v = \frac{1}{2}(v_R + v_L)$$

$$= \frac{1}{2}(r\dot{\theta}_R + r\dot{\theta}_L)$$

yields

$$\dot{x} = \frac{r}{2}(\cos\phi)(\dot{\theta}_R + \dot{\theta}_L) \tag{26.22}$$

Similarly, we get

$$\dot{y} = \frac{r}{2}(\sin\phi)(\dot{\theta}_R + \dot{\theta}_L) \tag{26.23}$$

Finally, Equation (26.10) yields

© 2002 by CRC Press LLC

$$\dot{\phi} = \frac{r}{l}(\dot{\theta}_R - \dot{\theta}_L) \tag{26.24}$$

From Equations (26.22 to 26.24) we obtain

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \\ \dot{\theta}_R \\ \dot{\theta}_L \end{bmatrix} = \begin{bmatrix} \frac{r}{2}\cos\phi & \frac{r}{2}\cos\phi \\ \frac{r}{2}\sin\phi & \frac{r}{2}\sin\phi \\ \frac{r}{l} & -\frac{r}{l} \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\theta}_R \\ \dot{\theta}_L \end{bmatrix} \tag{26.25}$$

We now observe that only two components of the vector $\dot{q}$ determine all others. These two components are the velocities of the wheels. Equation (26.25) in general has the form

$$\dot{q} = S(q)v \tag{26.26}$$

where $\dot{q}_0$ is a subvector of $\dot{q}$. There is an important relationship between R and S that can be easily derived by substituting Equation (26.26) into $R(q)\dot{q} = 0$:

$$RS = S^{\mathrm{T}}R^{\mathrm{T}} \tag{26.27}$$

This property is necessary to obtain the state-space model of the mobile robot. To derive this model we first have to find the acceleration vector $\ddot{q}$ by differentiating Equation (26.26) with respect to time:

$$\ddot{q} = S(q)\dot{v} + \dot{S}(q)v \tag{26.28}$$

Then we substitute the acceleration vector Equation (26.28) into the dynamic model Equation (26.21), and obtain the model in a simple form:

$$\bar{H}\dot{v} + \bar{h} = S^T T(q)\boldsymbol{\tau} \tag{26.29}$$

where $\bar{H} = S^T H S$, and $\bar{h} = S^T H \dot{S} v + S^T h$.

A number of important conclusions are based on the form of the model Equation (26.29). First, the time derivative of the vector v contains accelerations of the "core" variables: wheel coordinates. Matrix $\bar{H} = S^T H S$ depends on coordinates, and represents an inertia matrix as seen from the "wheels." Vector $\bar{h}$ depends on system coordinates and velocities. The dimension of the matrices in the model Equation (26.29) equals the number of core coordinates (in our example, the dimension of the matrix $\bar{H}$ is 2×2. This model is very useful for system simulation and control.

26.4 Control of Mobile Robots

A variety of control systems with mobile robots are currently in use. The simplest control systems were developed for so-called "teleoperators" more than 20 years ago. The teleoperators are remotely driven mobile platforms equipped with a manipulator aimed at performing various tasks in nuclear and hazardous environments. Radio or cable link is used to connect the teleoperator with the control

© 2002 by CRC Press LLC

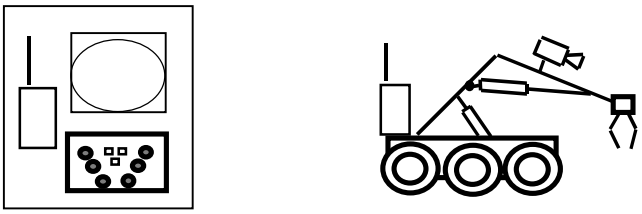


FIGURE 26.10 Teleoperator.

station (Figure 26.10). The control station consists of a TV monitor, and a control panel with joysticks and pushbutton commands. Bi-directional radio links allow simultaneous transmission of

1. Commands from the control panel (CP) to the robot actuators (those that drive the chassis, and control the configuration of the arm)
2. Robot-mounted camera signal to be shown on the TV monitor
3. Voice signal in both directions

Let us denote the control signals from the CP's joysticks as u_c, and assume that the chassis is of the type shown in Figure 26.9. In most applications the power drivers to the chassis motors are current-mode drivers, that is, motor current is proportional to the input signal. Since the electromagnetic torque produced by the motor current is proportional to the current, we can simply express the whole control system as

$$\begin{aligned} \bar{H}\dot{v} + \bar{h} &= S^T T(q)\boldsymbol{\tau} \\ \boldsymbol{\tau} + \boldsymbol{\tau}_f &= k_t u_c \end{aligned} \tag{26.30}$$

where the first equation is Equation (26.29). The friction torque in the gear train and between the wheels and the ground (tracks) is modeled by τ_f. The constant that relates torque and the input signal is denoted as k_t. Such control is very imprecise because of the significant amount of friction in the drive train that is not compensated by the control system. Typically, the vehicle poorly follows the commands especially in transition from a static to a dynamic state (motion). There is no way that small and precise increments in position/orientation can be precisely commanded. Still, such control has theoretical significance, and we will elaborate in more details how a torque control scheme can be developed. Let us assume that the force between the wheel and the chassis (F_R and F_L in Equation 26.19) is measured by strain gauges or force sensors. In such cases, control can be formulated as a PI control on the force-error signal:

$$\begin{aligned} \tau_R &= k_F \frac{1}{r}(F_{Rd} - F_R) + k_{FI}\int (F_{Rd} - F_R)dt \\ \tau_L &= k_F \frac{1}{r}(F_{Ld} - F_L) + k_{FI}\int (F_{Ld} - F_L)dt \end{aligned} \tag{26.31}$$

By tuning the gains k_F and k_{FI} the measured force will more or less follow the desired one (F_{Rd} and F_{Ld}). The error becomes smaller with higher gains, but stability will be affected at high gains due to the presence of noise. Assuming the prefect control of forces F_R and F_L, we obtain a model of the mobile platform as follows:

$$m\ddot{x} - (F_R + F_L)\cos\phi = 0$$

$$m\ddot{y} - (F_R + F_L)\sin\phi = 0$$

$$I\ddot{\phi} - \frac{l}{2}(F_R - F_L) = 0$$

© 2002 by CRC Press LLC

Here we can assume that F_R and F_L are control inputs. Or, alternatively, we can introduce the following control inputs

$$u_1 = F_R + F_L$$

$$u_2 = F_R - F_L$$

so that the model becomes

$$\begin{aligned} m\ddot{x} &= u_1 \cos\phi \\ m\ddot{y} &= u_1 \sin\phi \\ I\ddot{\phi} &= \frac{l}{2} u_2 \end{aligned} \tag{26.32}$$

This is a second-order dynamic model of the platform assuming that all damping and friction forces are compensated by the use of force-feedback. Although relatively simple, this model exhibits nonholonomic properties. This can be illustrated as follows: assume the system is in the state (x, y, ϕ) = (0,0,0). The model of the chassis in this position and orientation becomes

$$\begin{aligned} m\ddot{x} &= u_1 \\ m\ddot{y} &= 0 \\ I\ddot{\phi} &= \frac{l}{2} u_2 \end{aligned}$$

indicating that the system is impossible to control in the y-direction from the given state. Or, in other words, the platform behaves as a singular system in the y-direction.

This system Equation (26.32), as well as the original system Equation (26.19), has very important properties that are elaborated by Zheng.[4]

1. The system (Equations 26.32 and 26.19) is a nonholonomic controllable system.
2. The system (Equations 26.32 and 26.19) cannot be made asymptotically stable by a smooth-state feedback.

The term "controllable" refers to the following: if a system can be transferred from any state to any other state by finite control signals in a finite time, it is a controllable system. With linear systems it would automatically imply the existence of smooth feedback that guarantees asymptotic stability. This does not hold for nonlinear systems as stated by the second property.

The developed control scheme is not used in practice. The control scheme based on local-velocity feedback loops is used much more often. A tachometer exists on every wheel to measure the speed of rotation of the wheel with such vehicles. The right-hand side wheel dynamics and control are then described by

$$\begin{aligned} F_R r + I_W \ddot{\theta}_R &= \tau_R \\ \tau_R + \tau_f &= k_v(\dot{\theta}_{Rd} - \dot{\theta}_R) \end{aligned} \tag{26.33}$$

© 2002 by CRC Press LLC

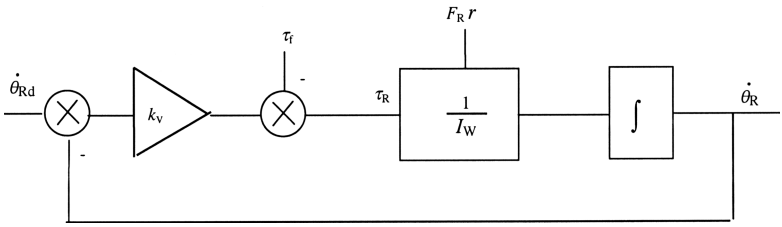


FIGURE 26.11 Control diagram based (Equation 26.33).

where k_v is the velocity feedback constant. The same model holds for the left wheel (the subscript R should be replaced by L). The control diagram based Equation (26.33) is shown in Figure 26.11.

Note that in this control scheme the friction torque and the torque due to the presence of driving force F_R act as a disturbance to the system. The goal of such control is to maintain the wheel velocity as close as possible to the desired velocity. The feedback gain k_v is usually tuned to a large value so that the velocity error

$$e = \dot{\theta}_{Rd} - \dot{\theta}_R$$

is small, while control is still stable. Here, the joystick control output equals

$$u_c = \dot{\theta}_{Rd}$$

Obviously, fine commands (small joystick increments) will be amplified by the gain k_v so that the wheel driver will be able to produce motion. This crucial and fine motion of the chassis is required from any static position. Assume, for example, that the vehicle is at rest and that the rotation by an angle ϕ of only one degree is required. This is usually impossible by the direct control scheme, and becomes possible with the velocity control loop. This is clear from Equation (26.20):

$$\dot{\phi} = \frac{r}{l}(\dot{\theta}_R - \dot{\theta}_L)$$

which shows that the better speed control in the wheels, the better overall turning ability of the chassis.

Having high-velocity gains in wheel controllers has another important implication on the behavior of the overall system — it simplifies the system's behavior if we assume that the real wheel velocity is equal to the desired (commanded) one. In this case, the above equation becomes

$$\dot{\phi} = \frac{r}{l}(u_R - u_L) \tag{26.34}$$

where u_R and u_L are left- and right-wheel control inputs. From (Equations 26.22 and 26.23) we obtain:

$$\dot{x} = \frac{r}{2}(\cos\phi)(u_R + u_L) \tag{26.35}$$

and

$$\dot{y} = \frac{r}{2}(\sin\phi)(u_R + u_L) \tag{26.36}$$

© 2002 by CRC Press LLC

The last three equations constitute the new model of the system. We can also introduce the control signals u_1 and u_2 instead of u_R and u_L such that

$$u_1 = u_R + u_L$$

$$u_2 = u_R - u_L$$

Now, the system model becomes

$$\begin{aligned} \dot{x} &= \frac{r}{2} u_1 \cos\phi \\ \dot{y} &= \frac{r}{2} u_1 \sin\phi \\ \dot{\phi} &= \frac{r}{l} u_2 \end{aligned} \tag{26.37}$$

We see that this model has a very similar form to Equation (26.32). The difference is in the order of the model: Equation (26.37) is a first-order model, while Equation (26.32) is a second-order one. Properties one and two listed above for the model Equation (26.32) hold also for model Equation (26.37). Another interesting property that also holds for Equations (26.37 and 26.32) is that the system is not full-state linearizable via static feedback loop. This means that no such (nonlinear) feedback loop can transform the system into a linear one. We recall that with manipulator models it is possible to introduce a nonlinear feedback loop that can fully linearize the system. Let us repeat, for clarity — if the dynamic model of a manpulator is

$$H\ddot{q} + h = u$$

then the nonlinear control

$$u = H(\ddot{q}^0 + k_v(\dot{q}^0 - \dot{q}) + k_p(q^0 - q)) + h$$

generates the error equation

$$(\ddot{q}^0 - \ddot{q}) + k_v(\dot{q}^0 - \dot{q}) + k_p(q^0 - q) = 0$$

so that the overall system becomes linear. As stated above, this is impossible with the models (Equations 26.32 and 26.37).

From the discussion above it is clear that no simple linear control can stabilize the system Equation (26.37). A number of different nonlinear control laws have been proposed in the literature, for example, tracking control, path following, stabilization about a desired posture, etc.

Let us illustrate control strategy in an example of tracking control. For example, tracking control is based on the reference model (vehicle)

$$\begin{aligned} \dot{x}_r &= \frac{r}{2} u_{1r} \cos\phi_r \\ \dot{y}_r &= \frac{r}{2} u_{1r} \sin\phi_r \\ \dot{\phi}_r &= \frac{r}{l} u_{2r} \end{aligned}$$

© 2002 by CRC Press LLC

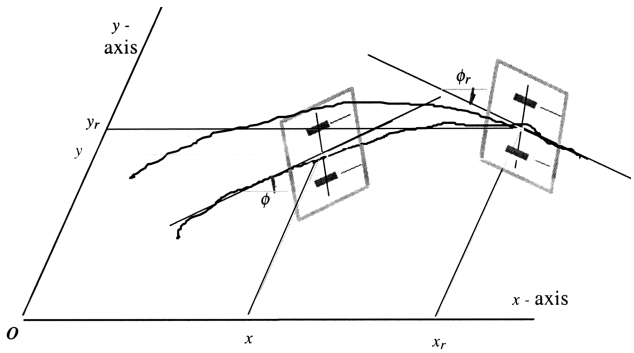


FIGURE 26.12 Tracking control illustration.

that has bounded control inputs that do not tend to zero when t approaches infinity. The control goal is to achieve zero asymptotic error in state difference (Δx, Δy, $\Delta\phi$) between the real and the reference model when $t \rightarrow$ infinity.

The control

$$u_1 = u_{1r}\cos(\phi_r - \phi) + \frac{2k_1}{r}e_1$$

$$u_2 = u_{2r} + \frac{k_2 l}{2}u_{1r}\frac{\sin(\phi_r - \phi)}{\phi_r - \phi}e_2 + \frac{k_3 l}{r}(\phi_r - \phi)$$

where

$$e_1 = (x_r - x)\cos\phi + (y_r - y)\sin\phi$$

$$e_2 = -(x_r - x)\sin\phi + (y_r - y)\cos\phi$$

and k_i are positive constants (i = 1, 2, 3).

It was shown[4] that this control globally asymptotically stabilizes the system so that the error in (Δx, Δy, $\Delta\phi$) tends to zero with time. As a consequence, the platform will follow the reference one and the error will tend toward zero with time. The proof is based on a suitable Liapunov function, which is nonincreasing along any system solution. An illustration of the tracking control is given in Figure 26.12.

Similar control laws have been developed for stabilization about a point and path-following problem. As a result, we see that the control problem is nonlinear and by no means a straightforward application of simple control theory.

There are many other practical issues related to the control of mobile robots. First is the sensorial system that can provide a good estimate of the platform position and orientation. Ultrasonic, infrared, laser-based, and camera-based sensors usually do this. Most of these sensors are used to detect a distance from an obstacle, and give information on the relative position of the vehicle with respect to the environment. Absolute coordinates are possible to get through GPS, which uses information on the geographic position of the robot obtained from a satellite by a radio link. One or more processors or microcontrollers process the sensor signals. The processors provide communication and control functions.

© 2002 by CRC Press LLC

References

1. Moravec, P.H., The Stanford Cart and the CMU Rover, in *Autonomous Robot Vehicles,* Cox, I.J. and Wilfong, G.T., Eds., Springer-Verlag, New York, 1990, 407–419.
2. Wang F.-Y. and Lever, P.J.A., An intelligent robotic vehicle for lunar and Martian resource assessment, in *Recent Trends in Mobile Robots,* Zheng, Y.F., Ed., World Scientific, Singapore, 293–313.
3. Campion, G., d'Andrea-Novel, B., and Bastin, G., Controllability and state feedback stabilization of nonholonomic mechanical systems, in *Lecture Notes in Control and Information Science,* de Wit, C.C., Ed., Springer-Verlag, New York, 1991, 106–124.
4. Zheng, Y.F., *Recent Trends in Mobile Robots,* World Scientific, Singapore, 1993.

© 2002 by CRC Press LLC

27

Humanoid Robots

27.1 Zero-Moment Point — Proper Interpretation
Introduction • The ZMP Notion • The Difference between ZMP and the Center of Pressure (CoP)

27.2 Modeling of Biped Dynamics and Gait Synthesis
Single-Support Phase • Double-Support Phase • Biped Dynamics • Example

27.3 Control Synthesis for Biped Gait
Synthesis of Control with Limited Accelerations • Synthesis of Global Control with Respect to ZMP Position • Example

27.4 Dynamic Stability Analysis of Biped Gait
Modeling of Composite Subsystems • Stability Analysis • Example

27.5 Realization of Anthropomorphic Mechanisms and Humanoid Robots
Active Exoskeletons • Humanoid Robots • Virtual Humanoid Robot Platform • New Application of the ZMP Concept in Human Gait Restoration

27.6 Conclusion

Miomir Vukobratović
Mihajlo Pupin Institute

Branislav Borovac
University of Novi Sad

Dragoljub Šurdilović
Fraunhofer Institute

Dragan Stokić
ATB Institute

During the last several years, significant stagnation has been observed in the development and application of industrial robots. The reason for this lies in the fact that in the last two or three decades a majority of simpler jobs in different industries and in workplaces presenting hostile environments to humans have been robotized.

We are now in an era of specialized, unconventional robots dedicated to complex tasks to be performed under specific and hazardous conditions. These robots are endowed with the elements of artificial intelligence. The objective is to initiate long-term multidisciplinary research with the goal of designing function-oriented devices equipped with proper onboard intelligence capable of autonomously performing common human work.

It is expected that the new generation of robots will yield explosive development that will have an impact comparable to that we witnessed with the appearance of personal computers. The present service robots will be replaced by personal robots. From the scientific point of view they will represent the continuation of the earlier research on anthropomorphic robots (now called humanoid robots) that are endowed with elements of artificial intelligence. The future service robots will work in the environment of humans, which imposes the requirement for human-like behavior in the area of motion, intelligence, and communication. Obviously, modeling, control, and design of such robots must be based on a wide range of disciplines such as system theory, artificial intelligence, material science, mechanics, and even biomechanics, physiology, and neuroscience.

© 2002 by CRC Press LLC

27.1 Zero-Moment Point — Proper Interpretation

27.1.1 Introduction

Biped locomotion has been at the focus of researchers for decades. Theoretical studies have been followed by simulations and realizations — from the simplest cases of planar mechanisms to humanoid robots, which are the most complex locomotion mechanisms constructed to date. Irrespective of their structures and complexities, the basic characteristics of all locomotion systems are: (1) the presence of unpowered degrees of freedom (DOFs) formed by the contact of the foot with the ground surface, (2) gait repeatability (symmetry), and (3) regular interchangeability of the number of legs that are simultaneously in contact with the ground. Two different situations arise in sequence during walking: the statically stable double-support phase in which the mechanism is supported on both feet simultaneously, and the statically unstable single-support phase when only one foot of the mechanism is in contact with the ground and the other is transferred from the back to front position. Thus, the locomotion mechanism changes its structure from an open to a closed kinematic chain during a single walking cycle. All these circumstances have to be taken into account in gait synthesis.

All of the biped mechanism joints are powered and directly controllable except for the joint formed by contact of the foot and the ground. This contact is essential for walking because this is the only point at which the mechanism interacts with the environment, and the mechanism's position relative to the environment depends on the regularity of its behavior. It is often called unpowered DOF because in case of an improper motion, the mechanism as a whole would start to rotate about the foot edge, and a new unpowered joint would appear. If such improper foot behavior occurred, the position of the entire mechanism relative to the environment would be jeopardized and the mechanism would overturn.

Foot behavior cannot be controlled directly; it is controlled in an indirect way by ensuring appropriate dynamics of the mechanism above the foot. Thus, the overall indicator of mechanism behavior is the ground reaction force: its intensity, direction, and particularly its acting point, termed the *zero-moment point* (ZMP).[1-5] Recognition of the significance and role of ZMP in the biped artificial walk was a turning point in gait planning and control. The methods for gait synthesis (semi-inverse method) were proposed in the two seminal works,[1-2] and for a long time they remained the only mechanisms for biped gait synthesis. Recently, another method has been reported,[6] which, among other criteria, takes into account the overall gait indicator: the ZMP position.

The concept of ZMP has recently found practical applications in humanoid, biped, and multi-legged robots. Numerous studies addressed the mathematical formalisms for computing the ZMP. Several algorithms for biped control and monitoring based on the ZMP concept have been proposed (e.g., Inoue et al.,[7] Huang et al.,[8] Yagi and Lumelsky.[9] Fujimoto et al.,[10] and Fukuda et al.[11]). As demonstrated recently,[12] the ZMP is also convenient for the analysis and control of human gait in rehabilitation robotics. The ZMP concept provides a useful dynamic criterion for the analysis and synthesis of human/humanoid robot locomotion. The ZMP indicates gait balance during the entire gait cycle and provides a quantitative measure for the unbalanced moment about the support foot and for the robustness (balancing margin) of the dynamic gait equilibrium.

27.1.2 The ZMP Notion

First, we would like to clarify the notion and, accordingly, the name of ZMP. Let us consider the single-support phase as shown in Figure 27.1, i.e., when only one foot is in contact with the ground (stance leg) while the other is in the swing phase, passing from the back to the front position. To maintain the mechanism's dynamic equilibrium, the ground reaction force $\bar{\mathbf{R}}$ should act at the appropriate point on the foot sole to balance all the forces acting on the mechanism during motion (inertial, gravitational, Coriolis, and centrifugal forces, and the corresponding moments), as shown in Figure 27.1.

© 2002 by CRC Press LLC

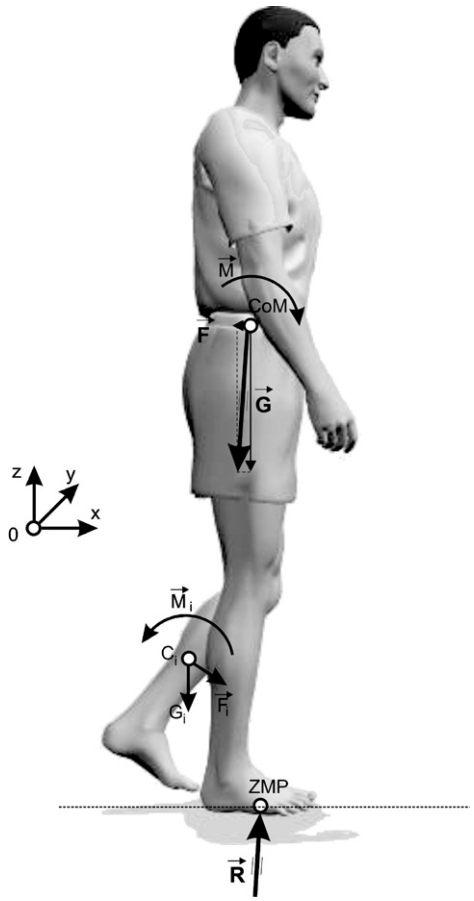


FIGURE 27.1 Single-support phase.

If we place the coordinate system at the point where $\vec{\mathbf{R}}$ is acting (let us assume that this point is under the foot), it is clear from the equilibrium conditions that the moments acting about the horizontal axes x and y will always be equal to zero, i.e., $M_x = 0$ and $M_y = 0$. The only moment component that may exist is M_z. It is realistic to assume that the friction coefficient between the ground and the foot is high enough and that M_z is balanced by friction forces. Thus, M_z will not cause foot motion and change in foot dynamics, and will not influence behavior of the mechanism above the foot. Since both moments relevant to the gait continuation (M_x and M_y) equal zero, a natural name for the ground reaction force acting at this point will be the zero moment point. Any change in the locomotion dynamics will change the vector of the ground reaction force, causing simultaneous changes in its direction, intensity, and acting point (ZMP). The following basic ZMP definition[1,2,13] reflects the above consideration:

Definition 1 (ZMP): The pressure under the supporting foot can be replaced by the appropriate reaction force acting at a certain point of the mechanism's foot. Since the sum of all moments of active forces with respect to this point is equal to zero, it is termed the zero-moment point (ZMP).

To define ZMP in a mathematical form, let us consider the dynamic model of the human, humanoid, or biped robot (the following analysis can be applied to all these systems). The human/humanoid dynamics will be modeled using the multibody system model consisting of N chains involving the body parts (head, arms, legs, trunk, and pelvis). Each chain consists of n_i-links (i = 1,…,N) interconnected with single DOF joints (the multiple DOF joints are decomposed into

© 2002 by CRC Press LLC

the single ones). For the sake of simplicity, let us consider the rigid body model that is a relatively good approximation of the humanoid dynamics, though it represents a very idealized model of the human gait. The multi-DOF structures of the human locomotion mechanism, joint flexibility, and structural and behavioral complexity of the foot support the realization of dynamic gait patterns that are difficult to achieve with the existing humanoid systems.

During locomotion the following active motion forces act on the body links:

$\vec{\mathbf{G}}_i$ = Gravitation force of the i-th link acting at the mass center C_i

$\vec{\mathbf{F}}_i$ = Inertial force of the i-th link acting at the mass center C_i

$\vec{\mathbf{M}}_i$ = Moment of the inertial force of the i-th link for C_i

$\vec{\mathbf{R}}$ = Resultant ground reaction force

All active motion forces (gravitational and inertial forces and moments) can be replaced by main resultant gravitation and inertial force and, in most cases, resultant inertial moment reduced at body center of mass (CoM). The ground reaction force and moment can be decomposed into vertical and horizontal components with respect to the reference frame. The horizontal reaction force represents the friction force essential for preserving the contact between the foot and the ground. The vertical reaction moment represents the moment of the friction reaction forces reduced at an arbitrary point P. We will assume a stable foot–floor contact without sliding. This means that the static friction forces compensate for the corresponding dynamic body reaction forces. Accordingly, the vertical reaction force and horizontal reaction moment components represent the dynamic reaction forces that are not compensated by the friction. The decomposition will be presented in the following form:

$$\begin{aligned}\vec{\mathbf{R}} &= \vec{\mathbf{R}}_v + \vec{\mathbf{R}}_f \\ \vec{\mathbf{M}} &= \vec{\mathbf{M}}_h + \vec{\mathbf{M}}_f\end{aligned} \tag{27.1}$$

where the indices h and v denote the horizontal and vertical components respectively, while f indicates the friction reaction force and moment components. Let us select the ZMP as the reduction point of interest, i.e., $P = ZMP$. Then the following equations express the dynamic equilibrium during the motion in the reference coordinate system:

$$\begin{aligned}&\vec{\mathbf{R}}_v + \vec{\mathbf{R}}_f + \sum_{j=1}^{N}\sum_{i=1}^{n_j}\left(\vec{\mathbf{F}}_i + \vec{\mathbf{G}}_i\right) = 0 \\ &\overrightarrow{OZMP} \times \vec{\mathbf{R}} + \sum_{j=1}^{N}\sum_{i=1}^{n_j}\overrightarrow{OC_i} \times \left(\vec{\mathbf{F}}_i + \vec{\mathbf{G}}_i\right) + \sum_{j=1}^{N}\sum_{i=1}^{n_j}\vec{\mathbf{M}}_i + \vec{\mathbf{M}}_{hZMP} + \vec{\mathbf{M}}_{fZMP} = 0\end{aligned} \tag{27.2}$$

where O denotes the origin of the reference frame (Figure 27.1). Then, based on the ZMP definition we have:

$$\vec{\mathbf{M}}_{hZMP} = 0\,. \tag{27.3}$$

Substituting the relation:

$$\overrightarrow{OC_i} = \overrightarrow{OZMP} + \overrightarrow{ZMPC_i} \tag{27.4}$$

into the second equation of Equation (27.2) and taking into account the first equation of (27.2) gives:

© 2002 by CRC Press LLC

$$\sum_{j=1}^{N}\sum_{i=1}^{n_j}\overrightarrow{ZMPC_i}\times\left(\vec{\mathbf{F}}_i+\vec{\mathbf{G}}_i\right)+\sum_{j=1}^{N}\sum_{i=1}^{n_j}\vec{\mathbf{M}}_i+\vec{\mathbf{M}}_{fZMP}=0\,. \tag{27.5}$$

Considering only the dynamic moment equilibrium in the horizontal ground plane (i.e., the moments that are not compensated by friction), we can write:

$$\left(\sum_{j=1}^{N}\sum_{i=1}^{n_j}\overrightarrow{ZMPC_i}\times\left(\vec{\mathbf{F}}_i+\vec{\mathbf{G}}_i\right)+\sum_{j=1}^{N}\sum_{i=1}^{n_j}\vec{\mathbf{M}}_i\right)_h=0\,. \tag{27.6}$$

Substituting Equation (27.4) in Equation (27.6) yields:

$$\left(\overrightarrow{OZMP}\times\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left(\vec{\mathbf{F}}_i+\vec{\mathbf{G}}_i\right)\right)_h=\left(\vec{\mathbf{R}}\times\overrightarrow{OZMP}\right)_h=\left(\sum_{j=1}^{N}\sum_{i=1}^{n_j}\overrightarrow{OC_i}\times\left(\vec{\mathbf{F}}_i+\vec{\mathbf{G}}_i\right)+\sum_{j=1}^{N}\sum_{i=1}^{n_j}\vec{\mathbf{M}}_i\right)_h \tag{27.7}$$

Equations (27.6) and (27.7) represent the mathematical interpretation of ZMP and provide the formalism for computing the ZMP coordinates in the horizontal ground plane.

The one-step cycle consists of the single- and double-support phases, taking place in sequence. A basic difference between these elemental motion phases is that during the motion in the single-support phase, the position of the free foot is not fixed relative to the ground. In the double-support phase, the positions of both feet are fixed. From the ZMP point of view, the situation is identical. In both cases, ZMP should remain within the support polygon in order to maintain balance. During the gait (let us call it *balanced gait* to distinguish it from the situation when equilibrium of the system is jeopardized and the mechanism collapses by rotating about the support polygon edge), the ground reaction force acting point can move only within the support polygon. The gait is balanced when and only when the ZMP trajectory remains within the support area. In this case, the system dynamics is perfectly balanced by the ground reaction force and overturning will not occur. In the single-support phase, the support polygon is identical to the foot surface. In the double-support phase, however, the size of the support polygon is defined by the size of the foot surface and by the distance between them (the convex hulls of the two supporting feet).

This ZMP concept is primarily related to the gait dynamics; however it can also be applied to consider static equilibrium when the robot maintains a certain posture. The only difference is in the forces inducing the ground reaction force vector. In the static case, there is only the mechanism weight, while the gait also involves dynamic forces. Accordingly, when equilibrium of a static posture (the mechanism is frozen in a certain posture and no gait is performed) is considered, the vertical projection of total active force acting at the mass center must be within the support polygon. This is a well-known condition for static equilibrium.

27.1.3 The Difference between ZMP and the Center of Pressure (CoP)

One can see from the above analysis that ZMP is apparently equivalent to the center of pressure (CoP), representing the application point of the ground reaction forces (GRFs). The CoP can be defined as:

Definition 2 (CoP): CoP represents the point on the support foot polygon at which the resultant of distributed foot ground reaction forces acts.

The CoP is commonly used in human gait analysis based on force platform or pressure mat measurements. In human locomotion, the CoP changes during the stance phase, generally moving from the heel toward a point between the first and second metatarsal heads. It is relatively simple to demonstrate that in the considered single-support phase and for balanced dynamic gait equilibrium

© 2002 by CRC Press LLC

(Figure 27.1), the ZMP coincides with the CoP. Let us again consider the equilibrium (Equation (27.2)) assuming that CoP is the reduction point P = CoP and ZMP and CoP do not coincide. According to the adopted notation, the force and moment reduced at CoP are denoted as $-\vec{\mathbf{R}}$ and $-\vec{\mathbf{M}}_{CoP}$ respectively, while the reaction force and moment are $\vec{\mathbf{R}}$ and $\vec{\mathbf{M}}_{CoP}$. Consider the equilibrium of the foot reaction forces, supposing that ZMP does not coincide with CoP. For this case we can write:

$$\left(\overrightarrow{ZMPCoP} \times \vec{\mathbf{R}} + \vec{\mathbf{M}}_{CoP}\right)_h = 0 \tag{27.8}$$

However, on the basis of CoP definition for the balanced gait, we have:

$$\left(\vec{\mathbf{M}}_{CoP}\right)_h = 0 \tag{27.9}$$

which can only be satisfied if:

$$\overrightarrow{ZMPCoP} = 0 \tag{27.10}$$

and it follows that $ZMP \equiv CoP$.

Let us discuss the justification of introducing a new term (ZMP) for a notion that has already been known in technical practice (CoP). While CoP is a general term encountered in many technical branches (e.g., fluid dynamics), ZMP expresses the essence of this point that is used exclusively for gait synthesis and control in the field of biped locomotion. It reflects much more clearly the nature of locomotion. For example, in the biped design we can compute ZMP on the assumption that the support polygon is large enough to encompass the calculated acting point of the ground reaction force. Then we can determine the form and dimension of the foot-supporting area encompassing all ZMP points or, if needed, we can change the biped dynamic parameters or synthesize the nominal gait and control the biped to constantly keep ZMP within the support polygon.

Furthermore, the ZMP has a more specific meaning than CoP in evaluating the dynamics of gait equilibrium. To show the difference between ZMP and CoP, let us consider the dynamically unbalanced single-support situation (the mechanism as a whole rotates about the foot edge and overturns) illustrated in Figure 27.2, which is characterized by a moment about CoP that could not be balanced by the sole reaction forces. The reaction moment that can be generated between the foot and the ground is limited due to the unilateral contact between each sole and the floor. The intensity of balancing moments depends on the foot dimension. Obviously, it is easier for a person with larger sole to balance the gait. The dynamic motion moments in specific cases may exceed the limit, causing the foot to leave the ground. In spite of the existence of a nonzero supporting area (soft human/humanoid foot), reaction forces cannot balance the system in such a case. The way in which this situation in human/humanoid gait can occur will be considered later. As is clear from Figure 27.2, the CoP and the ZMP do not coincide in this case. Using an analogy to fluid dynamics, we could determine CoP as the center of pressure distribution (e.g., obtained by a pressure mate). It should be mentioned that in regular human gait, in a dynamic transition phase (e.g., heel strike and toe off), it is difficult to estimate CoP on the basis of force plate measurements.

However, ZMP, even in the case illustrated in Figure 27.2, can be uniquely determined on the basis of its definition. Assuming that both reaction force and unbalanced moment are known, we can mathematically replace the force–moment pair with a pure force displaced from the CoP. In this situation, however, the ZMP and the assigned reaction force have a pure mathematical/mechanical meaning (obviously, the ZMP does not coincide with the CoP) and the ZMP does not represent a physical point. However, the ZMP location outside the support area (determined by the vector $\vec{\mathbf{r}}$ in Figure 27.2) provides very useful information for gait balancing. The fact that ZMP is instantaneously on the edge or has left the support polygon indicates the occurrence of an unbalanced

© 2002 by CRC Press LLC

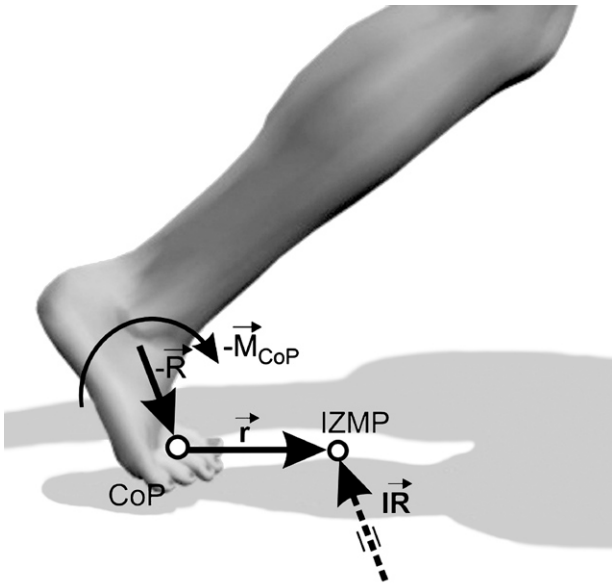


FIGURE 27.2 Action/reaction forces at CoP and ZMP (irregular case).

moment that cannot be compensated for by foot reaction forces. The distance of ZMP from the foot edge provides the measure for the unbalanced moment that tends to rotate the human/humanoid around the supporting foot and, possibly, to cause a fall. When the system encounters such a hazardous situation, it is still possible by means of a proper dynamic corrective action of the biped control system to bring ZMP into the area where equilibrium is preserved. To avoid this, a fast rebalancing by muscles or actuator actions (change of dynamic forces acting on the body) is needed. Several approaches to realization of this action have been discussed.[13]

On the basis of the above discussion, it is obvious that generally the ZMP does not coincide with the CoP

$$ZMP \neq CoP. \tag{27.11}$$

The CoP may never leave the support polygon. However, the ZMP, even in the single-support gait phase, can leave the polygon of the supporting foot when the gait is not dynamically balanced by foot reaction forces, e.g., in the case of a nonregular gait (even in the case of a degenerative gait). Hence, ZMP provides a more convenient dynamic criterion for gait analysis and synthesis.

The ZMP outside the support polygon indicates an unbalanced (irregular) gait and does not represent a physical point related to the sole mechanism. It can be referred to as imaginary ZMP (IZMP). Three characteristic cases for the nonrigid foot in contact with the ground floor, sketched in Figure 27.3, can be distinguished. In the so-called regular (balanced and repetitive) gait, the ZMP coincides with CoP (Figure 27.3a). If a disturbance brings the acting point of the ground reaction force to the foot edge, the perturbation moment will cause rotation of the complete biped locomotion system about the edge point (or a very narrow surface, under the assumption that the sole of the shoe is not fully rigid) and overturning (Figure 27.4). In that case we speak of IZMP, whose imaginary position depends on the intensity of the perturbation moment (Figure 27.3b). However, it is possible to realize the biped motion, for example, on the toe tips (Figure 27.3c) with special shoes having pinpoint areas (balletic locomotion), while keeping the ZMP position within the pinpoint area. Although it is not a regular (conventional, ordinary) gait, the ZMP also coincides with CoP in that case.

© 2002 by CRC Press LLC

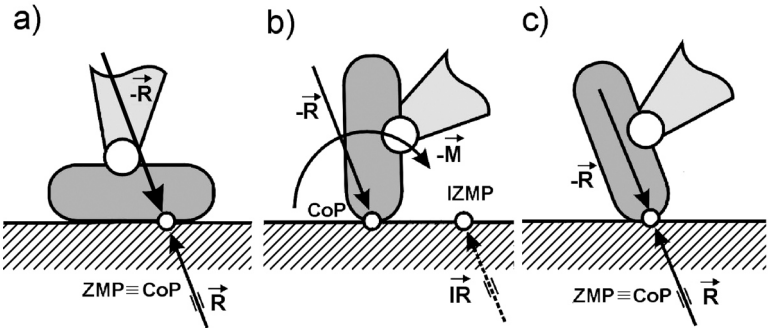


FIGURE 27.3 The possible relative positions of ZMP and CoP: (a) dynamically balanced gait, (b) unbalanced gait (the system rotates about the foot-edge and overturns), and (c) intentional foot-edge equilibrium (balletic locomotion).

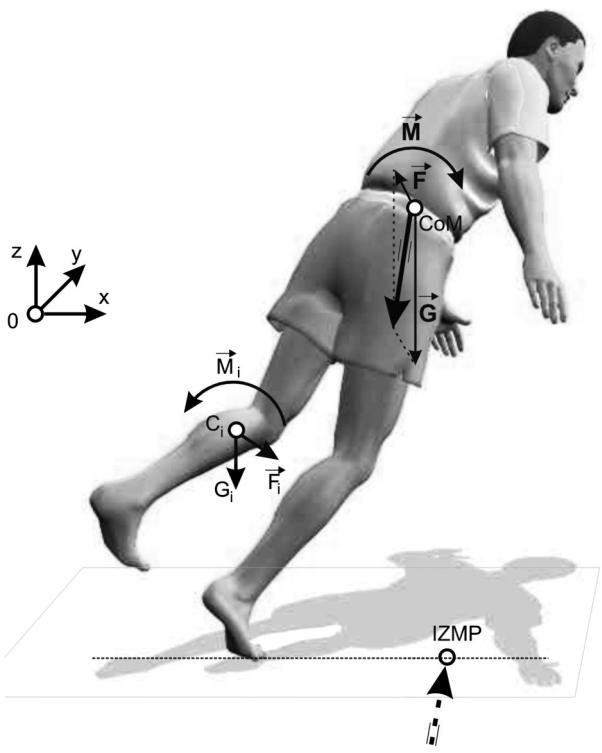


FIGURE 27.4 Imaginary ZMP in unbalanced human gait.

Because of foot elasticity and the complex form of the supporting area, the ZMP displacements outside the safe zone (Figure 27.2) in human locomotion are much more complex and difficult to model. Even in a regular human gait, ZMP leaves the support polygon dynamically during the transition from the single- to double-support phase, providing a smooth dynamic locomotion. The implementation of such gait patterns in humanoids with simple rigid feet is not practically possible.

In the double-support phase, and even more during transition from the single to the double phase, the ZMP leaves the foot-supporting polygon. Stable dynamic equilibrium in the double-support phase is characterized by the ZMP location within the enveloping polygon between the two feet.[13]

© 2002 by CRC Press LLC

The extent of ZMP dislocation from the enveloping polygon also provides a practical measure for the unbalanced moments. In previous works[13] our attention has mainly been focused on the problems of biped design and nominal motion synthesis, as well as stability analysis and biped dynamic control that will prevent the ZMP excursions close to the edges of the supporting polygon in spite of various disturbances and model uncertainties. Due to limitations of the sensory and control systems, the occurrence of a new unpowered joint (ZMP at the edges of the support polygon) has been considered as critical and undesirable in the past.

Hence, the situation when ZMP can arbitrarily be located in the foot plane was practical in designing the biped foot dimensions and nominal motion synthesis. When the ZMP approaches critical areas or even abandons the support polygon (Figure 27.3), balancing is focused primarily on compensating for the unbalanced dynamic moment using the posture control. One way of overcoming such critical situation is to switch to a new nominal trajectory that is closest to the momentary system state.[5] These nominals are synthesized to bring the system back to the stationary state and enable gait continuation. To do this, it is not necessary to have information about exact intensity of the disturbance moment. For such an approach (which is very close to the human behavior in similar situations), it suffices to detect the occurrence of such hazardous situations. Thus, there is no need for on-line computation of the IZMP location for the purpose of biped control. For these reasons the IZMP location has not gained more practical importance. However, the recent development of powerful control and sensory systems and the fast expansion of humanoid robots gives a new significance to the IZMP, particularly in rehabilitation robotics. The consideration of ZMP locations, including also the areas outside the supporting foot sole, becomes essential for rehabilitation devices.[12]

27.2 Modeling of Biped Dynamics and Gait Synthesis

The synthesis of the motion of humanoid robots requires realization of a human-like gait. There are several possible approaches, depending on the type of locomotion activity involved. It should be kept in mind that the skeletal activity of human beings is extremely complex and involves many automated motions. Hence, the synthesis of the artificial locomotion–manipulation motion has complexities related to the required degree of mimicking of the corresponding human skeletal activity.

If, however, we concentrate on the synthesis of a regular (repeatable) gait, then it is natural to copy the trajectories of the natural gait and impose them onto the artificial (humanoid) system. Of course, the transfer of trajectories (in this case of the lower limbs) from a natural to an artificial system can be realized with a higher or lower degree of similarity to the human gait. Hence, the anthropomorphism of artificial gait represents a serious problem. To explain the practical approach to solving this problem, let us assume we have adopted one of the possible gait patterns. By combining the adopted (prescribed) trajectories of the lower limb joints (method of prescribed synergy[1,2,5,13]) and the position (trajectory) of ZMP using the semi-inverse method,[2,5,13] it is possible to determine the compensation motion of the humanoid robot from the moments about the corresponding axes for the desired position of the ZMP (or ZMP trajectory). The equilibrium conditions can be written also for the arm joints. In fact, the unpowered arm joints represent additional points where moments are known (zero). These supplementary moment equations about the unpowered arm axes yield the possibility of including passive arms in the synthesis of dynamically balanced humanoid gait.

27.2.1 Single-Support Phase

Let us suppose the system is in single-support phase and the contact with the ground is realized by the full foot (Figure 27.5). It is possible to replace all vertical elementary reaction forces by the resultant force R_V. Only regular gait will be considered, and the ZMP position has to remain within the support area (polygon).

© 2002 by CRC Press LLC

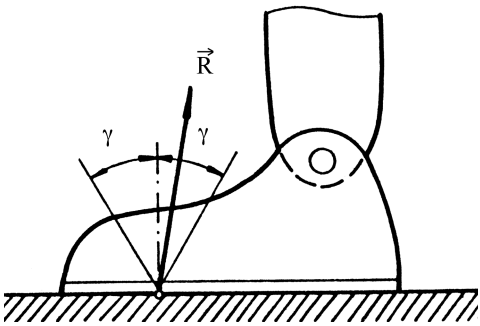


FIGURE 27.5 Longitudinal distribution of pressure on the foot and ZMP position.

The basic idea of artificial synergy synthesis is that the law of the change of total reaction force under foot is known in advance or prescribed. The prescribed segments of the dynamic characteristics which restrict the system in a dynamic sense are called dynamic connections. If a certain point represents the ZMP and the ground reaction forces $\vec{R}_V$ is reduced to it, then the moment $\vec{M}$ should be equal to zero. The vector $\vec{M}$ always has a horizontal direction and, hence, two dynamic conditions have to be satisfied: the projection of the moment on the two mutually orthogonal axes X and Y in the horizontal plane should be equal to zero.

$$M_X = 0 \quad M_Y = 0 \tag{27.12}$$

As far as friction forces are concerned, it is a realistic assumption that the friction coefficient is sufficiently large to prevent slippage of the foot over the ground surface. Thus, it can be stated that their moment with respect to the vertical axis V is equal to zero:

$$M_V = 0\,. \tag{27.13}$$

The axis V can be chosen to be in any place, but if it passes through the ZMP, then the axes X, Y, and V constitute an orthogonal coordinate frame, and V will be denoted by Z. The external forces acting on the locomotion system are the gravity, friction, and ground reaction forces. Let us reduce the inertial forces and moments of inertial forces of all the links to the ZMP and denote them by $\vec{F}$ and $\vec{M}_F$, respectively. The system equilibrium conditions can be derived using D'Alembert's principle and conditions (27.12) can be rewritten as:

$$(\vec{M}_G + \vec{M}_F)\cdot\vec{e}_X = 0\,, \quad (\vec{M}_G + \vec{M}_F)\cdot\vec{e}_Y = 0 \tag{27.14}$$

where $\vec{M}_G$ is the total moment of gravity forces with respect to ZMP, while $\vec{e}_X$ and $\vec{e}_Y$ are unit vectors of the x and y axes of the absolute coordinate frame. The third equation of dynamic connections, Equation (27.13), becomes:

$$(\vec{M}_F + \vec{\rho} \times \vec{F}\,)\cdot\vec{e}_V = 0 \tag{27.15}$$

where $\vec{\rho}$ is a vector from ZMP to the piercing point of the axis V through the ground surface; $\vec{e}_v$ is a unit vector of the axis V.

Let us adopt the relative angles between two links to be the generalized coordinates and denote them by q^i. Suppose the mechanism foot rests completely on the ground, so that the angle is zero, $q_0 \equiv 0$. The inertial force $\vec{F}$ and the moment $\vec{M}_F$, in general, can be represented in the linear forms of the generalized accelerations and quadratic forms of generalized velocities:

© 2002 by CRC Press LLC

$$F^k = \sum_{i=1}^{n} a_i^k \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}^k \cdot \dot{q}^i \dot{q}^j,\ \mathrm{k} = 1,\ 2,\ 3$$

$$M_F^k = \sum_{i=1}^{n} c_i^k \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^k \cdot \dot{q}^i \dot{q}^j,\ \mathrm{k} = 1,\ 2,\ 3 \tag{27.16}$$

where the coefficients $a_i^k, b_{ij}^k, c_i^k, d_{ij}^k$ (k = 1, 2, 3; and = 1, …, n; j = 1, … n) are the functions of the generalized coordinates, and F^k and M_F^k (k = 1, 2, 3) denote projections of the vectors $\vec{F}$ and $\vec{M}_F$ onto the coordinate axes. By introducing these expressions into Equations (27.14) and (27.15) one obtains:

$$\vec{M}_G \cdot \vec{e}_x + \sum_{i=1}^{n} c_i^1 \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^1 \cdot \dot{q}^i \dot{q}^j = 0$$

$$\vec{M}_G \cdot \vec{e}_y + \sum_{i=1}^{n} c_i^2 \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2 \cdot \dot{q}^i \dot{q}^j = 0 \tag{27.17}$$

$$\sum_{i=1}^{n} c_i^3 \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{i=j}^{n} d_{ij}^3 \cdot \dot{q}^i \dot{q}^j + \rho^x \left(\sum_{i=1}^{n} a_i^2 \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{i=1}^{n} b_{ij}^2 \cdot \dot{q}^i \dot{q}^i \right) - \rho^y \left(\sum_{i=1}^{n} a_i^1 \cdot \ddot{q}^i + \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}^1 \cdot \dot{q}^i \dot{q}^j \right) = 0$$

where the superscripts *x* and *y* denote the components in direction of the corresponding axis.

If the biped locomotion system has only three DOFs, the trajectories for all angles q^i can be computed from Equation (27.17). If the system has more than three DOFs (and this is actually the case), the trajectories for the rest (n-3) coordinates should be prescribed in such a way to ensure the desired legs trajectories (for example, measured from the human walk). The trajectories for this part of the system are prescribed, while the dynamics of the rest of the system (i.e., the trunk and arms) are determined in a such a way to ensure the dynamic balance of the overall mechanism.

The set of coordinates can be divided in two subsets: one containing all the coordinates whose motion is prescribed, denoted as q^{0i}, and the other comprising all the coordinates whose motion is to be defined using the semi-inverse method,[1,13] denoted as q^{xi}. Accordingly, the condition (27.17) becomes:

$$\sum_{i=1}^{n} c_i^k \cdot \ddot{q}^{xi} + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^k \cdot \dot{q}^{xi} \dot{q}^{xj} + g^k = 0,\ \mathrm{k} = 1,\ 2,\ 3 \tag{27.18}$$

where c_i^k and d_{ij}^k (k = 1, 2, 3) are the vector coefficients dependent on q^0 and q^x, whereas the vector g^k (k = 1, 2, 3) is a function of q^0, $\dot{q}^0$, $\ddot{q}^0$, and q^x. Since the gait is symmetric, the repeatability conditions can be written in the form:[1,13]

$$q^i(0) = \pm q^i\left(\frac{T}{2}\right),\quad \dot{q}^i(0) = \pm \dot{q}^i\left(\frac{T}{2}\right)$$

where the sign depends on the physical nature of the appropriate coordinates and their derivatives; $(T/2)$ is the duration of one half-step. As the motion of the prescribed part of the mechanism has been already defined (repeatability conditions are implicitly satisfied), the repeatability conditions:

$$q^{xi}(0) = \pm q^{xi}\left(\frac{T}{2}\right),\quad \dot{q}^{xi}(0) = \pm \dot{q}^{xi}\left(\frac{T}{2}\right) \tag{27.19}$$

© 2002 by CRC Press LLC

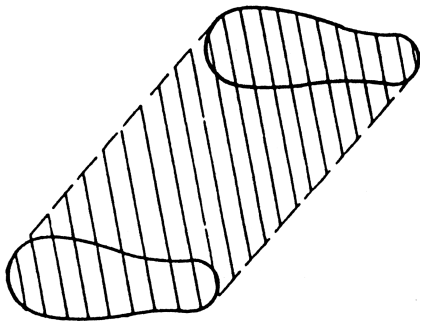


FIGURE 27.6 Double-support phase.

for the rest of the mechanism are to be added to the original set of equations describing the mechanism motion.

The system of Equation (27.18), together with the conditions (27.19), enables one to obtain the necessary trajectories of the coordinates q^{xi}, i.e., to carry out compensation synergy synthesis. After the synergy synthesis is completed, the driving torques that force the system to follow the nominal trajectories have to be computed.

27.2.2 Double-Support Phase

In the double-support phase, both mechanism feet are simultaneously in contact with the ground. The kinematic chain playing the role of the legs is closed, i.e., the unknown reaction forces to be determined act on both ends.

The procedure for the synergy synthesis is in the most part analogous to that for the single-support phase. Let the position of the axis V be selected within the dashed area in Figure 27.6. Then, by writing the equilibrium equations with respect to the three orthogonal axes passing through ZMP and setting the sum of all the moments of external forces to zero, the compensating movements for the corresponding part of the body can be computed.

The next problem is how to choose the position of the axis V with respect to the ZMP. The information on ZMP and axis V is insufficient for computation of the driving torques. For this reason, it is necessary to provide some additional relations concerning the ground reaction force. The total reaction force under one foot can be expressed as a sum of three reaction forces and moment components in the direction of coordinate axes. The components M_X and M_Y can be equal to zero since the vertical forces on the diagram are of the same sign. The third component M_V should also be equal to zero, according to the following consideration. Generally speaking, friction forces can produce moments, but in synergy synthesis, the moment M_V should also be equal to zero. Consequently, if the moments of friction forces are generated, they should be of the opposite sign under each foot. However, these moments do not affect the system motion but only load the leg drives and joints additionally. Because of that, it is reasonable to synthesize the gait in such a way to reduce each of these moments to zero. Thus it can be assumed that total moments of reaction forces under each foot are equal to zero:

$$\vec{M}_a = \vec{M}_b = 0 \tag{27.20}$$

where the subscripts a and b denote the left and right foot, respectively.

Characteristics of the friction between the foot and the ground can be represented by a friction cone (Figure 27.7). If the total ground reaction force $\vec{R}$ is within the cone of the angle 2γ, its horizontal component (i.e., friction force) will be of sufficient intensity to prevent an unwanted horizontal motion of the supporting foot over the ground surface. This can be expressed as:[13]

© 2002 by CRC Press LLC

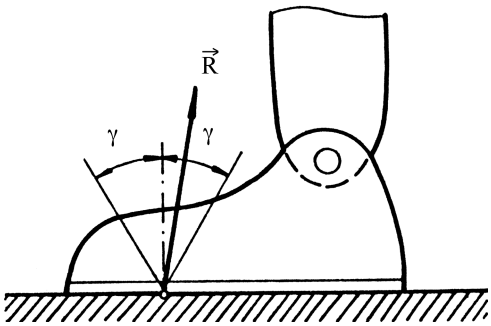


FIGURE 27.7 Friction cone.

$$\frac{\left|\vec{R}_x + \vec{R}_y\right|}{\left|\vec{R}_v\right|} \le tg\gamma = \mu \tag{27.21}$$

where μ is the friction coefficient of the surfaces in contact. Thus, it is reasonable to distribute the horizontal components of ground reaction forces per foot proportionally to the normal pressure. The vertical components are inversely proportional to the distances between the ZMP and the corresponding foot, so:

$$\frac{\left|\vec{R}_{v_a}\right|}{\left|\vec{R}_{v_b}\right|} = \frac{\ell_b}{\ell_a} \tag{27.22}$$

Then, from Equation (27.21) the relation:[4,13]

$$\frac{\left|\vec{T}_a\right|}{\left|\vec{T}_b\right|} = \frac{\ell_b}{\ell_a} \tag{27.23}$$

holds for the horizontal components, where $\vec{T}_a$ and $\vec{T}_b$ are the friction forces under the corresponding foot (Figure 27.8). On the basis of similarity of the triangles Δ OAD and Δ OBC, it can be concluded that the relation (27.23) does not depend on the direction of the force $\vec{T}$ (i.e., the distances ℓ'_a and ℓ'_b), but only on the distances between the feet, ℓ_a and ℓ_b. Thus, in order to have friction forces divided in proportion to the vertical pressures, a necessary and sufficient condition is that the axes ℓ'_a and ℓ'_b pass through the ZMP. Then, for synergy synthesis in the double-support phase, the following vector equation holds:

$$\sum_{i=1}^{n}\left(\vec{r}_i \times \left(\vec{G}_i + \vec{F}_i\right) + \vec{M}_i\right) = 0 \tag{27.24}$$

where $\vec{r}_i$ is a radius vector from the ZMP to the gravity center of the *i*-th link and $\vec{F}_i$ and $\vec{M}_i$ are the inertial force and corresponding moment of the *i*-th link reduced to its center of gravity.

When the synthesis of the compensating laws of motion is completed, it is possible to determine the total horizontal and vertical reactions:

$$R_Z = -\sum_{i=1}^{n}\left(\vec{F}_{iZ} + \vec{G}_i\right), \quad \vec{T} = \sum_{i=1}^{n}\left(\vec{F}_{iX}\cdot\vec{e}_X + \vec{F}_{iY}\cdot\vec{e}_Y\right) \tag{27.25}$$

© 2002 by CRC Press LLC

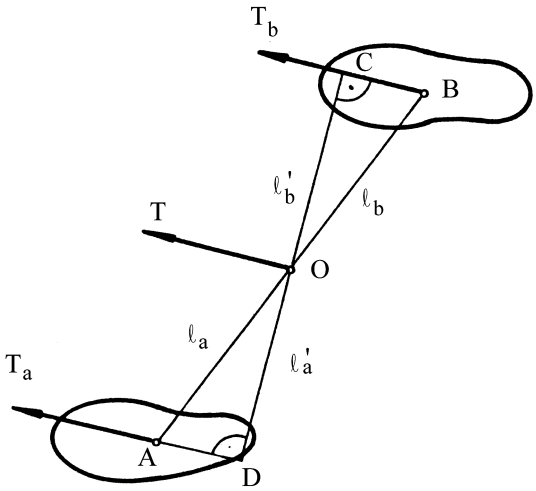


FIGURE 27.8 Determination of total friction force.

where $\vec{F}_{iZ}$ is the projection of $\vec{F}_i$ onto the vertical axis and $\vec{F}_{iX}$ and $\vec{F}_{iY}$ are projections onto the axes X and Y, respectively. Here, the axis Z corresponds to the vertical axis (previously denoted by V) passing through ZMP. The axes X, Y, and Z constitute the absolute orthogonal coordinate frame. Furthermore, the relations $\vec{T} = \vec{T}_a + \vec{T}_b$ and $\vec{R} = \vec{R}_a + \vec{R}_b$ are obvious, and they, together with the relations:

$$\vec{\ell}_a \times \vec{T}_a + \vec{\ell}_b \times \vec{T}_b = 0$$

$$\vec{\ell}_a \times \vec{R}_a + \vec{\ell}_b \times \vec{R}_b = 0 \tag{27.26}$$

extend the possibility of defining the vertical reactions $\vec{R}_a$ and $\vec{R}_b$, as well as the friction forces $\vec{T}_a$ and $\vec{T}_b$. The $\vec{\ell}_a$ and $\vec{\ell}_b$ are the vectors from the ZMP (denoted by 0) to the centers of the corresponding supporting surfaces A and B, respectively.

27.2.3 Biped Dynamics

The active spatial mechanism for realization of the artificial anthropomorphic gait belongs to the class of complex kinematic chains, as shown in Figure 27.9. During walking, the kinematic chain representing the legs changes its configuration from open to closed,[4,13] in the single- and double-support phases, respectively. Each phase involves a different procedure for forming dynamic equations, but it is based on the well-known procedure for dynamic modeling of simple open kinematic chains for robotic manipulators. This procedure enables us to obtain the following expression:[13,14]

$$P = H(q,\theta)\cdot\ddot{q} + h(q,\dot{q},\theta) \tag{27.27}$$

where $P = [P_1, \ldots, P_n]$ is a vector of driving moments at the joints, $H(q,\theta)$ is the $n \times n$ inertial matrix, $h(q,\dot{q},\theta)$ is the $n \times 1$ vector of Coriolis, centrifugal, and gravity forces, $q = [q_1, \ldots, q_n]$ is a vector of joint coordinates, and $\theta = [\theta_1, \ldots, \theta_n]$ is a geometric and dynamic parameter vector.

In the case of complex kinematic chains, the system has at least one link (branching link) belonging to more than two kinematic chains. Calculation of the elements of the matrix H and the vector h of complex kinematic chains can be carried out by introducing the corresponding number

© 2002 by CRC Press LLC

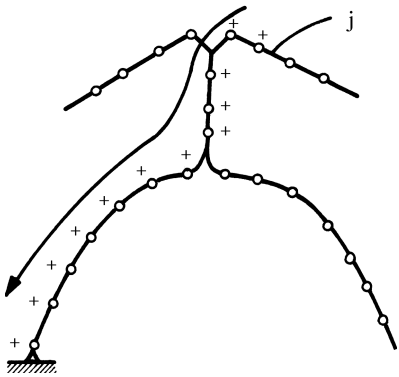


FIGURE 27.9 Complex kinematic chain.

of series of "+" joints.[4,13] A series of "+" joints is formed in such a way that in the case of chain rupture at a certain "+" joint, the *j*-th link should remain in the external part (not connected to the support) of the mechanism. In the course of forming differential equations, the inertial force and moment of the *j*-th mechanism link are reduced to its own "+" joint only, and the following procedure is possible: $\vec{F}_j^u$ and $\vec{M}_j^u$ (total external force and moment) corresponding to the *j*-th link are successively reduced to all "+" joints going from the *j*-th link toward the support. After projecting $\vec{F}_j^u$ and $\vec{M}_j^u$ onto the axis of the *i*-th joint, the resulting quantities denoted by ΔH_{ik}^j and Δh_i^j can be calculated in the following way:[4,13]

$$\Delta H_{ik}^j = -\vec{e}_i\left(\vec{b}_{jk} + \vec{r}_{ji} \times \vec{a}_{jk}\right), \quad \Delta h_i^j = -\vec{e}_i\left(\vec{r}_{ji} \times \left(\vec{a}_j^o + \vec{G}_j\right) + \vec{b}_j^o\right) \tag{27.28}$$

where $\vec{e}_i$ is the unit vector of the *i*-th joint axis, and $\vec{r}_{ji}$ is the radius vector from the *i*-th joint to the *j*-th link center of mass, while $\vec{a}$ and $\vec{b}$ are the corresponding vector coefficients. Note that the angular ($\vec{\varepsilon}_i$) and linear ($\vec{w}_i$) accelerations of the *i*-th link can be expressed as:

$$\vec{\varepsilon}_i = \left[\vec{\alpha}_{i1} \ldots \vec{\alpha}_{ii}\, 0 \ldots 0\right]\ddot{q} + \vec{\alpha}_i^0, \quad \vec{w}_i = \left[\vec{\beta}_{i1} \ldots \vec{\beta}_{ii}\, 0 \ldots 0\right]\ddot{q} + \vec{\beta}_i^0$$

where $\ddot{q} = \left[\ddot{q}^1, \ldots, \ddot{q}^n\right]$. The α_{ij} and β_{ij} are the vector coefficients depending on the generalized coordinates, while the vector coefficients $\vec{\alpha}_i^0$ and $\vec{\beta}_i^0$ depend on the generalized coordinates and velocities. Then, we have:[4,13]

$$\vec{a}_{ij} = -m_i \vec{\beta}_{ij}, \quad \vec{a}_i^0 = -m_i \vec{\beta}_i^0, \quad \vec{b}_{ij} = -T_i \vec{\alpha}_{ij}, \quad \vec{b}_j^0 = -T_i \vec{\alpha}_i^0 + \vec{\lambda}_i$$

with $T_i = \sum_{\ell=1}^{3} Q_{i\ell} J_{i\ell}$, $Q_{i\ell} = \left[q_{i\ell}^1 \vec{q}_{i\ell}^1 \;\; q_{i\ell}^2 \vec{q}_{i\ell}^2 \;\; q_{i\ell}^3 \vec{q}_{i\ell}^3\right]$, and $\vec{\lambda}_i = Q_i \begin{bmatrix} (\vec{\omega}_i \cdot \vec{q}_{i2})(\vec{\omega}_i \cdot \vec{q}_{i3})(J_{i2} - J_{i3}) \\ (\vec{\omega}_i \cdot \vec{q}_{i3})(\vec{\omega}_i \cdot \vec{q}_{i1})(J_{i3} - J_{i1}) \\ (\vec{\omega}_i \cdot \vec{q}_{i1})(\vec{\omega}_i \cdot \vec{q}_{i2})(J_{i1} - J_{i2}) \end{bmatrix}$

where $q_{i\ell}^j$ (j = 1, 2, 3) denotes the *j*-th component of the vector $\vec{q}_{i\ell}$, $\vec{q}_{i\ell}$ denotes the axes of the *i*-th (local) coordinate frame, i.e., the frame attached to the *i*-th link, and J_{ij} is the *j*-th component of the *i*-th link inertia tensor defined with respect to the local coordinate frame associated to each link of the kinematic chain.

The components of matrix *H* and vector *h* are obtained by summing the corresponding values from Equation (27.28) with respect to all series of "+" joints:

© 2002 by CRC Press LLC

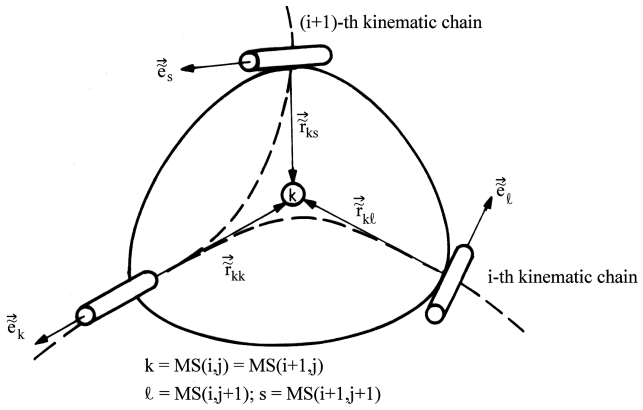


FIGURE 27.10 Branching link of a complex kinematic chain.

$$H_{ik}^{j} = \sum_{(j)} \Delta H_{ik}^{j}, \quad h_i^j = \sum_{(j)} \Delta h_i^j \tag{27.29}$$

Figure 27.10 illustrates a branching link that has three kinematic pairs and is a constituent of two series of "+" joints. The topological structure of the complex chain can be represented by the matrix MS. Each row of this matrix contains ordinal numbers of the corresponding series of "+" joints. The element MS(i,j) is the *j*-th joint in the *i*-th series of "+" joints. In addition, for each series of "+" joints, the ordinal number of the initial joint is also defined. The initial joint of the *i*-th series of "+" joints is the first joint of the *i*-th series differing from joints of the (i–1)-th series of "+" joints. The initial joint of the first series of "+" joints is MS(1,1). For the first link appearing in the first series of "+" joints, the matrix Q_o^o should be known. Q_o^o is a transformation matrix between the reference frame representing ground floor (or basis) and the first link of the kinematic chain resting on the ground floor. It is convenient to place the reference frame just at the contact point. Then, if a fixed support serves as a basis, the matrix Q_o^o is a unit matrix. If we proceed to another chain, then the matrix Q_i^o (Q_i^o is a transformation matrix of the *i*-th link coordinate frame into the reference frame) should be either stored or formed on the basis of the procedure for forming dynamic equations of motion for open kinematic chains. The transformation matrix of the branching link serves to calculate the vectors $\vec{r}$ and $\vec{e}$. Information is needed on the branching link velocities ($\vec{v}$) and accelerations ($\vec{w}$) and the vector coefficients $\vec{\alpha}$, $\vec{\beta}$, $\vec{\alpha}^o$, and $\vec{\beta}^o$.[4,13] All support vectors are equal to zero. These quantities for the mobile branching link should be stored when the preceding chain is analyzed.

To consider the biped dynamics in the double-support phase (Figure 27.11), an equivalent open kinematic chain should be employed. Let us suppose that the terminal link of the equivalent chain does not coincide with the basic link (Figure 27.12). It is possible to determine the translational ($\Delta\vec{\rho}$) and angular ($\Delta\vec{\sigma}$) displacements yielding the coincidence of the two coordinate systems Ψ and Ψ'. Since we consider an open chain equivalent to the closed chain in the dynamic analysis, it is possible to use the same basic procedure as for the open chain.[4,13] Thus, the algorithm should be supplied with an iterative procedure to calculate the position and additional velocities of the first chain. The procedure is repeated until the closure conditions are satisfied.

27.2.4 Example

The nominal dynamics synthesized for the biped locomotion mechanism shown in Figures 27.13 and 27.14 consists of 14 links and 14 revolute joints of the 5-th class for the single-support phase only. Links 5 and 10 are branching links. In the course of motion, the hands are fixed on the chest of the

© 2002 by CRC Press LLC

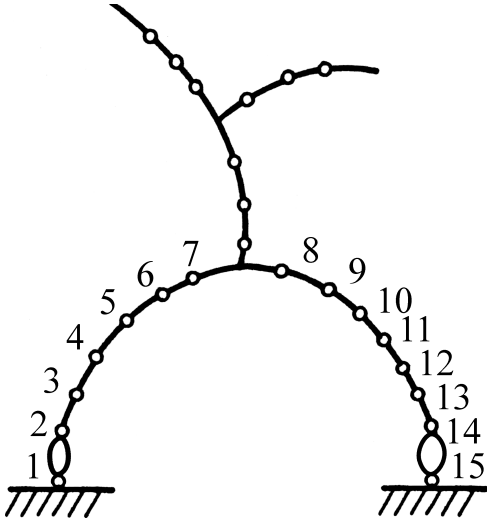


FIGURE 27.11 The anthropomorphic mechanism in double-support position.

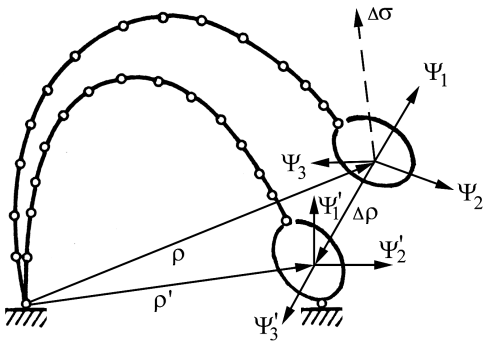


FIGURE 27.12 Terminal link of the equivalent chain.

mechanism. This structure can be split into three kinematic chains containing joints 1 through 8, 9 through 12, and 13 and 14, in the first, second, and third chains, respectively. The topological structure of the complex kinematic chain can be represented by a series of "+" joints in a matrix form:

$$MS = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 0 \\ 1 & 2 & 3 & 4 & 5 & 9 & 10 & 11 & 12 \\ 1 & 2 & 3 & 4 & 5 & 9 & 10 & 13 & 14 \end{bmatrix}.$$

Table 27.1 shows numerical values for the mechanical part of the mechanism. The prescribed portion of the mechanism motion is adopted for joints 1 through 8, i.e., for the first chain. This part of the synergy (prescribed synergy) is defined on the basis of the human gait measurements shown in Figure 27.15. A set of prescribed ZMP trajectories for the single-support gait phase is given in Figure 27.16 and the compensation has been synthesized for q^9 and q^{10} (Figure 27.17) to ensure dynamic equilibrium of the mechanism during the motion, in both the sagittal and frontal planes (Figure 27.18).

27.3 Control Synthesis for Biped Gait

Hierarchy is a basic principle on which control of large scale systems is generally based. This holds true for robots as well. The hierarchical organization of the control system is most often vertical, so that each control level deals with wider aspects of the overall system behavior than the lower level.[15-18] A higher control level always refers to the lower ones, and it controls system parameters that vary more slowly. A higher level communicates with a lower level, giving it instructions and receiving from it relevant information required for decision making. After obtaining the information from a lower level, each level makes decisions taking into account decisions obtained from higher levels and forwards them to the lower levels for execution.

27.3.1 Synthesis of Control with Limited Accelerations

Control synthesis is performed in two steps: (a) nominal regimes, and (b) perturbed regimes. For nominal regimes, the control is computed on the basis of the complete (nonlinear) model with the

© 2002 by CRC Press LLC

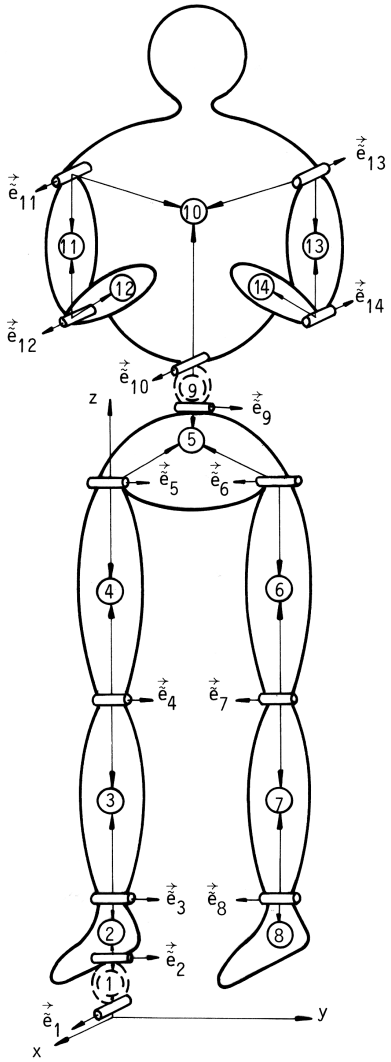


FIGURE 27.13 Mechanical scheme of the anthropomorphic mechanism with fixed arms.

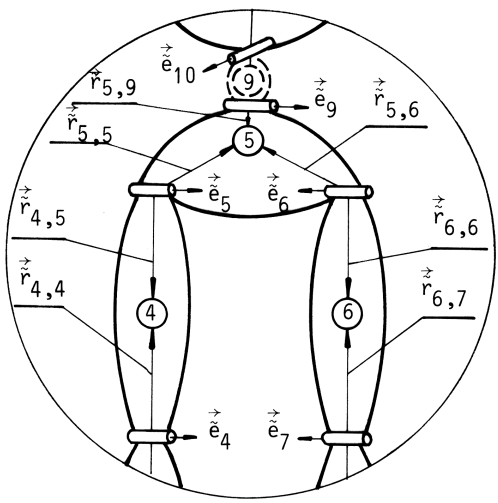


FIGURE 27.14 Notation of $\vec{r}_{i,j}$ vectors.

permanent requirement of satisfying dynamic equilibrium conditions for the overall mechanism. This control should enable the system (in the absence of disturbance) to follow the nominal trajectories. For perturbed regimes, the control should force the state vector to its nominal value, i.e., to the nominal programmed trajectory. The action should be smooth, with no significant change in link acceleration, to keep its influence on the unpowered DOFs within an acceptable range.[13,19] Let us consider the overall system model defined as:

$$S^i:\ \dot{x} = \hat{A}(x) + \hat{B}(x)N(u)$$

Assume that the part of the system corresponding to powered DOFs can be rearranged as a set of subsystems S_a^i coupled via the term $\left(f_c^i \cdot P_c^i\right)$:

$$S_a^i:\quad \dot{x}_c^i = A_c^i x_c^i + b_c^i N(u^i) + f_c^i P_c^i\,,\quad \forall i \in I_1$$

© 2002 by CRC Press LLC

TABLE 27.1 Kinematic and Dynamic Parameters of the Mechanism

Link	Mass (kg)	Moment of Inertia (kgm²) J_X	J_Y	J_Z	Distance of the Axes Centers of Joints from the Link Center (m)	Joint Unit Axes
1	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{1,1} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{1,2} = (0,0,-0.0001)^T$	$\vec{\tilde{e}}_1 = (1,0,0)^T$
2	1.53	0.00006	0.00055	0.00045	$\vec{\tilde{r}}_{2,2} = (0,0,0.030)^T$; $\vec{\tilde{r}}_{2,3} = (0,0,-0.070)^T$	$\vec{\tilde{e}}_2 = (0,1,0)^T$
3	3.21	0.00393	0.00393	0.00038	$\vec{\tilde{r}}_{3,3} = (0,0,0.210)^T$; $\vec{\tilde{r}}_{3,4} = (0,0,-0.210)^T$	$\vec{\tilde{e}}_3 = (0,1,0)^T$
4	8.41	0.01120	0.01200	0.00300	$\vec{\tilde{r}}_{4,4} = (0,0,0.220)^T$; $\vec{\tilde{r}}_{4,5} = (0,0,-0.220)^T$	$\vec{\tilde{e}}_4 = (0,1,0)^T$
5	6.96	0.00700	0.00565	0.00627	$\vec{\tilde{r}}_{5,5} = (0,0.135,0.1)^T$; $\vec{\tilde{r}}_{5,6} = (0,-0.135,0.1)^T$; $\vec{\tilde{r}}_{5,9} = (0,0,-0.05)^T$	$\vec{\tilde{e}}_5 = (0,1,0)^T$
6	8.41	0.01120	0.01200	0.00300	$\vec{\tilde{r}}_{6,6} = (0,0,-0.220)^T$; $\vec{\tilde{r}}_{6,7} = (0,0,0.220)^T$;	$\vec{\tilde{e}}_6 = (0,-1,0)^T$
7	3.21	0.00393	0.00393	0.00038	$\vec{\tilde{r}}_{7,7} = (0,0,-0.210)^T$; $\vec{\tilde{r}}_{7,8} = (0,0,0.210)^T$	$\vec{\tilde{e}}_7 = (0,-1,0)^T$
8	1.53	0.00006	0.00055	0.00045	$\vec{\tilde{r}}_{8,8} = (0,0,-0.070)^T$	$\vec{\tilde{e}}_8 = (0,-1,0)^T$
9	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{9,9} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{9,10} = (0,0,-0.0001)^T$	$\vec{\tilde{e}}_9 = (0,1,0)^T$
10	30.85	0.15140	0.13700	0.02830	$\vec{\tilde{r}}_{10,10} = (0,0,0.34)^T$; $\vec{\tilde{r}}_{10,11} = (0,0.2,-0.06)^T$; $\vec{\tilde{r}}_{10,13} = (0,-0.2,-0.06)^T$	$\vec{\tilde{e}}_{10} = (1,0,0)^T$
11	2.07	0.00200	0.00200	0.00022	$\vec{\tilde{r}}_{11,11} = (0,0,-0.154)^T$; $\vec{\tilde{r}}_{11,12} = (0,0,0.154)^T$	$\vec{\tilde{e}}_{11} = (1,0,0)^T$
12	1.14	0.00250	0.00425	0.00014	$\vec{\tilde{r}}_{12,12} = (0,0,-0.132)^T$	$\vec{\tilde{e}}_{12} = (1,0,0)^T$
13	2.07	0.00200	0.00200	0.00022	$\vec{\tilde{r}}_{13,13} = (0,0,-0.154)^T$; $\vec{\tilde{r}}_{13,14} = (0,0,0.154)^T$	$\vec{\tilde{e}}_{13} = (-1,0,0)^T$
14	1.14	0.00250	0.00425	0.00014	$\vec{\tilde{r}}_{14,14} = (0,0,-0.132)^T$	$\vec{\tilde{e}}_{14} = (-1,0,0)^T$

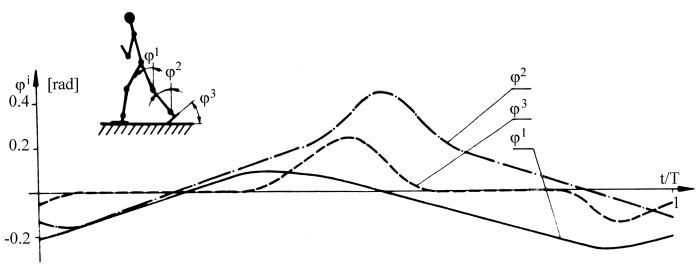


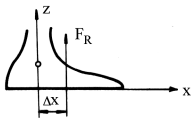
FIGURE 27.15 Synergy for walking upon level ground.

where $A_c^i \in R^{n_i x n_i}$ is the subsystem matrix, whereas $b_c^i \in R^{n_i}$ and $f_c^i \in R^{n_i}$ are distribution vectors of the input control signal and force, respectively, $x_c^i \in R^{n_i}$ is the subsystem state vector, $u^i \in R^1$ and $P_c^i \in R^1$ are the scalar values of control input and generalized force of the i-th subsystem, $i \in I_1$, $I_1 = \{i, i = 1,2,...,m\}$, $N(u^i)$ is the nonlinearity of the amplitude saturation type.

Let the nominal trajectory x_c^o, $x_c^o = \left(x_c^{o1^T}, x_c^{o2^T}, \ldots, x_c^{om^T}\right)^T$ and the nominal control u^o, $u^o = \left(u^{o1}, u^{o2}, \ldots, u^{om}\right)^T$ be introduced in such a way to satisfy:

$$S_a^i: \quad \dot{x}_c^{oi} = A_c^i x_c^{oi} + b_c^i N(u^{oi}) + f_c^i P_c^{oi}, \quad \forall i \in I_1. \tag{27.30}$$

© 2002 by CRC Press LLC



ZMP displacement

CASE	t(sec)	Δx (m)
I	0 ÷ T/2	0.0
II	0 ÷ 0.3 0.3 ÷ T/2	0.0 0.035
III	0 ÷ 0.5 0.5 ÷ T/2	0.0 0.035
IV	0 ÷ 0.2 0.2 ÷ 1.0 1.0 ÷ T/2	-0.02 0.0 0.0
V	0 ÷ 0.2 0.2 ÷ 0.6 0.6 ÷ T/2	-0.02 0.0 0.035

FIGURE 27.16 Set of ZMP trajectories for the single-support gait phase.

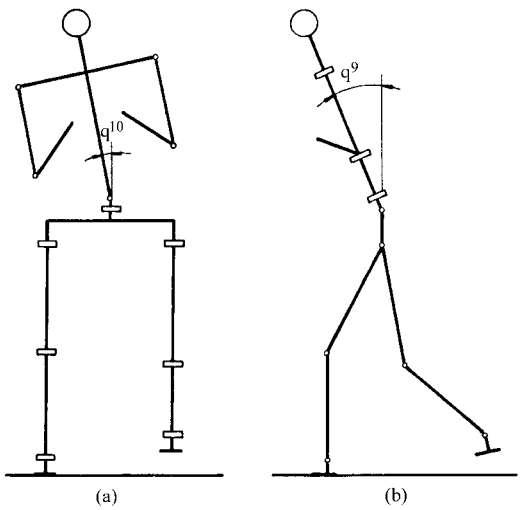


FIGURE 27.17 Compensating DOF in the frontal (a) and sagittal (b) planes.

Then, the model of subsystem deviation from the nominal is considered in the form:

$$\Delta\dot{x}_c^i = A_c^i \Delta x_c^i + b_c^i N(\Delta u^i) + f_c^i \Delta P_c^i, \quad \forall i \in I_1 . \tag{27.31}$$

The purpose of the synthesis of disturbance-compensating control (Δu) is to force the system deviation Δx_c^i, (i = 1, 2, …, m) to zero, to maintain the overall system dynamic balance.

We shall synthesize the local controller for the *i*-th actuator, i.e., for the S_a^i subsystem whose model of state deviation around the nominal trajectory is given by Equation (27.31). We want to define the controller for this subsystem that will reduce the state deviation $\Delta x_c^i(t)$ to zero, but in doing so we want to prevent the appearance of very high accelerations. Therefore, we shall synthesize a controller that will ensure the acceleration of the corresponding joint $\Delta\ddot{q}^i$ be limited. We start with the simple problem of the second order linear system with limited accelerations.[13,19]

Let us consider the classical time-minimum problem. Let the system be described by:

$$\Delta\dot{q}_1^i(t) = \Delta q_2^i(t)$$

$$\Delta\dot{q}_2^i(t) = u_i^*(t) \quad \left|u_i^*(t)\right| \le \Omega^i \tag{27.32}$$

© 2002 by CRC Press LLC

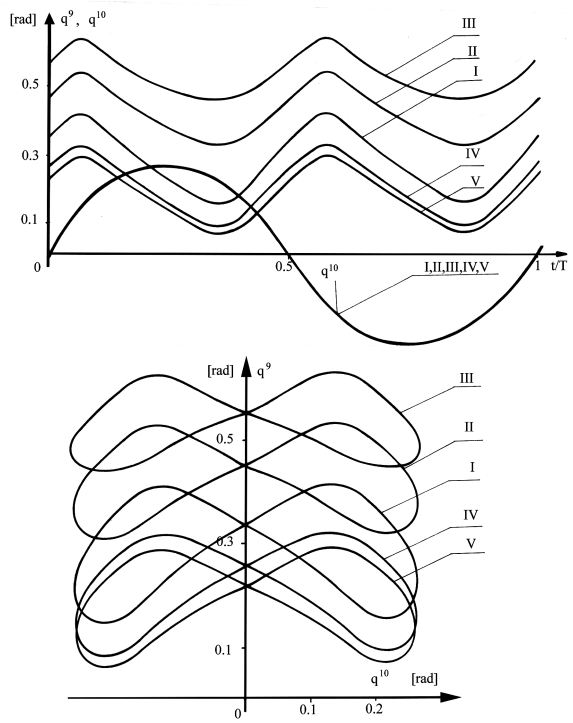


FIGURE 27.18 Compensating movements for the single-support gait upon level ground for T = 1.5, S = 0.6, and ZMP laws from Figure 27.16.

with the initial conditions $\Delta q_1^i(0)=\alpha^i$ and $\Delta q_2^i(0)=\beta^i$, where $\Omega^i \in R^1$, $\Delta q_1^i \in R^1$, and $\Delta q_2^i \in R^1$. The value $u_i^*(t)$ should be computed in such a way to ensure that the system (27.32) returns from $\left(\alpha^i, \beta^i\right)$ to the point (0, 0) in a minimal time interval.

Therefore, such a solution of Equation (27.32) should be obtained that the functional $J=\int_0^{T^*} dt$ is at minimum, where $T^* \in R^1$ defines an unspecified time interval. Such types of problems are well known, and for this particular case ($\Omega^i \neq 1$), its solution is given by the expression:

$$u_i^*=\begin{cases} +\Omega^i \quad \text{if} \quad \Delta q_1^i < \dfrac{\Delta q_2^i \left|\Delta q_2^i\right|}{2\Omega^i} \quad \text{or} \quad \Delta q_1^i = \dfrac{\left(\Delta q_2^i\right)^2}{2\Omega^i} \wedge \Delta q_2^i \leq 0 \\ -\Omega^i \quad \text{if} \quad \Delta q_1^i > \dfrac{-\Delta q_2^i \left|\Delta q_2^i\right|}{2\Omega^i} \quad \text{or} \quad \Delta q_1^i = \dfrac{-\left(\Delta q_2^i\right)^2}{2\Omega^i} \wedge \Delta q_2^i \geq 0 \end{cases} . \tag{27.33}$$

We shall apply this solution to control one single actuator, i.e., the subsystem S_a^i associated with the *i*-th joint. Let us suppose the mechanism is powered by DC motors whose models are given in the form (the state vector $\Delta x_c^i=(\Delta q^i, \Delta \dot{q}^i, \Delta i_R^i)^T$:

$$\begin{bmatrix} \Delta \dot{q}^i \\ \Delta \ddot{q}^i \\ \Delta \dot{i}_R^i \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & a_{22}^i & a_{23}^i \\ 0 & a_{32}^i & a_{33}^i \end{bmatrix} \begin{bmatrix} \Delta q^i \\ \Delta \dot{q}^i \\ \Delta i_R^i \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{f}_c^i \\ 0 \end{bmatrix} \Delta P_c^{i*} + \begin{bmatrix} 0 \\ 0 \\ \bar{b}_c^i \end{bmatrix} \Delta u^i . \tag{27.34}$$

© 2002 by CRC Press LLC

From the second equation of (27.34) we can write:

$$\Delta i_R^{i*} = (u^{i*} - a_{22}^i \, \Delta \dot{q}^i - \bar{f}_c^i \, \Delta P_c^i) / a_{23}^i \tag{27.35}$$

where $\Delta \ddot{q}^i$ is replaced by the value of the allowed link acceleration u^{i*} from (27.33). Then, Δi_R^{i*} in (27.35) is the corresponding rotor current, and its derivative can be computed from the following expression:

$$\Delta \dot{i}_R^i = \left(\Delta i_R^{i*} - \Delta i_R^i\right) / \Delta t \tag{27.36}$$

where $\Delta t \in R^1$ is the control sampling period. Now we can determine the control for Equation (27.34). Let us assume that we want to limit acceleration of the actuator (joint) within the limits $\Omega_{\min}^i$ and $\Omega_{\max}^i$. Starting from the time-minimum control (Equation (27.33)) we can adopt the following control. From the third equation of Equation (27.34) the compensation control signal for the i-th actuator is[13]

$$\Delta u^i = k_{i1}^L \Delta q^i + k_{i2}^L \Delta \dot{q}^i + k_{i3}^L \Delta \dot{i}_R^i + k_{i4}^G \Delta P_c^{i*} + k_{i5}^G \tag{27.37}$$

where the constant feedback gains are

$$k_{i1}^L = 0, \quad k_{i2}^L = \left(-a_{22}^i - a_{23}^i \cdot a_{32}^i \cdot \Delta t\right) / d$$

$$k_{i3}^L = -a_{23}^i \left(1 + a_{33}^i \cdot \Delta t\right) / d, \quad k_{i4}^G = -\bar{f}_c^i / d, \quad k_{i5}^G = k_{i5} / d \tag{27.38}$$

$$k_i^5 = \begin{cases} \Omega_{\max}^i & \text{if} \quad \Delta q^i < \dfrac{-0.5 \Delta \dot{q}^i \left|\Delta \dot{q}^i\right|}{\Omega^i} \quad \text{or} \quad \Delta q^i = \dfrac{0.5 \left(\Delta \dot{q}^i\right)^2}{\Omega^i} \ \wedge \ \Delta \dot{q}^i < 0 \\ \Omega_{\min}^i & \text{if} \quad \Delta q^i > \dfrac{-0.5 \Delta \dot{q}^i \left|\Delta \dot{q}^i\right|}{\Omega^i} \quad \text{or} \quad \Delta q^i = \dfrac{-0.5 \left(\Delta \dot{q}^i\right)^2}{\Omega^i} \ \wedge \ \Delta \dot{q}^i > 0 \end{cases}$$

where $d = a_{23}^i \bar{b}_c^i \Delta t$, a_{jk}^i, $\bar{b}_c^i$, $\bar{f}_c^i$ are the elements of the corresponding matrix and vectors of the actuator (27.34), $\Omega_{\max}^i$ and $\Omega_{\min}^i$ are the maximal and minimal values of the accelerations of the i-th link. The feedback gains synthesized in this way have to ensure compensating movements such that the accelerations do not exceed a certain predetermined limit. As a consequence, the induced inertial forces will not cause an undesirable motion of the unpowered DOFs, i.e., the displacement of ZMP out of a prescribed area. The proposed control law (27.37) consists of two parts: local control and global control (concerning feedback with respect to ΔP_c^{i*} from the rest of the system upon the i-th subsystem). The term k_{i5}^G may be conditionally associated with global control, although it is based upon local feedback information. The global control ($k_{i4}^G \Delta P_c^{i*}$) requires information on the coupling acting upon the i-th subsystem.

27.3.2 Synthesis of Global Control with Respect to ZMP Position

The decentralized control defined by Equation (27.37) applied at the mechanism's joints is not sufficient to ensure tracking of internal nominal trajectories with the addition of the appropriate behavior of the unpowered subsystem. An additional feedback must be introduced at one of the powered joints to ensure satisfactory motion of the complete mechanism. The task of this feedback is to reduce the destabilizing effect of the coupling acting upon the unpowered subsystems.

© 2002 by CRC Press LLC

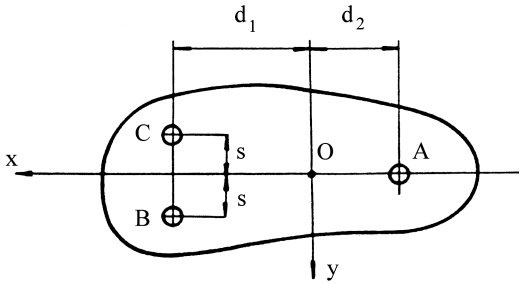


FIGURE 27.19 Disposition of the force sensors on the mechanism sole.

Since a dominant role in system stability is played by the unpowered DOFs it is necessary to reduce the destabilizing coupling effects acting upon them. Because the unpowered subsystem cannot compensate for its own deviation from the nominal state, one of the powered subsystems has to be chosen to accomplish it. As coupling of the subsystems S^j is a function of control input to the i-th subsystem S_a^i, it is clear that a feedback from the subsystem S^j to the inputs Δu^i of the subsystem S_a^i should be introduced.

In dealing with bipeds,[13,20] where the unpowered DOFs are formed by contact of the feet and the ground, it is possible to measure the ground reaction force with the aid of force sensors (at least three) to determine the acting point of total vertical reaction force. For the known motion of the overall mechanism, the ground reaction force (or force in double-support phase) is defined by the intensity, direction, and position of the acting point on the foot. If force sensors A, B, and C are introduced (Figure 27.19) and the system is performing gait, the measured values of vertical reaction forces R_A, R_B, and R_C correspond to their nominal values, and the nominal position of ZMP can be determined. Measurement of the vertical reaction forces R_A, R_B, and R_C when the mechanism is performing gait in the presence of disturbances enables the determination of the actual position of ZMP. If the nominal ZMP position corresponds to point 0, it can be written:

$$s(\Delta R_B - \Delta R_C) = M_x = R_z \cdot \Delta y$$

$$d_1(\Delta R_B + \Delta R_C) - d_2 \Delta R_A = M_y = R_z \cdot \Delta x \tag{27.39}$$

where ΔR_A, ΔR_B, and ΔR_C are the deviations of the corresponding measured forces from their nominal values, R_z is the total vertical reaction force, and Δx and Δy are the displacements of ZMP from its nominal position. These displacements can be computed from Equation (27.39), provided the sensor dispositions and vertical reaction forces are known. The actual position of the ZMP is the best indicator of overall biped behavior, and we can use it to achieve a dynamically balanced motion. Our aim is to synthesize such control that will ensure a stable gait. The primary task of the feedback with respect to ZMP position is to prevent its excursion out of the allowable region, i.e., to prevent the system from falling by rotation about the foot edge. If this is fulfilled, a further requirement imposed is to ensure that the actual ZMP position is as close as possible to the nominal.

Our further discussion will be limited to biped motion in the sagittal plane, which means that the ground reaction force position will deviate only in the direction of the x axis by Δx. Figure 27.20 illustrates the case when the vertical ground reaction force R_z deviates from the nominal position 0 by Δx; thus, the moment $R_z \cdot \Delta x = M_{ZMP}^x$ is a measure of the mechanism's overall behavior.

In the same way we can consider the mechanism motion in the frontal plane. $R_z \cdot \Delta y = M_{ZMP}^y$ is a measure of the mechanism's behavior in the direction of the y axis. Let us assume the correction of the R_z acting point in one direction is done by the action at only one joint, arbitrarily selected in advance. A basic assumption introduced for the purpose of simplicity is that the action at the chosen joint will not cause a change in the motion at any other joint. If we consider only this

© 2002 by CRC Press LLC

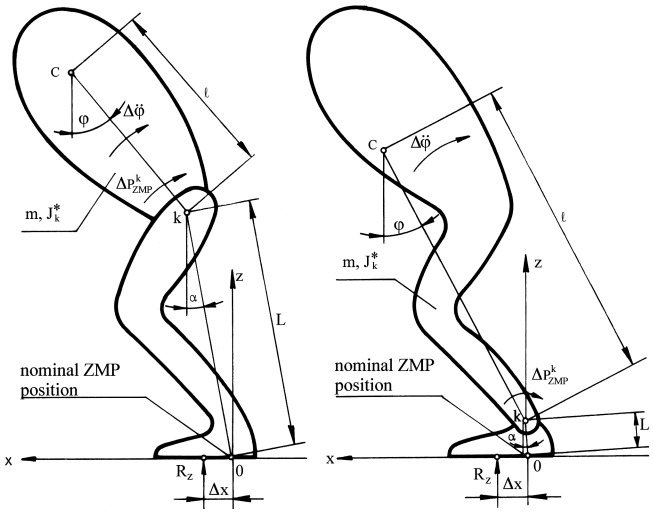


FIGURE 27.20 Compensation of ZMP displacement by (a) hip joint, and (b) ankle joint.

action, the system will behave as if composed of two rigid links connected at the joint *k,* as presented in Figure 27.20. In other words, the servo systems are supposed to be sufficiently stiff. In Figure 27.20, two situations are illustrated: (case a) when the hip of the supporting leg is the joint compensating for the ZMP displacement, and (case b) when the ankle joint is that joint. In both cases, this joint is denoted by *k,* and all links above and below it are considered as a single rigid body. The upper link is of the total mass *m* and inertia moment J_k^* for the axis of the joint *k*. Of course, numerical values are different for both cases.

The distance from the ground surface to *k* is denoted by L, from *k* to C (mass center of the upper link) by ℓ, whereas ΔP_{ZMP}^k stands for the additional correctional torque applied to the joint *k.* In Figure 27.20, the upper (compensating) link is presented as a single link above the joint *k*. In both cases presented, the compensating link includes the other leg (not drawn in the figure), which is in the swing phase. The calculation of the inertia moment J_k^* must include all the links found further onward with respect to the selected compensating joint. All the joints except the *k*-th joint are considered frozen, and, as a consequence, the lower link, representing the sum of all the links below the *k*-th joint, is considered a rigid body standing on the ground surface and performing no motion.

The procedure by which the correctional amount of global control is synthesized with respect to ZMP position is as follows. Assume the mechanism performs the gait such that displacement of the ground reaction force $\vec{R}$ in the *x* direction occurs, so $M_{ZMP}^x = R_z \cdot \Delta x$. The quantity ΔP_{ZMP}^k is to be determined on the basis of the value M_{ZMP}^x and the known mechanism and gait characteristics. Assume that the additional torque ΔP_{ZMP}^k will cause change in acceleration of the compensating link $\Delta\ddot{\varphi}$, while velocities will not change due to the action of ΔP_{ZMP}^k, $\Delta\dot{\varphi} \approx 0$. From the equation of planar motion of the considered system of two rigid bodies (Figure 27.20), which is driven by ΔP_{ZMP}^k, and under the assumption that the terms $(\dot{\varphi}\,\Delta\dot{\varphi})$ and $(\Delta\dot{\varphi})^2$ in the expression for normal component of angular acceleration of the upper link are neglected, it follows that:[13,19]

$$\Delta P_{ZMP}^k = \frac{M_{ZMP}^x}{1 + \dfrac{m \cdot \ell \cdot L \cdot \cos\varphi\,\cos\alpha}{J_k^*} + \dfrac{m \cdot \ell \cdot L \cdot \sin\varphi\,\sin\alpha}{J_k^*}}\,. \tag{27.40}$$

The control input to the actuator of the compensating joint that has to realize ΔP_{ZMP}^k can be computed from the model of the actuator deviation from the nominal. Thus:

© 2002 by CRC Press LLC

$$\begin{bmatrix} \Delta\dot{q}^k \\ \Delta\ddot{q}_T^k \\ \Delta\dot{i}_R^k \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & a_{22}^k & a_{23}^k \\ 0 & a_{32}^k & a_{33}^k \end{bmatrix} \begin{bmatrix} \Delta q^k \\ \Delta\dot{q}^k \\ \Delta i_R^k \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{f}_c^k \\ 0 \end{bmatrix} (\Delta P_c^k + \Delta P_{ZMP}^k) + \begin{bmatrix} 0 \\ 0 \\ \bar{b}_c^k \end{bmatrix} (\Delta u^k + \Delta u_{ZMP}^k). \quad (27.41)$$

This model differs from Equation (27.34) by the terms ΔP_{ZMP}^k and Δu_{ZMP}^k. From the second equation of (27.41), the change of the rotor current is:

$$\Delta i_R^k = \frac{\Delta\ddot{q}_T^k - a_{22}^k \cdot \Delta\dot{q}^k - \bar{f}_c^k (\Delta P_c^k + \Delta P_{ZMP}^k)}{a_{23}^k}. \quad (27.42)$$

Here the subscript T is used for the acceleration $\Delta\ddot{q}^k$ from Equation (27.41). It denotes the total change of link acceleration, which consists of two parts. The first part is the regular change of acceleration due to the control already applied to each powered joint defined by Equation (27.37) and corresponds to ΔP_c^k. The second part is a direct consequence of the compensation torque ΔP_{ZMP}^k. Thus:

$$\Delta\ddot{q}_T^k \approx \Delta\ddot{q}^k + \Delta\ddot{q}_{ZMP}^k \approx \Delta\ddot{q}^k + \frac{\Delta P_{ZMP}^k}{J_k^*} \quad (27.43)$$

where $\Delta\ddot{q}^k(t) \approx (\Delta\dot{q}^k(t) - \Delta\dot{q}^k(t-\Delta t)/\Delta t$. Then, from the third equation of (27.41) we have:

$$\Delta u_{ZMP}^k = \frac{\Delta\dot{i}_R^k - a_{32}^k \cdot \Delta\dot{q}^k - a_{33}^k \cdot \Delta i_R^k}{\bar{b}_c^k} - \Delta u^k \quad (27.44)$$

where Δu^k is the control defined by Equation (27.37), while $\Delta\dot{i}_R^k$ stands for $\Delta\dot{i}_R^k(t) \approx (\Delta i_R^k(t) - \Delta i_R^k(t+\Delta t))/\Delta t$. Equation (27.44) defines the control input to the k-th actuator that has to produce ΔP_{ZMP}^k. Taking into account that ΔP_{ZMP}^k is derived by introducing certain simplifications, an additional feedback gain $k_{ZMP}^{Gk} \in R^1$ has to be introduced into Equation (27.44). Thus, Equation (27.44) becomes:

$$\Delta u_{ZMP}^k = k_{ZMP}^{Gk} \left(\frac{\Delta\dot{i}_R^k - a_{32}^k \cdot \Delta\dot{q}^k - a_{33}^k \cdot \Delta i_R^k}{\bar{b}_c^k} - \Delta u^k \right) \quad (27.45)$$

The additional feedback and correctional input to the selected powered mechanism's subsystem have the purpose only of maintaining the ZMP position. It is quite possible that the feedback introduced could spoil the tracking of the internal nominal trajectory of the joint *k,* but the dynamic stability of the overall system would be preserved, which is the most important goal of a locomotion system. Which of the joints (ankle, hip, etc.) is most suitable for this purpose cannot be determined in advance, because the answer depends on the task imposed. In Figure 27.21 a scheme of the control is given with feedback introduced with respect to the ZMP position.

27.3.3 Example

The scheme of the biped structure used for the gait simulation is presented in Figure 27.22 and its mechanical parameters are given in Table 27.2. The joint with more than one DOF has been modeled as a set of corresponding numbers of simple rotational joints connected with light links of zero length ($\vec{r}_{i,k} = 0$). These are called fictitious links, and in Figure 27.22 are represented by a dashed

© 2002 by CRC Press LLC

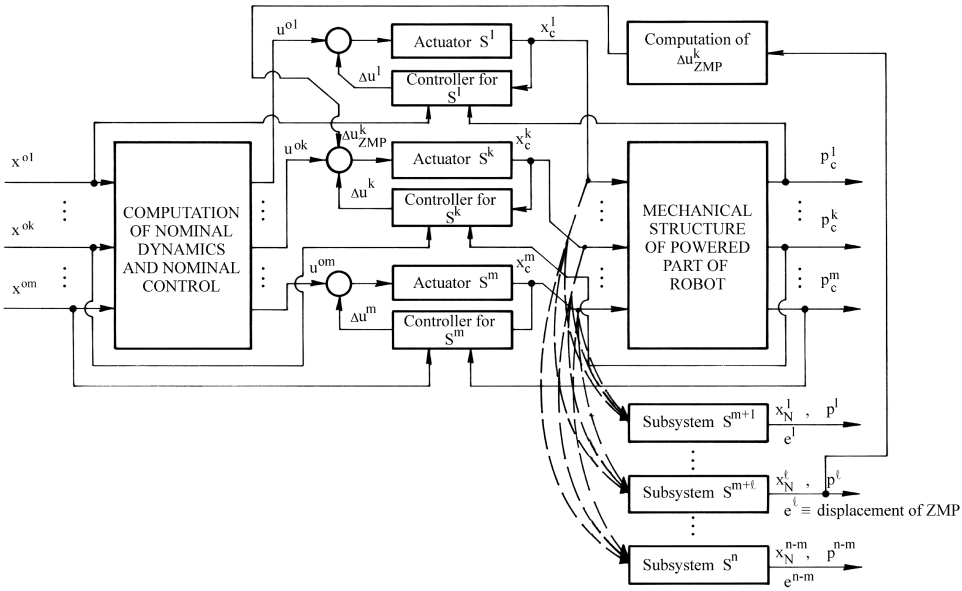


FIGURE 27.21 Control scheme with global feedback from the *l*-th unpowered subsystem to the *k*-th powered subsystem.

line. For example, simple rotational joints with the unit rotational axes $\vec{e}_3$ and $\vec{e}_4$ connected to the fictitious link 3 constitute the ankle joint of the right leg. In a similar way we can represent the hip and trunk joints. Other joints possess one DOF only.[13]

Since the mechanism in the single-support phase can rotate as a whole about the foot edges in the frontal and sagittal planes, these two DOFs are modeled as simple rotational joints with the axes $\vec{e}_1$ and $\vec{e}_2$ below the supporting foot. The mechanism rotation about the z axis is supposed to be prevented by sufficiently large friction between the sole mechanism and the ground surface. The nominal motion is synthesized using the prescribed synergy method. The compensating movements are executed by the trunk in the sagittal and frontal planes about $\vec{e}_{15}$ and $\vec{e}_{16}$.

The motion is simulated for one half-step period and for the single-support phase only. Duration time of the simulated motion was T = 0.75 s. The perturbed motion of the system around the nominal trajectory was simulated, and the trunk angular displacement from the nominal trajectory in the sagittal plane of $\Delta q^{15}(0) = 0.2\ [rad]$ was adopted as disturbance at the initial moment. Each gait is simulated using three different values of $\Omega^i_{\max}$ (i = 4 for the ankle joint, i = 15 for the trunk) defined by k^G_{i5}. They are $|\Omega^i_{\max}| = 3[rad/s^2]$, $|\Omega^i_{\max}| = 5[rad/s^2]$, and $|\Omega^i_{\max}| = 7[rad/s^2]$. Figure 27.23 presents the case when only local feedback gains defined by Equation (27.38) are employed but without feedback with respect to the overall system equilibrium. The trunk deviation Δq^{15} converges to the nominal value, but the deviations Δq^4 (ankle joint) and Δq^6 (hip joint) slightly diverge — the absolute values of these deviations are, however, very small.

Figure 27.24 gives the example of ZMP displacement compensated by the ankle joint. Again, the trunk inclination for 0.2 [rad] was adopted as initial disturbance and $k^{Gk}_{ZMP} = 0.5$. Figure 27.24e illustrates the ZMP behavior; maximal average deviation is about 1 cm, which can be considered very successful. Behavior of the other joints, especially of the ankle, is not affected much by keeping ZMP position strongly under control.

27.4 Dynamic Stability Analysis of Biped Gait

The system is considered a set of subsystems, each of which is associated with one powered joint. The stability of each subsystem is checked (neglecting the coupling) and then dynamic coupling

© 2002 by CRC Press LLC

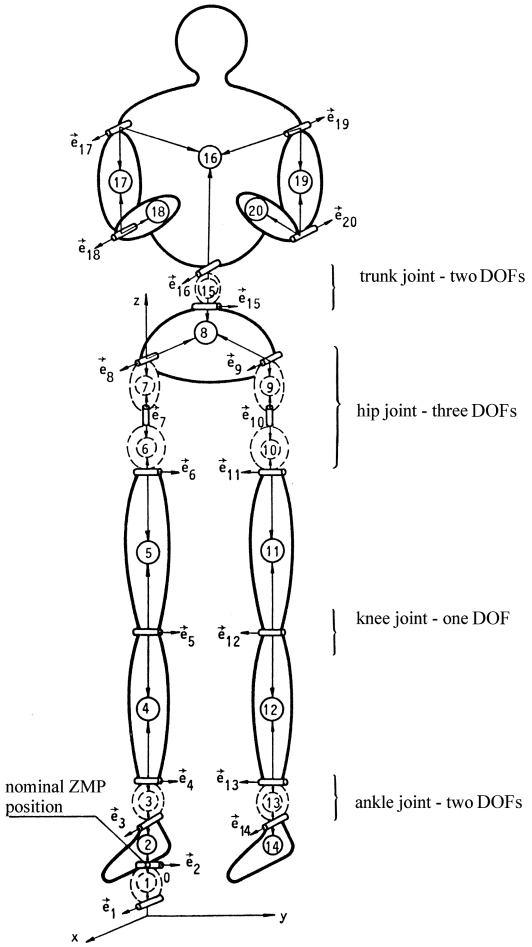


FIGURE 27.22 Scheme of mechanical biped structure.

between the subsystems is analyzed. The stability of the overall system is tested by taking into account all dynamic interconnections between the subsystems. However, these tests require that all subsystems are stable. To analyze stability of the mechanisms including unpowered joints, we introduced the so-called composite subsystems that consist of one powered and one unpowered joint. Thus we obtain a subsystem which, if considered decoupled from the rest of the system, might be stabilized. Further, the interconnections of the composite subsystem with the rest of subsystems are taken into account, and a test for stability of the overall mechanism is established.

To analyze stability of the locomotion mechanisms, we shall use the aggregation– decomposition method via Lyapunov vector functions in bounded regions of state space, originally developed for manipulation robots.[13,21,22] Since it is valid for the mechanism with all joints powered, this method cannot be directly applied to locomotion mechanisms containing unpowered DOFs. Because of that, we modified the subsystem modeling by incorporating the models of unpowered DOFs into the composite subsystem models. In this way, the entire mechanism is considered in the stability analysis.[13,22]

27.4.1 Modeling of Composite Subsystems

The mathematical model of the complete system S consists of two parts: the model of mechanical structure S^M and the model of actuators S_a^i. These models are:[13,17]

© 2002 by CRC Press LLC

TABLE 27.2 Kinematic and Dynamic Parameters of the Mechanism

Link	Mass (kg)	Moment of Inertia (kgm²)			Distance of the Axes Centers of Joints from the Link Center (m)	Joint Unit Axes
		J_X	J_Y	J_Z		
1	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{1,1} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{1,2} = (0,0,-0.0001)^T$	$\vec{\tilde{e}}_1 = (1,0,0)^T$
2	1.53	0.00006	0.00055	0.00045	$\vec{\tilde{r}}_{2,2} = (0,0,0.030)^T$; $\vec{\tilde{r}}_{2,3} = (0,0,-0.070)^T$	$\vec{\tilde{e}}_2 = (0,1,0)^T$
3	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{3,3} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{3,4} = (0,0,-0.0001)^T$	$\vec{\tilde{e}}_3 = (1,0,0)^T$
4	3.21	0.00393	0.00393	0.00038	$\vec{\tilde{r}}_{4,4} = (0,0,0.210)^T$; $\vec{\tilde{r}}_{4,5} = (0,0,-0.210)^T$	$\vec{\tilde{e}}_4 = (0,1,0)^T$
5	8.41	0.01120	0.01200	0.00300	$\vec{\tilde{r}}_{5,5} = (0,0,0.220)^T$; $\vec{\tilde{r}}_{5,6} = (0,0,-0.220)^T$	$\vec{\tilde{e}}_5 = (0,1,0)^T$
6	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{6,6} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{6,7} = (0,0,-0.0001)^T$;	$\vec{\tilde{e}}_6 = (0,1,0)^T$
7	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{7,7} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{7,8} = (0,0,-0.0001)^T$	$\vec{\tilde{e}}_7 = (1,0,0)^T$
8	6.96	0.00700	0.00565	0.00625	$\vec{\tilde{r}}_{8,8} = (0,0.135,0.1)^T$; $\vec{\tilde{r}}_{8,9} = (0,-0.135,0.1)^T$; $\vec{\tilde{r}}_{8,15} = (0,0,-0.05)^T$	$\vec{\tilde{e}}_8 = (1,0,0)^T$
9	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{9,9} = (0,0,-0.0001)^T$; $\vec{\tilde{r}}_{9,10} = (0,0,0.0001)^T$	$\vec{\tilde{e}}_9 = (1,0,0)^T$
10	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{10,10} = (0,0,-0.0001)^T$; $\vec{\tilde{r}}_{10,11} = (0,0,0.0001)^T$;	$\vec{\tilde{e}}_{10} = (0,0,1)^T$
11	8.41	0.01120	0.01200	0.00300	$\vec{\tilde{r}}_{11,11} = (0,0,-0.220)^T$; $\vec{\tilde{r}}_{11,12} = (0,0,0.220)^T$	$\vec{\tilde{e}}_{11} = (0,-1,0)^T$
12	3.21	0.00393	0.00393	0.00038	$\vec{\tilde{r}}_{12,12} = (0,0,-0.210)^T$; $\vec{\tilde{r}}_{12,13} = (0,0,0.210)^T$	$\vec{\tilde{e}}_{12} = (0,-1,0)^T$
13	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{13,13} = (0,0,-0.0001)^T$; $\vec{\tilde{r}}_{13,14} = (0,0,0.0001)^T$	$\vec{\tilde{e}}_{13} = (0,-1,0)^T$
14	1.53	0.00006	0.00055	0.00045	$\vec{\tilde{r}}_{14,14} = (0,0,-0.070)^T$	$\vec{\tilde{e}}_{14} = (1,0,0)^T$
15	0.0	0.0	0.0	0.0	$\vec{\tilde{r}}_{15,15} = (0,0,0.0001)^T$; $\vec{\tilde{r}}_{15,16} = (0,0,-0.0001)^T$	$\vec{\tilde{e}}_{15} = (0,1,0)^T$
16	30.85	0.15140	0.13700	0.02830	$\vec{\tilde{r}}_{16,16} = (0,0,0.34)^T$; $\vec{\tilde{r}}_{16,17} = (0,0.2,-0.06)^T$; $\vec{\tilde{r}}_{16,19} = (0,-0.2,-0.06)^T$	$\vec{\tilde{e}}_{16} = (1,0,0)^T$
17	2.07	0.00200	0.00200	0.00022	$\vec{\tilde{r}}_{17,17} = (0,0,-0.154)^T$; $\vec{\tilde{r}}_{17,18} = (0,0,0.154)^T$	$\vec{\tilde{e}}_{17} = (1,0,0)^T$
18	1.14	0.00250	0.00425	0.00014	$\vec{\tilde{r}}_{18,18} = (0,0,-0.132)^T$	$\vec{\tilde{e}}_{18} = (1,0,0)^T$
19	2.07	0.00200	0.00200	0.00022	$\vec{\tilde{r}}_{19,19} = (0,0,-0.154)^T$; $\vec{\tilde{r}}_{19,20} = (0,0,0.154)^T$	$\vec{\tilde{e}}_{19} = (-1,0,0)^T$
20	1.14	0.00250	0.00425	0.00014	$\vec{\tilde{r}}_{20,20} = (0,0,-0.132)^T$	$\vec{\tilde{e}}_{20} = (-1,0,0)^T$

$$S^M:\ P = H(q)\cdot\ddot{q} + h(q,\dot{q}) \tag{27.46}$$

$$S_a^i:\ \dot{x}_c^i = A_c^i x_c^i + b_c^i N(u^i) + f_c^i P_c^i \tag{27.47}$$

The mechanical structure of n DOFs is powered by m actuators. Since the $n - m$ joints are unpowered, the driving torques P^i about the axes of these joints are assumed to be zero, i.e., the vector of driving torques P has the following form $P = (P_c^1, P_c^1, ..., P_c^m, 0, ..., 0)^T$. In order to apply the method for stability analysis, we shall arrange the model in another way. The model of the ℓ-th unpowered joint follows from Equation (27.46):

$$-H_{\ell\ell}\ddot{q}_N^\ell = \sum_{\substack{j=1 \\ j\neq\ell}} H_{\ell j}\ddot{q}^j + h_\ell(q,\dot{q}) \tag{27.48}$$

© 2002 by CRC Press LLC

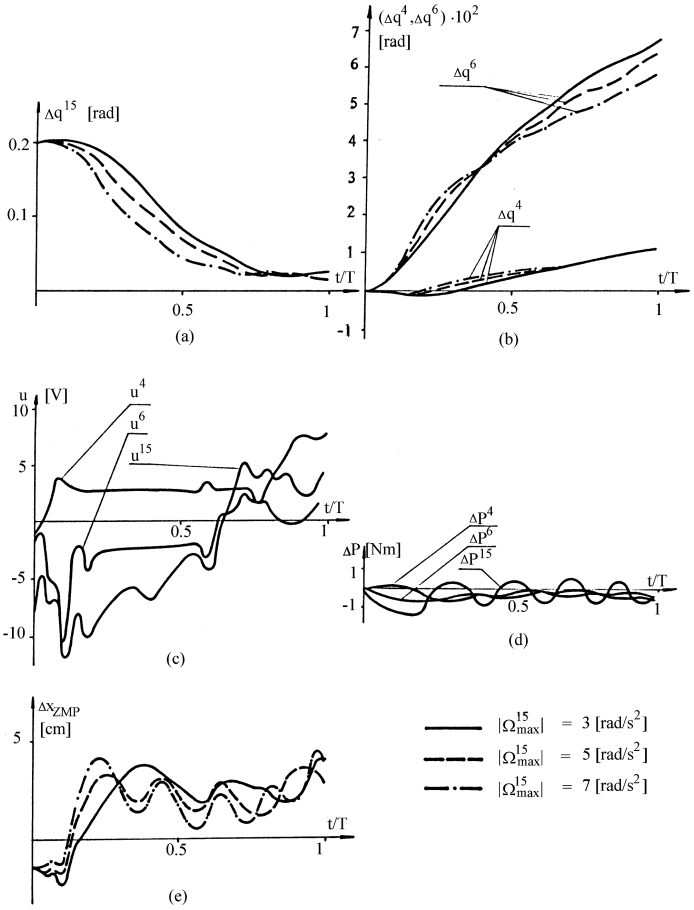


FIGURE 27.23 Walk simulation with local feedback gains defined by Equation (27.37) and with no feedback with respect to ZMP position, $\Delta q^{15}(0)=0.2[rad]$.

where $q_N^l \in R^1$ is the angle of the ℓ-th unpowered joint, $H_{\ell j}$ are the members of the matrix H(q), and h_ℓ is the member of the vector *h*. The subscript *N* denotes the unpowered DOF. However, instead of this model, let us describe the motion of the system about the axis of the ℓ-th unpowered joint as the motion of an inverted pendulum. The equation of the inverted pendulum motion in a plane is:

$$\ddot{q}_N^\ell = \frac{M}{I_o + M\rho^2} g \sin q_N^\ell + \frac{1}{I_o + M\rho^2} P_N^\ell \tag{27.49}$$

where M and I_o are the mass and inertia moment of the pendulum (the pendulum corresponds to the whole system), ρ is the distance from the supporting point to the pendulum mass center, and $P_N^\ell \in R^1$ is the resultant generalized force acting on the pendulum. If the angle q_N^ℓ is small, we can introduce the approximation $\left(\sin q_N^\ell\right) \approx q_N^\ell$. If the term multiplying q_N^ℓ is denoted by $c_o^{\ell *}$, and the term multiplying P_N^ℓ by $\bar{f}_N^\ell$, and $x_N^\ell = \left[q_N^\ell, \dot{q}_N^\ell\right]^T$ is adopted as state vector, then Equation (27.49) can be written in the matrix form:

$$\begin{bmatrix} \dot{q}_N^\ell \\ \ddot{q}_N^\ell \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ c_o^{\ell *} & 0 \end{bmatrix} \begin{bmatrix} q_N^\ell \\ \dot{q}_N^\ell \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{f}_N^\ell \end{bmatrix} P_N^\ell \tag{27.50}$$

© 2002 by CRC Press LLC

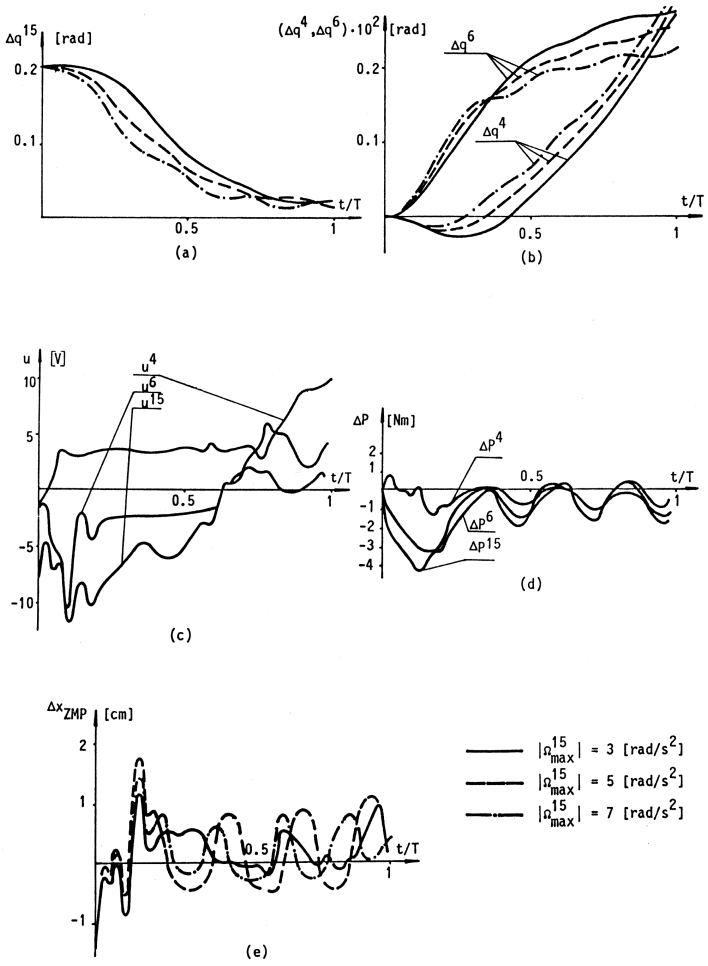


FIGURE 27.24 Walk simulation with added compensation for ZMP position (compensation is performed by ankle joint), $k_{ZMP}^{Gk}=0.5$, $\Delta q^{15}(0)=0.2[rad]$.

and in compact form

$$\dot{x}_N^\ell = A_N^\ell x_N^\ell + f_N^\ell P_N^\ell$$

Since we want the models (27.50) and (27.48) to coincide, we shall define the force P_N^ℓ as:

$$P_N^\ell = \frac{-H_{\ell\ell}^{-1}}{f_N^\ell}\left[\sum_{\substack{j=1\\ j\neq\ell}}^{n} H_{\ell j}\ddot{q}^j + h_\ell\left(q\ ,\dot{q}\ \right)\right] - \frac{C_o^{\ell *}}{\bar{f}_N^\ell} q_N^\ell \tag{27.51}$$

In this way we ensure that Equation (27.50) is an exact model of the system motion about the axis of the unpowered joint. Let us now form the composite subsystem containing one unpowered DOF (Equation (27.50)) and the *k*-th powered DOF (Equation (27.47)):

$$\begin{bmatrix}[\dot{x}_N^\ell]\\ [\dot{x}_c^k]\end{bmatrix} = \begin{bmatrix}[A_N^\ell] & 0\\ 0 & [A_c^k]\end{bmatrix}\begin{bmatrix}[x_N^\ell]\\ [x_c^k]\end{bmatrix} + \begin{bmatrix}[f_N^\ell] & 0\\ 0 & [f_c^k]\end{bmatrix}\begin{bmatrix}P_N^\ell\\ P_c^k\end{bmatrix} + \begin{bmatrix}0\\ [b_c^k]\end{bmatrix} N\left(u^k\right) \tag{27.52}$$

© 2002 by CRC Press LLC

The subscript N corresponds to the unpowered and subscript c to the powered DOFs. Here, $x_N^\ell \in R^{n_N^\ell}$ and $x_c^k \in R^{n_c^k}$ are the state vectors of the ℓ-th unpowered $x_N^\ell = \left(q_N^\ell, \dot{q}_N^\ell\right)$ k-th powered DOFs $x_c^k = \left(q_c^k, \dot{q}_c^k, i_R^k\right)$, i_R^k is the rotor current of the corresponding DC motor, and $n_N^\ell = 2$ and $n_c^k = 3$ are their orders. $A_N^\ell \in R^{n_N^\ell x n_N^\ell}$, $A_c^k \in R^{n_c^k x n_c^k}$, $f_N^\ell \in R^{n_N^\ell}$, and $f_c^k \in R^{n_c^k}$, $P_N^\ell \in R^1$, and $P_c^k \in R^1$ are the system matrices, force distribution vectors, and generalized forces of the unpowered and powered DOFs, respectively. A_N^ℓ and f_N^ℓ are defined by Equation (27.50). Taking into account the form of the actuator matrix A_c^k and the form of the unpowered DOF (Equation (27.50)), expression (27.52) can be written as:[13,22]

$$\begin{bmatrix} \begin{bmatrix} \dot{q}_N^\ell \\ \ddot{q}_N^\ell \end{bmatrix} \\ \begin{bmatrix} \dot{q}_c^k \\ \ddot{q}_c^k \\ \dot{i}_R^k \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ c_o^{\ell *} & 0 \end{bmatrix} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & a_{2,2}^k & a_{2,3}^k \\ 0 & a_{3,2}^k & a_{3,3}^k \end{bmatrix} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} q_N^\ell \\ \dot{q}_N^\ell \end{bmatrix} \\ \begin{bmatrix} q_c^k \\ \dot{q}_c^k \\ i_R^k \end{bmatrix} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \bar{f}_N^\ell & 0 \\ 0 & 0 \\ 0 & \bar{f}_c^k \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_N^\ell \\ P_c^k \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \bar{b}_c^k \end{bmatrix} N\left(u^k\right) \quad (27.53)$$

where $a_{i,j}^k$ are the elements of the matrix A_c^k or, in a compact form:

$$\dot{x}_z^k = A_z^k \cdot x_z^k + f_z^k \cdot P_z^k + b_z^k \cdot N(u^k) \qquad \forall k \in J$$

where $x_z^k \in R^{n_z^k}$ is the state vector of the composite subsystem, and $A_z^k \in R^{n_z^k x n_z^k}$, $f_z^k \in R^{n_z^k x 2}$, and $b_z^k \in R^{n_z^k}$ are the subsystem matrix, matrix of force distribution and vector of control distribution, respectively. Thus, $n_z^k = n_N^\ell + n_c^k$, $x_z^k = \left(x_N^{\ell T}, x_c^{kT}\right)^T$, and $P_z^k = \left(P_N^\ell, P_c^k\right)^T$. Obviously, Equation (27.53) defines only the k-th composite subsystem model. The set J is defined as J = {j, j = 2 m-n+1, …, m}. It is assumed that the k-th powered joint is associated with the ℓ-th unpowered joint.

In the stability analysis, all decoupled subsystems must be exponentially stable. If the subsystem corresponds to the joints of the kinematic chain, their coupling is represented by the moments about the joint axis. In fact, decoupling means an investigation of the subsystem model without the term that corresponds to the generalized force. In case of a composite subsystem this term is $\left(f_z^k \cdot P_z^k\right)$, i.e., the decoupled composite subsystem can be written as:

$$\dot{x}_z^k = A_z^k x_z^k + b_z^k N\left(u^k\right). \quad (27.54)$$

The interaction between these DOFs is the only way to control the motion of the unpowered DOFs. To preserve the integrity of the decoupled composite model, some additional elements should be introduced into the matrix A_z^i in the places representing the influence of powered DOFs on the unpowered DOFs and vice versa. The model of the composite subsystem has the final form:[13,22]

$$\begin{bmatrix} \begin{bmatrix} \dot{q}_N^\ell \\ \ddot{q}_N^\ell \end{bmatrix} \\ \begin{bmatrix} \dot{q}_c^k \\ \ddot{q}_c^k \\ \dot{i}_R^k \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ c_o^{\ell *} & 0 \end{bmatrix} & \begin{matrix} 0 & 0 & 0 \\ \left[D_1^{1k} & D_1^{2k} & D_1^{3k}\right] \end{matrix} \\ \begin{matrix} 0 & 0 \\ \left[D_2^{1k} & D_2^{2k}\right] \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & a_{2,2}^k & a_{2,3}^k \\ 0 & a_{3,2}^k & a_{3,3}^k \end{bmatrix} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} q_N^\ell \\ \dot{q}_N^\ell \end{bmatrix} \\ \begin{bmatrix} q_c^k \\ \dot{q}_c^k \\ i_R^k \end{bmatrix} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \bar{f}_N^\ell & 0 \\ 0 & 0 \\ 0 & \bar{f}_c^k \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P_N^\ell - \dfrac{D_1^k \cdot x_c^k}{\bar{f}_N^\ell} \\ P_c^k - \dfrac{D_2^k \cdot x_N^\ell}{\bar{f}_c^k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \bar{b}_c^k \end{bmatrix} N\left(u^k\right) \quad (27.55)$$

© 2002 by CRC Press LLC

or

$$\dot{x}_z^k = A_z^{k*} x_z^k + f_z^k P_z^{k*} + b_z^k N\left(u^k\right).$$

The vector $D_1^k = \left[D_1^{1k}\, D_1^{2k}\, D_1^{3k}\right]$ represents the influence of the powered DOF on the unpowered one, whereas $D_2^k = \left[D_2^{1k}\, D_2^{2k}\right]$ represents an opposite effect. Since the vectors D_1^k and D_2^k are chosen arbitrarily, $D_1^k \cdot \left[x_k\right]^T$ and $D_2^k \cdot \left[x_n^\ell\right]^T$ will be subtracted from $\left(f_z^k\ P_z^k\right)$, i.e., $P_z^{k*} = \left(P_N^{\ell*}, P_c^{k*}\right)$, $P_N^{\ell*} = P_N^\ell - \left[\left(D_1^k \cdot x_c^k\right) / \overline{f}_N^\ell\right]$, and $P_c^{k*} = P_c^k - \left[\left(D_2^k \cdot x_N^\ell\right) / \overline{f}_c^k\right]$.

The composite subsystem model formed in this way is suitable for stability investigation and enables the stability analysis of a system having joints without actuators. It should be emphasized that the models of composite subsystems (27.55) are exact, i.e., they contain no approximations. The model (27.55) coincides with the original model of the ℓ-th unpowered joint (27.48) and the model of the *k*-th powered joint with the actuator (27.47) that is driving the *k*-th joint. We rearranged the model in order to present it in a more convenient form. The mathematical model of the mechanism part that consists of composite subsystems is

$$\dot{x}_z = \hat{A}_z^* x_z + f_z P_z^* + b_z N(u_z) \tag{27.56}$$

where $x_z \in R^{N_z}$ is the state vector; $x_z = \left(x_z^{(2m-n+1)T}, \ldots, x_z^{mT}\right)^T$. $\hat{A}_z^* \in R^{N_z \times N_z}$, $\hat{A}_z^* = diag\left\{A_z^{k*}\right\}$ is the system matrix, while $b_z = diag\left\{b_z^k\right\}$ and $f_z = diag\left\{f_z^k\right\}$, $b_z \in R^{N_z \times (n-m)}$, are the distribution matrices of control force; $N(u_z) \in R^{n-m}$ and $P_z^* \in R^{2(n-m)}$ are the corresponding control and force defined by Equation (27.55), N_z and is the order of the model formed of composite subsystems:

$$N_z = \sum_{k=2m-n+1}^{m} n_z^k . \tag{27.57}$$

Thus, the mathematical model of a complete biped mechanism S with the composite subsystems included, can be obtained by uniting the model of composite subsystems (Equation (27.56)) and the powered DOFs:

$$\text{S:}\quad \dot{x} = Ax + FP + BN(u) \tag{27.58}$$

where $x \in R^N$, $x = (x_c^{1T}, \ldots, x_c^{(n-m)T}, x_z^T)^T$ is the system state vector $P = (P_c^1, P_c^2, \ldots P_c^{2m-n}, P_z^{*2m-n+1}, \ldots, P_z^{*m})^T$. Matrices $A \in R^{N \times N}, B \in R^{N \times m}$, and $F \in R^{N \times n}$ are $A = \begin{bmatrix} \hat{A}_c & 0 \\ o & \hat{A}_z \end{bmatrix}$, $B = \begin{bmatrix} \hat{b}_c & 0 \\ 0 & b_z \end{bmatrix}$, $F = \begin{bmatrix} \hat{f}_z & 0 \\ 0 & f_z \end{bmatrix}$ $N = N_z + \sum_{i=1}^{2m-n} n_i$, $\hat{A}_c = diag\left[A_c^i\right]$, $\hat{b}_c = diag\left[b_c^i\right]$, and $\hat{f}_c = diag\left[f_c^i\right]$, $\forall i \in I_2$, $I_2 = \left\{i, i = 1, \ldots, 2m-n\right\}$.

The complete system S (Equation (27.58)) is composed of *m* subsystems: (2 *m* – *n*) subsystems correspond to the powered joints modeled as in Equation (27.47), and (*n* – *m*) composite subsystems modeled as in Equation (27.56). In fact, all the subsystems can be written in the same form:

$$\dot{x}^i = A^i x^i + b^i N(u^i) + f^i \hat{P}_i(x), \qquad \forall i \in I_1 \tag{27.59}$$

© 2002 by CRC Press LLC

where x^i stands for x_c^i if i = 1,2, …, 2 $m - n$, and x_z^i if i = 2$m - n$ + 1, …,m. The same holds for A^i, b^i, f^i, and u^i, while $\hat{P}_i$ stands for P_c^i if i = 1,2, …, 2 $m - n$, and for P_z^{i*} if i = 2$m - n$ + 1, …, m. The order of the subsystems (Equation (27.59)) is denoted by n_i (though it might be either n_i or n_z^i, depending on i). Thus, we obtain a model of the system S in a form convenient for the application of the chosen method for stability analysis.

27.4.2 Stability Analysis

In regard to biped locomotion systems, the most suitable stability analysis seems to be the definition of practical stability.[13,22] Accordingly, the system is considered to be practically stable if $\forall x(0) \in X^I$ implies $x(t) \in X^F$, $\forall t \in T_s$ where $T_s = \{t : t \in (\tau_s, \tau)\}$ and $x(t) \in X^t(t)$, $\forall t \in T$, where $X^I \subseteq X^t(0)$ and $X^F \subseteq X^t(t)$, $\forall t \in T_s$.

Let us consider the overall system model S defined as in Equation (27.59), which can be considered as a set of m subsystems S^i (of the composite or powered joints) coupled through the term $\left(f^i \cdot \hat{P}_i\right)$.[13,22]

$$\dot{x}^i = A^i x^i + b^i N(u^i) + f^i \hat{P}_i(x) , \qquad \forall i \in I_1 .$$

Let us assume the nominal trajectory of the state vector $x^\circ(t)$, $\forall t \in T$ be given in such a way that it satisfies $x^\circ(0) \in X^I$, $x^\circ(t) \in X^F$, $\forall t \in T_s$, and $x(t) \in X^t(t)$, $\forall t \in T$. Further, let us assume the nominal trajectory $x^\circ(t)$ has been selected in such a way that we can find a nominal (programmed) control $u^\circ(t)$, which is a function of time, and which satisfies:

$$\dot{x}^{\circ i} = A^i x^{\circ i} + b^i N(u^{\circ i}) + f^i \hat{P}_i^\circ(x^\circ) , \qquad \forall i \in I_1 , \ \forall t \in T \tag{27.60}$$

where $x^\circ(t) = \left(x^{\circ 1T}(t), x^{\circ 2T}(t), \ldots, x^{\circ mT}(t)\right)^T$, $u^\circ = \left(u^{\circ 1}, u^{\circ 2}, \ldots, u^{\circ m}\right)^T$. $\hat{P}_i^\circ(x^\circ)$ denotes the nominal values of $\hat{P}_i(x)$. Because the subsystems (27.59) include the composite subsystems, the nominal trajectory $x^\circ(t)$ satisfies the composite subsystems. We assume that the nominal trajectory $x^\circ(t)$ and the corresponding nominal control $u^\circ(t)$, satisfying Equation (27.60), can be determined. However, due to the perturbation actions acting upon the system, a deviation of the system state from its nominal trajectory must appear. The model of deviation from the nominal trajectory can be written according to (27.59) and (27.60) as:

$$\Delta\dot{x}^i = A^i \Delta x^i + b^i N(t, \Delta u^i) + f^i \Delta\hat{P}_i(t, \Delta x, x^\circ(t)), \ \forall i \in I_1 \tag{27.61}$$

where $\Delta x^i = x^i - x^{\circ i}(t)$, $\Delta u^i = u^i - u^{\circ i}(t)$, $\Delta\hat{P}_i = \hat{P}_i - \hat{P}_i^\circ(x^\circ)$. The problem is to stabilize the model of deviation (27.61) from the nominal trajectory $x^\circ(t)$, i.e., we have to synthesize the control Δu^i such that the model of deviation from $x^\circ(t)$ (27.61) is stabilized. The aim is to ensure practical stability of the system around the nominal trajectory $x^\circ(t)$, so that for each $\Delta x(0) \in X^I - x^\circ(0)$, the conditions $\Delta x(t) \in X^F - x^\circ(t)$, $\forall t \in T_s$, and $\Delta x(t) \in X^t(t) - x^\circ(t)$, $\forall t \in T$ are fulfilled.

Let us synthesize a decentralized control. Consider an approximate model of deviation in its decoupled form (i.e., the model in which the coupling terms between subsystems $(f^i \Delta\hat{P}_i)$ are neglected):

$$\Delta\dot{x}^i = A^i \Delta x^i + b^i N(t, \Delta u^i) , \qquad \forall i \in I_1 . \tag{27.62}$$

The decoupled model of the system (Equation (27.62)) represents a set of decoupled linear subsystems that can be stabilized by simple linear feedback control:

© 2002 by CRC Press LLC

$$\Delta u^i = -k_i^{LT} \Delta x^i \,, \qquad \forall i \in I_1 \tag{27.63}$$

where $k_i^L \in R^{n_i}$ is the vector of local feedback gains selected so that the subsystem:

$$\Delta \dot{x}^i = (A^i - b^i k_i^{LT}) \Delta x^i = \tilde{A}^i \Delta x^i \,, \qquad \forall i \in I_1 \tag{27.64}$$

(where $\tilde{A}^i$ is a closed-loop subsystem matrix) is exponentially stable. In (27.64), we neglected the amplitude saturation upon the input $N(t, \Delta u)^i$. If this nonlinearity is taken into account, the subsystem Equation (27.64) is exponentially stabilized in the finite region X_i in the state space with a desired stability degree Π_i. If the decoupled subsystems Equation (27.64) are considered, it is obvious that this model will be exponentially stable in the region:[13,17]

$$X = X_1 \times X_2 \times \ldots \times X_n \,. \tag{27.65}$$

We shall analyze stability of the overall system (27.61) when the decentralized control (27.63) is applied. Let us express the subsystems characteristics by the Lyapunov functions, which, with their derivatives along with solutions for decoupled subsystems, have to satisfy:[2,13,23]

$$\Pi_{i1} \left\| \Delta x^i \right\| < V_i(\Delta x^i) < \Pi_{i2} \left\| \Delta x^i \right\| \tag{27.66}$$

$$-\Pi_{i3} \left\| \Delta x^i \right\| < \dot{V}_i\left(\Delta x^i\right) < -\Pi_{i4} \left\| \Delta x^i \right\| \quad \text{along solution of (27.64)} \tag{27.67}$$

for $\forall i \in I_1$, $\Pi_{ik} > 0$ are real numbers for k = 1,2,3,4, $V_i \rangle 0$, $V_i : R^{n_i} \to R^1$. The analysis concerning the stability on finite regions using the aggregation–decomposition method can be conservative. Functions V_i should be chosen to represent the best estimates of the degree of exponential stability Π_1 of the decoupled subsystems. Thus, we should select such Lyapunov function V_i that satisfies:[13,17]

$$\dot{V}_i\left(\Delta x^i\right) = \left(grad\, V^i\right)^T \Delta x^i \le -\Pi_{i4} \Pi_{i2}^{-1} V_i \le -\Pi_i V_i \,, \qquad \forall i \in I_1 \tag{27.68}$$

where $\dot{V}_i$ is taken along the trajectory of the decoupled subsystem (27.64). Let us select the Lyapunov function in the form:

$$V_i = \left(\Delta x^{iT} H^i \Delta x^i\right)^{1/2} \,, \qquad \forall i \in I_1 \tag{27.69}$$

where the matrix $H^i \in R^{n_i \times n_i}$ (symmetric and positive definite) can be derived as the solution of the Lyapunov matrix equation:[23]

$$\tilde{A}^{iT} H^i + H^i \tilde{A}^i = -G^i \tag{27.70}$$

where $G^i \in R^{n_i \times n_i}$ is an arbitrarily defined, symmetric, and positive definite matrix. If we select G^i to be equal to $\Pi_i H^i$, then the selected Lyapunov function (27.69) obviously satisfies Equation (27.68). Since the control signal is of limited amplitude, the condition of Equation (27.68) can be satisfied only in the finite region of initial conditions $x_i(0) \in X_i$, i.e., the decoupled system is asymptotically stable in the region X defined by Equation (27.65). The region X_i can be estimated by the regions $\tilde{X}_i$ via Lyapunov functions with an appropriate choice of $V_{i\circ}$:

© 2002 by CRC Press LLC

$$\tilde{X}_i = \left\{ \Delta x^i : V_i(\Delta x^i) < V_{i\circ} \ \ and \ \ \Delta x^i \in X_i \right\} \ , \quad \forall i \in I_1 \ , \quad \tilde{X}_i \subseteq X_i \tag{27.71}$$

where $V_{i\circ} > 0$ are positive numbers. Then, the region:

$$\tilde{X}(0) = \tilde{X}_1 \times \tilde{X}_2 \times \ldots \times \tilde{X}_m, \qquad \tilde{X}(0) \subseteq R^N \tag{27.72}$$

is the best estimate of the region of asymptotic stability X of the set of decoupled subsystems (27.64). However, in Equation (27.64) we neglected the coupling terms $\left(f^i \cdot \Delta \hat{P}_i \right)$.

We should investigate how coupling influences the stability of the overall system S. Since $\lim_{\Delta x \to 0} \Delta \hat{P}_i \to 0$, the coupling influence can be estimated by the numbers ξ_{ij} $\left(\xi_{ij} \geq 0 \ \ for \ \ i \neq j \right)$ satisfying:[24]

$$(grad V_i^T) \, f^i \, \Delta \hat{P}_i(t \, , \Delta x) \leq \sum_{j=1}^{m} \xi_{ij} V_j \, , \qquad \forall i \in I_1 \, , \quad \forall t \in T \, , \ \forall \Delta x \in \tilde{X} - x^{\circ}(t) \, . \tag{27.73}$$

A sufficient condition for asymptotic stability of the overall system S in the region $\tilde{X}(0)$ is:[25]

$$G V_o < 0 \tag{27.74}$$

where V_o is the m × 1 vector and $V_o = (V_{1o}, \ldots, V_{mo})^T$, $V_o \in R^m$, and the elements of the m × m matrix G are defined as:

$$G_{ij} = -\Pi_i \delta_{ij} + \xi_{ij} \tag{27.75}$$

where δ_{ij} is the Kronecker symbol.

It is necessary to point out that Equation (27.74) is a sufficient but not necessary condition. If this condition is not fulfilled, we cannot say anything about the system stability. If Equation (27.74) is fulfilled, then $\tilde{X}(0)$ is an estimate of the region of the overall system stability. It is possible to estimate the region $\tilde{X}(t)$ that contains the system state during the tracking of the nominal trajectory $x^{\circ}(t)$ by:

$$\max_{i \in I} (V_i(\Delta x^i(t)) / V_{i0}) < \max_{i \in I} (V_i(\Delta x^i(0)) / V_{i0}) \exp(-\beta t) \tag{27.76}$$

where $\beta > 0$ can be computed from:

$$\beta = \min_{i \in I} (-V_{i0}^{-1} \sum_{j=1}^{m} G_{ij} V_{j0}) = \min_{i \in I} (\beta_i) \tag{27.77}$$

where $\beta_i = -V_{i0}^{-1} \sum_{j=1}^{m} G_{ij} V_{j0}$.

The inequality (27.76) is an estimation of shrinkage of the region $\tilde{X}(t)$ that contains a solution of the system S. Now the practical stability of the system can be checked out. If

$$X^I \subseteq \tilde{X}(0) \ \ and \ \ \tilde{X}(t) \subseteq X^t(t) \, , \quad \forall t \in T, \ \ \tilde{X}(t) \subseteq X^F(t) \, , \quad \forall t \in T_s \tag{27.78}$$

is satisfied, then the system S is practically stable around $x^o(t)$. If the local linear feedback controllers defined by Equation (27.63) are not sufficient to stabilize the system, an additional control input should be introduced. We may introduce the global control in the form (we use $\Delta \hat{P}_i^*$ instead of ΔP_c^{i*}):

© 2002 by CRC Press LLC

$$\Delta u_i^G = k_{i4}^G \Delta \hat{P}_i^* + k_{i5}^G \tag{27.79}$$

where k_{i4}^G and k_{i5}^G are the scalar gains defined by Equation (27.38). $\Delta \hat{P}_i^*$ represents a quantity that corresponds to the coupling $\Delta \hat{P}_i$. By measuring forces at the contact point of the sole of the supporting leg and the ground, we get information on the effects of coupling upon the unpowered joint ΔP_N^ℓ. Therefore, we can establish a global control from the unpowered joint to the one of the powered joints (i.e., to its actuator) and compensate for the effects of coupling upon the unpowered joint. If a global control is introduced, the stability analysis can be performed as described above. However, the numbers ξ_{ij}^* estimating coupling are now defined to satisfy [instead of (27.73)] the following condition:[13,17]

$$(grad\, V_i)^T f^i \Delta \hat{P}_i + \left(grad\, V_i\right)^T b^i \Delta u_i^G \le \sum_{j=1}^{m} \xi_{ij}^* V_j\,, \quad \forall i \in I_1\,, \forall t \in T\,, \quad \forall \Delta x \in \tilde{X} - x^\circ(t)\,. \tag{27.80}$$

The next step is to check the conditions (27.74), i.e., to test whether the system with the applied local and global control is asymptotically stable in the region $\tilde{X}(0)$. The ξ_{ij} in Equation (27.75) must be replaced by ξ_{ij}^*. If the global control is properly selected, ξ_{ij}^* must satisfy:

$$\xi_{ij}^* \le \xi_{ij}, \ \forall i, j \in I_1\,.$$

Therefore, the fulfillment of stability test is easier when the global control is introduced than if only the local control is applied.

27.4.3 Example

The scheme of the locomotion mechanism is the same as in Figure 27.22. Each powered joint is modeled as one subsystem; the composite subsystem comprises the models of one powered and one unpowered joint. The inactive DOFs are not included in the subsystem modeling. To make the examples of stability analysis easier to follow, a redrawn scheme of the same mechanism is presented in Figure 27.25, with only those DOFs that will be included in the stability analysis. All the joints represented by the unit rotational axes $\vec{e}_i'$ (i = 1, . ., 9) and the corresponding links are renumbered. The link representing the upper body comprises the trunk and both hands. We shall investigate system stability in the sagittal plane only, so that there is only one unpowered DOF. Thus, the mechanism considered here has nine DOFs (n = 9), and eight of them (m = 8) are powered. The elements of matrices of the actuator models and their distribution per joint are given in Table 27.3.

The nominal trajectories are synthesized using the prescribed synergy method. The control input to the *i*-th actuator consists of two parts:

$$u^i = u^{oi} + \Delta u^i \tag{27.81}$$

where u^{oi} is the nominal control input to the *i*-th actuator while Δu^i is the corrective input to the same actuator, synthesized at the level of perturbed regimes. The control law (27.37) holds for the subsystems i = (1, 2, . ., $2m - n$), and a similar control is derived for the composite subsystems, taking into account that $\Delta \hat{P}_i^*$ values for the composite subsystems are the (2×1) vectors.

In Equation (27.37), the part depending only on local states of the *i*-th joint corresponds to the local and the rest to the global control. The global control is introduced in the form of feedback with respect to both the driving torques $\Delta \hat{P}_i^*$, and the part k_{i5}^G. $\Delta \hat{P}_i^*$ represents the force feedback (i.e., the measured moments about joints). Additional feedback with respect to the ZMP position defined by Equation (27.45) is also available.

Let us determine the stability region X_i of the decoupled subsystem. Consider first the local control (27.37 and 27.38) that has to stabilize the decoupled subsystem. If we assume that the

© 2002 by CRC Press LLC

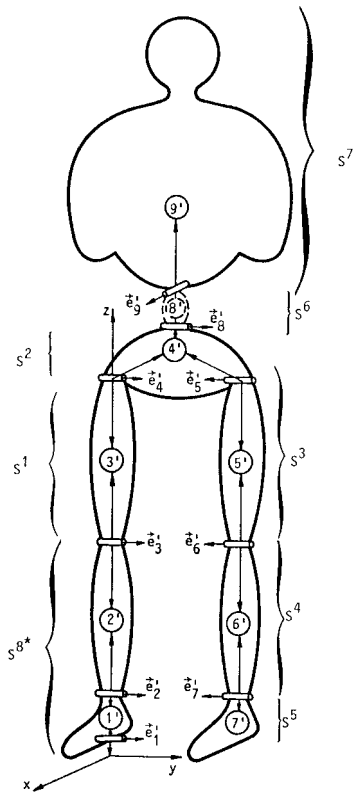


FIGURE 27.25 Simplified scheme of mechanical biped structure with disposition of the modeled subsystems.

TABLE 27.3 Actuator Parameters

	Term						
Actuator	$a_{2,2}$	$a_{2,3}$	$a_{3,2}$	$a_{3,3}$	b_3	f_2	Used at joint
M_1	–3.0	0.13	-10^5	–450	2000	$-7 \cdot 10^{-4}$	2, 3, 6, 7
M_2	–1.928	4.03	–6800	–264	400	–0.179	4, 5, 8, 9

complete state vector Δx^i is measurable, the closed-loop subsystem is given by (27.64). In the case of a stable subsystem, the poles must be at the left-hand side of the complex plane. If we denote the modulus of their real part by $|\sigma_p^i|$, the subsystem will be exponentially stable with a stability degree defined as:

$$\Pi_i = \min_{p=1,2,3} |\sigma_p^i| \tag{27.82}$$

which can be guaranteed only if the control inputs are within the limits:

$$|k_i^{LT} \Delta x^i| < \bar{u}_m^i = u_m^i - \max_{t \in T} |u^{\circ i}(t)| \ . \tag{27.83}$$

The actuator velocity–torque characteristics limit the values of the state coordinates. In view of these characteristics we can write:

$$|\bar{k}_i^2 \Delta \dot{q}^i + \bar{k}_i^3 \Delta i_R^i| \le \bar{k}_m^i \rightarrow |\bar{k}_i^T \Delta x^i| \le \bar{k}_m^i \tag{27.84}$$

© 2002 by CRC Press LLC

where $\bar{k}_i = (0, \bar{k}_i^2, \bar{k}_i^3)^T$ and $\bar{k}_m^i$ are defined by the motor characteristics. Further, the regions of allowable angle deviations for each DOF are introduced. In this way, the stability regions are constrained for both the powered and composite subsystems.

We may define a finite region X_i (according to (27.65)) in the state space R_i^n, in which the subsystem S^i is exponentially stable with a stability degree Π_i:

$$X_i = \left\{ \Delta x^i, \left| k_i^{LT} \Delta x^i \right| < \left| \bar{u}_m^i \right| \wedge \left| \bar{k}_i^T \Delta x^i \right| \le \bar{k}_m^i \right\}. \tag{27.85}$$

To investigate the stability of the whole system, the Lyapunov subsystem functions must be chosen according to Equation (27.69), taking into account the relation (27.68), which has to be satisfied in the region X_i. $\tilde{X}_i$ will be the estimate of X_i according to:

$$\tilde{X}_i = \left\{ \Delta x^i : V_i(\Delta x^i) \le V_{i\circ} \right\}, \qquad \forall i \in I_1 .$$

To investigate asymptotic stability of the overall system, the values ξ_{ik}^*, which estimate the subsystem coupling, have to be determined according to Equation (27.80). The expression for the composite subsystem is of the form:[13,22]

$$(grad\, V_i)^T \left[f_z^i \Delta P_z^{i*}(t, \Delta x) - b_z^i (k_{i4}^G \cdot \Delta \hat{P}_i^* + k_{i5}^G) \right] \le \sum_{k=1}^{m} \xi_{ik}^* V_k \,,\ i = 2m - n + 1, \ldots, m$$

where the global control by both ΔP_N^ℓ and ΔP_c^i is introduced. If for ξ_{ij}^* thus defined, the condition (27.74) is satisfied, it can be claimed that region $\tilde{X}$, defined by (27.72), is an estimate of the region of the overall system stability.

We will form the composite subsystem model of the models of the unpowered DOFs and ankle joint. The powered subsystem model S^1 corresponds to the model of joint 3 powered by the actuator, S^2 corresponds to joint 4, S^3 to joint 5, S^4 to joint 6, S^5 to joint 7, S^6 to joint 8, S^7 to joint 9. The last model of the powered subsystem, S^8, will be included in the composite subsystem. Thus, the model of joint 2 with the model of the corresponding actuator corresponds to S^8. The model of composite subsystem will be denoted S^{8*} and will comprise the models of the unpowered subsystem and S^8. The composite subsystem matrices A_z^{8*} and f_z^8 and vector b_z^8 are defined as:

$$A_z^{8*} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.882 & 0 & -80 & -10 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 300 & 100 & 0 & -3 & 0.13 \\ 0 & 0 & 0 & -100000 & -450 \end{bmatrix}, \quad f_z^8 = \begin{bmatrix} 0 & 0 \\ 0.01387 & 0 \\ 0 & 0 \\ 0 & 0.0007 \\ 0 & 0 \end{bmatrix}, \quad b_z^8 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 2000 \end{bmatrix}.$$

The vectors from Equation (27.55) are $D_1^8 = [-80, -10, 0]$ and $D_2^8 = [300, 100]$. The Lyapunov functions of all subsystems are selected in the form of Equation (27.69). The matrices H^i are selected to satisfy Equation (27.70) and they are obtained as:

$$H^8 = \begin{bmatrix} 100422\ .00 & 33545\ .10 & 8660\ .65 & -98\ .68 & 0.02 \\ 33545.10 & 11425.23 & -3273.52 & -335.46 & -0.07 \\ 8660.65 & -3273.52 & 183099.94 & 8965.84 & 1.95 \\ -98.68 & -335.46 & 8965.84 & 1097.767 & 0.086 \\ 0.02 & -0.07 & 1.95 & 0.086 & 0.0009 \end{bmatrix}.$$

© 2002 by CRC Press LLC

The Lyapunov matrices corresponding to the models of powered subsystems S^i are:

$$H^i = \begin{bmatrix} 62777.42 & 2291.71 & 0.56 \\ 2291.71 & 161.15 & 0.02 \\ 0.56 & 0.02 & 0.00011 \end{bmatrix} \quad i = 1, 4, 5 \quad H^i = \begin{bmatrix} 32912.207 & 408.621 & 4.535 \\ 408.621 & 6.482 & 0.065 \\ 4.535 & 0.065 & 0.00092 \end{bmatrix} \quad i = 2, 3, 6, 7.$$

The regions of joint angle deviations (the superscripts correspond to the subsystems model numbers) in which stability is investigated are (in radians):

$$\Delta q^1 = \pm 0.044, \quad \Delta q^2 = \pm 0.0422, \quad \Delta q^3 = \pm 0.0126, \quad \Delta q^4 = \pm 0.03,$$

$$\Delta q^5 = \pm 0.03, \quad \Delta q^6 = \pm 0.099, \quad \Delta q^7 = \pm 0.01, \quad \Delta q_z^1 = \pm 0.01, \quad \Delta q_z^2 = \pm 0.077$$

where Δq_z^1 and Δq_z^2 correspond to the joints comprising the composite subsystem, i.e., the unpowered and powered DOFs (the ankle joint with the axis of rotation $\vec{e}_2$).

The constants V_{io} defining the estimates of the stability regions $\tilde{X}_i$ are computed as:

$$V_{30} = 0.65, \quad V_{20} = 1.0399, \quad V_{30} = 0.4256, \quad V_{40} = 0.1545,$$

$$V_{50} = 0.75166, \quad V_{60} = 0.7395, \quad V_{70} = 0.1814, \quad V_{80} = 0.3291 .$$

The constant V_{8o} corresponds to the composite subsystem. The results of stability analysis are presented in Table 27.4. Three types of control laws are investigated:

1. The complete feedback structure defined by Equation (27.37) plus the global control with respect to ZMP displacement defined by Equation (27.44).
2. The local control is introduced (the corresponding gains k_{i1}^L, k_{i2}^L, and k_{i3}^L are defined in (27.38) plus global control with respect to ZMP position (27.44).
3. Only local control from Equation (27.37) is introduced.

Table 27.4 presents results (matrix G, and vectors Gv_o and η) for all three control laws. The first three rows correspond to laws 1 through 3, respectively. The first and second groups of three rows each correspond to the hip and knee of the supporting leg. The third, fourth, and fifth groups represent the hip, knee, and ankle of the leg in swing phase. The sixth and seventh groups correspond to trunk motion in the frontal and sagittal planes. The last group of three rows represents the composite subsystem.

To draw a final conclusion about system stability, the product $G_o \cdot V_o$ must be observed. If this product is negative, the stability under the given conditions is proved. The vector η represents shrinkage of the bounds of the regions $\tilde{X}(t)$.

27.5 Realization of Anthropomorphic Mechanisms and Humanoid Robots

Locomotion activity, especially the human gait, belongs to a class of highly automated motion. Bernstein[26] was the first to comprehend this fact. Humans have at their disposal to achieve complete skeletal (locomotion–manipulation) activity several hundred muscles that allow over 300 equivalent DOFs. In view of the high number of biological actuators through which humans exercise motor activity, the imitation of this activity seems to be a hopelessly difficult task. Understanding the mechanisms of gait control and other skeletal activities is extremely difficult, especially if one bears in mind the necessarily detailed insight into the multilevel structures to

© 2002 by CRC Press LLC

TABLE 27.4 Results of Stability Analysis (Composite System Consisting of Ankle and Unpowered Joint)

$$G = \begin{bmatrix}
-53.92 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-50.91 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-50.91 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -2404.16 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -2360.11 & 0 & 0 & 0 & 0.172 & 0.003 & 0 \\
0 & -2360.11 & 0 & 0 & 0 & 0 & 0 & 0 \\
6.021 & 0 & -76.15 & 0 & 0 & 0 & 0 & 0 \\
6.491 & 0 & -73.88 & 0 & 0 & 0 & 0 & 0 \\
6.491 & 0 & -73.88 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -5114.02 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -4761.15 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -4761.15 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -789.58 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -764.16 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -764.16 & 0 & 0 & 0 \\
6.351 & 0 & 0 & 0 & 0 & -67.56 & 0 & 0 \\
6.767 & 0 & 0 & 0 & 0 & -66.40 & 0 & 0 \\
6.767 & 0 & 0 & 0 & 0 & -66.40 & 0 & 0 \\
0 & 0.89 & 0 & 0 & 0 & 0 & -32.84 & 0 \\
0.045 & 1.547 & 0 & 0 & 0 & 0.198 & -28.86 & 0.109 \\
0.045 & 1.547 & 0 & 0 & 0 & 0 & -28.86 & 0.109 \\
1.236 & 16.534 & 0 & 16.375 & 7.77 & 11.14 & 70.52 & -218.31 \\
1.544 & 17.505 & 0 & 20.465 & 6.93 & 11.99 & 67.1 & -216.53 \\
3.242 & 22.87 & 3.8 & 43.052 & 2.29 & 16.72 & 47.86 & -206.7
\end{bmatrix}$$

$$G \cdot v_O = \begin{bmatrix}
-110.81 & -1564.4 & -20.03 & -790.28 & -593.5 & -36.91 & -5.38 & -34.82 \\
-104.63 & -1535.6 & -18.1 & -735.75 & -574.4 & -35.2 & -3.95 & -32.93 \\
-104.63 & -1535.6 & -18.1 & -735.75 & -574.4 & -35.2 & -3.95 & -20.84
\end{bmatrix}^T$$

$$\eta = \begin{bmatrix}
53.92 & 2404.16 & 47.08 & 5114.01 & 789.6 & 49.9 & 29.64 & 98.04 \\
50.91 & 2359.92 & 42.54 & 4761.14 & 764.16 & 47.6 & 21.78 & 92.7 \\
50.91 & 2359.92 & 42.54 & 4761.14 & 764.16 & 47.6 & 21.78 & 58.67
\end{bmatrix}^T$$

control human locomotion and manipulation activity. Despite these challenges, technological advances have allowed us to create mechanical counterparts to humans that are capable of performing some human tasks. This section will briefly describe active anthropomorphic mechanisms and humanoid robots.

© 2002 by CRC Press LLC

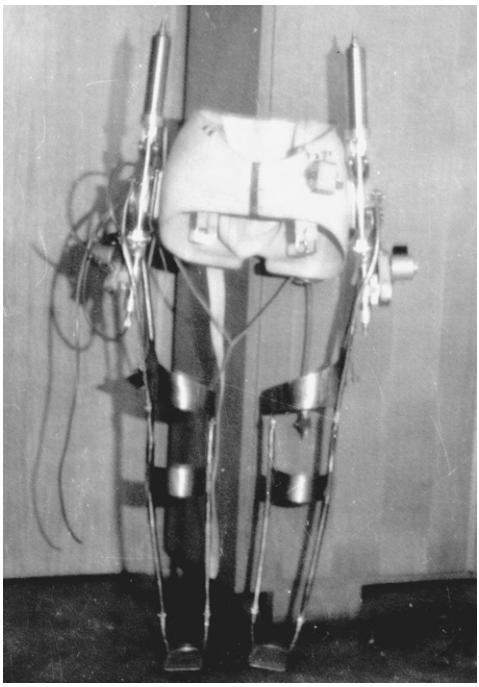


FIGURE 27.26 Kinematic walker (1969).

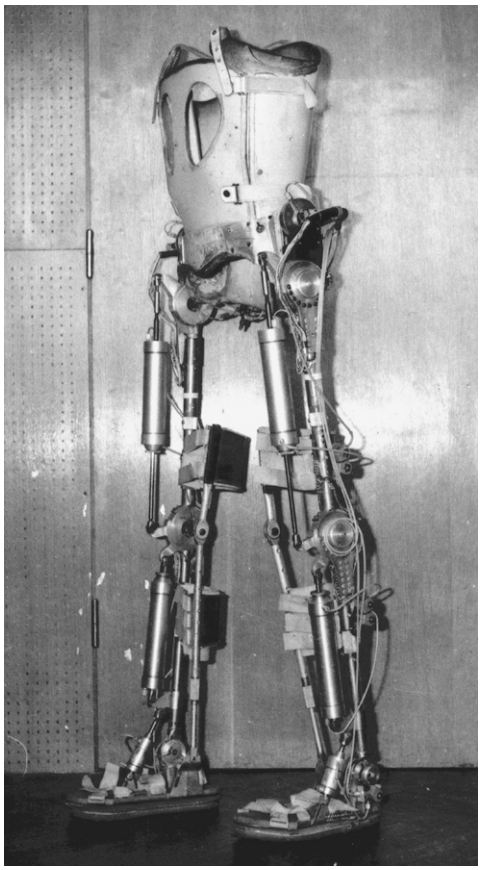


FIGURE 27.27 Complete pneumatically driven active exoskeleton.

27.5.1 Active Exoskeletons

The first pneumatically driven active anthropomorphic exoskeleton (Figure 27.26) was created in 1969 at the Mihajlo Pupin Institute (Belgrade).[13] Each leg had its own drive consisting of only one pneumatic cylinder, which, through kinematic linkage, actuated also the knee, while the foot was passive. The device produced a sliding-foot gait.

After this first attempt of anthropomorphic gait the researchers in the Robotics Laboratory of the Mihajlo Pupin Institute also developed new models of partial and complete exoskeletons dedicated to the restoration of basic locomotion activities of paraplegics. A complete pneumatically driven active exoskeleton for rehabilitation use was developed and manufactured in 1972.[13,27,28] It is depicted in Figure 27.27.

The first experiments with these exoskeletons were successfully carried out at the Belgrade Orthopedic Clinic during 1972 and 1973. It was concluded in 1973 that further development of pneumatically driven active exoskeletons for rehabilitation was not promising because of poor energy autonomy provided by the pneumatic source of energy. The decision was made to switch to the development of an active exoskeleton with electromechanical drives.[13,27,28]

In 1974, the first electromechanical active exoskeleton (Figure 27.28) was realized. It included a number of sophisticated solutions in the design of the exoskeleton structure. In addition to the basic motion of the leg in the sagittal plane, the hip joint performed two kinematically programmed

© 2002 by CRC Press LLC

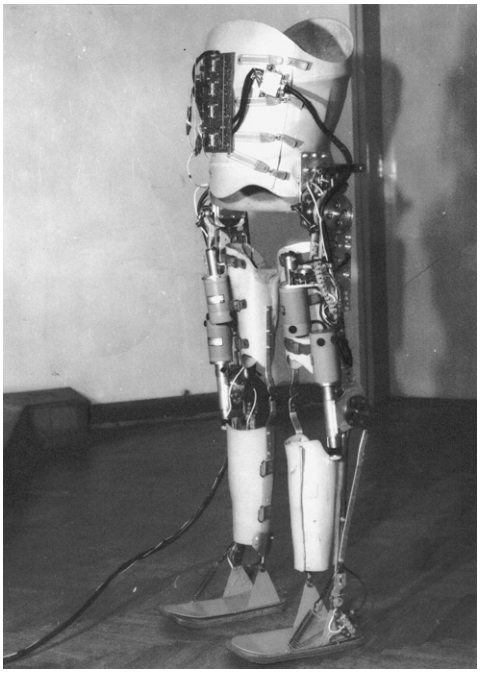


FIGURE 27.28 Complete electromechanical exoskeleton.

motions, pelvic twist and sideways motion, to achieve more stable and quasi-realistic gait. When this exoskeleton type was realized, light batteries were not available to power the devices and computer technology was in its pre-microprocessor stage. For these reasons, the power source and control system were outside the exoskeleton, and this greatly limited the utility of the device.

However, it should be pointed out that the above technological limitations have been overcome. An early conception of artificial gait for rehabilitation purposes, having in mind the real possibility of active exoskeleton autonomy with respect to control and energy, will be realizable in a much more successful way.

27.5.2 Humanoid Robots

High expectations accompanied the appearance of robots for personal use. Such robots coexist with humans and provide support for the aged and the physically handicapped.

An anthropomorphic form of robot is usually expected when the robot is intended for personal use. A human-like or humanoid robot that works with humans as a partner in the living environment needs to share the same workspace and possess the common thinking and behavioral modes of humans. A humanoid robot will integrate information from its sensors and coordinate its actions to realize high level and natural communication with humans by using speech, facial expression, and body motion.

The Waseda University has been one of the leading research centers working on anthropomorphic robots since I. Kato and his colleagues started the WABOT (WAseda roBOT) project and produced the first biped walking robot, WABOT-1, in 1973 (Figure 27.29).[29]

In 1984, A. Takanishi and his co-workers developed a hydraulic biped robot (Waseda Leg 10 Refined Dynamic or WL-10RD) that was able to walk dynamically. By 1985, their dynamic biped had the ability to walk on slopes and stairs. Takanishi's group achieved compensation by lower limbs and trunk in 1986. They introduced an effective computation algorithm to obtain the periodic

© 2002 by CRC Press LLC

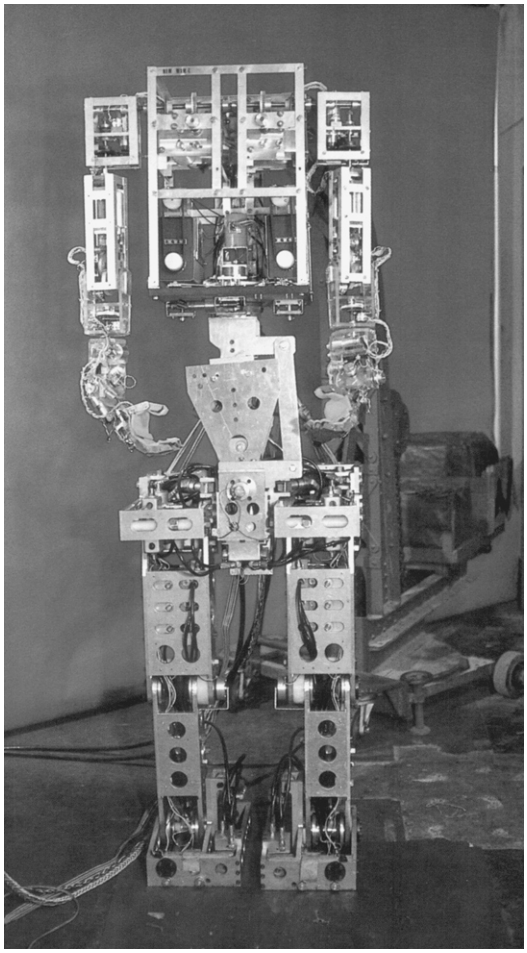


FIGURE 27.29 WABOT-1 (1973).

upper-body trajectory from the arbitrarily preset ZMP and leg trajectories. Their work was the first example of implementation of the ZMP concept into a humanoid robot. The WL-12 series robots achieved dynamic walking and turning under conditions of unknown external forces and on unknown surfaces (Figure 27.30).

Takanishi and his co-workers also developed a bipedal humanoid robot called WABIAN (WAseda BIpedal humANoid) and the method of its control (Figure 27.31). They accomplished the following design goals: (1) develop an electrically powered bipedal humanoid robot having upper limbs, a three DOF trunk, and a head (the total number of active DOFs is 35), (2) devise a motion pattern-generating program to study overall cooperative motion, and (3) support the effectiveness of a dynamic walking system that allows walking, dancing with the motion of the three DOF trunk and upper limbs, carrying of a load with both hands, and continuous bipedal walking in human living space based on a closed-loop dynamic walking control method.[29,31]

The most successful representation of a humanoid robot is the Honda humanoid robot.[32] The goal was to develop a robot capable of coexisting and collaborating with humans and performing tasks that humans cannot perform. Honda wanted to develop a new robot to meet consumers' needs — not a robot limited to specialized operations. Such a robot is capable of moving around the house, encountering various obstacles such as staircases, doors, furniture, etc.

© 2002 by CRC Press LLC

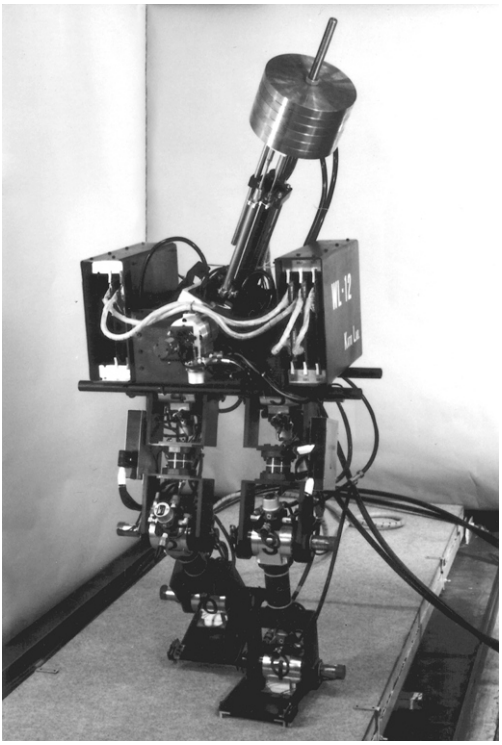


FIGURE 27.30 Dynamic walking with a compensating body: WL-12 (1986).

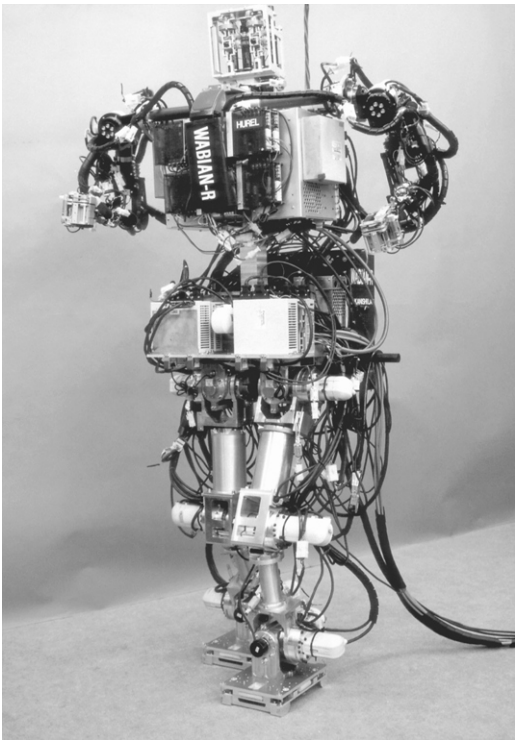


FIGURE 27.31 Bipedal humanoid robot WABIAN.

The legs of the Honda robot have 12 DOFs and the redundant arms have 14 DOFs. The hand is similar to a two-finger gripper with two DOFs. The head has two DOFs, bringing the total number of DOFs to 32 (Figure 27.32). Honda developed several types of biped robots. The maximum walking speed achieved by a specialized robot was 4.7 km/h. To maintain dynamical equilibrium and stabilize the perturbed work regime of the robot, Honda concluded that the robotic system required a body inclination sensor and a ground reaction sensor for each foot. The inclination sensor consists of three accelerometers and three angular speed sensors (optical fiber gyros). It is also used as a navigation sensor. Each foot and wrist has a six-axis (three component forces and three component moments) sensor. The robot is also equipped with an impact absorption mechanism to damp the landing impact ground reaction force. The overall height of the robot is 1820 mm, its width is 600 mm, and it weighs 210 kg.

To recover the robots's posture, the ground reaction force control shifts the total ground reaction force to an appropriate position by adjusting each foot's desired position and posture. Model ZMP is used to control the shifting of the desired ZMP to an appropriate position in order to recover posture. The foot landing position control corrects the relative position of the upper body and feet in conjunction with the ZMP control. By having the three controls working simultaneously Honda has realized the robot with a posture stabilizing control similar to that of the human.

Honda has continued its research on biped walking humanoid robots. In 2000, it developed the ASIMO (Advanced Step in Innovative Mobility) humanoid robot.[33] It has an overall height of 1200 mm, weighs 43 kg, and has 26 DOFs. The robot is compatible with human living environments. The walking technology includes behavior prediction such that the robot can change its walking motion freely and smoothly without interruption. A central role in the control strategy is played by the model ZMP control.

© 2002 by CRC Press LLC

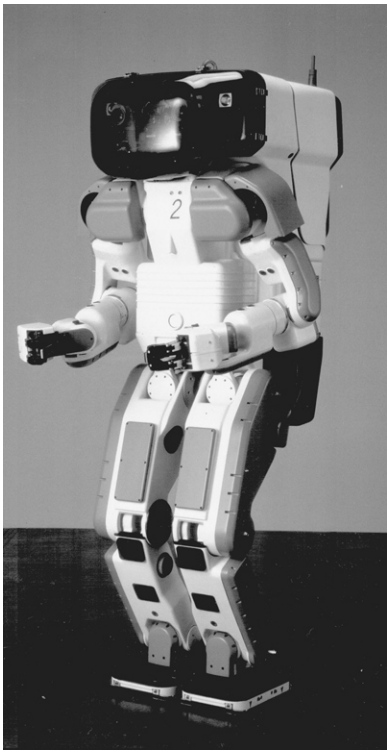


FIGURE 27.32 Honda humanoid robot P2.

27.5.3 Virtual Humanoid Robot Platform

Nakamura and Takanishi and a group of associates developed the most complete software tool for modeling and control of humanoid robots reported to date. They developed a simulator of humanoid robots and a controller of whole body motion.[34] The basic modules of this software include:

1. A *dynamic simulator* that executes efficient dynamics and kinematics and can accomodate structure changes of any open or closed kinematic chain, and even such kinematic chains as to change connectivity in operation. The connectivity change function is essential because it is often seen when a humanoid walks, touches or holds the environment, grasps an object with the both hands, and is even connected with another humanoid.
2. A *view simulator* or image synthesis that consists of modeling, illumination, shapes and materials of objects in a scene, and cameras. The shapes of artificial objects can be obtained from CAD data, but it is hard to produce material models of surfaces. The simulator can generate sequences of the fields of view from the eyes of the robot according to the dynamics simulation. When the view simulator is integrated with the dynamics simulator, visual feedback of humanoid robots can be simulated.
3. A *humanoid motion controller* that can handle biped locomotion, dynamic balance control at the standing position, and collision avoidance.

As part of the Virtual Humanoid Robot Platform (V-HRP) project, a motion controller has been developed to achieve biped locomotion adaptive to terrain, including walking straight, turning, going up or down the stairs, and walking on rugged ground.[34] With this programming library, complex locomotion can be realized as a sequence of basic motion patterns. The link between the basic motions of the robot is automatically generated for continuous motion control. Control data

© 2002 by CRC Press LLC

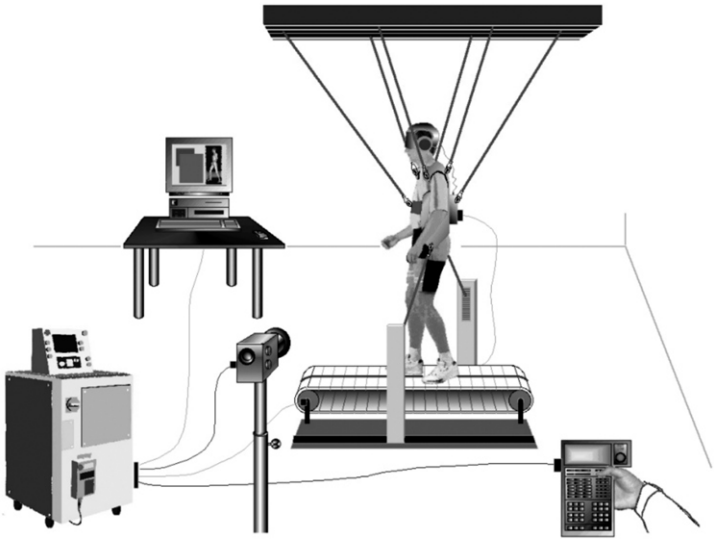


FIGURE 27.33 Global RehaRob concept.

for a walk adaptable to terrain generated by the library have been examined and found to be consistent with the mechanism's dynamics obtained using a dynamics simulator. The data's consistency has also been proven by experiment using the hardware model developed to verify the compatibility of the simulation model with the real world.

Concerning this module, two examples were presented. In the first example the robot is standing on two legs, and both legs are controlled in the same manner. With the proposed balance control, the robot can successfully sit down, reach the ground with its arms, and stand up again. To demonstrate 3-D balance, a kicking motion was tested. The robot can fully swing its left leg in one second while balancing with its right leg. With the proposed control, the robot is capable of successful kicking and balancing. The motions of the arms and body were added to provide a natural appearance. All compensation is done by the ankle actuators of the supporting leg.

These software modules are integrated via CORBA (Common Object Request Broker Architecture). This enables Internet clients to use the software. The modules are implemented as CORBA servers, and a client can utilize them if the servers are accessible via the IIOP (Internet Inter Orb Protocol).

The developers of the Humanoid Robot Platform expect it to be "the common base of humanoid robotics research focusing on software development for the community."

27.5.4 New Application of the ZMP Concept in Human Gait Restoration

A novel application of ZMP for human gait rehabilitation using treadmill training and partial body weight bearing (PWB) has been proposed.[12] This methodology has recently been successfully used for gait recovery by stroke patients.[35]

One of the Wisa-ROMED projects endowed by the Fraunhofer Community developed a demonstration system referred to as RehaRob which represents the first application of the ZMP concept for evaluation and control of the human gait. The RehaRob is a powerful robotic system for supporting gait rehabilitation and restoration of motor functions. It combines the advantages of PWB with a number of robotic and humanoid control functions. Safe, reliable, and dynamically controlled weight suspension and posture control systems support patients and allow them to autonomously recover their gaits early in the rehabilitation stage.

The global RehaRob architecture is presented in Figure 27.33. The system consists of an active weight-relieving robotic system (wire robot) that performs partial dynamic weight compensation

© 2002 by CRC Press LLC

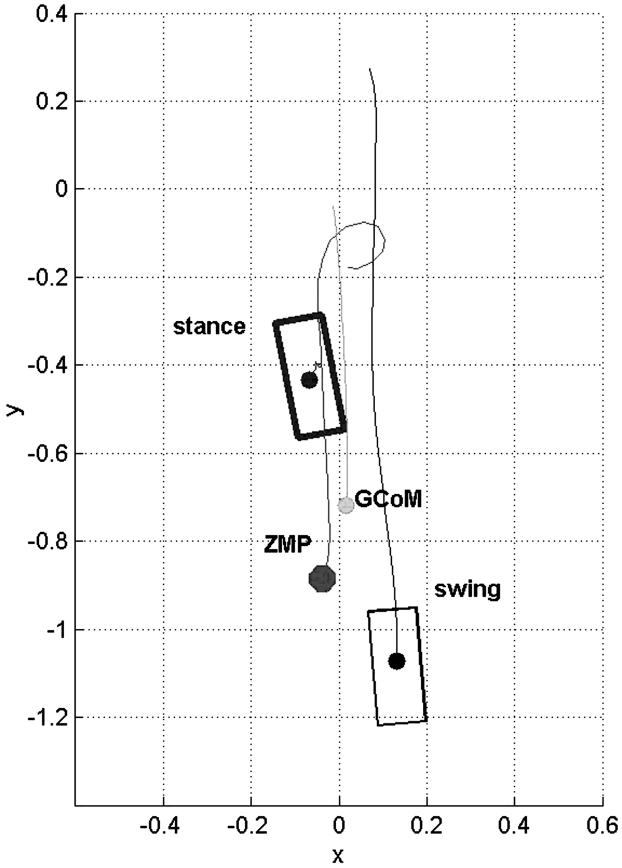


FIGURE 27.34 ZMP and GCoM trajectories during single limb phase.

and posture control synchronously with the human lower limb motion; a harness system (patient interface); a treadmill and/or lifting system axial motion device (for rehabilitation of orthopedic patients) that supports repetitive motion progression; a sensory system (motion camera, insole pressure sensors, force sensors, inclinometers, wire position sensors) that collects data about the human gait and provides feedback to the control system; a system controlling the wire robot, treadmill (the axial motion system), rehabilitation planning, and programming system (user interface); an AR or VR system providing visual feedback; computer safety control; and a mechanical system providing exception-handling functions.

The robot wires are connected to the trunk and pelvis at optimized attachment points (in the system under development, a total of ten wires are applied). The robot exerts active external forces upon the trunk and pelvis to reduce the weight on the lower limbs (reaction force) and balance the posture, thus essentially supporting the gait. Redundant wires are needed to ensure tension in all wires independent of dynamic loads. The rigid trunk–pelvis system connected by a spherical joint has nine DOFs.

The RehaRob control is based on the ZMP concept. It utilizes wire force, foot reactions measurements, and a model of wire robot and human interaction to estimate and control ZMP. The application of the human motion ZMP trajectory for controlling a biped robot has recently been proposed.[36] For body modeling, the RehaRob uses a rigid model of a human developed with MATLAB (MatMan). The model has 37 DOFs. Figure 27.34 presents the results of simulation of the ZMP and ground projection of CoM (GCoM) trajectories for the human gait during stance phase (stance and swing legs are denoted in the figure). Apparently, in a period of time the ZMP is within the stance foot supporting area, while in the remaining time it leaves this area following

© 2002 by CRC Press LLC

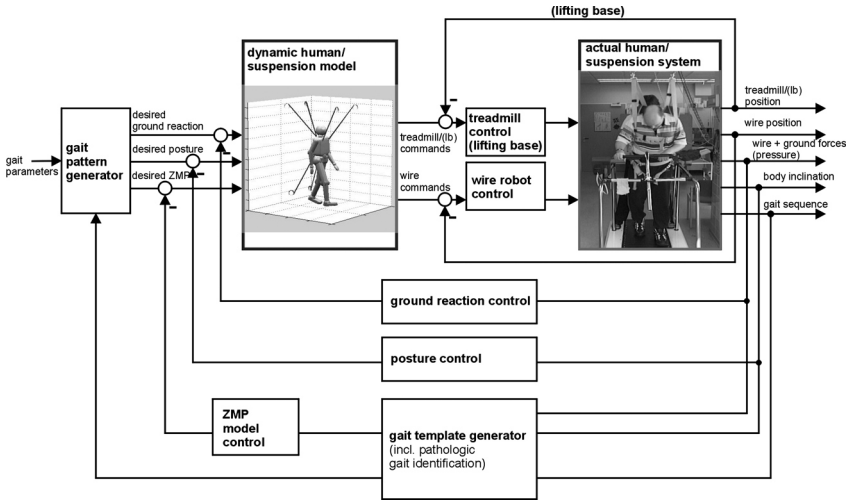


FIGURE 27.35 RehaRobot control concept.

the moving swing foot. In a stable gait, the ZMP remains within the enveloping area constrained by foot projections on the ground.

Unlike the ZMP model of human/humanoid walking (Equation 27.7), the RehaRob system includes additional wire-active forces affecting the equilibrium conditions (27.2). By means of the wire forces, it is possible to control both body reaction and ZMP location.

The motion of the relevant upper and lower limbs (e.g., knee) affecting system dynamics can be measured by relatively inexpensive sensors. The trunk and pelvis positions in the RehaRob system are directly measured and controlled using both wires and body sensors. To cope with model inaccuracies and ZMP estimation errors, the RehaRob system implements a relatively complex control structure closing several control loops (Figure 27.35) around reaction force and gait posture and uses the internal wire robot and treadmill control. This control scheme includes the basic gait pattern generator (initial contact, stance, swing, single limb), which, based on the captured gait state and required weight suspension percentage generates the desired ground reaction, as well as nominal posture and ZMP location data. These values are compared with the measured (i.e., estimated) ones, and the control feedback is closed around the dynamic human gait and wire robot models. This provides the inputs to the internal wire and treadmill control loops (treadmill velocity, pulley position, and wire forces). The local controllers control the wire robot system so that the posture can support and follow the joint motion of the desired walking pattern. This pattern is a combination of ideal walking patterns (templates) including desired weight suspension and the subject's gait performance estimated using the sensory system. Gait balance is achieved by the ZMP and posture controls for the generated pattern. The gait template generator includes data about the subject's abnormal gait, as well as the emergency and exception-handling strategies (to compensate for 100% weight upon the stance leg if the conditions for single-leg support are not available, for example, due to improper ankle or knee joint position, etc.). This control scheme is similar to the recent humanoid control approaches proposed in Hirai et al.[32] and Yamaguchi et al.[37]

27.6 Conclusion

Having at our disposal limited space in a thematically widely conceived handbook, it was a difficult task to present such a broad and attractive field of scientific and professional interest (one recently experiencing tremendous impetus) in a way that would be both an introduction and offer in-depth coverage of humanoid robotics.

© 2002 by CRC Press LLC

Because of this, some new challenging aspects and dilemmas concerning humanoid robotics had to be omitted, as well as new views on its current importance and role in the future. Some examples include issues such as the new frontiers of humanoid robotics, human–humanoid interactions, using humanoid robots to study human behavior, humanoid features in public places, a neurobiological perspective on humanoid robot design, humanoid robot cognition. Because the character of *Mechanical Design Handbook: Modeling, Measurement, and Control* is mostly determined by dynamics, dynamic control, and advanced design diverse types of objects and systems, the authors believe they need to mention, at least briefly, some of the phenomena pertaining to humanoid robots that deserve detailed studies to make these robots more suitable for use.

The above mostly relates to refining the trajectory of the zero-moment point, especially when the gait passes from the single-support to the double-support phase. It is then that the introduction of the semi-rigid foot, in contrast to its rigid version, offers the possibility of a more faithful representation of the perturbation state of the humanoid robot to prevent the robot instantaneously reaching its foot edge — the case that has been considered up to now.

Further very important improvements are related to a more reliable description of the constraint environment model, which enables more realistic insight into robot–environment interactions that offer the possibility of control dynamic performance, e.g., by reducing the dynamic impact of the robot's foot at the end of the swing phase, which is achieved by applying active dampers at the ankle joint contstruction, as well as by passive or semi-active dampers at the other joints of the mechanical construction.

Finally, let us emphasize once more that the gait of humanoid robots is an extremely complex contact task involving a mobile object whose dynamics include interaction with its environment's dynamics, which means that (among other things) it is necessary to ensure simultaneous control with respect to both position and contact force. Some preliminary results indicate justification of such an approach, [38,39] whereas more extensive results will be achieved by a more faithful analysis of dynamic contact and the synthesis of the appropriate laws of simultaneous dynamic position force control.[40]

References

1. Vukobratović, M. and Juričič, D., Contribution to the synthesis of biped gait, *IEEE Trans. Biomedical Eng.,* 16(1), 1969.
2. Juričič, D. and Vukobratović, M., *Mathematical Modeling of Biped Walking Systems,* ASME Publication 72-WA/BHF-13, 1972.
3. Vukobratović, M. and Stepanenko, Yu., On the stability of anthropomorphic systems, *Mathematical Biosciences,* 15, 1, 1972,
4. Vukobratović, M. and Stepanenko, Yu., Mathematical models of general anthropomorphic systems, *Mathematical Biosciences,* 17, 191, 1973.
5. Vukobratović, M., How to control the artificial anthropomorphic systems, *IEEE Trans. Syst., Man, Cybernetics, SMC-3,* 497, 1973.
6. Arakawa, T. and Fukuda, T., Natural motion of biped locomotion robot using hierarchical trajectory generation method consisting of GA, EP, layers, *Proc. IEEE Conf. Automation Robotics,* Albuquerque, NM, 211, 1997.
7. Inoue, K., Yoshida, H., Arai, T., and Mae, Y., Mobile manipulation of humanoids — real time control based on manipulability and stability, *Proc. IEEE Int. Conf. Robotics Automation,* San Francisco, 2217, 2000.
8. Huang, Q., Kajita, S., Koyachi, N., Kaneko, K., Yokoi, K., Arai, H., Komoriya, K., and Tanie, K., A high stability, smooth walking pattern for a biped robot, *Proc. IEEE Int. Conf. Automation Robotics,* Detroit, 65, 1999.
9. Yagi, M. and Lumelsky, Biped robot locomotion in scenes with unknown obstacles, *Proc. IEEE Int. Conf. Automation Robotics,* Detroit, 375, 1999.

© 2002 by CRC Press LLC

10. Fujimoto, Y., Obata, S., and Kawamura, A., Robust biped walking with active interaction control between foot and ground, *Proc. IEEE Int. Conf. Robotics Automation,* Leuven, Belgium, 2030, 1988.
11. Fukuda, T., Komata, Y., and Arakawa, T., Stabilization control of biped locomotion robot base learning with GAs having self-adaptive mutation and recurrent neutral networks, *Proc. IEEE Int. Conf. Robotics Automation,* Albuquerque, NM, 217, 1997.
12. Šurdilović, D. and Bernhardt, R., Robust control of dynamic interaction between robot and human: application in medical robotics, *Proc. German Robotic Conf.,* 429, 2000.
13. Vukobratović, M., Borovać, B., Surla, D., and Stokić, D., *Scientific Fundamentals of Robotics, Vol. 7, Biped Locomotion — Dynamics, Stability, Control, and Application*, Springer-Verlag, Berlin, 1990.
14. Stepanenko, Yu. and Vukobratović, M., Dynamics of articulated open chain active mechanisms, *Mathematical Biosciences,* 28(1/2), 1976.
15. Vukobratović, M. and Stokić, D., One engineering concept of dynamic control of manipulators, *Trans. ASME J. Dynamic Syst., Meas. Control,* 102, June 1981.
16. Vukobratović, M. and Stokić, D., Is dynamic control needed in robotic systems and if so, to what extent? *Int. J. Robotics Research,* 2(2), 18–34, 1983.
17. Vukobratović, M. and Stokić, D., *Scientific Fundamentals of Robotics, Vol. 2, Control of Manipulation Robots: Theory and Application*, Springer-Verlag, Berlin, 1982.
18. Vukobratović, M. and Stokić, D., Suboptimal synthesis of robot decentralized control for large-scale mechanical systems, *IFAC Automatica,* 20(6), 803, 1984.
19. Borovać, B., Vukobratović, M., and Surla, D., An approach to biped control synthesis, *Robotica,* 7, 231–241, 1989.
20. Vukobratović, M. and Stokić, D., Significance of the force-feedback in realizing movements of extremities, Trans. Biomedical Eng., 27(12), 705, 1980.
21. Stokić, D. and Vukobratović, M., Practical stabilization of robotic systems by decentralized control, *IFAC Automatica,* 20(3), 1984.
22. Borovać, B., Vukobratović, M., and Stokić, D., Stability analysis for mechanisms with unpowered degrees of freedom, *Robotica,* 7, 349, 1989.
23. Šiljak, D.D., *Large Scale Dynamic Systems: Stability and Structure*, North-Holland, Amsterdam, 1978.
24. Morari, M., Stephanopulos, G., and Aris, R, Finite stability regions for large scale systems with stable and unstable systems, *Int. J. Control,* 26(5), 1977.
25. Weissenberger, S., Stability regions of large-scale systems, *Automatica,* 9, 653, 1973.
26. Bernstein, N.A., *On the Motion Synthesis*, Medgiz Moscow, 1947 (in Russian).
27. Vukobratović, M., Hristić, D., and Stojiljović, Z., Development of active anthropomorphic exoskeletons, *Med. Biol. Eng.*, 12(1), 1974.
28. Vukobratović, M. and Hristić, D., Active orthoses of lower extremities, *Orthopedic Technic,* 4, 1985.
29. Takanishi, A., Humanoid robots — a new tide towards the next century for natural human-robot collaboration, *Proc. Int. Conf. Humanoids,* Tokyo, 2000.
30. Takanishi, A., Ishida, M., Yamazaki, Y., and Kato, I., The realization of dynamic walking by the biped walking robot WL-10RD, *Proc. 1985 ICAR,* 459, 1985.
31. Lim, H., Ishiji, A., and Takanishi, A., Emotion based walking of bipedal humanoid robot, *Proc. 13-th CISM-IFToMM Symp. Theory Practice Robots Manipulators.* Springer-Verlag, Berlin, 295, 2000.
32. Hirai, K., Hirose, M., Haikawa, Y., and Takenaka, T., The development of honda humanoid robot, *Proc. 1998 IEEE Int. Conf. Robotics Automation,* Leuven, Belgium, 1321, May 1998.
33. Technical Review, Honda R&D, 13(1), April 2000.
34. Nakamura, Y., Hirukawa, H., Yamane, K., Kajita, S., Yokoi, K., Tanie, K., Fujie, M., Takanishi, A., Fujiwara, K., Suehiro, T., Kita, N., Kita, Y., Hirai, S., Nagashima, F., Murase, Y., Inoba, M., and Inoue, H., V-HRP: virtual humanoid robot platform, *Proc. Int. Conf. Humanoids,* Tokyo, Japan, 2000.

© 2002 by CRC Press LLC

35. Kawamura, J., Ide, T., Hayashi, S., Ono, H., and Honda, T., Automatic suspension device for gait training, *Prosthetics and Orthotics Int.,* 120, 1993.
36. Dasgupta, A. and Nakamura, Y., Making feasible walking motion of humanoid robots from human motion capture data, *Proc. IEEE Int. Conf. Robotics Automation,* Detroit, 1044, 1999.
37. Yamaguchi, J., Soga, E., Inoue, S., and Takanishi, A., Development of a bipedal humanoid robot — control method of whole body cooperative dynamic biped walking, *Proc. IEEE Int. Conf. Robotics Automation,* Detroit, 368, 1999.
38. Park, J. H. and Chung, H., Hybrid control for biped robots using impedance control and computed torque control, *Proc. IEEE Int. Conf. Robotics Automation*, Detroit, Michigan, 1365–1370, 1999.
39. Silva, F. M. and Machado, J. A. T., Goal-oriented biped walking based on force interaction control, *Proc. IEEE Int. Conf. Robotics Automation*, Seoul, Korea, 4122–4129, 2001.
40. Vukobratović, M. and Ekalo, Yu., New approach to control of robotic manipulators interacting with dynamic environments, *Robotica*, 14, 3139, 1996.

© 2002 by CRC Press LLC

28

Present State and Future Trends in Mechanical Systems Design for Robot Application

28.1 Introduction
28.2 Industrial Robots
Definition and Applications of Industrial Robots • Robot Kinematic Design • Industrial Robot Application
28.3 Service Robots
From Industrial Robots to Service Robots • Examples of Service Robot Systems • Case Study: A Robot System for Automatic Refueling

Martin Hägele
Fraunhofer Institute

Rolf Dieter Schraft
Fraunhofer Institute

28.1 Introduction

In 1999 some 940,000 industrial robots were at work and major industrial countries reported growth rates in robot installation of more than 20% compared to the previous year (see Figure 28.1) The automotive, electric, and electronic industries have been the largest robot users; the predominant applications are welding, assembly, material handling, and dispensing. The flexibility and versatility of industrial robot technology have been strongly driven by the needs of these industries, which account for more than 75% of the world's installation numbers. Still, the motor vehicle industry accounts for some 50% of the total robot investment worldwide.[9]

Robots are now mature products facing enormous competition by international manufacturers and falling unit costs. A complete six-axis robot with a load capacity of 10 kg was offered at less than $60,000 in 1999. It should be noted that the unit price only accounts for about 30% of the total system cost. However, for many standard applications in welding, assembly, palletizing, and packaging, preconfigured, highly flexible workcells are offered by robot manufacturers, thus providing cost effective automation to small and medium sized operations.

A broad spectrum of routine job functions led to a robotics renaissance and the appearance of service robots. Modern information and telecommunication technologies have had a tremendous impact on exploiting productivity and profitability potentials in administrative, communicative, and consultative services. Many transportation, handling, and machining tasks are now automated. Examples of diverse application fields for robots include cleaning, inspection, disaster control, waste sorting, and transportation of goods in offices or hospitals. It is widely accepted that service robots can contribute significantly to better working conditions, improved quality, profitability, and availability of services. Statistics on the use and distribution of service robots are scarce and incomplete. Based on sales figures from leading manufacturers, the total service robot stock can

© 2002 by CRC Press LLC

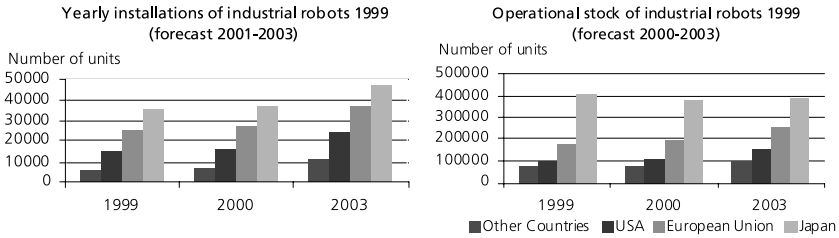


FIGURE 28.1 Yearly installations and operational stock of industrial robots worldwide.

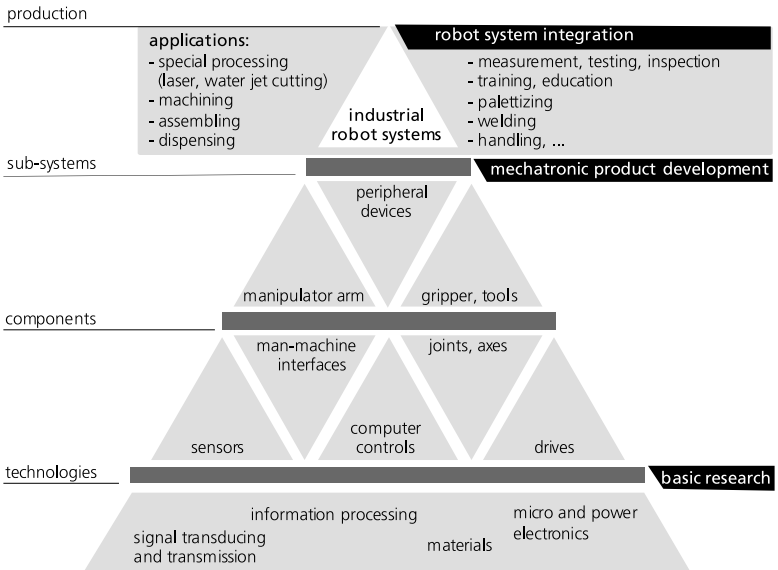


FIGURE 28.2 Robotics and mechatronics. (From Warnecke, H.-J. et al., in *Handbook of Industrial Robotics,* 1999, p. 42. Reprinted with permission of John Wiley & Sons.)

be estimated at a few thousand and certainly below 10,000 units. It is expected that within ten years, service robots may become commodities and surpass industrial robot applications.

Robots are representative of mechatronics devices which integrate aspects of manipulation, sensing, control, and communication. Rarely have so many technologies and scientific disciplines focused on the functionality and performance of a system as they have done in the fields of robot development and application. Robotics integrates the states of the art of many front-running technologies as depicted in Figure 28.2.

This chapter will give an overview of the state of the art and current trends in robot design and application. Industrial and service robots will be considered and typical examples of their system design will be presented in two case studies.

28.2 Industrial Robots

28.2.1 Definition and Applications of Industrial Robots

Large efforts have been made to define an industrial robot and to classify its application by industrial branches so that remarkably precise data and monitoring are available today.[9] According to ISO 8373, a manipulating industrial robot is defined as:

© 2002 by CRC Press LLC

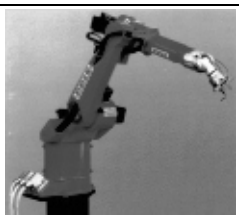
Specialization of robots			
universal robot	application specific	specialist (modular design)	specialist (customized design)
			
Examples: Reis RV6	ABB Flex Palettizer	CMB Modular Robot	IPA Robot Refuelling
• design fits standard applications • product variants according to payload, dexterity, working envelope • use of customized components • high manufacturing quantities	• application-oriented designs • integrated process-control functions • preconfigured workcells available • medium manufacturing quantities	• task specific design • integration of standard modules (axis, control, sensors) • preferred applications: material handling • small manufacturing quantities	• task specific designs • primary applications: nonmanufacturing fields (service robots) • task based kinematic structure • small to large manufacturing quantities

FIGURE 28.3 Examples of specialization of robot designs. (Courtesy of Reis Robotics, ABB Flexible Automation, and CMB Automation. From Warnecke, H.-J. et al., in *Handbook of Industrial Robotics,* 1999, p. 42. Reprinted with permission of John Wiley & Sons.)

An automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes (in three or more degrees of freedom, DOF), which may be either fixed in place or mobile for use in industrial automation applications.

The terms used in the definition above are:

- Reprogrammable: a device whose programmed motions or auxiliary functions may be changed without physical alterations.
- Multipurpose: capable of being adapted to a different application with physical alterations.
- Physical alterations: alterations of the mechanical structure or control system except for changing programming cassettes, ROMs, etc.
- Axis: direction used to specify motion in a linear or rotary mode.

A large variety of robot designs evolved from specific task requirements (see Figure 28.3). The specialization of robot designs had a direct impact on robot specifications and its general appearance. The number of multipurpose or universal robot designs was overwhelming. However, many applications are common enough that robot designs with specific process requirements emerged. Examples of the different designs and their specific requirements are shown in Figure 28.4.

28.2.2 Robot Kinematic Design

The task of an industrial robot in general is to move a body (workpiece or tool) with six maximal Cartesian spatial DOF (three translations, three rotations) to another point and orientation within a workspace. The complexity of the task determines the required kinematic configuration. The number of DOFs determines how many independently driven and controlled axes are needed to move a body in a defined way. In the kinematic description of a robot, we distinguish between:

- Arm: an interconnected set of links and powered joints that support or move a wrist, a hand or an end effector.
- Wrist: a set of joints between the arm and the hand that allows the hand to be oriented to the workpiece. The wrist is for orientation and small changes in position.

© 2002 by CRC Press LLC

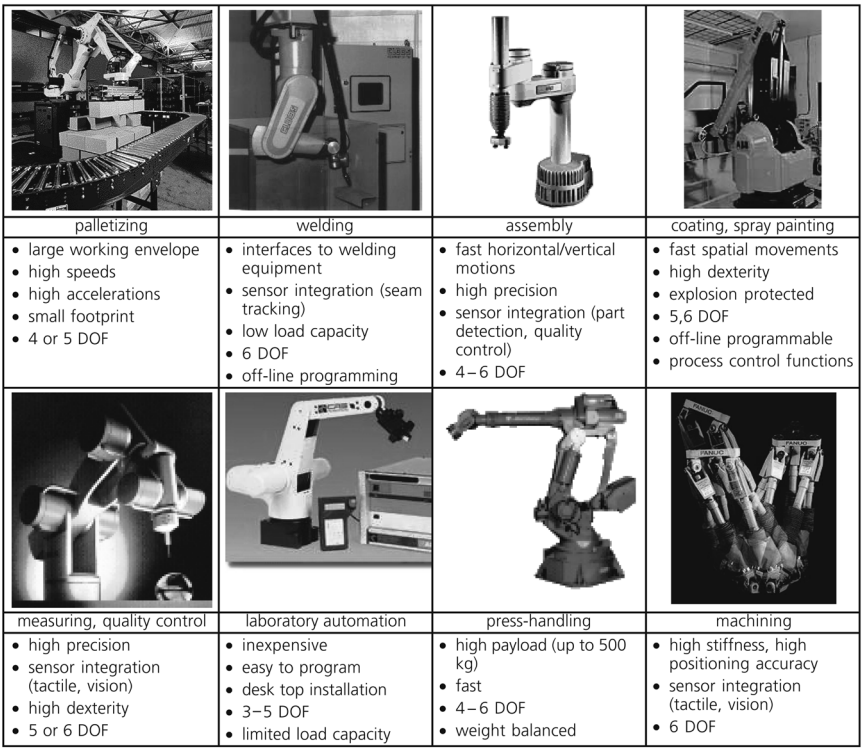


FIGURE 28.4 Application-specific designs of robots and their major functional requirements. (Courtesy of FANUC Robotics, CLOOS, Adept Technology, ABB Flexible Automation, Jenoptik, CRC Robotics, and Motoman Robotec. From Warnecke, H.-J. et al., in *Handbook of Industrial Robotics,* 1999, p. 42. Reprinted with permission of John Wiley & Sons.)

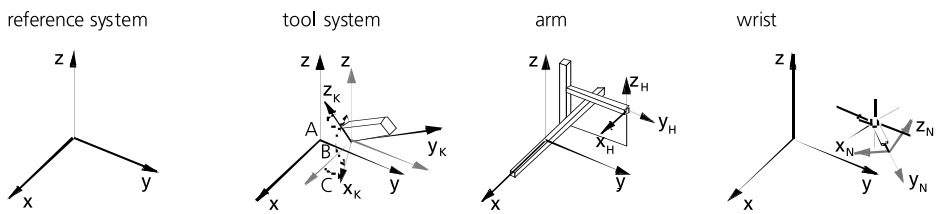


FIGURE 28.5 Definition of coordinate systems for the handling task and the robot.

Figure 28.5 illustrates the following definitions:

- The reference system defines the base of the robot and, also in most cases, the zero position of the axes and the wrist.
- The tools system describes the position of a work piece or tool with six DOFs (X_k, Y_k, Z_k, A, B, C).
- The robot (arm and wrist) is the link between the reference and tool systems.

Axes are distinguished as follows:

- Rotary axis: an assembly connecting two rigid members that enables one to rotate in relation to the other around a fixed axis.
- Translatory axis: an assembly between two rigid members enabling one to have linear motion in contact with the other.

© 2002 by CRC Press LLC

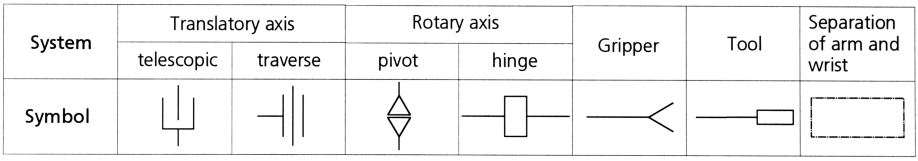


FIGURE 28.6 Symbols for the kinematic structure description of industrial robots according to VDI guideline 2681.

Figure 28.6 shows an overview of the symbols used in VDI guideline 2861 and in this chapter. Any kinematic chain can be combined by translatory and rotatory axes.

The manifold of possible variations of an industrial robot structure can be determined as follows:

$$V = 6^{DOF}$$

where V = number of variations and DOF = number of degrees of freedom. A large number of different chains can be built; for example, 46,656 different kinematic chains are possible for six axes. However, a large number is inappropriate for kinematic reasons:[1]

- Positioning accuracy generally decreases with the number of axes.
- Kinetostatic performance depends directly on the choice of kinematic configuration and its link and joint parameters.
- Power transmission becomes more difficult as the number of axes increases.

Industrial robots normally have up to four principal arm axes and three wrist axes. Figure 28.7 shows the most important kinematic chains. While many existing robot structures use serial kinematic chains (with the exception of closed chains for weight compensation and motion transmission), some parallel kinematic structures have been adopted for a variety of tasks. Most closed-loop kinematics are based on the so-called hexapod principle (Steward platform), which represents a mechanically simple and efficient design. The structure is stiff and allows excellent positioning accuracy and high speeds, but working volume is limited.

If the number of independent robot axes (arm and wrist) is greater than six, we speak of kinematically redundant arms. Because there are more joints than the minimum number required, internal motions may allow the manipulator to move while keeping the position of the end effector fixed.[14] The improved kinematic dexterity may be useful for tasks taking place under severe kinematic constraints. Redundant configuration such as a six-axis articulate robot installed on a linear axis (Figure 28.8) or even a mobile robot (automated guided vehicle, AGV) is quite common and used as a measure to increase the working volume of a robot.

28.2.2.1 Cartesian Robots

Cartesian robots have three prismatic joints whose axes are coincident with a Cartesian coordinate system. Most Cartesian robots come as gantries, which are distinguished by framed structures supporting linear axes. Gantry robots are widely used for handling tasks such as palletizing, warehousing, order picking, and special machining tasks such as water jet or laser cutting where robot motions cover large surfaces.

Most gantry robot designs follow a modular system. Their axes can be arranged and dimensioned according to the given tasks. Wrists can be attached to the gantry's z axis for end effector orientation (Figure 28.9). A large variety of linear axes can be combined. Numerous component manufacturers offer complete programs of different sized axes, drives, computer controls cable carriers, grippers, etc.

28.2.2.2 Cylindrical and Spherical Robots

Cylindrical and spherical robots have two rotary and one prismatic joint. A cylindrical robot's arm forms a cylindrical coordinate system, and a spherical robot arm forms a spherical coordinate

© 2002 by CRC Press LLC

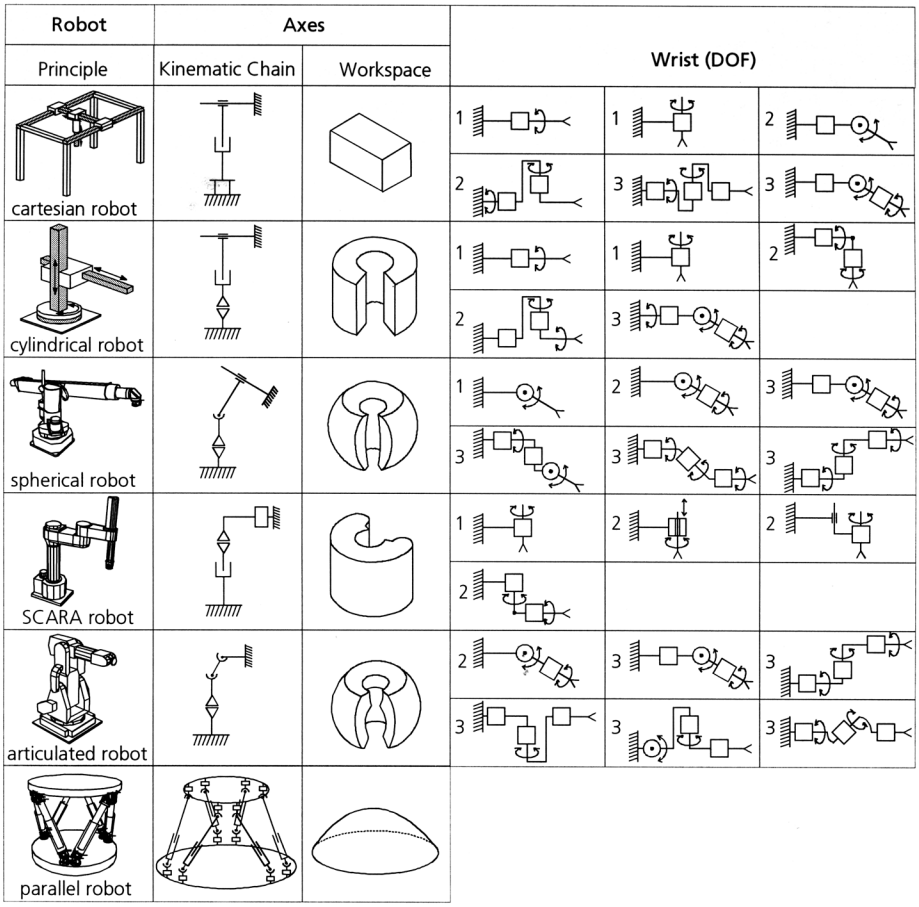


FIGURE 28.7 Typical arm and wrist configurations of industrial robots.

system. Today these robot types play only a minor role and are used for palletizing, loading, and unloading of machines. See Figure 28.10.

28.2.2.3 SCARA Type Robots

As a subclass of cylindrical robot, the SCARA (Selective Compliant Articulated Robot for Assembly) consists of two parallel rotary joints to provide selective compliance in a plane which is produced by its mechanical configuration. The SCARA was introduced in Japan in 1979 and has been adopted by numerous manufacturers. The SCARA is stiff in its vertical direction but, due to its parallel arranged axes, shows compliance in its horizontal working plane, thus facilitating insertion processes typical in assembly tasks. Furthermore, its lateral compliance can be adjusted by setting appropriate force feedback gains. SCARA's direct drive technology fulfills in all potentials: high positioning accuracy for precise assembly, fast and vibration-free motion for short cycle times, and advanced control for path precision and controlled compliance. Figure 28.11 shows the principle of a direct-drive SCARA.

28.2.2.4 Articulated Robots

The articulated robot arm, as the most common kinematic configuration, consists of at least three rotary joints by definition. High torque produced by the axes' own weight and relatively long reach can be counterbalanced by weights or springs. Figure 28.12 displays a typical robot design.

© 2002 by CRC Press LLC

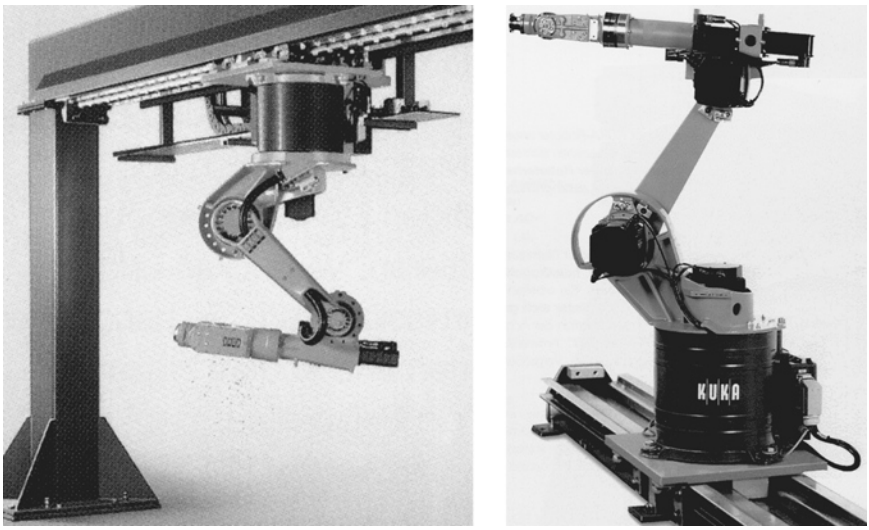


FIGURE 28.8 Floor and overhead installations of a six-DOF industrial robot on a translational axis, representing a kinematically redundant seven-DOF robot system. (Courtesy of KUKA.)

28.2.2.5 Modular Robots

For many applications, the range of tasks that can be performed by commercially available robots may be limited by their mechanical structures. Therefore, it may be advantageous to deploy a modular robotic system that can be reassembled for other applications. A vigorous modular concept that allows universal kinematic configurations has been proposed:

- Each module with common geometric interfaces houses power and control electronics, an AC servo-drive, and a harmonic drive reduction gear.
- Only one cable, which integrates the DC power supply and field bus signal fibers, connects the modules.
- The control software is configured for the specific kinematic configuration using a development tool.
- A simple power supply and a PC with appropriate field bus interfaces replace a switching cabinet.

Figure 28.13 illustrates the philosophy of this system and gives an example.

28.2.2.6 Parallel Robots

Parallel robots are distinguished by concurrent prismatic or rotary joints. Two kinematic designs have become popular:

- The tripod with three translatory axes connecting end effector, plate, and base plate, and a two-DOF wrist.
- The hexapod with six translatory axes for full spatial motion.

At the extremities of the link, we find a universal joint and a ball-and-pocket joint. Due to the interconnected links, the kinematic structure generally shows many advantages such as high stiffness, accuracy, load capacity, and damping.[11,21] However, kinematic dexterity is usually limited.

Parallel robots now work in many new applications where conventional serial chain robots reached shown their limits — machining, deburring, and part joining, where high process forces at high motion accuracy are overwhelming. Parallel robots can be simple in design and often rely on readily available, electrically or hydraulically powered, precision translatory axes.[12] Figure 28.14

© 2002 by CRC Press LLC

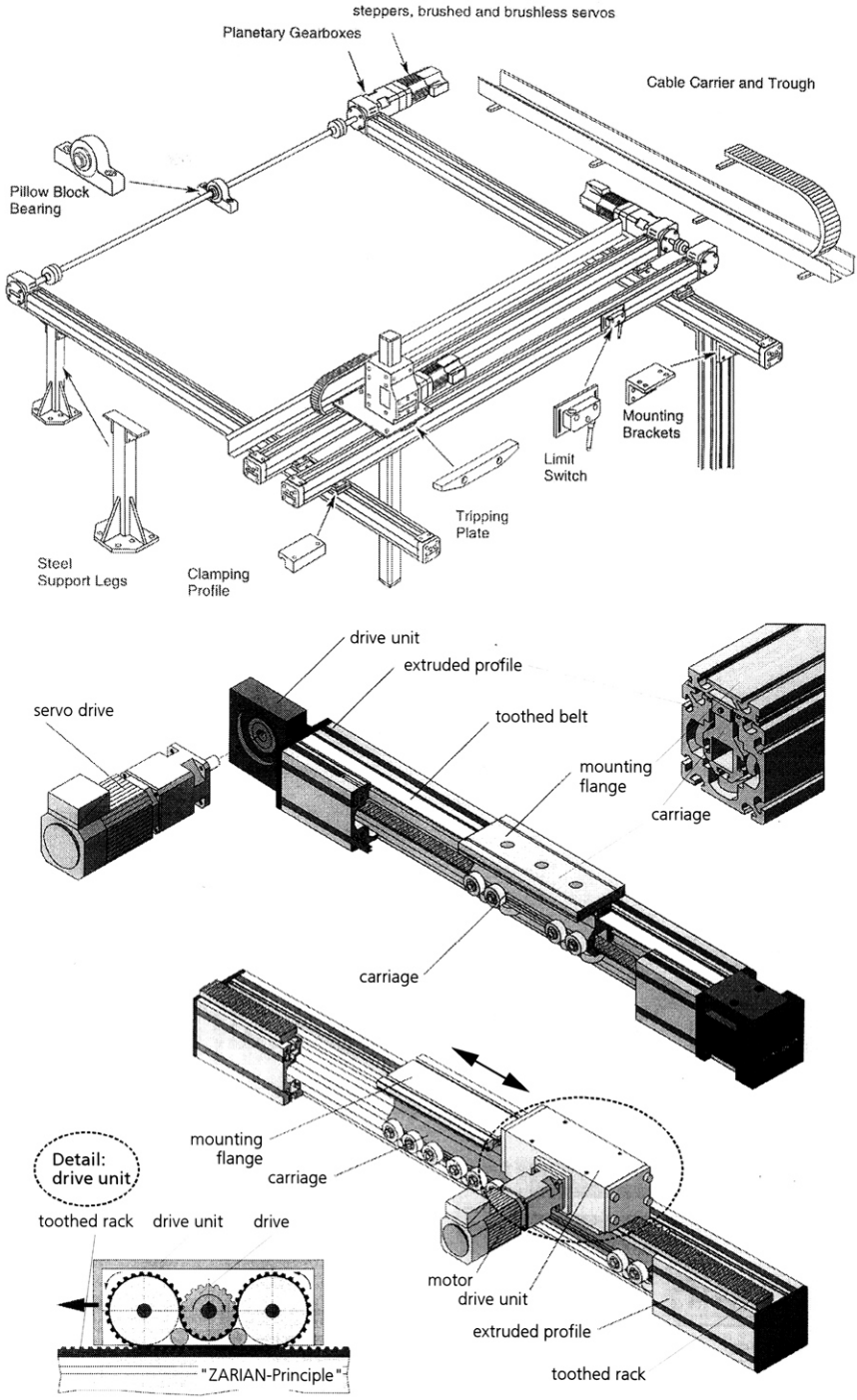


FIGURE 28.9 Modular gantry robot program with two principles of toothed belt-driven linear axes. (Courtesy of Parker Hannifin, Hauser division. From Warnecke, H.-J. et al., in *Handbook of Industrial Robotics,* 1999, p. 42. Reprinted with permission of John Wiley & Sons.)

© 2002 by CRC Press LLC

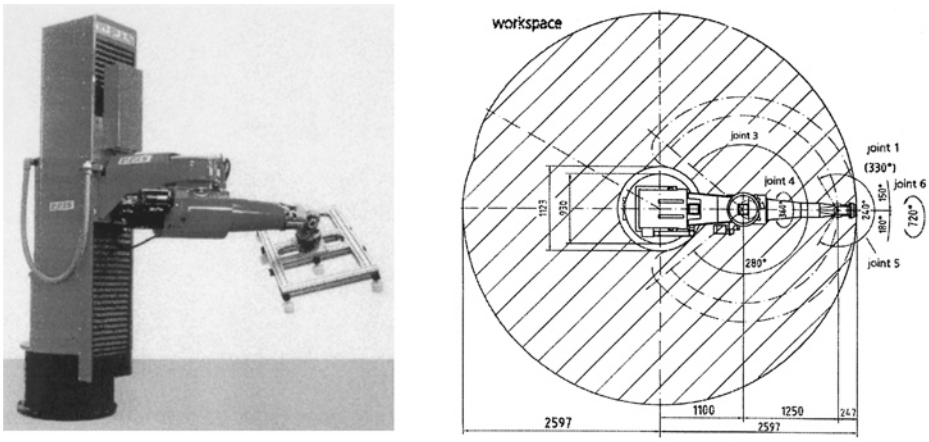


FIGURE 28.10 Five-DOF cylindrical robot with depiction of its workspace (top view, in millimeters). (Courtesy of Reis Robotics.)

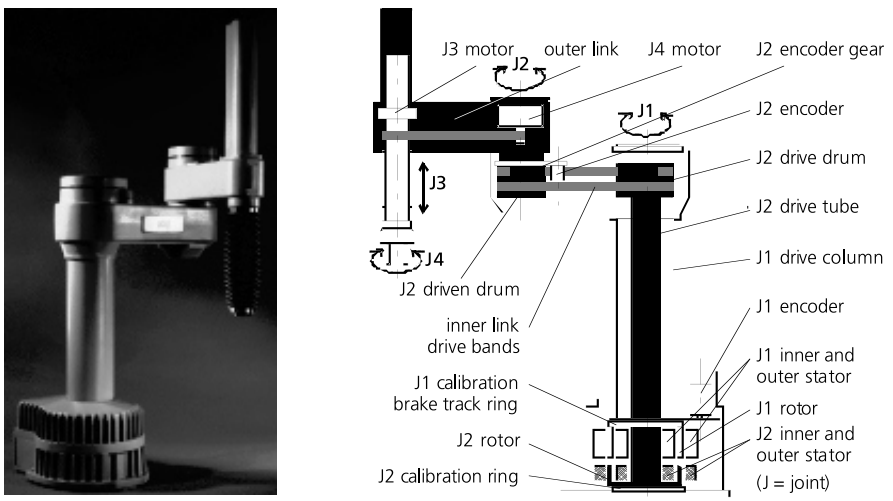


FIGURE 28.11 View of a SCARA robot (left) and cross-section through its direct drive arm transmission. (Courtesy of Adept.)

gives examples of tripod and hexapod platforms. Although parallel manipulators have been introduced recently and their designs are quite different from those of most classical manipulators, their advantage for many robotics tasks is obvious, and they will probably become indispensable.

28.2.3 Industrial Robot Application

28.2.3.1 Benefits of Robot Automation

The development of robot automation is characterized by a dramatic improvement in functional capabilities as well as rapidly falling price/performance ratios (technology push). There is also an increase in the demand for automation solutions, generated by the constant striving of industrial companies, in particular those subjected to international competition, to reduce costs and to improve

© 2002 by CRC Press LLC

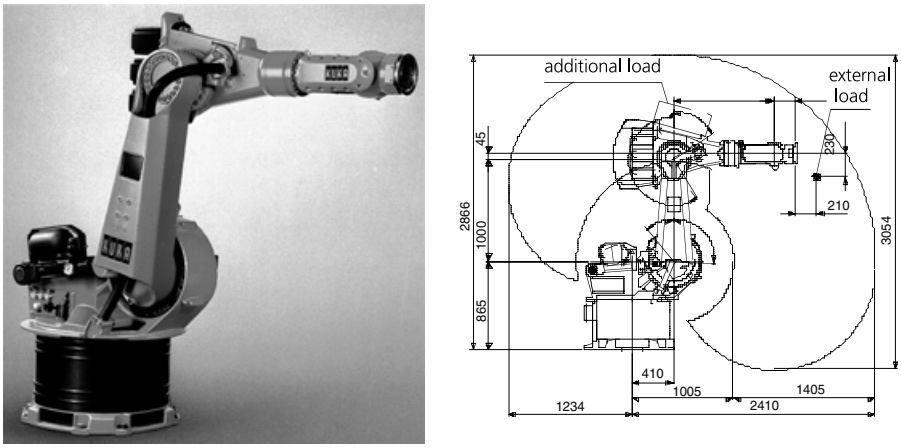


FIGURE 28.12 Articulated robot and its workspace. Note the gas spring that acts as a counterbalance to the weight produced by axis 2. (Courtesy of KUKA.)

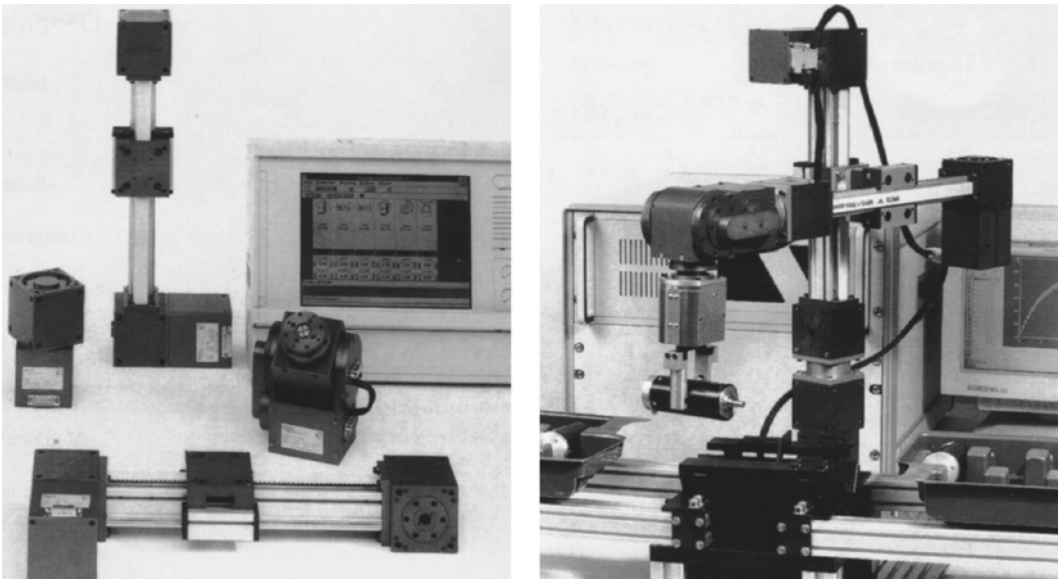


FIGURE 28.13 Modular robot system consisting of rotary and translatory axis modules, grippers, and configurable control software. (Courtesy of Amtec.)

product quality (market pull). Falling unit costs and improved robot system performance led to new automation solutions, many of them outside classical industrial robot applications, such as:

- Food industry (material flow automation with functions such as packaging, palletizing, order picking, sorting, warehousing, processing, etc.)
- Mail order and postal services (material flow automation)
- Airports, train stations, freight terminals, etc. (material flow automation)
- Consumer goods (processing, material flow automation)
- Chemical, pharmaceutical, and biotechnical industries (processing, material flow automation)

© 2002 by CRC Press LLC

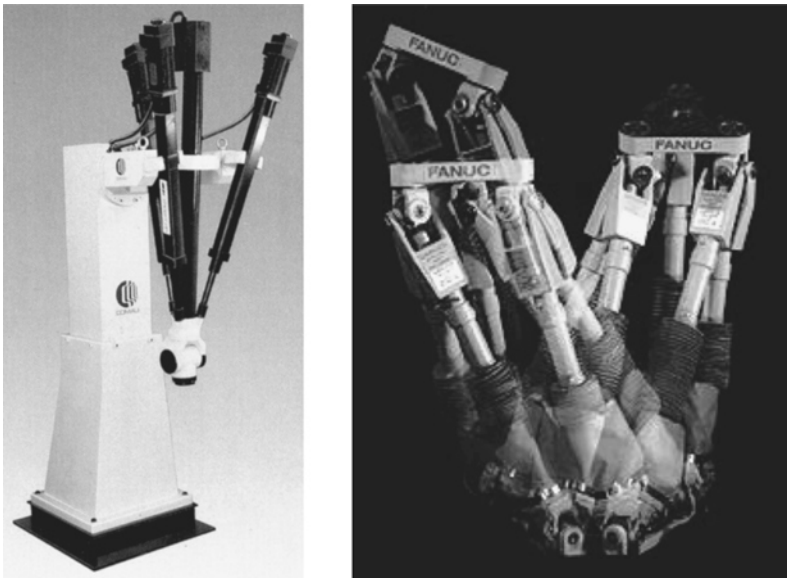


FIGURE 28.14 The COMAU Tricept, a six-DOF tripod and the FANUC FlexTool Steward platform with six servo-spindle modules connecting the bottom and moving plate. (Courtesy of COMAU and FANUC Robotics.)

Robot manufacturers and integrators now supply low-cost flexible workcells with standard configurations, which can be rapidly integrated into existing production systems for standard applications. Even small volume operations can be effectively automated for functions such as parts welding and cutting, flexible assembly, packaging, and palletizing.

A recent survey among German manufacturers reviewed the benefits realized from investing in robot automation (see Figure 28.15). Besides cost effectiveness, there are many other reasons a company considers in selecting a robot system, e.g., effect on parts quality, manufacturing productivity (faster cycle time), yield (less scrap), reduction in labor, improved worker safety, and reduction of work in progress.

28.2.3.2 Robot Workcell Planning and Design

Once the desired benefits and requirements are identified, specification, commissioning, and the process of putting the robot system into operation must be approached in a systematic manner. Installing a robot workcell is best done in a multistep process that involves consideration of robot, the products to be handled by the cell, other production equipment in the cell, layout, scheduling, material flow, safety, maintenance, and training. See Figure 28.16.

Numerous planning tools support the planning and design of the robot workcell. These so-called computer-aided production engineering (CAPE) tools assist in effectively designing, evaluating, and controlling production facilities. They help meet performance requirements and cost and time constraints. Suppliers can be selected on the basis of price and on their ability to offer integrated services during workcell planning, implementation, and operation. In fact, clients and robot system integrators often establish close partnerships that last over the life of the system. The case studies reviewed below clearly show the importance of such partnerships for the success of installation and operation of robot cells.

28.2.3.3 Case Study: Automated High-Frequency Sealing in Measuring Instruments

28.2.3.3.1 Introduction

The company Rohde & Schwarz is an established leader in the field of electronic systems and measuring instruments. It attained this position by successfully offering high quality standard products and custom-designed systems. Its production is characterized by small lots, short delivery

© 2002 by CRC Press LLC

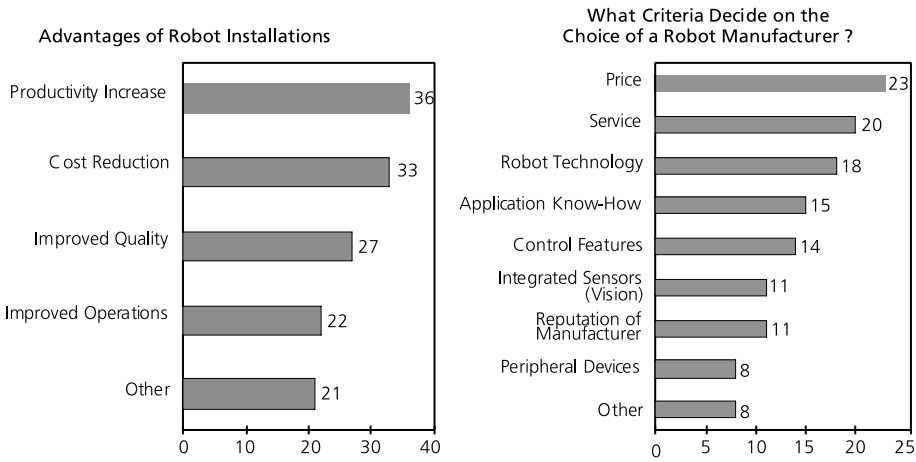


FIGURE 28.15 Survey of benefits from robot automation and criteria for selecting suppliers.

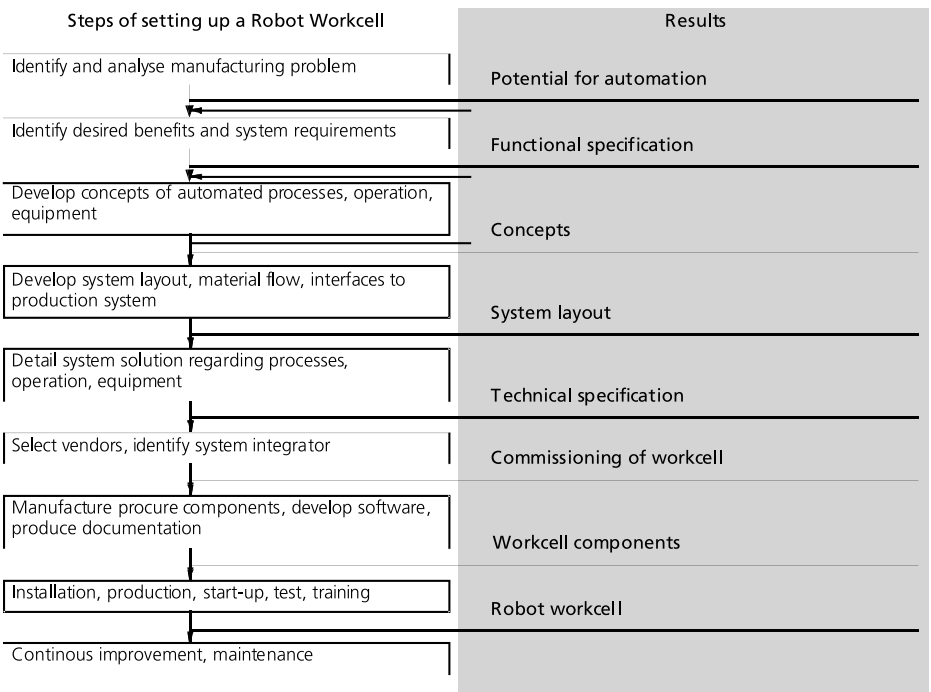


FIGURE 28.16 Typical steps for launching a robot workcell.

times, and short development lead times. By investing in an automated assembly system for measuring instrument cases, the company estimated that it could manufacture the products at lower costs. The cases are composed of several frame parts. Each frame part is separated by a metal cord for screening against high frequency (hf) radiation (see Figure 28.17).

The cases have various dimensions (six heights, three widths, and three depths). The company produces about 1000 product variants with customer-specific fastening positions for the insertion of the measuring instruments. The assembly line is split to allow order-independent preassembly and order-specific final-assembly (see Figure 28.18).

© 2002 by CRC Press LLC

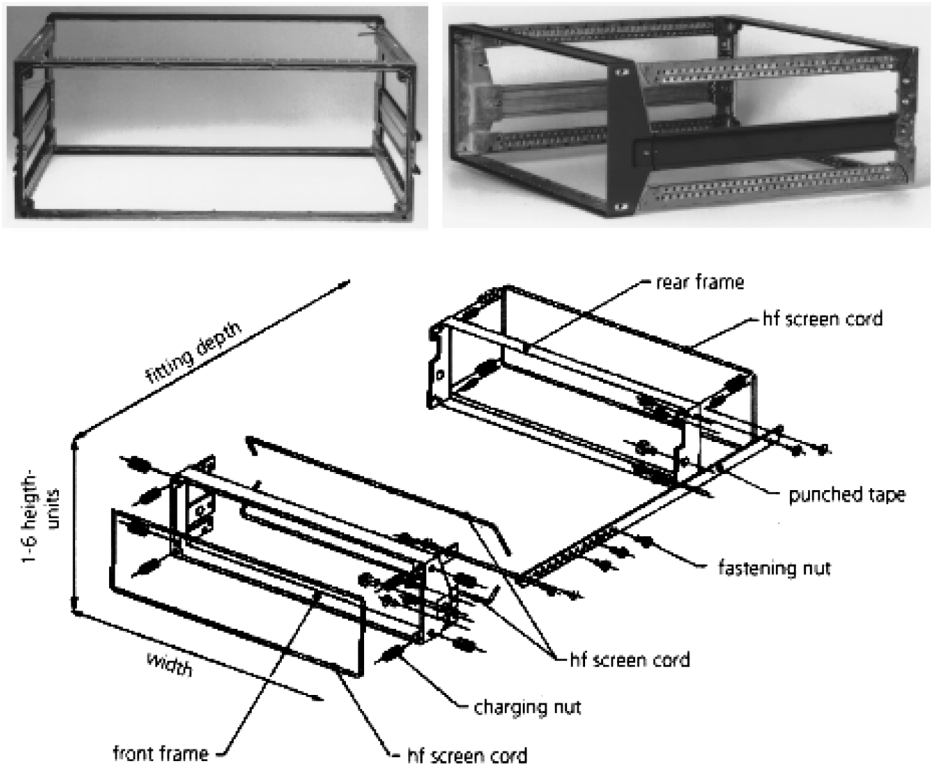


FIGURE 28.17 Typical frame design (top left and right) and frame-components (bottom) of cases for measuring instruments.

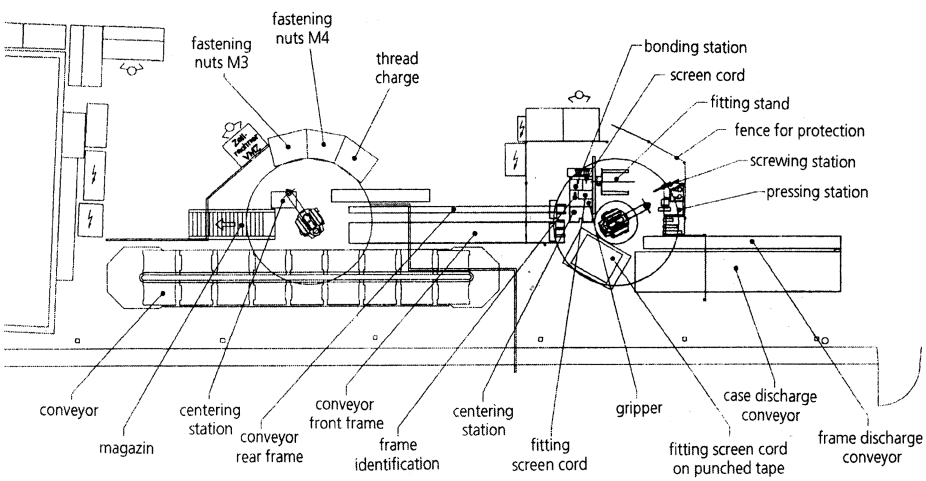


FIGURE 28.18 Layout of the preassembly (left) and final assembly cell (right).

28.2.3.3.2 Pre-Assembly of Cases

Automatic stations in the preassembly cell press in fastening nuts and screw in several threaded bolts. An industrial robot handles the frame parts. After removing the frames from the supply pallets, the robot brings the frame into a mechanical centering device for fine positioning and for eliminating tolerances in the pallets. It is possible to achieve exact positioning of the frame in the robot gripper.

© 2002 by CRC Press LLC

During the pressing operation, the robot positions and fixes the frame in the press station. The pressing operation requires a press force of about 3000 N. A compliance system is integrated into the gripper to eliminate tolerances in the frame dimensions. The robot also positions and fixes the frames in the screwing station. A system controls the rotation angle, screwing torque, and screwing depth to consistently reproduce a screwing depth of 0.1 mm.

After terminating the preassembly, the robot places the frames on a conveyer system. The conveyer links the preassembly and final assembly cells. The conveyer belts are separated into belt pairs for front and rear frames.

28.2.3.3.3 Final-Assembly of the Cases

The final assembly consists of:

- Fitting the metal cord for high frequency screening into the frames
- Order-specific pressing of fastening elements
- Screwing together all frames that form the finished case
- Lettering the finished case

One of the most interesting technical potentials for automation was the assembly of the metal cord for high frequency screening into each frame. The metal cords are nonrigid parts. At the beginning of the project, the company had little experience in automated assembly of cords. The cords have no rigidity; they can only transmit tensile forces. The results of other forces and torques and undefined deformations were unforeseen. The influence of temperature variations had to be considered. An additional problem is reproducibility of cord diameter. Two metal cords with different diameters (2.0 mm and 3.0 mm) have to be fit into four different running slots. The 250-m cords are supplied on coils. In the slot of the rear frame, it is necessary to insert adhesive points to give the cord the required stability.

Several basic principles for fitting the metal cord into the slots were investigated. Fitting with an oscillating plunger was the best method for assembling the cord into the slots.

Four geometrically different plungers were necessary for the different running slots to achieve a minimum of plunger changing time, all plungers were integrated in the robot tool. See Figure 28.19. Depending on the slot type, the right plunger is positioned and coupled with the oscillating motor. A cord cutting system is integrated into the robot tool to obtain the right length of the cord (depending on the dimensions of the frames). It also includes an adhesive-dispensing system to set the adhesive spots into the slots.

After fitting the metal cord into the front inside, front outside, rear and side ledge frames the fasteners for the inserts are pressed in order-specific positions into the side ledges. For this operation the robot takes a ledge with the required length from a magazine and brings it to a press station. A guide rail defines the exact position. The fasteners are blown automatically from the feeder through a feed pipe to the press position. Force and position of the press plunger are monitored during the press operation. Figure 28.20 shows how the sealing tool fits the cord into the rear frame of the case and the subsequent screwing of all frame components, which is also done by the robot.

The robot first moves to the screwing position. The screws are blown automatically through the feed pipe on the feeder to the screwing tool. To achieve a perfect result, rotation angle and screwing torque are monitored.

Depending on the construction of the case, it is important to have accessibility from four directions throughout the assembly process, and it was necessary to install a clamping device that can turn the case in all required positions. The result was a system consisting of standard components that can clamp more than 25 cases with different dimensions. Figure 28.21 shows the preassembly cell (left robot in the layout in Figure 28.18) and the final assembly cell with the flexible clamping system (right robot in Figure 28.18). For quick tool changes, the robot arm has an automatic tool changing system. Each tool can be picked up within a few seconds.

© 2002 by CRC Press LLC

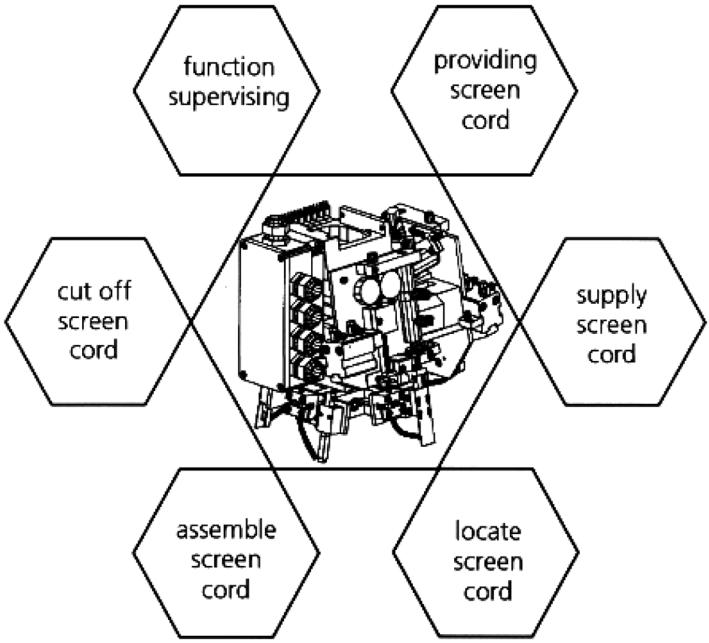


FIGURE 28.19 Tool system tasks in automated high-frequency sealing.

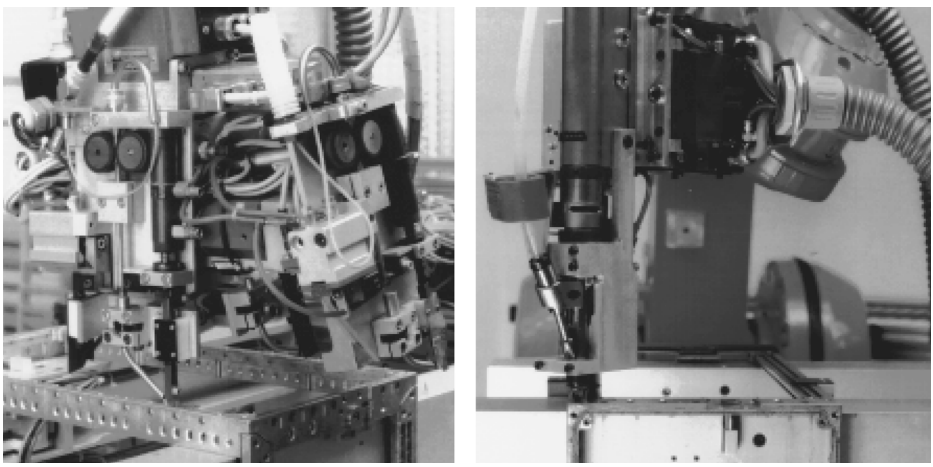


FIGURE 28.20 Robot tool for fitting of metal cord in front frame (left) and for screwing of frame parts.

28.2.3.3.4 Conclusion

Tasks with extensive numbers of assembly steps and production volumes of less than 10,000 units per year can be automated in a cost-effective way. Such automation projects are of special interest for small and medium sized companies in the electronic industry. Almost all components developed for this system can be used in other assembly systems with only small modifications. The main objective of the robot investment was to combine high product quality with improved cost effectiveness. A pay-back period of 3 to a maximum 4 years on the basis of 10,000 produced units per year, was set as the break-even for an investment of some 700,000 DM in equipment cost and 200,000 DM in engineering costs. Since the workcell was installed in 1993, production volume has, increased to more than 16,000 units per year so that a pay-back period of well below 3 years

© 2002 by CRC Press LLC

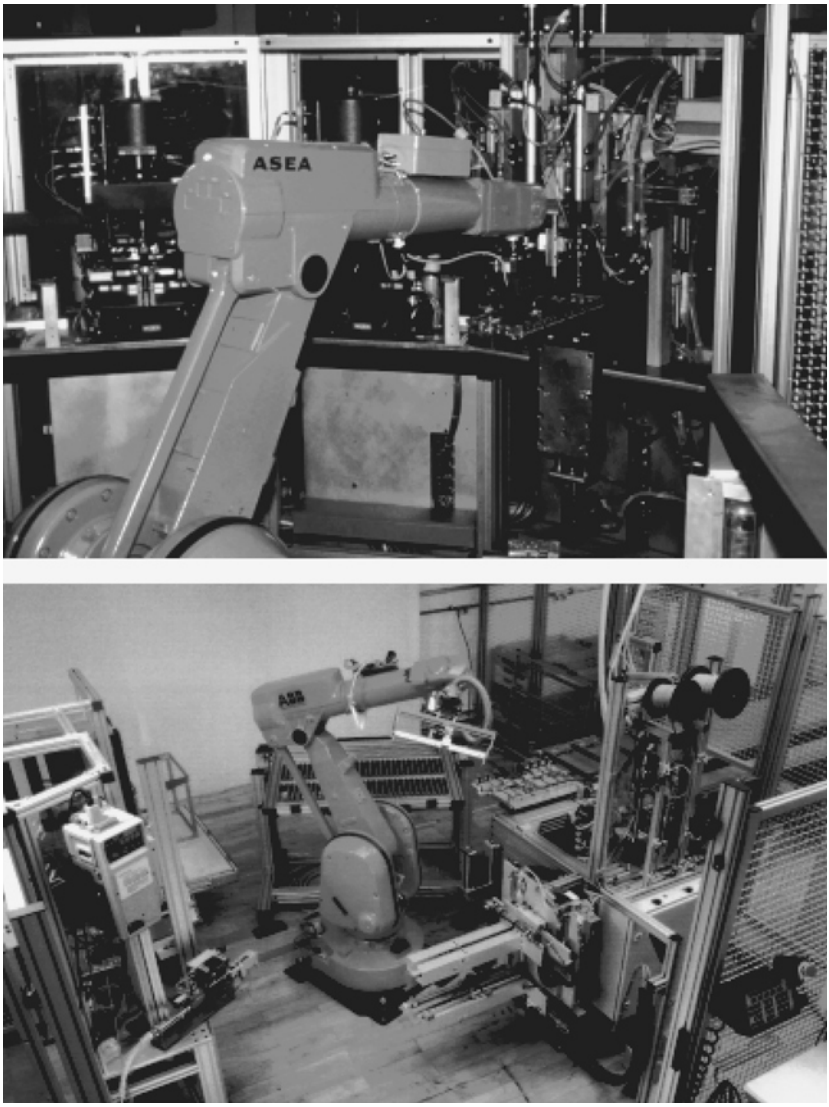


FIGURE 28.21 Preassembly cell (top) and final assembly cell (bottom) with flexible clamping system.

was achieved. The decision to invest in a robot system for the complicated process of sealing and assembling variable frames turned out to be more profitable than originally estimated.

28.3 Service Robots

28.3.1 From Industrial Robots to Service Robots

Early industrial robots were found in many nonmanufacturing applications:

- Inspection tasks in hazardous environments
- Laboratory automation
- Automated pharmacy warehousing
- Storage and retrieval of data cartridges in computing centers

© 2002 by CRC Press LLC

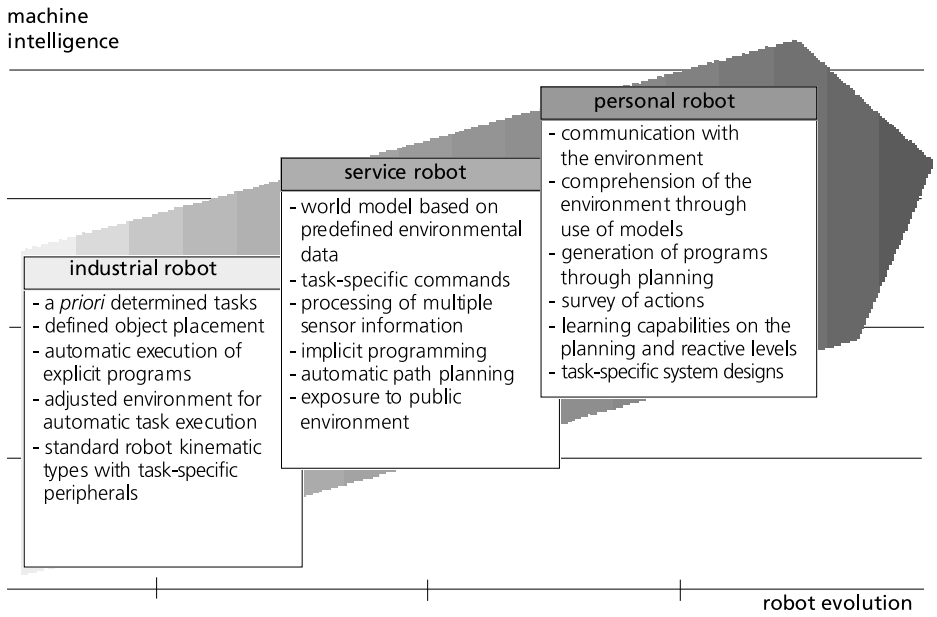


FIGURE 28.22 From industrial robots to service robots — the evolution of machine intelligence.

Robot application in nonmanufacturing fields has been on the rise as key technologies have become more available. Sensors in combination with advanced perception algorithms allow robots to function in partly or even completely unstructured environments. Fast interactions between sensing and action account for effective and robust task execution, even in dynamically changing situations.

A definition recently suggested by IFR (the International Federation of Robotics) offers a description of the main characteristics of service robots, their exposure to public, and task execution in unstructured environments.[15] Service robots are considered extensions of industrial robots.[19]

> Service robots are robots which operate semi or fully autonomously to perform services useful to the well being (hence, non-manufacturing) of humans and equipment. They are mobile or manipulative or combinations of both.

IFR has adopted a preliminary system for classifying service robots by application areas:

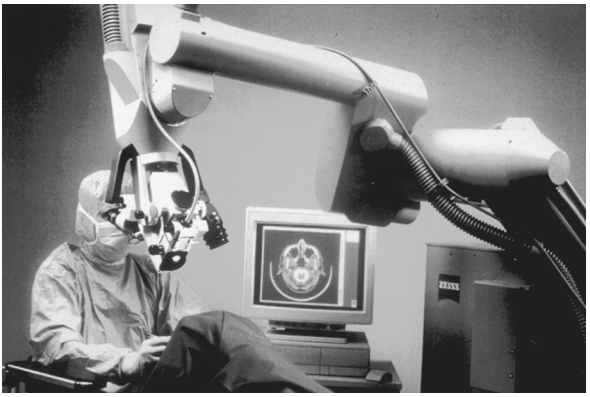
- Servicing humans (personal, safeguarding, entertainment, etc.)
- Servicing equipment (maintenance, repair, cleaning, etc.)
- Performing autonomous functions (surveillance, transport, data acquisition, etc.) including service robots that cannot be classified in the previous categories.

Some scientists and engineers even predict a future for "personal robots,"[5,7] and visions depict these robots as companions for household tasks, gardening, leisure, and even entertainment. The evolution of robots can be characterized by the level of machine intelligence implemented for task execution. See Figure 28.22.[17]

28.3.2 Examples of Service Robot Systems

Service robots are designed for the execution of specific tasks in specific environments. Unlike an industrial robot, a service robot system must be completely designed. New concepts stress the possibility of using preconfigured modules for mechanical components (joints) and information processing (sensors, controls). The following is a survey of different service robot systems, based on the IFR classification scheme.

© 2002 by CRC Press LLC



MKM

Servicing humans — The medical manipulator (MKM) produced by Carl Zeiss, Germany, consists of a weight-balanced servo-controlled six-DOF arm, a computer control, and a graphical workstation for visualization and programming. It carries a surgical microscope. Movements follow preprogrammed paths or are generated manually by a six-DOF input device (space-mouse) or voice.

MANUS

The MANUS arm of Exact Dynamics, The Netherlands, is a wheel-chair mountable six-DOF lightweight manipulator meant for persons with severe disabilities. The combination of wheelchair and manipulator helps in executing simple tasks such as opening doors, preparing coffee, etc. The arm folds discreetly while not in use. The man–machine interface for motion command can be individually adjusted to the person's abilities and can be a mouth whistle, voice, joystick, or any other adequate device.

© 2002 by CRC Press LLC

CASPAR

CASPAR (Computer Assisted Surgical Planning and Robotics) of ortoMAQUET, Germany, consists of an industrial robot mounted on a mobile base, a milling tool, and a calibration unit. The system assists the surgeon in orthopedic interventions such as hip surgery. On the basis of patient data, the placement of a hip prosthesis is simulated. All contours for a perfect fit are milled with remarkable precision under surgical supervision.

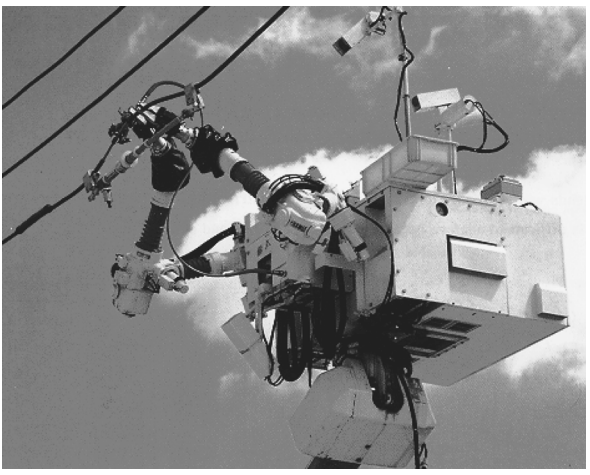
Electrolux

Electrolux, Sweden, introduced the first lawn mower powered by solar cells. Some 43 solar cells transform sunlight into electrical energy. The solar mower is fully automatic and eliminates emissions into air and makes almost no noise.

© 2002 by CRC Press LLC

Skywash

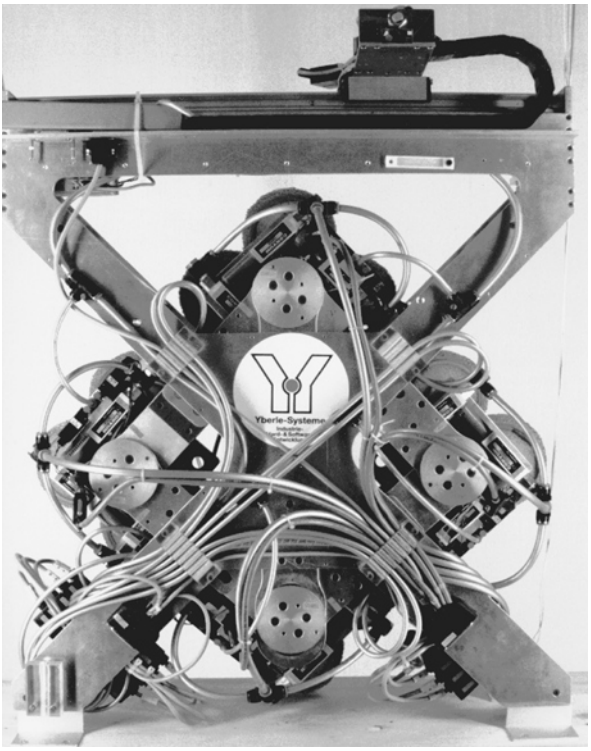
Servicing equipment — With two Skywash systems (Putzmeister Werke, Germany) in parallel operation, a reduction of ground times per washing event for factor 3 (wide body) aircraft and factor 2 (narrow body) can be achieved. Skywash integrates all features of an advanced robot system: pregeneration of motion programs by CAD aircraft models, object location by 3D-sensors, tactile sensor-controlled motion, redundant arm kinematics (11 DOFs) installed on a mobile base, and full safety features for maximum reliability. From a rough placement relative to the aircraft, Skywash operates under human supervision.



Master–slave two-armed robot

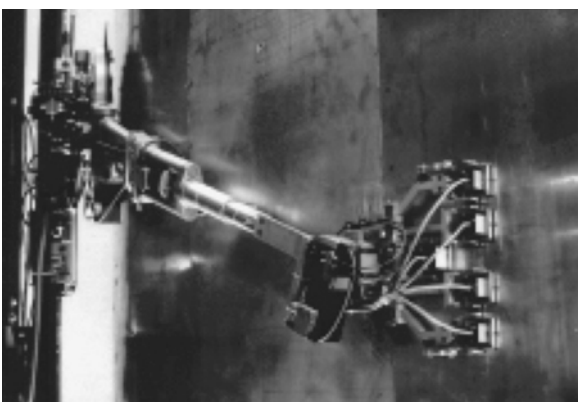
A master–slave two-armed robot (Yaskawa, Japan) carries out operations with live wires (cutting, repair, etc.) of up to 6600 V capacity. A truck-mounted boom carries the manipulator arms which are operated from a cabin.

© 2002 by CRC Press LLC



Rosy

Rosy produced by Robot System of Yberle, Germany, climbs surfaces on suction cups to perform cleaning, inspection, painting, and assembly tasks. Tools can be mounted on the upper transversal axis. Navigation facilities allow accurate and controlled movements.



Robot for nuclear reactor outer core inspection

A robot for nuclear reactor outer core inspection (Siemens KWU, Germany) follows a modular approach. Each joint module with common geometric interfaces houses power and control electronics, an AC servo drive and a reduction gear. The robot travels along existing rails and maps the core surface by its end effector-mounted ultrasound sensors. Material flaws can be detected and monitored during reactor operations.

© 2002 by CRC Press LLC

Cleaning robot

Performing autonomous functions — Cleaning robots have entered the market. Larger surfaces (central stations, airports, malls, etc.) can be cleaned automatically by robots with full autonomous navigation capability. The HACOmatic of Hako-Werke, Germany, is an example.

CyberGuard of Cybermotion Inc., United States, is a powerful tool that provides security, fire detection, environmental monitoring, and building management technology. The autonomous mobile robotic system features a rugged self-guided vehicle, autocharger docking station, array of survey instrumentation, and dispatcher software that provides system control over a secure digital spread-spectrum link.

The HelpMate of Pyxis, United States, is a mobile robot for courier services in hospitals, introduced in 1993. It transports meals, pharmaceuticals, and documents along normal corridors. Clear and simple user interfaces, robust robot navigation, and ability to open doors and operate elevators by remote control make it a pioneering system in terms of technology and user benefit. More than 100 installations are currently operating in hospitals with excellent acceptance by personnel.

The Care-O-Bot (Fraunhofer IPA, Germany) helps achieve greater independence for elderly or mobility-impaired persons and helps them remain at home. It offers multimedia communication, operation of home electronics, active guiding or support, and will fetch and carry objects such as meals or books.

28.3.3 Case Study: A Robot System for Automatic Refueling

Design and setup of service robot workcells require a vigorous systems approach when a robot is designed for a given task. Unlike industrial robot applications, a system environment or a task sequence generally allows little modification so that the robot system must be designed in depth. A good example of a service robot system design for automation of a simple task is the following.

28.3.3.1 Introduction

The use of a refueling robot should be convenient and simple, like entering a car park. Upon pulling up to the refueling station the customer inserts a credit card and enters a PIN code and refueling order. A touch on the start button of a touch screen activates the refueling. The robot opens the tank flap and docks on the tank cap. The robot then places the required grade and amount of fuel

© 2002 by CRC Press LLC

CyberGuard

HelpMate

Care-O-Bot

© 2002 by CRC Press LLC

in the open tank — automatically, emissions-free, and without losing a drop. The task was to develop a refueling robot geared to maximum customer convenience and benefit.

A consortium consisting of the ARAL mineral oil company and Mercedes-Benz and BMW set out to turn this vision into reality. Besides increasing comfort and safety, the system has significance in the future because of:

- Higher throughputs by shorter refueling cycles
- Reduced surface requirements of refueling stations
- No emissions or spillage
- Controlled and safe refueling

Customer benefits include

- Fully automatic vehicle refueling within 2 min
- Possibility of robotic refueling over 80% of all vehicles that have their filler caps on the rear right-hand sides
- Minimum conversion work on automobiles
- Up to five fuel grades available without producing emissions or odors
- Layout of refueling station that satisfies the appropriate ergonomic requirements
- Controlled, reliable system behavior in the event of unexpected human or vehicle movement or other disruptive factors
- Safe operating systems in areas at risk of explosion
- Economically viable equipment

Robot refueling is a typical use of an articulated service robot with characteristic properties:

- It can carry out its task safely without explicit knowledge of all possible situations and environmental conditions
- It can function when information on the geometric properties of the environment is imprecise or only partly known
- It creates confidence that encourages its use

28.3.3.2 Systems Design

Planning and design of service robot systems involves systematic design of mechatronic products (Schraft and Hägele,[18] Kim and Koshla,[94] and Schraft et al.[20]) followed by designing methods that will meet cost, quality, and life cycle objectives. The geometric layout and the overall configuration of the information processing architecture of the service robot are critical tasks. System design becomes more complex as requirements regarding dexterity, constraints, autonomy, and adaptivity increase. See Figure 28.23.

The technical specification of a service robot system can be divided into two successive phases: functional specification and system layout and architecture specification. This approach will be examined and applied to the development of the fuel refueling robot.

28.3.3.2.1 Functional Specification

Functionality is defined as the applicability of an object for the fulfillment of a particular purpose.[3] Various properties characterize an object and contribute to its definition of functionality. The works of Cutkosky[4] and Iberall[8] address the importance of understanding functionality when robots manipulate and interact with objects in a complex and dynamic environment. The functional specification phase develops:

- A list of the system's functional and economical requirements over its life cycle from manufacturing and operation to dismantling and recovery
- A formal description of the underlying processes in nominal and off-nominal modes

© 2002 by CRC Press LLC

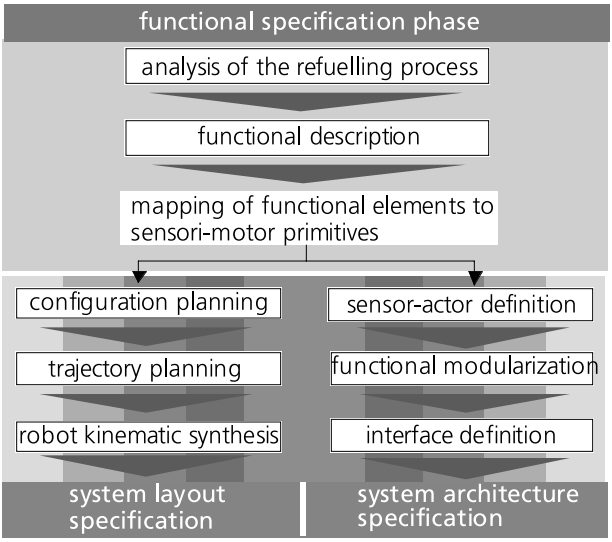


FIGURE 28.23 Technical specification of service robot systems. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

The analysis of service tasks is carried out similarly by process structuring and restructuring to define the necessary sequence and possible parallelism of all task elements. The focus lies in the analysis and observation of object motions and their immediate interactions as sensorimotor primitives.[2,13] Tasks are divided into:

- Elementary motions without sensor guidance and control (absolute motion control)
- Sensorimotor primitives defined as encapsulations of perception and motion that form domain general blocks for fast task strategies (reactive motion control)

The formalism for describing, controlling, and observing object motion in a dynamic environment concentrates on defining all relevant geometric, kinematic, and dynamic properties:

- Geometrical properties that identify quantifiable parameters (goal frames, dimensions, volumes, etc.)
- Kinematic properties that identify the mobilities of objects in trajectories
- Dynamic properties that describe how the object responds to forces or geometrical constraints

28.3.3.2.2 System Layout and Architecture Specification

The system layout specification comprises: the list of all devices required for task execution, trajectories and goal frames of analyzed objects, and robot kinematic parameters. After defining all devices, their geometry, spatial arrangement, and geometric constraints inside the workcell must be determined. The next step is trajectory planning of the automated task execution. It defines all geometric and kinetic entities such as goal frames, trajectories, permissible workspaces, and minimal distances to possible collision partners. Kinematic synthesis is the most complex step. It requires the optimal solution of a highly nonlinear and constrained problem. The task-based design requires the determination of:

- The number of degrees of freedom (DOFs)
- The kinematic structure
- The joint and link parameters
- Placement inside the robot workcell
- The location of the tool center point (TCP) relative to its last axes[16]

© 2002 by CRC Press LLC

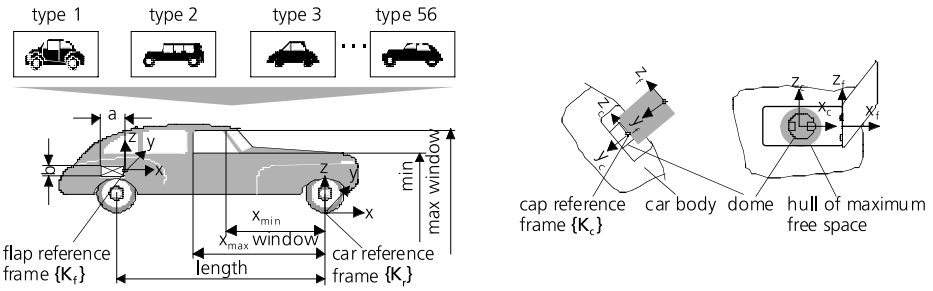


FIGURE 28.24 Registered car dimensions for automated refuelling. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

The quality of the manipulator design is expressed by objective functions such as dexterity, reachability, singularity avoidance, and kinematic simplicity.

The system architecture specification comprises the definitions of:

- All sensors and actuators with their logical interactions
- Logical interfaces between all data processing elements and their integration in a system architecture
- Man–machine interactions and their task level interfaces

Perceptive capabilities of the system result in the mapping of the task sequence into motion elements and sensorimotor primitives. The selection of the sensor depends on:

- The modality of information (force, distance, etc.)
- Dimensionality of the sensation
- Covering of the events defining possible transitions in the task execution
- Confidence in the observation that results from the observability of the event and the relevance of the sensor information.

28.3.3.3 Refueling Robot System Layout

The functional specification of the automated refueling describes the geometry, object motion, and its observability by perceptive elements in a straightforward manner:

Geometry — All robot movements must be limited to the car's rear section. The doors must not be obstructed or opened any time. The only reference for the coarse positioning of the car is the terminal. For 56 car types representing over 90% of Germany's car population, all relevant data regarding dimensions and flap and cap locations were registered (Figure 28.24).

Motion — The task sequence incorporates simple motion elements (e.g., move linearly, move circularly) and sensorimotor primitives like docking which requires a controlled approach toward dynamic goal frames (Figure 28.25).

Dynamic — Vertical vehicle movements may reach a frequency of over 1 Hz at a maximum velocity of 1 m/s. Sudden acceleration must result in safe emergency undocking.

The configuration of the system is shown in Figure 28.25. The concept of the refilling station suggests a simple layout and clear spatial perspective that should belie any complicated technology. The driver should simply have to drive up to the terminal, without having to stop the vehicle at a precise point. The robot is initially positioned out of sight. Only a refilling island 150-mm high is visible above the ground. All doors may swing open and people may exit the car any time. The terminal serves as a user-friendly customer interface and as a reference for the driver to conveniently position the car. The terminal can be reached, moved, and its height adjusted from the driver's window.

© 2002 by CRC Press LLC

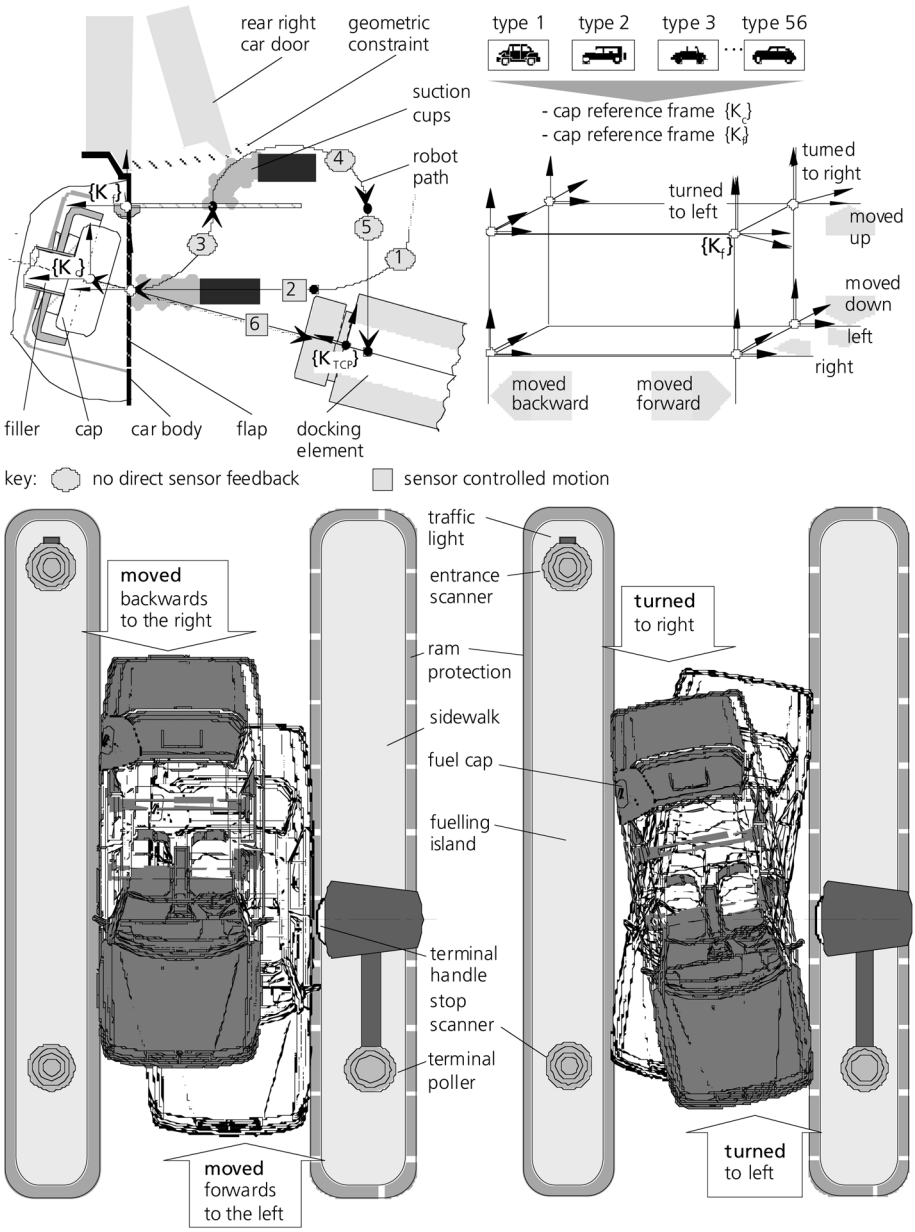


FIGURE 28.25 Layout of the automated fuel refilling system considering assumed extremal car locations in the filling station. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

Trajectory planning deals with the robot's movements covering one refilling cycle in nominal and off-nominal mode. Coming to a halt at the approach location the end effector (1) docks on to the flap, (2) turns the flap, (3) proceeds to the cap approach location, (4) docks on the cap, (5) turns the cap open, and (6) undocks and departs. All locations refer to the car's reference frame K_r. The range of car locations inside the refilling station is limited by the need to reach from the driver's window to the central axes of the terminal. Kinematic synthesis builds upon a trr:rrr structure. By numerical optimization, the best fitting arm kinematic must be found with respect to:

© 2002 by CRC Press LLC

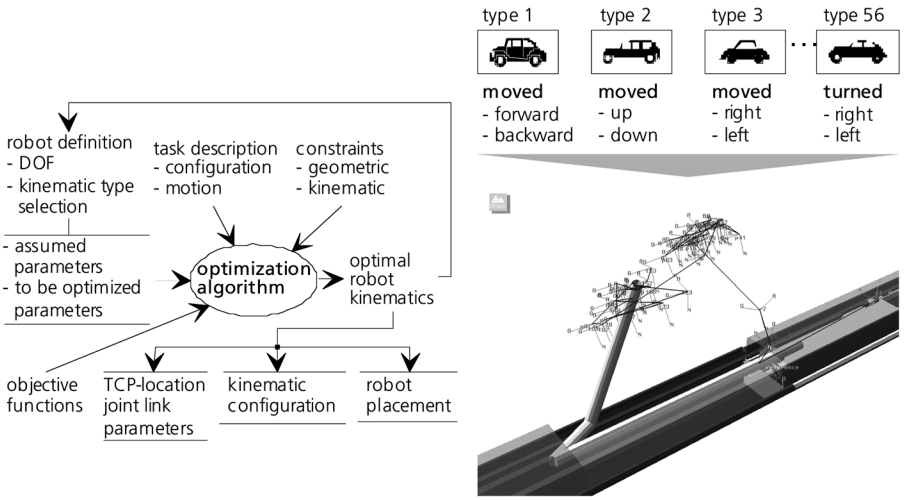


FIGURE 28.26 Robot kinematic optimization procedure (left) and goal frames and trajectories of a complete refilling cycle representing two extremal locations in the refilling station (right). (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

- Reachability suitable for any car type, considering assumed extremal positions or orientations
- Maximum clearance from potential collision objects
- Kinematic performance (dexterity)

The IGRIP CAPE tool was used as the kernel for the numerical simulation so that all motions, collision bodies, and geometric constraints could be interactively generated and visualized. See Figure 28.26.

28.3.3.4 Identification and Localization

Two general approaches for car identification were analyzed. See Figure 28.27.

Centralized data storage and retrieval — The car carries an individual serial number or type-specific code. After identification, the code is related to a specific motion program stored in a relational database. For any new car/type, the robot program and database reference will require immediate updates throughout all refilling stations. The transmitting of car-specific codes violates laws protecting personal data.

Decentralized data storage and retrieval — This preferred concept avoids these disadvantages. Vehicle identification takes place via a passive data carrier (transponder) located in the underfloor of the car. When the car is driven to the refilling station, the data stored in the transponder are scanned by a signal loop under the road surface. The data required include vehicle type, permitted fuel selection, maximum supply rate, and geometrical data, as Figure 28.28 depicts. These data are transferred into a standard robot motion program.

Since all trajectories and goal frames correspond to the car's reference frame, its location relative to the robot's base K_0 must be determined. Two laser scanners integrated into the entry and exit bollards scan a given surface of the filling station (Figure 28.29). Once the vehicle contour has been recognized and compared with the known dimensions of the detected car type, its exact position can be determined. The space defined by the vehicle contour and the curtain pattern of the scan define the safety zone. Any changes inside the zone such as human movements, opening doors, etc. are detected and temporarily freeze the robot.

© 2002 by CRC Press LLC

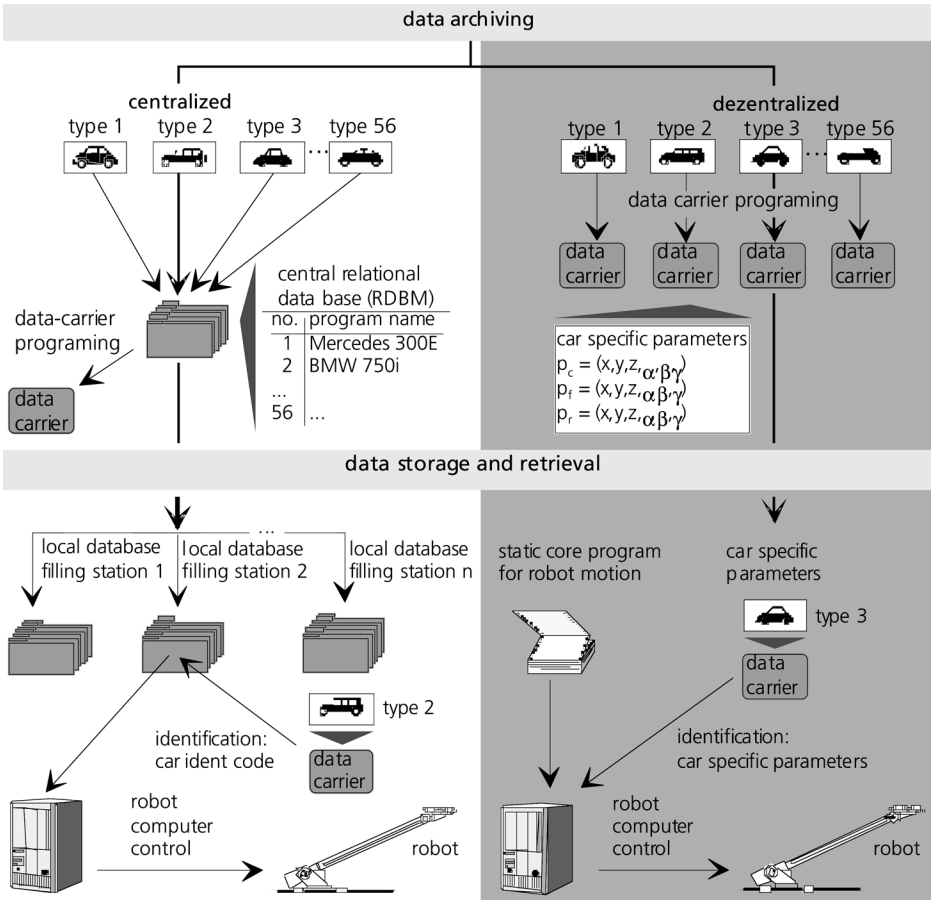


FIGURE 28.27 Centralized and decentralized data storage and retrieval for car-type identification and robot program selection. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

geometrical parameters	- location of flap and cap - opening trajectory of flap - flap type
fuel specific parameters	- allowable fuels - filling rate - tank volume
user data	- car type - standard fuel
auxiliary	- program version - program mode (standard/servicing) - service information - service programs for special car-types

FIGURE 28.28 Car-type specific data stored in the transponder. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

© 2002 by CRC Press LLC

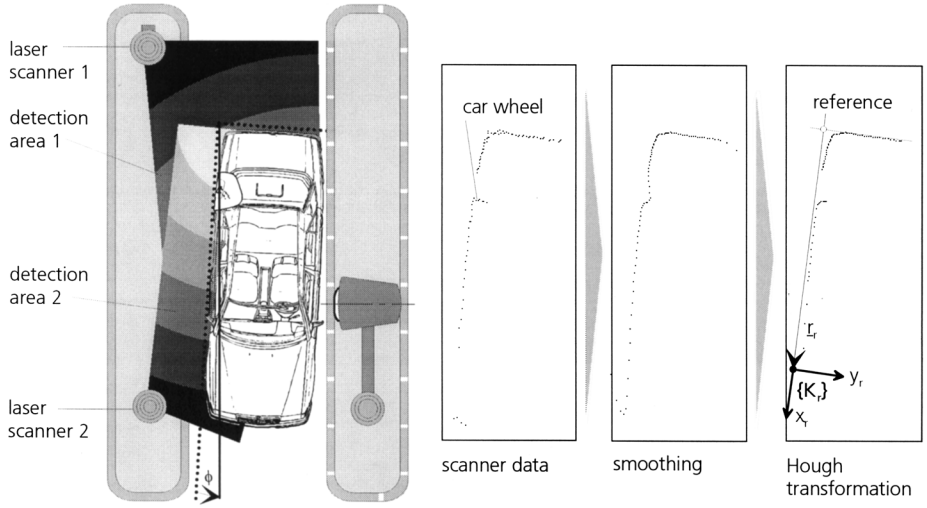


FIGURE 28.29 Vehicle location using two laser scanners with sensor data acquisition and modeling. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

28.3.3.5 Robot End-Effector

The end-effector as shown in Figure 28.30 is the interface between robot and filler flap or cap. The flap is lifted by two suction elements and opened by the robot's turning motion. A cylindrical docking-on element, the tank dome, establishes the mechanical connection and disconnection. When approaching the cap, the element is driven forward by a pneumatically powered tendon drive. The nozzle's entry and exit movements are driven by a second feed drive. The toothed ring recesses on the cap and turns it through 25° so that the fuel nozzle can be inserted. During refilling, the docking-on element's sealing action and integrated gas recirculation ensure that no emissions or odors are produced. The cap also permits refueling without difficulty.

In an emergency, for instance if the vehicle suddenly starts, the robot is disconnected instantly by the release of springs in the pneumatic cylinders. A graph representing the event structure and the step-wise increase in docking accuracy is depicted in Figure 28.31.

28.3.3.6 Docking Sensors

From their initial approach location (see Figure 28.25), the docking sensors detect and follow the reflectors on the filler flap and cap. LEDs pulse infrared or deep red light through fibers that illuminate the scene in front of the end effector. The line feed sensor receives the reflected light signal through a fiber-optic arrangement integrated in the docking-on element. To reach signal cycles of up to 200 Hz, the thresholds produced by the contrast between reflecting tape and its less reflecting vicinity are processed. The three fibers with their opening angles of some 60° cover three 120°-segmented lines about the optical axis as Figure 28.32 depicts.

The threshold of the reflected light produces signal peaks detected by corresponding pixel segments on a single line sensor. The positions of the peaks on each line sensor segment are measures of optical axis' displacement (ε_x, ε_y) from the reflectors center. This displacement is transmitted to the robot control which corrects end effector motion to the center of the reflector. The high sensor cycle time allows dynamic goals to be tracked effectively.

28.3.3.7 Experiments and Further Developments

The robot forms a compact functional unit with the refilling island and the delivery technology. The robot pulls out the correct fuel hose and nozzle based on the customer's choice of fuel.

© 2002 by CRC Press LLC

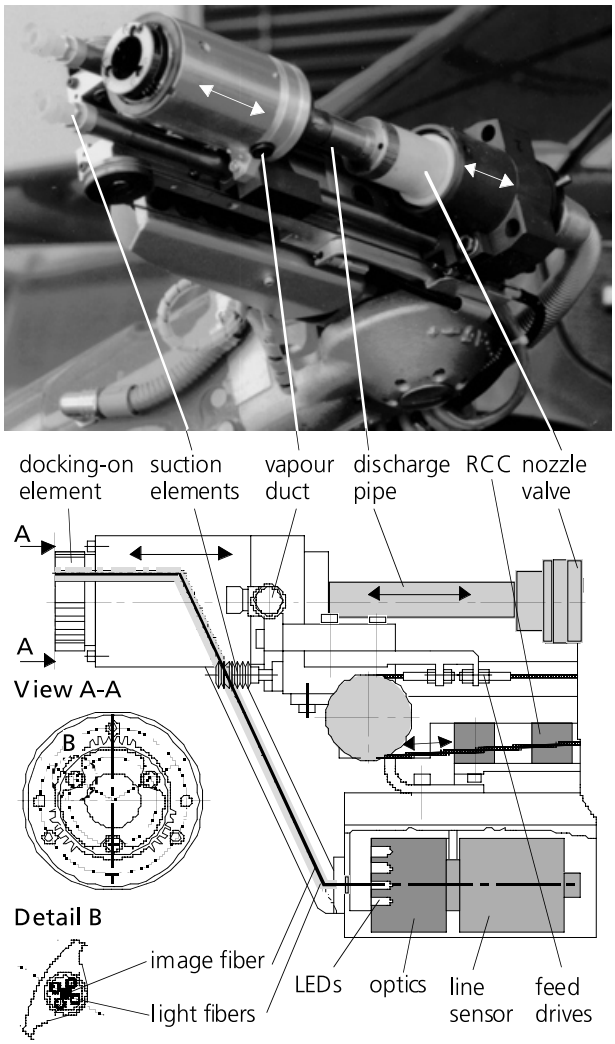


FIGURE 28.30 End-effector for automatic refueling. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

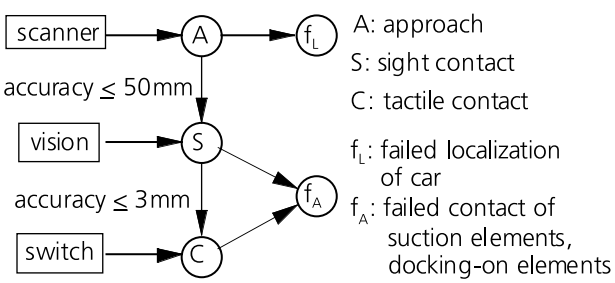


FIGURE 28.31 Event structure of the docking process. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

© 2002 by CRC Press LLC

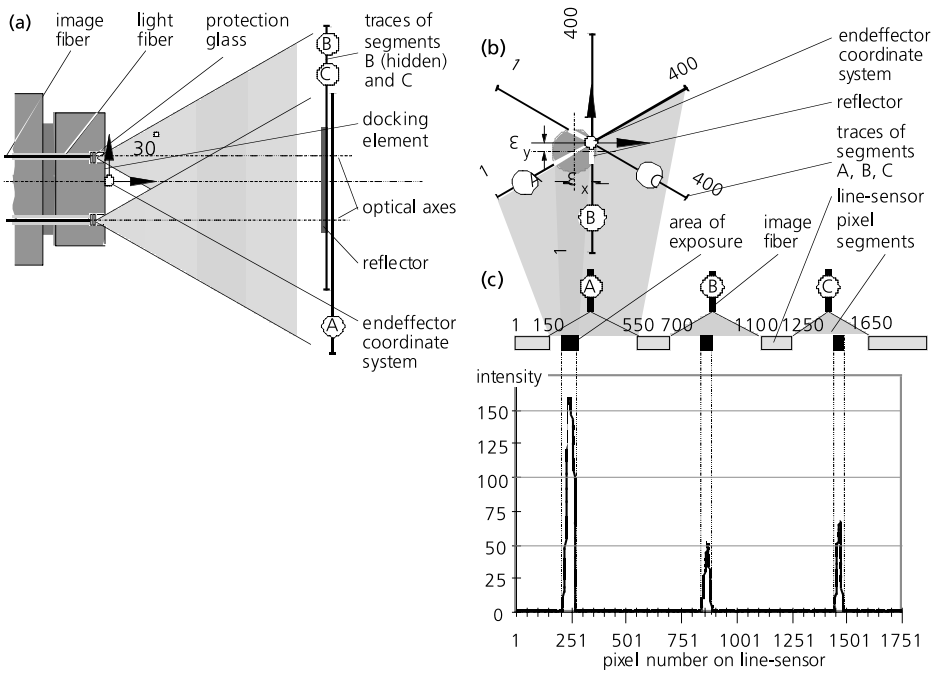


FIGURE 28.32 Working principle of the docking sensor. (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

FIGURE 28.33 View of a prototype installation at Fraunhofer IPA. A car being refueled by a robot (left) and a touch-screen terminal for inserting credit card, entering refilling order and printing (right). (From Leondes, C.T., *Mechatronic Systems Techniques and Applications,* Vol. 2, Gordon & Breach, Amsterdam, 2000. With permission.)

Underneath the refueling station, the robot moves into the initial position. It emerges from the opening in the refueling island and approaches the filler flap. The robot remains flexible when docked on, in other words, it can respond to vehicle movement even when subjected to a slight load.

Personal safety is enhanced by passive design measures and active optical sensors. During refueling, the area surrounding the robot is monitored for changes. Human movements, opening doors, etc. are detected during the docking-on process. The vehicle can be left at any time in an emergency, since nothing prevents the car door from opening. Safe access to the refueling island is guaranteed at all times. Figure 28.33 depicts a refilling station in operation since September 1995 at Fraunhofer IPA.

For more than 3 years, the robot has shown its reliability and robustness under even harsh conditions. The system is currently undergoing redesign to meet cost and operation requirements.

© 2002 by CRC Press LLC

References

1. Angeles, J., *Fundamentals of Robotic Mechanical Systems. Theory, Methods and Algorithms,* Springer–Verlag, New York, 1997.
2. Arbib, M.A. and Liaw, J.S., Sensori-motor transformations in the world of frogs and robots, *Artif. Intelligence,* 72, 53, 1995.
3. Bogoni, L. and Bajcsy, R., Functionality investigation using a discrete event system approach, *Robotics and Autonomous Syst.*, 13, 173, 1994.
4. Cutkosky, M.R., On grasp choice, grasp models, and the design of hands for manufacturing tasks, *IEEE Trans. Robotics Automation,* 5, 269, 1989.
5. Engelberger, J.F., Robotics R&D in the U.S.A., *Proc. 24th Int. Conf. Ind. Robots*, Tokyo, 1993.
6. Leondes, C.T. (Ed.)*, Mechatronic System Techniques and Applications, Transportation and Vehicular Systems,* Vol. 2, Gordon and Breach, Amsterdam, 2000.
7. Hirose, S., A code of conduct for robots coexisting with human beings, *Robotics Autonomous Syst.*, 18, 101, 1996.
8. Iberall, T., Jackson, L., Labbe, L., and Zampano, R., Knowledge-based prehension: capturing human dexterity, *Int. Conf. Robotic Res.*, 82, 1988.
9. UN/ECE, International Federation of Robotics (IFR), *World Robotics 2000*, United Nations Economic Commission for Europe (UN/ECE), Geneva, Switzerland, 2000.
10. Kim, J.-O. and Koshla, P., Design of space shuttle tile servicing robot: an application of task-based kinematic design, 10, 648, 1994.
11. Masory, O., Wang, J., and Zhuang, H., On the accuracy of a Stewart platform. Part II: Kinematic calibration and compensation*, Proc. IEEE Int. Conf. Robotics Automation*, Atlanta, 1993.
12. Merlet, J.-P., Designing a parallel robot for a specific workspace, *Res. Rep.* 2527, 1995.
13. Morrow, D.J., Sensori-motor primitives for robot assembly skills*, Proc. IEEE Int. Conf. Robotics Automation*, Nagoya, 1995.
14. Murray, R.M., Li, Z., and Sastry, S.S., *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, FL, 1993.
15. International Federation of Robotics, *Draft of the IFR Robot Statistics Documentation Package*, Revision 1, 1997.
16. Roth, B. and Mavroidis, C., Structural parameters which reduce the number of manipulator configurations, *ASME J. Mech. Design*, 116, 3, 1994.
17. Schraft, R.D., Degenhart, E., and Hägele, M., New robot application in production and service, *Proc. 1993 IEEE/Tsukuba Int. Workshop Adv. Robotics*, AIST Tsukuba Research Center, Tsukuba, Japan, 1993.
18. Schraft, R.D. and Hägele, M., Methods and tools for an efficient design of service robot applications, *Proc. 26th ISIR*, Singapore, 1995.
19. Schraft, R.D., Hägele, M., and Volz, H., Service robots: the appropriate level of automation and the role of users — operators in the task execution, *Proc. 2nd Fraunhofer IPA Technologie Forum F 17*, Stuttgart, Germany, 1996.
20. Schraft, R.D., Hägele, M., Heni, M., and Seid, R., *Mechatronic system techniques for robots for service applications*, in Leondes, C.T. (Ed.)*, Mechatronic System Techniques and Applications, Transportation and Vehicular Systems,* Vol. 2, Gordon and Breach, Amsterdam, 309, 2000.
21. Wang, J. and Masory, O., On the accuracy of a Stewart platform. Part I: The effect of manufacturing tolerances, *Proc. IEEE Int. Conf. Robotics Automation*, Atlanta, 1993.
22. Warnecke, H-J., Schraft, R.D., Hägele, M., Barth, O., and Schmierer, G., Manipulator Design in *Handbook of Industrial Robotics,* Nof, S. Y., Ed., John Wiley & Sons, New York, 42, 1999.

© 2002 by CRC Press LLC