Advances in Industrial Control

Tan Kok Kiong
Andi Sudjana Putra

Drives and Control for Industrial Automation

AIC

Springer

Advances in Industrial Control

For other titles published in this series, go to
www.springer.com/series/1412

Other titles published in this series:

Digital Controller Implementation and Fragility
Robert S.H. Istepanian and James F. Whidborne (Eds.)

Optimisation of Industrial Processes at Supervisory Level
Doris Sáez, Aldo Cipriano and Andrzej W. Ordys

Robust Control of Diesel Ship Propulsion
Nikolaos Xiros

Hydraulic Servo-systems
Mohieddine Mali and Andreas Kroll

Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques
Silvio Simani, Cesare Fantuzzi and Ron J. Patton

Strategies for Feedback Linearisation
Freddy Garces, Victor M. Becerra, Chandrasekhar Kambhampati and Kevin Warwick

Robust Autonomous Guidance
Alberto Isidori, Lorenzo Marconi and Andrea Serrani

Dynamic Modelling of Gas Turbines
Gennady G. Kulikov and Haydn A. Thompson (Eds.)

Control of Fuel Cell Power Systems
Jay T. Pukrushpan, Anna G. Stefanopoulou and Huei Peng

Fuzzy Logic, Identification and Predictive Control
Jairo Espinosa, Joos Vandewalle and Vincent Wertz

Optimal Real-time Control of Sewer Networks
Magdalene Marinaki and Markos Papageorgiou

Process Modelling for Control
Benoît Codrons

Computational Intelligence in Time Series Forecasting
Ajoy K. Palit and Dobrivoje Popovic

Modelling and Control of Mini-Flying Machines
Pedro Castillo, Rogelio Lozano and Alejandro Dzul

Ship Motion Control
Tristan Perez

Hard Disk Drive Servo Systems (2nd Ed.)
Ben M. Chen, Tong H. Lee, Kemao Peng and Venkatakrishnan Venkataramanan

Measurement, Control, and Communication Using IEEE 1588
John C. Eidson

Piezoelectric Transducers for Vibration Control and Damping
S.O. Reza Moheimani and Andrew J. Fleming

Manufacturing Systems Control Design
Stjepan Bogdan, Frank L. Lewis, Zdenko Kovačić and José Mireles Jr.

Windup in Control
Peter Hippe

Nonlinear H_2/H_∞*, Constrained Feedback Control*
Murad Abu-Khalaf, Jie Huang and Frank L. Lewis

Practical Grey-box Process Identification
Torsten Bohlin

Control of Traffic Systems in Buildings
Sandor Markon, Hajime Kita, Hiroshi Kise and Thomas Bartz-Beielstein

Wind Turbine Control Systems
Fernando D. Bianchi, Hernán De Battista and Ricardo J. Mantz

Advanced Fuzzy Logic Technologies in Industrial Applications
Ying Bai, Hanqi Zhuang and Dali Wang (Eds.)

Practical PID Control
Antonio Visioli

(continued after Index)

Tan Kok Kiong • Andi Sudjana Putra

Drives and Control for Industrial Automation

Springer

Tan Kok Kiong, Ph.D.
Department of Electrical and Computer Engineering
National University of Singapore
4 Engineering Drive 3
117576 Singapore
Singapore
kktan@nus.edu.sg

Andi Sudjana Putra
Engineering Dean's Office
National University of Singapore
9 Engineering Drive 1
Singapore 117576
Singapore
engpas@nus.edu.sg

ISSN 1430-9491
ISBN 978-1-84882-424-9 e-ISBN 978-1-84882-425-6
DOI 10.1007/978-1-84882-425-6
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Cover design: eStudio Calamar

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Advances in Industrial Control

Series Editors

Series Advisory Board

Professor (Emeritus) O.P. Malik
Department of Electrical and Computer Engineering
University of Calgary
2500, University Drive, NW
Calgary, Alberta
T2N 1N4
Canada

Professor K.-F. Man
Electronic Engineering Department
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong

Professor G. Olsson
Department of Industrial Electrical Engineering and Automation
Lund Institute of Technology
Box 118
221 00 Lund
Sweden

Professor A. Ray
Department of Mechanical Engineering
Pennsylvania State University
0329 Reber Building
University Park
PA 16802
USA

Professor D.E. Seborg
Chemical Engineering
University of California Santa Barbara
3335 Engineering II
Santa Barbara
CA 93106
USA

Doctor K.K. Tan
Department of Electrical and Computer Engineering
National University of Singapore
4 Engineering Drive 3
Singapore 117576
Singapore

Professor I. Yamamoto
Department of Mechanical Systems and Environmental Engineering
Faculty of Environmental Engineering
The University of Kitakyushu
1-1, Hibikino,Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0135
Japan

Series Editors' Foreword

The series *Advances in Industrial Control* aims to report and encourage technology transfer in control engineering. The rapid development of control technology has an impact on all areas of the control discipline. New theory, new controllers, actuators, sensors, new industrial processes, computer methods, new applications, new philosophies, ..., new challenges. Much of this development work resides in industrial reports, feasibility study papers and the reports of advanced collaborative projects. The series offers an opportunity for researchers to present an extended exposition of such new work in all aspects of industrial control for wider and rapid dissemination.

Monographs from the academic control community commonly have a strong focus on control system *design*, but this is only one aspect of industrial control, so it is pleasing to be able to introduce a monograph into the *Advances in Industrial Control* series that is concerned with a topic from the equally important area of control *technology*. Kok Kiong Tan and Andi Sudjani Putra from the National University of Singapore have worked for several years with industrial engineers and university students on the technology of drives and their applications. Their work has covered teaching, research, and applications, and now their experience has been captured in this comprehensive monograph *Drives and Control for Industrial Automation*. One focus of the book is to describe the hardware and working principles of hydraulic and pneumatic servo-drives, electric drives, and piezoelectric drives, all of which are presented and reviewed in one chapter each (Chapters 2–4, respectively). A general control system structure for these drives is then given in Chapter 5, where, being an industrially-oriented monograph, the control focus and discussion is on proportional-integral-derivative (PID) control. The use of a generic control system structure across the differing drive technologies reinforces the authors' approach to the industrial servo-drive as a packaged unit integrating sensors, actuators (prime mover), power moderation, and control system. To incorporate such a drive unit into a process application leads naturally to a consideration of industrial process communication technology and communication protocols; these are described in Chapter 6, where the focus is on fieldbus technology. The final chapter of the book reports on recent and future trends in motion control. The key developments identified are an industrial demand for miniaturisation and the growth of applications in the nano- and bio-technology fields.

Readers seeking an entry and introduction to the prevalent devices and current methods for servo-drive technology will find this monograph quite accessible. Such readers might include final-year undergraduate students, engineering postgraduates, industrial engineers, control engineers, and technologists, typically from the fields of electrical, mechanical, aviation, and process engineering. The breadth of the contents of the monograph means that it can also be used as a reference text for servo-drive technology.

Whilst this monograph from K.K. Tan and A.S. Putra has the advantage of comprehensiveness, readers seeking further specialist knowledge might find the following *Advances in Industrial Control* series monographs useful. In the field of hydraulics, the monograph *Hydraulic Servo-systems* (ISBN 978-1-85233-692-9) by M. Jelali and A. Kroll, for electric motors, the new monograph *Induction Motor Control Design* (ISBN 978-1-84996-283-4) by R. Marino, P. Tomei, and G.M. Verrelli. In the field of piezoelectric devices, the monograph *Piezoelectric Transducers for Vibration Control and Damping* (ISBN 978-1-84628-331-4) by S.O.R. Moheimani and A.J. Fleming, and finally in communications, the monograph *Measurement, Control and Communication* using IEEE1588 (ISBN 978-1-84628-250-8) by J.C. Eidson. For PID control, the *Advances in Industrial Control* series offers a number of seminal texts including: *Advances in PID Control* (ISBN 978-1-85233-138-2) by K.K. Tan, Q.-G. Wang, and C.C. Hang with T.J. Hägglund; *Precision Motion Control* (ISBN 978-1-84800-020-9) by K.K. Tan, T.H. Lee, and S. Huang; *Practical PID Control* (ISBN 978-1-84628-585-1) by A. Visioli, and finally for something a little different, *Model Predictive Control System Design and Implementation with MATLAB®* (ISBN 978-1-84882-330-3) by L. Wang.

Industrial Control Centre
Glasgow
Scotland, UK

M.J. Grimble
M.A. Johnson

Preface

Industrial automation has become an important feature today, especially in this age of rapid production an high precision. Automation allows industries to achieve the level of speed and quality unattainable by labour power; with affordable cost. While industrial automation is mostly profitable for mass manufacturing and homogeneous products, the bulk of industries produce goods in low quantity. In this situation, the challenge shifts into developing automation systems in industry that still justifies the installation cost. The knowledge and skill on this area has therefore become increasingly necessary.

This book recollects necessary materials related to servo control for industrial automation. It starts from a macroscopic view of servo control, especially for industrial automation, treating drives and control systems as inseparable entities. It then continues with detail discussions of major types of drives for precision control realization; namely servo hydraulic and pneumatic drives, electric drives, and piezoelectric drives. Each chapter contains detail discussions of the respective major components: actuators, sensors, and controllers—without going into the control theory. The techniques and theory of motion control itself is discussed in a separate chapter, considering that the control theory for all of the abovementioned drives is identical. For the same reason, digital communication protocol is also discussed in a separate chapter. This chapter is included as a recognition of the importance and growing trend of digitalization in motion and precision control. The more general trend in motion control is discussed in the closing chapter. Throughout the discussion, the integrity and nuance of mechatronics—a synergistic integration of the abovementioned components—are maintained, reflecting the reality of their synergy in today's industrial automation.

Despite its mechatronics nuance, the structure of this book allows traditional approach of step-by-step teaching to still be conducted should it be desired. Each chapter contains a material of its own that can be studied separately without compromising the understanding of the readers. This book is written for wide readership, from students, technicians, engineers, and researchers. The discussion is thorough, with concise basics yet sufficient details. Equations are provided as means to explain the certain concepts from the fundamentals such that it does not discourage inexperienced readers but is useful for those with prior knowledge. Readers who wish to know the applications of various sensors, actuators, and control systems in industrial automation will find this book of value. Readers will also find that the flow of the book reflects the current approach and view taken by the industry, yet is still sensible and is easy-to-read, which they can relate to the prior knowledge they have learned traditionally.

The inclusion of hydraulic and piezoelectric drives, as well as control and communication, is intended to ensure that the book covers all necessary aspects in control system. The discussion in the book starts from the history and the basic principle of each device, as well as the assembled systems. The synergistic integration of actuators, sensors, control systems, and communication protocols are maintained throughout the course of the book to reflect the current trend in industrial applications. This book is intended for professionals, engineers, and postgraduate students whose areas of interest are drives, sensors, and control system design. For teaching purpose, it is most suitable to courses such as: Control System, Mechatronic System Design, Industrial Drives, and Instrumentation and Sensors. For professionals, it is most suitable for those working in system design and control, which require broad perspective of drives and control system of plants.

This book is equipped with many illustrations, especially to present the working principles and structures of the abovementioned industrial systems. The combined usage of words and figures are prevalent in the entire book to convey clear concepts to the readers.

This book would not have been possible without the generous assistance of many colleagues and friends. The authors would like to express their sincere appreciation to Dr. Huang Sunan, Dr. Zhao Shao, Dr. Tang Kok Zuea, and Mr. Chen Silu, who have provided invaluable materials for this book. The authors would also like to thank Mr. Oliver Jackson, Ms. Aislinn Bunning, and Ms. Charlotte Cross, who have been extremely helpful in the editing and publication of the book. Finally, we would like to dedicate the book to our families for their love and support.

Singapore

Tan Kok Kiong
Andi Sudjana Putra

Contents

Chapter 1
Overview of Servo Control

The term *servo* originates from a Latin word *servus*, which means *servant* or *follower*. Along this perspective, a servo control system can be defined as a system that is able to control some variables of interest to track user-specified objectives closely. While the first contribution of servo control has generally been attributed to Ktesbios of Alexandria (*ca.* 200 B.C.E.) [75] for his invention of water clock, the continuous history of modern servo control started on 1788, when James Watt invented the fly-ball governor to regulate the speed of a steam engine. Subsequent development and invention of devices such as flow valves and pressure regulators contributed, historically, to the emergence of the Industrial Revolution and, technologically, to the servo control technology. Besides those “hardware” inventions, mathematical techniques and control algorithms were also devised, such as stability theory by Lyapunov around 1890 and frequency-domain analysis around 1920. A major boost to the development in this field came along with World War II, when servo control was used in diverse military applications, including the precise guidance and control of missiles, tracking of military targets, and development of navigational systems.

Today, servo control has become an integral part of almost every automation system or process, including in manufacturing, chemical, petrochemical, transportation, military, and biomedical. While the broad definition of servo control as mentioned above still holds, the expectations in terms of the tracking performance of servo control systems have risen significantly, in line with the ever tightening and stringent requirements associated with the products of today and the processes to achieve them.

This book will focus primarily on servo control in the application domain of *motion* control systems, although some of the topics covered will remain applicable to other application domains such as in process control systems.

1.1 Objectives of Servo Control

Generally, the objective of a servo control system is to make a controlled signal follow or track a reference input signal, sometimes also called the set-point, at cer-

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_1,

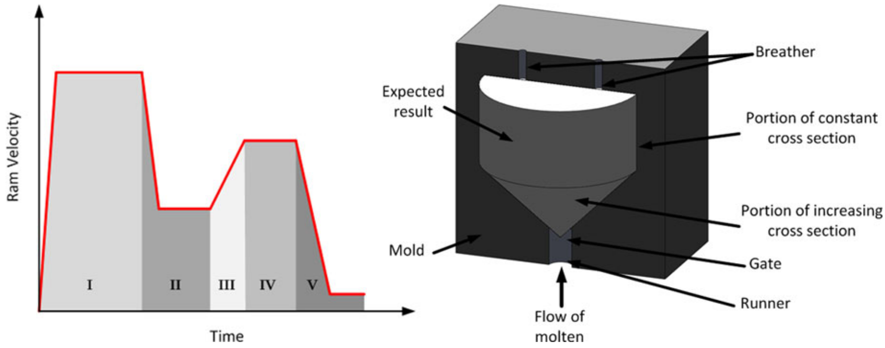

Fig. 1.1 Optimal ram velocity profile for injection molding system

tain speed and accuracy, and to remain robust to keep the controlled signal on track, despite possible undesirable disturbances affecting the system.

In the domain of motion control applications, servo control is concerned with the control of motion enablers, *i.e.* the actuators, to achieve desired motion profiles of the load to which they are attached to, in terms of direct motion variables such as proximity, position, velocity, and acceleration, or in terms of motion-induced variables such as force and torque. Disturbances to the achievement of this objective may arise in the form of load changes and the existence of motion impeding forces such as friction, backlash, and cogging forces.

The reference signal of a servo control system is the command input to the servo control system. It can be specified to optimize the quality of a product or the performance of a process. Two examples will be provided to illustrate this point.

Example 1.1 (Injection molding machine) The ram velocity profile of an injection molding machine affects the precision and consistency at which the plastic components are produced. The optimal profile depends on factors such as the specific resin material properties and the ram geometry, and it is typically designed and generated from computer simulation software with a model of the whole system. Figure 1.1 shows an example of a ram velocity profile. With the desired profile as the reference signal, the servo control of the injection molding system will work to force the ram velocity to track this profile with minimum offset.

The velocity profile presented in Figure 1.1 is a typical ram velocity profile to fill a mold of increasing cross-section initially, and having a constant cross-section thereafter. The profile for this part can be divided into five sections. The profile in Section I is used for filling the runner, when velocity is very rapidly increased and then held constant. In Section II, the velocity is rapidly reduced to eliminate jetting at the gate. Section III corresponds to filling of the increasing cross-section portion of the cavity, while Section IV is concerned with filling of the constant cross-section portion of the cavity. Finally, the velocity is reduced in Section V to eliminate flushing and over-packing. All of the above objectives have to be achieved in the short duration of time available for mold filling [116].

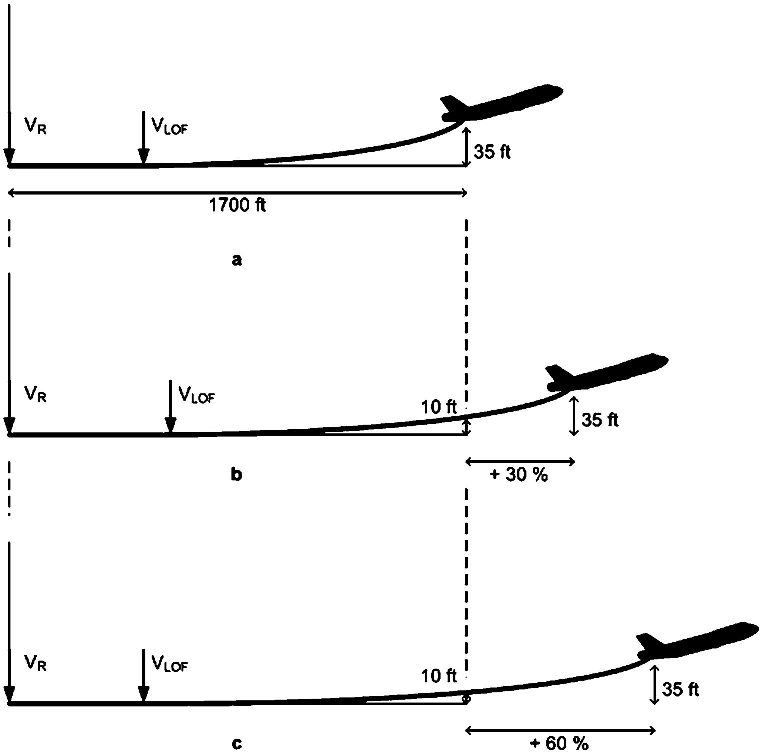

Fig. 1.2 Velocity profile for plane take-off, with rotation speed V_R and lift-off speed V_{LOF}; (**a**) normal rotation, (**b**) slow rotation, (**c**) under rotation

The application as provided above in *Injection molding machine* is an example where the operation parameters are noncontinuous. There are also applications where continuous motion is a demand for various reasons, as presented in the example below.

Example 1.2 (Flight control) An experienced pilot is able to take-off an airplane smoothly, causing minimal discomfort to the passengers and crews onboard. The velocity profile of the airplane along the runway determines how smooth the take-off will be. Figure 1.2 shows a possible velocity profile for a smooth take-off with three typical scenarios. The pilot tracks this profile via a servo control system, either on the actual airplane or for trainee pilots via a flight simulator.

The performance of a servo control system is therefore evaluated by indicators measuring how closely the objective reference signal is tracked, which is commonly based on the magnitude of the root-mean-square (RMS) tracking error over the profile. For point-to-point tracking (*i.e.* step change in reference signal), classical performance indicators can be used, including the rise time, overshoot, and steady state error. Figure 1.3 shows these indicators in relation to the response of the system to

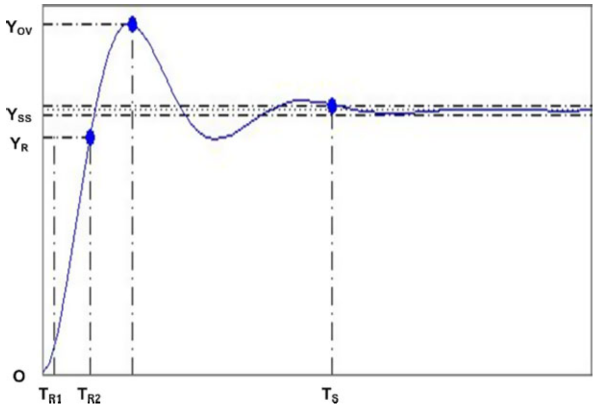

Fig. 1.3 Classical performance indicator using step response of a system; O origin point, Y_{SS} steady-state value, Y_{OV} overshoot value, Y_R rising value, T_R rise time, T_S settling time

a unit step change in the reference signal. This is a typical response of a second-order system to a step change, which causes the output to increase above the set point, oscillate for some time, and eventually settle down at a steady-state value according to the set point value. The amount of overshoot is usually indicated as a ratio of

$$\frac{Y_{OV} - Y_{SS}}{Y_{SS} - O},$$

with respect to Figure 1.3. Settling time is determined by the time required by the output to stay within a certain amount of oscillation from the steady-state value, *i.e.* typically within $\pm 1\%$ of the steady-state value. There are several ways to determine rise time, one of which is the time from 10% to 90% of the output, *i.e.* $T_{R2} - T_{R1}$.

1.2 Elements of a Servo Control

In this section, let us consider a flight simulator used for training a pilot.

The flight simulator is designed so that it will emulate the actual situations of an airplane on the ground and in the sky, as closely as possible to match various plausible scenarios which an aspiring pilot is likely to encounter in real life. Through the simulator, the trainee, or even a professional pilot, can experience the feel of flying an airplane at hardly a fraction of the risk incurred.

The flight simulator generates different motion profiles through a multiple-degree-of-freedom motion platform based on kinematics and dynamics models of the actual plane, and according to the pilot commands, current plane parameters, and possible exceptional situations. This complex system is a typical system which can be realized through a delicate application of servo control technology.

The flight simulator can be decomposed into several key functions, working in synchronization. The subsections to follow will briefly highlight the key components of a servo control system which will carry out these functions, with specific reference to the flight simulator. General details on these components will be covered in the subsequent chapters.

1.2.1 Measurement

In the cockpit, the pilot will be able to know the current flight parameters through arrays of display panels, meters, and alarms. These devices essentially display important measurements such as the current plane orientation, the current flight profile, and wind conditions. They will also show the operational status of key parts of the airplane, such as wings, ailerons, and tails and will sound/display alarm when faults and malfunctions are detected. This information is necessary for the pilot to make informed decisions on any necessary corrective actions. This measurement function is accomplished via a set of sensors and instrumentation circuits, which will collectively collect the data and convert them to a form that is amenable for display and control purposes.

1.2.2 Actuation

A framework of actuators is installed right below the cockpit to generate and deliver the appropriate motion profile of the entire platform according to the pilot's commands, and also to generate and deliver the effects of varying weather conditions and turbulences. The cockpit is the load which is driven by these actuators. The means powering the actuators can be in the form of air pressure, hydraulic pressure, or an electromagnetic force, respectively the mechanisms to operate a pneumatic, hydraulic, or electromagnetic actuator.

The actuators are capable of generating multiple degree-of-freedom motion to the cockpit to yield the same sense of flight to the pilot in a real airplane. Each actuator will therefore need to be controlled to such precision to create the overall effect which will emulate the real situation as closely as possible. In addition, the interaction and coupling between the individual actuators within the framework has to be adequately addressed, too. A high-performance control system will be necessary to fulfill this function.

1.2.3 Power Moderation

Power moderation is the intermediary function to match the pilot or control command signal to the final actuation signal for the actuator, both in form and energy level. The specific moderation varies considerably, depending on the overall system design. The moderation process may include amplification, conversion, or switching, depending on the overall system design, and the type and operational principle of the actuators used.

The role of power moderation is to enable adequate communication among different components of the servo control system, considering that different components at different levels operate at different magnitude order. Suppose that the flight simulator uses DC electric motors for actuation, requiring a few hundreds of volts. Measurement signals from sensors, on the other hand, are usually in lower energy forms of the order of millivolts (mV). Power moderation may then involve signal amplification to amplify the measurement signal to a few volt level to be processed by a computer or be displayed on a meter, and then to the hundreds of volt level to drive the motors. Amplification does not only apply to electrical signals; there are also pneumatic and hydraulic amplifiers.

For different types of actuators, such as pneumatic, hydraulic, or AC motors, other kinds of moderation may be necessary to convert the control signal to an alternate signal form which is needed to operate these actuators. AC motors, for example, may also require phase matching for their operation. Apart from amplification, other common moderation includes current-to-voltage conversion, current-to-pressure conversion, signal switching, and signal chopping.

1.2.4 Control

The control function in a servo control system is the core function which helps to integrate all other functions. Essentially, it generates and transmits the specific motion commands to the power moderator.

From the cockpit of the flight simulator, the pilot can give instructions and commands via the control panel on target flight parameters to achieve, such as speed and flight path. The motion commands can be directly transmitted to the power moderator, so that the pilot directly controls the actuators. Alternatively, the motion command can serve as the reference signal or set-point for an automatic control system. In this case, the trainee pilot may only need to specify the target speed and flight path, and the automatic control system uses the specifications to manipulate the actuators based on the feedback measurements. While the first case is called open-loop control, the latter case is called closed-loop control.

Apart from executing pilot commands for standard actions related to take off, landing, turning, and maneuvering, the control system will also need to respond adequately to exceptional conditions such as head wind, turbulence, storms, heavy rain, or even malfunction of some parts of the simulator. From a control system perspective, these are the disturbances which may disrupt the otherwise well-controlled systems. Thus, apart from tracking and executing commands, the control system in a flight simulator has to remain robust to the disturbances which a plane may experience.

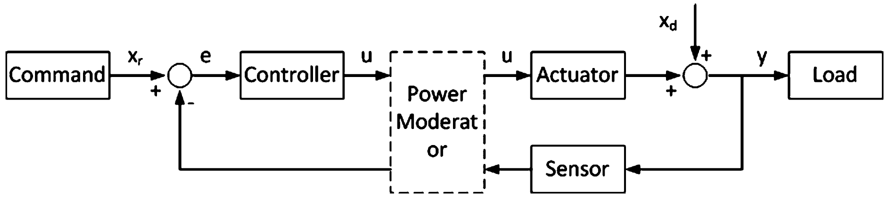

Fig. 1.4 Servo control system; x_r reference signal, e error, u control signal, x_d disturbance, y output

1.2.5 Putting Them All Together

These various components are now ready to be fit in place together to result in a complete system. Figure 1.4 presents the integration of these components in a simple control system; in reality, however, the servo control system may adopt more complex structures as will be covered in Chapter 5.

A reference signal or command is input into the system as the desired motion profile to be achieved. This command can be an input from the pilot or, in the case of an autopilot system, one from a higher-level supervisory control system. The present output of the system is the controlled motion variable, and it can be measured by sensors to yield the feedback signal. This feedback signal is then compared to the reference signal, the difference of which is the error signal. Based on the error signal, the controller will output a necessary control signal to the actuator, via the power moderator, so as to drive the load according to the desired motion profile.

In addition, disturbance signals X_d are also highlighted in Figure 1.4, representing the effects of extraneous signals seeping into the loop. The environment alone will expose the system to various conditions, which can be unpredictable. Disturbances, if not adequately compensated for, can cause the controlled variable to deviate from the reference signal. For example, due to tail wind, the plane can be accelerated beyond the desired profile. With a feedback controller, the control action will have to be reduced to bring the plane back on track.

The four key functions of a servo control system will be covered in detail in subsequent chapters. Collectively, the overall system, integrating these functions, is also often referred to as a servo drive. It will be a hydraulic servo drive if hydraulic pressure is used as the mean for actuation, or an electric servo drive if an electromotive force is used. However, it is important to clarify that this is not necessarily a universal definition of *drive*. To some manufacturers, the term *drive* may exclude the actuation function. In this book, however, the full system is referred to as the drive.

Chapter 2
Servo Hydraulic and Pneumatic Drive

Among the various drives available, the hydraulic drive ranks at the forefront in terms of application history. The history of hydraulic power dated to the beginning of civilization, where artefacts show the applications of water turbines in power generation [132]. However, more significant progress in this field is generally achieved following Pascal's pioneering work which was later known as Pascal's law of hydrostatic. Bernoulli's discovery of the hydrodynamic law in 1750, and the Industrial Revolution in 1850 further catalysed the development of servo hydraulic drives. They led to the first applications of hydraulic equipment as a power source to power industrial machines such as the press, the crane, and the jack, as they also contributed to the development of hydraulic pumps, driven by steam engines, which produce hydraulic energy to run hydraulic systems. In the early 20th century, a revolutionary design of hydraulic drives used the oil, instead of water, as the hydraulic liquid, which greatly expanded the applications of hydraulic drives to more devices. World War II also contributed to the development of hydraulic drives, especially in the development of submarine control systems, radar/sonar drives, and military cargo transportation.

Hydraulic drives are still being used today, and, in fact, their applications are expanding to loads of increasing mass and power requirements, yet with higher speed and control precision. The main attraction of a hydraulic drive is its high power-to-weight ratio, rendering it the natural servo drive to use for heavy applications found in the aircraft and space shuttle. Moreover, it possesses several desirable characteristics such as accuracy, flexibility in applications, and simplicity of operations, as it also allows for fast, smooth, and precise start, stop, and reversal actions.

2.1 Overview of Servo Hydraulic and Pneumatic Drive

Hydraulics has been around in human civilization since the ancient times. Historical records provide evidence that water mills had been used around 100 B.C.E. Another remarkable invention was a screw pump developed by ancient Greek's Archimedes around 300 B.C.E. The invention of water clock by Ktesbios around 250 is regarded

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_2,

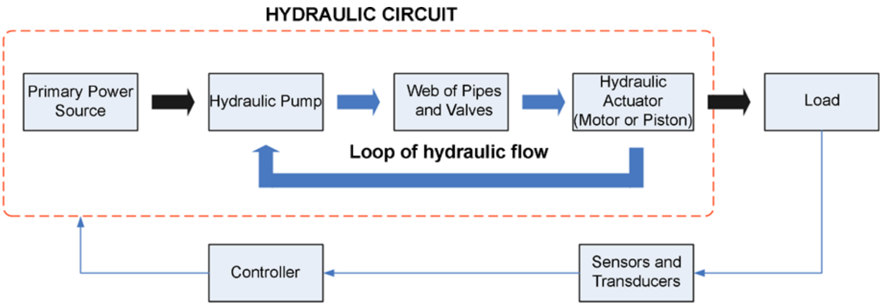

Fig. 2.1 Configuration of a servo hydraulic drive

not only as a remarkable invention in the field of hydraulic system, but also in the field of servo control.

The development of the theoretical foundation of hydraulics and pneumatics is spearheaded by Sir Isaac Newton with his viscosity theory, Daniel Bernoulli with his Bernoulli's equation, and Blaise Pascal with his Pascal's law.

The main events that boosted significantly the development of servo hydraulic and pneumatic drive were the Industrial Revolution and World War II. During the time of Industrial Revolution, the interest of using fluid power was growing rapidly, creating many inventions such as various pumps, mills, and steam engines. This provided a fertile ground for the development of pneumatic power system around 1900. World War II further accelerated the advancement of fluidic power system with the need to develop more powerful military aircrafts. Pneumatic drive proved to be the answer to this demand since it allowed the replacement of the cable-operated flight control, thereby allowing the pilot to operate the aircraft with a push of a button. This in turn enhanced the development of civil aviation technologies, whose working principles have been applied to less sophisticated system such as door-mechanism in public transportation.

Hydraulic and pneumatic systems share many similarities in terms of working principle and theoretical foundation; further discussed in Section 2.2. From this point onward, the term *servo fluidic drive* will be used to refer to both *servo hydraulic drive* and *servo pneumatic drive*.

A servo fluidic drive comprises several fluidic components which work in unison to deliver the functions highlighted in Chapter 1. A typical configuration of a servo hydraulic drive is shown in Figure 2.1, while that of pneumatic drive is shown in Figure 2.2.

In a hydraulic system, the primary power source, usually an electric motor, powers a hydraulic pump, which in turn generates a flow of the hydraulic liquid through a web of pipes, tubes, valves, and other hydraulic components, to a hydraulic actuator (a hydraulic motor or a hydraulic piston). The actuator will, in turn, drive the load. After transmitting the energy to the actuator, the hydraulic liquid will be returned to the reservoir to be recirculated by the pump. In a pneumatic system, a similar process also takes place via a pneumatic compressor, followed by a web of

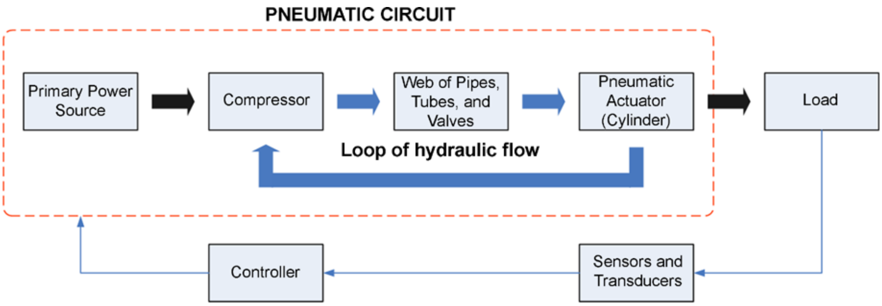

Fig. 2.2 Configuration of a servo pneumatic drive

Table 2.1 Comparison of characteristics of hydraulic and pneumatic drives

Hydraulic drive	Pneumatic drive
Instantaneous reaction	Delayed reaction
Hold load without unwanted movement	Hold load with some unwanted movement
Provide significant lubrication and cooling	Provide limited lubrication and cooling
More complicated design	Less complicated design
More expensive	Less expensive
Pose environmental problems (oil leakage)	No environmental problems

pneumatic circuitry similar to that of its hydraulic counterpart before powering up an actuator to finally drive the load.

The part of the drive within the dotted box is referred to as the hydraulic/pneumatic circuit, where the specific configuration depends on the fluidic characteristic required in each application. Common fluidic circuits will be presented in Section 2.4.

In a servo fluidic drive, sensors and transducers are also installed to yield the measurement signals for the controller. The controller will manipulate the controllable components in the fluidic circuit to achieve the control objectives.

2.2 Fundamentals of Hydraulic and Pneumatic Drives

Hydraulics is the principle of transmitting energy using liquids to achieve useful work. It shares many similarities with pneumatics, which uses gases to transmit energy. Collectively, liquids and gases are also called fluids, both of which exhibit a flow tendency when there is a pressure difference between two points along the flow path.

The main difference between liquids and gases lies in their compressibility; liquids are incompressible while gases are compressible. This difference influences the characteristics of a hydraulic and a pneumatic drive, and in turn, the applications suitable for each of the drives. Table 2.1 presents the main difference between hydraulic and pneumatic drives.

When a fluidic drive is complemented with a control system to achieve better performance, it becomes a servo fluidic drive. A servo fluidic drive measures its own outputs and takes control action so as to force the outputs to quickly and accurately follow a given command signal.

2.2.1 Basic Definitions and Principles

Below are fundamental definitions and concepts pertaining to fluidic drives.

2.2.1.1 Weight and Weight Density

Since fluid has mass, it also has weight due to gravity. The SI unit of weight is $\mathrm{kg{\cdot}m/s^2}$. Weight w is defined as follows:

$$w = mg, \tag{2.1}$$

where m is mass, and g is acceleration due to gravity.

The ratio of weight to volume is called weight density s and is a characteristic of fluids. It is defined as follows:

$$s = \frac{w}{V}, \tag{2.2}$$

where V is volume.

Most hydraulic liquids—the ones used in hydraulic systems—have a weight density of about 880–930kg/m^3.

2.2.1.2 Pressure

Pressure p is defined as the force applied per unit area, as follows:

$$p = \frac{F}{A}, \tag{2.3}$$

where F is force, and A is area.

The SI unit of pressure is N/m^2 or equal to Pascal (Pa), although the *bar* (equal to 10^5Pa) is also commonly used.

Since every component in a fluidic drive has a fixed area, pressure is equivalently the resultant force to produce useful work. Therefore, a variation of actuator force or shaft torque is achieved by a variation of the fluidic pressure.

2.2.1.3 Flow Rate

Flow rate refers to the volume of moving fluid per unit time. Variation of flow rate in a fluidic circuit results in variation of rod velocity or shaft speed. Flow rate is related to pressure, in the sense that flow is the result of pressure difference in a system. Likewise, without flow there cannot be pressure rise in a system. The SI unit of flow rate is m^3/s.

2.2.1.4 Pascal's Law

Pascal's law is a very important principle in fluidics since it describes quantitatively how fluid transmits power in the form of pressure. Pascal's law can be stated as follows:

> *Pressure applied to a* confined *fluid is transmitted undiminished in all directions.*

Pascal's law is often formulated in the form of the following equation:

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}, \tag{2.4}$$

where index 1 and index 2 denote object 1 and object 2, respectively. This equation essentially explains that heavy load can be lifted with small force using fluidics.

In hydraulic drives, the operating pressure ranges from 1MPa to 40MPa, while some special applications may require operating pressures as high as 70MPa. A higher operating pressure can deliver the same amount of power at a reduced equipment size and weight, desirable characteristics for many applications.

2.2.1.5 Bernoulli's Equation

Bernoulli's equation is another important equation in fluidic circuit analysis. While Pascal's law describes the behaviour of static fluids, Bernoulli's equation explains the behaviour of dynamic/moving fluids. Bernoulli's equation can be stated as follows:

$$f_1 + p_1 + h_1 = f_2 + p_2 + h_2, \tag{2.5}$$

where f is flow, and h is elevation.

Bernoulli's equation relates to the conservation of energy in a fluidic system. All three terms in the equation relates to different forms of energy, which depend on the liquid's flow rate, pressure, and position, respectively.

2.2.1.6 Bulk Modulus

Bulk modulus is the measure of fluid incompressibility, where higher bulk modulus means higher incompressibility. Bulk modulus B is defined as

$$B = -\frac{\Delta p}{\epsilon}, \tag{2.6}$$

where Δp is pressure change, and ϵ is volume strain.

The bulk modulus of a fluid changes with pressure and temperature.

2.2.1.7 Boyle's Law

This law is strictly valid for gases, although it is relevant to the discussion of the accumulator of a hydraulic drive. Boyle's law states that gases obey the following relationship:

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2}, \tag{2.7}$$

where T is temperature.

According to this law, under constant temperature, pressurization brings about compression, and compression brings about pressurization.

2.2.1.8 Viscosity

Viscosity is a concept which is especially useful in hydraulic systems. Theoretically, viscosity is a measure of the internal resistance of a liquid to shear and hence to flow. Lower viscosity signifies that the fluid can flow more easily.

Although the concept of viscosity is simple, the formulation is more complicated compared to other parameters. There are two measures of viscosity:

- absolute viscosity, defined as the force required to move a flat plane of one unit area, separated by a liquid by one unit distance apart from a fixed plane, at one unit velocity; measured in centipoises (cP)
- kinematic viscosity, defined as absolute viscosity divided by its mass density; measured in centistokes (cS)

The measurement of viscosity is conducted with a Saybolt viscosimeter. This device comprises of an inner chamber containing sample liquid and an outer chamber containing standard liquid. A standard orifice is located at the bottom of the inner chamber. The measurement is done by letting the sample liquid fill a standard 60cm^3 container through the orifice. The time (in seconds) to complete the filling is recorded as Saybolt Universal Seconds (SUS), which is the official unit of viscosity.

The viscosity of a liquid depends on the temperature. As the temperature increases, the viscosity decreases. Liquid's viscosity change with respect to temperature change is measured as viscosity index (VI). Oil with a low VI is more sensitive to changes in temperature.

2.2.2 Hydraulic Liquid

Hydraulic liquid plays a very important role in a hydraulic drive since it serves as the medium to transmit the power from the power source to the intended load in a manner according to the application requirements. In the early history of hydraulic drives, water was used as the hydraulic liquid due to its wide availability. However, with the ever strengthening requirements for higher pressure following the industrial revolution, the constraints with using water and problems relating to oxidation and corrosion can no longer be tolerated. Since the early 20th century, oil has replaced water as the transmission medium in hydraulic drives.

Hydraulic drives are designed to operate with a hydraulic liquid having a specified range of viscosity. When a drive is operated with a liquid which has viscosity above the tolerance band, the following problems may occur:

- difficulty in starting-up
- stiff/sluggish operations
- cavitation in the pump
- accelerated wear of pump
- sticky valves and higher pressure drop
- higher temperature and power consumption

On the other hand, when a liquid with a viscosity below the tolerance band is used, the following problems may arise:

- internal and external leakage
- slow and irregular operations of the actuators
- reduction in lubrication

Apart from the hydraulic properties, the chemical properties of the hydraulic liquid are also important features governing the ruggedness of a hydraulic drive. While adequate hydraulic properties will ensure the efficiency and effectiveness of a hydraulic drive, adequate chemical properties will ensure that the components of the drive will not be unduly damaged or corroded over prolonged operations. Typical chemical properties to be observed include the degree of oxidation, water content, and degree of contamination.

Over prolonged time, these properties may change. Various tests are available to check the condition of the hydraulic liquid, *e.g.* the *Fluid Analysis Test* and the *Particle Analysis Test*. These tests will help to determine if the key properties of the hydraulic liquid have deteriorated to a level which will necessitate a replacement.

2.2.3 Benefits of Fluidic Drives

Fluidic drives have remained an option for motion generation and power generation, despite the wide availability of electric drives. This is even more so in the case of hydraulic drives due to their ability to deliver higher power.

Advantages of the fluidic drives are as follows:

- higher power-to-weight ratio, compared to a servo electric drive
- fluid acting simultaneously as a lubricant and coolant, apart from a power transmitter; a distinct feature of fluidic drives
- flexible in nature, *i.e.* ability to operate under continuous or noncontinuous conditions, at variable (and reversible) speed, with step-less variations

- ability to be stalled (*e.g.* in the case of overloading) without affecting the whole system, with simple circuitry
- both linear and rotational actuators available
- flexibility for interconnection of various fluidic components
- low maintenance cost

Disadvantages of the hydraulic drives are as follows:

- not readily available, compared to electric power
- costly and complicated installation, compared to mechanical or electrical system

Even when fluidic drives have been selected as the driving mechanisms, there are considerations in choosing whether to use hydraulics or pneumatics based on their characteristics. For example, fluidic actuators incur a shorter response time; an important consideration in applications where time precision is critical. In terms of safety, hydraulic liquid poses fire, explosion, contamination, and leakage hazards; rendering requirements of continuous maintenance of hydraulic liquid. Due to their incompressibility, hydraulic drive has poor damping characteristics. Also, hydraulic drive is sensitive to leakage, filtration, and contamination.

2.3 Components of Fluidic Drives Systems

The components of a hydraulic drive can be represented by symbols. These symbols are commonly used in circuit diagram, which represents the connections of components in the system. Commonly used symbols of fluidic components are presented in Figure 2.3.

2.3.1 Primary Power Source

The primary power source normally consists of a prime mover, *e.g.* an electric motor, serving as the energy source of the fluidic drive. This component typically drives hydraulic pumps in the hydraulic drives or pneumatic compressors in the pneumatic drives.

2.3.2 Hydraulic Pump

The hydraulic pump converts mechanical energy into hydraulic energy, which can then be transmitted by the hydraulic liquid to the actuators (hydraulic motors or pistons). In practice, the pump also generates pressure due to hydraulic resistance of other components in the system, thereby maintaining the flow. The input energy source is usually an electric motor.

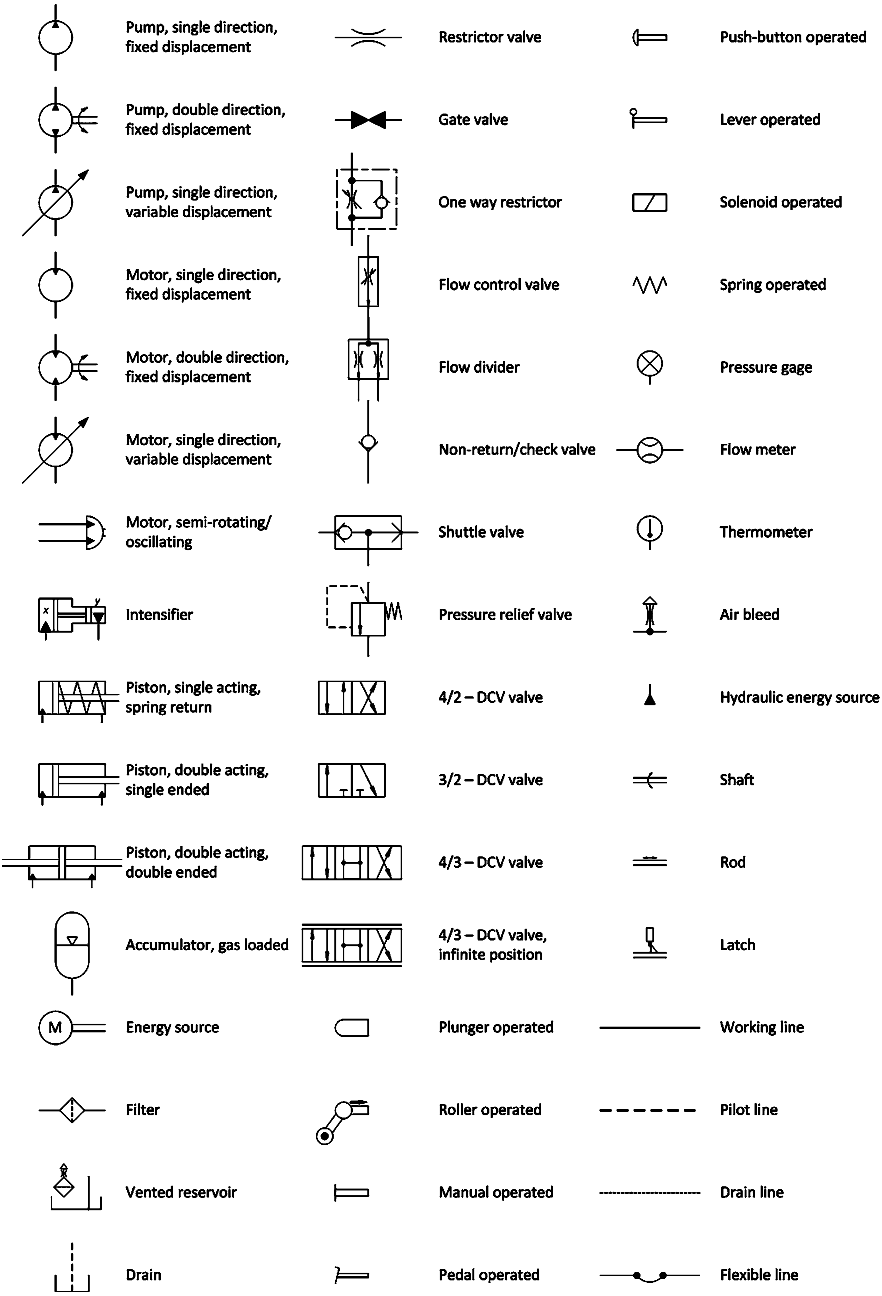

Fig. 2.3 Common symbols of fluidic components

Pumps can be classified based on the variability of the internal volume. Fixed pump refers to a pump with a fixed internal volume such that it always delivers the same flow rate of hydraulic liquid. Variable pump, on the contrary, has a variable internal volume, and hence it allows a variable flow rate of the hydraulic liquid.

Pumps can also be classified into positive displacement and nonpositive displacement pumps. Positive displacement pumps operate by displacing an amount of liquid from the low-pressure chamber (suction side) to the high-pressure chamber (delivery side) using its respective impeller, *i.e.* vane for vane pump, tooth for gear pump, *etc.* The flow is generated by varying the physical size of the pumping chamber, which has a larger volume for lower-pressure chamber and a smaller volume for higher-pressure chamber, so that the liquid is expelled at the high-pressure chamber accordingly. Inlet and outlet valves are installed at the inlet and outlet port, respectively, which is necessary to prevent a backward flow. Furthermore, since the operational pressure can be very high, sealing is important to prevent leakage. Valves and sealing increase the complexity of the construction of a positive displacement pump, and as such these parts have become the common sources of pump faults.

Nonpositive displacement pumps are generally used for generating large flow volume at low pressure. Examples of nonpositive displacement pumps are the centrifugal pump and the axial propeller pump.

Pumps used in hydraulic drives are positive displacement pumps, due to the requirement to generate and maintain a large pressure (typically up to about 13MPa) and high power at a specific certain flow rate.

2.3.2.1 Types of Positive Displacement Pumps

Examples of positive displacement pumps include gear pump, gerotor pump, screw pump, vane pump, and piston pump. In the selection of a pump for use in the hydraulic drive, several factors should be considered, including the load characteristics (*e.g.* the flow rate), the pump characteristics (*e.g.* the operational pressure), speed of the power source, cost, reliability, maintenance, noise generated, and the environmental condition.

1. Gear pump

   In an external gear pump, the pumping action is accomplished by a pair of gears meshed together on their external sides, where one gear is keyed to the motor shaft. As the electric motor drives the pump, the liquid enters the pump through the inlet port, and the teeth of the pump cause the liquid to circulate through the outer diameter of the gears. The liquid is then displaced to the higher-pressure chamber and leaves the pump through the outlet port. Teeth mesh prevents the liquid from bypassing the pump through the inner diameter. Figure 2.4 presents the construction and working principle of external gear pump.

   In an internal gear pump, the pair of gears is meshed on their internal sides instead, where the inside gear is keyed to the motor shaft. The centres of the two gears are then not aligned, although they are parallel. This creates a variation of physical volume of the pumping chamber, allowing positive displacement action

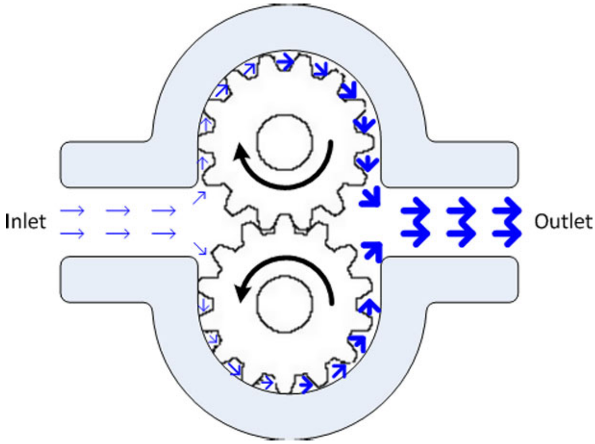

Fig. 2.4 External gear pump (*thin arrow* for low pressure, *thick arrow* for high pressure)

to take place. As the electric motor drives the pump, the liquid enters the pump and circulates along the pumping chamber. The displacement of the liquid from the larger chamber to the smaller chamber creates a liquid flow which effectively increases the pressure of the liquid.

Similar to other gear construction, gear pump, which typically uses spur gear, is not balanced. The lateral force created by the mesh of the gear creates a side force on the gears and shaft, thus reducing the pressure and speed at which the pump can be operated.

2. Gerotor pump

   The construction of a gerotor pump is rather similar to the internal gear pump, except that it has several, instead of only one, pumping chambers. The number of the internal teeth is one less than that of the external teeth and equal to the number of pumping chambers. The idea of the design of gerotor pump is to combine the characteristics of external and internal gear pumps. The profile of the teeth is designed inwards, so they can also act as the seal.

3. Screw pump

   The pumping action is accomplished by a screw, which displaces the liquid axially through the pumping chamber. The helical profile of the screw causes the liquid to be displaced without generation of a lateral force, although axial force is present. This is one desirable feature of screw pump, apart from its noiseless operations.

4. Vane pump

   A vane pump comprises of vanes, rotor, and cam ring. The pumping action is accomplished by the vanes, which are installed in an eccentric rotor. The rotation of the rotor, driven by a primary power source, causes the vanes to slide radially, and thereby varying the volume of the pumping chamber. When the volume increases, the liquid enters the pumping chamber due to suction effect. When the volume decreases, the liquid inside the chamber is forced to a high pressure and ejected out of the pump.

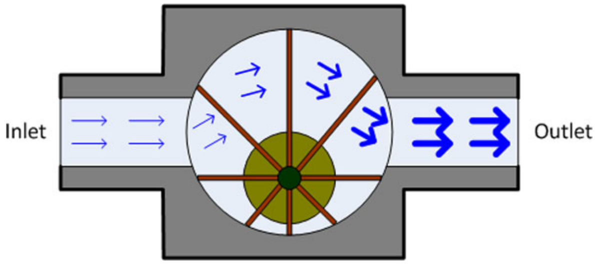

Fig. 2.5 Vane pump (*thin arrow* for low pressure, *thick arrow* for high pressure)

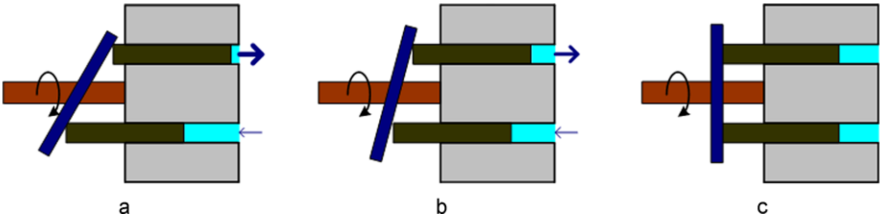

Fig. 2.6 Piston pump

Due to its eccentric design, the performance of the pump can be significantly affected. To reduce this effect, a balancing mechanism is usually implemented to the pump. This includes the inlet and outlet port arrangement, intra-vane, and dual vanes.

A variable displacement vane pump employs an unbalanced design. This is achieved by adjusting the eccentricity between the cam ring and the rotor.

Figure 2.5 presents the construction and working principle of vane pump.

5. Piston pump

In a radial piston pump, the pistons are located around the pump shaft at right angles, *i.e.* along the pump's radius. Each piston is seated on a roller or a sliding shoe, while each roller is located on a common cam. The rotation of the cam, along with the pump shaft, will then reciprocate the pistons movement inside the bore cylinder. When the piston moves toward the centre (suction), the liquid from the low pressure chamber will be drawn into the cylinder. When the piston moves away from the centre (discharge), the liquid from the cylinder will be discharged into the high-pressure chamber. Figure 2.6 presents the construction and working principle of radial piston pump. As the shaft of the pump turns, the liquid flows into the pump through the low-pressure inlet. The inclined plate pushes the piston out, pressurising the liquid out of the pump through the high-pressure outlet. The amount of outlet pressure is adjustable via the angle of the inclined plate. Figure 2.6(a) provides maximum pressure with maximum angle (denoted with thick arrow), which can be reduced by flatting the angle as presented by Figure 2.6(b), with complete flat plate results in zero pressure build-up as in Figure 2.6(c).

In an axial piston pump, the pistons are fixed in parallel around the pump, each of them seated on a shoe that is attached to a common swash plate. The swash plate is installed at an inclined angle with respect to the shaft. With the rotation of the shaft, the swash plate will reciprocate the movement of the pistons. The reciprocation of the pistons inside the cylinder will alternately draw and discharge the liquid, creating a pumping action.

2.3.2.2 Boost Pump

In a large hydraulic drive, an additional pump may be installed to assist the main pump. This additional pump, which is called boost pump, is installed before the main pump (between reservoir and main pump) with the following responsibilities:

- to fill the circuit with oil in initial operation
- to replace oil lost due to internal leakage during operation

Boost pump is usually a small, fixed displacement pump, driven directly from the main pump. In many cases, the boost pump is housed within the same installation as the main pump.

2.3.2.3 Performance Indicators of the Pump

A key indicator of the performance of the hydraulic pump is measured by its efficiency, which is defined as follows:

$$\eta = \frac{P_{\text{fluid}}}{P_{\text{input}}}, \tag{2.8}$$

where P_{fluid} represents the output fluidic power, and P_{input} represents the input power to the pump, which is usually electrical power.

The flow rate of the pump, defined as the volume of the pumped liquid per unit time, is also an important indicator of the pump performance. Possible factors which may lower the pump efficiency include:

- mechanical friction within the pump
- internal leakage of the liquid in the pump
- friction between the liquid and the pump components

A pump performance can deteriorate over time. Improper operating conditions in a hydraulic drive may result in pump cavitation. Pump cavitation occurs when the local pressure of the pump falls below the vapor pressure of the liquid, causing the liquid to start vaporizing and creating bubbles. Cavitation can be caused by:

- inadequate suction pressure
- poorly designed inlet port
- incorrect pump selection

The effects of cavitation are far reaching. Bubbles which are formed in the low-pressure region can move to the high-pressure region, disintegrate in this region, yielding a counter force in this region as a consequence, which can lead to severe erosion of the pump.

2.3.3 Hydraulic Motor

The hydraulic motor will convert the hydraulic energy from the hydraulic liquid back into mechanical energy in the form of shaft rotation, hence producing torque and speed. Various types of hydraulic motor are available, such as radial piston, axial piston, gear, and vane motor. Depending on the design, the hydraulic motor has an ability to rotate in both the clockwise and counterclockwise directions.

The operation principle of hydraulic motor is exactly the reverse of a positive displacement pump. Hydraulic liquid enters the high-pressure chamber through the inlet port and pushes the energy converter component (which can be vane, piston, *etc.*, depending on the type of the pump), which in turn rotates the motor shaft. The hydraulic liquid will then leave the motor from the outlet port with reduced pressure.

Like hydraulic pumps, fixed and variable motors are available, allowing adjustment of speed and torque according to the load condition.

The construction of a hydraulic motor determines the performance of the motor. Table 2.2 shows the different motor designs, their characteristics and benefits.

2.3.3.1 Types of Hydraulic Motors

Common types of hydraulic motors include the gear motor, the gerotor motor, the vane motor, and the piston motor.

1. Gear motor

   The construction of gear motor is very similar to a gear pump. A pair of gears is meshed onto each other, with one gear keyed to the output shaft. High-pressure liquid enters the motor and circulates around the outside of the gear teeth, since the inside is blocked by the mesh of the gear teeth. The circulation of the liquid from the high-pressure port to the low-pressure port around the outside diameter of both gears rotates the gear and, in turn, the shaft keyed to it. The favourable features of a gear pump is its high efficiency of up to 90%, speed range of up to 20,000rpm, power of up to 4kW; all these accomplished within a small size.

   Figure 2.7 presents the construction and working principle of gear motor.
2. Gerotor motor

   Gerotor motor operates in the exactly reversed way of a gerotor pump. The liquid enters the motor and moves into the chamber between the inner and outer gears. The high pressure of the liquid impels the gear and causes the motor to rotate, while at the same time displaces the liquid to a low-pressure chamber.

Table 2.2 Motor designs, characteristics, and benefits

Design	Characteristics	Benefits
Cam design	Multiple stroke, large displacement	High torque per weight (inertia) ratio
Even number of pistons	Main bearing unloaded from radial piston forces	Long service life of bearing
Optimal cam geometry	Constant torque, no ripple, no pressure pulsation	High torque, no contamination entering housing
Cam roller, solid, balanced	Reduced motor diameter	Long service life
Guide plate	Pistons unloaded from side forces, no stick slip	Limited piston bore wear
Side guide roller bearing	No stick slip	High mechanical efficiency
Oil distributing plate	Low volumetric loss	High volumetric efficiency, good low-speed performance, low noise level
Piston ring	Low volumetric loss	Improve volumetric efficiency
Stationary motor housing with torque arm	Compact design	Motor bed plate eliminated
Rotating cylinder block with hollow shaft	Compact design, no key ways	Ease in mounting
Through-hole in centre shaft	Air venting	Simplify mounting, possibility for cable through the motor

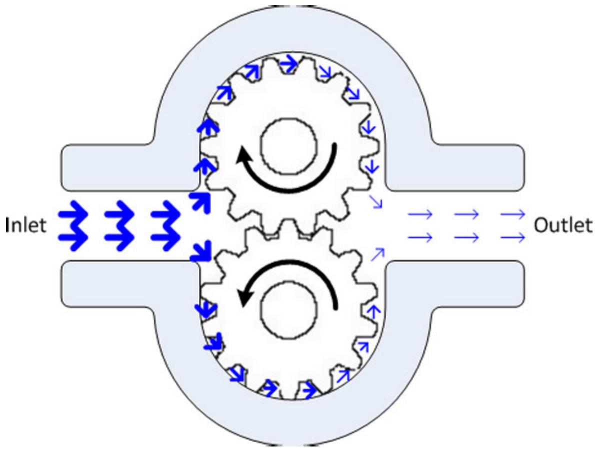

Fig. 2.7 Gear motor (*thin arrow* for low pressure, *thick arrow* for high pressure)

Contacts between teeth of inner and outer gears provide the seal between adjacent chambers. The displacement of the motor is determined by the space of the chamber.

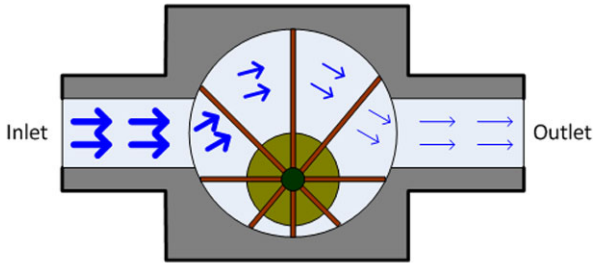

Fig. 2.8 Vane motor (*thin arrow* for low pressure, *thick arrow* for high pressure)

3. Vane motor

   The difference between the construction of a vane pump and a vane motor is the presence of a spring in vane motor to maintain the contact between the vanes and the cam ring. The liquid that enters the motor provides a high pressure and hence exerts a high force, against the vane. This causes rotation of the shaft, which is attached to the vanes. Figure 2.8 presents the construction and working principle of vane motor.
4. Piston motor

   In a radial piston motor, the pistons are arrayed in a radial construction, each of which is pressed against a cam roller. All cam rollers push against a cam ring that is keyed on the motor shaft. Piston ring is installed on every piston to prevent leakage. When the liquid, at high pressure, enters the piston bore, force is generated in the contact between the cam roller and cam ring along the tangential direction with respect to the shaft. This force produces torque and therefore rotates the shaft. While the role of the piston is to convert hydraulic energy into mechanical energy and the cam ring is to produce rotation, the role of the cam roller is to help absorbing the lateral force reaction against the piston, thereby increasing its efficiency. A torque arm is usually installed to anchor the motor within a static frame in order to prevent undesirable motor body rotation.

   Radial piston motor is usually used in heavy duty applications, *e.g.* crane hoisting, rolling mills, and railroad transportation. This is mainly because of its capability to sustain high radial and axial loads with short housing and length. Figure 2.9 presents the construction and working principle of piston motor, which is essentially the reverse of piston pump. As the high-pressure liquid flows into the motor, it pushes the piston, which in turns hits the inclined plate; thereby rotating the shaft. The liquid then flows out of the motor through low-pressure outlet. The output torque of the motor is adjustable via the angle of the inclined plate; high torque in Figure 2.9(a) with steep angle, moderate torque in Figure 2.9(b) with less angle, and no-torque in Figure 2.9(c) with flat angle.

   In an axial piston motor, as the name suggests, pistons are arrayed axially, each of which is pressed against a roller. All rollers push against a swash plate, instead of a cam ring, which is installed with an inclination angle with respect to the shaft. When the high-pressure liquid enters the piston bore, a tangential force is generated in the contact between the roller and the swash plate, which

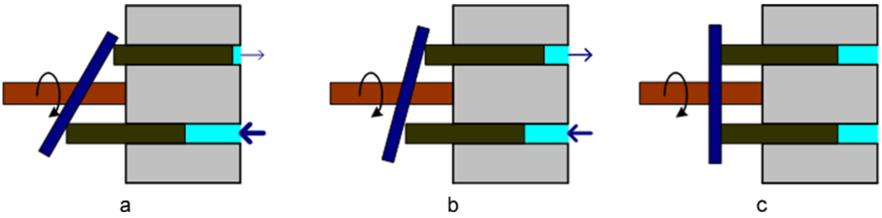

Fig. 2.9 Piston motor

will produce a torque and a shaft rotation. Adjustment of swash plate inclination angle allows a variation of the rotation speed; thus this kind of motor is also known as a variable hydraulic motor.

In a large hydraulic motor, which is usually required to handle a high load, double banks of pistons and their subsequent components are installed to yield a larger torque. The rotation, however, remains unchanged.

The distribution of liquid to the right piston is important to ensure proper operation of the motor. This is accomplished via valve plates installed in the inlet port, equipped with port timing. Also, in a double direction motor, a right-left connection is provided in the inlet port to allow the motor to operate in clockwise and counterclockwise directions.

Piston motor can also be used as a pump if the mechanical energy is applied to the output shaft; this device is commonly called the pump-motor. This is a flexibility the piston motor construction has to offer.

2.3.3.2 Performance Indicators of Hydraulic Motors

The performance of a hydraulic motor can be measured by indicators in terms of its torque, speed, power, and efficiency. Efficiency is a measure of how much fluid power can be converted into usable mechanical power of the shaft, as follows:

$$\eta = \frac{P_{\text{output}}}{P_{\text{fluid}}}, \tag{2.9}$$

where P_{output} represents shaft output power, and P_{fluid} represents fluid input power.

2.3.4 Hydraulic Piston/Cylinder

Hydraulic piston converts hydraulic energy back into mechanical energy in the form of linear motion and force. It is basically a piston contained in a cylinder, which can have one or two protruding rods out of it. While hydraulic piston is mainly for loading purpose, it can also be applied as a cushioning device as in a suspension system. The selection of a hydraulic piston is based on considerations such as the purpose of application, construction, technical requirements (force, duty cycle, action), and environmental conditions.

2.3.4.1 Loading Piston

The force exerted by a hydraulic piston is proportional to the liquid pressure and the cylinder's cross section area. Therefore, a piston with rods at both ends exerts the same force in both directions, while a piston with one rod exerts a larger force in one direction due to the difference in area.

Based on the activation approach, a hydraulic piston can be classified as either a single acting or a double acting piston.

A single acting hydraulic piston is activated by the hydraulic liquid from one side only. On the filling side, a channel is installed to allow oil to flow into the cylinder. When the liquid is directed to flow into the cylinder, it will push the piston by virtue of its pressure, and the load will be moved accordingly. The returning action is accomplished by external force, such as spring or load gravity. Alternatively, a drain valve can be installed to empty the liquid and thereby returns the load.

A double acting hydraulic piston has channels on both sides of the cylinder. These channels allow the liquid to flow into or out of the cylinder to generate motion. Depending on the direction to which the liquid is flowing, the piston will move either in forward or backward direction. A double acting piston can have one or two rods protruding out of the cylinder.

2.3.4.2 Cushioning Cylinder

A variation in the application of hydraulic piston is cushioning cylinder. As its name suggests, it prevents or reduces shock to the load by decelerating a moving load. A typical application of cushioning cylinder is a shock absorber.

The working principle of cushioning device is to slow down the piston movement before it contacts the end of the cylinder. It uses a taper or stepped rod that enters a sleeve mounted in one end of the cylinder. Through this design, the pressure applied to the piston increases, thereby reducing its velocity.

2.3.4.3 Sealing

Sealing is a very important issue to address in a hydraulic drive, especially in a hydraulic piston, because of its movable component (the rod) that extends from the pressure chamber. Since the rod is movable, there is a gap between the rod and the cylinder to allow motion. However, the gap must be as small as possible to prevent leakage. Another common problem relating to sealing lies in the seal itself, when it expands due to the repetitive movement of the rod.

2.3.4.4 Performance Indicators of Hydraulic Cylinder

The performance of a hydraulic cylinder can be measured by its capability to generate force, which is dependent on the pressure of the liquid and the pressure area of the cylinder as follows:

$$F = pA. \tag{2.10}$$

The velocity v at which the load is moved is also an important feature and formulated as follows:

$$v = \frac{f}{A}. \tag{2.11}$$

The stroke of the cylinder determines the total movement distance, and it is also an indicator of the performance of a cylinder.

2.3.5 Control Valves

Control valve is a device which is used to hydraulically regulate certain parameters, such as pressure and flow. Based on the nature of control, control valves can be classified into servo valves and proportional valves; the former referring to closed-loop control valves, while the latter referring to open-loop control valves. Based on their functions, they can be classified into a directional control valve (DCV), a flow control valve, or a pressure control valve.

2.3.5.1 Evolution of Control Valves

The first electrically operated control valve, which will later bring on the servo valve and proportional valve, was developed after World War II. This valve used a linear spool slide unit controlled by a linear motor, and it was used in aeronautics. Hydraulic pilot control was later introduced to the valve assembly via a jet flapper and jet pipe amplifier. A few years after this invention, mechanical feedback of the spool position was incorporated, yielding the servo valve. The servo valve has gained popularity mainly because of the precision and the power balancing features which it enables. From the early 1960s, there was an increasing application of servo valve in the industries; a noteworthy example was in injection molding applications.

However, that servo valves had significant drawbacks: very sensitive to contamination and costly. These issues became even more significant when a large volumetric flow and rapid control are required. These issues led to the revolutionary design and the conception of the proportional valve.

The control of proportional valve was done electrically by either direct control of the spool (with magnet or motor) or by hydraulic pilot control. The proportional valve soon gained popularity and wide acceptance because of its large flow rate and its robustness to contamination. Over the time, the proportional valve continued to evolve with much effort generated towards the simplification of its control equipment. It became possible to have a robust proportional control valve under servo control. This control valve is sometimes called the servo proportional valve.

2.3.5.2 Fundamentals of Control Valves

The key component of a control valve is a linear spool slide, whose role is to block or unblock the flow of liquid with a sliding motion. In a more advanced design, it can also regulate the flow of the liquid. In general, this component is termed *throttle*. One full sliding motion is called one stroke, which ranges from 0.4 to 10.0mm. The ratio of spool diameter to the stroke is from 8:1 to 15:1.

As a control device, the control valve has to achieve high precision in its operations and minimize the amount of leakage occurring. The effects of leakage are:

- reduced positioning accuracy
- reduced power delivery and hence efficiency
- introduction of higher lateral and frictional force

Therefore, a control valve requires high accuracy in its manufacturing. The allowable clearance between the spool and the housing is typically up to 1:2000 or equal to a nominal separation of not more than 2μm. The material used for the valve is also special (*e.g.* hardened spheroidal iron) to ensure that it has a high wear resistance. The roughness of the spool and cylinder bore is typically up to 0.4μm. O-rings are used to further prevent leakage in the control valve and are usually installed on the connection holes.

The connection holes of control valve are assigned with special terminology. They are usually termed P for pressure supply (pump), T for drain (tank), A, B, and so on for actuators, and X-outlet and Y-inlet for pilot control. The size and position of the connection holes are also mostly standard to promote interoperability of equipment from different manufacturers.

The pressure drop across the valve is an important factor to consider in the design and selection of control valves. Pressure drop induces thermal rise (typically 5°C per 100bar) which in turn reduces efficiency and power delivery. Typically, 75% of the pressure losses occur at the variable throttle points near the maximum stroke of the valve. Pressure drops are caused by:

- friction between the hydraulic liquid and the valve wall
- redirection of the flow inside the valve

Pressure drop is a nonlinear phenomenon (mainly saturation), and this poses a big challenge in control design, especially when high precision is a significant factor. A typical remedy of this problem is to limit the operations of the valve in the nonsaturated operation range by means of a mechanical structure.

Another cause of power reduction is due to the flow resistance force, which is in turn due to the longitudinal flow of the liquid across the valve. Since the flow induces pressure, a force will be generated across the cross-sectional area of the spool, resisting the flow.

The control of the valve's spool is accomplished electrically. Several mechanisms can be used, such as plunger coil, linear motor, torque motor, or proportional magnet. This mechanism converts electrical signal into electromagnetic force that is able to put the spool in motion. Hydraulic pilot control, however, can also be used to amplify and transmit the electromagnetic force into spool motion.

The operational principle of the mechanisms, which is based on the electromagnetic principle, can be briefly described as follows. When a current is applied through the armature which is contained in a magnetic field, a force is generated. Depending to the mechanical structure, this force may be transmitted directly or indirectly to the spool, resulting in a motion. The control is accomplished by adjustment of the current's amplitude and interval. The development of electromagnetic converters contributes to the advance in valve performance. Examples of these contributions are:

- new materials for magnet, *e.g.* samarium–cobalt and neodymium–iron–boron, which have larger specific energy
- low-power plunger, with smaller size and elimination of magnetic hysteresis
- combination with hydraulic amplifier to provide larger torque/force to the spool
- combination of magnetic and nonmagnetic materials to direct and to concentrate magnetic flux within the air gap

The spool configuration affects the flow characteristics significantly. Spool can be configured to have an open-centre or a close-centre configuration. An open-centre spool has both control ports open to return when the spool is centred, and it is typically used with fixed pumps. A close-centre spool, on the other hand, is typically used with variable pumps to form a constant pressure hydraulic circuit.

In the presence of large flow resistance force which cannot be overcome satisfactorily by electromagnetic force, a hydraulic amplifier can be applied. Hydraulic amplifier generates a hydraulic force that is typically up to 10 times larger than flow resistance force. Hydraulic amplifier is included in a hydraulic pilot control installation together with the electromagnetic converter to form an integrated pilot control. An example of hydraulic amplifier is a four-edge amplifier, which corresponds to the main stage of the flow control valve.

2.3.5.3 Proportional Control Valve

Proportional control valve can be controlled continuously during its operation via an electrical signal, which is activated through manual controllers (usually located in the control station), as it is controlled in an open-loop manner. This valve can direct and meter the flow of the liquid to actuators.

A proportional control valve comprises of a spool as the main regulating component, housed inside a valve bore. Valve bore forms valve cavity, connected to several channels linked to actuators, pump, and reservoir (tank). The control of the valve is executed by controlling the spool electrically with various electromagnetic converters.

Despite the open-loop control configuration, a feedback signal is provided from the spool to balance the action of the electromagnetic converter so that it can maintain the valve position according to the signal from the manual controller. Hydraulic amplifier may be installed to assist controlling of proportional control valve.

There are several types of common proportional control valves: jet-flapper, sliding spool pilot stage, jet-pipe, and jet-deflector valve. These are elaborated as follows.

1. Jet-flapper valve

   In a jet-flapper amplifier, two fixed and two variable throttles are arranged in a Wheatstone full-bridge; the fixed throttles are located at the upstream, and the variable throttles are located at the downstream. The flapper is arranged between the variable throttles, with its normal position in the centre. Two jets are directed towards the variable throttles. A flow channel, which houses the spool to be controlled, is connected to both fixed and variable throttles. The hydraulic liquid flow generates pressure drop at the throttle points.

   If the flapper is located at the centre, equal pressure drop occurs so that no force is exerted on the control spool. If the flapper is displaced, for example, towards the left jet, the cross section of the left outlet reduces while the right outlet increases. The pressure in the left branch will then increase, while that of the right branch will decrease. This pressure difference will exert a force on the spool, pushing it to the right. If the flapper is displaced to the opposite direction, an opposite effect will occur.

   The movement of the flapper is controlled by an electromagnetic converter, *e.g.* a torque motor. The armature of the converter is designed such that it has a fast dynamic response bandwidth and bidirectional control ability. A returning spring may be added to provide an even faster returning dynamic.

   When the spool is in motion, a balancing flow flows from one branch of the bridge to the other branch. If the flapper happens to close a jet completely, the flow of this branch can be diverted and used for the positioning velocity. An additional throttle can be installed between the variable throttles and the drain line to stabilize incoming flow from the jet. The balancing flow is proportional to the flapper position.

   If the actuators are blocked, no balancing flow will occur; a differential pressure will build up, instead, leading to leakage and power loss.

   The jet-flapper valve is mainly used in a high performance hydraulic system.

2. Sliding spool pilot stage valve

   This is a variation of a jet-flapper valve, except that the flow is across two meter-in and two meter-out control edges, connected to the controlling spool. The control of the valve is performed by the opening and closing of the meter-in and meter-out control edges.

   This valve is commonly used in a modest performance hydraulic system.

3. Jet-pipe valve

   In a jet-pipe amplifier, a jet-pipe with hydraulic liquid is integrated into the control element. A fluid jet flows from the pipe through a jet against two receiving holes, which are located close to each other.

   A connecting pipe from each of the hole leads to both ends of the controlling spool, housed in a slide unit. In the neutral position, the jet-pipe is situated at the centre, and the liquid is therefore distributed equally to both holes, creating equal pressure on the right and left of the spool. If the jet-pipe is displaced towards, for example, the left hole, more flow will be directed to this hole, creating a pressure difference on the spool and in turn moves to spool to the right. The opposite effect will occur when the jet-pipe moves towards the opposite direction. The velocity of the movement can be regulated by the displacement of the jet-pipe.

Due to its construction, leakage will occur in this valve, mainly at the cross section area of the holes, inducing a power loss.

4. Jet-deflector valve

   This is a variation of jet-pipe valve, whereby its jet is installed and fixed opposite to the two receiving holes. The control is provided by a thin jet-deflector plate, located between the jet and the holes, perpendicular to the jet flow. The jet-deflector plate is attached to the electromagnetic converter to perform the controlling action.

   In the neutral position, the jet-deflector plate is situated at the centre, and the liquid is therefore distributed equally to both holes, creating equal pressure on the right and left of the spool. Displacement of the jet-deflector plate towards the left will direct more flow to the left hole, creating a pressure difference on the spool and in turn moves to spool to the right. The opposite effect will occur when the jet-deflector plate moves along the opposite direction.

In short, the advantages of using a proportional control valve are as follows:

- accurate control of flow and direction
- installation of control valve can be near the actuators to simplify piping and plumbing
- reduction of fluid leakage
- several options of control available, *e.g.* with radio control

The disadvantages of using proportional control valve are as follows:

- less precise positioning as it uses open-loop control
- larger installation envelope

Proportional control valve is designed for a small pressure drop and therefore is suitable for a power delivery application. Typically, the pressure drop is up to 10% of the rated pressure, or equal to a nominal pressure drop of 10bar, with a flow dead zone of up to 30% of the spool stroke. Therefore, nonlinearity is a significant issue in the control of proportional control valve, especially when this valve is modified to operate in a closed-loop manner.

Closed-loop control can be implemented in a proportional control valve by replacing the manual controller with an electrically activated device. This is applied, for example, in a hydrostatic transmission, where manual controller is replaced by a potentiometer.

Proportional control valves are now commonly applied in the industries associated with material handling, construction equipment, and steel mill.

2.3.5.4 Servo Control Valve

A servo control valve produces continuous controlled output via feedback control loop, as it implements a closed-loop control system. The reference signal is compared to the actual output and is used to generate control signal, resulting in a precise positioning of the valve according to the reference signal. The positioning of the valve is accomplished by pilot flow, while the main flow of the liquid is carried out by the main ports. To increase its flow rate, the servo valve may be staged, which will provide a higher pressure to shift the spool. Because of its more complex control system, servo control valve is more expensive than proportional control valve.

The pressure drop of servo control valve can be large, reaching up to 30% of the rated pressure or equal to a nominal pressure of 70bar. The flow dead zone, however, is rather negligible. It is therefore very suitable for a precision metering application, where the effects of nonlinearity must be low.

Servo control valve can be used to measure various variables, such as position, velocity, pressure, or torque. The purpose of the valve determines the sensors to be used and also the controller structure. It is often required for a hydraulic drive, via its servo control valve, to satisfy the requirements relating to a number of variables simultaneously. These requirements will further increase the number of sensors needed, and complicate the controller structure.

With the advance in closed-loop control technology, many options are now available to simplify the structure of the valve's controller without sacrificing the performance. State observers, for example, can be implemented to reduce the number of sensors in an operation. The idea is to obtain a limited number of sensor signals of certain variables and then to estimate the other variables via computation.

2.3.5.5 Classification of Control Valves Based on Functions

Control valves can also be classified according to the function it performs in the overall system.

1. Directional control valve (DCV)

   As the name suggests, DCV serves as the direction controller of the liquid flow to the actuators, which can be motors or pistons. It can therefore control the direction of the actuators—whether the motor rotates in clockwise or counter-clockwise direction, or whether the piston extends or retracts but cannot control the speed or velocity of actuators.

   In the context of DCV, the following terms are commonly used:

   - *PORT*

     This refers to the number of connection of a DCV to external devices. A DCV must have at least two ports; one to a pressure source (pump) and another one to an actuator. For a DCV with more than two ports, the ports may be connected to a reservoir from the drain or to other actuators.
   - *POSITION*

     This refers to the number of its possible position. It may theoretically be infinite if the sliding of the DCV is not restricted by its housing.
   - *WAY*

     This refers to the number of flow path through the valve.

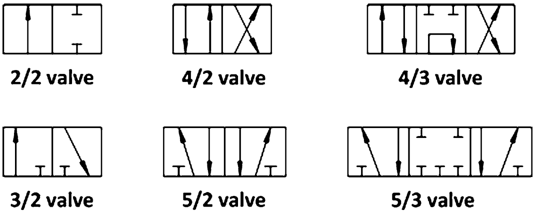

Fig. 2.10 Examples of how DCV is named

Table 2.3 Comparison of hydraulic systems with an open-centre and a close-centre valve

Features	Open-centre valve	Close-centre valve
System idle time	Long	Short
Typical power source	Fixed pump	Variable pump
Time characteristic of applications	Time insignificant	Time critical

A DCV with m ports and n positions is called an m/n DCV, as exemplified in Figure 2.10.

Directing action can be accomplished through the sliding spool, ball, plate, *etc.* A DCV can be controlled in various manners; with respect to its control (automatically or manually), its mechanism (mechanically, electrically, or hydraulically), and its distance from the controller (remote or direct-acting). The manner through which a DCV is controlled is presented in the symbol of the DCV.

When a pilot control is used in a DCV, the pilot fluid can be isolated from the hydraulic circuit and therefore permitting the usage of other substance to control the DCV.

A combination of several DCVs with proximity switches can develop into a sequencing hydraulic circuit, which can run automatically to perform a periodic task. The normal position of a DCV is important to determine the nature of the system at rest, especially in an automation system. Considerations over normal position of the valve also include the configuration of open-centre and close-centre valve. An open-centre valve connects the pump and the reservoir at rest; while a close-centre valve does not. The comparison between hydraulic systems with open-centre and close-centre valve is presented in Table 2.3.

2. Flow control valve

A flow control valve is used to control the flow rate of the hydraulic liquid. By controlling the flow rate of the liquid, it can effectively control the speed or velocity of the actuator. It can therefore be used as a speed control device to complement a DCV in a hydraulic system.

The operational principle of a flow control valve is based on flow restriction which will alter the flow rate of the liquid. However, flow restriction may affect the temperature and pressure, which will require timely compensation to maintain the operating conditions. For modest applications, however, compensation may not be necessary. The flow rate of the liquid across a flow control valve depends on the pressure difference, type of valve, and the characteristics of the hydraulic liquid.

With pressure compensation, the flow rate can be kept constant regardless of pressure changes in the system, which may occur due to the flow restriction itself. Similarly, with temperature compensation, the valve maintains the flow rate despite temperature changes, which also changes the viscosity of the hydraulic liquid.

The restriction of the liquid flow is conducted by a series of fixed or variable orifices. A fully open orifice allows the flow of the liquid at its rated flow rate, while a restricted orifice reduces the flow rate. Pressure drop is thus inevitable across a flow control valve.

A type of flow control valve, called deceleration valve, is used to alter the speed or velocity of the actuators at an intermediate position. This valve is typically manually operated but may be equipped with a proximity switch as a position indicator. An adjustable needle valve may also be employed to further control the flow rate.

3. Pressure control valve

   A pressure control valve serves as the controller of the pressure of a hydraulic system, often to limit the pressure below which the system should operate. Pressure control valves include the compound pilot drained valve, pressure reducing valve, sequence valve, unloading valve, counterbalance valve, and the pilot-operated check valve.

   A compound pilot drained valve is operated via a pilot control, resulting in a better pressure sensitivity. The main components are piston, poppet, and spring, which work together to open or close the discharge port when necessary.

   A pressure reducing valve is employed to supply low-pressure liquid. This valve steps down the pressure by restricting the flow when the preset pressure is reached.

   A sequence valve controls the order of operation that has to be performed by several actuators or circuits. This valve is connected to sensors (typically proximity sensors) and is activated to the respective sequencing position by the respective sensors.

   An unloading valve is used to discharge the liquid from the pump when the pump is not in use. This action will reduce power loss more efficiently than using pressure relief valve. A pilot control is employed to control this valve.

   A counterbalance valve is used to maintain the backward pressure in a vertically mounted hydraulic piston/cylinder. The application of pilot control in this valve will reduce energy consumption.

   A pilot-operated check valve is used to position varying loads on a hydraulic piston/cylinder by locking the cylinder upon shifting of DCV.

2.3.5.6 Performance Indicators of Control Valves

The performance of a control valve is measured with various indicators which determine the suitability of the valve to the overall application. Many of the parameters, such as responsiveness, backlash, and hysteresis, are common in closed-loop control and will be described further in Chapter 5.

There are also indicators unique to hydraulic control valves, *e.g.* hydraulic overlap, which represents the linearity deviation of the liquid flow in the vicinity of zero crossing.

2.3.6 Sensors

Measurements are necessary in the closed-loop control system of a servo hydraulic drive to monitor the condition of the system, such as its pressure, flow rate, and temperature. Motion control of the actuators will also require other measurements of position, velocity, force, and torque.

2.3.6.1 Pressure

Pressure measurement is one of the most widely measured physical parameters. In fact, a vast number of technological breakthroughs would not have been possible without an accurate measurement of pressure. Two common pressure sensors widely used is the aneroid barometer and the bourdon tube.

1. Aneroid barometer

   An aneroid barometer, as shown in Figure 2.11, is used for the measurement of the atmospheric pressure and consists of an evacuated metal capsule with flexible top and bottom faces. The shape of the capsules changes with variations in atmospheric pressure, and this deformation is usually amplified mechanically via a series of levers or gears. The pressure capsule can be fabricated in the form of bellows to provide further deflection. The measurement of deflection is done visually by a pointer connected to the levers aligned to an appropriate scale. To increase the effective stroke, more than one capsule can be connected together. This configuration of multiple capsules will have the same characteristics as a much larger single capsule, which makes it possible to construct low-pressure gauges and still maintain a reasonably small size.
2. Bourdon tube

   A bourdon tube operates on the same principle as an aneroid barometer, except that it uses a C-shaped or helical tube instead of an evacuated capsule or bellows arrangement, as shown in Figure 2.12. The tubes are closed at the one end and connected to the point of measurement at the other end, which is fixed in position. The Bourdon tube is made from a section of oval tube that has been formed into a specific shape. When pressure is applied to the tube, it tends to straighten. This movement at the end of the tube is used to move a pointer over a calibrated dial. The deflection of the Bourdon tube varies with the ratio of its major-to-minor cross-sectional axes, the tube length, the difference between the internal and external pressure, and its total tube angle.

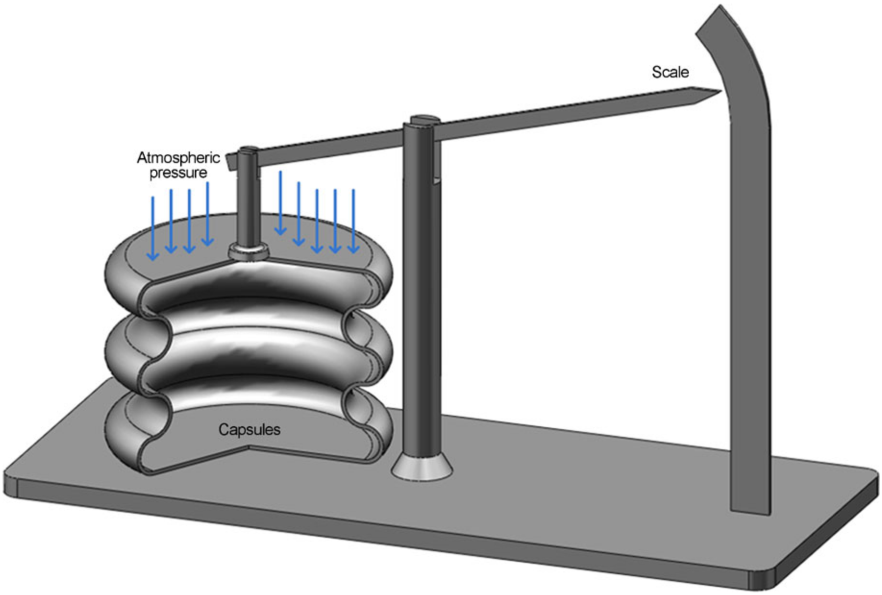

Fig. 2.11 Aneroid barometer sensing element

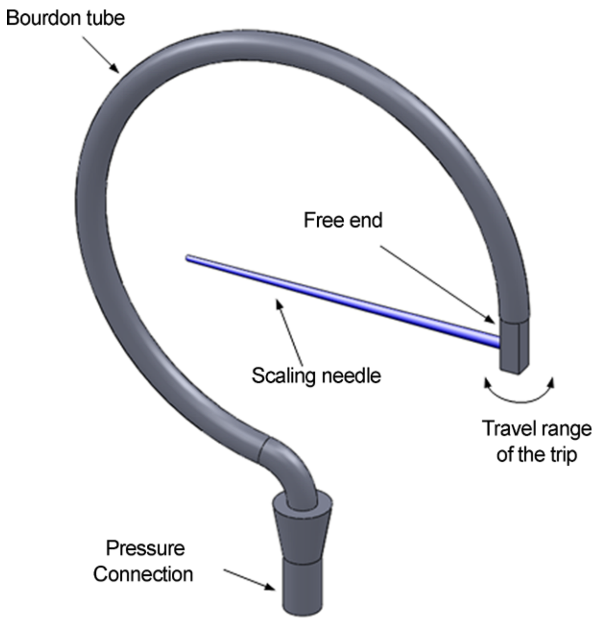

Fig. 2.12 Bourdon tube sensing element

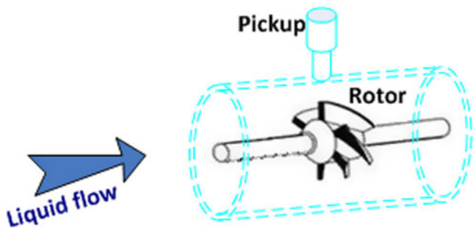

Fig. 2.13 Turbine meter

Bourdon tubes are usually used in the gauge pressure sensing application; however, sensing of differential pressure is possible by connecting two tubes to one pointer, thereby causing the pointer to measure the pressure difference between the two tubes. Helical tubes are more compact and reliable than the more traditional C-shaped devices. Bourdon tubes are used throughout the industry and are available in a wide range of pressure specifications.

2.3.6.2 Flow

Flow is often measured indirectly through the measurement of volume. Common flow sensors include the turbine meter and the paddle wheel meter.

1. Turbine meter

   Invented by Reinhard Woltman in the 18th century, a turbine meter is an accurate and reliable method used to measure flow rate for both liquids and gases. Figure 2.13 shows a typical design of a turbine meter. As fluid flows through it, the freely spinning turbine blade assembly would start to rotate. The pickup senses the movement of the blades and generates pulses to the meter. The rotating speed of the turbine is directly proportional to the flow rate of the fluid, thus the volumetric flow rate can be determined. Alternatively, a tachometer can be attached to the turbine to determine the number of revolutions of the turbine directly.

   The accuracy and sensitivity of the meter largely depend on the smoothness of the turbine rotations, which can be enhanced by reducing the friction between the blades and the shaft. Other factors affecting the accuracy include the nature of the fluid and the straightness of the fluid flow. There is a minimum flow rate needed for the turbine meter to function satisfactorily, and this is usually stated in the specification sheet. Turbine meters are expensive, but the measurements are highly accurate.

   Turbine flow meters are commonly used in the food and beverages, chemical, petroleum, and water distribution industries.
2. Paddle wheel meter

   A paddle wheel meter uses a rotor whose axis of rotation is parallel to the direction of flow as shown in Figure 2.14. In most cases, a paddle wheel meter is

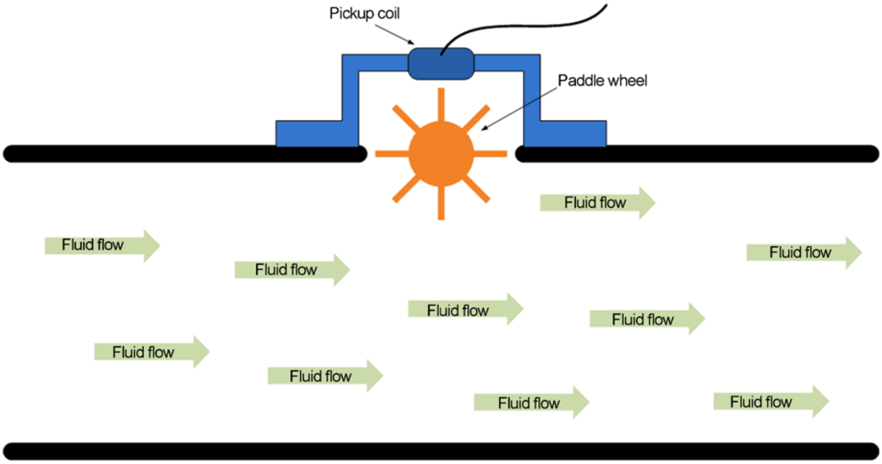

Fig. 2.14 Paddle wheel meter

bidirectional with flat-bladed rotors. However, some types of paddle wheel meter have crooked rotors, which restrict their usage to measure forward fluid flow.

When the fluid flows pass the paddle wheel, the paddle wheel will start to rotate along the direction of the flow. This will in turn activate the magnetic pickup coil and, as each paddle wheel passes through the magnetic field, a pulse is generated. The frequency of the generated pulse is proportional to the flow rate through the meter.

Paddle wheel flow meters are used in many applications including those found in the chemical industries, irrigation, air-conditioning, and refrigeration. Because paddle wheel flow meters are inexpensive and need little maintenance, they are used especially in applications where the cost to measure the variable is an important consideration.

2.3.6.3 Temperature

Temperature measurement and control are necessary in a hydraulic drive to maintain the viscosity of the hydraulic liquid. Two common temperature sensors in a hydraulic drive are the thermostat and the thermocouple.

1. Bimetallic thermostat

   A bimetallic thermostat is a device which can turn on or turn off a heating system to keep the hydraulic system at a range of desired temperature. The bimetallic strip is positioned so that when it is heated and bends, or when it is cooled and straightened, it will break or make an electric circuit. This effect is used to show if the temperature exceeds some threshold, or to control the heat produced by a heating system. The thermostat is thus only capable of on–off temperature measurement or control.

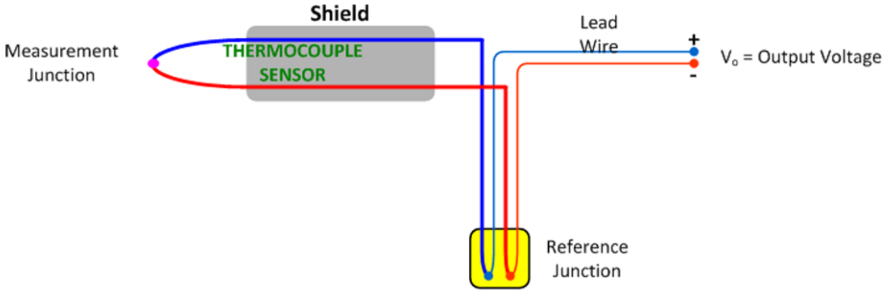

Fig. 2.15 Thermocouple

2. Thermocouple

   The thermocouple has two dissimilar metals joined together, forming a closed circuit. Figure 2.15 shows the construction of a thermocouple. One junction is held in a probe which is placed in the medium whose temperature is to be measured. The other junction is usually maintained at a fixed temperature. When the temperature at the two junctions differs, a resultant electromotive force (emf) is induced. Current will flow, and the direction and magnitude of the current flow depend on the temperature difference between the two junctions. This current can then be calibrated to give the required temperature measurement.

2.3.7 Auxiliary Equipment

Apart from the key components highlighted in the earlier sections, there is also auxiliary equipment in a hydraulic drive, whose role is to ensure safety, maintain pressure level, remove contaminants, and other additional functions.

2.3.7.1 Pressure Relief Valve

A pressure relief valve is used to avoid excessive pressure in a flow. It is usually installed near the outlet port of the pump to protect the pump and the system against pressure overload.

Several mechanisms can be employed in the pressure relief valve. In a spring actuated ball valve, which is the typical type of pressure relief valve, a ball element is placed on a conic seat and is pressed by a spring. The spring acts so that it tends to close the valve, while the flow of the liquid is directed in the opposite direction to open the valve. When the force of the liquid, caused by its pressure, exceeds the limit sets by the spring force, the valve is opened, and the liquid flows out of the channel to the reservoir until the pressure of the system returns to normal. As long as the pressure is below the limit, the force from the liquid does not exceed the spring force, and the valve remains close.

In relation to its operating mechanism, these two measures of pressure characterize a pressure relief valve as follows:

- cracking pressure; the pressure above which the valve passes liquid through
- closing pressure; the pressure below which the valve does not pass liquid through after being opened

2.3.7.2 Check Valve

A check valve is used to ensure that the flow of the liquid is only along one direction, hence preventing unwanted direction of flow. In a typical type of check valve called ball check valve, a ball is seated on a taper channel and is pushed by a spring against the flow of the liquid. When the liquid flows in the correct direction, *i.e.* against the spring, the liquid pushes the spring and opens the gap around the ball, allowing the liquid to flow. When the flow is in the opposite (*i.e.* incorrect) direction, the spring will shut the gap, and hence blocking the flow. This valving action can also be accomplished be a flapper or a poppet.

2.3.7.3 Shuttle Valve

A shuttle valve is a double-check valve with a cross bleed, permitting reverse flow between several connections in the circuit. It is a hydraulic version of an OR logic. It has two input connections and one output connections, making it suitable to implement an OR logic function.

2.3.7.4 Filter

A filter is used to remove harmful contaminants, especially solids, from the hydraulic liquid. A filter is usually installed with a bypass check valve to facilitate filter replacement. Various types of filter are available to serve different purposes.

Solid particles are often produced by high mechanical or hydraulic stress, which causes wear. Solid particles traveling in a hydraulic system will further cause more wear and therefore more contamination. The existence of solid in the system will cause the following:

- reduced efficiency and energy consumption due to internal leakage and heat generation
- slow cycle time and poor response dynamics
- sticky valve
- shorter life of bearing

A filter comprises a filter element that is contained inside a strong shell. The filter element may be blocked and has to be replaced from time to time. This is indicated by a differential pressure gage that is usually installed across the filter to serve as a clog indicator.

Common types of filter in a hydraulic drive are as follows:

- return line filter, installed in the return line prior to entering the reservoir
- boost filter, installed to filter the oil leaving boost pump, and therefore protects the main pump
- drain filter, to filter oil from the internal leakage of major components
- suction strainer, installed inside the reservoir to filter oil prior to reentering the system

2.3.7.5 Flow Divider

As its name suggests, a flow divider is used for dividing the flow of hydraulic liquid to desired components. A flow divider is employed when it is required to power several circuits with different flow simultaneously. The allocation of the flow is based on proportional or priority.

It is important to keep the flow divided evenly to prevent unwanted flow from a higher pressure discharge to a lower pressure discharge.

2.3.7.6 Accumulator

The main task of an accumulator is to keep certain amount of the hydraulic liquid under pressure and releases it to the system upon demand. Various types of accumulator are available, suited for specific applications, *e.g.* gas accumulator and spring accumulator.

The operational principle of a gas accumulator is based on Boyle's law. A compressed gas, usually nitrogen, is contained in a bladder against the liquid. Hydraulic pressure develops until a balance is achieved, hence keeping the pressure constant.

The operations of an accumulator can be divided into two stages:

1. When the pump in the system causes the liquid to enter the accumulator, the gas in the bladder compresses, and its pressure increases. The process of bladder compression stops when the pressures of the liquid and the gas are equal. At this stage, the bladder is not subject to abnormal mechanical stress, and due to its design, it compresses inwards.
2. When the system pressure falls, the stored liquid is returned to the system under pressure by the compressed gas, thereby increasing the system pressure closer to its normal level.

An accumulator can also be used to maintain system pressure, absorb hydraulic shock, cushion loads, dispense lubricants, and provide additional power source.

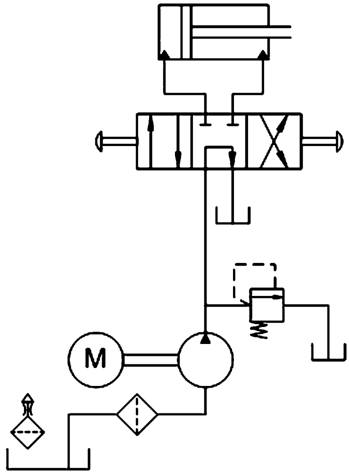

Fig. 2.16 Constant flow system

2.4 Basic Hydraulic Circuits

A hydraulic circuit refers to a combination of hydraulic components, connected to one another, to form a complete hydraulic system. There are various classifications of hydraulic systems. Based on its hydraulic characteristic, hydraulic circuits can be classified into the constant flow, the constant pressure, and the constant power system.

Several key issues in the design of hydraulic circuits, applicable to each of these categories, are as follows:

- A filter/strainer is installed before the main pump, after the tank.
- Check valves are installed before hydraulic actuator to maintain the flow direction and prevent cavitations.

2.4.1 Constant Flow System

As the name suggests, in a constant flow system (Figure 2.16), the volumetric flow of the system is kept constant throughout the operation. This system is usually used when constant displacement or velocity is required in the actuator; the reason being is that the volumetric flow is directly proportional to displacement and velocity. Constant volumetric flow can be obtained from a fixed pump turning at a relatively stable speed. Since the speed of the pump is independent of the load, the input power is proportional to the pressure due to the load. Constant flow system is applied in, for example, a hydraulic press and conveyor.

Despite its constant volumetric flow, infrequent changes of the flow may still occur. For this purpose, a flow control valve can be employed.

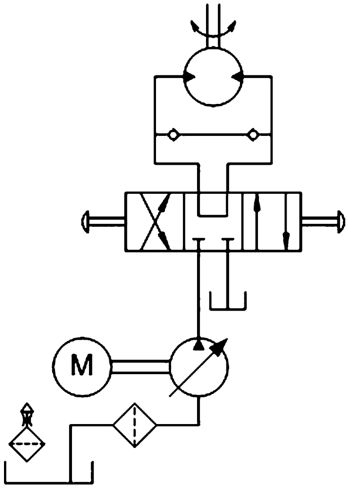

Fig. 2.17 Constant pressure system

2.4.2 Constant Pressure System

In a constant pressure system (Figure 2.17), the pressure of the system is kept constant throughout its operations. This system is usually used when constant force or torque is required in the actuator; the reason being is that the force is directly proportional to pressure. To allow the pressure to be kept constant, the volumetric flow from the pump has to be variable; thus a variable pump is usually employed in this system. A pressure-compensated variable pump is the typical pump used for a constant pressure system. This system is applied in, for example, unloading machine.

When the dynamic response becomes an important issue, a fixed pump can be used instead of a variable pump. The variation of the volumetric flow is then executed by a pressure relief valve.

2.4.3 Constant Power System

A constant power system (Figure 2.18) transmits constant power throughout the operations of the system, regardless of changes in the load. This requirement is achieved by a fixed displacement pump and a pressure relief valve, working in tandem with a variable displacement hydraulic motor. This arrangement provides constant power to the motor, while the motor adjusts its own displacement according to the requirement of the load. The constant power system is applied, for example, in a winch.

Constant power can also be achieved electrically by a constant power source (main drive), *e.g.* a storage battery providing a constant power. An increase in the load, for example, will increase the current and decrease the voltage, keeping the power constant.

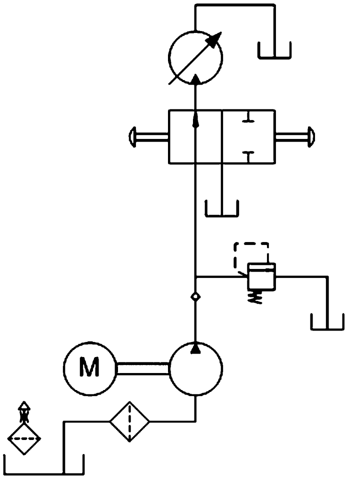

Fig. 2.18 Constant power system

2.4.4 Interlock of Hydraulic Circuits

A hydraulic drive, as any other systems, faces limitations in its operations, which can be limits of pressure, temperature, flow, *etc.* These can be high or low limits, beyond which the system will fail or even pose a health hazard.

To prevent such conditions, a hydraulic drive is usually equipped with an interlock system, which is a system to alarm, and when necessary, trip the whole hydraulic drive. To alarm means to give signs when certain limits have been breached, while to trip means to shut down certain functionalities or operations, usually when dangerous situations are about to occur. An interlock system operates automatically by means of instrumentation and can be overridden by manual action.

In many industrial applications, the limit before activation of alarm is indicated by LOW for low limit and HIGH for high limit, while the limit before trip is indicated by LOW LOW for low limit and HIGH HIGH for high limit.

Chapter 3
Electric Drives

Electric drives form an important and integral part of almost all industrial and automation processes. In addition, they also find many applications in our day-to-day lives, improving the quality of living. Modern trains are now powered by electrical drives, while cars employ electric drive components to enable basic and advanced functions. Household appliances also employ electric drives to fulfill a wide range of functions, from simple functions in electric shaver to more complicated functions necessary in an automatic washing machine.

Until today, electric drives have evolved in a significant way from what they were some 50 years ago. They are no longer constructions of windings and magnet that rotate when activated by electrical power; they are now equipped with controllers and converters that allow them to perform very precise functions almost independently. Indeed, in the last 20 years, many developments related to electric drives have taken place, enhancing electric drives capabilities to take on the challenges posed by emerging new industries.

3.1 Overview of Electric Drives

An electric drive is a system which performs a motoring function or a generating function, by converting electrical energy into mechanical energy in the former case and converting mechanical energy to electrical energy in the latter case. This book is focused on motion control systems that will leverage on the motoring function of electric drives. Figure 3.1 represents a general configuration of electric drives.

The electric motor is the core component of an electric drive as it is the motor that converts electrical energy into mechanical energy. One of the first electrical motors was invented in 1821 by Michael Faraday, which is a DC motor in nature. Ever since, the control of electric drives has been of great interest to researchers and engineers. Faraday's invention was then followed by the invention of modern DC motor by Zenobe Gramme in 1873. AC motors were later invented in 1889–1890.

The operation of a modern electric drive is mainly controlled via power electronics. One of the first inventions of power electronics is the mercury arc rectifier

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_3,

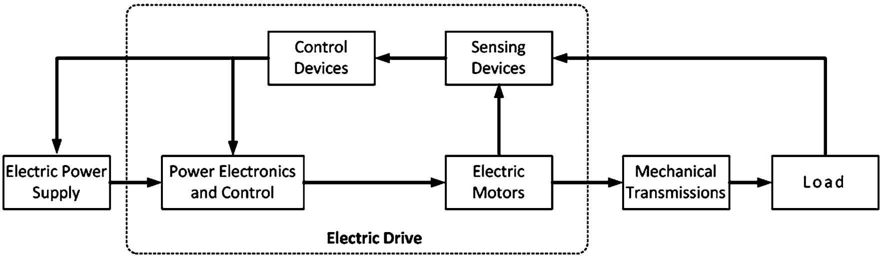

Fig. 3.1 Configuration of an electric drive

in 1920, followed by DC–AC converter in 1930. The invention of semiconductors and later on of transistors in the late 1940s by William Schokley, John Bardeen, and Walter Brattain further enhanced the development of power electronics. In the subsequent years, many more power electronic devices were invented, such as the thyristor and the power transistor. This contributed to the development of control circuitry of electric drives. In the 1980s, digital control technology begun its implementation in electric drives, riding on the advances brought forth by the application of microprocessors. This improved the performance of the drives in terms of accuracy, response speed, and cost. It has also allowed the implementation of modern complex control schemes such as self-learning and intelligent control approaches up until today.

Digital communication has further revolutionized the way a full electric drive application is developed and enabled efficient control strategies. Common and new design concepts of digital communication for drive control interface will be discussed in Chapter 6.

The applications of electric drives as motion enablers are wide ranging, from low-power applications in instrumentation, computer peripherals, and soccer robots, to high-power applications in ship propulsion and industrial mills. A classification according to precision and power requirements is presented in Figure 3.2.

From the perspectives of a motion system, the benefits offered by an electric drive can be summarized as follows:

- wide range of power, from milliwatts to megawatts
- wide range of speed, up to about 100,000 rpm
- configurable with a variety of components to suit specific application requirements
- wide range of environmental operation conditions, from controlled clean-room conditions to harsh and radioactive environment
- easy to control
- covering all quadrants/modes of operations (as will be described later); forward–backward motion, and motoring–braking operations
- adaptable to different load demands
- highly efficient in terms of energy conversion from primary energy forms
- very fast response to motion commands

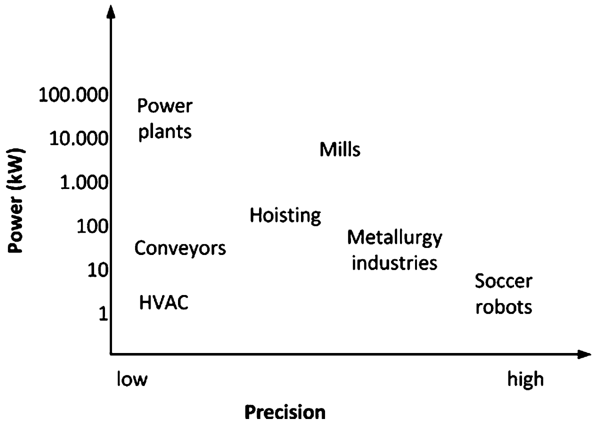

Fig. 3.2 Application range of electric drive

- significantly less noise compared to other prime movers
- simple maintenance requirements to continue operations over prolonged duration

However, an electric drive also come with constraints which may limit its use in certain applications, summarized as follows:

- continuous requirement of electric energy input, making it not feasible for mobile applications
- low power-to-weight ratio due to magnetic saturation and cooling problems

As depicted in Figure 3.1, an electric drive consists of the following main components:

- electric motors
- power electronics
- sensors

In the subsequent sections, the function and variety of these components are duly described.

3.2 Electric Motors

The principles of electric motors are based on the field of electromagnetism, which serves as a bridge between electric excitation and magnetic motion. The development of electric motors (and also generators) followed the discovery of electromagnetism by renowned scientists, such as Hans Christian Oersted, Michael Faraday, Hendrik Antoon Lorentz, and Heinrich Lenz.

Oersted first demonstrated the relationship between magnetism and electricity. When electric charge/current flows through a wire/conductor, a magnetic field is developed around the wire. Therefore, an electric current flowing through a winding

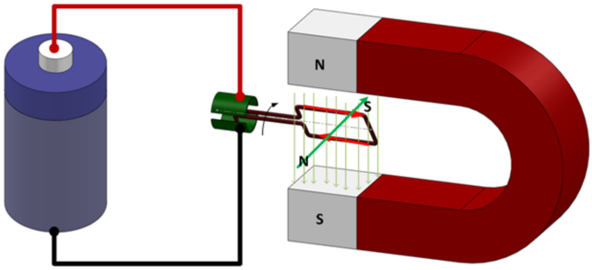

Fig. 3.3 Functional diagram of an electric motor, with N indicating north pole and S indicating south pole

of wire (known as coil) will create a magnet, more specifically, an electromagnet since it becomes a magnet only via electric excitation.

Faraday showed that the inverse phenomenon is true as well, *i.e.* current can be generated through magnetism. When a conductor is moved across a magnetic field, an electromotive force (emf) is induced across the conductor, and a current will flow if there is a closed circuit around the conductor. Lenz subsequently demonstrated that the direction of induced current will be such that its magnetic action tends to oppose/resist the motion through which it is produced.

Lorentz showed yet another permutation of these variables. If an electric current is transmitted along the conductor lying across a magnetic field, there will be a force generated on the wire, causing it to move along a direction perpendicular to the field. This force is now known as the Lorentz force.

With these physical principles, the operational principle behind an electric motor can now be explained, with respect to one of the simplest electromagnetic machine, a brush DC motor, as depicted in Figure 3.3.

This machine consists of a generated magnetic field and an armature winding. With the flow of current as indicated in Figure 3.3, the armature winding generates a magnetic field which is perpendicular to itself. This type of magnetic field, which is generated by electric current, is known as magnetomotive force (mmf). The interaction between field poles and armature poles, in the form of attraction and repulsion of opposite and similar poles, causes the motor to rotate. Commutator and brushes are used to maintain the direction of the current to force the motor to rotate continuously in one direction.

In an actual motor, the armature is fixed (and hence referred to as the stator), whereas the field winding is allowed to rotate (and hence referred to as the rotor) for ease of construction. Also, there can be numerous armature loops in the motor to provide a higher output power.

According to its excitation current, electric motors are generally classified into:

- DC excitation, including:
  - series DC motor
  - shunt DC motor
  - compound DC motor
  - brushless DC motor

- AC excitation, including:
  - synchronous AC motor
  - squirrel-cage AC motor
  - wound-rotor AC motor

3.2.1 Stepper Motor

A stepper motor is a device that converts electrical impulses into mechanical movement. Each electrical pulse input causes a rotation of the motor over a certain discrete angle. A stepper motor can therefore be considered as a digital motor, as opposed to a continuous motor. This motor allows precise control of load with regard to speed and position, with an error of less than 5% per angle. A stepper motor is commonly operated with a DC voltage, although operation with an AC voltage is also possible.

The main advantages of a stepper motor are as follows:

- compatibility with digital systems, eliminating the necessity of digital-to-analog converters
- easy implementation of closed-loop control
- wide range of step angles (from 1.8° to 90°) and torque (from a few Nm to tens of Nm)
- bidirectional
- covering low-speed operations without reduction gears
- low moment of inertia, reducing the acceleration or deceleration torque
- maximum torque at low pulse rate, reducing acceleration torque
- low starting current

The disadvantages of stepper motors are as follows:

- low efficiency
- very sensitive to misalignment of load and drive

The operational diagram of a stepper motor is depicted in Figure 3.4.

The operating principle of a stepper motor is based on attraction of opposite magnetic poles and repulsion of similar magnetic poles. In Figure 3.4, the rotor (the inner part) is a permanent magnet structure with poles as indicated. The stator is made of windings that will become an electromagnet when activated. With activation as indicated in Figure 3.4, poles 1, 2, and 3 will be north-polarized, while poles 4, 5, and 6 will be south-polarized. The rotor will thus rotate one step counterclockwise as pole N repels pole 1 and is attracted to pole 5, while pole S repels pole 4 and is attracted to pole 2.

In practice, the number of teeth of the stator and rotor determines the incremental angular rotation of the motor.

A stepper motor can be operated in either a full-stepping mode or a half-stepping mode. In a full-stepping mode, as presented in Figure 3.5, each switching action

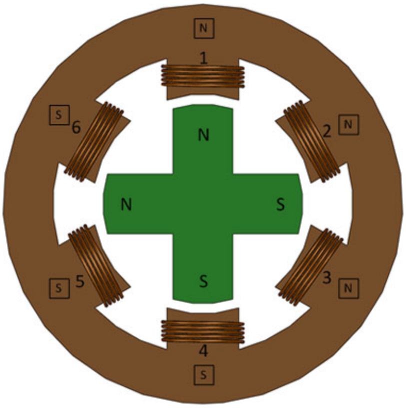

Fig. 3.4 Working diagram of stepper motor; with N indicating north pole, S indicating south pole, $\boxed{\text{N}}$ indicating north pole when activated, and $\boxed{\text{S}}$ indicating south pole when activated

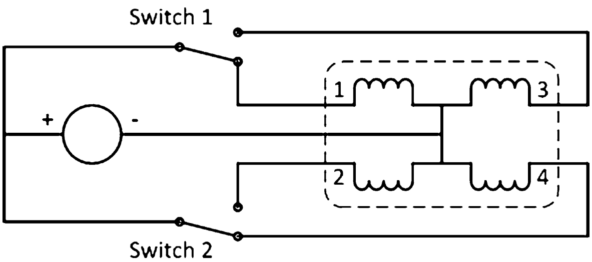

Fig. 3.5 Full-stepping switching configuration of a stepper motor

Table 3.1 Full-stepping switching sequence

Step	Switch 1	Switch 2
1	1	2
2	1	4
3	3	4
4	3	2

causes the rotor to advance one-fourth of a tooth. A full tooth is therefore achieved in four steps. The switching sequence under the full-stepping mode is presented in Table 3.1.

In a half-stepping mode, as presented in Figure 3.6, each switching action causes the rotor to advance one-eighth of a tooth. A full tooth is therefore achieved over eight steps, resulting in a finer step compared to a full-stepping mode. The switching sequence under a half-stepping mode is presented in Table 3.2.

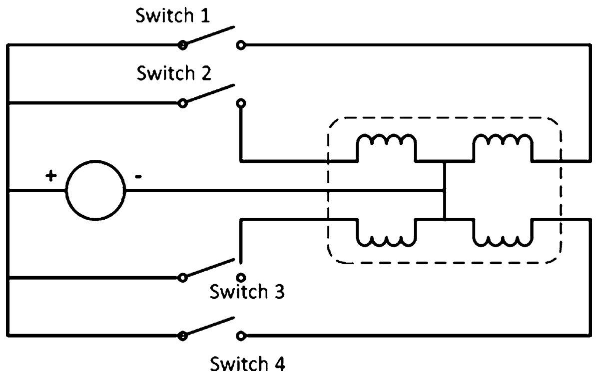

Fig. 3.6 Half-stepping switching configuration

Table 3.2 Half-stepping switching sequence

Step	Switch 1	Switch 2	Switch 3	Switch 4
1	ON	OFF	ON	OFF
2	ON	OFF	OFF	OFF
3	ON	OFF	OFF	ON
4	OFF	OFF	OFF	ON
5	OFF	ON	OFF	ON
6	OFF	ON	OFF	OFF
7	OFF	ON	ON	OFF
8	OFF	OFF	ON	OFF

Applications of stepper motors can be found in the textile industry, IC fabrication, and robotics, where positioning of parts or tools is an integral part of the requirements of such applications.

The common types of stepper motors are variable-reluctance motors, permanent-magnet motors, and hybrid motors.

3.2.2 DC Motor

DC motors have been implemented widely in the field of adjustable speed drive because of their simple control requirement. The simplicity of the control comes from the fact that the armature mmf and field mmf are independent of each other, so that controlling one mmf while keeping the other constant can be easily accomplished.

In the simplest definition, a DC motor is a motor that is driven by a DC electric supply. The speed of a DC motor is relatively easy to control, which is either via an applied armature voltage or an applied field voltage. Thus, DC motors are

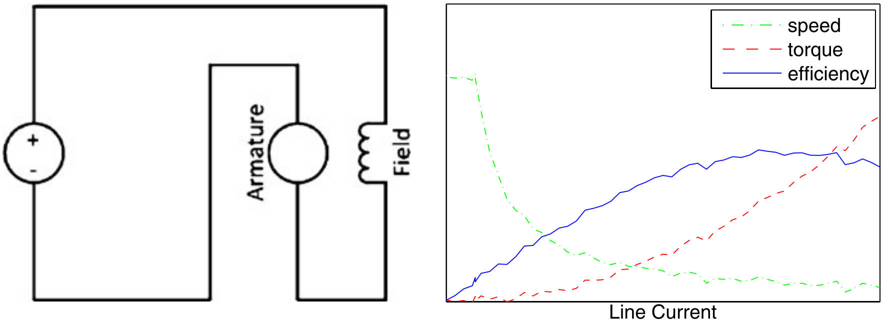

Fig. 3.7 Circuit and characteristics of a series DC motor

commonly used where variable speed and strong torque are required. Examples of applications of DC motors can be found commonly in a steel mill, printing press, crane, and hoist.

The main parts of a DC motor are the armature winding and field winding. The armature winding is the rotating component of the motor which is known as the rotor. This is attached to the shaft of the motor, which will in turn produce the rotating torque of the motor. The field winding is static and is connected to the frame of the motor. Field winding generates the primary magnetic field inside the motor.

The ends of the armature winding are connected to the commutator. The commutator is formed by insulated copper bars and acts as a switch which forces current to flow through the armature in the same direction. To improve commutation, commutating poles can be installed between regular field poles. Brushes are placed in contact with the commutator as the means to transmit the current in and out of the DC motors.

The construction of DC motors as explained above is a conventional construction known as brush DC motors. There is also another type of DC motors known as brushless DC motor, which, as the name suggests, does not have brushes and uses permanent magnets as rotors.

The magnetic fields of the armature and the field of DC motors interact with each other, causing the motor to rotate. Based on the generation and interaction of the armature and field winding, DC motors are classified as follows.

1. Series motor

   In a series motor (Figure 3.7), the field is generated by the full electrical current from the source. The torque varies as the square of the armature current. Thus, a series motor is suitable for applications that require a large torque generation with a small incremental current. On the other hand, the speed varies greatly, and thus this motor is not suitable for applications where the load may change drastically. The application of a series motor is, for example, in a crane operation.

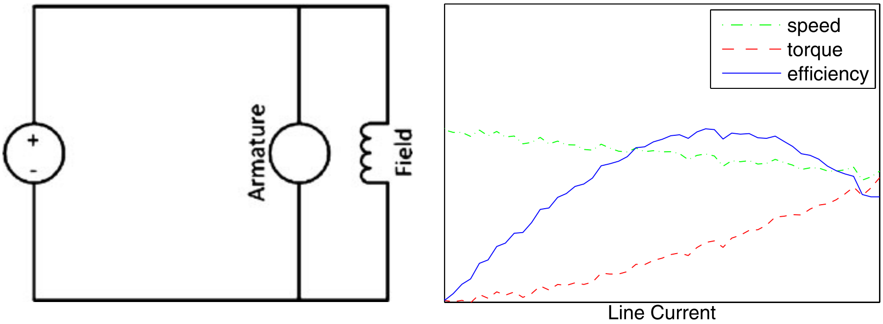

Fig. 3.8 Circuit and characteristics of a shunt DC motor

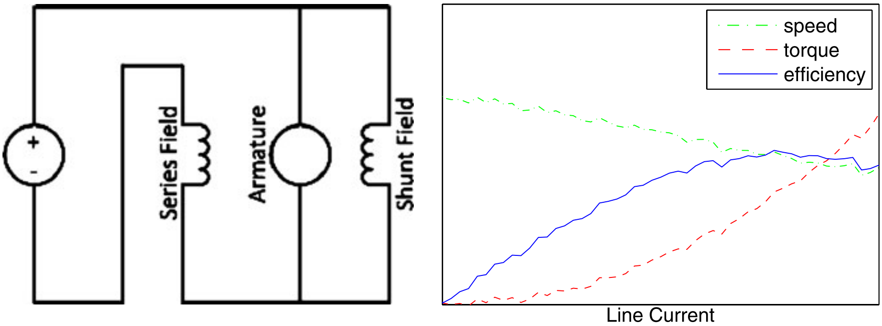

Fig. 3.9 Circuit and characteristics of a compound DC motor

2. Shunt motor

   In a shunt motor (Figure 3.8), the field is generated by a fraction of the electrical current. The speed is nearly constant within its operation range. The torque varies linearly with respect to the armature current and hence the load. Therefore, this motor is suitable for applications that require continuous operations at a constant speed.
3. Compound motor (combination of shunt and series)

   The characteristics of series motors and shunt motors are combined in compound motors (Figure 3.9). The speed changes with the load, but not as sharply as in a series motor. The applications of compound motors include elevators, conveyors, hoists, pumps, and presses. This motor is capable of starting heavy load like series motors, with safe operations at low torque like shunt motors.
4. Separately excited motor

   Permanent-magnet (PM) motors, which are typical constructions of separately excited motors (as presented in Figure 3.10), only have one winding, that is the armature winding. The magnetic field is generated using a permanent magnet, hence its name. The brushless DC motor is a type of PM motors. Brushless motors have the following advantages compared to the other types of DC motors:

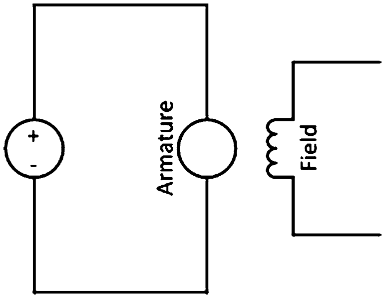

Fig. 3.10 Circuit of a separately excited DC motor

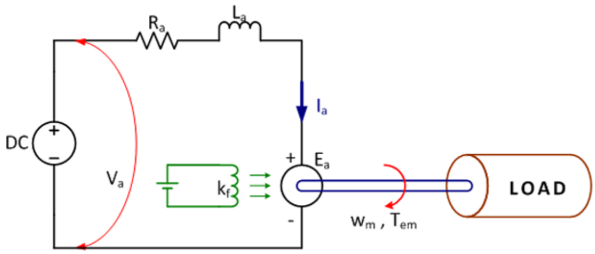

Fig. 3.11 DC motor equivalent circuit

- smaller size for a given power rating
- field strength not affected by the armature current (since they are separate in nature)
- linear relationship between speed and torque
- relatively easy to control
- high torque at low speed
- low maintenance
- self-braking mechanism

The disadvantages of brushless motors, on the other hand, include:

- relatively easy to experience overheating
- magnetic field distorted at extreme low and extreme high temperatures

DC motors can be represented as an electric circuit model as shown in Figure 3.11.

The performance characteristics of the DC motor are determined by the properties of the motor and the electrical variables applied to the motor. The properties of the motor include the resistance (R_a), the induction (L_a), and the field constant (k_ϕ). The speed of the motor (ω_m) is determined largely by its emf (which largely depends on the voltage), while the torque (T_{em}) is determined by its current. In short, the relationship among these parameters can be expressed as follows (referring to Figure 3.11):

$$\begin{aligned} E_\mathrm{a} &= k_\phi \omega_\mathrm{m}, \\ V_\mathrm{a} &= E_\mathrm{a} + I_\mathrm{a} R_\mathrm{a}, \\ T_\mathrm{em} &= k_\phi I_\mathrm{a}. \end{aligned} \tag{3.1}$$

Furthermore, the relationship between speed and torque is as follows:

$$\omega_\mathrm{m} = \frac{V_\mathrm{a}}{k_\phi} - \frac{R_\mathrm{a}}{k_\phi^2} T_\mathrm{em}. \tag{3.2}$$

The control of a DC motor constitutes the control of its speed and torque. While the control of torque can be accomplished by solely controlling the current, the control of speed can be achieved through a variety of approaches as follows:

1. Armature voltage control (controlling V_a)
   This method is preferred for speed control below the base speed, which is the maximum speed a motor can run while still delivering the rated torque.
2. Field flux control (controlling k_ϕ)
   This method is preferred for speed control above the base speed. If the speed builds up above the base speed, the insulation can fail and this will in turn burn the motor.
3. Armature resistance control (controlling R_a)
   This method, where the armature resistance is designed to be variable, is more complicated than the other two mentioned above and therefore is seldom used.

Therefore, when considering those two most widely used control methods, *i.e.* armature voltage control and field flux control, the following relationships between the two parameters and the speed motor can be used as a guideline:

- increasing armature voltage for increasing the speed of the motor, and vice versa
- increasing magnetic field for decreasing the speed of the motor, and vice versa

The approaches of controlling the speed of the motors are as follows:

1. Shunt-field control (as depicted in Figure 3.12)
   The magnetic field intensity is adjusted by a rheostat (variable resistor) installed in series to the field winding. The advantage of this approach is its high efficiency, since the unused energy is simply not delivered by the power source.
2. Armature-resistance control (as depicted in Figure 3.13)
   The armature voltage is adjusted by a rheostat installed in series to the armature winding. Increasing the rheostat resistance will subsequently reduce the armature voltage and then reduces the speed of the motor. This approach has a poor efficiency, since a large amount of unused energy is dissipated as heat through the rheostat.
3. Other special types of control
   Other types of DC motor control include voltage–voltage control, multi-voltage control, and Ward-Leonard control.

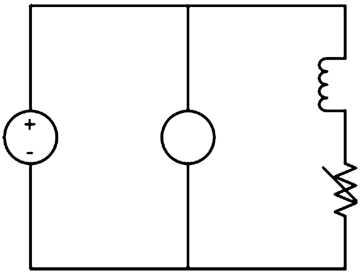

Fig. 3.12 Shunt-field control

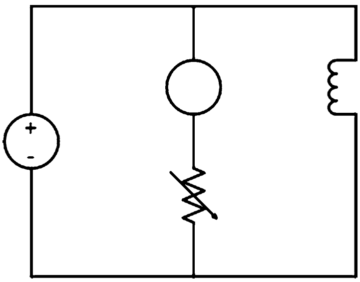

Fig. 3.13 Armature-resistance control

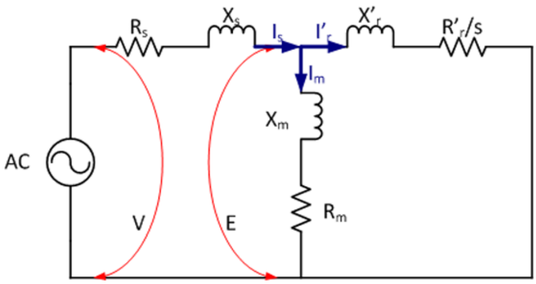

Fig. 3.14 AC motor equivalent circuit

3.2.3 AC Motor

AC motors are the economical workhorse of industries today, mainly because of their ruggedness, reliability, and low cost. Despite these merits, they are more complex than DC motors. AC motors, as the name implies, are driven by AC power supply, which requires both frequency and magnitude control. AC motors are mainly used in applications requiring a constant or slightly varying speed.

AC motors can be represented electrically as a circuitry of stator impedances, rotor impedances, and air gap impedances, each of which represents a component in its structure. Figure 3.14 represents the equivalent circuit of a one-phase AC motor.

For a three-phase AC motor, the same circuit is constructed according to the respective construction of the motor (a delta or a wye). In Figure 3.14, the parameters are according to the stator side of the motor; all other parameters are expressed with respect to the stator. The air gap impedances are included to take into account the magnetization effect that occurs at the air gap.

According to the relationship between the speed and the current phase, there are two types of AC motors:

1. synchronous motors
2. asynchronous motors (induction motors), including the squirrel-cage motors and wound-rotor motors

Between these two types, induction motors are by far more widely used in the industry.

3.2.3.1 Synchronous Motor

A synchronous motor is a motor which moves in tandem with the phase of the alternating current that drives it. A synchronous motor consists of two main elements:

- armature winding
- field winding

In practice, most synchronous motors will have a static armature winding (and hence called the stator) and a rotating field winding (and hence called the rotor). While the armature winding is connected to an alternating current supply, the field winding is connected to a direct current supply. This construction is typically implemented by installing a rectifier (explained in Section 3.3) in the motor, so that the motor effectively only requires one type of power supply, *i.e.* an AC supply.

The rotor is attached to the motor shaft, so that the resulting rotation of the rotor can be delivered to the load through this shaft.

The operational principles behind a synchronous motor are as follows. The alternating current supplied to the armature winding, which is sinusoidal in nature, produces a rotating magnetic field in the motor which rotates in synchronization with the frequency of the supply current. The direct current supplied to the field winding produces another non-rotating magnetic field, having fixed pole pairs. The interaction between the rotating and non-rotating magnetic field will thus cause the rotor to rotate at the speed synchronous to the supply frequency—hence its name.

A synchronous motor is not a self-starting machine, which means that it needs another mechanism to start the rotation from a stationary condition before it can rotate by itself at the synchronous speed. This can be explained as follows. In a stationary mode, while the armature rotating poles sweep across the field poles, they tend to pull the field poles alternately back and forth, resulting in no motion. Therefore, an additional starting mechanism is required to bring it to its operating speed. One way to start a synchronous motor is by using a squirrel cage winding (also called amortisseur winding), which is placed in the rotor to bring the rotor into

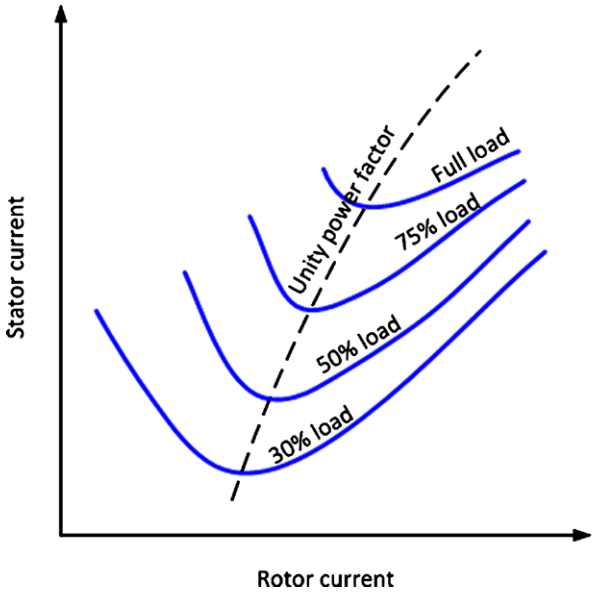

Fig. 3.15 Characteristics of a synchronous motor

its synchronous speed from stationary. After achieving the synchronous speed, the rotor is energized with the necessary DC voltage, and the motor can now operate independently.

The speed of a synchronous motor, as explained earlier, is determined by the frequency of the supply current. It is also determined by the number of the poles of the motor. Mathematically, the speed of a synchronous motor is given by

$$\omega = \frac{120 \times f}{p}, \tag{3.3}$$

where speed ω is in revolution-per-minute (rpm), frequency f is in Hertz, and p is the number of poles.

When a synchronous motor is required to operate at more than one speed, the manipulation of poles grouping is an approach that can be implemented. In this approach, the poles are grouped to effectively reduce the number of poles, which in turn increases the speed. The options of speed variation, however, are not flexible.

When a synchronous motor is pulling a constant load, the current that excites the armature winding and the field winding is varied in a certain fashion. The adjustment of the current depends on the power factor of the operating motor, as presented in Figure 3.15. At a given load, there is a unique field current that will give unity power factor, *i.e.* maximum power factor, other than which the power factor is reduced. As presented in Figure 3.15, unity power factor corresponds to the minimum armature current. However, when the motor is required to drive a variable load, the field current is usually kept constant at the maximum load.

The motor field can be excited by:

- a separate exciter set, driven by an induction motor
- a constant DC voltage supply

There are several torques associated with the operation of a synchronous motor as follows:

1. Starting torque (pull-in torque)
   The starting torque of a synchronous motor is the torque developed by the motor when the full-rated voltage is applied to the armature winding. A synchronous motor has a low starting torque of about 10% of the full torque.
2. Running torque
   The running torque is the torque developed by the motor during its full-rated operation. As the speed of the motor is largely fixed at its synchronous speed, the running torque is very much determined by the power of the motor.
3. Pull-out torque
   The pull-out torque is the maximum torque that the motor will develop without being pulled out of step, *i.e.* out of synchronization with the rotating field winding.

The performance of a synchronous motor is measured in terms of its power factor, which is the ratio of the actual power to the apparent power of the motor. A synchronous motor operates at a leading power factor, as opposed to the lagging power factor of an induction motor. Therefore, the application of synchronous motors in a system can improve the overall power factor.

Synchronous motors are typically used for the following purposes:

1. Power factor correction
   The application of synchronous motors as power factor correction is because of its leading power factor. In a large installation with many electrical devices, which typically have lagging power factors, the overall power factor can be very poor and can reduce the efficiency. The leading power factor of a synchronous motor improves the overall power factor of the installation. The power factor of a synchronous motor can be varied by adjusting the current of the field windings. This is a very desirable characteristic of a synchronous motor, as the power correction of power factor can be made adjustable. When a synchronous motor is employed exclusively for the purpose of correcting power factor, it is often called synchronous capacitor.
2. Voltage regulation
   At the end of a long transmission line, the voltage can vary greatly, especially with large inductive loads. Connection or disconnection of inductive loads may cause the voltage to rise or drop significantly. This may be overcome by installing a synchronous motor with a voltage regulator to control its field winding voltage. The action of voltage regulator is such that when the voltage drops, the field of the motor is strengthened so that its power factor rises and the line voltage is maintained. On the other hand, when the voltage rises, the field of the motor is weakened so that its power factor drops and again the line voltage is maintained.

3. Constant speed load

   As a synchronous motor runs at a constant speed, it is very suitable for applications that require a constant speed, for example in paper mills, centrifugal compressors, and DC generators.

3.2.3.2 Squirrel-cage Motor

Squirrel-cage motors are the workhorses among the motors used in the industry, having advantages because of their simple construction, operation, ruggedness, and ease of manufacture.

A squirrel-cage motor consists of:

- field winding as the stator
- squirrel-cage bars (hence the name) as the rotor

The stator is constructed by a laminated steel core with slots in which coils are located. The coils are grouped and connected to form a polar area, producing a rotating magnetic field. The rotor, on the other hand, is constructed from laminated steel, while the windings consist of short-circuited conductor bars. The gap between the stator and the rotor is called the air gap and is designed to be as small as possible to obtain the best power factor of the motor.

The operating principle of a squirrel-cage motor can be explained as follows. The field winding is energized by AC voltage, and this produces a rotating magnetic field inside the motor by virtue of the alternating current. As the magnetic field rotates, it cuts the squirrel cage bars and sets up voltage according to Faraday's principle. This voltage causes electrical current to flow through the squirrel-cage bars and develops another magnetic field, now on the rotor. The poles in the rotor interact with the rotation poles in the stator in an alternate attraction and repulsion, causing the rotor to rotate. However, the rotor does not rotate as fast as the rotating magnetic field, for if it was so, the conductor would be stationary rather than cutting across the field. Therefore, the rotation of a squirrel-cage motor is always less than its correspondent synchronous motor (hence, always less than the associated synchronous speed).

The difference between its operating speed and synchronous speed is known as slip, commonly expressed as a percentage of synchronous speed as follows:

$$s = \frac{\omega_{\text{synchronous}} - \omega_{\text{operating}}}{\omega_{\text{synchronous}}} \times 100\%, \tag{3.4}$$

where s is the slip of the motor.

Therefore, the speed of the motor is formulated as follows:

$$\omega = (1 - s) \times \omega_{\text{synchronous}}. \tag{3.5}$$

Slip is dependent on the load of the motor. When the load increases, the slip becomes larger.

As in synchronous motors, squirrel-cage motors also require a certain amount of starting torque. This ranges from 5% to 50% of the full rated-torque.

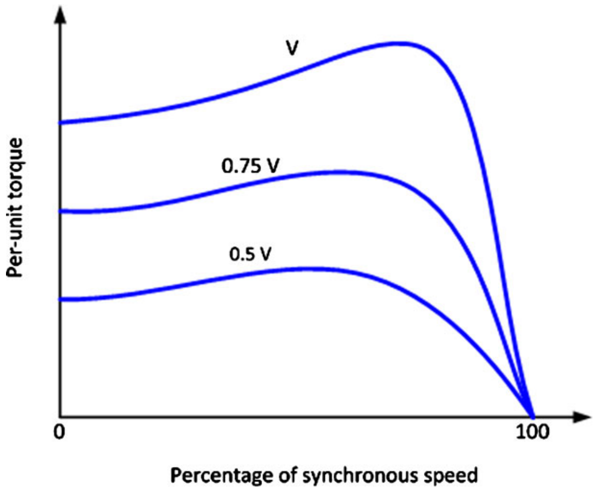

Fig. 3.16 Torque-speed characteristic of a squirrel-cage motor, where V is the rating voltage

A typical relationship between the torque and the speed of a squirrel-cage motor is presented in Figure 3.16.

To achieve satisfactory operations, suitable control schemes must be implemented to the motors, as will be discussed in Chapter 5. The control schemes must be able to allow the motors to perform the following functions satisfactorily:

- start and stop the motor
- speed regulation
- speed reversal
- motor protection

Squirrel-cage motors are applied in many diverse areas of applications in industry, due to their ruggedness and simplicity.

3.2.3.3 Wound-rotor Motor

A wound-rotor motor consists of:

- field winding as its stator
- armature winding as its rotor

The stator is constructed exactly as one used in squirrel-cage motor. The rotor, however, is different and consists of coils of wire connected in regular succession, having the same number of poles as the stator. The ends of the windings are connected to slip rings.

The operational principles of wound-rotor motors are as follows. As in a squirrel-cage motor, the stator produces a rotating magnetic field in the motor. This magnetic field cuts through the rotor winding. The current induced in the rotor is carried to an

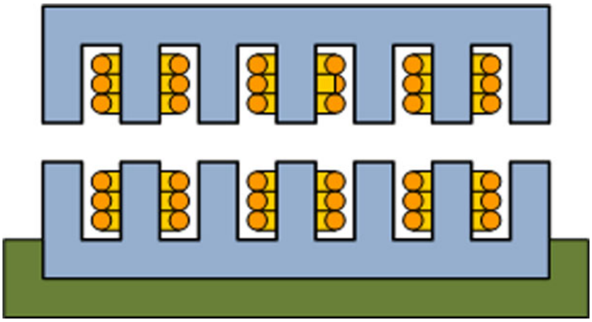

Fig. 3.17 Construction of linear motor

external resistance through slip rings. Subsequently, rotor current produces another magnetic field, now in the rotor. The interaction between these two magnetic fields causes rotation of motor, in the same manner as in a squirrel-cage motor.

As explained above, the induced current of the rotor is transmitted to an external resistance. The variation of this resistance can therefore be used to control the operations of the motor, since the intensity of the induced current directly affects the resulting magnetic field. The controller used in a wound-rotor motor is used for the following functions:

- to start and stop the motor in a satisfactory manner
- to regulate the speed of the motor, by variation of the external resistance

Wound-rotor motors are implemented in applications that do not require exact speed regulation, allows intermittent operation in low efficiency, and requires starting with a heavy load. Examples of such applications include hoists, cranes, elevators, pumps, and compressors.

3.2.4 Linear Motor

While rotary motors comprise almost 98% of motors applications, linear motors are also available with their unique merits. Linear motors are very much like rotary motors, having two windings that act as armature winding and field winding. The structure of a linear motor can be imagined as cutting and unrolling a rotary motor. The result is a flat linear motor that produces linear forces. The construction of a linear motor is presented in Figure 3.17.

The advantages of linear motors are as follows:

- high precision and accuracy in a high-speed implementation as compared to its rotary counterpart
- no requirement of coupling mechanism, as the output motion has already been a linear motion

- higher magnetic flux without significant heat (hence low thermal losses), as it incorporates earth permanent magnet
- high force density

The applications of linear motors are in industries or processes that require high accuracy and precision, such as semiconductor processes and precision metrology.

There are several designs of liner motor as follows:

1. Force-platen

   This motor consists of a moving platen and a stationary platen. The moving platen consists of electromagnetic coils (with winding and iron core). Permanent magnets are placed on the stationary platen oriented at a right angle to the thrust axis, but slightly skewed in the vertical plane to reduce the thrust ripple. Force-platen motors feature a low height profile and a wide range of available size. The applications include automobile and machine tools applications where high continuous and peak forces are required.
2. U-shaped

   This motor consists of a U-shaped motor armature and a permanent magnetic field generated by the track. The armature is a planar winding epoxy bonded to a plastic blade, which projects between double rows of magnet. The magnetic field works in conjunction with the electromagnetic field in the blade to produce linear motion. U-shaped motors are used in high-precision operations that require smooth motion, since these motors give a zero detent force. Furthermore, U-shaped motors are cost-effective and provide a long travel length. The drawbacks, however, include resonance in high-acceleration operation, high thermal losses, and inefficient magnetic utilization.
3. Tubular

   Tubular motors consist of a stationary thrust rod and a moving thrust block. The thrust rod is a permanent magnet, while the thrust block is an electromagnetic winding. Tubular motors feature a high force generation and high energy efficiency. The drawbacks include limited travel distance, tall height, and limited force range.

3.3 Power Electronics

The main task of power electronics is to process and control the flow of electric power from the source to the load by means of power semiconductor devices in such a way that the electric power is used efficiently. Power electronics devices should consume very little power to manipulate the flow of large electric power. Power conversion can be as follows:

- from DC to DC (chopper), whereby the output is a controlled DC voltage
- from DC to AC (inverter), whereby the output is a controlled AC voltage with controlled frequency
- from AC to AC (voltage regulator), whereby the output is a controlled AC voltage

- from AC to DC (rectifier), whereby the output is a controlled (variable) DC voltage

Conversion of electrical power, as well as variation of electrical parameters (such as current, voltage, and frequency), were carried out via other means, *e.g.* motor–generator. In recent years, however, power electronics has been taking over the role of power conversion, especially with advanced developments in microelectronics. The advantages of using power electronics are as follows:

- low maintenance
- fast response time
- wide control capability
- design compactness
- low cost
- high efficiency
- high reliability

Power electronic circuits consist of power semiconductor devices (PSD), which can be considered as on–off switches. Based on its controllability, PSD can be classified into the following:

- uncontrollable devices, where it is not possible to control the switch externally at any instance (*i.e.* the switching-on and the switching-off of the device fully determined by the operating conditions of the circuit)
- semi-controllable device, where it is possible to control the switching-on of the device externally, but not its switching-off (*i.e.* its switching-off determined by the operating conditions of the circuit)
- controllable device, where it is possible to control both the switching-on and the switching-off of the device externally

The following PSDs are often used in power electronic devices:

1. Diode
   This is a two-terminal uncontrollable device. A diode acts as a switch; it turns on when it is forward-biased and turned off when it is reversed-biased.
2. Thyristor
   It is also known as Silicon Controlled Rectifier (SCR). It is a three-terminal semi-controllable device. It is turned on when it is forward-biased and when the gating signal is positive, and is turned off when it is reversed-biased.
3. BJT (Bipolar Junction Transistor)
   It is a three-terminal fully controllable device, which is controlled by the base current. A BJT is turned on when a positive base current is applied and turned off when a negative base current is applied. The operating frequency is typically less than 1 kHz.
4. GTO (Gate Turn Off)
   A GTO is a controllable version of thyristor, with negative current to turn it off. The switching frequency is typically 1–2 kHz. The application of GTO is for UPS, VAR-compensator, and machine drive control.

5. MOSFET (Metal Oxide Semiconductor Field Effect Transistor)
   It is a three-terminal controllable device, which is controlled by the gate-to-source voltage. The typical switching frequency is of the order of MHz.
6. IGBT (Insulated Gate Bipolar-junction Transistor)
   It is a hybrid of MOSFET and BJT, therefore combining the advantages of both. It has a high impedance gate requiring small switching energy like MOSFET, yet small voltage drop like BJT.
7. MCT (MOS Controlled Thyristor)
   It is a hybrid of thyristor and MOSFET. MCT is a voltage-controlled device, and it has a faster switching frequency when compared to thyristor.

3.3.1 DC to DC Converter

A DC–DC converter, also termed as choppers, is a power electronic device that converts unregulated DC supply voltage to regulated (controlled) DC output voltage. With this converter, the output voltage is maintained constant regardless of fluctuation at the input side. As its name implies, this converter is used for DC motor controls. There are two types of DC–DC converter, based on the operating principle, as follows:

- linear regulator type
- switching converter type

Switching converter type is widely used because of its high efficiency compared to the linear regulator type.

However, in the discussion of DC motor control, it is more useful to classify DC–DC converter based on its output voltage and current, as it corresponds directly to motor speed and torque. The classifications are then as follows:

1. Class A Converter
   Class A converter outputs positive voltage and positive current; hence, it outputs positive power. As the output voltage is always less than the input voltage, this converter is also called step-down (buck) converter.
2. Class B Converter
   Class B converter outputs positive voltage and negative current; hence, it outputs negative power. As the output voltage is always higher than the input voltage, this converter is also called step-up (boost) converter.
3. Class C Converter
   Class C converter outputs positive voltage and either positive or negative current. This converter can be seen as a combination of the class A and the class B converters and is also called step-down-step-up (buck-boost) converter.
4. Class D Converter
   Class D converter outputs both positive or negative voltage and positive current.

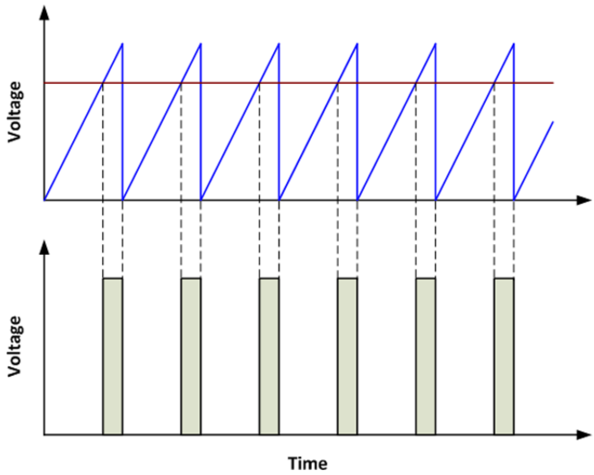

Fig. 3.18 Generation of PWM in DC–DC converters

5. Class E Converter

   Class E converter outputs voltage and current with either polarity, positive and negative. This converter is suitable for control of DC motor in all directions and all modes of operation.

The control of a DC–DC converter is accomplished by means of pulse-width-modulation (PWM). In this approach, the parameter being varied is the ratio of the on-duration to the switching time period; this ratio is known as the duty cycle. The switching is usually done at a constant frequency. The switch control signal is generated by comparing the signal control voltage, which is an amplified error, to a reference waveform, which is usually a sawtooth waveform. The frequency of the reference signal determines the switching frequency and is typically of the order of a few kHz. The switch will be turned on if the signal control voltage is greater than the reference waveform and will be turned off if otherwise. The generation of PWM signal is depicted in Figure 3.18.

Other than the PWM approach, a less-popular pulse-frequency-modulation (PFM) can also be implemented to control DC–DC converter. In this approach, the duty cycle is kept constant at a typical value of 0.5, but the frequency is varied. Higher frequency corresponds to a higher voltage, and lower frequency corresponds to a lower voltage. The significant switching loss makes it undesirable than the PWM approach.

3.3.1.1 Class A Converter

As the name suggests, a step-down (Class A) converter produces a lower output voltage than its input. It consists of a controlled switch and a diode, as depicted in

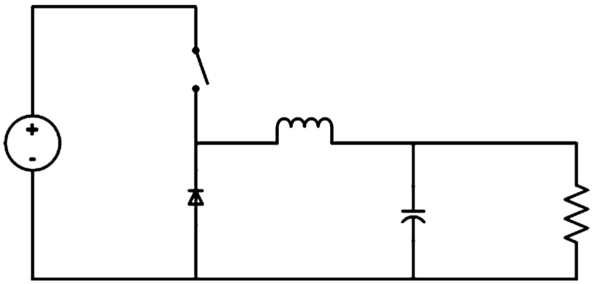

Fig. 3.19 Class A converter

Figure 3.19. The output voltage V_o can be determined from duty cycle D and input voltage V_i as follows:

$$V_o = DV_i. \tag{3.6}$$

The operations of a step-down DC–DC converter are explained as follows. During the period when the switch is on, the diode is reversed biased, and the source provides power to the load and inductor. During the period when the switch is off, the inductor releases its stored energy, while at the same time, the diode (called free-wheeling diode) is forward biased, providing power to the load. The capacitor stores electrical energy to prevent an output voltage ripple.

A step-down converter is mainly applied for the speed control of DC motors.

3.3.1.2 Class B Converter

As the name suggests, a step-up (Class B) converter produces a higher output voltage than its input. It consists of a controlled switch and a diode, as depicted in Figure 3.20. The output voltage V_o can be determined from duty cycle D and input voltage V_i as follows:

$$V_o = \frac{1}{1-D} V_i. \tag{3.7}$$

The operations of a step-up DC–DC converter are as follows. During the period when the switch is on, the diode is reversed biased, and this isolates the output from the input. The input then supplies the power to the inductor. During the period when the switch is off, both the inductor and the input supply electrical power to the load, making its average voltage always bigger than the supply. The capacitor stores electrical energy to prevent an output voltage ripple.

A step-up converter is mainly applied to the regenerative braking control of DC motors.

The other DC–DC converters are basically combinations of step-down and step-up converters.

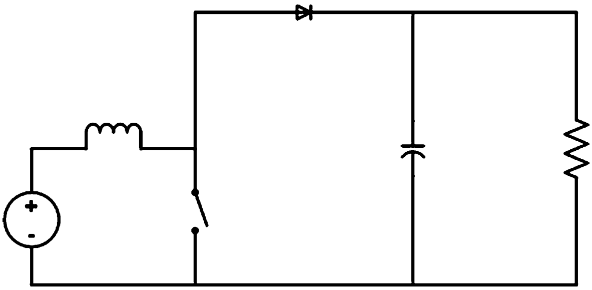

Fig. 3.20 Class B converter

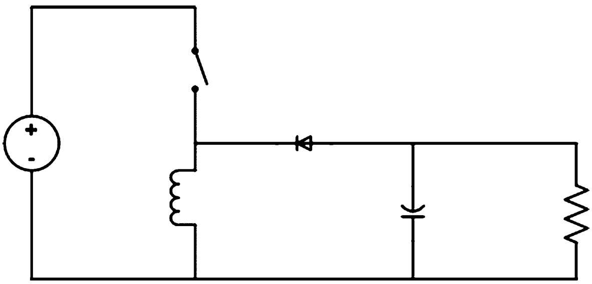

Fig. 3.21 Class C converter

3.3.1.3 Class C Converter

A step-down-step-up converter is a cascade-combination of step-down and step-up DC–DC converters. It produces positive voltage, so that it can keep the rotation of a DC motor in one direction. Furthermore, it can produce either positive or negative current, hence either positive (motoring) or negative (braking) torque. This arrangement makes it possible to drive a DC motor over two-quadrant modes, although it can only move a load in one direction. The basic circuitry of a buck-boost converter is presented in Figure 3.21. The output voltage V_{o} can be determined from duty cycle D and input voltage V_{i} as follows:

$$V_{\mathrm{o}} = \frac{D}{1-D} V_{\mathrm{i}}. \tag{3.8}$$

The operations of a step-down-step-up converter are as follows. During the period when the switch is on, the diode is reversed biased, and the input provides electrical power to the inductor. During the period when the switch is off, the inductor supplies electrical power to the load. With this arrangement, the load is never connected directly to the source; therefore a step-down-step-up converter is also referred to as an indirect converter. Again, the capacitor stores electrical energy to prevent an output voltage ripple.

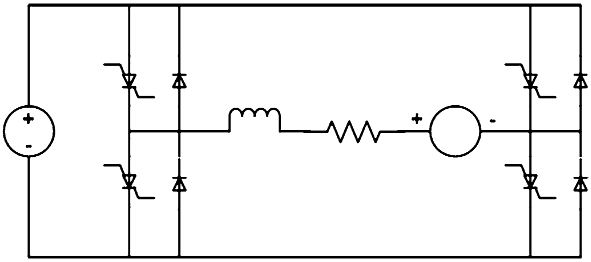

Fig. 3.22 Class E converter

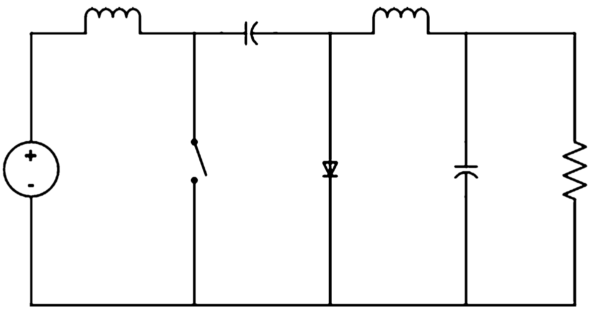

Fig. 3.23 Cuk converter

3.3.1.4 Class E Converter

A class E converter can provide all combinations of positive and negative voltage and current. This allows a DC motor to be controlled in four-quadrant mode; *e.g.* both motoring and braking in forward and backward directions. The circuitry of the Class E converter is presented in Figure 3.22.

3.3.1.5 Cuk Converter

Yet another DC–DC converter that does not fall under the classification above is a cuk converter. The circuitry of a cuk converter is presented in Figure 3.23. This converter can provide negative voltage with respect to the input. The output voltage V_o can be determined from duty cycle D and input voltage V_i as follows:

$$V_o = \frac{D}{1-D} V_i. \tag{3.9}$$

The operations of a cuk converter are as follows. During the period when the switch is on, the diode is reversed biased, and therefore both inductor currents flow through the switch and the capacitor discharges its power to the load. During the period when the switch is off, both inductor currents flow through the diode and therefore charge the capacitor.

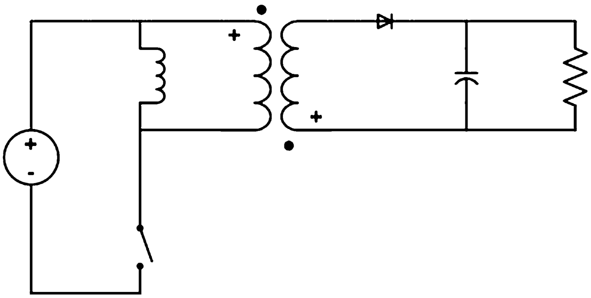

Fig. 3.24 Flyback converter

3.3.1.6 Flyback Converter

All DC–DC converters used in the control of DC motors and discussed above are conventional converters; the output load is not isolated from the input source. The advantage of providing some kind of isolation between the source and the load is that harsh fluctuation in the source will not affect the load.

The flyback converter is one way to provide the isolation, via a transformer. The operating principle of a flyback converter is similar to the rest of DC–DC converters. Electrical energy is stored in the inductor during the period when the switch is on and transferred to the load during the period when the switch is off. The circuitry of a flyback converter is presented in Figure 3.24.

3.3.2 DC to AC Converter

The DC–AC converter, also termed as an inverter, is a power electronic device that converts unregulated DC supply voltage to regulated (controlled) AC output voltage using static power semiconductor devices. With this converter, the output is a sinusoidal AC voltage with controlled magnitude and frequency. As its name implies, this converter is used for AC motor controls. The typical arrangement of a DC–AC converter is as follows, as depicted in Figure 3.25:

- a fixed voltage and frequency AC supply to be rectified to produce a DC voltage intermediate supply
- the resultant DC supply to be inverted to a controlled AC supply with a DC–AC converter

This arrangement allows regenerative braking of AC motor to take place and to improve the power factor of the motor installation.

DC–AC converters are classified into:

- Voltage Source Inverter (VSI), whose input is a DC voltage source; further classified into:
  - square wave inverter

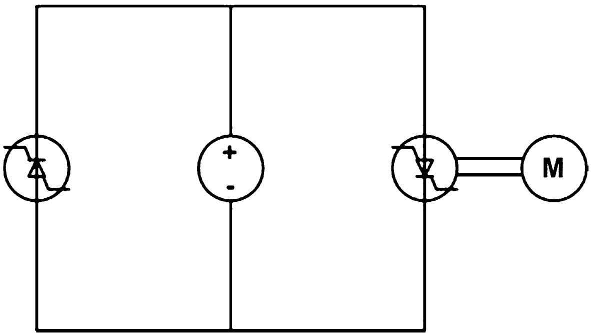

Fig. 3.25 DC–AC converter

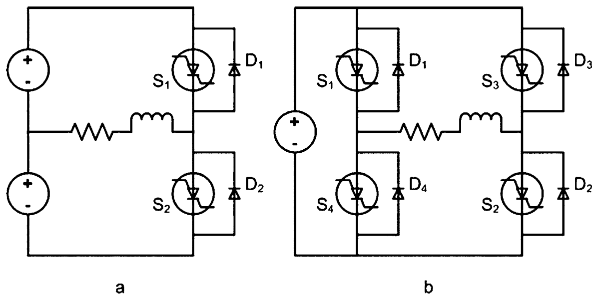

Fig. 3.26 Square wave inverter; (**a**) half-bridge inverter, (**b**) full-bridge inverter

  – pulse-width modulated (PWM) inverter
- Current Source Inverter (CSI), whose input is a DC current source

The square wave inverter can be constructed with two types of circuit: half-bridge and full-bridge, as presented in Figure 3.26. In a half-bridge inverter, the switches S_1 and S_2 are switched on and off alternately in such a way that they are not switched on simultaneously to avoid short-circuit. In a full-bridge inverter, the switches are activated as follows:

- switches S_1 and S_2 to provide positive voltage, in alternate fashion with
- switches S_3 and S_4 to provide negative voltage

The output voltage of a square-wave inverter is then alternating, but not sinusoidal. The frequency is controlled by varying the timing of the activation of the switches. The magnitude is controlled by varying the DC intermediate supply. Alternatively, the switching is modified so that it outputs zero voltage over certain periodical instances. This is only possible when a single-phase AC voltage is to be produced. This function is called voltage cancellation.

The application of PWM inverter can overcome the disadvantages of square-wave inverter. The characteristics of PWM inverter that makes it superior to square-wave inverter are:

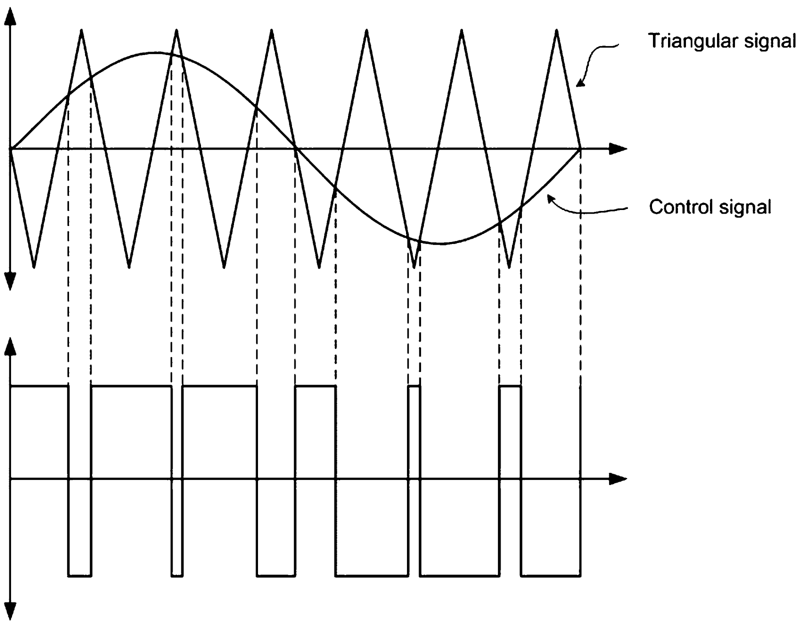

Fig. 3.27 Generation of PWM in DC–AC converters

- control of output voltage is internal, within the inverter itself
- low total harmonic distortion, reducing the necessity of installing additional filter

However, the circuit is more complex, and this contributes to higher switching losses.

PWM signal in DC–AC converter is generated as follows. A sinusoidal control signal, with the same frequency and magnitude as the desired output frequency and voltage, is compared to a triangular waveform. The frequency of the triangular waveform (also called carrier) determines the switching frequency of the inverter and is kept constant. The switch of the inverter is turned according to the following schemes (as also presented in Figure 3.27):

- switched on when the sinusoidal control signal is higher than the triangular reference signal
- switched off when the sinusoidal control signal is lower than the triangular reference signal

The following modulation ratios are defined:

- amplitude modulation ratio m_a, where

$$m_a = \frac{V_{\text{control}}}{V_{\text{triangular}}} \tag{3.10}$$

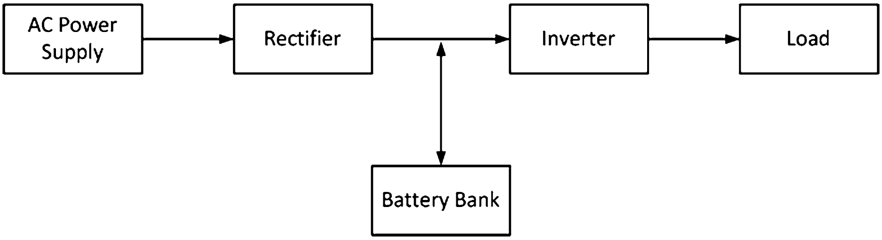

Fig. 3.28 Uninterruptible power supply

- frequency modulation ratio m_f, where

$$m_f = \frac{f_{\text{triangular}}}{f_{\text{control}}} \tag{3.11}$$

The output voltage will be a sinusoidal voltage with harmonics that appears as side band centred on the switching frequency. Therefore, to reduce the effect of harmonics, frequency modulation ratio is increased to push the harmonic frequency to a higher frequency for easy filtering out. Additionally, frequency modulation ratio can be chosen to be an odd integer, so that it only consists of odd harmonics. With V_d as the input voltage, the peak amplitude of the fundamental component of the output voltage V_o can be formulated to be

$$V_o = \frac{1}{2} m_a V_d. \tag{3.12}$$

As mentioned above, the main application of a DC–AC converter is in the speed control of an AC motor. By controlling the output frequency of the inverter, the synchronous speed of the motor can then be controlled, and from that, the speed of an induction motor can be controlled.

Another application of this converter is the Uninterruptible Power Supply (UPS), which is a power electronic device for AC supply with critical AC loads. UPS provides protection against power outage and also voltage regulation, as presented in Figure 3.28.

A rectifier, equipped with high-frequency flyback converter, supplies power to the inverter and batteries. The output of the inverter is regulated in such a way that it contains very little harmonic distortion. The inverter must allow almost instantaneous control over its output waveform to provide a high dynamic response.

3.3.3 AC to AC Converter

An AC–AC converter (also called a voltage regulator) converts an uncontrollable AC voltage to a controllable AC voltage, in terms of both the magnitude and frequency. The conversion can actually be performed with several approaches as follows:

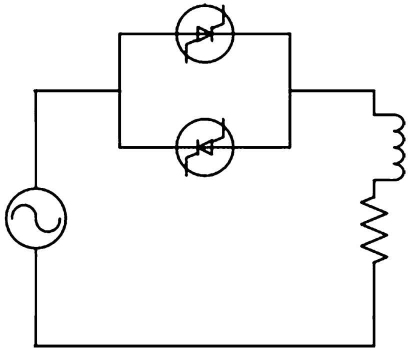

Fig. 3.29 AC–AC converter

- direct conversion, where the conversion is accomplished without any intermediate stage; consisting of:
  - AC controller, where the output frequency is the same as the input, only the magnitude is varied
  - cyclo-converter, where the output frequency and magnitude are both controlled
- indirect conversion, where the AC input voltage is first converted to DC intermediate voltage and then converted to AC output voltage with controlled frequency and magnitude

AC controller controls the amplitude of the output voltage with the following approaches:

- integral cycle control, where the converter is switched on and off periodically according to a certain duty cycle
- phase angle control, where the converter is controlled by the phase delay angle

The circuitry of AC–AC converter is presented in Figure 3.29.

AC–AC converter has limited applications in electric drives. One application is for soft starting of an induction motor, where the starting current of the motor is regulated via AC–AC converter such that it does not directly jump to the full operating current of the motor. In practice, this is accomplished by reducing the applied current while keeping the frequency constant, hence reducing the speed of the motor. With this operation, however, the efficiency is poor.

3.3.4 AC to DC Converter

An AC–DC converter, also termed rectifier, converts an uncontrollable AC voltage to uncontrollable or controllable DC voltage, depending on the requirement. An AC–DC converter is typically constructed using a full-bridge rectifier and consists of four diodes as presented in Figure 3.30.

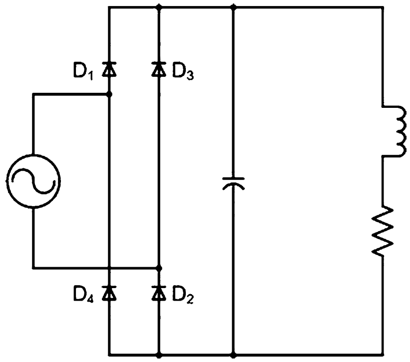

Fig. 3.30 AC–DC converter

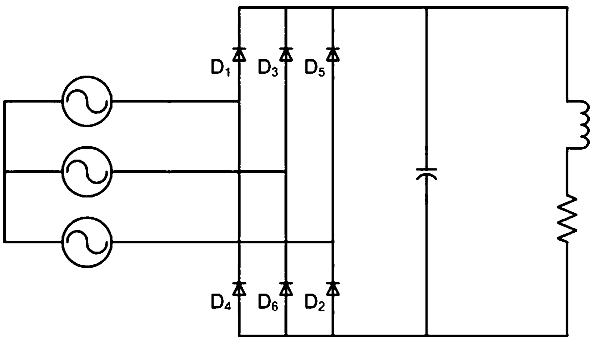

Fig. 3.31 Three-phase rectifier

The operation principle of an AC–DC converter is as follows. The AC power supply will output a sinusoidal voltage, which has an alternate polarity in nature. During the positive half-cycle, D_1 and D_2 conduct so that the output voltage has the same polarity as the supply; that is, the output voltage is positive. During the negative half-cycle, D_3 and D_4 conduct so that the output voltage has opposite polarity as the supply; that is, the output voltage is still positive. As a result, the output voltage will have the same polarity throughout the operation cycles. In an ideal condition, the magnitude of the output voltage is the same as the amplitude of the voltage source.

The conversion of sinusoidal voltage to non-sinusoidal (in this case constant) voltage brings about harmonic distortion that reduces the performance of the converter. Another distortion comes from commutation current, which occurs due to the alternate changes of current in the source side of the converter. The finite time required for current changes to take place also reduces the output voltage.

The conversion of a three-phase AC voltage can be accomplished by a three-phase rectifier, as depicted in Figure 3.31.

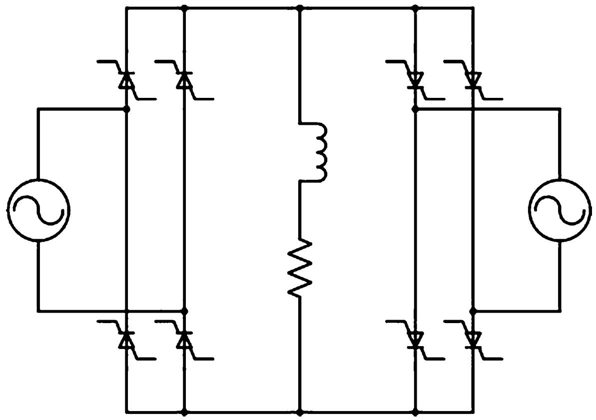

Fig. 3.32 Dual converter

A capacitor is placed in parallel to the load to stabilize the output voltage, resulting in a nearly constant output voltage.

A more advanced AC–DC converter is a controlled rectifier, where the output voltage is controllable. The control of the output voltage is accomplished by controlling the delay time of the gate pulse. Controlled rectifier can be classified into:

- full controlled rectifier, where the output voltage can be controlled from $-V_o$ to $+V_o$
- half controlled rectifier, where the output voltage can controlled from 0 to $+V_o$

The circuitry of a controlled rectifier is very similar to the conventional rectifier, except that the diodes are replaced with thyristors. The conduction of thyristors is controlled to regulate the output voltage. By doing so, the output voltage is chopped from the input voltage. Although the controlled rectifier is desirable in terms of controllability, it has a more significant harmonic distortion compared to the conventional rectifier.

Dual converter is yet another configuration of AC–DC converter where both the voltage and current can be reversed. The circuitry of a dual converter is presented in Figure 3.32.

AC–DC converters are applied in several areas, such as for DC motor drive control and battery charger circuit. As mentioned, it may also be implemented together with an inverter.

3.4 Sensors

In motion control applications, the primary sensors in electric drives consist of torque, position, speed/velocity, and acceleration sensors. An overview of these common sensors is covered in this section.

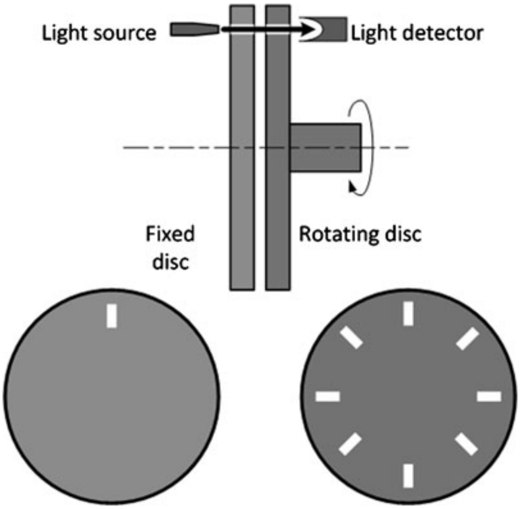

Fig. 3.33 Encoder

3.4.1 *Position Measurement*

Common types of position sensors include the encoder, potentiometer, reflective opto-transducer, linear variable differential transducer (LVDT), and the strain gauge.

3.4.1.1 Encoder

An encoder, as presented in Figure 3.33, is a device that provides a digital output as a result of a displacement. The positioning information of an encoder is categorized into:

- incremental encoder, detecting the displacement from some points
- absolute encoder, detecting the actual position of the encoder

The rotating disc has a number of apertures through which light beam can pass through from the light source to the light detector. When the disc rotates along with the shaft, a pulsed output is produced by the sensor. The angular displacement of the shaft can then be determined by the number of the output pulses. The typical resolution of encoder employed in industry varies between 6° and 0.3°.

3.4.1.2 Potentiometer

A potentiometer consists of a resistance element with a sliding contact that is movable over the length of the element. Figure 3.34 shows the working diagram of a potentiometer.

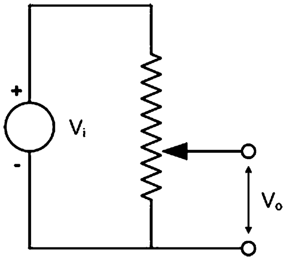

Fig. 3.34 Working diagram of a potentiometer

The operating principle of a potentiometer is based on the relationship between resistance and length. The sliding contact of the potentiometer varies the effective length of this device. The longer the effective length, the higher the resistance and hence the output voltage, V_0. The movable sliding contact can be attached to a mechanism, and therefore the position of that device can be determined from the output voltage.

3.4.1.3 Opto-transducer

Figure 3.35 shows the construction of a reflective opto-transducer, consisting of an infrared LED and phototransistor.

The operating principle of an opto-transducer is as follows. The beam from the three LEDs (in this case) is directed to hit the surface of the rotating Gray-coded disc, as presented in Figure 3.36. The beam will be reflected back to the phototransistor upon hitting the reflective surface of the disc, for which case a “1” will be its output. On the other hand, the beam will not be reflected upon hitting the non-reflective surface of the disc, for which case a “0” will be its output. The position of the rotating disc is then indicated by the combination of 1s and 0s, which can be interpreted with the help of a Gray-code table as exemplified by Table 3.3.

3.4.1.4 Linear Variable Displacement Transformer (LVDT)

The construction and circuit arrangement of an LVDT are as shown in Figure 3.37. It consists of three coils mounted on a common former with a movable magnetic core within the coils.

The operational principle of an LVDT is as follows. The center coil is the primary and is supplied from an AC supply. The coils on either side are the secondary coils, labeled A and B in Figure 3.37. Coils A and B have equal number of turns and are connected in series in anti-phase fashion so that the output voltage is the difference between the induced voltages in the coils. The amplitude of the output voltage increases as the coils move away from the neutral position, while the phase changes according to the direction of the movement.

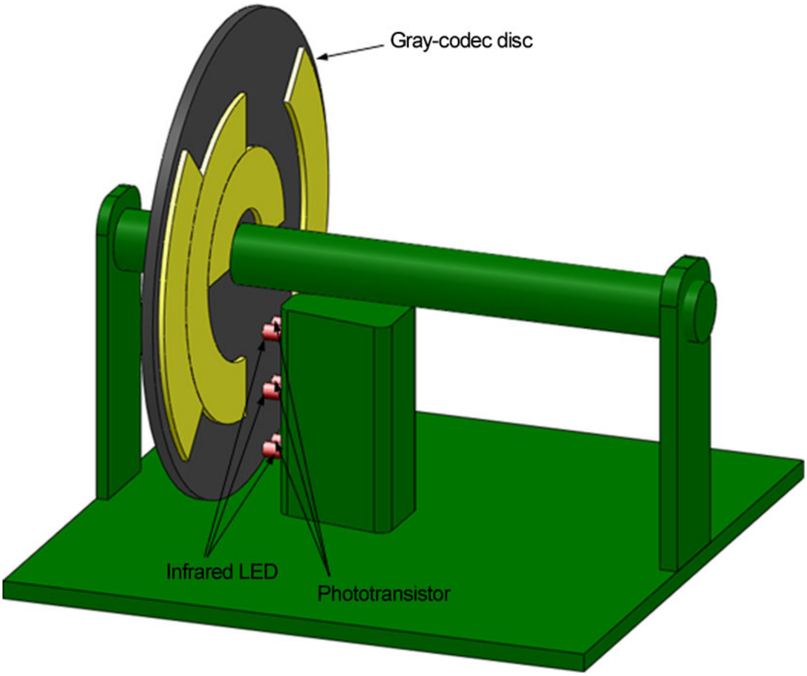

Fig. 3.35 Optotransducer

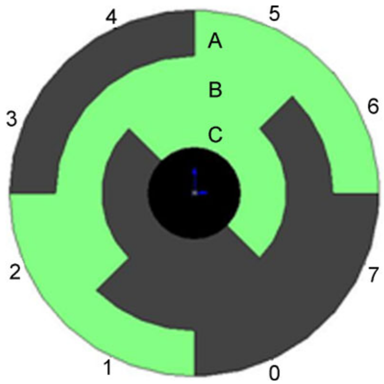

Fig. 3.36 Gray-coded disk

3.4.1.5 Strain Gauge

Strain gauge is a sensor that measures the strain of a material. It can thus be used as a displacement sensor of a position sensing device indirectly. Figure 3.38 shows the construction of a strain gauge, consisting of a grid of fine wire or semiconductor material bonded to a backing material.

Table 3.3 Gray-coded combination

Position	A	B	C
0	0	0	0
1	0	0	1
2	0	1	1
3	0	1	0
4	1	1	0
5	1	1	1
6	1	0	1
7	1	0	0

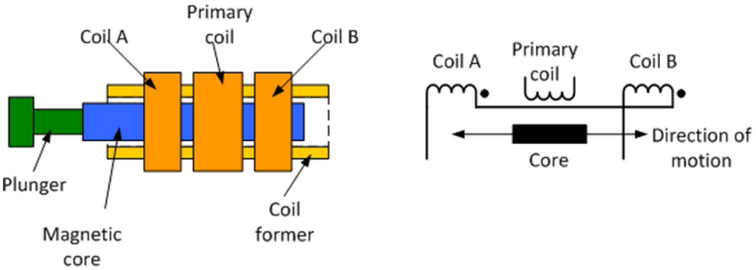

Fig. 3.37 LVDT

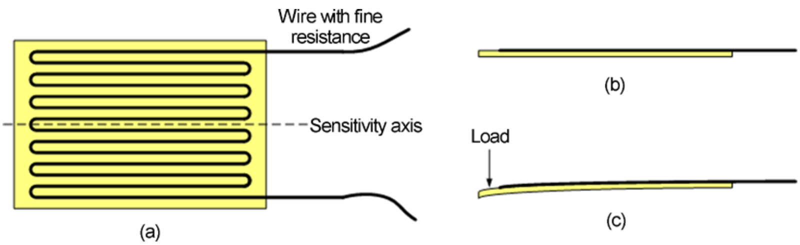

Fig. 3.38 Strain gauge; (**a**) top view, (**b**) side view when unloaded, (**c**) side view under loading

The operating principle of the strain gauge is as follows. The strain gauge unit is attached to the beam to be measured and is arranged so that the variation in length under loaded conditions is along the sensitive axis of the gauge. The loading on the beam increases the length of the gauge wire and also reduces its cross-sectional area. Both of these effects will increase the resistance of the wire.

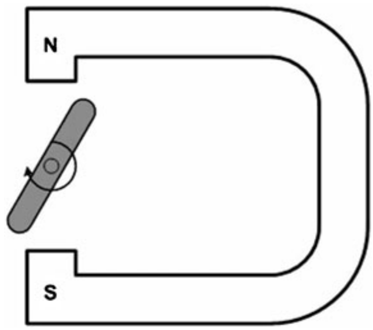

Fig. 3.39 Tachogenerator; with N indicating north pole and S indicating south pole

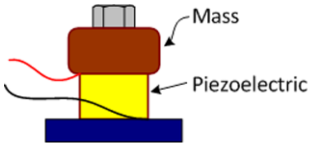

Fig. 3.40 Piezoelectric-actuated accelerometer

3.4.2 Velocity Measurement

The typical speed sensor in electric drives is tachogenerator, as presented in Figure 3.39.

Tachogenerator consists of a coil mounted in a magnetic field. When the coil rotates along with the shaft, electromagnetic induction results in an alternating induced emf across the coil. This emf is proportional to the rotational speed of the shaft and thus can be used as a measure of angular speed. Tachogenerator can typically measure up to 10,000rpm.

3.4.3 Acceleration Measurement

For applications where the acceleration measurement is critical, acceleration sensors can be installed in the drive systems. The typical acceleration sensor is the accelerometer, whose structure diagram of piezoelectric-actuated type is presented in Figure 3.40.

The main components of the piezoelectric-actuated accelerometer in Figure 3.40 are the piezoelectric crystal and the activation mass. The piezoelectric crystal is capable of converting mechanical displacement or force into electric charge. When the accelerometer is subject to an accelerated motion, *e.g.* vibration, the activation mass will compress and stretch the piezoelectric crystal due to the mass inertia. This force causes an electric charge to be generated (because of piezoelectric property).

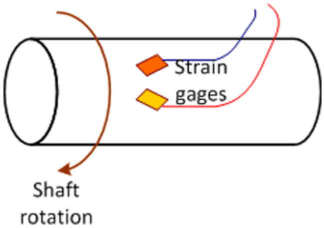

Fig. 3.41 Torque sensor

Based on the second Newton's law, the force is proportional to the acceleration, and therefore the electric charge is also proportional to the acceleration. By closing the measurement circuit and adding necessary amplifiers, an electric voltage will result.

While the abovementioned accelerometer is common, there are different types of accelerometer available as well. For example, the piezoelectric crystal may be replaced by a piezoresistive crystal or strain gauge, which has a better sensitivity for low-acceleration measurements.

3.4.3.1 Torque Measurement

Torque can be measured with strain gauges mounted on a rotating shaft, as presented in Figure 3.41. As the shaft rotates, deformation will take place on the shaft, and this will be measured by the strain gauges. A Wheatstone bridge that incorporates the strain gauges will convert the measurement into a calibrated signal. Because the deformation is proportional to the output torque, the calibrated signal is also proportional to the torque. This method of torque measurement is a direct measurement, since the torque is indeed measured via a circuitry of sensors.

Torque measurement can be done with non-contacting sensors or contacting sensors. Non-contacting sensors work based on radio telemetry communication between a stationary antenna in the frame and a loop antenna on the rotating shaft. The power from the loop antenna activates the strain gauges, allowing torque measurement to be conducted. The output signal is then retransmitted to the stationary antenna. When a contacting sensor is used, a slip ring is often used to assist signal transmission.

Torque can also be measured indirectly (inferred) by measuring the induced current of the motor. The direct relationship between current and torque allows this method to be used with satisfactory result. While this method is cost-effective, it is not as accurate as direct method with strain gauges. Furthermore, as implied by the above, it is only possible in the region where the current is proportional to the torque. Current measurement, however, can be used to infer a torque measurement when an accurate torque observation is not so critical.

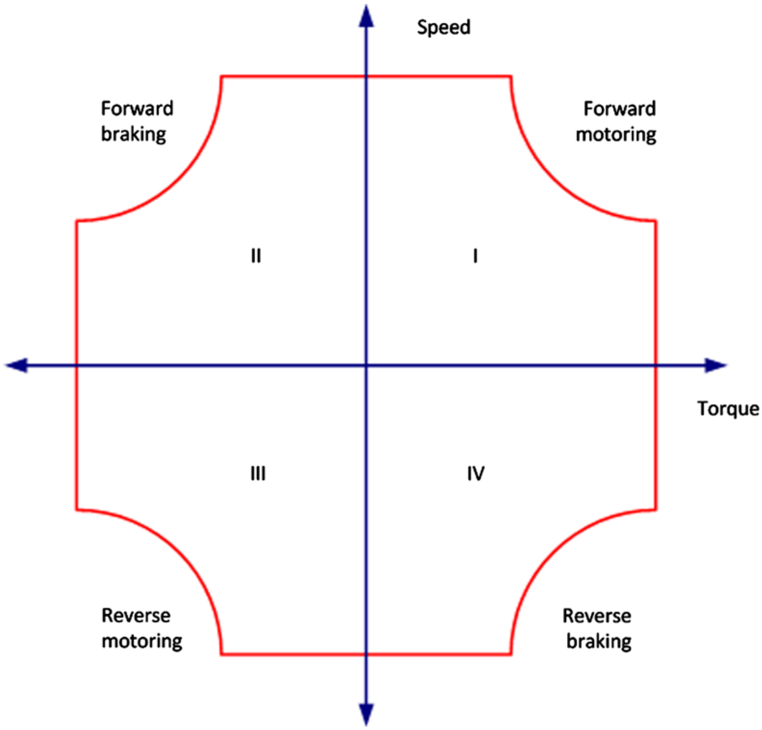

Fig. 3.42 Four quadrants operation of electric drives

3.5 Configuring an Electric Drive Application

Having discussed the actuators, sensors, and power moderators (power electronics), the configuration of these components to result in a complete system can now be discussed. Electric drives can be configured into various forms to fulfill different functions, depending on the desired applications and functions.

The modes of operation will therefore vary greatly according to the applications. Despite the numerous possibilities of tasks an electric drive may be required to perform, its operation can in general be classified into several classes as presented in Figure 3.42 as follows:

- forward motoring
- forward braking
- reverse motoring
- reverse braking

Operation in the motoring modes means that the drive draws power from the electrical supply and converts it to mechanical energy. Operation in the braking modes means that the drive produces power to an energy dissipation device or to a battery bank, hence working as a generator. Forward and backward operations are classified according to the direction of the drive.

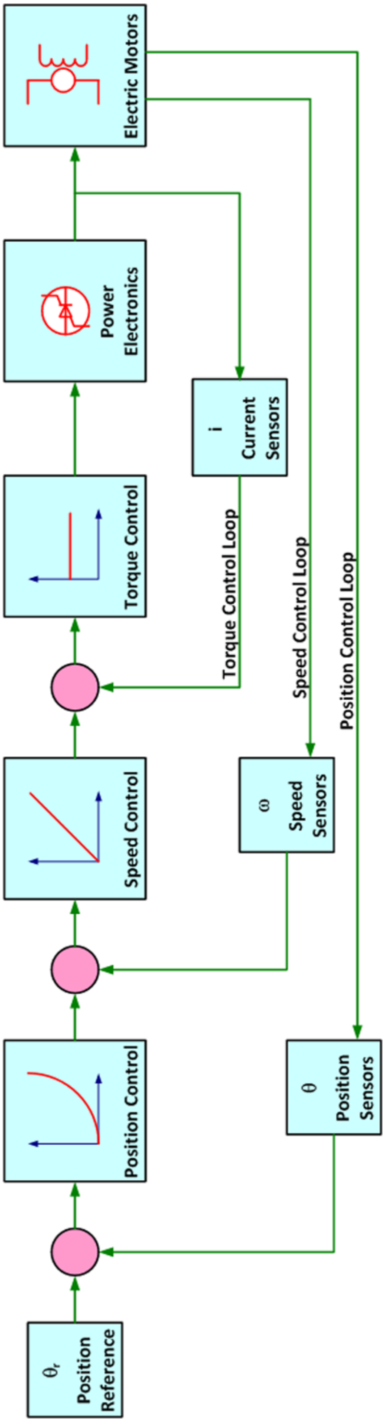

Fig. 3.43 Cascade control

In an application where the drive can operate in both motoring and braking modes, the control devices must be able to handle bidirectional flows of electrical current satisfactorily. Dynamic braking of a motor is accomplished with an external resistance to absorb the energy released by the motor during braking. The rotation of the motor under this operation mode causes voltage to be generated across the external resistance. The motor thus works as a generator and develops an opposing torque that will slow down the motor.

The type of components to be used for each specific application depends on the final objectives of the application. It is important to consider them collectively as a package when selecting/designing electric drives. The following factors should be considered:

- available electric power sources, including the type of electric current (AC or DC) and the amount of available power
- load requirements, including torque, speed, or power required by the load to perform the prescribed tasks
- characteristics of the controller, including its tasks, such as motor starting, stopping, overload protection, speed alteration/reversal, jogging, sequencing, and alarming
- environmental requirements, especially in terms of the reliability and safety of operations, including in harsh working conditions and in the presence of dangerous substance

The motion control variables of an electric drive mainly evolve around torque, speed, and position. Depending on the application, the requirements of those parameters may be stringent in terms of precision, response time, and robustness. To enable the drive system to perform according to the prescribed specifications, a closed-loop control system may be employed, equipped with necessary control schemes and sensors to provide feedback of the controlled variables. An example of a closed-loop control scheme utilizing simultaneous feedback of torque, speed, and position in a cascade control structure is shown in Figure 3.43.

Chapter 5 will address the motion control issues in more detail, for both the electric and hydraulic/pneumatic drives.

The final link from the electric drive to the load is via the mechanical transmission system. The role of mechanical transmission system is to convert the mechanical output of the electric motor into a useful mechanical action as is required to operate the load. In many applications, a smaller motor is desirable to yield a compact volume/size of the overall system. To match the motor performance to the torque and speed requirement of the load, a reduction gear head can be installed between the motor and the load to reduce the speed, hence increasing the torque. Apart from torque and speed adjustment, mechanical transmission is also employed to convert rotational motion into linear/reciprocating motion.

Chapter 4
Piezoelectric Drives

Piezoelectric actuators have received increasing attention in recent years along with the emergence of new technologies, such as nanotechnology and biotechnology, which require precision control in an unprecedented demand. Owing to many inherent merits of these actuators, such as high resolution of displacement, high stiffness, and fast response, piezoelectric actuators have been broadly used in many applications requiring fine precision control. The application of piezoelectric actuators is further fueled by the trend of miniaturization in applied research and in the industry nowadays.

The field of piezoelectric actuators is now an interesting subject of research worth spending millions of dollars annually. Because of the superior characteristics of piezoelectric actuators in terms of precision, the term piezoelectric actuator has been closely associated with high-precision actuators.

4.1 Solid-state Actuators and Piezoelectric Actuators

Piezoelectric actuators are a subset of solid-state actuators, which broadly refer to actuators that produce motion from their internal molecular action from a single solid body, without relative interaction with other bodies. Solid-state actuators are typically used for achieving motion of the order of nanometers. They include mechanisms of special transducer materials, such as magnetostrictive, electrostrictive, and piezoelectric materials.

Magnetostrictive materials have an ability to deform, *i.e.* to extend and to contract, in the presence of a magnetic field. These materials are often referred to as rare-earth materials, typically consisting of transition metals. Recently, magnetostrictive materials have been used as solid-state speakers, vibration tables, sensors, and actuators. The most appealing characteristic of this material is its lower hysteresis (which is an undesirable nonlinearity) compared to other high-precision transducers, which greatly relaxes actuation control. However, magnetostrictive materials are more sensitive to heat, and this makes them less preferable to, for example, electrostrictive materials.

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_4,

Electrostrictive materials have an ability to deform in the presence of an electric field. Electrostrictive materials are non-poled ceramics made from lead-magnesium-niobate, which have been applied in lifting mechanisms with high precision. Despite successful development of electrostrictive actuators and their low hysteresis, their quadratic, nonlinear relationship between voltage and deformation makes this type of actuator less popular than piezoelectric actuators.

Piezoelectric materials have an ability to deform in the presence of an electric field. Piezoelectric materials are poled ceramics made from lead-zirconium-titanate. The difference between piezoelectric and electrostrictive materials lies in the deformation process. In electrostrictive materials, an electric field separates the positively and negatively charged ions, expanding the materials. In piezoelectric materials, the electric charge excites the atoms of the materials according to the poling direction.

Piezoelectric actuators have become increasingly popular in many high-precision positioning applications, since they are able to provide very precise positioning (of the order of nanometer) and generate high forces (up to a few thousand Newtons).

Piezoelectric actuators have shown a high potential in applications that require manipulation within the range of sub-micrometer. Piezoelectric actuators have been applied in many areas, such as micro-electro-mechanical systems (MEMS), bioengineering, and nanotechnology.

The experiment conducted by the French' Curie brothers in the 1880s is widely considered as the birth of piezoelectricity. In their experiment, the Curies discovered that certain crystals generated an electric charge when they were subjected to a mechanical strain and, conversely, generated strain when they were subjected to an electric charge. It took several decades before this remarkable phenomenon was used commercially in ultrasonic submarine detectors developed during World War I. During World War II in the 1940s, scientists discovered that certain ceramics could be made piezoelectric by subjecting them in an electric field during their transformation phase. This boosted the implementation of piezoelectricity since the production cost could then be reduced. The birth of nanotechnology in the 1950s put the ultimate drive of its development and application.

Piezoelectric materials are currently used both as sensors and as actuators. They have the following beneficial characteristics that make them suitable for numerous high-precision applications:

- high resolution (unaffected by stiction and friction)
- not affected by magnetic fields
- low power consumption, since power is only required during motion and not during the holding phase
- clean operation compatibility
- fast response time

It is for these reasons that piezoelectric actuators are preferred to other solid-state actuators.

4.2 Piezoelectricity

The discovery of piezoelectricity is attributed to the research conducted by the French's Curie brothers in the 1880s. They discovered that certain crystals possess unique properties, which they called:

1. Piezoelectric effect
   This is an effect where a crystal becomes electrically polarized when a mechanical load is applied. This effect is useful in its application as a *sensor*.
2. Inverse piezoelectric effect
   This is an effect where a crystal becomes mechanically deformed when an electrical charge is applied. This effect is useful in its application as an *actuator*.

The analysis of piezoelectric effect is conducted based on the linear theory of piezoelectricity, where the equations of linear elasticity are coupled to the equation of electric charge by means of piezoelectric constants. A widely accepted approximation of the piezoelectric behaviour of a material is presented by a constitutive equation as follows [55]:

$$\sigma_{ij} = c^{E}_{ijkl} s_{kl} - e_{kij} V_k, \tag{4.1}$$

$$D_i = e_{ikl} s_{kl} + \epsilon^{S}_{ij} V_k, \tag{4.2}$$

where σ is the stress vector, s is the strain vector, V is the electric field vector, D is the electric displacement vector, c is the elastic stiffness constant matrix, e is the piezoelectric constant matrix, and ϵ is the permittivity constant matrix. Superscripts E and S denote constant electric field and constant strain, respectively.

In a piezoelectric material, stress/force, strain/displacement, and electric charge/voltage couple each other, as shown in (4.1) and (4.2). If an electric potential is applied across the electrodes of a piezoelectric material, there will be stress and strain induced in the material. In turn, this stress will lead to the build-up of an opposing electric charge which will affect the induced stress and strain of the material.

This iterated cause-effect phenomenon is evident from an autocorrelation plot of the output, as presented in Figure 4.1. This plot was obtained by measuring the displacement of a piezoelectric element in a relaxed position, *i.e.* only one end of the element fixed at a frame. Figure 4.1 shows that the small displacement variation contains a significant random portion and also a systematic component that can be attributed to the coupling of the various parameters in a piezoelectric material as explained.

There are three principle modes of deformation of a piezoelectric element, with regards to its direction of activation voltage and polarization. Figure 4.2 depicts the deformation of a piezoelectric element in cylindrical coordinate. The vectors E, P, and Δ denote the direction of activation voltage, polarization, and deformation, respectively. The hatched surface indicates the surface on which the electrodes are applied to the piezoelectric element.

In practical applications, however, it is sufficient to consider that upon activation a piezoelectric element will deform along all three principle directions. When, for example, a cylindrical coordinate is considered, the deformation can be depicted as

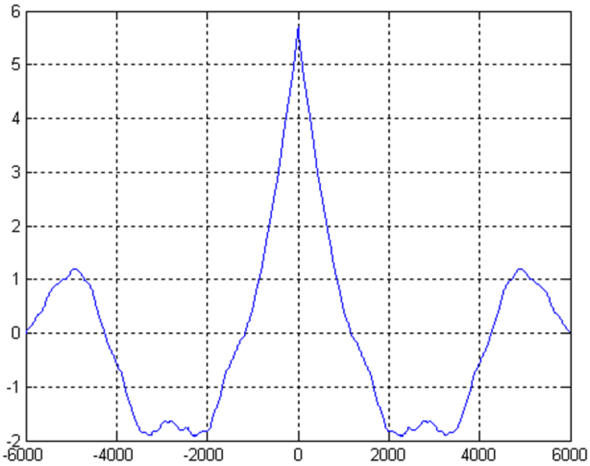

Fig. 4.1 Displacement spectrum showing iterated cause-effect phenomenon

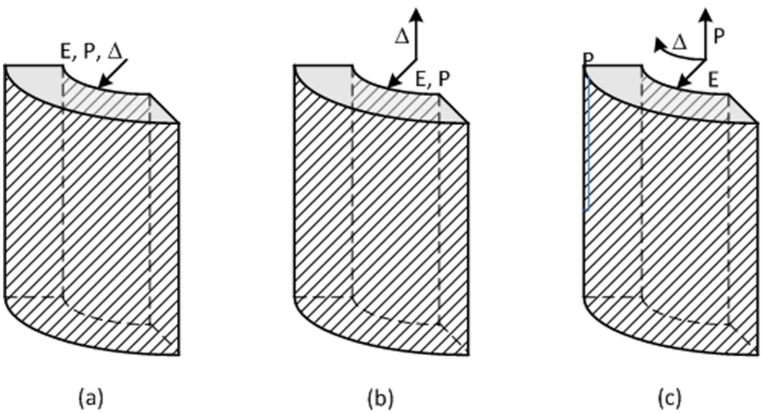

Fig. 4.2 Principle mode of deformation of piezoelectric element in cylindrical coordinate: (**a**) longitudinal, (**b**) transversal, (**c**) shear

in Figure 4.3. In Figure 4.3, the shadowed surface is fixed onto a platform, and $\hat{r}$, $\hat{\phi}$, and $\hat{z}$ denotes the three principle axes, with the activation voltage applied across the inner and outer surfaces.

4.3 Nonlinearity in Piezoelectric Actuators

Nonlinearity is known to occur and affect the performance of piezoelectric actuators. Because of their nonlinearity (which occurs between the displacement and the electric field), the response of piezoelectric actuators to an input becomes un-

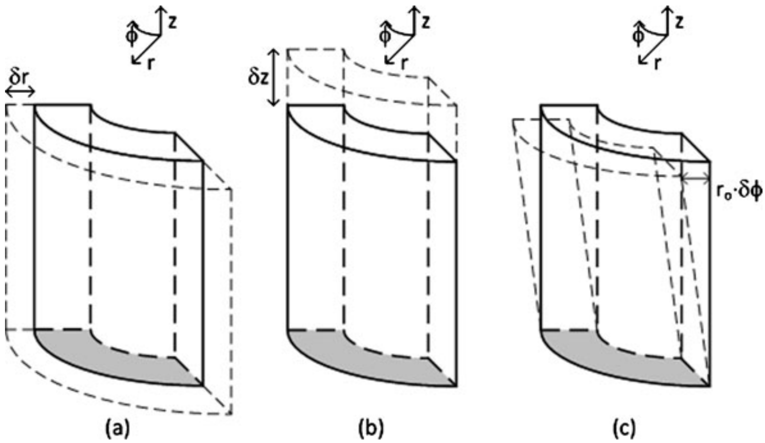

Fig. 4.3 Principle deformation of piezoelectric element in cylindrical coordinate: (**a**) radial, (**b**) axial, (**c**) tangential

predictable and uncontrollable. Among many types of nonlinearity, friction, force ripples, and hysteresis are the three most common types.

Friction and force ripples pose several difficulties to motion systems. Stiction, due to friction, for example, induces stick-slip motion. Limit cycle oscillations can also occur due to the discontinuous nature of the frictional force with respect to velocity. Force ripples produce "bumps" along the direction of motion, which will cause difficulties in achieving smooth, yet high-speed motion with linear control only. Model-based approaches have been proposed to compensate for friction and force ripples, such as with adaptive robust control scheme [138, 139], to achieve high-speed and high-accuracy motion control by compensating friction and force ripples [52]. Force ripple model is developed and identified with a force sensor, based on which a feedforward compensation component is designed [7].

Hysteresis is yet another type of nonlinearity in motion systems. Hysteresis is non-differential, multivalued, usually unknown, and commonly existing in physical systems, such as piezoelectric actuator [54]. The existence of hysteresis often severely limits the performance of a piezoelectric actuator, causing, among others, undesirable oscillation and instability. While its effect may be negligible in long-traveling motion systems, it significantly impedes the performance of short-traveling motion systems, such as direct-drive actuators. Hysteresis can be corrected using a large signal control [80], but it will lead to saturation and drift. Model-based approaches have also been proposed to compensate for hysteresis, such as feedforward and PID feedback controller [36].

The study and control of hysteresis require a mathematical model, with which the phenomenon is described quantitatively. While many models of hysteresis have been proposed, the hysteresis model developed by Preisach is widely used to model the hysteretic phenomena. Originally developed for magnetic materials, this model has been adjusted to suit piezoelectric materials [36]. Preisach's model is composed

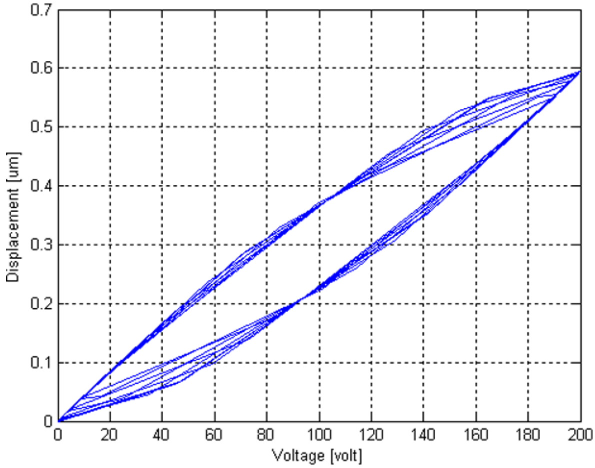

Fig. 4.4 Modeling of hysteresis with Preisach's model

of hysteresis operators $\gamma_{\alpha\beta}$, each of which exhibit one local memory hysteresis. The overall Preisach's model is expressed as follows:

$$x(t) = \int\int_{\alpha\geq\beta} \mu(\alpha,\beta)\gamma_{\alpha\beta}\big[u(t)\big]\,d\alpha\,d\beta, \tag{4.3}$$

where $x(t)$ is the output displacement of a piezoelectric actuator, $\mu(\alpha,\beta)$ is a weighing function, $u(t)$ is the input, and α and β represent the "up" and "down" switching values of the input.

For computational purposes, however, the following equations are used:

$$\begin{aligned} x(t) &= \sum_{k=1}^{n-1}\big[X(\alpha_k,\beta_{k-1}) - X(\alpha_k,\beta_k)\big] + X\big(u(t),\beta_{n-1}\big), \\ x(t) &= \sum_{k=1}^{n-1}\big[X(\alpha_k,\beta_{k-1}) - X(\alpha_k,\beta_k)\big] \\ &\quad + \big[X(\alpha_n,\beta_{n-1}) - X\big(\alpha_n,u(t)\big)\big], \end{aligned} \tag{4.4}$$

where $X(\alpha_k,\beta_{k-1})$ and $X(\alpha_k,\beta_k)$ are past input extrema. Figure 4.4 presents the hysteresis of piezoelectric actuators with Preisach's model.

There has been much research conducted with the objective of overcoming the adverse effects of hysteresis in piezoelectric actuators. This research can generally be classified into two categories:

1. Electrical approach

   The electrical approach concerns itself with using electrical techniques, *i.e.* hardware-based, as means to achieve the objective. This approach includes using electric charge, rather than electric voltage, to actuate the actuator [32, 33, 80].

This approach, however, is often impractical in real applications, as evident from the requirement of using less-available charge amplifier.

2. Control approach

   The control approach concerns itself with using control theory to determine the suitable activation signal such that the effect of hysteresis is overcome. This approach relies on mathematical computation, which is sometimes rigorous and intensive.

4.4 Mechanical Linkages for Piezoelectric Drives

Linkages are required to connect the prime movers (such as piezoelectric actuators) to the loads in a system in such a way that the output of the prime movers can drive the loads according to the requirements. Sliders, universal joints, and gears are examples of conventional linkages that are commonly used in drive systems. Their clearance and tolerance, however, which are integral to their working principle and viability, render them inapplicable in applications with sub-micrometer level precision. A kinematic chain of such joints makes matters worse, resulting in poorer accuracy and repeatability.

Therefore, piezoelectric drives, whose precision is of the order of sub-micrometer, require non-conventional linkages to suit the high precision they are required to perform.

For applications with high precision requirements where micrometer level tolerance is unacceptable, flexible joints, commonly known as flexures, offer an alternative to conventional mechanisms. This technology has been around for the past 50 years. Flexures utilize the compliance of a material in order to allow small transmission of motion, thereby eliminating the presence of backlash, which is a hindrance for high-precision motion. Furthermore, the use of flexures prevents friction and wear and allows sub-micrometer accuracy due to their monolithic construction [118].

Flexures are typically formed by one or a combination of two shapes as follows:

- notch
- beam

Notch joint is formed by a recess, often symmetrical, in a solid bar, so that a small rotational motion is allowed around the recess. Today, it is widely used for high-precision mechanism with small displacement [88]. Beam joint, typically applied as leaf spring, provides flexible translational motion by virtue of deflection over its length. It has been applied in many high-precision motion stages, including in medical instrumentation [107] and MEMS devices [95].

These shapes are often combined in assemblies and are most commonly used as revolute joints, universal joints, or parallel four-bar translational joints. Most commercially available flexures nowadays, however, are a combination of these joints, with the addition of a wide variety of connections to suit particular application needs.

Flexures, however, suffer from drawbacks that are relatively non-existent in conventional joints. These drawbacks are inherent from the structure of the flexures themselves, which completely rely on their rigidity and absence of links. These drawbacks are as follows:

1. Limited range of motion

   All flexures are limited to a finite range of motion, while their rigid counterparts rotate infinitely or translate long distances. The range of motion of a flexible joint is limited by the permissible stresses and strains in the material. When the yield stress is reached, elastic deformation becomes plastic, after which, joint behaviour is unstable and unpredictable. Therefore, the range of motion is determined by both the material and geometry of the joint.
2. Axis drift

   In addition to limited range of motion, most flexure joints also undergo imprecise motion referred to as axis drift or parasitic motion. For notch-type joints, the centre of rotation does not remain fixed with respect to the links it connects. With translational flexures, there can be considerable deviation from the axis of straight-line motion. For example, a simple four-bar leaf spring experiences curvilinear motion. Axis drift can be improved by adding symmetry to the design of a joint. However, this often increases the stiffness of the joint in the desired direction of motion. Furthermore, more space is required to accommodate any symmetric joint components. Having minimal axis drift is essential to preserving the kinematics of the original mechanism when doing conventional joint replacement with flexures. Naturally, it is desired to have a low axis drift for high precision.
3. Off-axis stiffness

   While most flexure joints deliver some degree of compliance in the desired direction, they typically suffer from low rotational and translational stiffness in other directions. A high ratio of off-axis to axial stiffness is considered a key characteristic of an effective compliant joint.
4. Stress concentration

   Most notch-type joints have areas of reduced cross section through which their primary deflection occurs. Depending on the shape of these reduced cross sections, the joints may be prone to high stress concentrations and hence a poor fatigue life.

With respect to the nature of motion, flexures are classified as follows (in which their names are self explanatory):

- rotational joints
- translational joints

Rotational joints are more numerous and more commonly used. Many other mechanisms using translation joints include sliders, rails, and linear bearings.

In the following sections, various types of flexures will be elaborated.

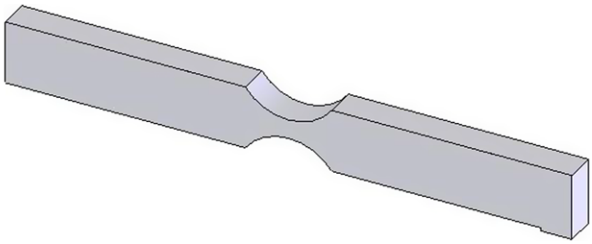

Fig. 4.5 Notch joints

4.4.1 Notch Joints

The basic shape of notch joints is presented in Figure 4.5. The main feature of a notch joint is a shape recess that allows small movement between one end to the other.

The range of motion highly depends on the stiffness of the material and the shape of the notch. Based on the relationship between the range of motion and notch parameters, a method for optimizing the performance of notch joints can be devised by considering the following inputs:

- material properties, such as modulus of elasticity
- notch joint geometry, such as joint length, thickness, and notch radius
- safety factor, which depends on the application of the notch joint

The outputs are as follows:

- notch joint stress
- rotational stiffness

Simple comparison and analysis can be generated by simulating many conditions in which for every analysis, only one parameter is varied.

The above list can be expanded based on the selected notch equations to include more parameters, such as off-axis stiffness or off-axis stress to further evaluate notch joints.

The shape of the joint, in particular its notch radius, affects the stiffness, stress concentration, and range of motion. More round notches provide higher stiffness and lower stress concentration despite smaller range of motion. Elliptical notches with appropriate radius are designed to provide acceptable stiffness and stress concentration, with a desirable range of motion.

4.4.2 Cross-strip Pivot and Cartwheel Hinge

Cross-strip pivot has been around since 1949. The advantage of using cross-strip pivot is its considerably wider range of motion (compared to notch joint); while its

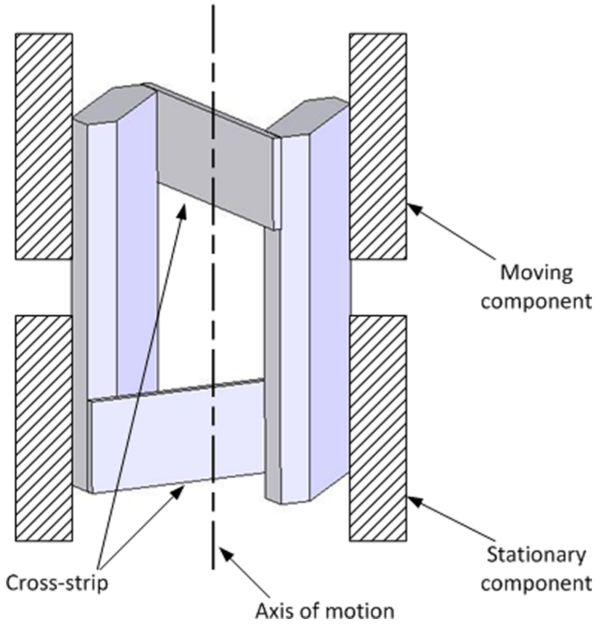

Fig. 4.6 Cross-strip pivot

disadvantage is its quite large axis drift (which even gets larger as the deflection increases).

The basic construction of cross-strip pivot is presented in Figure 4.6. There are two non-intersecting plates perpendicular to each other that connect two outer blocks. These pairs of plates and blocks constitute the main structure of cross-strip pivot. One end of the pivot is attached to the stationary component of the joint, while the other end is to the moving component.

Cartwheel hinge is a simple modification of cross-strip pivot, where the two strips of the joint are now made intersecting. This is to reduce the axis drift of the structure. While it performs better mechanically, it imposes more difficulty to the manufacturing.

Compared to cross-strip pivot, cartwheel hinge has the following characteristics:

- higher stiffness
- smaller range of motion
- smaller axis drift

The basic construction of cartwheel hinge is presented in Figure 4.7. As mentioned, the two plates intersect (compared to cross-strip joint) each other, while the two blocks take the same shape and position. The position of the stationary and moving components is similar to cross-strip joint. The shape of the block can also be made such that it matches the shape of the stationary and moving components to give better transmission of motion.

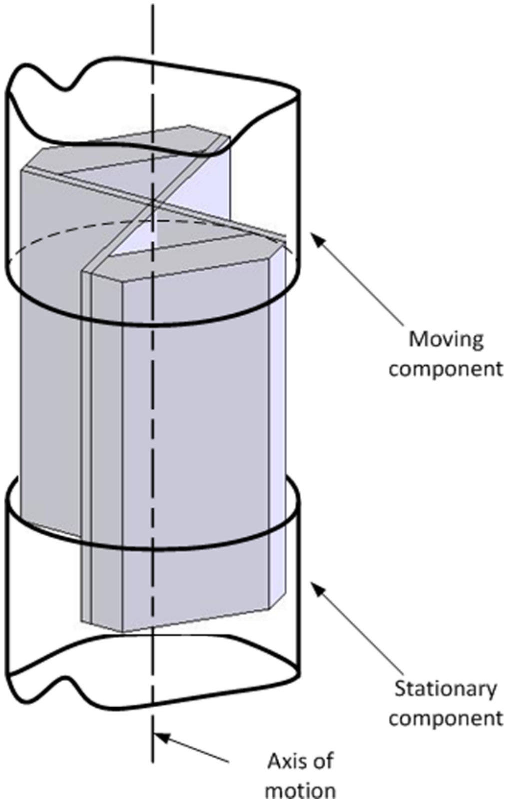

Fig. 4.7 Cartwheel hinge

4.4.3 Passive Joints

Passive joints are contact/sliding joints, which can be thought of as force-closed conventional hinge joints. They are easy to design but not ideal since the contact, which is particular to their working principle, causes friction. Furthermore, the kinematics of the joint may change along with the relative motion, often quite significantly from the original design. They can, however, greatly increase load carrying capability in some applications.

There are several examples of passive joints as follows:

1. Q-joint

   Q-joint, or quadrilateral-joint, is a parallelogram that is formed to constrain the angles of opposite links to be equal, thus transmitting at an equal angle across the joint.
2. Torsional hinges

   The difference between torsional and revolute joints is the characteristics of the compliance. Torsion-based joints are considered to have more distributed compliance than bending/revolute-based joints. Not only can rectangular beams serve as torsion joints, but square, circular, and hollow cross-sections—also known as closed shell—can also be used. An open shell, on the other hand, is formed by cutting a slit lengthwise along a hollow beam, resulting in a very low compliance while maintaining the high off-axis stiffness of a closed shell.

3. Split-tube joint

   The split-tube joint has an off-axis stiffness of a cylinder and a very little torsional stiffness. Furthermore, it has almost no centre of rotation drift when the connected links are fixed along the line of the centre of rotation. This joint offers the off-axis stiffness of a solid circular tube while having a low torsional stiffness. Perfect rigidity, however, would require infinitely thin line contact between the connecting link and the tube. Furthermore, this joint exhibits a trade-off between the range of motion and the off-axis stiffness. Under large displacements, the gap separation increases, and the tube warps out of circular shape, reducing the off-axis stiffness.

4. Disc couplings

   Disc couplings are three degree-of-freedom joints, usually used as compliant replacements to traditional universal joints, while also adding an axial degree-of-freedom. The purpose of these joints is to transmit torque from one shaft to another, even when the shafts connect at an angle. The axial degree-of-freedom allows for some play during assembly of a system and is also useful for self-alignment applications. Torque transmission performance is often rated in terms of energy efficiency.

5. Rotationally symmetric leaf-type hinge

   The joint can be created simply by machining notches in the sides of a tube.

4.4.4 Compliant Revolute Joint

A compliant revolute joint maintains zero axis-drift by virtue of two joint holders at the two ends of the flexible cross bar. It therefore has a large range of motion with a high ratio of off-axis stiffness to stiffness in the desired direction of motion.

The structure of a compliant revolute joint is presented in Figure 4.8, whose main structure is a cross bar. A fillet, however, is usually incorporated in the design to reduce the stress concentration and to increase the range of motion. Stress concentration is a consequence of the fixed joint at the two ends of the cross bar. The addition of a fillet comes with another consequence of a small increase in the torsional stiffness.

4.4.5 Compliant Translational Joint

A compliant translational joint is an improvement to other translational flexures. This joint uses redundancy to attain high off-axis stiffness ratios and zero axis drift.

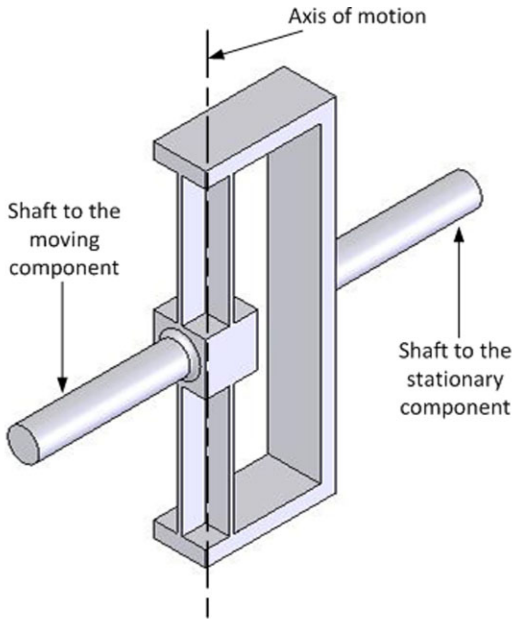

Fig. 4.8 Compliant revolute joint

The benefits of using compliant translational joint are as follows:

- larger range of motion
- more consistency of motion (in terms of parallelism)
- less concentration of stress, which increases the life time of the joint

The range of motion of the compliant joint is as follows:

$$x = \frac{2L^2\sigma}{3tE}, \tag{4.5}$$

where x is the range of motion, L is the length of the beam, t is the thickness of the beam, and σ and E are the shear stress and the elasticity of the beam material, respectively.

4.5 Example of Application

Now that the working principle, control, and mechanism of employing piezoelectric actuators have been explained, an application example of piezoelectric actuators is presented in this section to illustrate how piezoelectric actuation system is designed to execute a task.

This section discusses self-sensing actuation (SSA)—in which piezoelectric actuators find their superiority—to facilitate the implementation of a piezoelectric actuator in an intelligent mechatronic system. SSA is a technique to employ smart materials, such as piezoelectric materials, simultaneously as a sensor and an actuator; thereby increasing the level of integration of the system. The piezoelectric actuator is equipped with an exclusive adaptive controller amidst its nonlinearity and system's disturbance. The application area to be discussed is a microdispensing system, which is an example of a micromanufacturing process, combining a fluidic system and a positioning system.

A microdispensing system is a system to dispense fluid in a minute volume and in a precise manner, of the order of microliters. Because of the high demand of microdispensing systems and the wide variety of dispensing conditions, there have been many proposals and designs for microdispensing device. Various techniques have been employed, which in principle can be categorized into three techniques as follows [14]:

- time pressure
- rotary screw (continuous motion)
- positive displacement

Piezoelectric actuators are commonly used in this application for their ability to achieve high-precision motion of the order of sub-micrometer [9, 140]. The current achievement in microdispensing field is 25μm [44].

Despite the success of various existing microdispensing devices, production cost has been a persistent problem, especially when high-quality dispensing is desired and a variety of droplet dimension is required. The current practice is to use one unit of piezoelectric-coated injector to dispense one dimension of droplets, leading to the requirement of using many piezoelectric-coated injectors for various dimension requirements.

The objective of the proposed system is a microdispensing system with low production and maintenance cost, with the same level of precision and size of the existing system. The underlying idea is to use the same piezoelectric actuator and to only change the injector to obtain different size of droplets. In addition, drop-on-demand (DOD) [109] approach is applied in the proposed device, where one liquid droplet is formed by each movement of the piezoelectric actuator. This technique is generally preferred for its ability to form non-continuous shape/pattern.

The proposed microdispensing system employs a contacting method to dispense the liquid to the work piece. This method relies on several factors, such as the adhesive force between the liquid droplet and the work piece and their contact time. Figure 4.9 presents the structure of the proposed microdispensing system using a piezoelectric actuator. A flexural lever structure is employed as a motion modifier, where it can amplify the distance of motion of the injector several times of the actuator motion, depending on the setting of the amplification factor. The injector delivers the liquid onto the surface of the work piece, thereby delivering the final results.

The dimension of the droplet is therefore controllable via two settings as follows:

- the activation voltage of the piezoelectric actuator, which determines the contact time in a relatively small range
- the amplification factor of the flexural lever; which determines the contact time in a larger range

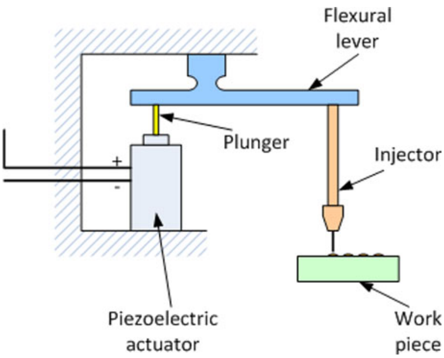

Fig. 4.9 Working principle of the microdispensing system

This arrangement eliminates the necessity of using several piezoelectric actuators for different dimensions of droplets, which effectively reduces the installation cost of the microdispensing system.

The dispensing system is attached to an X–Y table to allow pattern formation or trajectory tracking. Therefore, the work piece is maintained stationary while the injector moves according to the desired trajectory.

Proportional-integral-differential (PID) control system is incorporated to control the X–Y table. Each of the axes is controlled by separate controller and drive, realizing a distributed control system scheme. Although the focus of this description is in the control of the piezoelectric actuator, the motion of the X–Y table is also taken care of in the actual system. As a matter of fact, the controller architecture also includes a synchronization between the motion of the X-axis and Y-axis of the table and the Z-axis of the microdispensing device for optimum results. Figure 4.10 presents the complete architecture of the proposed microdispensing system.

The model of the system is given by the following equation:

$$\ddot{x} = a\dot{x} + bu + d(t), \tag{4.6}$$

where x is the position, u is the control signal, and $d(t)$ is the disturbance or system uncertainty. The parameters a and b are unknown and are to be identified. The initial identification is accomplished off-line by studying the response of the piezoelectric actuator and comparing it with a second-order model as given in (4.6), considering it having a high damping ratio and a high natural frequency. From the off-line identification, the values of a and b are obtained as -4.2×10^3 and 7.5×10^4, respectively. Figure 4.11 presents the step response of the actuator and the model, showing a satisfactory model.

The off-line identification of the parameters of the system is then used as the initial value for the adaptation of parameters.

The structure of the control system is presented in Figure 4.12. The PID control is tuned using the linear quadratic regulator (LQR) method to achieve a satisfactory tracking error and is mainly aimed to cater for the linear portion of the system. The feedforward control is an adaptive control law used to deal with system disturbance or uncertainty.

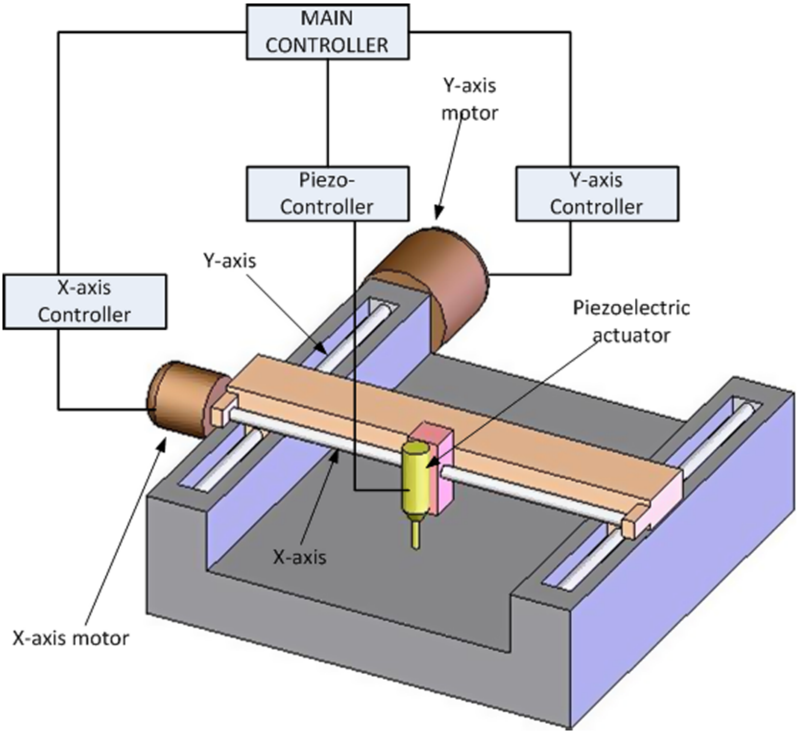

Fig. 4.10 Design of microdispensing system and its control system

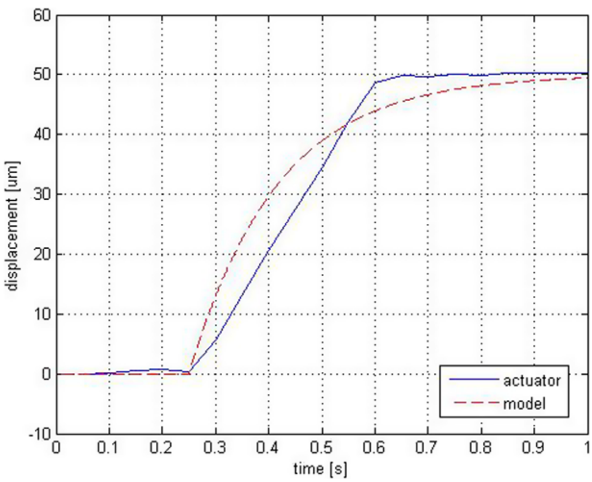

Fig. 4.11 Step response of the actuator and the model

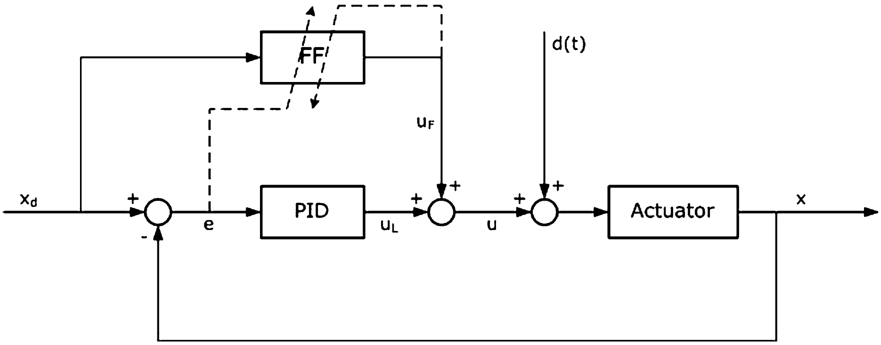

Fig. 4.12 Structure of adaptive feedforward control

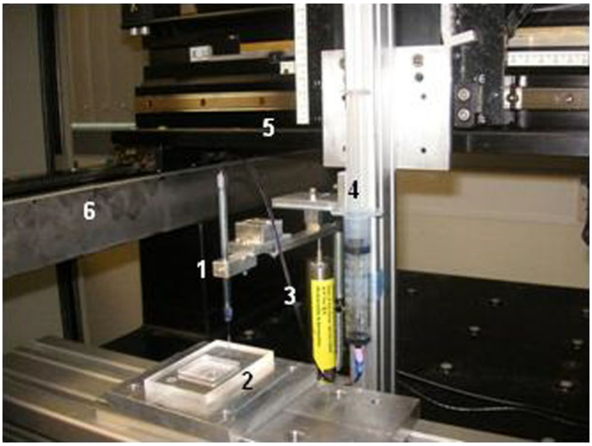

Fig. 4.13 Microdispensing installation in X–Y table; (*1*) injector, (*2*) work piece, (*3*) piezoelectric actuator, (*4*) liquid container, (*5*) X-axis, (*6*) Y-axis

The feedforward adaptive control is designed as follows (referring to Figure 4.12):

$$u_f = \frac{\ddot{x}_d - \hat{a}\dot{x}_d}{\hat{b}}. \tag{4.7}$$

Define the system error $e = x - x_d$, where x_d is the reference signal. The parameters $\hat{a}$ and $\hat{b}$ are obtained in an adaptive manner by the following on-line algorithm:

$$\dot{\hat{a}} = g_1 \dot{x} e, \tag{4.8}$$

$$\dot{\hat{b}} = g_2 u e, \tag{4.9}$$

where g_1 and g_2 represent the adaptation rates. In the adaptive algorithm, the values of a and b, which have been obtained from off-line identification, are used as the

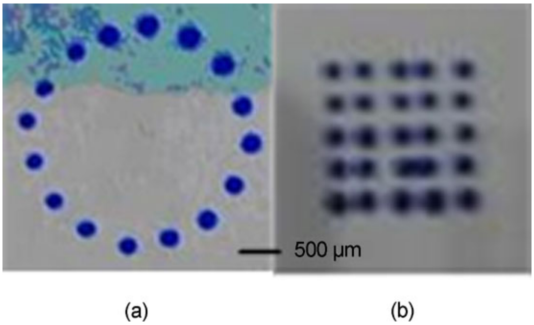

Fig. 4.14 Patterns of droplets; (**a**) circle pattern, (**b**) array pattern

initial values for $\hat{a}$ and $\hat{b}$. The control signal of the proposed PID with adaptive control can then be formulated as

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau)\, d\tau + K_d \frac{d}{dt} e(t) + \frac{\ddot{x}_d(t) - \hat{a}\dot{x}_d}{\hat{b}}, \tag{4.10}$$

where K_p, K_i, and K_d are the proportional, integral, and derivative gains of the PID controller, respectively.

Figure 4.13 presents the installation, which serves as an experimental setup for the proposed microdispensing system application. The dispensing system itself comprises of the piezoelectric actuator (3), a liquid container (4), and an injector (1). The piezoelectric actuator and the injector are connected via a delicate lever system, which is not visible in Figure 4.13. The dispensing system is placed on an X–Y table and is moved in a horizontal plane relative to a fixed work piece.

As a final note, the experiment of the microdispensing system was conducted to form an array of droplets, and in addition to that some trajectories, such as letters or geometrical shapes. Figure 4.14 presents patterns of droplets that show the viability of the proposed device to produce droplets. A tube with an inner diameter of 0.003 in (0.0762mm) was used as the dispensing tip, with a length of 10mm. The average dimension of the droplets is 247.45μm, with a standard deviation of 9%, which is generally considered satisfactory.

Chapter 5
Control System in Servo Drives

The core of a servo drive is the control system which orchestrates and links the different components to yield the highest level of performance. This chapter will address the control issues in a servo drive, such as various classifications and implementations of control systems, challenges in servo control of hydraulic and electric drives, common servo control structures, and typical control software standards.

5.1 Open-loop Versus Closed-loop Control

Up to the 1940s, sensors and actuators were standalone devices. There were neither signal transmission capabilities nor standards to allow the devices to communicate with each other. Even in the early preceding years following the birth of the transistor and the rise of electronics, sensors were still pure measurement devices. The users would manually establish a measurement connection, take measurement, and record it into a logbook or datasheet. The measurement data from the sensor ended in the sensor itself. Further and subsequent use of the data cannot be enabled by any function of the sensor.

Thus, control systems until then were manual-based, where the control operators would take a reading visually by looking at the display of the measurement instrument and then would manually adjust some actuators to bring the control variables to the desired levels. This form of control is called open-loop control, where the operators become the link between the sensors and the actuators, acting as the controller.

Open-loop control requires much experience and skills of the control operators to achieve good control performance, since the only form of feedback is via the operators. Such a form of control is seriously inadequate in applications with a huge number of control loops where it is simply inefficient to have the operators to tune individual loop.

The first innovation to move away from open-loop control was to simply retrofit a sensor with a relay at the output, so that the relay can be used for on–off control

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_5,

of the system. This was to become the early form of closed-loop control. However, it was the advances in analog transmission standards from the 1940s, such as the 3–15psi and 4–20mA standards, which provided the key steps to realize truly closed-loop control systems, where it is the system itself that acts as the controller. With different components of the control system adhering to the same transmission standards, it has become possible to integrate all of the components via a computer-based controller.

5.2 Servo Control Challenges

A high-performance servo control system has to adequately address several issues. This section will highlight the significant challenges facing the servo control system.

5.2.1 System Design

The achievable performance of a servo drive crucially depends on the overall design of the drive and the synergy in the integration of the various constituent components. A servo drive is a mechatronic system, comprising of mechanical, electrical, and control components, where the control system fulfills the task of a conductor, orchestrating the other components to work in synchronization. However, regardless of the quality of this conductor, the bottleneck to the overall achievable performance lies with the quality of the other members in the orchestra, namely the mechanical and electronic components.

5.2.1.1 Mechanical Components

The mechanical components will have to be properly sized and manufactured for a specific application. The motor should be able to sustain the maximum and continuous force/torque required by the load. The design of the mechanical structure should consider the kinematics and dynamics of the components to realize a stiff and repeatable motion platform. The assembly of these components is crucial as well. Geometrical errors often arise not just due to the constraints of manufacturing mechanical components precisely according to specifications, but also dependent on how each part is aligned and assembled in the right place at the highest precision. For example, in an XYZ cartesian robot, there will be 21 sources of geometrical errors arising from linear errors, angular errors, straightness and orthogonality errors. No matter how accurate the measurement system is, these geometrical errors will lead to positioning errors. They have to be reduced via improved manufacturing and assembly process or via a soft compensation by calibrating these errors with a laser interferometer.

5.2.1.2 Electronic Components

Similar to the mechanical components, the drive electronics will have to be properly sized and designed. The sampling frequency has to be sufficiently high to support the control bandwidth and avoid aliasing, and also to be able to resolve the intermediate positions when the load moves at the minimum velocity specified. The bandwidth of the control electronics will have to be sufficient to receive the encoder counts when the load is driven at the maximum velocity. One of the important issues facing the electronics of a servo drive is the requirement to keep the measurement and transmission noise to the minimum possible level. Proper cable shielding and grounding techniques should be used to reduce the influence of electromagnetic effects which will increase the noise level.

5.2.2 Nonlinear Dynamics

Control systems have to deal with nonlinearity which always exists in different elements of the system, in the sensors, actuators, and even the controller. In many cases when a high level in the control performance is required, these perturbations to the control system may be beyond tolerance threshold, and remedial measures are necessary. In this section, common types of nonlinearity present in a hydraulic/pneumatic or electric servo drive will be highlighted, with their compensation duly discussed.

5.2.2.1 Sensor Characteristics

Sensors are mostly nonlinear in the relationship between the output and the input of the device. The reason for this is that the sensors are often not measuring directly the quantity of interest, but rather an indirect quantity which is related to it. In what follows, examples of nonlinearities associated with some sensors will be highlighted.

1. Thermocouples

   The temperature sensor based on thermocouple is one example. In the thermocouple, a voltage is measured which is dependent on the temperature. The relationship between voltage and temperature is a nonlinear one as given in the following equation:

   $$v = a + b\tau + c\tau^2 + \cdots, \tag{5.1}$$

   where v is the voltage, τ is the temperature, and a, b, c, *etc.*, are constants. A calibration table tracking the relationship of the measured and actual variable is often provided along with the sensor. Many controllers are indeed fitted with such a calibration table so that one can linearize these signals. In this way, the controller algorithm really does obtain a calibrated signal which is proportional to temperature.

2. Flow sensor

   Flow is commonly measured by means of a differential pressure sensor. The pressure difference is not a linear function of flow, but the pressure is proportional to the square of the flow rate, given by the following equation:

$$\Delta p = kq^2, \tag{5.2}$$

   where Δp is the differential pressure, q is the volumetric flow, and k is a constant.

   In order to get a linear relationship between the input and output of the sensor, the square root of the signal has to be taken before it reaches the control algorithm. The sensors themselves are sometimes provided with a square root algorithm, so that the signals from the device are already linear. In order to cover those cases where the sensors themselves do not linearize the signal, most controllers have the ability to take the square root of the controlled variable.

5.2.2.2 Friction

Friction is inevitably present in any servo drive. It manifests itself prominently in the control valves and moving mechanisms of the servo drive.

1. Control valves

   In the hydraulic drive, friction in control valves is one of the major reasons for poor servo control performance. A valve with high friction causes stick-slip motion and oscillations in the control loop. In all sorts of valves, friction appears in the packing boxes around the valve stem, especially when they are tightened. In ball valves, ball segment valves, and throttle valves there is often also significant friction between the ball/throttle and the seat. The problem is that the static friction mostly increases gradually during operation. The packing boxes are, for example, often tightened after some period of operation in order to avoid leakage.

   Friction is varying, both in time and between different operating points. Temperature variations also cause friction variations. A high temperature means that the material expands and therefore that the frictional force increases. Some media give fouling that also increases the friction. Particles in the media may cause damage on the valve. The wear is often non-uniform, so that the friction is different at different valve positions.

   If the valve friction has become so high that stick-slip motion occurs, maintenance should be undertaken. The problem is that there are several possible causes of oscillations in control loops, and inappropriate action is often taken. Valve friction is the most likely reason, but it may also be external disturbances or poor controller tuning that causes the oscillations.

   Many users assume that oscillations in control loops are caused by bad controller tuning, and therefore they detune the controllers. This is not the appropriate action when valve friction is the cause. Most adaptive controllers behave in the same way. Oscillating disturbances with frequencies in the neighbourhood of the ultimate frequency will detune most adaptive controllers, since the adaptive controllers interpret the oscillations as a result of an excessive loop gain.

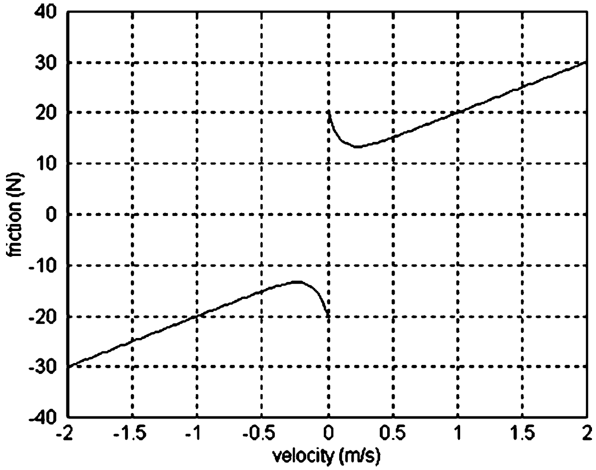

Fig. 5.1 Tustin friction model

2. Mechanical transmission

   In electric or hydraulic motors and cylinders, friction will feature at the interface of all moving mechanisms where the power generated is transformed into mechanical motion. In this case, several characteristic properties of friction have been observed, which can be broken down into two categories: static and dynamic. The static characteristics of friction, including the stiction friction, the kinetic force, the viscous force, and the Stribeck effect, are functions of steady-state velocity. The dynamic phenomena include pre-sliding displacement, varying breakaway force, and frictional lag. Many empirical friction models have been developed which attempt to capture specific components of observed friction behaviour, but generally it is acknowledged that a precise and accurate friction model is difficult to be obtained in an explicit form, especially for the dynamical component. A common friction model used in this case is the Tustin friction model, shown in Figure 5.1.

   Friction arising from mechanical transmission can be reduced using more efficient transmission mechanisms such as air and magnetic bearings.

5.2.2.3 Force Ripples

In almost all variations of electric motors where a ferromagnetic core is used for the windings, force ripples exist. The two primary components of the force ripple are the cogging (or detent) force and the reluctance force. The cogging force arises as a result of the mutual attraction between the magnets and iron cores of the translator/rotor. This force exists even in the absence of any winding current, and it exhibits a periodic relationship with respect to the position of the translator/rotor relative to the magnets. Cogging manifests itself by the tendency of the translator/rotor to align

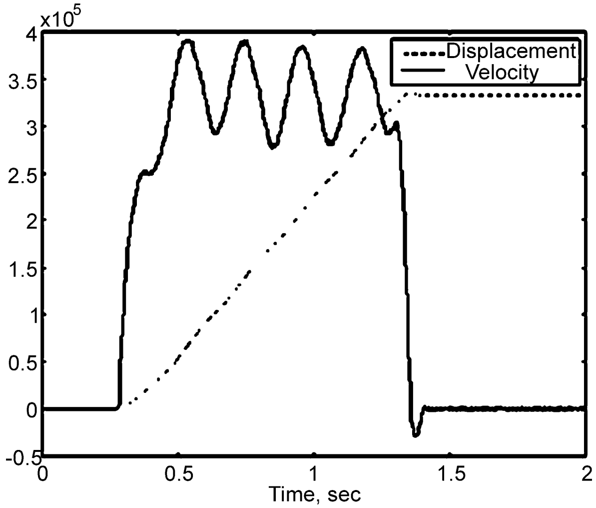

Fig. 5.2 Force ripple over a displacement ramp

in a number of preferred positions regardless of excitation states. There are two potential causes of the periodic cogging force in a permanent magnet motor, resulting from the slotting and the finite length of iron-core translator. The reluctance force is due to the variation of the self-inductance of the windings with respect to the relative position between the translator/rotor and the magnets. Thus, the reluctance force also has a periodic relationship with the translator/rotor-magnet position.

Collectively, the cogging and reluctant force constitutes the overall force ripple phenomenon. Even when the motor is not powered, force ripples are clearly existent when the translator/rotor is moved or rotated. There are discrete points where minimum/maximum resistance is experienced. At a lower velocity, the rippling effects are more fully evident due to the lower momentum available to overcome the magnetic resistance.

The force ripple has significant effects on the position accuracy achievable, and it may also cause oscillations and yield stability problems, particularly at low velocities or with a light load (low momentum). Figure 5.2 shows the real-time displacement ramp response of a tubular-type permanent magnet linear motor where the force ripples are especially evident over the portion with constant velocity.

5.2.2.4 Hysteresis and Backlash

Poorly machined gears or loose fitting linkages exhibit backlash. This manifests itself as a differing input/output relationship according to the direction of movement of the input shaft. Stiction and friction can also cause a similar effect. This can be visualized as in Figure 5.3 and is known in control engineering as hysteresis or backlash.

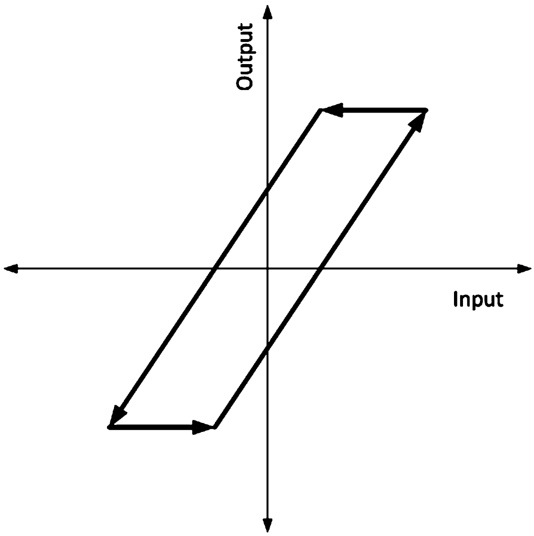

Fig. 5.3 Hysteresis and backlash

Hysteresis can be a source of problems and often manifests itself as a dither about the set point as the control system hunts in the dead band. It can also be self-reinforcing, as dither will lead to more wear and more backlash.

Hysteresis is best handled by design and careful manufacturing techniques such as spring-loaded gears and pre-tensioned linkages.

5.2.2.5 Saturation

The signals in control loops are always limited. The measuring sensors have their own working ranges. If the measured quantity falls outside the measurement range of the sensor, the signal will be limited. In the same way, the control signal is also limited, as shown in Figure 5.4.

On the actuator's side, saturation also occurs. A valve, for example, has its working range between fully closed and fully opened. The speed of a dc motor is limited to prevent centrifugal forces from damaging the motor and couplings. These phenomena are often referred to as actuator saturation. The effect of saturation is to effectively reduce the gain at high amplitudes, slowing the system response to disturbances.

Limitation of the control signal can cause special problems if the controller is not informed when this is occurring. In controllers, the limitations on the control signals are often specified, and the range should be adhered to as far as possible. Operation outside of the control signal range may risk the controller to a phenomenon known as integral wind-up.

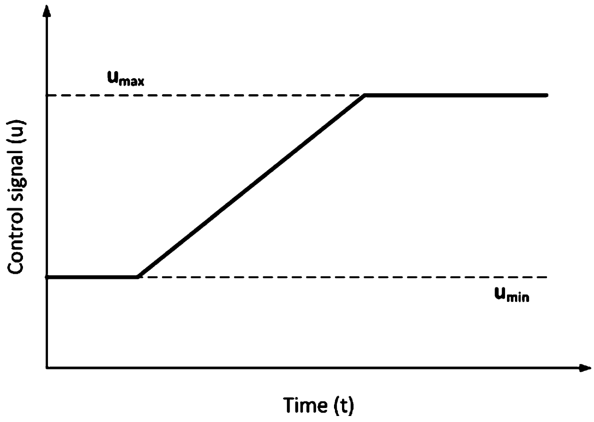

Fig. 5.4 Saturation

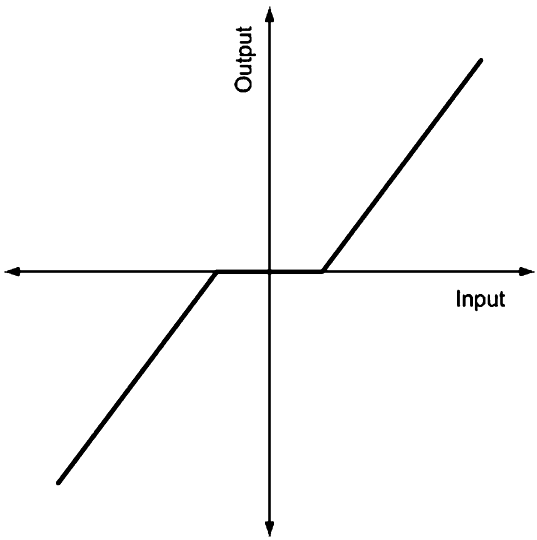

Fig. 5.5 Dead zone

5.2.2.6 Dead Zone

Dead zone has the response as depicted in Figure 5.5. It can arise due to friction, or it can be deliberately introduced into a controller characteristic after the error subtraction to prevent response to small errors. A typical application is a level control system in a hydraulic surge tank where the level is allowed to vary within limits without corrective action being taken, and it is undesirable for the control system to attempt to correct for ripples and hydraulic resonance.

5.2.3 Disturbances

Disturbances are extraneous signals occurring in the control systems, often out of the operator's control. They are mostly of three different kinds: disturbances due to set point changes, low-frequency load disturbance, and high-frequency measurement noise. A more exact knowledge of the disturbances may be used to improve the control.

5.2.3.1 Reference Signal Changes

In servo control, the reference signal changes according to the desired motion profile. This reference disturbance is pre-planned and initiated by the operator or controller. Thus, it is one of those known disturbances which the controller will be typically designed and well positioned to address.

5.2.3.2 Low-frequency Drift

Low-frequency disturbances such as slow drift in the load disturbance are compensated for using the controller. These may arise due to load change or changes in the drive mechanisms. The integral part in a PID controller ensures that a high gain is available in these frequency ranges and therefore effectively eliminates these disturbances.

Disturbances which lie in the same frequency range as the system own dynamics are probably the most difficult to handle. These oscillations cannot be easily filtered out without filtering out other useful information. In addition, the controller also has difficulty in rejecting these disturbances. If possible, these disturbances should be eliminated from the feedback loop. The best way, if possible, is to eliminate them at the source through proper design of the system and its components. If this is not possible, then feedforward from the source may be an effective method to eliminate their effects before they affect control performance.

5.2.3.3 High-frequency Noise

High-frequency disturbances at the input to the system are usually controlled effectively by the system itself, because nearly all physical systems have the characteristics of a low-pass filter. The high-frequency components present in the controlled variables are therefore usually generated in the sensor or the lead from the sensor. They can also arise due to switching power moderators such as pulse-width-moderator (PWM). These high-frequency disturbances or noise should not be addressed at the controller. In order to avoid wear and tear on the valves and actuators, the noise should be filtered out before the controlled variable is transmitted to the control algorithm.

If the noise level is so high that the normal filtering in the controller is not sufficient, the controlled signal should be filtered further. In most modern controllers, the facility is therefore provided for passing the controlled signal through a low-pass filter. The filter bandwidth should be chosen to be wide enough so that the filter does not affect the lower-frequency signals in the same band as the controller. The choice of filter bandwidth is a compromise between speed and stability. With hard filtering (long filter time constant), smooth signals may be obtained with little wear on the valves and actuators as a result. On the other hand, the consequence is a slower handling of load disturbances as the effect of these is also filtered.

5.3 Servo Control Structures

A generally applicable servo control structure used for servo control is shown in Figure 5.6. The structure is shown with reference to a position control problem for illustration, although it can just as efficiently be adopted for other controlled variables such as speed, force, and pressure. The various components under this structure will be described in this section.

5.3.1 Trajectory Generator

This component conditions the position reference signal to yield a smooth position profile with corresponding speed and acceleration profiles which the drive can handle. This conditioning may involve filtering of abrupt step changes in position to reduce jerks and limit the speed and acceleration according to the drive capacity.

5.3.2 Feedback Control

Feedback control used in the industry today is still mainly of the Proportional-Integral-Derivative (PID) type. PID is a familiar term of high significance to many engineers, technicians, and other practitioners involved in automatic control systems. Controllers of the PID type have existed for more than 60 years. Today, PID controllers can be found in virtually all control systems, with applications ranging from precision motion control for assembly and process automation to process conditions regulation. This is not surprising since the reliability of the PID controllers has been proven by decades of successful applications. The wide acceptance and massive support from control engineers all over the world ensure that they have remained the single most important tool in the control toolbox. On the other hand, the benefits derived from the implementation of more complex and advanced control systems have never been quite obvious, despite the remarkable theoretical advances and breakthroughs that have been achieved so far. This is especially true

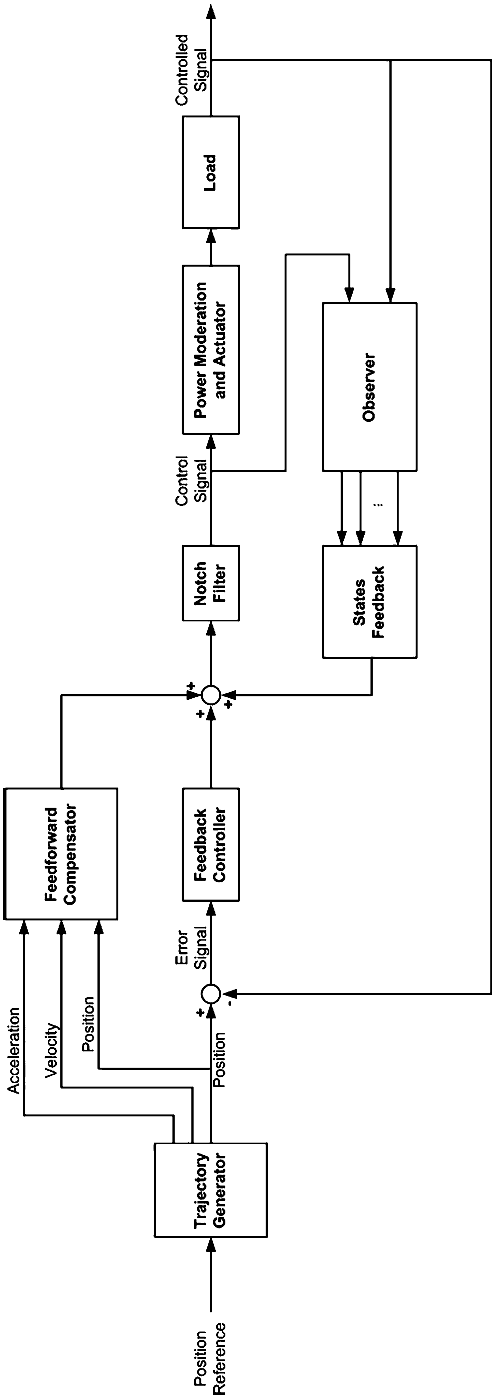

Fig. 5.6 Servo control structure

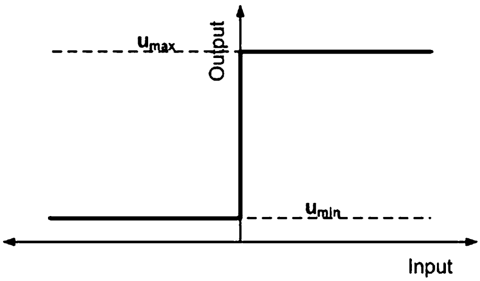

Fig. 5.7 On/off control

under practical operating conditions of industrial control systems, where harsh and varying environmental conditions, imperfect system models, generally modest operator expertise available, and short breakdown recovery tolerance usually render the application and implementation of advanced control very difficult.

5.3.2.1 Components of PID Controller

The PID control structure consists of three constituent components; Proportional, Integral, and Derivative part. This subsection will illustrate the functional principle of each of these components and explain their evolution from the simplest of all types of controller, namely the On/Off controller.

The On/Off controller is undoubtedly the most widely used type of control for both industrial and domestic services. It is the simplest and most intuitively akin to direct manual control. It has an output signal which may be changed to either a maximum or minimum value, depending on whether the process variable is greater or less than the set point. The control law is described by

$$u = \begin{cases} u_{\max}, & e \geq 0, \\ u_{\min}, & e < 0, \end{cases} \tag{5.3}$$

where e is the control error, $e = r - y$ (Figure 5.7), and u is the control signal. The minimum value of the control output is usually zero (off). It has been assumed that the process has a positive static gain, but On/Off control is equally applicable to one with a negative static gain as well, with a direct exchange of the switching conditions. The mechanism to generate On/Off control is usually a simple relay.

One renowned and significant disadvantage of this controller is that it oscillates around a constant set point. This directly affects the process variable which will also oscillate around the desired value. If, for example, On/Off control is applied to controlling the speed of a hydraulic drive with the aid of an On/Off valve which can only open or shut completely, the On/Off controller therefore naturally opens and shuts the valve alternatively. While this control method is attractively simple and easy to implement, it causes rapid wear and tear on the moving parts in the actuating device (the valve in this level-control case), and the *ringing* or oscillation phenomenon may not be tolerable in certain cases. The solution in most On/Off

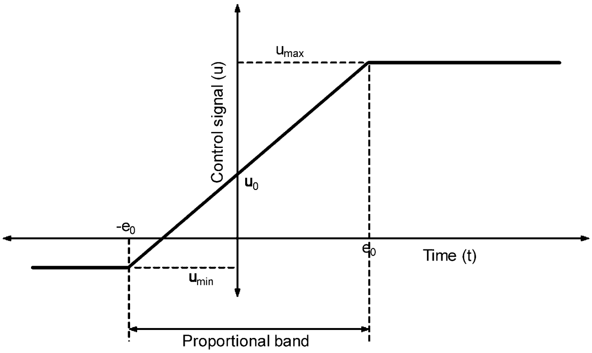

Fig. 5.8 Proportional control

controllers is to establish a dead zone or hysteresis of about 0.5% to 2.0% of the full range. This dead zone straddles the set point so that no control action takes place when the process variable lies within the dead zone.

1. Proportional

   Apart from the use of a small dead zone to reduce the phenomenon of signal oscillations, one alternative way to alleviate the oscillation phenomenon associated with On/Off control is to use a small gain for the controller when the error is small and conversely to use a large gain when the error is large. This can be achieved with a proportional or P controller. Proportional control is the basic continuous control mode. The control signal in a P controller is given by

$$u = \begin{cases} u_{\text{max}}, & e \geq 0, \\ u_0 + K_c e, & -e_0 < e < e_0, \\ u_{\text{min}}, & e \leq -e_0, \end{cases} \tag{5.4}$$

   where u_0 is the level of the control signal when there is no control error, and K_c is the proportional gain of the controller. K_c is also referred to as the proportional sensitivity of the controller. It indicates the change in the control signal per unit change in the error signal. It is indeed amplification and represents a parameter which may be adjusted by the operator. The P controller can also be described graphically as shown in Figure 5.8.

   There are several additional names for proportional control, such as correspondence control and modulating control. While the signal oscillations with On/Off control may be quenched, a new problem with proportional control now arises instead. With pure P control, it is possible and typical that steady state error will occur. In other words, after the transients have died down, there may remain a deviation between the set point and the process variable. This phenomenon may be easily observed from the proportional relationship between the control signal and a small control error:

$$u = u_0 + K_c e. \tag{5.5}$$

This means that the control error is given by

$$e = \frac{u - u_0}{K_c}. \tag{5.6}$$

In the steady state, the control error $e_{ss} = 0$ if and only if at least one of the two following conditions is true:

- K_c is infinitely large
- $u_{ss} = u_0$

Note that e_{ss} and u_{ss} denote the steady-state error and control signal, respectively.

The first condition implies an infinite gain for the controller which effectively reverts P control back to On/Off control with the associated problems P control is set to resolve. The second condition cannot generally be satisfied for all set points r. Even if u_0 can be adjusted relative to the set point, it is still necessary to know at least the process static gain before the adjustment can be done. Thus, for P control, the highest control gain may be used corresponding to an acceptable level of closed-loop stability to reduce the steady-state error. In addition, the largest steady-state error which can occur can be minimized if u_0 is selected to be right in the middle of the operating range of the control signal. In most controllers, therefore, u_0 is chosen to be 50%.

2. Integral

To resolve the steady-state error problem with P control without using excessive controller gain, the integral or reset action should be introduced. The integral action eliminates the problem of any remaining steady-state error. The I part is able to find the correct value for control signal automatically in response to any set point without having to know the process static gain. Integral control action usually is combined with proportional control action, although it is possible but less common that integral action is used by itself. The combination is referred to as PI control. The combination is favourable in that some of the advantages of both types of control action are made available.

The control signal in a PI controller is given by

$$u = K_c \left(\frac{1}{T_i} \int e \cdot dt + e \right), \tag{5.7}$$

where T_i is the integral time of the controller. The constant level of u_0 found in the P controller has thus been replaced by the integral

$$u_0 = \frac{K_c}{T_i} \int e \cdot dt. \tag{5.8}$$

The integral of the control error is effectively proportional to the area under the curve between the controlled variable and the reference (Figure 5.9).

To better understand the steady-state error elimination capability of the PI controller, assume that the closed-loop system is stable and that a steady-state control error exists, despite having a PI controller which implies $e_{ss} \neq 0$. The integrator will continuously accumulate the error signal at the input, and thus the control signal u will be either rising or falling, depending on whether the error is

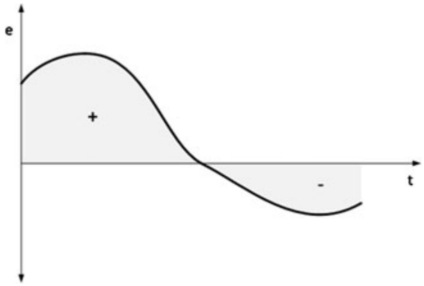

Fig. 5.9 Integral control

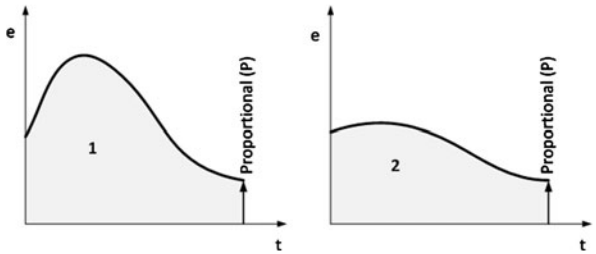

Fig. 5.10 Derivative control

positive or negative. The P part has a constant value corresponding to $K_{\mathrm{c}}e_{\mathrm{ss}}$ and thus will not affect the analysis. If the control signal rises (or falls), the process variable will also rise or fall. This in turn means that the error $e = r - y$ cannot be constant in steady state and thus contradicts the assumption that the error is stationary. Thus, it is not possible to have a non-zero steady-state error when the controller has an integral part and the closed-loop is stable.

A PI controller thus solves the problem of the remaining stationary error and the problem of oscillation associated with On/Off control. The PI controller is therefore an efficient controller without any significant faults and is often sufficient when the control requirements are low to modest.

3. Derivative

Both the P and I components of a PI controller operate on past control errors and do not attempt to predict future control errors. This characteristic limits the achievable performance of the PI controller. The problem is more clearly illustrated in Figure 5.10. The two curves in this figure show the time graph of the control error for two different processes. At time t, the P part, being proportional to the control error, is the same for both cases. Assuming that the I part, being proportional to the area under the two control error curves, is also equal in the two cases, this means that a PI controller gives exactly the same control signal at time t for the two processes. However, it is clear that there is a great difference between the two cases if the rate of change in the control error is considered. For Response 1, the control error changes rapidly, and here the controller should reduce its output to avoid an overshoot in the process variable occurring in the

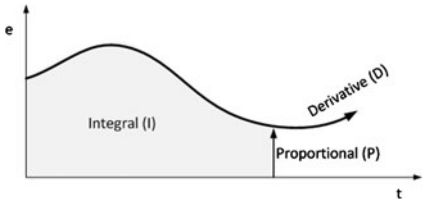

Fig. 5.11 Components of PID controller

future. For Response 2, the control error changes sluggishly, and the controller should react strongly in order to reduce the error more rapidly. Derivative or rate control indeed carries out just this type of compensation.

While it is possible theoretically to have a control action based solely on the rate of change of the error signal e, it is not practical since if the error is large but unchanging, the derivative controller output would be zero. Thus, derivative control is usually combined with at least a proportional control.

The D part of the PID controller is proportional to the change in the error, in other words, to its derivative (Figure 5.11). The D part is proportional to the predicted error at time $t + T_\mathrm{d}$, where T_d is the derivative time of the controller. The control law for the PID controller is

$$u = K_\mathrm{c}\left(e + \frac{1}{T_\mathrm{i}}\int e \cdot dt + T_\mathrm{d}\frac{de}{dt}\right). \tag{5.9}$$

In practice, the derivative is not usually taken of the control error, but only of the controlled variable. This is because the reference signal r is normally constant with abrupt changes. It will thus normally not contribute to the derivative term. Moreover, the term $\frac{dr}{dt}$ will change drastically when the reference is changed. For this reason, it is common practice to apply the derivative action only to the system output. The control structure is then

$$u = K_\mathrm{c}\left(e + \frac{1}{T_\mathrm{i}}\int e \cdot dt - T_\mathrm{d}\frac{dy}{dt}\right). \tag{5.10}$$

There is generally no noticeable difference in control performance between these options; stability or the ability to deal with disturbances or load changes are unaffected, and derivative on the process variable is normally the preferred choice. The only clear occasion when true derivative on error is advantageous is where the controlled variable is required to track a continually changing set point as in military gunnery control.

The derivative action may also result in difficulties if high-frequency measurement noise is present. In practice, these difficulties are normally resolved using additional derivative filtering techniques, such as connecting a low-pass filter to the derivative part. The main objective of the low-pass filter is to ensure that the derivative part affects signals only within the frequency range of interest. The low-pass filter must therefore have different time constants depending on the dy-

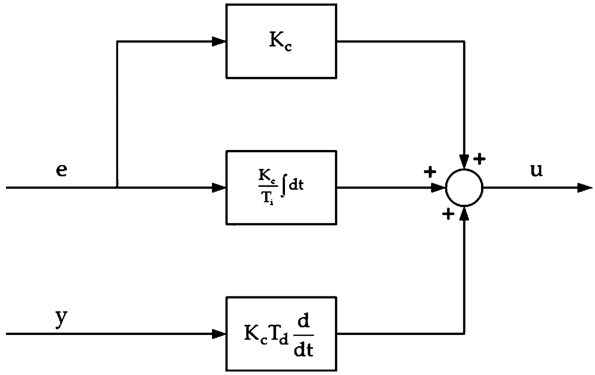

Fig. 5.12 Structure of PID controller

namics of the system. The most common method of solving this problem is to link the time constant of the filter to the derivative time:

$$T_{\text{filter}} = \frac{T_\text{d}}{N}, \tag{5.11}$$

so that the derivative part is given instead by

$$\frac{T_\text{d}}{N}\frac{du_\text{d}(t)}{dt} + u_\text{d}(t) = K_\text{c}T_\text{d}\frac{de}{dt}. \tag{5.12}$$

In most products where N is fixed, it lies in the range 5–10. In those cases where the operator himself has to set the value of the filter constant, this range is usually found to be a suitable choice. In a number of cases, it is not the filter time constant which is to be set, but the maximum gain at high frequencies. This gain is equivalent to the numerical value of N.

When measurement noise is significant, the derivative part is really not necessary. In this situation, one can usually manage well with just a PI controller.

4. Overall structure

   Collectively, the structure of the PID controller (without the derivative filter) is shown in Figure 5.12, where y represents the controlled signal, e is the error signal, and u is the control signal.

5.3.2.2 Tuning the PID Controller

The design objectives or the way the controller is expected to control naturally depends on the application in point. The general dilemma faced in controller tuning is the compromise between the desire for speed versus the desire for stability. Fast control is usually accompanied by poor stability and oscillations. A step change in the set point may well result in a severe overshoot. On the other hand, very stable control without overshoot is usually achieved at the expense of a more sluggish response to set point changes and load disturbances. The solution to this compromise may also depend on the type of disturbance present.

Table 5.1 Effects of PID parameters on speed and stability

	Speed	Stability
K_c increases	Increases	Decreases
T_i increases	Decreases	Increases
T_d increases	Increases	Increases

The robustness of the controller is also related to the demand for speed. Robustness means the sensitivity to variations in the process dynamics. Controllers which are tuned to give fast control are usually more sensitive to variations in the process than controllers which are more conservative in their settings. There is no single set of objectives behind the tuning of PID controller. In certain cases, it is important to tune the controller such that there is no overshoot; in certain cases, a slow and smooth response is desired; other cases may warrant fast response, and significant oscillations are no problem. The point is that the definition of good control depends on specific requirements.

1. Trial and Error Method

   The PID controller has three adjustable parameters, the proportional gain K_c, the integral time T_i, and the derivative time T_d. In certain cases, the operator can also set a filter time constant in order to be able to use the derivative part.

   In this section, the general and typical effect from the individual parameters on the performance in terms of response speed and closed-loop stability will be elaborated. It is naturally necessary and important to know this *cause and effect* when the controller is to be tuned without the help of systematic methods, *i.e.* when tuning a controller purely by trial and error. Even when the systematic methods are used, it is still important to know how the different controller parameters affect the control, because these methods do not always result in the desired control performance and subsequent fine tuning and manual adjustments are necessary.

   The selection of the controller parameters means finding a compromise between the requirement for fast control and the need for stable control. Table 5.1 shows how stability and speed change when the parameters are changed. Note that the table only contains rules of thumb and there are exceptions. For example, an increased gain often results in more stable control for low-order processes such as those involving liquid level control.

   When a PID controller is tuned manually, it is usual to tune the parameters in the order P, I, D. Initially, the I and D parts are effectively disconnected by setting T_i very high (if the integral part cannot be switched off altogether) and $T_d = 0$. After K_c has been adjusted according to the rules-of-the-thumb in Table 5.1 such that the performance approaches the desired one, T_i is reduced to a suitable setting for the integral part. Referring to Table 5.1, this will result in a reduction in stability, which in turn means that the gain K_c has to be reduced. One significant exception to this is in liquid flow control. Liquid flow control loops are very fast and quite often tend to be very noisy. As a result, integral action is often added to the feedback controller in liquid flow control loops to provide a dampening or filtering action for the loop. The advantage of eliminating offset is still present in these cases, but it is not the principal motivating factor.

When the PI controller exhibits satisfactory performance, the adjustment for T_d may begin. Increase in T_d will normally result in an improvement in stability, which in turn means that the gain K_c may be further increased and T_d further reduced.

The derivative part, according to the table, produces both faster and more stable control when T_d is increased. This is only true up to a certain limit, and if the signal is sufficiently free of noise. Raising T_d above this limit will result in reduced stability in control. As mentioned previously, the function of the derivative part is to estimate the change in the control a time T_d ahead. This estimation will naturally be poor for large values of T_d. The reasoning above has also been made without taking into account noise or other disturbances. The noise is amplified to a greater extent when increases. It is thus often the noise which sets the upper limit for the magnitude of T_d. In some cases, the noise level can be so high, despite filtering, that one may be unwilling to use the derivative part because it gives such a bad control signal. A PI controller would then be preferred with its smoother control signal, even at the cost of poorer control performance.

2. Systematic methods

Until 1940, the tuning of the PID controllers was still more of an art conducted by *ad hoc* methods on controllers that were a hodge–podge of techniques or add-on components that defied any logical rules that could be universally applied.

When two engineers at Taylor Instrument Company tried tuning the Fulscope controller, they discovered that tuning two parameters was already difficult, not to mention three for the full PID controller. The engineers, John Ziegler and Nathaniel Nichols, decided to work on the tuning problem. By 1941, they had a relatively straightforward method for tuning the PID controller, the results were the now famous *Ziegler–Nichols* methods of tuning controllers—methods that survived the slings and arrows of its early detractors, withstood the test of time, and work just as well as many of the later more sophisticated optimizing forms on a great majority of process applications.

The Ziegler–Nichols methods use information from the step response or the frequency response of the system for tuning the PID controller. They are more applicable to process control applications and are thus not reproduced here. For velocity control, a *Ziegler–Nichols*-like set of rules is also available as follows:

$$K_c = \frac{1}{K T_d \tilde{T} \omega_o^2}, \tag{5.13}$$

$$T_i = \frac{2\xi}{\omega_o}, \tag{5.14}$$

$$T_i T_d = \frac{1}{\omega_o^2}. \tag{5.15}$$

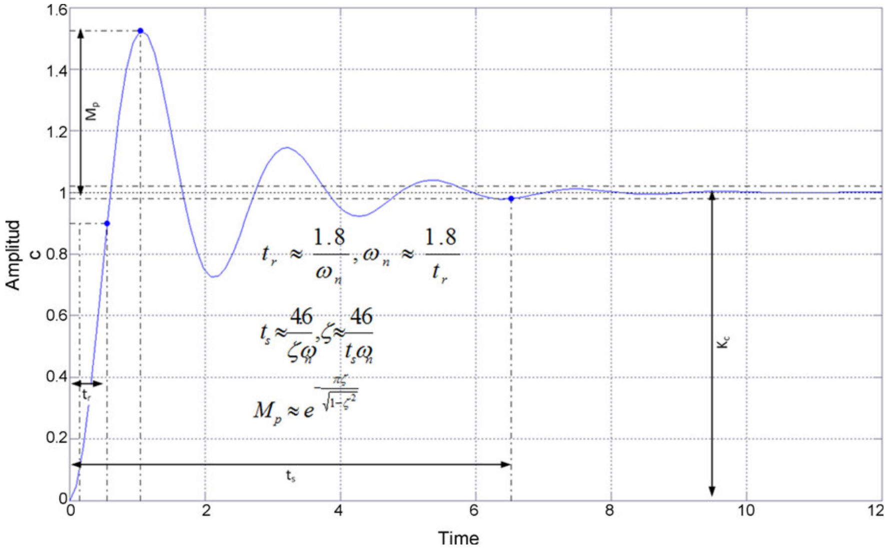

Fig. 5.13 Velocity step response

The gain K, damping ratio ξ, and natural frequency ω_o of the system are identified from a velocity step response as shown in Figure 5.13. With these parameters and the desired closed-loop time constant $\tilde{T}$, the PID parameters can be obtained.

For position control, PD controller can be used with the position feedback. Integration action is not necessary as the position control loop already has an inherent integrator. The PD gains are:

$$K_c = \frac{1}{K\tilde{T}}, \tag{5.16}$$

$$T_d = T, \tag{5.17}$$

where T is the open-loop time constant, and $\tilde{T}$ is the closed-loop time constant.

There are many other rules available for tuning the PID controllers. All of these rules would require a model of the system under control, either in a parametric form or non-parametric form. Automatic tuning methods for PID controller will typically involve automated procedures to efficiently extract a model for the system, based on which the PID parameters can be automatically computed.

5.3.2.3 Integrator Wind-up

Control saturation leads to an integrator wind-up phenomenon. Figure 5.14 shows the control signal, the controlled variable (process variable), and the reference (setpoint) in a case where the control signal is limited. After the first change in set

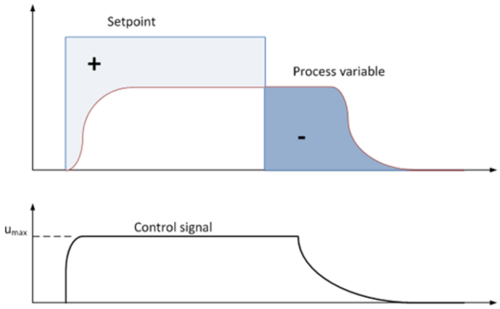

Fig. 5.14 Integrator wind-up

point, the control signal increases up to its upper limit $u_{\max}$. This control signal is not large enough to eliminate the control error. The integral part of the PID controller is proportional to the area under the control error curve. This area is marked in Figure 5.14. Hence, the integral part will continue to rise since the control signal is unable to eliminate the control error.

Figure 5.14 shows what happens when after a certain time the reference is changed to a lower level where the controller is able to eliminate the control error. Because the integral part was allowed to rise and reach a high level while the control signal was limited, the control signal remains at its limit for a longer time before it *unwinds* in accordance to the negative errors.

The problem is known as integrator or reset wind-up, and the consequence is a response with large overshoot and long settling time. Integrator wind-up commonly occurs in override, surge, batch, and pH control loops. The duration of the windup is longer for those systems with longer time constants and large process gains because the corresponding small proportional and integral action prolongs the reversal of controller output.

There are several ways to avoid integrator wind-up. The simplest way of overcoming the problem is to stop updating the integral part when the control signal is limited. For this to be possible, the controller naturally has to know what the limits are. In Figure 5.15, this point in time corresponds to point B. The question now is at what point it is re-enabled again. Point C is obviously far too late although it is still better than point D in the unprotected controller). A common solution is to de-saturate the integral term at the point where the rate of increase of the integral action equals the rate of decrease of the proportional and derivative terms. This occurs when the slope of the PID output is zero, *i.e.* when

$$e = -T_{\mathrm{i}}\left(\frac{de}{dt} + T_{\mathrm{d}}\frac{d^2e}{dt^2}\right). \tag{5.18}$$

This brings the controller out of saturation at the earliest possible moment, but this can, in some cases, be too soon leading to an unnecessarily damped response. Most controllers are equipped with methods, in one form or another, for avoiding integrator wind-up.

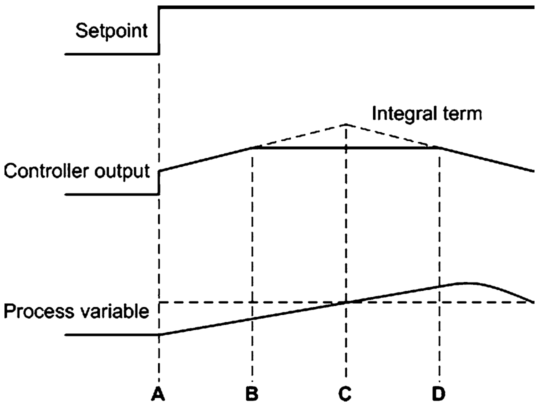

Fig. 5.15 Effect of integrator wind-up

5.3.3 Feedforward Compensator

Common nonlinearities occurring in servo drives have been highlighted. If a model of a specific nonlinearity is available, a feedforward controller can be designed to compensate for the effects. For example, if a friction model is available, such as the Tustin model shown in Figure 5.1, and then based on the velocity reference signal from the trajectory generator, a counter-force can be generated to eliminate or minimize the frictional effects, so that to the feedback controller, the system appears to be friction-free.

Feedforward compensation is applicable to tackle measurable disturbances. For example, the pressure in a hydraulic drive may fluctuate due to demand for hydraulic pressure from different parts of the plant. The variation will affect the hydraulic flow and also the force transmitted to the actuator, leading further to poor motion control. A pressure sensor can detect this fluctuation and input it to a feedforward compensator to take corrective action before it affects the final control objective.

There is an alternative and commonly used method, called gain scheduling, to compensate nonlinearities, which can be considered as a form of feedforward compensation as well. The idea behind gain scheduling is to divide the operating range of the controller into several sub-ranges and to use different controller parameters within the different sub-ranges.

Suppose, for example, that the nonlinearity is due to a nonlinear valve. One possibility is to tune the controller for the worst case, which means that the controller is tuned to handle the operating condition where the system gain is large. This may work, but it will typically result in sluggish control in other operating ranges.

Using gain scheduling, a table with a number of sets of controller parameters is built. The controller parameters that are suitable for the actual operating condition are automatically selected from the table based on a suitable signal. For the nonlinear valve, this selection is performed using the control signal. It is the control signal that will determine the position of the valve characteristic, and it can thus be used to initiate the feedforward transmission of a suitable set of gains from the scheduling table for the feedback controller.

For other types of nonlinearity, other signals are used to determine when to change parameters in the gain scheduling table. Here are some examples of reference signals selected for gain scheduling:

- Nonlinear valves and actuators
  The control signal is used to form the reference signal for gain scheduling.
- Nonlinear sensors
  The system variable is used to produce the reference signal for gain scheduling.
- Production-dependent variations
  Some signal that is related to the production is used as reference for the gain scheduling. It may be the measurement signal, but it may also be some external signal that has to be connected to the controller.

5.3.4 States Feedback with Observers

A servo control system may feedback multiple signals to achieve a high level of performance. Using Figure 5.6 as an example of a position control problem, to achieve a faster response to a changing position reference, feedback of the velocity or even acceleration signals can be used for low-frequency servo drives. With this feedback, the control system is also more robust to low-frequency disturbances, since these disturbances can be addressed earlier before they affect the position. These internal variables are not directly controlled variables, although they will affect it. They are also called state variables. Full state feedback control will require the measurements of all the states to be available.

A particular case of a states feedback controller is the common configuration where velocity feedback is used by a secondary and inner controller and position feedback is used by a primary and outer controller as shown in Figure 5.16. PID control can be adopted for either of the controllers, and the tuning rules suggested earlier remain applicable.

Another example of the application of additional measurements for compensation of nonlinearities is the valve positioner, the working diagram of which is presented in Figure 5.17. As discussed previously, friction is very commonly present in valves, giving rise to a dead zone in the valve dynamics. This means that the valve may *stick* and does not respond to the control signal until the control signal is increased to overcome this dead zone. Then, the valve will overcome its *stickiness* and opens or closes by an unnecessary big margin. Such nonlinear characteristics cause problems for control.

To overcome this, some valves have built-in valve positioners which operate based on cascade control, by introducing a built-in inner position control loop to control the valve step position.

Sometimes, measurements may not be available for certain state variables to implement a full state feedback control. The reason may be due to difficulty to mount

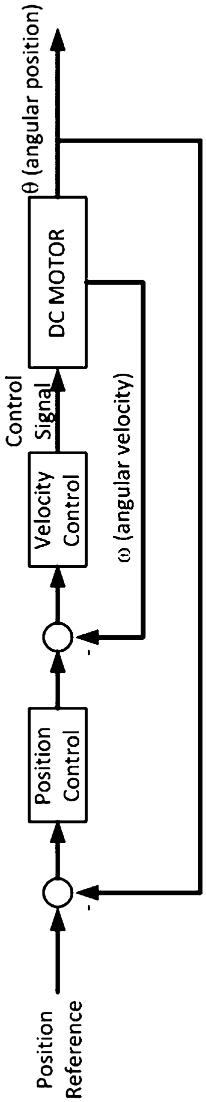

Fig. 5.16 Cascade control

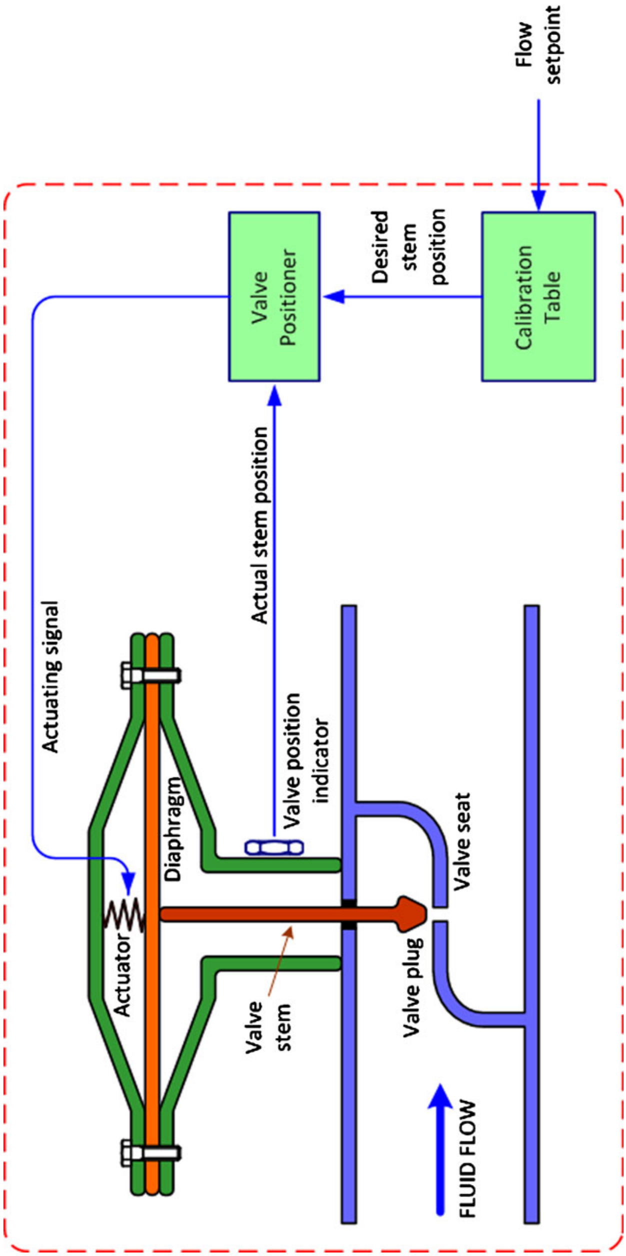

Fig. 5.17 Valve positioner

a sensor in the overall system for that particular measurement, or simply there is no suitable sensor for those variables. In these situations, a state observer can be designed as a soft sensor to estimate these states. A state observer is designed based on a model of the system. The noise present has to be properly accounted to avoid cumulative error in state estimation. For example, under practical situations, it is unrealistic to estimate velocity from position measurements through a differentiator as it will amplify the measurement noise. For that matter, it is even more unlikely to be able to obtain acceleration signal this way.

Many approaches are now available to design an observer; one of the most common approaches is based on the Kalman filter which is capable of periodically flushing the cumulative prediction errors due to measurement noise. Apart from states estimation, an observer can also be used to yield disturbances which are not measurable so that they can be compensated for.

5.3.5 Notch Filter

Mechanical vibration in servo drives and equipment can occur due to many factors, such as unbalanced inertia, bearing failures on turbines, motors, generators, pumps, drives, turbofans, *etc.*, poor kinematic design resulting in a non-rigid support structure, component failure, and/or operations outside prescribed load ratings. The machine vibration signal can be typically characterized as a narrow-band interference signal anywhere in the range from 1 Hz to 500 kHz. To prevent equipment damage from the severe shaking that occurs when machines malfunction or vibrate at resonant frequencies, a real-time monitoring or control device will be very useful. When the servo drive is used to perform highly precise positioning functions, undue vibrations can lead to poor repeatability properties, impeding any systematic error compensation effort. This results directly in a loss of precision and accuracy achievable.

The task of eliminating/suppressing undesirable narrow-band frequencies can be efficiently accomplished using a notch filter (also known as a band-stop filter). Ideally, the filter highly attenuates a particular frequency component and leaves the others relatively unaffected. Thus, an ideal notch filter has a unity gain at all frequencies and a zero gain at the null frequencies. A single-notch filter is effective in removing a single frequency or a narrow-band interference; a multiple-notch filter is useful for the removal of multiple narrow-bands, necessary in applications requiring harmonics cancellation.

Digital notch filters are widely used to retrieve sinusoids from noisy signals, eliminate sinusoidal disturbances, and track and enhance time-varying narrow-band signals in wide-band noise. They have found extensive applications in the areas of radar, signal processing, communications, biomedical engineering, and control/instrumentation systems.

Complete narrow-band disturbance suppression requires an exact adjustment of the filter parameters to align the notches with the resonant frequencies. If the true frequency of the narrow-band interference to be rejected is stable and known a priori, a notch filter with fixed null frequency and fixed bandwidth can be used. However, if no information is available a priori or when the resonant frequencies drift with time, the fixed notch may not coincide exactly with the desired null frequency if the bandwidth is too narrow. In this case, a tunable/adaptive notch filter is highly recommended.

5.4 Implementation

This section will be focused on the two main approaches to implement a control system in an actual application.

5.4.1 Digital Control

Most PID controllers today are based on microprocessors or implemented in computer-based distributed control systems (DCS). There are associated consequences which will be briefly addressed here. Comparing microprocessor-based PID controllers with the analog equivalents, there appear to be mainly advantages. The use of a microprocessor means that it has become possible to incorporate more complex functions in the controller such as automatic tuning, alarm handling, filtering, digital control, *etc.* The parameter values can also be tuned more accurately, since the user settings will be precisely the true control settings unlike analog PID where they may be different due to aging of analog components.

There is, however, one important disadvantage. Digital controllers do not process the analog signals directly. They are sampled instead for analysis. Between consecutive samples, the controller gets no information on the intermediate value of the analog signal. This is not a problem as long as the time interval between the samplings is much shorter than the time constant of the controlled process.

Sampling is also effectively equivalent to introducing an additional dead-time into the feedback loop. A sudden disturbance is only detected in the controller after an average time of half a sampling period. The sampled version is thus delayed compared with the actual signal, and the delay is approximately half the sampling period. The selection of the sampling interval is often crucial to the performance of digital control systems.

Figure 5.18 illustrates the effects of aliasing. The original signal is a sinusoid with a period of T_{sine}. Sampling is done every $1.25T_{\text{sine}}$ interval. The period of the signal, reconstructed from the samples, is $5T_{\text{sine}}$ instead.

The sampling frequency should thus be chosen to avoid aliasing; otherwise such frequencies have to be filtered out before sampling by using a low-pass filter with cut-off frequency at the Nyquist frequency. Such a filter is called an anti-aliasing or pre-sampling filter.

In certain single loop controllers, the sampling period can be fixed. For example, in pressure controllers, the sampling interval is usually between 0.1 and 0.3s. In instrumentation systems, on the other hand, it is often possible to specify the sampling period.

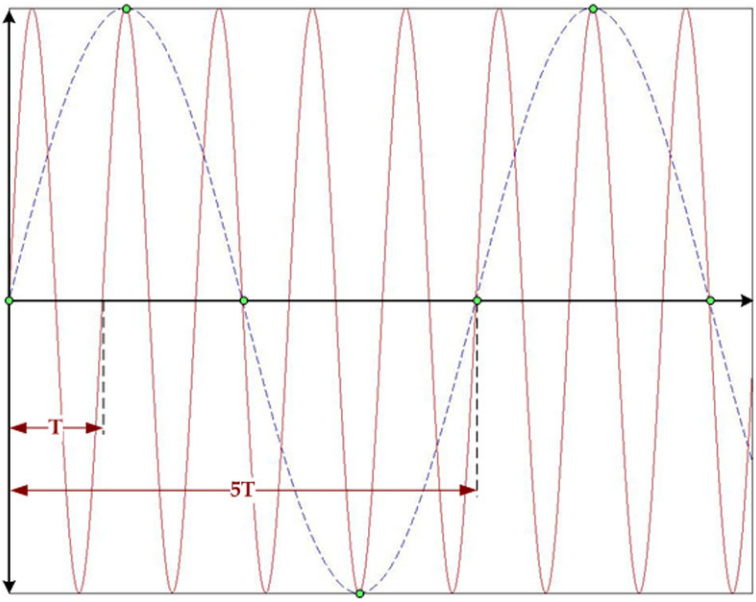

Fig. 5.18 Aliasing

5.4.1.1 Discretization

To implement a continuous-time control law, involving continuous-time signal differentiation and integration such as the PID controller, in a digital computer, it will be necessary to approximate the continuous-time control law by difference equations.

In the case of PID control:

1. Proportional control

   In continuous-time, the proportional term is

$$u_{\mathrm{p}}(t) = K_{\mathrm{c}} e(t). \tag{5.19}$$

   This term is implemented simply by replacing the continuous variables with their samples:

$$u_{\mathrm{p}}(kh) = K_{\mathrm{c}} e(kh), \tag{5.20}$$

   where h is the sampling interval, and kh thus denotes the kth sampling instants.
2. Integral control

   In continuous-time, the integral term is

$$u_{\mathrm{i}}(t) = \frac{K_{\mathrm{c}}}{T_{\mathrm{i}}} \int_0^t e(\tau) \cdot d\tau, \tag{5.21}$$

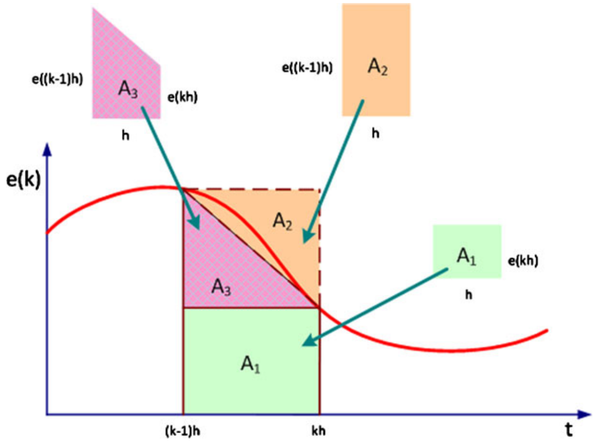

Fig. 5.19 Discretization of PID controller

where $e(t) = r(t) - y(t)$. In discrete time, the integral may be approximated by the area under the $e(t)$ curve from $t = 0$ to $t = kh$ iteratively as

$$u_{\mathrm{i}}(kh) = u_{\mathrm{i}}\big((k-1)h\big) + \frac{K_{\mathrm{c}}}{T_{\mathrm{i}}} A, \tag{5.22}$$

where A is the area under the $e(t)$ curve from $t = (k-1)$ to kh.

As shown in Figure 5.19, A may be approximated in various ways:

$$\begin{aligned} A_1 &= he(kh), \\ A_2 &= he\big((k-1)h\big), \\ A_3 &= h\frac{e(k-1)h + e(kh)}{2}. \end{aligned} \tag{5.23}$$

The three discretization methods may also be respectively obtained by differentiating the integral equation and applying forward, backward, and Tustin's approximation to the differential term.

3. Derivative control

In continuous time, the derivative term with filtering is

$$\frac{T_{\mathrm{d}}}{N}\frac{du_{\mathrm{d}}(t)}{dt} + u_{\mathrm{d}}(t) = K_{\mathrm{c}} T_{\mathrm{d}} \frac{de}{dt}. \tag{5.24}$$

Integrating this equation, it follows that

$$\begin{aligned} \frac{T_{\mathrm{d}}}{N} u_{\mathrm{d}}(kh) + \int_{(k-1)h}^{kh} u_{\mathrm{d}}(t) \cdot dt &= \frac{T_{\mathrm{d}}}{N} u_{\mathrm{d}}\big((k-1)h\big) \\ &\quad + K_{\mathrm{c}} T_{\mathrm{d}}\Big(e(kh) - e\big((k-1)h\big)\Big). \end{aligned} \tag{5.25}$$

Denoting $A = \int_{(k-1)h}^{kh} u_{\mathrm{d}}(t) \cdot dt$, A may be computed in various ways, similar to the integral action case:

$$\begin{aligned}
A_1 &= hu(kh), \\
A_2 &= hu\big((k-1)h\big), \\
A_3 &= h\frac{u(k-1)h + u(kh)}{2}.
\end{aligned} \tag{5.26}$$

Thus, the D part can be discretized so that the corresponding difference equation for $u_\mathrm{d}(kh)$ in terms of $u_\mathrm{d}((k-1)h)$, $e(kh)$, and $e((k-1)h)$ can be obtained accordingly as follows:

$$\begin{aligned}
u_\mathrm{d}(kh) &= \frac{T_\mathrm{d}}{hN + T_\mathrm{d}} u_\mathrm{d}\big((k-1)h\big) \\
&\quad + \frac{K_\mathrm{c} T_\mathrm{d} N}{hN + T_\mathrm{d}} \big(e(kh) - e\big((k-1)h\big)\big), \\
u_\mathrm{d}(kh) &= \frac{T_\mathrm{d} - hN}{T_\mathrm{d}} u_\mathrm{d}\big((k-1)h\big) \\
&\quad + \frac{K_\mathrm{c} T_\mathrm{d} N}{T_\mathrm{d}} \big(e(kh) - e\big((k-1)h\big)\big), \\
u_\mathrm{d}(kh) &= \frac{2T_\mathrm{d} - hN}{hN + 2T_\mathrm{d}} u_\mathrm{d}\big((k-1)h\big) \\
&\quad + \frac{2K_\mathrm{c} T_\mathrm{d} N}{hN + 2T_\mathrm{d}} \big(e(kh) - e\big((k-1)h\big)\big).
\end{aligned} \tag{5.27}$$

These discrete approximations will be stable if the sampling interval h is sufficiently small.

5.4.1.2 Inputs/Outputs

The inputs and outputs of a digital control system can be of various kinds. They can be direct analog signals, direct digital signals, or a bus interface. Modern control systems are designed to be modular in nature, so that apart from the core control module, the number and types of I/O modules can be optimized to the requirements.

5.4.2 Analog Control

The field of control is rooted together with analog computing. Prior to the wide proliferation of digital computers from the 1970s, a control engineer had to use analog computing technology to simulate control systems, and analog technology in mechanics, pneumatics, and electronics was the core of all control systems.

Analog control systems were huge, bulky, and poorly reconfigurable. A PID controller based on analog amplifiers is shown in Figure 5.20.

Due to the rapid developments in microelectronics and digital computers, digital control systems have now almost completely replaced the analog counterparts, both

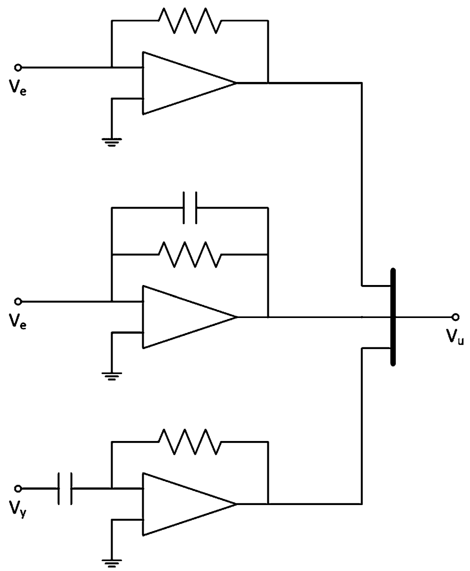

Fig. 5.20 Analog PID controller

in the central control room for supervisory control and in the field for direct front-end control using microprocessor-based digital controllers.

Still, there is a place for analog control. From an academic perspective, analog control systems should still be used in classical control education which has focused on systems with natural time continuous signals. From an application perspective, an all-digital control system is still non-existent. A digital control system has to work with analog components, in the form of sensors, actuators, and power moderators, and this is unlikely to change as systems and signals in real-life are predominantly of an analog nature. The digital computer is used mainly in the rapid and convenient processing of this analog information.

Far from becoming an item soon to be found only in the museum, the use of pure analog control is being re-looked. It has one single and strongest feature, which corresponds to the weakest link in a digital control system. Analog control does not require sampling, and this single feature which digital control can never possess has far fetching implications as far as accuracy and control speed in a servo control loop are concerned. With the advance in nanotechnology, a nanotechnology application may soon require speed and precision higher than what a digital controller can deliver, and analog control may be injected a fresh dose of life. Already, there are efforts to fabricate general-purpose analog computers using modern VLSI chips.

5.5 IEC 61131-3 Programming Standards

The IEC 61131-3 standard was published in 1993 as an attempt to unify the numerous and different languages and dialects associated with Programmable Logic

Controllers (PLCs) from different manufacturers. With the PLC being used in practically all industries (automobile production lines, brewing, water treatment, assembly lines, *etc.*), a standard for PLC programming languages is necessary. Although the ladder program is a widely accepted language due to its intuitive and similar structure to relay logic programs, the ladder symbols and facilities also vary between different PLC products. Furthermore, due to the limitations associated with the ladder programs considering the very expansive industrial control applications in the present day, other programming languages are included in the standard, each armed with specific features to address different issues of the control problem, either individually or collectively.

Essentially, IEC 61131-3 improves control software quality. Supported by numerous PLC members and rapidly extending to other industrial controllers, the standard offers many benefits to the users of compliant controllers.

IEC 61131-3 compliant software offers the following:

- a well-structured program development
- strong data typing
- full execution control
- support for complex sequential behaviour
- well-defined data structures
- flexible language selection. (The system designer is free to select the language that is most suitable to solve a part of an application program. Different parts of a program can be expressed in any of the languages)
- vendor-independent software element. (Software written for IEC 61131-3 compliant controller can be portable and run on controllers from different vendors)

There are five languages included under this standard. They are:

- ladder diagrams
- instruction list
- structured text
- sequential function chart
- function block diagram

The sixth one, called the Continuous Function Chart (CFC), is beginning to be adopted by some control manufacturers. In the following subsections, a short presentation of the key features structure of these languages will be given.

5.5.1 Ladder Diagrams

A ladder diagram/program is rather similar to a relay ladder diagram/program. It implements a logic via a combination of switches, which can be normally open or normally closed, and an associated set of coils.

The rungs in a PLC program implement the logic for the sequence control. Switches are connected such that the coil is activated when the necessary logical conditions are satisfied.

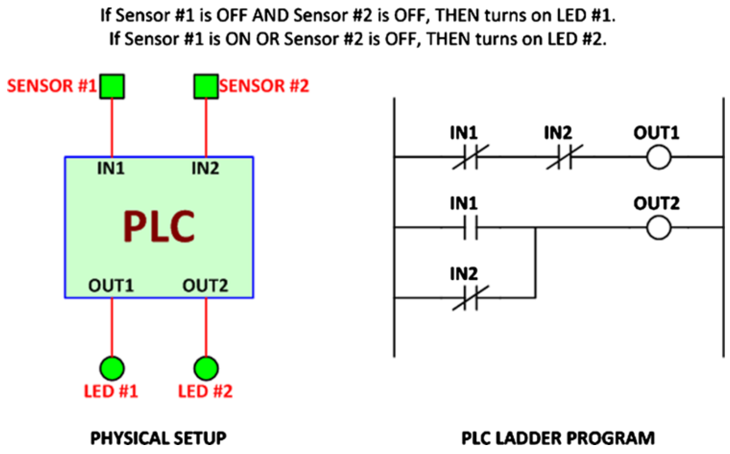

Fig. 5.21 Ladder implementation of inverse logic

The following example, shown in Figure 5.21, will illustrate the implementation of this logic.

5.5.2 Instruction List (IL)

IL is a low-level language which has a structure similar to a simple machine assembler. It is ideal for solving straightforward problems where there are few decision points and where there are a limited number of changes in program execution flow. It may be used to write tight, optimized code for performance critical sections of a program. As with other machine assembler code, it can be difficult to follow the flow and structure of the program.

Conversion from IL to other languages is not always possible, but converting other languages to IL can always be achieved. An example of an IL program (illustrated alongside the ladder program) is shown in Figure 5.22.

5.5.3 Structured Text (ST)

The Structured Text (ST) is a procedural language (Pascal-like) which can be used for programming a wide range of industrial applications. It is a high-level language with strong data type checking and a formal syntax especially useful to implement calculations which can involve simple or complex expressions. ST has facilities for conditional evaluation of statements, for repeating sections of code, and for calling functions and function blocks. ST is typically used to program:

- an entire program
- actions within a Sequential Function Charts (SFC)

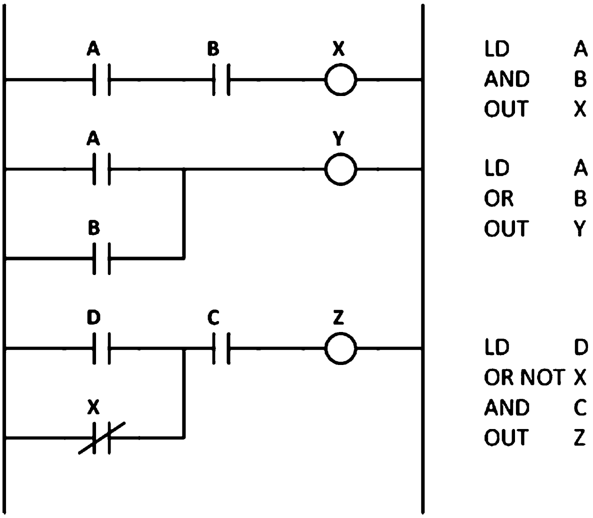

Fig. 5.22 Instruction list compared to ladder diagram

- transition conditions for SFC transitions
- functions or function blocks

A fragment of ST is given below.

Fragment of structure text

$tempA := b * b - 4 * a * c;$
IF $tempA >= 0.0$ THEN
$tempB := SQRT(tempA);$
$x_{\mathrm{Root1}} := (-b + tempB)/(2 * a);$
$x_{\mathrm{Root2}} := (-b - tempB)/(2 * a);$
ELSE
$msg :=$ '*Imaginary Roots*';
ENDIF

5.5.4 Sequential Function Charts (SFC)

SFC is a graphical language for dictating sequences. It is mostly derived from the French Grafcet Standard. SFC is based on established standards and accepted industrial practice. It allows alternative sequences to be selected using divergent paths. There is support for running sequences in parallel using simultaneous sequences (concurrency). Steps and transitions can be programmed using any of the IEC lan-

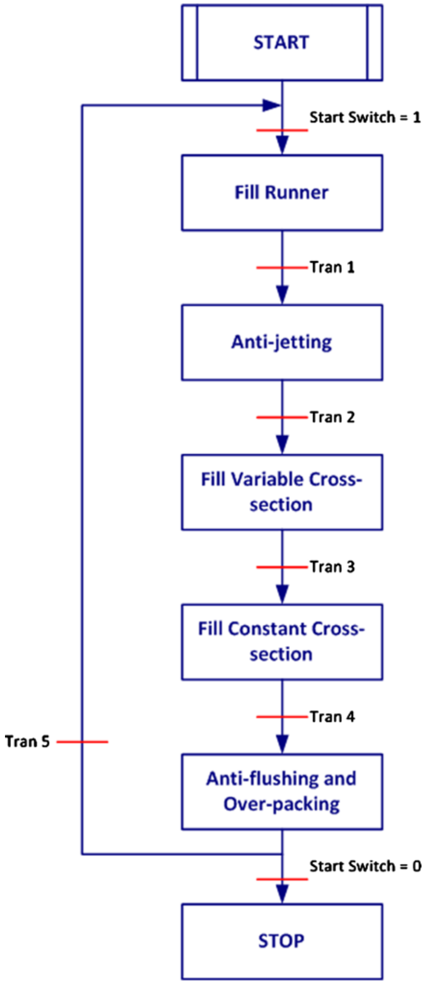

Fig. 5.23 Sequential function chart organization of an injection molding cycle

guages. SFC can be used at any design level, at the top level within a program, or at a lower level within a function block.

An example of an SFC used to organize the different phases in the control of the ram velocity over an injection molding cycle is given in Figure 5.23.

Each of the rectangular boxes corresponds to a step in the SFC, and each of the horizontal bar separating two steps corresponds to a transition condition. The transition condition can be based on an event happening, such as a sensor turning on or off, or it can be based on specified time duration. The transition from one step to another in an SFC can only occur when the transition condition holds. For example, the **Start** step can only transit to the *Fill Runner* step when the condition **Start Switch** = **1** holds, *i.e.* a start switch is activated.

Each step of the SFC is associated with a number of actions to be executed. These actions can be programmed using any of the IEC 61131-3 languages.

5.5.5 Function Block Diagrams (FBD)

The main features of an FBD are:

- graphics-based, software building blocks for constructing control systems using proven solutions
- well-defined I/O
- interconnection between blocks may be written using ST
- same block may be used several times, allowing PLC memory to be used efficiently
- IEC 61131-3 defines a small collection of basic and very useful function blocks including bi-stables, timers, and counters
- standard blocks can be used to build powerful composite function blocks
- these blocks may be written in any IEC languages described earlier

Examples of standard function blocks are given in Figure 5.24.

These languages may be used concurrently. Figure 5.25 shows a PID function block which can be realized using these languages.

Below is the realization of this PID block using ST (in the gray box) and using FBD (Figure 5.26).

Realization of the PID block using ST

FUNCTION BLOCK PID
*VAR*INPUT
PV: *REAL*;
SP: *REAL*;
KP: *REAL*;
TI: *REAL*;
TD: *REAL*;
CYCLE: *TIME*; (* Function block cycle time *)
*END*VAR
*VAR*OUTPUT
XOUT: *REAL*;
*END*VAR
VAR
ERROR: *REAL*;
ITERM: *INTEGRAL*; (* Integral comp *)
DTERM: *DERIVATIVE*; (* Derivative comp *)
*END*VAR
(* PID algorithm expressed FBD *)

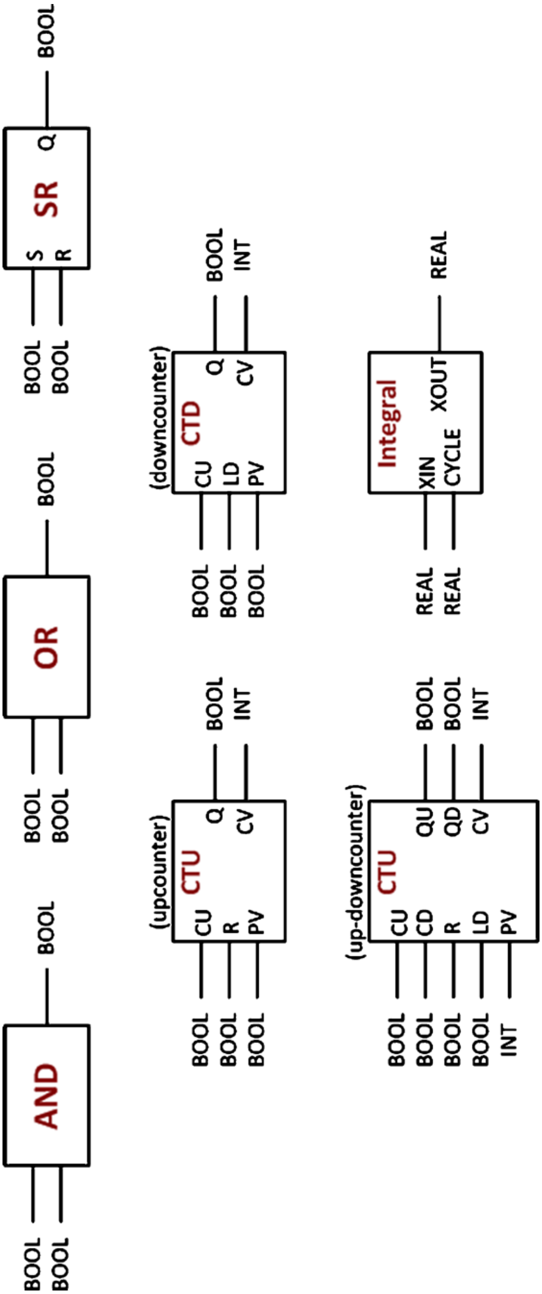

Fig. 5.24 Examples of standard function blocks

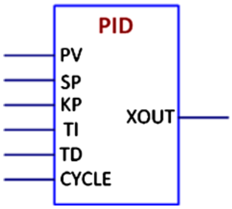

Fig. 5.25 PID function block

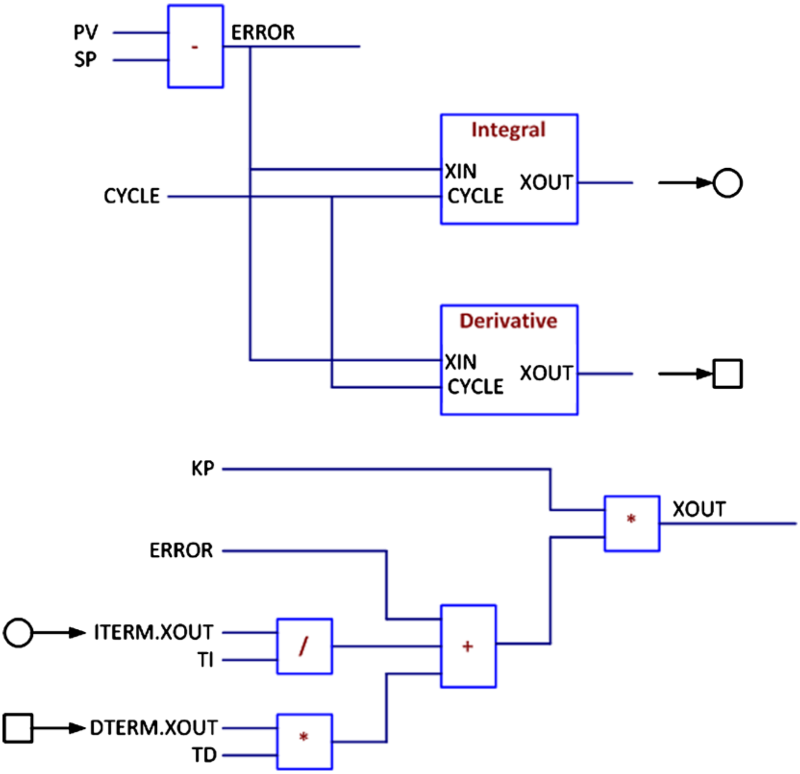

Fig. 5.26 Realization of the PID function block

5.5.6 Continuous Function Chart (CFC)

CFC facilitates efficient configuration of control structures. Standard and pre-programmed function blocks are easily put in place graphically. Their inputs and outputs are simply soft-wired together to realize the overall system. There are minimum requirements on the skills of the programmer with background syntax checks and automatic dialogs which allow a modest programmer to generate a program in a short period of time.

Chapter 6
Digital Communication Protocols

Preceding chapters have focused specifically at sensors, actuators, and controllers, components which are necessary to form a servo drive. In order to integrate these components to work as a system or to interconnect multiple control systems, signal communication protocols and standards will be necessary.

Currently, many servo drive components are equipped with a digital fieldbus interface. Fieldbus is a digital, bidirectional, serial bus communications network that links various instruments, transducers, controllers, final control elements, and other devices. It serves as a spinal column of distributed real time systems and tremendously simplifies the wiring among field devices.

In this chapter, the evolution of field communication protocols, along with the development in control and computer systems will be highlighted. The typical fieldbus protocol stack will be described, and the common field protocols currently used in servo drives will be presented, along with applications to servo drives.

6.1 Evolution of Fieldbuses

The evolution of fieldbuses can be elaborated from a control systems perspective. In this section, the changes in control system configuration over the years will be described to set the background and rationale of the transition in field signal communication from pneumatic and analog transmission to the digital and field protocols of today, as shown in Figure 6.1.

6.1.1 Distributed Control Systems

In the early days of the 1920s, controllers are mainly localized to specific control loops for two reasons. Firstly, control systems are predominantly manual-based, requiring the control operators to be close to the field devices (*i.e.* sensors and actuators). Secondly, the field devices were mechanical systems working in close proximity to each other. Sensing variables were converted to force which drove some

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_6,

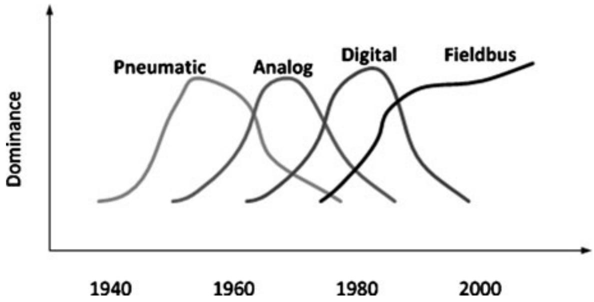

Fig. 6.1 Evolution of field communication

spring-loaded pointer systems, and the control operator would have to be close to the meters to read the measurements and adjust the actuators. Such configurations posed great difficulties in the coordination of control loops to achieve good overall control performance. The advances in analog transmission and standards were a response to the requirements for centralized control. The 3–15psi pneumatic standards and 4–20mA current standards evolved in the 1940s to allow the implementation of the concept of a central control room. With analog transmission, the field and the control room can be separated physically over a long distance. Signals were transmitted to and from the control room via analog transmission. Control systems began to deploy automatic analog control on a large scale.

In the 1960s, the advances in computer systems brought about the first direct digital control. Analog controllers in the control room can be replaced by a central computer which realizes the first digital control. With all the information contained in the computer, configuration and optimization, both at the direct and supervisory control level, can be achieved efficiently. Nevertheless, because the early computers were massive, slow, and limited in capacity, direct digital control was a customized and costly solution. Reliability and cost issues were big bottlenecks of a direct digital control. When the central computer broke down, production would grind to a stop.

A major breakthrough then came about with the invention of the transistor and, subsequently, the integrated circuit (IC). It also led to the first microprocessor produced by Intel in 1971. The invention of this tiny device has expanded the whole world. With the IC, computer immediately took on a new look. Costs, size, and power requirement took a dive, while power and capacity increased separately to be able to handle some real-time computations. It became possible to fit small pieces of the computer into other devices; field controllers included. The embedded computers can fulfill not just control functions, but digital communication functions as well. They helped to bring about the first distributed control system by Honeywell in the early 1970s.

In a distributed control system (DCS), the front-end controllers, which are microprocessor-based digital controllers, are used for direct control. A central computer fulfilled the supervisory and coordination role for the entire network of front-end controllers. Communication between the central computer and the front-end controllers is via digital transmission. With higher noise immunity, transmission dis-

Table 6.1 Comparison of analog and digital communication

Analog	Digital
Generally susceptible to noise	Improved noise rejection
Point to point	Multi-drop communication, allows many devices to connect to a single communication system
High cabling costs	Reduced cabling cost
Single parameter access of the same device; only transducer signal or command is transmitted	Multiple parameter access of the same device (*e.g.* transducer signal, device status, device identifiers, configuration parameters, *etc.*)
Interface to analog devices	Interface to more devices including intelligent digital field devices (smart devices)

tance can be greatly increased. With multi-drop communication, the costs in terms of communication medium can be lower than the analog equivalent. No longer is transmission limited to a single analog signal. Multiple and different information can be transmitted on the same transmission line as long as they can be encoded digitally, resulting in the development of multivariable and self-contained smart sensors. The dividing line between sensor, actuator, and controller becomes more and more vague.

A DCS integrates various controllers of a process line into a coordinated, interactive system. It enables the users to manage the process as a complete system, with control over the interrelationship of the various subsystems. A DCS thus allows the users see the overall system and improve the overall efficiency and quality of the plant. The most important feature of a DCS is that the inter-communication is performed via a serial digital bus.

Table 6.1 presents the comparison between analog and digital communication.

DCS also offers advantageous features, among which is its ability to achieve both short-loop fast response advantage of localized-loop control and full monitoring advantage of centralized coordinated control. Furthermore, the failure of a piece of equipment has a local effect only, and it does not affect the entire system operation. Figure 6.2 shows an example of the configuration of a DCS.

DCS facilitates the integration of different control systems, which comprises continuous control and discrete event control. The DCS structure can involve more levels of hierarchy in a hierarchical DCS as shown in Figure 6.3 and further explained in Table 6.2. Hierarchical DCS is suitable for large-scale systems, where central computer of DCS may be overloaded, and facilitates the decomposition of responsibilities. The lowest two-level buses are related to communication on the part of the direct control systems (more commonly referred to as fieldbuses).

6.1.2 Issues of Proprietary Protocols

The fieldbus is the landmark feature of a distributed control system. It represents the communication system at the field level of the control hierarchy. It is essentially a

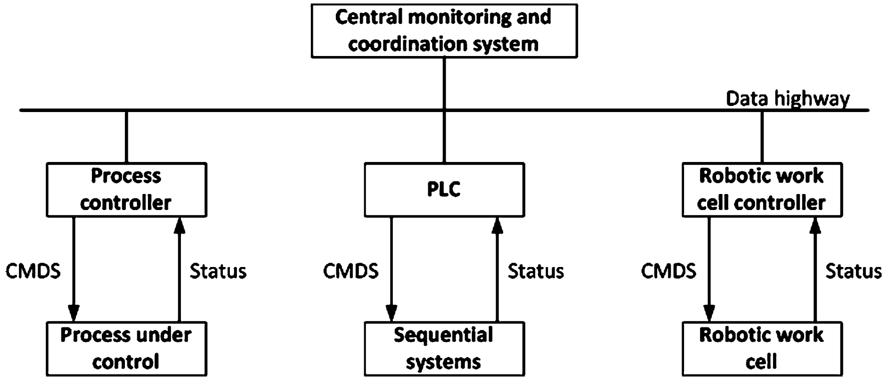

Fig. 6.2 Configuration of DCS

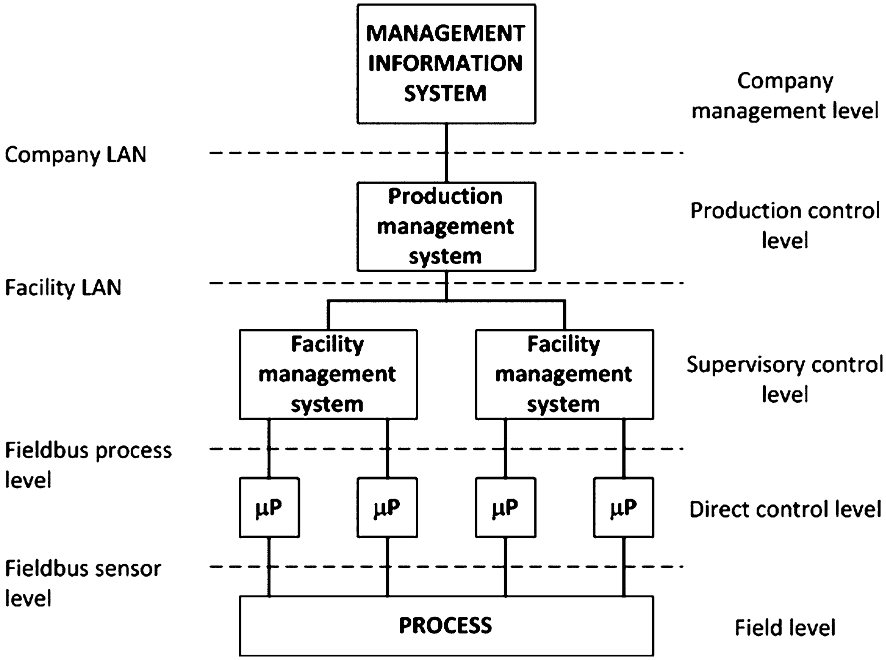

Fig. 6.3 Hierarchical of DCS

serial bus running on a low-cost physical medium such as unshielded twisted pair (UTP). It employs a serial interface standard such as the EIA-485 which allows for multi-drop communication. A serial protocol is sufficient to meet the communication requirements at the field level for two main reasons. Firstly, at the field level, typically short messages of a modest volume are expected, so that serial mode of

Table 6.2 Function of DCS levels

DCS level	Functions
Company management	Market analysis, orders processing, financial reports, *etc.*
Production control	Production scheduling, inventory control, production reporting, *etc.*
Supervisory control	Process coordination, adaptive control, performance monitoring, *etc.*
Direct control	Data acquisition, process monitoring, control, *etc.*
Field	Execution

communication is sufficient. Secondly, at this level, the protocol must be simple so that messages are sent and received quickly with minimum protocol overheads.

Currently, there are quite a number of fieldbuses offered on the market, most of them proprietary and usually associated with one or a few manufacturers. Unfortunately, to a user, this wide range of alternatives does not translate into more choices. Once one fieldbus is selected, the user will be limited by the range of compatible devices to use on that bus. The user may be bounded to a particular brand and will have to pay the proprietary margins associated with the brand, although his primary concern may be on the performance. Thus, to the users, it is good to have a standard for field communication. By standards, they need not involve state-of-the-art technology but rather widely-accepted guidelines which promote interoperability so that equipment from different vendors and manufacturers can communicate with each other over the standard fieldbus.

In the early 1980s, there were two main parallel attempts to establish such a standard. A number of mainly German companies developed suitable technologies and released them to public domain, *e.g.* Profibus and Interbus-S . These designs have been quickly standardized through German standardizing body, DIN. At the same time, a consortium of co-operating companies was created to develop common standards, the most notable one being Fieldbus Foundation in USA. The first approach resulted in relatively early release of useful standards. The European industry adopted these standards quite willingly. The most successful among them is arguably Profibus. The consortium approach was slow to deliver. In 1985, IEC formed a working group (Technical Committee 65C) to develop a fieldbus standard. The group also found it difficult to achieve useful results.

In 1992, a group of vendors left the IEC process and formed a new consortium, known as Interoperable Systems Project (ISP). The most prominent ISP members were Fisher Rosemount, Siemens, and Yokogawa. ISP largely followed Profibus. Other large vendors, including Honeywell, Allen–Bradley, and Bailey, who did not joined ISP, formed a rival group known as WorldFIP. They formed a standard based on FIP.

In September 1994, the two competing organizations merged into Fieldbus Foundation (FF). However, while there is progress with the number of FF installations increasing steadily, the FF cannot be said to have represented the de-facto standards. The progress is slow, mainly because it is difficult to reach a consensus among different manufacturers in the consortium. Furthermore, there are fieldbuses which have been around for a long time and have their niche group of followers in specific applications. For example, Profibus has been around since the mid 1980s, and it has its own followings, especially from the food and pharmaceutical industries.

Meanwhile, the Controller Area Network (CAN) bus is also rapidly gaining popularity very fast. Intel and Philips have produced CAN controller chips on a large scale with many automotive companies as strong backers, such as BMW and Mercedes.

A standard fieldbus has not arrived; rather, a number of fieldbuses, each attracting specific groups of users and industries, appears to be a more likely steady-state scenario.

6.2 Fieldbus Protocol Stack

The Open Systems Interconnect (OSI) model defines the main aspects of any communication system. The seven layers of the model are given below in Table 6.3.

This OSI model is a reference model towards standardization. It does not define the exact services or protocols, but just what each layer should do. The communication functions are implemented by the lowest three layers, and the host functions are implemented by the top four layers. The lowest two layers deal with intermediate points. The remaining layers have end-to-end significance.

Figure 6.4 shows the flow of a package of information between two points through the 7-layers protocol stack.

The OSI model is a general communication model. The protocol stacks for fieldbuses usually only involve the following layers:

- physical layer
- link layer
- network layer
- application layer

There may be a user layer above the application layer implementing high-level user functions, or these functions may be implemented as part of the application

Table 6.3 OSI model

Layers	Function
Application layer	Rules for interpreting data at source and destination
Presentation layer	Information representation and transformation
Session layer	Dialog management between sub-application
Transport layer	Method of achieving reliable data transfer across network
Network layer	Method of arranging communication across a complex network
Data-link layer	Rules for coordination between two directly connected units
Physical layer	Physical nature of links and signals

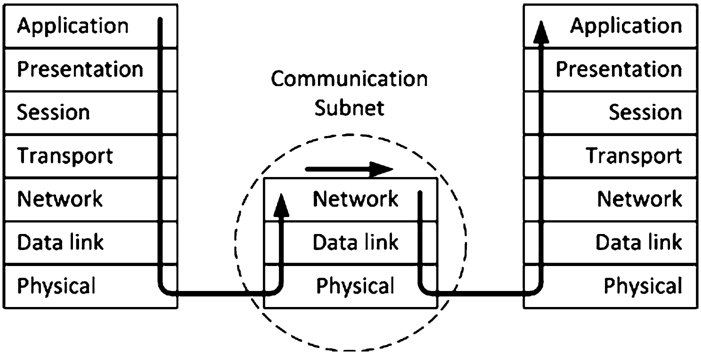

Fig. 6.4 Flow of information package

layer. The remaining layers do not matter, except in more complex systems for the following reasons:

- protocol overheads involved in traversing the full seven layers not acceptable for a real-time response solution
- field communication requirements not intensive (typically short to medium size messages), while protocol overheads low
- requirements of simplicity, ease of installation and maintenance, and low cost

In addition, the link and network layers are often closely tied together, with the network component being relatively simple. Some technical descriptions tend to treat such a combination of link/network layers, with the network part being only rudimentary, as a link layer.

6.2.1 Physical Layer

In this layer, the physical nature of data links and data signal are defined, *i.e.* the type of data link used, the physical topology, the nature of signal on the link (voltage, current, or frequency), the signaling rate, unit of data (bytes or words), the signal to logic translation, and the type of connectors to be used.

Examples of topology are given in Figures 6.5 and 6.6.

For fieldbuses, the most popular technologies at physical layer are as follows:

1. EIA-232 over ordinary wire

   EIA-232 is an American standard, published by Electronics Industries Association (EIA). It defines all the fundamental aspects of the physical layer. EIA-232 link is depicted in Figure 6.7.

   The following describes the characteristics of EIA-232:

   - EIA-232 uses ordinary wire, one for the signal and one for common ground.
   - All signals are defined relative to ground, called common mode; therefore it is sensitive to noise.

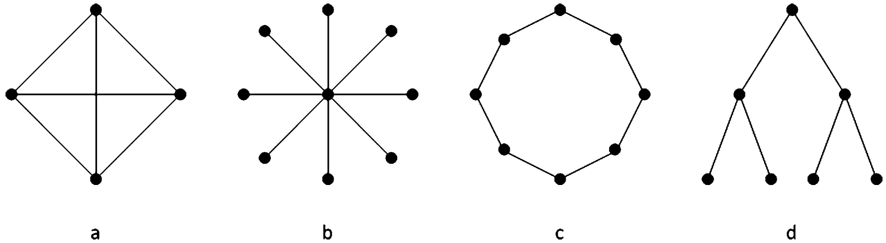

Fig. 6.5 Point-to-point topology: (**a**) complete topology, (**b**) star topology, (**c**) ring topology, (**d**) tree topology

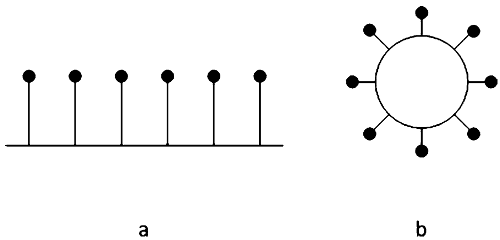

Fig. 6.6 Broadcast topology: (**a**) bus topology, (**b**) ring topology

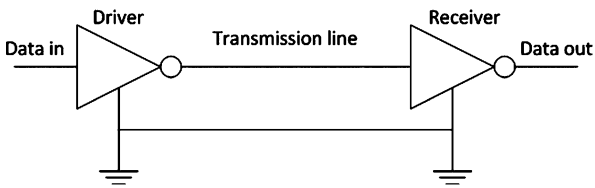

Fig. 6.7 EIA-232

- The topology of EIA-232 is point-to-point only.
- The driver should generate signal not exceeding 25V and in the range 5–15V for a load in the range 3000–7000.
- The receiver shall accept any signal lower than −3V as 1 and any higher than +3V as 0.
- The signaling rate is up to about 20kbps, with the cable length of up to about 15m.
- EIA-232 links can be used for full-duplex communication, *i.e.* both devices transmitting at the same time.
- EIA-232 does not define format of data (*e.g.* data bits, stop bit, and parity bit), which control signals to use, how to use the control signal (*e.g.* RTS or CTS), or whether the transmission is synchronous or asynchronous.
- EIA-232 is very popular and inexpensive, but its use in large-scale control is limited.

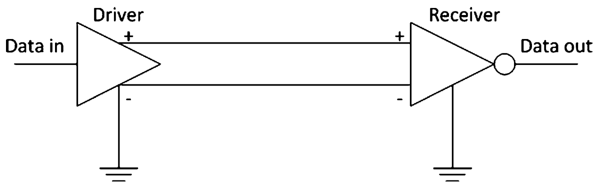

Fig. 6.8 EIA-485

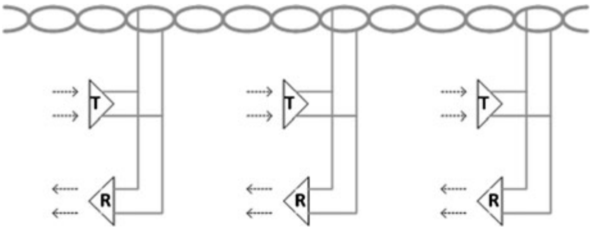

Fig. 6.9 Multi-drop link in RS-485

2. EIA-485 over twisted wire

   EIA-485 has been developed to overcome the deficiencies of EIA-232; the structure of which is depicted in Figure 6.8.

   The following describes the characteristics of EIA-485:

- EIA-485 contains twisted pair wires for each signal. The level of signal is defined as voltage difference between these wires, called differential mode.
- The level of signals is not affected by difference in ground potential at both ends, therefore less sensitive to noise.
- The topology of EIA-485 allows multi-drop and compliant devices to be connected to the same multi-drop link, as depicted in Figure 6.9.
- The driver should generate differential signal at least 1.5V for a load of 60Ω or more.
- The receiver shall accept any signal of 200mV or more.
- The signaling rate is up to 10Mbps.
- The permissible length depends on the transmission speed. It ranges from 4000ft for slow speed to 60ft for speed at 2Mbps.
- EIA-485 standard does not define method of encoding 0 and 1, format of data units, and link level protocol. For EIA-485 devices to communicate, these additional aspects must be identical.
- EIA-485 is very popular in process control area because of its low cost of implementation and wiring, multi-drop topology, good resistance to noise, and high data transfer rate.

For high-speed communication, coaxial cable or fiber optics can be used at the physical layer. Coaxial cable allows multi-drop connection and offers very high noise immunity. There are a number of physical standards for signals on a coaxial cable, such as Ethernet and Arcnet. Fiber optics is used for very high-speed

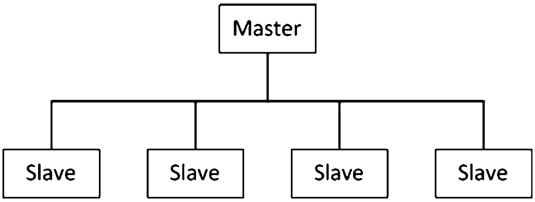

Fig. 6.10 Master–slave configuration

communications and in installations requiring extremely high noise immunity. Only point-to-point connections are possible.

6.2.2 Link Layer

The link layer defines rules of communication between a communicating entity (source or destination) and the element of the communication system that is directly connected to that entity.

The key elements involved at the link layer include:

- the establishment of a link between two points
- media-access control, *i.e.* the rules of 'the right to speak' (flow control), whether the link is half-duplex or full-duplex
- segregation of the link data and link commands
- error detection and correction techniques
- indication of link failure and the methods of recovery

Media access control may be centred on popular schemes such as master–slave, token bus, or carrier sense multiple access/collision detect (CSMA/CD).

A master–slave configuration for media access control is depicted in Figure 6.10.

Although this structure is simple, it has several disadvantages in control as follows:

- slow response
- only one unit as a master
- hard to provide prioritized responses

A more efficient control of the traffic of information can be accomplished with a token-bus or CSMA/CD approach. The token-bus method is shown in Figure 6.11.

The characteristics of token-bus traffic control are as follows:

- Token (special message) is passed from one user to another on the bus.
- Only the station with token has the right to transmit.
- Upon completion of message transmission, the token is passed to the next station.
- The timer keeps track of the maximum time a station can hold the token.

With this structure, the upper bound of time before transmission is known, making the overall system deterministic. The main difficulty, however, is when the token

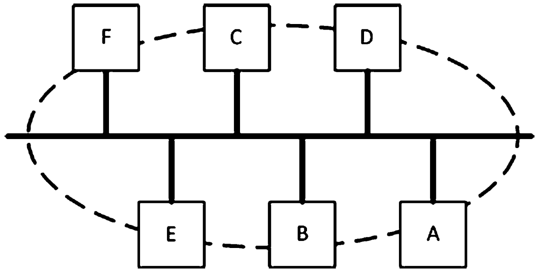

Fig. 6.11 Token-bus configuration

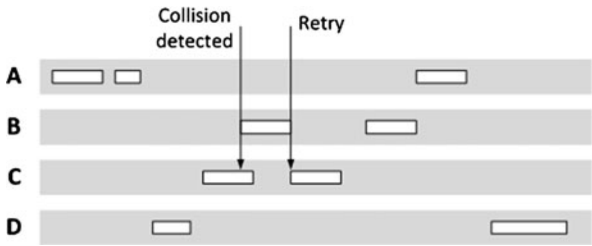

Fig. 6.12 CSMA/CD structure

is lost, which leads to temporary suspension of all communication until a recovery action is taken.

The CSMA/CD approach is depicted in Figure 6.12.

The node that wants to send a packet will first check if no node in the bus is sending packets. It will only send when no node is sending. When two nodes start at exactly the same time, they will detect collision and break off by checking if the received signal is the same as the signal they sent. After a collision, both nodes wait for an increasingly random amount of time.

The application of CSMA/CD traffic control is limited by the fact that it is impossible to guarantee how long a unit may have to try before it can succeed. In other words, CSMA/CD is less deterministic than token-bus.

Apart from media access control, error detection and correction are two important features that the link layer must be able to perform, since errors are ever present and happen in every system. Errors may be small (just one bit inverted) or larger (loss of whole command or response).

Error detection can be organized on the following levels:

- byte level, which is a parity bit
- data packet level, which is a checksum byte or a number of checksum bytes

If a whole message is lost, there is only one method to recover, which is retransmission. Self-recovery protocols are risky and usually unreliable.

6.2.3 Network Layer

The network layer defines the rules for end-to-end communication. It takes care of the method of addressing in the network and the routing of data.

To provide orderly communication, all elements of a network must have their addresses assigned. Addressing is an important issue to consider because of the following reasons:

- Address space must be sufficient. (If eight bits are allocated, the network cannot have more than 256 members. If too many bits are used than necessary, network performance will be degraded.)
- An address must be set in all network components, from sensors to controllers.
- Each address must be unique.

Practically, all fieldbuses allow for messages that are not addressed to any specific device but instead, are addressed to every device. Such messages are called broadcast messages.

6.2.4 Application Layer

The application layer defines the meaning of data transmitted and received. It is right at the access level for applications which require the network services. This layer will provide file, print, message, and application database services. It will identify and initiate the services necessary for a user's request. In a control system application, at this level, the code for various commands is defined with the format of response codes and data. This layer provides services for the various control functions.

6.2.5 User Layer

There may be a user layer which aims to implement high-level control functions more efficiently. It connects the individual plant areas and provides a high-level environment for applications. Function blocks are now commonly used at this level to realize the control strategies. Each fieldbus device is described with a Device Description (DD). The DD can be viewed as the driver for the specific device. It includes all variables descriptions and corresponding operating procedures to use the device, thus making the DD truly interoperable. Any control system or host will be able to communicate with the device once it has the DD for that device.

6.2.6 Traversing the Stack

When a transmitting device sends a message as a Packet Data Unit (PDA), the message will travel down through the layers to the physical medium at the physical level and subsequently move up the layers to the receiving device. When there is a request in the message, the receiving device will attend to it and responds in the reverse manner. While traversing down the stack, the original message from the transmitting device is added with a piece of information from each layer, and the same information is stripped off in the corresponding layer of the receiving device.

6.3 Common Fieldbuses

The selection of field devices is driven by the supporting protocols and vice versa. Devices from a particular manufacturer are usually conforming to particular protocols, so that there is usually no real option what protocol and supporting devices to use. In this section, some common fieldbuses, with wider user bases, will be presented.

6.3.1 CANopen

Controller Area Network (CAN) is a serial bus network of microcontrollers that connects devices, sensors, and actuators in a system or sub-system for real-time control applications. CAN provides many powerful features, including multi-master functionality and the ability to broadcast or multicast telegrams. CAN offers many other advantages, among which are the low cost, high data reliability, short response time, and a huge user base. These strong points put CAN among the leaders in fieldbus technology, especially in the automotive and textile industries.

In this protocol, the message is broadcast to all nodes in the network using an identifier unique to the network. Based on the identifier, the individual node decides whether or not to process the message and also determines the priority of the message in terms of competition for bus access. This method allows for uninterrupted transmission when a collision is detected, unlike Ethernet that will stop transmission upon collision detection.

CANopen is a CAN-based higher-layer protocol. It was developed as a standardized embedded network and designed for motion-oriented machine control networks, although it is now used in many other fields, such as medical equipment, off-road vehicles, maritime electronics, public transportation, building automation, *etc.*

CANopen provides a mechanism so that devices of different types can be integrated and can communicate in a standardized fashion. By making use of the device profile information, CANopen devices will ensure common operating functions. For example, two digital modules from two different manufacturers will have common functionality such as setting the outputs and reading the inputs. The profile specifies the functionality which must be common for devices to be interoperable. With CANopen, the manufacturers are not constrained in making various features in their devices.

The fundamental part of device profile is object dictionary, consisting of a mixture of data objects, communication objects, and commands/actions. CANopen services give the user full access to the object dictionary, allowing reading and writing of data and commands. Data and commands are implemented using a 16-bit index addressing mechanism together with 8-bit sub-index, giving an address range from 0000H to FFFFH. Parts of the object dictionary are divided into different areas based on functionality. Index 6000H, for example, is reserved for reading.

CANopen provides an open protocol and allows direct data exchange between nodes on the network without participation of a bus master unit. It allows full broadcast/multicast features and a variety of communication modes designed to keep bus loading minimal and predictable. Therefore, CANopen is well suited to the concept of remote intelligence and is ideal for distributed control solutions.

6.3.2 Profibus

Profibus is an open digital communication system which is mainly used in factory automation. It is now one of the market leaders in data communication, with a 20% share and over 2000 products. Profibus is now widely applied in the food and pharmaceutical industries.

The protocol structure of Profibus comprises three layers: physical layer (layer 1), data link layer (layer 2), and application layer (layer 7).

Physical layer describes standard serial communication EIA-485 using twisted and shielded pair, with a bus topology tree expandable using repeaters and a digital transmission NRZ coded.

Data link layer defines a logical model of the network nodes. The network nodes consist of passive stations (which use LAN transmission medium only under an active station request) and active stations (which communicate among themselves and with passive stations). In the active stations, the medium access follows the token passing rule, *i.e.* a defined bit sequence that allows mutual communication between devices and grants the use of LAN. A logical ring connects active stations and guarantees defined transmission time and lack of collision with other data in the bus cable, as only the station with the token can transmit, while the others remain inhibited. Medium access control in passive stations follows master–slave method, where the masters are the active stations. The data link layer is further divided into two sub-layers, one controlling medium access and the other giving higher-level interface with synchronous/asynchronous transmission services.

Application layer allows objects manipulation, with objects being variables, arrays, matrix, variable lists, program calls, subroutines, *etc.* It is divided into two sub-layers as follows:

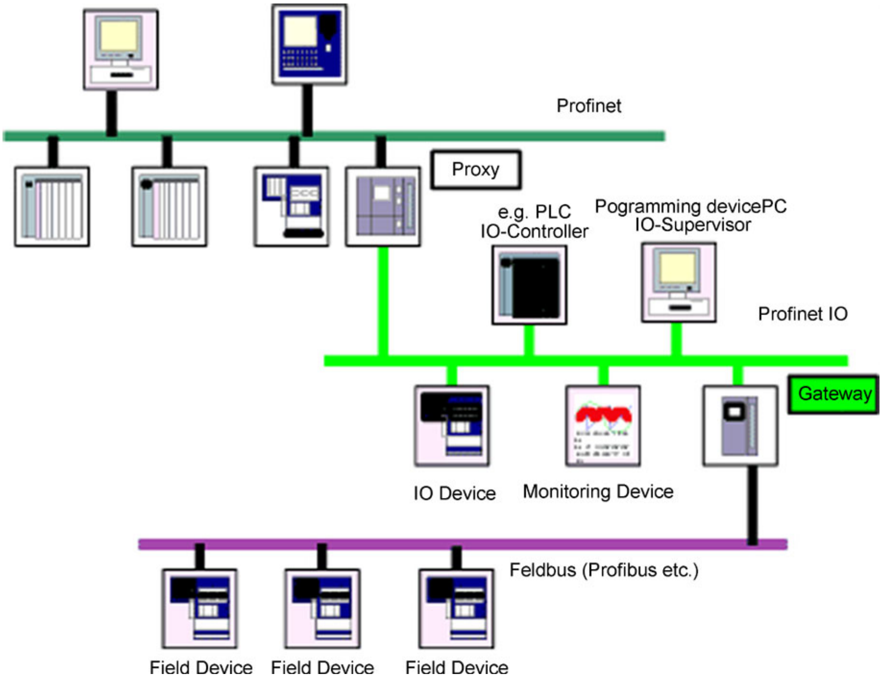

Fig. 6.13 Configuration of profibus

- Fieldbus Message Specification (FMS)
- Lower Layer Interface (LLI)

Interconnecting drives and steering unit with Profibus yields the possibility to realize control loops through LAN. In an electric drive operation, for example, the active station handles path generation and position loops, while the passive station handles speed and torque loops.

Figure 6.13 presents the typical configuration of Profibus.

6.3.3 Foundation Fieldbus

Foundation Fieldbus (FF) is a serial, two-way communication system that serves as the base-level network in factory automation. FF is typically implemented together with Ethernet as its hardware platform. FF defines two level networks, H1 and HSE, with 31.25Kbps and 100Mbps transfer rates, respectively.

The key concept of FF is a schedule timetable to ensure that all messages are transferred to correct destination nodes within a prescribed time. This contributes to FF capability of timely information access. The schedule timetable acts as a manager that controls information traffic; it determines when a message is sent, what message is to be sent, where it should be sent, *etc.* Schedule timetable is in essence an algorithm of information execution, and the content can differ from one algorithm to another; however, the massages usually include:

- periodic data (synchronous)
- request from user (asynchronous)
- request from devices (asynchronous)

All messages are transferred within one bus. Since there are several messages to be transferred, token mechanism is used to determine which should be transferred first, with priority given to synchronous message.

The structure and characteristics of FF make it suitable for diagnostics purposes, for example in valve diagnostics in a hydraulic application, which requires timely information access. In this application, process information, such as pressure, temperature, and level, is monitored. This information is then processed with statistical tools and compared to a threshold value, so that the condition of the process can be diagnosed. For example, an exponential rise in the statistical distribution of temperature may indicate that the system is overheating.

6.3.4 Firewire

Firewire (also known as i.Link or IEEE 1394) is a personal computer and digital video serial bus interface standard offering high-speed communications and real-time data services. Firewire is a successor technology to SCSI Parallel Interface. It was developed by Apple computer and has been widely used in the computer and consumer electronics industries.

The advantages of Firewire are as follows:

- It is a low-cost, self-powered, high-speed, digital single cable serial bus suited for real-time motion control applications.
- The bandwidth can be determined overhead.
- It is compatible with peer-to-peer communications.
- The speed ranges from 100 to 400Mbps.
- It supports 63 devices in one bus.
- Up to 10 meters cable length repeaters can be used to extend distance.
- It is compatible with digital plug-n-play setup with all parameters software driven.
- The interface has typically been included with every PC, with easy set-up and configuration.

The disadvantages of Firewire are as follows:

- There is no standard protocol.
- Firewire only addresses a small proportion of the real problems in developing an industry standard.

6.3.5 Sercos

Serial real-time communication system (Sercos) interface is a digital motion control bus that interconnects motion controls, drives, I/O and sensors. It is an open controller digital drive interface which is designed for high-speed serial communication of closed-loop data in real-time over a noise-immune, fibre optic cable. Sercos takes advantage of digital drive capabilities by not only replacing the standard ± 10 volt analog standard interfaces, but also providing two-way communications between control and drive.

The controls and drives use a standard medium for transmission, topology, connection techniques, signal levels, message structures, timing, and data formats. This allows devices from different Sercos' manufacturers to communicate with each other in the same platform.

The advantages of Sercos are as follows:

- Data is exchanged between control and drives via fibre optic rings, which eliminates the electromagnetic interference.
- The time taken to transmit command and actual values is very short, which guarantees an exact synchronization with axes.
- It supports four operating modes including torque, velocity, position control, and block mode.
- Sercos machines have plug-n-play capability, allowing easy adaptation to different applications.
- Implementation of Sercos improves system flexibility as one identical drive can be parameterized to handle multiple prime movers.
- The Sercos interface products from various vendors are interoperable.
- The master controls each ring, assigning timeslots to ensure proper transmission of data.

The disadvantages of Sercos are as follows:

- The speed of the Sercos interface has been pointed out as negative. The capability to transmit data 4 to 10 times faster than required does little to improve machine performance.
- The interface has 32,767 identification numbers for standard commands, and it also has the capability to incorporate 32,767 identification numbers. This sometimes makes one Sercos interface incompatible with the other.
- The cost of Sercos interface may be higher than that of other standards.

6.3.6 Ethernet

Ethernet was initially mainly used for office automation purposes with the flexibility in layout and interoperability with a majority of office networks. The development of Ethernet has brought the present technology to communicate at 100Mbps, allowing it to be used as a host for industrial networking.

Ethernet is a computer networking technique for local area network (LAN). The word itself comes from *ether*, which was perceived by scientists to be the medium in the outer space; Ethernet is in fact the hardware medium of communication of various devices in one network. Ethernet defines the connection in the physical layer and data link layer.

With Ethernet, devices with different protocols (but same Ethernet platform) can exist in the same local area network without conflict, although no communication is established due to different protocols. This simplifies the electrical installation of the network. Furthermore, Ethernet addressing schemes further eliminates the possibility of conflict among devices.

The topology of Ethernet follows the physical layer topology (although star topology is now commonly used). Signal attenuation due to the physics of the cable limits the length of the network, and in this situation hubs and repeaters are required. Their role is essentially to refresh the signal, so that the signal can now be transmitted over a long distance. There is still, however, a limit imposed in the number of hubs and repeaters by the data transfer speed. For example, in a speed of 10Mbps, the number of hubs and repeaters may not exceed four between two nodes (devices). Hubs and receivers broadcast a message to all ports so that each port only gets a part of the available bandwidth.

In an industrial networking, it is very likely that the network grows bigger that it becomes as if the entire system consists of several local networks. Another similar case occurs when there arises a merger among several small networks, each with their own Ethernet-based network, required to form a high-performance system. Ethernet bridge allows connection of several networks to be accomplished, even though each may work at a different speed. Ethernet bridge has an ability to direct the message to the intended recipient by making use of the Ethernet addressing scheme. This is the difference between Ethernet bridge and hubs, as hubs do not filter a message.

The connection of devices into Ethernet is executed via Network Interface Card (NIC). NIC can automatically select the correct speed it should work on.

Ethernet, from being a standard for office-level automation, is permeating into the industrial control environment. With an Industrial Ethernet backbone in place, diagnostics can easily be expanded, and service functions will be available network-wide and on a location-independent basis. Despite much development of Ethernet, the future of Ethernet is still full of challenges. Data capture is one of the many challenges which have to be overcome. It would be of great advantage if the plant data can be extracted from the field devices and local controllers. This requires a great deal of space in the controller memory. Current development in Ethernet technology includes the use of fibre optics and wireless communication.

Wireless Ethernet usually employs infrared or radio frequency communication and can be applied for both peer-to-peer and infrastructure communications. Although wireless Ethernet offers a great deal of flexibility, its effectiveness is significantly reduced by its incapability of transferring messages over a long distance (about 100–200m). In addition, wireless Ethernet is more prone to message collision as some nodes may not be able to recognize another transmission due to physical obstruction (unlike the case of wire Ethernet). This problem is usually solved by virtual checking, whereby a transmitting node sends a request to all nodes and waits for permission prior to sending a message. This, however, is still an ongoing research area.

Wireless Ethernet can significantly boost the performance of motion control systems, as it can reach the parts of network unreachable by cables, for example when chemical conditions prevents cables/wires to be used in the network.

6.4 Applications in Hydraulic/Pneumatic and Electric Drives

The various communication protocols as described above have been widely used in hydraulic and electric drives. It is now common to apply one among many fieldbuses, *e.g.* CAN, in the controller to provide multi-axis position control via high-speed serial bus. This allows communication of a far higher density of information to and from the controller and also enables the realization of multivariable sensors and integrated sensing, control, and actuation in a single system.

In both hydraulic/pneumatic and electric drives, this capability opens an opportunity to automate processes, covering larger range of parameters, such as position, pressure, speed, flow, *etc.* In addition, communication with external devices and systems is also possible, without additional cablings, which further enhances the implementation of a fieldbus-based controller.

In addition, the use of fieldbuses at various levels of automation allows the realization of smart devices. Smart devices refer to devices which have capability of making decision with minimum intervention from the user. A preliminary direction from the user is still necessary, but subsequently the devices will take care of the process by themselves by extracting information from process parameters according to certain statistical or mathematical tools; hence the term smart. Digital communication supports the operation of smart devices since it enables timely information and control, which is not possible with analog communication. The digital components of smart devices usually contain the following information:

- process parameters, such as set point, pressure (in hydraulic/pneumatic drives), or current (in electric drives)
- control functions, which control various process parameters as contained in the previous points

These functionalities are built internally in the devices themselves, so it is as if that the devices are able to perform the control function and seem to have their own intelligence. One key advantage of having a built-in controller is that the requirement of information handling can be compromised since the location of the sensors, actuators, and control elements has become very close to each other. Furthermore, with the speed achievable by fieldbuses, these smart devices can outperform their analog counterparts.

6.4.1 Fieldbuses in Hydraulic/Pneumatic Drives

A number of recent developments in hydraulics are enabled through the use of digital control electronics along with fieldbus interfaces. One example of such a development is an axis-control valve which offers complete axis controllers, representing another step in the direction of decentralized controller intelligence. The valve is controlled via fieldbuses, allowing for decentralization of control with a fast cycle time.

In terms of hardware, the incorporation of smart devices into an existing system—be it hydraulic or electric drive—should not change the system drastically; in fact, the changes should be as low as possible. By using fieldbuses, cabling requirement will be reduced significantly.

The main processor performs supervisory control tasks and takes care of the fieldbus communication. The second processor is a digital signal processor which executes all direct input and output operations, the coil-current control, and the actuation and evaluation of the position sensor such as LVDT.

The axis-control valve's interfaces are configured in such a way that all major axis-control functions in servo hydraulics can be realized from the compact device. Such high-density functionalities can be achieved via the fieldbus interface, as additional analog cablings will be far too extensive to realize the valve in this compact form. Essentially, with a fieldbus interface, different and more types of information can be transmitted both to and from the valve, thus enabling sensing, control, and actuation to be realized from within the same housing.

6.4.2 Fieldbuses in Electric Drives

By implementing digital control loop with high-bandwidth response and advanced signal processing algorithms, greater output torque can be achieved, contributing to higher-power density actuators. Integral motion control and programmable I/O enable complex functionalities to be performed, reducing the necessity to employ separate motion control hardware. This is because the use of a fieldbus, like CANopen, simplifies coordinated multi-axis motion control.

This design allows users to easily implement synchronized control to various components with same control architecture, with a cycle time of 500μs. The compromise made in the cycle time is compensated by the modularity and flexibility of the controller.

Chapter 7
Trends in Motion Control

The trends in motion control will be expected to be primarily driven by developments in precision control and engineering, which has been steadily gathering momentum and attention over the last century. The driving force of this development arises from the requirements of miniaturization in products, where more functionality is to be contained in a smaller dimension, as well as the delicate processes to be able to realize them.

The historical roots of precision engineering are arguably in the field of horology (developed since the 1300s) and optics. Major contributions were made in the late 1800s and early 1900s during the development of ruling engines for the manufacturing of scales, reticules, and spectrographic diffraction gratings. Today, precision engineering has come a long way, encompassing multidisciplinary technologies such as nanotechnology and biotechnology, which together are opening up new frontiers in precision applications, including motion control applications.

7.1 Background

7.1.1 Nanotechnology

The concept of nanotechnology was first enunciated by an American physicist, Dr. Richard Feynman, in 1959, while the term itself was coined by Professor Norio Taniguchi of Tokyo Science University, in 1974. The term was used to refer to the processing of a material to a nanometer-scale precision, primarily (at that time) using ultrasonic machining.

Nanotechnology is the science of manipulating atoms and molecules to fabricate materials, devices, and systems. The name *nano* comes from the size of atoms and molecules, which is measured in nanometer. The functioning of the devices in nanotechnology depends on each individual atom and molecule. Nanotechnology is becoming significantly important to many industrial applications and is poised to revolutionize new trends in the advancement of technology. With current progress in

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6_7,

precision engineering reaching sub-micrometer level, future development will be to achieve nanometer level, making nanotechnology the future of precision engineering.

There can be no doubt that many new and interesting developments and products will arise from today's nanoscience and nanotechnology work. Waves of product miniaturization to follow will see existing macro products replaced by Micro-system Technologies (MST) and nanotechnology products, produced by new nanotechnology-based manufacturing facilities. Nanotechnology is a major new technological force that will have substantial socio-economic effects throughout the world, and many benefits in standards of living and quality of life can be confidently expected.

7.1.2 Biotechnology

Biotechnology is the technology to manipulate the structure and function of biological systems, especially when used in food science, agriculture, and medicine. Modern biotechnology is often related to genetic alteration of living materials, such as microorganisms, plants, and animals.

Examples of biotechnology applications include the following:

- Minimally invasive surgery, assisted by remotely operated surgical instruments and diagnostic tools, *e.g.* micro-catheters down to 100-μm diameter incorporating optical fibers for delivery and retrieval of light images for high-resolution cameras, nano-scale sensors for measuring blood chemistry, and tip-mounted micro-turbine rotary cutters for arterial plaque removal
- Intracytoplasmic sperm injection (ICSI), a method to help fertilization by the injection of sperm to an egg cell, which requires high-precision actuator to minimize the damage imposed to the cell
- Accurate and efficient drug targeting and delivery by nano-particle technology, acting as medicinal bullets
- Replacement of damaged nerves by artificial equivalents
- Improved adhesion growth of living tissue cells on to prosthetic implants by micro- and nano-surface patterning of implant materials

There have been efforts to combine biotechnology and nanotechnology to result in yet more effective and powerful devices and technologies. An example is the fabrication of biochips, sometimes referred to as laboratories-on-a-chip, where arrays of 10^3 to 10^6 of pharmaceutical compounds are fabricated on a small chip.

The development of motion control is thus tightly coupled with the advancement of those enabling and application technologies. While the requirements from these technologies will set the pace of motion control, the progress will hinge on how motion control technology can suit their purpose.

In the rest of this chapter, precision technologies and applications which are direct or indirect beneficiaries of precision motion control, and/or enabling new challenges to the field will be elaborated.

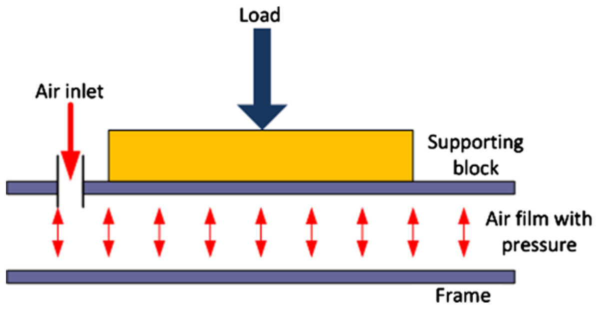

Fig. 7.1 Air bearing

7.2 Ultra-precision Machining

7.2.1 Ultra-precision Spindles

Ultra-precision spindles are used to drive loads with high speed and moderate or low torque. They are implemented in high-precision manufacturing devices such as high-speed turning and milling machine, as well as non-manufacturing devices such as higher performance magnetic memory disk file systems, high-definition large-scale projection television, and video cassette recorders. These applications call for highly precise positioning, which poses a challenge since it is to be accomplished at high speed.

To achieve the required specification, air-bearing is typically employed. The characteristic of interest of air-bearing is its low asynchronous error motion, making it possible to achieve high rotational accuracy. The disadvantage, however, is its low stiffness and dampening ability. Figure 7.1 shows the working diagram of an air-bearing, where pressurized air is used to keep the gap between the rotating and static parts of the machine (*e.g.* spindles).

The cutting edge technology being developed is the integration of air-bearing with conventional oil bearing to achieve high precision with yet high stiffness.

7.2.2 Excimer Laser Micromachining

Lasers, in particular excimer lasers, are today widely used for micromachining of different kinds of materials due to their unique pulsed ultra violet (UV) emission. They have been applied in the research laboratories since 1977, and about 10 years later, they were successfully introduced into industrial processing and manufacturing. Excimer lasers have been used for the highly precise marking of glass (such as in eyeglasses) and ceramics, especially in surface-mounted devices (SMD). In microelectronics production lines, drilling into printed-circuit boards can be performed with this technique. In semiconductor processes, it can be used as a direct writing

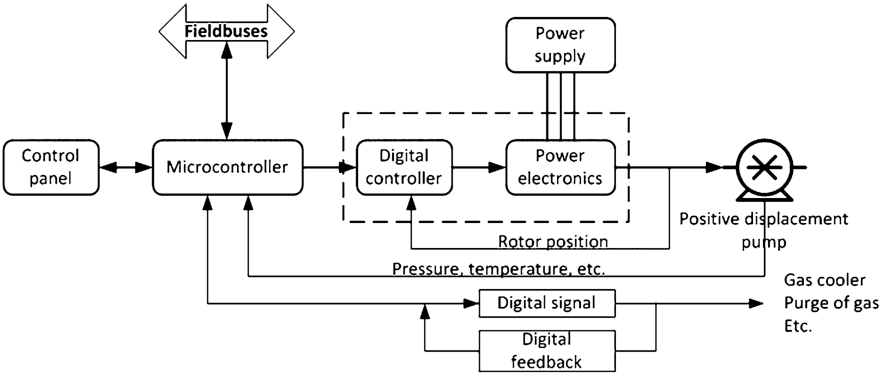

Fig. 7.2 Control of vacuum pump in excimer laser

tool to replace photomasks, as a micro-drill for multilayer chip, and as an ablation tool for non-chemical etching and repair in semiconductor processes.

Excimer laser is excited by a rare-gas halide or rare-gas metal vapor, often employing noble gas due to their stability. Controlling the flow and pressure of the gas is therefore necessary in order to maintain the precision. Excimer laser control includes controlling gas exhaust filter, vacuum pump, and gas mixer. Figure 7.2 shows an example of the control structure of a vacuum pump.

7.3 Micro-fabrication

7.3.1 Lithography

The semiconductor and microelectronics industry have led the development and application of the photo and electron beam lithography techniques which are expected to serve as the main basis for continuing miniaturization in large-scale production in the future. Features and dimensions are printed on silicon chips using a process called photolithography, in which UV light from a mercury vapor lamp is shone through a mask containing the features of the chip and projected onto the surface of the silicon wafers in a machine known as a photolithographic *stepper*, so called because it prints an image of one chip and then *steps* to the next location on the wafer to print the pattern for the next and so on. For feature size of smaller than 0.1μm, shorter wavelength radiation in the form of electron beam or X-ray lithography can be used.

Ultra-large-scale integration (ULSI) chips will be the harvest of precision lithography. These are fast becoming smaller, faster, and cheaper and are now equipped with more memory. They are expected to bring further massive improvements to the performance of microprocessors and computers and will, in turn, lead to direct benefits for telecommunications, domestic, automotive, and medical products and services.

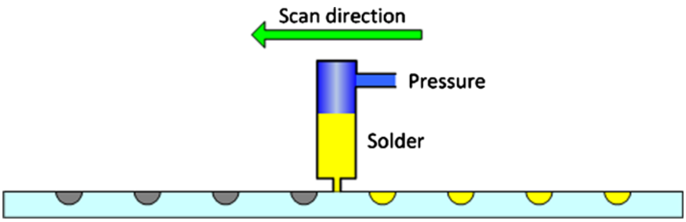

Fig. 7.3 Flip chip

7.3.2 Micro-electro-mechanical Systems (MEMS)

MEMS is the integration of mechanical and electronic elements, including sensors and actuators a common substrate, usually silicon. MEMS components are fabricated using micromachining processes that selectively etch away parts of the silicon wafer or add new structural layers to form the intended structures. With MEMS, it is possible to develop a system-on-chip; a term commonly used to refer to a multifunctional chip.

From the early examples of accelerometers and gyroscopes, MEMS products with micro-mechanical features, such as specialized sensors, arrays of sensors, and actuators fully integrated into the same silicon chip, are already burgeoning; applications are expected to expand in the navigational, automotive, biomedical, and pharmaceutical industries.

7.4 Micro-assembly

Another process involving high precision is in the area of pick-and-place micro-assembly. One example of micro-assembly process is a flip chip assembly.

A flip chip is a chip mounted on the substrate with various interconnect materials and methods, such as tape-automated bonding, flux-less solder bumps, wire interconnects, isotropic and anisotropic conductive adhesives, metal bumps, compliant bumps, and pressure contacts, as long as the chip surface (active area or I/O side) is facing the substrate.

One of the earliest flip chip technologies was solder-bumped flip-chip technology, as a possible replacement for the expensive, unreliable, low productivity, and manually operated face-up wire-bonding technology. Bumps are formed by injecting molten solder into etched cavities in a glass mold plate across a wafer. The mold plate is heated to just below melting point of the solder. The injector includes a slightly pressurized reservoir of molten solder of any composition. Figure 7.3 illustrates the process of solder bump deposition.

The use of flip-chip technologies in the manufacture of IC devices has increased tremendously in recent years. As the size of devices gets smaller, the precision required to align the solder bumps on the chip to the pads on the substrate becomes more crucial.

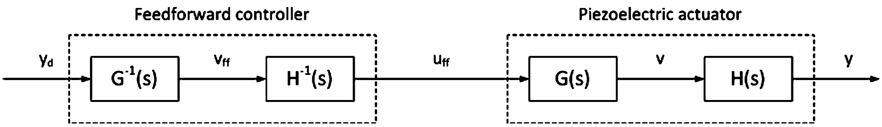

Fig. 7.4 A feedforward control to compensate for the hysteresis in a piezoelectric actuator

Besides flip-chip assembly, high-precision robots are also used to assemble micro-electronic and mechanical components.

7.5 Precision Metrology and Test

The measurement precision associated with Co-ordinate Measuring Machines (CMM) has been ever increasing over the years. When these machines are fitted with precision tools, such as a probe, vision device, or a microscope, special applications can be set up in the area of metrology and tests. One of the applications is Scanning Probe Microscopy (SPM).

An effective method to overcome the hysteresis in a piezoelectric actuator is using feedforward control compensation, which, unlike feedback, compensates for the deficit in its performance, as explained in Chapter 5. This is generally accomplished by selecting a suitable model of the hysteresis and then inverting the selected model to compensate for the dynamics and hysteresis effects. Figure 7.4 presents the concept of using feedforward control to compensate for the hysteresis.

7.6 Driving Technologies

7.6.1 Micromanufacturing

Manufacturing is an essential process in the industry. One of manufacturing process is machining, which is a process of removing excess or unwanted stock of material by the use of machine tools, such as cutting, grinding, and finishing. Conventional machining is executed via turning machine, drilling machine, milling machine, *etc.* While they are still in use, the development of machining to provide high-precision components has introduced non-conventional machining via laser cutting, hydrodynamic fluid, chemical substance, *etc.* Nowadays, there has been a trend towards non-contact machining as opposed to contact machining.

Micro-fabrication covers a range of manufacturing processes that produce patterns or layers of material to form microstructures. Lithography and MEMS (or MST) are common examples of micro-fabrication processes.

7.6.2 Microassembly

Microassembly is generally performed in one of two ways as follows:

- robotic manipulation through the use of macroscale robotic manipulators with microscale end effectors [26, 97]
- parallel self-assembly where structures are aggregated through stochastic interactions of components [73, 130]

7.6.3 Micrometrology

The final stage of production includes inspection of products, which essentially is a measurement and assessment whether to accept or reject the results of the production. This is where metrology comes into play.

One example of micrometrology is AFM.

The first generation of SPM is the Scanning Electron Microscope (SEM), where an electron beam is focused into a small spot on the object, and an electromagnetic raster is scanned across it. Images can be formed by collecting the secondary electrons generated by the impact of the impinging electron beam, by detecting the backscattered electrons or by detecting the generated X-rays. In this way, several different aspects of the object can be characterized, including morphology, average atomic number, and composition.

In another technique, Scanning Tunneling Microscope (STM), the evanescent wave is an electron wave function with an intrinsic wavelength of about 1 nm which extends beyond the surface of a sharp metal tip. If a conducting surface is brought to within about 1 nm of the tip and a potential difference is applied between them, then a tunneling current will be induced. The magnitude of this current is an exponentially decaying function of distance and is also dependent upon the difference between the work functions of the two materials. Thus, information can be derived of both the topography of the surface and its chemical composition. The limitation of STM is that it can only work with conducting surfaces.

Atomic force microscopy (AFM) is one of the foremost nanotechnology tools for measurement in the nanoscale [96], which is developed to overcome the limitation of STM. AFM employs a diamond stylus on a gold foil cantilever to lightly scan the surface of the specimen. The van der Walls attraction/repulsion force between the specimen's surface and the tip affects the resonance frequency of the cantilever—especially when the tip approaches the surface.

The change in cantilever resonance frequency is sensed as the tip approaches the sample surface and is affected by the van der Waals attraction. This type of microscopy has been used for a very wide range of surface characterization, including imaging and topography.

Figure 7.5 presents the working principle of AFM, showing mainly the control configuration of the process.

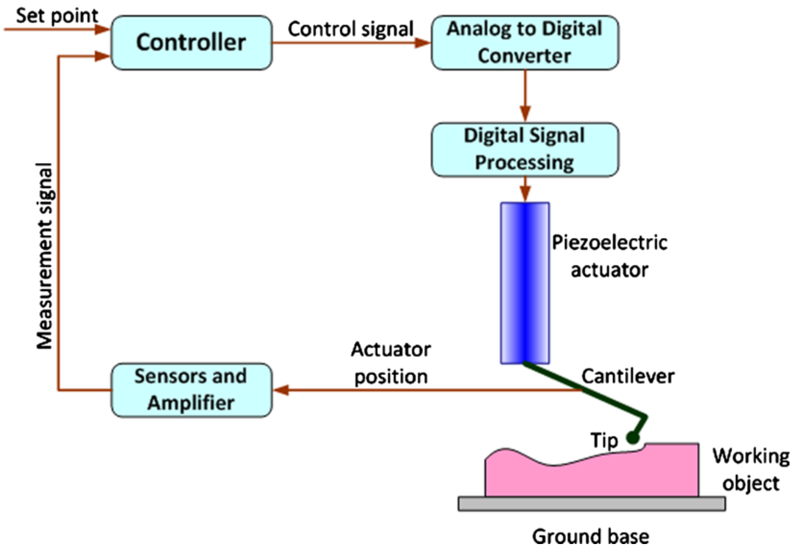

Fig. 7.5 AFM

In the process shown in Figure 7.5, a piezoelectric actuator is employed. Piezoelectric is a material, usually made from ceramic, with a capability of transforming an electrical signal (voltage) into a motion of the order of nanometer level, as explained in Chapter 4. The piezoelectric tube actuator drives the cantilever, which in turn drives a silicon probe tip. This tip is the source of electron that will develop the pattern on the working object. Therefore, the performance of the overall system is very much dictated by the motion precision of the piezoelectric actuator. The application of piezoelectric material in AFM is an example of how material science will also enhance and influence the developments of motion systems.

The use of piezoelectric actuator in AFM is, as explained in Chapter 4, limited by the hysteresis of the actuator, especially when the measurement involves a large range [6].

References

1. Ahmadi A, Ng SC, Liow SL, Ali J, Bongso A, Ratnam SS (1995) Intracytoplasmic injection of mouse oocytes with 5 mM Ca^{2+} at different intervals. Hum Reprod 10:431–435
2. Anderson EH (1994) Self-sensing piezoelectric actuator: Analysis and application to controlled structure. In: Proc 33rd struc dynam mater conf, pp 2141–2155
3. Aseltine JA, Mancini AR, Sartune CW (1958) A survey of adaptive control systems. IRE Trans Automat Contr 3(6):102–108
4. ASRM (2001) Patient's fact sheet: Intra-cytoplasmic sperm injection (ICSI). American Society for Reproductive Medicine
5. ATSDR (1999) Toxicological profile of mercury. Agency for Toxic Substances and Disease Registry, Atlanta, GA, USA
6. Barrett RC, Quale CF (1991) Optical scan-correction system applied to atomic force microscopy. Rev Sci Instrum 62(5):1393–1399
7. Braembussche PV, Swevers J, Brussel HV, Vanherck V (1996) Accurate tracking control of linear synchronous motor machine tool axes. Mechatronics 6(5):507–521
8. Braude P, Rowell P (2003) Assisted conception: In vitro fertilization and intra-cytoplasmic sperm injection. British Med J 327:852–855
9. Budinger M, Rouchon JF, Nogarede B (2004) Analytical modeling for the design of piezoelectric rotating-mode motor. IEEE/ASME Trans Mechatron 9(1):1–9
10. Cahyadi A, Yamamoto Y (2006) Hysteretic modeling of piezoelectric actuator attached on flexure hinge mechanism. In: Proc IEEE/ASME int conf intel robots syst, Beijing, China, pp 5437–5440
11. Caldwell WI (1950) Control system with automatic response adjustment. American Patent no 2,517,081
12. Canudas-de-Wit C, Olsson H, Astrom K, Lischinsky P (1995) A new model for control of systems with friction. IEEE Trans Automat Contr 40(3):419–425
13. Chen F, Xie H, Fedder GK (2001) A MEMS-based monolithic electrostatic microactuator for ultra-low magnetic disk head fly height control. IEEE Trans Magn 37(4):1915–1918
14. Chen XB, Kai J (2004) Modeling of positive-displacement fluid dispensing processes. IEEE Trans Electron Packag Manuf 27(3):157–163
15. Choi GH, Oh JH, Choi GS (1999) Repetitive tracking control of a coarse-fine actuator. In: Proc IEEE/ASME int conf advanced intel mechatronics, Atlanta, GA, USA, pp 335–340
16. Choi YH, Love CC, Chung YG, Varner DD, Westhusin ME, Burghardt RC, Hinrichs K (2002) Production of nuclear transfer horse embryos by piezo-driven injection of somatic cell nuclei and activation with stallion sperm cytosolic extract. Biol Reprod 67:561–567
17. Cohen J (1992) A review of clinical microsurgical fertilization. In: Cohen J, Malter HE, Talansky BE, Grifo J (eds) Micromanipulation of human gametes and embryos. Raven Press, New York

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6,

18. Cole DG, Clark RL (1994) Adaptive compensation of piezoelectric sensori-actuators. J Intel Mater Syst Struct 5:665–672
19. Cox WR (2001) Low-cost fiber collimation for MOEMS switches by ink-jet printing. SPIE MOEMS & Miniaturized Syst
20. Cox GF, Burger J, Lip V, Mau UA, Sperling K, Wu BL, Horsthemke B (2002) Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet 71:162–164
21. Crawley EF, de Luis J (1987) Use of piezoelectric actuators as elements of intelligent structures. AIAA J 25:1373–1385
22. Cruz-Hernandez JM, Hayward V (1997) On the linear compensation of hysteresis. In: Proc 36th IEEE conf decision contr, San Diego, CA, USA, pp 1956–1957
23. Cruz-Hernandez JM, Hayward V (1998) Reduction of major and minor hysteresis loops in a piezoelectric actuator. In: 37th IEEE conf decision control, Tampa, FL, USA, pp 4320–4325
24. Cuttino JF, Miller AC, Schinstock DE (1999) Performance of a fast tool servo for single-point diamond turning machines. IEEE/ASME Trans Mechatron 4:169–179
25. Davison DE, Gaudette D (2005) Tumor-tracking in radiotherapy: Parameterization of sensor time-delay compensators and associated performance limitations. In: Proc IEEE conf cont appl, Toronto, Canada, pp 131–136
26. Dechev N, Cleghorn WL, Mills JK (2004) Tether and joint design for micro-components used in microassembly of 3D microstructures. In: Proc SPIE—micromachining and microfabrication, Photonics West, pp 134–146
27. Denai MA, Attia SA (2002) Intelligent control of an induction motors. Electr Power Comput Syst 30(4):409–427
28. de Silva CW (2005) Mechatronics—an integrated approach. CRC Press, Boca Raton
29. Dosch JJ, Inman DJ, Garcia E (1992) A self-sensing piezoelectric actuator for colocated control. J Intell Mater Syst Struct 3(1):166–185
30. Du D, Wu C, Luo X, Zuo X (2006) Delay time identification and dynamic characteristics study on ANN soft sensor. In: Proc 6th int conf intel sys des appl, pp 42–45
31. Ediz K, Olgac N (2005) Effect of mercury column on the microdynamics of the piezo-driven pipettes. J Biomech Eng 127:531–535
32. Fleming AJ, Moheimani SOR (2003) Precision current and charge amplifiers for driving highly capacitive piezoelectric loads. Electron Lett 39(3):282–284
33. Fleming AJ, Moheimani SOR (2005) A grounded load charge amplifier for reducing hysteresis in piezoelectric tube scanners. Rev Sci Instrum 76(7)
34. Fuller CR, Gibbs GP (1994) Active control of interior noise in a business jet using piezoelectric actuators. In: Proc nat conf noise control eng, pp 389–394
35. Ge P, Jouaneh M (1995) Modeling hysteresis in piezoceramic actuators. Precis Eng 17: 211–221
36. Ge P, Jouaneh M (1996) Tracking control of a piezoceramic actuator. IEEE Trans Control Syst Technol 4:209–216
37. Gilles R, Dragan D, Nava S (2001) Separation of nonlinear and friction-like contributions to the piezoelectric hysteresis, pp 699–702
38. Grabowski PZ, Kazmierkowski MP, Bose BK, Blanbjerg F (2000) A simple direct-torque neuro-fuzzy control of PWM-inverter-fed induction motor drive. IEEE Trans Ind Electron 47(4):863–870
39. Goldfarb M, Celanovic N (1997) Modeling piezoelectric stack actuators for control of micromanipulation. IEEE Trans Control Syst Mag 17(3):69–79
40. Goto K (1997) Current status and future of micromanipulation-assisted fertilization in animals and human. J Reprod Dev 43(2):107–119
41. Guo C, Tani J (1994) Sensorless vibration control of a cantilever beam. In: Proc 71st ann conf Japanese soc mech eng II, pp 669–671
42. Haertling GH (1994) Rainbow ceramics: A new type of ultra-high-displacement actuator. Am Ceram Soc Bull 73(1):93–96
43. Hagood NW, Chung WH, von Flotow A (1990) Modeling of piezoelectric actuator dynamics for active structure control. J Intel Mater Syst Struct 1:327–354

44. Hayes DJ, Wallace DB, Cox WR (1999) MicroJet printing of solder and polymers for multi-chip modules and chip-scale packages. In: IMAPS intern conf high density packaging MCMs, pp 1–6
45. He JB, Wang QG, Lee TH (1998) PI/PID controller tuning via LQR approach. In: Proc 37th IEEE conf decision contr, vol 1, pp. 1177–1182
46. Hilbing JH, Heister SD (1996) Droplet size control in liquid jet breakup. J Phys Fluids 8:1574–1581
47. Hillenbrand S, Pandit M (1999) Discrete-time iterative learning control law with exponential rate of convergence. In: Proc 38th conf decision & control, Phoenix, AR, USA
48. Hirabayashi M, Kato M, Aoto T, Sekimoto A, Ueda M, Miyoshi I, Kasai N, Hochi S (2002) Offspring derived from intracytoplasmic injection of transgenic rat sperm. Transgenic Res 11:221–228
49. Hiremane R (2005) From Moore's law to Intel innovation—prediction to reality. Tech at Intel Mag, pp 1–9
50. Howell LL (2001) Compliant mechanisms. Wiley, New York
51. Hrapko M, van Dommelen JAW, Peters GWM, Wismans JSHM (2005) The mechanical behaviour of brain tissue: Large strain responses and constitutive modeling. In: Int IRCOBI conf, Prague, Czech Rep, pp 56–59
52. Huang S, Lee TH, Tan KK (2002) Robust adaptive numerical compensation for friction and force ripple in permanent magnet linear motors. IEEE Trans Magn 38(1):221–228
53. Hu AP, Register A, Sadegh N (1999) Using a learning controller to achieve accurate linear motor motion control. In: Proc IEEE/ASME inter conf advanced intell mechatron, Atlanta, GA, USA, pp 611–616
54. Hwang CL, Jan C (2003) A reinforcement discrete neuro-adaptive control for unknown piezoelectric actuator systems with dominant hysteresis. IEEE Trans Neural Netw 14(1): 66–78
55. IEEE Standard Board (1987) IEEE standard on piezoelectricity
56. Ikeda T (1990) Fundamentals of piezoelectricity. Oxford science publications. Oxford University Press, London
57. Iula A, Lamberti N, Carotenuto R, Pappalardo M (1997) A 3-D model of the classical Langevin transducer. IEEE Ultrason Symp 2:987–990
58. Jayacharndran V, Sun JQ (1998) Modeling shallow-spherical-shell piezoceramic actuators as acoustic boundary control elements. Smart Mater Struct 7:72–84
59. Jones L, Garcia E, Waites H (1994) Self-sensing control as applied to a PZT stack actuator used as a micropositioner. Smart Mater Struct 3(2):147–156
60. Jung SB, Kim SW (1994) Improvement of scanning accuracy of PZT piezoelectric actuator by feed-forward model-reference control. Precis Eng 16(1):49–55
61. Karray F, de Silva CW (2008) Introduction to the focused section on smart mechatronic systems and embedded design. IEEE/ASME Trans Mechatron 13(1):1–2
62. Kim JD, Nam SR (1995) Development of a micro-positioning grinding table using piezoelectric voltage feedback. IME Proc I J Syst Contr Eng 209:469–474
63. Kimura Y, Yanagimachi R (1995) Intracytoplasmic sperm injection in the mouse. Biol Reprod 52:709–720
64. Kurosawa M, Nakamura K, Okamoto T, Ueha S (1989) An ultrasonic motor using bending vibrations of a short cylinder. IEEE Trans Ultrason Ferroelectr Freq Control 36(5):517–521
65. Kwok WW, Davison DE (2006) A separation principle associated with sensor time delay compensation in feedback control. In: Proc IEEE conf cont appl, Munich, Germany, pp 3194–3199
66. Land JA, Evers JL (2003) Risk and complications in assisted reproduction techniques: Report of an ESHRE consensus meeting. Hum Reprod 18:455–457
67. Law WW, Liao WH, Huang J (2003) Vibration control of structures with self-sensing piezoelectric actuators incorporating adaptive mechanism. Smart Mater Struc 12(5):720–730
68. Leang KK, Zou Q, Devasia S (2009) Feedforward control of piezoactuators in atomic force microscope systems. IEEE Control Syst Mag 29:70–82

69. Lee TH, Tan KK, Lim SY, Dou HF (2000) Iterative learning of permanent magnet linear motor with relay automatic tuning. Mechatronic 10(1–2):169–190
70. Lee BR, Yang SY, Ahn KK (2003) Precision control of piezoelectric actuator using inverse hysteresis model and neuro control. In: 7th Korea–Russia intern symp sci tech, vol 1, pp 237–238
71. Leonard D, Krishnamurty M, Reaves CM, Denbaars SP, Petroff PM (1993) Direct formation of quantum-sized dots from uniform coherent islands of InGaAs on GaAs surfaces. Appl Phys Lett 63(23):3203–3205
72. Li X, Wang W, Chen Z (2005) New challenges in precision positioner development. In: Proc IEEE/ASME int conf adv intel mechatron, Monterey, CA, USA, pp 408–413
73. Liang SH, Xioang X, Böhringer K-F (2004) Towards optimal designs for self-alignment in surface tension driven micro-assembly. In: Proc IEEE conf MEMS, pp 9–12
74. Low TS, Guo W (1995) Modeling of a three-layer piezoelectric bimorph beam with hysteresis. J MEMS 4(4):230–237
75. Mayr O (1970) Origins of feedback control. MIT Press, Cambridge
76. Meintjes M, Graff KJ, Paccamonti D, Eilts BE, Cochran R, Sullivan M, Fall H, Godke RA (1996) In vitro development and embryo transfer of sperm-injected oocytes derived from pregnant mares. Theriogenology 1:304
77. Moriwaki T, Shamoto E (1997) Ultraprecision feed system based on walking drive. Ann CIRP 46(1):505–508
78. Moskalik AJ, Brei D (1996) Force-deflection behavior of individual unimorph piezoceramic C-block actuators. In: Proc ASME aerospace division conf, vol 52, pp 679–687
79. Nagy A, Gertsenstein M, Vintersten K, Behringer R (2003) Manipulating the mouse embryo: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
80. Newcomb CV, Flinn I (1982) Improving the linearity of piezoelectric ceramic actuators. Electron Lett 18(11):442–444
81. Olsson A, Astrom KJ, de Wit CC, Gafcert M, Lischinsky P (1998) Friction models and friction compensation. Eur J Control 3:176–195
82. Otten G, de Vries JA, van Amerongen J, Rankers AM, Gaal EW (1997) Linear motor motion control using a learning forward controller. IEEE Trans Mechatron 2(3):179–187
83. Palazzolo AB, Jagannathan S, Kascak AF, Montague GT, Kiraly LJ (1993) Hybrid active vibration control of rotor bearing systems using piezoelectric actuators. Trans ASME J Vibr Acoust 115:111–119
84. Palermo G, Joris H, Derde MP, van Steirteghem AC (1992) Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet 340:17–18
85. Pang CK, Guo G, Chen BM, Lee TH (2006) Self-sensing actuation for nanopositioning and active-mode damping in dual-stage HDDs. IEEE/ASME Trans Mechatron 11(3):328–338
86. Paros JM, Weisbord L (1965) How to design flexure hinges. Mach Des 37:151–156
87. Percin G, Khuri-Yakub BT (2002) Piezoelectrically actuated flextensional micromachined ultrasound droplet ejectors. IEEE Trans Ultrason Ferroelect Freq Control 49(6):756–766
88. Pernette E, Henein S, Magnani I, Clavel R (1997) Design of parallel robots in microrobotics. Robotica 15:417–420
89. Pervozvanski AA, de Wit CC (2002) Asymptotic analysis of the dither effect in systems with friction. Automatica 38:105–113
90. Physik Instrumente (1998) NanoPositioning
91. Pillarisetti A, Anjum W, Desai JP, Friedman G, Brooks AD (2005) Force feedback interface for cell injection. In: Proc joint europhatics conf symp haptic interfaces virtual environ teleoperator syst
92. Pilch M, Erdman C (1987) Use of break-up time data and velocity history data to predict the maximum size of stable fragments for acceleration-induced break-up of a liquid drop. Int J Multiph Flow 13:741–757
93. Pota HR, Alberts TE (1995) Multivariable transfer functions for a slewing piezoelectric laminate beam. Trans ASME J Dyn Syst Meas Control 117:352–359
94. Precup RE, Preitl S, Korondi P (2007) Fuzzy controller with maximum sensitivity for servo systems. IEEE Trans Ind Electron 54(3):1298–1310

95. Saggere L, Kota S, Crary SB (1994) A new design for suspension of linear microactuators. Proc Dyn Syst Control 55(2):671–675
96. Salapaka SM, Salapaka MV (2008) Scanning probe microscopy. IEEE Control Syst Mag 28(2):65–83
97. Skidmore G, Ellis M, Geisberger A, Tsui K, Tuck K, Saini R, Udeshi T, Nolan M, Stallcup R, Von Ehr J II (2004) Assembly technology across multiple length scales from the micro-scale to the nano-scale. In: Proc 17th IEEE int conf MEMS, pp 588–592
98. Sofikitis NV, Miyagawa I, Agapitos E, Pasyianos P, Toda T, Hellstrom WJG, Kawamura H (1994) Reproductive capacity of the nucleus of the male gamete after completion of meiosis. J Assist Reprod Genet 11:335–341
99. Song JK, Washington G (2000) Mechatronic design and control of singly and doubly curved composite mesoscale actuator systems. IEEE/ASME Trans Mechatron 5(1):49–57
100. Santa K, Mews M, Riedmiller M (1998) A neural approach for the control of piezoelectric micromanipulation robots. J Intell Robot Syst 22:351–374
101. Sinoda H, Sasaki S, Nakamura K (2000) Instantaneous evaluation of friction based on ARTC tactile sensor. In: Proc IEEE int conf robotics automation, pp 2173–2178
102. Shah VG, Hayes DJ (2002) Trimming and printing of embedded resistors using demand-mode ink-jet technology and conductive polymer. IPC Printed Circuit Expo
103. Shetty D, Kolk RA (1997) Mechatronics system design. PWS Publishing, Boston
104. Sira-Ramirez H (1993) Nonlinear pulse width modulation controller design, variable structure control for robotics and aerospace applications. Elsevier, Amsterdam
105. Sitti M (2003) Piezoelectrically-actuated four-bar mechanism with two flexible links for micromechanical flying insect thorax. IEEE/ASME Trans Mechatron 8(1):26–36
106. Sloane AJ (2002) High throughput peptide mass fingerprinting and protein microarray analysis using chemical printing strategies. Mol Cell Proteomics 1:490–499
107. Smith S (2000) Flexures, elements of elastic mechanisms. Taylor & Francis, London
108. Stepanenko Y, Su CY (1998) Intelligent control of piezoelectric actuators. In: Proc 37th IEEE conf decision contr, Tampa, FL, USA, pp 4234–4239
109. Szczech JB, Megaridis CM, Gamota DR, Zhang J (2002) Fine-line conductor manufacturing using drop-on-demand PZT printing technology. IEEE Trans Electron Packag Manuf 25(1):26–33
110. Takahashi M, Kurosawa M (1995) Direct friction driven surface acoustic wave motor. In: 8th inter conf solid-state sens act, Stockholm, Sweden, pp 401–404
111. Takemura K, Maeno T (2001) Design and control of an ultrasonic motor capable of generating multi-DOF motion. IEEE/ASME Trans Mechatron 6(4):499–506
112. Takigami T, Oshima K, Hayakawa Y, Ito M (1998) Application of self-sensing actuator to control of a soft-handling gripper. In: Proc IEEE int conf cont appl, Trieste, Italy, pp 902–906
113. Tan KK, Wang QG, Hang CC (1999) Advances in PID control. Springer, New York
114. Tan KK, Ng SC, Xie Y (2002) Optimal intra-cytoplasmic sperm injection with a piezo micromanipulator. In: Proc 4th world cong intel contr auto, pp 1120–1123
115. Tan KK, Zhao S (2003) PI control of ram velocity in injection molding machines with iterative learning. Comput Control Eng J 14(6):6–9
116. Tan KK, Xie Y, Lee TH (2003) Automatic friction identification and compensation with a self-adapting dual relay. Intell Auto Soft Comp
117. Tao G, Kokotovic PV (1995) Adaptive control of plants with unknown hysteresis. IEEE Trans Automat Control 40:200–212
118. Trease BP, Moon Y-M, Kota S (2005) Design of large-displacement compliant joints. Trans ASME 127:788–798
119. Tsai KY, Yen JY (1999) Servo system design of a high-resolution piezodriven fine stage for step-and-repeat microlithography systems. In: Proc IEEE IECON'99, vol 1, pp 11–16
120. Tzou HS (1993) Piezoelectric shells, distributed sensing and control of continua. Kluwer Academic, Dordrecht
121. US Patent 4894579 (1990)
122. US Patent 5225750 (1993)

123. US Patent 5229679 (1993)
124. US Patent 6415995
125. Vinson JR (1993) The behavior of shells composed of isotropic and composite materials. Kluwer Academic, Dordrecht
126. Vipperman JS, Clark RL (1995) Hybrid analog and digital adaptive compensation of piezoelectric sensori-actuators. In: Proc AIAA/ASME adaptive struct forum, New Orleans, LA, USA, pp 2854–2859
127. Wai RJ, Lin CM, Peng YF (2004) Adaptive hybrid control for linear piezoelectric ceramic motor drive using diagonal recurrent CMAC network. IEEE Trans Neural Net 15(6):1491–1506
128. Wang J, Hu X (2000) Current situation of nano-positioning technology. Mech Des Study 2:43–44
129. Wang Z, Ho DWC, Liu X (2004) Robust filtering under randomly varying sensor delay with variance constraints. IEEE Trans Circ Syst II:51 (6):320–326
130. Whitesides GM, Grzybowski B (2002) Self-assembly at all scales. Science 295(5564):2418–2421
131. Wikander J (1993) Autonomous mechatronic actuators as a basis for modular intelligent machines. In: Inter conf indust electron cont instr, Maui, HI, USA, pp 21–26
132. Wikander O (2000) Handbook of ancient water technology: technology and change in history. Brill, Leiden, pp 371–400
133. Woldringh GH, Kremer JAM, Braat DDM, Meuleman EJH (2005) Intracytoplasmic sperm injection: A review of risk and complications. Br J Urol 96:749–753
134. Wulp H (1997) Piezo-driven stages for nanopositioning with extreme stability: Theoretical aspects and practical design considerations. Delft Univ Tech
135. Xie WF (2007) Sliding-mode-observer-based adaptive control for servo actuator with friction. IEEE Trans Ind Electron 54(3):1517–1527
136. Xu QC, Dogan A, Tressler J, Yoshikawa S, Newnham RE (1991) Ceramic-metal composite actuator. In: Proc IEEE ultrasonics symp, pp 923–928
137. Yanagida K, Katayose H, Yazawa H, Kimura Y, Konnai K, Sato A (1998) The usefulness of a piezo-micromanipulator in intra-cytoplasmic sperm injection in humans. Hum Reprod 14(2):448–453
138. Yao B, Xu L (2000) Adaptive robust precision motion control of linear motors with ripple force compensations: Theory and experiments. In: Proc 2000 IEEE inter conf contr app, Anchorage, Alaska, USA
139. Yao B, Xu L (2002) Adaptive robust motion control of linear motors for precision manufacturing. Mechatronics 12:595–616
140. Yoon KJ, Park KH, Lee SK, Goo NS, Park HC (2004) Analytical design model for a piezo-composite unimorph actuator and its verification using lightweight piezo-composite curved actuators. Smart Mater Struct 13(3):459–467
141. Yu GR, Hwang RC (2004) Optimal PID speed control of brushless DC motors using LQR approach PID speed control of brushless DC motors using LQR approach. Proc IEEE Int Conf Syst Man Cybern 1:473–478
142. Zhang B, Zhu Z (1997) Developing a linear piezomotor with nanometer resolution and high stiffness. IEEE/ASME Trans Mechatron 2(1):22–29
143. Zhou H, Tan KK, Lee TH (2000) Micro-positioning of linear piezoelectric motors based on a learning nonlinear PID controller. In: Proc 39th IEEE conf decision cont, Sydney, Australia, vol 12, pp 913–918

Index

K.K. Tan, A.S. Putra, *Drives and Control for Industrial Automation*,
Advances in Industrial Control,
DOI 10.1007/978-1-84882-425-6, © Springer-Verlag London Limited 2011

Other titles published in this series (continued):

Soft Sensors for Monitoring and Control of Industrial Processes
Luigi Fortuna, Salvatore Graziani, Alessandro Rizzo and Maria G. Xibilia

Adaptive Voltage Control in Power Systems
Giuseppe Fusco and Mario Russo

Advanced Control of Industrial Processes
Piotr Tatjewski

Process Control Performance Assessment
Andrzej W. Ordys, Damien Uduehi and Michael A. Johnson (Eds.)

Modelling and Analysis of Hybrid Supervisory Systems
Emilia Villani, Paulo E. Miyagi and Robert Valette

Process Control
Jie Bao and Peter L. Lee

Distributed Embedded Control Systems
Matjaž Colnarič, Domen Verber and Wolfgang A. Halang

Precision Motion Control (2nd Ed.)
Tan Kok Kiong, Lee Tong Heng and Huang Sunan

Optimal Control of Wind Energy Systems
Julian Munteanu, Antoneta luliana Bratcu, Nicolaos-Antonio Cutululis and Emil Ceangă

Identification of Continuous-time Models from Sampled Data
Hugues Gamier and Liuping Wang (Eds.)

Model-based Process Supervision
Arun K. Samantaray and Belkacem Bouamama

Diagnosis of Process Nonlinearities and Valve Stiction
M.A.A. Shoukat Choudhury, Sirish L. Shah and Nina F. Thornhill

Magnetic Control of Tokamak Plasmas
Marco Ariola and Alfredo Pironti

Real-time Iterative Learning Control
Jian-Xin Xu, Sanjib K. Panda and Tong H. Lee

Deadlock Resolution in Automated Manufacturing Systems
ZhiWu Li and MengChu Zhou

Model Predictive Control Design and Implementation Using MATLAB®
Liuping Wang

Predictive Functional Control
Jacques Richalet and Donal O'Donovan

Fault-tolerant Flight Control and Guidance Systems
Guillaume Ducard

Fault-tolerant Control Systems
Hassan Noura, Didier Theilliol, Jean-Christophe Ponsart and Abbas Chamseddine

Detection and Diagnosis of Stiction in Control Loops
Mohieddine Jelali and Biao Huans (Eds.)

Stochastic Distribution Control System Design
Lei Guo and Hong Wang

Dry Clutch Control for Automotive Applications
Pietro J. Dolcini, Carlos Canudas-de-Wit and Hubert Béchart

Advanced Control and Supervision of Mineral Processing Plants
Daniel Sbárbaro and René del Villar (Eds.)

Active Braking Control Design for Road Vehicles
Sergio M. Savaresi and Mara Tanelli

Active Control of Flexible Structures
Alberto Cavallo, Giuseppe de Maria, Ciro Natale and Salvatore Pirozzi

Induction Motor Control Design
Riccardo Marino, Patrizio Tomei and Cristiano M. Verrelli

Fractional-order Systems and Controls
Concepcion A. Monje, YangQuan Chen, Blas M. Vinagre, Dingyu Xue and Vincente Feliu

Model Predictive Control of Wastewater Systems
Carlos Ocampo-Martinez