



THIRD EDITION

INSTRUMENTATION AND CONTROL SYSTEMS

WILLIAM BOLTON



INSTRUMENTATION AND CONTROL SYSTEMS

THIRD EDITION

INSTRUMENTATION AND CONTROL SYSTEMS

THIRD EDITION

WILLIAM BOLTON



Newnes is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2021 Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

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

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress.

ISBN: 978-0-12-823471-6

For Information on all Newnes publications
visit our website at <https://www.elsevier.com/books-and-journals>

Publisher: Mara Conner
Acquisitions Editor: Sonnini R. Yura
Editorial Project Manager: Alice Grant
Production Project Manager: Prasanna Kalyanaraman
Cover Designer: Matthew Limbert

Typeset by MPS Limited, Chennai, India



Contents

Preface	ix		
Acknowledgement	xiii		
1. Measurement Systems	1	2.5 Fluid Flow	29
1.1 Introduction	1	2.5.1 Differential Pressure Methods	29
1.1.1 Systems	1	2.5.2 Turbine Meter	31
1.2 Instrumentation Systems	2	2.5.3 Ultrasonic Time of Flight Flow Meter	32
1.2.1 The Constituent Elements of an Instrumentation System	2	2.5.4 Vortex Flow Rate Method	32
1.3 Performance Terms	4	2.5.5 Coriolis Flow Meter	33
1.3.1 Resolution, Accuracy, and Error	4	2.6 Liquid Level	33
1.3.2 Range	6	2.6.1 Floats	34
1.3.3 Precision, Repeatability, and Reproducibility	7	2.6.2 Displacer Gauge	34
1.3.4 Sensitivity	7	2.6.3 Differential Pressure	34
1.3.5 Stability	8	2.6.4 Load Cell	34
1.3.6 Dynamic Characteristics	9	2.6.5 Electrical Conductivity Level Indicator	35
1.4 Dependability	9	2.6.6 Capacitive Level Indicator	35
1.4.1 Reliability	10	2.6.7 Ultrasonic Level Gauge	35
1.5 Requirements	11	2.6.8 Nucleonic Level Indicators	36
1.5.1 Calibration	12	2.7 Temperature Sensors	36
1.5.2 Safety Systems	13	2.7.1 Bimetallic Strips	36
Problems	14	2.7.2 Liquid in Glass Thermometers	36
2. Instrumentation System Elements	17	2.7.3 Resistance Temperature Detectors (RTDs)	36
2.1 Introduction	18	2.7.4 Thermistors	37
2.2 Displacement Sensors	18	2.7.5 Thermocouples	37
2.2.1 Potentiometer	18	2.7.6 Thermodiodes and Transistors	40
2.2.2 Strain-Gauged Element	19	2.7.7 Pyrometers	40
2.2.3 Capacitive Element	20	2.8 Sensor Selection	41
2.2.4 Linear Variable Differential Transformer	21	2.9 Signal Processing	42
2.2.5 Optical Encoders	21	2.9.1 Resistance to Voltage Converter	42
2.2.6 Moiré Fringes	22	2.9.2 Temperature Compensation	45
2.2.7 Optical Proximity Sensors	23	2.9.3 Thermocouple Compensation	45
2.2.8 Mechanical Switches	24	2.9.4 Protection	47
2.2.9 Capacitive Proximity Sensor	25	2.9.5 Analogue-to-Digital Conversions	47
2.2.10 Inductivity Proximity Sensor	26	2.9.6 Digital-to-Analogue Conversions	50
2.3 Speed Sensors	26	2.9.7 Microcontroller Systems	51
2.3.1 Optical Methods	26	2.9.8 Op-Amps	52
2.3.2 Incremental Encoder	26	2.9.9 Pressure-to-Current Converter	56
2.3.3 Tachogenerator	26	2.10 Signal Transmission	56
2.4 Fluid Pressure Sensors	26	2.10.1 Noise	59
2.4.1 Diaphragm Sensor	27	2.11 Smart Systems	60
2.4.2 Piezoelectric Sensor	28	2.11.1 MEMS	61
2.4.3 Bourdon Tube	28	2.12 Data Presentation Element	62
		2.12.1 Indicator	62
		2.12.2 Illuminative Displays	62
		2.12.3 Graphical User Interface (GUI)	64
		2.12.4 Data Loggers	64
		2.12.5 Printers	65
		Problems	66

3. Measurement Case Studies	71	5.3.1 Proportional Band	107
3.1 Introduction	71	5.3.2 Limitations of Proportional Control	109
3.2 Case Studies	72	5.4 Derivative Control	110
3.2.1 A Temperature Measurement	72	5.4.1 PD Control	110
3.2.2 An Absolute Pressure Measurement	73	5.5 Integral Control	112
3.2.3 Detection of the Angular Position of a Shaft	74	5.5.1 PI Control	112
3.2.4 Air Flow Rate Determination	74	5.6 PID Control	114
3.2.5 Fluid Level Monitoring	75	5.6.1 PID Process Controller	116
3.2.6 Measurement of Relative Humidity	75	5.7 Tuning	117
3.2.7 Dimension Checking	76	5.7.1 Process Reaction Tuning Method	118
3.2.8 Temperature of a Furnace	77	5.7.2 Ultimate Cycle Tuning Method	120
3.2.9 Automobile Tyre Pressure Monitoring	77	5.7.3 Quarter Amplitude Decay	121
3.2.10 Control System Sensors with Automobiles	77	5.7.4 Lambda Tuning	121
3.3 Data Acquisition Systems	78	5.7.5 Software Tools	121
3.3.1 Data Acquisition Software	80	5.7.6 Adaptive Controllers	122
3.3.2 Data Loggers	80	5.8 Digital Systems	122
3.4 Testing	80	5.8.1 Embedded Systems	123
3.4.1 Maintenance	80	5.9 Fuzzy Logic Control	124
3.4.2 Common Faults	81	5.9.1 Fuzzy Logic	124
Problems	82	5.9.2 Fuzzy Logic Control Systems	126
		5.9.3 Fuzzy Logic Controller	127
		5.9.4 Fuzzy Logic Tuning of PID Controllers	130
4. Control Systems	85	5.10 Neural Networks	130
4.1 Introduction	85	5.10.1 Neural Networks for Control	132
4.2 Control Systems	86	Problems	133
4.2.1 Open- and Closed-Loop Control	87		
4.3 Basic Elements	89	6. Correction Elements	137
4.3.1 Basic Elements of a Closed-Loop System	89	6.1 Introduction	137
4.4 Case Studies	91	6.1.1 The Range of Actuators	138
4.4.1 Control of the Speed of Rotation of a Motor Shaft	91	6.2 Pneumatic and Hydraulic Systems	138
4.4.2 Control of the Position of a Tool	92	6.2.1 Current to Pressure Converter	138
4.4.3 Power Steering	92	6.2.2 Pressure Sources	138
4.4.4 Control of Fuel Pressure	93	6.2.3 Control Valves	139
4.4.5 Antilock Brakes	93	6.2.4 Actuators	140
4.4.6 Thickness Control	94	6.3 Directional Control Valves	141
4.4.7 Control of Liquid Level	95	6.3.1 Sequencing	143
4.4.8 Robot Gripper	95	6.3.2 Shuttle Valve	145
4.4.9 Machine Tool Control	97	6.4 Flow Control Valves	146
4.4.10 Fluid Flow Control	97	6.4.1 Forms of Plug	147
4.5 Discrete-Time Control Systems	98	6.4.2 Rangeability and Turndown	149
4.6 Digital Control Systems	99	6.4.3 Control Valve Sizing	149
4.7 Hierarchical Control	100	6.4.4 Valve Positioners	151
Problems	101	6.4.5 Other Forms of Flow Control Valves	151
		6.4.6 Fail-Safe Design	152
		6.5 Motors	152
5. Process Controllers	103	6.5.1 D.C. Motors	152
5.1 Introduction	103	6.5.2 Brushless Permanent Magnet D.C. Motor	154
5.1.1 Direct and Reverse Actions	104	6.5.3 Stepper Motor	154
5.1.2 Dead Time	104	6.6 Case Studies	159
5.1.3 Capacitance	104	6.6.1 A Liquid Level Process Control System	159
5.2 On–Off Control	104	6.6.2 Milling Machine Control System	159
5.2.1 Relays	106	6.6.3 A Robot Control System	159
5.3 Proportional Control	107	Problems	160

7. PLC Systems	165	9.6 Sensitivity	221
7.1 Introduction	165	9.6.1 Sensitivity to Changes in Parameters	221
7.2 Logic Gates	166	9.6.2 Sensitivity to Disturbances	223
7.2.1 Field-Programmable Gate Arrays	169	Problems	223
7.3 PLC System	170	10. System Response	227
7.4 PLC Programming	171	10.1 Introduction	227
7.4.1 Logic Gates	173	10.2 Inputs	227
7.4.2 Latching	174	10.3 Determining Outputs	228
7.4.3 Internal Relays	174	10.3.1 Partial Fractions	230
7.4.4 Timers	175	10.4 First-Order Systems	233
7.4.5 Counters	176	10.4.1 First-Order System Parameters	235
7.5 Testing and Debugging	178	10.5 Second-Order Systems	237
7.6 Case Studies	179	10.5.1 Second-Order System Parameters	240
7.6.1 Signal Lamp to Monitor Operations	179	10.6 Stability	245
7.6.2 Cyclic Movement of a Piston	179	10.6.1 The s Plane	246
7.6.3 Sequential Movement of Pistons	179	10.7 Steady-State Error	250
7.6.4 Central Heating System	181	Problems	252
Problems	182	11. Frequency Response	257
8. System Models	189	11.1 Introduction	257
8.1 Introduction	189	11.1.1 Sinusoidal Signals	258
8.1.1 Static Response	189	11.1.2 Complex Numbers	258
8.1.2 Dynamic Response	189	11.2 Sinusoidal Inputs	260
8.2 Gain	190	11.2.1 Frequency Response Function	260
8.2.1 Gain of Systems in Series	191	11.2.2 Frequency Response for First-Order Systems	262
8.2.2 Feedback Loops	191	11.2.3 Frequency Response for Second-Order Systems	264
8.3 Dynamic Systems	193	11.3 Bode Plots	265
8.3.1 Mechanical Systems	193	11.3.1 Transfer Function a Constant K	266
8.3.2 Rotational Systems	194	11.3.2 Transfer Function $1/s^n$	266
8.3.3 Electrical Systems	196	11.3.3 Transfer Function s^m	267
8.3.4 Thermal Systems	198	11.3.4 Transfer Function $1/(1 + \tau s)$	267
8.3.5 Hydraulic Systems	200	11.3.5 Transfer Function $(1 + \tau s)$	269
8.4 Differential Equations	202	11.3.6 Transfer Function $\omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$	270
8.4.1 First-Order Differential Equations	202	11.3.7 Transfer Function $(s^2 + 2\zeta\omega_n s + \omega_n^2)/\omega_n^2$	272
8.4.2 Second-Order Differential Equations	204	11.4 System Identification	274
8.4.3 System Identification	205	11.5 Stability	276
Problems	206	11.5.1 Stability Measures	277
9. Transfer Function	209	11.6 Compensation	278
9.1 Introduction	209	11.6.1 Changing the Gain	278
9.2 Transfer Function	210	11.6.2 Phase-Lead Compensation	280
9.2.1 Transfer Function	211	11.6.3 Phase-Lag Compensation	281
9.2.2 Transfer Functions of Common System Elements	212	Problems	283
9.2.3 Transfer Functions and Systems	213	12. Nyquist Diagrams	287
9.3 System Transfer Functions	214	12.1 Introduction	287
9.3.1 Systems in Series	214	12.2 The Polar Plot	287
9.3.2 Systems with Feedback	215	12.2.1 Nyquist Diagrams	289
9.4 Block Manipulation	216	12.3 Stability	290
9.4.1 Blocks in Series	217	12.4 Relative Stability	292
9.4.2 Moving Takeoff Points	217	Problems	293
9.4.3 Moving a Summing Point	217		
9.4.4 Changing Feedback and Forward Paths	217		
9.5 Multiple Inputs	220		

13. Control Systems	297	13.8.2 Networks	319
13.1 Introduction	297	13.8.3 Control Area Network (CAN)	320
13.2 Controllers	298	13.8.4 Automated Assembly Lines	321
13.2.1 Proportional Steady-State Offset	300	13.8.5 Automated Process Plant Networks	321
13.2.2 Disturbance Rejection	301	13.8.6 PLC Networks	323
13.2.3 Integral Wind-Up	301	13.8.7 Supervisory Control and Data Acquisition (SCADA)	325
13.2.4 Bumpless Transfer	302	13.8.8 The Common Industrial Protocol (CIP)	326
13.3 Frequency Response	302	13.8.9 Security Issues	326
13.4 Systems with Dead Time	303	Problems	327
13.5 Cascade Control	305	Answers	329
13.6 Feedforward Control	306	Appendix A: Errors	341
13.7 Digital Control Systems	307	Appendix B: Differential Equations	347
13.7.1 The z -Transform	309	Appendix C: Laplace Transform	353
13.7.2 The Digital Transfer Function $G(z)$	311	Appendix D: The z-Transform	361
13.7.3 PID Controller	314	Index	367
13.7.4 Software Implementation of PID Control	317		
13.8 Control Networks	317		
13.8.1 Data Transmission	318		

Preface

This book provides a first-level introduction to instrumentation and control engineering and as such is suitable for the BTEC units of *Industrial Process Controllers* and *Industrial Plant and Process Control* for the National Certificates and Diplomas in Engineering, and the unit *Control Systems and Automation* for the Higher National Certificates and Diplomas in Engineering and also providing a basic introduction to instrumentation and control systems for undergraduates. The book aims to give an appreciation of the principles of industrial instrumentation and an insight into the principles involved in control engineering.

The book integrates actual hardware with theory and analysis, aiming to make the mathematics of control engineering as readable and approachable as possible.

STRUCTURE OF THE BOOK

The book has been designed to give a clear exposition and guide readers through the principles involved in the design and use of instrumentation and control systems, reviewing background principles where necessary. Each chapter includes worked examples, multiple-choice questions and problems; answers are supplied to all questions and problems. There are numerous case studies in the text indicating applications of the principles.

PERFORMANCE OUTCOMES

The following indicate the outcomes for which each chapter has been planned. At the end of the chapters the reader should be able to:

Chapter 1: Measurement systems

Read and interpret performance terminology used in the specifications of instrumentation.

Chapter 2: Instrumentation system elements

Describe and evaluate sensors commonly used with instrumentation used in the measurement of position, rotational speed, pressure, flow, liquid level, temperature and the detection of the presence of objects.

Describe and evaluate methods used for signal processing and display.

Chapter 3: Measurement case studies

Explain how system elements are combined in instrumentation for some commonly encountered measurements.

Chapter 4: Control systems

Explain what is meant by open and closed-loop control systems, the differences in performance between such systems.

Explain the principles involved in some simple examples of open and closed-loop control systems.

Describe the basic elements of digital control systems.

Chapter 5: Process controllers

Describe the function and terminology of a process controller and the use of two-step, proportional, derivative and integral control laws.

Explain PID control and how such a controller can be tuned.

Explain what is meant by fuzzy logic and how it can be used for control applications.

Explain what is meant by artificial neural networks and how they can be used for control applications.

Chapter 6: Correction elements

Describe common forms of correction/regulating elements used in control systems.

Describe the forms of commonly used pneumatic/hydraulic and electric correction elements.

Chapter 7: PLC systems

Describe the functions of logic gates and the use of truth tables.

Describe the basic elements involved with PLC systems.

Devise programs to enable PLCs to carry out simple control tasks.

Chapter 8: System models

Explain how models for physical systems can be constructed in terms of simple building blocks.

Chapter 9: Transfer function

Define the term transfer function and explain how it is used to relate outputs to inputs for systems.

Use block diagram simplification techniques to aid in the evaluation of the overall transfer function of a number of system elements.

Chapter 10: System response

Use Laplace transforms to determine the response of systems to common forms of inputs.

Use system parameters to describe the performance of systems when subject to a step input.

Analyse systems and obtain values for system parameters.

Explain the properties determining the stability of systems.

Derive the steady-state error for a basic closed-loop control system.

Chapter 11: Frequency response

Explain how the frequency response function can be obtained for a system from its transfer function.

Construct Bode plots from a knowledge of the transfer function.

Use Bode plots for first and second-order systems to describe their frequency response.

Use practically obtained Bode plots to deduce the form of the transfer function of a system.

Compare compensation techniques.

Chapter 12: Nyquist diagrams

Draw and interpret Nyquist diagrams.

Chapter 13: Control systems

Explain the reasons for the choices of P, PI, or PID controllers.

Explain the effect of dead time on the behaviour of a control system.

Explain the uses of cascade control and feedforward control.

Explain the principles of digital control systems and the use of the z-transform to analyse them.

Describe the principles involved in control networks.

Describe the principles involved in Fieldbus.

Describe the principles of CAN, SCADA, DSC and CIP control networks.

Identify the issues involved in maintaining a secure system.

SOFTWARE TOOLS

Details of programs and methods suitable for their development have not been included in this book. It was felt to be more appropriate to leave such development to more specialist texts such as MATLAB and SIMULINK for Engineers by Agam Kumar Tyagi (Oxford Higher Education 2011), A Guide to MATLAB: For Beginners and Experienced Users by B. R. Hunt and R. L. Lipsman (Cambridge University Press 2014), Hands-On Introduction to LabView for Scientists and Engineers by John Essick (Oxford University Press 2012), and Labview for Everyone: Graphical Programming Made Easy and Fun by Jeffrey Travis (Prentice Hall, 2006).

CHANGES FOR THE 3RD EDITION

The major changes introduced to the third edition are a discussion of dependability that has been included in Chapter 1, the discussion of smart systems extended and an introduction to radio telemetry for data transmission. A discussion of interactive and non-interactive forms of PID control and integrator windup has been added to Chapter 5, and it also now includes a revised discussion of steady-state error and fuzzy logic and artificial neural networks for control applications. Chapter 10 extends the discussion of the steady-state error. Chapter 13 extends the discussion of the z-transform and bus systems used with control networks, introducing the HART Communication Protocol, Fieldbus and CIP control networks, and also extends the discussion of security issues. An appendix has been included on the basic features of the z-transform.

W. Bolton

Acknowledgement

I am grateful to all those who reviewed the previous edition and made very helpful suggestions for this new edition.

1

Measurement Systems

OUTLINE

1.1 Introduction	1	1.3.5 Stability	8
1.1.1 Systems	1	1.3.6 Dynamic Characteristics	9
1.2 Instrumentation Systems	2	1.4 Dependability	9
1.2.1 The Constituent Elements of an Instrumentation System	2	1.4.1 Reliability	10
1.3 Performance Terms	4	1.5 Requirements	11
1.3.1 Resolution, Accuracy, and Error	4	1.5.1 Calibration	12
1.3.2 Range	6	1.5.2 Safety Systems	13
1.3.3 Precision, Repeatability, and Reproducibility	7	Problems	14
1.3.4 Sensitivity	7		

1.1 INTRODUCTION

This chapter is an introduction to the instrumentation systems used for making measurements and deals with the basic elements of such systems and the terminology used to describe their performance in use.

1.1.1 Systems

The term *system* will be freely used throughout this book, and so here is a brief explanation of what is meant by a system and how we can represent systems.

If you want to use an amplifier then you might not be interested in the internal working of the amplifier but what output you can obtain for a particular input. In such a situation we can talk of the amplifier being a system and describe it by means of specifying how the output is related to the input. With an engineering system an engineer is often more interested in the inputs and outputs of a system than the internal workings of the component elements of that system.

A *system* can be defined as an arrangement of parts within some boundary which work together to provide some form of output from a specified input or inputs. The boundary divides the system from the environment and the system interacts with the environment by means of signals crossing the boundary from the environment to the system, i.e. inputs, and signals crossing the boundary from the system to the environment, i.e. outputs (Figure 1.1).

A useful way of representing a system is as a *block diagram*. Within the boundary described by the box outline is the system, and inputs to the system are shown by arrows entering the box and outputs by arrows leaving the box. Figure 1.2 illustrates this for an electric motor system; there is an input of electrical energy and an output of

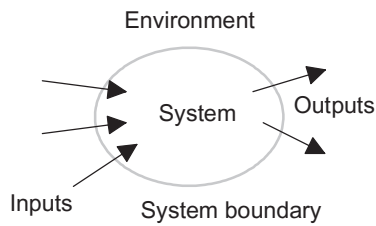


FIGURE 1.1 A system.

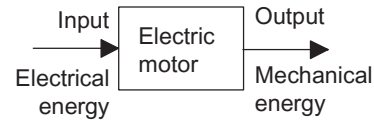


FIGURE 1.2 Electric motor system.

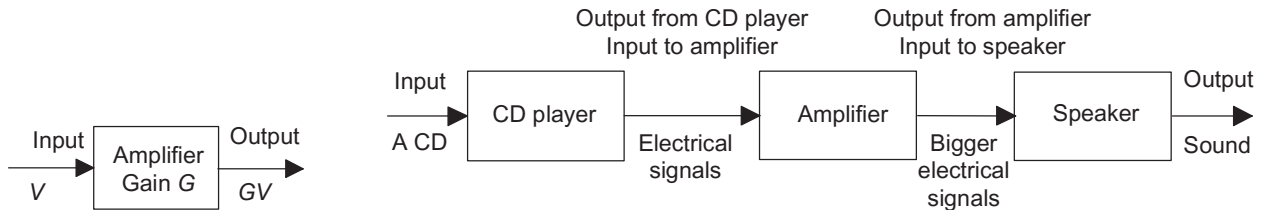


FIGURE 1.3 Amplifier system.

FIGURE 1.4 Interconnected systems.

mechanical energy, though you might consider there is also an output of waste heat. The interest is in the relationship between the output and the input rather than the internal science of the motor and how it operates. It is convenient to think of the system in the box operating on the input to produce the output. Thus, in the case of an amplifier system (Figure 1.3) we can think of the system multiplying the input V by some factor G , i.e. the amplifier gain, to give the output GV .

Often we are concerned with a number of linked systems. For example, we might have a CD player system linked to an amplifier system, which, in turn, is linked to a loudspeaker system. We can then draw this as three interconnected boxes (Figure 1.4) with the output from one system becoming the input to the next system. In drawing a system as a series of interconnected blocks, it is necessary to recognise that the lines drawn to connect boxes indicate a flow of information in the direction indicated by the arrow and not necessarily physical connections.

1.2 INSTRUMENTATION SYSTEMS

The purpose of an *instrumentation system* used for making measurements is to give the user a numerical value corresponding to the variable being measured. Thus a thermometer may be used to give a numerical value for the temperature of a liquid. We must, however, recognise that, for a variety of reasons, this numerical value may not actually be the true value of the variable. Thus, in the case of the thermometer, there may be errors due to the limited accuracy in the scale calibration, or reading errors due to the reading falling between two scale markings, or perhaps errors due to the insertion of a cold thermometer into a hot liquid, lowering the temperature of the liquid and so altering the temperature being measured. We thus consider a measurement system to have an input of the true value of the variable being measured and an output of the measured value of that variable (Figure 1.5). Figure 1.6 shows some examples of such instrumentation systems.

An *instrumentation system* for making measurements has an input of the true value of the variable being measured and an output of the measured value. This output might be then used in a control system to control the variable to some set value.

1.2.1 The Constituent Elements of an Instrumentation System

An instrumentation system for making measurements consists of several elements which are used to carry out particular functions. These functional elements are:

1. Sensor

This is the element of the system which is effectively in contact with the process for which a variable is being measured and gives an output which depends in some way on the value of the variable and which can

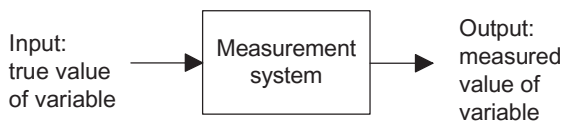


FIGURE 1.5 An instrumentation/measurement system.

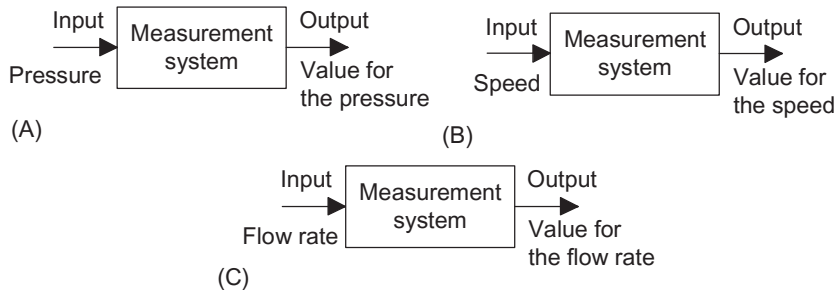


FIGURE 1.6 Examples of instrumentation systems: (A) pressure measurement, (B) speedometer, (C) flow rate measurement.

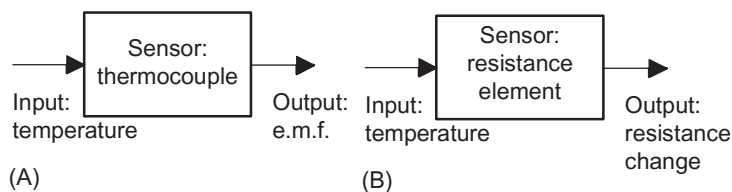


FIGURE 1.7 Sensors: (A) thermocouple, (B) resistance thermometer.

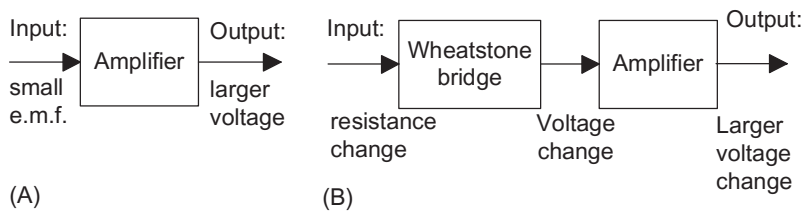


FIGURE 1.8 Examples of signal processing.

be used by the rest of the measurement system to give a value to it. For example, a thermocouple is a sensor which has an input of temperature and an output of a small e.m.f. (Figure 1.7A) which in the rest of the measurement system might be amplified to give a reading on a meter. Another example of a sensor is a resistance thermometer element which has an input of temperature and an output of a resistance change (Figure 1.7B).

2. Signal processor

This element takes the output from the sensor and converts it into a form which is suitable for display or onward transmission in some control system. In the case of the thermocouple this may be an amplifier to make the e.m.f. big enough to register on a meter (Figure 1.8B). There often may be more than an item, perhaps an element which puts the output from the sensor into a suitable condition for further processing and then an element which processes the signal so that it can be displayed. The term *signal conditioner* is used for an element which converts the output of a sensor into a suitable form for further processing. Thus in the case of the resistance thermometer there might be a signal conditioner, such as a Wheatstone bridge, which transforms the resistance change into a voltage change, then an amplifier to make the voltage big enough for display (Figure 1.8B) or for use in a system used to control the temperature.

3. Data presentation

This presents the measured value in a form which enables an observer to recognise it (Figure 1.9). This may be via a display, e.g. a pointer moving across the scale of a meter or perhaps information on a visual display unit (VDU). Alternatively, or additionally, the signal may be recorded, e.g. in a computer memory, or transmitted to some other system such as a control system.

Figure 1.10 shows how these basic functional elements form a measurement system.

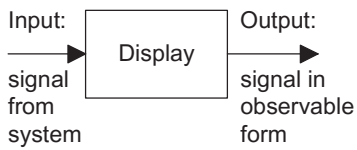


FIGURE 1.9 A data presentation element.

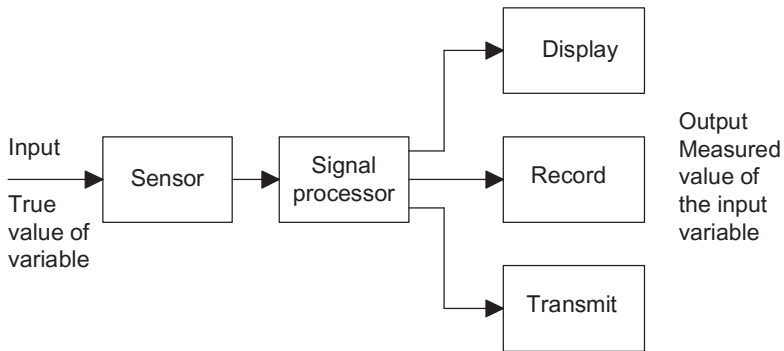


FIGURE 1.10 Measurement system elements.

The term *transducer* is often used in relation to measurement systems. Transducers are defined as an element that converts a change in some physical variable into a related change in some other physical variable. It is generally used for an element that converts a change in some physical variable into an electrical signal change. Thus sensors can be transducers. However, a measurement system may use transducers, in addition to the sensor, in other parts of the system to convert signals in one form to another form.

EXAMPLE

With a resistance thermometer, element A takes the temperature signal and transforms it into resistance signal, element B transforms the resistance signal into a current signal, element C transforms the current signal into a display of a movement of a pointer across a scale. Which of these elements is (a) the sensor, (b) the signal processor, (c) the data presentation?

The sensor is element A, the signal processor element B, and the data presentation element is C. The system can be represented by Figure 1.11.

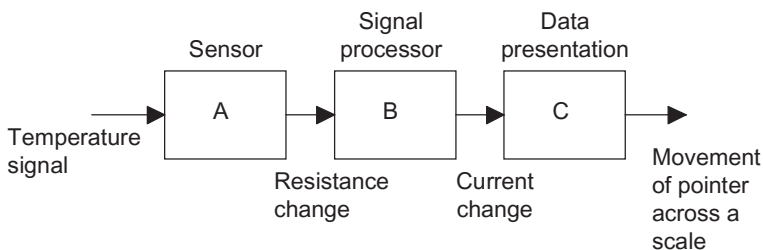


FIGURE 1.11 Example.

1.3 PERFORMANCE TERMS

The following are some of the more common terms used to define the performance of measurement systems and functional elements.

1.3.1 Resolution, Accuracy, and Error

Resolution is the smallest amount of an input signal change that can be reliably detected by an instrument. Resolution as stated in a manufacturer's specifications for an instrument is usually the least-significant digit (LSD) of the instrument or in the case of a sensor the smallest change that can be detected. For example, the OMRON ZX-E displacement sensor has a resolution of $1\ \mu\text{m}$.

Accuracy is the extent to which the value indicated by a measurement system or element might be wrong. For example, a thermometer may have an accuracy of $\pm 0.1^\circ\text{C}$. Accuracy is often expressed as a percentage of the full

range output or full-scale deflection (f.s.d). For example, a system might have an accuracy of $\pm 1\%$ of f.s.d. If the full-scale deflection is, say, 10 A, then the accuracy is ± 0.1 A. The accuracy is a summation of all the possible errors that are likely to occur, as well as the accuracy to which the system or element has been calibrated. As an illustration, the accuracy of a digital thermometer is quoted in the specification as: full scale accuracy – better than 2%.

The term *error* is used for the difference between the result of the measurement and the true value of the quantity being measured, i.e.

$$\text{Error} = \text{Measured value} - \text{True value}$$

Thus if the measured value is 10.1 when the true value is 10.0, the error is +0.1. If the measured value is 9.9 when the true value is 10.0, the error is -0.1.

See Appendix A for a discussion of how the accuracy of a value determined for some quantity can be computed from values obtained from a number of measurements, e.g. the accuracy of the value of the density of some material when computed from measurements of its mass and volume, both the mass and volume measurements having errors.

Errors can arise in a number of ways and the following describes some of the errors that are encountered in specifications of instrumentation systems.

1. Hysteresis error

The term *hysteresis error* (Figure 1.12) is used for the difference in outputs given from the same value of quantity being measured according to whether that value has been reached by a continuously increasing change or a continuously decreasing change. Thus, you might obtain a different value from a thermometer used to measure the same temperature of a liquid if it is reached by the liquid warming up to the measured temperature or it is reached by the liquid cooling down to the measured temperature.

2. Non-linearity error

The term *non-linearity error* (Figure 1.13) is used for the error that occurs as a result of assuming a linear relationship between the input and output over the working range, i.e. a graph of output plotted against input is assumed to give a straight line. Few systems or elements, however, have a truly linear relationship and thus errors occur as a result of the assumption of linearity. Linearity error is usually expressed as a percentage error of full range or full scale output. As an illustration, the non-linearity error for the OMRON ZX-E displacement sensor is quoted as $\pm 0.5\%$. As a further illustration, a load cell is quoted in its specification as having: non-linearity error $\pm 0.03\%$ of full range, hysteresis error $\pm 0.02\%$ of full range.

3. Insertion error

When a cold thermometer is put in to a hot liquid to measure its temperature, the presence of the cold thermometer in the hot liquid changes the temperature of the liquid. The liquid cools and so the thermometer ends up measuring a lower temperature than that which existed before the thermometer was introduced. The act of attempting to make the measurement has modified the temperature being measured. This effect is called *loading* and the consequence as an *insertion error*. If we want this modification to be small, then the thermometer should have a small heat capacity compared with that of the liquid. A small heat capacity means that very little heat is needed to change its temperature. Thus the heat taken from the liquid is minimised and so its temperature little affected.

Loading is a problem that is often encountered when measurements are being made. For example, when an ammeter is inserted into a circuit to make a measurement of the circuit current, it changes the resistance of the circuit and so changes the current being measured (Figure 1.14). The act of attempting to make such a

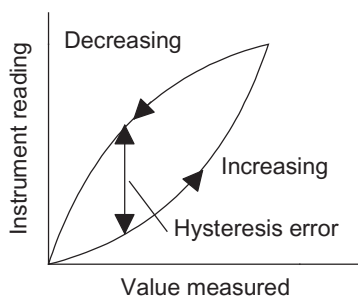


FIGURE 1.12 Hysteresis error.

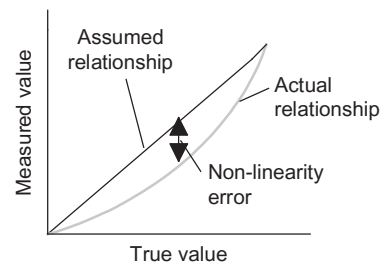


FIGURE 1.13 Non-linearity error.

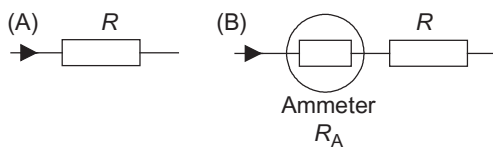


FIGURE 1.14 Loading with an ammeter: (A) circuit before meter introduced, (B) extra resistance introduced by meter.

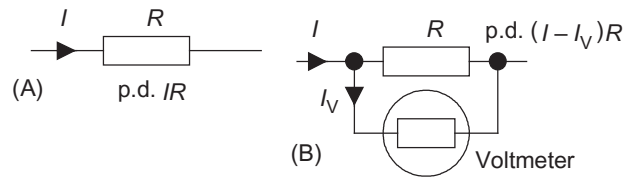


FIGURE 1.15 Loading with a voltmeter: (A) before meter, (B) with meter present.

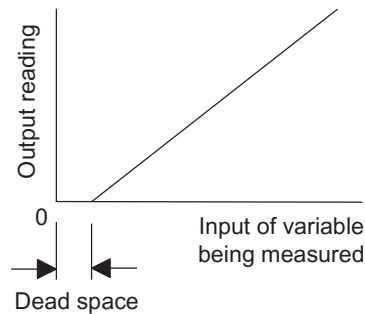


FIGURE 1.16 Dead space.

measurement has modified the current that was being measured. If the effect of inserting the ammeter is to be as small as possible and for the ammeter to indicate the original current, the resistance of the ammeter must be very small when compared with that of the circuit.

When a voltmeter is connected across a resistor to measure the voltage across it, then what we have done is connected a resistance, that of the voltmeter, in parallel with the resistance across which the voltage is to be measured. If the resistance of the voltmeter is not considerably higher than that of the resistor, the current through the resistor is markedly changed by the current passing through the meter resistance and so the voltage being measured is changed (Figure 1.15). The act of attempting to make the measurement has modified the voltage that was being measured. If the effect of inserting the voltmeter in the circuit is to be as small as possible, the resistance of the voltmeter must be much larger than that of the resistance across which it is connected. Only then will the current bypassing the resistor and passing through the voltmeter be very small and so the voltage not significantly changed.

EXAMPLE

Two voltmeters are available, one with a resistance of $1\text{ k}\Omega$ and the other $1\text{ M}\Omega$. Which instrument should be selected if the indicated value is to be closest to the voltage value that existed across a $2\text{ k}\Omega$ resistor before the voltmeter was connected across it?

The $1\text{ M}\Omega$ voltmeter should be chosen. This is because when it is in parallel with $2\text{ k}\Omega$, less current will flow through it than if the $1\text{ k}\Omega$ voltmeter had been used and so the current through the resistor will be closer to its original value. Hence the indicated voltage will be closer to the value that existed before the voltmeter was connected into the circuit.

1.3.2 Range

The *range* of variable of system is the limits between which the input can vary. For example, a resistance thermometer sensor might be quoted as having a range of -200°C to $+800^\circ\text{C}$. The term *dead band* or *dead space* is used if there is a range of input values for which there is no output. Figure 1.16 illustrates this. For example, bearing friction in a flow meter using a rotor might mean that there is no output until the input has reached a particular flow rate threshold.

1.3.3 Precision, Repeatability, and Reproducibility

The term *precision* is used to describe the degree of freedom of a measurement system from random errors. Thus, a high precision measurement instrument will give only a small spread of readings if repeated readings are taken of the same quantity. A low precision measurement system will give a large spread of readings. For example, consider the following two sets of readings obtained for repeated measurements of the same quantity by two different instruments:

20.1 mm, 20.2 mm, 20.1 mm, 20.0 mm, 20.1 mm, 20.1 mm, 20.0 mm
 19.9 mm, 20.3 mm, 20.0 mm, 20.5 mm, 20.2 mm, 19.8 mm, 20.3 mm

The results of the measurement give values scattered about some value. The first set of results shows a smaller spread of readings than the second and indicates a higher degree of precision for the instrument used for the first set.

The terms repeatability and reproducibility are ways of talking about precision in specific contexts. The term *repeatability* is used for the ability of a measurement system to give the same value for repeated measurements of the same value of a variable. Common causes of lack of repeatability are random fluctuations in the environment, e.g. changes in temperature and humidity. The error arising from repeatability is usually expressed as a percentage of the full range output. For example, a pressure sensor might be quoted as having a repeatability of $\pm 0.1\%$ of full range. Thus with a range of 20 kPa, this would be an error of ± 20 Pa. The term *reproducibility* is used to describe the ability of a system to give the same output when used with a constant input with the system or elements of the system being disconnected from its input and then reinstalled. The resulting error is usually expressed as a percentage of the full range output.

Note that precision should not be confused with accuracy. High precision does not mean high accuracy. A high precision instrument could have low accuracy. Figure 1.17 illustrates this.

1.3.4 Sensitivity

The *sensitivity* indicates how much the output of an instrument system or system element changes when the quantity being measured changes by a given amount, i.e. the ratio output/input. For example, a thermocouple might have a sensitivity of $20 \mu\text{V}/^\circ\text{C}$ and so give an output of $20 \mu\text{V}$ for each 1°C change in temperature. Thus, if we take a series of readings of the output of an instrument for a number of different inputs and plot a graph of output against input (Figure 1.18), the sensitivity is the slope of the graph. For example, an iron–constantan thermocouple might be quoted as having a sensitivity at 0°C of $0.05 \text{ mV}/^\circ\text{C}$.

The term is also frequently used to indicate the sensitivity to inputs other than that being measured, i.e. environmental changes. For example, the sensitivity of a system or element might be quoted to changes in temperature or perhaps fluctuations in the mains voltage supply. Thus a pressure measurement sensor might be quoted as having a temperature sensitivity of $\pm 0.1\%$ of the reading per $^\circ\text{C}$ change in temperature.

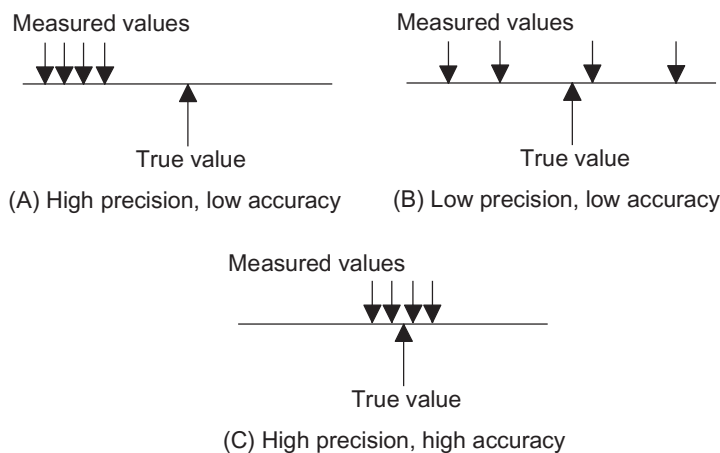


FIGURE 1.17 Precision and accuracy.

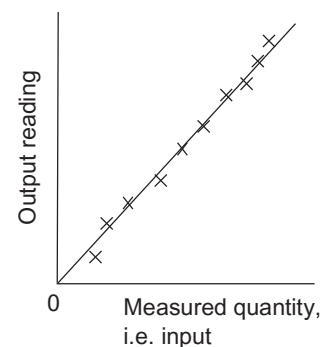


FIGURE 1.18 Sensitivity as slope of input–output graph.

As an illustration of the type of information available in a specification, a commercial pressure measurement system is quoted in the manufacturer's specification as having:

Range 0 to 10 kPa
 Supply Voltage ± 15 V dc
 Linearity error $\pm 0.5\%$ FS
 Hysteresis error $\pm 0.15\%$ FS
 Sensitivity 5 V dc for full range
 Thermal sensitivity $\pm 0.02\%/^{\circ}\text{C}$
 Thermal zero drift $0.02\%/^{\circ}\text{C}$ FS
 Temperature range 0 to 50°C

EXAMPLE

A spring balance has its deflection measured for a number of loads and gave the following results. Determine its sensitivity.

Load in kg	0	1	2	3	4
Deflection in mm	0	10	20	30	40

Figure 1.19 shows the graph of output against input. The graph has a slope of 10 mm/kg and so this is the sensitivity.

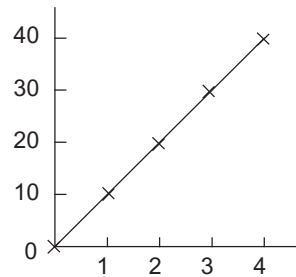


FIGURE 1.19 Example.

EXAMPLE

A pressure measurement system (a diaphragm sensor giving a capacitance change with output processed by a bridge circuit and displayed on a digital display) is stated as having the following characteristics. Explain the significance of the terms:

Range: 0 to 125 kPa and 0 to 2500 kPa
 Accuracy: $\pm 1\%$ of the displayed reading
 Temperature sensitivity: $\pm 0.1\%$ of the reading per $^{\circ}\text{C}$

The range indicates that the system can be used to measure pressures from 0 to 125 kPa or 0 to 2500 kPa. The accuracy is expressed as a percentage of the displayed reading, thus if the instrument indicates a pressure of, say, 100 kPa then the error will be ± 1 kPa. The temperature sensitivity indicates that if the temperature changes by 1°C the displayed reading will be in error by $\pm 0.1\%$ of the value. Thus for a pressure of, say, 100 kPa the error will be ± 0.1 kPa for a 1°C temperature change.

1.3.5 Stability

The *stability* of a system is its ability to give the same output when used to measure a constant input over a period of time. The term *drift* is often used to describe the change in output that occurs over time. The drift may be expressed as a percentage of the full range output. The term *zero drift* is used for the changes that occur in output when there is zero input.

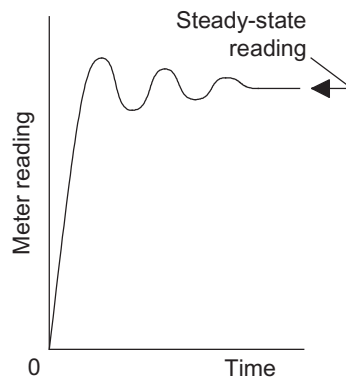


FIGURE 1.20 Oscillations of a meter reading.

1.3.6 Dynamic Characteristics

The terms given above refer to what can be termed the *static characteristics*. These are the values given when steady-state conditions occur, i.e. the values given when the system or element has settled down after having received some input. The *dynamic characteristics* refer to the behaviour between the time that the input value changes and the time that the value given by the system or element settles down to the steady-state value. For example, Figure 1.20 shows how the reading of an analogue ammeter might change when the current is switched on. The meter pointer oscillates before settling down to give the steady-state reading.

The following are terms commonly used for dynamic characteristics.

1. *Response time*

This is the time which elapses after the input to a system or element is abruptly increased from zero to a constant value up to the point at which the system or element gives an output corresponding to some specified percentage, e.g. 95%, of the value of the input.

2. *Rise time*

This is the time taken for the output to rise to some specified percentage of the steady-state output. Often the rise time refers to the time taken for the output to rise from 10% of the steady-state value to 90% or 95% of the steady-state value.

3. *Settling time*

This is the time taken for the output to settle to within some percentage, e.g. 2%, of the steady-state value.

1.4 DEPENDABILITY

The term *dependability* (see the paper *Dependability and Its Threats: A Taxonomy* by Algurdis Avizienis, Jean-Claude Laprie and Brian Randell – freely available on-line) is here used to describe the ability of a system to deliver a service that can be trusted, service being a system's behaviour as perceived by the user. Other definitions that have been used for dependability include the ISO definition as availability performance and its influencing factors, namely reliability performance, maintainability, performance and maintenance support performance. An IEC definition involves the extent to which the system can be relied upon to perform exclusively and correctly the system tasks under defined operational and environmental conditions over a defined period of time or at a given time.

Dependability encompasses the following attributes:

- Availability, i.e. readiness for correct service;
- Reliability, i.e. the ability to continue with correct service;
- Safety, i.e. the ability to deliver a service which is safe to the user and the environment;
- Maintainability, i.e. the ability to undergo repairs such as the removal of faulty components, preventive maintenance and modifications; and
- Integrity, i.e. the absence of improper system alterations.

The dependability specification for a system needs to include the requirements for the above attributes in terms of the acceptable frequency and severity of failures for the specified use environment.

In general, the means to attain dependability include:

- Fault prevention, i.e. the ability to prevent the occurrence or introduction of faults;
- Fault tolerance, i.e. the means to avoid service failures in the presence of faults;
- Fault removal, i.e. the means to reduce the number and severity of faults; and
- Fault forecasting, i.e. the means to estimate the future occurrence and consequences of faults.

Fault prevention and fault tolerance aims involve the giving to the system of the ability to deliver a service that can be trusted while fault removal and fault forecasting aim to give confidence in that ability and that the dependability specifications are adequate and the system is likely to meet them. Faults can arise during the development of the system or during its operation and may be internal faults within the system or result from faults external to the system which propagates errors into the system. Faults may originate in the hardware of the system or be faults that affect software used with the system. The cause of a fault may be a result of human actions, possibly malicious or simply omissions such as wrong setting of parameters. Malicious actions can be designed to disrupt service or access confidential information and involve such elements as a Trojan horse or virus. The paper referred to earlier, i.e. *Dependability and Its threats: A Taxonomy*, gives a classification of faults that can occur as:

- The phase of system life during which faults occur during the development of the system, during maintenance when it is in use, and procedures used to operate or maintain the system;
- The location of faults: internal to the system or external;
- The phenomenological cause of the faults: natural faults that naturally occur without human intervention, and human-made faults as a result of human actions;
- The dimension in which faults occur in hardware or software;
- How the faults were introduced: malicious or non-malicious;
- The intent of the human or humans who introduced the faults: deliberate or non-deliberate;
- How the human introduced the faults: accidental or incompetence; and
- The persistence of the faults: permanent or transient.

Maintainability for a system involves both corrective maintenance with repairs for the removal of faults and preventative maintenance in which repairs are carried out in anticipation of failures. Maintenance also involves adjustments in response to environmental changes and augmentation of the system's function.

1.4.1 Reliability

If you toss a coin ten times you might find, for example, that it lands heads uppermost six times out of the ten. If, however, you toss the coin for a very large number of times then it is likely that it will land heads uppermost half of the times. The probability of it landing heads uppermost is said to be half. The *probability* of a particular event occurring is defined as being

$$\text{Probability} = \frac{\text{Number of occurrences of the event}}{\text{Total number of trials}}$$

When the total number of trials is very large. The probability of the coin landing with either a heads or tails uppermost is likely to be 1, since every time the coin is tossed this event will occur. A probability of 1 means a certainty that the event will take place every time. The probability of the coin landing standing on edge can be considered to be zero, since the number of occurrences of such an event is zero. The closer the probability is to 1 the more frequent an event will occur; the closer it is to zero the less frequent it will occur.

Reliability is an important requirement of a measurement system. The *reliability* of a measurement system, or element in such a system, is defined as being the probability that it will operate to an agreed level of performance, for a specified period, subject to specified environmental conditions. The agreed level of performance might be that the measurement system gives a particular accuracy. The reliability of a measurement system is likely to change with time as a result of perhaps springs slowly stretching with time, resistance values changing

as a result of moisture absorption, wear on contacts and general damage due to environmental conditions. For example, just after a measurement system has been calibrated, the reliability should be 1. However, after perhaps 6 months the reliability might have dropped to 0.7. Thus the system cannot then be relied on to always give the required accuracy of measurement, it typically only gives the required accuracy seven times in ten measurements, seventy times in a hundred measurements.

A high reliability system will have a low failure rate. *Failure rate* is the number of times during some period of time that the system fails to meet the required level of performance, i.e.:

$$\text{Failure rate} = \frac{\text{Number of failures}}{\text{Number of systems observed} \times \text{Time observed}}$$

A failure rate of 0.4 per year means that in one year, if ten systems are observed, 4 will fail to meet the required level of performance. If 100 systems are observed, 40 will fail to meet the required level of performance. Failure rate is affected by environmental conditions. For example, the failure rate for a temperature measurement system used in hot, dusty, humid, corrosive conditions might be 1.2 per year, while for the same system used in dry, cool, non-corrosive environment it might be 0.3 per year.

Failure rates are generally quantified by giving the *mean time between failures* (MTBF). This is a statistical representation of the reliability in that while it does not give the time to failure for a particular example of the system it does represent the time to failure when the times for a lot of the examples of that system are considered.

With a measurement system consisting of a number of elements, failure occurs when just one of the elements fails to reach the required performance. Thus in a system for the measurement of the temperature of a fluid in some plant we might have a thermocouple, an amplifier, and a meter. The failure rate is likely to be highest for the thermocouple since that is the element in contact with the fluid while the other elements are likely to be in the controlled atmosphere of a control room. The reliability of the system might thus be markedly improved by choosing materials for the thermocouple which resist attack by the fluid. Thus it might be in a stainless steel sheath to prevent fluid coming into direct contact with the thermocouple wires.

EXAMPLE

The failure rate for a pressure measurement system used in factory A is found to be 1.0 per year while the system used in factory B is 3.0 per year. Which factory has the most reliable pressure measurement system?

The higher the reliability the lower the failure rate. Thus factory A has the more reliable system. The failure rate of 1.0 per year means that if 100 instruments are checked over a period of a year, 100 failures will be found, i.e. on average each instrument is failing once. The failure rate of 3.0 means that if 100 instruments are checked over a period of a year, 300 failures will be found, i.e. instruments are failing more than once in the year.

1.5 REQUIREMENTS

The main requirement of a measurement system is *fitness for purpose*. This means that if, for example, a length of a product has to be measured to a certain accuracy that the measurement system is able to be used to carry out such a measurement to that accuracy. For example, a length measurement system might be quoted as having an accuracy of ± 1 mm. This would mean that all the length values it gives are only guaranteed to this accuracy, e.g. for a measurement which gave a length of 120 mm the actual value could only be guaranteed to be between 119 and 121 mm. If the requirement is that the length can be measured to an accuracy of ± 1 mm then the system is fit for that purpose. If, however, the criterion is for a system with an accuracy of ± 0.5 mm then the system is not fit for that purpose.

In order to deliver the required accuracy, the measurement system must have been calibrated to give that accuracy. *Calibration* is the process of comparing the output of a measurement system against standards of known accuracy. The standards may be other measurement systems which are kept specially for calibration duties or some means of defining standard values. In many companies some instruments and items such as standard resistors and cells are kept in a company standards department and used solely for calibration purposes.

1.5.1 Calibration

Calibration should be carried out using equipment which can be traceable back to national standards with a separate calibration record kept for each measurement instrument. This record is likely to contain a description of the instrument and its reference number, the calibration date, the calibration results, how frequently the instrument is to be calibrated, and probably details of the calibration procedure to be used, details of any repairs or modifications made to the instrument, and any limitations on its use.

The *national standards* are defined by international agreement and are maintained by national establishments, e.g. the National Physical Laboratory in Great Britain and the National Bureau of Standards in the United States. There are seven such *primary standards*, and two *supplementary ones*, the primary ones being:

1. *Mass*

The kilogram is defined by setting Planck's constant h to exactly $662\,607\,015 \times 10^{-34}$ J s given the definitions of the metre and second. Then 1 kg is $h/(662\,607\,015 \times 10^{-34})$.

2. *Length*

The length standard, the metre, is defined as the distance travelled by light in a vacuum in $1/(299\,792\,458)$ second.

3. *Time*

The time standard, the second, is defined as a time duration of 9 192 631 770 periods of oscillation of the radiation emitted by the caesium-133 atom under precisely defined conditions of resonance.

4. *Current*

The current standard, the ampere, is defined as the flow of $1/(602\,176\,634 \times 10^{-19})$ times the elementary charge e per second.

5. *Temperature*

The kelvin (K) is the unit of temperature and is defined by setting the numerical value of the Boltzmann constant k to be $1\,380\,649 \times 10^{-23}$ J/K given the definitions of the kilogram, metre and second.

6. *Luminous intensity*

The candela is defined as the luminous intensity, in a given direction, of a specified source that emits monochromatic radiation of frequency 540×10^{12} Hz and that has a radiant intensity of $1/683$ watt per unit steradian (a unit solid angle, see later).

7. *Amount of substance*

The mole is defined as the amount of substance of exactly $602\,214\,076 \times 10^{23}$ elementary entities.

The *supplementary standards* are:

1. *Plane angle*

The radian is the plane angle between two radii of a circle which cuts off on the circumference an arc with a length equal to the radius (Figure 1.21).

2. *Solid angle*

The steradian is the solid angle of a cone which, having its vertex in the centre of the sphere, cuts off an area of the surface of the sphere equal to the square of the radius (Figure 1.22).

Primary standards are used to define national standards, not only in the primary quantities but also in other quantities which can be derived from them. For example, a resistance standard of a coil of manganin wire is defined in terms of the primary quantities of length, mass, time, and current. Typically these national standards

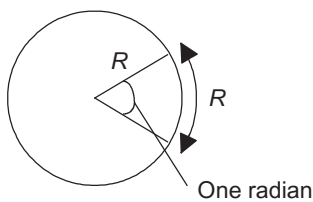


FIGURE 1.21 The radian.

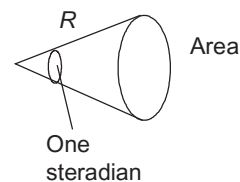


FIGURE 1.22 The steradian.

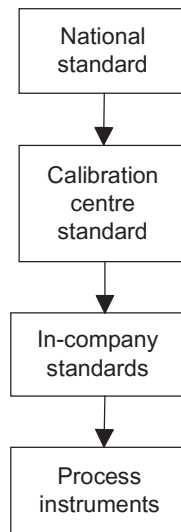


FIGURE 1.23 Traceability chain.

in turn are used to define reference standards which can be used by national bodies for the calibration of standards which are held in calibration centres.

The equipment used in the calibration of an instrument in everyday company use is likely to be *traceable* back to national standards in the following way:

1. National standards are used to calibrate standards for calibration centres.
2. Calibration centre standards are used to calibrate standards for instrument manufacturers.
3. Standardised instruments from instrument manufacturers are used to provide in-company standards.
4. In-company standards are used to calibrate process instruments.

There is a simple traceability chain from the instrument used in a process back to national standards (Figure 1.23). In the case of, say, a glass bulb thermometer, the traceability might be:

1. National standard of fixed thermodynamic temperature points.
2. Calibration centre standard of a platinum resistance thermometer with an accuracy of $\pm 0.005^{\circ}\text{C}$.
3. An in-company standard of a platinum resistance thermometer with an accuracy of $\pm 0.01^{\circ}\text{C}$.
4. The process instrument of a glass bulb thermometer with an accuracy of $\pm 0.1^{\circ}\text{C}$.

1.5.2 Safety Systems

Statutory safety regulations lay down the responsibilities of employers and employees for safety in the work place. These include for employers the duty to:

- Ensure that process plant is operated and maintained in a safe way so that the health and safety of employees is protected.
- Provide a monitoring and shutdown system for processes that might result in hazardous conditions.

Employees also have duties to:

- Take reasonable care of their own safety and the safety of others.
- Avoid misusing or damaging equipment that is designed to protect people's safety.

Thus, in the design of measurement systems, due regard has to be paid to safety both in their installation and operation. Thus:

- The failure of any single component in a system should not create a dangerous situation.
- A failure which results in cable open or short circuits or short circuiting to ground should not create a dangerous situation.

- Foreseeable modes of failure should be considered for fail-safe design so that, in the event of failure, the system perhaps switches off into a safe condition.
- Systems should be easily checked and readily understood.

The main risks from electrical instrumentation are electrocution and the possibility of causing a fire or explosion as a consequence of perhaps cables or components overheating or arcing sparks occurring in an explosive atmosphere. Thus it is necessary to ensure that an individual cannot become connected between two points with a potential difference greater than about 30 V and this requires the careful design of earthing so that there is always an adequate earthing return path to operate any protective device in the event of a fault occurring.

PROBLEMS

Questions 1 to 5 have four answer options: A, B, C, and D. Choose the correct answer from the answer options.

- Decide whether each of these statements is True (T) or False (F).
Sensors in a measurement system have:
 - An input of the variable being measured.
 - An output of a signal in a form suitable for further processing in the measurement system.
 Which option BEST describes the two statements?
 - (i) T (ii) T
 - (i) T (ii) F
 - (i) F (ii) T
 - (i) F (ii) F
- The signal conditioner element in a measurement system:
 - Gives an output signal dependent on the temperature.
 - Changes the temperature signal to a current signal.
 - Takes the output from the sensor and makes it bigger.
 - Gives an output display.
- Decide whether each of these statements is True (T) or False (F).
The discrepancy between the measured value of the current in an electrical circuit and the value before the measurement system, an ammeter, was inserted in the circuit is bigger the larger:
 - The resistance of the meter.
 - The resistance of the circuit.
 Which option BEST describes the two statements?
 - (i) T (ii) T
 - (i) T (ii) F
 - (i) F (ii) T
 - (i) F (ii) F
- Decide whether each of these statements is True (T) or False (F).
A highly reliable measurement system is one where there is a high chance that the system will:
 - Have a high mean time between failures.
 - Have a high probability of failure.
 Which option BEST describes the two statements?
 - (i) T (ii) T
 - (i) T (ii) F
 - (i) F (ii) T
 - (i) F (ii) F
- Decide whether each of these statements is True (T) or False (F).
A measurement system which has a lack of repeatability is one where there could be:
 - Random fluctuations in the values given by repeated measurements of the same variable.
 - Fluctuations in the values obtained by repeating measurements over a number of samples.
 Which option BEST describes the two statements?

- A. (i) T (ii) T
 B. (i) T (ii) F
 C. (i) F (ii) T
 D. (i) F (ii) F

6. List and explain the functional elements of a measurement system.
 7. Explain the terms (a) reliability and (b) repeatability when applied to a measurement system.
 8. Explain what is meant by calibration standards having to be traceable to national standards.
 9. Explain what is meant by 'fitness for purpose' when applied to a measurement system.
 10. The reliability of a measurement system is said to be 0.6. What does this mean?
 11. The measurement instruments used in the tool room of a company are found to have a failure rate of 0.01 per year. What does this mean?
 12. Determine the sensitivity of the instruments that gave the following readings:

(a)

Load kg	0	2	4	6	8
Deflection mm	0	18	36	54	72

(b)

Temperature °C	0	10	20	30	40
Voltage mV	0	0.59	1.19	1.80	2.42

(c)

Load N	0	1	2	3	4
Charge pC	0	3	6	9	12

13. Calibration of a voltmeter gave the following data. Determine the maximum hysteresis error as a percentage of the full-scale range.

Increasing input:

Standard mV	0	1.0	2.0	3.0	4.0
Voltmeter mV	0	1.0	1.9	2.9	4.0

Decreasing input:

Standard mV	4.0	3.0	2.0	1.0	0
Voltmeter mV	4.0	3.0	2.1	1.1	0

Instrumentation System Elements

OUTLINE

2.1 Introduction	18	2.6.8 Nucleonic Level Indicators	36
2.2 Displacement Sensors	18	2.7 Temperature Sensors	36
2.2.1 Potentiometer	18	2.7.1 Bimetallic Strips	36
2.2.2 Strain-Gauged Element	19	2.7.2 Liquid in Glass Thermometers	36
2.2.3 Capacitive Element	20	2.7.3 Resistance Temperature Detectors (RTDs)	36
2.2.4 Linear Variable Differential Transformer	21	2.7.4 Thermistors	37
2.2.5 Optical Encoders	21	2.7.5 Thermocouples	37
2.2.6 Moiré Fringes	22	2.7.6 Thermodiodes and Transistors	40
2.2.7 Optical Proximity Sensors	23	2.7.7 Pyrometers	40
2.2.8 Mechanical Switches	24	2.8 Sensor Selection	41
2.2.9 Capacitive Proximity Sensor	25	2.9 Signal Processing	42
2.2.10 Inductivity Proximity Sensor	26	2.9.1 Resistance to Voltage Converter	42
2.3 Speed Sensors	26	2.9.2 Temperature Compensation	45
2.3.1 Optical Methods	26	2.9.3 Thermocouple Compensation	45
2.3.2 Incremental Encoder	26	2.9.4 Protection	47
2.3.3 Tachogenerator	26	2.9.5 Analogue-to-Digital Conversions	47
2.4 Fluid Pressure Sensors	26	2.9.6 Digital-to-Analogue Conversions	50
2.4.1 Diaphragm Sensor	27	2.9.7 Microcontroller Systems	51
2.4.2 Piezoelectric Sensor	28	2.9.8 Op-Amps	52
2.4.3 Bourdon Tube	28	2.9.9 Pressure-to-Current Converter	56
2.5 Fluid Flow	29	2.10 Signal Transmission	56
2.5.1 Differential Pressure Methods	29	2.10.1 Noise	59
2.5.2 Turbine Meter	31	2.11 Smart Systems	60
2.5.3 Ultrasonic Time of Flight Flow Meter	32	2.11.1 MEMS	61
2.5.4 Vortex Flow Rate Method	32	2.12 Data Presentation Element	62
2.5.5 Coriolis Flow Meter	33	2.12.1 Indicator	62
2.6 Liquid Level	33	2.12.2 Illuminative Displays	62
2.6.1 Floats	34	2.12.3 Graphical User Interface (GUI)	64
2.6.2 Displacer Gauge	34	2.12.4 Data Loggers	64
2.6.3 Differential Pressure	34	2.12.5 Printers	65
2.6.4 Load Cell	34	Problems	66
2.6.5 Electrical Conductivity Level Indicator	35		
2.6.6 Capacitive Level Indicator	35		
2.6.7 Ultrasonic Level Gauge	35		

2.1 INTRODUCTION

This chapter discusses the sensors, signal processors, and data presentation elements commonly used in engineering. The term *sensor* is used for an element which produces a signal relating to the quantity being measured. The term *signal processor* is used for the element that takes the output from the sensor and converts it into a form which is suitable for data presentation. *Data presentation* is where the data is displayed, recorded, or transmitted to some control system.

2.2 DISPLACEMENT SENSORS

A displacement sensor is here considered to be one that can be used to:

1. Measure a linear displacement, i.e. a change in linear position. This might, for example, be the change in linear displacement of a sensor as a result of a change in the thickness of sheet metal emerging from rollers.
2. Measure an angular displacement, i.e. a change in angular position. This might, for example, be the change in angular displacement of a drive shaft.
3. Detect motion, e.g. this might be as part of an alarm or automatic light system, whereby an alarm is sounded or a light switched on when there is some movement of an object within the 'view' of the sensor.
4. Detect the presence of some object, i.e. a proximity sensor. This might be in an automatic machining system where a tool is activated when the presence of a work piece is sensed as being in position.

Displacement sensors fall into two groups: those that make direct contact with the object being monitored, by spring loading or mechanical connection with the object, and those which are non-contacting. For those linear displacement methods involving contact, there is usually a sensing shaft which is in direct contact with the object being monitored, the displacement of this shaft is then being monitored by a sensor. This shaft movement may be used to cause changes in electrical voltage, resistance, capacitance, or mutual inductance. For angular displacement methods involving mechanical connection, the rotation of a shaft might directly drive, through gears, the rotation of the sensor element, this perhaps generating an e.m.f. Non-contacting proximity sensors might consist of a beam of infrared light being broken by the presence of the object being monitored, the sensor then giving a voltage signal indicating the breaking of the beam, or perhaps the beam being reflected from the object being monitored, the sensor giving a voltage indicating that the reflected beam has been detected. Contacting proximity sensors might be just mechanical switches which are tripped by the presence of the object, the term limit switch being used. The following are examples of displacement sensors.

2.2.1 Potentiometer

A *potentiometer* consists of a resistance element with a sliding contact which can be moved over the length of the element and connected as shown in [Figure 2.1](#). With a constant supply voltage V_s , the output voltage V_o between terminals 1 and 2 is a fraction of the input voltage, the fraction depending on the ratio of the resistance R_{12} between terminals 1 and 2 compared with the total resistance R of the entire length of the track across which the supply voltage is connected. Thus $V_o/V_s = R_{12}/R$. If the track has a constant resistance per unit length, the output is proportional to the displacement of the slider from position 1. A rotary potentiometer consists of a coil

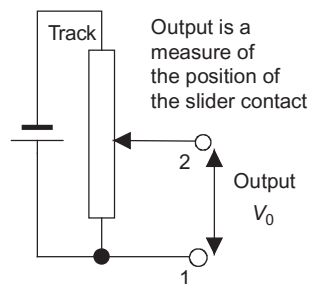


FIGURE 2.1 Potentiometer.

of wire wrapped round into a circular track, or a circular film of conductive plastic or a ceramic–metal mix termed a cermet, over which a rotatable sliding contact can be rotated. Hence an angular displacement can be converted into a potential difference. Linear tracks can be used for linear displacements.

With a wire-wound track the output voltage does not continuously vary as the slider is moved over the track but goes in small jumps as the slider moves from one turn of wire to the next. This problem does not occur with a conductive plastic or the cermet track. Thus, the smallest change in displacement which will give rise to a change in output, i.e. the resolution, tends to be much smaller for plastic tracks than wire-wound tracks. Errors due to non-linearity of the track for wire tracks tend to range from less than 0.1% to about 1% of the full range output and for conductive plastics can be as low as about 0.05%. The track resistance for wire-wound potentiometers tends to range from about 20 Ω to 200 k Ω and for conductive plastic from about 500 Ω to 80 k Ω . Conductive plastic has a higher temperature coefficient of resistance than wire and so temperature changes have a greater effect on accuracy. The resolution of such a sensor depends on its construction. If it is a wire-wound coil with a rotatable sliding contact then the finer the wire the higher the resolution. Thus a sensor with 25 turns per mm would have a resolution of $\pm 40 \mu\text{m}$. Such a sensor has a fast response time and a low cost.

The following is an example of part of the specification of a commercially available displacement sensor using a plastic conducting potentiometer track:

Ranges from 0 to 10 mm to 0 to 2 m
 Non-linearity error $\pm 0.1\%$ of full range
 Resolution $\pm 0.02\%$ of full range
 Temperature sensitivity ± 120 parts per million/ $^{\circ}\text{C}$
 Resolution $\pm 0.02\%$ of full range

An application of a potentiometer is to sense the position of the accelerator position in an automobile and feed the information to the engine control system. Another potentiometer might be used as the throttle position sensor.

2.2.2 Strain-Gauged Element

Strain gauges consist of a metal foil strip (Figure 2.2A), flat length of metal wire (Figure 2.2B), or a strip of semiconductor material which can be stuck onto surfaces like a postage stamp. When the wire, foil, strip, or semiconductor is stretched, its resistance R changes. The fractional change in resistance $\Delta R/R$ is proportional to the strain ε , i.e.:

$$\frac{\Delta R}{R} = G$$

where G , the constant of proportionality, is termed the *gauge factor*.

Metal strain gauges typically have gauge factors of the order of 2.0. When such a strain gauge is stretched its resistance increases, and when compressed its resistance decreases. Strain is 'change in length/original length' and so the resistance change of a strain gauge is a measurement of the change in length of the gauge and hence the surface to which the strain gauge is attached. Thus a displacement sensor might be constructed by attaching strain gauges to a cantilever (Figure 2.3), the free end of the cantilever being moved as a result of the linear

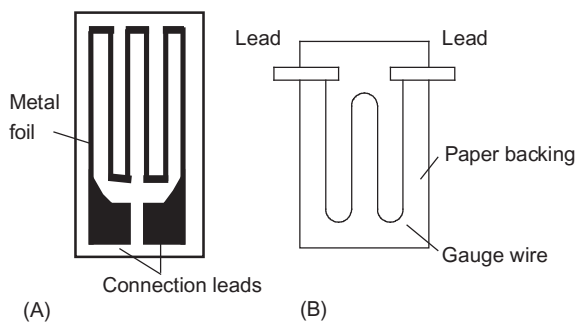


FIGURE 2.2 Strain gauges.

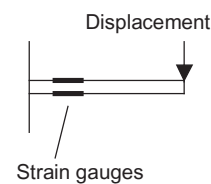


FIGURE 2.3 Strain-gauged cantilever.

displacement being monitored. When the cantilever is bent, the electrical resistance strain gauges mounted on the element are strained and so give a resistance change which can be monitored and which is a measure of the displacement. With strain gauges mounted as shown in Figure 2.3, when the cantilever is deflected downwards the gauge on the upper surface is stretched and the gauge on the lower surface is compressed. Thus the gauge on the upper surface increases in resistance while that on the lower surface decreases. Typically, this type of sensor is used for linear displacements of the order of 1 mm to 30 mm, having a non-linearity error of about $\pm 1\%$ of full range. A commercially available displacement sensor, based on the arrangement shown in Figure 2.3, has the following in its specification:

- Range 0 to 100 mm
- Non-linearity error $\pm 0.1\%$ of full range
- Temperature sensitivity $\pm 0.01\%$ of full range/ $^{\circ}\text{C}$

A problem that has to be overcome with strain gauges is that the resistance of the gauge changes when the temperature changes and so methods have to be used to compensate for such changes in order that the effects of temperature can be eliminated. This is discussed later in this chapter when the circuits used for signal processing are discussed.

2.2.3 Capacitive Element

The capacitance C of a parallel plate capacitor (Figure 2.4) is given by:

$$C = \frac{\epsilon_r \epsilon_0 A}{d}$$

where ϵ_r is the relative permittivity of the dielectric between the plates, ϵ_0 a constant called the permittivity of free space, A the area of overlap between the two plates and d the plate separation. The capacitance will change if the plate separation d changes, the area A of overlap of the plates changes, or a slab of dielectric is moved into or out of the plates, so varying the effective value of ϵ_r (Figure 2.5). All these methods can be used to give linear displacement sensors.

A commercially available capacitor displacement sensor based on the use of the sliding capacitor plate (Figure 2.5B) includes in its specification:

- Ranges available from 0 to 5 mm to 0 to 250 mm
- Non-linearity and hysteresis error $\pm 0.01\%$ of full range

One form that is often used is shown in Figure 2.6 and is referred to as a *push-pull* displacement sensor. It consists of two capacitors, one between the movable central plate and the upper plate and one between the central movable plate and the lower plate. The displacement x moves the central plate between the two other plates.

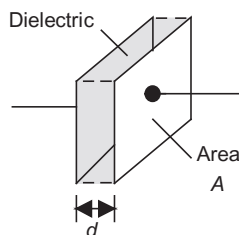


FIGURE 2.4 Parallel plate.

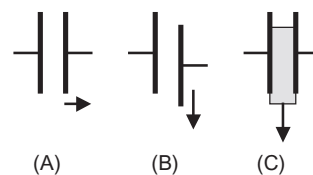


FIGURE 2.5 Capacitive sensors.

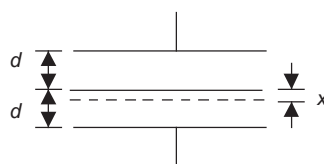


FIGURE 2.6 Push-pull displacement capacitor sensor.

Thus when the central plate moves upwards it decreases the plate separation of the upper capacitor and increases the separation of the lower capacitor. Thus the capacitance of the upper capacitor is increased and that of the lower capacitor is decreased. When the two capacitors are incorporated in opposite arms of an alternating current bridge, the output voltage from the bridge is proportional to the displacement. Such a sensor has good long-term stability and is typically used for monitoring displacements from a few millimetres to hundreds of millimetres. Non-linearity and hysteresis errors are about $\pm 0.01\%$ of full range.

2.2.4 Linear Variable Differential Transformer

The *linear variable differential transformer*, generally referred to by the abbreviation *LVDT*, is a transformer with a primary coil and two secondary coils. Figure 2.7 shows the arrangement, there being three coils symmetrically spaced along an insulated tube. The central coil is the primary coil and the other two are identical secondary coils which are connected in series in such a way that their outputs oppose each other. A magnetic core is moved through the central tube as a result of the displacement being monitored. When there is an alternating voltage input to the primary coil, alternating e.m.f.s are induced in the secondary coils. With the magnetic core in a central position, the amount of magnetic material in each of the secondary coils is the same and so the e.m.f.s induced in each coil are the same. Since they are so connected that their outputs oppose each other, the net result is zero output. However, when the core is displaced from the central position there is a greater amount of magnetic core in one coil than the other. The result is that a greater e.m.f. is induced in one coil than the other and then there is a net output from the two coils. The bigger the displacement the more of the core there is in one coil than the other, thus the difference between the two e.m.f.s increases, the greater the displacement of the core.

Typically, LVDTs have operating ranges from about ± 2 mm to ± 400 mm. Non-linearity errors are typically about $\pm 0.25\%$. LVDTs are very widely used for monitoring displacements. A commercially available displacement sensor using a LVDT has the following in its specification:

- Ranges ± 0.125 mm to ± 470 mm
- Non-linearity error $\pm 0.25\%$ of full range
- Temperature sensitivity $\pm 0.01\%$ of full range
- Signal conditioning incorporated within the housing of the LVDT

2.2.5 Optical Encoders

An *encoder* is a device that provides a digital output as a result of an angular or linear displacement. Position encoders can be grouped into two categories: incremental encoders, which detect changes in displacement from some datum position, and absolute encoders, which give the actual position. Figure 2.8 shows the basic form of an *incremental encoder* for the measurement of angular displacement of a shaft. It consists of a disc which rotates along with the shaft. In the form shown, the rotatable disc has a number of windows through which a beam of light can pass and be detected by a suitable light sensor. When the shaft rotates and disc rotates, a pulsed output is produced by the sensor with the number of pulses being proportional to the angle through which the disc rotates. The angular displacement of the disc, and hence the shaft rotating it, can thus be determined by the

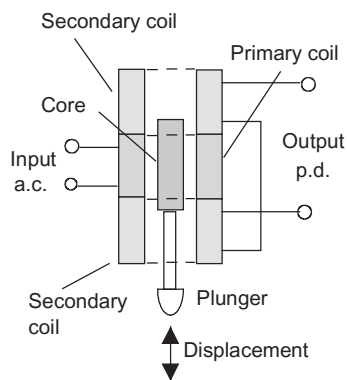


FIGURE 2.7 LVDT.

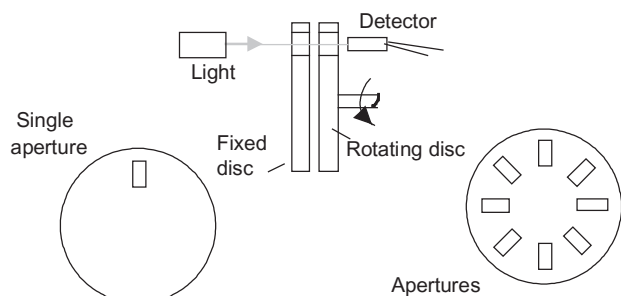


FIGURE 2.8 Optical incremental encoder.

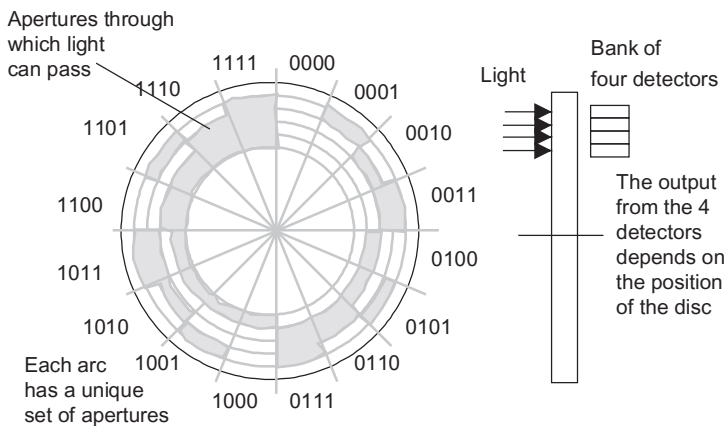


FIGURE 2.9 The rotating wheel of the absolute encoder. Note that though the normal form of binary code is shown in the figure, in practice a modified form of binary code called the Gray code is generally used. This code, unlike normal binary, has only one bit changing in moving from one number to the next. Thus we have the sequence 0000, 0001, 0011, 0010, 0011, 0111, 0101, 0100, 1100, 1101, 1111.

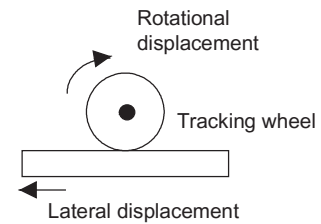


FIGURE 2.10 Tracking wheel.

number of pulses produced in the angular displacement from some datum position. Typically the number of windows on the disc varies from 60 to over a thousand with multi-tracks having slightly offset slots in each track. With 60 slots occurring with 1 revolution then, since 1 revolution is a rotation of 360° , the minimum angular displacement, i.e. the resolution, that can be detected is $360/60 = 6^\circ$. The resolution thus typically varies from about 6° to 0.3° or better.

The OMRON E6A2-C incremental encoder can be used for 100, 200, 300, or 500 pulses per rotation, and the E6A2-CS3E for 10, 20, 60, 100, 200, 300, 360, or 500 pulses per rotation.

With the incremental encoder, the number of pulses counted gives the angular displacement, a displacement of, say, 50° giving the same number of pulses whatever angular position the shaft starts its rotation from. However, the *absolute encoder* gives an output in the form of a binary number of several digits, each such number representing a particular angular position. Figure 2.9 shows the basic form of an absolute encoder for the measurement of angular position.

With the form shown in the figure, the rotating disc has four concentric circles of slots and four sensors to detect the light pulses. The slots are arranged in such a way that the sequential output from the sensors is a number in the binary code, each such number corresponding to a particular angular position. A number of forms of binary code are used. Typical encoders tend to have up to 10 or 12 tracks. The number of bits in the binary number will be equal to the number of tracks. Thus with 10 tracks there will be 10 bits and so the number of positions that can be detected is 2^{10} , i.e. 1024, a resolution of $360/1024 = 0.35^\circ$. The OMRON absolute encoder E6CP-AG3C has an output of Gray code and gives an 8-bit output.

The incremental encoder and the absolute encoder can be used with linear displacements if the linear displacement is first converted to a rotary motion by means of a tracking wheel (Figure 2.10).

2.2.6 Moiré Fringes

Moiré fringes are produced when light passes through two gratings which have rulings inclined at a slight angle to each other. Movement of one grating relative to the other causes the fringes to move. Figure 2.11A illustrates this. Figure 2.11B shows a transmission form of instrument using Moiré fringes and Figure 2.11C a reflection form. With both, a long grating is fixed to the object being displaced. With the transmission form, light passes through the long grating and then a smaller fixed grating, the transmitted light being detected by a photocell. With the reflection form, light is reflected from the long grating through a smaller fixed grating and onto a photocell.

Coarse grating instruments might have 10 to 40 lines per millimetre, fine gratings as many as 400 per millimetre. Movement of the long grating relative to the fixed short grating results in fringes moving across the view of the photocell and thus the output of the cell is a sequence of pulses which can be counted. The displacement is

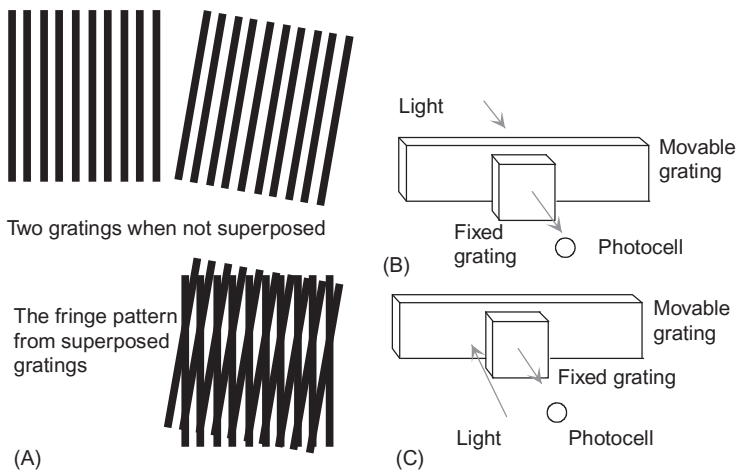


FIGURE 2.11 (A) Moiré fringes, (B) transmission, and (C) reflection forms of instrument.

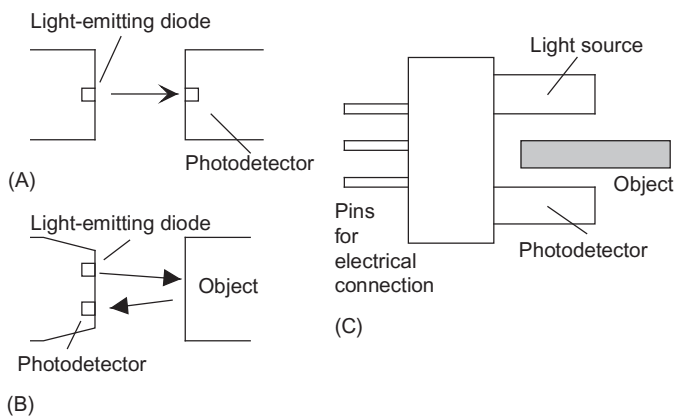


FIGURE 2.12 Photoelectric proximity sensors.

thus proportional to the number of pulses counted. Displacements as small as $1\ \mu\text{m}$ can be detected by this means. Such methods have high reliability and are often used for the control of machine tools.

2.2.7 Optical Proximity Sensors

There are a variety of optical sensors that can be used to determine whether an object is present or not. Photoelectric switch devices can either operate as *transmissive types* where the object being detected breaks a beam of light, usually infrared radiation, and stops it reaching the detector (Figure 2.12A) or *reflective types* where the object being detected reflects a beam of light onto the detector (Figure 2.12B).

In both types the radiation emitter is generally a *light-emitting diode (LED)*. The LED is a basic pn junction diode which emits light when suitable voltage is applied to it. The LED emits radiation which is reflected by the object. The reflected radiation is then detected by a phototransistor. In the absence of the object there is no detected reflected radiation; when the object is in the proximity, there is. The radiation detector might be a *photodiode* or a *phototransistor*, often a pair of transistors, known as a *Darlington pair*; using the pair increases the sensitivity. Depending on the circuit used, the output can be made to switch to either high or low when light strikes the transistor. Such sensors are supplied as packages for sensing the presence of objects at close range, typically at less than about 5 mm.

The OMRON Z4W-V displacement sensor uses a red light LED as the light source and this gives a beam of diameter 2 mm to be reflected from an object and the reflection detected (as in Figure 2.12B). The range is $\pm 4\ \text{mm}$, resolution is 10 microns and response time 5 ms. Indicators on the sensor package show green if the object being sensed is in range and red if out of range. The output from the sensor is an analogue signal of 4 to 20 mA.

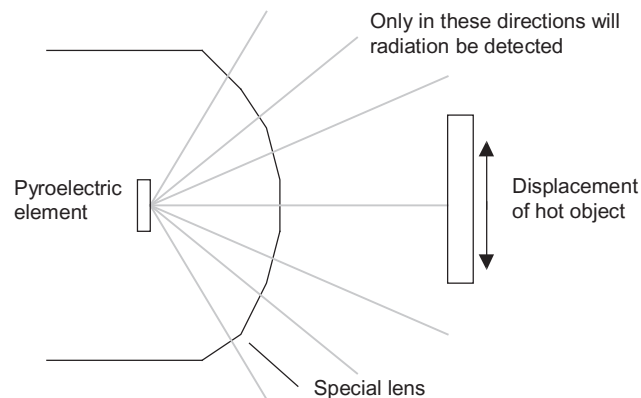


FIGURE 2.13 Pyroelectric sensor.

Figure 2.12C shows a U-shaped form where the object breaks the light beam and so changes the response from a photodetector.

Another form of optical sensor is the *pyroelectric sensor*. Such sensors give a voltage signal when the infrared radiation falling on them changes, no signal being given for constant radiation levels. Lithium tantalate is a widely used pyroelectric material. Figure 2.13 shows an example of such a sensor. Such sensors can be used with burglar alarms or for the automatic switching on of a light when someone walks up the drive to a house. A special lens is put in front of the detector. When an object which emits infrared radiation is in front of the detector, the radiation is focused by the lens onto the detector. But only for beams of radiation in certain directions will a focused beam fall on the detector and give a signal. Thus when the object moves then the focused beam of radiation is effectively switched on and off as the object cuts across the lines at which its radiation will be detected. Thus the pyroelectric detector gives a voltage output related to the changes in the signal.

2.2.8 Mechanical Switches

There are many situations where a sensor is required to detect the presence of some object. The sensor used in such situations can be a mechanical switch, giving an on–off output when the switch contacts are opened or closed by the presence of an object. Switches are used for such applications as where a work piece closes the switch by pushing against it when it reaches the correct position on a work table, such a switch being referred to as a *limit switch* because they were first used to determine the limit of travel of a moving part and when an object has reached its end of travel. The switch might then be used to switch on a machine tool to carry out some operation on the work piece. Another example is a light being required to come on when a door is opened, as in a refrigerator. The action of opening the door can be made to close the contacts in a switch and trigger an electrical circuit to switch on the lamp. A similar application is to switch on the interior light in an automobile when a door is opened. Another application is to sense the end-of-travel position for electric windows and switch off the motor moving the windows. Limit switches are widely used, being rugged, reliable and easily installed.

The basic part of a limit switch is an electrical switch which is switched on or off by means of a plunger (Figure 2.14A), the movement of this plunger being controlled by the actuator head of the limit switch which transfers the external force and movement to it. Depending on the object and movement being detected, the actuator head can take a number of forms. A common form is a roller rotating an arm to move a plunger inside the switch to activate the switch contacts and having a spring return (Figure 2.14B). The way in which the plunger moves when making the switch contacts and when breaking them is determined by springs to enable the switch to be slow or fast acting when making or breaking contacts.

As an illustration of the types of limit switches commercially available, switches are available with a roller actuator lever and a range of different contact operations. For example, these might be slow action with spring return or snap action with spring return. With snap action, movement of the actuator creates a fast change in contact state once the actuator has reached the operating position, with the slow action relay the contacts are operated at a speed proportional to the speed of operation of the actuator. With the switches the angle through which the lever has to be rotated to activate the switch can be selected and can range from just a few degrees to about

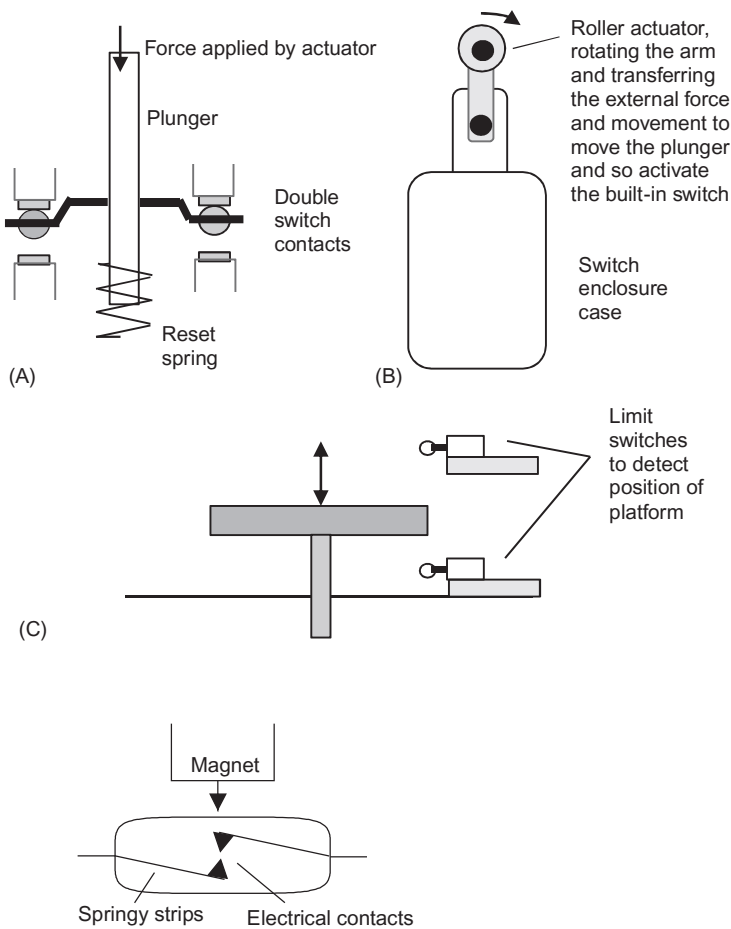


FIGURE 2.14 The basic form of a limit switch: (A) the built-in switch with the contacts being switched by the movement of the plunger, (B) a roller-actuated limit switch, rotation of the roller arm moving the plunger, and (C) a simple application of limit switches.

FIGURE 2.15 Reed switch.

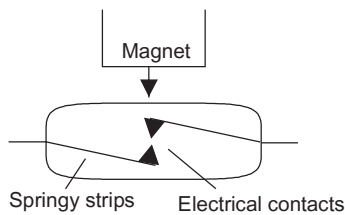
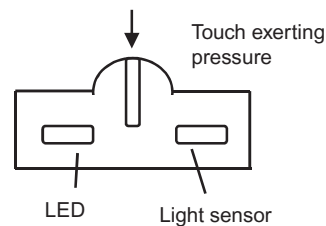


FIGURE 2.16 Touching causes the plunger to interrupt the light beam and so the signal detected by the light sensor.



25°. The Honeywell GLHC33AA4J limit switch is side rotary and has one normally closed (NC) and one normally open (NO) slow action break-before-make switches.

Figure 2.15 shows a form of a non-contact switch sensor, a *reed switch*. This consists of two overlapping, but not touching, strips of a spring magnetic material sealed in a glass or plastic envelope. When a magnet or current carrying coil is brought close to the switch, the strips become magnetised and attract each other. The contacts then close. Typically a magnet closes the contacts when it is about 1 mm from the switch.

There are situations where detectors are required to be touch-sensitive, giving outputs when touched. One example of this used as a resistance element is a doped conductive elastomer. When it is touched, the pressure causes it to elastically indent and become compressed. This increases the doping density in that region and so results in a resistance change. Such sensors have been developed in the form of the doped elastomer in the form of cords and laid in a grid pattern with the resistances being monitored at the points of intersection to yield information about which ones have been touched and so give information about the size and shape of the object in contact with the grid. A sensor that has been used with robot hands to monitor the force exerted when the hand grips an object involves a plunger being moved by the force and interrupting the light beam from a LED and hence the output from a light sensor (Figure 2.16).

2.2.9 Capacitive Proximity Sensor

A proximity sensor that can be used with metallic and non-metallic objects is the *capacitive proximity sensor*. This is a non-contact sensor which has a probe which essentially consists of a pair of parallel capacitor plates between which a potential difference is applied to produce an electric field. This field is disturbed when it is

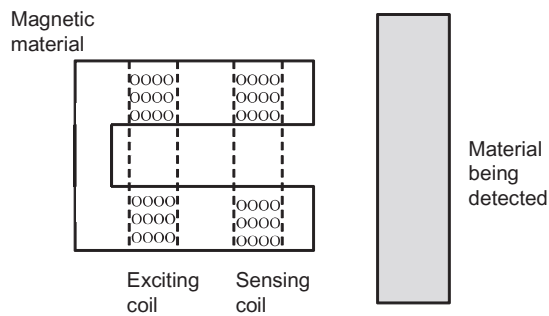


FIGURE 2.17 Inductive proximity switch.

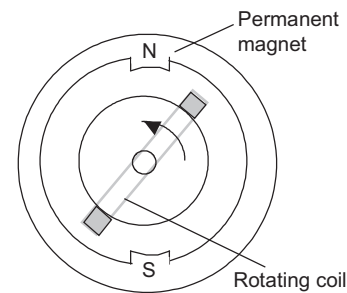


FIGURE 2.18 The tachogenerator.

moved close to an object. Thus the proximity of the object is detected by a change in capacitance. The sensor can also be used to detect a wide variety of non-metallic and metallic objects and typically operate over a range of 3 to 30 mm.

2.2.10 Inductivity Proximity Sensor

This is basically a coil which has an alternating current applied to it to produce an alternating magnetic field which is sensed by another coil (Figure 2.17). When a metallic object is moved close to the coil it disturbs the magnetic field and so the magnetic field detected by the sensing coil. Such a sensor can be used to detect metallic objects up to about 8 mm away. The OMRON ZX-EDRT has a measurement range of 0 to 5 mm with an accuracy of 1 μm .

2.3 SPEED SENSORS

The following are examples of sensors that can be used to monitor linear and angular speeds.

2.3.1 Optical Methods

Linear speeds can be measured by determining the time between when the moving object breaks one beam of radiation and when it breaks a second beam some measured distance away. Breaking the first beam can be used to start an electronic clock and breaking the second beam to stop the clock.

2.3.2 Incremental Encoder

The incremental encoder described above in Section 2.2.5 can be used for a measurement of angular speed or a rotating shaft, the number of pulses produced per second being counted.

2.3.3 Tachogenerator

The basic tachogenerator consists of a coil mounted in a magnetic field (Figure 2.18). When the coil rotates electromagnetic induction results in an alternating e.m.f. being induced in the coil. The faster the coil rotates the greater the size of the alternating e.m.f. Thus the size of the alternating e.m.f. is a measure of the angular speed. Typically such a sensor can be used up to 10 000 revs per minute and has a non-linearity error of about $\pm 1\%$ of the full range.

2.4 FLUID PRESSURE SENSORS

Many of the devices used to monitor fluid pressure in industrial processes involve the monitoring of the elastic deformation of diaphragms, bellows, and tubes. The following are some common examples of such sensors.

The term *absolute pressure* is used for a pressure measured relative to a vacuum, *differential pressure* when the difference between two pressures is measured and *gauge pressure* for the pressure measured relative to some fixed pressure, usually the atmospheric pressure.

2.4.1 Diaphragm Sensor

The movement of the centre of a circular diaphragm as a result of a pressure difference between its two sides is the basis of a pressure gauge (Figure 2.19A). For the measurement of the absolute pressure, the opposite side of the diaphragm is a vacuum; for the measurement of pressure difference, the pressures are connected to each side of the diaphragm; for the gauge pressure, i.e. the pressure relative to the atmospheric pressure, the opposite side of the diaphragm is open to the atmosphere. The amount of movement with a plane diaphragm is fairly limited; greater movement can, however, be produced with a diaphragm with corrugations (Figure 2.19B).

The movement of the centre of a diaphragm can be monitored by some form of displacement sensor. Figure 2.20 shows the form that might be taken when strain gauges are used to monitor the displacement, the strain gauges being stuck to the diaphragm and changing resistance as a result of the diaphragm movement. Typically such sensors are used for pressures over the range 100 kPa to 100 MPa, with an accuracy up to about $\pm 0.1\%$.

One form of diaphragm pressure gauge uses strain gauge elements integrated within a silicon diaphragm and supplied, together with a resistive network for signal processing, on a single silicon chip as the Motorola MPX pressure sensor (Figure 2.21). With a voltage supply connected to the sensor, it gives an output voltage directly proportional to the pressure. In one form it has a built-in vacuum on one side of the diaphragm and so the deflection of the diaphragm gives a measure of the absolute pressure applied to the other side of the diaphragm. The output is a voltage which is proportional to the applied pressure with a sensitivity of 0.6 mV/kPa. Other versions are available which have one side of the diaphragm open to the atmosphere and so can be used to measure gauge pressure, others allow pressures to be applied to both sides of the diaphragm and so can be used to measure differential pressures. Such sensors are available for use for the measurement of absolute pressure, differential pressure or gauge pressure, e.g. MPX2100 has a pressure range of 100 kPa and with a supply voltage of 16 V d.c. gives a voltage output over the full range of 40 mV.

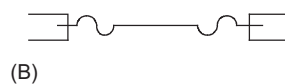
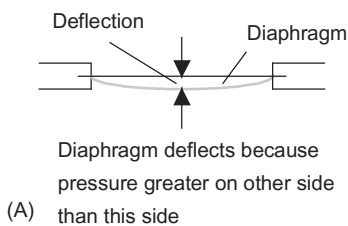


FIGURE 2.19 Diaphragm pressure sensors.

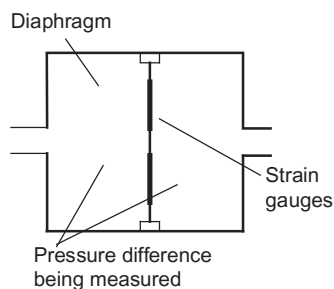


FIGURE 2.20 Diaphragm pressure.

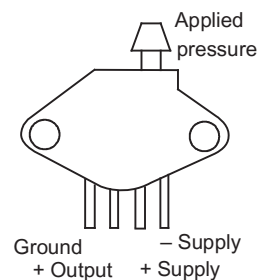


FIGURE 2.21 MPX100AP using strain gauges.

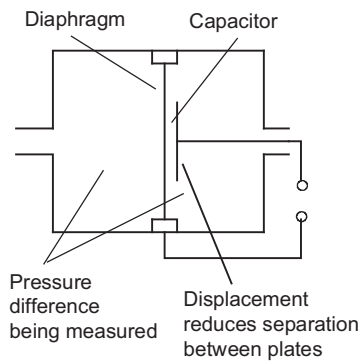


FIGURE 2.22 Diaphragm gauge using capacitance.

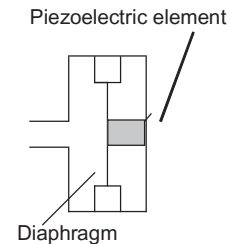


FIGURE 2.23 Basic form of a piezoelectric sensor.

The specification of a MPX pressure sensor typically includes:

- Pressure range 0 to 100 kPa
- Supply voltage 10 V
- Sensitivity 0.4 mV/kPa
- Linearity $\pm 0.25\%$ of full scale
- Pressure hysteresis $\pm 0.1\%$ of full scale
- Response time (10% to 90%) 1.0 ms

Figure 2.22 shows the form that might be taken by a capacitance diaphragm pressure gauge. The diaphragm forms one plate of a capacitor, the other plate being fixed. Displacement of the diaphragm results in changes in capacitance. The range of such pressure gauges is about 1 kPa to 200 kPa with an accuracy of about $\pm 0.1\%$.

Another form of diaphragm sensor uses a LVDT to monitor the displacement of the diaphragm. Such an arrangement is typically used for low pressure measures where high stability is required. The total error due to non-linearity, hysteresis, and repeatability can be of the order of $\pm 0.5\%$ of full scale.

2.4.2 Piezoelectric Sensor

When certain crystals are stretched or compressed, charges appear on their surfaces. This effect is called *piezoelectricity*. Examples of such crystals are quartz, tourmaline, and zirconate-titanate.

A piezoelectric pressure gauge consists essentially of a diaphragm which presses against a piezoelectric crystal (Figure 2.23). Movement of the diaphragm causes the crystal to be compressed and so charges produced on its surface. The crystal can be considered to be a capacitor which becomes charged as a result of the diaphragm movement and so a potential difference appears across it. The amount of charge produced and hence the potential difference depends on the extent to which the crystal is compressed and hence is a measure of the displacement of the diaphragm and so the pressure difference between the two sides of the diaphragm. If the pressure keeps the diaphragm at a particular displacement, the resulting electrical charge is not maintained but leaks away. Thus the sensor is not suitable for static pressure measurements. Typically such a sensor can be used for pressures up to about 1000 MPa with a non-linearity error of about $\pm 1.0\%$ of the full range value.

A commercially available piezoelectric diaphragm pressure gauge has in its specification:

- Ranges 0 to 20 MPa, 0 to 200 MPa, 0 to 500 MPa, 0 to 1000 MPa,
- Non-linearity error $\pm 0.5\%$,
- Sensitivity -0.1 pC/kPa,
- Temperature sensitivity $\pm 0.5\%$ of full scale for use $+20^\circ\text{C}$ to $+100^\circ\text{C}$

2.4.3 Bourdon Tube

The *Bourdon tube* is an almost rectangular or elliptical cross-section tube made from materials such as stainless steel or phosphor bronze. With a C-shaped tube (Figure 2.24A), when the pressure inside the tube increases the closed end of the C opens out, thus the displacement of the closed end becomes a measure of the pressure.

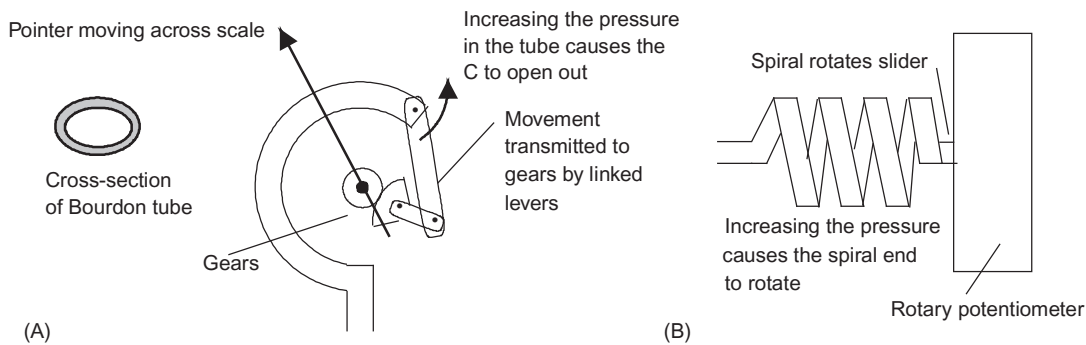


FIGURE 2.24 Bourdon tube instruments: (A) geared form, (B) potentiometer form.

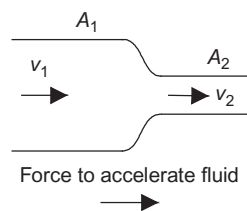


FIGURE 2.25 Pressure drop at a constriction.

A C-shaped Bourdon tube can be used to rotate, via gearing, a shaft and cause a pointer to move across a scale. Such instruments are robust and typically used for pressures in the range 10 kPa to 100 MPa with an accuracy of about $\pm 1\%$ of full scale.

Another form of Bourdon instrument uses a helical-shaped tube (Figure 2.24B). When the pressure inside the tube increases, the closed end of the tube rotates and thus the rotation becomes a measure of the pressure. A helical-shaped Bourdon tube can be used to move the slider of a potentiometer and so give an electrical output related to the pressure. Helical tubes are more expensive but have greater sensitivity. Typically they are used for pressures up to about 50 MPa with an accuracy of about $\pm 1\%$ of full range.

2.5 FLUID FLOW

For a fluid flowing through a pipe of cross-sectional area A_1 with a velocity v_1 (Figure 2.25), in 1 s the fluid advances a distance v_1 and so the volume of fluid passing a particular point per second is A_1v_1 . If the fluid then flows through a constriction of cross-sectional area A_2 in the pipe then we must have $A_2v_2 = A_1v_1$ and so there must be an increase in velocity. An increase in velocity means an acceleration and therefore a force is required to move the fluid through the constriction. This force is provided by the pressure in the fluid dropping at the constriction.

The traditional methods used for the measurement of fluid flow involve devices based on the measurement of the pressure difference occurring at a constriction and using it as a measure of the flow rate. The relationship between the pressure drop and the volume rate of flow is non-linear, i.e. the flow rate is not directly proportional to the pressure difference but to the square root of the pressure difference. The venturi tube and the orifice plate described below are common examples.

Other methods have, however, been developed which more rapidly and efficiently record the flow rate and often with less interference to the flow.

2.5.1 Differential Pressure Methods

There are a number of forms of differential pressure devices based on the above equation and involving constant size constrictions, e.g. the venturi tube, nozzles, Dall tube and orifice plate. In addition there are other

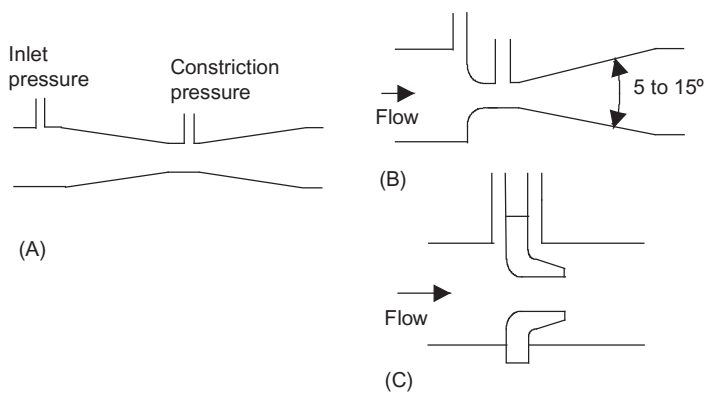


FIGURE 2.26 (A) Venturi tube, (B) venturi nozzle, (C) flow nozzle.

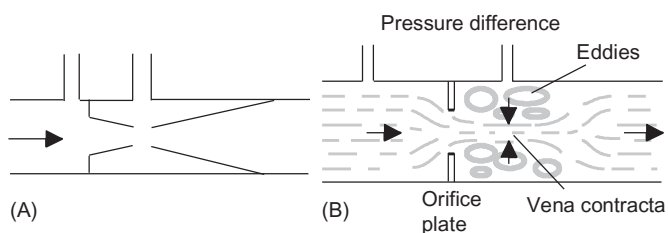


FIGURE 2.27 (A) Dall Tube, (B) Orifice plate.

devices involving variable size constrictions, e.g. the rotameter. The following are discussions of the characteristics of the above devices.

The *venturi tube* (Figure 2.26) has a gradual tapering of the pipe from the full diameter to the constricted diameter. The presence of the venturi tube results in a pressure loss occurring in the system of about 10 to 15%, a comparatively low value. The pressure difference between the flow prior to the constriction and the constriction can be measured with a simple U-tube manometer or a differential diaphragm pressure cell. The instrument can be used with liquids containing particles, is simple in operation, capable of accuracy of about $\pm 0.5\%$, has a long-term reliability, but is comparatively expensive and has a non-linear relationship between pressure and the volume rate of flow.

A cheaper form of venturi is provided by the *nozzle flow meter* (Figure 2.26). Two types of nozzle are used, the venturi nozzle and the flow nozzle. The venturi nozzle (Figure 2.26A) is effectively a venturi tube with an inlet which is considerably shortened. The flow nozzle (Figure 2.26B) is even shorter. Nozzles produce pressure losses of the order of 40 to 60%. Nozzles are cheaper than venturi tubes, give similar pressure differences, and have an accuracy of about $\pm 0.5\%$. They have the same non-linear relationship between the pressure and the volume rate of flow.

The *Dall tube* (Figure 2.27) is another variation of the venturi tube. It gives a higher differential pressure and a lower pressure drop. The Dall tube is only about two pipe diameters long and is often used where space does not permit the use of a venturi tube. The *orifice plate* (Figure 2.27) is simply a disc with a hole. The effect of introducing it is to constrict the flow to the orifice opening and the flow channel to an even narrower region downstream of the orifice. The narrowest section of the flow is not through the orifice but downstream of it and is referred to as the *vena contracta*. The pressure difference is measured between a point equal to the diameter of the tube upstream of the orifice and a point equal to half the diameter downstream. The orifice plate has the usual non-linear relationship between the pressure difference and the volume rate of flow. It is simple, reliable, produces a greater pressure difference than the venturi tube and is cheaper but less accurate, about $\pm 1.5\%$. It also produces a greater pressure drop. Problems of silting and clogging can occur if particles are present in liquids.

The *rotameter* (Figure 2.28) is an example of a *variable area flow meter*; a constant pressure difference is maintained between the main flow and that at the constriction by changing the area of the constriction. The rotameter has a float in a tapered vertical tube with the fluid flow pushing the float upwards. The fluid has to flow through the constriction which is the gap between the float and the walls of the tube and so there is a pressure drop at that point. Since the gap between the float and the tube walls increases as the float moves upwards, the pressure drop decreases. The float moves up the tube until the fluid pressure is just sufficient to balance the weight of the

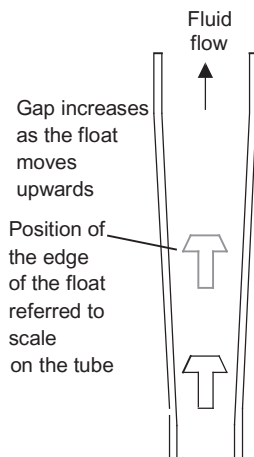


FIGURE 2.28 Rotameter.

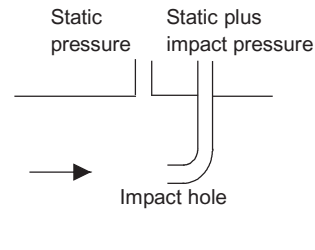


FIGURE 2.29 Pitot tube.

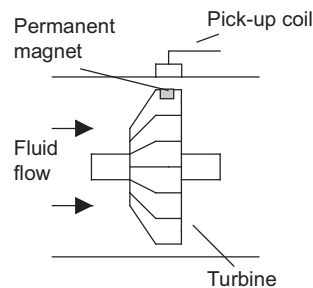


FIGURE 2.30 Basic principle of the turbine flowmeter.

float. The greater the flow rate the greater the pressure difference for a particular gap and so the higher up the tube the float moves. A scale alongside the tube can thus be calibrated to read directly the flow rate corresponding to a particular height of the float. The rotameter is cheap, reliable, has an accuracy of about $\pm 1\%$ and can be used to measure flow rates from about $30 \times 10^{-6} \text{ m}^3/\text{s}$ to $1 \text{ m}^3/\text{s}$.

The *Pitot tube* can be used to directly measure the velocity of flow of a fluid, rather than the volume rate of flow and consists essentially of just a small tube inserted into the fluid with an opening pointing directly upstream (Figure 2.29). The fluid impinging on the open end of the tube is brought to rest and the pressure difference measured between this point and the pressure in the fluid at full flow. The difference in pressure between where the fluid is in full flow and the point where it is stopped is due to the kinetic energy of the fluid being transformed to potential energy, this showing up as an increase in pressure. Because kinetic energy is $\frac{1}{2}mv^2$, the velocity is proportional to the square root of the pressure difference.

2.5.2 Turbine Meter

The *turbine flowmeter* (Figure 2.30) consists of a multi-bladed rotor that is supported centrally in the pipe along which the flow occurs. The rotor rotates as a result of the fluid flow, the angular velocity being approximately proportional to the flow rate. The rate of revolution of the rotor can be determined by attaching a small permanent magnet to one of the blades and using a pick-up coil. An induced e.m.f. pulse is produced in the coil every time the magnet passes it. The pulses are counted and so the number of revolutions of the rotor can be determined. The meter is expensive, with an accuracy of typically about $\pm 0.1\%$. Another form uses helical screws which rotate as a result of the fluid flow.

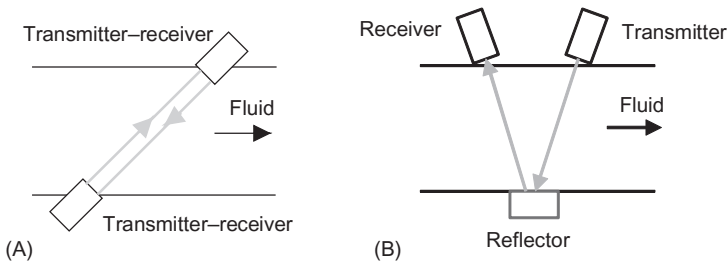


FIGURE 2.31 Ultrasonic flow meter.

2.5.3 Ultrasonic Time of Flight Flow Meter

Figure 2.31 shows one way ultrasonic waves can be used to determine the flow rate of a fluid. There are either a pair of ultrasonic receivers and transmitters with one pair on each side of the pipe through which the fluid flows and the other pair on the other side or a transmitter and receiver on one side of the pipe and a reflector on other. Either way, there is one beam of ultrasonic waves travelling in the direction of the flow and one beam in the direction against the flow. The flow meter can either measure the average fluid velocity by averaging the difference in measured transit times between the pulses of ultrasound going across the tube into and against the direction of flow or by measuring the frequency shift using the Doppler effect, this depending on the fluid containing discontinuities that reflect the ultrasonic wave, e.g. slurries and liquids with bubbles. Ultrasonic methods can be used for pipes from 75 mm to 1500 mm diameter, with fluid velocities from about 0.2 m/s to 12 m/s with an accuracy of about $\pm 1\%$ or better.

The OMEGA Doppler ultrasonic flow meter FD-400, of the form in Figure 2.31B, includes the following in its specification:

- Non-intrusive, clamp-on transducer for most pipes from 6 mm to 3 m
- Flow velocity range 0.05 to 9 m/s
- Accuracy $\pm 2\%$ FS over calibrated range
- Monitor display: 2-line 8 character LCD

2.5.4 Vortex Flow Rate Method

When a fluid flow encounters a body, the layers of fluid close to the surfaces of the body are slowed down. With a streamlined body, these boundary layers follow the contours of the body until virtually meeting at the rear of the object. This results in very little wake being produced. With a non-streamlined body, a so-called *bluff body*, the boundary layers detach from the body much earlier and a large wake is produced. When the boundary layer leaves the body surface it rolls up into vortices. These are produced alternately from the top and bottom surfaces of the body (Figure 2.32). The result is two parallel rows of vortices moving downstream with the distance between successive vortices in each row being the same, a vortex in one row occurring half way between those in the other row. For a particular bluff body, the number of vortices produced per second f , i.e. the frequency, is proportional to the flow rate. A number of methods are used for the measurement of the frequency. For example, a thermistor might be located behind the face of the bluff body (Figure 2.32B(i)). The thermistor, heated as a result of a current passing through it, senses vortices due to the cooling effect caused by their breaking away. Another method uses a piezoelectric crystal mounted in the bluff body (Figure 2.32B(ii)). Flexible diaphragms react to the pressure disturbances produced by the vortices and are detected by the crystal.

Vortex flow meters are used for both liquids and gases, having an output which is independent of density, temperature or pressure, and having an accuracy of about $\pm 1\%$. They are used at pressures up to about 10 MPa and temperatures of 200°C.

Another form of flowmeter uses a MEMS chip. MEMS are micro-electro-mechanical systems and are very small devices. They usually consist of a central unit that processes data (the microprocessor) and several components that interact with the surroundings such as microsensors. The OMRON D6F MEMS flow meter is a chip only 1.5 mm² by 0.4 mm thick and has two thermopiles on either side of a tiny heater element. When a gas flows past the sensor a difference in temperature occurs between the two thermopiles and this gives an output potential difference between the two which is related to the flow rate.

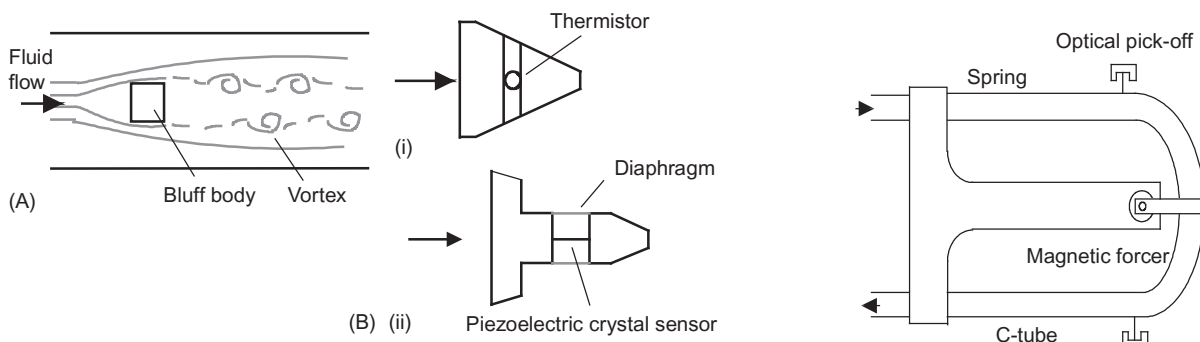


FIGURE 2.32 (A) Vortex shedding, (B) detection systems: (i) thermistor, (ii) piezoelectric crystal. FIGURE 2.33 Coriolis flow meter.

2.5.5 Coriolis Flow Meter

If a skater is spinning with arms outstretched and then pulls in his or her arms, they spin faster. As a consequence we can think of there being a torque acting on the skater's body to result in the increased angular velocity. This torque is considered to arise from a tangential force called the *Coriolis force*. When we move an object in a rotating system, it seems to be pushed sideways. For a body of mass M moving with constant linear radial velocity v and subject to an angular velocity ω the Coriolis force is $2M\omega v$.

The *Coriolis flow meter* consists basically of a C-shaped pipe (Figure 2.33) through which the fluid flows. The pipe, and fluid in the pipe, is given an angular acceleration by being set into vibration, this being done by means of a magnet mounted in a coil on the end of a tuning fork-like leaf spring. Oscillations of the spring then set the C-tube into oscillation. The result is an angular velocity that alternates in direction. At some instant the Coriolis force acting on the fluid in the upper limb is in one direction and in the lower limb in the opposite direction, this being because the velocity of the fluid is in opposite directions in the upper and lower limbs. The resulting Coriolis forces on the fluid in the two limbs are thus in opposite directions and cause the limbs of the C to become displaced. When the direction of the angular velocity is reversed then the forces reverse in direction and the limbs become displaced in the opposite direction. These displacements are proportional to the mass flow rate of fluid through the tube. The displacements are monitored by means of optical sensors, their outputs being a pulse with a width proportional to the mass flow rate. The flow meter can be used for liquids or gases and has an accuracy of $\pm 0.5\%$. It is unaffected by changes in temperature or pressure.

2.6 LIQUID LEVEL

Methods used to measure the level of liquid in a vessel include those based on:

1. Floats whose position is directly related to the liquid level.
2. Archimedes' principle and a measurement of the upthrust acting on an object partially immersed in the liquid; the term displacer is used.
3. A measurement of the pressure at some point in the liquid, the pressure due to a column of liquid of height h being $h\rho g$, where ρ is the liquid density, and g the acceleration due to gravity.
4. A measurement of the weight of the vessel containing the liquid plus liquid. The weight of the liquid is $Ah\rho g$, where A is the cross-sectional area of the vessel, h the height of liquid, ρ its density, and g the acceleration due to gravity and thus changes in the height of liquid give weight changes.
5. A change in electrical conductivity when the liquid rises between two probes.
6. A change in capacitance as the liquid rises up between the plates of a capacitor.
7. Ultrasonic and nuclear radiation methods.

The following gives examples of the above methods used for liquid level measurements.

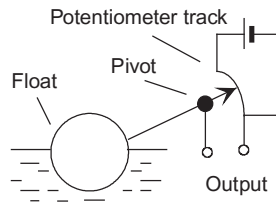


FIGURE 2.34 Potentiometer float gauge.

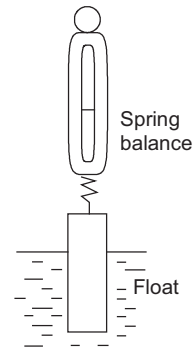


FIGURE 2.35 Displacer gauge.

2.6.1 Floats

Figure 2.34 shows a simple *float system*. The float is at one end of a pivoted rod with the other end connected to the slider of a potentiometer. Changes in level cause the float to move and hence move the slider over the potentiometer resistance track and so give a potential difference output related to the liquid level. A problem with floats is that there is the problem of fluids and deposits coating the floats and changing the level at which they float.

2.6.2 Displacer Gauge

When an object is partially or wholly immersed in a fluid it experiences an upthrust force equal to the weight of fluid displaced by the object. This is known as *Archimedes' principle*. Thus a change in the amount of an object below the surface of a liquid will result in a change in the upthrust. The resultant force acting on such an object is then its weight minus the upthrust and thus depends on the depth to which the object is immersed. For a vertical cylinder of cross-sectional area A in a liquid of density ρ , if a height h of the cylinder is below the surface then the upthrust is $hA\rho g$, where g is the acceleration due to gravity, and so the apparent weight of the cylinder is $(mg - hA\rho g)$, where m is the mass of the cylinder. Such *displacer gauges* need calibrating for liquid level determinations for particular liquids since the upthrust depends on the liquid density. Figure 2.35 shows a simple version of a displacer gauge. A problem with displacers is that there is the problem of fluids coating the floats and apparently changing the buoyancy.

2.6.3 Differential Pressure

The pressure due to a height h of liquid above some level is $h\rho g$, where ρ is the liquid density and g the acceleration due to gravity. With a tank of liquid open to the atmosphere, the pressure difference can be measured between a point near the base of the tank and the atmosphere. The result is then proportional to the height of liquid above the pressure measurement point (Figure 2.36A). With a closed tank, the pressure difference has to be measured between a point near the bottom of the tank and in the gases above the liquid surface (Figure 2.36B). The pressure gauges used for such measurements tend to be diaphragm instruments.

2.6.4 Load Cell

The weight of a tank of liquid can be used as a measure of the height of liquid in the tank. Load cells are commonly used for such weight measurements. One form of *load cell* consists of a strain-gauged cylinder (Figure 2.37) which is included in the supports for the tank of liquid. When the level of the liquid changes, the weight changes and so the load on the load cell changes and the resistances of the strain gauges change. The resistance changes of the strain gauges are thus a measure of the level of the liquid. Since the load cells are completely isolated from the liquid, the method is useful for corrosive liquids.

A typical specification for such a load cell is that for the Omega LCM304 which is a small 25 mm diameter cell available with a number of ranges, e.g. 0 to 500 N, 0 to 200 N, 0 to 5000 N, 0 to 10 000 N, 0 to 20 000 N, 0 to 50 000 N. The accuracy is $\pm 0.5\%$ of the full-scale-output, linearity, hysteresis, repeatability combined.

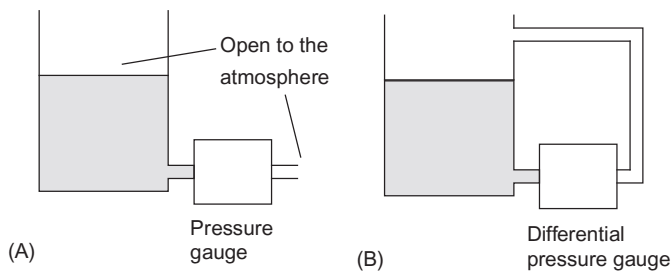


FIGURE 2.36 Pressure level gauges.

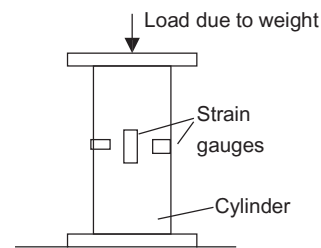


FIGURE 2.37 Load cell.

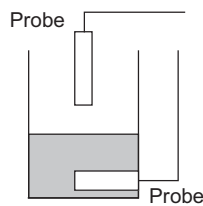


FIGURE 2.38 Conductivity level indicator.

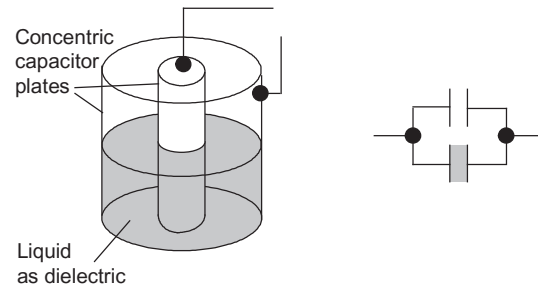


FIGURE 2.39 Capacitive gauge.

2.6.5 Electrical Conductivity Level Indicator

Conductivity methods can be used to indicate when the level of a high electrical conductivity liquid reaches a critical level. One form has two probes, one probe mounted in the liquid and the other either horizontally at the required level or vertically with its lower end at the critical level (Figure 2.38). When the liquid is short of the required level, the resistance between the two probes is high since part of the electrical path between the two probes is air. However, when the liquid level reaches the critical level, there is a path entirely through the liquid and so the conductivity drops. Foaming, splashing, and turbulence can affect the results.

An integrated circuit LM1830N can be used for signal processing with conductivity probes so that an output is given which can be used to activate a loudspeaker or a LED. The circuit compares the resistance of the liquid with the IC's internal reference resistance.

2.6.6 Capacitive Level Indicator

A common form of *capacitive level gauge* consists of two concentric conducting cylinders, or a circular rod inside a cylinder, acting as capacitor plates with the liquid between them acting as the dielectric of a capacitor (Figure 2.39). If the liquid is an electrical insulator then the capacitor plates can be bare metal, if the liquid is conducting then they are metal coated with an insulator, e.g. Teflon. The arrangement consists essentially of two capacitors in parallel, one formed between the plates inside the liquid and the other from that part of the plates in the air above the liquid. A change in the liquid level changes the total capacitance of the arrangement. Errors can arise as a result of temperature changes since they will produce a change in capacitance without any change in level. Errors can also arise if, when the liquid level drops, the electrodes remain coated with liquid. The system can be used, with suitable choice of electrode material, for corrosive liquids and is capable of reasonable accuracy.

2.6.7 Ultrasonic Level Gauge

In one version of an *ultrasonic level indicator*, an ultrasonic transmitter/receiver is placed above the surface of the liquid (Figure 2.40). Ultrasonic pulses are produced, travel down to the liquid surface and are then reflected back to the receiver. The time taken from emission to reception of the pulses can be used as a measure of the position of the liquid surface. Because the receiver/transmitter can be mounted outside the liquid, it is particularly useful for corrosive liquids. Errors are produced by temperature changes since they affect the speed of the sound wave. Such errors are typically about 0.18% per °C.

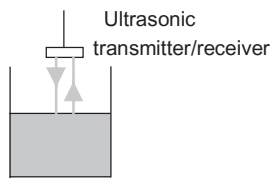


FIGURE 2.40 Ultrasonic gauge.

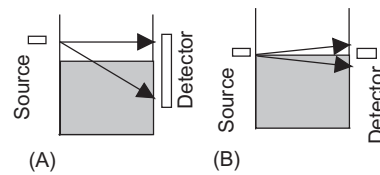


FIGURE 2.41 Radionic gauges.

2.6.8 Nucleonic Level Indicators

One form of level indicator uses gamma radiation from a radioactive source, generally cobalt-60, caesium-137, or radium-226. A detector is placed on one side of the container and the source on the other. The intensity of the radiation depends on the amount of liquid between the source and detector and can be used to determine the level of the liquid. Figure 2.41 shows two possible arrangements. With a compact source and extended detector, level changes over the length of the detector can be determined. A compact source and a compact detector can be used where small changes in a small range of level are to be detected. Such methods can be used for liquids, slurries, and solids, and, since no elements of the system are in the liquid, for corrosive and high temperature liquids.

2.7 TEMPERATURE SENSORS

The expansion or contraction of solids, liquids, or gases, the change in electrical resistance of conductors and semiconductors, thermoelectric e.m.f.s, and the change in the current across the junction of semiconductor diodes and transistors are all examples of properties that change when the temperature changes and can be used as basis of temperature sensors. The following are some of the more commonly used temperature sensors.

2.7.1 Bimetallic Strips

A *bimetallic strip* consists of two different metal strips of the same length bonded together (Figure 2.42). Because the metals have different coefficients of expansion, when the temperature increases the composite strip bends into a curved strip, with the higher coefficient metal on the outside of the curve. The amount by which the strip curves depends on the two metals used, the length of the composite strip and the change in temperature. If one end of a bimetallic strip is fixed, the amount by which the free end moves is a measure of the temperature. This movement may be used to open or close electric circuits, as in the simple thermostat commonly used with domestic heating systems. Bimetallic strip devices are robust, relatively cheap, have an accuracy of the order of $\pm 1\%$, and are fairly slow reacting to changes in temperature.

2.7.2 Liquid in Glass Thermometers

The *liquid in glass thermometer* involves a liquid expanding up a capillary tube. The height to which the liquid expands is a measure of the temperature. With mercury as the liquid, the range possible is -35°C to $+600^{\circ}\text{C}$, with alcohol -80°C to $+70^{\circ}\text{C}$, with pentane -200°C to $+30^{\circ}\text{C}$. Such thermometers are direct reading, fragile, capable of reasonable accuracy under standardised conditions, fairly slow reacting to temperature changes, and cheap.

2.7.3 Resistance Temperature Detectors (RTDs)

The resistance of most metals increases in a reasonably linear way with temperature (Figure 2.43) and can be represented by the equation:

$$R_t = R_0(1 + \alpha t)$$

where R_t is the resistance at a temperature $t^{\circ}\text{C}$, R_0 the resistance at 0°C and α a constant for the metal, termed the temperature coefficient of resistance. *Resistance temperature detectors* (RTDs) are simple resistive elements in the form of coils of metal wire, e.g. platinum, nickel, or copper alloys. Platinum detectors have high linearity,

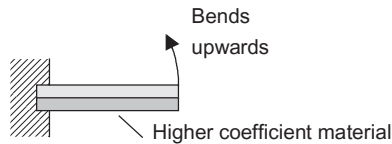


FIGURE 2.42 Bimetallic strip.

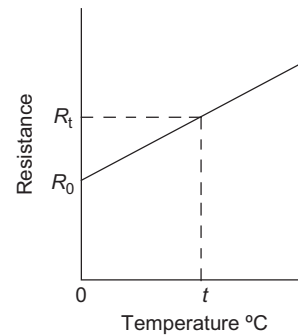


FIGURE 2.43 Resistance variation with temperature for metals.

good repeatability, high long-term stability, can give an accuracy of $\pm 0.5\%$ or better, a range of about -200°C to $+850^{\circ}\text{C}$, can be used in a wide range of environments without deterioration, but are more expensive than the other metals. They are, however, very widely used. Nickel and copper alloys are cheaper but have less stability, are more prone to interaction with the environment, and cannot be used over such large temperature ranges.

A commercially available platinum resistance thermometer includes the following in its specification:

Range -200°C to $+800^{\circ}\text{C}$
 Accuracy $\pm 0.01^{\circ}\text{C}$
 Sensitivity $0.4 \Omega/^{\circ}\text{C}$ for 100Ω

2.7.4 Thermistors

Thermistors are semiconductor temperature sensors made from mixtures of metal oxides, such as those of chromium, cobalt, iron, manganese, and nickel. The resistance of thermistors decreases in a very non-linear manner with an increase in temperature, Figure 2.44 illustrating this. The change in resistance per degree change in temperature is considerably larger than that which occurs with metals. For example, a thermistor might have a resistance of $29 \text{ k}\Omega$ at -20°C , $9.8 \text{ k}\Omega$ at 0°C , $3.75 \text{ k}\Omega$ at 20°C , $1.6 \text{ k}\Omega$ at 40°C , $0.75 \text{ k}\Omega$ at 60°C . The material is formed into various forms of element, such as beads, discs, and rods (Figure 2.45). Thermistors are rugged and can be very small, so enabling temperatures to be monitored at virtually a point. Because of their small size they have small thermal capacity and so respond very rapidly to changes in temperature. The temperature range over which they can be used will depend on the thermistor concerned, ranges within about -100°C to $+300^{\circ}\text{C}$ being possible. They give very large changes in resistance per degree change in temperature and so are capable, over a small range, of being calibrated to give an accuracy of the order of 0.1°C or better. However, their characteristics tend to drift with time. Their main disadvantage is their non-linearity. Thermistors are commonly used to monitor the coolant temperatures and the outside and internal air temperatures in automobiles.

The following is part of the specification for a bead thermistor temperature sensor:

Accuracy $\pm 5\%$
 Maximum power 250 mW
 Dissipation factor $7 \text{ mW}/^{\circ}\text{C}$
 Response time 1.2 s
 Thermal time constant 11 s
 Temperature range -40°C to $+125^{\circ}\text{C}$

2.7.5 Thermocouples

When two different metals are joined together, a potential difference occurs across the junction. The potential difference depends on the two metals used and the temperature of the junction. A *thermocouple* involves two such junctions, as illustrated in Figure 2.46.

If both junctions are at the same temperature, the potential differences across the two junctions cancel each other out and there is no net e.m.f. If, however, there is a difference in temperature between the two junctions,

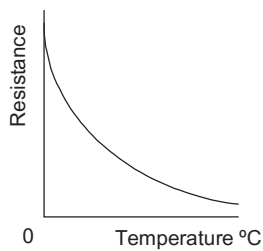


FIGURE 2.44 Variation of resistance.

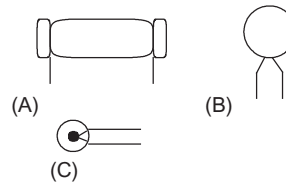


FIGURE 2.45 Thermistors: (A) rod with temperature for thermistors (B) disc, (C) bead.

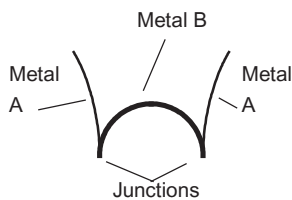


FIGURE 2.46 Thermocouple.

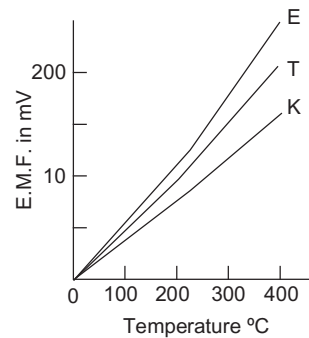


FIGURE 2.47 Thermocouples: chromel-constantan (E), chromel-alumel (K), copper-constant (C).

TABLE 2.1 Thermocouples

	Materials	Range °C	Sensitivity mV/°C
E	Chromel-constantan	0 to 980	63
J	Iron-constantan	-180 to 760	53
K	Chromel-alumel	-180 to 1260	41
R	Platinum-platinum/ rhodium 13%	0 to 1750	8
T	Copper-constantan	-180 to 370	43

there is an e.m.f. The value of this e.m.f. E depends on the two metals concerned and the temperatures t of both junctions. Usually one junction is held at 0°C and then, to a reasonable extent, the following relationship holds:

$$E = at + bt^2$$

where a and b are constants for the metals concerned.

Figure 2.47 shows how the e.m.f. varies with temperature for a number of commonly used pairs of metals. Standard tables giving the e.m.f.s at different temperatures are available for the metals usually used for thermocouples. Commonly used thermocouples are listed in Table 2.1, with the temperature ranges over which they are generally used and typical sensitivities. These commonly used thermocouples are given reference letters. The base-metal thermocouples, E, J, K, and T, are relatively cheap but deteriorate with age. They have accuracies which are typically about ± 1 to 3%. Noble-metal thermocouples, e.g. R, are more expensive but are more stable with longer life. They have accuracies of the order of $\pm 1\%$ or better. Thermocouples are generally mounted in a sheath to give them mechanical and chemical protection. The response time of an unsheathed thermocouple is very fast. With a sheath this may be increased to as much as a few seconds if a large sheath is used.

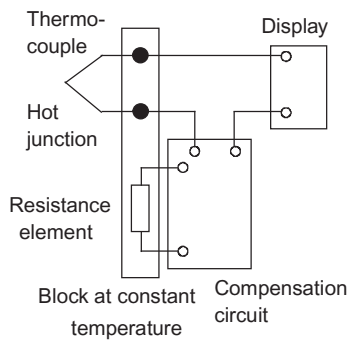


FIGURE 2.48 Cold junction compensation.

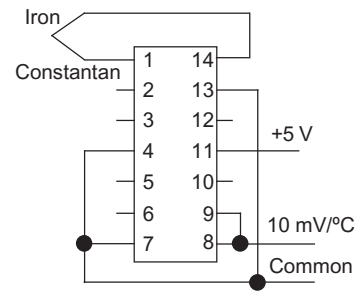


FIGURE 2.49 AD594.

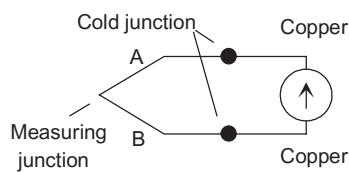


FIGURE 2.50 The junctions with a measurement instrument.

A thermocouple can be used with the reference junction at a temperature other than 0°C . However, the standard tables assume that the junction is at 0°C junction and hence a correction has to be applied before the tables can be used. The correction is applied using what is known as the *law of intermediate temperatures*, namely:

$$E_{t,0} = E_{t,I} + E_{I,0}$$

The e.m.f. $E_{t,0}$ at temperature t when the cold junction is at 0°C equals the e.m.f. $E_{t,I}$ at the intermediate temperature I plus the e.m.f. $E_{I,0}$ at temperature I when the cold junction is at 0°C . Consider a type E thermocouple. The following is data from standard tables.

Temp. ($^{\circ}\text{C}$)	0	20	200
e.m.f. (mV)	0	1.192	13.419

Thus, using the law of intermediate temperatures, the thermoelectric e.m.f. at 200°C with the cold junction at 20°C is:

$$E_{200,20} = E_{200,0} - E_{20,0} = 13.419 - 1.192 = 12.227 \text{ mV}$$

Note that this is not the e.m.f. given by the tables for a temperature of 180°C with a cold junction at 0°C , namely 11.949 mV.

To maintain one junction of a thermocouple at 0°C , it needs to be immersed in a mixture of ice and water. This, however, is often not convenient and a *compensation circuit* (Figure 2.48) is used to provide an e.m.f. which varies with the temperature of the 'cold' junction in such a way that when it is added to the thermocouple e.m.f. it generates a combined e.m.f. which is the same as would have been generated if the cold junction had been at 0°C (see Section 2.9.3), e.g. the Analog Devices AD594 (Figure 2.49). This, when used with a +5 V supply and a constantan–iron thermocouple, gives an output of $10 \text{ mV}/^{\circ}\text{C}$.

When a thermocouple is connected to a measuring circuit, other metals are involved (Figure 2.50). Thus we can have as the 'hot' junction that between metals A and B and the 'cold' junction effectively extended by the introduction of copper leads and the measurement instrument. Provided the junctions with the intermediate materials are at the same temperature, there is no extra e.m.f. involved and we still have the e.m.f. as due to the junction between metals A and B.

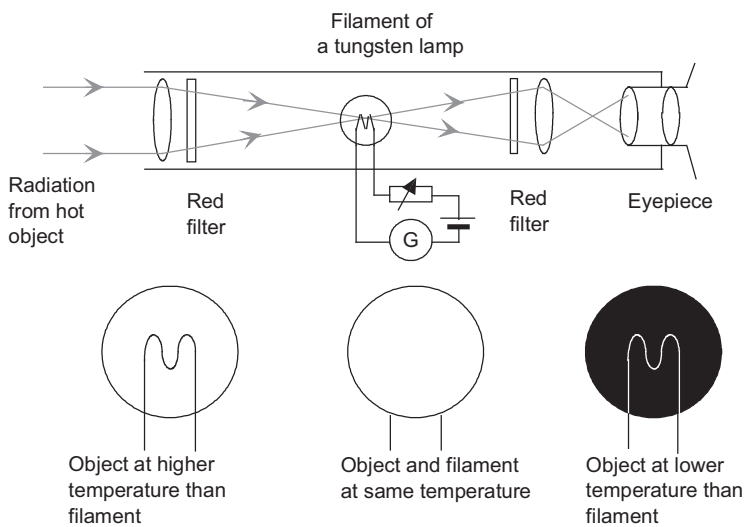


FIGURE 2.51 Disappearing filament pyrometer.

2.7.6 Thermodiodes and Transistors

When the temperature of doped semiconductors changes, the mobility of their charge carriers change. As a consequence, when a p-n junction has a potential difference across it, the current through the junction is a function of the temperature. Such junctions for use as temperature sensors are supplied, together with the necessary signal processing circuitry as integrated circuits, e.g. LM3911 which gives an output voltage proportional to temperature. In a similar manner, transistors can be used as temperature sensors. An integrated circuit temperature sensor using transistors is LM35. This gives an output, which is a linear function of temperature, of $10 \text{ mV}/^\circ\text{C}$ when the supply voltage is 5 V . The specification for LM35 includes: Accuracy at 25°C $\pm 0.4\%$, Non-linearity 0.2°C , Sensitivity $10 \text{ mV}/^\circ\text{C}$.

2.7.7 Pyrometers

Methods used for the measurement of temperature which involve the radiation emitted by the body include:

1. Optical pyrometer

This is based on comparing the brightness of the light emitted by the hot body with that from a known standard.

2. Total radiation pyrometer

This involves the measurement of the total amount of radiation emitted by the hot body by a resistance element or a thermopile.

The *optical pyrometer*, known generally as the *disappearing filament pyrometer*, involves just the visible part of the radiation emitted by a hot object. The radiation is focused onto a filament so that the radiation and the filament can both be viewed in focus through an eyepiece (Figure 2.51). The filament is heated by an electrical current until the filament and the hot object seem to be the same colour, the filament image then disappearing into the background of the hot object. The filament current is then a measure of the temperature. A red filter between the eyepiece and the filament is generally used to make the matching of the colours of the filament and the hot object easier. Another red filter may be introduced between the hot object and the filament with the effect of making the object seem less hot and so extending the range of the instrument.

The disappearing filament pyrometer has a range of about 600°C to 3000°C , an accuracy of about $\pm 0.5\%$ of the reading and involves no physical contact with the hot object. It can thus be used for moving or distant objects.

The *total radiation pyrometer* involves the radiation from the hot object being focused onto a radiation detector. Figure 2.52 shows the basic form of an instrument which uses a mirror to focus the radiation onto the detector. Some forms use a lens to focus the radiation. The detector is typically a thermopile with often up to 20 or 30 thermocouple junctions, a resistance element or a thermistor. The detector is said to be *broad band* since it detects

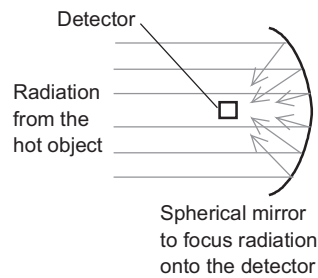


FIGURE 2.52 Total radiation pyrometer.

radiation over a wide band of frequencies and so the output is the summation of the power emitted at every wavelength. It is proportional to the fourth power of the temperature (the Stefan-Boltzmann law). The accuracy of broad band total radiation pyrometers is typically about $\pm 0.5\%$ and ranges are available within the region 0°C to 3000°C . The time constant (a measure of how fast the system responds to a change in temperature and is the time taken to reach about 63% of the final value) for the instrument varies from about 0.1 s when the detector is just one thermocouple or small bead thermistor to a few seconds with a thermopile involving many thermocouples. Some instruments use a rotating mechanical chopper to chop the radiation before it impinges on the detector. The aim is to obtain an alternating output from the detector, since amplification is easier with an alternating voltage. It is thus of particular benefit when the level of radiation is low. However, choppers can only be used with detectors which have a very small time constant and thus tend to be mainly used with small bead thermistor detectors.

2.8 SENSOR SELECTION

The selection of a sensor for a particular application requires a consideration of:

1. The nature of the measurement required, i.e. the sensor input. This means considering the variable to be measured, its nominal value, the range of values, the accuracy required, the required speed of measurement, the reliability required, and the environmental conditions under which the measurement is to be made.
2. The nature of the output required from the sensor, this determining the signal processing required. The selection of sensors cannot be taken in isolation from a consideration of the form of output that is required from the system after signal processing, and thus there has to be a suitable marriage between sensor and signal processing.

Then possible sensors can be identified, taking into account such factors as their range, accuracy, linearity, speed of response, reliability, life, power supply requirements, ruggedness, availability, and cost.

EXAMPLE

Select a sensor which can be used to monitor the temperature of a liquid in the range 10°C to 80°C to an accuracy of about 1°C and which will give an output which can be used to change the current in an electrical circuit.

There are a number of forms of sensor that can be used to monitor such a temperature in the range, and to the accuracy, indicated. The choice is, however, limited by the requirement for an output which can change the current in an electrical circuit. This would suggest a resistance thermometer. In view of the limited accuracy and range required, a thermistor might thus be considered.

EXAMPLE

Select a sensor which can be used for the measurement of the level of a corrosive acid in a circular vessel of diameter 1 m and will give an electrical output. The acid level can vary from 0 to 3 m and the minimum change in level to be detected is 0.1 m. The empty vessel has a weight of 50 kg. The acid has a density of 1050 kg/m^3 .

Because of the corrosive nature of the acid there could be problems in using a sensor which is inserted in the liquid. Thus a possibility is to use a load cell, or load cells, to monitor the weight of the vessel. Such cells would give an electrical output. The weight of the liquid changes from 0 when empty to, when full, $1050 \times 3 \times \pi(1^2/4) \times 9.8 = 24.3$ kN. Adding this to the weight of the empty vessel gives a weight that varies from about 0.5 kN to 4.9 kN. A change of level of 0.1 m gives a change in weight of $0.10 \times 1050 \times \pi(1^2/4) \times 9.8 = 0.8$ kN. If the load of the vessel is spread between three load cells, each will require a range of about 0 to 5 kN with a resolution of about 0.3 kN.

2.9 SIGNAL PROCESSING

The output signal from the sensor of a measurement system has generally to be processed in some way to make it suitable for display or use in some control system. For example, the signal may be too small and have to be amplified, be analogue and have to be made digital, be digital and have to be made analogue, be a resistance change and have to be made into a current change, be a voltage change and have to be made into a suitable size current change, be a pressure change and have to be made into a current change, etc. All these changes can be referred to as *signal processing*. For example, the output from a thermocouple is a very small voltage, a few millivolts. A signal processing module might then be used to convert this into a larger voltage and provide cold junction compensation (i.e. allow for the cold junction not being at 0°C). Note that the term *signal conditioning* is sometimes used for the conversion of the output from a sensor into a suitable form for signal processing. The following gives some of the elements that are used in signal processing.

An important issue in signal processing is that though most sensors give analogue signals, there are great advantages in converting their outputs to digital signals. Digital signals are really just numbers and have the great advantage that they can be manipulated mathematically by software in microcontroller systems. This manipulation is easier to do by programming than by the electronics that would be needed if the signals had remained as analogue signals. Digital signals can also be easily transmitted and stored. If a digital output is required to operate an analogue signal input device then the output from the microcontroller can be readily converted back to analogue.

2.9.1 Resistance to Voltage Converter

Consider how the resistance change produced by a thermistor when subject to a temperature change can be converted into a voltage change. Figure 2.53 shows how a *potential divider circuit* can be used. A constant voltage, of perhaps 6 V, is applied across the thermistor and another resistor in series. With a thermistor with a resistance of 4.7 k Ω , the series resistor might be 10 k Ω . The output signal is the voltage across the 10 k Ω resistor. When the resistance of the thermistor changes, the fraction of the 6 V across the 10 k Ω resistor changes.

The output voltage is proportional to the fraction of the total resistance which is between the output terminals. Thus:

$$\text{Output} = \frac{R}{R + R_t} V$$

where V is the total voltage applied, in Figure 2.53 this is shown as 6 V, R the value of the resistance between the output terminals (10 k Ω), and R_t the resistance of the thermistor at the temperature concerned. The potential divider circuit is thus an example of a simple resistance to voltage converter.

Another example of such a converter is the *Wheatstone bridge*, Figure 2.54 shows the basic form of the bridge. The resistance element being monitored forms one of the arms of the bridge. When the output voltage V_o is zero, then there is no potential difference between B and D and so the potential at B must equal that at D. The potential difference across R_1 , i.e. V_{AB} , must then equal that across R_3 , i.e. V_{AD} . Thus:

$$I_1 R_1 = I_2 R_2$$

We also must have the potential difference across R_2 , i.e. V_{BC} , equal to that across R_4 , i.e. V_{DC} . Since there is no current through BD then the current through R_2 must be the same as that through R_1 and the current through R_4 the same as that through R_3 . Thus:

$$I_1 R_2 = I_2 R_4$$

Dividing these two equations gives:

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$

The bridge is said to be *balanced*.

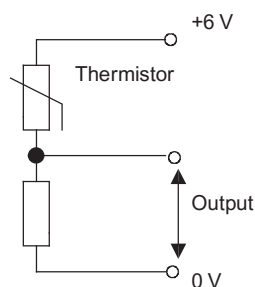


FIGURE 2.53 Resistance to voltage conversion for a thermistor.

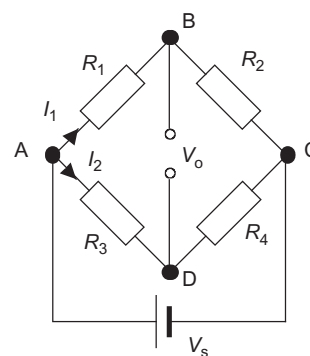


FIGURE 2.54 Wheatstone bridge.

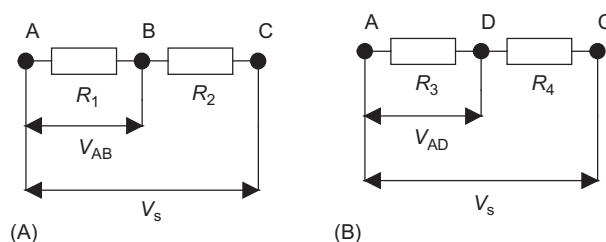


FIGURE 2.55 Potential drops across AB and AD.

Now consider what happens when one of the elements has a resistance which changes from this balanced condition. The supply voltage V_s is connected between points A and C and thus the potential drop across the resistor R_1 is the fraction $R_1/(R_1 + R_2)$ of the supply voltage (Figure 2.55A). Hence:

$$V_{AB} = \frac{V_s R_1}{R_1 + R_2}$$

Similarly, the potential difference across R_3 (Figure 2.55B) is:

$$V_{AD} = \frac{V_s R_3}{R_3 + R_4}$$

Thus the difference in potential between B and D, i.e. the output potential difference V_o , is:

$$V_o = V_{AB} - V_{AD} = V_s \left(\frac{R_1}{R_1 + R_2} - \frac{R_3}{R_3 + R_4} \right)$$

This equation gives the balanced condition when $V_o = 0$.

Consider resistance R_1 to be a sensor which has a resistance change, e.g. a strain gauge which has a resistance change when strained. A change in resistance from R_1 to $R_1 + \delta R_1$ gives a change in output from V_o to $V_o + \delta V_o$, where:

$$V_o + \delta V_o = V_s \left(\frac{R_1 + \delta R_1}{R_1 + \delta R_1 + R_2} - \frac{R_3}{R_3 + R_4} \right)$$

Hence:

$$(V_o + \delta V_o) - V_o = V_s \left(\frac{R_1 + \delta R_1}{R_1 + \delta R_1 + R_2} - \frac{R_1}{R_1 + R_2} \right)$$

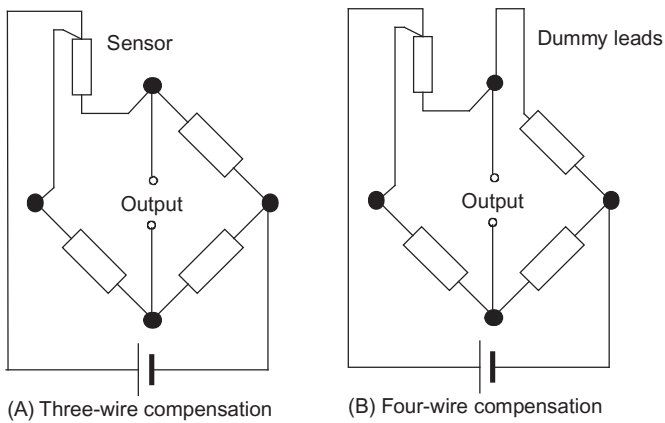


FIGURE 2.56 Temperature compensation.

If δR_1 is much smaller than R_1 then the denominator $R_1 + \delta R_1 + R_2$ approximates to $R_1 + R_2$ and so the above equation approximates to:

$$\delta V_o \approx V_s \left(\frac{\delta R_1}{R_1 + R_2} \right)$$

With this approximation, the change in output voltage is thus proportional to the change in the resistance of the sensor. We thus have a resistance to voltage converter. Note that the above equation only gives the output voltage when there is no load resistance across the output. If there is such a resistance then the loading effect has to be considered (see Section 1.3.1).

As the output voltage is proportional to the bridge excitation voltage, voltage drops along the cables from the voltage supply and the cable resistance between resistors in the bridge can affect the output, this being a particular problem if temperature changes cause resistance changes in these cables. Three-wire compensation (Figure 2.56A) can be used to help overcome the problem of cable resistance between, say, a temperature sensor and the bridge. Figure 2.56B shows a four-wire form of compensation, there being two parallel, dummy leads.

EXAMPLE

A platinum resistance coil is to be used as a temperature sensor and has a resistance at 0°C of $100\ \Omega$. It forms one arm of a Wheatstone bridge with the bridge being balanced at this temperature and each of the other arms also being $100\ \Omega$. If the temperature coefficient of resistance of platinum is $0.0039\ \text{K}^{-1}$, what will be the output voltage from the bridge per degree change in temperature if the supply voltage is $6.0\ \text{V}$?

The variation of the resistance of the platinum with temperature can be represented by:

$$R_t = R_0(1 + \alpha t)$$

where R_t is the resistance at $t^\circ\text{C}$, R_0 the resistance at 0°C and α the temperature coefficient of resistance. Hence:

$$\text{Change in resistance} = R_t - R_0 = R_0 \alpha t$$

Thus, for a one degree change in temperature:

$$\text{Change in resistance} = 100 \times 0.0039 \times 1 = 0.39\ \Omega$$

Since this resistance change is small compared to the $100\ \Omega$, the approximate equation for the output voltage can be used. Hence, the change in output per degree change in temperature is:

$$\delta V_o \approx V_s \left(\frac{\delta R_1}{R_1 + R_2} \right) = \frac{6.0 \times 0.39}{100 + 100} = 0.012\text{V}$$

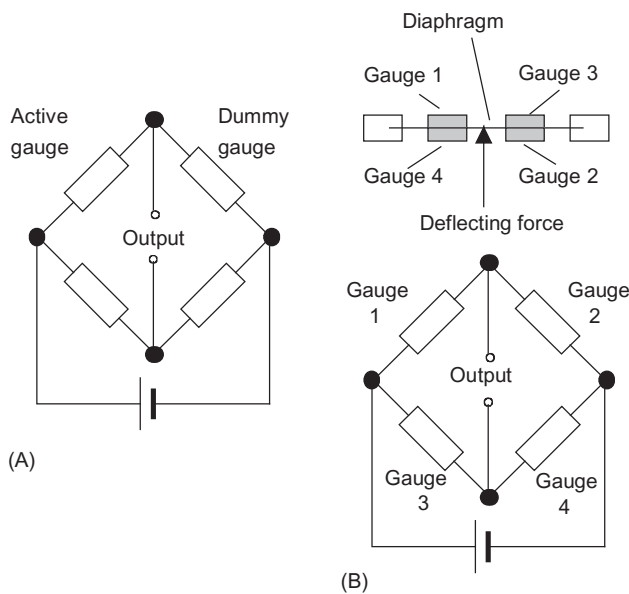


FIGURE 2.57 Temperature compensation: (A) dummy gauge, (B) with four active strain gauges.

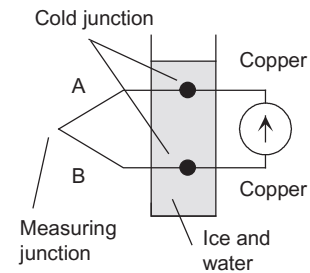


FIGURE 2.58 Cold junction at 0°C .

2.9.2 Temperature Compensation

The electrical resistance strain gauge is a resistance element which changes resistance when subject to strain. However, it will also change resistance when subject to a temperature change. Thus, in order to use it to determine strain, compensation has to be made for temperature effects. One way of eliminating the temperature effect is to use a *dummy strain gauge*. This is a strain gauge identical to the one under strain, the active gauge, which is mounted on the same material as the active gauge but not subject to the strain. It is positioned close to the active gauge so that it suffers the same temperature changes. As a result, a temperature change will cause both gauges to change resistance by the same amount. The active gauge is mounted in one arm of a Wheatstone bridge (Figure 2.57A) and the dummy gauge in an opposite arm so that the effects of temperature-induced resistance changes cancel out.

Strain gauges are often used with other sensors such as diaphragm pressure gauges or load cells. Temperature compensation is still required. While dummy gauges could be used, a better solution is to use four strain gauges with two of them attached so that the applied forces put them in tension and the other two in compression. The gauges that are in tension, gauges 1 and 3, will increase in resistance while those in compression, gauges 2 and 4, will decrease in resistance. The gauges are connected as the four arms of a Wheatstone bridge (Figure 2.57B). As all the gauges and so all the arms of the bridge will be equally affected by any temperature changes the arrangement is temperature compensated. The arrangement also gives a much greater output voltage than would occur with just a single active gauge.

2.9.3 Thermocouple Compensation

With a thermocouple, one junction should be kept at 0°C ; the temperature can then be obtained by looking up in tables the e.m.f. produced by the thermocouple. Figure 2.58 illustrates what is required.

This keeping of a junction at 0°C , i.e. in a mixture of ice and water, is not always feasible or very convenient and the cold junction is often allowed to be at the ambient temperature. To take account of this a compensation voltage has to be added to the thermocouple. This voltage is the same as the e.m.f. that would be generated by the thermocouple with one junction at 0°C and the other at the ambient temperature. We thus need a voltage which will depend on the ambient temperature. Such a voltage can be produced by using a resistance temperature sensor in a Wheatstone bridge. Figure 2.59 illustrates this.

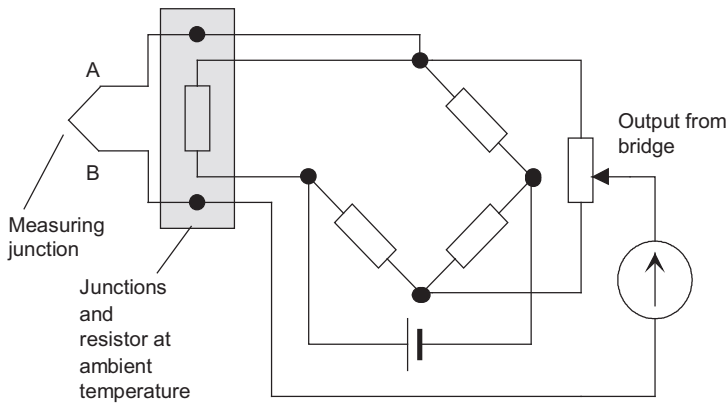


FIGURE 2.59 A Wheatstone bridge compensation circuit.

The bridge is balanced at 0°C and the output voltage from the bridge provides the correction potential difference at other temperatures. By a suitable choice of resistance temperature sensor, the appropriate voltage can be obtained.

The resistance of a metal resistance temperature sensor is given by:

$$R_t = R_0(1 + \alpha t)$$

where R_t is the resistance at $t^{\circ}\text{C}$, R_0 the resistance at 0°C and α the temperature coefficient of resistance. When there is a change in temperature:

$$\text{Change in resistance} = R_t - R_0 = R_0 \alpha t$$

The output voltage for the bridge, taking R_1 to be the resistance temperature sensor, is given by:

$$\delta V_o \approx V_s \left(\frac{\delta R_1}{R_1 + R_2} \right) = \frac{V_s R_0 \alpha t}{R_0 + R_2}$$

This voltage must be the same as that given by the thermocouple with one junction at 0°C and the other at the ambient temperature. The thermocouple e.m.f. e is likely to vary with temperature t in a reasonably linear manner over the small temperature range being considered, i.e. from 0°C to the ambient temperature. Thus we can write $e = at$, where a is a constant, i.e. the e.m.f. produced per degree change in temperature. Hence for compensation we must have:

$$at = \frac{V_s R_0 \alpha t}{R_0 + R_2}$$

and so the condition:

$$aR_2 = R_0(V_s \alpha - a)$$

EXAMPLE

Determine the value of the resistance R_2 in Figure 2.59 if compensation is to be provided for an iron – constantan thermocouple giving $51 \mu\text{V}/^{\circ}\text{C}$. The compensation is to be provided by a nickel resistance element with a resistance of 10Ω at 0°C and a temperature coefficient of resistance of 0.0067 K^{-1} . Take the supply voltage for the bridge to be 2.0 V .

Using the equation developed above, $aR_2 = R_0(V_s \alpha - a)$, then:

$$51 \times 10^{-6} \times R_2 = 10(2 \times 0.0067 - 51 \times 10^{-6})$$

Hence R_2 is 2617Ω .

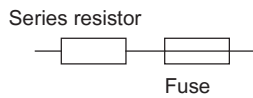


FIGURE 2.60 Protection against high currents.

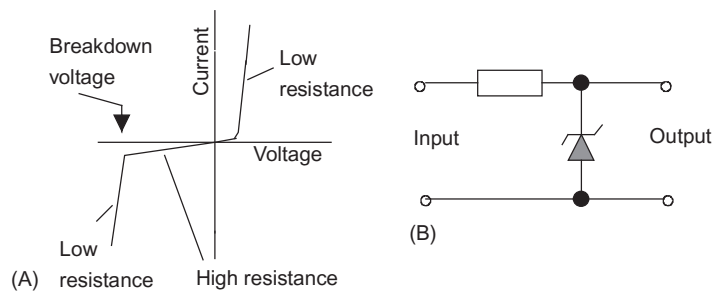


FIGURE 2.61 Zener diode: (A) current–voltage relationship, (B) protection circuit.

2.9.4 Protection

An important element that is often present with signal processing is *protection against high currents or high voltages*. For example, sensors when connected to a microprocessor can damage it if high currents or high voltages are transmitted to the microprocessor. A high current can be protected against by the incorporation in the input line of a series resistor to limit the current to an acceptable level and a fuse to break if the current does exceed a safe level (Figure 2.60).

Protection against high voltages and wrong polarity voltages may be obtained by the use of a *Zener diode circuit* (Figure 2.60). The Zener diode with a reverse voltage connected across it has a high resistance up to some particular voltage at which it suddenly breaks down and becomes conducting (Figure 2.61A). Zener diodes are given voltage ratings, the rating indicating at which voltage they become conducting. For example, to allow a maximum voltage of 5 V but stop voltages above 5.1 V being applied to the following circuit, a Zener diode with a voltage rating of 5.1 V might be chosen. For voltages below 5.1 V the Zener diode, in reverse voltage connection, has a high resistance. When the voltage rises to 5.1 V the Zener diode breaks down and its resistance drops to a very low value. Thus, with the circuit shown in Figure 2.61B, with the applied voltage below 5.1 V, the Zener diode, in reverse voltage connection, has a much higher resistance than the other resistor and so virtually all the applied voltage is across the Zener diode. When the applied voltage rises to 5.1 V, the Zener diode breaks down and has a low resistance. As a consequence, most of the voltage is then dropped across the resistor, the voltage across the diode drops and so the output voltage drops. Because the Zener diode is a diode with a low resistance for current in one direction through it and a high resistance for the opposite direction, it also provides protection against wrong polarity.

To ensure protection, it is often necessary to completely isolate circuits so that there are no electrical connections between them. This can be done using an *optoisolator*. Such a device converts an electrical signal into an optical signal, transmits it to a detector which then converts it back into an electrical signal (Figure 2.62). The input signal passes through an infrared light-emitting diode (LED) and so produces a beam of infrared radiation. This infrared signal is then detected by a phototransistor. To prevent the LED having the wrong polarity or too high an applied voltage, it is likely to be protected by the Zener diode circuit (of the type shown above in Figure 2.61). Also, if there is likely to be an alternating signal in the input a diode would be put in the input line to rectify it.

2.9.5 Analogue-to-Digital Conversions

The electrical output from sensors such as thermocouples, resistance elements used for temperature measurement, strain gauges, diaphragm pressure gauges, LVDTs, etc. is in analogue form. Microcontrollers require digital inputs. They can then perform mathematical operations such as adding and subtracting. Also digital signals are more readily transmitted and stored.

An *analogue signal* (Figure 2.63A) is one that is continuously variable, changing smoothly over a range of values. The signal is an analogue, i.e. a scaled version, of the quantity it represents. Thus, where a microprocessor is used it has to be converted into a digital form before it can be used as an input to the microprocessor. A *digital signal* increases in jumps, being a sequence of pulses, often just on–off signals (Figure 2.63B). The value of the quantity instead of being represented by the height of the signal, as with analogue, is represented by the sequence of on–off signals. The output from a microprocessor is digital. Most control elements require an

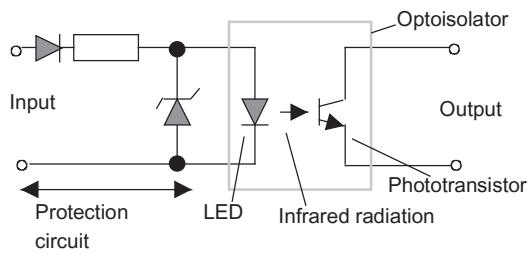


FIGURE 2.62 An optoisolator.

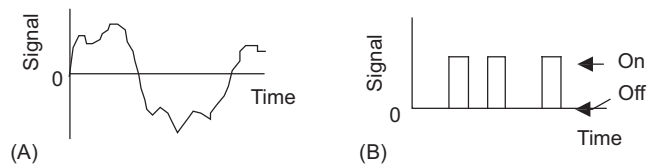


FIGURE 2.63 Signals: (A) analogue, (B) digital.

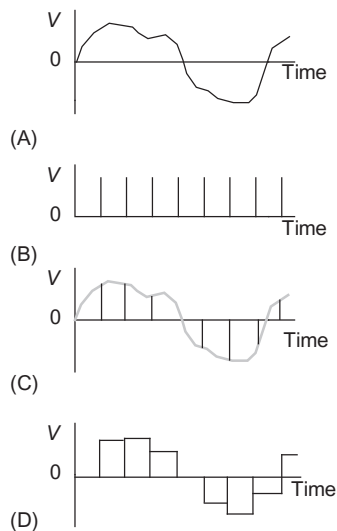


FIGURE 2.64 (A) Analogue signal, (B) time signal, (C) sampled signal, (D) sampled and held signal.

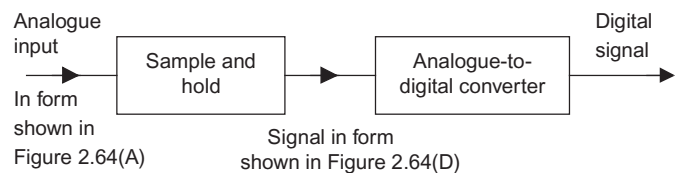


FIGURE 2.65 Analogue-to-digital conversion.

analogue input and so the digital output from a microprocessor has to be converted into an analogue form before it can be used by them. Thus there is a need for analogue-to-digital converters (ADC) and digital-to-analogue converters (DAC). Microcontrollers are microprocessors with input and output signal processing incorporated on the same chip and often incorporate analogue-to-digital and digital-to-analogue converters.

Analogue-to-digital conversion involves a number of stages. The first stage is to take samples of the analogue signal (Figure 2.64A). A clock supplies regular time signal pulses (Figure 2.64B) to the analogue-to-digital converter and every time it receives a pulse it samples the analogue signal. The result is a series of narrow pulses with heights which vary in accord with the variation of the analogue signal (Figure 2.64C). This sequence of pulses is changed into the signal form shown in Figure 2.64D by each sampled value being held until the next pulse occurs. It is necessary to hold a sample of the analogue signal so that conversion can take place to a digital signal at an analogue-to-digital converter. This converts each sample into a sequence of pulses representing the value. For example, the first sampled value might be represented by 101, the next sample by 011, etc. The 1 represents an 'on' or 'high' signal, the 0 an 'off' or 'low' signal. Analogue-to-digital conversion thus involves a sample and hold unit followed by an analogue-to-digital converter (Figure 2.65).

Because all ADCs sample the value of the input at discrete intervals of time, errors can arise if the input is sampled below a rate, termed the *Nyquist rate*. This is twice the highest frequency of interest. If lower frequencies are sampled then some signal cycles can be missed.

To illustrate the action of the analogue-to-digital converter, consider one that gives an output restricted to three bits. The binary digits of 0 and 1, i.e. the 'low' and 'high' signals, are referred to as *bits*. A group of bits is called a *word*. Thus the three bits give the *word length* for this particular analogue-to-digital converter. The word is what represents the digital version of the analogue voltage. The position of bits in a word has the significance that the least significant bit is on the right end of the word and the most significant bit on the left. This is just like

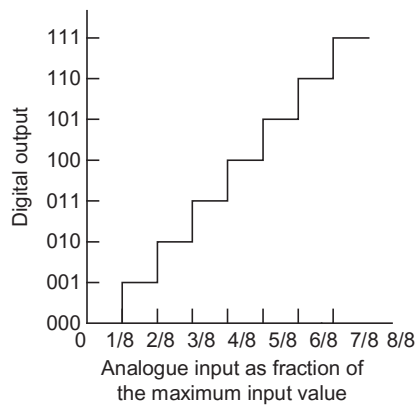


FIGURE 2.66 Digital output from an ADC.

counting in tens, 435 has the 5 as the least significant number and the 4 as the most significant number, the least significant number contributing least to the overall value of the 435 number. The position of the digit in a decimal number is significant; the least significant digit having its value multiplied by 10^0 , the next by 10^1 , the next by 10^2 , and so on. Likewise, the position of bits in a binary word is significant; the least significant bit having its value multiplied by 2^0 , the next by 2^1 , the next by 2^2 , and so on. For a binary word of n bits :

$$\begin{array}{cccc}
 2^{n-1}, & \dots, & 2^3, & 2^2, & 2^1, & 2^0 \\
 \uparrow & & & & \uparrow & \\
 \text{Most significant} & & & & \text{Least significant} & \\
 \text{bit (MSB)} & & & & \text{bit (LSB)} &
 \end{array}$$

With binary numbers we have the basic rules: $0 + 0 = 0$, $0 + 1 = 1$, $1 + 1 = 10$. Thus if we start with 000 and add 1 we obtain 001. If we add a further 1 we have 010. Adding another 1 gives 011. With three bits in a word we thus have the possible words of:

000 001 010 011 100 101 110 111

There are eight possible words which can be used to represent the analogue input; the number of possible words with a word length of n bits is 2^n . Thus we divide the maximum analogue voltage into eight parts and one of the digital words corresponds to each. Each rise in the analogue voltage of $(1/8)$ of the maximum analogue input then results in a further bit being generated. Thus for word 000 we have 0 V input. To generate the next digital word of 001 the input has to rise to $1/8$ of the maximum voltage. To generate the next word of 010 the input has to rise to $2/8$ of the maximum voltage. Figure 2.66 illustrates this conversion of the sampled and held input voltage to a digital output.

Thus if we had a sampled analogue input of 8 V, the digital output would be 000 for a 0 V input and would remain at that output until the analogue voltage had risen to 1 V, i.e. $1/8$ of the maximum analogue input. It would then remain at 001 until the analogue input had risen to 2 V. This value of 001 would continue until the analogue input had risen to 3 V. The smallest change in the analogue voltage that would result in a change in the digital output is thus 1 V. This is termed the *resolution* of the converter.

The word length possible with an analogue-to-digital converter determines its *resolution*. With a word length of n bits the maximum, or full scale, analogue input V_{FS} is divided into 2^n pieces. The minimum change in input that can be detected, i.e. the *resolution*, is thus $V_{FS} / 2^n$. With an analogue-to-digital converter having a word length of 10 bits and the maximum analogue signal input range 10 V, then the maximum analogue voltage is divided into $2^{10} = 1024$ pieces and the resolution is $10/1024 = 9.8$ mV. All ADCs suffer from non-linearity errors, these arising from physical imperfections and causing their output to deviate from a linear function of their input. Typically this might be $\pm 1/2$ the least significant bit.

There are a number of forms of analogue-to-digital converter; the most commonly used being successive approximations, ramp, and flash. The *flash form* is much faster than either the successive approximations form or the ramp form. The term *conversion time* or *settling time* is used to specify the time it takes a converter to generate

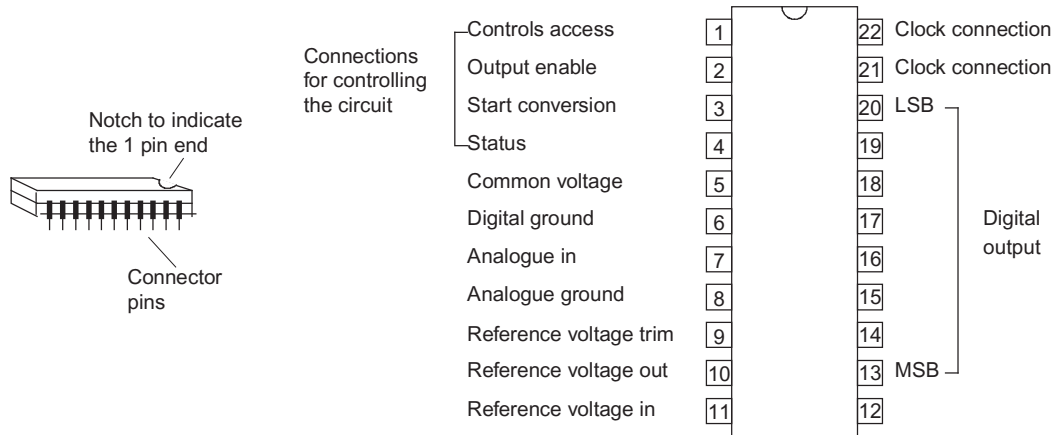


FIGURE 2.67 The GEC Plessey ZN439E 8-bit analogue-to-digital converter. Its specification includes: Resolution 8 bits, Conversion time $5 \mu\text{s}$, Linearity error $\pm \frac{1}{2}$ LSB, Linearity temperature coefficient typically $\pm 3 \text{ ppm}/^\circ\text{C}$, Power dissipation 150 mW Operating temperature -55 to $+125^\circ\text{C}$, Supply voltage typically 5.0 V, current typically 30 mA.

a complete digital word when supplied with the analogue input. Typically successive approximations converters might have conversion times of 1 to $10 \mu\text{s}$ for an 8-bit conversion. Analogue-to-digital converters are generally purchased as integrated circuits. Figure 2.67 shows an example of the pin connections for the ZN439E, a successive approximation form of ADC.

EXAMPLE

A thermocouple gives an output of 0.4 mV for each degree change in temperature. What will be the word length required when its output passes through an analogue-to-digital converter if temperatures from 0 to 200°C are to be measured with a resolution of 0.5°C ?

The full-scale-output from the sensor is $200 \times 0.4 = 80 \text{ mV}$. With a word length n there are 2^n digital numbers. Thus this voltage will be divided into 2^n levels and so the minimum voltage change that can be detected is $80/2^n \text{ mV}$. For a resolution of 0.5°C we must be able to detect a signal from the sensor of $0.5 \times 0.4 = 0.20 \text{ mV}$. Hence:

$$0.20 = \frac{80}{2^n}$$

and so $2^n = 400$ and $n = 8.6$. Thus a 9-bit word length is required.

2.9.6 Digital-to-Analogue Conversions

The input to a digital-to-analogue converter is a binary word and the output its equivalent analogue value. For example, if we have a full-scale-output of 7 V then a digital input of 000 will give 0 V, 001 give 1 V, ... and 111 the full scale value of 7 V. Figure 2.68 illustrates this.

The basic form of a *digital-to-analogue converter* involves the digital input being used to activate electronic switches such that a 1 activates a switch and a 0 does not, the position of the 1 in the word determining which switch is activated. Thus when, with say a 3-bit converter, 001 is received we have a voltage of, say, 1 V switched to the output, when 010 is received we have 2 V, switched to the output, and when 100 is received we have 4 V switched to the output. Hence if we have the digital word 011 we have the least significant bit 001 switching 1 V to the output and the 010 bit 2 V to the output to give a summed output of 3 V (Figure 2.69).

Figure 2.70 shows details of the GEC Plessey ZN558D 8-bit latched input digital-to-analogue converter. After the conversion is complete, the 8-bit result is placed in an internal latch until the next conversion is complete. A latch is just a device to retain the output until a new one replaces it. The settling time is the time taken for the

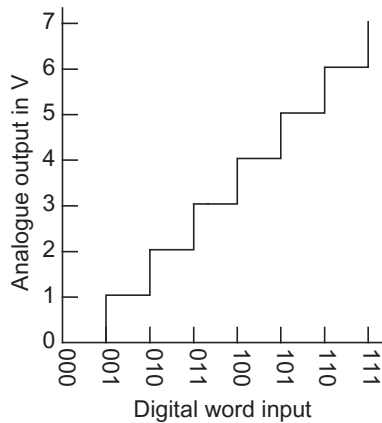


FIGURE 2.68 Digital-to-analogue conversion.

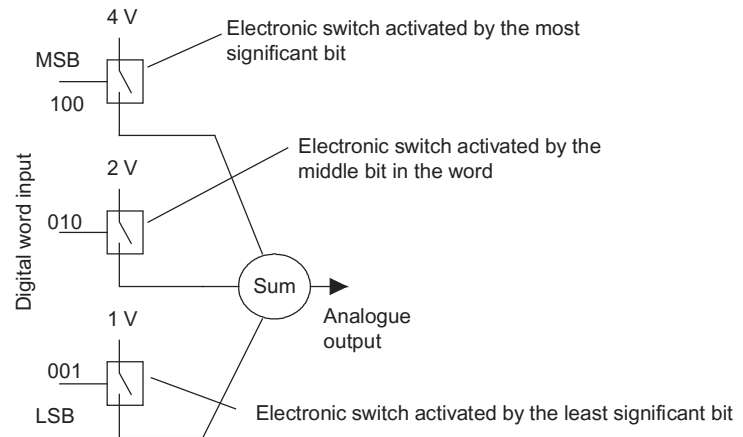


FIGURE 2.69 The principle of a 3-bit digital-to-analogue converter.

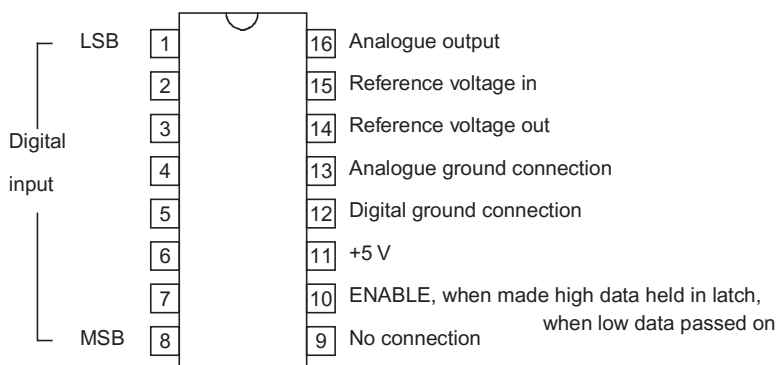


FIGURE 2.70 The GEC Plessey ZN558D 8-bit latched digital-to-analogue converter. Its specification includes: Resolution 8 bits, Settling time 800 ns, Non-linearity 0.5% of full scale, Power dissipation 100 mW.

EXAMPLE

A microprocessor gives an output of an 8-bit word. This is fed through an 8-bit digital-to-analogue converter to a control valve which requires 6.0 V to be fully open. If the fully open state is to be indicated by the output of the digital word 11111111 what will be the change in output to the valve when there is a change of 1 bit?

The output voltage will be divided into 2^8 intervals. Since there is to be an output of 6.0 V when the output is 2^8 of these intervals, a change of 1 bit is a change in the output voltage of $6.0/2^8 = 0.023$ V.

analogue output voltage to settle within a specified band, usually $\pm \text{LSB}/2$, about its final value when the digital word suddenly changes.

2.9.7 Microcontroller Systems

Many control systems involve sensors supplying inputs to microcontrollers and it is their outputs which might then be used to exercise control functions. Microcontrollers contain a microprocessor, memory, and programmable inputs and outputs on a single integrated circuit. Figure 2.71 shows the basic elements of a microcontroller. It contains a microprocessor, volatile memory (RAM) for data storage, non-volatile memory (ROM) for storing program and operating parameters, a clock generator, and input/output interfaces. They have input/output ports for connection to external inputs and outputs and digital or analogue signals, having such items as ADCs and DACs built in. They can include channels for serial as well as parallel signals, some possibly having specific bus interfaces such as CAN which is used with automobiles.

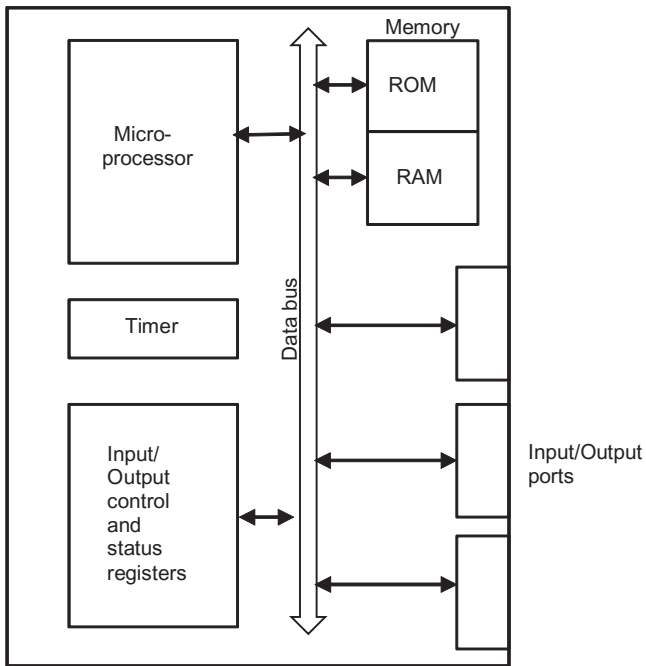


FIGURE 2.71 The basic elements of a microcontroller.

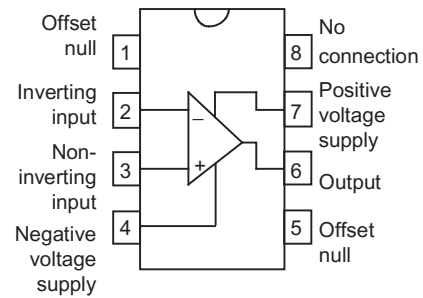


FIGURE 2.72 Pin connections for a 741 op-amp.

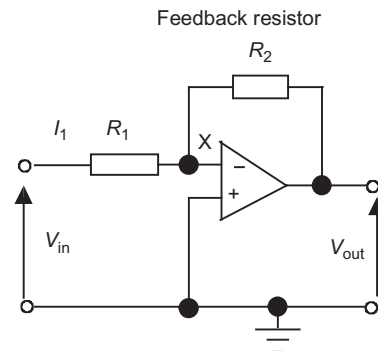


FIGURE 2.73 Inverting amplifier.

2.9.8 Op-Amps

The *operational amplifier* (op-amp) is a very high gain d.c. amplifier, the gain typically being of the order of 100 000 or more, which is supplied as an integrated circuit on a silicon chip. It has two inputs, known as the inverting input (–) and the non-inverting input (+). In addition there are inputs for a negative voltage supply, a positive voltage supply and two inputs termed offset null, these being to enable corrections to be made for the non-ideal behaviour of the amplifier. Figure 2.72 shows the pin connections for a 741 type operational amplifier with the symbol for the operational amplifier shown superimposed. On the symbol the + sign indicates the non-inverting input and the – sign the inverting input.

The operational amplifier is a very widely used element in signal conditioning and processing circuits and the following indicates common examples of such circuits.

Figure 2.73 shows the connections made to the amplifier when it is used as an *inverting amplifier*, such a form of amplifier giving an output which is an inverted form of the input, i.e. it is out of phase by 180° . The input is taken to the inverting input through a resistor R_1 with the non-inverting input being connected to ground. A feedback path is provided from the output, via the resistor R_2 to the inverting input. The operational amplifier has a very high voltage gain of about 100 000 and the change in output voltage is limited to about ± 10 V. Thus the input voltage at point X must be between $+0.0001$ V and -0.0001 V. This is virtually zero and so point X is at virtually earth potential and hence is termed a *virtual earth*. With an ideal operational amplifier with an infinite gain, point X is at zero potential. The potential difference across R_1 is $(V_{in} - V_X)$. Hence, for an ideal operational amplifier with an infinite gain, and hence $V_X = 0$, the input potential V_{in} can be considered to be across R_1 . Thus:

$$V_{in} = I_1 R_1$$

The operational amplifier has a very high impedance between its input terminals, for a 741 this is about $2\text{ M}\Omega$. Thus virtually no current flows through X into it. For an ideal operational amplifier the input impedance is taken to be infinite and so there is no current flow through X into the amplifier input. Hence, since the current entering the junction at X must equal the current leaving it, the current I_1 through R_1 must be the current through R_2 .

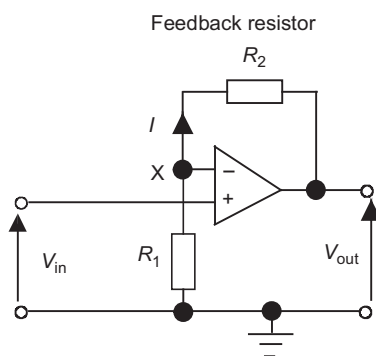


FIGURE 2.74 Non-inverting amplifier.

The potential difference across R_2 is $(V_X - V_{out})$ and thus, since V_X is zero for the ideal amplifier, the potential difference across R_2 is $-V_{out}$. Thus:

$$-V_{out} = I_1 R_2$$

Dividing these two equations gives the ratio of the output voltage to the input voltage, i.e. the voltage gain of the circuit. Thus:

$$\text{Voltage gain of circuit} = \frac{V_{out}}{V_{in}} = -\frac{R_2}{R_1}$$

The voltage gain of the circuit is determined solely by the relative values of R_2 and R_1 . The negative sign indicates that the output is inverted, i.e. 180° out of phase, with respect to the input.

To illustrate the above, consider an inverting operational amplifier circuit which has a resistance of $10 \text{ k}\Omega$ in the inverting input line and a feedback resistance of $100 \text{ k}\Omega$. The voltage gain of the circuit is:

$$\text{Voltage gain of circuit} = \frac{V_{out}}{V_{in}} = -\frac{R_2}{R_1} = -\frac{100}{10} = -10$$

Figure 2.74 shows the operational amplifier connected as a *non-inverting amplifier*. Since the operational amplifier has a very high input impedance, there is virtually no current flowing into the inverting input. The inverting voltage input voltage is V_{in} . Since there is virtually no current through the operational amplifier between the two inputs there can be virtually no potential difference between them. Thus, with the ideal operational amplifier, we must have $V_X = V_{in}$. The output voltage is generated by the current I which flows from earth through R_1 and R_2 . Thus:

$$I = -\frac{V_{out}}{R_1 + R_2}$$

But X is at the potential V_{in} , thus the potential difference across the feedback resistor R_2 is $(V_{in} - V_{out})$ and we must have:

$$I = \frac{V_{in} - V_{out}}{R_2}$$

Thus:

$$-\frac{V_{out}}{R_1 + R_2} = \frac{V_{in} - V_{out}}{R_2}$$

Hence:

$$\text{Voltage gain of circuit} = \frac{V_{out}}{V_{in}} = \frac{R_1 + R_2}{R_1} = 1 + \frac{R_2}{R_1}$$

A particular form of this amplifier which is often used has the feedback loop as a short circuit, i.e. $R_2 = 0$. The voltage gain is then just 1. However, the input voltage to the circuit is across a large resistance, the input resistance of an operational amplifier such as the 741 typically being $2 \text{ M}\Omega$. The resistance between the output

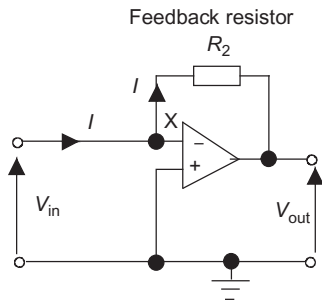


FIGURE 2.75 Current-to-voltage.

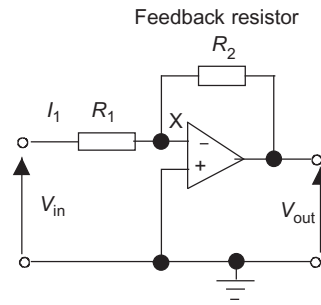


FIGURE 2.76 Voltage-to-current converter.

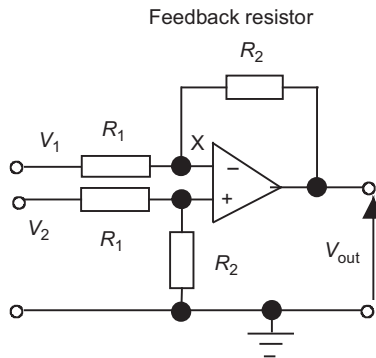


FIGURE 2.77 Differential amplifier.

terminal and the ground line is, however, much smaller, typically 75Ω . Thus the resistance in the circuit that follows is a relatively small one compared with the resistance in the input circuit and so affects that circuit less. Such a form of amplifier circuit is referred to as a *voltage follower* and is typically used for sensors which require high impedance inputs such as piezoelectric sensors.

Figure 2.75 shows how the standard inverting amplifier can be used as a *current-to-voltage converter*. Point X is the virtual earth. Thus any input current has to flow through the feedback resistor R_2 . The voltage drop across R_2 must therefore produce the output voltage and so $V_{out} = -IR_2$. So the output voltage is just the input current multiplied by the scaling factor R_2 . The advantage of this method of converting a current to a voltage, compared with just passing a current through a resistor and taking the potential difference across it, is that there is a high impedance across the input and so there is less likelihood of loading problems.

Situations often arise where the output needs to be a current in order to drive perhaps an electromechanical device such as a relay or possibly give a display on a moving coil meter. A *voltage-to-current converter* is provided by the basic inverting amplifier circuit with the device through which the current is required being the feedback resistor (Figure 2.76). Since X is a virtual earth, the potential difference across R_1 is V_{in} and the current through it I_1 . Hence $I_1 = V_{in}/R_1$. The current through R_2 is I_1 . Thus the input voltage has been converted to the current I_1 through the feedback resistor, with the current being V_{in}/R_1 .

Figure 2.77 shows how an operational amplifier can be used as a *differential amplifier*, amplifying the difference between two input signals. Since the operational amplifier has high impedance between its two inputs, there will be virtually no current through the operational amplifier between the two input terminals. There is thus no potential difference between the two inputs and therefore both will be at the same potential, that at X. The voltage V_2 is across resistors R_1 and R_2 in series. We thus have a potential divider circuit with the potential at the non-inverting input, which must be the same as that at X of V_X , as:

$$\frac{V_X}{V_2} = \frac{R_2}{R_1 + R_2}$$

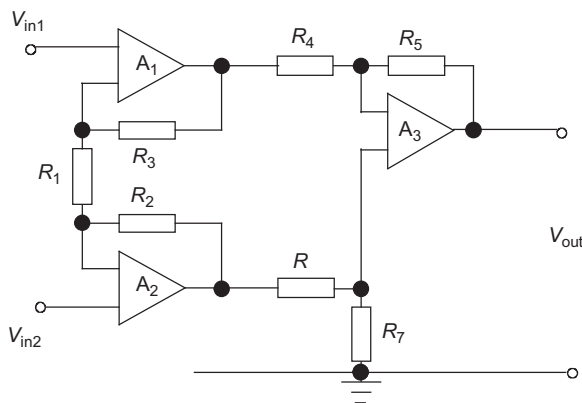


FIGURE 2.78 Three op-amp form of an instrumentation amplifier.

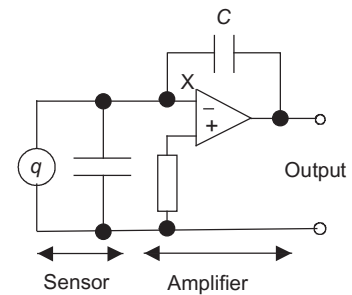


FIGURE 2.79 Charge amplifier.

The current through the feedback resistance must be equal to that from V_1 through R_1 . Hence:

$$\frac{V_1 - V_X}{R_1} = \frac{V_X - V_{\text{out}}}{R_2}$$

This can be rearranged to give:

$$\frac{V_{\text{out}}}{R_2} = V_X \left(\frac{1}{R_2} + \frac{1}{R_1} \right) - \frac{V_1}{R_1}$$

Hence, substituting for V_X using the earlier equation:

$$V_{\text{out}} = \frac{R_2}{R_1} (V_2 - V_1)$$

The output is thus proportional to the difference between the two input voltages. Such a circuit might be used with a thermocouple to amplify the difference in e.m.f.s between the hot and cold junctions. Suppose we require there to be an output of $1 \text{ mV}/^\circ\text{C}$. With an iron – constantan thermocouple with the cold junction at 0°C , the e.m.f. produced between the hot and cold junctions is about $53 \mu\text{V}/^\circ\text{C}$. Thus, for a 1°C temperature difference between the junctions, the above equation gives:

$$1 \times 10^{-3} = \frac{R_2}{R_1} \times 53 \times 10^{-6}$$

Hence we must have $R_2/R_1 = 18.9$. Thus if we take for R_1 a resistance of $10 \text{ k}\Omega$ then R_2 must be $189 \text{ k}\Omega$.

The differential amplifier is the simplest form of what is often termed an *instrumentation amplifier*. A more usual form involves three operational amplifiers (Figure 2.78). Such a circuit is available as a single integrated circuit. The first stage involves the amplifiers A_1 and A_2 . These amplify the two input signals without any increase in the common mode voltage before amplifier A_3 is used to amplify the differential signal. The differential amplification produced by A_1 and A_2 is $(R_1 + R_2 + R_3)/R_1$ and that produced by A_3 is R_5/R_4 and so the overall amplification is the product of these two amplifications. The overall gain is usually set by varying the value of R_1 . Normally the circuit has $R_2 = R_3$, $R_4 = R_6$ and $R_5 = R_7$.

A *charge amplifier* provides an output voltage which is proportional to the charge stored on a device connected to its input terminals and is widely used with sensors employing piezoelectric crystals. Basically, a charge amplifier can be considered to be an op-amp with a capacitor in the feedback path (Figure 2.79). The potential difference across the capacitor is $(v_X - v_{\text{out}})$ and since v_X is effectively zero, being the virtual earth, it is $-v_{\text{out}}$. The charge on this capacitor is Cv_{out} and thus the current through it is the rate of movement of charge and so $-Cd v_{\text{out}}/dt$. But this is the same current as supplied by the sensor and this is dq/dt , where q is the charge on sensor. Thus, $-Cd v_{\text{out}}/dt = dq/dt$ and the output voltage is $-(1/C)$ times the charge on the sensor.

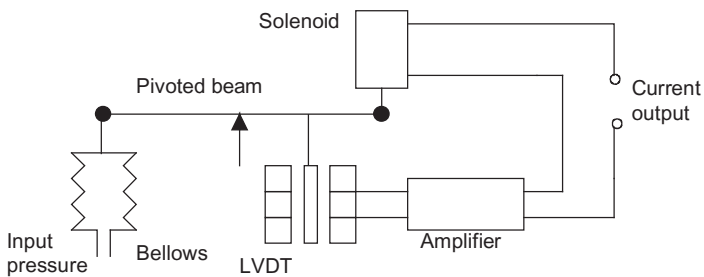


FIGURE 2.80 Pressure-to-current converter.

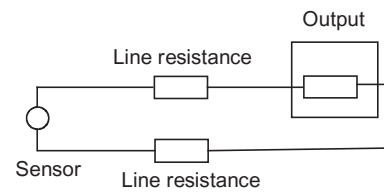


FIGURE 2.81 Analogue voltage transmission.

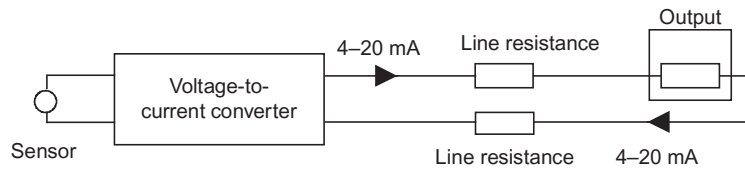


FIGURE 2.82 Current loop transmission.

2.9.9 Pressure-to-Current Converter

Control systems generally are electrical and so there is often a need to convert a pressure into an electrical current. Figure 2.80 shows the basic principle of such a converter. The input pressure causes the bellows to extend and so apply a force to displace the end of the pivoted beam. The movement of the beam results in the core being moved in a linear variable differential transformer (LVDT). This gives an electrical output which is amplified. The resulting current is then passed through a solenoid. The current through the solenoid produces a magnetic field which is used to attract the end of the pivoted beam to bring the beam back to its initial horizontal position. When the beam is back in the position, the solenoid current maintaining it in this position is taken as a measure of the pressure input.

2.10 SIGNAL TRANSMISSION

Measurement signals often have to be transmitted over quite large distances from the place of measurement to a display unit and/or a process control unit. Methods used for such transmission are:

1. Analogue voltage transmission

Analogue voltage signals can suffer corruption due to induced noise and the resistance of the connecting cables can result in attenuation of the voltage, the voltage drop across the output being reduced by that across the line resistance (Figure 2.81). Such effects can be reduced by the use of signal amplification and shielding with the connecting cables. However, because of these problems, such signals are not generally used for large distance transmission.

2. Current loop transmission

The attenuation which occurs with voltage transmission can be minimised if signals are transmitted as varying current signals. This form of transmission is known as current loop transmission and uses currents in the range 4 mA to 20 mA to represent the levels of the analogue signal. The level of 4 mA, rather than 0 mA, is used to indicate the zero signal level since, otherwise, it would not be possible to distinguish between a zero value signal and a break in the transmission line. Figure 2.82 shows the basic arrangement with the signal from the sensor being converted to a current signal by a voltage-to-current converter, e.g. that shown in Figure 2.76, transmitted and then converted to a voltage signal at the display.

TABLE 2.2 ASCII Code

	ASCII		ASCII		ASCII
A	100 0001	N	100 1110	0	011 0000
B	100 0010	O	100 1111	1	011 0001
C	100 0011	P	101 0000	2	011 0010
D	100 0100	Q	101 0001	3	011 0011
E	100 0101	R	101 0010	4	011 0100
F	100 0110	S	101 0011	5	011 0101
G	100 0111	T	101 0100	6	011 0110
H	100 1000	U	101 0101	7	011 0111
I	100 1001	V	101 0110	8	011 1000
J	100 1010	W	101 0111	9	011 1001
K	100 1011	X	101 1000		
L	100 1100	Y	101 1001		
M	100 1101	Z	101 1010		

3. Digital voltage signals

Digital signals can be transmitted over transmission lines using either serial or parallel communication. With serial communication, the sequence of bits used to describe a value is sent in sequence along a single transmission line. With parallel transmission, each of the bits is sent along a separate parallel transmission line. For long-distance communication, serial communication is used.

In order to transfer data, both the sender and receiver have to agree on the meaning of the transmitted binary digital patterns. The most commonly used character set is the American Standard Code for Information Interchange (ASCII), thus using 7 bits to represent each character (Table 2.2). The format used for sending such data has to be standardised. For example, with the RS-232 form of serial transmission, a sequence of 10 bits is used, the first bit being a start of message signal, then seven bits for the data, then a parity bit to identify whether errors have occurred in the transmission, and finally a stop bit to indicate the end of the message.

The RS-232 is a widely used serial data interface standard in which the electrical signal characteristics, such as voltage levels, the forms of plug and sockets for interconnections, and the interchange circuits are all specified. Because RS-232 is limited to distances less than about 15 m, another standard such as RS-485 tends to be used in many control systems with distances up to about 1200 m being possible.

Digital signal transmission has a great advantage when compared with analogue transmission in that signal corruption effects can be considerably reduced. With digital transmission, error coding is used to detect whether corruption has occurred. These are bits added to the sequence of bits used to represent the value and are check values which are not likely to tally with the received bits if corruption has occurred, the receiver can then request the message be sent again. For example, the sequence 1010 might be transmitted and corruption result in 1110 being received. In order to detect such errors one form of check uses a parity bit which is added at transmission. With even-parity, the bit is chosen so that the total number of 1s in the transmission, including the parity bit, is an even number. This 1010 would be transmitted as 10100. If it is corrupted and received as 11100 then the parity bit shows there is an error.

4. Pneumatic transmission

Pneumatic transmission involves converting the sensor output to a pneumatic pressure in that range 20 to 100 kPa or 20 to 180 kPa. The lower limit gives the zero sensor signal and enables the zero value to be distinguished from a break in the circuit. Such pressure signals can then be transmitted through plastic or metal piping, the distances being limited to about 300 m because of the limitations of speed of response at larger distances.

5. Fibre-optic transmission

An optical fibre is a light conductor in the form of a long fibre along which light can be transmitted by internally being reflected off the sides of the fibre (Figure 2.83). The light sources used are LEDs or



FIGURE 2.83 Optical fibres.

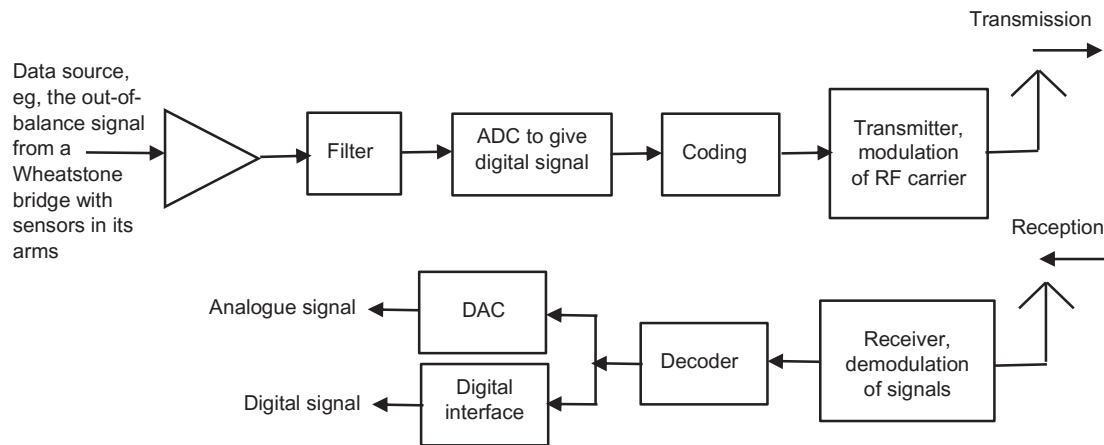


FIGURE 2.84 Digital telemetry, the basic elements for one data channel.

semiconductor laser diodes. Digital electrical signals are converted into light pulses which travel down the fibre before being detected by a photodiode or phototransistor and converted back into an electrical signal. Fibre optics has the advantages that they are immune to electromagnetic interference, data can be transmitted with much lower losses than with electrical cables, the fibres are smaller and less heavy than copper cables, and are more inert in hazardous areas.

6. Radio telemetry

Radio transmission for getting signals from measurement systems can be used when there are problems in installing electrical or fibre optics links or the transmission distances involved are large or the sensors are mounted on rotating shafts and direct wire connections between the sensors and the data display unit present problems. The transmission of analogue data usually involves frequency modulation (FM), each analogue sensor voltage modulating the carrier frequency into a varying frequency around the carrier frequency. Frequency modulation has, however, now given way to digital telemetry. With this, the output of the sensors is, after filters are used to smooth out any 'jaggies' that are present in what should be a smooth input, then converted into digital data. Multiple signals are often combined into one composite digital data stream for transmission. To guarantee a digital transmission is error free, blocks of code can be checked by also sending a series of bits representing their binary sum. If the transmission becomes corrupted the sum is not likely to be valid. This digital stream then modulates the radio frequency carrier which is then transmitted from the transmitter aerial to be picked up by the receiver aerial where it is then decoded to give the signals relevant to the data stream transmitted. In its simplest form amplitude shift keying (ASK) is used for the modulation of digital signals when the amplitude is modulated with the term on-off keying (OOK) for the simplest form of amplitude-shift keying modulation that represents digital data as the presence or absence of a carrier wave. The term frequency shift keying (FSK) is used when the frequency is modulated and phase shift modulation (PSK) when the phase is modulated. Figure 2.84 illustrates the radio telemetry transmission process and Figure 2.85 a simplistic sketch of the modulation methods. As an illustration of what is commercially available, there are radio telemetry instruments with frequency modulation of the radio frequency carrier and a transmission range ranging from 5 to over 50 m and which interface with strain gauge bridges and thermocouples. Such instruments are used with automobiles with sensors which can be attached to a wheel and transmit data when the wheel is rotating to the control unit in the auto.

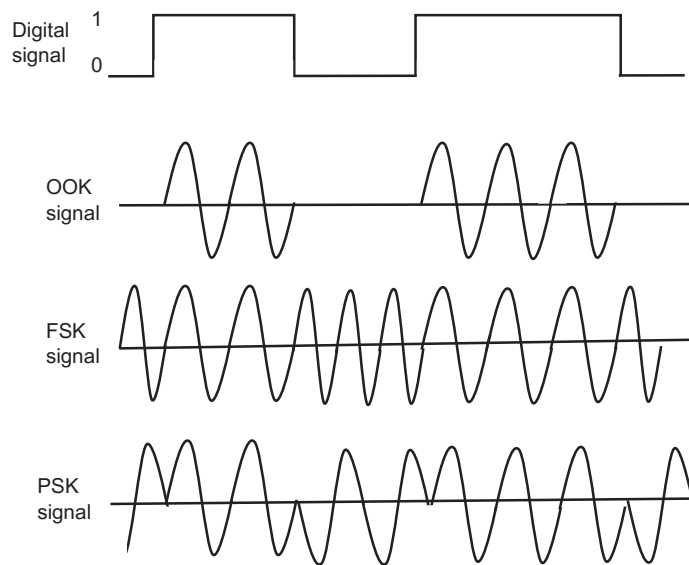


FIGURE 2.85 The basic elements of modulation methods.

2.10.1 Noise

The term *noise* is used, in this context, for the unwanted signals that may be picked up by a measurement system and interfere with the signals being measured. There are two basic types of electrical noise:

1. *Interference*

This is due to the interaction between external electrical and magnetic fields and the measurement system circuits, e.g. the circuit picking up interference from nearby mains power circuits.

2. *Random noise*

This is due to the random motion of electrons and other charge carriers in components and is determined by the basic physical properties of components in the system.

The three main types of interference are:

1. *Inductive coupling*

A changing current in a nearby circuit produces a changing magnetic field which can induce e.m.f.s, as a result of electromagnetic induction, in conductors in the measurement system.

2. *Capacitive coupling*

Nearby power cables, the earth, and conductors in the measurement system are separated from each other by a dielectric, air. There can thus be capacitance between the power cable and conductors, and between the conductors and earth. These capacitors couple the measurement system conductors to the other systems and thus signals in the other systems affecting the charges on these capacitors can result in interference in the measurement system.

3. *Multiple earths*

If the measurement system has more than one connection to earth, there may be problems since there may be some difference in potential between the earth points. If this occurs, an interference current may arise in the measurement system.

Methods of reducing interference are:

1. *Twisted pairs of wires*

This involves the elements of the measurement system being connected by twisted wire pairs (Figure 2.86). A changing magnetic field will induce e.m.f.s in each loop, but because of the twisting the directions of the e.m.f.s will be in one direction for one loop and in the opposite direction for the next loop and so cancel out.

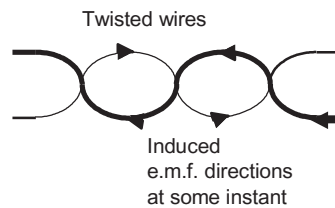


FIGURE 2.86 Twisted pairs.

2. Electrostatic screening

Capacitive coupling can be avoided by completely enclosing the system in an earthed metal screen. Problems may occur if there are multiple earths. Coaxial cable gives screening of connections between elements, however, the cable should only be earthed at one end if multiple earths are to be avoided.

3. Single earth

Multiple earthing problems can be avoided if there is only a single earthing point.

4. Differential amplifiers

A differential amplifier can be used to amplify the difference between two signals. Thus if both signals contain the same interference, then the output from the amplifier will not have amplified any interference signals.

5. Filters

A filter can be selected which transmits the measurement signal but rejects interference signals.

2.11 SMART SYSTEMS

It is possible to have a measurement system where the sensor and signal processing such as amplification and analogue-to-digital conversion are carried out with separate components. However, these are often available combined in a single integrated sensor circuit. Often the output from such a system needs further data processing and the resulting combination of sensor, signal processing and a microprocessor to give 'intelligent' processing of sensor inputs results in what is termed a *smart* or *intelligent sensor*. Such a microprocessor-equipped sensor can have the functions to give compensation for random errors, automatic calculation of measurement accuracy, automatic self-calibration, adjustment for non-linearities to give a linear output and self-diagnosis of faults. Smart sensors have the ability to 'talk', to 'listen', and to interact with data. Figure 2.87 is a block diagram representation of a smart sensor. A smart sensor will require power to operate its components and so a rechargeable battery is likely to be included in the sensor package. The output channel may include the facility to offer multi-channel wireless communication, so enabling outputs does not require wiring to communicate to an external data collection point. An example of such an intelligent sensor is one to detect hazardous fuel leak detection, this just needing to be placed where leaks may occur and not require any external wiring. The term 'lick-and-stick' has been used to describe such sensors as they can be just placed in the appropriate place like applying a postage stamp to an envelope.

Smart sensors have their own standard, IEEE 1451.4, so that smart sensors to this standard can be used in a 'plug-and-play' manner in control systems, holding and communicating data in a standard defined manner. With this standard, a sensor is considered as divided into two parts: a smart transducer interface module (STIM) containing the sensing element, appropriate signal conditioning and an analogue to digital converter (ADC), plus a TEDS (transducer electronic datasheet) in which information is stored, generally in a digital read-only memory chip embedded in the sensor system or as a file downloadable from the internet, and this identifies the device, its interface requirements, and calibration data so that the signal from an analogue sensor can be correctly used. Sensors defined in this way eradicate a common source of error – incorrectly transcribed calibration data from sensor data sheets. TEDS also describes the sensor's electrical interface requirements so that data acquisition hardware is correctly configured to properly read the sensor. Such sensors can perform self-diagnosis by monitoring internal signals for evidence of faults.

As an illustration of smart sensors that are available, OMEGA has smart infrared temperature sensors (OS35) which are centred round a microprocessor that continuously monitors and adjusts the system. It not only determines the output but calibrates itself once every second and so eliminates the problems of drift. The OS35-3 has an accuracy of $\pm 1^\circ\text{C}$.

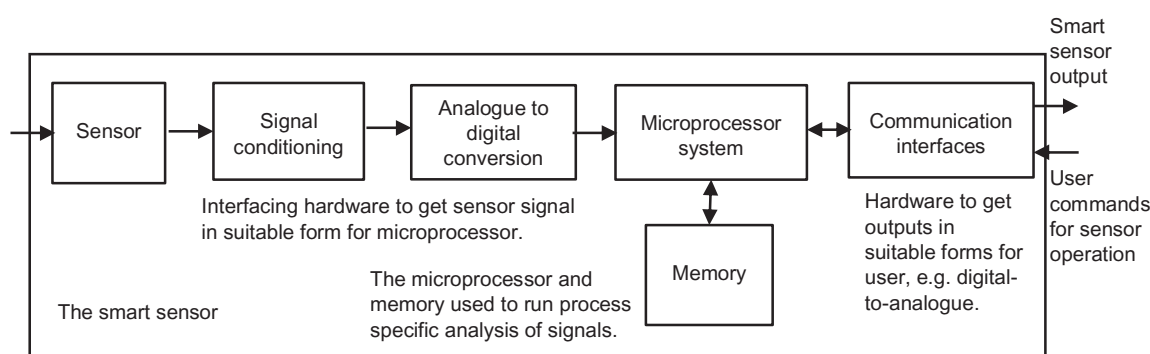


FIGURE 2.87 Smart sensor.

An example of the use of smart sensors in a production system is in smart valve positioners (see Section 6.4.4). A valve positioner is a device which adjusts a control valve actuator's position based on the control signal supplied to it in order to adjust the airflow through the valve and so exercise control in a pneumatic system. The valve positioner is mounted in the valve so that it is connected mechanically to the valve stem, movement of which controls the airflow in the pneumatic system. The sensor then enables the position of the valve stem to be monitored and compared with the position requested by the controller. When the control differs from the valve actuator's position, the valve positioner sends a feedback signal to move the actuator until the correct position is attained. A smart valve positioner is a digital valve controller using a microprocessor to exercise control by its current output being converted to a pressure signal for the pneumatic system. Such a smart positioner has advantages over non-smart positioners of automatic calibration and configuration, can carry out on-line valve diagnostics and so be aware of how a valve is functioning and improved process control accuracy.

Another example of a smart sensor is one that can be attached to the outer casing of a motor and configured to pick up data on temperature, vibration and other parameters related to the running of the motor. This information can be sent to a computer, or even a smartphone, to enable the running of the motor to be monitored and maintenance planned to prevent unscheduled motor downtime.

Because smart sensors include sensing and their own signal conditioning, analogue-to-digital conversion, bus interfacing and data processing, a designer is able to use a single electronic system to support many such sensing elements. In a process plant there are likely to be a large number of smart sensors, each providing information which has to be fed back to a control panel. To avoid using separate cables for each sensor to transmit their data, a *bus* system can be used. A *bus* is a common highway for signals which is used to link components. Thus, each sensor would put its information onto the common highway for transmission to the control panel. The *Hart communication protocol* is widely used for such transmissions, the term Hart standing for *Highway Addressable Remote Transducer* (see Section 13.8.5).

Smart sensors are also well suited to safety-critical situations where there is a need for high reliability but access for maintenance can present problems, the smart sensor can self-diagnose and repair itself. A range of smart sensors can be used in a factory for production tracking. Smart sensors can also be used to detect and initiate an alarm and maintenance request procedure when faults on a production line occur. These types of arrangements can then lead to the 'smart factory'.

2.11.1 MEMS

MEMS stands for micro-electro-mechanical systems and refers to mechanical and electro-mechanical systems that are manufactured using the techniques of microfabrication that are used to produce transistors, integrated circuits, and other miniature electronic components. Such MEMS devices consist of microsensors, microelectronics and micro actuators all integrated on the same electronic chip. The microsensors detect the changes in the sensed environment, e.g. changes in temperature or pressure, the microelectronics processes this information and send signals to the micro actuators to react to the sensed environment changes. MEMS devices are very small.

As an illustration, a MEMS device is used with the modern airbag systems used in automobiles. This was one of the first commercial devices using MEMS. The sensor is an accelerometer (Figure 2.88A) which in one form is an integrated micromechanical capacitive device consisting of a suspended plate which moves in response to rapid deceleration and causes a change in capacitances with fixed plates, this then being detected by the MEMS

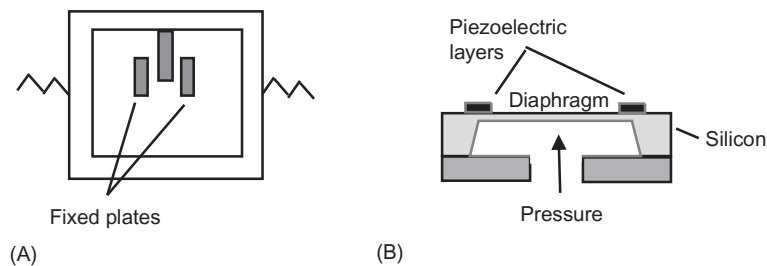


FIGURE 2.88 A basic element of a MEMS device used to (A) trigger the firing of an automobile airbag, (B) measure blood pressure.

electronics and so actuates the airbag control unit and causes it to fire and inflate the airbag. There will usually be of a number of such capacitor plate arrangements in order to give a larger capacitance change when deceleration occurs. Another example of a MEMS application is that of the miniature disposable pressure sensor used to monitor blood pressure in hospitals by being connected to a patient's intravenous line and so monitors the blood pressure through the IV solution. The disposable sensor consists of a silicon substrate which is etched to produce a membrane and is bonded to a substrate (Figure 2.88B). A piezoresistive layer is applied on the membrane surface near the edges to convert the mechanical stress into an electrical voltage. Pressure corresponds to deflection of the membrane. The sensing element is mounted on a plastic or ceramic base with a plastic cap over it, designed to fit into a manufacturer's housing. Another example of a MEMS sensor is the ultrasonic time-of-flight sensor CH-101 from TDK Corporation which is just $3.5 \text{ mm} \times 3.5 \text{ mm} \times 1.25 \text{ mm}$ and can be used for range sensing of objects up to 1 m from it with a millimetre precision, regardless of the colour of the object or whether it is transparent, and can be used to trigger a programmed response. The sensor package contains an ultrasonic piezoelectric MEMS transducer with a digital signal processor with a tiny ultrasonic transducer chip to send out a pulse of ultrasonic waves and then detect the echo signal reflected back from objects, the echo time then being a measure of the distance of the object from the sensor.

2.12 DATA PRESENTATION ELEMENT

The elements that can be used for the presentation of data can be classified into three groups: indicators, illuminative displays and recorders. *Indicators* and *illuminative displays* give an instant visual indication of the sensed variable while *recorders* record the output signal over a period of time and give automatically a permanent record. A recorder will be the most appropriate choice if the event is high speed or transient and cannot be followed by an observer, or there are large amounts of data, or it is essential to have a record of the data. The following are some brief notes about some of the characteristics of commonly used data presentation methods.

2.12.1 Indicator

The *moving coil meter* is an analogue data presentation element involving a pointer moving across a fixed scale. The basic instrument movement is a d.c. microammeter with shunts, multipliers and rectifiers being used to convert it to other ranges of direct current and alternating current, direct voltage and alternating voltage. With alternating current and voltages, the instrument is restricted to frequencies between about 50 Hz and 10 kHz. The overall accuracy is generally of the order of ± 0.1 to $\pm 5\%$. The time taken for a moving coil meter to reach a steady deflection is typically in the region of a few seconds. The low resistance of the meter can present loading problems.

2.12.2 Illuminative Displays

Commonly used illuminative display systems generally use *LEDs* or *liquid crystal displays*. LEDs are basically pn junction diodes which emit light when a suitable voltage is applied. The voltage enables electrons to recombine with holes within the device and release the energy in the form of photons. They just require low voltages and low currents in order to emit light and are cheap. The most commonly used LEDs can give red, yellow or green colours.

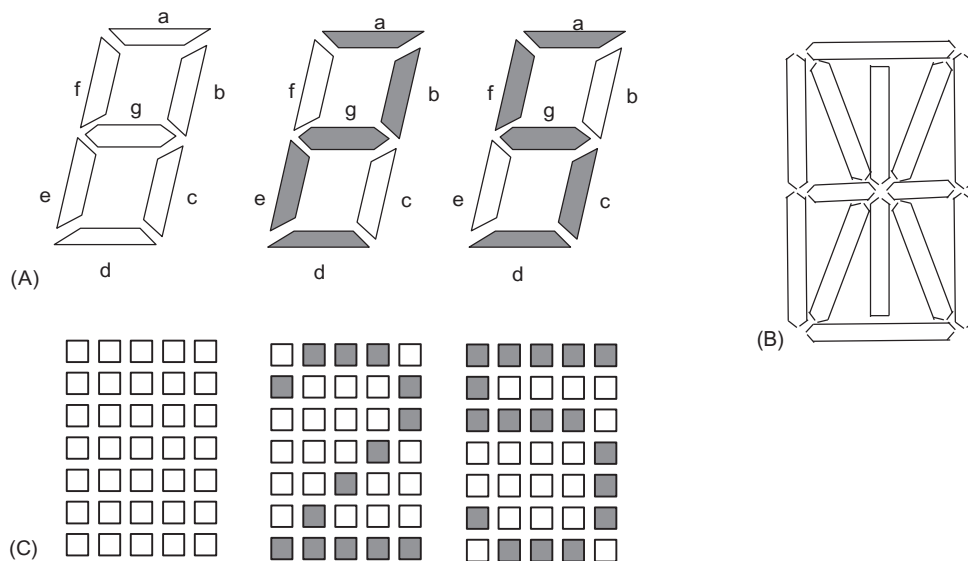


FIGURE 2.89 (A) 7-segment format with examples of the numbers 2 and 5, (B) 14-segment display, (C) 7 by 5 dot-matrix format with examples of the numbers 2 and 5, (C) 14-segment display.

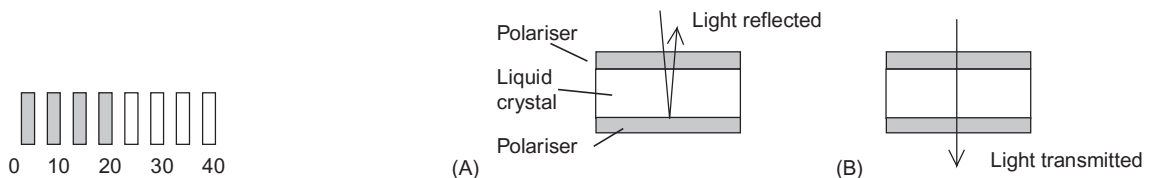


FIGURE 2.90 Bar type of display.

FIGURE 2.91 Electric field: (A) applied, (B) not applied.

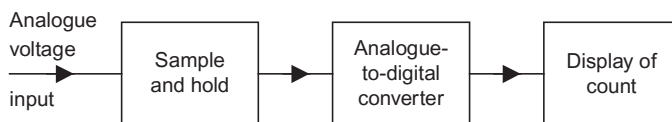


FIGURE 2.92 Principle of the digital voltmeter.

The term *alphanumeric display* is used for one that can display the letters of the alphabet and numbers. Two basic types of array are used to generate alphanumeric displays, segmented and dot matrix. The 7-segment display (Figure 2.89A) is a common form. By illuminating different segments of the display the full range of numbers and a small range of alphabetical characters can be formed. For example, to form a 2 the segments a, b, d, e, and g are illuminated. Fourteen-segment displays (Figure 2.89B), sometimes referred to as a *starburst display*, are now becoming used. The 5×7 dot-matrix (Figure 2.89C) display enables a full range of numbers and alphabetical characters to be produced by illuminating different segments in a rectangular array.

LEDs can also be arranged in other formats. For example, they can be arranged in the form of bars, the length of the illuminated bar then being a measure of some quantity (Figure 2.90). A speedometer might use this form of display.

Liquid crystal displays (LCD) do not produce any light of their own but use reflected light and can be arranged in segments like the LEDs shown above. The crystal segments are on a reflecting plate (Figure 2.91). When an electric field is applied to a crystal, light is no longer passed through it and so there is no reflected light. That segment then appears dark.

An example of an instrument using LED or LCD forms of display is the *digital voltmeter*. This gives its reading in the form of a sequence of digits and is essentially just a sample and hold unit feeding an analogue-to-digital converter with its digital output counted and the count displayed (Figure 2.92). It has a high resistance, of the order of $10 \text{ M}\Omega$, and so loading effects are less likely than with the moving coil meter with its lower resistance. The sample and hold unit takes samples and thus the specification of the sample rate with such an instrument gives the time

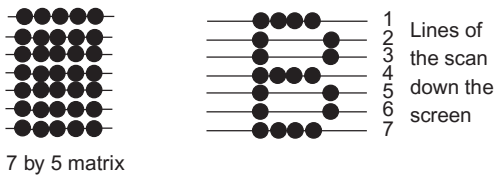


FIGURE 2.93 Character build-up by pixels.

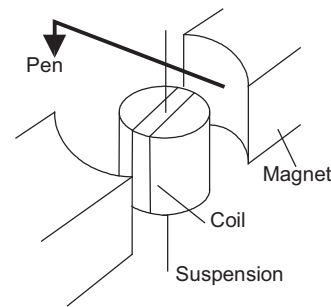


FIGURE 2.94 The principle of the galvanometric chart recorder.

taken for the instrument to process the signal and give a reading. Thus, if the input voltage is changing at a rate which results in significant changes during the sampling time the voltmeter reading can be in error.

Large screen displays, termed *visual display units* (VDUs), have developed from the cathode ray tube (CRT) to thin film liquid crystal display (LCD), light-emitting diode (LED), organic light emitting diode (OLED), and plasma displays.

With the CRT an electron gun forms images by firing electrons onto a phosphor-coated screen. The beam of electrons is deflected in the Y direction by a potential difference applied between Y-deflection plates and in the X direction by a potential difference between X-deflection plates.

An organic light-emitting diode (OLED) consist of a layer of organic semiconductor material between two electrodes and emits light in response to a current. OLED displays provide higher contrast and better viewing angles than LCDs and are widely used for mobile phone and digital camera displays.

Plasma displays use small cells containing charged ionised gases, termed plasmas, sandwiched between two panels of glass. A high voltage is applied across a cell to ionise the gas and form the plasma. This resulting freed electrons strike mercury particles as the electrons move through the plasma and cause ultraviolet light to be emitted. This cells are coated with a phosphor which converts this ultraviolet light into visible light.

The term pixel is used for the smallest addressable dot on a display device. Character generation is by the selective illuminations of these pixels. Thus for a 7 by 5 matrix, [Figure 2.93](#) shows how characters are built up by selective activation of individual display elements.

The input data to the VDU is usually in digital *ASCII* (*American Standard Code for Information Interchange*) format so that as the electron beam sweeps across the screen it is subject to on–off signals which end up ‘painting’ the characters on the screen. The ASCII code is a 7-bit code and so can be used to represent $2^7 = 128$ characters. This enables all the standard keyboard characters to be covered, as well as some control functions such as RETURN which is used to indicate the return from the end of a line to the start of the next line (see [Table 2.2](#) for an abridged list of the code).

2.12.3 Graphical User Interface (GUI)

The term *graphical user interface* (GUI) is used for an interface on a computer screen that allows the user to interact with electronic devices by manipulation of graphical elements on the screen. The WIMP (window, icon, menu, pointing device) presents information in windows, representing items as icons with available commands listed in menus. It involves the use of an input device such as a mouse to position a pointer on the screen. Such a system can be much easier to use than command-line interfaces which require commands to be typed on a keyboard.

Mimic displays are often used to present data on a computer screen from control systems. The screen displays mimic diagram of the key parts of a plant with animated symbols, updated values of controlled variables, trends in data, and analysis in the form of bar graphs. The displays also allow operators to interact with the controlled system by setting demand values.

2.12.4 Data Loggers

The *galvanometric type* of chart recorder ([Figure 2.94](#)) which worked on the same principle as the moving coil meter movement has largely been replaced as a recorder of data by the *data logger*. The galvanometric recorder

was essentially a coil suspended between two fixed points by a suspension wire and in the magnetic field produced by a permanent magnet. When a current passes through the coil a torque acts on it, causing it to rotate and twist the suspension. The coil rotates to an angle at which the torque is balanced by the opposing torque resulting from the twisting of the suspension. The rotation of the coil results in a pen being moved across a chart.

Omni Instruments have a general purpose chart recorder BD 300 in 4, 6, and 8 pen versions. It has an optional floppy disk for independent storage and an optional RS-232C interface for connection to a PC or network.

Data loggers are electronic systems that record data over time in internal memory from either built-in sensors or external sensors. The stored data can then be downloaded to a computer.

For example, OMEGA offers a wide range of data loggers with some dedicated to a certain type of input while others are programmable for different types of inputs. The OMEGA Temperature and Humidity data logger OM-DVTH has a graphical display (LCD) as well as data storage for the ambient temperature and humidity. The display shows the real time, temperature, humidity and dew point, a logging summary, sample rate, memory used, and memory left. Over 40 000 data points can be stored for later downloading to a computer. The instrument can sample as fast as once per second. The temperature accuracy is $\pm 0.5^\circ\text{C}$ with a range of -10 to 65°C , the relative humidity range being 0 to 99%. The interface to a computer is via a USB port. Another OMEGA data logger is the RDXL6SD which is a 6-channel temperature data logger for use with 4 thermocouples and 2 resistance temperature sensors. Another data logger is the OMRON SQ2010 which is a general purpose data logger with 4 to 8 analogue input channels to measure current, voltage, resistance and temperature plus 8 digital channels. It has an LCD display and a USB interface for downloading.

Omni Instruments have a general purpose data logger SQ2010 for 4 to 8 analogue inputs for use with current, voltage, resistance and temperature inputs and 8 digital inputs, gives a LCD display and has interfaces for RS232 and USB.

2.12.5 Printers

Alphanumeric printers give records in the form of numbers, letters, or special characters These can be dot-matrix, ink-jet, or laser printers.

The *dot-matrix printer* has a print head of either 9 or 24 pins in a vertical line (Figure 2.95). Each pin is controlled by an electromagnet which when turned on presses the pin onto the inking ribbon and so gives a small blob of ink on the paper behind the ribbon. The alphanumeric characters are formed by moving the print head across the paper and firing the appropriate pins.

The *ink jet printer* uses conductive ink being forced through a small nozzle to produce a jet of drops of ink. Very small, constant diameter, drops of ink are produced at a constant frequency and so regularly spaced in the jet. With one form a constant stream of ink passes along a tube and is pulsed to form fine drops by a piezoelectric crystal which vibrates at a frequency of about 100 kHz (Figure 2.96). Another form uses a small heater in the print head with vaporised ink in a capillary tube, so producing gas bubbles which push out drops of ink (Figure 2.97). In one form each drop of ink is given a charge as a result of passing through a charging electrode. The charged drops then pass between deflection plates which can deflect the stream of drops in a vertical direction, the amount of deflection depending on the charge on the drops. In another form, a vertical stack of nozzles is used and the jets of each just switched on or off on demand.

The *laser printer* has a photosensitive drum which is coated with a selenium-based material that is light sensitive (Figure 2.98). In the dark the selenium has a high resistance and becomes charged as it passes close to the



FIGURE 2.95 9-pin print head.

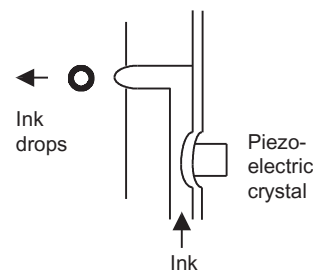


FIGURE 2.96 Continuous flow printing.

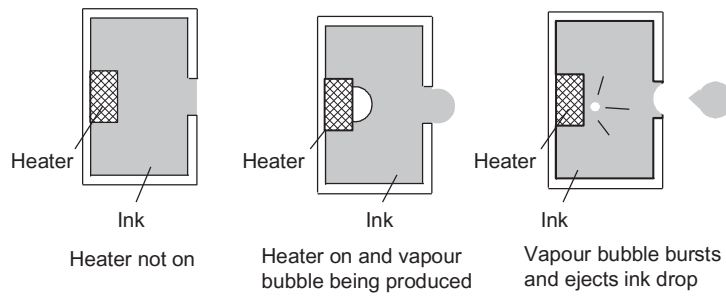


FIGURE 2.97 Principle of the on-demand system.

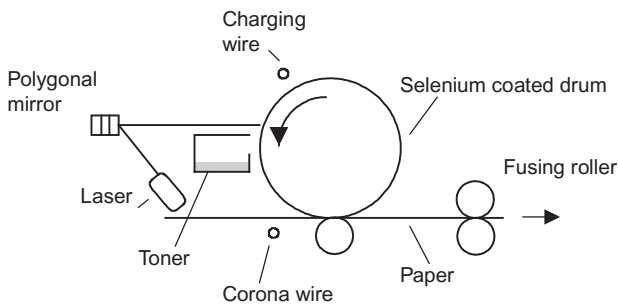


FIGURE 2.98 Basic elements of a laser printer.

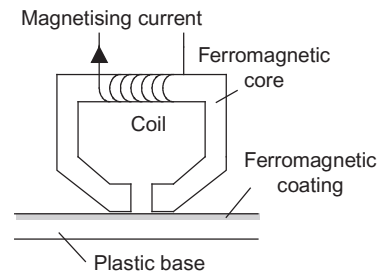


FIGURE 2.99 A magnetic recording write/read head.

charging wire. This is a wire which is held at a high voltage and charge leaks off it. A light beam is made to scan along the length of the drum by a small eight-sided mirror which rotates and so reflects the light that it scans across the drum. When light strikes the selenium its resistance drops and it can no longer remain charged. By controlling the brightness of the beam of light, so points on the drum can be discharged or left charged. As the drum passes the toner reservoir, the charged areas attract particles of toner which then stick to its surface to give a pattern of toner on the drum with toner on the areas that have not been exposed to light and no toner on the areas exposed to light. The paper is given a charge as it passes another charging wire, the so-called corona wire, with the result that as the paper passes close to the drum it attracts the toner off the drum. A hot fusing roller is then used to melt the toner particles so that, after passing between rollers, they firmly adhere to the paper.

Magnetic recording involves data being stored in a thin layer of magnetic material as a sequence of regions of different magnetism. The material may be in the form of magnetic tape or disks, the disks being referred to as hard disks or floppy disks. Figure 2.99 shows the basic principle of magnetic recording. The recording current is passed through a coil wrapped round a ferromagnetic core. This core has a small non-magnetic gap. The proximity of the magnetic tape or disk to the gap means that the magnetic flux in the core is readily diverted through it. The magnetic tape or disk consists of a plastic base coated with a ferromagnetic powder. When magnetic flux passes through this material it becomes permanently magnetised. Thus a magnetic record can be produced of the current through the coil. The patterns of magnetism on a tape or disk can be read by passing it under a similar head to the recording head. The movement of the magnetised material under the head results in magnetic flux passing through the core of the head and electromagnetic induction producing a current through the coil wrapped round the head.

PROBLEMS

Questions 1 to 19 have four answer options: A, B, C and D. Choose the correct answer from the answer options.

Questions 1 to 6 relate to the following information. The outputs from sensors can take a variety of forms. These include changes in:

- A. Displacement
- B. Resistance
- C. Voltage
- D. Capacitance

Select the form of output from above which is concerned with the following sensors:

1. A thermocouple which has an input of a temperature change.
2. A thermistor which has an input of a temperature change.
3. A diaphragm pressure cell which has an input of a change in the pressure difference between its two sides.
4. A LVDT which has an input of a change in displacement.
5. A strain gauge which has an input of a change in length.
6. A Bourdon gauge which has an input of a pressure change.
7. Decide whether each of these statements is True (T) or False (F).
In selecting a temperature sensor for monitoring a rapidly changing temperature, it is vital that the sensor has:
 - i. A small thermal capacity.
 - ii. High linearity.
 Which option BEST describes the two statements?
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
8. A copper-constantan thermocouple is to be used to measure temperatures between 0 and 200°C. The e.m.f. at 0°C is 0 mV, at 100°C it is 4.277 mV and at 200°C it is 9.286 mV. If a linear relationship is assumed between e.m.f. and temperature over the full range, the non-linearity error at 100°C is:
 - A. - 3.9°C
 - B. - 7.9°C
 - C. + 3.9°C
 - D. + 7.9°C
9. The change in resistance of an electrical resistance strain gauge with a gauge factor of 2.0 and resistance 50 Ω when subject to a strain of 0.001 is:
 - A. 0.0001 Ω
 - B. 0.001 Ω
 - C. 0.01 Ω
 - D. 0.1 Ω
10. An incremental shaft encoder gives an output which is a direct measure of:
 - A. The absolute angular position of a shaft.
 - B. The change in angular rotation of a shaft.
 - C. The diameter of the shaft.
 - D. The change in diameter of the shaft.
11. A pressure sensor consisting of a diaphragm with strain gauges bonded to its surface has the following information in its specification:
 - Range: 0 to 1000 kPa
 - Non-linearity error: $\pm 0.15\%$ of full range
 - Hysteresis error: $\pm 0.05\%$ of full range
 - The total error due to non-linearity and hysteresis for a reading of 200 kPa is:
 - A. ± 0.2 kPa
 - B. ± 0.4 kPa
 - C. ± 2 kPa
 - D. ± 4 kPa
12. The water level in an open vessel is to be monitored by a diaphragm pressure cell responding to the difference in pressure between that at the base of the vessel and the atmosphere. The range of pressure differences across the diaphragm that the cell will have to respond to if the water level can vary between

zero height above the cell measurement point and 1 m above it is (take the acceleration due to gravity to be 9.8 m/s^2 and the density of the water as 1000 kg/m^3):

- A. 102 Pa
- B. 102 kPa
- C. 9800 Pa
- D. 9800 kPa

13. Decide whether each of these statements is True (T) or False (F).

A float sensor for the determination of the level of water in a container is cylindrical with a mass 1.0 kg, cross-sectional area 20 cm^2 and a length of 0.5 m. It floats vertically in the water and presses upwards against a beam attached to its upward end.

- i. The maximum force that can act on the beam is 9.8 N.
- ii. The minimum force that can act on the beam is 8.8 N.

Which option BEST describes the two statements?

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

14. A Wheatstone bridge when used as a signal processing element can have an input of a change in resistance and an output of:

- A. A bigger resistance change.
- B. A digital signal.
- C. A voltage.
- D. A current.

15. The resolution of an analogue-to-digital converter with a word length of 8 bits and an analogue signal input range of 10 V is:

- A. 39 mV
- B. 625 mV
- C. 1.25 V
- D. 5 V

16. A sensor gives a maximum analogue output of 5 V. The word length is required for an analogue-to-digital converter if there is to be a resolution of 10 mV is:

- A. 500 bits
- B. 250 bits
- C. 9 bits
- D. 6 bits

17. Decide whether each of these statements is True (T) or False (F).

A cold junction compensator circuit is used with a thermocouple if it has:

- i. No cold junction.
- ii. A cold junction at the ambient temperature.

Which option BEST describes the two statements?

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

18. Decide whether each of these statements is True (T) or False (F).

A data presentation element which has an input which results in a pointer moving across a scale is an example of:

- i. An analogue form of display.
- ii. An indicator form of display.

Which option BEST describes the two statements?

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

19. Suggest sensors which could be used in the following situations:
 - a. To monitor the rate at which water flows along a pipe and given an electrical signal related to the flow rate.
 - b. To monitor the pressure in a pressurised air pipe, giving a visual display of the pressure.
 - c. To monitor the displacement of a rod and give a voltage output.
 - d. To monitor a rapidly changing temperature.
20. Suggest the type of signal processing element that might be used to:
 - a. Transform an input of a resistance change into a voltage.
 - b. Transform an input of an analogue voltage into a digital signal.
21. A potentiometer with a uniform resistance per unit length of track is to have a track length of 100 mm and used with the output being measured with an instrument of resistance 10 k Ω . Determine the resistance required of the potentiometer if the maximum error is not to exceed 1% of the full-scale reading.
22. A platinum resistance coil has a resistance at 0°C of 100 Ω . Determine the change in resistance that will occur when the temperature rises to 30°C if the temperature coefficient of resistance is 0.0039 K⁻¹.
23. A platinum resistance thermometer has a resistance of 100.00 Ω at 0°C, 138.50 Ω at 100°C and 175.83 Ω at 200°C. What will be the non-linearity error at 100°C if a linear relationship is assumed between 0°C and 200°C?
24. An electrical resistance strain gauge has a resistance of 120 Ω and a gauge factor of 2.1. What will be the change in resistance of the gauge when it experiences a uniaxial strain of 0.0005 along its length?
25. A capacitive sensor consists of two parallel plates in air, the plates each having an area of 1000 mm² and separated by a distance of 0.3 mm in air. Determine the displacement sensitivity of the arrangement if the dielectric constant for air is 1.0006.
26. A capacitive sensor consists of two parallel plates in air, the plates being 50 mm square and separated by a distance of 1 mm. A sheet of dielectric material of thickness 1 mm and 50 mm square can slide between the plates. The dielectric constant of the material is 4 and that for air may be assumed to be 1. Determine the capacitance of the sensor when the sheet has been displaced so that only half of it is between the capacitor plates.
27. A chromel – constantan thermocouple has a cold junction at 20°C. What will be the thermoelectric e.m.f. when the hot junction is at 200°C? Tables give for this thermocouple: 0°C, e.m.f. 0.000 mV; 20°C, e.m.f. 1.192 mV; 200°C, e.m.f. 13.419 mV.
28. An iron – constantan thermocouple has a cold junction at 0°C and is to be used for the measurement of temperatures between 0°C and 400°C. What will be the non-linearity error at 100°C, as a percentage of the full-scale reading, if a linear relationship is assumed over the full range? Tables give for this thermocouple: 0°C, e.m.f. 0.000 mV; 100°C, e.m.f. 5.268 mV; 400°C, e.m.f. 21.846 mV.
29. Show that the output voltage for a Wheatstone bridge with a single strain gauge in one arm of the bridge and the other arms all having the same resistance as that of the unstrained strain gauge is $\frac{1}{4}V_s G \epsilon$, where V_s is the supply voltage to the bridge, G the gauge factor of the strain gauge and ϵ the strain acting on the gauge.
30. A Wheatstone bridge has a platinum resistance temperature sensor with a resistance of 120 Ω at 0°C in one arm of the bridge. At this temperature the bridge is balanced with each of the other arms being 120 Ω . What will be the output voltage from the bridge for a change in temperature of 20°C? The supply voltage to the bridge is 6.0 V and the temperature coefficient of resistance of the platinum is 0.0039 K⁻¹.
31. A diaphragm pressure gauge employs four strain gauges to monitor the displacement of the diaphragm. A differential pressure applied to the diaphragm results in two of the gauges on one side of the diaphragm being subject to a tensile strain of 1.0×10^{-5} and the two on the other side a compressive strain of 1.0×10^{-5} . The gauges have a gauge factor of 2.1 and resistance 120 Ω and are connected in the bridge with the gauges giving subject to the tensile strains in arms 1 and 3 and those subject to compressive strain in arms 2 and 4 (Figure 2.57). If the supply voltage for the bridge is 10 V, what will be the voltage output from the bridge?
32. A thermocouple gives an e.m.f. of 820 μ V when the hot junction is at 20°C and the cold junction at 0°C. Explain how a Wheatstone bridge incorporating a metal resistance element can be used to compensate for when the cold junction is at the ambient temperature rather than 0°C and determine the parameters for the bridge if a nickel resistance element is used with a resistance of 10 Ω at 0°C and a temperature coefficient of resistance of 0.0067 K⁻¹ and the bridge voltage supply is 2 V.
33. An operational amplifier circuit is required to produce an output that ranges from 0 to -5 V when the input goes from 0 to 100 mV. By what factor is the resistance in the feedback arm greater than that in the input?

34. What will be the feedback resistance required for an inverting amplifier which is to have a voltage gain of 50 and an input resistance of $10\text{ k}\Omega$?
35. What will be the feedback resistance required for a non-inverting amplifier which is to have a voltage gain of 50 and an input resistance of $10\text{ k}\Omega$?
36. A differential amplifier is to have a voltage gain of 100 and input resistances of $1\text{ k}\Omega$. What will be the feedback resistance required?
37. A differential amplifier is to be used to amplify the voltage produced between the two junctions of a thermocouple. The input resistances are to be $1\text{ k}\Omega$. What value of feedback resistance is required if there is to be an output of 10 mV for a temperature difference between the thermocouple junctions of 100°C with a copper-constantan thermo-couple. The thermocouple can be assumed to give an output of $43\text{ }\mu\text{V}/^\circ\text{C}$?
38. What is the resolution of an analogue-to-digital converter with a word length of 12 bits?
39. A sensor gives a maximum analogue output of 5 V . What word length is required for an analogue-to-digital converter if there is to be a resolution of 10 mV ?
40. What is the voltage resolution of an 8-bit DAC when it has a full-scale input of 5 V ?

Measurement Case Studies

O U T L I N E

3.1 Introduction	71	3.2.9 Automobile Tyre Pressure Monitoring	77
3.2 Case Studies	72	3.2.10 Control System Sensors with Automobiles	77
3.2.1 A Temperature Measurement	72	3.3 Data Acquisition Systems	78
3.2.2 An Absolute Pressure Measurement	73	3.3.1 Data Acquisition Software	80
3.2.3 Detection of the Angular Position of a Shaft	74	3.3.2 Data Loggers	80
3.2.4 Air Flow Rate Determination	74	3.4 Testing	80
3.2.5 Fluid Level Monitoring	75	3.4.1 Maintenance	80
3.2.6 Measurement of Relative Humidity	75	3.4.2 Common Faults	81
3.2.7 Dimension Checking	76	Problems	82
3.2.8 Temperature of a Furnace	77		

3.1 INTRODUCTION

In designing measurement systems, there are a number of steps that need to be considered:

1. Identification of the *nature of the measurement* required.
 For example, what is the variable to be measured, its nominal value, the range of values that might have to be measured, the accuracy required, the required speed of measurement, the reliability required, the environmental conditions under which the measurement is to be made, etc.
2. Identification of *possible sensors*.
 This means taking into account such factors as their range, accuracy, linearity, speed of response, reliability, maintainability, life, power supply requirements, ruggedness, availability, cost. The sensor needs to fit the requirements arrived at in 1 and also be capable, with suitable signal processing, to give the required output for use in a control system and/or display.
3. Selection of appropriate *signal processing*.
 This element needs to take the output signal from the sensor and modify it in such a way as to enable it to drive the required display or be suitable for control of some device. For example, control applications might require a 4–20 mA current to drive an actuator.
4. Identification of the *required display/storage/transmission*.
 This means considering the form of display and any data storage and transmission that is required. What is the purpose of the display? Why store the data? How far does it have to be transmitted for display and analysis?

Digital data frequently has to be transmitted over some distance from a measurement system before it can be used. There are two basic methods of transmission, namely parallel and serial. With parallel data transfer, each bit of the data is transmitted along its own conducting path. With serial data transfer, the data is transferred bit by bit along a single conducting path. This is slower but requires less conducting paths and so is much cheaper. For this reason, other than for transmissions within a computer where the distances involved are very small, serial data transmission is used (see Chapter 2.10). A common standard that is used with instrumentation and control systems for serial transmission is the RS232 standard. The standard defines the electrical signal characteristics such as voltage and signalling rate, the mechanical means needed to connect transmitters and receivers, i.e. pluggable connectors, and the functions of circuits in the interface connectors. RS232 is good for signal rates up to 20 kbit/s but has the limitation that it can only support communication to one device connected to the bus at a time. The standard does not define a maximum cable length but 15 m is about the maximum length without special cables being used if the resulting capacitance is not to become too large. For short distance connection to a computer, the universal series bus (USB) is likely to be used for serial transmission. For USB 2.0 the maximum cable length is likely to be about 5 m and for USB 3.0 about 3 m. USB 2.0 can achieve a transmission rate of 480 Mbits/s while USB 3.0 achieves up to 5 Gbits/s. USB is ideal for short distance data acquisition since it has the ability to provide power to the peripheral device.

This chapter is a consideration of some examples of measurement systems and the selection of the elements in such systems.

3.2 CASE STUDIES

The following gives case studies of measurement systems.

3.2.1 A Temperature Measurement

Requirement: Determination of temperature of a liquid in the range 0–100°C where only rough accuracy is required. The situation might be the determination of the temperature of the cooling water for a car engine and its display as a pointer moving across a scale marked to indicate safe and unsafe operating temperatures.

Sensor: A solution might be to use a *thermistor* as a sensor. It is cheap and robust. This is the commonly used solution with car engine coolant.

Signal processing: The resistance change of the thermistor has to be converted into a voltage which can then be applied across a meter and so converted to a current through it and hence a reading on the meter related to the temperature. [Figure 3.1](#) shows a possible solution involving a *potential divider circuit* to convert the resistance change into a voltage change.

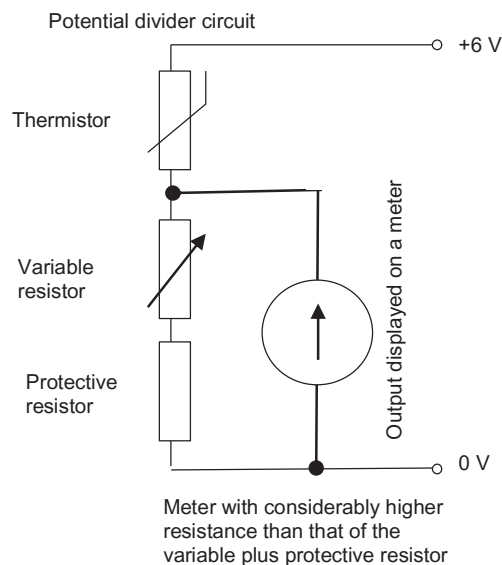


FIGURE 3.1 Temperature measurement.

Suppose we use a $4.7\text{ k}\Omega$ bead thermistor. This has a resistance of $4.7\text{ k}\Omega$ at 25°C , $15.28\text{ k}\Omega$ at 0°C and $0.33\text{ k}\Omega$ at 100°C . The variable resistor might be $0\text{--}10\text{ k}\Omega$. It enables the sensitivity of the arrangement to be altered. However, if the variable resistor was set to zero resistance then, without a protective resistor, we could possibly have a large current passed through the thermistor. The *protective resistor* is there to prevent this occurring. The maximum power that the thermistor can withstand is specified as 250 mW . Thus, with a 6 V supply, the variable resistor set to zero resistance, the protective resistance of R , and the thermistor at 100°C , the current I through the thermistor is given by $V = IR$ as $6 = I(0 + R + 330)$, and so:

$$I = \frac{6}{R + 330}$$

The power dissipated by the thermistor is $I^2 \times 330$ and so if we want this to be significantly below the maximum possible, say 100 mW , then we have

$$0.100 = \left(\frac{6}{R + 330} \right)^2 \times 330$$

Hence R needs to be about $15\ \Omega$.

Display: When the temperature of the thermistor is 0°C its resistance is $15.28\text{ k}\Omega$. If we set the variable resistor as, say, $5\text{ k}\Omega$ and the protective resistor as $15\ \Omega$ then the voltage output when the supply is 6 V is

$$\text{Output voltage} = \frac{5.015}{12.58 + 5.015} \times 6 = 1.48\text{ V}$$

When the temperature rises to 100°C the output voltage becomes

$$\text{Output voltage} = \frac{5.015}{0.33 + 5.015} \times 6 = 5.63\text{ V}$$

Thus, over the required temperature range, the voltage output varies from 1.48 V to 5.63 V . A voltmeter to cover this range could be used to display the output.

In general, calibration of thermometers is by determining their response at temperatures which are specified as the standard values for freezing points and boiling points for pure materials. Alternatively, calibration at these and other temperatures within the range can be obtained by comparison with the readings given by a standard thermometer.

3.2.2 An Absolute Pressure Measurement

Requirement: Measurement of the manifold absolute pressure in a car engine as part of the electronic control of engine power.

Sensor: A sensor that is used for such a purpose is a *diaphragm pressure gauge*. Figure 3.2A shows the basic form of a diaphragm pressure gauge which is often used in such circumstances. The diaphragm is made of silicon with the strain gauges diffused directly into its surface. Four strain gauges are used and so arranged that when two are in tension, the other two are in compression (Figure 3.2B).

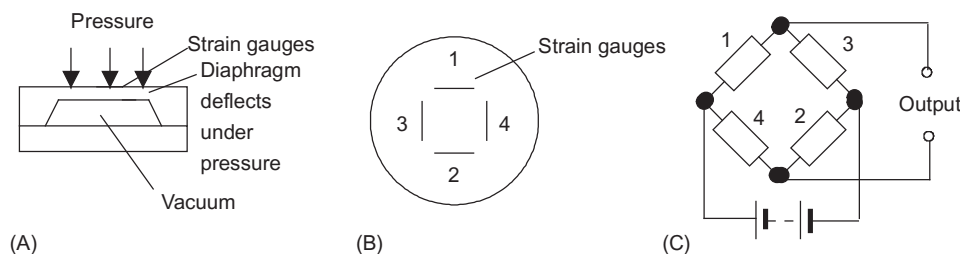


FIGURE 3.2 Diaphragm pressure gauge: (A) basic form of sensor, (B) possible arrangement of strain gauges on diaphragm and (C) the Wheatstone bridge signal processor.

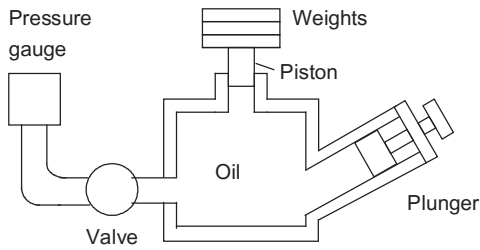


FIGURE 3.3 Dead-weight pressure gauge calibration.

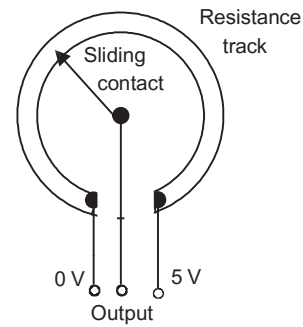


FIGURE 3.4 Rotary potentiometer.

Signal processing: The four gauges are so connected as to form the arms of a Wheatstone bridge (Figure 3.2C). This gives temperature compensation since a change in temperature affects all the gauges equally. Thus the output from the sensor with its signal processing is a voltage which is a measure of the pressure.

Display: If required, the output voltage could be displayed on a meter, possibly following some amplification. Calibration of pressure gauges is usually with a *dead-weight pressure system*. Figure 3.3 shows the basic form of such a system. The calibration pressures are generated by adding standard weights to the piston tray, the pressure then being W/A , where W is the total weight of the piston, tray, and standard weights and A the cross-sectional area of the piston. After the weights have been placed on the tray, the screw-driven plunger is screwed up to force the oil to lift the piston-weight assembly. Then the oil is under the pressure given by the piston-weight assembly since it is able to support that weight. By adding weights to the piston tray, a gauge can be calibrated over its range.

Data transmission: In the case of the pressure output being required as an element in a control process, there is a need to transmit the data to the controller. This will generally be done by serial transmission in which the data is converted to digital form and then sent bit by bit to the controller (see the discussion of the CAN bus in Chapter 13).

3.2.3 Detection of the Angular Position of a Shaft

Requirement: Detection of the angular position of the throttle shaft of an automobile to give an indication of the throttle opening, and hence the driver's power demand on the engine, as part of a car engine management system.

Sensor: A rotary potentiometer (Figure 3.4) is generally used with the potentiometer wiper being rotated over the potentiometer track.

Signal processing: For a 5 V d.c. voltage connected across the potentiometer, with the throttle closed and the engine idling the wiper can be at a position close to the 0 V terminal and so give a small voltage output, typically about 0.5 V. As the throttle is opened, the shaft rotates and the wiper moves over the track so that at wide-open throttle the wiper is nearly at the end of its track and the output voltage has risen to about 4.3 V. The engine management system uses an operational amplifier to compare the output from the potentiometer with a fixed voltage of 0.5 V so that the op-amp gives a high output when the potentiometer output is 0.5 V or lower and a low output when higher. This high–low signal, together with signals from other sensors, is fed to a micro-processor which then can give an output to control the engine idle speed.

Data transmission: As with the previous case, the data is likely to be transmitted to the controller by serial transmission and so needs conversion to a digital signal before transmission.

3.2.4 Air Flow Rate Determination

Requirement: Measurement of the flow rate for the inflow of air in a car manifold in an electronic controlled engine. A simple, and cheap, measurement of the mass rate of flow of air is required with the output being an electrical signal which can be used for control purposes.

Sensor: One method that is used with cars is the *hot-wire anemometer*. This sensor consists of a platinum wire which is heated by an electrical current passing through it to about 100–200°C. The temperature of the wire will depend on the cooling generated by the flow of air over the wire. Thus, since the electrical resistance of the wire will depend on its temperature, the resistance is a measure of the rate of flow of air over the heated wire. Figure 3.5A shows the basic form of such a sensor.

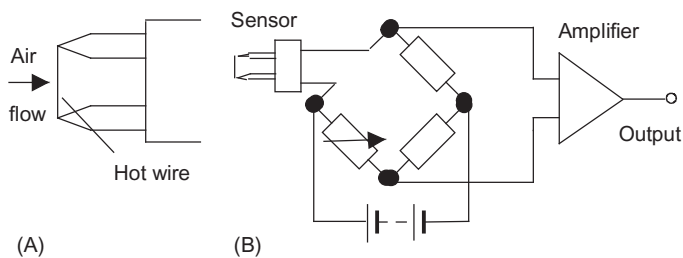


FIGURE 3.5 Hot-wire anemometer and its signal processing.

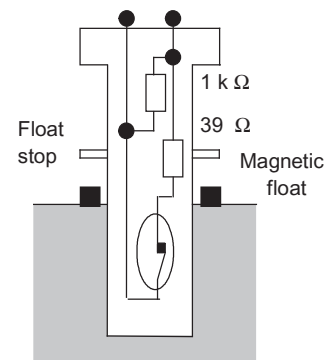


FIGURE 3.6 Liquid level monitoring.

Signal processing: The resistance change is transformed into a voltage change by incorporating the sensor as one of the arms of a Wheatstone bridge (Figure 3.5B). The bridge is balanced at zero rate of flow and then the out-of-balance voltage is a measure of the rate of flow. This voltage is fairly small and so has generally to be amplified.

An alternative arrangement which is used is

Sensor: The vortex flow sensor (see Section 2.5.4) with the vortex frequency measured by means of a semiconductor pressure sensor.

Signal processing: The output of the pressure sensor is typically a frequency of about 100 Hz when the engine is idling and rising to about 2000 Hz at high engine speed. Signal conditioning is used to transform this output into a square-wave signal which varies between 0.6 V and 4.8 V and can then be processed by the engine control unit.

3.2.5 Fluid Level Monitoring

Requirement: Monitoring the level of a liquid to indicate when the level falls below some critical value.

Sensor: One method would be to use a magnetic float (Figure 3.6) which rises with the liquid level and opens a reed switch (see Section 2.2.8) when the level falls too low.

Signal conditioning: The reed switch is in series with a $39\ \Omega$ resistor so that this is switched in parallel with a $1\ \text{k}\Omega$ resistor by the action of the reed switch. Opening the reed switch thus increases the resistance from about $37\ \Omega$ to $1\ \text{k}\Omega$. Such a resistance change can be further transformed by signal conditioning to give suitable light on–off signals.

3.2.6 Measurement of Relative Humidity

Requirement: Direct measurement of relative humidity without the need for the operator to use tables to convert temperature values to relative humidity. The traditional method of measuring relative humidity involves two thermometers, one with its bulb directly exposed to the air and giving the ‘dry temperature’ and the other with its bulb covered with muslin which dips into water. The rate of evaporation from the wet muslin depends on the amount of water vapour present in the air; when the air is far from being saturated then the water evaporates quickly, when saturated there is no net evaporation. This rate of evaporation affects the temperature indicated by the thermometer, so giving the ‘wet temperature’. Tables are then used to convert these readings into the humidity.

Sensor: Rather than use a ‘wet’ thermometer element, a capacitive humidity sensor can be used. The sensor (Figure 3.7A) consists of an aluminium substrate with its top surface oxidised to form a porous layer of aluminium oxide. On top of the oxide a very thin gold layer is deposited, this being permeable to water vapour. Electrical connections are made to the gold layer and the aluminium substrate, the arrangement being a capacitor with an aluminium oxide dielectric. Water vapour enters the pores of the aluminium oxide and changes its dielectric constant and hence the capacitance of the capacitor. The capacitance thus gives a measure of the amount of water vapour present in the air.

Signal processing: Figure 3.7B shows the type of system that might be used with such a sensor. For the capacitive sensor, signal conditioning is used to transform the change in capacitance to a suitable size voltage signal. A temperature sensor is also required since the maximum amount of water vapour that air can hold depends on the temperature and thus to compute the humidity the microprocessor needs to know the temperature, this also requiring signal conditioning to get a signal of the right size. An ADC is then used to convert the signals to

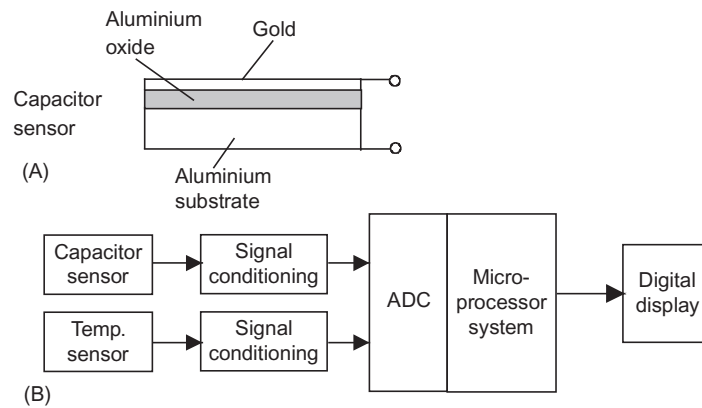


FIGURE 3.7 Relative humidity measurement.

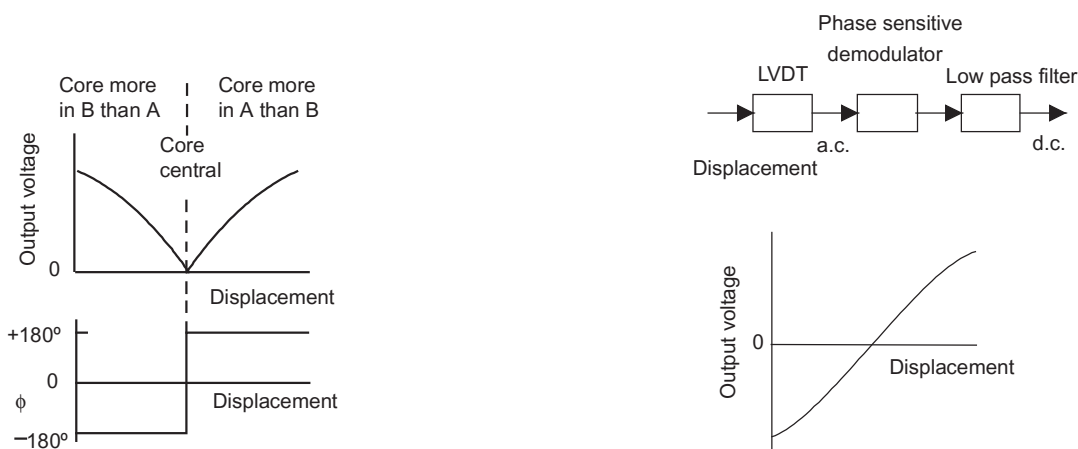


FIGURE 3.8 LVDT output.

FIGURE 3.9 LVDT d.c. output.

digital for processing by a microprocessor system; a microcontroller is likely to be used with an integrated ADC, microprocessor and memory on a single chip, there then being a number of input connections for analogue signals to the system. The microprocessor takes the values for the two inputs and can use a 'look-up' table in its memory to determine the value of the relative humidity. This is then outputted to a digital meter.

3.2.7 Dimension Checking

Requirement: A method by which the dimensions of components can be checked.

Sensor: LVDT displacement sensors can be used (see Section 2.2.4).

Signal processing: The e.m.f. induced in a secondary coil by a changing current I in the primary coil is given by $e = M di/dt$, where M is the mutual inductance, its value depending on the number of turns on the coils and the magnetic linkage between the coils, thus the material in the core of the coils. Thus, for a sinusoidal input current to the primary coil of an LVDT, alternating e.m.f.s are induced in the two secondary coils A and B. The two outputs are in series so that their difference is the output. Figure 3.8 shows how the size and phase of the alternating output changes with the displacement of the core. The same amplitude output voltage is produced for two different displacements. To give an output voltage which distinguishes between these two situations, a phase sensitive demodulator, with a low pass filter, is used to convert the output into a d.c. voltage which gives a unique value for each displacement (Figure 3.9).

The coils with simple LVDT sensors and parallel-sided coils exhibit non-linearities as the ferrite core approaches the ends of the coils. This can be corrected for by using stepped windings or, more cheaply, by using a microprocessor system and programming it to compensate for such non-linearities. Such a microprocessor

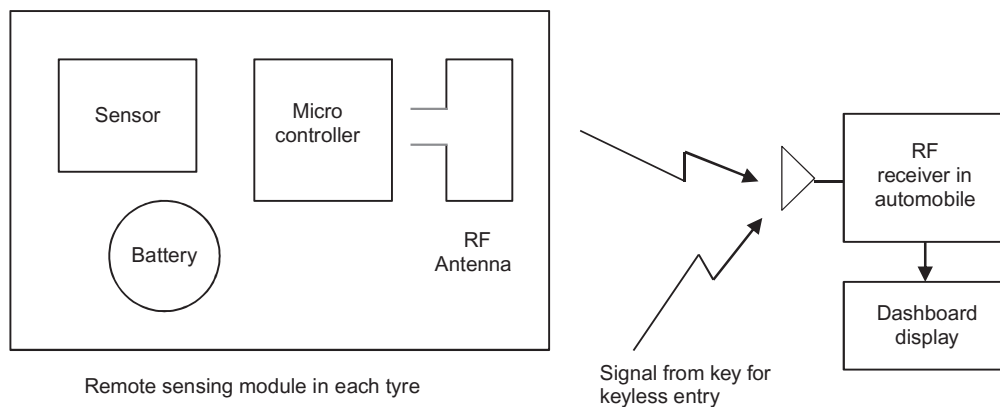


FIGURE 3.10 Monitoring of tyre pressure.

system could also be used to receive the inputs from a number of such LVDT sensors and compare the outputs with the required dimensions of the components and give outputs indicating divergence.

3.2.8 Temperature of a Furnace

Requirement: To monitor the temperature of a furnace operating from room temperature up to 500°C with an accuracy of plus or minus a few degrees.

Sensor: A chromel–aluminium thermocouple (see Table 2.1), with a sensitivity of 41 mV/°C and range –180°C to 1260°C, can be used.

Signal processing: Cold junction compensation and amplification are required (see Section 2.7.5) for the output to be displayed on a meter. Calibration can be against the fixed temperature points or a secondary standard thermometer.

3.2.9 Automobile Tyre Pressure Monitoring

Requirement: To monitor tyre pressures and give signals about them on the dashboard.

Sensor and signal conditioning: A sensor that is used is the Motorola MPXY8020A smart sensor. This includes a capacitance pressure sensing element, a resistive temperature sensing element and interface circuits for signal conditioning on a single chip. The sensor together with a microcontroller is combined with an ultra-high frequency radio transmitter on a single chip (Motorola 68HC08RF2) and also mounted in the tyre (Figure 3.10). The signals from such transmissions are decoded by the receiver and at its simplest might just activate indicator lights on the dashboard to indicate that the tyre pressure needs checking or with a more refined system give indications of the actual tyre pressures. In most automobiles, the receiver also is used for the remote keyless entry signal.

3.2.10 Control System Sensors with Automobiles

Modern automobiles use a wide variety of sensors to provide the signals required to operate the vehicle and exercise control over the wide variety of systems involved. The following are just a few of the sensors used.

To monitor the ambient air temperature

Thermistor with a range –40 to 100°C with an accuracy of $\pm 2^\circ\text{C}$.

To monitor the coolant temperature

Thermistor with a range –40 to 100°C with an accuracy of $\pm 2^\circ\text{C}$.

To monitor the inlet manifold pressure

Piezoresistive silicon strain gauged diaphragm with a range 0–105 kPa and an accuracy of $\pm 3\%$.

To monitor the brake fluid pressure

Piezoresistive silicon strain gauged diaphragm with a range 250 bar and an accuracy of $\pm 5\%$.

To monitor the throttle position

Potentiometer with a range of 0–4 k Ω from closed to fully open throttle and an accuracy of $\pm 3\%$.

To monitor the vehicle speed

The angular speed of the drive-shaft is monitored using a magnetic reluctance coil sensor to give pulses each time tabs mounted on a disk attached to the shaft move past it. The sensor consists of a coil wound round a magnet. When the tab, of ferromagnetic material, on the rotating shaft moves past the end of the coil it changes the magnetic flux linked by the coil and so induces an e.m.f. pulse in the coil which can be detected and counted.

To monitor the gear selector position

Microswitch with 8 position selection.

To monitor the fuel level in the fuel tank

A floating-arm level sensor which converts the liquid level to an angular position using a floating arm and this then is converted to an electrical signal by a potentiometer.

To operate the internal light when a door is opened

A microswitch.

To determine the end of travel of a window when opened

A microswitch.

To monitor the oxygen in the exhaust gas

The zirconium dioxide oxygen sensor is widely used. When hot the zirconium dioxide element produces a voltage that depends on the amount of oxygen in the exhaust compared to that in the outside air. Figure 3.11 shows the basic principle. The zirconium dioxide is the electrolyte between two platinum electrodes, one being located at the exhaust side and one the air side. If the levels of oxygen are different at the two electrodes, oxygen ions will flow from the air-side electrode to the exhaust-side electrode and a potential difference results can be monitored.

3.3 DATA ACQUISITION SYSTEMS

Frequently, modern instrumentation systems involve the use of PCs and the term data acquisition is used. *Data acquisition* is the process by which data from sensors is transformed into electrical signals that are converted into digital form for processing and analysis by a computer. A data acquisition system (DAQ) will thus include:

1. *Sensors*
2. *Signal processing* to get the signal into the required form and size for the data acquisition hardware
3. *Data acquisition hardware* to collect and convert analogue signals to digital format and digital signals for transfer to the computer
4. *A computer* which is loaded with *data acquisition software* to enable analysis and display of the data.

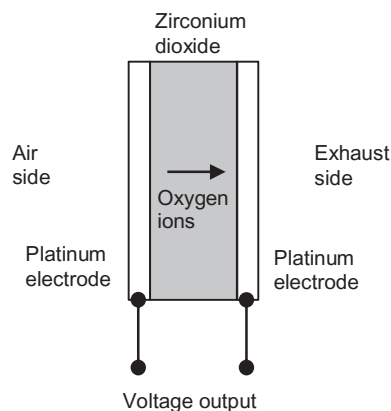


FIGURE 3.11 Oxygen sensor.

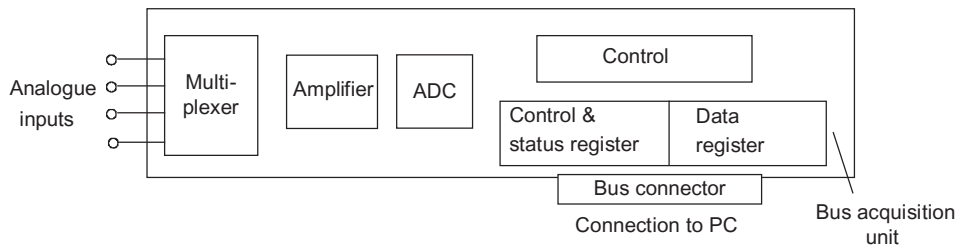


FIGURE 3.12 Basic elements of a digital acquisition card.

Personal computer-based systems frequently use *plug-in boards* for the interface between the computer and the sensors. These boards carry the components for the data acquisition so that the computer can store data, carry out data processing, control the data acquisition process and issue signals for use in control systems. The plug-in board is controlled by the computer and the digitised data transferred from the board to the memory of the computer. The processing of the data is then carried out within the computer according to how it is programmed.

Data acquisition boards are available for many computers and offer, on a plug-in board, various combinations of analogue, digital and timing/counting inputs and outputs for interfacing sensors with computers. The board is a printed circuit board that is inserted into an expansion slot in the computer, so connecting it into the computer bus system. Figure 3.12 shows the basic elements of a simple data acquisition board.

Analogue inputs from the sensors are accessed through a multiplexer. A multiplexer is essentially an electronic switching device which enables a number of inputs to be sampled and each in turn fed to the rest of the system. A typical board might have 8 or 16 analogue input channels. An analogue-to-digital converter then converts the amplified sampled signal to a digital signal. The control element can be set up to control the multiplexer so that each of the inputs is sequentially sampled or perhaps samples are taken at regular time intervals or perhaps just a single sensor signal is used. A typical board might have a sampling rate of 30,000 to 250,000 samples per second. The other main element is the bus interface element which contains two registers, the control and status register and the data register. The term register is used for memory locations within a microprocessor.

Consider now what happens when the computer is programmed to take a sample of the voltage input from a particular sensor. The computer first activates the board by writing a control word into the control and status register. This word indicates the type of operation that is to be carried out. The result of this is that the multiplexer is switched to the appropriate channel for the sensor concerned and the signal from that sensor passed, after amplification at an instrumentation amplifier, to the analogue-to-digital converter. Typically a board will have 12 or 16 bit resolution. The outcome from the ADC is then passed to the data register and the word in the control and status register changed to indicate that the conversion is complete. Following that signal, the computer sets the address bus to select the address of where the data is stored and then issues the signal on the control bus to read the data and so take it into storage in the computer for processing. In order to avoid the computer having to wait and do nothing while the board carries out the acquisition, an interrupt system can be used whereby the board signals the computer when the acquisition is complete and then the computer microprocessor can interrupt any program it was carrying out and jump from that program to a subroutine which stores its current position in the program, reads the data from the board and then jumps back to its original program at the point where it left it.

With the above system, the acquired data moves from the board to the microprocessor which has to interrupt what it is doing, set the memory address and then direct the data to that address in the memory. The microprocessor thus controls the movement. A faster system is to transfer the acquired data directly from the board to memory without involving the microprocessor. This is termed *direct memory access* (DMA). To carry this out, a DMA controller is connected to the bus. This controller supplies the memory address locations where the data is to be put and so enables the data to be routed direct to memory.

Plug-in data acquisition boards with 12 or 16 bit resolution are suitable for most applications. However, boards with greater resolution are available. Special data acquisition boards are available for example for the direct connection of thermocouples and resistance temperature detectors without the need for external signal conditioning.

3.3.1 Data Acquisition Software

The software for use with a data acquisition system can be specially written for the data acquisition hardware concerned, developed using off-the-shelf software that is provided with the hardware used, or developed using a general package which provides a graphical interface for programming, e.g. LabVIEW. LabVIEW is a programming language and set of subroutines developed by National Instruments for data acquisition and scientific programming. It is a graphical programming language which has each subprogram and program structure represented by icons. All the programming is then done graphically by drawing lines between connection points on the various icons. The resulting picture shows the data flow.

3.3.2 Data Loggers

Because sensors have often to be located some distance from the computer in which the data processing occurs, instead of running long wires from the sensors to the computer with degradation of the signal enroute, data loggers (see Chapter 2.12.3) are often located close to the sensors and provide digital signal outputs on a display. They often can provide feeds to send data to computers over a common bus or transfer to them on a portable memory card.

3.4 TESTING

In order to ensure that when a measurement system is installed it will correctly function, testing is required. Testing a measurement system on installation can be considered to fall into three stages:

1. *Pre-installation testing*

This is the testing of each instrument and element for correct calibration and operation prior to it being installed as part of a measurement system.

2. *Cabling and piping testing*

Cables and/or piping will be used to connect together the elements of measurement systems. The display might, for example, be in a control room. All the instrument cables should be checked for continuity and insulation resistance prior to the connection of any instruments or elements of the system. When the system involves pneumatic lines the testing involves blowing through with clear, dry, air prior to connection and pressure testing to ensure the lines are leak free.

3. *Pre-commissioning*

This involves testing that the measurement system installation is complete, all the instrument and other components are in full operational order when interconnected and all control room panels or displays function.

3.4.1 Maintenance

When the system is in operation, maintenance will be required to ensure it continues to operate correctly.

If you own a car or motorcycle, then you will be involved with maintenance and testing, whether you carry out the procedures yourself or a garage does it for you. The function of maintenance is to keep the car/motorcycle in a serviceable condition to that it can continue to carry out its function of transporting you from one place to another. Maintenance is likely to take two forms. One form is *breakdown or corrective maintenance* in which repairs are only made when the car/motorcycle fails to work. Thus breakdown maintenance might be used with the exhaust with it only being replaced when it fails. The other form is *preventative maintenance* which involves anticipating failure and replacing or adjusting items before failure occurs. Preventative maintenance involves inspection and servicing. The inspection is intended to diagnose impending breakdown so that maintenance can prevent it. Servicing is an attempt to reduce the chance of breakdown occurring. Thus preventative maintenance might be used with the regular replacement of engine oil, whether it needs it or not at that time. Inspection of the brakes might be carried out to ascertain when new brake linings are going to be required so that they can be replaced before they wear out. In carrying out maintenance, testing will be involved. Testing might involve checking the coolant level, brake fluid level, etc. and diagnosis tests in the case of faults to establish where the fault is.

In carrying out the maintenance of a measurement system, the most important aid is the *maintenance manual*. This includes such information as:

1. A description of the measurement system with an explanation of its use.
2. A specification of its performance.
3. Details of the system such as block diagrams illustrating how the elements are linked; photographs, drawings, exploded views, etc. giving the mechanical layout; circuit diagrams of individual elements; etc.
4. Preventative maintenance details, e.g. lubrication, replacement of parts, cleaning of parts and the frequency with which such tasks should be carried out.
5. Breakdown/corrective maintenance details, e.g. methods for dismantling, fault diagnosis procedures, test instruments, test instructions, safety precautions necessary to protect the service staff and precautions to be observed to protect sensitive components. With electrical systems the most commonly used test instruments are multi-range meters, cathode ray oscilloscopes and signal generators to provide suitable test signals for injections into the system.
6. Spare parts list.

Maintenance can involve such activities as:

1. Inspection to determine where potential problems might occur or where problems have occurred. This might involve looking to see if wear has occurred or a liquid level is at the right level.
2. Adjustment, e.g. of contacts to prescribed separations or liquid levels to prescribed values.
3. Replacement, e.g. routine replacement of items as part of preventative maintenance and replacement of worn or defective parts.
4. Cleaning as part of preventative maintenance, e.g. of electrical contacts.
5. Calibration. For example, the calibration of an instrument might drift with time and so recalibration becomes necessary.

A record should be maintained of all maintenance activities, i.e. a maintenance activity log. With preventative maintenance this could take the form of a checklist with items being ticked as they are completed. Such a preventative maintenance record is necessary to ensure that such maintenance is carried out at the requisite times. The maintenance log should also include details of any adjustments made, recalibrations necessary or parts replaced. This can help in the diagnosis of future problems.

3.4.2 Common Faults

The following are some of the commonly encountered tests and maintenance points that can occur with measurement systems.

1. *Sensors*

A test is to substitute a sensor with a new one and see what effect this has on the results given by the measurement system. If the results change then it is likely that the original sensor was faulty. If the results do not change then the sensor was not at fault and the fault is elsewhere in the system. Where a sensor is giving incorrect results it might be because it is not correctly mounted or used under the conditions specified by the manufacturer's data sheet. In the case of electrical sensors their output can be directly measured and checked to see if the correct voltages/currents are given. They can also be checked to see whether there is electrical continuity in connecting wires.

2. *Switches and relays*

A common source of incorrect functioning of mechanical switches and relay is dirt and particles of waste material between the switch contacts. A voltmeter used across a switch should indicate the applied voltage when the contacts are open and very nearly zero when they are closed. If visual inspection of a relay discloses evidence of arcing or contact welding then it might function incorrectly and so should be replaced. If a relay fails to operate checks can be made to see if the correct voltage is across the relay coil and, if the correct voltage is present, then there is electrical continuity within the coil with an ohmmeter.

3. *Hydraulic and pneumatic systems*

A common cause of faults with hydraulic and pneumatic systems is dirt. Small particles of dirt can damage seals, block orifices and cause moving parts to jam. Thus, as part of preventative maintenance, filters need to be regularly checked and cleaned. Also oil should be regularly checked and changed. Testing

with hydraulic and pneumatic systems can involve the measurement of the pressure at a number of points in a system to check that the pressure is the right value. Leaks in hoses, pipes and fittings are common faults. Also, damage to seals can result in hydraulic and pneumatic cylinders leaking, beyond that which is normal, and result in a drop in system pressure.

PROBLEMS

Questions 1–6 have four answer options: A, B, C and D. Choose the correct answer from the answer options.

1. Decide whether each of these statements is True (T) or False (F).

As part of the electronic control system for a car engine, a thermistor is to be used to monitor the air temperature. The signal processing circuit that could be used with the thermistor in order to give an electrical voltage output is:

- i. A Wheatstone bridge.
- ii. A potential divider circuit.

- A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F

2. Decide whether each of these statements is True (T) or False (F).

It is proposed to monitor the exhaust temperature of a diesel engine by using a thermocouple. In order to give an output which is a few volts in size and which is independent of the temperature of the surrounding air temperature, the output from the thermocouple:

- i. Requires amplification.
- ii. Requires cold junction compensation.

- A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F

3. Decide whether each of these statements is True (T) or False (F).

It is proposed to monitor the transmission oil pressure in a car by using a diaphragm pressure gauge with the movement of the diaphragm monitored by means of a linear variable differential transformer (LVDT).

- i. The output from the LVDT will be a resistance change which can be converted into a voltage change by a Wheatstone bridge.
- ii. The input to the LVDT is the displacement of the diaphragm.

- A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F

4. Decide whether each of these statements is True (T) or False (F).

The signal processing needed for a system where the output from a thermocouple is to be fed into a microprocessor/computer includes:

- i. A digital-to-analogue converter.
- ii. Amplification.

- A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F

5. Decide whether each of these statements is True (T) or False (F).

The signal processing needed for a system where the output from an optical encoder is to be fed into a microprocessor/computer includes:

- i. An analogue-to-digital converter.
- ii. A resistance-to-voltage converter.

- A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F
6. Decide whether each of these statements is True (T) or False (F).
The term preventative maintenance is used when:
i. Systems are inspected to diagnose possible points of failure before failure occurs.
ii. Systems are regularly maintained with such things as items being lubricated and cleaned.
A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F
7. A driverless vehicle is being designed for operation in a factory where it has to move along prescribed routes transporting materials between machines. Suggest a system that could be used to direct the vehicle along a route.
8. Identify the requirements of the measurement system and hence possible functional elements that could be used to form such a system for the measurement of:
a. The production of an electrical signal when a package on a conveyor belt has reached a particular position.
b. The air temperature for an electrical meter intended to indicate when the temperature drops below freezing point.
c. The production of an electrical signal which can be displayed on a meter and indicate the height of water in a large storage tank.
9. A data acquisition board has a 12-bit analogue-to-digital converter and is set for input signals in the range 0–10 V with the amplifier gain at 10. What is the resolution in volts?
10. A load cell has a sensitivity of 25 mV/kN and is connected to a digital acquisition board which has a 0–10 V, 12-bit analogue-to-digital converter. What amplifier gain should be used if the cell is to give an output for forces in the range 0.1–10 kN?

4

Control Systems

OUTLINE

4.1 Introduction	85	4.4.5 Antilock Brakes	93
4.2 Control Systems	86	4.4.6 Thickness Control	94
4.2.1 Open- and Closed-Loop Control	87	4.4.7 Control of Liquid Level	95
4.3 Basic Elements	89	4.4.8 Robot Gripper	95
4.3.1 Basic Elements of a Closed-Loop System	89	4.4.9 Machine Tool Control	97
4.4 Case Studies	91	4.4.10 Fluid Flow Control	97
4.4.1 Control of the Speed of Rotation of a Motor Shaft	91	4.5 Discrete-Time Control Systems	98
4.4.2 Control of the Position of a Tool	92	4.6 Digital Control Systems	99
4.4.3 Power Steering	92	4.7 Hierarchical Control	100
4.4.4 Control of Fuel Pressure	93	Problems	101

4.1 INTRODUCTION

The term *automation* is used to describe the automatic operation or control of a process. In modern manufacturing, there is an ever increasing use of automation, e.g. automatically operating machinery, perhaps in a production line with robots, which can be used to produce components with virtually no human intervention. Also, in appliances around the home and in the office there is an ever increasing use of automation. Automation involves carrying out operations in the required sequence and controlling outputs to required values.

The following are some of the key historical points in the development of automation, the first three being concerned with developments in the organisation of manufacturing which permitted the development of automated production:

1. Modern manufacturing began in England in the eighteenth century when the use of water wheels and steam engines meant that it became more efficient to organise work to take place in factories, rather than it occurring in the home of a multitude of small workshops. The impetus was thus provided for the development of machinery.
2. The development of powered machinery in the early 1900s meant improved accuracy in the production of components so that instead of making each individual component to fit a particular product, components were fabricated in identical batches with an accuracy which ensured that they could fit any one of a batch of a product. Think of the problem of a nut and bolt if each nut has to be individually made so that it fitted the bolt and the advantages that are gained by the accuracy of manufacturing nuts and bolts being high enough for any of a batch of nuts to fit a bolt.

3. The idea of production lines followed from this with Henry Ford, in 1909, developing them for the production of motor cars. In such a line, the production process is broken up into a sequence of set tasks with the potential for automating tasks and so developing an automated production line.
4. In the 1920s, developments occurred in the theoretical principles of control systems and the use of feedback for exercising control. A particular task of concern was the development of control systems to steer ships and aircraft automatically.
5. In the 1940s, during the Second World War, developments occurred in the application of control systems to military tasks, e.g. radar tracking and gun control.
6. The development of the analysis and design of feedback amplifiers, e.g. the paper by Bode in 1945 on Network Analysis and Feedback Amplifier design, was instrumental in further developing control system theory.
7. Numerical control was developed in 1952 whereby tool positioning was achieved by a sequence of instructions provided by a program of punched paper tape, these directing the motion of the motors driving the axes of the machine tool. There was no feedback of positional data in these early control systems to indicate whether the tool was in the correct position, the system being open-loop control.
8. The invention of the transistor in 1948 in the United States led to the development of integrated circuits, and, in the 1970s, microprocessors and computers which enabled control systems to be developed which were cheap and able to be used to control a wide range of processes. As a consequence, automation has spread to common everyday processes such as the domestic washing machine and the automatic focusing, automatic exposure, camera.

The automatic control of machines and processes is now a vital part of modern industry. The benefits of such control systems include greater consistency of product, reduced operating costs due to improved utilisation of plant and materials and a reduction in manpower, and greater safety for operating personnel.

This chapter is an introduction to the basic idea of a control system and the elements used.

4.2 CONTROL SYSTEMS

As an illustration of what control systems can do, consider the following:

Control a variable to obtain the required value

1. The required temperature for a room is obtained by setting to the required temperature the room thermostat of a central heating system. This is an example of a control system with the controlled variable being the room temperature.
2. In a bottling plant, the bottles are automatically filled to the required level. The variable being controlled is the liquid level in a bottle and control is exercised to ensure no difference between the required level and that which occurs.
3. A computer-numerical-control (CNC) machine tool is used to automatically machine a workpiece to the required shape, the control system ensuring that there is no difference between the required dimensions and that which occurs.
4. Packets of biscuits moving along a conveyor belt have their weights checked and those that are below the required minimum weight limit are automatically rejected. Control is being exercised over the weight.

Control the sequence of events

5. A belt is used to feed blanks to a pressing machine. As a blank reaches the machine, the belt is stopped, the blank positioned in the machine, the press activated to press the required shape, then the pressed item is ejected from the machine and the entire process repeated. A sequence of operations is being controlled with some operations controlled to occur only if certain conditions are met, e.g. activation of the press if there is a blank in place.
6. The dials on an automatic washing machine are set to indicate that 'whites' are being washed and the machine then goes through the complete washing cycle appropriate to that type of clothing. This is an example of a control system where a controlled sequence of events occurs.

Control whether an event occurs or not

7. The automatic clothes washing machine has a safety lock on the door so that the machine will not operate if the door is open. The control is of the condition which allows the machine to operate.

A *control system* can be thought of as a system which for some particular input or inputs is used to control its output to some particular value (Figure 4.1A), give a particular sequence of events (Figure 4.1B) or give an event if certain conditions are met (Figure 4.1C).

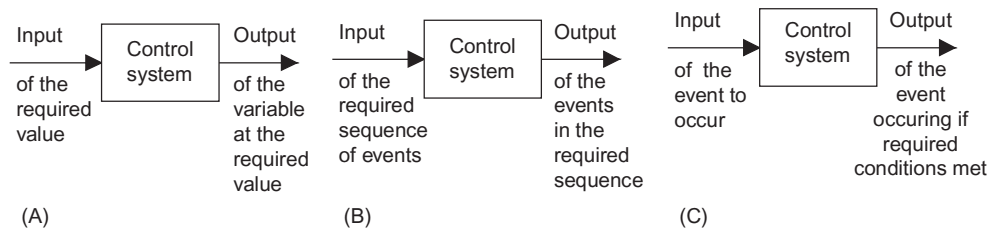


FIGURE 4.1 Control systems (A) to control a variable, (B) to control a sequence of events and (C) control whether an event is to be allowed.

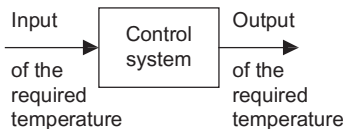


FIGURE 4.2 Central heating system.

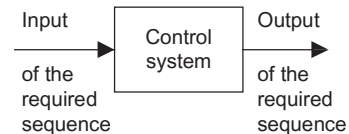


FIGURE 4.3 Clothes washing machine system.

As an example of the type of control system described in Figure 4.1A is a central heating control system which has as its input the temperature required in the house and as its output the house at that temperature (Figure 4.2). The required temperature is set on the thermostat and the control system adjusts the heating furnace to produce an output to give that temperature. The control system is used to control a variable to some set value.

As an example of the type of control system described by Figure 4.1B, a clothes washing machine has as its input a set of instructions as to the sequence of events required to wash the clothes, e.g. fill the drum with cold water, heat the water to 40°C, tumble the clothes for a period of time, empty the drum of water, etc. The manufacturers of the machine have arranged a number of possible sequences which are selected by pressing a button or rotating a dial to select the appropriate sequence for the type of wash required. Thus the input is the information determining the required sequence and the output is the required sequence of events (Figure 4.3). The control system is used to control a sequence of events.

The term single input-single output (SISO) is used for a control system for which there is one control output for one input variable. The term multiple input-multiple output (MIMO) is used when there are several control outputs that are affected by several input variables.

4.2.1 Open- and Closed-Loop Control

Consider two alternative ways of heating a room to some required temperature. In the first instance, there is an electric fire which has a selection switch which allows a 1 kW or a 2 kW heating element to be selected. The decision might be made that to obtain the required temperature it is only necessary to switch on the 1 kW element. The room will heat up and reach a temperature which is determined by the fact the 1 kW element is switched on. The temperature of the room is thus controlled by an initial decision and no further adjustments are made. This is an example of *open-loop control*. Figure 4.4 illustrates this. If there are changes in the conditions, perhaps someone opening a window, no adjustments are made to the heat output from the fire to compensate for the change. There is no information *fed back* to the fire to adjust it and maintain a constant temperature.

Now consider the electric fire heating system with a difference. To obtain the required temperature, a person stands in the room with a thermometer and switches the 1 kW and 2 kW elements on or off, according to the difference between the actual room temperature and the required temperature in order to maintain the temperature of the room at the required temperature. There is a constant comparison of the actual and required temperatures. In this situation there is *feedback*, information being fed back from the output to modify the input to the system. Thus if a window is opened and there is a sudden cold blast of air, the feedback signal changes because the room temperature changes and so is fed back to modify the input to the system. This type of system is called *closed loop*. The input to the heating process depends on the deviation of the actual temperature fed back from the output of the system from the required temperature initially set, the difference between them being determined by a comparison element. In this example, the person with the thermometer is the comparison element. Figure 4.5 illustrates this type of system.

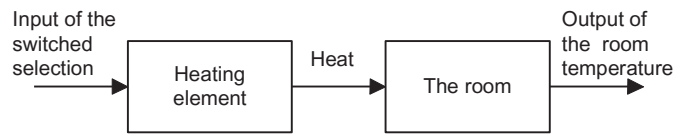


FIGURE 4.4 Open-loop control.

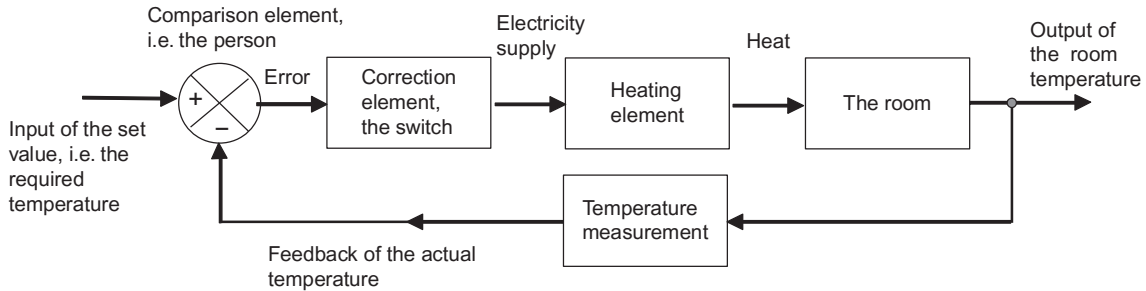


FIGURE 4.5 The electric fire closed-loop system.

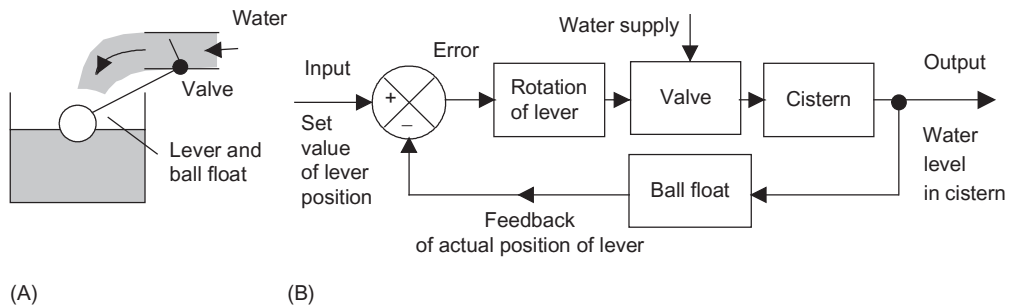


FIGURE 4.6 Ball valve used to control water level in a cistern.

Note that the comparison element in the closed-loop control system is represented by a circular symbol with a + opposite the set value input and a - opposite the feedback signal. The circle represents a summing unit and what we have is the sum.

$$+ \text{ set value} - \text{ feedback value} = \text{error}$$

This difference between the set value and feedback value, the so-called error, is the signal used to control the process. If there is a difference between the signals then the actual output is not the same as the desired output. When the actual output is the same as the required output then there is zero error. Because the feedback signal is subtracted from the set value signal, the system is said to have *negative feedback*.

Consider an example of a ball valve in a cistern used to control the height of the water (Figure 4.6A). The set value for the height of the water in the cistern is determined by the initial setting of the pivot point of the lever and ball float to cut the water off in the valve. When the water level is below that required, the ball moves to a lower level and so the lever opens the valve to allow water into the tank. When the level is at the required level the ball moves the lever to a position which operates the valve to cut off the flow of water into the cistern. Figure 4.6B shows the system when represented as a block diagram.

In an open-loop control system, the output from the system has no effect on the input signal to the plant or process. The output is determined solely by the initial setting. In a closed-loop control system, the output does have an effect on the input signal, modifying it to maintain an output signal at the required value.

Open-loop systems have the advantage of being relatively simple and consequently cheap with generally good reliability. However, they are often inaccurate since there is no correction for errors in the output which might

result from extraneous disturbances. Closed-loop systems have the advantage of being relatively accurate in matching the actual to the required values. They are, however, more complex and so more costly with a greater chance of breakdown as a consequence of the greater number of components.

4.3 BASIC ELEMENTS

Figure 4.7 shows the basic elements of an open-loop control system. The system has three basic elements: control, correction and the process of which a variable is being controlled.

1. Control element

This determines the action to be taken as a result of the input of the required value signal to the system.

2. Correction element

This has an input from the controller and gives an output of some action designed to change the variable being controlled.

3. Process

This is the process of which a variable is being controlled.

With open loop, there is no changing of the control action to account for any disturbances which change the output variable. Figure 4.8 shows an example of an open-loop system which a motor is used to drive a gearbox and so rotate an aerial dish picking up signals from a satellite. With open loop, changes in set point have to be used to accommodate disturbances or changes in loading. So if the aerial is required to track a satellite while a person could keep adjusting the input set value what is really required is a closed-loop system in which feedback about the actual position of the satellite is fed back to automatically adjust the input to the system.

4.3.1 Basic Elements of a Closed-Loop System

Consider the task of driving an automobile along a road. The driver observes the direction of the road and sets the direction by use of the steering wheel. If there are disturbances, such as perhaps traffic, then the driver observes the disturbances and corrects the direction of travel by means of the steering wheel. This is an example of a closed-loop control system and is illustrated in Figure 4.9.

Figure 4.10 shows the general form of a basic closed-loop system.

The following are the functions of the constituent elements:

1. Comparison element

This element compares the required value of the variable being controlled with the measured value of what is being achieved and produces an error signal:

$$\text{error} = \text{required value signal} - \text{measured actual value signal}$$

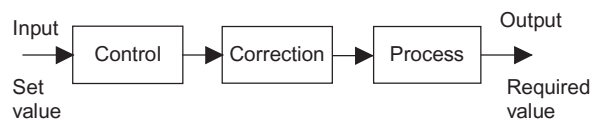


FIGURE 4.7 Basic elements of an open-loop control system.

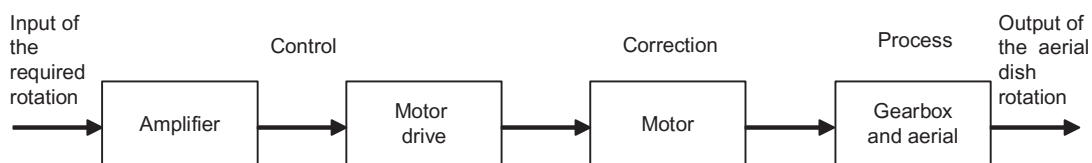


FIGURE 4.8 An open-loop control system for an aerial dish.

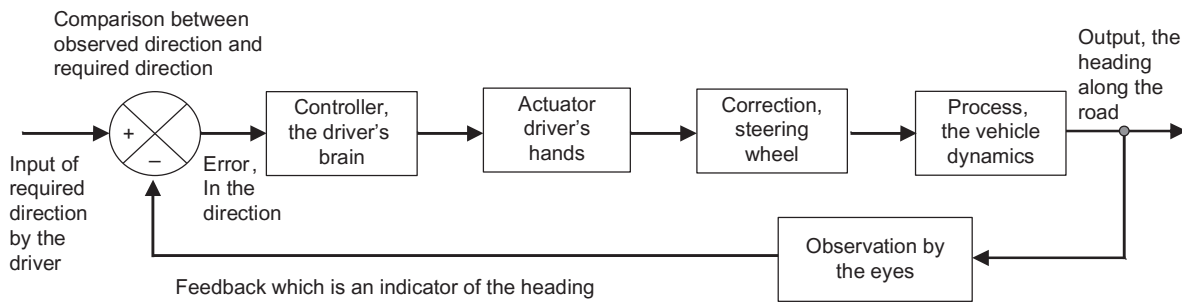


FIGURE 4.9 An example of a closed-loop control system.

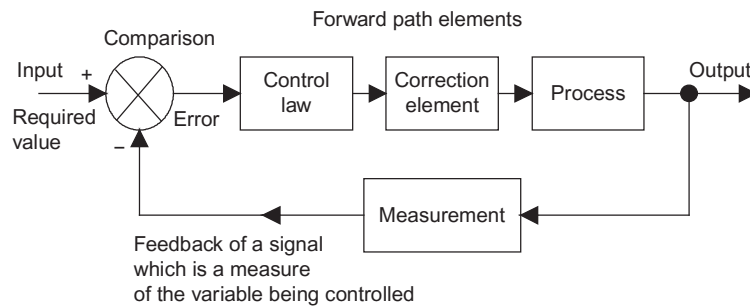


FIGURE 4.10 Basic elements of a closed-loop control system.

Thus if the output is the required value then there is no error and so no signal is fed to initiate control. Only when there is a difference between the required value and the actual values of the variable there will be an error signal and so control action initiated.

2. Control law implementation element

The control law element determines what action to take when an error signal is received. The control law used by the element may be just to supply a signal which switches on or off when there is an error, as in a room thermostat, or perhaps a signal which is proportional to the size of the error so that if the error is small a small control signal is produced and if the error is large a large proportional control signal is produced. Other control laws include *integral mode* where the control signal continues to increase as long as there is an error and *derivative mode* where the control signal is proportional to the rate at which the error is changing.

The term *control unit* or *controller* is often used for the combination of the comparison element, i.e. the error detector, and the control law implementation element. An example of such an element is a differential amplifier which has two inputs, one the set value and one the feedback signal, and any difference between the two is amplified to give the error signal. When there is no difference, there is no resulting error signal.

3. Correction element

The correction element or, as it is often called, the *final control element*, produces a change in the process which aims to correct or change the controlled condition. The term *actuator* is used for the element of a correction unit that provides the power to carry out the control action. Examples of correction elements are directional control valves which are used to switch the direction of flow of a fluid and so control the movement of an actuator such as the movement of a piston in a cylinder. Another example is an electric motor where a signal is used to control the speed of rotation of the motor shaft.

4. Process

The process is the system in which there is a variable that is being controlled, e.g. it might be a room in a house with the variable of its temperature being controlled.

5. Measurement element

The measurement element produces a signal related to the variable condition of the process that is being controlled. For example, it might be a temperature sensor with suitable signal processing.

The following are terms used to describe the various paths through the system taken by signals:

1. *Feedback path*

Feedback is a means whereby a signal related to the actual condition being achieved is fed back to modify the input signal to a process. The feedback is said to be *negative* when the signal which is fed back subtracts from the input value. It is negative feedback that is required to control a system. *Positive feedback* occurs when the signal fed back adds to the input signal.

2. *Forward path*

The term *forward path* is used for the path from the error signal to the output. In Figure 4.10, these forward path elements are the control law element, the correction element and the process element.

The term *process control* is often used to describe the control of variables, e.g. liquid level or the flow of fluids, associated with a process in order to maintain them at some value. Note also that the term *regulator* is sometimes used for a control system for maintaining a plant output constant in the presence of external disturbances. Hence the term *regulator* is sometimes applied to the correction unit.

4.4 CASE STUDIES

The following are examples of closed-loop control systems to illustrate how, despite the different forms of control being exercised, the systems all have the same basic structural elements.

4.4.1 Control of the Speed of Rotation of a Motor Shaft

Consider the motor system shown in Figure 4.11A for the control of the speed of rotation of the motor shaft and its block diagram representation in Figure 4.11B.

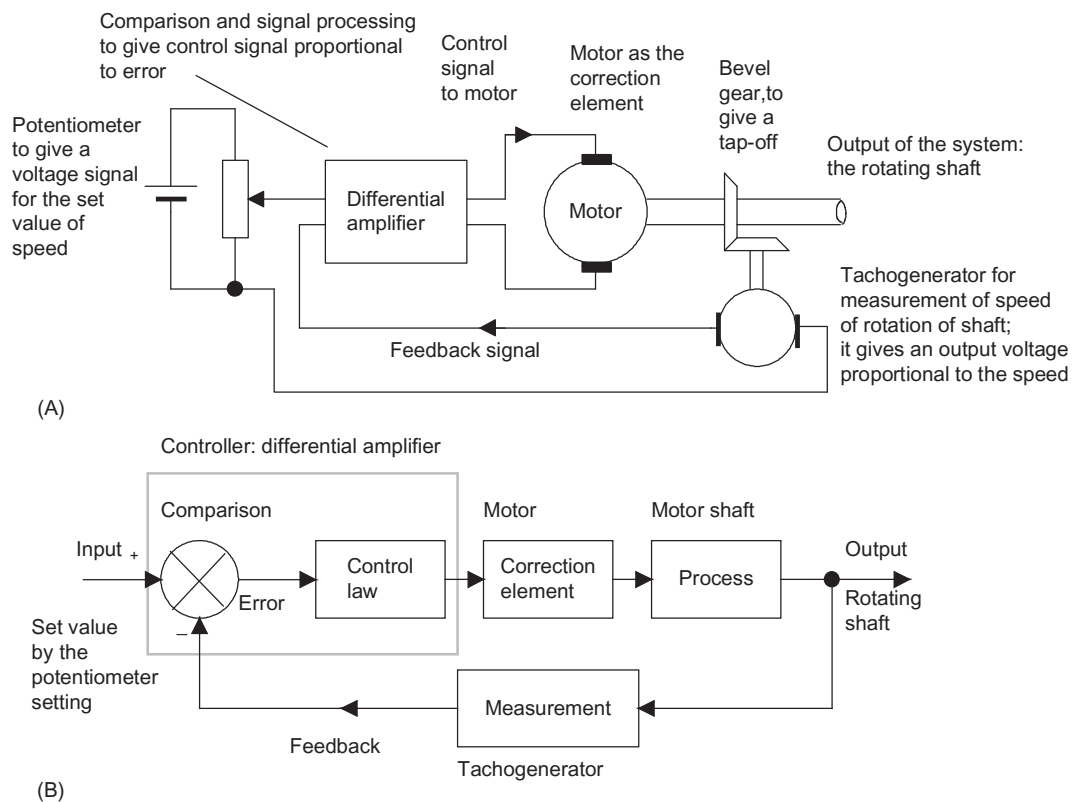


FIGURE 4.11 Control of the speed of rotation of a shaft.

The input of the required speed value is by means of the setting of the position of the movable contact of the potentiometer. This determines what voltage is supplied to the comparison element, i.e. the differential amplifier, as indicative of the required speed of rotation. The differential amplifier produces an amplified output which is proportional to the difference between its two inputs. When there is no difference then the output is zero. The differential amplifier is thus used to both compare and implement the control law. The resulting control signal is then fed to a motor which adjusts the speed of the rotating shaft according to the size of the control signal. The speed of the rotating shaft is measured using a tachogenerator, this being connected to the rotating shaft by means of a pair of bevel gears. The signal from the tachogenerator gives the feedback signal which is then fed back to the differential amplifier.

4.4.2 Control of the Position of a Tool

Figure 4.12 shows a position control system using a belt driven by a stepper motor to control the position of a tool and Figure 4.13 its block diagram representation.

The inputs to the controller are the required position voltage and a voltage giving a measure of the position of the workpiece, this being provided by a potentiometer being used as a position sensor. Because a microprocessor is used as the controller, these signals have to be processed to be digital. The output from the controller is an electrical signal which depends on the error between the required and actual positions and is used, via a drive unit, to operate a stepper motor. Input to the stepper motor causes it to rotate its shaft in steps, so rotating the belt and moving the tool.

4.4.3 Power Steering

Control systems are used to not only maintain some variable constant at a required value but also to control a variable so that it follows the changes required by a variable input signal. An example of such a control system is the power steering system used with a car. This comes into operation whenever the resistance to turning the

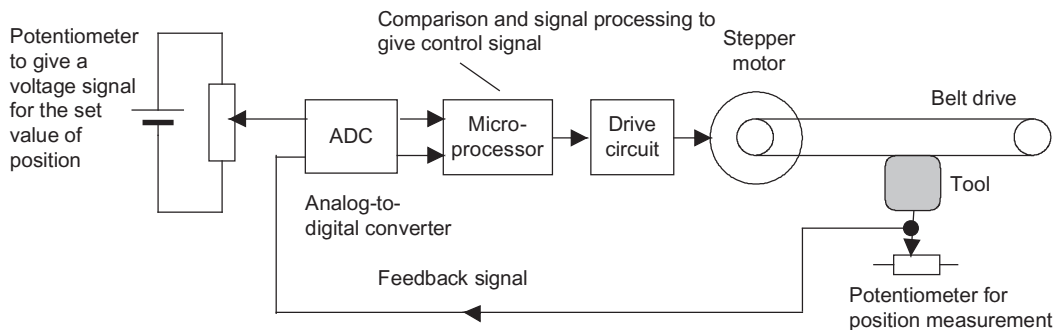


FIGURE 4.12 Position control system.

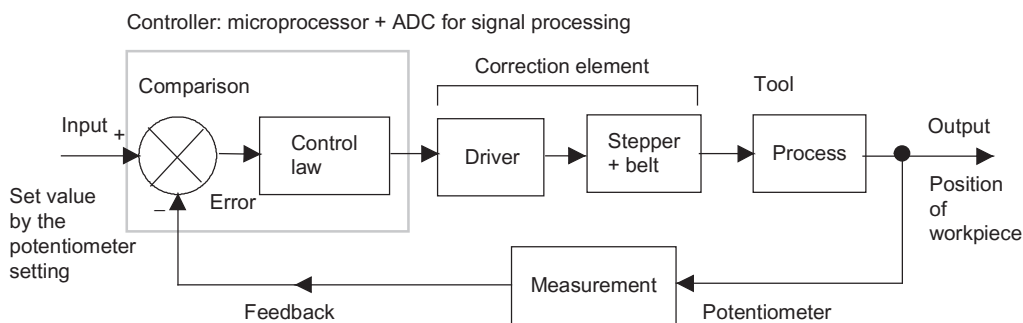


FIGURE 4.13 Position control system.

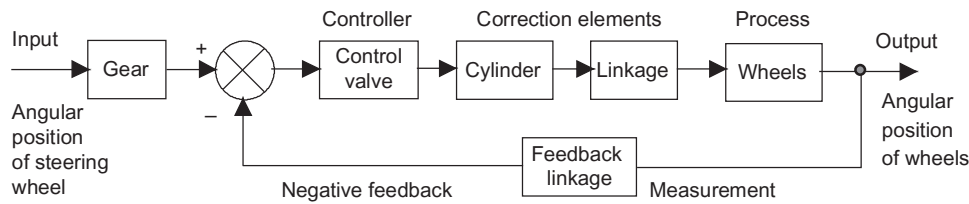


FIGURE 4.14 Power assisted steering.

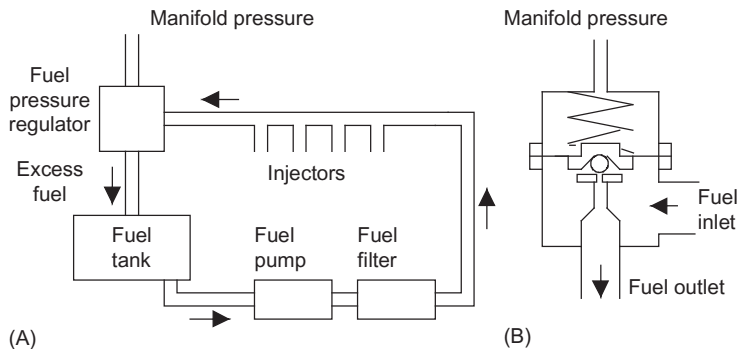


FIGURE 4.15 (A) Fuel supply system and (B) fuel pressure regulator.

steering wheel exceeds a predetermined amount and enables the movement of the wheels to follow the dictates of the angular motion of the steering wheel. The input to the system is the angular position of the steering wheel. This mechanical signal is scaled down by gearing and has subtracted from it a feedback signal representing the actual position of the wheels. This feedback is via a mechanical linkage. Thus when the steering wheel is rotated and there is a difference between its position and the required position of the wheels, there is an error signal. The error signal is used to operate a hydraulic valve and so provide a hydraulic signal to operate a cylinder. The output from the cylinder is then used, via a linkage, to change the position of the wheels. Figure 4.14 shows a block diagram of the system.

4.4.4 Control of Fuel Pressure

The modern car involves many control systems. For example, there is the *engine management system* aimed at controlling the amount of fuel injected into each cylinder and the time at which to fire the spark for ignition. Part of such a system is concerned with delivering a constant pressure of fuel to the ignition system. Figure 4.15A shows the elements involved in such a system. The fuel from the fuel tank is pumped through a filter to the injectors, the pressure in the fuel line being controlled to be 2.5 bar (2.5×0.1 MPa) above the manifold pressure by a regulator valve. Figure 4.15B shows the principles of such a valve. It consists of a diaphragm which presses a ball plug into the flow path of the fuel. The diaphragm has the fuel pressure acting on one side of it and on the other side is the manifold pressure and a spring. If the pressure is too high, the diaphragm moves and opens up the return path to the fuel tank for the excess fuel, so adjusting the fuel pressure to bring it back to the required value.

The pressure control system can be considered to be represented by the closed-loop system shown in Figure 4.16. The set value for the pressure is determined by the spring tension. The comparator and control law is given by the diaphragm and spring. The correction element is the ball in its seating and the measurement is given by the diaphragm.

4.4.5 Antilock Brakes

Another example of a control system used with a car is the *antilock brake system (ABS)*. If one or more of the vehicle's wheels lock, i.e. begins to skid, during braking, then braking distance increases, steering control is lost and tyre wear increases. Antilock brakes are designed to eliminate such locking. The system is essentially a

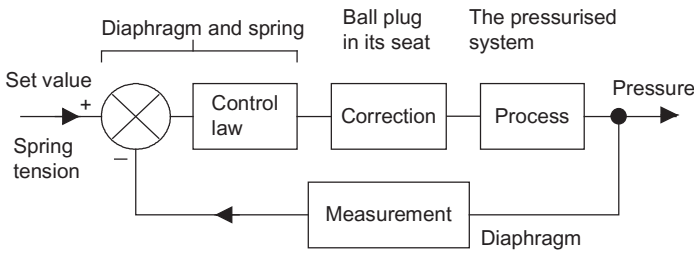


FIGURE 4.16 Fuel supply control system.

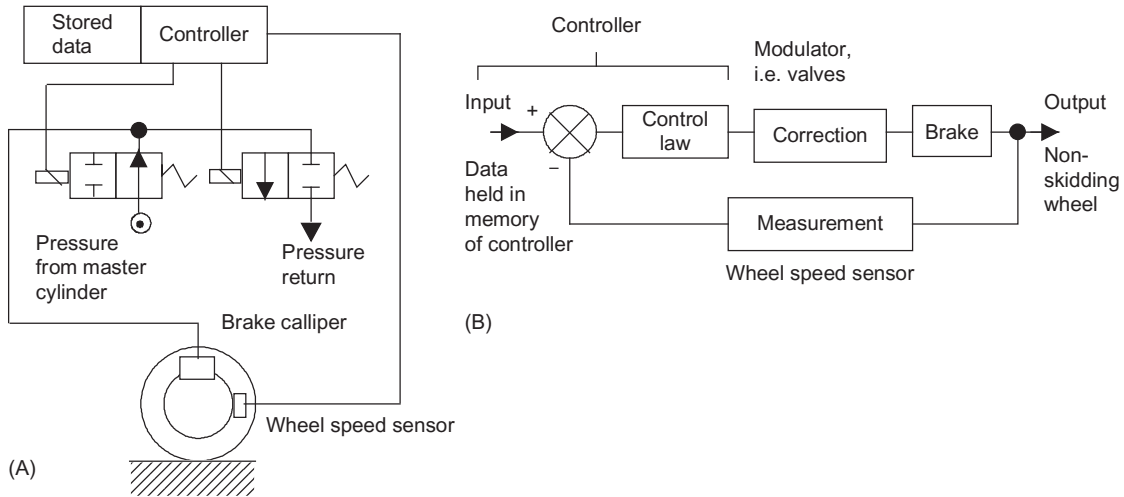


FIGURE 4.17 Antilock brakes: (A) schematic diagram and (B) block form of the control system.

control system which adjusts the pressure applied to the brakes so that locking does not occur. This requires continuous monitoring of the wheels and adjustments to the pressure to ensure that, under the conditions prevailing, locking does not occur. Figure 4.17 shows the principles of such a system.

The two valves used to control the pressure are solenoid-operated directional-control valves, generally both valves being combined in a component termed the modulator. When the driver presses the brake pedal, a piston moves in a master cylinder and pressurises the hydraulic fluid. This pressure causes the brake calliper to operate and the brakes to be applied. The speed of the wheel is monitored by means of a sensor. When the wheel locks, its speed changes abruptly and so the feedback signal from the sensor changes. This feedback signal is fed into the controller where it is compared with what signal might be expected on the basis of data stored in the controller memory. The controller can then supply output signals which operate the valves and so adjust the pressure applied to the brake.

4.4.6 Thickness Control

As an illustration of a *process control system*, Figure 4.18 shows the type of system that might be used to control the *thickness of sheet* produced by rollers and Figure 4.19 shows the block diagram description.

The thickness of the sheet is monitored by a sensor such as a linear variable differential transformer (LVDT). The position of the LVDT probe is set so that when the required thickness sheet is produced, there is no output from the LVDT. The LVDT produces an alternating current output, the amplitude of which is proportional to the error. This is then converted to a d.c. error signal which is fed to an amplifier. The amplified signal is then used to control the speed of a d.c. motor, generally being used to vary the armature current. The rotation of the shaft of the motor is likely to be geared down and then used to rotate a screw which alters the position of the upper roll, hence changing the thickness of the sheet produced.

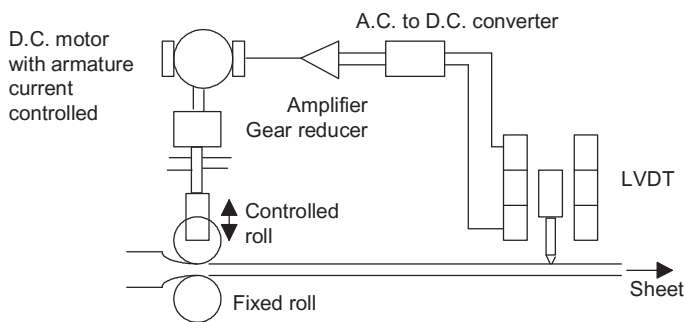


FIGURE 4.18 Sheet thickness control system.

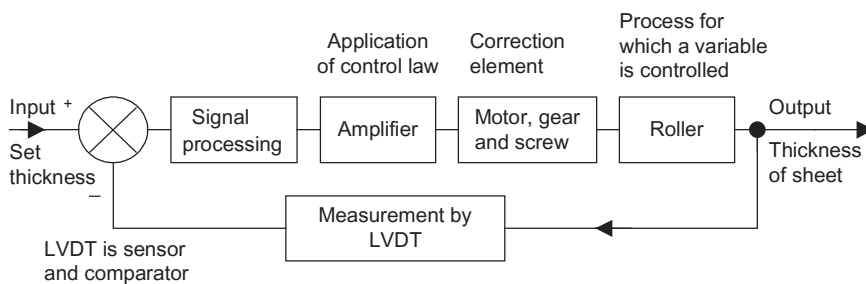


FIGURE 4.19 Sheet thickness control system.

4.4.7 Control of Liquid Level

Figure 4.20 shows a control system used to control the level of liquid in a tank using a float-operated pneumatic controller and Figure 4.21 shows a block diagram of the system.

When the level of the liquid in the tank is at the required level and the inflow and outflows are equal, then the controller valves are both closed. If there is a decrease in the outflow of liquid from the tank, the level rises and so the float rises. This causes point P to move upwards. When this happens, the valve connected to the air supply opens and the air pressure in the system increases. This causes a downward movement of the diaphragm in the flow control valve and hence a downward movement of the valve stem and the valve plug. This then results in the inflow of liquid into the tank being reduced. The increase in the air pressure in the controller chamber causes the bellows to become compressed and move that end of the linkage downwards. This eventually closes off the valve so that the flow control valve is held at the new pressure and hence the new flow rate.

If there is an increase in the outflow of liquid from the tank, the level falls and so the float falls. This causes point P to move downwards. When this happens, the valve connected to the vent opens and the air pressure in the system decreases. This causes an upward movement of the diaphragm in the flow control valve and hence an upward movement of the valve stem and the valve plug. This then results in the inflow of liquid into the tank being increased. The bellows react to this new air pressure by moving its end of the linkage, eventually closing off the exhaust and so holding the air pressure at the new value and the flow control valve at its new flow rate setting.

4.4.8 Robot Gripper

The term *robot* is used for a machine which is a reprogrammable multi-function manipulator designed to move tools, parts, materials, etc. through variable programmed motions in order to carry out specified tasks. Here just one aspect will be considered, the gripper used by a robot at the end of its arm to grip objects. A common form of gripper is a device which has 'fingers' or 'jaws'. The gripping action then involves these clamping on the object. Figure 4.22 shows one form such a gripper can take if two gripper fingers are to close on a parallel-sided object. When the input rod moves towards the fingers they pivot about their pivots and move closer together. When the rod moves outwards, the fingers move further apart. Such motion needs to be controlled so that the grip exerted by the fingers on an object is just sufficient to grip it, too little grip and the object will fall out of the grasp of the gripper and too great might result in the object being crushed or otherwise deformed. Thus there needs to be feedback of the forces involved at contact between the gripper and the object. Figure 4.23 shows the type of closed-loop control system involved.

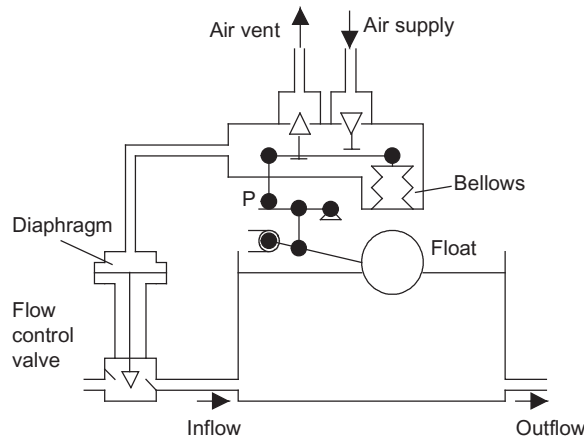


FIGURE 4.20 Level control system.

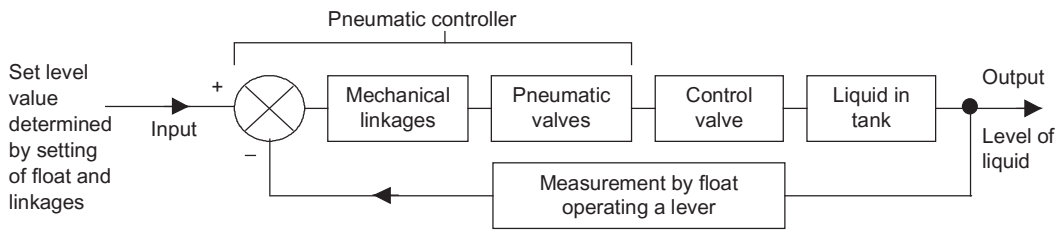


FIGURE 4.21 Level control system.

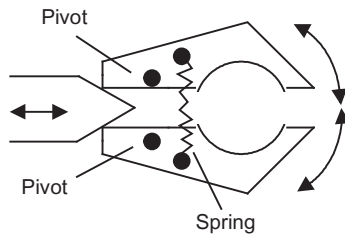


FIGURE 4.22 An example of a gripper.

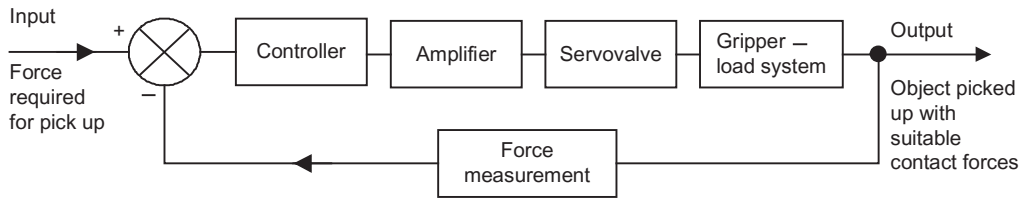


FIGURE 4.23 Gripper control system.

The drive system used to operate the gripper can be electrical, pneumatic or hydraulic. Pneumatic drives are very widely used for grippers because they are cheap to install, the system is easily maintained and the air supply is easily linked to the gripper. Where larger loads are involved, hydraulic drives can be used. Sensors that might be used for measurement of the forces involved are piezoelectric sensors or strain gauges. Thus when

strain gauges are stuck to the surface of the gripper and forces applied to a gripper, the strain gauges will be subject to strain and give a resistance change related to the forces experienced by the gripper when in contact with the object being picked up.

The robot arm with gripper is also likely to have further control loops to indicate when it is in the right position to grip an object. Thus the gripper might have a control loop to indicate when it is in contact with the object being picked up; the gripper can then be actuated and the force control system can come into operation to control the grasp. The sensor used for such a control loop might be a microswitch which is actuated by a lever, roller or probe coming into contact with the object.

4.4.9 Machine Tool Control

Machine tool control systems are used to control the position of a tool or workpiece and the operation of the tool during a machining operation. Figure 4.24 shows a block diagram of the basic elements of a closed-loop system involving the continuous monitoring of the movement and position of the work tables on which tools are mounted while the workpiece is being machined.

The amount and direction of movement required in order to produce the required size and form of workpiece is the input to the system, this being a program of instructions fed into a memory which then supplies the information as required. The sequence of steps involved is then:

1. An input signal is fed from the memory store.
2. The error between this input and the actual movement and position of the work table is the error signal which is used to apply the correction. This may be an electric motor to control the movement of the work table. The work table then moves to reduce the error so that the actual position equals the required position.
3. The next input signal is fed from the memory store.
4. Step 2 is then repeated.
5. The next input signal is fed from the memory store and so on.

4.4.10 Fluid Flow Control

Figure 4.25 shows the elements of a control system used to control the rate of flow of liquid to some required value, regardless of any fluctuations in supply pressure or back pressure. Figure 4.26 shows a block diagram of the system.

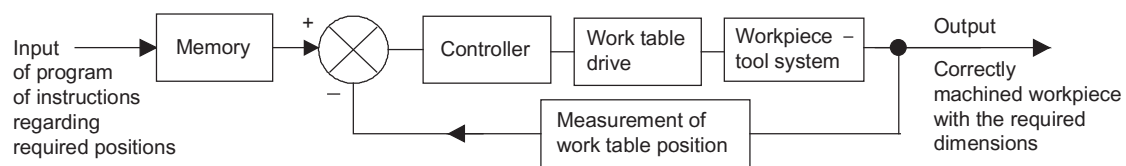


FIGURE 4.24 Closed-loop machine tool control system.

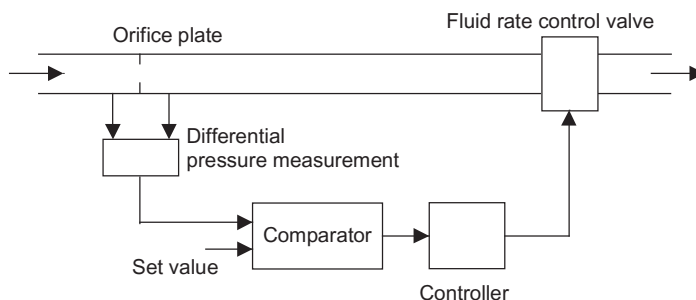


FIGURE 4.25 Control system for controlling the rate of flow of a liquid.

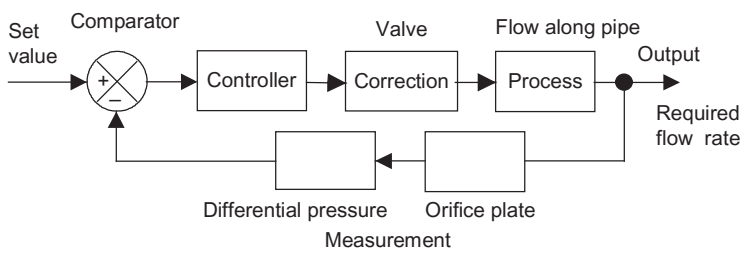


FIGURE 4.26 Control system for controlling the rate of flow of a liquid.

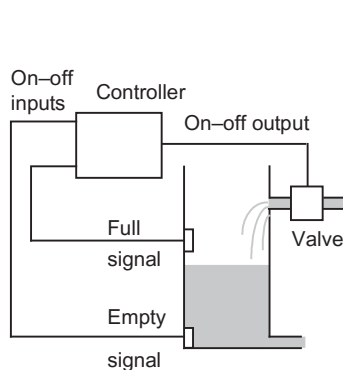


FIGURE 4.27 Discrete-event control with the controller switching the valve open when empty signal received and closed when the full signal.

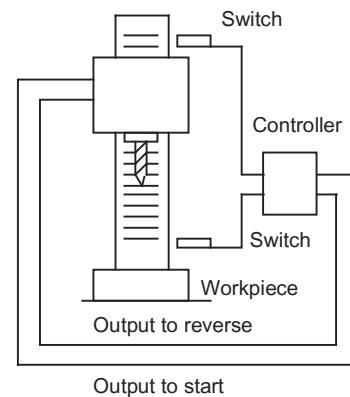


FIGURE 4.28 An automatic drill.

4.5 DISCRETE-TIME CONTROL SYSTEMS

Discrete-time control systems are control systems in which one or more inputs can change only at discrete instants of time, i.e. the inputs are effectively on–off signals and so in digital form rather than the analog form which has been discussed earlier in this chapter. This form of control is often called *sequential control*. It describes control systems involving logic control functions, e.g. is there or is there not a signal from sensor A or perhaps an AND logic system where the issue is whether there is an input from sensor A and an input from sensor B, in order to determine whether to give an output and so switch some device on or off.

As a simple illustration of sequential control, consider the automatic kettle. When the kettle is switched on, the water heats up and continues heating until a sensor indicates that boiling is occurring. The sensor is just giving an on–off signal. The kettle then automatically switches off. The heating element of the kettle is not continuously controlled but only given start and stop signals.

As another example, the filling of a container with water might have a sensor at the bottom to register when the container is empty and give an input to the controller to switch the water flow on and a sensor at the top to register when the container is full and give an input to the controller to switch off the flow of water (Figure 4.27). We have two sensors giving on–off signals in order to obtain the required sequence of events.

As an illustration of the type of control that might be used with a machine, consider the system for a drill which is required to automatically drill a hole in a workpiece when it is placed on the work table (Figure 4.28). A switch sensor can be used to detect when the workpiece is on the work table, such a sensor being an on–off sensor. This then gives an on input signal to the controller and it then gives an output signal to actuate a motor to lower the drill head and start drilling. When the drill reaches the full extent of its movement in the workpiece, the drill head triggers another switch sensor. This provides an on input to the controller and it then reverse the direction of rotation of the drill head motor and the drill retracts.

See Chapter 7 for a further discussion of sequential control systems and how they can be realised.

4.6 DIGITAL CONTROL SYSTEMS

With analog control systems all the signals are analogs, i.e. a scaled version, of the quantities they represent. Digital signals, however, are sequences of pulses, i.e. on–off signals, with the value of the quantity being represented by the sequence of on–off signals. Most of the signals being controlled are analog and thus it is necessary with a digital control system to convert analog inputs into digital signals for the controller and then the digital outputs from the controller to analog for the process being controlled. Thus analog-to-digital converters (ADCs) (see Chapter 2.9.5) and digital-to-analog converters (DACs) (see Chapter 2.9.6) are used. Figure 4.29 shows the basic form of a closed-loop digital control system.

Digital controllers can be microcontrollers or computers. A microcontroller is an integration of a micro-processor with memory, input/output interfaces and other peripherals such as timers on a single chip (see Chapter 2.9.7). The microcontroller effectively incorporates the comparison element, the controller, the DAC and the ADC. At some particular instant the microcontroller samples its input signals. There is then an input of the required analog value to the microcontroller and the measured analog value of the output. It then carries out its program and gives an analog output to the correction element. The microcontroller then repeats its program for the next sample of signals. The program followed by the microcontroller is thus:

Take a sample of the signals at its inputs, i.e. read the required value input and read the actual measured value of the output, i.e. the feedback signal, at its ADC input port
 Calculate the error signal
 Calculate the required controller output
 Send the controller output to its DAC output port
 Wait for the next sampling interval
 Repeat the process.

Digital control has advantages over analog control in that digital operations can be easily controlled by a program, i.e. a piece of software, information storage is easier, accuracy can be greater and digital circuits are less affected by noise.

As an illustration of a digital control system, Figure 4.30 shows the system for the control of the speed of rotation of a motor shaft which was represented by an analog system in Section 4.4.1.

The control systems so far considered are essentially concerned with a single loop for which there is a single set point, a single controlled variable and a single actuator. However, in many control situations there are more than one variable being controlled. Modern automobiles have engine management microcontrollers which are used to exercise control over the large number of systems in such automobiles. The engine control unit aims to ensure that the engine operates at optimal conditions at all times and thus exercises control over such items as fuel injection, spark timing, idle speed and anti-knock. It does this by taking inputs from a number of sensors, interpreting the values and then giving output to adjust the operation of the engine. There are also systems for such items as braking and traction control, suspension control, cruise control, air conditioning, air-bag systems and security systems.

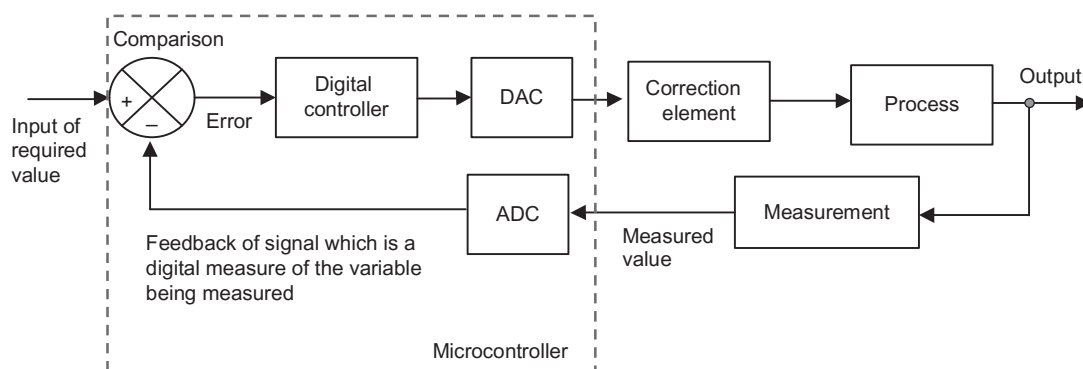


FIGURE 4.29 The basic elements of a digital closed-loop control system.

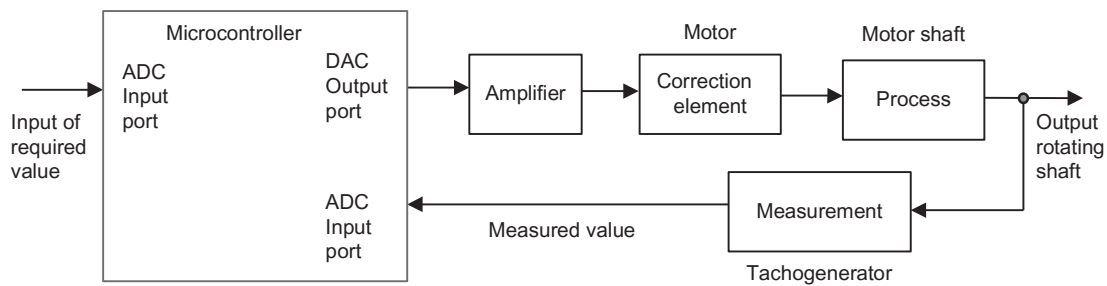


FIGURE 4.30 Control of the speed of rotation of a shaft.

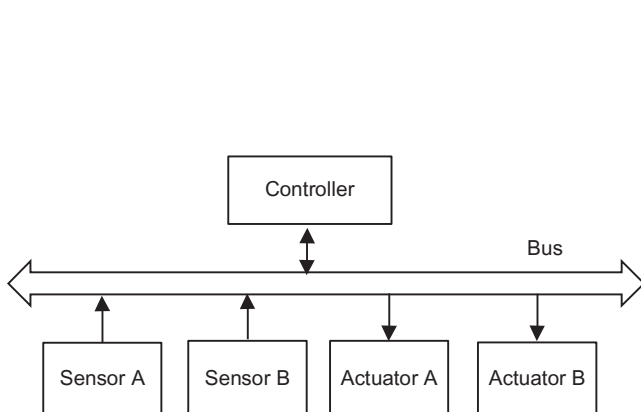


FIGURE 4.31 A bus system.

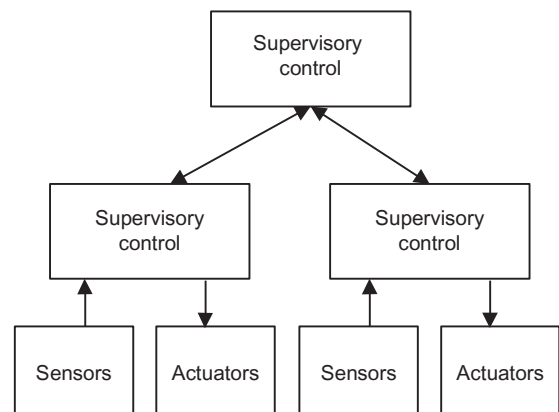


FIGURE 4.32 Hierarchical control.

While with control systems having more than one variable, more than one sensor and more than one actuator it would be possible to wire each situation independent of the other with its own controller, there are advantages in using a single controller to control the inputs for more than one sensor. This does allow more than one variable to be taken into account in arriving at the optimum outputs for sensors but also it is possible to simplify the wiring by using a bus connector so that each sensor and actuator is connected locally into a common connection termed a *bus* (Figure 4.31). Sensors and actuators are then in communication with the bus controller via a common connection, each such sensor and actuator being uniquely identified to the controller so that information can be supplied from supervisory control to specific sensors and specific actuators and receive information back from them.

The term *fieldbus* is used for such a bus system that conforms to the IEC 61158 specification, there being a number of such bus systems. The standard defines the physical means by which the buses can be realised and the means by which devices connected to a bus can communicate with a controller and receive control instructions. Fieldbus are bi-directional, digital serial networks with the data being put on bit by bit along a single connector. In process control, the main buses used are the FOUNDATION Fieldbus and PROFIBUS PA. FOUNDATION Fieldbus is a digital, serial, two-way communication system widely used as the network in process industries. PROFIBUS was designed specifically for use with programmable logic controllers in control networks (see Chapter 7). The bus most used in automobiles is the CAN bus (see Chapter 13 for more discussion of bus systems).

4.7 HIERARCHICAL CONTROL

The term *hierarchical control* is used for the form of control systems where we have one control system being used to control other control systems to give a pyramid of control (Figure 4.32).

Thus we might have one control system controlling one aspect of a manufacturing process and another controlling a different aspect, the two control systems then being supervised. Indeed there may be a number of such supervised controls with the supervisory controls in turn being supervised by a management controller. Such supervision might be used to change the set points of control systems in order to cope with some change in circumstances (see Chapter 13 for further discussion).

PROBLEMS

Questions 1–4 have four answer options: A, B, C and D. Choose the correct answer from the answer options.

1. Decide whether each of these statements is True (T) or False (F).
An open-loop control system:
 - i. Has negative feedback.
 - ii. Responds to changes in conditions.

A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F
2. Decide whether each of these statements is True (T) or False (F).
A closed-loop control system:
 - i. Has negative feedback.
 - ii. Responds to changes in conditions.

A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F
3. Decide whether each of these statements is True (T) or False (F).
A closed-loop control system:
 - i. Has a measurement system which gives feedback of a signal which is a measure of the variable being controlled.
 - ii. Has a controller which has an input based on the difference between the set value and the fed back value for the variable being controlled.

A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F
4. Decide whether each of these statements is True (T) or False (F).
Negative feedback with a control system is when:
 - i. The fed back signal is added to the input signal.
 - ii. The fed back signal is a measure of the output value of the control system.

A. (i) T (ii) T
B. (i) T (ii) F
C. (i) F (ii) T
D. (i) F (ii) F
5. Suggest the possible form control systems might take for the following situations:
 - a. Controlling the thickness of sheet steel produced by a rolling mill.
 - b. A conveyor belt is to be used to transport packages from a loading machine to a pick-up area. The control system must start the belt when a package is loaded onto the belt, run the belt until the package arrives at the pick-up area, then stop the belt until the package is removed. Then the entire sequence can start again.
 - c. Monitoring breathing in an intensive care unit, sounding an alarm if breathing stops.
 - d. Controlling the amount of a chemical supplied by a hopper into sacks.
 - e. Controlling the volume of water supplied to a tank in order to maintain a constant level.
 - f. Controlling the illumination of the road in front of a car by switching on the lights.
 - g. Controlling the temperature in an automobile by the driver manually selecting the heater controls, switching between them as necessary to obtain the required temperature.

6. Figure 4.33 shows two systems that might be used to control the temperature of a room. Explain how each operates.

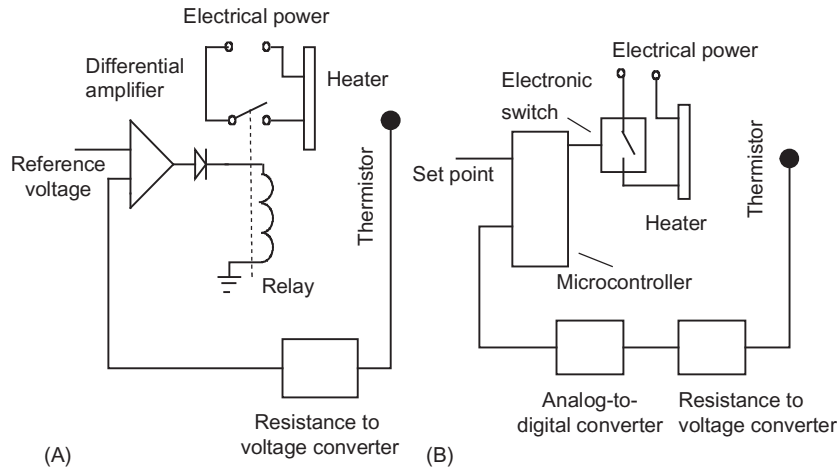


FIGURE 4.33 Problem 6.

7. Draw a block diagram of a domestic central heating system which has the following elements:
 - a. A thermostat which has a dial which is set to the required temperature and has an input of the actual temperature in the house and which operates as a switch and gives an output electrical signal which is either on or off.
 - b. A solenoid valve which has the input of the electrical signal from the thermostat and controls the flow of oil to the central heating furnace.
 - c. A heating furnace where the input is the flow of oil and the output is heat to the rooms via water flowing through radiators in the house.
 - d. The rooms in the house where the input is the heat from the radiators and the output is the temperature in the rooms.
8. Figure 4.34 shows a water level control system. Identify the basic functional elements of the system.

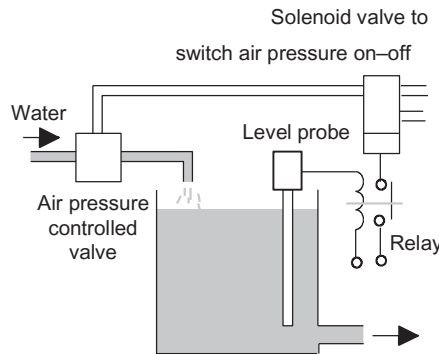


FIGURE 4.34 Problem 8.

9. Draw a block diagram for a negative feedback system that might be used to control the level of light in a room to a constant value.
10. Draw block diagrams which can be used to present the operation of a toaster when it is (a) an open-loop system and (b) a closed-loop system.

5

Process Controllers

O U T L I N E

5.1 Introduction	103		
5.1.1 Direct and Reverse Actions	104	5.7.1 Process Reaction Tuning Method	118
5.1.2 Dead Time	104	5.7.2 Ultimate Cycle Tuning Method	120
5.1.3 Capacitance	104	5.7.3 Quarter Amplitude Decay	121
5.2 On–Off Control	104	5.7.4 Lambda Tuning	121
5.2.1 Relays	106	5.7.5 Software Tools	121
5.3 Proportional Control	107	5.7.6 Adaptive Controllers	122
5.3.1 Proportional Band	107	5.8 Digital Systems	122
5.3.2 Limitations of Proportional Control	109	5.8.1 Embedded Systems	123
5.4 Derivative Control	110	5.9 Fuzzy Logic Control	124
5.4.1 PD Control	110	5.9.1 Fuzzy Logic	124
5.5 Integral Control	112	5.9.2 Fuzzy Logic Control Systems	126
5.5.1 PI Control	112	5.9.3 Fuzzy Logic Controller	127
5.6 PID Control	114	5.9.4 Fuzzy Logic Tuning of PID Controllers	130
5.6.1 PID Process Controller	116	5.10 Neural Networks	130
5.7 Tuning	117	5.10.1 Neural Networks for Control	132
		Problems	133

5.1 INTRODUCTION

Process controllers are control system components which basically have an input of the error signal, i.e. the difference between the required value signal and the feedback signal, and an output of a signal to modify the system output. The ways in which such controllers react to error changes are termed the *control laws*, or more often, the *control modes*. The simplest form of controller is an *on–off device* which switches on some correcting device when there is an error and switches it off when the error ceases. However, such a method of control has limitations and often more sophisticated controllers are used. While there are many ways a controller could be designed to react to an error signal, a form of controller which can give satisfactory control in a wide number of situations is the *three-term* or *PID controller*. The three basic control modes are *proportional* (P), *integral* (I) and *derivative* (D). The three-term controller is a combination of all three modes, the principles of these methods of control being introduced in this chapter. Newer modes of control which are becoming increasingly used are fuzzy logic control and neural networks, and an introduction to these is also given in this chapter.

5.1.1 Direct and Reverse Actions

In discussing the elements of a control system, the term *direct action* is used for an element that for an increase in its input gives an increase in its output, e.g. a domestic central heating furnace where an increase in the controlled input to the system results in an increase in temperature. The term *reverse action* is used when an increase in input gives a decrease in output, e.g. an air conditioner where an increase in the energy input to it results in a decrease in temperature.

5.1.2 Dead Time

In any control system with feedback the system cannot respond instantly to any change and thus there are delays while the system takes time to accommodate the change. Such delays are referred to as *dead time* or *lags*. For example, in the control of the temperature in a room by means of a central heating system, if a window is suddenly opened and the temperature drops or the thermostat is suddenly set to a new value, a lag will occur before the control system responds, switches on the heater and gets the temperature back to its set value. A source of dead time in a control system is the response time of the measurement sensor. Thus, for a system using a temperature sensor, a resistance temperature detector (RTD) has a slower response time than a thermocouple. A thermocouple has typically a response time of about 0.5 s while an RTD is a few seconds.

Transfer delays are a common event with control systems where flow is concerned, e.g. water flowing along a pipe from point A where the control valve is to point B where the rate of flow is required and monitored. Any change made at some point A will take some time before its effects are apparent at a point B, the *time delay* depending on the distance between A and B and the rate of flow between them. The term *distance–velocity lag* is sometimes used to describe such delays. An example of a transfer delay is where a hopper is loading material onto a conveyor belt moving with a velocity v . The rate at which the material leaves the hopper is controlled by a valve with feedback from a weight sensor. If the weight of deposited material per unit length of belt is monitored a distance L from the hopper discharge point, then there will be a time delay of L/v in the control system.

5.1.3 Capacitance

In the level control of the level of water in a tank, an important attribute of the system is its *capacitance*. If we have water leaving the tank and the control signal used to determine the rate of flow of water into the tank, then the greater the surface area of the water in the tank the longer it will take the controlled inflow of water to respond and restore a drop in level. We talk of the system having capacitance and the greater the capacitance the longer it takes to react to changes. If the capacitance was decreased then the system would react quicker to make the changes necessary to restore the required level. As an illustration of the effect of capacitance consider an electrical circuit with a capacitor C and resistor R in series. When a voltage V is switched on, the voltage across the capacitor increases with time until it eventually reaches a steady state value. The initial rate of change of voltage across the capacitor is V/RC . Thus the bigger the capacitance the smaller the rate of change and so the longer it takes for the capacitor to become fully charged.

Another example is a domestic heating system controlled by a thermostat. The larger the space being heated the longer it will take the controller to respond and restore a drop in temperature. Again we talk of the capacitance of the system.

5.2 ON–OFF CONTROL

With on–off control, the controller is essentially a switch which is activated by the error signal and supplies just an on–off correcting signal (Figure 5.1). The controller output has just two possible values, equivalent to on and off. For this reason the controller is sometimes termed a *two-step controller*.

An example of such a controller is the bimetallic thermostat (Figure 5.2A) used with a simple temperature control system. The bimetallic strip consists of two strips of different metals, e.g. steel and copper or steel and brass, which expand at different rates. The two strips are joined together throughout their length and so, as a consequence, a rise in temperature causes one strip to expand more than the other and as they are joined together the only way this can occur is if they curve. This curving can be used to close an electrical switch. Thus, if the actual temperature is above the required temperature, the bimetallic strip is in an off position and the heater is switched off; if the actual temperature is below the required temperature, the bimetallic strip moves into the on position and the heater is switched on. The controller output is thus just on or off and so the correcting signal on or off.

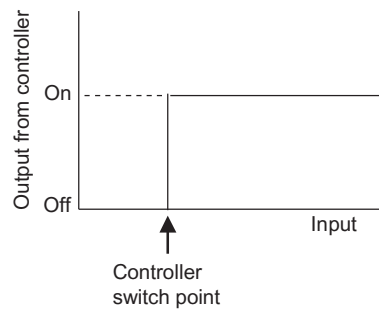


FIGURE 5.1 On-off control.

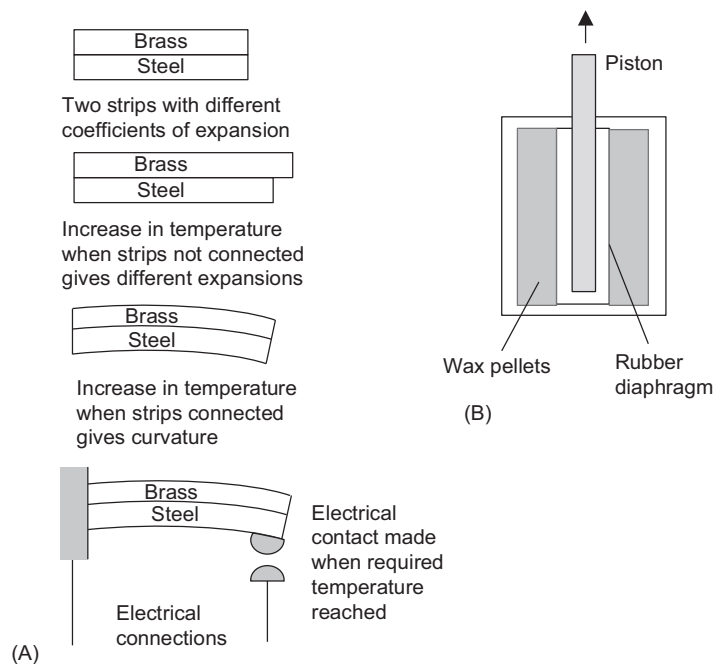


FIGURE 5.2 The basic principles of (A) a metallic thermostat and (B) a wax thermostat.

Another commonly used on-off controller is the wax thermostat (Figure 5.2B). This uses a plug of wax inside a sealed chamber. When the temperature increases, the wax melts and as a result expands greatly and is able to push a rod out from the chamber and open a valve. Such thermostats are used to control the engine cooling system in an automobile. They are also used in the thermostat valves fitted to central heating radiators.

Because the control action is discontinuous and there are time lags in the system, oscillations, i.e. cycling, of the controlled variable occur about the required condition. Thus, with temperature control using the bimetallic thermostat, when the room temperature drops below the required level there is a significant time before the heater begins to have an effect on the room temperature and, in the meantime, the temperature has fallen even more. When the temperature rises to the required temperature, since time elapses before the control system reacts and switches the heater off and it cools, the room temperature goes beyond the required value. The result is that the room temperature oscillates above and below the required temperature (Figure 5.3).

There is also a problem with the simple on-off system in that when the room temperature is hovering about the set value the thermostat might be reacting to very slight changes in temperature and almost continually switching on or off. Thus, when it is at its set value a slight draught might cause it to operate. This problem can be reduced if the heater is switched on at a lower temperature than the one at which it is switched off (Figure 5.4). The term *dead band* or *neutral zone* is used for the values between the on and off values. For example, if the set value on a thermostat is 20°C , then a dead band might mean it switches on when the temperature falls to 19.5° and off when it is 20.5° . The temperature has thus to change by one degree for the controller to switch the heater on or off and thus smaller changes do not cause the thermostat to switch. A large dead band results in

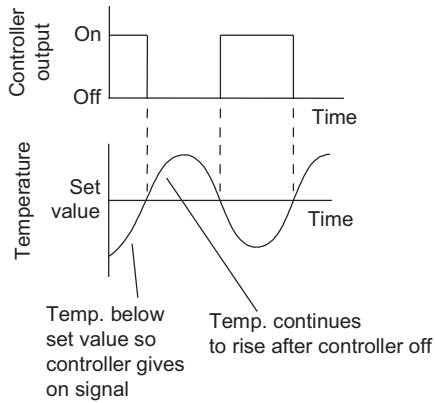


FIGURE 5.3 Fluctuation of temperature about set value.

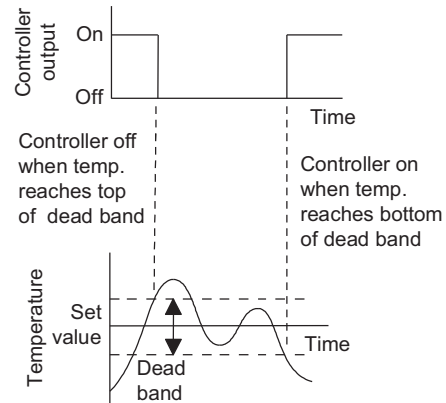


FIGURE 5.4 On-off controller with a dead band.

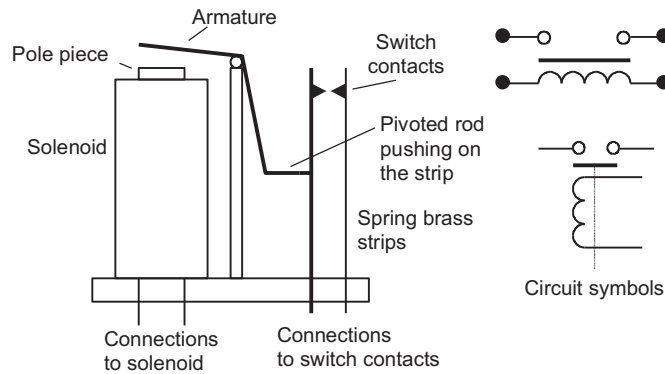


FIGURE 5.5 Relay.

large fluctuations of the temperature about the set temperature; a small dead band will result in an increased frequency of switching. The bimetallic thermostat shown in Figure 5.2 has a permanent magnet on one switch contact and a small piece of soft iron on the other; this has the effect of producing a small dead band in that, when the switch is closed, a significant rise in temperature is needed for the bimetallic element to produce sufficient force to separate the contacts.

On-off control is not too bad at maintaining a constant value of the variable when the capacitance of the system is very large, e.g. a central heating system heating a large air volume, and so the effect of changes in, say, a heater output results in slow changes in the variable. It also involves simple devices and so is fairly cheap. On-off control can be implemented by mechanical switches such as bimetallic strips or relays, with more rapid switching being achieved with electronic circuits, e.g. thyristors or transistors used to control the speed of a motor. On-off control is simple and inexpensive and is often used where cycling can be reduced to an acceptable level. The normal domestic central heating system has an on-off controller, though a modern one is more likely to be an electronic on-off sensor rather than a bimetallic strip. Another example of on-off control is the control of a car radiator cooling fan using a wax element thermostat.

5.2.1 Relays

A widely used form of on-off controller is a *relay*. Figure 5.5 shows the basic form of an electromagnetic relay. A small current at a low voltage applied to the solenoid produces a magnetic field and so an electromagnet. When the current is high enough, the electromagnet attracts the armature towards the pole piece and in doing so operates the relay contacts. A much larger current can then be switched on. When the current through the solenoid drops below the critical level, the springy nature of the strip on which the contacts are mounted pushes the armature back to the off position. Thus if the error signal is applied to the relay, it trips on when the error reaches a certain size and can then be used to switch on a much larger current in a correction element such as a heater or a motor.

5.3 PROPORTIONAL CONTROL

With the on–off method of control, the controller output is either an on or an off signal and so the output is not related to the size of the error. With *proportional control* the size of the controller output is proportional to the size of the error (Figure 5.6), i.e. the controller input. Thus we have: controller output \propto controller input. We can write this as

$$\text{controller output} = K_p \times \text{controller input}$$

where K_p is a constant called the gain. This means the correction element of the control system will have an input of a signal which is proportional to the size of the correction required. Thus, for example, if we have a heater being controlled to give a set temperature then as the temperature approaches the set point the power supplied to the heater will be reduced and so slow down the heating so that it will not result in the temperature overshooting the set value and end up oscillating about the set point.

The float method of controlling the level of water in a cistern (Figure 5.7) is an example of the use of a proportional controller, the control mode being determined by the lever.

The output is proportional to the error, the gain being x/y . The error signal is the input to the ball end of the lever, the output is the movement of the other end of the lever. Thus, we have output movement = $(x/y) \times$ the error.

Another example of a proportional mode controller is an amplifier which gives an output which is proportional to the size of the input. Figure 5.8 illustrates, for the control of temperature of the outflow of liquid from a tank, the use of a differential amplifier as a comparison element and another amplifier as supplying the proportional control mode.

5.3.1 Proportional Band

Note that it is customary to express the output of a controller as a percentage of the full range of output that it is capable of passing on to the correction element. Thus, with a valve as a correction element, as in the float operated control of level in Figure 5.9, we might require it to be completely closed when the output from the controller is 0% and fully open when it is 100% (Figure 5.10). Because the controller output is proportional to the error, these percentages correspond to a zero value for the error and the maximum possible error value. When the error is 50% of its maximum value then the controller output will be 50% of its full range.

Some terminology that is used in describing controllers:

1. Range

The range is the two extreme values between which the system operates. A common controller output range is 4–20 mA.

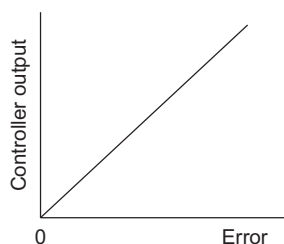


FIGURE 5.6 Proportional control.

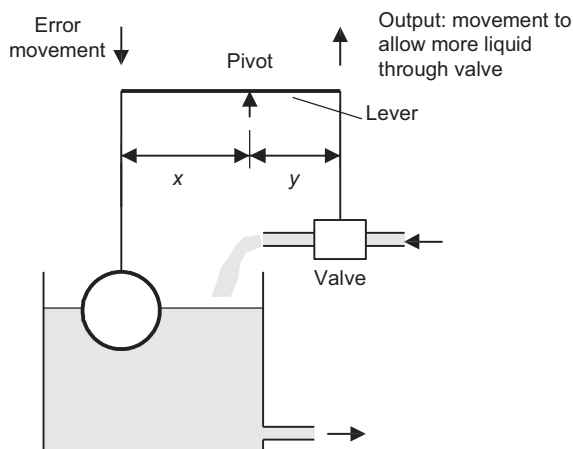


FIGURE 5.7 The float-lever proportional controller.

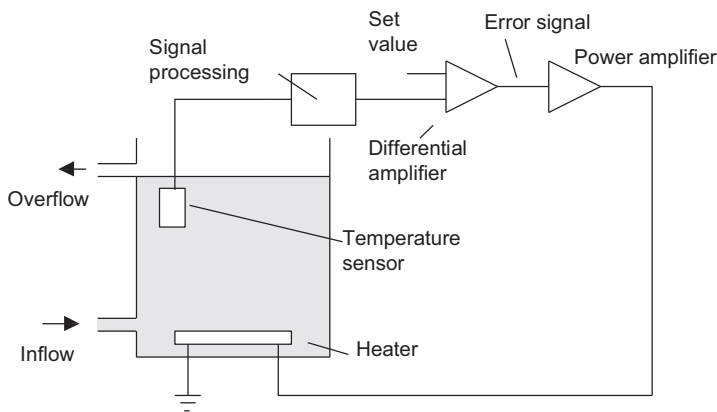


FIGURE 5.8 Proportional controller for the control of temperature.

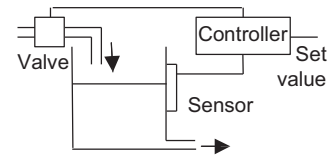


FIGURE 5.9 Water level control system.

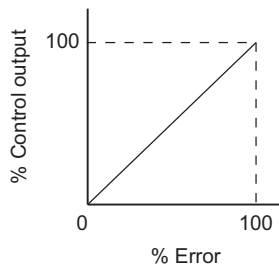


FIGURE 5.10 Percentages.

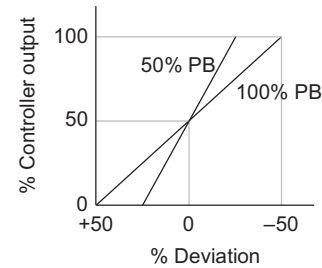


FIGURE 5.11 Proportional band.

2. Span

The span is the difference between the two extreme values within which the system operates, e.g. a temperature control system might operate between 0°C and 30°C and so have a span of 30°C.

3. Absolute deviation

The set point is compared to the measured value to give the error signal, this being generally termed the *deviation*. The term *absolute deviation* is used when the deviation is just quoted as the difference between the measured value and the set value, e.g. a temperature control system might operate between 0°C and 30°C and have an absolute deviation of 3°C.

4. Fractional deviation

The deviation is often quoted as a *fractional* or *percentage deviation*, this being the absolute deviation as a fraction or percentage of the span. Thus, a temperature control system operating between 0°C and 30°C with an error of 3°C has a percentage deviation of $(3/30) \times 100 = 10\%$. When there is no deviation then the percentage deviation is 0% and when the deviation is the maximum permitted by the span it is 100%.

Generally with process controllers, the proportional gain is described in terms of its *proportional band* (PB). The proportional band is the fractional or percentage *deviation* that will produce a 100% change in controller output (Figure 5.11):

$$\%PB = \frac{\% \text{ deviation}}{\% \text{ change in controller output}} \times 100$$

The 100% controller output might be a signal that fully opens a valve, the 0% being when it fully closes it. A 50% proportional band means that a 50% error will produce a 100% change in controller output; 100% proportional band means that a 100% error will produce a 100% change in controller output.

Since the percentage deviation is the error e as a percentage of the span and the percentage change in the controller output is the controller output y_c as a percentage of the output span of the controller:

$$\%PB = \frac{e}{\text{measurement span}} \times \frac{\text{controller output span}}{y_c} \times 100$$

Since the controller gain K_p is y_c/e :

$$\%PB = \frac{1}{K_p} \frac{\text{controller output span}}{\text{measurement span}} \times 100$$

EXAMPLE

What is the controller gain of a temperature controller with a 60% PB if its input range is 0–50°C and its output is 4–20 mA?

$$\%PB = \frac{1}{K_p} \frac{\text{controller output span}}{\text{measurement span}} \times 100$$

and so

$$K_p = \frac{1}{60} \frac{20 - 4}{50 - 0} \times 100 = 0.53 \text{ mA}/^\circ\text{C}$$

5.3.2 Limitations of Proportional Control

Proportional controllers have limitations. Consider a simple unity feedback control system (Figure 5.12). When there is no error, we must have the output value equal to the set value. For many closed-loop control systems, this is what is wanted. But will this be achieved with a proportional control system? The error signal fed to the proportional controller K_p is the difference between the set value and the output value from the system.

$$\text{Error} = \text{set value} - \text{output value}$$

Thus when there is an error and the plant is not delivering the required output, the correcting input applied to the plant is $K_p \times \text{error}$. This correcting signal will steadily reduce the error. However, as the error becomes smaller, the signal applied to the plant to correct the output becomes smaller and will eventually become too small for the output to be maintained at the required set value. We can only have a correcting input to the plant if the error is not zero. Thus this use of a proportional controller leaves the output with a *steady-state error*.

All proportional control systems have a steady-state error. The proportional mode of control tends to be used in processes where the gain K_p can be made large enough to reduce the steady-state error to an acceptable level. However, the larger the gain, the greater the chance of the system oscillating. The oscillations occur because of time lags in the system—the higher the gain, the bigger will be the controlling action for a particular error and so the greater the chance that the system will overshoot the set value and oscillations occur. Zero steady-state error can be obtained when there is an integrator with the proportional control, see Section 5.5.1 for a discussion of PI control. Steady-state errors are also discussed in Section 10.7.

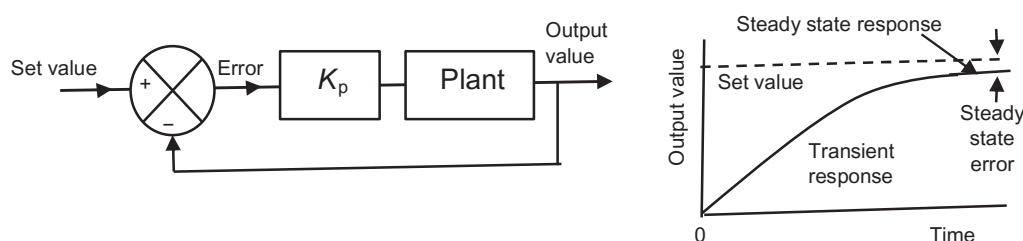


FIGURE 5.12 A closed-loop control system.

EXAMPLE

A proportional controller has a gain of 4. What will be the percentage steady state error signal required to maintain an output from the controller of 20% when the normal set value is 0%?

With a proportional controller we have

$$\% \text{controller output} = \text{gain} \times \% \text{error} = 20 = 4 \times \% \text{error}$$

Hence the percentage error is 5%.

EXAMPLE

For the water level control system described in Figure 5.9, the water level is at the required height when the linear control valve has a flow rate of $5 \text{ m}^3/\text{h}$ and the outflow is $5 \text{ m}^3/\text{h}$. The controller output is then 50% and operates as a proportional controller with a gain of 10. What will be the controller output and the offset when the outflow changes to $6 \text{ m}^3/\text{h}$?

Since a controller output of 50% corresponds to $5 \text{ m}^3/\text{h}$ from the linear control valve, then $6 \text{ m}^3/\text{h}$ means that the controller output will need to be 60%. To give a change in output of $60 - 50 = 10\%$ with a controller having a gain of 10 means that the error signal into the controller must be 1%. There is thus an offset of 1%.

5.4 DERIVATIVE CONTROL

With *derivative control* the change in controller output from the set point value is proportional to the rate of change with time of the error signal, i.e. controller output \propto rate of change of error. Thus we can write

$$\text{derivative controller output} = K_d \times \text{rate of change of error}$$

It is usual to express these controller outputs as a percentage of the full range of output and the error as a percentage of full range. K_d is the constant of proportionality and is commonly referred to as the *derivative time* since it has units of time. Figure 5.13 illustrates the type of response that occurs when there is a steadily increasing error signal.

Because the rate of change of the error with time is constant, the derivative controller gives a constant controller output signal to the correction element. With derivative control, as soon as the error signal begins to change there can be quite a large controller output since it is proportional to the rate of change of the error signal and not its value. Thus with this form of control there can be rapid corrective responses to error signals that occur.

5.4.1 PD Control

Derivative controllers give responses to changing error signals but do not, however, respond to constant error signals, since with a constant error the rate of change of error with time is zero. Because of this, derivative control D is combined with proportional control P. Then:

$$\text{PD controller output} = K_p \times \text{error} + K_d \times \text{rate of change of error with time}$$

Figure 5.14 shows how, with proportional plus derivative control, the controller output can vary when there is a constantly changing error. There is an initial quick change in controller output because of the derivative action followed by the gradual change due to proportional action. This form of control can thus deal with fast process changes better than just proportional control alone. It still, like proportional control alone, needs a steady state error in order to cope with a constant change in input conditions or a change in the set value.

The above equation for PD control is sometimes written as

$$\text{PD controller output} = K_p \left(\text{error} + \frac{K_d}{K_p} \text{rate of change of error} \right)$$

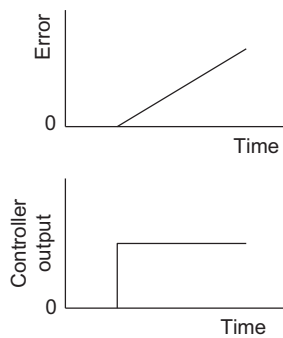


FIGURE 5.13 Derivative control.

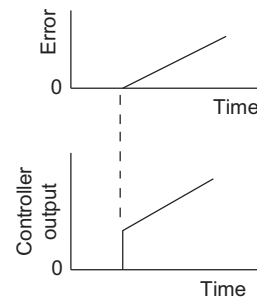


FIGURE 5.14 PD control.

K_d/K_p is called the *derivative action time* T_d and so

$$\text{PD controller output} = K_p(\text{error} + T_d \times \text{rate of change of error})$$

EXAMPLE

A derivative controller has a derivative constant K_d of 0.4 s. What will be the controller output when the error (a) changes at 2%/s, (b) is constant at 4%?

- a. Using the equation given above, i.e. controller output = $K_d \times$ rate of change of error, then we have

$$\text{controller output} = 0.4 \times 2 = 0.8\%$$

This is a constant output.

- b. With a constant error, there is no change of error with time and thus the controller output is zero.

EXAMPLE

What will the controller output be for a proportional plus derivative controller (a) initially and (b) 2 s after the error begins to change from the zero error at the rate of 2%/s (Figure 5.15). The controller has $K_p = 4$ and $T_d = 0.4$ s.

- a. Initially the error is zero and so there is no controller output due to proportional action. There will, however, be an output due to derivative action since the error is changing at 2%/s. Since the output of the controller, even when giving a response due to derivative action alone, is multiplied by the proportional gain, we have

$$\text{controller output} = T_d \times \text{rate of change of error} = 4 \times 0.4 \times 2\% = 3.2\%$$

- b. Because the rate of change is constant, after 2 s the error will have become 4%. Hence, then the controller output due to the proportional mode will be given by

$$\text{controller output} = K_p \times \text{error}$$

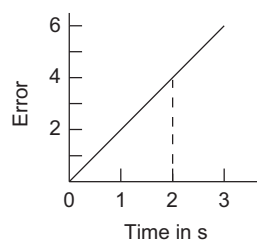


FIGURE 5.15 Example.

and so that part of the output is

$$\text{controller output} = 4 \times 4\% = 16\%$$

The error is still changing and so there will still be an output due to the derivative mode. This will be given by

$$\text{controller output} = K_p T_d \times \text{rate of change of error}$$

and so

$$\text{controller output} = 4 \times 0.4 \times 2\% = 3.2\%$$

Hence the total controller output due to both modes is the sum of these two outputs and $16\% + 3.2\% = 19.2\%$.

5.5 INTEGRAL CONTROL

Integral control is the control mode where the controller output is proportional to the integral of the error with respect to time, i.e.

$$\text{controller output} \propto \text{integral of error with time}$$

and so we can write

$$\text{I controller output} = K_i \times \text{integral of error with time}$$

where K_i is the constant of proportionality and, when the controller output is expressed as a percentage and the error as a percentage, has units of s^{-1} .

To illustrate what is meant by the integral of the error with respect to time, consider a situation where the error varies with time in the way shown in Figure 5.16. The value of the integral at some time t is the area under the graph between $t = 0$ and t . Thus we have

$$\text{controller output} \propto \text{area under the error graph between } t = 0 \text{ and } t$$

Thus as t increases, the area increases and so the controller output increases. Since, in this example, the area is proportional to t then the controller output is proportional to t and so increases at a constant rate. Note that this gives an alternative way of describing integral control as

$$\text{rate of change of controller output} \propto \text{error}$$

A constant error gives a constant rate of change of controller output.

5.5.1 PI Control

The integral mode I of control is not usually used alone but generally in conjunction with the proportional mode P. When integral action is added to a proportional control system the controller output is given by

$$\text{PI controller output} = K_p \times \text{error} + K_i \times \text{integral of error with time}$$

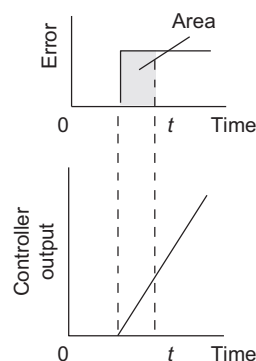


FIGURE 5.16 Integral control.

EXAMPLE

An integral controller has a value of K_i of 0.10 s^{-1} . What will be the output after times of (a) 1 s, (b) 2 s, if there is a sudden change to a constant error of 20%, as illustrated in Figure 5.17?

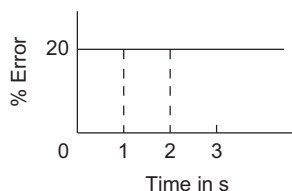


FIGURE 5.17 Example.

We can use the equation:

$$\text{controller output} = K_i \times \text{integral of error with time}$$

- The area under the graph between a time of 0 and 1 s is 20% s. Thus the controller output is $0.10 \times 20 = 2\%$.
- The area under the graph between a time of 0 and 2 s is 40% s. Thus the controller output is $0.10 \times 40 = 4\%$.

where K_p is the proportional control constant and K_i the integral control constant.

Figure 5.18 shows how a system with PI control reacts when there is an abrupt change to a constant error. The error gives rise to a proportional controller output which remains constant since the error does not change. There is then superimposed on this a steadily increasing controller output due to the integral action.

The combination of integral mode with proportional mode has one great advantage over the proportional mode alone: the steady state error can be eliminated. This is because the integral part of the control can provide a controller output even when the error is zero. The controller output is the sum of the area all the way back to time $t = 0$ and thus even when the error has become zero, the controller will give an output due to previous errors and can be used to maintain that condition. Figure 5.19 illustrates this.

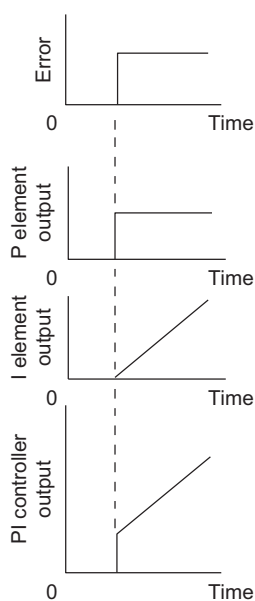


FIGURE 5.18 PI control.

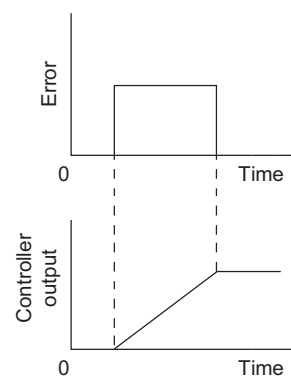


FIGURE 5.19 Controller output when error becomes zero.

The above equation for PI controller output is often written as

$$\text{PI controller output} = K_p \left(\text{error} + \frac{K_i}{K_p} \times \text{integral of error} \right)$$

K_p/K_i is called the *integral action time* T_i and so

$$\text{PI controller output} = K_p \left(\text{error} + \frac{1}{T_i} \times \text{integral of error} \right)$$

Because of the lack of a steady state error, a PI controller can be used where there are large changes in the process variable. However, because the integration part of the control takes time, the changes must be relatively slow to prevent oscillations.

5.6 PID CONTROL

Combining all three modes of control (proportional, integral and derivative) enables a controller to be produced which has no steady state error and reduces the tendency for oscillations. Such a controller is known as a *three-mode controller* or *PID controller*: A PID controller can be considered to be a proportional controller which has integral control added to eliminate offset error and derivative control to reduce time lags. Figure 5.20 shows a number of forms such a control system can take. For the parallel arrangement (Figure 5.20A), the equation describing its action is:

$$\text{PID controller output} = K_p \times \text{error} + K_i \times \text{error} + K_d \times \text{error}$$

where K_p is the proportionality constant, K_i the integral constant and K_d the derivative constant. With this arrangement the proportional gain does not influence the integral and derivative control, each acting independently on the error, and this version is thus rarely used. Figure 5.20B shows what is termed the interactive form of controller and Figure 5.20C the non-interactive form. With these forms the controller gain element does affect the proportional, integral and derivative control, and these are the PID control forms which are generally used. The interactive form has the problem that the derivative control is acting on the output of the integrator whereas the non-interactive form, sometimes called the ideal form, has the integral and derivative controls just acting directly on the output of the proportional element. For the interactive controller the equation describing its action is:

$$\text{PID controller output} = K_c \left(\text{error} + \frac{1}{T_i} \times \text{integral of error} \right) \times (1 + T_d \times \text{rate of change of error})$$

where K_c is the controller gain. For the non-interactive form:

$$\text{PID controller output} = K_c \left(\text{error} + \frac{1}{T_i} \times \text{integral of error} + T_d \times \text{rate of change of error} \right)$$

If no derivative control is used, the interactive and non-interactive forms are equal.

A problem encountered with PID controllers is *integrator windup*. For example, there are typically limitations with actuators in that the output cannot always be proportional to the input and can reach a saturation or limiting output. For example, a motor will have a limited speed and a valve cannot be more than fully opened or fully closed. Thus even when the actuator has reached its limit, the integrator in the control system can continue integrating, and the required value of the process variable reached and passed before the integral term starts to decrease (Figure 5.21). The integral term and the controller output can then become very large and thus lead to large overshoots and times needed for the system to recover. This situation is referred to integrator windup. To correct for integrator windup, commercial PID controllers include anti-windup circuits. Essentially, these detect when actuator saturation occurs and then switch off the integral action.

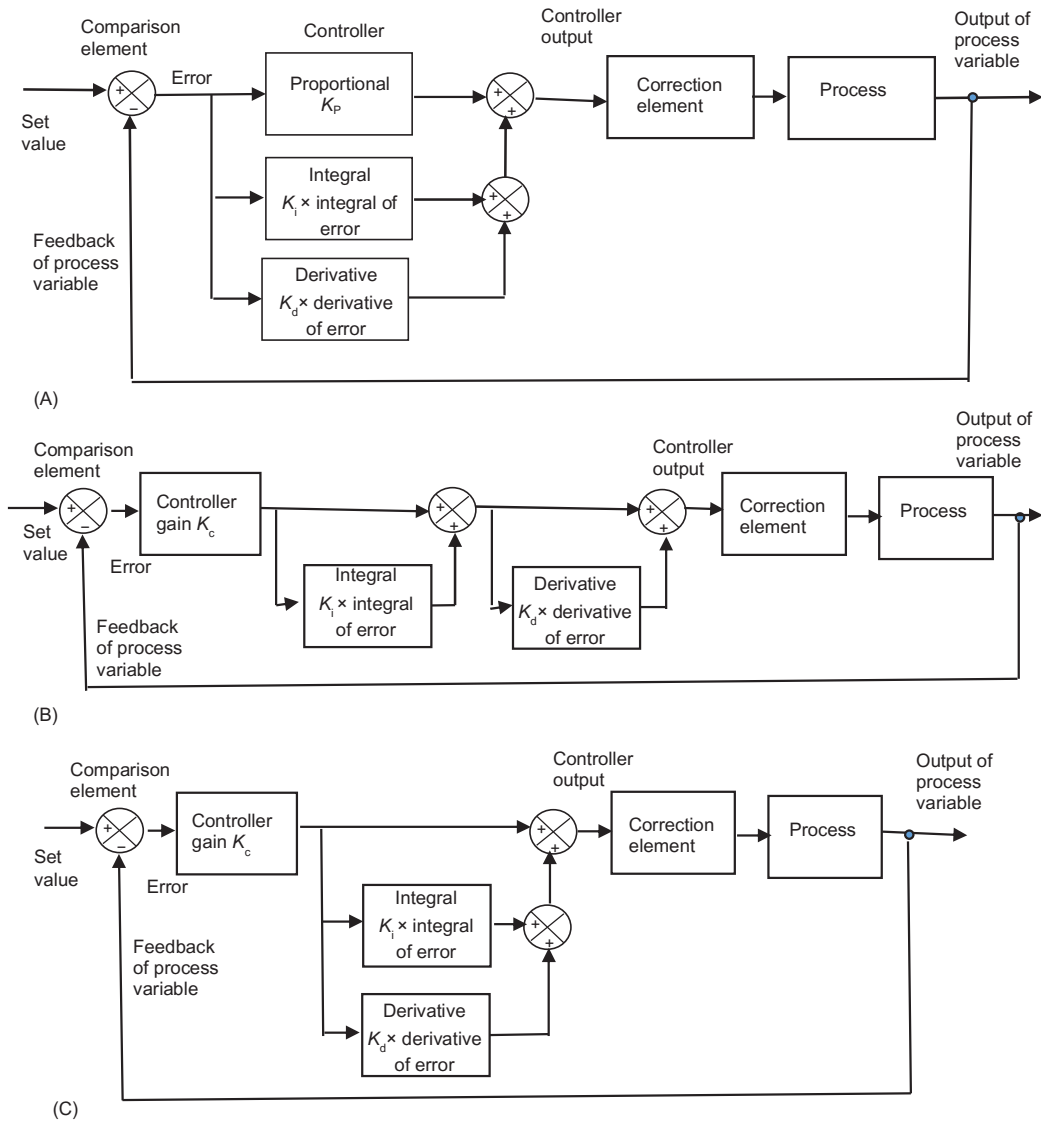


FIGURE 5.20 PID controllers: (A) parallel, (B) interactive and (C) non-interactive forms.

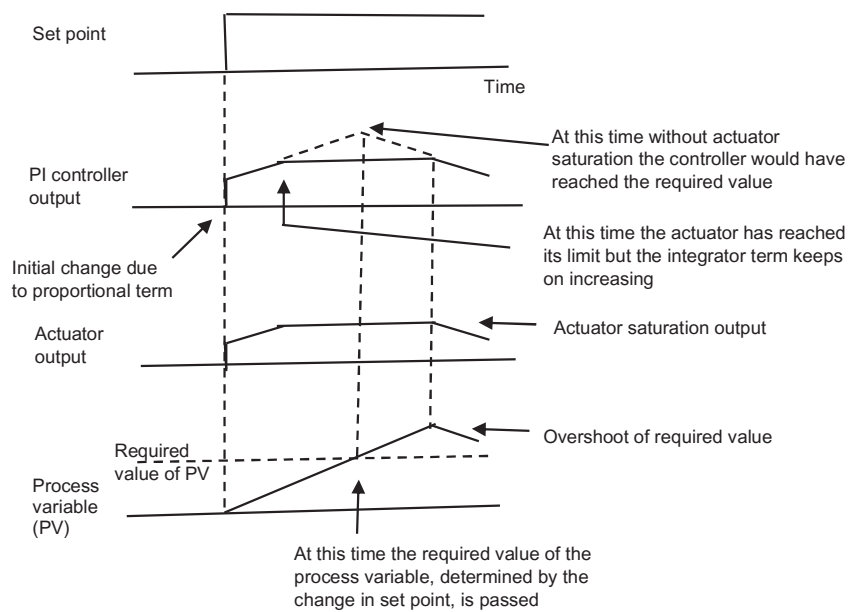


FIGURE 5.21 Integrator windup.

EXAMPLE

Determine the controller output of a three-mode controller having K_p as 4, T_i as 0.2 s, T_d as 0.5 s at time (a) $t = 0$ and (b) $t = 2$ s when there is an error input which starts at 0 at time $t = 0$ and increases at 1%/s (Figure 5.22).

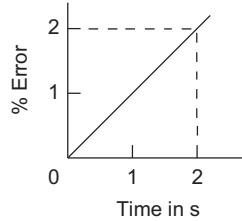


FIGURE 5.22 Example.

a. Using the equation:

$$\text{controller output} = K_p(\text{error} + (1/T_i) \times \text{integral of error} + T_d \times \text{rate of change of error})$$

we have for time $t = 0$ an error of 0, a rate of change of error with time of 1 s^{-1} , and an area between this value of t and $t = 0$ of 0. Thus

$$\text{controller output} = 4(0 + 0 + 0.5 \times 1) = 2.0\%$$

b. When $t = 2$ s, the error has become 1%, the rate of change of the error with time is 1%/s and the area under between $t = 2$ and $t = 0$ is 1%/s. Thus

$$\text{controller output} = 4(1 + (1/0.2) \times 1 + 0.5 \times 1) = 26\%$$

5.6.1 PID Process Controller

Figure 5.23 shows the basic elements that tend to figure on the front face of a typical three-term process controller.

As an example of such a controller, the OMEGA CN63100-R1 temperature controller is listed as having:

- PID Control with Reduced Overshoot
- On-Demand Auto-Tuning of PID Control Settings
- Dual LED Displays for Simultaneous Indication of Temperature and Setpoint

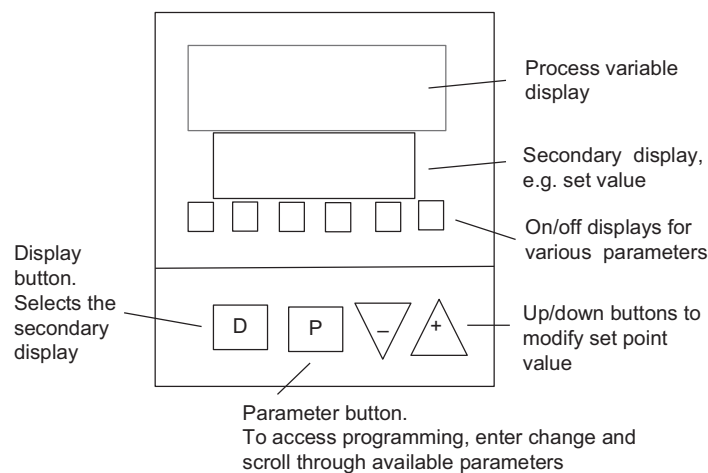


FIGURE 5.23 The facia of a typical PID controller.

Status Indicators for Outputs and Control Modes
 Thermocouple or RTD Inputs
 Optional Dual Alarm Inputs
 Manual/Automatic Control Modes
 Setpoint Ramping for Process Startup
 Sensor Error Compensation (Offset) and Break Detection.

The controller operates in the PID control mode for both heating and cooling, with on-demand auto-tune, which will establish the tuning constants. The PID tuning constants may be fine-tuned by the operator at any time and then locked-out from further modification.

Another example is the TEMPATRON PID330 temperature controller which is listed as having:

Dual display of set point and process variable
 Can be configured as 1 or 2 On/Off or PID control outputs
 Auto-tuning of PID parameters
 Up to 3 relay or SSR outputs
 Alarm output
 Thermocouple, RTD and analog inputs
 Sensor break detection
 Optional RS-485 communication.

5.7 TUNING

The design of a controller for a particular situation involves selecting the control modes to be used and the control mode settings. This means determining whether proportional control, proportional plus derivative, proportional plus integral or proportional plus integral plus derivative is to be used and selecting the values of K_P , K_I and K_D . These determine how the system reacts to a disturbance or a change in the set value, how fast it will respond to changes, how long it will take to settle down after a disturbance or change to the set value, and whether there will be a steady state error.

Figure 5.24 illustrates the types of response that can occur with the different modes of control when subject to a step input, i.e. a sudden change to a different constant set value or perhaps a sudden constant disturbance. Proportional control gives a fast response with oscillations which die away to leave a steady state error. Proportional plus integral control has no steady state error but is likely to show more oscillations before settling down. Proportional but integral plus derivative control has also no steady state error, because of the integral element, and is likely to show less oscillations than the proportional plus integral control. The inclusion of derivative control reduces the oscillations.

The term *tuning* is used to describe methods used to select the best controller settings to obtain the optimum performance when changes occur, so the controller gives a fast response time, oscillations decay quite rapidly, and there is the smallest overshoot. Some commercial controllers offer a self-tuning feature, auto-tuning, in which small set point changes are sent to the process, and from the response the controller itself calculates optimum tuning values and applies them.

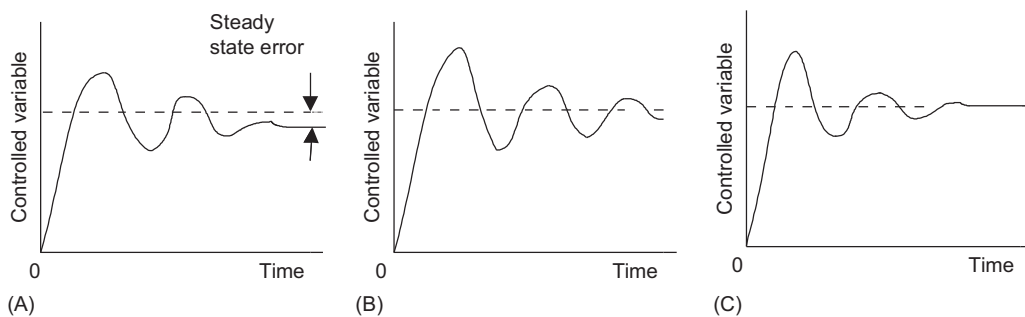


FIGURE 5.24 Responses to (A) P, (B) PI, and (C) PID control.

The following is a description of some of the methods that can be used for tuning by an operator directly doing the tuning. Methods that are widely used are the process reaction method and the ultimate cycle methods, both by *Ziegler and Nichols*. Various software tools are also available for tuning.

5.7.1 Process Reaction Tuning Method

This method uses certain measurements made from testing the system with the control loop open so that no control action occurs. Generally the break is made between the controller and the correction unit (*Figure 5.25*). A test input signal is then applied to the correction unit and the response of the controlled variable determined.

The test signal is a step signal with a step size expressed as the percentage change P in the correction unit (*Figure 5.26*). The output response of the controlled variable, as a percentage of the full-scale range, to such an input is monitored and a graph (*Figure 5.27*) of the variable plotted against time. This graph is called the *process reaction curve*. A tangent is drawn to give the maximum gradient of the graph. The time between the start of the test signal and the point at which this tangent intersects the graph time axis is termed the lag L . If the value of the maximum gradient is M , expressed as the percentage change of the set value of the variable per minute, *Table 5.1* shows the criteria given by Ziegler and Nichols to determine the controller settings. The basis behind these criteria is to give a closed-loop response for the system which exhibits a *quarter amplitude decay* (*Figure 5.28*), i.e. the amplitude of the response of the system shows oscillations which decay with time so that the amplitude decreases by a quarter on each oscillation.

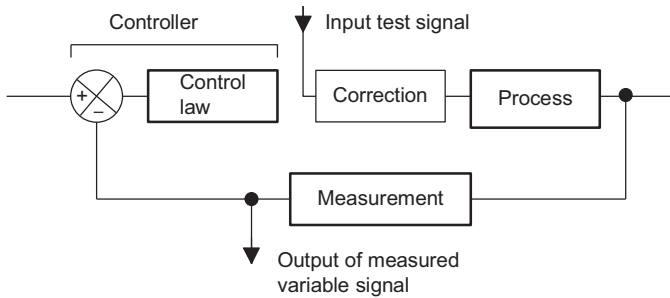


FIGURE 5.25 Test arrangement.

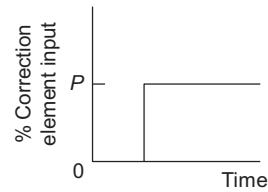


FIGURE 5.26 Test signal.

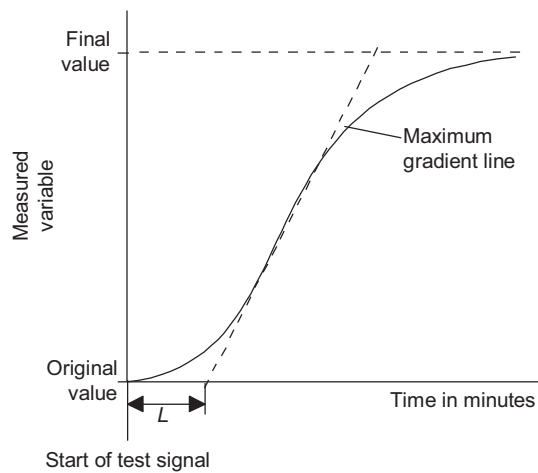


FIGURE 5.27 The process reaction graph.

TABLE 5.1 Settings from the Process Reaction Curve Method

Type of controller	K_p	T_i	T_d
P	P/RL		
PI	$0.9P/RL$	$3.3L$	
PID	$1.2P/RL$	$2L$	$0.5L$

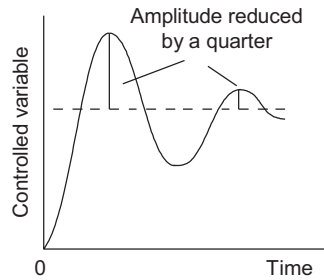


FIGURE 5.28 Quarter amplitude decay.

EXAMPLE

Determine the settings of K_p , K_i and K_d required for a three-mode controller which gave a process reaction curve shown in Figure 5.29 when the test signal was a 10% change in the control valve position.

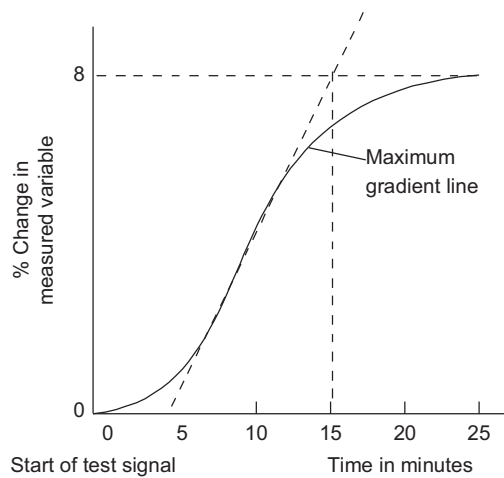


FIGURE 5.29 Example.

Drawing a tangent to the maximum gradient part of the graph gives a lag L of 5 min and a gradient M of $8/10 = 0.8\%/min$. Hence

$$K_p = \frac{1.2P}{ML} = \frac{1.2 \times 10}{0.8 \times 5} = 3$$

$$T_i = 2L = 2 \times 5 = 10 \text{ min}$$

$$T_d = 0.5L = 0.5 \times 5 = 2.5 \text{ min}$$

Since $K_i = K_p/T_i$, then $K_i = 0.3 \text{ min}^{-1}$. Since $K_d = K_p T_d$, then $K_d = 7.5 \text{ min}$.

TABLE 5.2 Settings for the Ultimate Cycle Method for Quarter Amplitude Decay

Type of controller	K_p	T_i	T_d
P	$0.5K_{pu}$		
PI	$0.45K_{pu}$	$T_u/1.2$	
PID	$0.6K_{pu}$	$T_u/2$	$T_u/8$

TABLE 5.3 Settings for the Ultimate Cycle Method for PI Control

Type of controller	K_p	T_i	T_d
PID, $1/4$ decay	$0.6K_{pu}$	$T_u/2$	$T_u/8$
PID, some overshoot	$0.33K_{pu}$	$T_u/2$	$T_u/3$
PID, no overshoot	$0.2K_{pu}$	$T_u/3$	$T_u/2$

5.7.2 Ultimate Cycle Tuning Method

With this method, with the closed-loop system, the integral and derivative actions are first reduced to their least effective values. The proportional constant K_p is then set low and gradually increased until oscillations in the controlled variable start to occur. The critical value of the proportional constant K_{pc} at which this occurs is noted and the periodic time of the oscillations T_c measured. The procedure is thus:

1. Set the controller to manual operation and the plant near to its normal operating conditions
2. Turn off all control modes but proportional
3. Set K_p to a low value, i.e. the proportional band to a wide value
4. Switch the controller to automatic mode and then introduce a small set-point change, e.g. 5–10%
5. Observe the response
6. Set K_p to a slightly higher value, i.e. make the proportional band narrower
7. Introduce a small set-point change, e.g. 5–10%
8. Observe the response.
9. Keep on repeating 6, 7 and 8, until the response shows sustained oscillations which neither grow nor decay. Note the value of K_p giving this condition (K_{pu}) and the period (T_u) of the oscillation
10. The Ziegler and Nichols recommended criteria for controller settings for a system to have quarter amplitude decay are given by Table 5.2. For a PID system with some overshoot or with no overshoot, the criteria are given by Table 5.3.

EXAMPLE

When tuning a three-mode control system by the ultimate cycle method it was found that, with derivative and integral control switched off, oscillations begin when the proportional gain is increased to 3.3. The oscillations have a periodic time of 500 s. What are the suitable values of K_p , K_i and K_d for quarter amplitude decay?

Using the equations given above:

$$K_p = 0.6K_{pc} = 0.6 \times 3.3 = 1.98$$

$$T_i = 500/2 = 250 \text{ s and so } K_i = K_p/T_i = 1.98/250 = 0.00792 \text{ s}^{-1}$$

$$T_d = 500/8 = 62.5 \text{ s and so } K_d = K_p T_d = 1.98 \times 62.5 = 123.75 \text{ s}$$

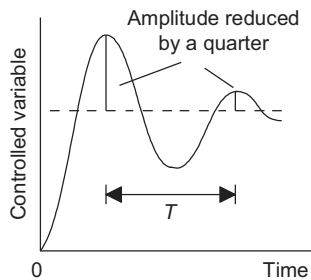


FIGURE 5.30 Quarter amplitude decay.

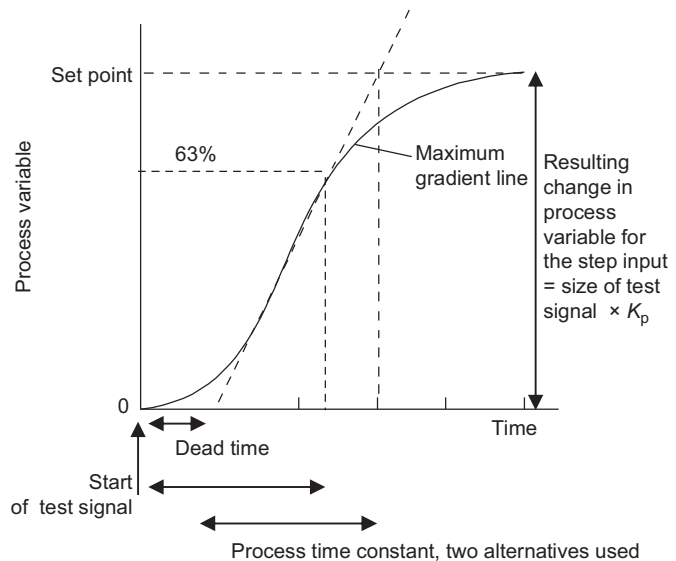


FIGURE 5.31 Response to sudden step input, i.e. a bump.

5.7.3 Quarter Amplitude Decay

A variation on the ultimate cycle method involves not adjusting the control system for sustained oscillations but for oscillations which have a quarter amplitude decay (Figure 5.30). The controller is set to proportional only. Then, with a step input to the control system, the output is monitored and the amplitude decay determined. If the amplitude decay is less than a quarter the proportional gain is increased, if less than a quarter it is decreased. The step input is then repeated and the amplitude decay again determined. By a method of trial and error, the test input is repeated until a quarter wave amplitude decay is obtained. We then have the value for the proportional gain constant. The integral time constant is then set to be $T/1.5$ and the derivative time constant to $T/6$.

5.7.4 Lambda Tuning

Lambda tuning has been developed to give a non-oscillatory response to set point changes and load disturbances. The first part of the test is to switch the controller to manual mode, wait until the process variable has settled to a steady state value and then apply an abrupt step up, a so-called bump, in the set point. The process variable is then monitored in response to this step input until it settles out again (Figure 5.31). The response is then specified by extracting from the response graph the dead time D , the process time constant τ , which is defined as being the lambda value, and the process gain K_p , where $K_p = (\text{the resulting change in the process variable}) / (\text{the size of the step input producing the change})$. The PID controller is then set to have the values:

$$\text{Proportional gain} = \frac{\tau}{K_p(\lambda + T_d)}$$

$$\text{Integral time} = \tau$$

$$\text{Derivative time} = 0$$

5.7.5 Software Tools

Software packages are available for tuning. They gather data from the system, develop process models and then suggest optimal tuning values. An impulse can be induced in the system and then the control system's response used to determine optimum loop values.

5.7.6 Adaptive Controllers

Adaptive controllers change the controller parameters to adapt to changes in the process that occur with time, e.g. a change in plant load. If the transfer function of the plant changes then there needs to be retuning of the controller if optimum control conditions are to occur. An adaptive controller thus determines the values of K_p , K_i and K_d needed to adapt to new process conditions and makes the necessary changes.

The self-tuning, often termed auto-tuning that exists with many commercially available controllers is a form of adaptive control. When the operator presses a button on the controller, the controller injects a small disturbance into the system and determines the system's response. This is then compared with the desired response and the control parameters adjusted to bring the actual response closer that required.

5.8 DIGITAL SYSTEMS

The term *direct digital control* is used to describe the use of digital computers/microcontrollers in the control system to calculate the control signal that is applied to the actuators to control the plant. Such a system is of the form shown in Figure 5.32, control using a microcontroller being discussed in Section 4.6 and illustrated in Figure 4.29. Samples are taken at regular intervals of time, via the analog-to-digital converter (ADC), the feedback signal from the plant. This, together with the input of the required value which has been inputted as a digital signal to the memory of the computer, is then processed by the computer according to the required control law to give the required correction signal which is then sent via the digital-to-analog converter (DAC) to provide the correcting action to the plant to give the required control. Direct digital control laws are computer programs that take the set value and feedback signals and operate on them to give the output signal to the actuator. The program might thus be designed to implement PID control.

The program involves the computer carrying out operations on the fed back measurement value occurring at the instant it is sampled and also using the values previously obtained. The program for proportional control thus takes the form of setting initial values to be used in the program and then a sequence of program instructions which are repeated every sampling period:

Initialise

Set the required value

Set the value of the proportional gain.

Loop

Sample the measured value and determine the error at the instant concerned

Calculate the output by multiplying the error by the set value of the proportional gain

Output the value of the calculated output which the DAC will hold until the next sample is taken

Wait for the end of the sampling period

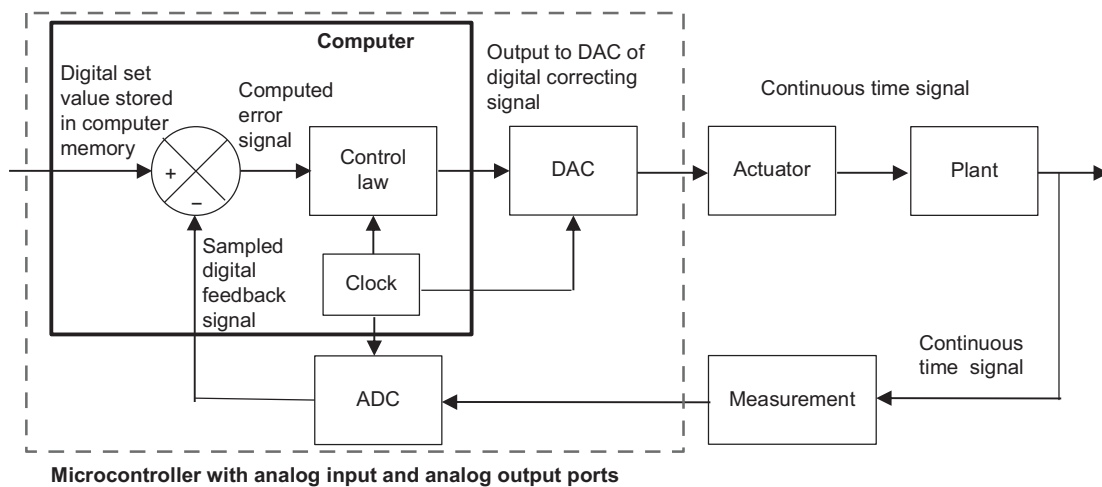


FIGURE 5.32 Direct digital control.

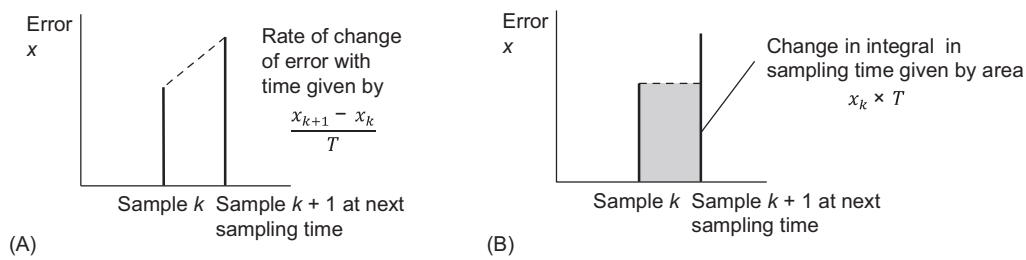


FIGURE 5.33 Approximations with samples taken at a sampling time interval of T : (A) derivative and (B) integration.

Go back to Loop and repeat the program.

For PD control the program is

Initialise

- Set the required value
- Set the initial value of the error that is assumed to have occurred in the previous sampling period
- Set the value of the proportional gain
- Set the value of the derivative gain.

Loop

- Input the error at the instant concerned
- Calculate the proportional part of the output by multiplying the error by the set value of the proportional gain
- Calculate the derivative part of the output by subtracting the value of the error at the previous sampling instant from the value at the current sampling instant (the difference is a measure of the rate of change of the error since the signals are sampled at regular intervals of time) and multiply it by the set value of the derivative gain (Figure 5.33A)
- Calculate the output by adding the proportional and derivative output elements
- Output the value of the calculated output
- Wait for the end of the sampling period
- Go back to Loop and repeat the program.

For PI control the program is:

Initialise

- Set the required value
- Set the value of the output that is assumed to have occurred in the previous sampling period
- Set the value of the proportional gain
- Set the value of the integral gain.

Loop

- Input the error at the instant concerned
- Calculate the proportional part of the output by multiplying the error by the set value of the proportional gain
- Calculate the integral part of the output by multiplying the value of the error at the current sampling instant by the sampling period and the set value of the integral gain (this assumes that the output has remained constant over the previous sampling period and so multiplying its value by the sampling period gives the area under the output–time graph) and add to it the previous value of the output (Figure 5.33B)
- Calculate the output by adding the proportional and integral output elements
- Output the value of the calculated output
- Wait for the end of the sampling period
- Go back to Loop and repeat the program.

5.8.1 Embedded Systems

The term *embedded system* is used for control systems involving a microprocessor being used as the controller and located as an integral element, i.e. embedded, in the system. Such a system is used with engine management

control systems in modern cars, exposure and focus control in modern cameras, the controlling of the operation of modern washing machines and indeed is very widely used in modern consumer goods.

5.9 FUZZY LOGIC CONTROL

PID control is a well-used method of control used to drive a system to a set point, particularly in temperature control systems. However, there are problems in situations where the target control value changes. If the value changes instantaneously, the PID controller applies a large correcting factor that can lead to overshooting. Tracking a gradual change in the set point can also lead to problems. A more flexible method of control which can overcome such problems is now becoming widely used, this being termed *fuzzy logic control*. Such controllers use a flexible set of control rules. Some commercial controllers already use this method of control, e.g. Omega's Platinum Series Temperature, Process and Strain controllers which are described as fully auto-tune PID with fuzzy logic adaptive control mode that allows the PID algorithm to be continuously optimised.

Fuzzy logic controllers are now increasingly being used for domestic appliances such as washing machines and vacuum cleaners. Such controllers are able to solve control problems for mathematical models cannot be easily developed.

5.9.1 Fuzzy Logic

With a conventional controller a central heating system might be designed to switch on heating when the temperature input is a particular temperature, say 15°C, but not when it is, say, 20°C. We can say that the controller is responding to a 'crisp' input, and such controllers are what have so far been considered in this chapter. However, you probably would like the heating system to switch on when you feel it is cold, thus responding to a range of temperatures. The term cold represents a 'fuzzy' concept of the temperature and requires a controller which responds to a range of temperatures, such as a range identified as cold, rather than a specific temperature. Incidentally, this means that the information which humans use in their everyday lives on which to base intuitive decisions and apply general rules of thumb is what is being asked of fuzzy logic control situations.

When the term fuzzy is used, the decisions are not made on the basis of YES there is a particular signal or NO there is no such signal but also on possibilities between YES and NO, and so we can have for descriptions of the input to the controller:

CERTAINLY YES

MAYBE YES or NO

CERTAINLY NO

and thus there is a degree of fuzziness about data fitting a proposition. Propositions need not be classified as definitely true or false, i.e. 0 or 1, but their truth or falsehood can be weighted so that it can be classified as being between true and false on a 0 to 1 scale. This accords with real-life situations in that they do not always assume a true or false situation with a crisp dividing line between them, but the dividing line might be fuzzy. We have the concept of partial truth.

We can put data into sets, a set being a collection of things that belong to some definition. With non-fuzzy sets, we would have distinct unique grade sets so that data is either in one set or the other, i.e. yes or no. With fuzzy logic, we can have the two sets overlapping so that data can be in the distinctly separate set regions yes or no, or in the overlapping region maybe yes or no (Figure 5.34). A fuzzy set arrangement has a continuation of grades of membership, not a unique grade.

As an illustration, consider the heights of people. Suppose, we have people 1.8 m or taller classified as tall and people less than 1.8 m as short. This set can be represented by Figure 5.35A showing how the degree of membership of the tall set, or membership function as it is usually called, varies with height. A person with a height of



FIGURE 5.34 Data sets: (A) definitely in the yes or no sets and (B) in the yes, maybe yes or no, or no sets.

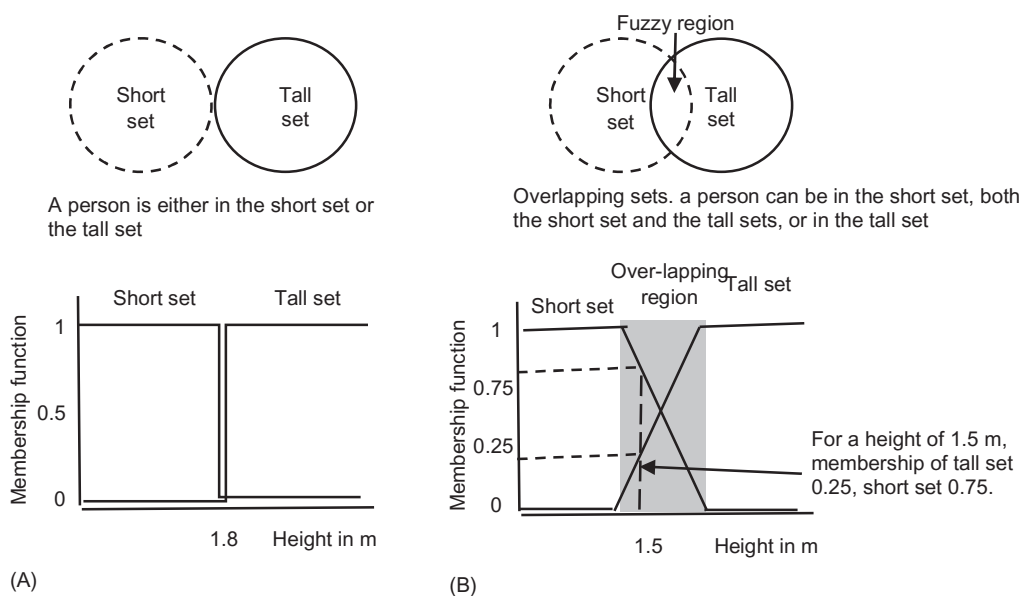


FIGURE 5.35 (A) Classical set theory and (B) fuzzy set theory.

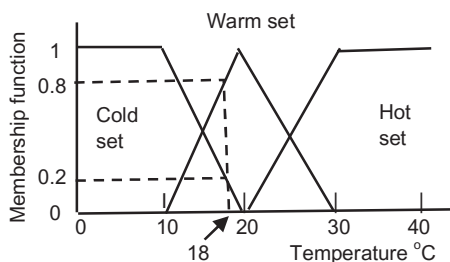


FIGURE 5.36 Fuzzy set membership functions.

1.8 m or more has a membership function of 1.0 and is definitely classed as tall and in the tall set. A person with a height less than 1.8 m has a membership function of 0 and is definitely not tall and not in the tall set. Thus even if the person was only 1 cm less than 1.8 m, they would not be in the tall set. With classical sets we have defined to sets, one for tall people and one for short people and the sets do not overlap. A person is either in one set or the other. A person is thus either in or not in the tall set.

With fuzzy sets, we have overlapping sets so the membership function is of the form shown in Figure 5.35B. The membership function defines the fuzzy set for the possible values underneath it on the horizontal axis. The vertical axis, on a scale of 0–1, provides the membership value of the height in the tall fuzzy set. A membership function of 1 means that the corresponding element is definitely a member of that set. A membership function of 0 means that the element is not an element of that set. For Figure 5.35B with a person of height 1.5 m, there is a membership of 0.25 in the tall set and 0.75 in the short set, so having a low degree of membership of the set of tall people, and so is not very tall. A person with a membership function of 0.95 in the tall set and 0.05 in the short set would have a high degree of membership in the set of tall people. A membership greater than 0 and less than 1 corresponds to an element in the fuzzy region. Fuzzy set theory thus presents a better representation of the tallness of the people, recognising not only clear-cut alternative heights but also the graduations in between.

As a further example, consider the temperature of some object. We might define with fuzzy logic overlapping sets of cold 0°C–20°C, warm 10°C–30°C and hot 20°C–40°C. A temperature of 18°C is thus a member of two sets. If the fuzzy set membership functions are defined as shown in Figure 5.36, then 18°C has a cold membership of 0.2, a warm membership of 0.8 and a hot membership of 0. A given input may thus belong to more than one fuzzy set.

5.9.2 Fuzzy Logic Control Systems

Figure 5.37 illustrates the basic elements of a fuzzy logic-controlled process, it consists of fuzzification of the input data, using rules to make decisions relating to the fuzzy data, then defuzzification to give outputs. Such a controller is designed to respond to a range of values of input rather than a specific input. The procedure adopted is thus described in the following:

1. **Fuzzification:** Determine a set of fuzzy rules for the fuzzy set membership and use them to fuzzify the inputs using the input membership functions and so convert crisp data to fuzzy values using the membership function. Inputs are fuzzified by finding the intersection of the crisp input value with the input membership function. An input might be classified as being in one or more sets. Thus, for a temperature input, we might have for the sets for temperature ranges defined as being VERY COLD, COLD, WARM, VERY WARM and HOT. Thus a crisp temperature value of 18°C might be fuzzified by being defined as falling within the membership function defined as being WARM with a degree of membership of 0.9 and also falling within the membership function defined as COLD with a degree of membership of 0.2, as illustrated in Figure 5.36. The purpose of fuzzification is thus to map the inputs from a set of sensors to values from 0 to 1 using a set of input membership functions.
2. **Using rules to make decisions:** At the heart of a fuzzy logic rule-based system are a set of rules. These, when combined with facts, e.g. this could be inputs from sensors and users, enable decisions to be made which are then used to actuate actuators and control outputs. Fuzzy logic control rules can be developed by simply writing down common sense rules without having a comprehensive theoretical knowledge of control theory. The rule evaluation then involves the input rule strength being considered with the output membership function. Rules are written in the form: IF ... THEN. Thus, for example, for the control of temperature, we might have rules of the form:
 - IF temperature high THEN action 1
 - IF temperature medium THEN action 2
 There is the possibility that a particular temperature could occur in different rule strengths in both the high and medium sets.
3. **Defuzzification:** Finally, there is defuzzification to give a single output, i.e. a crisp value. If action 1 is defined by fuzzy set F_1 and action 2 by fuzzy set F_2 , then for defuzzification the two sets can be aggregated if the input, e.g. temperature in the high and medium sets, occurs in both sets. Figure 5.38 shows how actions can be represented by fuzzy sets. A number of methods exist for defuzzification, but a commonly used one is *Mamdani's inference method* (Figure 5.39A). With this method, each output membership function is truncated at the corresponding membership value for each function and the resulting truncated membership functions added together with overlapping areas just counting as a single-area layer, the OR function being used for the areas. What is then often referred to as the value at the centre of gravity of the consolidated area, outlined in the figure, is then taken as the crisp output value. It is the value at which the outlined area to the left of the point equals the outlined area to the right. Thus, for Figure 5.39A, the centre of gravity of the consolidated truncated membership areas for which there is membership, i.e. functions 2 and 3 and so the area outlined, is taken as the crisp output value. An alternative method is similar but when the truncated membership functions are added together, the overlapping areas are not counted as a single-area layer but two separate layers (Figure 5.39B). The centre of gravity of the summed separate areas is then taken as the crisp output value. The centre of gravity of the truncated areas for which there is membership, i.e. functions 2 and 3 and so the two areas outlined, is taken as the crisp output value.

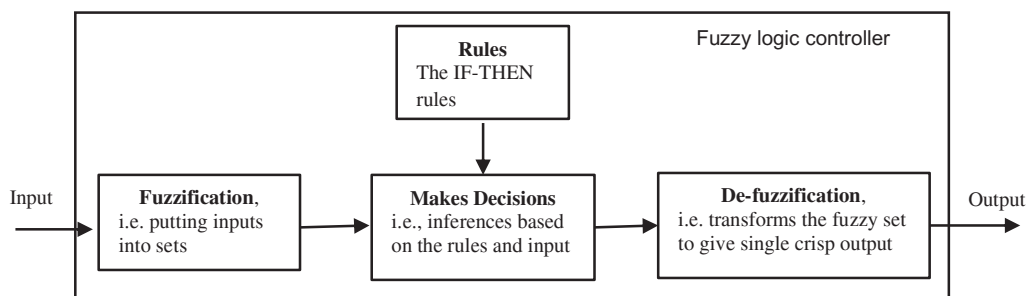


FIGURE 5.37 The basic elements of a fuzzy logic controller.

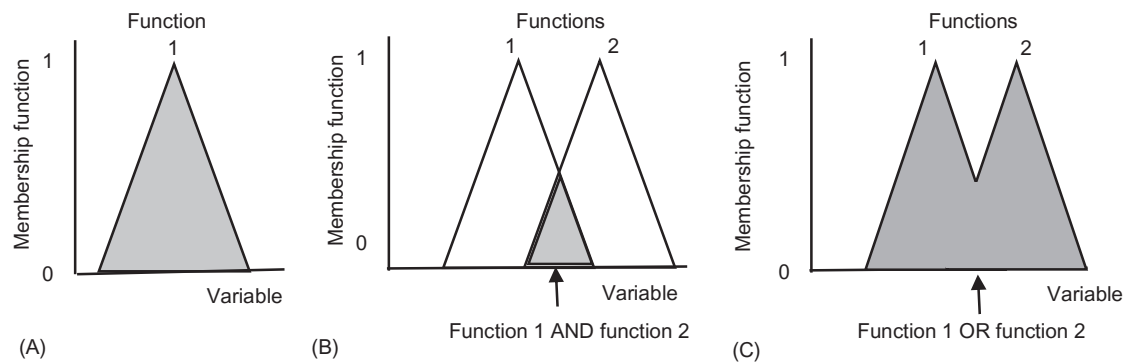


FIGURE 5.38 Membership functions for (A) a single membership function, (B) the AND of two membership functions and (C) the OR of two membership functions.

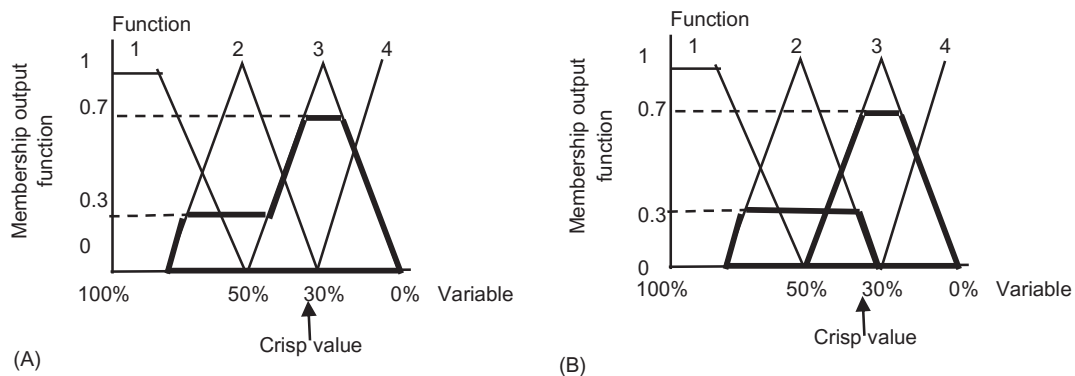


FIGURE 5.39 Defuzzification using (A) Mamdani's method and (B) simpler centre of gravity method.

$$\text{Centre of gravity of the truncated areas} = \frac{\sum \text{For each area, centre of gravity} \times \text{truncated area}}{\sum \text{For each area, truncated area}}$$

The area under a symmetric triangular curve has its centre of gravity under its apex. The area of a trapezium cut from a triangle is $1/2 \text{ height} \times \text{sum of its upper and lower lengths}$.

5.9.3 Fuzzy Logic Controller

Fuzzy logic has been used to control temperature by the action of varying the on-time of a heater, the system implementing the temperature control consisting of a temperature sensor, a microcontroller and actuator (Figure 5.40). The temperature sensor could be a thermo-transistor such as LM35 (see Section 2.7.6) and the actuator an electric heater. The analogue output of the sensor is converted by the microcontroller ADC input to a digital signal which can be processed by the fuzzy logic program stored in the microcontroller and the output used to operate the heater by switching it off or on according to the output from the fuzzy logic program, the fuzzy rules being:

- IF the error is negative THEN timer switched on for very long period
- IF the error is small negative THEN timer switched on for long period
- IF the error is zero THEN timer switched on for medium period
- IF the error is small positive THEN timer switched on for short period
- IF the error is large positive THEN timer switched on very short period

Figure 5.41 shows how the basic components of a feedback control loop can appear when using fuzzy logic control in place of a traditional PID controller for closed-loop control. The input to the fuzzy logic system is 'crisp' inputs of the error, and the set of crisp inputs is transformed to the corresponding fuzzy set, probably a

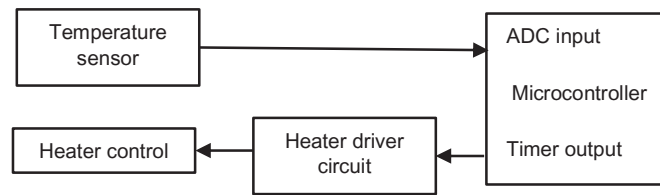


FIGURE 5.40 Control system for heater.

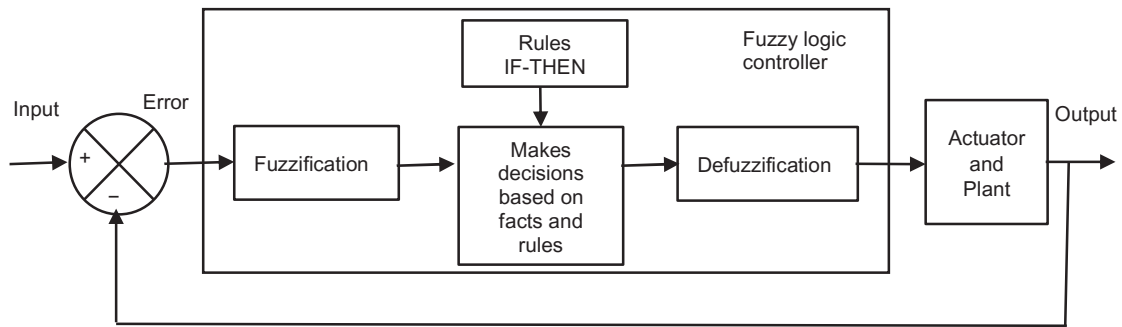


FIGURE 5.41 Fuzzy logic controller in feedback control loop.

triangular membership function, which is then operated on by the fuzzy logic rule. The fuzzy logic controller rules can include rules such as

IF error is zero and change in error is zero THEN output is zero

IF error is zero and change in error is positive small THEN output is negative small

IF error is zero and change in error is negative small THEN output is positive small

The IF–THEN rule describes the desired relationship between inputs and output to be implemented using fuzzy logic control and so make it possible to design a controller for a system with an unknown transfer function. The fuzzy set output is then defuzzified to give a ‘crisp’ output control signal to the plant. The defuzzification can be obtained by a method such as Mamdani’s inference method.

As an illustration of a fuzzy control system, consider one which might be used for automobile cruise control. Figure 5.42 shows the basic system with fuzzy controller having inputs of the error signals for the velocity and the rate of change of velocity, i.e. acceleration. It is a fuzzy PD control system. The derivative part, i.e. the rate of change of error, is obtained by using backward difference, i.e. the change in error during the last time interval divided by the length of the interval. Possible rules for the controller output for the application of the driving force by the engine are as follows:

IF velocity error means too fast and acceleration is negative THEN apply minimum force

IF velocity error means too fast and acceleration is zero THEN apply minimum force

IF velocity error means too fast and acceleration is positive THEN apply minimum force

IF velocity error is zero and acceleration is negative THEN apply normal force.

IF velocity error is zero and acceleration is zero THEN apply normal force.

IF velocity error is zero and acceleration is positive THEN apply normal force

IF velocity error means too slow and acceleration is negative THEN apply maximum force

IF velocity error means too slow and acceleration is zero THEN apply maximum force

IF velocity error means too slow and acceleration is positive THEN apply maximum force

Modern washing machines are quite often programmed to provide the optimum washing conditions using fuzzy logic. The inputs used to provide the facts for the fuzzy control can be the weight of the washing load, the type of fabric being washed and the degree of dirtiness of the load. The outputs can be the amount of detergent, the washing time, water level and temperature. One of the rules might then be something like:

IF amount of dirt is low and fabric type is less sensitive and weight is low THEN wash time is low and spin speed of machine is low and amount of water little and water temperature low

Fuzzy logic can also be used to implement a controller such as the conventional PD, PI or PID. For example, for a PD controller, fuzzification can be used to associate the linguistic terms positive, negative etc. with corresponding input values. Thus a positive error is one where the set value is greater than the feedback value and a

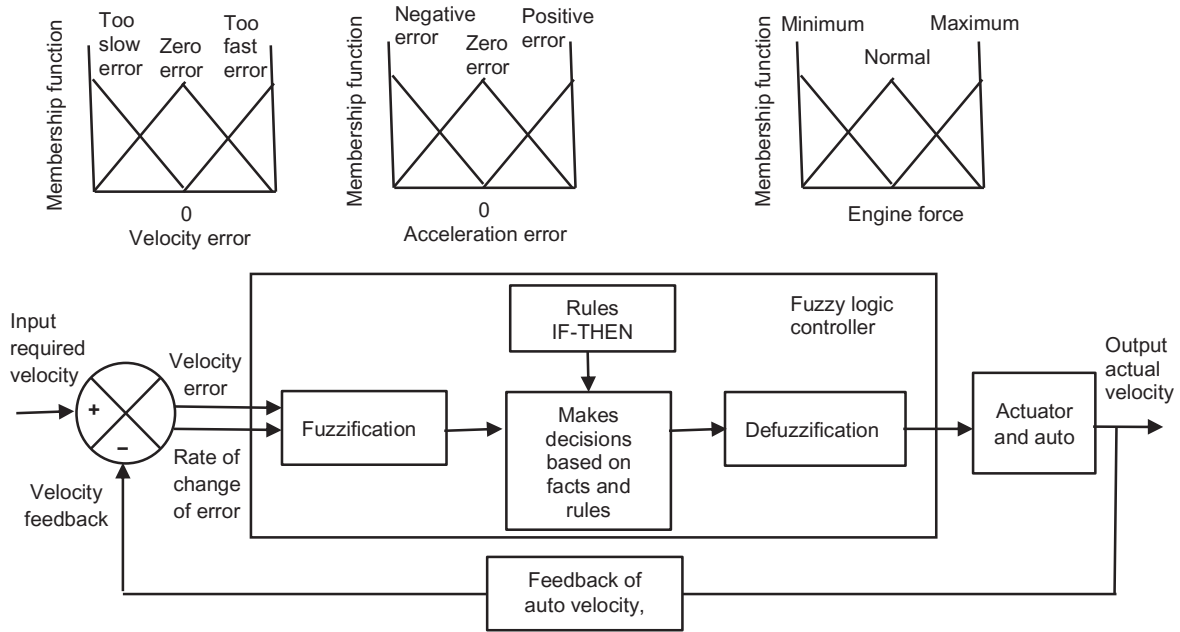


FIGURE 5.42 Illustrative example of a fuzzy logic controller in the feedback loop for an automobile cruise control system.

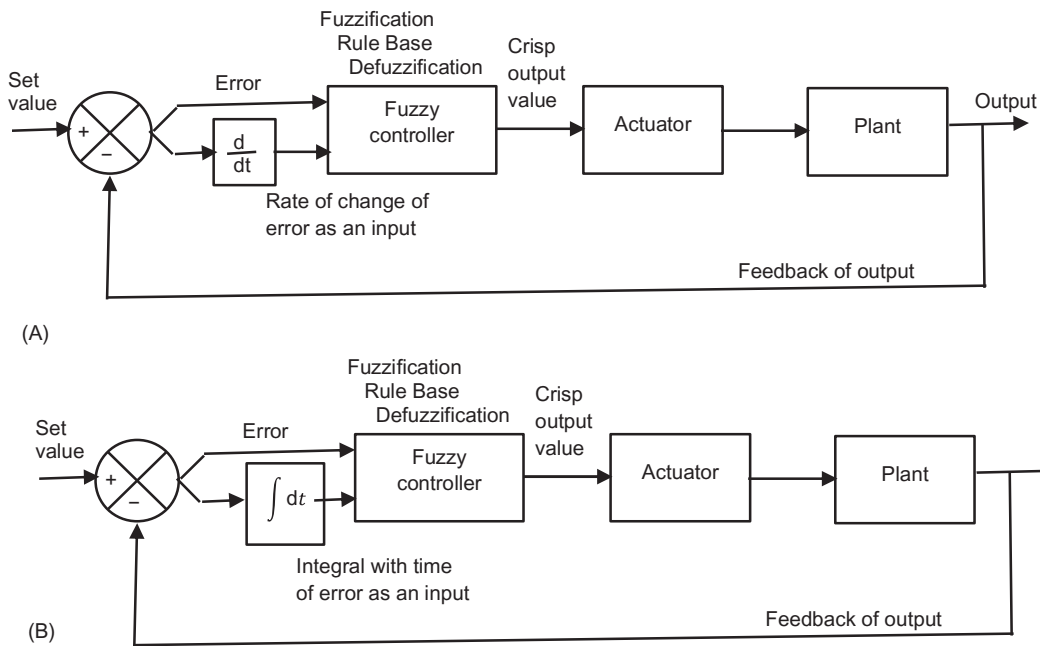


FIGURE 5.43 (A) Fuzzy PD control system and (B) fuzzy PI control system.

negative one where it is less than the feedback value. A positive rate of change of error is where the error is increasing and a negative where the error is decreasing. The control rules might then include:

- IF the error is positive and the rate of change of error is positive THEN the controller output is zero.
- IF the error is positive and the rate of change of error is negative THEN the controller output is positive.
- IF the error is negative and the rate of change of error is positive THEN the controller output is zero.
- IF the error negative and the rate of change of error is negative THEN the controller output is zero.

Figure 5.43A illustrates such a PD control system and Figure 5.43B a PI control system.

Fuzzy logic programs, e.g. MATLAB's Fuzzy Logic Toolbox, can be used to simulate fuzzy logic control systems and analyse the results.

5.9.4 Fuzzy Logic Tuning of PID Controllers

Modern PID controllers generally include the ability to auto-tune the PID parameters when in use (see Section 5.6.1). Such self-tuning may be needed when plant parameters or operating conditions change. Fuzzy logic can be used to give such auto-tuning. Rules which can be used by the fuzzy system might include:

IF the steady-state error is large THEN increase the proportional gain

IF the response is oscillatory THEN increase derivative gain

IF the response is sluggish THEN increase the proportional gain

IF the overshoot is too large THEN decrease the proportional gain

Compared with the conventional PID controller, a self-tuning PID controller can give better performance for both transient and steady-state response.

5.10 NEURAL NETWORKS

The conventional methods of control discussed earlier in this chapter rely on mathematical models in order to implement control. With fuzzy control, we rely on a set of rules in order to implement control. Another control method that is becoming increasingly used is neural networks. These are mathematical models inspired by our understanding of the way in which the human brain operates. The human brain contains nerve cells, termed neurons, which are interconnected by links so that inputs to some neurons can be passed to other neurons in the network. Out of processing of inputs by such a network of neurons emerges the human mind that is capable of analysing and solving complex situations. *Artificial neural networks* (ANNs) have been developed which can begin to analyse input data and solve problems in a similar manner. Such ANNs are made up of 'neurons' connected by links. Each 'neuron' can input data and perform simple operations on the data with the results being passed via links to other 'neurons' to carry out further operations. Figure 5.44 illustrates this.

A neural network is made up of input neurons, hidden neurons and output neurons. Input neurons are designed to receive information from the outside world and pass it into the network so that the network can process it. Output neurons are on the opposite side of the network and signal how the network has responded to the information it has received. In-between the input neurons and output neurons is a layer of hidden neurons, generally just one such hidden layer as in Figure 5.41. These hidden neurons are connected to neurons in the layers on either side. Incoming information thus passes from the input neurons through connection links to neurons in the hidden layer and then passed from them through connection links to neurons in the outer layer. Each connection link has a weight factor that multiplies the signal it is transmitting. The higher the weight, the more influence one neuron has on a linked neuron. Thus for a neuron in the hidden layer, it will have input links from various input neurons and will receive signals from them (Figure 5.45). Each of these signals will be multiplied by a weight associated with the link concerned. The sum of these weighted link signals is then the input to the hidden-layer neuron. This hidden-layer neuron thus performs the weighted sum of its inputs and compares this to its threshold, if it is exceeded it gives an output of 1 otherwise 0. Figure 5.45 shows one form of threshold function, other forms exist. The issue is then how do we determine the weighting factors to use to obtain firing?

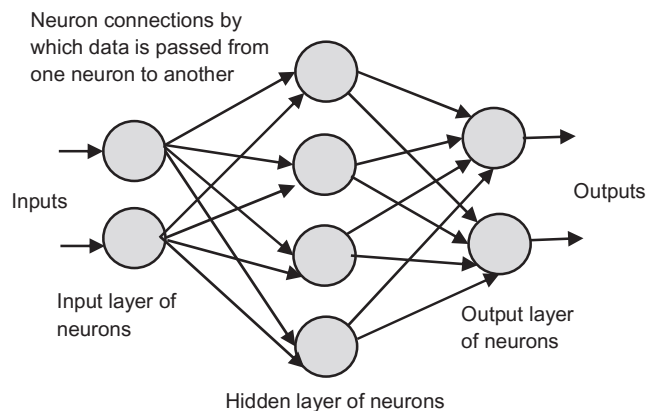


FIGURE 5.44 The basic elements of a simple neural network.

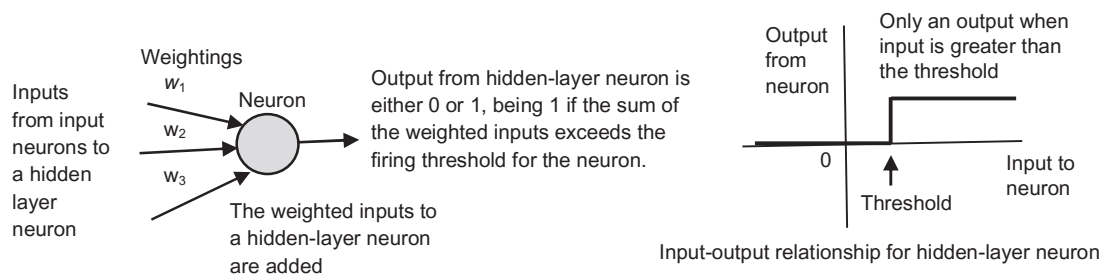


FIGURE 5.45 Inputs and output from a hidden-layer neuron.

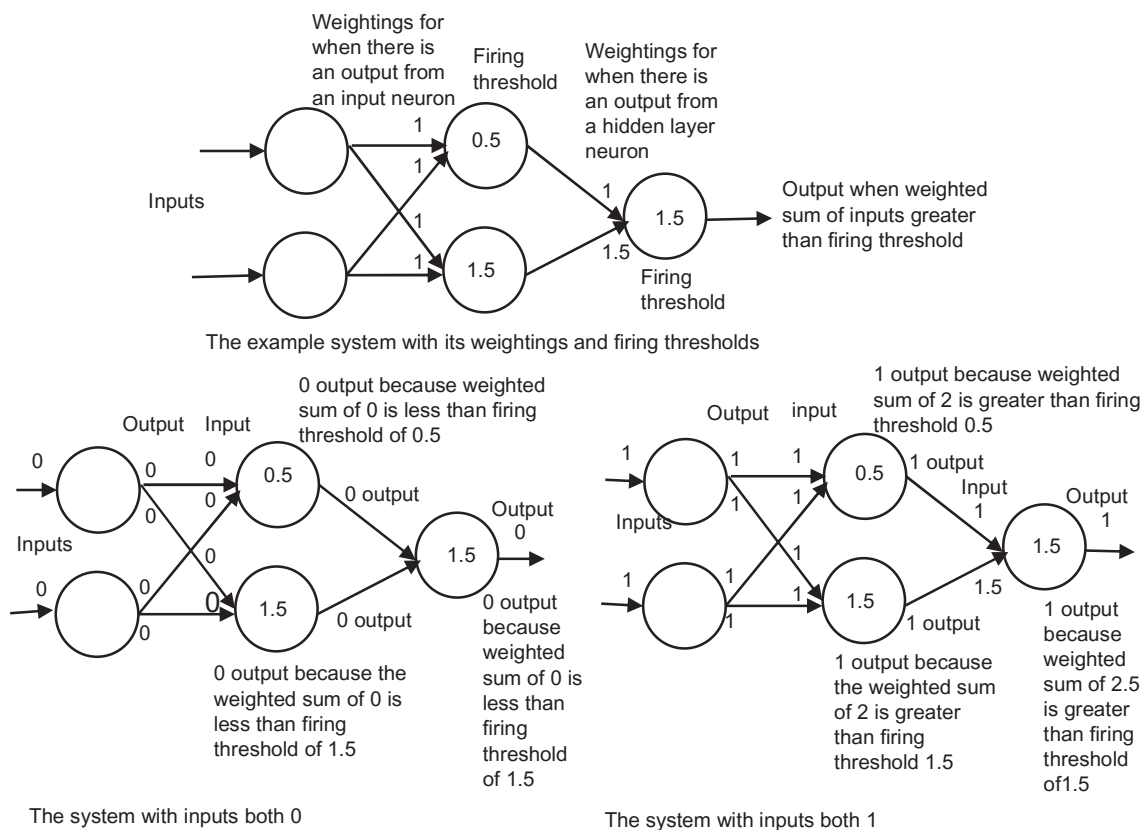


FIGURE 5.46 Illustration of the function of hidden-layer neurons.

Supervised learning can be used with the information being supplied to the neurons in the form of a training set consisting of known input–output pairs. The weight of the connection between any two neurons is then gradually adjusted as the weights are changed so that the actual output of the system corresponds to the desired output. The network learns. Thus when a neural network is learning, or operating normally after being trained, patterns of information are fed into the network via the input neurons. As an illustration of the function of the hidden neurons, consider the example given by Figure 5.46 with two input neurons feeding two hidden-layer neurons which in turn feed a single output neuron.

Neural networks need to be trained from examples before they can be used. Neural networks learn by a feedback process that involves comparing the output a network produces with the output it was meant to produce and using the difference between them to modify the weights of the connections between the units in the network. They ‘learn’ as more and more data is analysed by them and their outcomes compared to the real answers, so establishing when a pathway produces a correct answer and when it does not. The pathway can then be remembered for use when the same inputs occur again. Once the network has been trained with enough learning

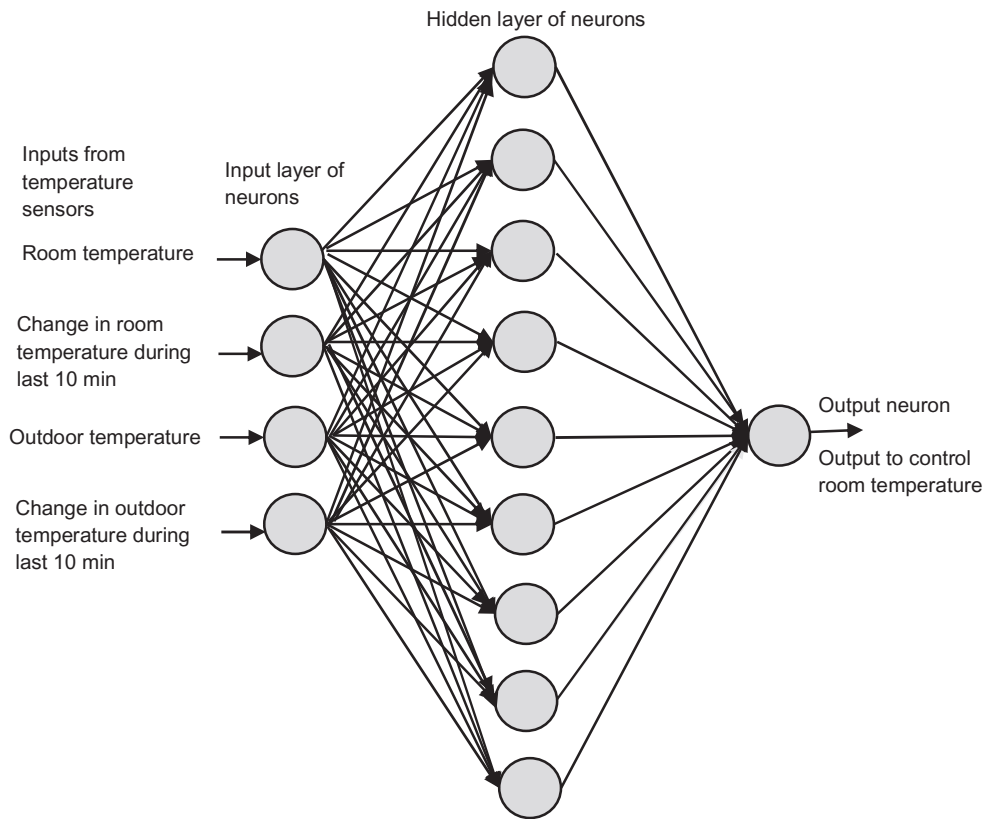


FIGURE 5.47 A neural controller as heater controller for a room.

examples, it reaches a point where it can be presented with an entirely new set of inputs it is never seen before and its response observed.

5.10.1 Neural Networks for Control

Neural networks find an application in a control system as a controller. When used in this manner, the neural network is said to be in *direct neural control*. In a conventional on/off control system the on and off conditions are fixed and do not change if conditions change. If, however, a neural network is used as the controller, it can operate as an on/off controlling switch and be trained to recognise when conditions change and adjust its control action accordingly. It can be considered to act rather like a relay on–off controller (see Section 5.2.1). For example, a heater control system for a room might have a neural controller with four input neurons, one for the room temperature, one for the change in room temperature during the last 10 minutes, one for outdoor temperature and one for the change in outdoor temperature during the last 10 minutes (Figure 5.47). After processing by a single layer of nine hidden-layer neurons, there is just one output neuron which supplies a signal to change the room temperature in an on/off control. The neural network controller must be trained to provide the desired control actions for a wide range of variability in its inputs.

In the above-mentioned discussion the neural network is the controller in the control network. Another way a neural network can be used for control is for it to model the system being controlled and so predict the plant behaviour with time. The predictions are then used to determine the control signals to be used with the actual system. With the actual system the controller itself can be a conventional controller, but its parameters are derived from a system that is modelled by a neural network. This is known as *indirect neural control*. Neural networks used in this way are particularly useful when the mathematical model of the system is not known, and the behaviour of the system is only known in the form of sample performance data.

PROBLEMS

Questions 1–20 have four answer options: A, B, C and D. Choose the correct answer from the answer options.

1. Decide whether each of these statements is True (T) or False (F).
 An on–off temperature controller must have:
 - i. An error signal input which switches the controller on or off.
 - ii. An output signal to switch on or off the correction element.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
2. Decide whether each of these statements is True (T) or False (F).
 Oscillations of the variable being controlled occur with on–off temperature controller because:
 - i. There is a time delay in switching off the correction element when the variable reaches the set value.
 - ii. There is a time delay in switching on the correction element when the variable falls below the set value.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
3. Decide whether each of these statements is True (T) or False (F).
 With a proportional controller:
 - i. The controller output is proportional to the error.
 - ii. The controller gain is proportional to the error.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
4. A steady state error will not occur when there is a change to the set value with a control system operating in the mode:
 - A. Proportional
 - B. Proportional plus derivative
 - C. Derivative
 - D. Proportional plus integral

Questions 5–8 concern the error input to a controller shown in Figure 5.48A and the possible controller outputs shown in Figure 5.48B.
5. Which one of the outputs could be given by a proportional controller?

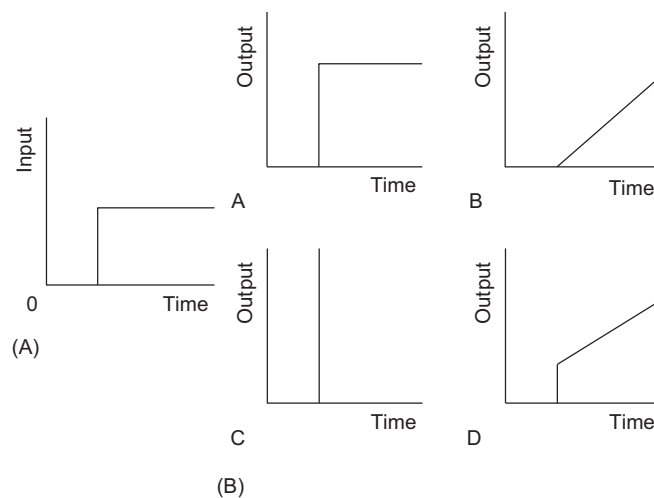


FIGURE 5.48 Problems 5–8.

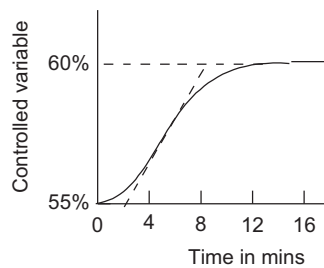


FIGURE 5.49 Problem 10.

6. Which one of the outputs could be given by a derivative controller?
7. Which one of the outputs could be given by an integral controller?
8. Which one of the outputs could be given by a proportional plus integral controller?

9. Decide whether each of these statements is True (T) or False (F).
 With a PID process control system tested at start-up using the ultimate cycle method with the derivative mode turned off and the integral mode set to its lowest setting, the period of oscillation was found to be 10 min with a proportional gain setting of 2. The optimum settings, using the criteria of Ziegler and Nichols, will be:
 - i. An integral constant of 0.2 min^{-1} .
 - ii. A proportional gain setting of 1.2.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
10. Decide whether each of these statements is True (T) or False (F).
 With a PI process control system tested by the process reaction method and the controller output changed by 10%, the response graph obtained was as shown in Figure 5.49. The optimum settings, using the criteria of Ziegler and Nichols, will be:
 - i. Proportional gain constant 7.2.
 - ii. Integral constant 0.15 min^{-1} .
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
11. Decide whether each of these statements is True (T) or False (F).
 With a PI process control system tested at start-up using the ultimate cycle method with the derivative mode turned off and the integral mode set to its lowest setting, the period of oscillation was found to be 20 min with a proportional gain setting of 1.2. The optimum settings, using the criteria of Ziegler and Nichols, will be:
 - i. An integral constant of 0.06 min^{-1} .
 - ii. A proportional gain setting of 1.2.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
12. A control system is designed to control temperatures between -10°C and $+30^{\circ}\text{C}$. What is (a) the range, (b) the span?
13. A temperature control system has a set point of 20°C and the measured value is 18°C . What is (a) the absolute deviation, (b) the percentage deviation?
14. What is the controller gain of a temperature controller with a 80% PB if its input range is $40-90^{\circ}$ and its output is $4-20 \text{ mA}$?

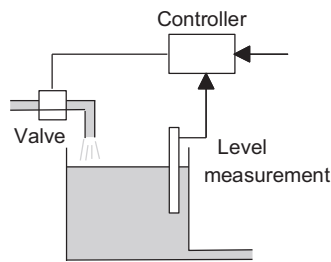


FIGURE 5.50 Problem 16.

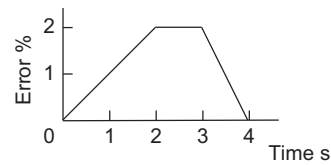


FIGURE 5.51 Problem 17.

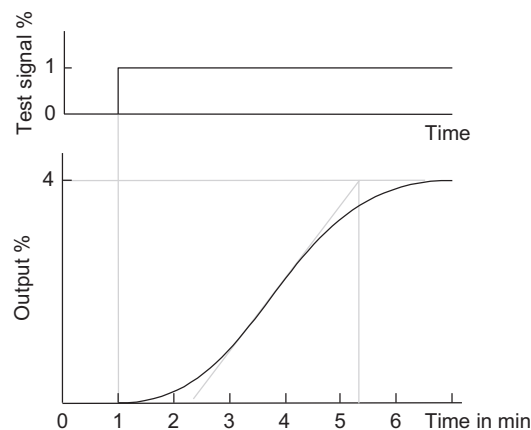


FIGURE 5.52 Problem 20.

15. A controller gives an output in the range 4–20 mA to control the speed of a motor in the range 140–600 rev/min. If the motor speed is proportional to the controller output, what will be the motor speed when the controller output is (a) 8 mA, (b) 40%?
16. Figure 5.50 shows a control system designed to control the level of water in the container to a constant level. It uses a proportional controller with K_p equal to 10. The valve gives a flow rate of $10 \text{ m}^3/\text{h}$ per percent of controller output, its flow rate being proportional to the controller input. If the controller output is initially set to 50% what will be the outflow from the container? If the outflow increases to $600 \text{ m}^3/\text{h}$, what will be the new controller output to maintain the water level constant?
17. Sketch graphs showing how the controller output will vary with time for the error signal shown in Figure 5.51 when the controller is set initially at 50% and operates as (a) just proportional with $K_p = 5$, (b) proportional plus derivative with $K_p = 5$ and $K_d = 1.0 \text{ s}$, (c) proportional plus integral with $K_p = 5$ and $K_i = 0.5 \text{ s}^{-1}$.
18. Using the Ziegler–Nichols ultimate cycle method for the determination of the optimum settings of a PID controller, oscillations began with a 30% proportional band and they had a period of 11 min. What would be the optimum settings for the PID controller?
19. Using the Ziegler–Nichols ultimate cycle method for the determination of the optimum settings of a PID controller, oscillations began with a gain of 2.2 with a period of 12 min. What would be the optimum settings for the PID controller?
20. Figure 5.52 shows the open-loop response of a system to a unit step in controller output. Using the Ziegler–Nichols data, determine the optimum settings of the PID controller.
21. Domestic washing machines are increasingly using fuzzy logic control to implement the washing of dirty clothes. Suggest fuzzy logic rules that might be used.
22. Suggest fuzzy logic rules that might be used for a program for the motion control of a mobile robot on a sloping rough terrain.

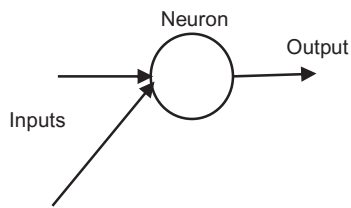


FIGURE 5.53 Problem 23.

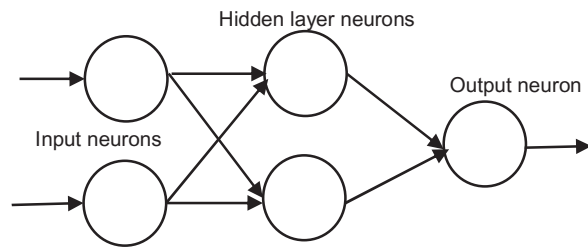


FIGURE 5.54 Problem 24.

23. Suggest a possible specification for a neuron (Figure 5.53), which has two inputs and is to give an output which has the neuron acting as an AND gate for those inputs. An AND gate will have an output of 1 only when both inputs are 1, otherwise, the output will be 0.
24. Suggest a specification for a neuron network with two neuron input layers, two neuron hidden layers and one neuron output (Figure 5.54) which will act as an XOR system and give 0 output when both inputs are 0 and when both inputs are 1 but when one input is 0 and the other is 1 will give a 1 output.

6

Correction Elements

O U T L I N E

6.1 Introduction	137		
6.1.1 The Range of Actuators	138	6.4.5 Other Forms of Flow Control Valves	151
6.2 Pneumatic and Hydraulic Systems	138	6.4.6 Fail-Safe Design	152
6.2.1 Current to Pressure Converter	138	6.5 Motors	152
6.2.2 Pressure Sources	138	6.5.1 D.C. Motors	152
6.2.3 Control Valves	139	6.5.2 Brushless Permanent Magnet D.C. Motor	154
6.2.4 Actuators	140	6.5.3 Stepper Motor	154
6.3 Directional Control Valves	141	6.6 Case Studies	159
6.3.1 Sequencing	143	6.6.1 A Liquid Level Process Control System	159
6.3.2 Shuttle Valve	145	6.6.2 Milling Machine Control System	159
6.4 Flow Control Valves	146	6.6.3 A Robot Control System	159
6.4.1 Forms of Plug	147	Problems	160
6.4.2 Rangeability and Turndown	149		
6.4.3 Control Valve Sizing	149		
6.4.4 Valve Positioners	151		

6.1 INTRODUCTION

The *correction element* or *final control element* is the element in a control system which is responsible for transforming the output of a controller into a change in the process and so correct the change in the controlled variable. Thus, for example, it might be a valve which is operated by the output from the controller and used to change the rate at which liquid passes along a pipe and so change the controlled level of the liquid in a cistern. It might be a motor which takes the electrical output from the controller and transforms it into a rotatory motion in order to move a load and so control its position. It might be a switch which is operated by the controller and so used to switch on a heater to control temperature.

The term *actuator* is used for the part of a correction/final control element that provides the power, i.e. the bit which moves, grips or applies forces to an object, to carry out the control action. Thus a valve might have an input from the controller and be used to vary the flow of a fluid along a pipe and so make a piston move in a cylinder and result in linear motion. The piston–cylinder system is termed an actuator.

6.1.1 The Range of Actuators

The following are some of the types of actuators that can be encountered in control systems:

- *Hydraulic*: These consist of a cylinder along which a piston can be moved by the action of pressure applied by means of an hydraulic fluid.
- *Pneumatic*: These consist of a cylinder along which a piston can be moved by the action of pressure applied by air pressure.
- *Electric*: These use electrical energy to power a motor and hence actuate equipment.
- *Mechanical*: These involve gears, pulleys, chains, screws and other mechanical devices to execute movement by converting rotary motion into linear motion. For example, a screw can be used to transform circular to linear motion.

In this chapter, pneumatic/hydraulic and electric correction control elements are discussed.

6.2 PNEUMATIC AND HYDRAULIC SYSTEMS

Process control systems frequently require control of the flow of a fluid. The valves used as the correction elements in such situations are frequently pneumatically operated, even when the control system is otherwise electrical. This is because such pneumatic devices tend to be cheaper and more easily capable of controlling large rates of flow. The main drawback with pneumatic systems is, however, the compressibility of air. This makes it necessary to have a storage reservoir to avoid changes in pressure occurring as a result of loads being applied. Hydraulic signals do not have this problem and can be used for even higher power control devices. They are, however, expensive and there are hazards associated with oil leaks which do not occur with air leaks.

6.2.1 Current to Pressure Converter

Generally the signals required by a pneumatic correction element are in the region of 20–100 kPa gauge pressure, i.e. pressure above the atmospheric pressure. Figure 6.1 shows the principle of one form of a *current to pressure converter* that can be used to convert a current output from a controller, typically in the range 4–20 mA, to a pneumatic pressure signal of 20–100 kPa to operate a final control element. The current from the controller passes through coils mounted on a pivoted beam. As a consequence, the coils are then attracted towards a magnet, the extent of the attraction depending on the size of the current. The movement of the coils causes the lever to rotate about its pivot and so change the separation of a flapper from a nozzle. The position of the flapper in relation to the nozzle determines the size of the output pressure in the system.

6.2.2 Pressure Sources

With a pneumatic system, a source of pressurised air is required. This can be provided by an electric motor driving an air compressor (Figure 6.2). The air is drawn from the atmosphere via a filter. Since the air compressor increases the temperature of the air, a cooling system is likely to follow and, since air also contains a significant

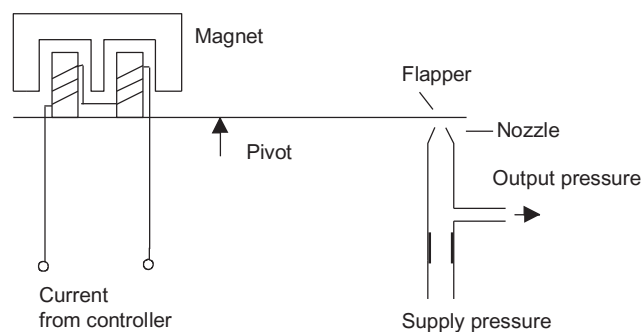


FIGURE 6.1 Current to pressure converter.

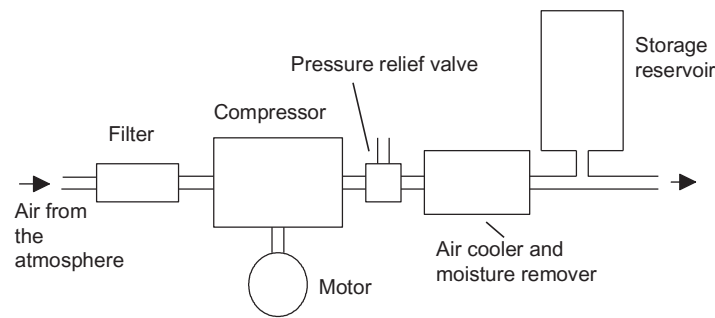


FIGURE 6.2 A pressurised air source.

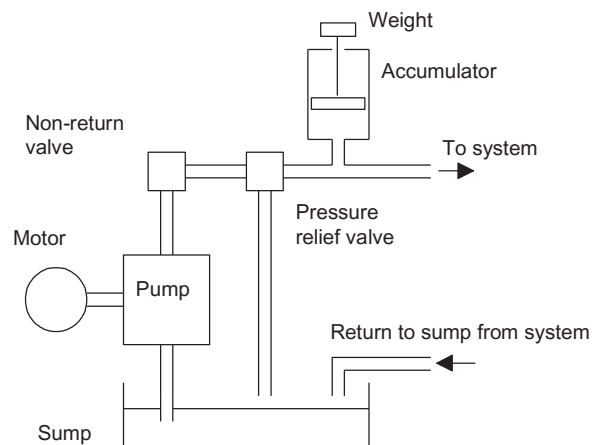


FIGURE 6.3 A source of pressurised oil.

amount of moisture, a moisture separator to remove the moisture from the air. A storage reservoir, by virtue of its volume, is used to smooth out any pressure fluctuations due to the compressibility of air. A pressure relief valve provides protection against the pressure in the system rising above a safe level.

With a hydraulic system, a source of pressurised oil is required. This can be provided by a pump driven by an electric motor. The pump pumps oil from a sump through a non-return valve and an accumulator and back to the sump (Figure 6.3). The non-return valve is to prevent the oil being back-driven to the pump. A pressure relief valve is included so that the pressure is released if it rises above a safe level. The accumulator is essentially just a container in which the oil is held under pressure against an external force and is there to smooth out any short-term fluctuations in the output oil pressure. If the oil pressure rises then the piston moves to increase the volume the oil can occupy and so reduces the pressure. If the oil pressure falls then the piston moves in to reduce the volume occupied by the oil and so increase its pressure.

6.2.3 Control Valves

Pneumatic and hydraulic systems use control valves to give direction to the flow of fluid through a system, control its pressure and control the rate of flow. These types of valve can be termed *directional control valves*, *pressure control valves* and *flow control valves*. Directional control valves, sometimes termed *finite position valves* because they are either completely open or completely closed, i.e. they are on/off devices, are used to direct fluid along one path or another. They are equivalent to electric switches which are either on or off. Pressure control valves, often termed pressure regulator valves, react to changes in pressure in switching a flow on or off, or varying it. Flow control valves, sometimes termed *infinite position valves*, vary the rate at which a fluid passes through a pipe and are used to regulate the flow of material in process control systems. Valves are discussed in more detail later in this chapter.

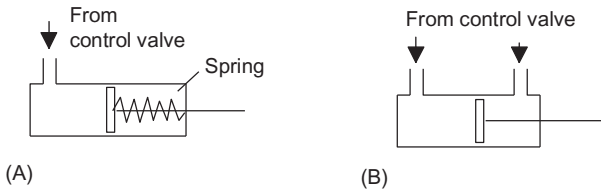


FIGURE 6.4 (A) Single acting and (B) double acting cylinder.

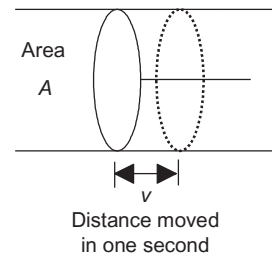


FIGURE 6.5 Movement of a piston in a cylinder.

6.2.4 Actuators

Fluid power actuators can be classified into two groups: *linear actuators* which are used to move an object or apply a force in a straight line and *rotary actuators* which are used to move an object in a circular path.

The *hydraulic* or *pneumatic cylinder* is a linear actuator, the principles and form being the same for both versions with the differences being purely a matter of size as a consequence of the higher pressures used with hydraulics. The hydraulic/pneumatic cylinder consists of a hollow cylindrical tube along which a piston can slide. Figure 6.4A shows the single acting form and Figure 6.4B the double acting form. The *single acting* form has the control pressure applied to just one side of the piston, a spring often being used to provide the opposition to the movement of the piston. The piston can only be moved in one direction along the cylinder by the signal from the controller. The *double acting* form has control pressures that can be applied to each side of the piston. When there is a difference in pressure between the two sides the piston moves, the piston being able to move in either direction along the cylinder. As an illustration of what is available commercially, the Parker Pneumatics PID series have single acting or double acting cylinders in a range of bores, the standard version being from 32 to 125 mm, and the operating pressure is a maximum of 10 bar. There is a choice of piston material and whether it is to be cushioned at the end of its travel. Cushioning is where the air flow out of the cylinder becomes restricted when the piston nears the end of its travel and this slows down the movement of the piston.

The choice of cylinder is determined by the force required to move the load and the speed required. Hydraulic cylinders are capable of much larger forces than pneumatic cylinders. However, pneumatic cylinders are capable of greater speeds.

Since pressure is force per unit area, the force produced by a piston in a cylinder is equal to the cross-sectional area of the piston, this being effectively the same as the internal cross-sectional area of the cylinder, multiplied by the difference in pressure between the two sides of the piston. Thus for a pneumatic cylinder with a pressure difference of 500 kPa and having an internal diameter of 50 mm,

$$\text{force} = \text{pressure} \times \text{area} = 500 \times 10^3 \times \frac{1}{4}\pi \times 0.050^2 = 982 \text{ N}$$

A hydraulic cylinder with the same diameter and a pressure difference of 15,000 kPa, hydraulic cylinders being able to operate with higher pressures than pneumatic cylinders, will give a force of 2.95 kN. Note that the maximum force available is not related to the flow rate of hydraulic fluid or air into a cylinder but is determined solely by the pressure and piston area.

The speed with which the piston moves in a cylinder is determined by the rate at which fluid enters the cylinder. If the flow rate of hydraulic liquid into a cylinder is a volume of Q per second, then the piston must sweep out a volume of Q . If a piston moves with a velocity v then, in 1 s, it moves a distance of v (Figure 6.5).

But for a piston of cross-sectional area A , this must mean that the volume swept out by the piston in 1 s is Av . Thus we must have

$$Q = Av$$

Thus the speed v of a hydraulic cylinder is equal to the flow rate of liquid Q through the cylinder divided by the cross-sectional area A of the cylinder. The speed is determined by just the piston area and the flow rate. For example, for a hydraulic cylinder of diameter 50 mm and a hydraulic fluid flow of $7.5 \times 10^{-3} \text{ m}^3/\text{s}$:

$$\text{speed } v = \frac{Q}{A} = \frac{7.5 \times 10^{-3}}{\frac{1}{4}\pi \times 0.050^2} = 3.8 \text{ m/s}$$

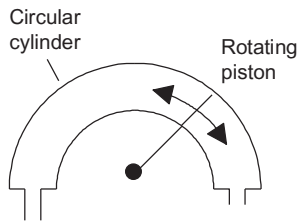


FIGURE 6.6 Rotary actuator.

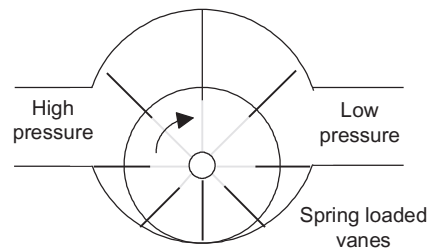


FIGURE 6.7 Vane motor.

Rotary actuators give rotary motion as a result of the applied fluid pressure. Figure 6.6 shows a rotary actuator which gives partial rotary movement. Continuous rotation is possible with some forms and then they are the equivalent of electric motors. Figure 6.7 shows one form known as a *vane motor*. The vanes are held out against the walls of the motor by springs or hydraulic pressure. Thus, when there is a pressure difference between the inlet and outlet of the motor, rotation occurs.

EXAMPLE

A hydraulic cylinder is to be used in a manufacturing operation to move a workpiece through a distance of 250 mm in 20 s. If a force of 50 kN is required to move the workpiece, what is the required pressure difference and hydraulic liquid flow rate if a cylinder with a piston diameter of 150 mm is to be used?

As derived above, the force produced by the cylinder is equal to the product of the cross-sectional area of the cylinder and the working pressure. Thus the required pressure is:

$$\text{pressure} = \frac{F}{A} = \frac{50 \times 10^3}{\frac{1}{4}\pi \times 0.150^2} = 2.8 \times 10^6 \text{ Pa} = 2.8 \text{ MPa}$$

The average speed required is $250/20 = 12.5$ mm/s. As derived above, the speed of a hydraulic cylinder is equal to the flow rate of liquid through the cylinder divided by the cross-sectional area of the cylinder. Thus the required flow rate is:

$$\text{flow rate} = \text{speed} \times \text{area} = 0.0125 \times \frac{1}{4}\pi \times 0.150^2 = 2.2 \times 10^{-4} \text{ m}^3/\text{s}$$

6.3 DIRECTIONAL CONTROL VALVES

Directional control valves are widely used in control systems as elements for switching on or off hydraulic or pneumatic pressures which can then, via some actuator, control the movement of some item. A *directional control valve* on the receipt of some external signal, which might be a mechanical, electrical or pressure signal, changes the direction of, or stops, or starts the flow of fluid in some part of a pneumatic/hydraulic circuit.

The basic symbol for a control valve is a square. With a directional control valve, two or more squares are used with each square representing the positions to which the valve can be switched. Thus, Figure 6.8A represents a valve with two switching positions, Figure 6.8B a valve with three switching positions. Lines in the boxes are used to show the flow paths with arrows indicating the direction of flow (Figure 6.9A) and shut-off positions indicated by terminated lines (Figure 6.9B). The pipe connections, i.e. the inlet and outlet ports of the valve, are indicated by lines drawn on the outside of the box and are drawn for just the 'rest/initial/neutral position', i.e. when the valve is not actuated (Figure 6.9C). You can imagine each of the position boxes to be moved by the action of some actuator so that it connects up with the pipe positions to give the different connections between the ports.

Directional control valves are described by the number of ports and the number of positions. Thus, a 2/2 valve has 2 ports and 2 positions, a 3/2 valve 3 ports and 2 positions, a 4/2 valve 4 ports and 2 positions, a 5/3 valve 5 ports and 3 positions. Figure 6.10 shows some commonly used examples and their switching options and Figure 6.11 the means by which valves can be switched between positions.

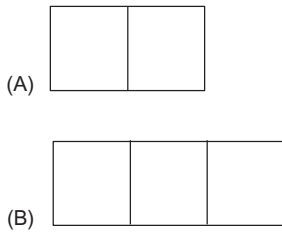


FIGURE 6.8 (A) Two position and (B) three position valves.

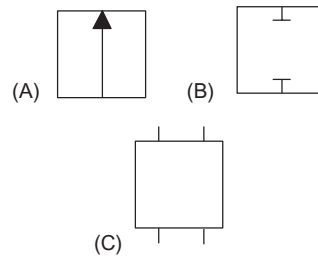


FIGURE 6.9 (A) Flow path, (B) shut-off and (C) input connections.

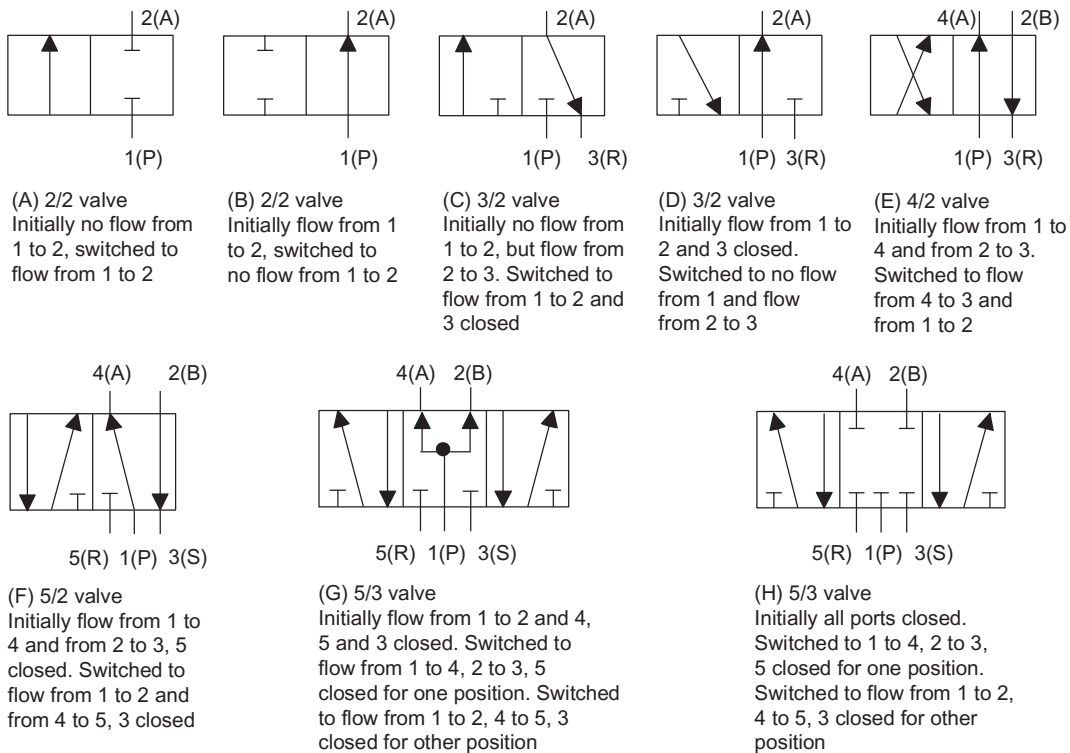


FIGURE 6.10 Commonly used direction valves: P or 1 indicates the pressure supply ports, R and S or 3 and 5 the exhaust ports, A and B or 2 and 4 the signal output ports.

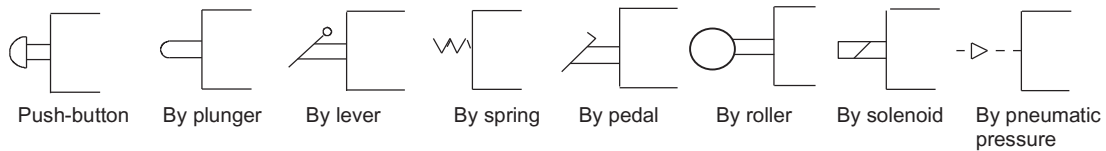


FIGURE 6.11 Examples of valve actuation methods.

As an illustration, Figure 6.12 shows the symbol for a 3/2 valve with solenoid activation and return by means of a spring. Thus, when the solenoid is not activated by a current through it, the signal port 2 is connected to the exhaust 3 and so is at atmospheric pressure. When the solenoid is activated, the pressure supply P is connected to the signal port 2 and thus the output is pressurised. In Section 4.4.5, an antilock brake system (ABS) for a car is discussed and Figure 4.17 shows how valves are used.

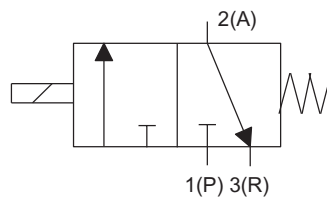


FIGURE 6.12 Symbol for a solenoid-activated valve with return spring.

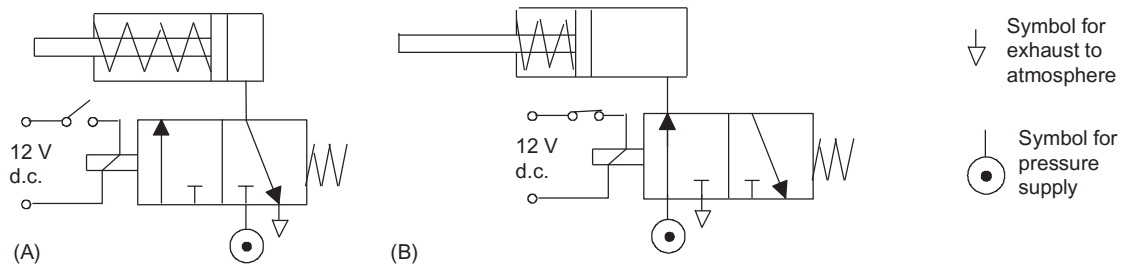


FIGURE 6.13 Control of a single-acting cylinder: (A) before solenoid activated and (B) when solenoid activated.

Figure 6.13 shows how such a valve might be used to cause the piston in a single-acting cylinder to move; the term single-acting is used when a pressure signal is applied to only one side of the piston. When the switch is closed and a current passes through the solenoid, the valve switches position and pressure is applied to extend the piston in the cylinder.

Figure 6.14 shows how a double-solenoid activated valve can be used to control a double-acting cylinder. Momentary closing switch S1 causes a current to flow through the solenoid at the left-hand end of the valve and so result in the piston extending. On opening S1, the valve remains in this extended position until a signal is received by the closure of switch S2 to activate the right-hand solenoid and return the piston.

As an illustration of valves that are available, Continental Hydraulics include in their range VSNG valves from which firstly the basic valve format is selected from: Single operator 2 position spring offset, double operator 3 position spring centre, single operator 2 position spring centred and single operator 2 position 3 way spring operated. A single operator valve is one like the ones in Figure 6.13 which can only have an operator input at one position, double operator have operator inputs possible at both ends, as in Figure 6.14. The next selection is for the spool to be used, i.e. the ports and the connections, as in Figure 6.10.

6.3.1 Sequencing

Situations often occur where it is necessary to activate a number of cylinders in some sequence. Thus event 2 might have to start when event 1 is completed, event 3 when event 2 has been completed. For example, we might have: only when cylinder A is fully extended (event 1) can cylinder B start extending (event 2), and cylinder A can only start retracting (event 3) when cylinder B has fully extended (event 2). In discussions of sequential control, it is common practice to give each cylinder a reference letter A, B, C, D, etc., and to indicate the state of each cylinder by using a + sign if it is extended or a - sign if retracted. Thus a sequence of operations might be shown as A +, B +, A -, B -. This indicates that the sequence of events is cylinder A extends, followed by cylinder B being extended, followed by cylinder A retracting, followed by cylinder B retracting. Figure 6.15 illustrates this with a displacement step diagram. Figure 6.16 shows a circuit that could be used to generate this displacement-event diagram for two cylinders A and B.

In order to generate the displacement-event diagram of Figure 6.15 the sequence of operations with Figure 6.16 is:

Event 1

1. Start push button pressed.
2. Cylinder A extends, releasing limit switch a - .

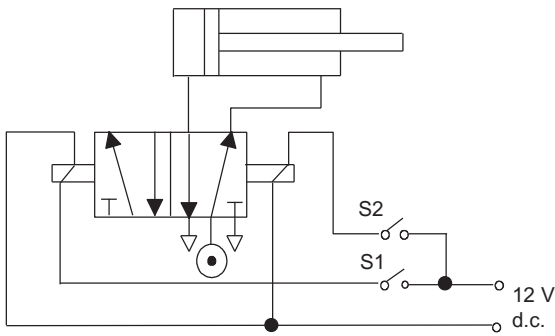


FIGURE 6.14 Control of a double-acting cylinder.

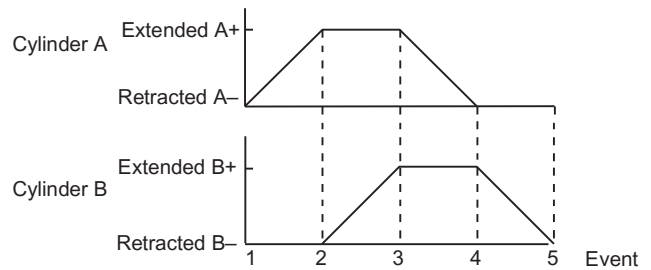


FIGURE 6.15 Displacement-event diagram.

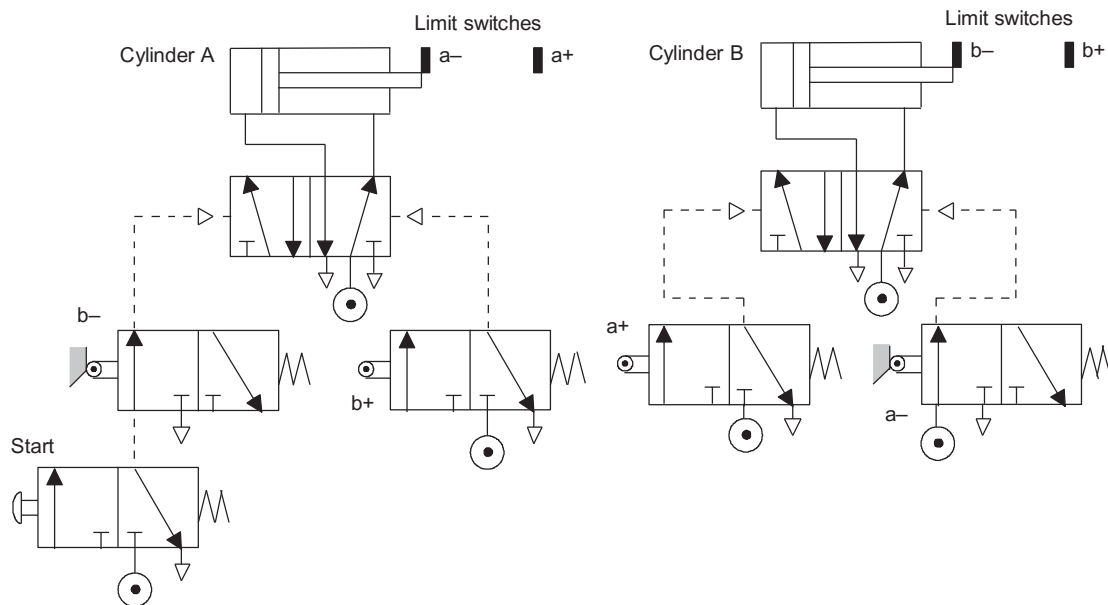


FIGURE 6.16 Two-actuator sequential operation.

Event 2

3. Cylinder A fully extended, limit switch a + operated to start B extending.
4. Cylinder B extends, limit switch b – released.

Event 3

5. Cylinder B fully extended, limit switch b + operated to start cylinder A retracting.
6. Cylinder A retracts, limit switch a + released.

Event 4

7. Cylinder A fully retracted, limit switch a – operated to start cylinder B retracting.
8. Cylinder B retracts, limit switch b + released.

Event 5

9. Cylinder B fully retracted, limit switch b – operated to complete the cycle.

The cycle can be started again by pushing the start button. If we wanted the system to run continuously then the last movement in the sequence would have to trigger the first movement.

As an illustration of the types of problems that sequential control can be used for, consider the control required with an automatic machine to perform a number of sequential actions such as positioning objects, operating clamps and then operating some machine tool (Figure 6.17). This requires the switching in sequence of a number of cylinders, the movements of the cylinder pistons being the mechanisms by which the actions are initiated.

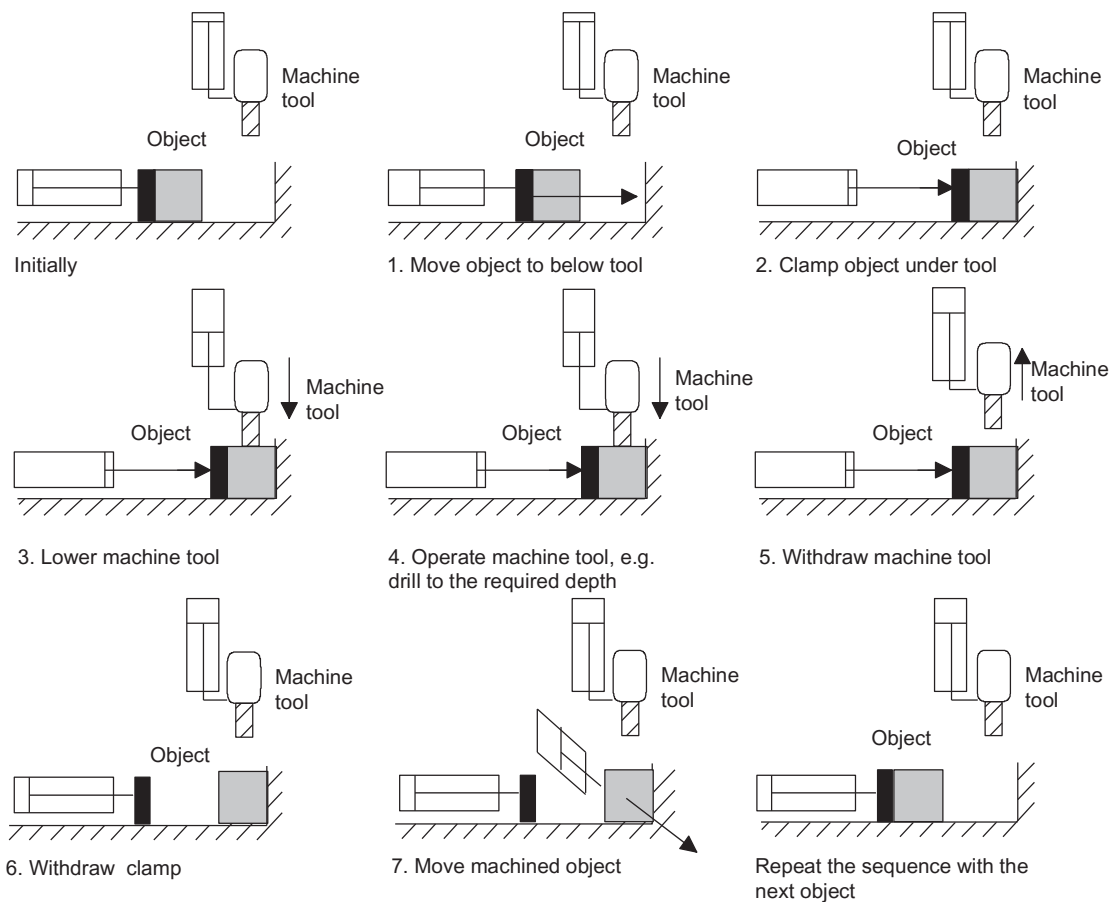


FIGURE 6.17 Operations required for an automatic machine operation.

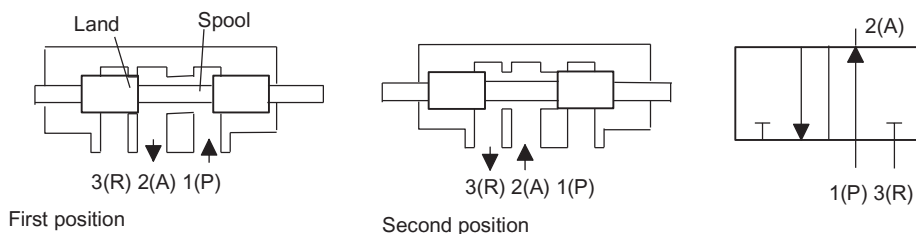


FIGURE 6.18 3/2 shuttle valve.

6.3.2 Shuttle Valve

The most common form of directional control valve is the *shuttle* or *spool valve*. Shuttle valves have a spool moving horizontally within the valve body. Raised areas, termed *lands*, block or open ports to give the required valve operation. Figure 6.18 illustrates these features with a 3/2 valve. In the first position, the shuttle is located so that its lands block off the 3 port and leave open, and connected, the 1 and 2 ports. In the second position, the shuttle is located so that it blocks off the 1 port and leaves open, and connected, the 2 and 3 ports. The shuttle can be made to move between these two positions by manual, mechanical, electrical or pressure signals applied to the two ends of the shuttle.

Figure 6.19 shows an example of a 4/3 shuttle valve. It has the rest position with all ports closed. When the shuttle is moved from left to right, the pressure is applied to output port 2 and port 4 is connected to the exhaust port. When the shuttle moves from right to left, pressure is applied to output port 4 and port 2 is connected to the exhaust port.

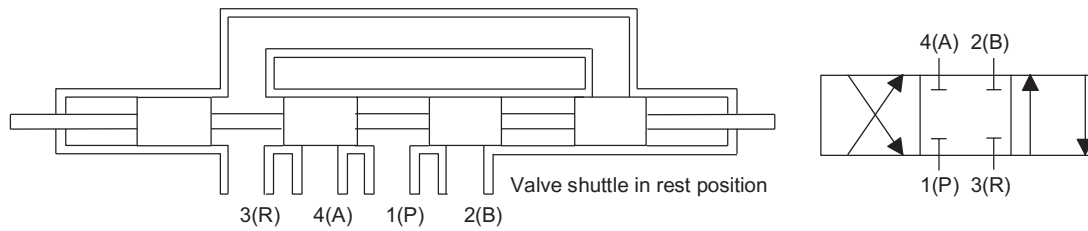


FIGURE 6.19 4/3 shuttle valve.

EXAMPLE

State what happens for the pneumatic circuit shown in Figure 6.20 when the push button is pressed and then released.

The right-hand box shows the initial position with the pressure source, i.e. the circle with the dot in the middle, connected to a closed port and the output from the right-hand end of the cylinder connected to the exhaust port, i.e. the open triangle. When the push button is pressed the connections between the ports become those indicated in the left-hand box. The pressure source is then connected to the output port and hence to the right-hand end of the cylinder and forces the piston back against its spring and so from left to right. When the push button is released, the connections between the ports become those in the right-hand box and the right-hand end of the cylinder is exhausted. The piston then moves back from left to right.

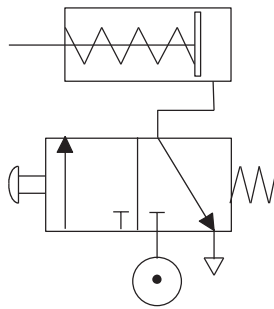


FIGURE 6.20 Example.

6.4 FLOW CONTROL VALVES

In many control systems, the rate of flow of a fluid along a pipe is controlled by a valve which uses pneumatic action to move a valve stem and hence a plug or plugs into the flow path, so altering the size of the gap through which the fluid can flow (Figure 6.21). The term *single seated* is used where just one plug is involved and *double seated* where there are two. A single-seated valve has the advantage compared with the double-seated valve of being able to close more tightly but the disadvantages that the force on the plug is greater from the fluid and so a larger area diaphragm may be needed.

Figure 6.22 shows the basic elements of a common form of such a control valve. The term *globe valve* is used because of their spherical bodies, though some modern valves no longer have much of a spherical shape. The two halves of their body are separated by an internal baffle which has an opening that forms a seat through which a movable plug can be moved to close the valve. The plug is connected to a stem which is moved by an actuator assembly.

The movement of the stem, and hence the position of the plug or plugs in the fluid flow, results from the use of a diaphragm moving against a spring and controlled by air pressure. The air pressure from the controller exerts a force on one side of the diaphragm, the other side of the diaphragm being at atmospheric pressure, which is opposed by the force due to the spring on the other side. When the air pressure changes then the

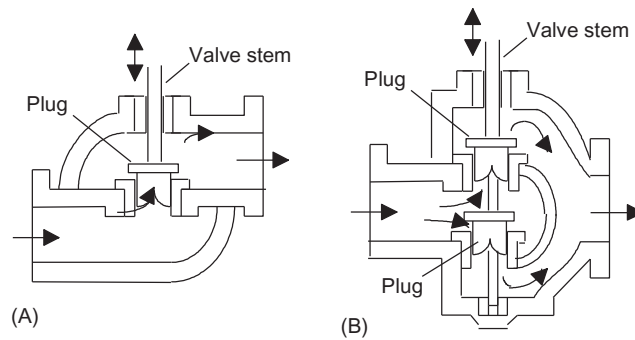


FIGURE 6.21 Body globe valve: (A) single and (B) double seated.

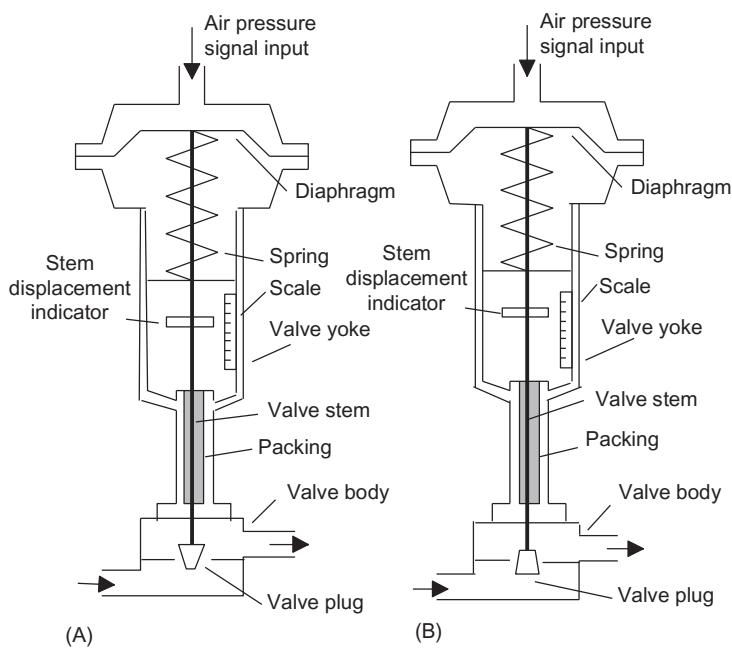


FIGURE 6.22 Direct action: (A) air pressure increase to close and (B) air pressure increase to open.

diaphragm moves until there is equilibrium between the forces resulting from the pressure and those from the spring. Thus the pressure signals from the controller result in the movement of the stem of the valve. There are two alternative forms, *direct* and *reverse action* forms (Figure 6.23) with the difference being the position of the spring. The valve body is joined to the diaphragm element by the *yoke*.

6.4.1 Forms of Plug

There are many forms of valve body and plug. The selection of the form of body and plug determines the characteristic of the control valve, i.e. the relationship between the valve stem position and the flow rate through it. For example, Figure 6.24 shows how the selection of plug can be used to determine whether the valve closes when the controller air pressure increases or opens when it increases and Figure 6.24 shows how the shape of the plug determines how the rate of flow is related to the displacement of the valve stem:

1. Linear plug

The change in flow rate is proportional to the change in valve stem displacement, i.e.

$$\text{change in flow rate} = k(\text{change in stem displacement})$$

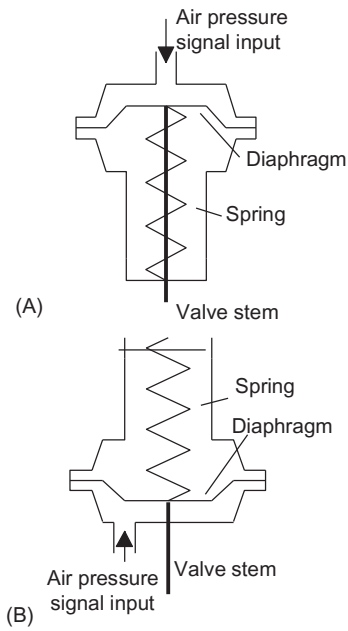


FIGURE 6.23 (A) Direct action and (B) reverse action.

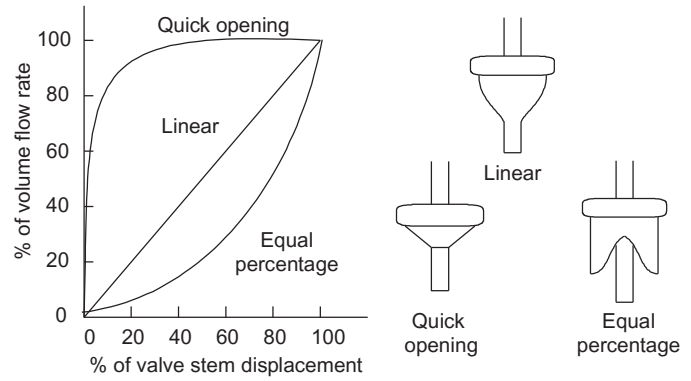


FIGURE 6.24 Effect of plug shape on flow.

where k is a constant. If Q is the flow rate at a valve stem displacement S and Q_{\max} is the maximum flow rate at the maximum stem displacement S_{\max} then we have:

$$\frac{Q}{Q_{\max}} = \frac{S}{S_{\max}}$$

or the percentage change in the flow rate equals the percentage change in the stem displacement. Such valves are widely used for the control of liquids entering cisterns when the liquid level is being controlled.

2. Quick-opening plug

A large change in flow rate occurs for a small movement of the valve stem. This characteristic is used for on-off control systems where the valve has to move quickly from open to closed and vice versa.

3. Equal percentage plug

The amount by which the flow rate changes is proportional to the value of the flow rate when the change occurs. Thus, if the amount by which the flow rate changes is ΔQ for a change in valve stem position ΔS , then it is proportional to the value of the flow Q when the change occurs, i.e.:

$$\frac{\Delta Q}{\Delta S} \propto Q$$

Hence we can write

$$\frac{\Delta Q}{\Delta S} = kQ$$

where k is a constant. Generally this type of valve does not cut off completely when at the limit of its stem travel, thus when $S = 0$ we have $Q = Q_{\min}$. If we write this expression for small changes and then integrate it we obtain:

$$\int_{Q_{\min}}^Q \frac{1}{Q} dQ = k \int_0^S dS$$

Hence

$$\ln Q - \ln Q_{\min} = kS$$

If we consider the flow rate Q_{\max} is given by S_{\max} then

$$\ln Q_{\max} - \ln Q_{\min} = kS_{\max}$$

Eliminating k from these two equations gives

$$\frac{\ln Q - \ln Q_{\min}}{\ln Q_{\max} - \ln Q_{\min}} = \frac{S}{S_{\max}}$$

$$\ln \frac{Q}{Q_{\min}} = \frac{S}{S_{\max}} \ln \frac{Q_{\max}}{Q_{\min}}$$

and so

$$\frac{Q}{Q_{\min}} = \left(\frac{Q_{\max}}{Q_{\min}} \right)^{S/S_{\max}}$$

EXAMPLE

A valve has a stem movement at full travel of 30 mm and has a linear plug which has a minimum flow rate of 0 and a maximum flow rate of 20 m³/s. What will be the flow rate when the stem movement is 15 mm?

The percentage change in the stem position from the zero setting is $(15/30) \times 100 = 50\%$. Since the percentage flow rate is the same as the percentage stem displacement, then a percentage stem displacement of 50% gives a percentage flow rate of 50%, i.e. 10 m³/s.

EXAMPLE

A valve has a stem movement at full travel of 30 mm and an equal percentage plug. This gives a flow rate of 2 m³/s when the stem position is 0. When the stem is at full travel there is a maximum flow rate of 20 m³/s. What will be the flow rate when the stem movement is 15 mm?

Using the equation:

$$\frac{Q}{Q_{\min}} = \left(\frac{Q_{\max}}{Q_{\min}} \right)^{S/S_{\max}}$$

$$\frac{Q}{2} = \left(\frac{20}{2} \right)^{15/30}$$

gives $Q = 6.3 \text{ m}^3/\text{s}$.

6.4.2 Rangeability and Turndown

The term *rangeability* R is used for the ratio Q_{\max}/Q_{\min} , i.e. the ratio of the maximum to minimum rates of controlled flow. Thus, if the minimum controllable flow is 2.0% of the maximum controllable flow, then the rangeability is $100/2.0 = 50$. Valves are often not required to handle the maximum possible flow and the term *turndown* is used for the ratio:

$$\text{turndown} = \frac{\text{normal maximum flow}}{\text{minimum controllable flow}}$$

For example, a valve might be required to handle a maximum flow which is 70% of that possible. With a minimum flow rate of 2.0% of the maximum flow possible, then the turndown is $70/2.0 = 35$.

6.4.3 Control Valve Sizing

The term *control valve sizing* is used for the procedure of determining the correct size, i.e. diameter, of the valve body. A control valve changes the flow rate by introducing a constriction in the flow path. But introducing such

TABLE 6.1 Flow Coefficients and Valve Size

Flow coefficients	Valve size in mm							
	480	640	800	960	1260	1600	1920	2560
C_V	8	14	22	30	50	75	110	200
$A_V \times 10^{-5}$	19	33	52	71	119	178	261	474

a constriction introduces a pressure difference between the two sides of the constriction. The basic equation (from an application of Bernoulli's equation) relating the rate of flow and pressure drop is

$$\text{rate of flow } Q = K\sqrt{\text{pressure drop}}$$

where K is a constant which depends on the size of the constriction produced by the presence of the valve. The equations used for determining valve sizes are based on this equation. For a liquid, this equation is written as

$$Q = A_V \sqrt{\frac{\Delta p}{\rho}} \text{ m}^3/\text{s}$$

where A_V is the *valve flow coefficient*, Δp the pressure drop in Pa across the valve and ρ the density in kg/m^3 of the fluid. Because the equation was originally specified with pressure in pounds per square inch and flow rate in American gallons per minute, another coefficient C_V based on these units is widely quoted. With such a coefficient and the quantities in SI units, we have

$$Q = 2.37 \times 10^{-5} C_V \sqrt{\frac{\Delta p}{\rho}} \text{ m}^3/\text{s}$$

or

$$Q = 0.75 \times 10^{-6} C_V \sqrt{\frac{\Delta p}{G}} \text{ m}^3/\text{s}$$

G is the specific gravity (relative density) and Δp is the pressure difference. Other equations are available for gases and steam. For gases:

$$Q = 6.15 \times 10^{-4} C_V \sqrt{\frac{\Delta p \times p}{TG}} \text{ mm}^3/\text{s}$$

where T is the temperature on the Kelvin scale and p the inlet pressure. For steam:

$$Q = 27.5 \times 10^{-6} C_V \sqrt{\frac{\Delta p}{V}} \text{ kg/s}$$

where V is the specific volume of the steam in m^3/kg , the specific volume being the volume occupied by 1 kg. Table 6.1 shows some typical values of A_V , C_V and the related valve sizes.

EXAMPLE

Determine the valve size for a valve that is required to control the flow of water when the maximum flow rate required is $0.012 \text{ m}^3/\text{s}$ and the permissible pressure drop across the valve at this flow rate is 300 kPa.

Taking the density of water as $1000 \text{ kg}/\text{m}^3$, we have

$$A_V = Q \sqrt{\frac{\rho}{\Delta p}} = 0.012 \sqrt{\frac{1000}{300 \times 10^3}} = 69.3 \times 10^{-5}$$

Thus, using Table 6.1, this value of coefficient indicates that the required valve size is 960 mm.

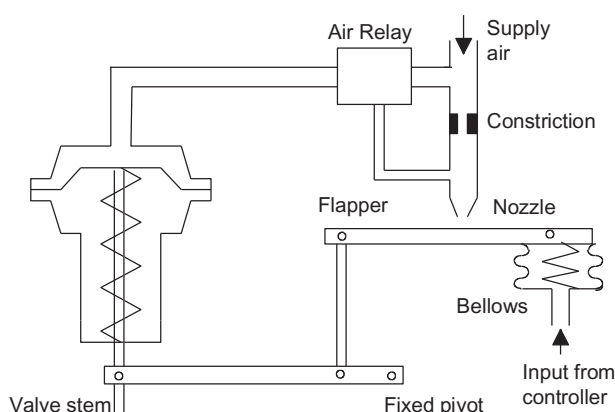


FIGURE 6.25 Valve positioner.

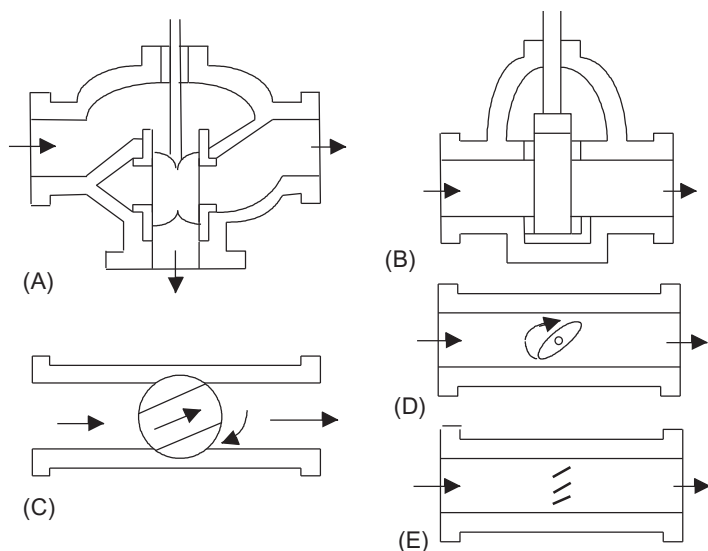


FIGURE 6.26 (A) Three-way globe, (B) gate, (C) ball, (D) butterfly, and (E) louvre.

6.4.4 Valve Positioners

Frictional forces and unbalanced forces on the plug may prevent the diaphragm from positioning the plug accurately. In order to overcome this, *valve positioners* may be fitted to the control valve stem. They position the valve stem more accurately and also provide extra power to operate the valve and so increase the speed of valve movement. Figure 6.25 shows the basic elements of a positioner.

The output from the controller is applied to a spring-loaded bellows. A flapper is attached to the bellows and is moved by pressure applied to the bellows. An increase in this pressure brings the flapper closer to the nozzle and so cuts down the air escaping from it. As a consequence, the pressure applied to the diaphragm is increased. The resulting valve stem displacement takes the flapper away from the nozzle until the air leakage from the nozzle is just sufficient to maintain the correct pressure on the diaphragm.

6.4.5 Other Forms of Flow Control Valves

The type of control valve described in the earlier parts of this section is basically the *split-body globe* valve body with a plug or plugs. This is the most commonly used form. There are, however, other forms. Figure 6.26A shows a three-way globe. Other valve types are the *gate* (Figure 6.26B), the *ball* (Figure 6.26C), the *butterfly* (Figure 6.26D) and the *louvre* (Figure 6.26E). All exercise control by restricting the fluid flow. Ball valves use a ball

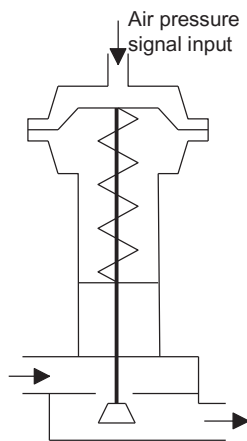


FIGURE 6.27 Shut down if air pressure fails.

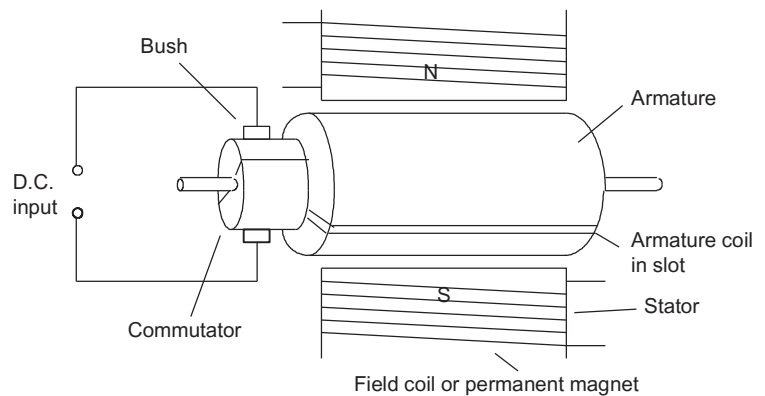


FIGURE 6.28 Basic elements of a d.c. motor.

with a through-hole which is rotated; they have excellent shut-off capability. Butterfly valves rotate a vane to restrict the air flow and, as a consequence, suffer from the problem of requiring significant force to move from the full-open position and so can 'stick' in that position.

6.4.6 Fail-Safe Design

Fail-safe design means that the design of a plant has to take account of what will happen if the power or air supply fails so that a safe shut down occurs. Thus, in the case of a fuel valve, the valve should close if failure occurs, while for a cooling water valve the failure should leave the valve open. Figure 6.27 shows a direct acting valve which shuts down the fluid flow if the air supply to the diaphragm fails.

6.5 MOTORS

Electric motors are frequently used as the final control element in position or speed control systems. The basic principle on which motors are based is that a force is exerted on a conductor in a magnetic field when a current passes through it. For a conductor of length L carrying a current I in a magnetic field of flux density B at right angles to the conductor, the force F equals BIL .

There are many different types of motor. In the following, discussion is restricted to those types of motor that are commonly used in control systems, this including d.c. motors and the stepper motor. A *stepper motor* is a form of motor that is used to give a fixed and consistent angular movement by rotating an object through a specified number of revolutions or fraction of a revolution.

6.5.1 D.C. Motors

In the d.c. motor, coils of wire are mounted in slots on a cylinder of magnetic material called the *armature*. The armature is mounted on bearings and is free to rotate. It is mounted in the magnetic field produced by *field poles*. This magnetic field might be produced by permanent magnets or an electromagnet with its magnetism produced by a current passing through the, so-termed, *field coils*. Whether permanent magnet or electromagnet, these generally form the outer casing of the motor and are termed the *stator*. Figure 6.28 shows the basic elements of d.c. motor with the magnetic field of the stator being produced by a current through coils of wire. In practice there will be more than one armature coil and more than one set of stator poles. The ends of the armature coil are connected to adjacent segments of a segmented ring called the *commutator* which rotates with the armature. Brushes in fixed positions make contact with the rotating commutator contacts. They carry direct current to the armature coil. As the armature rotates, the commutator reverses the current in each coil as it moves between the field poles. This is necessary if the forces acting on the coil are to remain acting in the same direction and so continue the rotation.

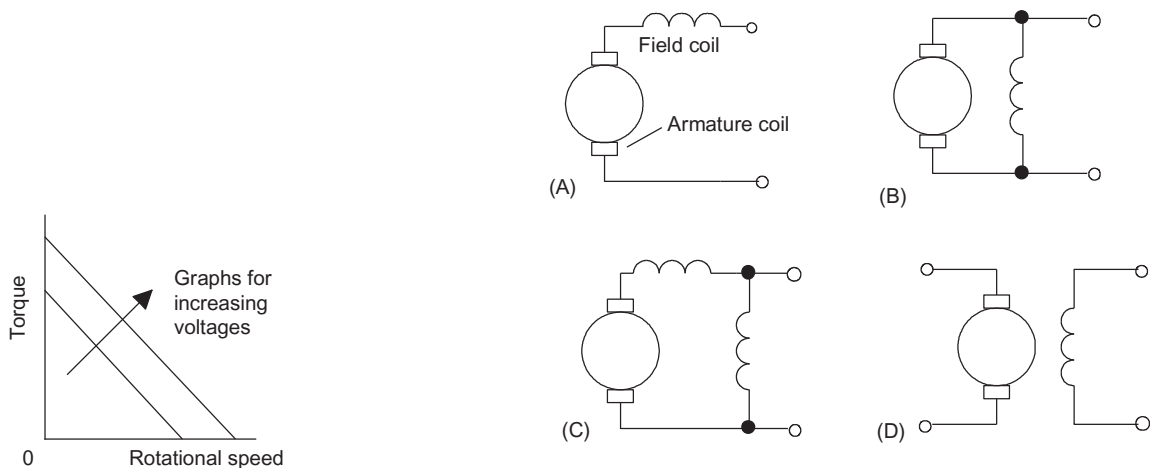


FIGURE 6.29 Permanent magnet motor characteristic.

FIGURE 6.30 (A) Series, (B) shunt, (C) compound, and (D) separately wound.

For a d.c. motor with the field provided by a permanent magnet, the speed of rotation can be changed by changing the size of the current to the armature coil, the direction of rotation of the motor being changed by reversing the current in the armature coil. Figure 6.29 shows how, for a permanent magnet motor, the torque developed varies with the rotational speed for different applied voltages. The starting torque is proportional to the applied voltage and the developed torque decreases with increasing speed.

D.C. motors with field coils are classified as series, shunt, compound and separately excited according to how the field windings and armature windings are connected.

1. Series-wound motor

With the *series-wound motor* the armature and field coils are in series (Figure 6.30A). Such a motor exerts the highest starting torque and has the greatest no-load speed. However, with light loads there is a danger that a series-wound motor might run at too high a speed. Reversing the polarity of the supply to the coils has no effect on the direction of rotation of the motor, since both the current in the armature and the field coils are reversed.

2. Shunt-wound motor

With the *shunt-wound motor* (Figure 6.30B) the armature and field coils are in parallel. It provides the lowest starting torque, a much lower no-load speed and has good speed regulation. It gives almost constant speed regardless of load and thus shunt-wound motors are very widely used. To reverse the direction of rotation, either the armature or field current can be reversed.

3. Compound motor

The *compound motor* (Figure 6.30C) has two field windings, one in series with the armature and one in parallel. Compound-wound motors aim to get the best features of the series and shunt-wound motors, namely a high starting torque and good speed regulation.

4. Separately excited motor

The *separately excited motor* (Figure 6.30D) has separate control of the armature and field currents. The direction of rotation of the motor can be obtained by reversing either the armature or the field current.

Figure 6.31 indicates the general form of the torque–speed characteristics of the above motors. The separately excited motor has a torque–speed characteristic similar to the shunt-wound motor. The speed of such d.c. motors can be changed by either changing the armature current or the field current. Generally it is the armature current that is varied. The choice of d.c. motor will depend on what it is to be used for. Thus, for example, with a robot manipulator the robot wrist might use a series-wound motor because the speed decreases as the load increases. A shunt-wound motor might be used if a constant speed was required, regardless of the load.

The speed of a permanent magnet motor can be controlled by varying the current through the armature coil, with a field coil motor by either varying the armature current or the field current though generally it is the armature current that is varied. Thus speed control can be obtained by controlling the voltage applied to the armature.

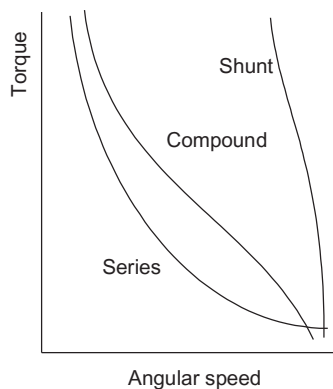


FIGURE 6.31 Torque–speed characteristics of d.c. motors.

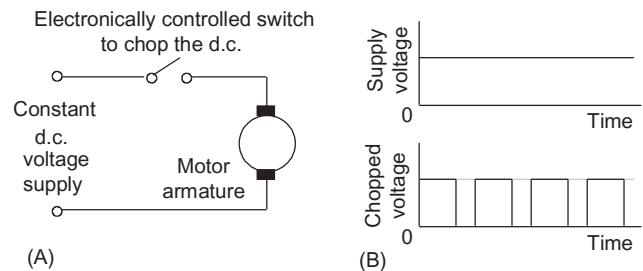


FIGURE 6.32 PWM: (A) principle of PWM circuit and (B) varying the average armature voltage by chopping the constant d.c. voltage.

Rather than just try to directly vary the input voltage, a more convenient method is to use *pulse width modulation* (PWM). This basically involves taking a constant d.c. supply voltage and using an electronic circuit to chop it so that the average value is varied (Figure 6.32).

6.5.2 Brushless Permanent Magnet D.C. Motor

A problem with the d.c. motors described in the previous section is that they require a commutator and brushes in order to periodically reverse the current through each armature coil. Brushes have to be periodically changed and the commutator resurfaced because the brushes make sliding contacts with the commutator and suffer wear. Brushless d.c. motors do not have this problem.

A current-carrying conductor in a magnetic field experiences a force and with the conventional d.c. motor the magnet is fixed and the current-carrying conductors consequently made to move. However, as a consequence of Newton's third law of motion, the magnet will experience an opposite and equal force to that acting on the current-carrying conductors and so, with the brushless permanent magnet d.c. motor, the current-carrying conductors are fixed and the magnet moves. With just one current-carrying coil, the resulting force on the magnet would just cause it to deflect. In order to keep the magnet moving, a sequence of current-carrying coils have to be used and each in turn switched on.

Figure 6.33 shows the basic form of such a motor. The rotor is a ferrite or ceramic permanent magnet. The current to the stator coils AA, BB and CC is electronically switched by transistors in sequence round them, the switching being controlled by the position of the rotor so that there are always forces acting on the magnet causing it to rotate in the same direction. Hall sensors (a magnetic field input to the sensor gives a voltage output) are generally used to sense the position of the rotor and initiate the switching by the transistors, the sensors being positioned around the stator. Figure 6.34 shows the transistor switching circuits that might be used with the motor shown in Figure 6.33.

To switch the coils in sequence, we need to supply signals to switch the transistors on in the right sequence. This is provided by the outputs from the three sensors operating through a decoder circuit to give the appropriate base currents. Thus when the rotor is in the vertical position, i.e. 0° , there is an output from sensor c but none from a and b and this is used to switch on transistors A+ and B-. When the rotor is in the 60° position there are signals from the sensors b and c and transistors A+ and C- are switched on. Table 6.2 shows the entire switching sequence. The entire circuit for controlling such a motor is available as a single integrated circuit.

6.5.3 Stepper Motor

The *stepper or stepping motor* produces rotation through equal angles, the so-called *steps*, for each digital pulse supplied to its input. For example, if with such a motor 1 input pulse produces a rotation of 1.8° then 20 input pulses will produce a rotation through 36.0° , 200 input pulses a rotation through one complete revolution of 360° . It can thus be used for accurate angular positioning. By using the motor to drive a continuous belt, the angular rotation of the motor is transformed into linear motion of the belt and so accurate linear positioning can be achieved. Such a motor is used with computer printers, x - y plotters, robots, machine tools and a wide variety of instruments for accurate positioning.

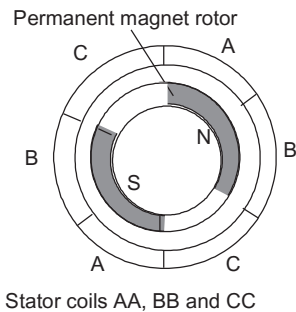


FIGURE 6.33 Brushless permanent magnet d.c. motor.

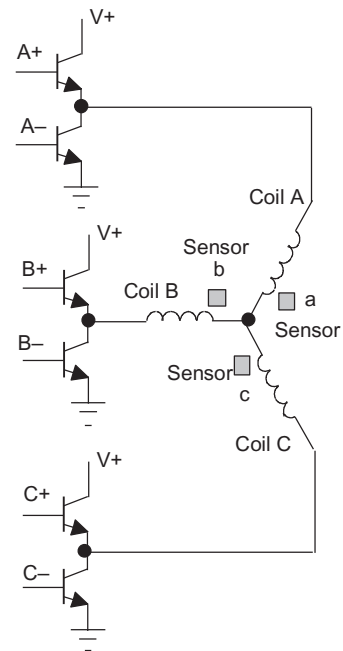


FIGURE 6.34 Transistor switching.

TABLE 6.2 Switching Sequence

Rotor position	Sensor signals			Transistors on	
	a	b	c		
0°	0	0	1	A+	B-
60°	0	1	1	A+	C-
120°	0	1	0	B+	C-
180°	1	1	0	B+	A-
240°	1	0	0	C+	A-
360°	1	0	1	C+	B-

There are two basic forms of stepper motor, the *permanent magnet* type with a permanent magnet rotor and the *variable reluctance* type with a soft steel rotor. Figure 6.35 shows the basic elements of the permanent magnet type with two pairs of stator poles.

Each pole of the permanent magnet motor is activated by a current being passed through the appropriate field winding, the coils being such that opposite poles are produced on opposite coils. The current is supplied from a d.c. source to the windings through switches. With the currents switched through the coils such that the poles are as shown in Figure 6.35, the rotor will move to line up with the next pair of poles and stop there. This would be, for Figure 6.35, an angle of 45°. If the current is then switched so that the polarities are reversed, the rotor will move a step to line up with the next pair of poles, at angle 135° and stop there. The polarities associated with each step are:

Step	Pole 1	Pole 2	Pole 3	Pole 4
1	North	South	South	North
2	South	North	South	North
3	South	North	North	South
4	North	South	North	South
5	Repeat of steps 1–4			

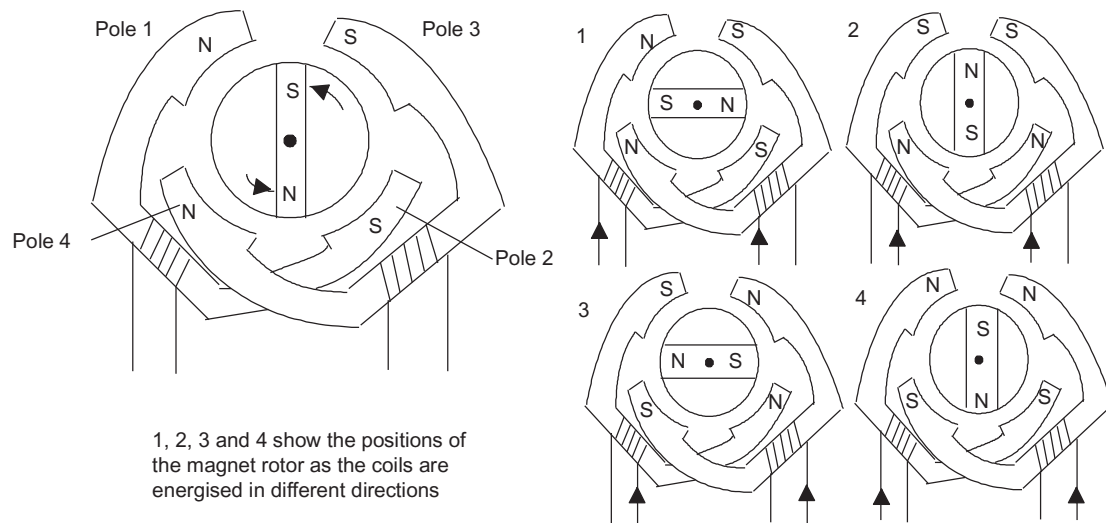


FIGURE 6.35 The basic principles of the permanent magnet stepper motor (2-phase) with 90° steps.

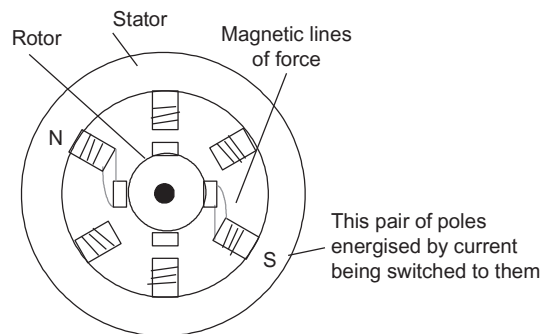


FIGURE 6.36 Basic principles of a three-phase variable reluctance stepper motor.

There are thus, in this case, four possible rotor positions: 45°, 135°, 225° and 315°.

Figure 6.36 shows the basic form of the *variable reluctance* type of stepper motor. With this form the rotor is made of soft steel and is not a permanent magnet. The rotor has a number of teeth, the number being less than the number of poles on the stator. When an opposite pair of windings on stator poles has current switched to them, a magnetic field is produced with lines of force which pass from the stator poles through the nearest set of teeth on the rotor. Since lines of force can be considered to be rather like elastic thread and always trying to shorten themselves, the rotor will move until the rotor teeth and stator poles line up. This is termed the position of minimum reluctance. Thus by switching the current to successive pairs of stator poles, the rotor can be made to rotate in steps. With the number of poles and rotor teeth shown in Figure 6.36, the angle between each successive step will be 30°. The angle can be made smaller by increasing the number of teeth on the rotor.

There is another version of the stepper motor and that is a *hybrid stepper*. This combines features of both the permanent magnet and variable reluctance motors. They have a permanent magnet rotor encased in iron caps which are cut to have teeth. The rotor sets itself in the minimum reluctance position in response to a pair of stator coils being energised.

The following are some of the terms commonly used in specifying stepper motors:

1. Phase

This is the number of independent windings on the stator, e.g. a four-phase motor. The current required per phase and its resistance and inductance will be specified so that the controller switching output is specified.

Figure 6.35 is an example of a two-phase motor, such motors tending to be used in light-duty applications.

Figure 6.36 is an example of a three-phase motor. Four-phase motors tend to be used for higher power applications.

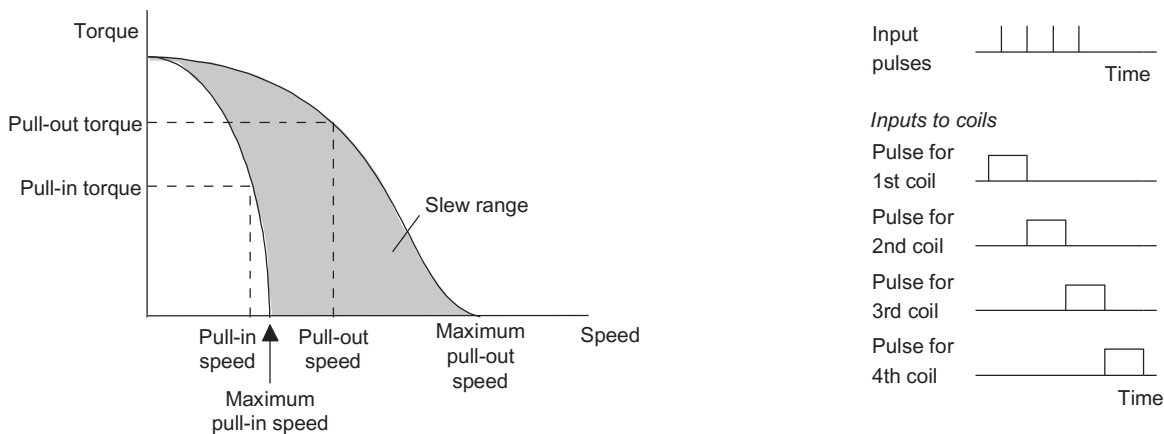


FIGURE 6.37 Stepper motor characteristics.

FIGURE 6.38 Input and output for a drive system for a stepper motor.

2. Step angle

This is the angle through which the rotor rotates for one switching change for the stator coils.

3. Holding torque

This is the maximum torque that can be applied to a powered motor without moving it from its rest position and causing spindle rotation.

4. Pull-in torque

This is the maximum torque against which a motor will start, for a given pulse rate, and reach synchronism without losing a step.

5. Pull-out torque

This is the maximum torque that can be applied to a motor, running at a given stepping rate, without losing synchronism.

6. Pull-in rate

This is the maximum switching rate or speed at which a loaded motor can start without losing a step.

7. Pull-out rate

This is the switching rate or speed at which a loaded motor will remain in synchronism as the switching rate is reduced.

8. Slew range

This is the range of switching rates between pull-in and pull-out within which the motor runs in synchronism but cannot start up or reverse.

Figure 6.37 shows the general characteristics of a stepper motor.

The variable reluctance stepper motor does not contain a magnet and this helps to make it cheaper and lighter and so able to accelerate more quickly. However, this lack of a magnet means that when it is not powered there is nothing to hold the rotor in a fixed position. The permanent magnet motor tends to have larger step angles, 7.5° or 15° , than the variable reluctance motor. The hybrid motor typically has 200 rotor teeth and rotates at 1.8° step angles. They have high static and dynamic torque and can run at very high step rates. As a consequence they are very widely used.

To drive a stepper motor, so that it proceeds step by step to provide rotation, requires each pair of stator coils to be switched on and off in the required sequence when the input is a sequence of pulses (Figure 6.38). Driver circuits are available to give the correct sequencing and Figure 6.39 shows an example, the SAA1027 for a four-phase unipolar stepper. Motors are termed *unipolar* if they are wired so that the current can only flow in one direction through any particular motor terminal, *bipolar* if the current can flow in either direction through any particular motor terminal. The stepper motor will rotate through one step each time the trigger input goes from low to high. The motor runs clockwise when the rotation input is low and anticlockwise when high. When the set pin is made low the output resets. In a control system, these input pulses might be supplied by a microprocessor.

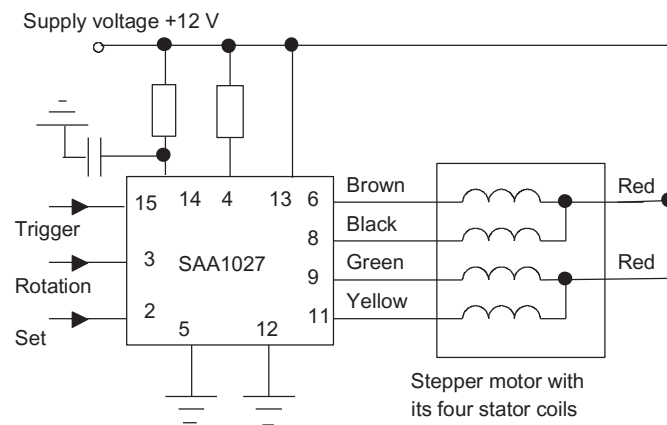


FIGURE 6.39 Driver circuit SAA1027 for a 12 V 4-phase stepper motor.

Some applications require very small step angles. Though the step angle can be made small by increasing the number of rotor teeth and/or the number of phases, generally more than four phases and 50–100 teeth are not used. Instead a technique known as *mini-stepping* is used with each step being divided into a number of equal size substeps by using different currents to the coils so that the rotor moves to intermediate positions between normal step positions. For example, this method might be used so that a step of 1.8° is subdivided into 10 equal steps.

Section 4.4.2 shows the application of a stepper motor to the control of the position of a tool. A manufacturer's data for a stepper motor includes: 12 V 4-phase, unipolar, Step angle 7.5° , Suitable driver SAA1027.

EXAMPLE

A stepper motor is to be used to drive, through a belt and pulley system (Figure 6.40), the carriage of a printer. The belt has to move a mass of 500 g which has to be brought up to a velocity of 0.2 m/s in a time of 0.1 s. Friction in the system means that movement of the carriage requires a constant force of 2 N. The pulleys have an effective diameter of 40 mm. Determine the required pull-in torque.

The force F required to accelerate the mass is

$$F = ma = 0.500 \times (0.2/0.1) = 1.0 \text{ N.}$$

The total force that has to be overcome is the sum of the above force and that due to friction. Thus the total force that has to be overcome is $1.0 + 2 = 3 \text{ N}$.

This force acts at a radius of 0.020 m and so the torque that has to be overcome to start, i.e. the pull-in torque, is

$$\text{torque} = \text{force} \times \text{radius} = 3 \times 0.020 = 0.06 \text{ Nm}$$

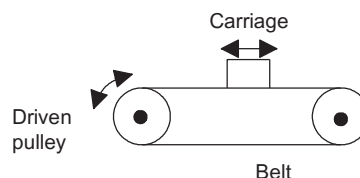


FIGURE 6.40 Example.

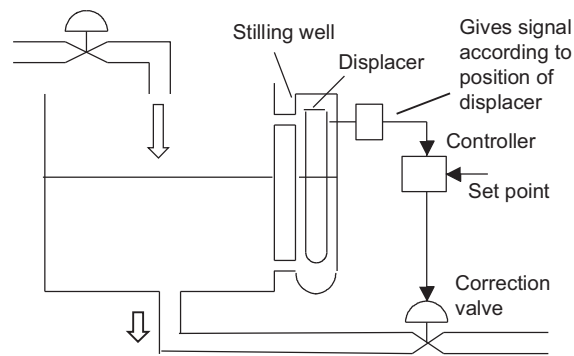


FIGURE 6.41 Liquid level control.

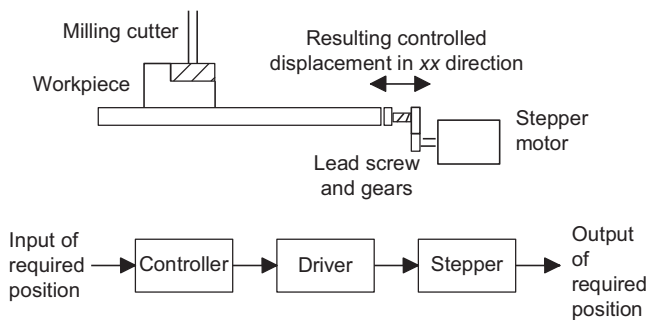


FIGURE 6.42 Automatic milling machine.

6.6 CASE STUDIES

The following are case studies designed to illustrate the use of correction elements discussed in this chapter.

6.6.1 A Liquid Level Process Control System

Figure 6.41 shows one method of how a flow control valve can be used to control the level of a liquid in a container. Because there may be surface turbulence as a result of liquid entering the container or stirring of the liquid or perhaps boiling, such high frequency 'noise' in the system is often filtered out by the use of a *stilling well*, as shown in Figure 6.41. However, it must be recognised that the stilling well constitutes a U-tube in which low frequency oscillations of the liquid level can occur. Vertical movement of the displacer results in a signal being passed to the controller. This may be achieved by the movement causing the slider of a potentiometer to move across its track.

6.6.2 Milling Machine Control System

Figure 6.42 shows how a stepper motor can be used to control the movement of the workpiece in an automatic milling machine. The stepping motor rotates by controlled steps and gives, via a lead screw and gears, controlled displacements of the worktable in the *xx* direction. A similar arrangement is used for displacement in the *yy* direction. The system is open-loop control with no feedback of the worktable position. The system relies on the accuracy with which the stepper motor can set the position of the worktable.

6.6.3 A Robot Control System

Figure 6.43 shows how directional control valves can be used for a control system of a robot. When there is an input to solenoid A of valve 1, the piston moves to the right and causes the gripper to close. If solenoid B is energised, with A de-energised, the piston moves to the left and the gripper opens. When both solenoids are de-energised, no air passes to either side of the piston in the cylinder and the piston keeps its position without change. Likewise, inputs to the solenoids of valve 2 are used to extend or retract the arm. Inputs to the solenoids of valve 3 are used to move the arm up or down. Inputs to the solenoids of valve 4 are used to rotate the base in either a clockwise or anticlockwise direction.

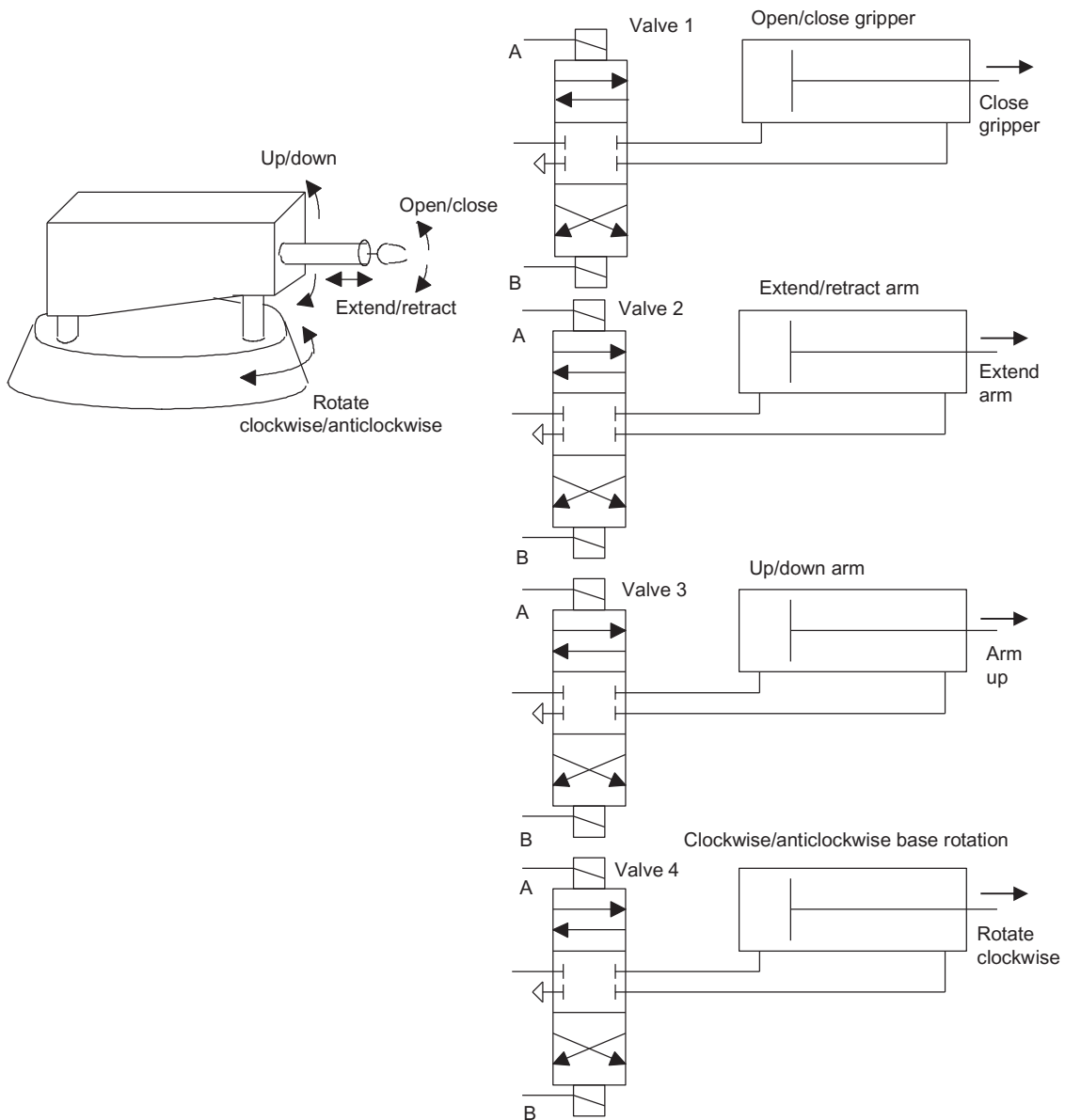


FIGURE 6.43 Robot controls.

PROBLEMS

Questions 1–19 have four answer options: A, B, C and D. Choose the correct answer from the answer options.

1. Decide whether each of these statements is True (T) or False (F).

For a hydraulic cylinder:

- i. The force that can be exerted by the piston is determined solely by the product of the pressure exerted on it and its cross-sectional area.
- ii. The speed with which the piston moves is determined solely by the product of the rate at which fluid enters the cylinder and the cross-sectional area of the piston.

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

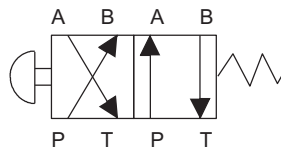


FIGURE 6.44 Problems 4–6.

2. A pneumatic cylinder has a piston of cross-sectional area 0.02 m^2 . The force exerted by the piston when the working pressure applied to the cylinder is 2 MPa will be:
 - A. 100 MN
 - B. 40 MN
 - C. 40 kN
 - D. 20 kN
3. A hydraulic cylinder with a piston having a cross-sectional area of 0.01 m^2 is required to give a workpiece an average velocity of 20 mm/s . The rate at which hydraulic fluid should enter the cylinder is:
 - A. $4 \times 10^{-6} \text{ m}^3/\text{s}$
 - B. $2 \times 10^{-4} \text{ m}^3/\text{s}$
 - C. $0.2 \text{ m}^3/\text{s}$
 - D. $2 \text{ m}^3/\text{s}$

Questions 4–6 refer to Figure 6.44 which shows a valve symbol.

4. Decide whether each of these statements is True (T) or False (F).
The valve has:
 - i. 2 ports
 - ii. 4 positions
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
5. Decide whether each of these statements is True (T) or False (F).
When the push button is pressed:
 - i. Hydraulic fluid from the supply is transmitted through port B.
 - ii. The hydraulic fluid in the line to port A is returned to the sump.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
6. Decide whether each of these statements is True (T) or False (F).
When the press button is released:
 - i. Hydraulic fluid from the supply is transmitted through port A.
 - ii. The hydraulic fluid in the line to port B is returned to the sump.
 - A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F

Questions 7–10 refer to Figure 6.45 which shows a pneumatic circuit involving two valves and a single acting cylinder.

7. Decide whether each of these statements is True (T) or False (F).
When push button 1 is pressed:
 - i. The load is lifted.
 - ii. Port A is closed.

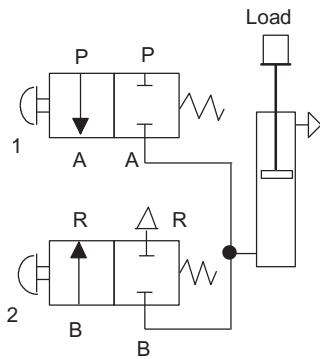


FIGURE 6.45 Problems 7–10.

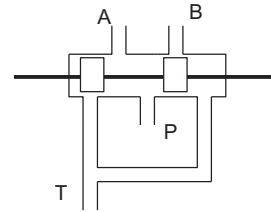


FIGURE 6.46 Problem 11.

- A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
8. Decide whether each of these statements is True (T) or False (F).
When push button 1, after being pressed, is released:
- i. The load descends.
 - ii. Port A is closed.
- A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
9. Decide whether each of these statements is True (T) or False (F).
When push button 2 is pressed:
- i. The load is lifted.
 - ii. Port B is vented to the atmosphere.
- A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
10. Decide whether each of these statements is True (T) or False (F).
When push button:
- i. 1 is pressed the load is lifted.
 - ii. 2 is pressed the load descends.
- A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
11. Decide whether each of these statements is True (T) or False (F).
Figure 6.46 shows a two-way spool valve. For this valve, movement of the shuttle from left to right:
- i. Closes port A.
 - ii. Connects port P to port B.
- A. (i) T (ii) T
 - B. (i) T (ii) F
 - C. (i) F (ii) T
 - D. (i) F (ii) F
12. A flow control valve has a diaphragm actuator. The air pressure signals from the controller to give 0–100% correction vary from 0.02 to 0.1 MPa above the atmospheric pressure. The diaphragm area

needed to 100% open the control valve if a force of 400 N has to be applied to the stem to fully open the valve is:

- A. 0.02 m^3
- B. 0.016 m^3
- C. 0.004 m^3
- D. 0.005 m^3

13. Decide whether each of these statements is True (T) or False (F).

A quick-opening flow control valve has a plug shaped so that:

- i. A small change in the flow rate occurs for a large movement of the valve stem.
- ii. The change in the flow rate is proportional to the change in the displacement of the valve stem.

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

14. A flow control valve with a linear plug gives a minimum flow rate of 0 and a maximum flow rate of $10 \text{ m}^3/\text{s}$. It has a stem displacement at full travel of 20 mm and so the flow rate when the stem displacement is 5 mm is:

- A. $0 \text{ m}^3/\text{s}$
- B. $2.5 \text{ m}^3/\text{s}$
- C. $5.0 \text{ m}^3/\text{s}$
- D. $7.5 \text{ m}^3/\text{s}$

15. A flow control valve with an equal percentage plug gives a flow rate of $0.1 \text{ m}^3/\text{s}$ when the stem displacement is 0 and $1.0 \text{ m}^3/\text{s}$ when it is at full travel. The stem displacement at full travel is 30 mm. The flow rate with a stem displacement of 15 mm is:

- A. $0.32 \text{ m}^3/\text{s}$
- B. $0.45 \text{ m}^3/\text{s}$
- C. $1.41 \text{ m}^3/\text{s}$
- D. $3.16 \text{ m}^3/\text{s}$

16. Decide whether each of these statements is True (T) or False (F).

A flow control valve has a minimum flow rate which is 1.0% of the maximum controllable flow. Such a valve is said to have a:

- i. Rangeability of 100.
- ii. Turndown of 100.

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

17. Decide whether each of these statements is True (T) or False (F).

A stepper motor is specified as having a step angle of 7.5° . This means that:

- i. The shaft takes 1 s to rotate through 7.5° .
- ii. Each pulse input to the motor rotates the motor shaft by 7.5° .

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

18. Decide whether each of these statements is True (T) or False (F).

For a series-wound d.c. motor:

- i. The direction of rotation can be reversed by reversing the direction of the supplied current.
- ii. The speed of rotation of the motor can be controlled by changing the supplied current.

- A. (i) T (ii) T
- B. (i) T (ii) F
- C. (i) F (ii) T
- D. (i) F (ii) F

19. Decide whether each of these statements is True (T) or False (F).

With a shunt-wound d.c. motor:

- i. The direction of rotation can be changed by reversing the direction of the armature current.
- ii. The direction of rotation can be changed by reversing the direction of the current to the field coils.

A. (i) T (ii) T

B. (i) T (ii) F

C. (i) F (ii) T

D. (i) F (ii) F

20. A force of 400 N is required to fully open a pneumatic flow control valve having a diaphragm actuator. What diaphragm area is required if the gauge pressure from the controller is 100 kPa?
21. An equal percentage flow control valve has a rangeability of 25. If the maximum flow rate is $50 \text{ m}^3/\text{s}$, what will be the flow rate when the valve is one-third open?
22. A stepper motor has a step angle of 7.5° . What digital input rate is required to produce a rotation of 10.5 rev/s?
23. A control valve is to be selected to control the rate of flow of water into a tank requiring a maximum flow of $0.012 \text{ m}^3/\text{s}$. The permissible pressure drop across the valve at maximum flow is 200 kPa. What valve size is required? Use Table 6.1. The density of water is $1000 \text{ kg}/\text{m}^3$.
24. A control valve is to be selected to control the flow of steam to a process, the maximum flow rate required being $0.125 \text{ kg}/\text{s}$. The permissible pressure drop across the valve at maximum flow is 40 kPa. What valve size is required? Use Table 6.1. The specific volume of the steam is $0.6 \text{ m}^3/\text{s}$.