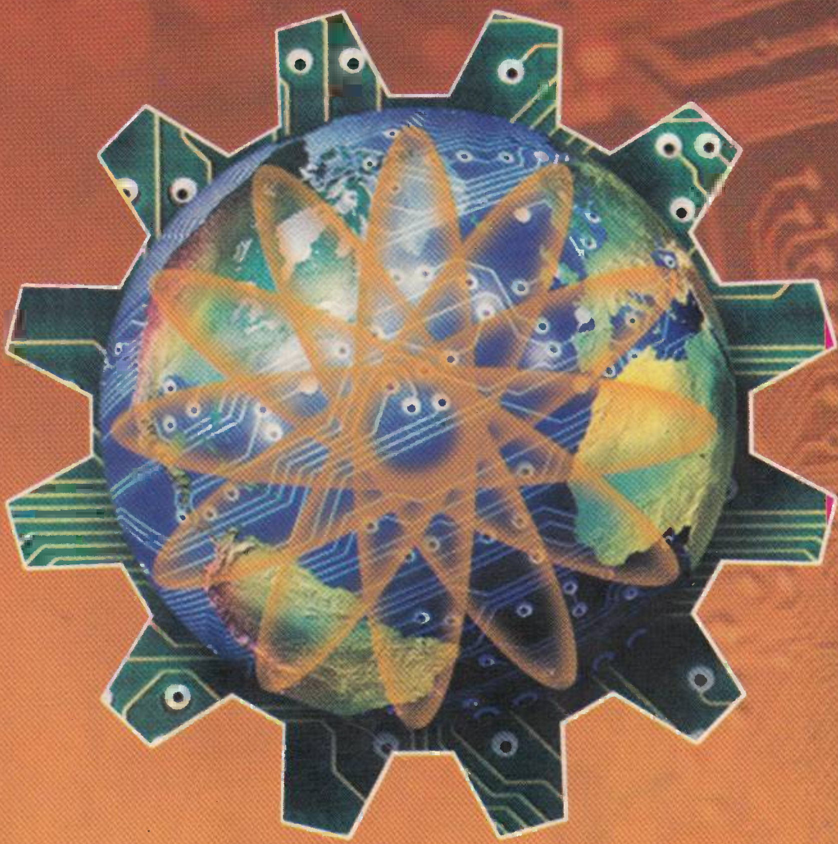
A TEXTBOOK OF

MECHATRONICS



R.K. RAJPUT

S. CHAND

A TEXTBOOK OF
MECHATRONICS

A TEXTBOOK OF

MECHATRONICS

For Engineering Students of B.Tech/B.E. Courses

R.K. RAJPUT

M.E. (Hons.) Gold Medalist; Grad. (Mech. Engg. & Elect. Engg.); M.I.E. (India);
M.S.E.S.I; M.I.S.T.E; C.E. (India)

Recipient of :
"Best Teacher (Academic) Award"
"Distinguished Author Award"
"Jawahar Lal Nehru Memorial Gold Medal"
for an outstanding research paper
(Institution of Engineers–India)
PRINCIPAL (Formerly)
Punjab College of Information Technology
PATIALA (Punjab)

S. CHAND & COMPANY LTD.

(AN ISO 9001 : 2000 COMPANY)
RAM NAGAR, NEW DELHI-110055

S. CHAND & COMPANY LTD.
(An ISO 9001 : 2000 Company)
Head Office : 7361, RAM NAGAR, NEW DELHI - 110 055
Phones : 23672080-81-82, 9899107446, 9911310888;
Fax : 91-11-23677446
Shop at: **schandgroup.com**; E-mail: **schand@vsnl.com**

Branches :

- 1st Floor, Heritage, Near Gujarat Vidhyapeeth, Ashram Road, **Ahmedabad**-380 014. Ph. 27541965, 27542369, ahmedabad@schandgroup.com
- No. 6, Ahuja Chambers, 1st Cross, Kumara Krupa Road, **Bangalore**-560 001. Ph : 22268048, 22354008, bangalore@schandgroup.com
- 238-A M.P. Nagar, Zone 1, **Bhopal** - 462 011. Ph : 4274723. bhopal@schandgroup.com
- 152, Anna Salai, **Chennai**-600 002. Ph : 28460026, chennai@schandgroup.com
- S.C.O. 2419-20, First Floor, Sector- 22-C (Near Aroma Hotel), **Chandigarh**-160022, Ph-2725443, 2725446, chandigarh@schandgroup.com
- 1st Floor, Bhartia Tower, Badambadi, **Cuttack**-753 009, Ph-2332580; 2332581, cuttack@schandgroup.com
- 1st Floor, 52-A, Rajpur Road, **Dehradun**-248 001. Ph : 2740889, 2740861, dehradun@schandgroup.com
- Pan Bazar, **Guwahati**-781 001. Ph : 2514155, guwahati@schandgroup.com
- Sultan Bazar, **Hyderabad**-500 195. Ph : 24651135, 24744815, hyderabad@schandgroup.com
- Mai Hiran Gate, **Jalandhar** - 144008 . Ph. 2401630, 5000630, jalandhar@schandgroup.com
- A-14 Janta Store Shopping Complex, University Marg, Bapu Nagar, **Jaipur** - 302 015, Phone : 2719126, jaipur@schandgroup.com
- 613-7, M.G. Road, Ernakulam, **Kochi**-682 035. Ph : 2381740, cochin@schandgroup.com
- 285/J, Bipin Bihari Ganguli Street, **Kolkata**-700 012. Ph : 22367459, 22373914, kolkata@schandgroup.com
- Mahabeer Market, 25 Gwynne Road, Aminabad, **Lucknow**-226 018. Ph : 2626801, 2284815, lucknow@schandgroup.com
- Blackie House, 103/5, Walchand Hirachand Marg , Opp. G.P.O., **Mumbai**-400 001. Ph : 22690881, 22610885, mumbai@schandgroup.com
- Karnal Bag, Model Mill Chowk, Umrer Road, **Nagpur**-440 032 Ph : 2723901, 2777666 nagpur@schandgroup.com
- 104, Citicentre Ashok, Govind Mitra Road, **Patna**-800 004. Ph : 2300489, 2302100, patna@schandgroup.com

First Edition 2007

ISBN : 81-219-2859-1
Code : 10 343

PRINTED IN INDIA

By Rajendra Ravindra Printers (Pvt.) Ltd., 7361, Ram Nagar, New Delhi-110 055 and published by S. Chand & Company Ltd, 7361, Ram Nagar, New Delhi-110 055

PREFACE TO THE FIRST EDITION

This treatise on the subject "**Mechatronics**" contains comprehensive treatment of the subject matter in simple, lucid and direct language. It covers the syllabi of the various Indian Universities in this subject exhaustively.

The book contains **nine** chapters in all, namely :

1. *Introduction to mechatronics, measurement systems and control systems* ; 2. *Basic and digital electronics; 3. Sensors and transducers ; 4. Signal conditioning, data acquistion, transmission and presentation/display ; 5. Microprocessors ; 6. System models and controllers ; 7. Actuators–Mechanical, electrical, hydraulic, pneumatic ; 8. Mechatronic systems ; 9. Elements of CNC machines.*

All these chapters are saturated with much needed text, supported by simple and self-explanatory figures, and worked examples, wherever required. At the end of each chapter "*Highlights*", "*Objective Type Questions*", "*Theoretical Questions*" and "*Unsolved Examples*" have been added to make the book a comprehensive and complete unit in all respects.

The author's thanks are due to his wife Ramesh Rajput for extending all cooperation during preparation and proof reading of the manuscript.

As ever before, I take this opportunity to thank my publisher Sh. Ravindra Kumar Gupta, Managing Director, and Sh. Navin Joshi, GM (Sales & Marketing) of S.Chand & Company Ltd for the personal interest they took in printing this book.

Any suggestions for improvement of this book will be thankfully acknowledged and incorporated in the next edition.

R.K. RAJPUT
(Author)

CONTENTS

6. SYSTEM MODELS AND CONTROLLERS 343–373

7. ACTUATORS – MECHANICAL, ELECTRICAL, HYDRAULIC, PNEUMATIC 374–486

APPENDIX–B : BASIC ELECTRICAL CONCEPTS 590–616

Introduction to SI Units and Conversion Factors

A. INTRODUCTION TO SI UNITS

SI, the international system of units are divided into three classes :

1. Base units
2. Derived units
3. Supplementary units.

From the scientific point of view division of SI units into these classes is to a certain extent arbitrary, because it is not essential to the physics of the subject. Nevertheless the General Conference, considering the advantages of a single, practical, world-wide system for international relations, for teaching and for scientific work, decided to base the international system on a choice of six well-defined units given in Table 1 below :

Table 1. SI Base Units

Quantity	*Name*	*Symbol*
length	metre	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	kelvin	K
luminous intensity	candela	cd
amount of substance	mole	mol

The second class of SI units contains derived units, *i.e.*, units which can be formed by combining base units according to the algebraic relations linking the corresponding quantities. Several of these algebraic expressions in terms of base units can be replaced by special names and symbols can themselves be used to form other derived units.

Derived units may, therefore, be classified under three headings. Some of them are given in Tables 2, 3 and 4.

Table 2. Examples of SI Derived Units Expressed in Terms of Base Units

Quantity	*SI Unit*	
	Name	*Symbol*
area	square metre	m^2
volume	cubic metre	m^3
speed, velocity	metre per second	m/s
acceleration	metre per second squared	m/s^2
density, mass density	kilogram per cubic metre	kg/m^3
concentration (of amount of substances)	mole per cubic metre	mol/m^3
activity (radioactive)	1 per second	s^{-1}
specific volume	cubic metre per kilogram	m^3/kg
luminance	candela per square metre	cd/m^2

Table 3. SI Derived Units with Special Names

Quantity	*SI Units*			
	Name	*Symbol*	*Expression in terms of other units*	*Expression in terms of SI base units*
frequency	hertz	Hz	–	s^{-1}
force	newton	N	–	$m.kg.s^{-2}$
pressure	pascal	Pa	N/m^2	$m^{-1}.kg.s^{-2}$
energy, work, quantity of heat power	joule	J	N.m	$m^2.kg.s^{-2}$
radiant flux, quantity of electricity	watt	W	J/s	$m^2.kg.s^{-3}$
electric charge	coloumb	C	A.s	s.A
electric tension, electric potential	volt	V	W/A	$m^2.kg.s^{-3}.A^{-1}$
capacitance	farad	F	C/V	$m^{-2}.kg^{-1}.s^4$
electric resistance	ohm	Ω	V/A	$m^2.kg.s^{-3}.A^{-2}$
conductance	siemens	S	A/V	$m^{-2}.kg^{-1}.s^3.A^{-2}$
magnetic flux	weber	Wb	V.S.	$m^2.kg.s^{-2}.A^{-1}$
magnetic flux density	tesla	T	Wb/m^2	$kg.s^{-2}.A^{-1}$
inductance	henry	H	Wb/A	$m^2.kg.s^{-2}.A^{-2}$
luminous flux	lumen	lm	–	cd.sr
illuminance	lux	lx	–	m^{-2}.cd.sr

Table 4. Examples of SI Derived Units Expressed by Means of Special Names

Quantity	*SI Units*		
	Name	*Symbol*	*Expression in terms of SI base units*
dynamic viscosity	pascal second	Pa-s	$m^{-1}.kg.s^{-1}$
moment of force	metre newton	N.m	$m^{-2}.kg.s^{-2}$
surface tension	newton per metre	N/m	$kg.s^{-2}$
heat flux density, irradiance	watt per square metre	W/m^2	$kg.s^{-2}$
heat capacity, entropy	joule per kelvin	J/K	$m^{-2}.kg.s^{-2}.K^{-1}$
specific heat capacity, specific entropy	joule per kilogram kelvin	J/(kg.K)	$m^{-2}.s^{-2}.K^{-1}$
specific energy	joule per kilogram	J/kg	$m^{-2}.s^{-2}$
thermal conductivity	watt per metre kelvin	W/(m.K)	$m.kg.s^{-3}.K^{-1}$
energy density	joule per cubic metre	J/m^3	$m^{-1}.kg.s^{-2}$
electric field strength	volt per metre	V/m	$m.kg.s^{-3}.A^{-1}$
electric charge density	coloumb per cubic metre	C/m^3	$m^{-3}.s.A$
electric flux density	coloumb per square metre	C/m^2	$m^{-2}.s.A$
permitivity	farad per metre	F/m	$m^{-3}.kg^{-1}.s^4.A^4$
current density	ampere per square metre	A/m^2	–
magnetic field strength	ampere per metre	A/m	–
permeability	henry per metre	H/m	$m.kg.s^{-2}.A^{-2}$
molar energy	joule per mole	J/mol	$m^{-2}.kg.s^{-2}mol^{-1}$
molar heat capacity	joule per mole kelvin	J/(mol.K)	$m^{-2}.kg.s^{-2}.K^{-1}$.mol–1

The SI units assigned to third class called "Supplementary units" may be regarded either as base units or as derived units. Refer to Table 5 and Table 6.

Table 5. SI Supplementary Units

Quantity	*SI Units*	
	Name	*Symbol*
plane angle	radian	rad
solid angle	steradian	sr

Table 6. Examples of SI Derived Units Formed by Using Supplementary Units

Quantity	*SI Units*	
	Name	*Symbol*
angular velocity	radian per second	rad/s
angular acceleration	radian per second squared	rad/s^2
radiant intensity	watt per steradian	W/sr
radiance	watt per square metre steradian	$W\text{-}m^{-2}.sr^{-1}$

second per second. Since acceleration due to gravity equals 9.81 m/s^2, one kilogram force equals 9.81 newtons.

Joule. The joule (J) is a derived unit of energy, work or quantity of heat and is defined as the work done when a force of one newton acts so as to cause a displacement of one metre. Energy is defined as the capacity to do work. A unit of energy in nuclear physics is the electron volt (eV) which is defined as the energy gained by an electron in rising through a potential difference of one volt.

$$1 \text{ eV} = 1.6021 \times 10^{-19} \text{ J}.$$

Watt. The watt (W) is a unit of power (*i.e.*, rate of doing work)

$$\text{Power in watts} = \frac{\text{work (or energy) in joules}}{\text{time in seconds}}$$

Thus 1 watt equals 1 Joule/sec.

1 kilo watt-hour (kWh) = 1000 watt-hours = 3600000 joules.

Coulomb. The coulomb (C) is the derived unit of charge. It is defined as *the quantity of electricity passing a given point in a circuit when a current of 1 A is maintained for 1 second.*

$$Q = I.t$$

where Q = charge in coulombs,
I = current in ampees, and
t = time in seconds.

1 coulomb represents 6.24×10^{18} electrons.

Ohm. The ohm (Ω) is the unit of electric resistance and is defined as *the resistance in which a constant current of 1 A generates heat at the rate of 1 watt.*

Siemen. The siemen is a unit of electric conductance (*i.e.*, reciprocal of resistance). If a circuit has a resistance of 5 ohms, its conductance is 0.2 siemen. A more commonly used name for siemen is *mho* (℧).

Volt. The volt is a unit of potential difference and electromotive force. It is defined as the *difference of potential across a resistance of 1 ohm carrying a current of 1 ampere.*

Hertz. The hertz (Hz) is a unit of frequency. 1 Hz = 1 cycle per second.

Horse-power. It is a practical unit of mechanical output. BHP (British horse power or brake horse power) equals 746 watts. The metric horse power equals 735.5 watts. To avoid confusion between BHP and metric horse power, the mechanical output of machines in SI units, is expressed in watts or kilowatts.

C. SALIENT FEATURES OF SI UNITS

The salient features of SI units are as follows :

1. It is a coherent system of units, i.e., product or quotient of any two base quantities results in a unit resultant quantity. For example, unit length divided by unit time gives unit velocity.
2. It is a rationalised system of units, applicable to both, magnetism and electricity.
3. It is a non-gravitational system of units. It clearly distinguishes between the units of mass and weight (force) which are kilogram and newton respectively.
4. All the units of the system can be derived from the base and supplementary units.
5. The decimal relationship between units of same quantity makes possible to express any small or large quantity as a power of 10.

6. For any quantity there is one and only one SI unit. For example, joule is the unit of energy of all forms such as mechanical, heat, chemical, electrical and nuclear. However, kWh will also continue to be used as unit of electrical energy.

Advantages of SI Units :

1. Units for many different quantities are related through a series of simple and basic relationship.
2. Being an absolute system, it avoids the use of factor '*g*' *i.e.*, acceleration due to gravity in several expressions in physics and engineering which had been a nuisance in all numericals in physics and engineering.
3. Being a rationalised system, it ensures all the advantages of rationalised MKSA system in the fields of electricity, magnetism, electrical engineering and electronics.
4. Joule is the only sole unit of energy of all forms and watt is the sole unit of power hence a lot of labour is saved in calculations.
5. It is a coherent system of units and involves only decimal co-efficients. Hence it is very convenient and quick system for calculations.
6. In electricity, all the practical units like volt, ohm, ampere, henry, farad, coulomb, joule and watt accepted in industry and laboratories all over the world for well over a century have become absolute in their own right in the SI system, without the need for any more practical units.

Disadvantages :

1. The non-SI time units 'minute' and 'hour' will still continue to be used until the clocks and watches are all changed to kilo seconds and mega seconds etc.
2. The base unit kilogram (kg) includes a prefix, which creates an ambiguity in the use of multipliers with gram.
3. SI units for energy, power and pressure (i.e., joule, watt and pascal) are too small to be expressed in science and technology, and, therefore, in such cases the use of larger units, such as MJ, kW, kPa, will have to be made.
4. There are difficulties with regard to developing new SI units for apparent and reactive energy while joule is the accepted unit for active energy in SI systems.

D. CONVERSION FACTORS

1. Force :

$$1 \text{ newton} = \text{kg-m/sec}^2 = 0.012 \text{ kgf}$$
$$1 \text{ kgf} = 9.81 \text{ N}$$

2. Pressure :

$$1 \text{ bar} = 750.06 \text{ mm Hg} = 0.9869 \text{ atm} = 10^5 \text{ N/m}^2$$
$$= 10^3 \text{ kg/m-sec}^2$$
$$1 \text{ N/m}^2 = 1 \text{ pascal} = 10^{-5} \text{ bar} = 10^{-2} \text{ kg/m-sec}^2$$
$$1 \text{ atm} = 760 \text{ mm Hg} = 1.03 \text{ kgf/cm}^2 = 1.01325 \text{ bar}$$
$$= 1.01325 \times 10^5 \text{ N/m}^2$$

3. Work, Energy or Heat :

$$1 \text{ joule} = 1 \text{ newton metre} = 1 \text{ watt-sec}$$
$$= 2.7778 \times 10^{-7} \text{ kW-h} = 0.239 \text{ cal}$$
$$= 0.239 \times 10^{-3} \text{ kcal}$$

$$1 \text{ cal} = 4.184 \text{ joule} = 1.1622 \times 10^{-6} \text{ kWh}$$
$$1 \text{ kcal} = 4.184 \times 10^{3} \text{ joule} = 427 \text{ kgf m}$$
$$= 1.1622 \times 10^{-3} \text{ kWh}$$
$$1 \text{ kWh} = 8.6042 \times 10^{5} \text{ cal} = 860.42 \text{ kcal}$$
$$= 3.6 \times 10^{6} \text{ joule}$$
$$1 \text{ kgf-m} = \left(\frac{1}{427}\right) \text{kcal} = 9.81 \text{ joules}$$

4. **Power :**

$$1 \text{ watt} = 1 \text{ joule/sec} = 0.860 \text{ kcal/h}$$
$$1 \text{ h.p.} = 75 \text{ m kgf/sec} = 0.1757 \text{ kcal/sec} = 735.3 \text{ watts}$$
$$1 \text{ kW} = 1000 \text{ watts} = 860 \text{ kcal/h}$$

5. **Specific heat :**

$$1 \text{ kcal/kg-°K} = 0.4184 \text{ joules/kg-K}$$

6. **Thermal conductivity :**

$$1 \text{ watt/m-K} = 0.8598 \text{ kcal/h-m-°C}$$
$$1 \text{ kcal/h-m-°C} = 1.16123 \text{ watt/m-K} = 1.16123 \text{ joules/s-m-K.}$$

7. **Heat transfer co-efficient :**

$$1 \text{ watt/m}^2\text{-K} = 0.86 \text{ kcal/m}^2\text{-h-°C}$$
$$1 \text{ kcal/m}^2\text{-h-°C} = 1.163 \text{ watt/m}^2\text{-K.}$$

The following conversion factors may be used to convert the quantities in non-SI units into SI units.

To convert	*To*	*Multiply by*
angstroms	m	10^{-10}
atmospheres	kg/m^2	10332
bars	kg/m^2	1.02×10^4
Btu	joules	1054.8
Btu	kWh	2.928×10^{-4}
circular mils	m^2	5.067×10^{-10}
cubic feet	m^3	0.02831
dynes	newtons	10^{-5}
ergs	joules	10^{-7}
ergs	kWh	0.2778×10^{-13}
feet	m	0.3048
foot-pounds	joules	1.356
foot-pounds	kg-m	0.1383
gauss	tesla	10^{-4}
grams (force)	newton	9.807×10^{-3}
horse power (metric)	watts	735.5
lines/sq. inch	tesla	1.55×10^{-5}
Maxwell	webers	10^{-8}
mho	siemens	1
micron	metre	10^{-6}
miles	km	1.609
mils	cm	2.54×10^{-3}

poundals	newton	0.1383
pounds	kilogram	0.454
pounds (force)	newtons	0.448
pounds/sq. ft.	N/m^2	47.878
pounds/sq. inch	N/m^2	6894.43

E. IMPORTANT ENGINEERING CONSTANTS AND EXPRESSIONS IN S.I. UNITS

Engineering Constants and Expressions	*M.K.S. System*	*SI Units*
1. Value of g_0	9.81 kg-m/kgf-sec^2	1 kg-m/N-sec^2
2. Universal gas constant	848 kgf-m/kg mol-°K	$848 \times 9.81 = 8314$ J/kg-mole-K ($\because$ 1 kgf-m=9.81 joules)
3. Gas constant (R)	29.27 kgf-m/kg-°K for air	$\frac{8341}{29} = 287$ joules/kg-K for air
4. Specific heat (for air)	$c_v = 0.17$ kcal/kg-°K $c_V = 0.24$ kcal/kg-°K	$c_v = 0.17 \times 4.184 = 0.71128$ kJ/kg-K $c_p = 0.24 \times 4.184 = 1$ kJ/kg-K
5. Flow through nozzle-Exit velocity (C_2)	91.5 U where U is in kcal	44.7 $\sqrt{U}$ where U is the kJ
6. Refrigeration 1 ton	= 50 kcal/min	= 210 kJ/min
7. **Heat transfer** The Stefan Boltzmann Law is given by :	$Q = \sigma T^4$ kcal/m^2-h where $\sigma = 4.9 \times 10^{-8}$ kcal/h-m^2-°K^4	$Q = \sigma T^4$ kcal/m^2-h where $\sigma = 5.67 \times 10^{-8}$ W/m^{2}K^4

F. DIMENSIONS OF QUANTITIES

Different units can be represented dimensionally in terms of units of length L, mass M, time T and current I. The dimensions can be derived as under :

1. Velocity = length/time = $L/T = LT^{-1}$
2. Acceleration = velocity/time = $LT^{-1}/T = LT^{-2}$
3. Force = mass × acceleration = MLT^{-2}
4. Charge (coulomb) = current × time = IT
5. Work or energy = force × distance = ML^2T^{-2}
6. EMF or potential = work/charge = $ML^2T^{-2}/IT = ML^2I^{-1}T^{-3}$
7. Power = work/time = $ML^2T^{-2}/T = ML^2T^{-3}$
8. Current density = current/area = $I/L^2 = IL^{-2}$
9. Resistance = emf/current = $ML^2I^{-1}T^{-3}/I = ML^2I^{-2}T^{-3}$
10. Electric flux density = electric flux or charge/area = $IT/L^2 = ITL^{-2}$
11. MMF = current × number of turns = I
12. Conductance = 1/resistance = $1/ML^2I^{-2}T^{-3} = I^2T^3M^{-1}L^{-2}$

13. Electric field intensity $=$ volt/metre $= ML^2I^{-1}T^{-3}/L = MLI^{-1}T^{-3}$

14. Resistivity $= \dfrac{\text{resistant} \times \text{area}}{\text{length}}$

$= (ML^2I^{-2}T^{-3})(L^2)/L$

$= ML^3I^{-2}T^{-3}$

15. Magnetic field intensity (H) $=$ MMF/length

$= I/L = IL^{-1}$

16. Magnetic flux $=$ emf $\times$ time $= (ML^2I^{-1}T^{-3})(T) = ML^2I^{-1}T^{-2}$

17. Magnetic flux intensity $=$ magnetic flux/area

$= (ML^2I^{-1}T^{-2})/L^2 = MI^{-1}T^{-2}$

18. Impedence $=$ emf/current $= ML^2I^{-2}T^{-3}$

19. Admittance $= I$/impedence $= I^2T^3M^{-1}L^{-2}$

20. Inductance $=$ magnetic flux/current

$= ML^2T^{-2}I^{-1}/I = ML^2T^{-2}I^{-2}$

21. Capacitance $=$ electric charge/potential

$= IT/ML^2T^{-3}I^{-1} = M^{-1}L^{-2}T^4I^2$

CHAPTER

1 Introduction to Mechatronics, Measurement Systems and Control Systems

1.1. Introduction to mechatronics and measurement systems – Definition and scope – Advantages and disadvantages of mechatronics – Components of a mechatronic system – Examples of mechatronic systems – Introduction to measurement systems – Functions of instruments and measurement systems – Applications of measurement systems – Measurement system performance.

1.2. Control systems – Introduction – System – Control system – Classification of control systems – Open loop control systems (Non-feedback systems) – Closed loop control systems (Feedback control systems) – Automatic control systems – Servo-mechanism – Regulator – Representation through model – Analogous systems – Block diagram – Mathematical block diagram – Signal flow graph – Time response of control system – Stability – Frequency response – Error detector – LVDT – Servo Amplifier – Sampled data systems – Industrial controllers – Pneumatic control systems – Hydraulic control system – Highlights – Objective Type Questions – Theoretical Questions.

1.1. INTRODUCTION TO MECHATRONICS AND MEASUREMENT SYSTEMS

1.1.1. Definition and Scope

*"*Mechatronics*" may be defined as follows :

"The synergistic combination of precision mechanical engineering, electronic control and systems thinking in the design of products and manufacturing processes."

Or

"Integration of electronics, control engineering and mechanical engineering".

Or

"The interdisciplinary field of engineering dealing with the design of products whose function relies on the synergistic integration of mechanical and electronic components co-ordinated by a control architecture."

- *"Mechatronics"* involves a number of technologies such as :
  - — Mechanical engineering ;
  - — Electronic engineering ;
  - — Electrical engineering ;
  - — Computer technology ;
  - — Control engineering.

This can be *considered to be the application of computer-based digital control techniques, through electronic and electric interfaces to mechanical engineering problems.*

* This word was coined in Japan in the late sixties, spread through Europe and is now being commonly used elsewhere in the world.

It represents the next generation machines, robots and smart mechanisms for carrying out work in a variety of environments – predominantly factory automation, office automation and home automation.

Evolution levels of mechatronics :

Following are the evolution levels of mechatronics:

1. ***Primary level mechatronics*** : This level incorporates *I/O devices* such as sensors, and actuators that integrates electrical signals with mechanical action at the basic control level.

   Examples : Electrically controlled fluid valves and relays.
2. ***Secondary level mechatronics*** : This level *integrates microelectronics into electrically controlled devices.*

   Example : Cassette player.
3. ***Third level mechatronics*** : This level incorporates *advanced feed back functions into control strategy thereby enhancing the quality in terms of sophistication* – called *"Smart system"*.

   — The control strategy includes microelectronics, microprocessor and other 'Application Specific Integrated Circuits' (ASIC).

   Examples : Control of electrical motor used to activate industrial robots, hard disk, CD drives, automatic washing machines.
4. ***Fourth level mechatronics*** : This level incorporates *intelligent control* in mechatronic system. It introduces intelligence and Fault Detection and Isolation (FDI) capability systems.

1.1.2. Advantages and Disadvantages of Mechatronics

Following are the *advantages* and *disadvantages* of mechatronics :

Advantages :

1. The products produced are cost effective and of very good quality.
2. The performance characteristics of mechatronics products are such which are otherwise very difficult to achieve without the synergistic combination.
3. High degree of flexibility.
4. A mechatronics product can be better than just sum of its parts.
5. Greater extent of machine utilisation.
6. Due to the integration of sensors and control systems in a complex system, capital expenses are reduced.
7. Owing to the incorporation of intelligent, self correcting sensory and feedback systems, the *mechatronic approach results in* :

   — greater productivity ;

   — higher quantity and producing reliability.

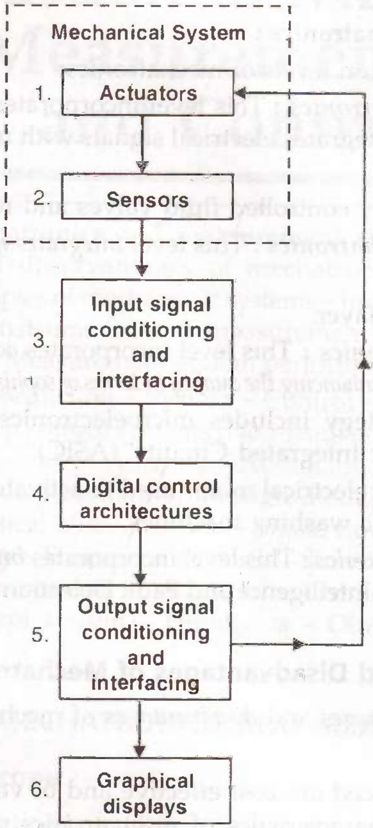
Disadvantages :

1. High initial cost of the system.
2. Imperative to have knowledge of different engineering fields for design and implementation.
3. Specific problems for various systems will have to be addressed separately and properly.
4. It is expensive to incorporate mechatronics approach to an existing/old system.

1.1.3. Components of a mechatronic system :

The term *mechatronic system* (sometimes referred to as 'smart device') encompasses a myriad of devices and systems. Increasingly, *microcontrollers are embedded in the* *electromechanical devices, creating much more flexibility and control possibilities in system design.*

Fig. 1.1 shows all components in a typical *"mechatronic system"*.



1. **Actuators :** Solenoids, voice coils ; D.C. motors ; Stepper motors ; Servomotor ; hydraulics; preumatics.
2. **Sensors :** Switches ; Potentiometer ; Photoelectrics ; Digital encoder ; Strain guage ; Thermocouple ; accelerometer etc.
3. **Input signal conditioning and interfacing :** Discrete circuits ; Amplifiers, Filters ; A/D, D/D.
4. **Digital control architectures :** Logic circuits ; Microcontroller ; SBC ; PLC ; Sequencing and timing ; Logic and arithmetic ; Control algorithms ; Communication.
5. **Output signal conditioning and interfacing :** D/A, D/D ; Amplifiers ; PWM ; Power transistors ; Power Op-amps.
6. **Graphical displays :** LEDs ; Digital displays ; LCD ; CRT.

Fig. 1.1. Components of a typical "mechatronic system".

- The **actuators** *produce motion or cause some action ;*
- The **sensors** *detect the state of the system parameters, inputs and outputs ;*
- **Digital devices** *control the system ;*
- **Conditioning and interfacing circuits** *provide connection between the control circuits and the input/output devices ;*
- **Graphical displays** *provide visual feedback to users.*

1.1.4. Examples of Mechatronic Systems :

Following are the examples of mechatronics systems :

1. *Home appliances :*
   — Washing mchines
   — Bread machines etc.
2. *Automobile :*
   — Electrical fuel injection
   — Antilock brake system.
3. *Aircraft :*
   — Flight control
   — Navigation system.
4. *Automated manufacturing :*
   — Robots
   — Numerically controlled (NC) machine tools.

- An automatic production line, an automatic camera and a truck suspension are examples of synergistic combination of electronic control systems and mechanical engineering. Such control systems generally use **microprocessors** *as controllers and have electrical sensors extracting information from the mechanical inputs and outputs via electrical actuators to mechanical systems.*

"Copy machine" – Example of mechatronic system.

Major components :

(*i*) *Analog circuit :*
— Controlling lamps
— Heaters
— Other power circuits.

(*ii*) *Digital circuit :*
— Control digit displays
— Indicator lights
— Buttons
— Switches.

(*iii*) *Microprocessor*—Co-ordinates all of the functions in the machine.

(*iv*) *Servo and stepper motors*—Loading and transporting the paper, turning the drum, and indexing the sorter.

Copying process :

An original in a loading bin

↓

Scanning

↓

Metal drum with charge distribution

↓

The paper from a loading cartridge with an electrostatic deposition of ink tone powder

↓

Heated the paper

↓

Delievered the copy to an appropriate bin by a sorting mechanism.

1.1.5. Introduction to Measurement Systems

Following are the *elements* of a measuring system :

1. Transducer
2. Signal processor
3. Recorder.

Fig. 1.2. Elements of a measurement system.

- **Transducer** *is a sensing device that converts a physical input into output, usually voltage.*
- **Signal processor** *performs filtering and amplification functions.*
- **Recorder** *records or displays the output of signal processor.*

Example : Measurement—Digital thermometer.

Refer to Fig. 1.3.

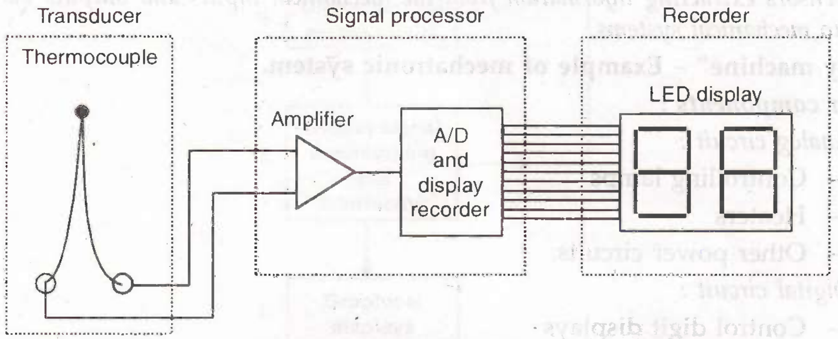


Fig. 1.3. Digial thermometer.

— **Thermocouple** *converts temperature to a small voltage.*
— **Amplifier** *increases the magnitude of the voltage.*
— **A/D** (analog to digital) *converts the analog voltage to a digital signal.*
— **LEDS** (Light emitting diodes) *display the value of temperature.*

1.1.6. Functions of Instruments and Measurement Systems

Following are the three main functions of instruments and measurement systems :

1. **Indicating function :**

   Examples : — A pressure gauge is used for indicating pressure.
   — The deflection of a pointer of a speedometer indicates the speed of the automobile at that moment.

2. **Recording function :**

   Examples : — A potentiometer type of recorder used for monitoring temperature records the instantaneous values of temperatures on a strip chart recorder.

3. **Controlling function :**

   This is one of the most important functions specially in the field of *industrial control processes*. In this case, the information is used by the instrument or the system to control the original measured quantity.

1.1.7. Applications of Measurement Systems

The instruments and measurement systems are used for different applications as mentioned below :

1. **Monitoring of processes and operations :**

   Examples : An *ammeter or a voltmeter* indicates the value of current or voltage being monitored (measured) at a particular instant.

   — *Water and electric energy meters* installed in homes keep track of commodity used so that later on its cost may be computed to be realised from the user.

2. **Control of processes and operations :**

   Examples : — Typical refrigeration system which *employs a thermostatic control. A temperature measuring device* (often a *bimetallic element*) senses the room temperature, thus providing the information necessary for proper functioning of the control system.

3. **Experimental engineering analysis :**

   Experimental engineering analysis has several uses, some of which are listed below :

   - Determination of system parameters, variables and performance indices.
   - Testing the validity of theoretical predictions.
   - Solutions of mathematical relationships with the help of analogies.
   - Formulation of generalised empirical relationships in cases where no proper theoretical backing exists.
   - For development in important spheres of study where there is ample scope of study.

1.1.8. Measurement System Performance

Following are the main two distinct categories of instruments and measurements characteristics :

1. **Static characteristics.** The main static characteristics are :

   (*i*) Accuracy
   (*ii*) Sensitivity
   (*iii*) Reproducibility
   (*iv*) Drift
   (*v*) Static error
   (*vi*) Dead zone.

2. **Dynamic characteristics.** The dynamic characteristics of a measurement system are :

   (*i*) Speed of response
   (*ii*) Measuring lag
   (*iii*) Fidelity
   (*iv*) Dynamic error.

1.2. CONTROL SYSTEMS

1.2.1. Introduction

Automatic control has played a significant role in the advance of engineering science. Besides its extreme importance in space-vehicle systems, missile-guidance systems, etc., automatic control has become an important and integral part of modern manufacturing and industrial processes. Automatic control, for example, is essential in ;

— Design of auto pilot systems in *aero space industries* ;
— Design of cars and trucks in the *automobile industries* ;
— Industrial operations as controlling pressure, temperature, humidity, viscosity, and flow in the *process industries*.

1.2.2. System

A system may be defined as follows :

- *"A **system** is an arrangement, set or collection of things connected or related in such a manner as to form an entirely or whole".*

Or

*"A **system** is an arrangement of physical components connected or related in such a manner as to form and / or act as entire unit.*

- A system consists of a sequence of components in which *each coponent has some cause as inpout and its effect will be its output. Broadly it is a sequential set of cause and effects.*

Each system may have a *large number of subsytems; "Examples"* :

(*i*) This universe is itself a system consisting of large number of subsystems.
(*ii*) Human body as a system has digestive system, respiratory system etc.

1.2.3. Control System

A ***control system*** *is an arrangement of physical components connected or related in such a manner as to command, direct or regulate itself or another system.*

Elements of a control system:

The elements of a control system are enumerated and defined below :

Element	*Definition*
1. *Controlled variable*	The quantitly or condition of the controlled system which can be directly measured and controlled is called *controlled varibale.*
2. *Indirectly controlled variable*	The quantity or condition related to controlled variable, but cannot be directly measured is called *indirectly controlled variable.*
3. *Command*	The input which can be independently varied is called *command.*
4. *Reference input*	A standard signal used for comparison in the close-loop system.
5. *Actuating signal*	The difference between the feedback signal and reference signal is called *actuating signal.*
6. *Disturbance*	Any signal other than the reference which affects the system performance is called *disturbance.*
7. *System error*	The difference between the actual value and ideal value is called *system error.*

Examples of control system applications:

Following are some examples of control system applications:

1. Steering control of automobile.
2. Print wheel control system.
3. Industrial sewing machine.
4. Sun-tracking control of solar collectors.
5. Speed control system.
6. Temperature control of an electric furnace.

1.2.4. Classification of Control systems

Control systems are classified into the following *two basic types* :

1. Open-loop control systems (Unmonitored or non–feedback control systems)
2. Closed–loop control systems (Monitored or feedback control systems).

Comparison between Open-loop and Closed-loop Systems

Open-loop	*Closed-loop*
1. Less accurate.	1. More accurate.
2. Generally build easily.	2. Generally complicated and costly.
3. Stability can be ensured.	3. May become unstable at times.
4. Presence of non-linearities cause malfunctioning.	4. It usually perfoms accurately even in the presence of non-linearities.
5. Any change is system component cannot be taken care of automatically.	5. Change in system component is automatically taken care of.
6. Input cammand is the sole factor responsible for providing the control action.	6. The control action is provided by the difference between the input command and the corresponding output.
7. The control adjustment depends upon human judgement and estimate.	7. The control adjustment depends on output and feedback element.
Examples : (*i*) Automatic washing machine. (*ii*) The electric switch. (*iii*) An automatic toaster. **Note:** All control systems *operated by present timing mechanisms are open-loop.*	*Examples :* (*i*) Liquid level control system. (*ii*) Traffic signal system. (*iii*) Human being reaching for an object.

1.2.5. Open-loop Control systems (Non-feedback Systems)

- *An Open-loop control system is one in which the control action is independent of the desired output.* The actuating signal depends only on the input command and output has *no control* over it.
- The *elements* of an open-loop control system can usually be divided into the following two parts (Refer to Fig. 1.4):

  (*i*) Controller;

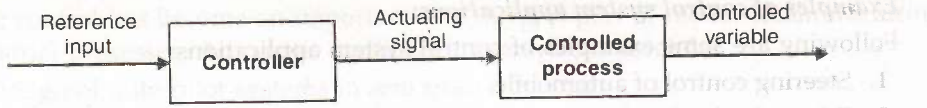


Fig. 1.4. Elements of an open-loop control system.

(*ii*) Controlled process.

— An input signal or command is applied to the controller, whose output acts as the actuating signal; the actuating signal then controls the controlled process so that the controlled variable will perform according to prescribed standards.

— *In simple cases,* the controller can be an *amplifier, mechanical linkage, filter,* or *other control element,* depending on the nautre of the system. In ***more sophisticated cases,*** the controller can be a computer such as a *microprocessor.*

— Because of the *simplicity and economy* of open-loop control systems we find this type of system in many non-critical applications.

Examples :

1. ***Idle-speed control system :***

- The following are the main objectives of the idle-speed control system of automobile :
  (*i*) To eleminate or minimize the speed drop when engine loading is applied.
  (*ii*) To maintain the engine speed at a desired value.
- Fig. 1.5 shows an idle-speed control system from the stand point of inputs-system-outputs. In this case the throttle angle and the load torque (due to the application of air conditioning, power steering, transmission, power brake, etc.) are the inputs, and the engine speed is the output. The engine is the controlled process of the system.

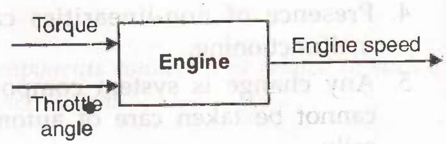


Fig. 1.5. Idle-speed control system.

2. ***Print wheel control system:***

Fig. 1.6 shows an example of the printwheel control system of a word processor or electronic typewriter (and also shows a typical input-ouput set for the system).

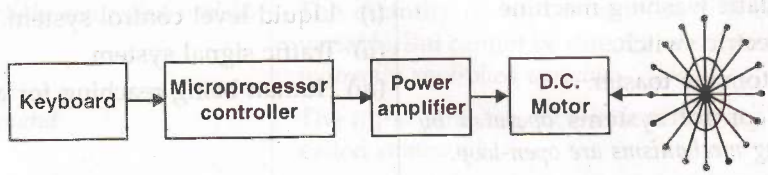


Fig. 1.6. Open-loop word processor control system.

— When a reference command input is given, the signal is represented as a step function. Since the electric windings of the motor have inductance and the mechanical load has inertia, the printwheel cannot respond to the input instantaneously. Typically it will follow the response and settle at the new position after sometime. Printing should not begin until the printwheel has come to complete stop; otherwise, the character will be smeared.

Advantages and limitations of open-loop control system :

Advantages :

1. Simple construction.
2. Easy maintenance.
3. Less costly than a closed-loop system.
4. No stability problem.
5. Convenient when output is difficult to measure or measuring the output precisely is economically not feasible.

Limitations/Disadvantages :

1. Since the system is affected by internal and external disturbances, the *output may differ from the desired value.*
2. For getting accurate results, this system needs frequent and careful calibrations.
3. Any change in system component cannot be taken care of automatically.
4. Presence of non-linearities causes malfuctioning.

1.2.6. Closed-loop Control System (Feedback Control System)

- **A closed-loop system** is one in *which control action is somehow dependent on the output.* In this case the controlled output is fed back through a *feedback element* and compared with the reference input. Thus *the actuating signal is the difference of desired output and reference input.*
- **Feedback** is *that property of a closed-loop system which permits the output or some other controlled variable of the system, to be compared with the input to the system, so that the appropriate control action may be formed as some function of the output and input. A feedback is said to exist in system when a closed sequence of cause and effect relations exists between system variables.*

The *Characteristics of feedback are as follows :*

(*i*) Increased bandwidth.
(*ii*) Increased accuracy.
(*iii*) Tendency towards oscillation or instability.
(*iv*) Reduced effects or non-linearities and distortion.
(*v*) Reduced sensitivity of the ratio of output to input to variations in system characteristics.

- *A closed-loop idle-speed control system* is shown in Fig. 1.7.

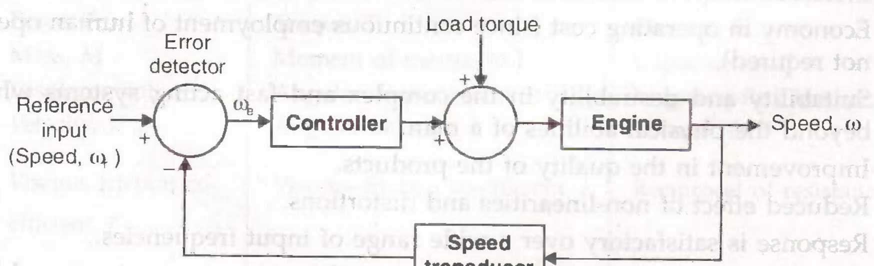


Fig. 1.7. Closed-loop idle-speed control system.

— The reference input (ω_r) sets the desired idling speed. The engine speed at idle should agree with reference value (ω_r), and any difference such as load torque is sensed by the speed transducer and the error detector. The controller will operate on the difference and provide a signal to adjust the throttle angle to correct the error.

Advantages and limitations :

Advantages :

1. More accurate comparatively.
2. Usually performs accurately even in the presence of non-linearities.
3. Change in system component is automatically taken care of.
4. The use of feedback system response is relatively insensitive to external distrubanes and internal variations in system parameters. It is thus *possible to use relatively inaccurate and inexpensive components to obtain the accurate control of a given plant* (whereas doing so is impossible in the open-loop case).

Limitations/Disadvantages :

1. Generally complicated in construction.
2. Generally higher in cost and power.
3. May become unstable at times.

1.2.7. Automatic Control Systems

- *A closed-loop control system operating without human operator is called an* **automatic control** system.

Examples :

(*i*) Centrifugal watt governor, where the lift of the rotating balls is used as speed monitor. The supply of steam is automatically controlled as speed tends to increase or decrease beyond a set point.

(*ii*) *A pressure control system* where the pressure inside the furnace is automatically controlled by affecting changes in the position of the damper.

(*iii*) *The level control system* where the inflow of water to the tank is dependent on the water level in the tank. The automatic controller maintains the liquid level by comparing the actual level with a desired level and correcting any error by adjusting the opening of the control valve.

Advantages and limitations :

Advantages :

1. Increased output.
2. Economy in operating cost (since continuous employment of human operator is not requred).
3. Suitability and desirability in the complex and fast acting systems which are beyond the physical abilities of a man.
4. Improvement in the quality of the products.
5. Reduced effect of non-linearities and distortions.
6. Response is satisfactory over a wide range of input frequencies.

Limitation : Automatic control system has a tendency to *overcorrect errors* which may *result in oscillations of constant or changing amplitude.*

1.2.8. Servo-Mechanism

A servo-mechanisim is a feedback control system used to *control position or its derivative.* It has the following essential *features :*

1. It is a closed-loop system.
2. It is used to control position, velocity or acceleration.
3. Its characteristics include :
   — automatic control;
   — remote operation;
   — high accuracy.
4. It has high power amplifying stages to operate the system from very small error to signal.

1.2.9. Regulator

A regulator is a system employed to *control quality which is to be kept constant for a fairly long interval.*

Example : Voltage regulator or speed regulator.

1.2.10. Representation Through Model

In order to solve a system problem, the specifications or description of the system configuration and its components must be put into a form amenable to analysis, design and evolution. Following three basic models may be used for various system :

1. Differential equations and other mathematical solutions.
2. Block diagrams.
3. Sign flow graphs (SFG).

1.2.11. Analogous Systems

For mathematical relations analogies are drawn between features of a system and features of some known elements or properties; some analogous systems are given below:

Table 1.1. Force-Current Analogy

S.No.	*Mechanical System*		*Electrical System*
	Translational	*Rotational*	
1.	Force, *F*	Torque, *T*	Current, *I*
2.	Mass, *M*	Moment of inertia, M.I.	Capacitance, C
3.	Displacement, *x*	Angular displacement, θ	Magnetic flux linkage, *f*
4.	Velocity, *V*	Angular velocity, ω	Voltage, *E*
5.	Viscous friction co-efficient, *f*	Viscous friction co-efficient, *f*	Reciprocal of resistance, $\frac{1}{R}$
6.	Spring stiffness, *K*	Torsional spring stiffness, *K*	Reciprocal of inductance, $\frac{1}{L}$

Table 1.2. Force-Voltage Analogy

S.No.	*Mechanical System*		*Elecrical System*
	Translational	*Rotational*	
1.	Force, F	Torque, T	Voltage, E
2.	Mass, M	Moment of inertia, M.I.	Inductance, L
3.	Displacement, x	Angular displacement, θ	Charge, q
4.	Velocity, U	Angular velocity, ω	Current, I
5.	Spring stiffness, K	Torsional spring stiffness, K	Reciprocal of capacitance, $\frac{1}{C}$
6.	Viscous friction co-efficient, f	Visocus friction Co-efficient, F	Resistance R

Table 1.3. Electrical, Thermal, Liquid level and Pneumatic Systems

S.No.	*Electrical systems*	*Thermal systems*	*Liquid-level systems*	*Pneumatic systems*
1.	Charge, coloumbs (C)	Heat flow, joules (J)	Liquid flow cum. (m^3)	Air flow, cum. (m^3)
2.	Current, amperes (A)	Heat flow rate, joules/sec. (J/s)	Liquid flow rate, cum/sec (m^3/s)	Air flow rate, cum/sec. (m^3/s)
3.	Voltage, volts (V)	Temperature, °C	Heat, meters (m)	Pressure, N/m^2
4.	Resistance, ohms (Ω)	Resistance, $°CsJ^{-1}$	Resistance, $m^{-2}s$	Resistance $N\text{-}ms^{-1}$
5.	Capacitance, farad (F)	Capacitance, J/°C	Capacitance, m^3/m	Capacitance, m^3/Nm^2

1.2.12. Block Diagram

A **block diagram** is the *diagrammatic representation of a physical system. The follwing steps* are worth noting :

— Firstly a functional block diagram is drawn to represent the functions of the system.

— Then it is *converted into a mathematical block diagram* by expressing the transfer function for each block.

— Finally is is *reduced to an equivalent simpler block diagram* for system analysis.

Fig. 1.8 shows a block diagram of the feedback control system.

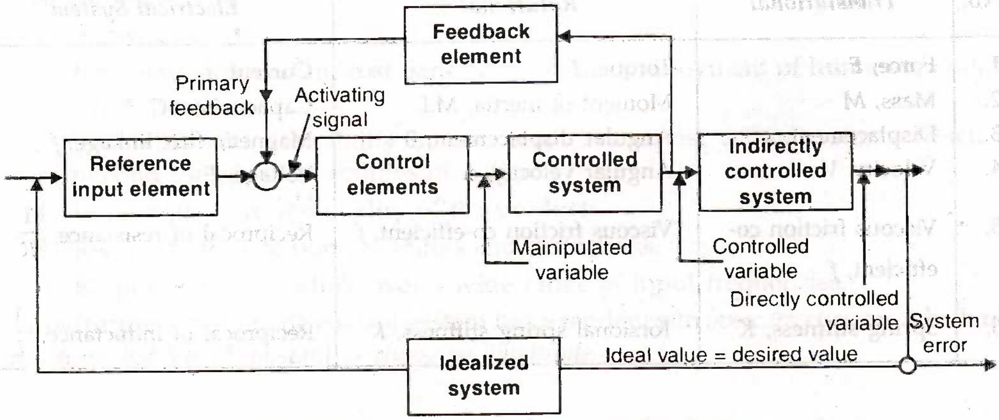


Fig. 1.8. Block diagram of the feedback control system.

1.2.13. Mathematical Block Diagram

Fig. 1.9 shows the block diagram of a closed-loop system. The various quantities shown are as follows :

$R(s)$ = Laplace transform of the reference input;

$C(s)$ = Laplace transform of the output;

$H(s)$ = Transfer function of the feedback path;

$B(s)$ = Laplace transform of the feedback signal = $C(s)\ H(s)$;

$E(s)$ = Laplace transform of the actuating signal = $R(s) - B(s) = R(s) - C(s)\ H(s)$;

$G(s)$ = Laplace transform of the formed path

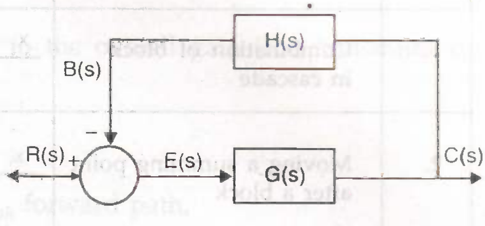


Fig. 1.9. Closed-loop system.

$$\therefore \quad C(s) = G(s)\ E(s) = G(s)\ R(s) - G(s)\ H(s)\ C(s)$$

$$\text{or,} \quad C(s) + G(s)\ H(s)\ C(s) = G(s)\ R(s)$$

$$\text{or,} \quad C(s)\ [1 + G(s)\ H(s)] = G(s)\ R(s)$$

$$\text{or,} \quad \frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)\ H(s)}$$

Hence the transfer function of the system,

$$M_s' = \frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)\ H(s)}$$

In the above equation the following points are worth noting :

(*i*) Product of transfer function of forward path and feedback path $G(s) \times H(s)$, sometimes expressed as $GH(s)$.

(*ii*) The system performance depends on its *characteristic eqation* (it is a key equation in the control system analysis) which is given as under :

$$1 + G(s)H(s) = 0.$$

Block reductions :

By using the rules (derived by simple algebraic manipulation of the equations representing the blocks) of block diagram algebra, a complex block diagram configuration can be simplified by certain rearrangements of block diagrams; such rules are given in the Table 1.4.

1.2.14. Signal Flow Graph

The block diagram reduction process, for complicated systems, becomes tedious and time consuming. For this purpose signal flow graphs (developed by S.J. Mason) are used.

A singal flow graph is a pictorial representation of the simultaneous equations describing a system.

Some important definitions relating to *signal flow graph* are given below :

1. **Input and output nodes.** A node having only outgoing branches is called *input node* while a node having only incoming branches is called *output node or sink.*

2. **Path.** Any continuous unidirectional succession of branches traversed in the indicated direction of branch is called *path.*

When the system has some roots with real parts equal to zero, but none with positive real parts, the system is said to be *"marginally stable"* which in unstable.

Routh stability criterion :

Routh stability criterion is a method for determining system stability that can be applied to an *n*th order characteristic equation of the form

$$a_n S^n + a_{n-1} S^{n-1} + \ldots\ldots + a_1 S + a_0 = 0$$

The Routh table is prepared as defined below :

s_n	a_n	a_{n-2}	a_{n-4}
S_{n-1}	a_{n-1}	a_{n-3}	a_{n-5}
:	b_1	b_2	b_3
:	c_1	c_2	c_3

After the array is completed the following criterion is applied :

"The number of changes in sign for the terms in the first column equals the number of roots of the characteristic equation with positive real parts.

Hence by the Routh criterion, for a system to be stable the array resulting from its characteristic equation must have a first column with terms of the same sign.

Deficiencies of Routh's criterion :

1. It does not provide the facility for selecting in a simple and direct fashion the parameters of a system component to stabilize the system when it is found to be absolutely unstable.
2. It assumes that characteristic equation is available in polynomial form; which is not necessarily always true.
3. The Routh array may show no change in sign in the first column but the ensuing dynamic response may be characterised by overshoots so excessive as to render the system useless for control purposes. Thus the system may be relatively unstable inspite of the fact that it is absolutely stable.
4. Although this criterion gives information about absolute stability, it conveys little or no information about how close the system may be to become unstable.

1.2.17. Frequency Response

The analysis of the *system* whose input is frequency and amplitude is dealt under frequency response. *The system is actuated by a sinusodal input and allowed to settle. The output amplitude and its phase with respect to input are measured. The phase difference and amplitude change indicate the nature of the system.*

Graphical methods :

The following four graphical methods are available to control systems analyses which are simpler and more direct than the time domain method for practical linear models of feedback control systems :

1. Bode's-Plot-Representation
2. Nyquist Diagrams
3. Nichols Charts
4. The Root Locus method

The first three are *frequency-domain techniques.*

Bode's Plot. This method has the following *advantages :*

(*i*) It is the simplest method.

(*ii*) The multiplication of magnitudes can be converted into addition.

(*iii*) Transfer function can be determined easily.

Nyquist method :

- This method *handles* systems with time delays without the necessity of approximations and hence yields exact results about both absolute and relative stability of the system.
- It is also useful for obtaining information about transfer functions of components or systems from experimental frequency response data.

Root locus method :

This method permits accurate computations of the time-domain response in addition to yielding readily available frequency response information.

1.2.18. Error Detector

An **error detector** is *a sensor to sense the error between the reference input and the desired output.*

- It gives an input to the amplifier and actuator in proportion to the error.
- Its output should be directly electrical or a transducer should be cascaded to give electrical output.
- An error-cum-transducer is obtained by connecting two potentiometers in parallel to a voltage source. Their movable points are brought out to give output voltage in proportion to the difference between the positions of the movable contacts.

1.2.19. LVDT

LVDT (Linear-Variable-Differential Transformer) is a transformer having one primary, and two secondary windings and movable core. The secondary windings are connected in series opposition, so as to have output which is difference of the two induced secondary voltages. The movable core is connected to the shaft and a normal position output voltage is zero. When the core moves the *output voltage is a function of the shaft position.*

1.2.20. Servo-Amplifier

A **servo-amplifier** is the amplifier used to *amplify the small output of the error detector to directly operate the actuator.*

- It can be electronic, magnetic or rotating.
- It should have high input impedance, low output impedance, frequency response curve should be flat in the range of operating frequencies, phase sensitive, small residual voltage and minimum noise.

1.2.21. Sampled Data Systems

These systems (also called *discrete time systems*) are *dynamic systems, in which one or more variables change at the discrete instant of time.* The time interval between two discrete instants is very small so that the data during this interval can be approximated by interpolation.

These systems find application in :

(*i*) Numerically controlled machine tool operations.

(*ii*) Pulse control or digital control of electric drives.

(*iii*) High spped tin plate rolling mill using quantized data for control.

(*iv*) Large complex systems employing telemetry links based on pulse modulation transmission of data.

1.2.22. Industrial Controllers

Industrial controllers may be classified *according to their control action* as follows :

1. Two-position or on-off controllers.
2. Proportional controllers.
3. Integral controllers.
4. Proportional-plus-integral controllers.
5. Proportional-plus-derivative controllers.
6. Proportional-plus-integral-plus-derivative controllers.

— Most industrial controllers use pressurised fuel such as oil or air or electricity as power sources. Consequently, controllers may also be classified according to the kind of power employed in the operation, such as *"pneumatic controllers"*, *"hydraulic controllers" or "electronic controllers"*. However, the kind of controllers to be used must be decided based on the nature of the plant and operating conditions, including such considerations as safety, cost, availability, reliability, accuracy, weight, and size.

1.2.23. Pneumatic Control Systems

- Pneumatic controllers use air control medium to provide an output signal which is a function of an input error signal.
- Fig 1.12 shows the schematics of a pneumatic control system, the major components are :

Error detector; Flopper nozzle (controller mechanism); Amplifier or Pilot relay.

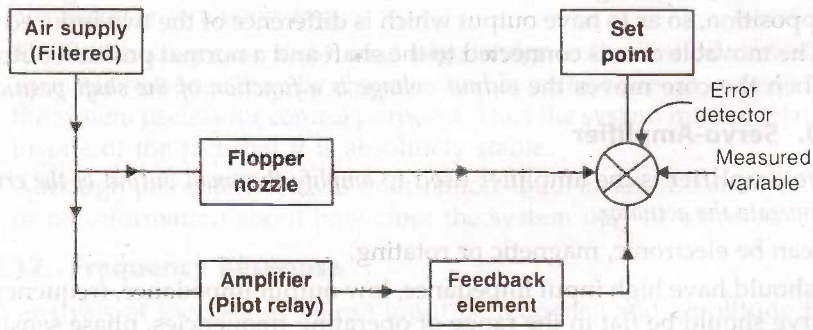


Fig. 1.12. Schematics of a pneumatic control system.

— The controller mechanisms are of two types : *Free balance* and *motion balance.*

Advantages :

1. Simple construction and easy maintenance.
2. Relatively high power amplification for operating the final control elements.
3. Relatively inexpensive power system.
4. No return pipes are required when air is used.
5. Insensitive to temperature changes.
6. Fire-and explosion-proof.
7. The normal operating pressure of pneumatic system is very much lower than that of hydraulic systems.

Limitations/Disadvantages :

1. Output powers are considerably less (than those of hydraulic systems).
2. Accuracy of pneumatic actuators is poor at low velocities.
3. Slow response of final control elements, and transmission lag.
4. Operation difficult under freezing conditions.
5. Lubrication of the mating parts is difficult.

Uses :

- The pneumatic systems are employed for majority of the plant and process control actions in petroleum, petrochemical, chemical, paper, textile and food industries.
- They are also sometimes used in the aircraft systems and guided missiles.

1.2.24. Hydraulic Control System

- Compressed air has seldom been used (except for low-pressure controllers) for the continuous control of the motion of devices having significant mass under external load forces. For such a case, hydraulic controllers are generally preferred.
- The widespread use of hydraulic circuitry in *"machine tool applications"*, *"aircraft control systems"* and *"similar operations"* occurs because of such factors as *accuracy, positiveness, flexibility, high power-to-weight ratio, fast starting, stopping, and reversal with smoothness and precision and simplicity of operations.*
  - — The operating pressure in hydraulic systems lies between 1 and 35 MPa; in some special applications the operating pressure may go upto 70 MPa.
  - — For the same power requirement, the weight and size of the hydraulic unit can be made smaller by increasing the supply pressure. *Very large force can be obtained with hydraulic systems.*
  - — With *hyraulic systems, rapid-acting, accurate positioning of heavy loads is possible.*
  - — A combination of electronic and hydraulic systems is widely used because it combines the advantages of both electronic control and hydraulic power.
- **Hydraulic controllers** *employ a liquid control medium to provide an output signal which is a function of an input error signal.*

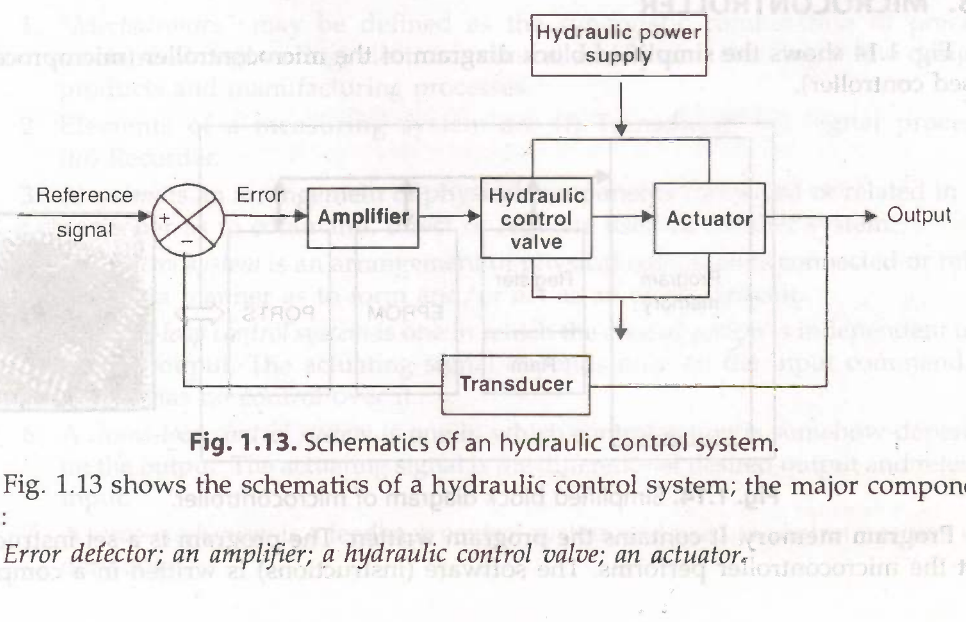


Fig. 1.13. Schematics of an hydraulic control system.

Fig. 1.13 shows the schematics of a hydraulic control system; the major components are :

Error detector; an amplifier; a hydraulic control valve; an actuator.

— Hydraulic power supply system is of the following two types : *"Constant flow arrangement"* and *"Constant pressure arrangement"*

Advantages :

1. Because of low leakages in hydraulic actuators, *speed drop* when loads are applied is *small.*
2. Hydraulie actuators have a *higher speed of response with fast starts, stops, and speed reversals.*
3. Availability of both linear and rotary actuators gives *flexibility in design.*
4. Simplicity of actuator system.
5. Operation of hydraulic actuators under continuous, intermittent, reversing and stalled conditions without damage is possible.
6. Large forces or torques can be developed by the comparatively small sized hydraulic actuators.
7. Long life due to self lubricating properties of the hydraulic liquids.

Disadvantages/Limitations :

1. In order to prevent the leakage of hydraulic fluid, the proper seals and connections are needed.
2. Unless fire-resistant fluids are used, fire and explosion hazards exist.
3. For keeping the fluid clean and pure careful maintenance of the system is required.
4. As a result of the non-linear and other complex characteristics involved, the design of sophisticated hydraulic systems is quite complicated.
5. Contaminated oil may cause failure in the proper functioning of a hydraulic system.

Uses : The hydraulic systems, because of their high power-to-weight ratio find a wide range of use in :

— Machine tools;

— Speed governing systems;

— Position control systems.

1.3. MICROCONTROLLER

Fig. 1.14 shows the simplified block diagram of the microcontroller (microprocessor based controller).

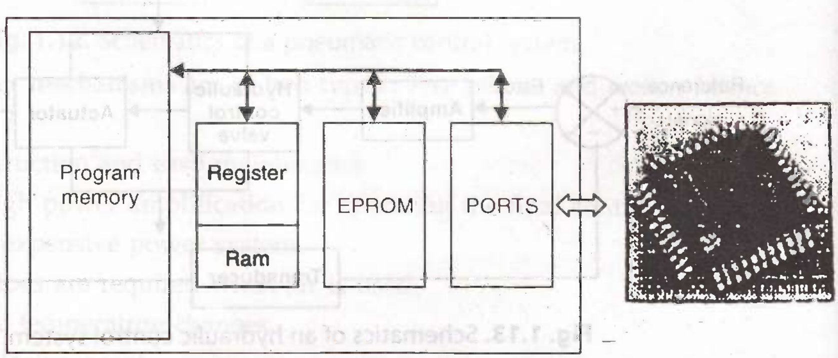


Fig. 1.14. Simplified block diagram of microcontroller.

Program memory. It contains the program written. The program is a set instruction that the microcontroller performs. The software (instructions) is written in a computer and then programmed (burned) into the "program memory". This memory is a EPROM memory which can be rewritten thousand times.

Register and Ram box. It contains all the *internal registers and a small Ram* where data can be stored temporarily. *There are several registers with different functions.*

— The Ram memory is not large about 64-128 byte.

— *The content in the Register and Ram-info will disappear when the power is off.*

EPROM-memory. It is a small memory where data can be read as well as written, but the *data will not disappear when the power is off.* Next time the power is on we can go into this memory and fetch the data again.

Ports. The port is *input and output pins of the actual circuit.* We can define the pins as input or outputs. By writing or reading to the port we can conrol each pin as we wish.

"Microprocessors" are fastly replacing the mechanical controllers (e.g. cam-operated switch) and being used in general to carry out control functions. They have the great *advantage* that *a great variety of programs become feasible.*

— In several simple systems there might be just an embedded microcontroller, this being a microprocessor with memory all integrated on one clip, which has been specifically programmed for the task concerned.

Programmable logic controller (Fig 1.15) is a more adoptable form. This is a microprocessor based controller which uses programmable memory to store instructions and to implement functions such as *logic, sequence, timing, counting* and *arithmetic* to *control events and can be readily programmed for different tasks.*

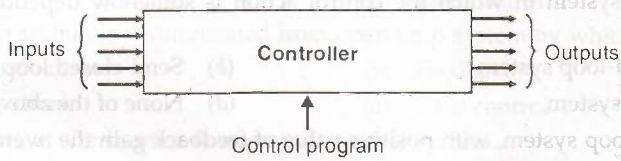


Fig. 1.15. Programmable logic controller

HIGHLIGHTS

1. *"Mechatronics"* may be defined as the synergistic combination of precision mechanical engineering, electronic control and system thinking in the design of products and manufacturing processes.
2. Elements of a measuring system are (*i*) Transducer, (*ii*) Signal processor, (*iii*) Recorder.
3. A *system* is an arrangement of physical components connected or related in such a manner as to command, direct or regulate itself or another system.
4. An *control system* is an arrangement of physical components connected or related in such a manner as to form and/or act as an entire circdcuit.
5. An *open-loop control system* is one in which the control action is independent of the desired ouput. The actuating signal depends only on the input command and output has no control over it.
6. A *closed-loop control system* is one in which control action is somehow dependent on the output. The actuating signal is the difference of desired output and reference input.
7. A *servo-mechanism* is a feedback control system and used to control position or its derivative.

8. A *regulator* is a system employed to control quality which is to be kept constant for a fairly long interval.
9. A *block diagram* is the diagrammatic representation of a physical system.
10. A *signal flow graph* is a pictorial representation of the simultaneous equations describing a system.
11. The responese of a system to input or disturbances determines its *stability*.

OBJECTIVE TYPE QUESTIONS

Chosse the Correct Answer :

1. In an open-loop control system
   (a) output is independent of control input
   (b) output is dependent on control input
   (c) only system parameters have effect on the control output
   (d) none of the above.
2. For open control system which of the following statements is *incorrect* ?
   (a) Less expensive.
   (b) Recalibration is not required for maintaining the required quality of the ouput.
   (c) Construction is simple and maintenance easy.
   (d) Errors are caused by distrubances.
3. A control system in which the control action is somehow dependent on the output is known as
   (a) Closed-loop system (b) Semi-closed loop system
   (c) Open-system (d) None of the above.
4. In closed-loop system, with positive value of feedback gain the overall gain of the system will
   (a) decrease (b) increase
   (c) be unaffected (d) any of the above.
5. Which of the following is an open-loop control system ?
   (a) Field-controlled D.C. motor (b) Ward leonard control
   (c) Metadyne (d) Stroboscope.
6. Which of the follwing statements is *not* necessarily correct for open control system?
   (a) Input command is the sole factor responsible for providing the control action.
   (b) Presence of non-linearities causes malfunctioning.
   (c) Less expensive.
   (d) Generally free from problems of non-linearities.
7. In open-loop system
   (a) the control action depends on the size of the system.
   (b) the control action depends on system variables.
   (c) The control action depends on the input signal.
   (d) the control action is independent of the output.
8. .... has tendency to oscillate.
   (a) Open-loope system (b) Closed-loop system
   (c) Both (a) and (b) (d) Neither (a) nor (b).

9. A good control system has all the following features *except*
(*a*) good stability (*b*) slow response
(*c*) good accuracy
(*d*) sufficient power handling capacity.

10. A car is running at a constant speed of 50 km/h, which of the following is the feedback element for the driver?
(*a*) Clutch (*b*) Eyes
(*c*) Needle of the seedometer (*d*) Steering wheel
(*e*) None of the above.

11. The initial response when the output is not equal to input is called.
(*a*) Transient response (*b*) Error response
(*c*) Dynamic response (*d*) Any of the above.

12. A control system working under unknown random actions is called ....
(*a*) Computer control system (*b*) Digital data system
(*c*) Stochastic control system (*d*) Adaptive control system.

13. An automatic toaster is a .... loop control system.
(*a*) open (*b*) closed
(*c*) partially closed (*d*) any of the above.

14. Any externally introduced signal affecting the controlled output is called a
(*a*) feedback (*b*) stimulus
(*c*) signal (*d*) gain control.

15. A closed-loop system is distinguished from open-loop system by which of the following?
(*a*) Servo-mechanism (*b*) Feedback
(*c*) Output pattern (*d*) Gain control.

16. .... is a part of the human temperature control system.
(*a*) Digestive system (*b*) Perspiration system
(*c*) Ear (*d*) Leg movement.

17. By which of the following the control action is determined when a man walks along a path?
(*a*) Brain (*b*) Hands
(*c*) Legs (*d*) Eyes.

18. .... is a closed-loop system.
(*a*) Auto-pilot for an aircraft (*b*) Direct current generator
(*c*) Car starter (*d*) Electric switch.

19. Which of the following devices are commonly used as error detectors in instruments?
(*a*) Vernisats (*b*) Microsyns
(*c*) Resolvers (*d*) Any of the above.

20. Which of the following should be done to make an unstable system stable ?
(*a*) The gain of the system should be decreased.
(*b*) The gain of the system should be increased.
(*c*) The number of poles to the loop transfer function should be increased.
(*d*) The number of zeros to the loop transfer function should be increased.

21. .... increases the steady state accuracy.
(*a*) Integrator (*b*) Differentiator
(*c*) Phase lead compensator (*d*) Phase lag compensator.

22. A.C. servomotor resembles ....
(a) two-phase induction motor
(b) three-phase induction motor
(c) direct current series motor
(d) universal motor.

23. As a result of introduction of negative feedback which of the following will *not* decrease?
(a) Bandwidth
(b) Overall gain
(c) Distortion
(d) Instability.

24. Regenerative feedback implies feedback with
(a) oscillations
(b) step input
(c) negative sign
(d) positive sign.

25. The output of a feedback control system must be a function of
(a) reference and output
(b) reference and input
(c) input and feedback singnal
(d) output and feedback signal.

26. .... is an open-loop control system
(a) Ward Leonard control
(b) Field-controlled D.C. motor
(c) Stroboscope
(d) Metadyne.

27. A control system with excessive noise, is likely to suffer from
(a) saturation in amplifying stages
(b) loss or gain
(c) vibrations
(d) oscillations.

28. Zero initial condition for a system means ....
(a) input reference signal is zero
(b) zero stored energy
(c) no initial movement of moving parts
(d) system is at rest and no energy is stored in any of its components.

29. Transfer function of a system is used to calculate which of the following?
(a) The order of the system
(b) The time constant
(c) The output for any given input
(d) The steady state gain.

30. The bandwidth, in a feedback amplifier,
(a) remains unaffected
(b) decreases by the same amount as the gain increase
(c) increases by the same amount as the gain decrease
(d) decreases by the same amount as the gain decrease.

31. On which of the following factors does the sensitivity of a closed-loop system to gain changes and load disturbances depend?
(a) Frequency
(b) Loop gain
(c) Forward gain
(d) All of the above.

32. The transient response, with feedback system
(a) rises slowly
(b) rises quickly
(c) decays slowly
(d) decays quickly.

33. The second derivative input signals modify which of the following?
(a) The time constant of the system.
(b) Damping of the system.
(c) The gain of the system.
(d) The time constant and suppress the oscillations.
(e) None of the above.

34. Which of the following statements is *correct* for any closed-loop system ?
(a) All the co-efficients can have zero value.
(b) All the co-effecients are always non-zero.
(c) Only one of the static error coefficients has a finite non-zero value.
(d) None of the above.

35. Which of the following statements is *correct* for a system with gain margin close to unity or a phase margin close to zero?
(*a*) The system is relatively stable. (*b*) The system is highly stable.
(*c*) The system is highly oscillatory. (*d*) None of the above.
36. Due to which of the following reasons excessive band-width in control system should be avoided?
(*a*) It leads to slow speed of response. (*b*) It leads to low relative stability.
(*c*) Noise is proportional to bandwidth. (*d*) None of the above.
37. In a stable control system backlash can cause which of the following?
(*a*) Underdamping (*b*) Overdamping
(*c*) Poor stability at reduced values of open-loop gain
(*d*) Low-level oscillations.
38. In an automatic control system which of the follwing elements is *not* used?
(*a*) Error detector (*b*) Final control element
(*c*) Sensor (*d*) Oscillator.
39. In a control system the output of the controller is given to
(*a*) final control element (*b*) amplifier
(*c*) comparator (*d*) sensor
(*e*) none of the above.
40. A controller, essentially, is a
(*a*) sensor (*b*) clipper
(*c*) comparator (*d*) amplifier.
41. Which of the follwing is the input to a controller?
(*a*) Servo signal (*b*) Desired variable value
(*c*) Error signal (*d*) Sensed signal.
42. The on-off controller is a .... system.
(*a*) digital (*b*) linear
(*c*) non-linear (*d*) discontinuous.
43. The capacitance, in force-current analogy, is analogous to
(*a*) momentum (*b*) velocity
(*c*) displacement (*d*) mass.
44. The temperature, under thermal and electrical system analogy, is considered analogous to
(*a*) voltage (*b*) current
(*c*) capacitance (*d*) charge
(*e*) none of the above.
45. In electrical-pneumatic system analogy the current is considered analogous to
(*a*) velocity (*b*) pressure
(*c*) air flow (*d*) air flow rate.
46. In liquid level and electrical system analogy, voltage is considered analogous to
(*a*) head (*b*) liquid flow
(*c*) liquid flow rate (*d*) none of the above.
47. The viscous friction co-efficient, in force-voltage analogy, is analogous to
(*a*) charge (*b*) resistance
(*c*) reciprocal of inductance (*d*) reciprocal of conductance
(*e*) none of the above.
48. In force-voltage analogy, velocity is analogous to
(*a*) current (*b*) charge

(c) inductance (d) capacitance.

49. In thermo-electrical analogy charge is considered analogous to
(a) heat flow (b) reciprocal of heat flow
(c) reciprocal of temperature (d) temperature
(e) none of the above.

50. Mass, in force-voltage analogy, is analogous to
(a) charge (b) current
(c) inductance (d) resistance.

51. The transient response of a system is mainly due to
(a) inertia forces (b) internal forces
(c) stored energy (d) friction.

52. .... signal will become zero when the feedback signal and reference signs are equal.
(a) Input (b) Actuating
(c) Feedback (d) Reference.

53. A signal other than the reference input that tends to affect the value of controlled variable is known as ....
(a) disturbance (b) command
(c) control element (d) reference input.

54. The transfer function is applicable to which of the following?
(a) Linear and time-invariant systems (b) Linear and time-variant systems
(c) Linear systems (d) Non-linear systems
(e) None of the above.

55. From which of the following transfer function can be obtained?
(a) Signal flow graph (b) Analogous table
(c) Output-input ratio (d) Standard block systems
(e) None of the above.

56. .... is the reference input minus the primary feedback.
(a) Manipulated variable (b) Zero sequence
(c) Actuating signal (d) Primary feedback.

57. The term backlash is associated with
(a) servomotors (b) induction relays
(c) gear trains (d) any of the above.

58. With feedback .... increases.
(a) system stability (b) sensitivity
(c) gain (d) effects of disturbing signals.

59. By which of the following the system response can be tested better?
(a) Ramp input signal (b) Sinusoidal input signal
(c) Unit impulse input signal (d) Exponetially decaying signal.

60. In a system zero initial condition means that
(a) the system is at rest and no energy is stored in any of its components
(b) the system is working with zero stored energy
(c) the system is working with zero reference signal.
(d) none of the above.

61. In a system low friction co-efficient facilitates
(a) reduced velocity lag error (b) increased velocity lag error
(c) increased speed of response
(d) reduced time constant of the system.

62. Hydraulic torque transmission system is analog of
(a) amplidyne set
(b) resistance-capacitance parallel circuit
(c) motor-generator set (d) any of the above.

63. Spring constant in force-voltage analogy is analgous to
(a) capacitance (b) reciprocal of capacitance
(c) current (d) resistance.

64. The frequency and time domain are related through which of the following?
(a) Laplace Transform and Fourier Integral
(b) Laplace Transform (c) Fourier Integral
(d) Either (b) or (c).

65. An increase in gain, in most systems, leads to
(a) smaller damping ratio (b) larger damping ratio
(c) constant damping ratio (d) none of the above.

66. Static error co-efficients are used as a measure of the effectiveness of closed-loop systems for specified .... input signal.
(a) acceleration (b) velocity
(c) position (d) all of these.

67. A conditionally stable system exhibits poor stability at
(a) low frequencies (b) reduced values of open-loop gain
(c) increased values of open-loop gain (d) none of the above.

68. The type 0 system has .... at the origin
(a) no pole (b) net pole
(c) simple pole (d) two poles
(e) none fo the above.

69. The type 1 system has .... at the origin.
(a) no pole (b) net pole
(c) simple pole (d) two poles.

70. The type 2 system has .... at the origin.
(a) no net pole (b) net pole
(c) simple pole (d) two poles.

71. The position and velocity errors of a type-2 system are
(a) constant, constant (b) constant, infinity
(c) zero, constant (d) zero, zero.

72. Velocity error constant of a system is measured when the inut to the system is unit .... function.
(a) parabolic (b) ramp
(c) impulse (d) step.

73. In case of type-1 system steady state acceleration is
(a) unity (b) infinity
(c) zero (d) 10.

74. If a step function is applied to the input of a system and the output remains below a certain level for all the time, the system is
(a) not necessarily stable (b) stable
(c) unstable (d) always unstable.
(e) any of the above.

75. Which of the following is the best method for determining the stability and transient response?
(*a*) Root locus (*b*) Bode plot
(*c*) Nyquist plot (*d*) None of the above.

76. Phase margin of a system is used to specify which of the following?
(*a*) Frequency response (*b*) Absolute stability
(*c*) Relative stability (*d*) Time response.

77. Addition of zeros in transfer function causes which of the follwing?
(*a*) Lead-compensation (*b*) Lag-compensation
(*c*) Lead-lag compensation (*d*) None of the above.

78. .... technique is *not* applicable to non-linear system?
(*a*) Nyquist Criterion (*b*) Quasi linearization
(*c*) Functional analysis (*d*) Phase-plane representation.

79. In order to increase the damping of a badly underdamped system which of following compensators may be used?
(*a*) Phase-lead (*b*) Phase-lag
(*c*) Both (*a*) and (*b*) (*d*) Either (*a*) or (*b*)
(*e*) None of the above.

80. The phase-lag produced by transportation relays
(*a*) is independent of frequency
(*b*) is inversely proportional to frequency
(*c*) increases linearly with frequency
(*d*) decreases linearly with frequency.

81. In a stable control system saturation can cause which of the follwing?
(*a*) Low-level oscillations (*b*) High-level oscillations
(*c*) Conditional stability (*d*) Overdamping.

82. Which of the follwing can be measured by the use of a tacho-generator?
(*a*) Acceleration (*b*) Speed
(*c*) Speed and acceleration (*d*) Displacement
(*e*) None of the above.

83. .... is *not* a final control element.
(*a*) Control valve (*b*) Potentiometer
(*c*) Electro-pneumatic converter (*d*) Servomotor.

84. Which of the following is the definition of proportional band of a controller?
(*a*) The range of air output as measured variable varies from maximum to minimum.
(*b*) The range of measured variables from set value.
(*c*) The range of measured variables through which the air output changes from maximum to minimum.
(*d*) Any of the above (*e*) None of the above.

85. In pneumatic control systems the control valve used as final control element converts ....
(*a*) pressure signal to electric signal (*b*) pressure signal to position change
(*c*) electric signal to pressure signal (*d*) position change to pressure signal
(*e*) none of the above.

ANSWERS

1. (*a*)	2. (*b*)	3. (*a*)	4. (*a*)	5. (*a*)	6. (*b*)	7. (*d*)
8. (*b*)	9. (*b*)	10. (*c*)	11. (*a*)	12. (*c*)	13. (*a*)	14. (*b*)
15. (*b*)	16. (*b*)	17. (*d*)	18. (*a*)	19. (*d*)	20. (*b*)	21. (*a*)

22. (a)	23. (a)	24. (d)	25. (a)	26. (b)	27. (a)	28. (d)
29. (c)	30. (c)	31. (d)	32. (d)	33. (d)	34. (c)	35. (c)
36. (c)	37. (d)	38. (d)	39. (a)	40. (c)	41. (c)	42. (c)
43. (d)	44. (a)	45. (d)	46. (a)	47. (b)	48. (a)	49. (d)
50. (c)	51. (c)	52. (b)	53. (a)	54. (a)	55. (a)	56. (c)
57. (c)	58. (a)	59. (c)	60. (a)	61. (a)	62. (c)	63. (b)
64. (a)	65. (a)	66. (d)	67. (b)	68. (a)	69. (c)	70. (d)
71. (d)	72. (b)	73. (b)	74. (a)	75. (d)	76. (c)	77. (b)
78. (a)	79. (a)	80. (c)	81. (c)	82. (b)	83. (b)	84. (c)
85. (b).						

THEORETICAL QUESTIONS

1. What is "Mechatronics"?
2. Define the term "Mechatronics" and give four examples of mechatronic systems.
3. What are the elements of a measuring system?
4. Enumerate and explain briefly the elements of a measuring system, with an example.
5. State the functions of instruments and measurement systems.
6. List the applications of measurement systems.
7. What are the main two distinct categories of instruments and measurement characteristics?
8. Define a 'system'.
9. What is a 'Control system'?
10. Enumerate and define the elements of a control system.
11. List four examples of control system applications?
12. How are control systems classified?
13 What is an'open-loop' control system?
14. What are the elements of an 'open-loop' control system?
15. Explain briefly two examples of 'open-loop' control system.
16. State the advantages and disadvantages of open-loop control system.
17. What is 'closed-loop' control system?
18. Define the term 'feedback'.
19. State the characteristics of 'feedback'.
20. Explan briefly a 'closed-loop' control system with an example.
21. State the advantages and limitations/disadvantages of a 'closed-loop' control system.
22. What is an 'automatic control system'? What are its advantages and limitations?
23. What is a block diagram?
24. What is a signal flow graph?
25. What do you understand by the term 'stability'?
26. How are industrial controllers classified?
27. Explain briefly a 'Pneumatic control system'. State its advantages and disadvantages.
28. Describe briefly 'Hydraulic control system'. State its advantages and disadvantages.
29. What is a microcontroller?
30. Explain briefly a microcontroller, with a simplified block diagram.

CHAPTER

2 Basic and Digital Electronics

1.1 Electronic Components : Introduction – Active components – Passive components. 2.2. Electronic Devices : General aspects - Semiconductors - Intrinsic semiconductor - Extrinsic semiconductor - PN junction diode - Zener diode - Tunnel diode - Bipolar junction transistor (BJT) - Field-effect transistor (FET) - Unijunction transistor (UJT) - Thyristor - Optoelectronic devices - Rectifiers. 2.3. Digital Electronics : Introduction - Advantages and disadvantages of digital electronics - Digital circuit - Logic gates - Universal gates - Half adder - Full adder - Boolean algebra - Boolean laws - De Morgan's theorems - Operator precedence - Duals - Logic system - Flip - flop circuits - Counters - Register - Logic families - Integrated circuits - Operational amplifiers. Highlights - Objective Type Questions – Theoretical Questions.

2.1 ELECTRONIC COMPONENTS

2.1.1. Introduction

In order to obtain a particular function electronics circuits are designed with a number of electronic components suitably connected. A few *basic components* used in all the electronic circuits are :

Tube devices ; Semiconductor devices ..... called ***Active components.***

Resistors ; Capacitors ; Inductors ; ..... called ***Passive components.***

2.1.2. Active Components

The electronic components which are capable of amplifying or processing an electrical signal are called ***active components.***

Examples :

(*i*) *Tube devices :*

— *Vacuum tubes* (*e.g.*, vacuum diode, vacuum triode, vacuum pentode, etc.)

— *Gas tubes* (*e.g.*, gas diode, thyratron etc.)

(*ii*) *Semiconductor* (*solid state*) *devices* (e.g., junction diode, zener diode, transistor, FET. UJT, SCR, etc.)

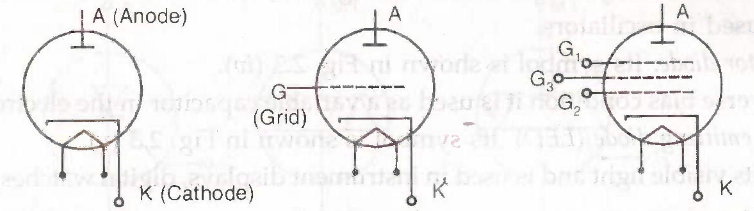
2.1.2.1. Tube devices

The various types of tube devices are discussed below :

(*a*) **Vacuum tubes :**

(*i*) *Vacuum diode.* Its symbol is shown in Fig. 2.1 (*i*).

- It is used as a rectifier and detector.



(i) Vacuum diode (ii) Vacuum triode (ii) Vacuum pentode

Fig. 2.1. Vacuum tubes.

(*ii*) *Vacuum triode.* Its symbol is shown in Fig. 2.1 (*ii*).

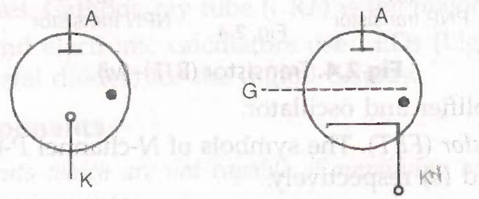
- It is used as amplifier and oscillator.

(*iii*) *Vacuum pentode.* Its symbol is shown in Fig. 2.1 (*iii*).

- It is used as amplifier and oscillator.

(*b*) **Gas tubes :**

(*i*) *Gas diode.* Its symbol is shown in Fig. 2.2 (*i*).



(i) Gas diode (ii) Gas triode (Thyratron)

Fig. 2.2. Gas tubes.

- It is used as voltage regulator and in neon signs.

(*ii*) *Gas triode.* (*thyratron*). Its symbol is shown in Fig. 2.2 (*ii*).

- It is used as controlled rectifier.

2.1.2.2. Semiconductor devices

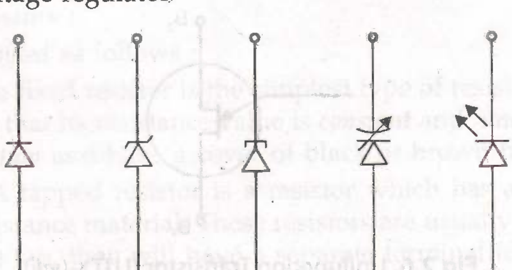
The various semiconductor devices are discussed as follows :

(*i*) *Junction diode.* Its symbol is shown in Fig. 2.3 (*i*).

- It is used as rectifier, detector and in switching circuits.

(*ii*) *Zener diode.* Its symbol is shown in Fig. 2.3. (*ii*).

- It is used as voltage regulator.



(i) Junction diode (ii) Zener diode (iii) Tunnel diode (iv) Varactor diode (v) LED

Fig. 2.3. Diodes.

(*iii*) *Tunnel diode.* Its symbol is shown in Fig. 2.3 (*iii*).

- It is used in oscillators.

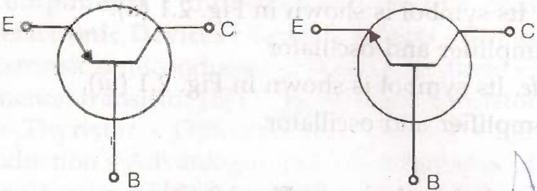
(*iv*) *Varactor diode.* Its symbol is shown in Fig. 2.3 (*iv*).

- In reverse bias condition it is used as a variable capacitor in the electronic circuits.

(*v*) *Light emitting diode* (*LED*). Its symbol is shown in Fig. 2.3 (*v*).

- It emits visible light and is used in instrument displays, digital watches, calculators, etc.

(*vi*) *Bipolar Junction Transistor* (*BJT*). The symbols of PNP and NPN transistors are shown in Fig. 2.4 (*a*) and (*b*) respectively.

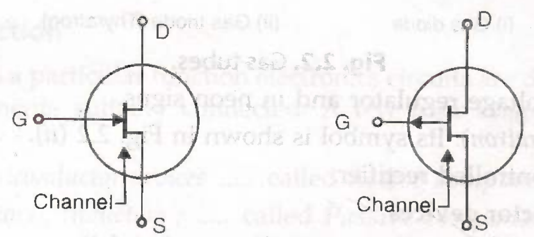


(a) PNP transistor   Fig. 2.4   (b) NPN transistor

Fig 2.4. Transistor (BJT)–(*vi*)

- It is used as amplifier and oscillator.

(*vii*) *Field Effect Transistor* (*FET*). The symbols of *N*-channel *P*-channel FET are shown in Fig. 2.5. (*a*) and (*b*) respectively.



(a) N-channel FET   (a) P-channel FET

Fig 2.5. Field effect transistor (FET)–(*vii*)

- It is used as amplifier and oscillator.

(*viii*) *Unijunction Transistor* (*UJT*). Its symbol is shown in Fig. 2.6.

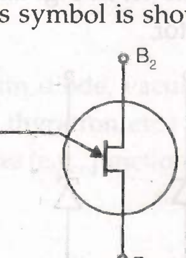


Fig 2.6. Unijunction Transistor (UJT)–(*viii*)

- It is used in power controls and switching circuits.

(*ix*) *Silicon Controlled Rectifier* (*SCR*). Its symbol is shown in Fig. 2.7 (*a*).

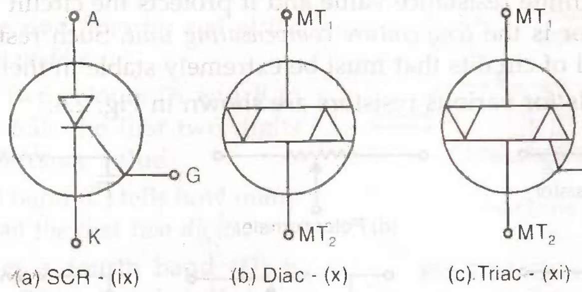


Fig. 2.7. SCR, Diac, Triac.

- It is used for speed control of motors and power controls.

(*x*) *Diac.* Its symbol is shown in Fig. 2.7 (*b*).

- Generally it is used to give a pulse to the gate of triac.

(*xi*) *Triac.* Its symbol is shown in Fig. 2.7 (*c*).

- It is a bidirectional device and is used to obtain regulated A.C. at the output.

(*xii*) *Visual display devices.* Cathode ray tube (*CRT*) is the major visual display device.

— Digital watches and electronic calculators use LEDs (Light emitting diodes) or LCDs (Liquid crystal diodes) for the digital display.

2.1.3. Passive Components

The electronic components which are not capable of amplifying or processing an electrical signal are called **passive components.**

Examples. Resistors ; inductors ; capacitors.

These components are as important as active ones, since the active devices cannot process the electrical signals without their assistance.

2.1.3.1. Resistors

A resistor entails the following *two main characteristics* :

(*i*) Its *resistance* (*R*) in ohms. ..... The resistors are available from a fraction of an ohm to many mega ohms.

(*ii*) The *wattage rating.* ..... The power rating may be as high as several hundred watts or as low as $\frac{1}{10}$ watt. *Power rating indicates the maximum wattage the resistor can dissipate without excessive heat* (Too much heat can make the resistor burn open).

Classification of resistors :

The resistors are *classified* as follows :

1. *Fixed resistors.* The fixed resistor is the simplest type of resistor. Fixed means that the unit is so constructed that its resistance value is *constant* and *unchangeable.* These are made of a *carbon composition* and have a cover of black or brown hard plastics.

2. *Tapped resistors.* A tapped resistor is a resistor which has a tap, or connection, somewhere along the resistance material. These resistors are usually wire wound type. If they have more than one tap, they will have a separate terminal for each.

3. *Variable resistors.* A variable resistor has a movable contact that is used to adjust or select the resistance value between two or more terminals. A variable resistor is commonly called a *control.*

4. *Special resistors.* The most common type of special resistor is the fusible type. A

fusible resistor has a definite resistance value and it protects the circuit much like a fuse. Another special resistor is the *temperature compensating unit*. Such resistors are used to provide special control of circuits that must be extremely stable in their operation.

Schematic symbols for various resistors are shown in Fig. 2.8.

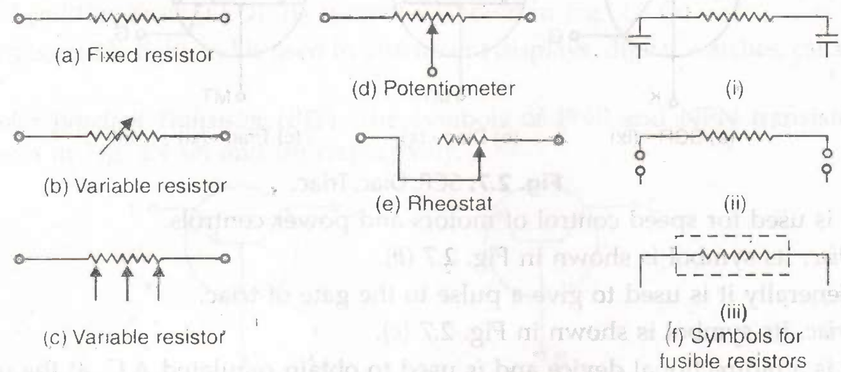


Fig. 2.8. Schematic symbols for various resistors.

The following types of resistors are used in electrical circuits :

(*i*) Carbon resistors.

(*ii*) Wire-wound resistors on ceramic or plastic forms (as in case of rheostats etc.).

(*iii*) Deposited carbon resistors on ceramic base.

(*iv*) Deposited metal resistors on ceramic base.

(*v*) Printed, painted or etched circuit resistors.

Resistor colour coding :

— Resistance is measured in units called *ohms*.

— *Wire wound resistors normally have their values in ohms and tolerance in percent stamped on them.*

— *For carbon or composition resistors a colour code is used.*

The resistance values, for several years have been coded by *three coloured bands* painted around the body of the resistors. If the tolerance is either 5 or 10 percent, a *fourth colour* band is added. Position of the bands is shown in Fig. 2.9.

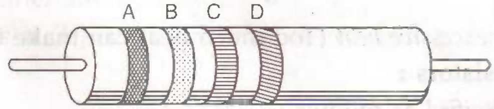


Fig. 2.9. The colour code system : colour bands indicate resistance value.

Colours and numbers :

Each of the colours represents one of the ten digits—0 through 9—as follows :

Colour	*Number*	*Colour*	*Number*
Black	0	Green	5
Brown	1	Blue	6
Red	2	Violet	7
Orange	3	Grey	8
Yellow	4	White	9

The *bands are read from the end of the resistor toward the middle.*

— The first two colours (*A* and *B* in Fig. 2.9.) tells the first two digits in the resistance value.

— The third band (C) tells how many *zeros follow the first two digits.*

— Sometimes a fourth band (*D*) is present. This band tells the tolerance and will be either gold or silver. A gold band means 5% tolerance, silver 10% and no fourth band, 20%. The tolerance band tells how close the resistance should be to the value shown by the other three bands.

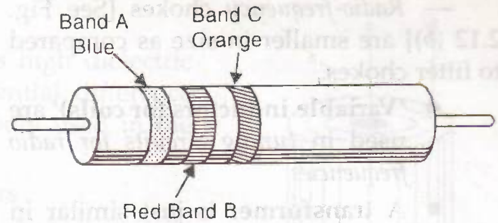


Fig 2.10. Colour code used on a 62000-ohm resistor.

The procedure of reading the bands is given below. Refer to Fig. 2.10 :

Band	*A*	*B*	*C*	*D*
Colour Numbers	Blue 6	Red 2	Orange 3 zeros	No band 20% tolerance

The blue-red-orange bands signify 62 followed by three zeros and would be read as 62000 ohms ± 20%.

2.1.3.2. Inductors

*An **inductor** is an electronic component (usually a coil) which opposes the change of current in the circuit.*

— *The property of the coil due to which it opposes any increase or decrease of current or flux through it, is known as **Self-inductance**.*

— Self induction is sometimes analogously called electromagnetic or electrical inertia.

— The unit of inductance (L) is *henry* (H).

- An inductor offers high impedance (opposition) to A.C. but very low impedance to D.C.
- In an electronic circuit the usual function of an inductor is to *block* A.C. signal *but to pass D.C. signal or voltage.*

Classification of inductors :

The inductors can be broadly *classified* as follows .

1. Fixed inductors.
2. Variable inductors.

The schematic symbols of fixed and variable inductors are shown in Fig. 2.11. (*a*) and (*b*) respectively.

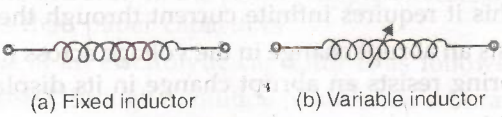


Fig 2.11. Inductors.

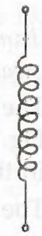
Filter chokes and Radio-frequency (RF) chokes :

— A *filter choke* [See Fig. 2.12 (*a*)] is an inductor used in the filter section of a D.C. power supply. Most of the power supplies use filter chokes having inductane ranging from about 1 H to 50 H, capable of carrying current upto 0.5 A.

— *Radio-frequency* chokes [See Fig. 2.12 (*b*)] are smaller in size as compared to filter chokes.

- **'Variable inductors (or coils)'** are used in *tuning circuits for radio frequencies.*
- A **transformer** is just similar in appearance to an inductor (filter choke). In electronic circuits, the transformers which are generally used are known as : *power transformers, output transformers* and *intermediate - frequency transformers.*

(a) Schematic symbol of filter choke.



(b) Schematic symbol of RF choke.

Fig. 2.12

2.1.3.3. Capacitors

A **capacitor** *is a device capable of storing electric charge.*

- It consists of two conducting surfaces (may be in the form of either circular or rectangular plates or of spherical or cylindrical shape) separated by an insulating material called a *dielectric.*
- *Capacitance is a measure of ability of a capacitor to store an electric charge.* It is the ratio of the charge (Q) that can be stored to the voltage applied (V) across the plates.

Mathematically, $C = \frac{Q}{V}$. Energy stored in a capacitor $= \frac{1}{2}CV^2$.

— The capacitance may be expressed in F (*Farads*) or μF or pF.

- This component (*i.e.,* capacitor) offers low impedance to *A.C.* but very high impedance (resistance) to D.C. The usual function of a capacitor is to *block D.C. voltage but pass the A.C. signal voltage,* by means of charging and discharging. These applications include *coupling, by passing* and *filtering for an A.C. signcl.*

Fig. 2.13 (*a*) and (*b*) shows the fixed and variable capacitors respectively.

(a) Fixed capacitor (b) Variable capacitor

Fig. 2.13 Fixed and variable capacitors.

- The *variable capacitors are mostly air-gang capacitors.*

Some important properties of capacitors :

1. The capacitor never dissipates energy, but only stores it.
2. A capacitor is sort of open circuit to D.C.
3. It the voltage across a capacitor is not changing with time, the current through it is zero.
4. It is not possible to change the voltage across a capacitor by a finite amount in zero time, for this it requires infinite current through the capacitor.
5. A capacitor resists an abrupt change in the voltage across it in a manner analogous to the way a spring resists an abrupt change in its displacement.

Types of capacitors :

The various types of capacitors are enumerated and discussed below :

1. Paper capacitors	2. Mica capacitors
3. Plastic film capacitors	4. Electrolytic capacitors
5. Ceramic capacitors	6. Air capacitors.

1. ***Paper capacitors :***

- Dry paper is good insulator and has high dielectric strength. It can withstand high potential difference without breaking down. It is commonly used in the manufacture of capacitors.
- There are *two* basic forms of capacitors :

— In one form it consists of two rolls of aluminium foils or tin foils sandwiching at tissue paper rolled by a machine so that the final shape is that of a small cylindrical tube. The entire cylinder is generally placed in a cardboard coated with wax or encased in a plastic paper. These capacitors are available in a wide range of capacitance values and voltage ratings. The physical size for 0.05 μF is typically 2.5 cm long with 1 cm diameter.

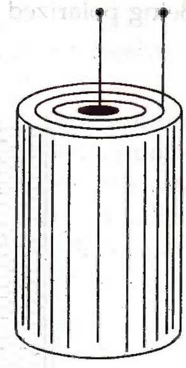


Fig. 2.14. Paper capacitor.

— In another form a "metallised" paper is used. A long strip of paper is metallised with aluminium by a special process. The strip is rolled to form a small cylinder. The capacitor is inserted into waxed cardboard case or plastic case.

- These capacitor should *not* be used in radio-frequency tuned circuits because they are not electrically stable enough.

2. ***Mica capacitors :***

Mica capacitors or *paraffined-capacitors* are widely used in radio circuit where *fixed volue capacitors are required.* Both these have metal foil sheets forming the coating and separated by a flat mica sheet or paraffin paper ; the dielectric paraffin paper capacitor of fairly large value is made by placing alternatively sheets or paraffin paper and the foil one above the other. Alternated tin foil sheets are connected together to form the two coatings.

- These capacitors are very small in size having 10 mm length and 3 mm thickness. These are often used for small capacitance values ranging from 50 to 500 pF, with voltage ratings ranging from 200 V to 1000 V.
- Silver mica capacitors are more stable electrically than foil-type capacitors and are used in high stability frequency determining circuits.

3. ***Plastic film capacitors :***

- Polyester is a thermoplastic material. It has better performance at high frequencies.
- The method of manufacture is same as in the case of paper capacitors, *i.e,* two strips of aluminium foils are separated by a thin film of polystyrene, then rolled and placed in aluminium container.

4. ***Electrolytic capacitors :*** Refer to Fig. 2.15

Electrolytic capacitors are used in the power supply circuits of *"Radio and TV circuits.* They have higher losses than paper capacitors.

The *working principle* of an electrolytic capacitor is as follows :

"When current is passed through a solution of aluminium borate or sodium phosphate with aluminium electrode, a layer of aluminium oxide forms at the positive electrode. This film acts as dielectric between the plates. As the film is very thin, a very high capacitance may be obtained. In the wet type oxide layer is reformed after being broken by a large potential difference applied. As this type has to be mounted vertically, the solution gets evaporised as such it has been replaced by dry type of electrolytic capacitor".

Being polarized they are *suitable only on D.C. supply.*

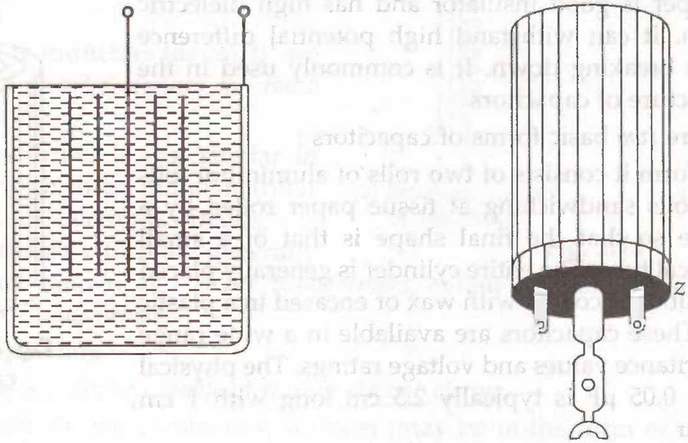


Fig. 2.15. Electrolytic capacitor.

- *Tantalum-electrolytic capacitor* consists of two tantalum foils with a tissue paper integrated with a non-corrosive electrolyte. The dielectric is pentaoxide layer which is electrochemically formed on the anode. The solid tantalum capacitors are available only in polarised form.

5. *Ceramic capacitors :* Refer to Fig. 2.16.

Deelectric constant of ceramic is high so that large capacitors can be obtained in a comparatively small space. It, however, suffers from the disadvantage of having higher losses than mica.

Ceramic capacitors are available in the following forms and shapes :

(*i*) Disc ceramics (*ii*) Tubular ceramics
(*iii*) Moulded ceramics (*iv*) Button ceramics.

The general construction of disc type consists of application of silver coatings on both sides of ceramic plates, in tubular type silver coating is applied on the inside and outside of hollow ceramic tube.

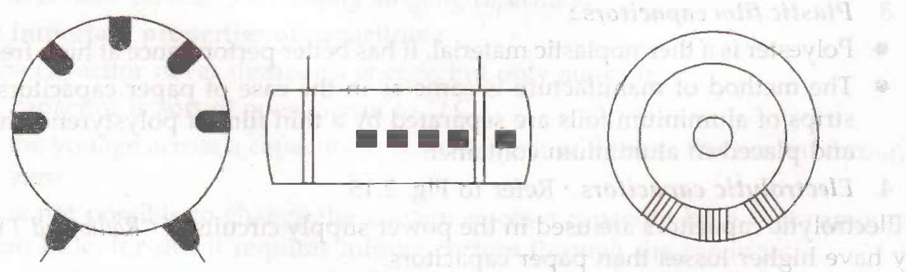


Fig. 2.16. Ceramic capacitors.

- Ceramic capacitors are used primarily as *coupling and bypass portions of radio frequency circuits* rather frequency determining elements. Specially designed ceramic capacitors are used in resonant circuits.

6. *Air capacitors :*

These capacitors (variable) are used in *radio receivers* for tuning the receiver to a particular transmitting station.

- Such a capacitor consists of a number of semicircular plates of sheet aluminium mounted together by metal rod and capable of moving in between a number of fixed aluminium semi-circular sheets. The *capacity increases* when the rotating sheets are moving into the fixed sheets. The set of rotating sheets is called the rotor while the set of fixed sheets is called the stator (Fig. 2.17). A circular dial and a pointer is used to read the value of the capacity for any position of the rotating plates.

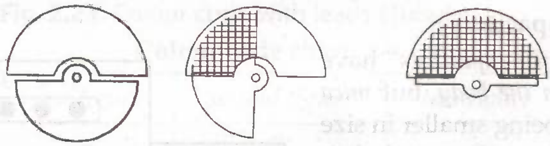


Fig. 2.17. Air capacitor—rotor and stator.

- Such condensers commonly used have a value of capacity varying almost from zero to 500 μF.

Variable capacitor :

*A capacitor whose capacity can be varied is called **'variable capacitor'**.* This is done by varying the thickness of the dielectric.

The various variable capacitors used in radio receiver are :

(*i*) ***Trimmers.*** Refer to Fig 2.18 Number of metal plates are inleaved with mica dielectric. The distance between the plates is controlled by a screw which is insulated from the plates stacked in a ceramic block.

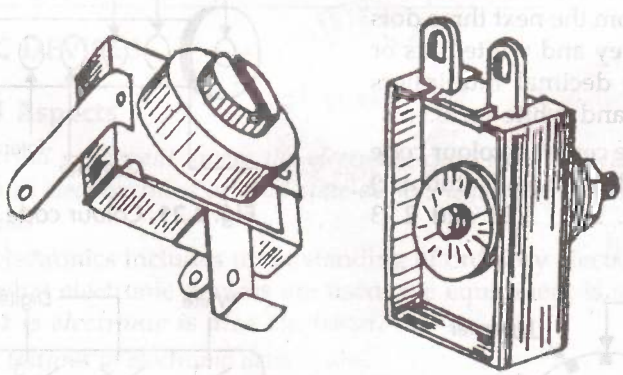


Fig. 2.18. Trimmers.

- They are available in the values of 30 pF to 70pF.

(*ii*) ***Padders.*** Refer to Fig. 2.19.

Padder is a mica capacitor (variable type). Its capacity is 600 pF.

These are continuously varying types. There are two sets of plates, fixed metal plates connected together form the stator set. Another set of movable plates form rotor. Rotor plates mesh with stator plates and can be moved with a shaft. Capacitance varies with the varying distance between the plates. Air is the dielectric. Usually two capacitors are ganged over a common shaft.

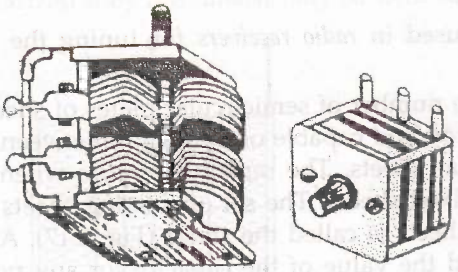


Fig. 2.19. Padders

Colour code of capacitors :

"Electrolyte and paper capacitors" have their values *printed on the body,* but *mica and tubular capacitors'* being smaller in size are colour coded. The *colour and their values are the same as in resistors.*

"Mica capacitors " (See Fig. 2.20) have six colour dots. Dots are marked from left to right in clockwise directon. Second and third dots indicate the digit and 4th dot is the multiplier. Dot 5 reads tolerance.

"Ceramic capacitors" (Fig. 2.21) have colour dots or bands. The wide colour band on the left specifies temperature coefficient. Capacitance value is read from left to right from the next three dots or colour strips. Grey and white dots or strips are used as decimal multipliers with grey for 0.01 and white for 0.1.

- Colour code ceramic, colour code mica and colour code with leads (tubular) are shown in Figs. 2.21, 2.22 and 2.23 respectively.

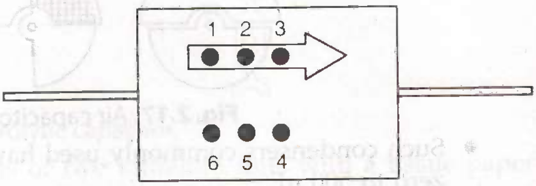


Fig. 2.20. Mica capacitor

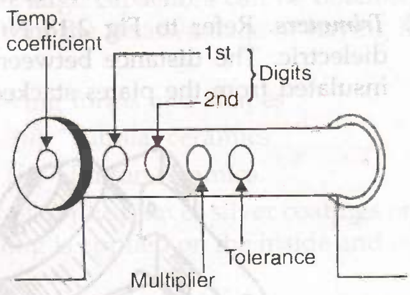


Fig. 2.21. Colour code ceramic.

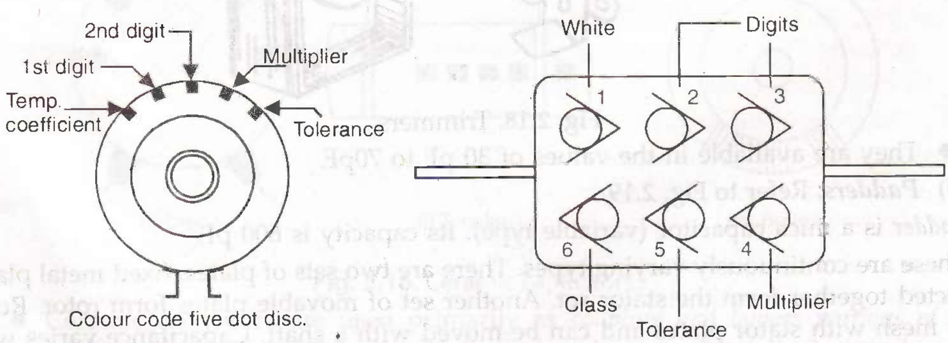


Fig. 2.22. Colour code mica.

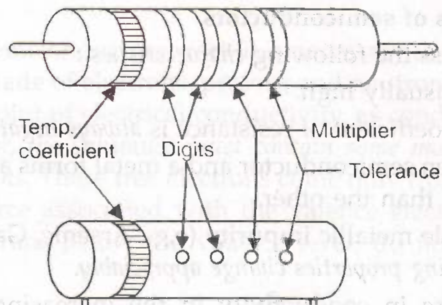


Fig. 2.23. Colour code with leads (Tubular).

Colour code chart

S.No.	*Colour*	*First figure*	*Second figure*	*Multiplier*	*Tolerance*
1.	Black	—	0	1	C < 10 pF
2.	Brown	1	1	10	± 1 pF
3.	Red	2	2	100	C > 10 pF
4.	Orange	3	3	1000	± 20%
5.	Yellow	4	4	10000	
6.	Green	5	5	—	
7.	Blue	6	6	—	
8.	Violet	7	7	—	
9.	Grey	8	8	0.01	
10.	White	9	9	0.1	

2.2. ELECTRONIC DEVICES

2.2.1. General Aspects

An ordinary electrical equipment enters the electronic class whenever its circuit includes electronic devices such as electron tubes or solid state devices which are formed by junctions of semiconducting materials.

Understanding electronics includes understanding of ordinary electrical devices and circuits. No matter what electronic devices are used, the equipment is still electrical. In other words, ***all that is electronic is also electrical.***

The characteristic features of electronic devices are

1. The electronic devices can rectify A.C. into D.C.
2. They can amplify input signals.
3. They can respond at speeds far beyond the speeds that one comes across in electrical and mechanical devices.
4. Some electronic devices are photosensitive. Some of these devices can produce radiations such as X-xays

2.2.2. Semiconductors

Semiconductors *are solid materials, either non-metallic elements or compounds, which allow electrons to pass through them so that they conduct electricity in much the same way as a metal.*

2.2.2.1 Characteristics of semiconductors

Semiconductors possess the following *characteristics* :

1. The resistivity is usually high.
2. The temperature coefficient of resistance is *always negative.*
3. The contact between semiconductor and a metal forms a layer which has a higher resistance in one direction than the other.
4. When some suitable metallic impurity (e.g., Arsenic, Gallium, etc.) is added to a semiconductor, its *conducting properties change appreciably.*
5. They exhibit a rise in conductivity in the increasing temperature, with the decreasing temperatures their conductivity falls off, and at low temperatures semiconductors become dielectrics.
6. They are usually metallic in appearance but (unlike metals) are generally hard and brittle.

Both the resistivity and the contact effect are as a rule very sensitive to small changes in physical conditions, and the *great importance* of semiconductors for a wide range of uses apart from rectification depend on the *sensitiveness.*

Examples of semiconducting materials :

Of all the elements in the periodic table, *eleven* are *semiconductors* which are listed below :

S. No.	*Element*	*Symbol*	*Group in the perodic table*	*Atomic No.*
1.	Boron	B	III	15
2.	Carbon	C	IV	6
3.	Silicon	Si	IV	14
4.	Germanium	Ge	IV	32
5.	Phosphorus	P	V	15
6.	Arsenic	As	V	33
7.	Antimony	Sb	V	51
8.	Sulphur	S	VI	
9.	Sellinum	Se	VI	
10.	Tellurium	Te	VI	
11.	Iodine	I	VIII	

Examples of semiconducting *compounds* are given below :

(*i*) Alloys : Mg_3Sb_2, ZnSb, Mg_2Sn, CdSb, AlSb, InSb, GeSb.

(*ii*) Oxide : ZnO, Fe_3O_4, Fe_2O_3, Cu_2O, CuO, BaO, CaO, NiO, Al_2O_3, TiO_2, UO_2, Cr_2O_3, WO_2, MoO_3.

(*iii*) Sulphides : Cu_2S, Ag_2S, PbS, ZnS, CdS, HgS, MoS_2.

(*iv*) Halides : AgI, CuI.

(*v*) Selenides and Tellurides.

PbS is *used in photo-conductive devices, BaO in oxide coated cathodes, caesium antimonide in photomultipliers, etc.*

Atomic structure :

To understand how semiconductors work it is necessary to study briefly the structure of matter. All atoms are made of electrons, protons and neutrons. Most solid materials are classed, from the stand point of electrical conductivity, as conductors, semiconductors or insulators. *To be conductor, the substance must contain some mobile electrons—one that can move freely between the atoms.* These free electrons come only from the valence (outer) orbit of the atom. Physical force associated with the valence electrons bind adjacent atoms together. The inner electrons below the valence level, do not normally enter into the conduction process.

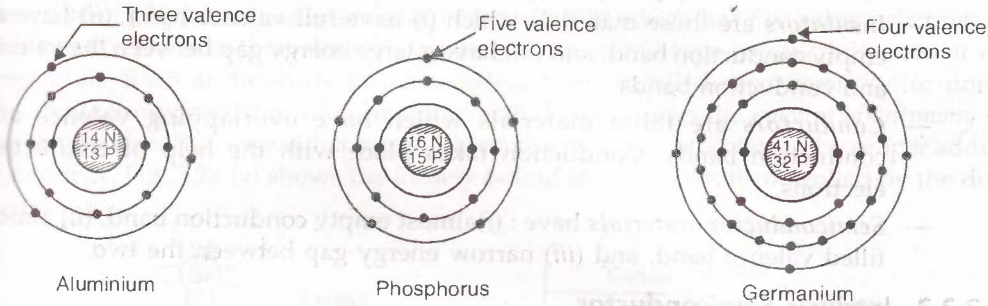


Fig. 2.24 **Fig. 2.25** **Fig. 2.26**

Conductivity depends on the number of electrons in the valence orbit. Electron diagrams for three typical elements, aluminium, phosphorus and germanium are shown in Figs. 2.24, 2.25, 2.26.

These elements can all be used in semiconductor manufacture. The *degree of conductivity* is determined as follows :

1. Atoms with *fewer than four valence electrons* are *good conductors.*
2. Atoms with *more than four valence electrons* are *poor conductors.*
3. Atoms with *four valence electrons* are *semiconductors.*

Fig. 2.24 shows aluminium which has *three valence electrons.* When there are less than four valence electrons they are loosely held so that at least one electron per atom is normally free ; hence aluminium is a good conductor. This ready availability of free electrons is also true of copper and most other metals.

Fig. 2.25 shows Phosphorus with *five valence electrons.* When there are more than four valence electrons, they are lightly held in orbit so that normally *none are free.* Hence phosphorus and similar elements are poor conductors (insulators).

Germanium (Fig. 2.26) has four *valence electrons.* This makes it neither a good conductor nor a good insulator, hence its name "semiconductor". Silicon also has four valence electrons and is a semiconductor.

Note. The energy level of an electron increases as its distance from the nucleus increases. Thus an electron in the second orbit possess more energy than electron in the first orbit ; electrons in the third orbit have higher energy than in the second orbit and so on. It follows, therefore, that electrons in the last orbit will possess very high energy. These high energy electrons are less bound to the nucleus and hence they are more mobile. *It is the mobility of last orbit electrons that they acquire the property of combining with other atoms.* Further it is due to this combining power of last orbit electrons of an atom that they are called *valence electrons.*

- Following points are worth noting:
  - *Conduction electrons* are those valence electrons which have gained enough energy to take part in conduction of electricity through a solid.
  - *Valence band* is the band of energy occupied by valence electrons. It is the highest occupied band and it may be completely or partially filled with electrons.
  - *Conduction band* is the higher energy band to the valence band. It is occupied by conduction electrons. It may be empty or partially filled. It is the lowest unfilled or unoccupied energy band.
  - *Insulators* are those materials which (*i*) have full valence band, (*ii*) have an empty conduction band, and (*iii*) have a large energy gap between the valence and conduction bands.
  - *Conductors* are those materials which have overlapping valence and conduction bands. Conduction takes place with the help of conduction electrons.
  - *Semiconductor materials* have : (*i*)almost empty conduction band, (*ii*) almost filled valence band, and (*iii*) narrow energy gap between the two.

2.2.3. Intrinsic Semiconductor

A *pure semiconductor is called "intrinsic semiconductor"*. Here no free electrons are available since all the covalent bonds are complete. A *pure semiconductor, therefore behaves as an insulator.* It exhibits a peculiar behaviour even at room temperature or with rise in temperature. The *resistance of a semiconductor decreases with increase in temperature.*

When an electric field is applied to an intrinsic semiconductor at a temperature greater than 0°K, conduction electrons move to the anode and, the holes (when an electron is liberated into the conduction band a positively charged hole is created in valence band) move to cathode. *Hence semiconductor current consists of movement of electrons in opposite direction.*

Fig. 2.27 shows the energy diagram for intrinsic (pure) semiconductor at absolute zero.

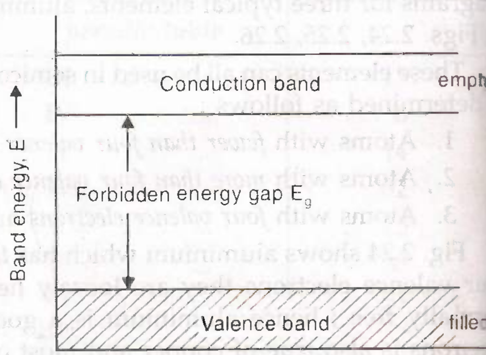


Fig. 2.27. Energy diagram for intrinsic (pure) semiconductor at absolute zero.

2.2.4. Extrinsic Semiconductor

In a pure semiconductor, which behaves like an insulator *under ordinary conditions*, if a small amount of certain *metallic impurity* is *added* it attains *current conducting properties*. The impure semiconductor is then called *"impurity semiconductor" or "extrinsic semiconductor". The process of adding impurity (extremely in small amounts, about 1 part in 10* *to a semiconductor to make it extrinsic (impurity) semiconductor is called* **Doping.**

Generally following doping agents are used :

(*i*) *Pentavalent atom* having five valence electrons (arsenic, antimony, phosphorus) ... called *donor atoms.*

(*ii*) *Trivalent atoms* having three valence electrons (gallium aluminium, boron) ... called *acceptor atoms.*

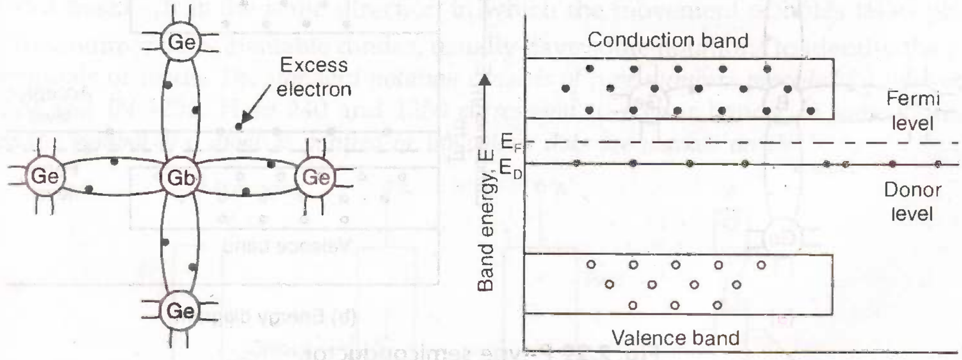
With the addition of suitable impurities to semiconductor, two type of semiconductors are :

(*i*) N-type semiconductor.

(*ii*) P-type semiconductor.

N-type semiconductor :

The presence of *even a minute quantity of impurity*, can produce N-type semiconductor. If the impurity atoms has *one valence electron more* than the semiconductor atom which it has substituted, this *extra electron* will be loosely bound to the atom. For example, an atom of *Germanium* possesses *four valence electrons* ; when it is replaced in the crystal lattice of the substance by an impurity atom of antimony (Sb) which has *five valence electrons*, the fifth valence electron (free electron) produces extrinsic N-type conductivity *even at room temperature.* Such an impurity into a semiconductor is called *donor impurity* (or donor). The conducting properties of germanium will depend upon the *amount of antimony* (*i.e.*, impurity) *added*. This means that controlled conductivity can be obtained by proper addition of impurity. Fig. 2.28 (*a*) shows the loosely bound excess electron controlled by the donor atom.



(a) (b) Energy diagram

Fig. 2.28. N-type semiconductor

- It may be noted that by giving away its one electron, the *donor atom becomes possitvely charged ion.* But it cannot take part in conduction because it is firmly fixed or tied into the crystal lattice. In addition to the electrons and holes *intrinsically available in germanium*, the addition of antimony greatly increases the number of conduction electrons. Hence, *concentration of electrons in the conduction band is increased and exceeds the concentration of the holes in the valence band.* Consequently, *Fermi level shifts upwards towards the bottom of the conduction band* as shown in Fig. 2.28 (*b*). [Since the number of electrons as compared to the number of holes increases with temperature, the *position of Fermi level also changes considerably with temperature*].
- It is worth noting that even though N-type semiconductor has excess of electrons, still it is *electrically neutral.* It is so because by addition of donor impurity, number of electrons available for conduction purposes becomes more than the number of holes available intrinsically. But the total charge of the semiconductor does not change because the donor impurity brings in as much negative charge (by way of electrons) as positive charge (by way of protons).

Note. In terms of energy levels, the fifth antimony electron has as energy level (called donor level) just below the conduction band. Usually, the donor level is 0.01 eV below conduction band for germanium and 0.054 eV for silicon.

P-type semiconductor:

- P-type extrinsic semiconductor can be produced if the impurity atom has *one valence electron less* than the semiconductor atom that it has replaced in the crystal lattice. This impurity atom cannot fill all the *interatomic* bonds, and the free bond can accept an electron from the neighbouring bond ; leaving behind a vacancy of *hole*. Such an impurity is called an *acceptor impurity* (*or acceptor*). Fig. 2.29 (*a*) shows structure of P-type semiconductor (Germanium and Boron).
- In this type of semiconductor, conduction is by means of holes in the valence band. Accordingly, *holes form the majority carriers whereas electrons constitute minority carriers.* The process of conduction is called *deficit conduction.*
- Since the concentration of holes in the valence band is more than the concentration of electrons in the conduction band, Fermi level shifts nearer to the valence band [Fig. 2.29 (*b*)]. The acceptor level lies immediately above the Fermi level. *Conduction is by means of hole movement at the top of valence band, the acceptor level readily accepting electrons from the valence band.*

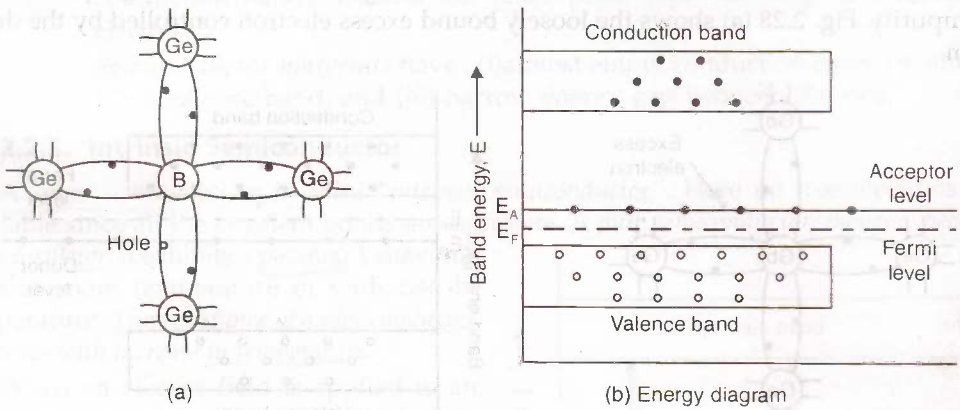


Fig. 2.29 P-type semiconductor.

It may be noted again that even though P-type semiconductor has excess of holes for conduction purposes, as a whole it is *electrically neutral* for the same reasons as dicusse earlier.

2.2.5. P-N Junction Diode

In an **N-type material** (Fig. 2.30) *the electron is called the majority carrier and the hole as the minority carrier.*

In a **P-type material** (Fig. 2.31) *the hole is the majority crrier and the electron is the minority carrier.* The N- and P-type materials represent the *basic building blocks of semiconductor devices.*

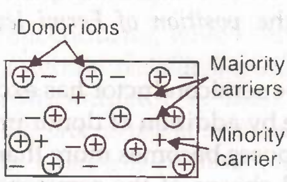


Fig. 2.30 N-type material.

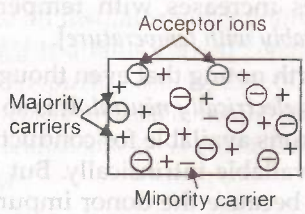


Fig. 2.31 P-type material.

The semiconductor diode is simply bringing these materials together (constructed from the same base-Ge or Si). At instant the two materials are *"joined"* the electrons and holes in the region of the junction will combine resulting in a lack of carriers in the region near the junction. *This region of unconvered positive and negative ions is called the* ***depletion region*** *due to the depletion of carriers in this region.*

Construction and types of P-N junction diodes :

The most extensively used elements in the manufacture of junction diodes are *germanium and silicon* (although some other materials are also assuming importance in recent years).

A *P-N* junction diode (known as a semiconductor or crystal diode) consists of a *P-N* junction, formed either in germanium or silicon crystal. The diode has two terminals namely *anode and cathode.* The anode refers to the *P*-type region and cathode refers the *n*-type region as shown in Fig. 2.32 (*a*)

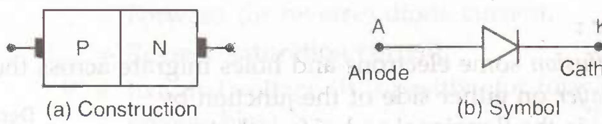
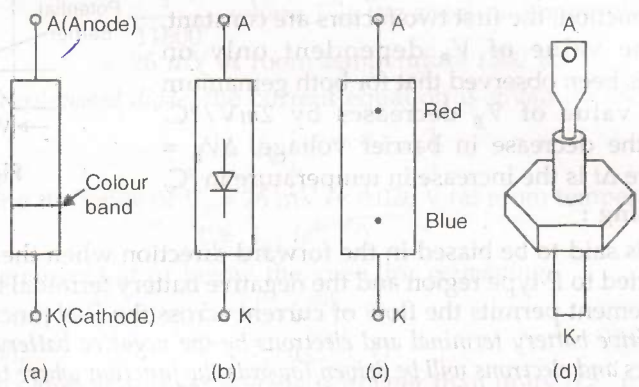


Fig. 2.32 *P-N* junction diode.

The arrow head, shown in the circuit symbol, points the direction of current flow, when it is "forward biased" (It is the same direction in which the movement of holes takes place).

The commercially available diodes, usually have some notations to identify the *P* and *N* terminals or leads. *The standard notation consists of type numbers preceded by IN,* such as *IN* 240 and *IN* 1250. Here 240 and 1250 *correspond* to colour bands. *In some diodes, the schematic symbol of a diode is painted or the colour dots are marked on the body.*



(*a*), (*b*) = Low current diodes ; (*c*) = Medium current diode ;
(*d*) = High current or power diode.

Fig. 2.33 Low, medium and high current diodes.

Fig. 2.33 shows low, medium and high current diodes.

— *Refer to Fig. 2.33 (a).* The diode shown has a colour band located near one of the ends. The end, which is near the colour band, is identified as cathode, and other end, obviously, is the anode (*A*).

— *Refer to Fig 2.33 (b).* The diode has a schematic symbol actually painted at its cathode (*K*) and the other end as anode.

The diodes of Fig. 2.33 (*a*) and (*b*) *can pass a forward current of 100 mA and are known as* ***low current diodes.***

— *Refer to Fig. 2.33 (c).* The diode has colour dots marked on its body. The end lying near the blue dot is a cathode, while the other end is anode. Sometimes this diode is *shown bigger in size* than that of diodes shown in Fig. 2.33 (*a*) and (*b*). The diodes of this size can *pass a forward current of 500 mA and are known as* ***medium current diodes.***

— *Refer to Fig. 2.33 (d).* It shows a diode, which *can pass a forward current of several amperes.* Therefore it known as a ***power diode*** or a ***high current diode.***

- *The outstanding property of P-N junction / crystal diode to conduct current in one direction only permits it to be used as a* **rectifier.**

Potential barrier and biasing :

A *P-N* junction diode which consists of *P-* and *N*-type semiconductors formed together to make a *P-N* junction is shown in the Fig 2.34. The *place dividing the two zones* is known as a *"junction".*

Potential barrier :

As a result of *diffusion* some electrons and holes migrate across the junction thereby forming a *depletion layer* on either side of the junction by neutralisation of holes in the *P*-regional and of free electrons in the *N*-region. *This diffusion of holes and electrons across the junction continues till* ***potential barrier*** *is developed in the depletion layer which then prevents further diffusion.* By the application of an external voltage this potential barrier is either *increased* or *decreased.*

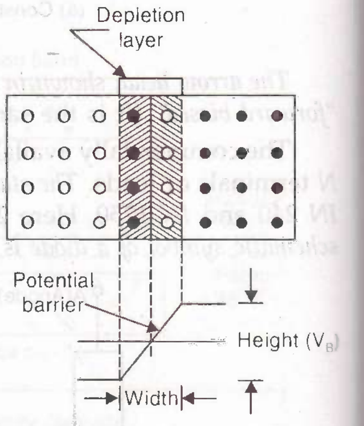


Fig. 2.34

The barrier voltage of a *P-N* junction depends upon three factors namely *density, electonic charge* and *temperature.* For a given *P-N* junction, the first two factors are constant, thus making the value of V_B dependent only on temperature. It has been observed that for both gemanium and silicon the value of V_B decreases by $2mV/°C$. Mathematically, the decrease in barrier voltage, $\Delta V_B = -0.002 \times \Delta t$, where Δt is the increase in temperature in °C.

Forward biasing :

The junction is said to be biased in the forward direction when then positive battery terminal is connected to *P*-type region and the negative battery terminal to the *N*-type (Fig. 2.35). This arrangement permits the flow of current across the *P-N* junction. The *holes are repelled by the positive battery terminal and electrons by the negative battery terminal with the result that both holes and electrons will be driven towards the junction where they will recombine.* Hence as long as the battery voltage is applied large current flows. In other words, the *forward bias lowers the potential barrier across the depletion layer* thereby *allowing more current to flow* across the junction.

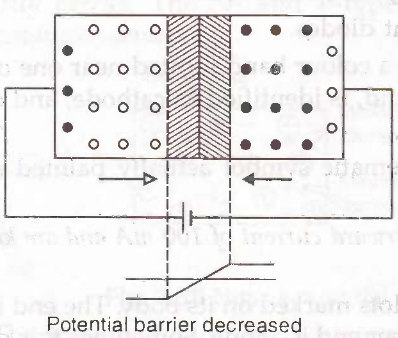


Fig. 2.35 Forward biasing.

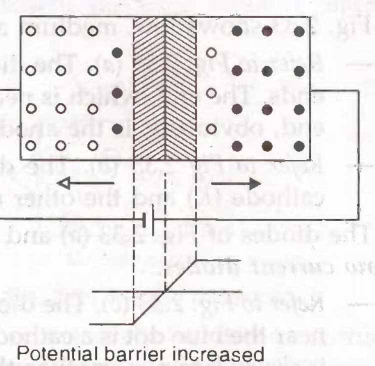


Fig. 2.36 Reverse biasing

Reverse biasing (Zener diode) :

The junction is said to be reversed biased when battery connection to the battery are reversed as shown in Fig. 2.36. In this arrangement holes are attracted by the negative battery terminal and electrons by the positive battery terminal so that both holes and electrons move *away* from the junction. Since there is no recombintion of electron-hole pairs, diode current is negligible and the junction has high resistance. Reverse biasing *increases the potential barrier* at the junction, thereby *allowing very little current to flow through* the junction.

Diode current equation :

The mathematical equation, which describes the forward and reverse characteristics of a semiconductor diode is called the ***diode current equation.***

Let I = Forward (or reverse) diode current,

I_{RS} = Reverse saturation current,

V = External voltage (It is positive for forward bias and negative for reverse bias),

η = A constant

= 1 for germanium diodes, 2 for silicon diodes for relative *low value of diode current* (*i.e.*, at or below the knee of the curve)

= 1 for germanium and silicon for *higher levels of diode current* (*i.e.*, in the rapidly increasing section of the curve), and

V_T = Volt-equivalent of temperature. Its value is given by the relation,

$\frac{T}{11600}$, where T is the absolute temperature

= 26 mV at room temperature (300 K).

For a *forward-biased diode*, the current equation is given by the relation,

$$I = I_{RS}\left[e^{V/(\eta\times V_T)} - 1\right] \quad ...(i)$$

Substituting the value of V_T = 26 mV or 0.026 V (at room temperature) in eq. (*i*), we get

$$I = I_{RS}\,(e^{40V/\eta})$$

$\therefore$ Diode current at or below the knee, for *germanium*,

$$I = I_{RS}\,(e^{40V} - 1) \qquad (\because \ \eta = 1)$$

and, for *silicon*,

$$I = I_{RS}\,(e^{20V} - 1) \qquad (\because \ \eta = 2)$$

When the value of applied voltage is greater than unity (*i.e.*, for the diode current in the rapidly increasing section of curve), the equation of diode current for germanium or and silicon,

$$I = I_{RS} \cdot e^{20V} \qquad (\because \ \eta = 2)$$

The current equation for a reverse biased diode may be obtained from eqn. (*i*) by *changing the sign of the applied voltage* (V). Thus the *diode current for reverse bias,*

$$I = I_{RS}\,[e^{-V/(\eta\times V_T)} - 1]$$

When $V >> V_T$, then the term $e^{-V/(\eta\times V_T)} << 1$. Therefore $I = I_{RS}$. Thus the *diode current under reverse bias is equal to the reverse saturation current as long as the external voltage is below its breakdown value.*

Example 2.1. *The current flowing in a certain P-N junction diode at room temperature is 1.8×10^{-7} A, when large reverse voltage is applied. Calculate the current flowing, when 0.12 V forward bias is applied at room temperature.*

Solution. *Given :* $I_{RS} = 1.8 \times 10^{-7}$ A; $V_F = 0.12$ V

The current flowing through the diode under forward bias is given by,

$$I = I_{RS}\,(e^{40V_F} - 1)$$

or, $$I = 1.8 \times 10^{-7}\,(e^{40 \times 0.12} - 1) = 21.69 \times 10^{-6}\ \text{A}$$

$$= \mathbf{21.69\ \mu A.\ (Ans.)}$$

Example 2.2. *Determine the germanium P-N junction diode current for the forward bias voltage of 0.2 V at room temperature 24°C with reverse saturation current equal to 1.1 mA. Take $\eta = 1$.*

Solution. *Given :* $V_F = 0.2$ V; $T = 24 + 273 = 297$ K;

$I_{RS} = 1.1$ mA $= 1.1 \times 10^{-3}$ A; $\eta = 1$

We know that, $$V_T = \frac{T}{11600} = \frac{297}{11600} = 0.0256\ \text{V}\ (i.e.,\ 25.6\ \text{mV})$$

$\therefore$ The diode current, $$I = I_{RS}\,[e^{V_F/(\eta \times V_T)} - 1]$$

$$= 1.1 \times 10^{-3}\,[e^{0.2/(1 \times 0.0256)} - 1] = \mathbf{2.717\ A.}\quad \textbf{(Ans.)}$$

Static and dynamic resistance of a diode : Refer to Fig. 2.37.

Static forward resistances (R_F). A diode has a definite value of resistance when forward biased. It is given by the *ratio of the D.C. voltage across the diode to D.C. current flowing through it.*

Mathematically, $$R_F = \frac{V_F}{I_F}.$$

R_F may be obtained *graphically* from the diode forward characteristics as shown in Fig 2.37. From the operating point P, the static forward resistance,

$$R_F = \frac{0.8}{16 \times 10^{-3}} = 50\ \Omega.$$

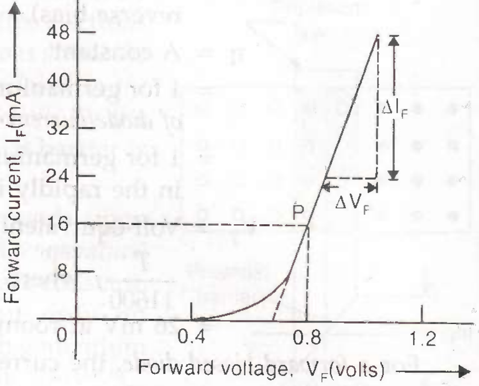


Fig. 2.37 Static and dynamic forward resistances of a diode from the characteristic curve.

Dynamic or A.C. resistance. In practice we don't use static forward resistance, instead, we use the dynamic or A.C. resistance. The A.C. resistance of a diode, at a particular D.C. voltage, is *equal to the reciprocal of the slope of the characteristic at that point; i.e.,* the A.C. resistance,

$$r_{A.C.} = \frac{1}{\Delta I_F / \Delta V_F} = \frac{\Delta V_F}{\Delta I_F} = \frac{\text{Change in voltage}}{\text{Resulting change in current}}.$$

Owing to the non-linear shape of the forward characteristic, the value of A.C. resistance of a diode is in the range of 1 to 25 Ω. Usually it is *smaller than D.C. resistance of a diode.*

Reverse resistance. When a diode is *reverse biased,* besides the forward resistance, it also possesses another resistance known as *reverse resistance.* It can be either D.C. or A.C. depending upon whether the reverse bias is direct or alternating voltage. Ideally, the reverse resistance of a diode is infinite. However, in actual practice, the reverse resistance is never infinite. It is *due to the existence of leakage current* in a reverse biased diode.

Its value for germanium and silicon diodes is of *several megaohms.*

The A.C. resistance of a diode may also be determined from the following *two resistances:*

1. Bulk resistance.
2. Junction resistance.

1. Bulk resistance r_B. *The resistance of P- and N-semiconductor materials of which the diode is made of,* is known as *"bulk or body resistance"*. It also includes the resistance introduced by the connection between the semiconductor material and external metallic conductor, contact resistance.

Mathematically $\quad r_B = r_P + r_N$

where, $\quad r_P$ = Ohmic resistance of *P*-type semiconductor, and

$\quad r_N$ = Ohmic resistance of *N*-type semiconductor.

The typical values of bulk resistance may be :

For high power devices..0.1 Ω

Low-power general purpose diodes..............................2 Ω

The total voltage drop across the diode,

$$V_F = V_B + I_F \cdot r_B \quad \text{...(2.1)}$$

$$= 0.6 + I_F \cdot r_B \quad \text{...For silicon diode ...[2.1 (a)]}$$

$$= 0.2 + I_F \cdot r_B \quad \text{...For germanium diode ...[2.1 (b)]}$$

2. Junction resistance r_J. The value of junction resistance for a forward-biased *P-N* junction depends upon the value of forward D.C. current and is given by relation,

$$r_J = \frac{26}{I_F} \quad \text{...(2.2)}$$

where, $\quad I_F$ = Forward current in 'milliampers'

Mathematically, the A.C. resistance,

$$r_{A.C.} = r_J + r_B \quad \text{...(2.3)}$$

Example 2.3. *A silicon diode has a bulk resistance of 2.2 Ω and a forward current of 11 mA. What is the actual value of V_F for the device?*

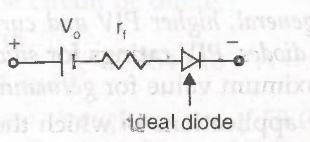
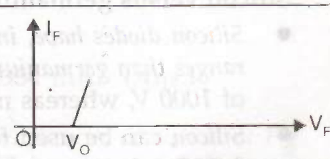
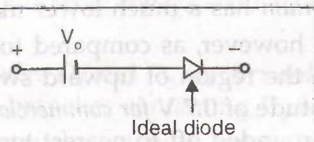
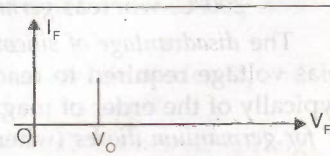
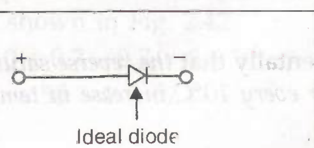
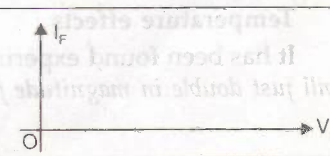
Solution. *Given :* $\quad r_B = 2.2\ \Omega;\ I_F = 11\ \text{mA} = 11 \times 10^{-3} = 0.011\ \text{A}$

Now, $\quad V_F = 0.6 + I_F \cdot r_B \quad$ ...[Eqn. 2.1 (*a*)]

$\quad = 0.6 + 0.011 \times 2.2 =$ **0.6242 Ω.** **(Ans.)**

Equivalent circuits of P-N junction diode :

The equivalent circuits of various models of *P-N* junction diode in a tabular form is given below :

S. No.	*Type*	*Model*	*Characteristic*
1.	*Approximate model*		
2.	*Simplified model*		
3.	*Ideal model*		

- An **ideal diode** is *a device, which conducts with zero resistance when forward biased and appears as an infinite resistance when reverse biased.* As matter of fact, an ideal diode cannot be manufactured in actual practice. It is only a theoretical approximation of a real diode. However, in a *well designed electric circuit, a real diode behaves almost like an ideal diode because the forward voltage across the diode is small as compared to the input and output stages.*

Power and current ratings of a diode :

The power dissipation for a *forward biased diode is given by,*

$$P_{DF} = V_F \times I_F \quad \text{...(2.4)}$$

where,
P_{DF} = Power dissipated by the diode,
V_F = Forward voltage drop, and
I_F = Forwrd current.

Similarly, power dissipation for a *reverse biased diode,*

$$P_{DR} = V_R \times I_R$$

where,
V_R = Reverse voltage drop, and
I_R = Reverse current.

The maximum value of power, which a diode can dissipate without failure, is called its **rating.** Thus the power dissipation should not exceed power rating in any case, otherwise the diode will get destroyed.

The diode manufacturers more oftenly list the maximum current, which a device can handle, (called *current rating*), rather than power rating. It is because of the fact that it is *easy to measure current rating than power rating.*

Applications of a diode :

An important characteristic of the *P-N* junction diode that it conducts well in forward direction and poorly in reverse direction has made it useful in several applications listed below :

1. As zener diodes in voltage stabilising circuits.
2. As rectifiers or power diodes in D.C. power supplies.
3. As a switch in logic circuits in computers.
4. As signal diodes in communication circuits.
5. As varactor diodes in radio and T.V. receivers.

Silicon versus germanium :

- *Silicon diodes have, in general, higher PIV and current rating and wider temperature ranges than germanium diodes. PIV* ratings for *silicon* can be in the neighbourhood of *1000 V,* whereas maximum value for *germanium* is closer to *400 V.*
- *Silicon* can be used for applications in which the temperature may rise to about *200°C,* whereas *germanium* has a much lower maximum rating (*100°C*).

The *disadvantage of silicon,* however, as compared to germanium is higher forward-bias voltage required to reach the region of upward swing of characteristic curve. It is typically of the order of magnitude of 0.7 V *for commercially available silicon diodes, and 0.3 V for germanium diodes* (when rounded off to nearest tenths).

Temperature effects :

It has been found experimentally that *the reverse saturation current I_{RS} of a silicon diode will just double in magnitude for every 10°C increase in temperature.*

Typical values of I_{RS} for silicon are *much lower* than that of germanium for similar power and current levels—a very important reason that silicon devices enjoy a significantly higher level of development and utilisation in design. Fundamentally, the *open-circuit equivalent in the reversebias region is better realised at any temperature with silicon than with germanium.*

Example 2.4. *Determine the resistance levels for the diode characteristics of Fig. 2.38 at.*

(i) $I_D = 2\ mA$

(ii) $I_D = 20\ mA$

(iii) $V_D = -10\ V$

where V_D and I_D are bias voltages and diode currents respectively.

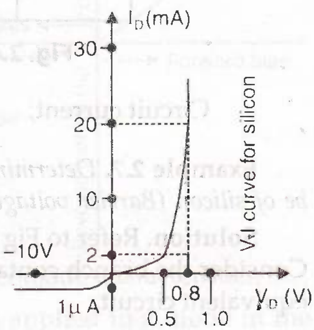


Fig. 2.38

Solution. (i) $R_1 = \dfrac{0.5}{2\times10^{-3}} = \dfrac{5}{2}\times10^2 = \mathbf{250\ \Omega}$

(ii) $R_2 = \dfrac{0.8}{20\times10^{-3}} = \dfrac{8}{2}\times10 = \mathbf{40\ \Omega}$

(iii) $R_3 = \dfrac{1.0}{1\times10^{-6}} = 1.0\times10^6 = \mathbf{10\ M\Omega.}$

Example 2.5. *Determine the current flowing through the silicon diode (Barrier voltage = 0.7 V) shown in Fig. 2.39. Assume forward resistance to be zero.*

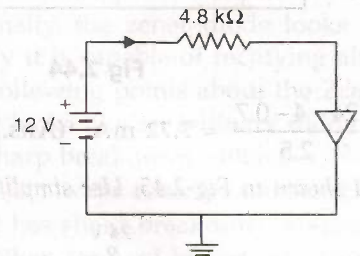


Fig. 2.39

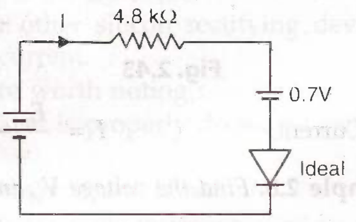


Fig. 2.40

Solution. The *barrier potential acts in opposite direction to the supply voltage.* A simplified circuit is shown in Fig 2.40.

∴ Current flowing through the circuit or diode,

$$I = \frac{12-0.7}{4.8\times10^3} = 2.354\times10^{-3}\ A = \mathbf{2.354\ mA.}\quad \textbf{(Ans.)}$$

Example 2.6. *Calculate the current through resistor of 50 Ω in the circuit shown in Fig. 2.41. Assume the diodes to be of silicon (Barrier voltage = 0.7 V) and forward resistance of each diode is 1 Ω.*

Solution. Refer to Fig. 2.41. Diodes D_1 and D_3 are *forward biased* while diodes D_2 and D_4 are *reverse biased.* Consider the branches containing D_2 and D_4 as *'open'*.

Replacing D_1 and D_3 by their equivalent circuits and making the branches containing D_2 and D_4 open we get the circuit shown in Fig. 2.42.

Net circuit voltage = 10 – 0.7 – 0.7 = 8.6 V

Total circuit resistance = 1 + 50 + 1 = 52 Ω

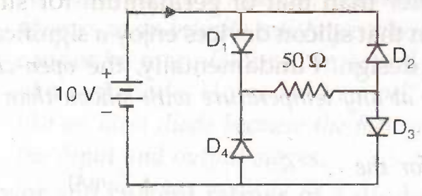


Fig. 2.41

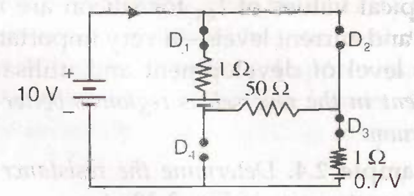


Fig. 2.42

$\therefore$ Circuit current; $I = \dfrac{8.6}{52} = 0.165$ A or **165 mA. (Ans.)**

Example 2.7. *Determine the current in the circuit shown in Fig. 2.43. Assume the diodes to be of silicon (Barrier voltage = 0.7 V) and forward resistance of the diodes to be zero.*

Solution. Refer to Fig 2.43. Diode D_1 is forward biased and diode D_2 is reverse biased. Consider the branch containing diode D_2 as open and D_1 can be replaced by its simplified equivalent circuit.

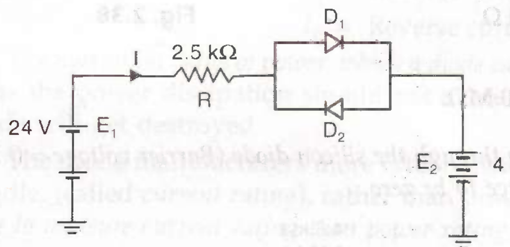


Fig. 2.43

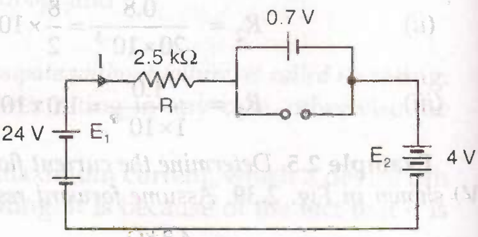


Fig 2.44

$\therefore$ Current, $I = \dfrac{E_1 - E_2 - 0.7}{R} = \dfrac{24 - 4 - 0.7}{2.5}$ = **7.72 mA. (Ans.)**

Example 2.8. *Find the voltage V_A in the circuit shown in Fig 2.45. Use simplified model.*

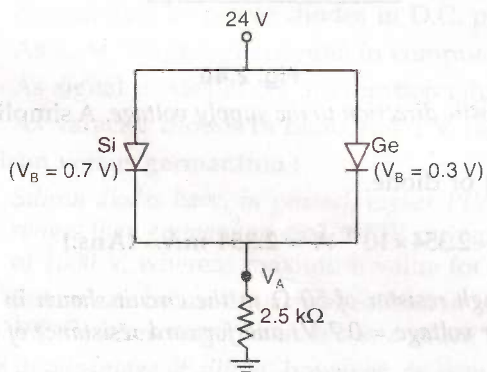


Fig. 2.45

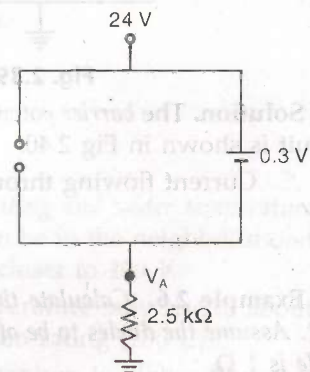


Fig. 2.46

Solution. Refer to Figs. 2.45 and 2.46. It appears that when voltage is switched on, both the sides will turn on, but it does not happpen. When voltage is applied, germanium diode (Barrier voltage = 0.3 V) will turn on first and a level of 0.3 V is maintained across the parallel circuit. The silicon diode never gets the opportunity to have 0.7 V across it and therefore remain in open state (Fig. 2.46)

$\therefore$ $V_A = 24 - 0.3 = $ **23.7 V (Ans.)**

2.2.6 Zener Diode

A properly doped P-N junction crystal diode which has a sharp breakdown voltage is known as **Zener diode.**

The voltage-regulator diode is commonly called a 'Zener' diode. It is a *voltage limiting diode* that has some applications in common with the older voltage-regulator gas tubes but serves a much wider field of application, because the devices cover a wide spectrum of voltages and power levels.

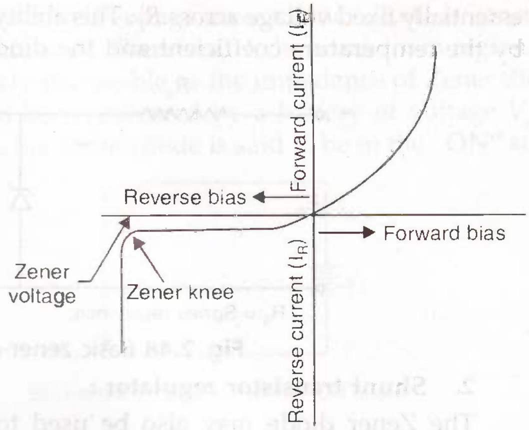


Fig. 2.47 Zener diode characteristics.

Performance/Operation :

The electrical performance of a zener diode is based on the *avalanche characteristics* of the *P-N* junction. When a source of voltage is applied to a diode in the *reversse direction* (negative to anode), a reverse current I_R is observed (see Fig 2.47). As the reverse potential is increased beyond the *"Zener knee"* avalanche breakdown becomes well developed at zener voltage V_Z. At voltage V_Z, the high counter resistance drops to a low value and the junction current increases rapidly. The current must of necessity be limited by an external resistance, since the voltage V_Z developed across the zener diode remains essentially constant. *Avalanche breakdown of the operating zener diode is not destructive as long as the rated power dissipation of the junction is not exceeded.*

Externally, the zener diode looks much like other silicon rectifying devices, and electrically it is capable of rectifying alternating current.

The following points about the *Zener diode* are worth noting :

(*i*) It looks like an ordinary diode except that it is properly doped so as to have a sharp breakdown voltage.

(*ii*) It is always reverse connected *i.e.,* it is *always reverse biased.*

(*iii*) It has sharp breakdown voltage, called Zener voltage V_Z.

(*iv*) When forward biased, its characteristics are just those of ordinary diode.

(*v*) It is not immediately burnt just because it has entered the breakdown region (The current is limited only by both external resistance and power dissipation of Zener diode).

- The location of Zener region can be controlled by varying the doping levels. An *increase in doping, producing an increase in the number of added impurities, will decrease the Zener potential.*
- Zener diodes are available having Zener potentials of 1.8 to 200 V with power ratings from $\frac{1}{4}$ to 50 W. *Because of its higher temperature and current capability, silicon is usually preferred in the manufacture of Zener diodes.*

Applications of zener diode :

Zener diode serves in the following variety of *applications :*

1. **Voltage reference or regulator element :**

The primary use of a zener diode is as a *voltage reference or regulator element.* Fig. 2.48 shows the fundamental circuit for the Zener diode employed as a shunt regulator. In the circuit, diode element and load R_L draw current through the series resistance R_S. If E_{in} increases, the current through the Zener element will increase and thus maintain an essentially fixed voltage across R_L. This ability to maintain the desired voltage is determined by the temperature coefficient and the diode impedance of the zener device.

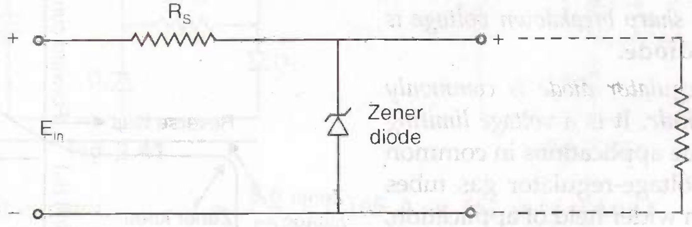


Fig. 2.48 Basic zener-diode regulator circuit.

2. **Shunt transistor regulator :**

The Zener diode may also be used to control the reference voltage of a transistor regulated power supply. An example of this in a shunt transistor regulator is shown in Fig. 2.49, where Zener element is used to control the operating point of the transistor. The advantages of this circuit over that shown in Fig. 2.48 are *increased power handling capability and a regulating factor improved by utilizing the current gain of the transistor.*

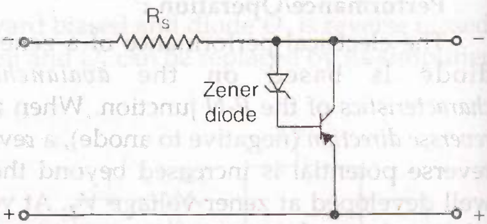


Fig. 2.49. Shunt transistor regulator.

3. **Audio or r-f application :**

The Zener diode also finds use in audio or *r-f* (radio frequency) applications where a source of stable reference voltage is required, as in bias supplies. Frequently, *Zener diodes are connected in series package,* with, for example, one junction operating in the reverse within a single direction and possessing a positive temperature V_Z coefficient; the remaining diodes are connected to operate in the forward direction and exhibit negative temperature V_Z coefficient characteristics. The net result is close neutralization of V_Z drift versus temperature change; such reference units are frequently used to *replace standard voltage cells.*

4. **Computer circuits :**

Zener diodes also find use in *computer circuits designed for switching about the avalanche voltage of the diode.* Design of the Zener diode permits it to absorb *overload surges and thereby serves the function of protecting delicate circuitry from overvoltage.*

— The usual *voltage specifications* V_Z of Zener diodes are 3.3 to 200 V with ± 1, 2, 5, 10 or 20% tolerances.

— Typical *power dissipation ratings* are 500 mW, 1, 10 and 50 W.

— The temperature coefficient range on V_Z is as low as 0.001% °C.

Equivalent circuit of zener diode :

The complete equivalent circuit of the Zener diode in the Zener region includes a small dynamic resistance and D.C. battery equal to the Zener potential, as shown in Fig. 2.50.

"ON" state. When reverse voltage across a Zener diode is equal to or more than breakdown voltage V_Z, the current increases very sharply. In

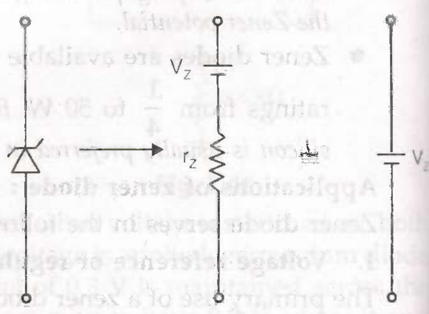
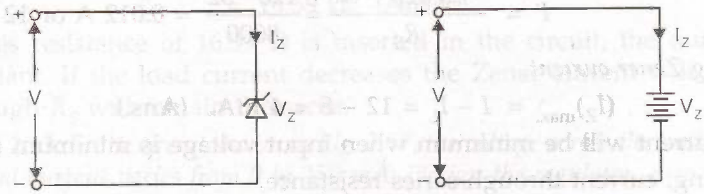


Fig. 2.50. Zener equivalent circuit : (*a*) Complete; (*b*) Approximately.

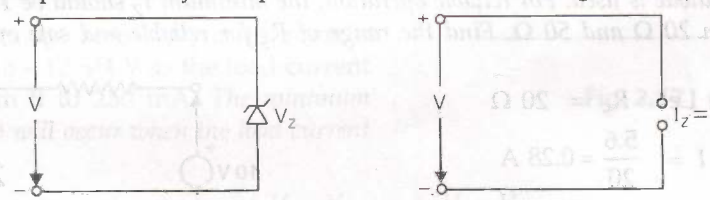
this region curve is almost vertical; it means that voltage across Zener diode is constant at V_Z even though the current through it changes. Therefore, in the breakdown region, an ideal Zener diode (this assumption is fairly reasonable as the impedance of Zener diode is *quite small in the breakdown region*) can be represented by a battery of voltage V_Z as shown in 2.51 (*b*). Under such conditions, the Zener diode is said to be in the **"ON" state.**



$V \geq V_Z$ (a)   Equivalent circuit of zener for "ON" state. (b)

Fig. 2.51

"OFF" state. When the reverse voltage across the Zener diode is less than V_Z but greater than 0 V, the Zener diode is in the "OFF" stage. Under such conditions, the Zener diode can be represented by an open circuit as shown in Fig. 2.52 (*b*).



$V_Z > V > 0$ (a)   Equivalent circuit of zener for "OFF" state. (b)

Fig. 2.52

Example 2.9. *Determine the current flowing through the Zener diode for the circuit shown in Fig. 2.53, if R_L = 4000 Ω, input voltage is 50 volts, R_S = 1800 Ω and output voltage is 32 volts.*

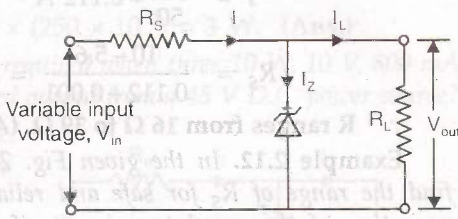


Fig. 2.53

Solution. Input voltage, V_{in} = 50 V

Output voltage, V_{out} = 32 V

Voltage drop in series resistor,

$$R_S = V_{in} - V_{out} = 50 - 32 = 18 \text{ V}$$

Current through series resistance,

$$I = \frac{V_{in} - V_{out}}{R} = \frac{18}{1800} = .01 \text{ A or } 10 \text{ mA}$$

Load current, $$I_L = \frac{V_{out}}{R_L} = \frac{32}{4000} = 0.008 \text{ A or } 8 \text{ mA}$$

Current through Zener diode,

$$I_z = I - I_L = 10 - 8 = 2 \text{ mA. } \textbf{(Ans.)}$$

Example 2.10. *Determine the maximum and minimum values of Zener current if in the circuit shown in Fig. 2.53 the load resistance, R_L = 4000 Ω, series resistance = 8000 Ω, output voltage = 32 V and source voltage varies between 100 V and 128 V.*

Solution. Refer to Fig. 2.53. *Given :*

$$R_L = 4000\ \Omega;\ R_S = 8000\ \Omega;\ V_{out} = 32 \text{ V}$$

Load current, $I_L = \dfrac{V_{out}}{R_L} = \dfrac{32}{4000} = 0.008$ or 8 mA

The Zener current will be maximum when input voltage is maximum *i.e.*, 128 V.

Corresponding current through series resistance,

$$I' = \frac{V_{in(max)} - V_{out}}{R_S} = \frac{128-32}{8000} = 0.012 \text{ A or } 12 \text{ mA}$$

Corresponding Zener current,

$$(I_Z)_{max.} = I - I_L = 12 - 8 = \mathbf{4\ mA.}\ \textbf{(Ans.)}$$

The zener current will be minimum when input voltage is minimum *i.e.*, 100 V.

Corresponding, current through series resistance,

$$I' = \frac{V_{in(min)} - V_{out}}{R_S} = \frac{100-32}{800} = 0.0085 = 8.5 \text{ mA}$$

Corresponding Zener current,

$$(I_Z)_{min.} = I' - I_L = 8.5 - 8 = \mathbf{0.5\ mA.}\ \textbf{(Ans.)}$$

Example 2.11. *In the simple Zener-diode based voltage regulator shown in Fig. 2.54 a 5.6 V, 0.25 W Zener diode is used. For reliable operation, the minimum I_Z should be 1 mA. The load R_L varies between 20 Ω and 50 Ω. Find the range of R_S for reliable and safe operation of the voltage regulator.*

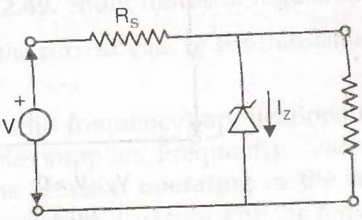


Fig. 2.54

Solution. (*i*) Let $R_S = 20\ \Omega$

$$I = \frac{5.6}{20} = 0.28 \text{ A}$$

$$R_S = \frac{10-5.6}{0.28+0.001} = 15.66\ \Omega \simeq 16\ \Omega.$$

(*ii*) Let, $R_S = 50\ \Omega$

$$I = \frac{5.6}{50} = 0.112 \text{ A}$$

$$R_S = \frac{10-5.6}{0.112+0.001} = 38.93\ \Omega \simeq 39\ \Omega.$$

∴ **R ranges from 16 Ω to 39 Ω. (Ans.)**

Example 2.12. *In the given Fig. 2.55 find the range of R_S for safe and reliable operation of the regulator circuit, if the minimum Zener-diode current is 1mA.*

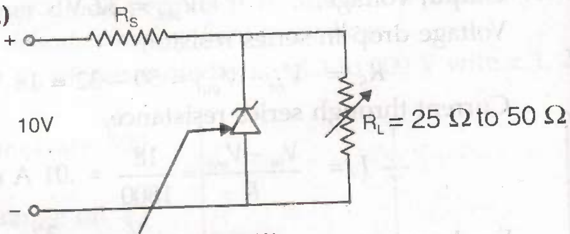


Fig. 2.55

Solution. The equivalent circuit is shown in Fig. 2.56.

The value of load current will be minimum, when the load is maximum *i.e.*, 50 Ω

$$\therefore \quad (I_L)_{min.} = \frac{6}{50} = 120 \text{ mA}$$

The value of load current will be maximum, when the load is minimum *i.e.*, 25 Ω

$$(I_L)_{max.} = \frac{6}{25} = 240 \text{ mA}$$

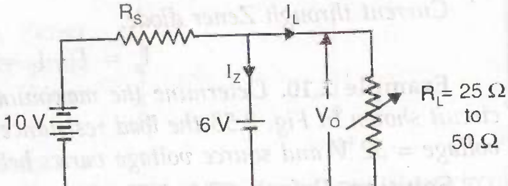


Fig. 2.56

As the load current changes from 120 to 240 mA the Zener current will be minimum, when the load current is maximum.

$$\therefore \quad R_S = \frac{V_{in} - V_0}{(I_Z)_{min} + (I_L)_{max}} = \frac{10 - 6}{(1 + 240)10^{-3}} = \frac{4 \times 10^3}{241}$$

$$= \mathbf{16.59\ \Omega.\ \ (Ans.)}$$

If a series resistance of 16.59 Ω is inserted in the circuit, the output voltage will remain constant. If the load current decreases the Zener current will increase, but the current through R_S will remain the same.

Example 2.13. *In the circuit shown in Fig. 2.57, the voltage across the load is to be maintained at 12 V as load current varies from 0 to 250 mA. Design the regulator. Also find the maximum voltage rating of Zener diode.*

Solution. Refer to Fig. 2.57. By designing the regulator here means to find the values of V_Z and R_S. Since the load voltage is to be maintained at 12 V, we will use a Zener diode of Zener voltage 12 V, *i.e.*, V_Z = **12 V. (Ans.)**

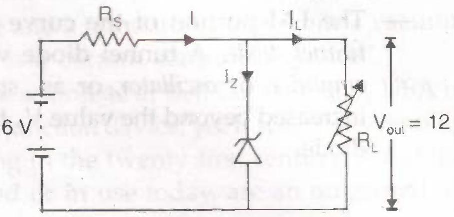


Fig. 2.57

The voltages across R_S is to remain constant at 16 − 12 = 4 V as the load current changes from 0 to 250 mA. *The minimum Zener current will occur when the load current is maximum.*

$$\therefore \quad R_S = \frac{V_{in} - V_{out}}{I} = \frac{V_{in} - V_{out}}{(I_Z)_{min.} + (I_L)_{max.}}$$

$$= \frac{(16 - 12)}{(0 + 250)\text{mA}} = \frac{4}{250 \times 10^{-3}} = \mathbf{16\ \Omega.\ \ (Ans.)}$$

Maximum power rating of Zener diode = 12 × (250 × 10^{-3}) = **3 W. (Ans.)**

Example 2.14. *What value of series resistance is required when three 10 W, 10 V, 800 mA diodes are connected in series to obtain a 30 V regulated output from a 45 V D.C. power source?*

Solution. Fig. 2.58 shows the desired circuit.

The worst case is at no load because then Zener diodes carry the maximum current.

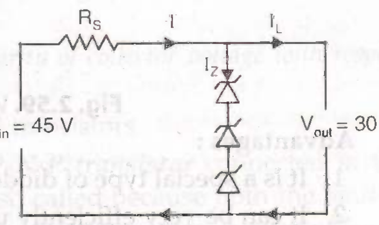


Fig. 2.58

Voltage rating of each Zener diode = 10 V

Current rating of each Zener diode = 800 mA

Input unregulated voltage,

$$V_{in} = 45\ \text{V}$$

Regulated output voltage,

$$V_{out} = 10 + 10 + 10 = 30\ \text{V}$$

Now series resistance, $$R_S = \frac{V_{in} - V_{out}}{I_Z} = \frac{45 - 30}{800 \times 10^{-3}} = \mathbf{18.75\ \Omega.\ \ (Ans.)}$$

2.2.7. Tunnel Diode

Tunnel diode is a heavily doped P-N junction type germanium having an extremely narrow junction. Because the junction is extremely narrow, the electrons can tunnel through it from one side of the junction to the other. The electrons are able to tunnel through it even if they have insufficient energies to overcome the barrier.

V/I Characteristic :

The voltage current (*V/I*) characteristic of such a diode is shown in Fig. 2.59.

— The tunnel diode conducts even during the reverse bias (less than Zener voltage) and a reverse current is produced. For low forward voltages the current is high, and at a certain value of (low) voltage V_1, the current reaches its peak value. When the forward bias increases beyond V_1, the tunnel diode current begins to decrease and reaches a minimum value for a voltage V_2.

— The LM portion of the curve represents a *negative resistance characteristic of the tunnel diode*. A tunnel diode when operated in this region may be used as an *amplifier*, or *oscillator*, or as *switch for timing circuits*. When forward voltage is increased beyond the value V_2, the current starts increasing just as in a conventional diode.

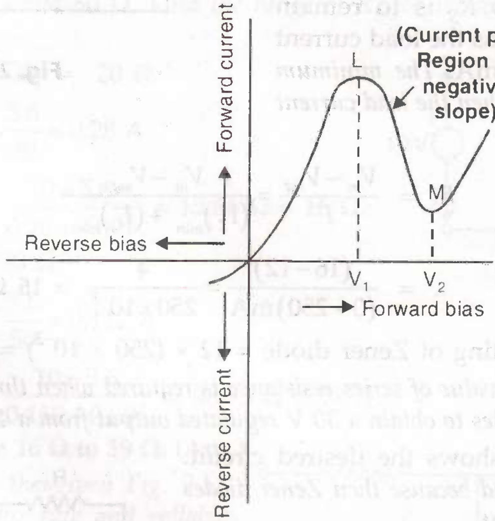


Fig. 2.59. *V/I* characteristic of a tunnel diode.

Advantages :

1. It is a special type of diode which can withstand very large temperature changes.
2. It can be very efficiently used in microwave region.
3. Its consumption is very low (about $\frac{1}{1000}$ th of a transistor)
4. Its cost is low.
5. It is of small size
6. It has a long life.

2.2.8. Biepolar junction transistor (BJT)

Introduction :

A transistor may be *defined* as follows :

— The word *transistor* was derived from the two word combination, *transfer-resistance* (Transfer + resistor ⟶ Transistor). A **transistor** *is a device to transfer a low resistance into a circuit having a high resistance.*

— A *'transistor' is a semiconductor device in which current flows in semiconductor materials.*

— When a thin layer of *P*-type or *N*-type semiconductor is between a pair of opposite types it constitutes a transistor.

- The **transistor** *is a solid state device, whose operation depends upon the flow of electric charge carriers within the solid.*

A **transistor** *is a semiconductor device having both rectifying and amplifying properties.*

"The main difference between a vacuum triode and a transistor is that while a vacuum triode is a voltage controlled device, a transistor is a current controlled device".

The transistor was invented by a team of three scientists at Bell Laboratories, USA in 1947. Although the first transistor was not a bipolar junction device, yet it was the beginning of a technological revolution that is still continuing in the twenty first century. All of the complex electronic devices and systems developed or in use today, are an outgrowth of early developments in semiconductor transistors.

The two basic types of transistors are :

1. Bipolar junction transistor (BJT)
2. Field-effect transistor (FET)

The *bipolar junction transistor is used in the following two broad area of electronics :*

(*i*) *As a* **linear amplifier** *to boost an electric signal.*

(*ii*) *As an* **electronic switch.**

P-N-P and N-P-N transistors. To understand the basic mechanism of transistor operation the following facts need to be kept in mind.

1. Since *emitter is to provide charge carriers, it is always "forward biased".*
2. *First letter of transistor type indicates the polarity of the emitter voltage with respect to base.*
3. Collector's job is *to collect or attract those carriers through the base, hence it is always "reverse-biased".*
4. *Second letter* of transistor type *indicates the polarity of collector voltage with respect to the base.*

The above points apply both to *P-N-P* and *N-P-N* transistors.

Working of P-N-P tansistor. Fig. 2.60 shows a ***P-N-P transistor*** connected in the common-base (or grounded-base) configuration (it is so called because both the emitter and collector are returned to the base terminals). The *emitter junction is forward-biased whereas the collector junction is reverse-biased.* The holes in the emitter are repelled by the positive battery terminal towards the *P-N* or emitter junction. The potential barrier at the junction is reduced due to the forward-biased, hence holes cross the junction and enter the *N*-type base. Because the base is thin and lightly-doped, majority of the holes (about 95%) are able to drift across the base without meeting electrons to combine with. The balance of 5% of holes are lost in the base region due to recombination with electrons. The holes which after crossing the *N-P* collector junction enter the collector region are swept up by the negative collector voltage V_C.

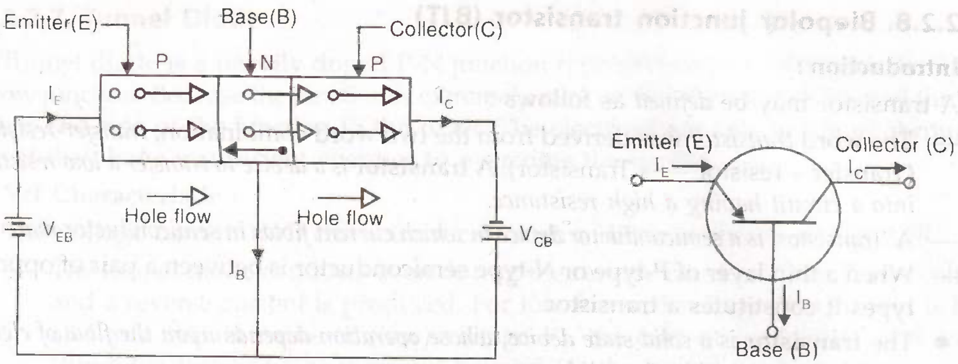


Fig. 2.60. *P-N-P* transistor.

The following points are worth noting :

1. In a *P-N-P* transistor *majority charge carriers are holes.*
2. *The collector current is always less than the emitter current because some recombination of holes and electrons take place.*

$$(I_C = I_E - I_B).$$

3. The *current amplification* (α) (or gain of P-N-P transistor) for steady conditions when connected in common base configuration is expressed as :

$$\alpha = \frac{I_C \text{ (collector current)}}{I_E \text{ (emitter current)}} < 1.$$

4. *Emitter arrow shows the direction of flow of conventional current.* Evidently, electron flow will be in the opposite direction.

Working of N-P-N transistor. Fig 2.61 shows a *N-P-N junction transistor.* The emitter is forward-biased and the collector reverse-biased. The electrons in the emitter region are repelled by the negative battery terminal towards the emitter or *N-P* junction. The *electrons cross over into the P-type base region because potential barrier is reduced due to forward bias.* Since the base is thin and lightly doped, most of the electrons (about 95%) cross over to the collector junction and enter the collector region where they are readily swept up by the positive collector voltage V_C. Only about 5% of the emitter electrons combine with the holes in the base and are lost as charge carriers.

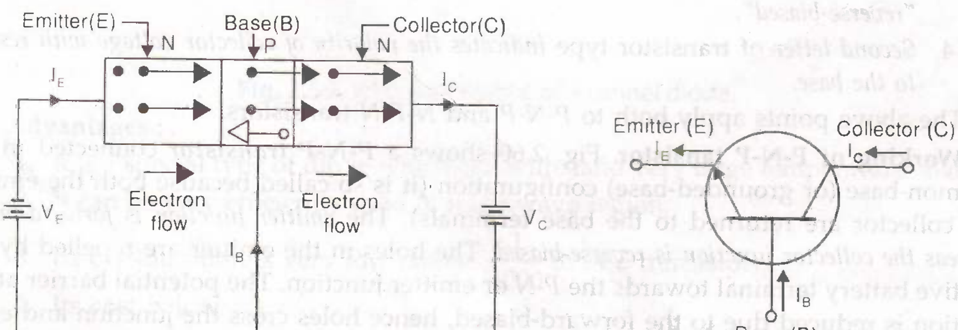


Fig. 2.61. *N-P-N* transistor.

The following points are worth noting :

1. In a *N-P-N* transistor, *majority change carriers are electrons.*
2. I_C *(collector current) is less than* I_E (emitter current) *so that* $\alpha < 1$.

3. *Emitter arrow shows the direction of flow of conventional current.*

- The choice of *N-P-N* transistor is *made more often because majority charge carriers are electrons whose mobility is much more than that of holes.*

Note. *The junction transistors have been made in power ranges from a few milliwatts to tens of watts. The tiny junction transistor is unparalleled in that it can be made to work at power level as I microwatt.*

Transistor circuit configurations. A transistor is a three-terminal device (having three terminals namely *emitter, base* and*collector*) but we require four terminals-two for the input and two for the output for connecting it in a circuit. Hence one of the terminals of the transistor is made common to the input and output circuits. Thus there are three types of configurations for operation of a transistor. These configurations are :

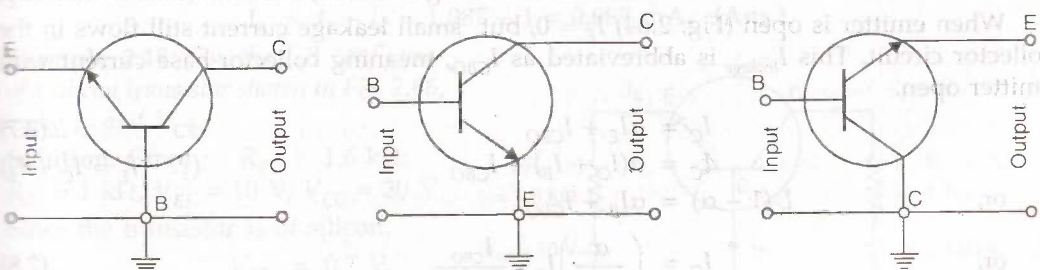
(*i*) Common-base (CB) configuration.

(*ii*) Common-emitter (CE) configuration.

(*iii*) Common-collector (CC) configuration.

The term 'common' is used to denote the electrode that is common to the input and output circuits. Because the common electrodes is generally grounded, these modes of operation are frequently referred to as ground-base, ground-emitter and grounded-collector configurations as shown in Fig. 2.62 for a *N-P-N* transistor.

Each circuit configuration has specific advantages and disadvantages. It may be noted here that regrdless of circuit connection, the *emitter is always biased in the forward direction, while the collector always has a reverse biase.*



(a) CB configuration (b) CE configuration (c) CC configuration

Fig. 2.62. Different circuit configurations for *N-P-N* transistor.

I. Common-base (CB) configuration :

In this circuit configuration, input is applied between emitter and base and output is taken from collector and base. Here, base of the transistor is common to both input and output circuits and hence the name common base configuration. A common-base configuration for *N-P-N* transistor is shown in Fig 2.63.

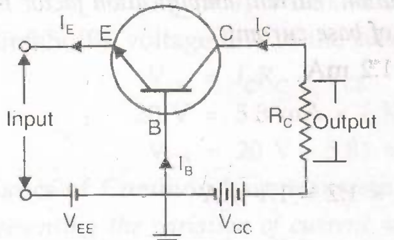


Fig. 2.63. Common-base *N-P-N* transistor.

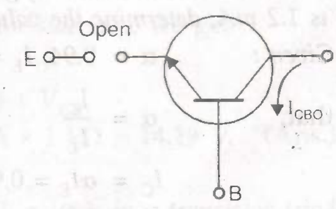


Fig. 2.64

Current amplification factor (α). *It is the ratio of output current to input current.* In CB configuration, the input current is the emitter current I_E and output current is the collector current I_C.

The ratio of change in collector current to the change in emitter current at constant collector-base voltage V_{CB} is known as ***current amplification factor*** *i.e.,*

$$\alpha = \frac{\Delta I_C}{\Delta I_E} \text{ at constant } V_{CB} \qquad ...(2.5)$$

If *only D.C. values are considered, then* $\alpha = \frac{I_C}{I_E}$ ...(2.6)

α is *less than unity.* This value can be increased (not more than unity) by decreasing the base current. This is accomplished by *making the base thin and doping it lightly.*

In commercil transistors, practical value of α varies from 0.9 to 0.99.

Collector current (I_C) :

Total collector current, $I_C = \alpha I_E + I_{leakage}$

(αI_E is the part of emitter current that reaches the collector terminal)

where, I_E = Emitter current, and

$I_{leakage}$ = Leakage current (This current is due to movement of minority carriers across base-collector junction on account of it being reversed; it is *much smaller* than αI_E)

When emitter is open (Fig. 2.64) $I_E = 0$, but small leakage current still flows in the collector circuit. This $I_{leakage}$ is abbreviated as I_{CBO}, meaning collector-base current with emitter open.

$$I_C = \alpha I_E + I_{CBO} \qquad ...(2.7)$$

$$\therefore \quad I_C = \alpha(I_C + I_B) + I_{CBO} \qquad (\because \; I_E = I_C + I_B)$$

or, $$I_C(1 - \alpha) = \alpha I_B + I_{CBO}$$

or, $$I_C = \left(\frac{\alpha}{1-\alpha}\right) I_B + \frac{I_{CBO}}{(1-\alpha)} \qquad ...(2.8)$$

— In view of improved construction techniques, the magnitude of I_{CBO} for general-purpose and *low-powered transistors* (especially silicon transistors) is usually very small and *may be neglected in calculations.*

— For *high power calculations,* I_{CBO} *appears in* μA *range.*

— I_{CBO} *is temperature dependent, therefore, at high temperature it must be considered in calculations.*

Example 2.15. *In a common-base configuration, current amplification factor is 0.92. If the emitter current is 1.2 mA, determine the value of base current.*

Solution. *Given :* $\alpha = 0.94; I_E = 1.2$ mA

We know that, $$\alpha = \frac{I_C}{I_E}$$

or, $$I_C = \alpha I_E = 0.92 \times 1.2 = 1.1 \text{ mA}$$

Also, $$I_E = I_C + I_B$$

$$\therefore \quad I_B = I_E - I_C$$

$$= 1.2 - 1.1 = \mathbf{0.1\ mA.} \quad \textbf{(Ans.)}$$

Example 2.16. *In a common-base configuration, the emitter current is 0.9 mA. If the emitter circuit is open, the collector current is 45 μA. Find the total collector current. Given that α = 0.9.*

Solution. *Given :* $I_E = 0.9$ mA; $I_{CBO} = 45\ \mu A = 45 \times 10^{-3}$ mA; $\alpha = 0.9$.

Collector current, $I_C = \alpha I_E + I_{CBO}$ ...[Eqn. (2.7)]

$= 0.9 \times 0.9 + 45 \times 10^{-3}$ = **0.855 mA. (Ans.)**

Exmple 2.17. *In a CB configuration, α = 0.92. The voltage drop across 2.5 kΩ resistance which is connected in the collector is 2.5 V. Find the base current.*

Solution. *Given :* The common-base configuration of the transistor is shown in Fig. 2.65.

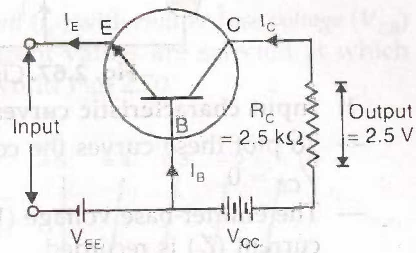


Fig. 2.65

The voltage drop across R_C (= 2.5 kΩ)

= 2.5 V ...(Given)

$$\therefore \quad I_C = \frac{2.5\ V}{2.5\ k\Omega} = 1\ mA$$

Now, $$\alpha = \frac{I_C}{I_E}$$

$$\therefore \quad I_E = \frac{I_C}{\alpha} = \frac{1}{0.92} = 1.087\ mA$$

Also, $I_E = I_C + I_B$

$\therefore$ $I_B = I_E - I_C = 1.087 - 1 =$ **0.087 mA. (Ans.)**

Example 2.18. *For the CB configuration of a silicon transistor shown in Fig. 2.66, determine I_C and V_{CB}.*

Solution. *Given :* $R_E = 1.6$ kΩ; $R_C = 1$ kΩ, $V_{EE} = 10$ V; $V_{CC} = 20$ V

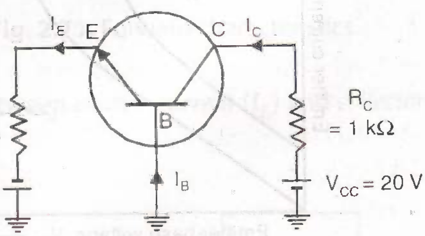


Fig. 2.66

Since the transistor is of silicon,

$V_{BE} = 0.7$ V.

Applying Kirchhoff's voltage law to the *emitter-side loop*, we get

$$V_{EE} = I_E R_E + V_{BE}$$

$$10\ V = I_E \times 1.6\ (k\Omega) + 0.7\ V$$

$$\therefore \quad I_E = \frac{10 - 0.7}{1.6} = 5.81\ mA$$

$\therefore$ $I_C \simeq I_E$ = **5.81 mA. (Ans.)**

Applyig Kirchhoff's voltage law to the *collector-side loop*, we get

$$V_{CC} = I_C R_C + V_{CB}$$

$$20\ V = 5.8\ mA \times 1\ K\Omega + V_{CB}$$

$\therefore$ $V_{CB} = 20\ V - 5.81\ mA \times 1\ k\Omega$ = **14.19 V. (Ans.)**

Characteristics of Common-base transistor :

Curves representing the variation of current with voltage in a transistor triode circuit are called **transistor characteristic curves.** There are the following two types of characteristic curves:

1. **Input characteristic curves** of I_E *versus emitter-base voltage* (V_{EB}).
2. **Output characteristic curves** of *collector current* (I_C) *versus collector-base voltage* (V_{CB}).

Fig. 2.67 shows the circuit of an *N-P-N* junction triode (common-base) studying characteristic curve.

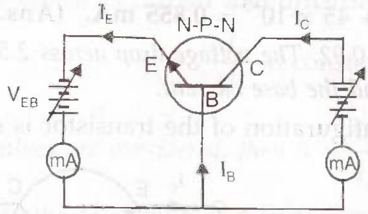


Fig. 2.67. Circuit of an *N-P-N* junction triode.

1. **Input characteristic curves :**

— To plot these curves the collector voltage is first put at zero potential (say), *i.e.*, $V_{CB} = 0$.

— The emitter-base voltage (V_{EB}) is now increased from zero onwards and emitter current (I_E) is recorded.

— A graph is plotted between I_E and V_{EB} as shown in Fig. 2.68.

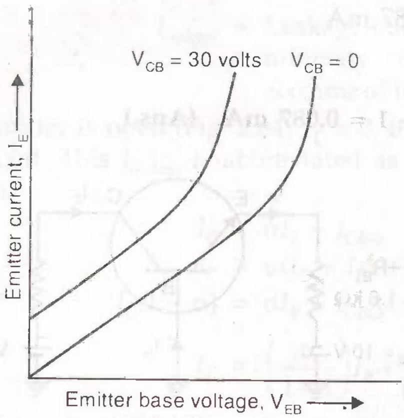


Fig. 2.68. Input characteristic curves.

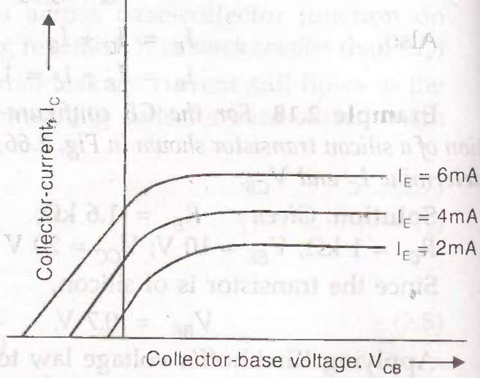


Fig. 2.69. Output characteristic curves.

— Another similar graph is plotted for V_{CB} = 30 volts (say).

From the graph we observe that :

(*i*) For a given collector voltage, the emitter current rises rapidly even with a very small increase in emitter potential. It means that *the input resistance* R_i $\left(= \dfrac{\Delta V_{EB}}{\Delta I_E} \text{ at constant } V_{CB}\right)$ *of the emitter-base circuit is very low.*

(*ii*) The emitter current is nearly *independent* of collector-base voltage.

2. **The output characteristic curves :**

These curves obtained by plotting the variation of collector current (I_C) with collector-base voltage (V_{CB}) at different constant values of emitter current (I_E).

— These *curves* shown in Fig. 2.69 indicate that some collector current is present even when the collector votge is zero. To make the collector current zero, we have to give a certain amount of *negative potential to the collector.*

— The *curves* also indicate that the collector current attains a high value even at a very low collector voltage and further increase in collector voltage does not produce any appreciable increase in collector current. It means that the *output resistance* $\left(R_o = \dfrac{\Delta V_{CB}}{\Delta I_C} \text{ at constant } I_E\right)$ *of the collector-base circuit is very large.*

The *collector current is always a little less than the emitter current* because of the neutralisation of a few holes and electrons within the base due to recombination.

3. **Feed back characteristic curves :**

These curves represent the variation of *collector current* (I_C) with *emitter-base voltage* (V_{EB}) for constant-emitter current. A number of emitter current values are selected at which measurements are made. The nature of curves is shown in Fig. 2.70.

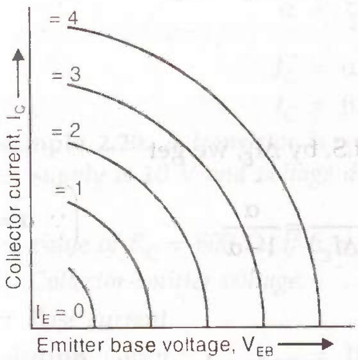


Fig. 2.70. Feed back characteristics.

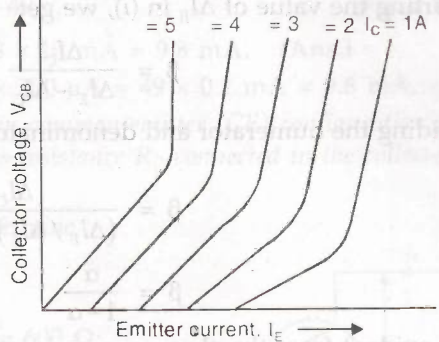


Fig. 2.71. Forward characteristics.

4. **Forward characteristic curves :**

Refer to Fig. 2.71. This type of curve is a graph between *emitter current* (I_E) and *collector-base voltage* at constant value of collector current.

II. Common-emitter (CE) configuration :

In CE configuration, input is applied between base and emitter and output is taken from the collector and emitter. Here, emitter of the transistor is common to both input and outpout circuits and hence the name common-emitter configuration. Fig. 2.72 shows common-emitter *N-P-N* transistor circuit.

Base current amplification factor (β). In CE configuration, input current is I_B and output current is I_C. *The ratio of change in collector current (ΔI_C) to the change in base current (ΔI_B) is known as "base current amplification factor" i.e.,*

$$\beta = \frac{\Delta I_C}{\Delta I_B} \quad ...(14)$$

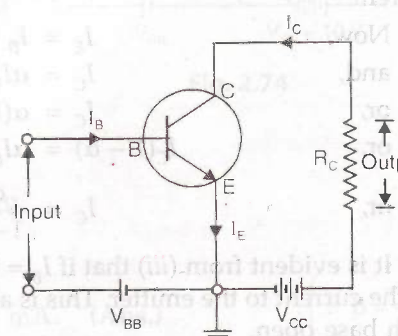


Fig. 2.72. Common emitter *N-P-N* transistor.

If D.C. values are considered,

$$\beta = \frac{I_C}{I_B} \quad ...[2.9\ (a)]$$

In almost every transistor 5% of emitter current flows as the base current. Therefore the value of β is generally greater than 20, β usually varies from 20 to 500.

- *CE configuration is frequently used as it gives appreciable current gain as well as voltage gain.*

Relation between β and α. The relation between β and α is derived as follows :

$$\beta = \frac{\Delta I_C}{\Delta I_B} \quad ...(i)$$

$$\alpha = \frac{\Delta I_C}{\Delta I_E} \quad ...(ii)$$

Now, $I_E = I_B + I_C$

or, $\Delta I_E = \Delta I_B + \Delta I_C$ or $\Delta I_B = \Delta I_E - \Delta I_C$

Inserting the value of ΔI_B in (*i*), we get

$$\beta = \frac{\Delta I_C}{\Delta I_E - \Delta I_C} \quad ...(iii)$$

Dividing the numerator and denominator of R.H.S. by ΔI_E, we get

$$\beta = \frac{\Delta I_C / \Delta I_E}{(\Delta I_E / \Delta I_E) - (\Delta I_C / \Delta I_E)} = \frac{\alpha}{1-\alpha} \quad \left(\because \alpha = \frac{\Delta I_C}{\Delta I_E}\right)$$

$$\therefore \quad \beta = \frac{\alpha}{1-\alpha} \quad ...(2.10)$$

It is evident from the above expression that when α approaches unity, β approaches infinity. In other words the *current gain in CE configuration is very high.* It is due to *this reason that this circuit arrangement is used is about* ***90 to 95 percent of all transistor applications.***

Collector current. In *CE* configuration, I_B is the input current and I_C is the output current :

Now, $I_E = I_B + I_C$ ...(*i*)

and, $I_C = \alpha I_E + I_{CBO}$ ...(*ii*)

or, $I_C = \alpha(I_B + I_C) + I_{CBO}$

or, $I_C(1-\alpha) = \alpha I_B + I_{CBO}$

or, $$I_C = \frac{\alpha}{1-\alpha} I_B + \frac{1}{1-\alpha} I_{CBO} \quad ...(iii)$$

It is evident from (*iii*) that if $I_B = 0$ (*i.e.*, base circuit is open), the collector current will be the current to the emitter. This is abbreviated as I_{CEO} meaning collector-emitter current with base open.

Inserting the value of $\frac{1}{1-\alpha} I_{CBO} = I_{CEO}$ in (*iii*), we get

$$I_C = \frac{\alpha}{1-\alpha} I_B + I_{CEO}$$

or, $$I_C = \beta I_B + I_{CEO} \quad ...(2.11) \quad \left(\because \beta = \frac{\alpha}{1-\alpha}\right)$$

It may be noted that,

$$I_{CEO} = (\beta + 1)I_{CBO} \quad ...(2.12)$$

Example 2.19. *Find the α rating of the transistor shown in Fig. 2.73. Hence determine the value of I_C using both α and β.*

Solution. Refer to Fig. 2.73.

$$\beta = \frac{\alpha}{1-\alpha} \quad ...[\text{Eqn. } (2.10)]$$

or, $\beta(1-\alpha) = \alpha$

or, $\beta - \alpha\beta = \alpha$

or, $\beta = \alpha(1+\beta)$

$\therefore$ $$\alpha = \frac{\beta}{1+\beta} = \frac{49}{1+49} = 0.98$$

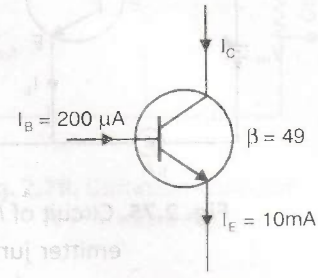


Fig. 2.73

$\therefore$ $I_C = \alpha I_E = 0.98 \times 10 \text{ mA} = \mathbf{9.8 \text{ mA.}}$ **(Ans.)**

Also, $I_C = \beta I_B = 49 \times 200 \,\mu\text{A} = 49 \times 0.2 \text{ mA} = \mathbf{9.8 \text{ mA.}}$ **(Ans.)**

Example 2.20. *A transistor is connected in common-emitter (CE) configuration in which collector supply is 10 V and voltage drop across resistance R_C connected in the collector circuit is 0.6 V.*

The value of R_C = 600 Ω. If α = 0.95, determine :

(*i*) *Collector-emitter voltage.*

(*ii*) *Base current.*

Solution. *Given :* $V_{CC} = 10$ V; $R_C = 600\ \Omega$; $\alpha = 0.95$.

The required *CE* configuration with various values is shown in Fig. 2.74.

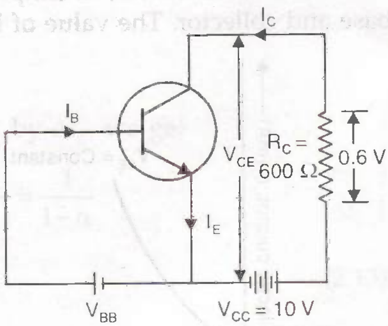


Fig. 2.74

(*i*) **Collector-emitter voltage V_{CE} :**

$$V_{CE} = V_{CC} - 0.6 = 10 - 0.6 = \mathbf{9.4\ V.}\ \textbf{(Ans.)}$$

(*ii*) **Base current I_B :**

$$I_C = \frac{0.6\text{ V}}{600\ \Omega} = 1\text{ mA}$$

Now, $$\beta = \frac{\alpha}{1-\alpha} = \frac{0.95}{1-0.95} = 19$$

$\therefore$ Base current, $$I_B = \frac{I_C}{\beta} = \frac{1}{19} = \mathbf{0.0526\ mA.}\ \textbf{(Ans.)} \qquad \left(\because \beta = \frac{I_C}{I_B}\right)$$

Characteristics of common-emitter transistor :

Fig. 2.75 shows the circuit of a *N-P-N* common-emitter junction transistor for the study of characteristic curves.

1. **Input characteristic curves.** It is the curve between base current I_B and the base-emitter voltage V_{BE} at constant collector-emitter voltage V_{CE} (Refer to Fig. 2.76).

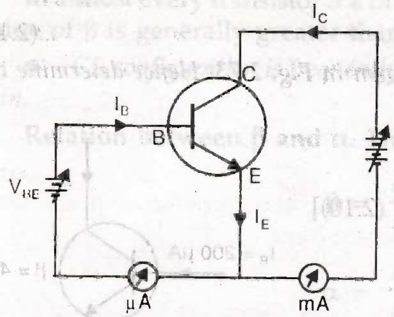


Fig. 2.75. Circuit of *N-P-N* common emitter junction transistor.

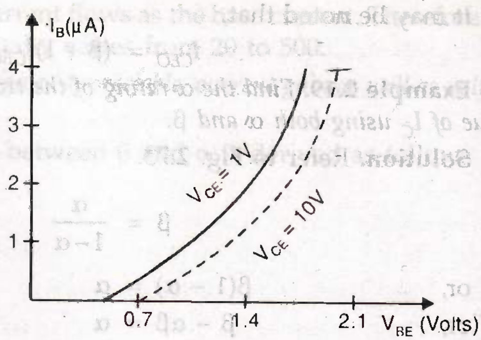


Fig. 2.76

Input resistance, $R_i = \dfrac{\Delta V_{BE}}{\Delta I_B}$ at constant V_{CE}. Its value is of the order of a few hundred ohms.

- Fig. 2.77 shows the graph of collector current (I_C) with base current (I_B) at constant collector-emitter voltage. It may be noted from the curve that there is a collector current even when the basic current is zero. This is known as *collector leakage current* :

It increases with rise in temperature and also arises due to the reverse biasing between base and collector. The value of leakage current ranges from 100 μA to 500 μA.

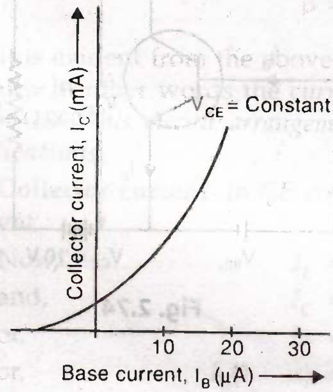


Fig. 2.77

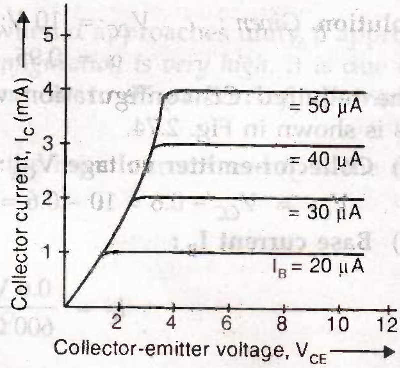


Fig. 2.78

2. **Output characteristic curves.** The collector-emitter voltage (V_{CE}) is varied and the corresponding collector current (I_C) is noted for various fixed values of base current (I_B).

The shape of the curves is shown in Fig. 2.78.

*Such common-emitter characteristics are widely used **for design purpose**.*

It may be noted $I_C >> I_B$

Output resistance $R_o = \dfrac{\Delta V_{CE}}{\Delta I_C}$ at constant I_B. Its value is of the order of 50 kΩ (less than that of *CB* circuit).

III. Common-collector (CC) configuration :

In this type of configuration, input is applied between base and collector while output is taken between the emitter and collector. Here, collector of the transistor is common to both input and output circuits and hence the name common-collector connection.

Fig. 2.79 shows the common-collector *N-P-N* transistor.

Current amplification factor γ. In CC configuration, the input current is the base current I_B and output current is the emitter current I_E. *The ratio of change in emitter current (ΔI_E) to the change in base current (ΔI_B) is known as "current amplification factor" i.e.,*

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

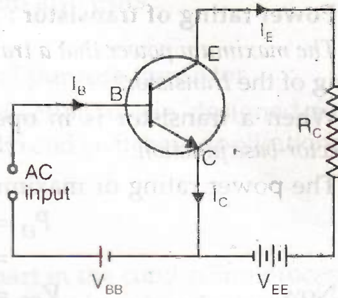


Fig. 2.79. Common-collector *N-P-N* transistor.

This circuit provides the same gain as the common-emitter configuration as $\Delta I_E = \Delta I_C$. However, *its voltage gain is always less than one.*

Relation between γ and α :

$$\gamma = \frac{\Delta I_E}{\Delta I_B} \quad \text{...(i)}$$

$$\alpha = \frac{\Delta I_C}{\Delta I_E} \quad \text{...(ii)}$$

Now, $I_E = I_B + I_C$,

or, $\Delta I_E = \Delta I_B + \Delta I_C$ or $\Delta I_B = \Delta I_E - \Delta I_C$

Inserting the value of ΔI_B in (*i*), we get

$$\gamma = \frac{\Delta I_E}{\Delta I_E - \Delta I_C}$$

Dividing the numerator and denominator of R.H.S. by ΔI_E, we get

$$\gamma = \frac{\Delta I_E / \Delta I_E}{(\Delta I_E / \Delta I_E) - (\Delta I_C / \Delta I_E)} = \frac{1}{1-\alpha} \qquad \left(\because \alpha = \frac{\Delta I_C}{\Delta I_E} \right)$$

$$\therefore \quad \gamma = \frac{1}{1-\alpha} \quad \text{...(2.13)}$$

Collector current :

We know that, $I_C = \alpha I_E + I_{CBO}$

Also, $I_E = I_B + I_C = I_B + (\alpha I_E + I_{CBO})$

or, $I_E(1-\alpha) = I_B + I_{CBO}$

or, $I_E = \dfrac{I_B}{1-\alpha} + \dfrac{I_{CBO}}{1-\alpha}$

or, $I_C, I_E = (\beta + 1)I_B + (\beta + 1)I_{CBO}$ ...(2.14)

$$\left[\beta = \frac{\alpha}{1-\alpha} \quad \because \beta + 1 = \frac{\alpha}{1-\alpha} + 1 = \frac{1}{1-\alpha} \right]$$

☞ **Commonly used transistor connection :**

Out of the three configuration, the **CE configuration** *is the most efficient.* It is used in about 90 to 95% of all transistor applications. This is due to following *reasons* :

1. *High current gain;* it may range from 20 to 500.
2. *High voltage and power gain.*
3. *Moderate output to input impedance ratio* (this ratio is small, to the tune of 50). This *makes this configuration an ideal one for coupling between various transistor stages.*

Power rating of transistor :

The maximum power that a transformer can handle without deterioration is known as **power rating** of the *transistor.*

When a transistor is in operation, almost all power is dissipated at *reverse biased *collector-base junction.*

The power rating or maximum power dissipation is given by,

$$P_D = \text{Collector current} \times \text{Collector-base voltage}$$

$$= I_C \times V_{CB}$$

Now, $V_{CE} = V_{CB} + V_{BE}$,

Since V_{BE} is very small,

$$V_{CB} \simeq V_{CE}$$

$$\therefore \quad P_D = I_C \times V_{CE} \qquad ...(20)$$

- *While connecting a transistor in the circuit it must be ensured that its power rating is not exceeded otherwise it may* ***get destroyed due to overheating.***

Semiconductor devices numbering system :

From the day the semiconductor devices come into existence different numbers were used in different countries. However, the numbering system announced by *Protection Standardisation Authority in Belgium* has been accepted and adopted internationally.

According to this numbering system :

(*i*) Every conductor device is numbered by *five* alpha-numeric symbol, comprising either two letters and three numbers (*e.g.* BF 194) or three letters and two numbers (*e.g.* BFX 63).

— The devices comprising *two letters and three numbers* (*e.g.* BF 194) are intended for *entertainment or consumer equipment.*

— The devices comprising *three letters and two numbers* (*e.g.* BFX 63) are intended for *industrial or professional equipment.*

(*ii*) The *first letter* indicates the *nature of semiconductor material.*

Example. *A* = Germanium, *B* = Silicon, C = Gallium arsenide, *R* = compound material (*e.g.* cadmium sulphate).

Thus AC 125 is a *germanium* transistor whereas BC 149 is a *silicon* transistor.

(*iii*) The second letter indicates the device and its circuit function *e.g.*,

A—Diode
B—Varactor (variable capacitance diode)
C—Audio-frequency (AF) low power transistor
D—AF power transistor
E—Tunnel diode
F—High frequency (HF) low power transistor
G—Multiple device
H—Magnetic sensitive device
K—Hall -effect device
L—High-frequency (HF) power transistor
M—Hall effect modulator
P—Radiation sensitive diode
Q—Radiation generating diode
R—Thyristor (SCR or Triac)
S—Low power switching transistor
T—High power transistor
U—Power switching transistor
X—Diode, multiplier
Y—Power device
Z—Zener diode.

* Power dissipated at the base-emitter junction is *negligible* [The base-emitter junction conducts about the same current as the collection-base junction ($I_E \approx I_C$), but V_{BE} is very small (0.3 V and 0.7 V for Ge and Si transistors respectively.)]

In addition to the above system, other numbering system also exits :

Examples : 1 N 4001 ...Silicon diode

2 N 3903 ...Silicon *N-P-N* general purpose transistor

2 N 5457 ...N-Channel FET deflection mode designed for general purpose audio and switching applications.

2.2.9 Field-Effect Transistor (FET)

Introduction :

In an ordinary transistor both holes and electrons play part in the conduction process and so it is sometimes called a *bipolar transistor.* This ordinary transistor has the following two main *disadvantages* : (*i*) It has a low input impedance (because of forward biased emitter junction), and (*ii*) It has considerable noise level. The field-effect transistor (FET) has, by virtue of its construction and biasing, large input impedance (which may be more than 100 MΩ. The FET is generally *much less noisy* than the BJT).

Types of Field-effect Transistors :

A **field-effect transistor** (*FET*) *is a three terminal (namely* **drain, source** *and* **gate**) *semiconductor device in which current conduction is by only one type of majority carriers (electrons in case of an N-channel FET or holes in a P-channel FET).* It is also sometimes called the *unipolar transistor.*

In a broad sense, following are two main types of field-effect transistors :

1. *Junction field-effect transistor (JFET)*

(*i*) *N*-channel

(*ii*) *P*-channel.

2. *Metal oxide semiconductor field-effect transistor (MOSFET) or insulated gate field-effect transistor (IGFET).*

(*i*) Depletion type :

(*a*) *N*-channel

(*b*) *P*-channel

(*ii*) Enhancement type :

(*a*) *N*-channel

(*b*) *P*-channel

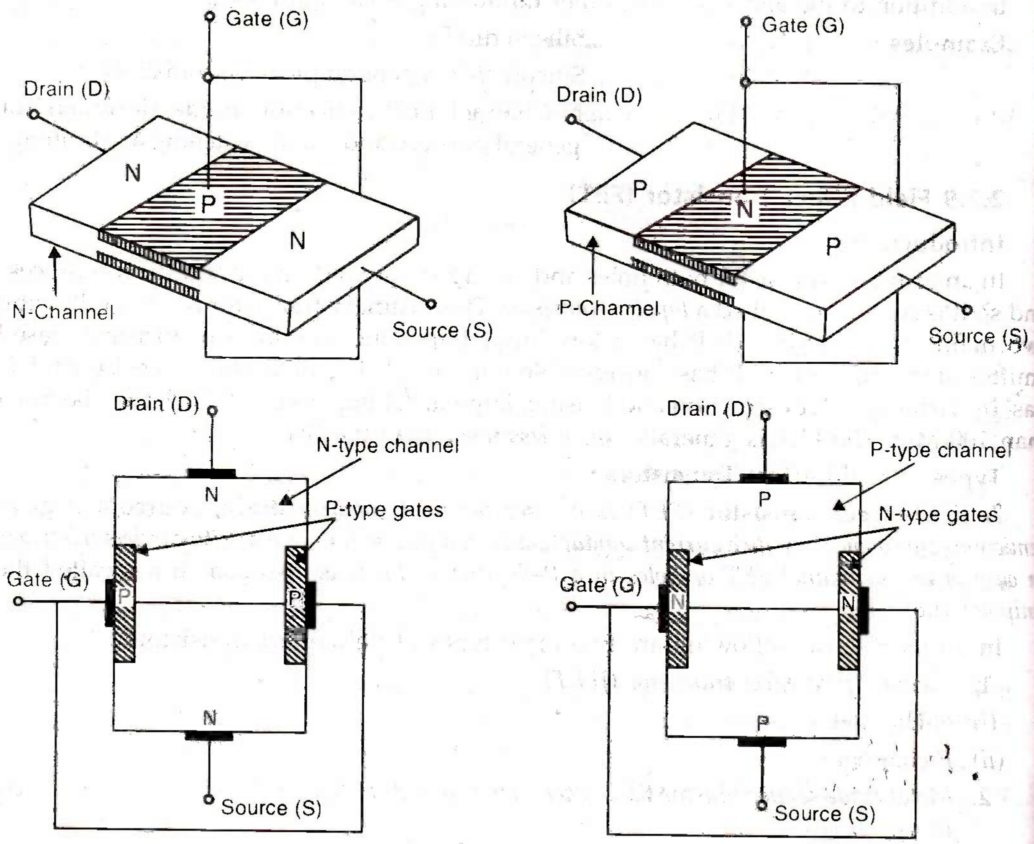
1. Junction field-effect transistors (JFET) :

The junction field-effect transistors (JFETs) can be divided depending upon their structure into the two following categories :

1. *N*-channel JFET
2. *P*-channel JFET

Construction :

- The basic construction of a *N*-channel JFET is as shown in Fig. 2.80 (*a*). It consists of an *N-type semiconductor bar with two P-type heavily doped regions diffused on opposite sides of its middle part.* The *P*-type regions form two *P-N* junctions. The *space between the junctions (i.e., N-type regions) is called a* **channel.** Both the *P-type* regions are connected internally and a single wire is taken out in the form of a terminal called the **gate (G).** The electrical connections (called *ohmic contacts*) are made to both ends of the N-type semiconductor and are taken out in the form of two "terminals called **drain (D)** and **source (S).** *The "drain (D)" is a terminal through which the electrons leave the semiconductor and "source (S)" is a terminal through which the electrons enter the semiconductor.*

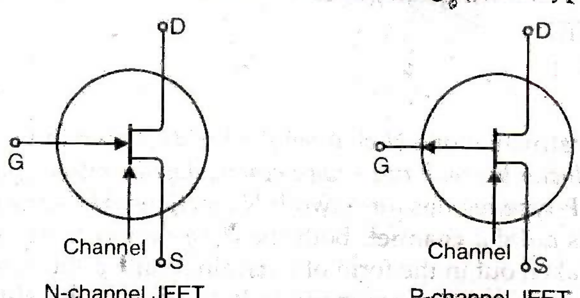


(a) N-Channel JFET (b) P-Channel JFET

Fig. 2.80. JFETs.

Whenever a voltage is applied across the drain and source terminals, a current flows through the *N*-channel. The current consists of only one type of carriers (*i.e.*, electrons), therefore, the FET is called a *unipolar device.* (This distinguishes FET from BJT where the current consists of the flow of both the electrons and holes).

- A *P*-channel JFET is shown in Fig. 2.80 (*b*). Its construction is similar to that of *N*-channel JFET, except that it consists of a *P-channel and N-type junctions.* The current carriers in JFET are the *holes,* which flow through the *P*-type channel.



N-channel JFET (a) P-channel JFET (b)

Fig. 2.81. Symbols for JFETs.

— Fig. 2.81 (*a*) shows the schematic symbol for a *N*-channel *JFET*. The arrow points *towards* the vertical line. The *vertical line* represents the *N-channel*.

— Similarly, Fig. 2.81 (*b*) shows the schematic symbol for a *P*-channel JFET.

The arrow points *away from* the vertical line. Here the vertical line represents the *P*-channel.

JFET polarities :

Fig. 2.82 (*a*) shows *N*-channel JFET polarities whereas Fig. 2.82 (*b*) shows the *P*-channel JFET polarities. It may be noted that in each case, the voltage between gate and source is such that the *gate is reverse biased*. This is the normal way of JFET connection. The drain and source terminals are interchangeable (This is generally valid for low frequencies but not for high frequencies applications).

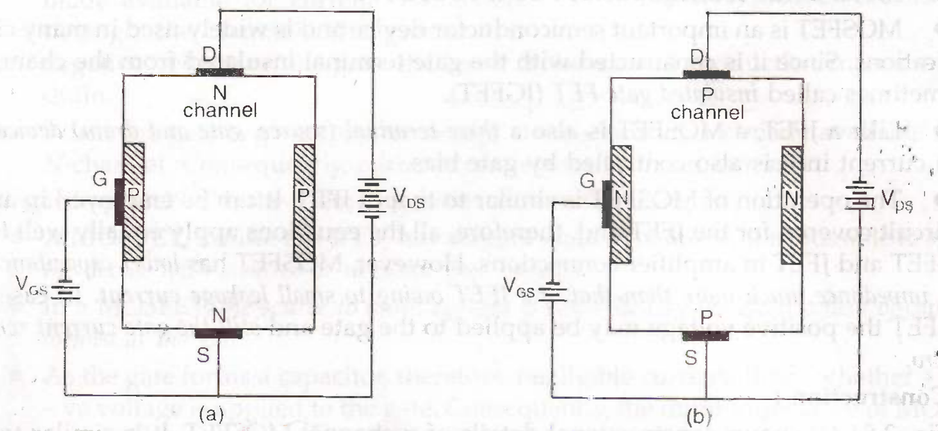


Fig. 2.82. JFET polarities.

Working :

Fig. 2.83 shows the circuit *N*-channel JFET with normal polarities. The circuit action is as follows :

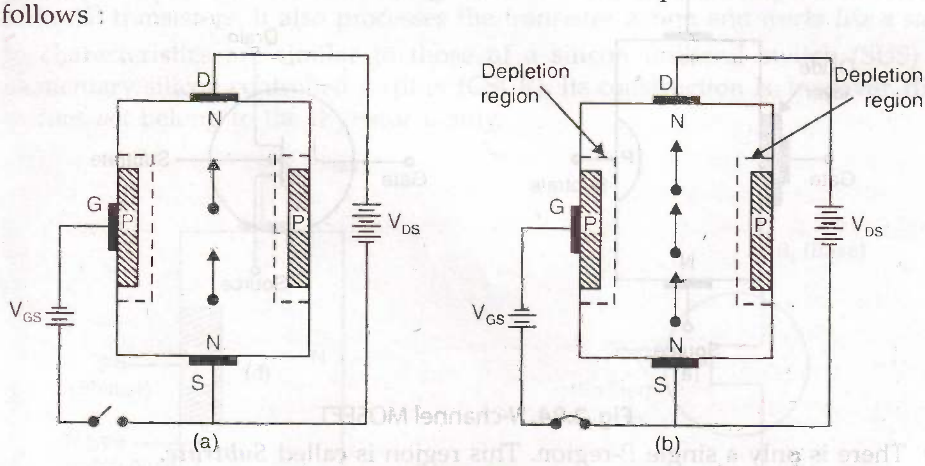


Fig. 2.83

(*a*) When a voltage V_{DS} is applied between drain and source terminals and voltage on the gate is zero [Fig. 2.83 (*a*)], the two *P-N* junctions at the sides of the bar establish deplection layers. The electrons will flow from source to drain through a channel between the depletion layers. *The size of these layers determines the width of the channel* and hence the current conduction through the bars.

(*b*) When a reverse voltge V_{GS} is applied between the gate and source [Fig. 2.83 (*b*)], the *width of the depletion layers is increased*. This *reduces* the width of conducting channel, thereby increasing the resistance of *N*-type bar. Consequently, the current from the source to drain is *decreased*. On the other hand if the reverse voltage on the gate is decreased, the width of depletion layers also decreases. This increases the width of the channel and source to drain current increases.

From the above discussion it is evident that current from source to drain can be controlled by the application of potential or electric field on the gate. It is due to this reason that this device is called field-effect transistor. Note that a *P-channel* JFET operates in the same manner as an *N*-channel JFET *except that channel current carriers will be the holes instead of electrons and the polarities of* V_{GS} *and* V_{DS} are reversed.

2. Metal oxide semiconductor FET (MOSFET) :

- MOSFET is an important semiconductor device and is widely used in many circuit applications. Since it is constructed with the gate terminal insulated from the channel, it is sometimes called *insulated gate FET* (IGFET).
- Like a JFET, a MOSFET is also a *three-terminal (source, gate and drain) device* and drain current in it is also controlled by gate bias.
- The operation of MOSFET is similar to that of JFET. It can be employed in any of the circuit covered for the JFET and, therefore, all the equations apply equally well to the MOSFET and JFET in amplifier connections. However, MOSFET has *lower capacitance and input impedance much more than that of a JFET owing to small leakage current*. In case of a MOSFET the positive voltage may be applied to the gate and *still the gate current remains the zero*.

Construction :

Fig. 2.84 (*a*) shows constructional details of *n*-channel MOSFET. It is similar to FET except with the following modifications :

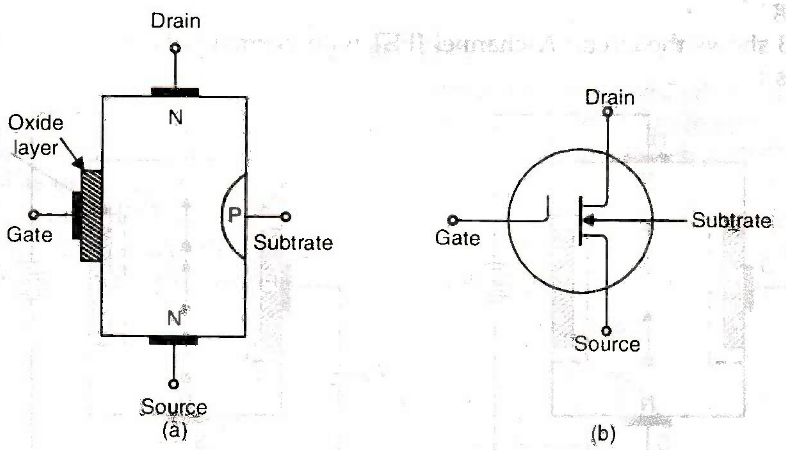


Fig. 2.84. *N*-channel MOSFET.

(*i*) There is only a single *P*-region. This region is called ***Subtrate.***

(*ii*) A thin layer of metal oxide (usually silicon oxide) is deposited over the left side of the channel. A *metalic **gate** is deposited over the oxide layers.* As silicon dioxide is an insulator, therefore, *gate is insulated from the channel.*

(*iii*) Like JFET, a MOSFET has three terminals *viz, source, gate* and *drain.*

Fig. 2.84 (*b*) shows the symbolic symbol of *N*-channel MOSFET.

Working :

Fig. 2.85 shows the MOSFET circuit. Instead of gate diode as in JFET, here gate is formed as a small capacitor. *One plate of this capacitor is the gate and the other plate is the channel with metal oxide as the dielectric.*

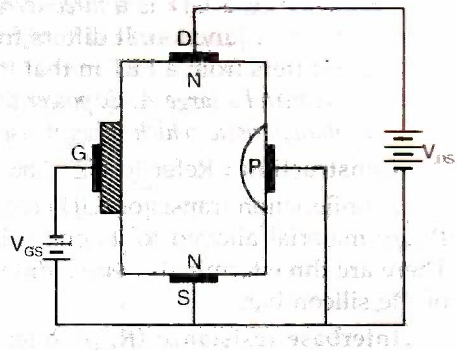


Fig. 2.85. MOSFET circuits.

— When *negative voltage* is applied to the gate, electrons accumulate on it. These electrons repel the conduction band electrons in the *N*-channel. Therefore, lesser number of conduction electrons are made available for current conduction through the channel. The greater the negative voltage on the gate, the lesser is the current conduction from source to drain.

— When the gate is given *positive voltage,* more electrons are made available in the *N*-channel. Consequently, current from source to drain *increases.*

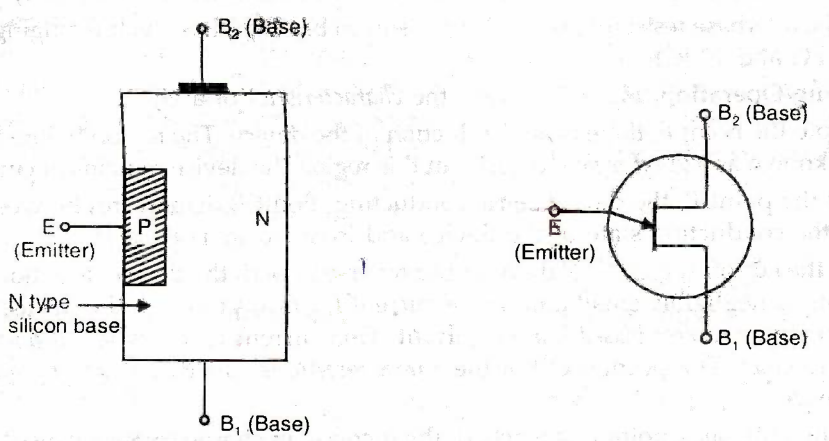
Regarding MOSFET, the following *points* are worth noting :

- **A MOSFET, unlike the JFET, has *no gate diode.*** Therefore, it is possible *to operate the device with positive or negative gate voltage.*
- In a MOSFET, *the source to drain current is controlled by the electric field of capacitor formed at the gate.*
- As the gate forms a capacitor, therefore, negligible currents flows whether + ve or – ve voltage is applied to the gate. Consequently, the input impedance of MOSFET is *very high* (varying from 10^4 MΩ to 10^6 MΩ).

2.2.10 Unijunction Transistor (UJT)

A unijunction transistor, unlike a bipolar transistor has *only one junction.* Like other conventional transistors, it also processes the transistor action and *works like a switch.*

Its characteristics are similar to those of a silicon unilteral switch (SUS) and a complementary silicon controlled rectifier (CSCR). Its construction is, however, different and it does *not* belong to the thyristor family.



(a) Construction of a UJT (b) Symbolic diagram of a UJT

Fig. 2.86. Unijunction Transistor (UJT).

- Basically, a UJT is a *three-terminal silicon diode.* As its name indicates, it has only one *P-N* junction. It differs from an ordinary diode in that it has *three* leads and it differs from a FET in that it has *no ability to amplify.* However, it has the ability to *control a large A. C. power with a small signal.* It also exhibits a *negative resistance characteristic which mkes it useful as an oscillator.*

Construction : Refer to Fig. 2.86

A unijunction transistor (UJT) consists of a *lightly doped silicon* bar with a *heavily doped P-type* material alloyed to its one side (closer to B_2) for producing single *P-N* junction. There are three terminals : one emitter, E and two bases B_1 and B_2 at the bottom and top of the silicon bar.

Interbase resistance (R_{BB}): Refer to Fig. 2.87.

The interbase resistance (R_{BB}) is the total resistance of the silicon bar from one end to the other with emitter terminal open; from equivalent circuit (see Fig. 2.87), we have

$$R_{BB} = R_{B_2} + R_{B_1}$$

The point C is such that $R_{B_1} > R_{B_2}$ (usually R_{B_1} is 60 percent of R_{BB}). R_{B_1} has been shown as a variable resistor because its value varies inversely as I_E.

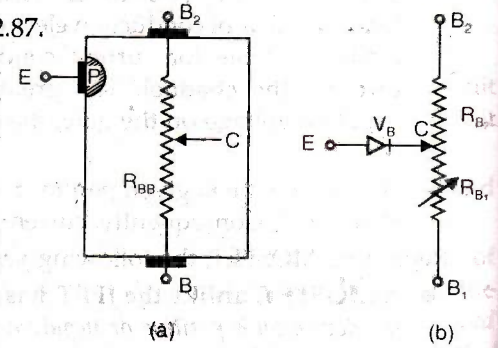


Fig. 2.87

Let the voltage drop across R_{B_1} is V_C. Then,

$$V_C = V_{BB} \times \frac{R_{B_1}}{R_{B_1} + R_{B_2}} \quad \text{...using voltage binder relationship}$$

$$= \eta . V_{BB}$$

where, $$\eta = \frac{R_{B_1}}{R_{B_1} + R_{B_2}}$$

η is called the ***instrinsic stand ratio.***

— The value of η depends on two factors namely : (*i*) Construction of the UJT, and (*ii*) Spacing between the emitter junction and the two base contacts.

— The value of η is always less than unity (lies between 0.51 and 0.81)

— The interbase resistance of the *N*-type silicon bar (R_{BB}) has a value ranging between 4 kΩ and 12 kΩ.

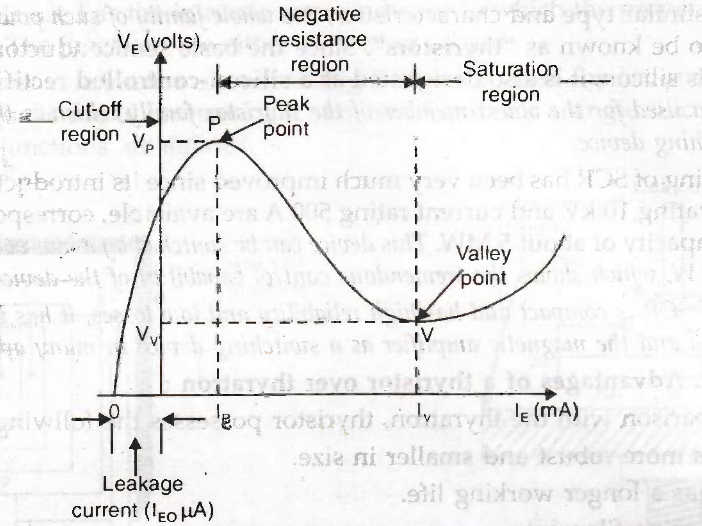
Working/Operation. Fig. 2.88 shows the *characteristics* of a UJT.

— Upto the point *P*, there is no conduction of the device. The region before this point is known as *'cut-off region'* because in this region the device remains in cut-off state.

Just at the point *P*, the device starts conducting. Point *P* demarcates between cut-off state and the conduction state of the device and is called its *peak point.*

— In the cut-off region, *P-N* diode being reverse biased, the device does not conduct. Only a negligibly small amount of current I_{EO} flows through the device which is known as *reverse biased leakage* current. This current is not *sufficient for the device to conduct.* The portion *OP* of the characteristic is called the *cut-off region* of the device.

— When the peak point *P* is reached, the increase in charge carriers causes decrease in resistance R_{B_1} and the device starts *conducting.* In the conduction state, the device depicts a negative resistance charcteristics. This means, as the emitter voltage V_E is further increased, the voltage across the device decreases, but the current increases. This region of conduction is called the *negative resistance region.* This region continues till the valley point V is reached in the characteristic. The portion PV of the characteristic is called the *negative resistance region.*



I_{EO} = Leakage current ; V_P = Peak point voltage; I_P = Peak point current; V_V = Valley point voltage; I_V = Valley point current; V_E = Emitter voltage ; I_E = Emitter current.

Fig. 2.88. Characteristics of UJT.

— After reaching the valley point, the device goes to its saturation state where further fall in the voltage across the device does not take place. The device voltage and current both reach standard values and do not change any more even if the emitter voltage is changed. This portion of the characteristic beyond the valley point is called *'saturation region'*.

A set of *V-I* characteristic for UJI can be obtained for different values of interbase voltage V_{BB}.

- It is seen that only terminals *E and B_1 are active terminals whereas B_2 is the bias terminal i.e.,* it is meant only for applying external voltage across UJT.
- Generally, *UJT is triggered into conduction by applying a suitable positive pulse at its emitter.* It can be brought *back to OFF state by applying a negative trigger pulse.*

Applications. One significant property of UJT is that it *can be triggered by* (or an output can be taken from) *any one of its three terminals.* Once triggered, the emitter current I_E of the UJT increases regeneratively till it reaches a limiting value determined by the external power supply. Owing to their particular behaviour, UJT is used in variety of circuit applications; some of these are :

1. Pulse generation;
2. Sine wave generator;
3. Saw tooth wave generator;
4. Switching;
5. Phase control;
6. Voltage or current regulated supplies;
7. Timing and trigger circuits.

2.2.11. Thyristor

2.2.11.1. Introduction :

Ample pioneering work on theory and fabrication of the *power-switching device,* which later came to be known as a *thyristor* (because its characteristics are *similar to those of the* *gas-tube thyratron*), was done at the Bell Laboratories in the U.S.A. The first prototype was introduced by the General Electric Company (USA) in 1957. Since then, many improvements have been made, both in the technique of its fabrication and in adapting it to numerous industrial appplications. With the development of a number of other devices of similar type and characteristics, the *whole family of such power-switching devices* has come to be known as **"thyristors"**. Since the basic semiconductor material used for the device is silicon, it is also designated as a **silicon-controlled rectifier (SCR).** *The term SCR is often used for the oldest member of the thyristor family which is the most widely used power-switching device.*

The rating of SCR has been very much improved since its introduction and now SCRs of voltage rating 10 kV and current rating 500 A are available, corresponding to a power-handling capacity of about 5 MW. *This device can be switched by a low-voltage supply of about 1 A and 10 W, which shows the tremendous control capability of the device.*

Because SCR is compact and has high reliability and low losses, it has more or less replaced the thyratron and the magnetic amplifier as a switching device in many applications.

2.2.11.2. Advantages of a thyristor over thyratron :

It comparison with the thyratron, thyristor possesses the follwing *advantages* :

1. It is more robust and smaller in size.
2. It has a longer working life.
3. It has no filament.
4. The voltage drop in the forward direction is only about 1 to 2 volts, compared 10 to 15 volts for the thyratron.
5. The triggering and recovery periods are much shorter, so that it is more suitable for *high-frequency switching operations.*
6. The arc ionizing and deionizing times for a thyratron are comparatively large and so the device applications are limited to a frequency of 1 kHz. *A thyristor can operate over a much greater range of frequency.*

Comparison between transistors and thyristors :

The comparison between transistors and thyristors is given in Table 2.1.

Table 2.1. Comparison between "Transistors" and "Thyristors"

S. No.	Aspects	Transistors	Thyristors
1.	*Type of device*	3-layers, 2-junction devices	4-layer, 2-or more junction devices
2.	*Reponse*	Fast	Very fast
3.	*Efficiency*	High	Very high
4.	*Reliability*	Highly reliable	Very highly reliable
5.	*Voltage drop*	Small voltage drop	Very small voltage drop
6.	*Life*	Long life	Very long life
7.	*Power ratings*	Small to medium power ratings	Very small to very large power ratings
8.	Conducting state	Require a continuous flow of current to remain in conducting state	Require only small pulse for triggering and thereafter remaining in conducting state.
9.	*Power consumption*	Low power consumption	Very low power consumption
10.	*Control capability*	Low control capability	High control capability
11.	*ON, OFF timings*	Small turn-ON and turn-OFF timings	Very small turn-ON and turn-OFF timings.

2.2.11.3. Construction, operation and characteristics of a thyristor :

Construction :

- The cross-sectional view of a typical SCR is shown in Fig. 2.89. Basically, the SCR consists of a four-layer pallet of *P*-type and *N*-type semiconductor materials. Silicon is used as the intrinsic semiconductor to which the proper impurities are added. The junctions are either *diffused* or *alloyed*.
- The *planer construction* shown in Fig 2.89 (*a*) is used for *low-power* SCRs. This technique is useful for making a number of units for a single silicon wafer. Here, all the junctions are *diffused*.

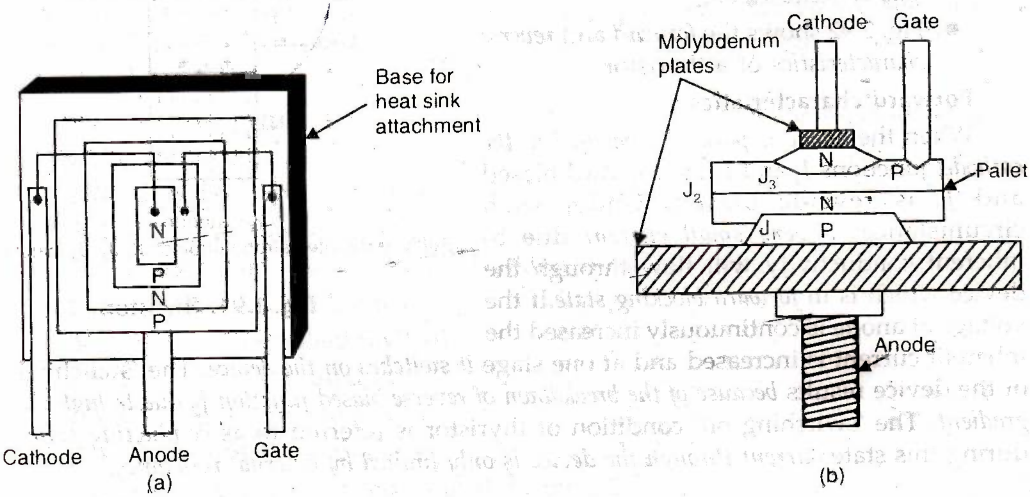


Fig. 2.89. (*a*) Planer type (all diffused), (*b*) Mesa type (alloy diffused).

- The other technique the *mesa construction* is shown in Fig. 2.89 (*b*). This is used for *high-power* SCRs. Here, the inner junction J_2 is obtained by diffusion, and then the outer two layers are alloyed to it. Because the *PNPN* pallet is required to handle large currents, it is properly braced with tungsten or molybdenum plates to provide greter mechanical strength. One of these plates is handsoldered to a copper or an aluminium stud, which is threaded for attachment to a heat sink. This provides an efficient thermal path for conducting the internal losses to the surrounding medium. The use of hand solder between the pallet and back up plates *minnimises thermal fatigue* when the SCRs are subjected to temperature-induced stresses. For medium and low-power SCRs, the pallet is mounted directly on the copper stud or casing, using soft-solder which absorbs the thermal stresses set-up by differential expansion and provides a good thermal path for heat transfer. When a larger cooling arrangement is required for highpower SCRs, the press-pack or hockey pack construction is used, which provides for double-sided air or water cooling.
- Fig. 2.90 shows the terminal configuration and symbolic diagram of a SCR.

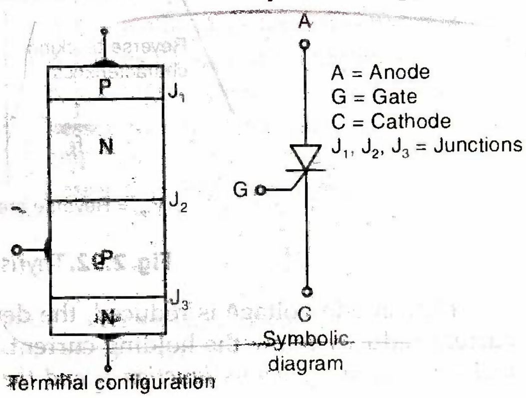
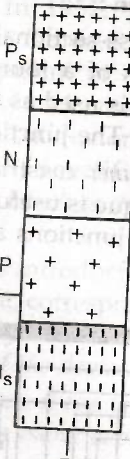


Fig. 2.90. Schematic diagram of a SCR.

- Fig. 2.91 shows a thyristor. It has four layers alternately of P and N with *outer layers heavely doped.* The control current is generally applied *via* middle P layer and N emitter. *Junctions J_1 and J_3 are forward biased while junction J_2 is reverse biased. If the voltage gradient between anode and cathode is too high then due to increased inherent current the thyristor may be switched on.*
- Fig. 2.92 shows the *forward* and *reverse characteristics* of a thyristor.



Suffix 's' denotes strong doping; J_1, J_2, J_3–junctions.

Fig. 2.91. Thyristor

Forward characteristics :

When the *anode is positive compared to the cathode,* junctions J_1 and J_3 are forward biased and J_2 is reverse biased. Under such circumstances *a very small current* due to inherent conductivity will flow through the device which is in *forward blocking state.*If the voltage at anode is continuously increased the inherent current is increased and at one stage *it switches on the device.* The 'switching on' of the device results *because of the breakdown of reverse biased junction J_2 due to high voltage gradient.* The 'switching on' condition of thyristor is referred to as *conducting state* and during this state *current through the device is only limited by external resistance.*

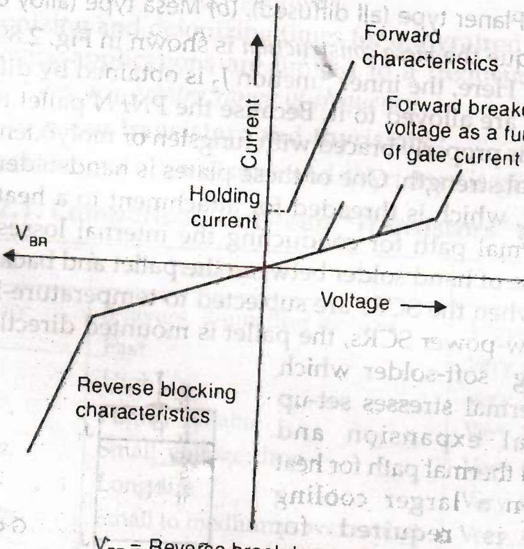


Fig. 2.92. Thyristor characteristics.

If the anode voltage is reduced, the device will continue to conduct till the forward current reduces below the holding current I_h. *Below this value of current the depletion layer will start appearing across junction J_2 and the device will be 'blocked'.*

Reverse blocking characteristics :

If the *cathode is positive as compared to anode,* the junction J_2 *is reverse biased* and only a small current flows through the device and the characteristics are called *reverse blocking characteristics.* If the voltage is continuously increased at one stage it may result in breaking of depletion layers at junctions J_1 and J_3 and the current through the device suddenly increases to a very high value. This is called *reverse breakdown* and the voltage is called reverse breakdown voltage V_{BR}.

As the outer layers of thyristor are highly doped compared to inner layers, the thickness of depletion layers at J_2 during forward bias is much more as compared to the total thickness of depletion layer at junctions J_1 and J_3 during reverse bias. Hence, the forward breakdown voltage V_{BF} *is normally greater than reverse breakdown voltge* V_{BR}.

Thyristor Family :

There are several members in the thyristor family, some of them are mentioned below:

1. DIAC (Bidirectional Diode Thyristor)
2. TRIAC (Bidirectional Triode Thyristor)
3. SCR (Silicon Controlled Rectifier)
4. SUS (Silicon Unilateral Switch), also known as complementry SCR (CSCR)
5. SBS (Silicon Bilateral Switch)
6. SCS (Silicon Controlled Switch)
7. LASCR (Light Activated SCR)
8. LASCS (Light Activated SCS).

2.2.11.4. Typical SCR parameters :

Typical SCR parameters are given in the table 2.2

Table 2.2. Typical SCR parameters

S. No.	Parameters	Typical
1.	Forward breakover voltage, V_{FBO}	50 to 500 volts
2.	Maximum on-state voltage	About 1.5 V
3.	Peak reverse voltage, PRV	Upto 2.5 kV
4.	Holding voltage, V_H	0.5 to 20 volts
5.	Forward breakover current	Less than a few hundred μA
6.	Peak forward current	30 A to over 100 A
7.	Holding current	A few mA to few hundred mA
8.	Turn-on and turn-off times	A few tenths of μs for fast acting SCRs; A few μs for slow acting SCRs
9.	Dynamic resistance in cut-off region	A few MΩ to a few hundred MΩ
10.	Dynamic resistance in saturation region.	Lesser than 1 Ω for currents of several amperes; Lesser than 10 Ω for large currents.

2.2.11.5. Diac

Refer to Fig. 2.93.

A **Diac** *is a two terminal, three layer bi-directional device which can be switched to ON state for either polarity of the applied voltage.* It is, therefore, also known as a *'bi-directional avalanche diode'.*

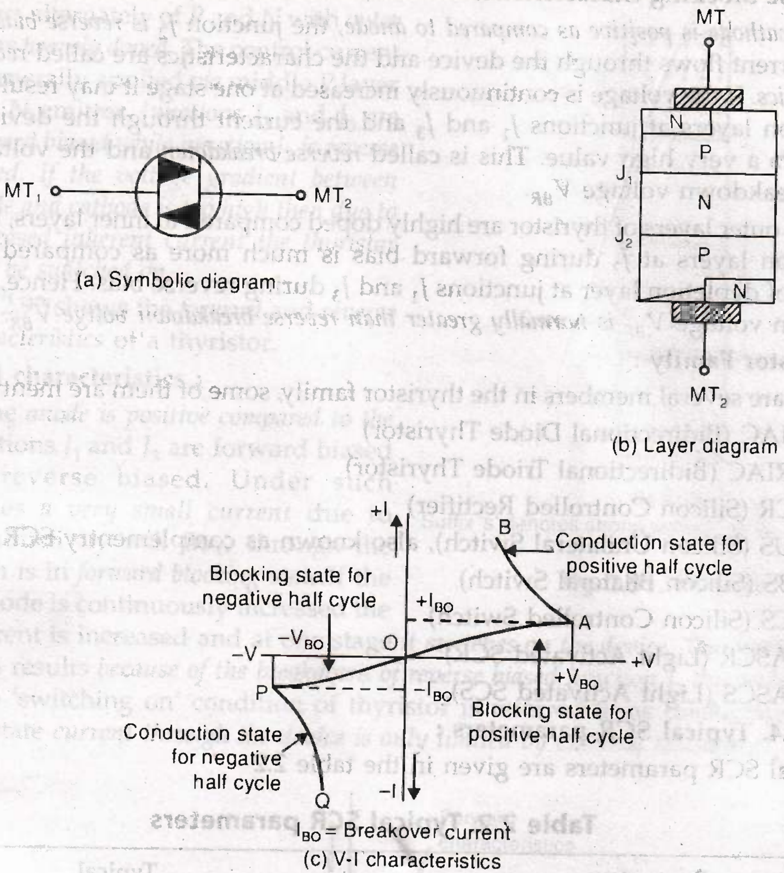


Fig. 2.93. Diac.

Fig. 2.93 (*a*, *b*) shows the construction of diac.

— A *diac* is a *PNPN* structured four layer, two terminal semconductor device. MT_1 and MT_2 are the two main terminals of the device. There is *no* control terminal in this device.

— It has two junctions J_1 and J_2.

— It is evident from the layer diagram (Fig. 2.93 (*b*)) that, a diac unlike a diode resembles bipolar transistor. However, the central layer of the diac is free from any connection with the terminals. The doping level at the two ends of the device is the same which leads to identical *V-I* characteristics in both Ist and IIIrd quadrants. Fig. 2.93 (*c*) shows the *V-I* characteristics of a diac.

— *When positive or negative voltage is applied across the terminals of a diac, only small leakage current will continue to flow through the device. As the applied voltage is increased, the leakage current will continue to flow until the voltage reaches the breakdown point. At this point, avalanche breakdown of the reverse biased junctions occurs and current through the device increases sharply.*

Applications. Diacs are used primarily for *triggering triacs in adjustable phase control of A.C. mains supply.*

2.2.11.6. Triac

A **triac** *is a three terminal semiconductor switching device which can control alternating current in load.*

One major difference between an SCR and triac is that whereas SCR is a unidirectional switch and can conduct in one direction only, a triac is *bi-directional switch and can conduct in either direction.*

Construction :

The **triac** is *a three terminal, four layer semiconductor device.* Its three terminals are MT_1, MT_2 and the 'Gate'. Its symbol, layer diagram and pin configuration are shown in Fig. 2.94 (*a*), (*b*) (*c*) respectively.

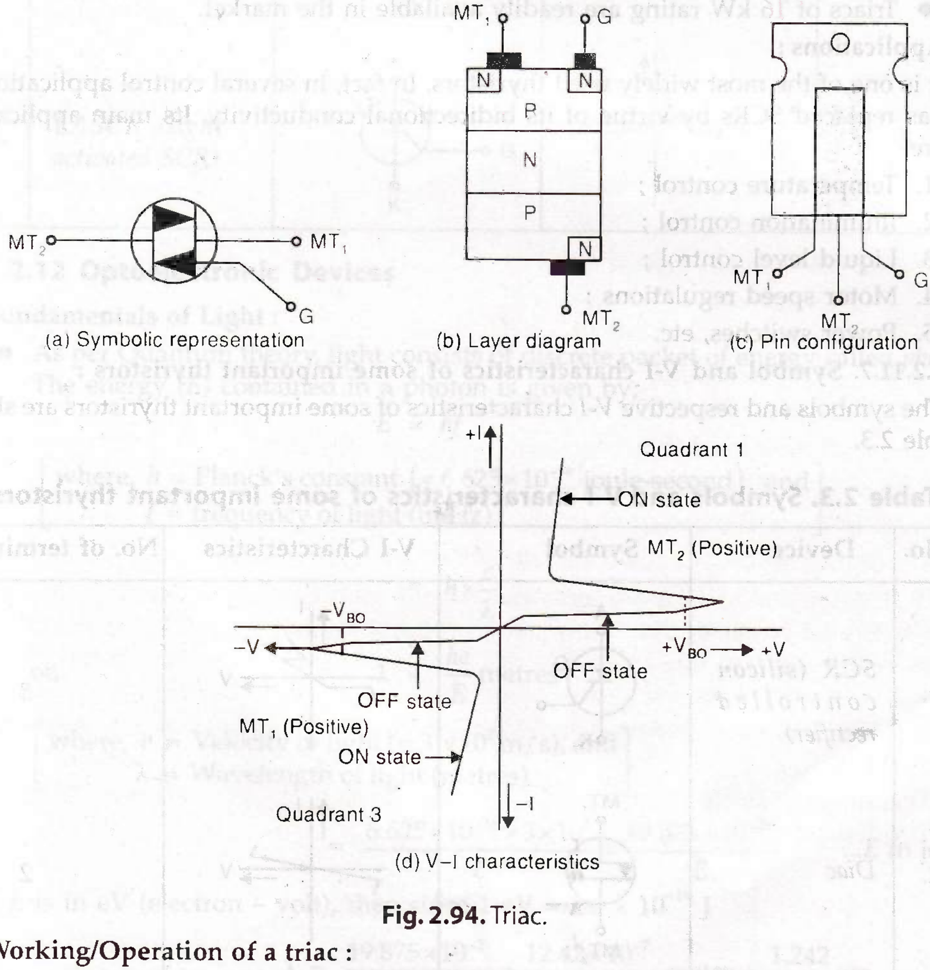


Fig. 2.94. Triac.

Working/Operation of a triac :

Fig. 2.94 shows the *V-I* characteristics of a triac.

— A triac, like an SCR, also starts conducting only when the breakover voltage is reached. Earlier to that, the leakage current which is very small in magnitude, flows through the device and therefore it remains in the OFF state.

— The 1st quadrant characteristic is just like an SCR, but 3rd quadrant characteristic of a triac is identical to its 1st quadrant, *except that, as the polarities of the main terminals change, the direction of current changes.*

— MT_2 is positive with respect to MT_1 in the 1st quadrant and it is negative in the 3rd quadrant. The device, when starts conducting, allows very heavy amount of current to flow through it. This high inrush of current must be limited by using external resistance, or it may otherwise damage the device. The **'gate'** is the *control terminal* of the device. By applying proper signal at the gate, the firing angle of the device can be changed thus, the phase control process can be changed.

- The *great advantage of triac* is that by adjusting the gate current to a proper value, any portion of both positive and negative half cycles of A.C. supply can be made to flow through the load. This *permits to adjust the transfer of A.C. power from the source to the load.*

— Its *main limitation* in comparison to SCRs is, its *low power handling capacity.*

- Triacs of 16 kW rating are readily available in the market.

Applications :

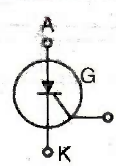
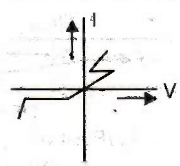
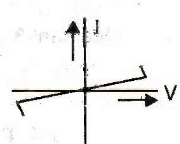
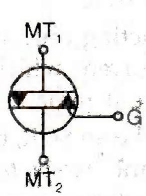
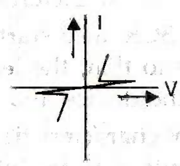
It is one of the most widely used thyristors. In fact, in several control applications, it has replaced SCRs by virtue of its bidirectional conductivity. Its main applications are:

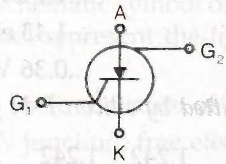
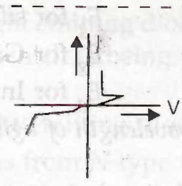
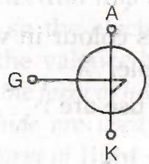
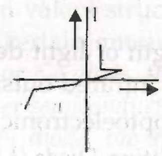
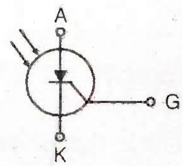
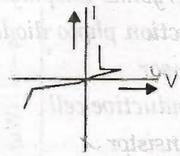
1. Temperature control ;
2. Illumination control ;
3. Liquid level control ;
4. Motor speed regulations ;
5. Power switches, etc.

2.2.11.7. Symbol and V-I characteristics of some important thyristors :

The symbols and respective *V-I* characteristics of some important thyristors are shown in table 2.3.

Table 2.3. Symbols and V-I characteristics of some important thyristors

S.No.	Device	Symbol	V-I Charcteristics	No. of terminals
1.	*SCR (silicon controlled rectifier)*			3
2.	*Diac*			2
3.	*Triac*			3

4.	*SCS (silicon controlled switch)*			4
5.	*SUS (silicon unilateral switch)*			3
6.	*LASCR (light activated SCR)*			3

2.2.12 Optoelectronic Devices

Fundamentals of Light :

- As per Quantum theory, light consists of discrete packet of energy called *photons.* The energy (E) contained in a photon is given by;

$$E = hf$$

$$\left[\begin{array}{l}\text{where, } h = \text{Planck's constant } \left(= 6.625\times10^{-34} \text{ joule-second}\right)\text{, and} \\ \qquad f = \text{frequency of light (in Hz)}\end{array}\right]$$

$$= h\times\frac{c}{\lambda}$$

or, $$\lambda = \frac{hc}{E} \text{ metres}$$

$$\left[\begin{array}{l}\text{where, } c = \text{Velocity of light } (= 3\times10^{8}\text{ m/s})\text{, and} \\ \qquad \lambda = \text{Wavelength of light (metres)}\end{array}\right]$$

$$= \frac{6.625\times10^{-34}\times3\times10^{8}}{E} = \frac{19.875\times10^{-26}}{E} \quad \text{...}E \text{ in joules}$$

If E is in eV (electron – volt), then since 1 eV = 1.6×10^{-19} J

$$\therefore \quad \lambda = \frac{19.875\times10^{-26}}{E\times1.6\times10^{-19}} = \frac{12.42\times10^{-7}}{E} \text{ metre} = \frac{1.242}{E}\ \mu\text{m}$$

- When the *P-N* junction is forward biased, both the electrons as well as holes cross the junction. During this process some electrons recombine with holes, consequently some energy is lost by the electrons. The *amount of energy lost (given off in the form of light energy) is equal to the difference in energy between the conduction and valence bands, this being known as the semiconductor energy band gap* E_g.

E_g for *silicon* ...1.1 eV

E_g for Ga As ...1.43 eV

E_g for In As ...0.36 V

Example. *The wavelength of light emitted by silicon P-N junction,*

$$\lambda = \frac{1.242}{E_g} = \frac{1.242}{1.1} = 1.13\ \mu m$$

- The wavelength of light determines its colour in visible range and whether it is ultraviolet or infrared outside the visible.
- The various optoelectronic devices in use are :
  - — *Light Emitting Diode (LED)*
  - — *Liquid Crystals Displays (LCD)*
  - — *P-N junction photo diode*
  - — *Dust Sensor*
  - — *Photoconductive cell*
  - — *Phototransistor*
  - — *Photodarlington*
  - — *Photovoltaic or Solar cell*
  - — *Laser Diode*
  - — *Optical Disks*
  - — *Hologram Scanners*
  - — *Light activated SCR (LASCR)*
  - — *Optical Isolators*
  - — *Optimal Modulators etc.*

Some of these devices are discusses hence forth.

1. Light Emitting Diode (LED) :

A *P-N* junction can absorb light energy and produce electric current. The opposite process is also possible, that is a junction diode can emit light. The emission of light occurs under forward bias condition due to recombination of electrons and holes. The emitted light may be visible or invisible.

A P-N junction diode, which emits light when forward biased is known as a **light emitting diode (LED).**

The amount of light output is *directly proportional to the foward current.* Thus, higher the forward current, higher is the light output.

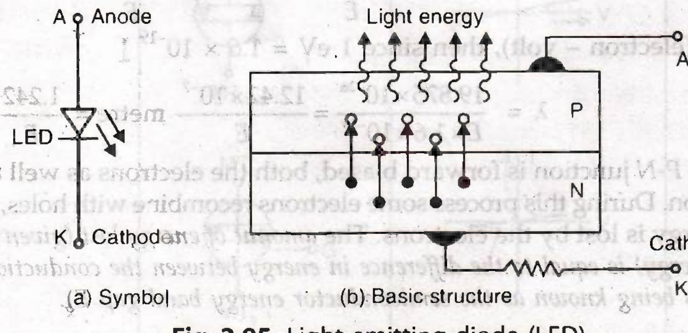


Fig. 2.95. Light emitting diode (LED).

Fig. 2.95 (*a*) shows the schematic symbol of a light emitting diode. The arrows pointing away from the diode symbol represent the *light*, which is being transmitted away from the junction.

Fig. 2.95 (*b*) shows the basic structure of a light emitting diode.

In a forward based *P-N* junction, free electrons from *N*-type material diffuse into *P*-region. Once in *P*-region these free electrons encounter holes and eventually recombine. Thus the free conduction electron fills a vacancy in valent structure and thus becomes a valence electron. In doing so the electron loses a certain amount of energy as it jumps from conduction band to the valence band. In *Si or Ge diode, the energy that recombining electrons lose is dissipated in the form of heat. But if other semiconductor material such as gallium arsenide and gallium phosphide are used to form P-N diode, the energy lost by recombining electrons is given off in the form of* light energy.

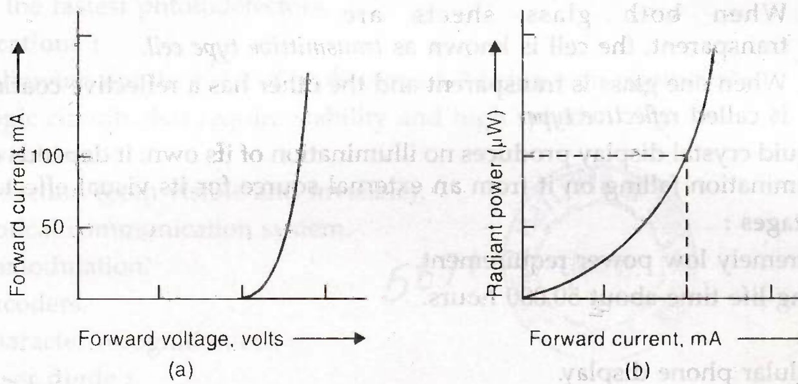


Fig. 2.96. Operating characteristics—LED.

— Diodes made of gallium arsenide (GaAs) emit infrared radiation invisible to eyes and such diodes are referred to as IRED—Infrared emitting diodes.

Fig. 2.96 shows two curves used to determine LED operating characteristics.

— Fig. 2.96 (*a*) is forward bias *V-I* curve for a typical IRED, the type used in burglar alarms. Forward bias of around 1 V is required to produce significant forward current.

— Fig. 2.96 (*b*) is a plot of radiant output power *vs* forward current. The radiant output power is rather small (μW) and *indicates a very low efficiency of electrical to radiant energy conversion.*

Applications :

Since LEDs operate at voltage levels 1.5 V to 3.3 V, they are highly compatible with solid state circuitry. Their uses include the following :

(*i*) Panel indicator
(*ii*) Digital watches
(*iii*) Calculators
(*iv*) Multimeters
(*v*) Intercoms
(*vi*) Switch boards
(*vii*) Burglar-alarm systems
(*viii*) Optical fibre communication system etc.

2. **Liquid Crystal Displays (LCD)**

- A liquid crystal is a material, usually an organic compound, which flows like a liquid at room temperature; its molecular structure has some properties normally associated with solids (*e.g.* chloesteryl nonanoate and *p*-azoxyanisole).
- When light is incident on an activated layer of a liquid crystal, it is *either absorbed or else is scattered by the disoriented molecules.*
- A liquid crystal 'cell' (Fig. 2.97) consists of a thin layer (about 10 μm) of a liquid crystal sandwiched between two glass sheets with transparent electrodes deposited on their inside faces.
  - — When both glass sheets are transparent, the cell is known as *transmittive type cell.*
  - — When one glass is transparent and the other has a reflective coating, the cell is called *reflective type.*

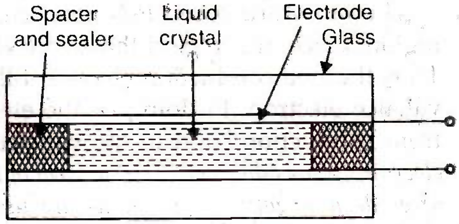


Fig. 2.97. A liquid crystal 'cell'.

- Liquid crystal display produces no illumination of its own; it depends entirely on illumination falling on it from an external source for its visual effect.

Advantages :

1. Extremely low power requirement.
2. Long life time-about 50,000 hours.

Uses :

1. Cellular phone display.
2. Desk top LCD monitors.
3. Note book computer display.
4. Watches and portable instruments.
5. Pocket T.V. receiver.

3. **Photo-voltaic cell :**

- In this cell sensitive element is a semiconductor (not metal) which generates voltage in proportion to the light or any radiant energy incident on it. The most commonly used photo-voltaic cells are barrier layer type like *iron-selenium* cells or *Cu-CuO*$_2$ cells.
- Fig. 2.98 shows a typical widely used photo-voltaic cell—*Selenium cell.* It consists of a metal electrode on which a layer of selenium is deposited; on the top of this a barrier layer is formed which is coated with very thin layer of gold. The latter serves as a transluscent electrode. When light falls, a negative charge will build up on the gold electrode and a positive charge on the bottom electrode.

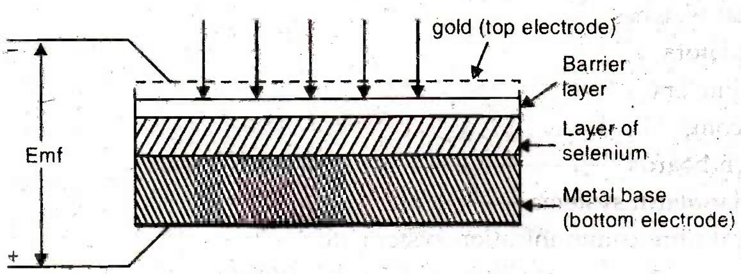


Fig. 2.98. Photo-voltaic cell.

Uses :

Photo-voltaic cells are widely used in the following fields :

(*i*) Automatic control systems.

(*ii*) Television circuits.

(*iii*) Sound motion picture and reproducing equipment.

4. **P-N Junction photodiode :**

- It is a two-terminal junction device which is operated first by reverse-biasing the junction and then illuminating it.
- The active diameter of these devices is about 2.5 mm but they are mounted in standard TO-5 packages with a window to allow maximum incident light.
- A photodiode can turn its current ON and OFF in nanoseconds, hence it is one of the fastest phtotodetectors.

Applications :

The following are the *fields of application* of *P-N* junction photodiode :

(*i*) Logic circuits that require stability and high speed.

(*ii*) Switching.

(*iii*) Detection (both visible and invisible).

(*iv*) Optical communication system.

(*v*) Demodulation.

(*vi*) Encoders.

(*vii*) Character recognition etc.

5. **Laser diode :**

The word LASER is an acronym for *Light Amplification by Stimulated Emission of Radiation.*

Laser diodes, like LED, are typical *P-N* junction devices used under a *forward bias.*

Laser diodes are of the following *two types :*

1. **Surface-emitting laser diodes.** These diodes emit light in a direction *perpendicular* of the *P-N* junction plane.
2. **Edge-emitting laser diodes.** These diodes emit light in a direction *parallel* to the *P-N* junction plane.

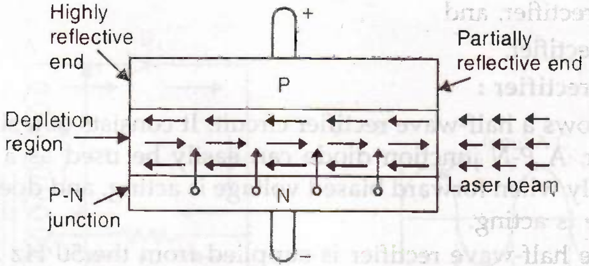


Fig. 2.99. Edge-emitting laser diode.

Fig. 2.99 shows an edge-emitting laser diode (called Fabry-Petrot type laser).

When an external voltage forward biases the *P-N* junction the electons move across the junction and usual combination takes place in the depletion region, resulting in the *production of photons.* With the increase in forward current, more photons are produced which drift at random in the depletion region. Some of these photons strike the reflective surface in the perpendicular direction. These reflected photons enter the depletion region, strike other atoms and release more photons. These photons move back and forth between reflective surfaces. The photon activity becomes so intense that at some point, a strong beam of laser light comes out of the partially reflective surface of the diode.

The beam of laser light is *coherent, monochromatic* and is *collimated.*

- The schematic symbol (Fig. 100) of a laser diode is similar to that of LED; a filler or lens is necessary to view the laser beam.

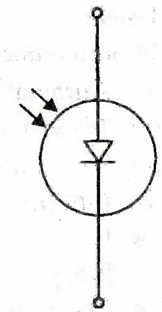


Fig. 2.100. Schematic symbol of a laser diode.

Applications :

1. Medical equipment used in surgery.
2. Compact disk (CD) players.
3. Laser printers.
4. Hologram scanners.
5. Parallel processing of information.
6. Parallel interconnections between computers etc.

6. Light Activated SCR (LASCR) :

- It is just an ordinary SCR except that it *can also be light-triggered.*
- Most LASCRs also have a gate terminal for being triggered by an electronic pluse just as conventional SCR. Fig. 2.101 shows the two LASCR symbols commonly used.
- These are manufactured mostly in relatively low-current ranges.

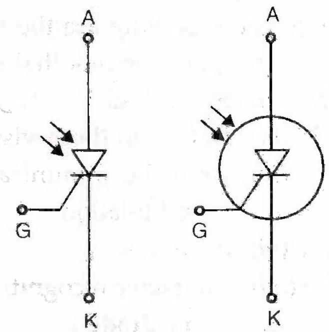


Fig. 2.101. LASCR symbols.

Applications :

1. Used for *triggering laser SCRs and triac.*
2. Used in *optical light controls, relays, motor control and a variety of computer applications.*

2.2.13. Rectifiers

A **rectifier** is *a circuit, which uses one or more diodes to convert A.C. voltage into pulsating D.C. voltage.*

A rectifier my be broadly categorized in the followign *two types :*

1. Half-wave rectifier, and
2. Full-wave rectifier.

1. Half-wave rectifier :

Fig. 2.102 (*a*) shows a half-wave rectifier circuit. It consists of a single diode in series with a load resistor. A *P-N* junction diode can easily be used as a rectifier because it conducts current only when forward biased voltage is acting, and does not conduct when reverse bias voltage is acting.

The input to the half-wave rectifier is supplied from the 50 Hz A.C. supply, whose wave form is shown in Fig. 2.102 (*b*).

Operation :

When an A.C. voltage source is connected across the junction diode as shown in Fig. 2.102 (*a*) the *positive half cycle of the input acts as forward bias voltage* and the output across the load resistance varies correspondingly. The *negative half cycle* of the input acts as a *reverse bias* and practically no current flows in the circuit. The output is, therefore, *intermittent, pulsating* and *unidirectional.*

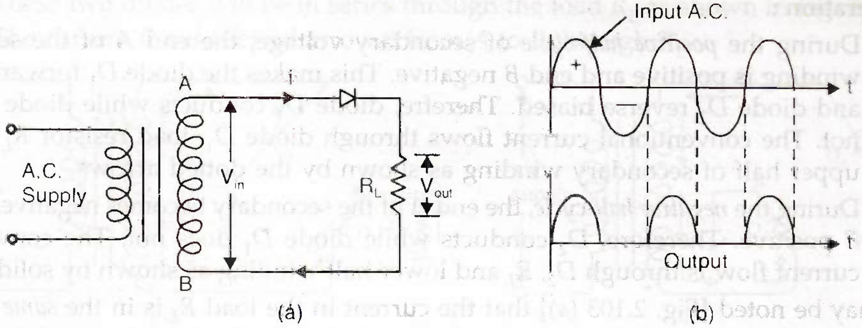


Fig. 2.102. Half-wave rectifier.

It is evident from the above discussion, that as the *circuit uses only one-half cycle of the A.C. input voltage, therefore, it is popularly known as a "half-wave rectifier"*.

Disadvantages :

The main *disadvantages* of a half-wave rectifier are :

(*i*) The A.C. supply delivers power only half the time; therefore, its *output is low.*

(*ii*) The pulsating current in the load contains alternating component whose basic frequency is equal to the supply frequency. Therefore, an elaborate filtering is required to produce steady direct current.

2. Full-wave rectifier :

A ***full-wave rectifier** is a circuit, which allows a unidirectional current to flow through the load during the entire input cycle.* This can be achieved with *two diodes working alternately.* For the positive half-cycle of input voltage, one diode supplies current to the load and for the negative half-cycle, the other doide does so; *current being always in the same direction through the load.*

For full-wave rectification the following *two circuits* are commonly used :

1. Centre-tapped full-wave rectifier.
2. Full-wave bridge rectifier.

Centre-tapped full-wave rectifier :

Fig. 2.103 shows the circuit of a centre-tapped full-wave rectifier. The circuit uses two diodes (D_1 and D_2) which are connected to the centre-tapped secondary winding *AB* of the transformer.

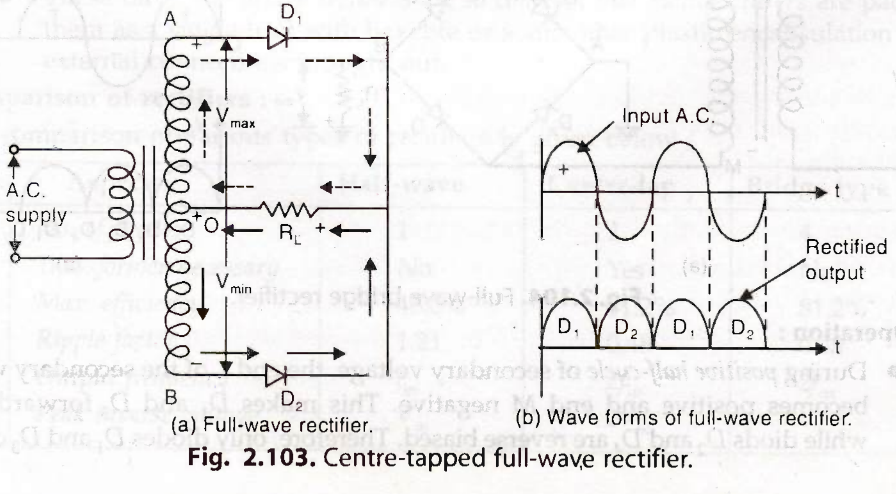


Fig. 2.103. Centre-tapped full-wave rectifier.

Operation :

- During the *positive half-cycle* of secondary voltage, the end A of the secondary winding is positive and end B negative. This makes the diode D_1 forward biased and diode D_2 reverse biased. Therefre, diode D_1 conducts while diode D_2 does not. The conventional current flows through diode D_1, load resistor R_L and the upper half of secondary winding as shown by the dotted arrows.
- During the *negative half-cycle,* the end A of the secondary becomes negative and end B positive. Therefore, D_2 conducts while diode D_1 does not. The conventional current flow is through D_2, R_L and lower half winding as shown by solid arrows.

It may be noted [Fig. 2.103 (*a*)] that the current in the load R_L is in the *same direction for both the cycles* of input A.C. voltage. Therefore, D.C. is obtained from the load R_L.

Also, Peak inverse voltage (PIV)

= Twice the maximum voltage across the half-secondary winding

i.e., $$PIV = 2\,V_{max.}$$

Advantages :

1. The D.C. output voltage and load current values are twice than those of half-wave rectifiers.
2. The ripple factor is much less (0.482) than that of half-wave rectifier (1.21).
3. The efficiency is twice that of half-wave rectifier.

For a full-wave rectifier, the maximum possible value of efficiency is 81.2% while that of half-wave rectifier is 40.6%.

Disadvantages :

1. The diodes used must have high peak inverse voltage.
2. It is difficult to locate the centre tap on the secondary winding.
3. The D.C. output is small as each diode utilises only one-half of transformer secondary voltage.

Full-wave bridge rectifier. It uses four diodes (D_1, D_2, D_3, D_4) across the main supply, as shown in Fig. 2.104 (*a*). The A.C. supply to be rectifier is appplied to the diagonally opposite ends of the bridge through the transformer. Between other two ends of the bridge, the load resistance R_L is connected.

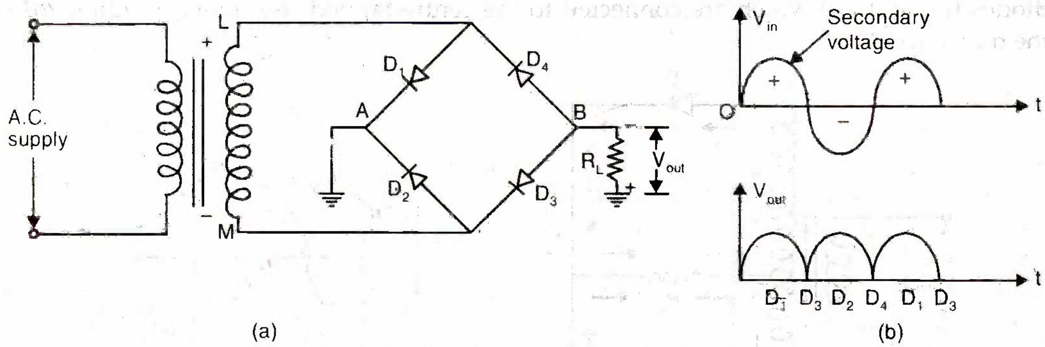


Fig. 2.104. Full-wave bridge rectifier.

Operation :

- During *positive half-cycle* of secondary voltage, the end L of the secondary winding becomes positive and end M negative. This makes D_1 and D_3 forward biased while diods D_2 and D_4 are reverse biased. Therefore, only diodes D_1 and D_3 conduct.

These two diodes will be in series through the load R_L as shown in Fig. 2.105 (*a*). The current flows (dotted arrows) from *A* to *B* through R_L.

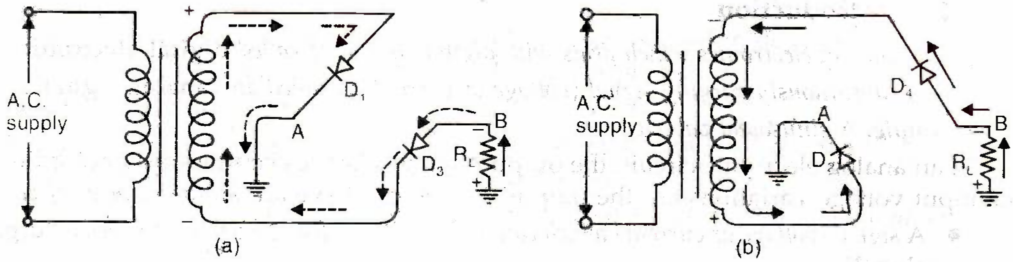


Fig. 2.105.

- During the *negative half-cycle* of the secondary voltage, end *L* becomes negative and *M* positive. This makes D_2 and D_4 forward biased whereas diodes D_1 and D_3 are reverse biased. Therefore, only diodes D_2 and D_4 conduct. These two diodes will be in series with R_L as shown in Fig. 2.105 (*b*). The current flows (*solid arrows*) from *A* to *B* through R_L *i.e.*, in the same direction as for positive half-cycle. Therefore, D.C. output is obtained across R_L.

Further it may noted that *peak inverse voltage (PIV) of each diode is equal to the maximum secondary voltage of transformer.*

Advantages :

1. It can be used with advantage in applications allowing floating input terminals *i.e.*, no output terminal is grounded.
2. The transformer is less costly as it is required to provide only half the voltage of an equivalent centre-tapped transformer used in a full-wave rectifier circuit.
3. No centre-tap is required on the transfomer.
4. The output is *twice* that of the centre-tapped circuit for the secondary voltage.

Disadvantages :

1. It uses four diodes as compared to two diodes for centre-tapped full wave rectifier.
2. Since during each half-cycle of A.C. input two diodes that conduct are in series, therefore, voltage drop in the internal resistance of the rectifying unit will be twice as great as in the centretapped circuit. This is objectionable when secondary voltage is small.
   - These days, the *bridge rectifiers are so common that manufacturers* are packing them as a single unit with bakelite or some other plastic encapsulation with external connections brought out.

Comparison of rectifiers :

The comparison of various types of rectifiers is given below :

S. No.	Aspects	Half-wave	Centre-tap	Bridge type
1.	*No. of diodes*	1	2	4
2.	*Transformer necessary*	No	Yes	No
3.	*Max. efficiency*	40.6%	81.2%	81.2%
4.	*Ripple factor*	1.21	0.48	0.48
5.	*Output frequency*	f_{in}	$2f_{in}$	$2f_{in}$
6.	*Peak inverse voltage*	V_m	$2V_m$	V_m

2.3. DIGITAL ELECTRONICS

2.3.1 Introduction

The *branch of electronics which deals with digital circuits is called* **digital electronics.**

- A *continuously varying signal (voltage or current) is called an* **"analog signal".**

Example. A *sinusoidal voltage.*

In an analog electronic circuit, the output voltage changes continuously according to the input voltage variations *i.e.,* the output voltage can have an *infinite number of values.*

- A *signal (voltage or current) which can have only two discrete values is called a* **"digital signal".**

Example. A *square wave.*

- *An electronic circuit that is designed for two-state operation is called a* **digital circuit.**

These days digital circuits are being used in many electronic products such as *video games, microwave ovens, oscilloscopes etc.*

2.3.2. Advantages and Disadvantages of Digital Electronics

The advantages and disdvantages of digital electronics are listed below :

Advantages :

1. Digital system can be normally *easily designed.*
2. Digital circuits are less *affected by noise.*
3. Storage of information is easy with digital circuits.
4. Digital circuits provide greater accuracy and precision.
5. More digital circuitry can be fabricated on integrated chips.

Disadvantages :

1. The digital circuits can handle only digital signals ; it requires encoders and decorders, due to which *cost of the equipment is increased.*
2. Under certain situations the use of only the analog techniques is simpler and economical (*e.g.* the process of signal amplification).

However, since the advantages outweigh the disadvantages, therefore, we are switching to digital techniques at a faster pace.

2.3.3. Digital Circuit

An electronic circuit that handles only a digital signal is called a **digital circuit.**

Or

An electronic circuit in which a state switches between the two states with time or with the change of the input states, and it is its state at the inputs and the outputs which has a significance is called a **digital circuit.**

"Digital" is derived from *"digitus".* In Latin, the latter means "finger". A finger is either up or down. Similarly an electronic circuit may have one of the states as :

(*i*) 'ON' (conduction) or 'OFF' (poor conduction), or

(*ii*) *'High'* voltage or *'Low'* voltage between two terminals, or

(*iii*) *'High'* current through a circuit or *'Low'* current through a circuit, or

(*iv*) *'High'* frequency signal or *'Low'* frequency signal, or

(*v*) 'Negative' potential difference or *'Positive'* potential difference, or

(*vi*) "1" or "0" etc.

Therefore, *digital circuit is one that expresses the values in digits 1's or 0's, hence the name "digital'*. The number concept that uses only the two digits 1 and 0 is the *binary numbering system.*

- As a digital circuit is based opon the two states, it is used in dealing with binary numbers ; digital circuit is therefore used in *computers.*

Advantages of digital circuit :

1. *Noise free* as output is measured in terms of its state, not in terms of a voltage, or a current, or a frequency. A *state has a difiniteness.*
2. *Capabilities of logical decision, arithmetic and Boolean operation on the binary numbers.*

Disadvantages :

1. *Slower speed* due to greater number of components to represent a state.
2. The circuits have *complexities* also. To represent a big decimal number, a *large number of components needed.*

Advantages of "Analog circuit" :

1. More close to physical system values.
2. A voltage level may represent temperature, wind, speed etc.

Disadvantages :

Lack of definiteness, preciseness and reliability.

2.3.4. Number Systems

In the field of digital electronics and computers, the number systems are used quite frequently. However, the type of number system used in computers could be different at different stages of the usage.

In digital circuits the following *four systems of arithmetic* are often used :

1. **Decimal.** It has a *base* (or radix) *of* **10** *i.e.,* it *uses 10 different sysmbols* to represent the number.
2. **Binary.** It has a *base of 2 i.e.,* it uses *only two different sysmbols.*
3. **Octal.** It has a base of **8** *i.e.,* it uses *eight different symbols.*
4. **Hexadecimal.** It has *base of* **16** *i.e.,* it uses *sixteen different symbols.*

All the above mentioned systems use the same type of ***positional notation*** except that :

— Decimal system uses *powers of* **10**

— Binary system uses *powers of* **2**

— Octal system uses *powers of* **8**

— Hexadecimal system uses *powers of* **16**

- ***Decimal numbers*** *are used to represent quantities which are outside the digital system.*
- ***Binary system*** is extensively used by *digital system like digital computers which operate on binary information.*
- ***Octal system*** has certain *advantages in digital work because it requires less circuitry to get information into and out of a digital system.* Moreover *it is easier to read, record and print out octal numbers than binary numbers.*
- ***Hexadecimal number system*** is *particularly suited for micro-computers.*

2.3.4.1. Decimal number system

The dicimal number system has a base of 10 and is a ***'position-value system'*** (meaning that value of digit depends on it ***position***). It has the following *characteristics* :

(*i*) **Base or radix.** It is *defined as the number of different digits which can occur in each position in the number system.*

The statement 'The decimal number system has a base of 10' implies that it contains ten unique symbols (or digits) *i.e.*, 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. Any one of these may be used in each position of the number. The ten digits do not limit us to express only ten different quantities because we use the various digits in appropriate positions within a number to indicate the magnitude of the quantity.

(*ii*) **Position value.** The absolute value of each digit is fixed but its *position value* (or *place value or weight*) is determined by its *position* in the overall number. For example, value of 4 in 4000 is not the same as in 400.

Consider the number 7654 (seven thousand six hundred and fifty four). The total value of this number is obtained by adding 4 unit values, 5 tens, 6 hundreds, and 7 thousands. Expressed more formally, it can be written as :

$$7654 = 7 \times 10^3 + 6 + 10^2 + 5 \times 10^1 + 4 \times 10^0$$

It will be noted that in this number, 4 is the ***least significant digit*** (LSD) whereas 7 is the ***most significant digit*** (MSD).

Again, the number 7654.358 can be written as

$$7654.358 = 7 \times 10^3 + 6 \times 10^2 + 5 \times 10^1 + 4 \times 10^0 + 3 \times 10^{-1} + 5 \times 10^{-2} + 8 \times 10^{-3}$$

It may be noted that *position values are found by raising the base to the number system (i.e., 10 in this case) to the power of the position.* Also *powers are numbered to the left of the decimal point starting with zero and to the right of the decimal point with –1.*

2.3.4.2. Binary number system

The binary number system, like decimal number (or denary) system, has a radix and uses the same type of position value system.

(*i*) ***Radix.*** The base or radix of the system is **2** because it uses only two digits 0 and 1 (the word 'binary digit' is contracted to **bit**).

All binary numbers consists of strings of 0s and 1s.

Examples. 10, 101 and 1011-reads one-zero-one-one respectively (to avoid confusion with decimal numbers).

Confusion can also be avoided by adding a subscript of 10 for decimal numbers and 2 for binary numbers as mentioned below :

10_{10}, 101_{10}, 6785_{10} .... Decimal numbers

10_2, 101_2, 110001_2 .... Binary numbers.

- *Binary numbers need more places for counting because their base is small.*

(*ii*) ***Position value.*** The binary system, like the decimal system, is also positionally-weighted. In this case, however, the position value of each bit corresponds to some power of 2. In each binary number, the value *increases* in powers of 2 starting with 0 to the *left* of the binary point and *decreases* to the *right* of the binary point starting with power of –1.

The decimal equivalent of the binary number may be found as under :

$$1101.011_2 = (1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0) + (0 \times 2^{-1}) + (1 \times 2^{-2}) + (1 \times 2^{-3})$$

$$= 8 + 4 + 0 + 1 + 0 + \frac{1}{4} + \frac{1}{8} = 13.375_{10}$$

2.3.4.3. Binary-to-decimal conversion

In order to convert a given binary integer (whole number) into its equivalent decimal number, the following *four* **steps** are involved.

Step 1. Write the binary number *i.e.*, all its bits in *a row.*

Step 2. Directly under the bits, write $1(2^0)$, $2(2^1)$, $4(2^2)$, $8(2^3)$, $16(2^4)$, ... starting from *right to left.*

Step 3. *Cross out the decimal weights which lies under 0 bits.*

Step 4. Add the remaining weights to get the decimal equivalent.

Example 2.21. *Convert 10011 to its equivalent decimal number.*

Solution. The *four steps* involved in conversion are:

Step 1.	1	0	0	1	1
Step 2.	16	8	4	2	1
Step 3.	16	$\cancel{8}$	$\cancel{4}$	2	1

Step 4. $16 + 2 + 1 = 19$

$\therefore$ $10011_2 = \mathbf{19_{10}}$. **(Ans.)**

It is seen that number contains 1 sixteen, 0 eight, 0 four's, 1 two's and 1 one.

Table 1 shows the equivalent binary numbers of decimal numbers.

Table 1. Equivalnent binary numbers of decimal numbers

Decimal	Binary	Decimal	Binary	Decimal	Binary
1	1	11	1011	21	10101
2	10	12	1100	22	10110
3	11	13	1101	23	10111
4	100	14	1110	24	11000
5	101	15	1111	25	11001
6	110	16	10000	26	11010
7	111	17	10001	27	11011
8	1000	18	10010	28	11100
9	1001	19	10011	29	11101
10	1010	20	10100	30	11110

- In binary number system, some terms like *bit, nibble* and *byte* are used.

— **Bit** is used for a *single binary digit.*

— **Nibble** is a binary number with *four bits.*

— **Byte** is a binary number with *eight bits.*

Binary fractions. The procedure is same as for binary integers except that the following weights are used for different bit positions :

$2^n \ldots 2^2$ $\quad 2^1 \quad 2^0 \quad \bullet \quad 2^{-1} \quad 2^{-2} \quad 2^{-3} \quad 2^{-4} \longrightarrow$

$\uparrow$ Binary point $\quad \frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{16} \longrightarrow$

Example 2.22. *Convert the binary fraction 0.101 into its decimal equivalent.*

Solution. The following *four steps* will be used :

Step 1. 0 1 0 1

Step 2. $\frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8}$

Step 3. $\frac{1}{2} \quad \cancel{\frac{1}{4}} \quad \frac{1}{8}$

Step 4. $\frac{1}{2} + \frac{1}{8} = 0.625$

$$0.101_2 = \mathbf{0.625_{10}} \text{ (Ans.)}$$

2.3.4.4. Decimal-to-binary coversion

A decimal-to-binary conversion can be achieved by using the so-called ***double-dabble method.*** It is also known as ***divide-by-two*** method.

(*a*) **Integers.** In this case, we progressively *divide* the given dicimal number by 2 and write down the remainders after each division. These remainders taken in the **reverse** *order* (*i.e.*, from *bottom-to-top*) form the required number.

Example 2.23. *Convert 19_{10} into its binary equivalent.*

Solution.

$$\begin{aligned} 19 \div 2 &= 9 + \text{remainder of } 1 \\ 9 \div 2 &= 4 + \text{remainder of } 1 \\ 4 \div 2 &= 2 + \text{remainder of } 0 \\ 2 \div 2 &= 1 + \text{remainder of } 0 \\ 1 \div 2 &= 0 + \text{remainder of } 1 \end{aligned}$$

Top

Bottom

$\therefore \quad 19_{10} = \mathbf{10011}$ **(Ans.)**

The above process may be simplified as under

2	19
2	9 – 1
2	4 – 1
2	2 – 0
2	1 – 0
	0 – 1

Reading the remainders from bottom to top, we get : $19_{10} = 10011$.

(*b*) **Fractions.** In this, ***Multiply-by-two*** rule is used ***i.e.***, we *multiply* each bit by 2 and record the carry in the integer form. These carries taken in the ***forward*** (*top-to-bottom*) direction give the required binary fraction.

Example 2.24. *Convert 0.65_{10} into its binary equivalent.*

Solution.

$$\begin{aligned} 0.65 \times 2 &= 1.3 = 0.3 \text{ with a carry of } 1 \\ 0.3 \times 2 &= 0.6 = 0.6 \text{ with a carry of } 0 \\ 0.6 \times 2 &= 1.2 = 0.2 \text{ with a carry of } 1 \\ 0.2 \times 2 &= 0.4 = 0.4 \text{ with a carry of } 0 \\ 0.4 \times 2 &= 0.8 = 0.8 \text{ with a carry of } 0 \\ 0.8 \times 2 &= 1.6 = 0.6 \text{ with a carry of } 1 \\ 0.6 \times 2 &= 1.2 = 0.2 \text{ with a carry of } 1 \\ 0.2 \times 2 &= 0.4 = 0.4 \text{ with a carry of } 0 \end{aligned}$$

$0.65_{10} = \mathbf{0.10100110_2}$ **(Ans.)**

Example 2.25. *Convert the following decimal number into binary : 12.0625.*

Solution. We shall carry out the conversion in two steps, (*i*) First for *integer* and (*ii*) then for *fraction.*

(*a*) **Integer**

2	12
2	6 – 0
2	3 – 0
2	1 – 1
2	0 – 1

(*b*) **Fraction**

$0.0625 \times 2 = 0.125$ with a carry of 0

$0.125 \times 2 = 0.25$ with a carry of 0

$0.25 \times 2 = 0.5$ with a carry of 0

$0.5 \times 2 = 1.0$ with a carry of 1

$\therefore \quad 12_{10} = 1100_2$ $\qquad \therefore \quad 0.0625_{10} = 0.0001_2$

Considering the complete number, we have : 12.0625_{10} = $\mathbf{1100.0001_2}$ **(Ans.)**

Example 2.26. *Convert 25.625_{10} into its binary equivalent.*

Solution. **Integer**

2	25
2	12 – 1
2	6 – 0
2	3 – 0
2	1 – 1
	0 – 1

Fraction

$0.625 \times 2 = 1.25 = 0.25 + 1$

$0.25 \times 2 = 0.5 = 0.5 + 0$

$0.5 \times 2 = 1.0 = 0.0 + 1$

$\therefore \quad 25_{10} = 11001_2$ $\qquad \therefore \quad 0.625_{10} = 0.101_2$

Considering the complete number, we have 25.625_{10} = $\mathbf{11001.101_2}$ **(Ans.)**

2.3.4.5. Binary Operations

In a decimal number system, we are familiar with the arithmetic operations such as addition, subtraction, multiplication and division. Similar operations can be performed on binary numbers, infact, binary arithmetic is much simpler than decimal arithmetic because here only two digits, 0 and 1 are involved.

The *addition,* in binary number system, is *the most important* of the four operation of addition, subtraction, multiplication and division.By using *'complements'*, subtraction can be reduced to addition. Most digital computers subtract by complements. *It leads to reduction in hardware because circuitry is required only for addition operation.* Similarly, *multiplication is nothing but repeated addition and, finally division is nothing but repeated subtraction.*

(*i*) *Binary addition* :

There are four rules/cases, described below for *addition* of binary numbers :

(1) $0 + 0 = 0$

(2) $0 + 1 = 1$

(3) $1 + 0 = 1$

(4) $1 + 1 = 10$ (This sum is *not 'ten'* but *'one-zero'*)

Example 2.27. *Add 110011_2 to 101101_2.*

Solution.

$$\begin{array}{r} 110011 \\ 101101 \\ \hline 1100000 \end{array}$$

1st column : 1 + 1 = 0 with a carry of 1

2nd column : 1 + 0 = 1 combined with carry 1 = 0 with carry 1

3rd coumn : 0 + 1 = 1 combined with carry 1 = 0 with carry 1

4th column : 0 + 1 = 1 combined with carry 1 = 0 with carry 1

5th column : 1 + 0 = 1 combined with carry 1 = 0 with carry 1

6th coumn : 1 + 1 = carry of 1 = $\mathbf{11_2}$ **(Ans.)**

(ii) Binary subtraction :

The *four rules* for binary subtraction are :

1. 0 – 0 = 0
2. 1 – 0 = 1
3. 1 – 1 = 0
4. 10 – 1 = 1

Example 2.28. *Subtract 0111_2 from 1001_2.*

Solution.

$$\begin{array}{r} 1001 \\ -\ 0111 \\ \hline 0010 \end{array}$$

1st column : 1 – 1 = 0

2nd column : 0 – 1 = 1 with a borrow of 1

3rd column : 1 (after borrow) – 1 = 0

4th column : 0 (after borrow) – 0 = 0.

(iii) Binary multiplication :

The *four rules* are :

1. 0 × 0 = 0
2. 0 × 1 = 0
3. 1 × 0 = 0
4. 1 × 1 = 1

Example 2.29. *Multiply 111_2 by 101_2 using binary multiplication method.*

Solution.

$$\begin{array}{rl} 111 & \\ \times\ 101 & \\ \hline 111 & \\ 000 & \text{.....shift left no add} \\ 111 & \text{.....shift left and add} \\ \hline 100011 & \textbf{(Ans.)} \end{array}$$

Example 2.30. *Multiply 11.01_2 by 10.11_2.*

Solution.

$$\begin{array}{r} 11.01 \\ 10.11 \\ \hline 1101 \\ 1101 \\ 0000 \\ 1101 \\ \hline 1000.1111 \end{array} \quad \textbf{(Ans.)}$$

Example 2.31. *Multiply* $(10001.101)_2 \times (111.001)_2$.

Solution.

$$
\begin{array}{r}
10001.101 \\
\underline{111.001} \\
10001101 \\
00000000 \\
00000000 \\
10001101 \\
10001101 \\
\underline{10001101} \\
\underline{1111101.100101}
\end{array}
\quad \textbf{(Ans.)}
$$

(iv) Binary division :

The rules of binary division are :

1. $0 \div 1 = 0$ or $\frac{0}{1} = 0$
2. $1 \div 1 = 1$ or $\frac{1}{1} = 1$.

Example 2.32. *Divide 1110101 by 1001.*

Solution.

$$
\begin{array}{r}
1001\,)\,1110101\,(\,1101 \\
\underline{1001} \\
1011 \\
\underline{1001} \\
1001 \\
\underline{1001} \\
\times
\end{array}
$$

Shifting a number to left or right :

Shifting binary numbers one step to the left or right corresponds respectively to multiplication or division by decimal 2.

Example. When binary number 11100_2 corresponding to decimal 28 is shifted one step *to the left*, it becomes 111000 which corresponds to decimal number 56 *i.e.*, it is *doubled*. If the given number is shifted one step *to the right* it becomes 1110 which corresponds to decimal number 14, *i.e.*, it is *halved*.

Shifting the place point :

As in a decimal system, moving of a dicimal point from one place to the right or left corresponds respectively to multiplication or division by 10, similarly shift of the binary point by one place to the right or left *multiplies or divides by 2*.

Example. 1011.0_2 corresponds to 11_{10} but 10110_2 corresponds to 22_{10} while 101.1_2 corresponds to 5.5_{10}.

Complement of a number :

In digital work, two types of complements of a binary number are used for *complemental subtraction :*

(*i*) **1's complement.** The 1's complement of a binary number is obtained by changing its each 0 *into a 1 and each 1 into a 0.* It is also called *'radix-minus-one'* complement.

Example. 1's complement of 100_2 is 011_2 and of 1110_2 is 0001_2.

(*ii*) **2's complement.** The 2's complement of a binary number is obtained by *adding 1 to its 1's complement.*

2's complement = 1's complement + 1

It is also known as **true** complement.

Example. *2's complement of 1011_2 is found as follows :*

— 1's component of 1011 is 0100.

— Next adding 1 we get 2's complement or 0101_2.

Hence 2's complement of 1011_2 is 0101_2

The *complement method of subtraction reduces* **subtraction to an addition process.**

This method is popular in digital computers because of the following reasons :

1. With digital circuits, is is easy *to get the complements.*
2. Only adder circuits are needed, thus *circuitry is simplified.*

1's complemental subtraction :

In this method, instead of subtracting a number, we *add its 1's complement to the minuend.* The last carry (whether 0 or 1) is then added to get the final answer.

The steps for subtraction by 1's complement are as under :

Step 1. Compute the 1's complement of the subtrahend by changing all its 1's to 0's and all its 0s to 1s.

Step 2. Add this complement to the minuend.

Step 3. Perform the end-around carry of the last 1 or 0.

Step 4. If there is no end-around carry (*i.e.*, 0 carry), then the answer must be recomplemented and negative sign attached to it.

Step 5. If the end-around carry is 1, no recomplementing is necessary.

Example 2.33. *Subtract 101_2 from 111_2.*

Solution.

```
   111
 + 010     ←— 1's complement of subtrahend (i.e., 101₂)
  ----
  1001
     1     ←— end-round carry
  ----
   010
```

Since end-around carry is 1, the *final* **answer** (step 5) is **010.**

Example 2.34. *Subtract 1101_2 from 1010.*

Solution

```
   1010
   0010    ←— 1's complement of 1101
   ----
   1100
 No carry
```

Since there is no end-around carry in this case, therefore, answer must be recomplemented (step-5) to get 0011 and a negative sign attached to it.

∴ *Final* **answer** is : **– 0011.**

Example 2.35. *Using 1's complement method, subtract 01101_2 from 11011_2.*

Solution.

```
   11011
 + 10010   ←— 1's complement of subtrahend (i.e., 01101₂)
  ------
  101101
       1   ←— end-around carry
  ------
    1110
```

Since end-around carry is 1, the *final* **answer** is **1110.**

2's Complemental subtraction :

The steps for subtraction by 2's complement are as under :

Step 1. Find the 2's complement of the subtrahend.

Step 2. Add this complement to the minuend.

Step 3. Drop the final carry.

Step 4. If the carry is 1, the answer is positive and needs no recomplementing.

Step 5. If there is no carry, recomplement the answer and attach minus sign.

Example 2.36. *Using 2's complement subtract 1010_2 from 1101_2.*

Solution. The 1's complement of 1010 is 0101. The 2's complement is 0101 + 1 = 0110. We will add it to 1101

$$\begin{array}{r} 1101 \\ +\ 0110 \\ \hline 10011 \\ \hline \end{array} \quad \longleftarrow \text{2's complement}$$

Since the carry is 1, the answer is positive and needs no recomplementing (step-4), therefore the *final* **answer** is **0011_2.**

Example 17. *Using 2's complement subtract 1101_2 from 1010_2.*

Solution. The 1's complement of 1101 is 0010. The 2's complement is 0010 + 1 = 0011.

$$\begin{array}{r} 1010 \\ +\ 0011 \\ \hline 1101 \\ \hline \end{array} \quad \longleftarrow \text{2's complement of } 1101_2$$

In this case there is *no carry,* hence we have to recomplement the answer. For this purpose, *we first subtract 1 from it to get 1100.*

Next we *complement it* to get 0011. After *attaching the minus sign,* the *final* **answer** becomes – **0011_2.**

(Taking in terms of decimal numbers, we have subtracted 13 from 10 *i.e.,* 10 – 13 = – 3).

2.3.4.6. Octal number system

The *number system with base (or radix) "eight" is known as the* ***octal number system.*** The octal number system entails the following ***merits.***

1. In digital systems, it is highly *inconvenient to handle long strings of binary numbers.* The octal number system requires *one-third in length* as compared to binary numbers. Thus from users' point of view it would be comparatively much *easier to handle the input and output data of a digital computer in octal form.*
2. The *print-outs are more compact and easy to read.*
3. *Conversion* from binary-to-octal and octal-to-binary is *quick and simple.*

- Since digital circuits can process only zeros and ones, the *octal numbers have to be converted into binary form* employing *special circuits known as octal-to-binary converters* before being processed by the digital circuits.

(*i*) **Radix a base.** It has radix or base of 8 which means that it has eight distinct counting digits :

0, 1, 2, 3, 4, 5, 6, 7

These digits 0 through 7, have *exactly the same physical meaning as in decimal system.*

For counting beyond 7, 2 *digit combinations are formed taking the second digit followed by the first, the second digit followed by the second and so on.* Hence after, 7, the next octal nuumber is 10 (second digit followed by the first), 11 (second digit followed by second) and so on. Hence different octal numbers are :

0, 1, 2, 3, 4, 5, 6, 7,

10, 11, 12, 13, 14, 15, 16, 17,
20, 21, 22, ...

(*ii*) **Position value.** The position value or (or weight) for each digit is given by different powers of 8 as shown below.

$$\longleftarrow \quad 8^3 \quad 8^2 \quad 8^1 \quad 8^0 \quad 8^{-1} \quad \bullet \quad 8^{-1} \quad 8^{-2} \quad 8^{-3} \longrightarrow$$

↑
Octal point

For example, decimal equivalent of octal 314 is

3	1	4	0
8^2	8^1	8^0	
64	8	1	$= 3 \times 64 + 1 \times 8 + 4 = 204_{10}$

or,
$$314_8 = 3 \times 8^2 + 1 \times 8^1 + 4 \times 8^0 = 192 + 8 + 4 = 204_{10}$$

Similarly decimal equivalent of 127.24 is

$$127.24 = 1 \times 8^2 + 2 \times 8^1 + 7 \times 8^0 + 2 \times 8^{-1} + 4 \times 8^{-2}$$
$$= 64 + 16 + 7 + \frac{2}{8} + \frac{4}{64} = 87.3125_{10}$$

2.3.4.7. Octal-to-decimal conversion

An octal number can be easily converted to its decimal equivalent by *multiplying each octal digit by its positional weight.*

Example 2.38. *Convert 206.104 into its decimal equivalent number.*

Solution.

2	0	6	1	0	4
8^2	8^1	8^0	8^{-1}	8^{-2}	8^{-3}

$$\therefore \quad 206.104_8 = 2 \times 8^2 + 0 \times 8^1 + 6 \times 8^0 + 1 \times 8^{-1} + 0 \times 8^{-2} + 4 \times 8^{-3}$$
$$= 128 + 6 + \frac{1}{8} + \frac{1}{128} = \left(\mathbf{134\frac{17}{128}}\right)_{10} \quad \textbf{(Ans.)}$$

2.3.4.8. Decimal-to-octal coversion

A decimal integer can be converted to octal by using the same repeated-division method called the *double-dabble method,* that was used in the decimal-to-binary conversion, but with a *division factor of 8 rather than 2.*

Example 2.39. *Convert 1375_{10} into its octal equivalent.*

Solution.

8	1375
8	171 - 7
8	21 – 3
8	2 – 5
	0 – 2

Taking the remainders in the *reverse order,* we have,

Equivalent octal number of 1375_{10} = $\mathbf{2537_8}$ **(Ans.)**

(Note that first remainder becômes the least significant digit (LSD) of the total number, and the last remainder becomes the most significant digit (MSD).

Example 2.40. *What is octal equivalent of 0.15_{10}?*

Solution.
$$0.15 \times 8 = 1.20 = 0.20 \text{ with a carry of } 1$$
$$0.20 \times 8 = 1.60 = 0.60 \text{ with a carry of } 1$$
$$0.60 \times 8 = 4.80 = 0.80 \text{ with a carry of } 4 \text{ etc.}$$
$$\therefore \quad 0.15_{10} \simeq \mathbf{0.114_8} \quad \textbf{(Ans.)}$$

(Here carries have been taken in the forward direction *i.e.,* from top to bottom).

Example 2.41. *Find the octal equivalent of the decimal fraction 0.685.*

Solution.

$0.685 \times 8 = 5.48 = 0.48$ with a carry of 5
$0.48 \times 8 = 3.84 = 0.84$ with a carry of 3
$0.84 \times 8 = 6.72 = 0.72$ with a carry of 6
$0.72 \times 8 = 5.76 = 0.76$ with a carry of 5
$0.76 \times 8 = 6.08 = 0.08$ with a carry of 6

The carries read in the *forward direction i.e., from top to bottom* give the octal fraction 0.53656 *i.e,* 0.685_{10} = $\mathbf{0.53656_8}$ **(Ans.)**

2.3.4.9. Octal-to-binary conversion

Since 8 (the base of octal numbers) is third power of 2 (the base of binary number), the conversion from octal to binary can be performed by *converting each octal digit to its 3-bit binary equivalent.* The eight possible digits are converted as indicated in the table 2 below :

Table 2

Octal digit	0	1	2	3	4	5	6	7
Binary equivalent	000	001	010	011	100	101	110	111

Using these coversions, any octal number can be converted to binary by individually converting each digit.

Example 2.42. *Convert 4161_8 into binary.*

Solution.

4	1	6	1
↓	↓	↓	↓
100	001	110	001

Hence 4161_8 = $\mathbf{(100\ 001\ 110\ 001)_2}$ **(Ans.)**

Example 2.43. *Convert 37.13_8 into binary.*

Solution.

3	7	.	1	3
011	111	.	001	011

Hence 37.13_8 = $\mathbf{(001\ 111.001\ 011)_2}$ **(Ans.)**

Using positional notation, the first few octal numbers and their decimal equivalents are shown in the table 3 below :

Table 3

Octal	Decimal	Octal	Decimal	Octal	Decimal
0	0	12	10	24	20
1	1	11	11	25	21
2	2	14	12	26	22
3	3	15	13	27	23
4	4	16	14	30	24
5	5	17	15	31	25
6	6	20	16	32	26
7	7	21	17	33	27
10	8	22	18	34	28
11	9	23	19	35	29

2.3.4.10. Binary-to-octal conversion

The conversion of a binary number to octal number is simply the reverse of the foregoing process. The bits of the binary number are grouped into groups of three bits starting at the LSB (least significant bit). Then each group is converted to its octal equivalent.

Example 2.44. *Convert the binary number 101011_2 to its octal equivalent.*

Solution.

$$\begin{array}{ccccc} 101011_2 & \rightarrow & 101 & & 011 \\ & & \downarrow & & \downarrow \\ & & 5 & & 3 \end{array}$$

$\therefore$ $101\ 011_2 = \mathbf{53_8}$ **(Ans.)**

Example 2.45. *Convert binary number 10101.11_2 into its octal equivalent.*

Soution. Here we will have to add one 0 infront of the integer part as well as to the fractional part

$$\begin{array}{cccccc} 10101.11_2 & \rightarrow & 010 & 101 & \bullet & 110 \\ & & \downarrow & \downarrow & \bullet & \downarrow \\ & & 2 & 5 & \bullet & 6 \end{array}$$

$\therefore$ $10101.11_2 = \mathbf{25.6_8}$ **(Ans.)**

Example 2.46. *Convert the binary number 11011100.101010_2 to octal equipment.*

Solution.

$$\begin{array}{cccccccc} 11011100.101010 & \rightarrow & 011 & 011 & 100 & \bullet & 101 & 010 \\ & & \downarrow & \downarrow & \downarrow & & \downarrow & \downarrow \\ & & 3 & 3 & 4 & \bullet & 5 & 2 \end{array}$$

i.e., $11011100.101010_2 = \mathbf{334.52_8}$ **(Ans.)**

Example 2.47. *Perform $1766_8 - 23_8$.*

Solution.

$$\begin{array}{rcrrrr} 1766_8 & = & 001 & 111 & 110 & 110_2 \\ 23_8 & = & 010 & 011_2 & & \\ \therefore \quad 1766_8 - 23_8 & = & 001 & 111 & 110 & 110 \\ & & - & & 010 & 011 \\ \hline & & 001 & 111 & 100 & 011 \\ & & \downarrow & \downarrow & \downarrow & \downarrow \\ & & 1 & 7 & 4 & 3 \end{array}$$

i.e., $1766_8 - 23_8 = \mathbf{1743_8}$ **(Ans.)**

2.3.4.11. Hexadecimal number system

For the two-state systems, the binary number system forms the natural choice. But in hexadecimal number system, the numbers *tend to get short rather long.* Hence *to reduce the length of a given number it is quite common to use hexadecimal system.*

- The chief use of this system is in connection with ***byte-organised machines.***
- It is used for *specifying addresses* of different binary numbers stored in computer memory.
- This system is extensively used in *microprocessor work.*

This system has the following *characteristics* :

1. It has *base of* **16.** Hence it uses sixteen distinct counting digits 0 through 9 and A through *F* as detailed below :

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, *A*, *B*, *C*, *D*, *E*, *F*.

2. The *place value* (or weight) for each digit is in *'ascending powers of 16' for integers* and *'descending powers of 16' for fractions.*

This system is an *alphanumeric system* since numeric digits and alphabets both are used to represent the digits.

Table 4 shows the *relationship between hexadecimal, decimal and binary.*

Table 4. Decimal and Binary Equivalents of Hexadecimal Number

Hexadecimal	decimal	Binary
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111
8	8	1000
9	9	1001
A	10	1010
B	11	1011
C	12	1100
D	13	1101
E	14	1110
F	15	1111

Counting beyond F in Hex number system :

A usual, we resort to "*2-digit combinations*". After reaching *F*, we take the *second digit followed by the first digit, the second followed by second, then second followed* by *third and so on,* as mentioned below :

10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 1*A*, 1*B*, 1*C*, 1*D*, 1*E*, 1*F*

20, 21, 22, 23 , 24, 25, 26, 27, 28, 29, 2*A*, 2*B*, 2*C*, 2*D*, 2*E*, 2*F*

30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 3*A*, 3*B*, ..., ..., ..., ...

— With two hexadecimal digits, we can count upto FF_{16} which is equal to 255_{10}.

— For counting beyond this, three hexadecimal digits are required

Example. $100_{16} = 256_{10}$, $101_{16} = 257_{10}$ and so on.

— The maximum three-digit hexadecimal number is FFF_{16} which is equal to 4095_{10}.

2.3.4.12. Hexadecimal-to-decimal conversion

A hexadecimal number can be converted to its decimal equivalent by *multiplying each hexdigit by its weight and then taking the sum of these products.* The weights of a hex number are increasing powers of 16 (from right to left).

For a *four-digit hex number* the weights are as follows :

16^3	16^2	16^1	16^0
4096	256	16	1

Example 2.48. *Convert* $F6D9_{16}$ *into decimal equivalent.*

Solution. $$F6D9_{16} = F(16^3) + 6(16)^2 + D(16)^1 + 9(16)^0$$

$$= 15 \times 16^3 + 6 \times 16^2 + 13 \times 16^1 + 9 \times 16^0$$
$$= 61440 + 1536 + 208 + 9 = \mathbf{63193_{10}} \quad \textbf{(Ans.)}$$

Example 2.49. *Convert 2B.1FA into decimal equivalent.*

Solution. $2B.1FA_{16} = 2 \times 16^1 + 11 \times 16^0 + 1 \times 16^{-1} + 15 \times 16^{-2} + 10 \times 16^{-3}$

$$= 32 + 11 + \frac{1}{16} + \frac{15}{256} + \frac{10}{4096}$$
$$= \mathbf{43.123535156_{10}} \quad \textbf{(Ans.)}$$

2.3.4.13. Decimal-to-hexadecimal conversion

Repeated division of a decimal number by 16 will produce the equivalent hex number formed by the remainder of each division. This is similar to the repeated division by 2 for decimal-to-binary conversion and repeated division by 8 for decimal-to-octal conversion.

Example 2.50. *Convert 1983_{10} into hexadecimal.*

Solution.

16	1983	
16	123 – 15	*F*
16	7 – 11	*B*
	0 – 7	7

Hence 1983_{10} = $\mathbf{7BF_{16}}$ **(Ans.)**

Example 2.51. *Convert decimal number 374.37 to hexadecimal.*

Solution. 1. *Integer 374 :*

16	374
16	23 – 6
16	1 – 7
	0 – 1

∴ Equivalent hex number of $374_{10} = 176_{16}$.

2. *Fraction 0.37 :*

$0.37 \times 16 = 5.92 = 0.92$ with a carry of 5
$0.92 \times 16 = 14.72 = 0.72$ with a carry of 14
$0.72 \times 16 = 11.52 = 0.52$ with a carry of 11
$0.52 \times 16 = 8.32 = 0.32$ with a carry of 8

∴ Equivalent hex number of 0.37 = 0.5*EB*8

Hence 374.37_{10} = $\mathbf{176.5EB8_{16}}$ **(Ans.)**

2.3.4.14. Hexadecimal-to-binary conversion

Hex numbers can be converted into equivalent binary number by *replacing each hex digit by its equivalent 4-bit binary number.*

Example 2.52. *Convert $23A_{16}$ into its binary equivalent.*

Solution.

2	4	*A*
↓	↓	↓
0010	0011	1010

∴ $23A_{16}$ = $\mathbf{0010\ 0011\ 1010_2}$ **(Ans).**

Example 2.53. *Convert 524.36_{16} into its binary equivalent.*

Solution.

5	2	4	3	6
↓	↓	↓	↓	↓
0101	0010	0100	0011	0110

Hence 524.36_{16} = $\mathbf{0101\ 0010\ 0100\ 0011\ 0110_2}$ **(Ans.)**

2.3.4.15. Binary-to-hexadecimal conversion

Conversion from binary to hex is first the reverse of the process discussed in Art. 2.3.14.14. The binary number is grouped into groups of 4-bits starting from LSB and moving toward MSB for *"integer part"* and then each group of four bits is replaced by its hex representation. *Zeros are added, as required to complete a 4-bit group.* For the *"fractional part"*, the above procedure is repeated from *the bit next to the binary point and moving towards the right.*

Example 2.54. *Convert 1011010111_2 to hexadecimal.*

Solution. $1011010111_2 \rightarrow$

0010	1101	0111
↓	↓	↓
2	D	7

$\therefore$ $1011010111_2 = \mathbf{2D7_{16}}$ **(Ans.)**

It may be noted that two 0s have between added to complete the 4-bit groups.

Example 2.55. *Convert 1001011010101_2 to hexadecimal.*

Solution. $1001011010101_2 \rightarrow$

0001	0010	1101	0101
↓	↓	↓	↓
1	2	D	5

$\therefore$ $1001011010101_2 = \mathbf{12D5_{16}}$ **(Ans.)**

Example 2.56. *Convert 1010.0111 to hexadecimal.*

Solution. 1010.0111

1010	0111
↓	↓
A	7

$\therefore$ $1010.0111_2 = \mathbf{A.7_{16}}$ **(Ans.)**

2.3.4.16. Conversion from Hex-to-octal and vice-versa

Hexadecimal numbers can be converted to equivalent octal numbers and octal numbers can be *converted to equivalent hex numbers by converting the hex/octal number to equivalent binary and then to octal/hex respectively.* The procedure is illustrated in the following examples.

Example 2.57. *Convert 1375_8 = ... $_2$ = ... $_{16}$.*

Solution. 1375 →

1	3	7	5
001	011	111	101

i.e., $1375_8 = 001\ 011\ 111\ 101_2$. **(Ans.)**

Now, 001011111101 →

0010	1111	1101
↓	↓	↓
2	F	D

i.e., $1375_8 = 001011111101_2 = \mathbf{2FD_{16}}$ **(Ans.)**

Example 2.58. *Convert ABCD hexadecimal number to octal through binary.*

Solution. $ABCD_{16} \rightarrow$

A	B	C	D
↓	↓	↓	↓
1010	1011	1100	1101

→

001	010	101	111	001	101
1	2	5	7	1	5

$\therefore$ $ABCD_{16} = \mathbf{125715_8}$ **(Ans.)**

Example 2.59. *Perform the operation : $A.5936_{16} - B.3158_{16}$.*

Solution.

$A.5936_{16}$ =	1010.0101	1001	0011	0110	
→	1010.0101	1001	0011	0110	
− $B.3158_{16}$ =	− 1011.0011	0001	0101	1000	
→	+ 0100.1100	1110	1010	0111	2's complement of B 3158
	1111.0010	0111	1101	1101	No carry, 2's complement of result
	− 0000.1101	1000	0010	0010	
	↓ ↓	↓	↓	↓	
	− 0 D	8	2	2	

i.e., $A.5936_{16} - B.3158_{16} = -\mathbf{0.D822_{16}}$ **(Ans.)**

2.3.5. Digital Coding

In digital circuits, each number of piece of information is *defined by an equivalent combination of binary digits. A complete group of these combinations which represent numbers, letters or symbols is called a* ***digital code.***

The group of 0s and 1s in the binary number can be thought of as a code representing the decimal numbers. When a decimal number is represented by its equivalent binary number, it is called a *straight binary coding.*

In modern digital equipment, *codes are used to represent and process numerical information.*

Types of codes. The various types of codes are enumerated and briefly discussed below :

1. **BCD Code**

It is also known as 'natural BCD' and is very convenient for representing decimal digits in digital circuits.

- It consists of four bits from 0000 to 1001 representing the decimal numbers from 0 to 9.1010 to 1111 are don't care conditions since they do not have any meaning in BCD.

2. **Excess-3 Code**

- The code can be derived from BCD by *adding 3 to each coded number.*
  It is useful when it is desired to obtain the 9's complement of a decimal digit represented by this code. The 9's complement is obtained simply by complementing each bit.
- This code can be *conveniently used for performing substracting operations in digital computers.*

3. **Gray Code**

- In this code ony one bit changes between any two successive numbers.
- It is mainly *used in the location of angular positions of a rotating shaft.*

4. **Octal Code**

- The octal system is a 8 base system.
- It uses 3 bits to represent one octal digit.

5. **Hexadecimal Code**

- The hexadecimal system is a base 16 system.
- It uses four bits to represent one hexadecimal digit.
- The hexadecimal digits are represented as 0 to 9 continued by aphabetical characters from A to F.

2.3.6. Logic Gates

General aspects :

A *digital circuit with one or more input signals but only output signal is called a* **logic gate.**

A ***logic gate*** *is an electronic circuit which makes logic decision.*

- Logic gates are the *basic building blocks* from which most of the digital systems are built up. They implement the hardware logic function based on the logical algebra developed by George Boolean which is called *Boolean algebra* in his honour.
  - — A unique characteristic of Boolean algebra is that variables used in it can assume only one of the two values *i.e., either 0 or 1.* Hence, every variable is either a 0 or a 1 (Fig. 2.106–limits on TTLIC's).
- Each gate has *distinct graphic symbol and its operation can be described by means of Boolean algebraic function.*

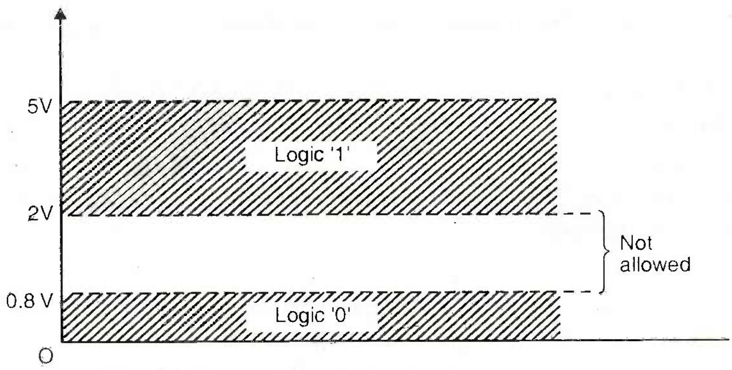


Fig. 2.106. Voltage assignment in a digital system.

- The *table which indicates output of gate for all possible combinations of input is known as a* ***truth table.***
- These gates are available today in the form of various IC families. The *most popular families are :*

(*i*) Transistor-transistor logic (TTL)
(*ii*) Emitter-coupled logic (ECL)
(*iii*) Metal-oxide-semiconductor (MOS)
(*iv*) Complementary metal-oxide-semiconductor (CMOS).

Applications of logic gates :

The following are the *fields of application of logic gates :*

1. Calculators and computers.
2. Digital measuring techniques.
3. Digital processing of communications.
4. Musical instruments.
5. Games and domestic appliances, etc.

6. The logic gates are also employed for decision making in *automatic control of machines and various industrial processes and for building more complex devices such as binary counters* etc.

Positive and negative logic :

The number symbols 0 and 1 represent, in computing systems, two possible states of a circuit or device. It does not make any difference if these two states are referred to as

'ON' and *'OFF'*, *'Closed'* and *'Open'*, *'High'* and *'Low'*, *'Plus'* and *'Minus'* or *'True'* and *'False'* depending upon the situations. The main point is they must be symbolized by *two opposite conditions.* In *positive logic* a **'1'** *represents* : an *'ON circuit'* ; a *'Closed switch'*; a *'High voltage'*, a *Plus sign'*, *'True statement'*. *Consequently*, a **0** *represent* : an 'OFF circuit' ; an *'Open switch'*, a *'Low voltage'* ; a *'Minus sign'*, a 'False statement'.

In negative logic, the just opposite conditions prevail.

Example. A digital system has two voltage levels of 0 V and 5 V. If we say that symbol 1 stands for 5 V and symbol 0 for 0 V, then we have *positive logic system.* If on the other hand, we decide that a 1 should represent 0 V and 0 should represent 5 V, then we will get negative logic system.

Main point is that in *'positve logic'* the ***more positive*** of the two voltage levels represents the 1 while in *'negative logic'* the ***more negative*** voltage represents the 1.

Types of Logic Gates : Refer to table 2.4 (page 126)

In the *complex circuits,* the following *six* different digital electronics gates are used as *basic elements :*

1. NOT Gate
2. NAND Gate
3. AND Gate
4. OR Gate
5. NOR Gate
6. XOR Gate.

— A truth table has 2^n rows. It gives in each of its row *m* outputs for a given combination of *n* inputs.

1. NOT Gate :

- *Not operation means that the output is the complement of input.* If input is logic '1', the output is logic '0' and if input is logic '0', the output is logic '1'.
- Fig. 2.107 shows the *symbol of NOT Gate.* It is generally represented by a *triangle followed by a bubble* (or a bubble followed by a triangle).
- NOT gate is *used when an output is desired to be complement of the input.*
- If all inputs of NAND gates are joined it shall act as NOT gate.
- NOT gate is also *called 'inverting logic circuit.* It is also *called a 'complementing circuit'.*

2. **NAND Gate :**

- A NAND gate can said to be *basic building block* of the *all digital TTL logic gates and other digital circuits.*
- It is represented by the symbol shown in Fig. 2.108.
- Its *unique property is that output is high '1' if any of the input is at low '0' logic level.*

Let us consider two inputs with the states *A* and *B* at the NAND gate. The answer (output) $X = \overline{A.B}$. Bar denotes a NOT log operation on A.B. The meaning of A.B, called AND operation, is given in 3 below.

3. **AND Gate :**

- *A NAND gate followed by a NOT gate gives us AND gate.*
- It is represented by a symbol in Fig. 2.109. Its symbol differs from NAND only by *omission of a bubble (circle).*
- Its *unique property is that its output is '0' unless all the inputs to it are at the logic 1's.*
- A two inputs, AND gate has $X = A,B$. *Dot* between the two states indicates 'AND' logic operation using these.

4. **OR Gate :**

- An 'OR' operation means that the *output is '0' only if all the inputs are '0s'.*
- It is represented by a symbol shown in Fig. 2.110.

- If any of the inputs is '1' the output is '1'. A two inputs 'OR' gate has $X = A + B$. Sign + between the two states indicates an 'OR' logic operation.

5. **NOR Gate :**

- An 'OR' circuit followed by a NOT circuit gives a 'NOR' gate (Fig. 2.111).
- ts *unique property is that its output is '0' if any of input is '1'.*
- A NOR gate is a *basic building block for other types of the logic gates than TTLs.* In the TTL circuits, a NOR is fabricated in an IC by the several NANDs.

A two input NOR has $X = \overline{A + B}$.

6. **XOR Gate :**

- A XOR gate (Fig. 2.112) is called 'Exclusive OR' gate.
- Its *unique property is that the output is '1' only if odd number of the inputs at it are '1's.*
- The 'Exclusive OR' can be written as : $X = A \cdot \overline{B} + \overline{A} \cdot B$ or $A \oplus B$.
- Exclusive OR gate is important in the circuits *for addition of two binary numbers.*

7. **Coincidence Gate :**

- This gate (Fig. 2.113) can be written as : $X = \overline{A} \cdot \overline{B} + A \cdot B$.
- Output available to those states when the inputs are identical.

Basic building blocks. AND, OR and NOT gates are called *basic* building blocks or basic gates because they are *essential to realize any boolean expression.*

Universal gates. NAND and NOR gates are known as *universal gates* because any logic gate can be constructed either by using NAND gates only or by using NOR gates only.

2.3.7. Universal Gates

NAND and NOR gates are known as **universal gates.**

The AND, OR, NOT gates can be realized using only NAND or NOR gates.

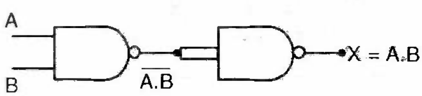
— Demorgan's theorem afford a convenient method to use these two gates in *logic design.* The entire logic system can be implemented by using any of these two gates.

— These two gates are *easier to realize and consume less power than other gates.*

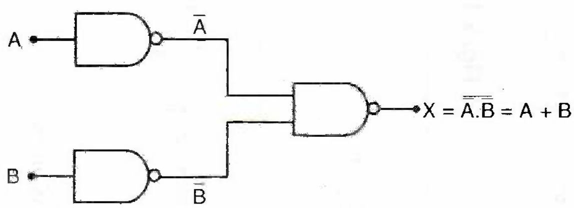
(*i*) **Realization of logic gates using NAND gates :**

Fig. 2.114 (*a*), (*b*), (*c*) shows realization of NOT, AND, OR gates using NAND gates respectively, which is self explanatory.

(a) Realization of NOT gate using NAND gate



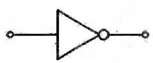
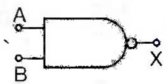
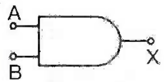
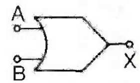
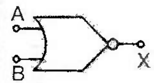
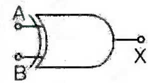
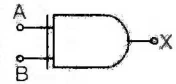
(b) Realization of AND gate using NAND gate



(c) Realization of OR gate using NAND gate

Fig. 2.114. Realization of NOT, AND and OR gates using NAND gates.

Table 2.4. Symbols, Boolean Expressions and Truth Tables of various Logic Gates

Logic	NOT	NAND	AND	OR	NOR	EX.OR	COINCIDENCE
Symbol	 **Fig. 2.107**	 **Fig. 2.108**	 **Fig. 2.109**	 **Fig. 2.110**	 **Fig. 2.111**	 **Fig. 2.112**	 **Fig. 2.113**
Boolean expression	$X = \overline{A}$	$X = \overline{A . B}$	$X = A.B$	$X = A + B$	$X = \overline{A + B}$	$X =$ $A . \overline{B} + \overline{A} . B$ $= A \oplus B$	$X = \overline{A} . \overline{B} + A.B$
Truth table	*A X* 0 1 1 0	*A B X* 0 0 1 0 1 1 1 0 1 1 1 0	*A B X* 0 0 0 0 1 0 1 0 0 1 1 1	*A B X* 0 0 0 0 1 1 1 0 1 1 1 1	*A B X* 0 0 1 0 1 0 1 0 0 1 1 0	*A B X* 0 0 1 0 1 1 1 0 1 1 1 0	*A B X* 0 0 1 0 1 0 1 0 0 1 1 1
Definition	Output available – when there is no innput.	Output available – to all states except when all the inputs are available.	Output available– when all inputs available.	Output available– when only one or more inputs available.	Output available– when no input is available.	Output available– when the inputs are *not* identical.	Output available– to those states when the inputs are identical.

(*ii*) **Realization of logic gates using NOR gates :**

The realization of NOT, OR and AND gates using NOR gates is shown in Fig. 2.115 (*a*), (*b*), (*c*) respectively.

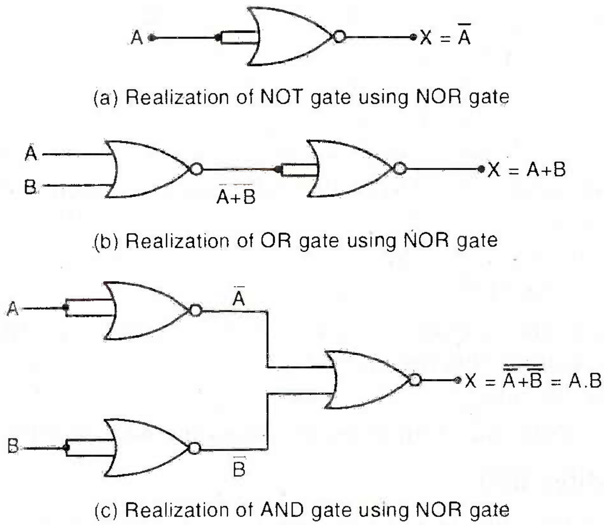


Fig. 2.115. Realization of NOT, OR and AND gates using NOR gates.

2.3.8. Half Adder (HA)

It is a *1-bit adder* and carries out binary addition with the help of XOR and AND gates. It has *two inputs and two outputs.*

It can add 2 binary digits at a time and produce a 2-bit data *i.e.*, 2-bit data *i.e.*, SUM and CARRY according to binary addition rules.

The circuit of a half adder is shown in Fig. 2.116. (*a*). It consists of an Ex-OR gate and AND gate. The outputs of the Ex-OR gate is called the SUM (S), while the output of the AND gate is known as CARRY (C). As the AND gate produces a *high* output only when both inputs are *high* and Ex-OR gate produces a *high* output if either input (not both) is high, the truth table of a half adder is developed by writing the truth table output of AND gate in the CARRY column and the output truth table of Ex-OR gate in SUM column. Truth table for half adder is given in table 2.5.

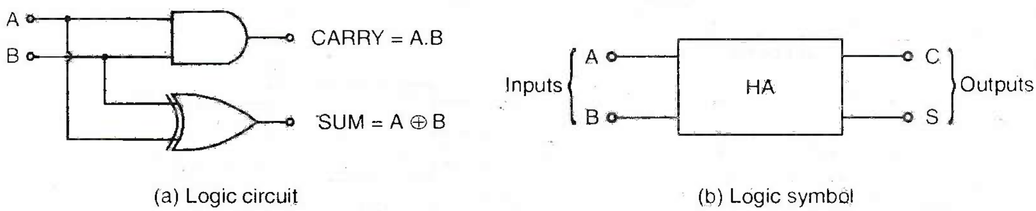


Fig. 2.116. Half adder.

Table 2.5. Truth table for Half Adder

Inputs		Outputs	
A	**B**	**C**	**S**
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

The logical expressions for CARRY and SUM can be written from the truth table for a half adder as follows :

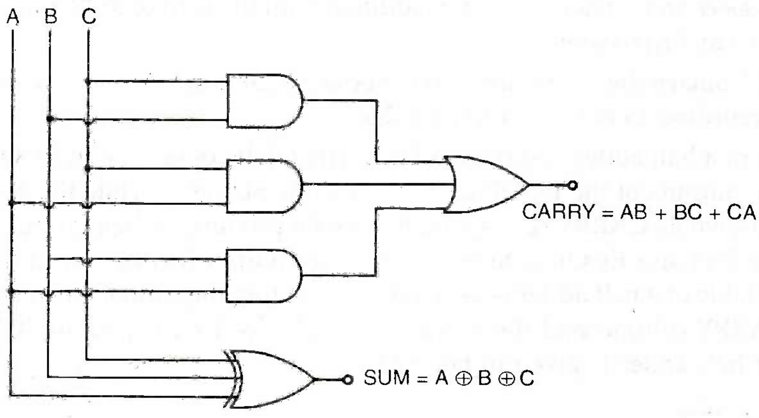
$$\text{CARRY, C} = \text{A.B}$$
$$\text{SUM, S} = \text{A} \oplus \text{B}$$

- This circuit is called *half-adder,* because it *cannot* accept a CARRY-IN from previous additions. Owing to this reason the half-adder circuit can be *used for binary addition of lower most bit only.*

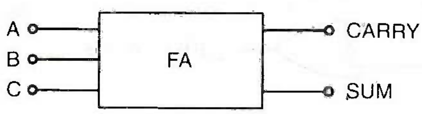
For higher-order columns, 3-input adder called *full adder* are used.

2.3.9. Full Adder (FA)

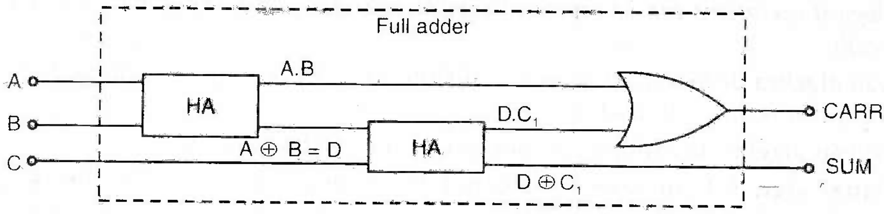
A full adder has *three inputs* and *two outputs.* It *can add 3 digits (or bits) at a time.* The bits A and B which are to be added come from the two registers and the third input C comes from the *'carry'* generated by the previous addition. It produces two outputs, SUM and CARRY-OUT (going to next higher column).



(a) Logic circuit



(b) Logic symbol



(c) Full adder circuit

Fig. 2.117. Full adder.

Table 2.6. Truth table for Full Adder

A	B	C	CARRY	SUM
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

A simple circuit of a full adder is shown in Fig. 2.117 (*a*), though other designs are also possible. It uses 3 AND gates, one Ex-OR gate and one OR gate. The final CARRY is given by the OR gate while the final SUM is given out by the Ex-OR gate.

Fig. 2.117(*b*) shows the logic symbol for a full adder.

Truth table for full adder for all possible inputs/outputs is given in Table 2.6. Truth table can be checked easily for its validity.

A full adder can be made by using two half adders and an OR gate. The circuit is shown in Fig. 2.117(*c*).

- The full adder can do more than a million additions per second. Besides that, it never get tired or bored or asks for a rest.

Note : *Binary additions :* Following are the four rules/cases for addition of binary numbers :

(1) $0 + 0 = 0$

(2) $0 + 1 = 1$

(3) $1 + 0 = 1$

(4) $1 + 1 = 10$ (This sum is *not ten* but *one-zero*).

2.3.10. Boolean Algebra

George Boolean in 1854 developed a mathematics now referred as *Boolean algebra.* It is the *algebra of logic presently applied to the operation of computer devices.* The rules of this algebra are based on human reasoning.

Digital circuits perform the binary arithmetic operations with binary digits 1 and 0. These operations are called logic functions or logic operations. *The algebra used to symbolically describe logic functions is called* **Boolean algebra.** Boolean algebra is *a set of rules and theorems* *by which logical operations can be expressed symbolically in equation form and be manipulated mathematically.*

Boolean algebra differs from ordinary algebra in that Boolean constant and variables *can have only two values* : '0' and '1' :

In Boolean algebra the following *four connecting symbols* are used :

1. **Equal sign (=).** In Boolean algebra the 'equal sign' refers to the standard mathematical equality. In other words, the logical value on one side of the sign is identical to the logical value on the other side of the sign.

Example. We are given two logical variables such that $A = B$. Then if $A = 1$, then $B = 1$ and if $A = 0$ then $B = 0$.

2. **Plus sign (+).** In Boolean algebra the 'plus sign' refers to logical *OR operation.*

The statement $A + B = 1$ means A ORed with B equals 1. Consequently, either $A = 1$ or $B = 1$ or both equal to 1.

3. **Multiply sign (.).** In Boolean algebra the 'multiply sign' refers to *AND operation.*

The statement $A.B = 1$ means A ANDed with B equals 1. Consequently, $A = 1$ and $B = 1$.

The function $A \,.\, B$ often written as AB, *omitting the dot* for convenience.

4. **Bar sign (–).** In Boolean algebra the 'bar sign' refers to *NOT operations.* The NOT has the effect of inverting (complementing) the logic value.

Thus if A = 1, then $\overline{A} = 0$.

2.3.11. Boolean Laws (For Outputs from Logic Inputs).

The following Laws can said to be associated with Boolean algebra :

1. **'OR' Laws**

The 'OR' Laws are described by the following equations :

$$A + 1 = 1 \quad ...[2.14(a)]$$

$$A + 0 = A \quad ...[2.14(b)]$$

$$A + A = A \quad ...[2.14(c)]$$

$$A + \overline{A} = 1 \quad ...[2.14(d)]$$

- An 'OR' operation is denoted by *plus sign.*
- 'OR' Law means :

(*i*) Any number (0 or 1) is a first input to an OR gate and another member at the second input is 1 then answer is 1,

(*ii*) If another is 0 then answer is as first input, and

(*iii*) If two inputs to an OR gate *complement* then output is '1'.

2. **'AND' Laws**

'AND' operation is denoted by the *dot* sign.

- True and true make true
- True and false make false
- False and false make false.

$$A \,.\, 1 = A \quad ...[2.15(a)]$$

$$A \,.\, 0 = 0 \quad ...[2.15(b)]$$

$$A \,.\, A = A \quad ...[2.15(c)]$$

$$A \,.\, \overline{A} = 0 \quad ...[2,15(d)]$$

3. **'NOT' Laws (Laws of Complementation)**

A NOT operation is denoted by putting a *bar* over a number.

- The NOT true means false.
- The NOT false means true.

$$\bar{1} = 0 \qquad ...[2.16(a)]$$

$$\bar{\bar{A}} = A \qquad ...[2.16(b)]$$

Eqn. [2.16(*b*)] means that if *A is inverted (complemented) and then again inverted, we get the original number.*

4. **Commutative Laws**

These Laws mean that *order of a logical operation is immaterial.*

$$A + B = B + A \qquad ...[2.17(a)]$$

$$A \,.\, B = B \,.\, A \qquad ...[2.17(b)]$$

5. **Associative Laws**

These laws *allow a grouping of the Boolean variables.*

$$A + (B + C) = (A + B) + C \qquad ...[2.18(a)]$$

$$A \,.\, (B \,.\, C) = (A \,.\, B) \,.\, C \qquad ...[2.18(b)]$$

6. **Distributive Laws**

These laws *simplify the problems in the logic disigns.*

$$A \,.\, (B + C) = (A \,.\, B) + (B \,.\, C) \qquad ...[2.19(a)]$$

$$A + (B \,.\, C) = (A + B) \,.\, (A + C) \qquad ...[2.19(b)]$$

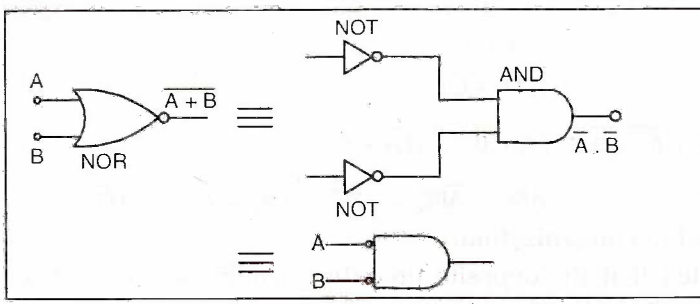
$$A + (\bar{A} \,.\, B) = A + B \qquad ...[2.19\ (c)]$$

The last two equations are typical to the Boolean algebra, and are not followed in the usual algebra.

2.3.12. De Morgan's Theorems

First theorem shows an equivalence of a NOR gate with an AND gate having bubbled inputs (Fig. 2.118), and is given by the equation :

$$\overline{A + B} = \bar{A} \,.\, \bar{B} \qquad ...(2.20)$$

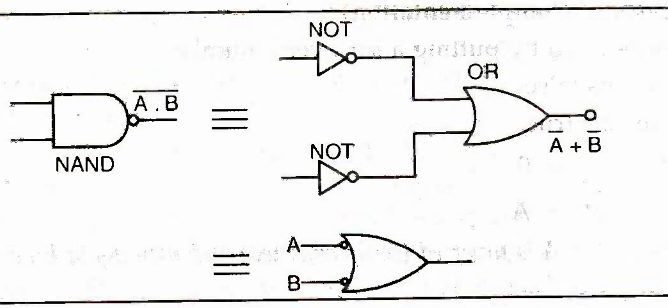


A	B	X
0	0	1
1	0	0
0	1	0
1	1	0

Fig. 2.118. De Morgan's First theorem showing an equivlence of a NOR gate (same holds for multiple inputs).

Second theorem shows an equivalence of a NAND gate with an OR having bubbled inputs as shown in Fig. 2.119 and is given by the equation :

$$\overline{A \,.\, B} = \bar{A} + \bar{B} \qquad ...(2.21)$$



A	B	X
0	0	1
1	0	1
0	1	1
1	1	0

Fig. 2.119. De Morgan's Second theorem showing the equivalence of a NAND gate (same holds for the multiple inputs).

In fact the eqns. (2.20) and (2.21) also hold for the cases of the multiple (more than two) inputs.

$$\overline{A+B+C+\ldots\ldots} = \overline{A}\,.\,\overline{B}\,.\,\overline{C} \qquad \ldots[2.22(a)]$$

$$\overline{A\,.\,B\,.\,C\,\ldots\ldots} = \overline{A}+\overline{B}+\overline{C} \qquad \ldots[2.22(b)]$$

The *purpose of these theorems is to enable digital circuit designers to implement all the other logic gates with the help of either NOR gates only or NAND gates only.* For example, a NOT gate is implementable by a NAND or a NOR as shown in the left part or lower right part of Fig. 2.118 respectively. This theorem finds wide use in the digital logic circuits as these are implementable on one single basic logic gate considered as a basic building unit.

- The *'first statement'* (De Morgan's) says that *the complement of a sum equals the product of the complements.* The *'second statement'* says that the *complement of a product equals the sum of the complements.* In fact, it allows transformation from a sum-of-products form to a product-of-sum form.

— The procedure required for *taking out an expression from under a NOT sign is as follows :*

1. *Complement the given expression i.e., remove the overall NOT sign.*
2. *Change all and ANDs to ORs and all the ORs to ANDs.*
3. Complement or negate all individual variables.

Examples : (*i*) $\overline{A + BC} = A + BC$ ...*Step 1*

$= A(B + C)$ ...*Step 2*

$= \overline{A}(\overline{B}+\overline{C})$ ...*Step 3.*

(*ii*) $\overline{\overline{(\overline{A}+B+\overline{C})}\,\overline{(\overline{A}+B+C)}} = (\overline{A}+B+\overline{C})(\overline{A}+B+C)$

$= \overline{A}B\overline{C} + \overline{A}BC = \overline{\overline{A}}\,\overline{B}\,\overline{\overline{C}} + \overline{\overline{A}}\,\overline{B}\,\overline{C} = A\overline{B}C + A\overline{B}\,\overline{C}$

This *process is called* **demorganization.**

— It may be noted that the opposite procedure would be followed to *bring an expression under the NOT sign :*

Example : $\overline{A}+\overline{B}+\overline{C} = \overline{\overline{\overline{A}}+\overline{\overline{B}}+\overline{\overline{C}}}$ ...*Step 3*

$= \overline{A + B + C}$

$= ABC$ ...*Step 2*

$= \overline{ABC}$ ...*Step 1*

2.3.13. Operator Precedence

For evaluating Boolean expression, the operator precedence is :
(*i*) parenthesis, (*ii*) NOT, (*iii*) AND and (*iv*) OR. In other words :

— The expression inside the *parenthesis* must be evaluated before all other operations,
— The next operation that holds precedence is the *complement*,
— Then follows the AND, and
— Finally the OR.

Example. In the Boolean expression $A + \overline{B}(C + D)$, and expression inside the parenthesis will be evaluated first, then $\overline{B}$ will be evaluated, then the results of the two [*i.e.*, $\overline{B}$ *and* $(C + D)$] will be ANDed and finally, the result of the product ORed with A.

Example 2.60. *Prove the following identity : $AC + ABC = AC$.*

Solution. Taking the left hand expression as X, we get

$$X = AC + ABC = AC(1 + B)$$

Now, $\quad 1 + B = 1 \qquad$ [Eqn. 2.14(*a*)]

$\therefore \quad X = AC.\,1 = AC$

$\therefore \quad AC + ABC = AC$ **...Proved.**

Example 2.61. *Prove the following Boolean identity : $(A + B)(A + C) = A + BC$.*

Solution. Putting the left hand side expression equal to X, we get

$$\begin{aligned} X &= (A + B)(A + C) \\ &= AA + AC + AB + BC && \text{...[Eqn. 2.19(a)]} \\ &= A + AC + AB + BC && [AA = A \text{ ... } (2.15(c))] \\ &= A + AB + AC + BC \\ &= A(1 + B) + AC + BC && (\because 1 + B = 1) \\ &= A + AC + BC \\ &= A(1 + C) + BC \\ &= A + BC && (\because 1 + C = 1) \end{aligned}$$

$\therefore \quad (A + B)(B + C) = A + BC.$ **...Proved.**

Example 2.62. *Prove the following identity : $A + \overline{A}B = A + B$.*

Solution. Putting the left hand expression equal to X, we get

$$\begin{aligned} X &= A + \overline{A}B = A.1 + \overline{A}B && \text{[Eqn. 2.15(a)]} \\ &= A(1 + B) + \overline{A}B && \text{[Eqn. 2.14(a)]} \\ &= A.1 + AB + AB && \text{[Eqn. 2.19(a)]} \\ &= A + BA + B\overline{A} \\ &= A + B(A + \overline{A}) && \text{[Eqn. 2.19(a)]} \\ &= A + B.1 && \text{[Eqn. 2.14(d)]} \\ &= A + B && \text{[Eqn. 2.15(a)]} \end{aligned}$$

$\therefore \quad A + \overline{A}B = A + B.$ **...Proved.**

Example 2.63. *Simplify the following Boolean expression to a minimum of literals :*

$$X = AB + \overline{A}C + BC.$$

Solution.

$$\begin{aligned} X &= AB + \overline{A}C + BC \\ &= AB + \overline{A}C + BC(A + \overline{A}) && \text{...[Eqn. 2.14 (d)]} \end{aligned}$$

$$= AB + \overline{A}\,C + ABC + \overline{A}\,BC$$

$$= AB\,(1 + C) + \overline{A}\,C(1 + B)$$

$$= AB + \overline{A}\,C \qquad \text{...[Eqn. 2.14(}a\text{)]}$$

$$X = AB + \overline{A}\,C. \quad \textbf{(Ans.)}$$

Example 2.64. *Simplify the following Boolean expression :*

$$AB\overline{C} + A\overline{B}\,\overline{C} + \overline{A}BC + ABC + A\overline{B}C.$$

Solution. Let, $X = AB\,\overline{C} + A\,\overline{B}\,\overline{C} + \overline{A}\,BC + ABC + A\,\overline{B}\,C$

Bringing together those terms which have two cammon letters, we get

$$X = ABC + AB\,\overline{C} + A\,\overline{B}\,\overline{C} + A\overline{B}\,C + \overline{A}\,BC$$

$$= AB(C + \overline{C}) + A\overline{B}\,(\overline{C} + C) + \overline{A}\,BC$$

$$= AB + A\overline{B} + \overline{A}\,BC \qquad \text{...[Eqn. 2.14(}d\text{)]}$$

$$= A\,(B + \overline{B}) + \overline{A}\,BC$$

$$= A + \overline{A}\,BC = A + BC. \quad \textbf{(Ans.)} \qquad \text{...[Eqn. 2.19(}c\text{)]}$$

Example 2.65. *Using Boolean algebra techniques, simplify the following expression :*

$$X = A.B\,.\,\overline{C}.\overline{D} + \overline{A}.B.\overline{C}.\overline{D} + \overline{A}.B.C.\overline{D} + A.B.C.\overline{D}.$$

Soution. $X = B\,\overline{C}\,\overline{D}\,(A + \overline{A}) + BC\,\overline{D}\,(A + \overline{A})$

...Taking out the common factors

$$= B\,\overline{C}\,\overline{D} + B\,C\,\overline{D} \qquad \text{...[Eqn. 2.14 (}d\text{)]}$$

$$= B\,\overline{D} + (C + \overline{C}) \qquad \text{...Again factorize}$$

$$= B\,\overline{D}.\,1 = B\,\overline{D} \quad \text{...(Simplified form)} \ \textbf{(Ans.)} \qquad \text{...[Eqn. 2.14(}d\text{)]}$$

Example. 2.66. *Simplify the following expression and show the minimum gate implementation.*

$$X = A\,.\,B\,.\,\overline{C}\,.\,\overline{D} + \overline{A}\,.\,B\,.\,\overline{C}\,\overline{D} + B.\,\overline{C}.\,D$$

Solution. $X = B\,.\,\overline{C}\,.\,\overline{D}\,(\overline{A} + B) + B\,.\,\overline{C}\,.\,D$

$$= B\,.\,\overline{C}\,.\,\overline{D}\,.\,1 + B\,.\,\overline{C}\,.\,D$$

...[Eqn. 2.14 (*d*)]

$$= B\,.\,\overline{C}\,.\,\overline{D} + B\,.\,\overline{C}\,.\,D$$

$$= B\,.\,\overline{C}\,.\,(D + \overline{D}) = B\,.\,\overline{C}\,.\,1 = B\,.\,\overline{C}$$

...[Eqn. 2.14 (*d*)]

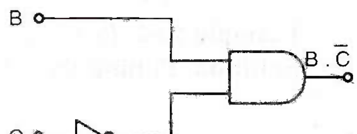


Fig. 2.120

Minimum gate implementation is shown in Fig. 2.120

Example 2.67. *Determine output expression for the circuit shown in Fig. 2.121.*

Solution. The output expression for the circuit shown in Fig. 2.121 is :

$$X = \overline{[\overline{(A + B)}\,.\,C\,.\,\overline{D}]}.$$

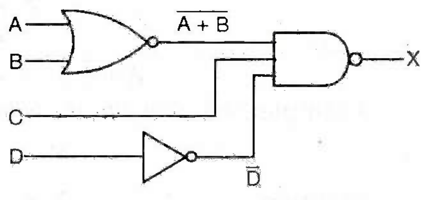


Fig. 2.121

Example 2.68. *Simplify the following Boolean expression and draw the logic circuit for simplified expression :*

$$X = \overline{B}\,(A + C) + C(\overline{A} + B) + AC.$$

Solution.

$$X = \overline{B}(A+C) + C(\overline{A}+B) + AC$$
$$= A\overline{B} + \overline{B}C + \overline{A}C + BC + AC$$
$$= A\overline{B} + C(\overline{B} + \overline{A} + B + A)$$
$$= A\overline{B} + C.1 = A\overline{B} + C \quad \text{...Simplified expression. (Ans.)}$$

Logic circuit for the simplified expression is shown in Fig. 2.122

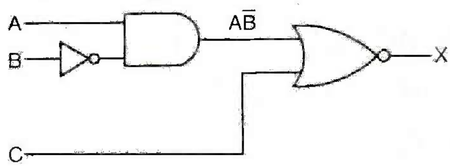


Fig. 2.122

Example 2.69. *Simplify the expression :* $(AB + C)(AB + D)$.

Solution. Let

$$X = (AB + C)(AB + D)$$
$$= AB\,AB + ABD + ABC + CD \quad \text{...[Eqn. 2.19(a)]}$$
$$= AABB + ABD + ABC + CD$$
$$= AB + ABD + ABC + CD \quad \text{...[Eqn. 2.15(c)]}$$
$$= AB(1 + D) + ABC + CD$$
$$= AB + ABC + CD \quad \text{...[Eqn. 2.14(a)]}$$
$$= AB(1 + C) + CD$$
$$\therefore \quad (AB + C)(AB + D) = AB + CD. \quad \textbf{(Ans.)}$$

Example 2.70. *Draw the logic circuit represented by the expression :*

$$X = AB + \overline{A}.\overline{B} + \overline{A}.B.C.$$

Solution. A circuit using gates can simply be designed by looking at the expression and finding out the basic gates which can be used to realize the various terms and then correct these gates appropriately.

In the given expression there are three input logical variables and X is the output.

- The *first* term $A.B$ is obtained by ANDing A with B as shown in Fig. 2.123 (*i*).
- The *second* term $\overline{A}.\overline{B}$ is obtained by using two INVERTERs and one AND gate and connecting them as shown in Fig. 2.123.

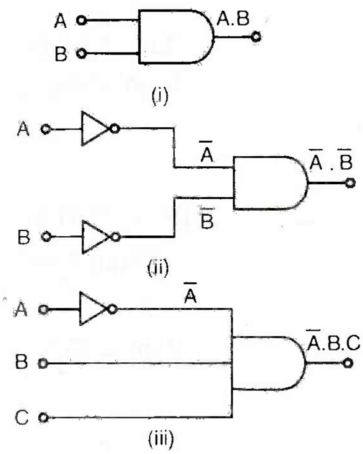


Fig. 2.123.

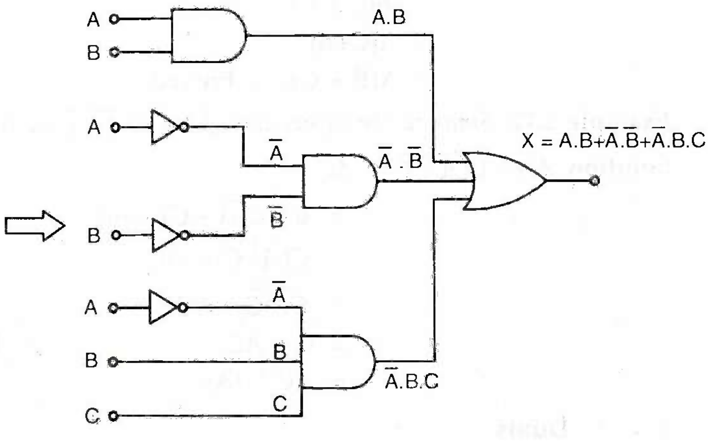


Fig. 2.124. Logic gate implementation of expression $A.B + \overline{A}.\overline{B} + \overline{A}.B.C.$

- The *last* term is used by using one INVERTER, one AND gate and connecting them as shown in Fig. 2.123(*iii*).

Now, the complete logic expression is realised by ORing the three outputs of the arrangements explained above *i.e.*, by ORing. $A.B$, $\bar{A}.\bar{B}$ and $\bar{A}.B.C$. The logic gate implementation for the given expression is shown in Fig. 2.124.

Example 2.71. *Show that :*

(*i*) $(A+B)(\bar{A}+C) = AC+\bar{A}B$ (*ii*) $AB+ABC+\bar{A}B+A\bar{B}C = B+AC$

(*iii*) $ABC+A\bar{B}C+AB\bar{C} = A(B+C)$.

Solution. (*i*) $(A+B)(\bar{A}+C)$

$$= A\bar{A}+AC+B\bar{A}+BC \qquad (\because \; A.\bar{A}=0)$$

$$= AC+B\bar{A}+BC$$

Multiplying the third term by $(A+\bar{A})$, we get

$$= AC+B\bar{A}+BC(A+\bar{A}) \quad [A+\bar{A}, \text{ being equal to 1 does not make any effect}]$$

$$= AC+B\bar{A}+ABC+\bar{A}BC \qquad \text{...[Eqn. 2.14(}d\text{)]}$$

$$= AC(1+B)+BA(1+C)$$

$$= AC+B\bar{A} \qquad \text{...[Eqn. 2.14(}d\text{)]}$$

$= AC+\bar{A}B$. **...Proved.**

(*ii*) $AB+ABC+\bar{A}B+A\bar{B}C$

$$= AB+\bar{A}B+AC(B+\bar{B})$$

$$= B(A+\bar{A})+AC(B+\bar{B})$$

$= B+AC$. **...Proved** ...[eqn. 2.14(*d*)]

(*iii*) $ABC+A\bar{B}C+AB\bar{C}$

$$= AC(B+\bar{B})+AB\bar{C}$$

$$= AC+AB\bar{C} \qquad \text{...[Eqn. 2.14(}d\text{)]}$$

$$= A(C+B\bar{C})$$

$$= A(C+B) \qquad \text{...[Eqn. 2.14(}d\text{)]}$$

$= A(B+C)$. **...Proved.** ...[Eqn. 2.19(*c*)]

Example 2.72. *Simplify the expression* $A\bar{A}+C(\overline{A+C})+AC$.

Solution. $A\bar{A}+C(\overline{A+C})+AC$

$$= 0+C(\overline{A+C})+AC \qquad \text{...[Eqn. 2.15.(}d\text{)]}$$

$$= C(\bar{A}.\bar{C})+AC \qquad \text{...[Eqn. 2.20]}$$

$$= C\bar{A}\bar{C}+AC$$

$$= 0+AC \qquad \text{...[Eqn. 2.15(}d\text{)]}$$

$= AC$. **(Ans.)**

2.3.14. Duals

In Boolean algebra *each expression has its dual which is as true as the original expression.* For getting the dual of a given Boolean expression, the procedure involves conversion of (*i*) all 1s to 0s and all 0s to 1s.

(*ii*) all ANDs to ORs and all ORs to ANDs.

The dual so obtained is *also found to be true.*

Some of the Boolean relations and their duals are given in Table 2.7.

Table 2.7.

Relation	Dual relation
$A \,.\, 0 = 0$	$A + 1 = 1$
$A \,.\, A = 1$	$A + A = A$
$A \,.\, \overline{A} = 0$	$A + \overline{A} = 1$
$A \,.\, 1 = A$	$A + 0 = A$
$A \,.\, (A + B) = A$	$A + AB = 1$
$A + (\overline{A} + B) = AB$	$A + \overline{A}\,B = A + B$

Example 2.73. *Determine the Boolean expression for the logic circuit shown in Fig. 2.125. Simplify the Boolean expression using Boolean laws and De Morgan's theorem. Redraw the logic circuit using the simplified Boolean expression.*

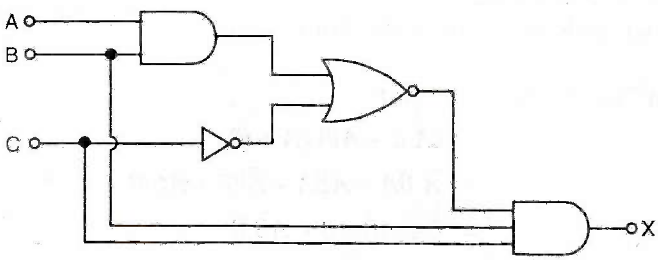


Fig. 2.125

Solution. The output of a given circuit can be obtained by *determining the output of each logic gate while working from left to right.*

With reference to Fig. 2.126, the output of the circuit is :

$$X = BC\left(\overline{AB + \overline{C}}\right)$$

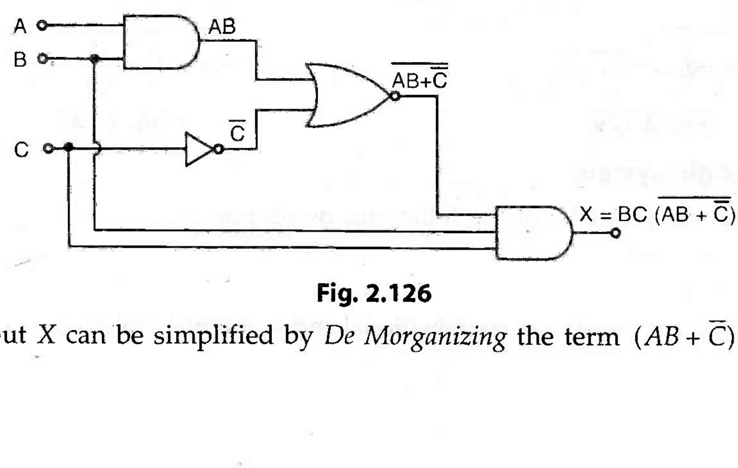


Fig. 2.126

The output X can be simplified by *De Morganizing* the term $(AB + \overline{C})$ as follows :

$$
\begin{aligned}
BC(\overline{AB + \overline{C}}) &= BC(AB + \overline{C}) && \text{...Step-1}\\
&= BC(A + B) \,.\, \overline{C} && \text{...Step-2}\\
&= BC(\overline{A} + \overline{B}) \,.\, \overline{\overline{C}} && \text{...Step-3}\\
&= BC(\overline{A} + \overline{B})C && \text{...[Eqn. 2.16(b)]}\\
&= BC(\overline{A} + \overline{B}) && \text{...[Eqn. 2.15(c)]}\\
&= \overline{A}\,BC + BC\overline{B} &&\\
&= \overline{A}\,BC + 0 && \text{...[Eqn. 2.15(d)]}\\
&= \overline{A}\,BC && \text{...[Eqn. 2.14(b)]}
\end{aligned}
$$

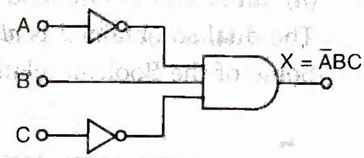


Fig. 2.127

The logic circuit with a simplified Boolean expression $X = \overline{A}\,BC$ is as shown in Fig. 2.127.

Example 2.74. *Determine the output X of a logic circuit shown in Fig. 2.128. Simplify the output expression using Boolean Laws and theorems. Redraw the logic circuit with the simplified expression.*

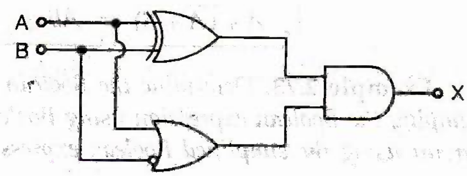


Fig. 2.128

Solution. The output of the given logic circuit can be obtained by determining the output of each logic gate while working from left to right.

As seen from Fig. 2.129, the output ;

$$
\begin{aligned}
X &= (\overline{A}\,B + A\overline{B})\,(A + \overline{B}) &&\\
&= \overline{A}\,BA + A\overline{B}A + \overline{A}B\overline{B} + A\overline{B}\,\overline{B} &&\\
&= 0 + A\overline{B} + 0 + A\overline{B}\,\overline{B} && \text{...[Eqn. 2.15(d)]}\\
&= A\overline{B} + A\overline{B} && \text{...[Eqn. 1(b), 2.15(c)]}\\
&= A\overline{B} && \text{...[Eqn. 2.14(c)]}
\end{aligned}
$$

Using the simplified Boolean expression, the logic circuit is as shown in Fig. 2.130.

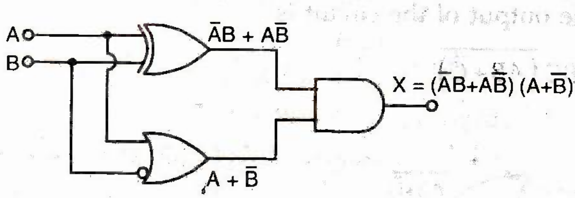


Fig. 2.129

Fig. 2.130

2.3.15. Logic System

The logic system may be of the following two types :

1. Combinational.
2. Sequential.

The essential charateristics of combinational and sequential logic systems are compared as follows :

Combinational	Sequential
1. Possesses no memory or storage capacity.	Possesses memory or storage capacity.
2. Simple logic gates only carry out the implementation.	To carry out the implementation alongwith the logic gates, flip-flops, counters, registers memory cores are also used.
3. The system is described by a set of output functions only.	It is described by a set of output functtions and also next state functions.
4. Output of the system depends only on the present input.	Output of the system depeds on the present input as well as on the present state of the system.

Combinational circuits :

A ***combinational circuit*** *consists of logic gates whose outputs at any time are determined directly from the combination of inputs without regard for previous input.*

The circuit possesses a set of *inputs*, a *memoryless logic network* to operate on the inputs and a *set of outputs* as shown in Fig. 2.131. Moreover, output combinational networks are used to make logical decisions and control the operation of different circuits in digital electronic systems. For a given set of input conditions, the output of such a circuit is the same. Consequently, truth table can fully describe the operation of such a circuit.

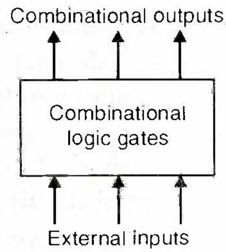


Fig. 2.131. Combinational logic circuit.

Examples. Examples of a combinational circuit are :

(*i*) Decoders (*ii*) Adders
(*iii*) Multiplexers (*iv*) Demultiplexers etc.

- **Multiplexers and demultiplexers :**
  — *Transmission of a large number of information units over a small number of lines is known as* **"Multiplexing".**
  — **"Demultiplexing"** is a *reverse operation and denots receiving information from a small number of channels and distributing it over a large number of destinations.*

Design procedure of combinational circuit :

Following *operations* are involved in the design procedure :

1. To state the problem.
2. To determine the number of available input variables and required output variables.
3. To assign letter symbol to each input and output variable.
4. To derive the truth table that defines the required relationship between inputs and outputs.
5. To obtain the simplified Boolean function for each output.
6. To draw the logic diagram.

- A circuit that adds two bits is called a ***half adder.***
- A ***full adder*** consists of three inputs and two outputs. The outputs are designated by the symbol S for sum and C for carry.
- A ***two bit subtractor*** has two inputs X (minuend) and Y (subtrahend).

- A *full subtractor* **(FS)** is a *combinational circuit that performs a subtraction between two bits.* This circuit has *three inputs and two outputs.*

Code conversion :

- A variety of codes are used by different digital systems. It is sometimes necessary to use the output of one system as input to the other.
- A conversion circuit must be inserted between the two systems if each uses different codes for the same information.
- To convert from binary code A to binary code B, the input lines must supply the bit combinations of elements as specified by code A and the output lines must generate the corresponding bit combination of code B. A combination circuit performs this transformation by means of *logic gates.*

Comparator. A comparator is a combinational circuit that compares two number A and B and determine their relative magnitude. The outcome of the comparison is displayed in three outputs that indicate $A > B \equiv X$, $A = B = Y$, $A < B \equiv Z$.

Decoders and encoders :

— A *decoder* is a combination circuit that converts a binary code of n variables into m output lines, one for each discrete element of information.

— An ***encoder*** is a combination circuit that accepts m input lines, one for each element of information, and generates a binary code of n output lines.

Sequential circuits :

Such circuits have *inputs, logic network, outputs* and a *memory,* as shown in Fig. 2.132. Their present output depends not only on their present inputs but also on the pevious logic states of the outputs.

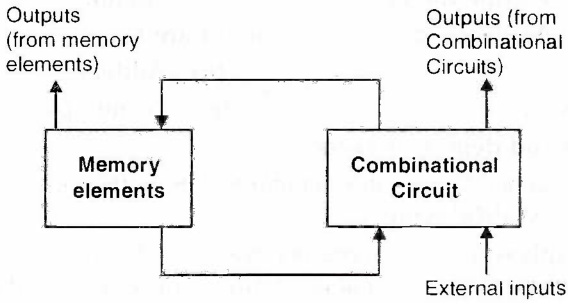


Fig. 2.132. Block diagram of a sequential circuit.

Examples. Examples of sequential circuits are :

(*i*) *Latches* (*ii*) *Flip-flops.*

The two main *types* of sequential circuits are :

1. Synchronous sequential circuits ..... referred to as *clocked-sequential circuits*
2. Asynchronous sequential circuits.

- The ***synchronous sequential circuits*** are *built to operate at a clocked rate whereas* **asynchronous ones** *are without clocking.*

2.3.16. Flip-Flop Circuits

The memory elements used in clocked sequential circuits are called **flip-flops.** These circuits are binary cells capable of storing one bit of information. It has two *outputs, one for the normal value and one for the complement value of the bit stored in it.* Binary information can enter a flip-flop in a variety of ways. Hence there different types of flip-flops.

A number of flip-flops are available in *IC* form. Some of these are *SR* (Set-Reset), *J-K* and *D* flip-flops. They are widely used as *switches, latches, counters, registers* and *memory cells* in computers.

- A salient feature of the flip-flop is that *output can exist in one of the two stable states, logic 1 and logic 0, simultaneously.* This is ensured by the appropriate crossed feedback connections associated with the most elementary form of the flip-flop known as a *latch.*

The following flip-flops will be discussed in the following articles :

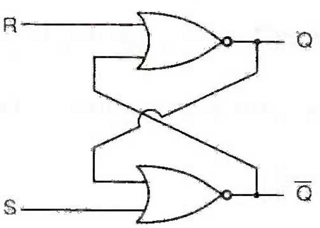
1. *R-S* flip-flop.
2. Clocked *R-S* flip-flop.
3. *D* flip-flop.
4. *J-K* flip flop.
5. *T* flip-flop.

R-S flip-flop :

Fig. 2.133 shows a *R-S flip-flop using* **NOR** *gates.* There are two inputs to the flip-flops called *S* (set) and *R* (*reset*). *The* cross-coupled connection from the output of one gate and input of the other constitutes a feedback path. For that reason, the *circuit is classified as synchronous circuit.*

- A low *R* and a high *S* results in the *set* state.
- A high *R* and a low *S* give the *reset* state.
- If both *R* and *S* are high, the output becomes *indeterminate* and this is known as *'race condition'*. This condition is *avoided by proper design.*

The truth table is shown in Table 2.8.



(*a*) Circuit diagram

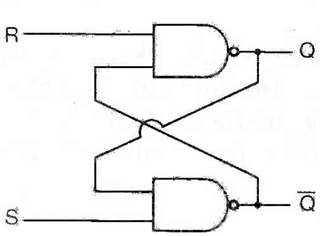
Table 2.8. Truth table for NOR latch

R	S	Q	Comment
0	0	NC	No change
0	1	1	Set
1	0	0	Reset
1	1		Race

(*b*) Truth table

Fig. 2.133. *R-S* flip-flop using NOR gates.

Fig. 2.134 shows a *R-S* flip-flop using **NAND** *gates.* Table 2.9 shows the truth table. It is seen that the inactive and race conditions are **reversed.**



(*a*) Circuit diagram

Table 2.9.Truth table for NAND latch

R	S	Q	Comment
0	0		Race
0	1	1	Set
1	0	0	Reset
1	1	NC	No change

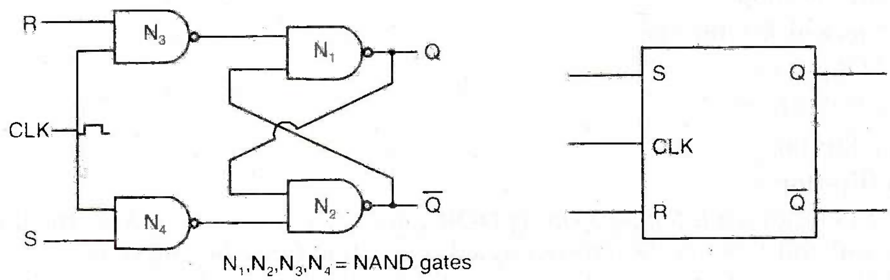
(*b*) Truth table

Fig. 2.134. *R-S* flip-flop using NAND gates.

- When R is *low*, output Q is *high*.
- When R is *high*, output Q is *low*.
- When both R and S are low, we get race condition which must be *avoided.*
- When both R and S are high - *no change condition.*

Clocked R-S flip-flop :

A large number of flip-flops are used in a computer. In order *to coordinate their working a square wave signal* known as *clock* is applied to the flip-flop. This *clock signal (indicated as CLK) pevents the flip-flop changing state till the right instant occurs.*



(a) Circuit diagram (b) Symbol

Fig. 2.135. Clocked *R-S* flip-flop

Fig. 2.135(*a*) shows a clocked R-S flip-flop using NAND gats (N_1 and N_2). This circuit uses two NAND gates N_3 and N_4 to apply CLK signal.

- When CLK is *low* the flip-flop *output Q ndicates no change.*
- If S is high and R is *low*, the *flip-flop must wait till CLK becomes high before Q can be set on 1.*
- If S is low and R is low, the flip-flop must wait for CLK to be high before Q is reset to low (0).

Clocked *R-S* flip-flop is a *synchroneous sequential logic circuit* because output state of the circuit changes at *discrete clocked instant of time.*

Fig. 2.135(*b*) shows a symbol for clocked *R-S flip-flop.*

Level clocking and edge triggering :

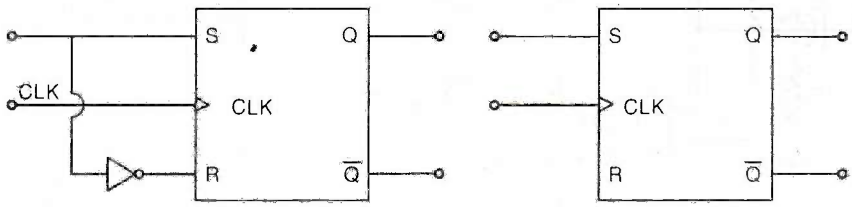
In a clocked flip-flop, the output can change state when CLK is high. When CLK is low, the output remains in the same state. Thus, the output can change state during the entire half cycle when CLK is high. This may be a *disadvantage in several situations. It is necessary that the output should change state only at one instant in the positive half cycle of the clock.* This *is known as* ***edge triggering*** *and the resulting flip-flop is known as* ***edge triggered flip-flop.***

Edge triggering can be made feasible by the use of an *RC* circuit. The time constant *RC* is made much smaller than the width of the clock pulse. Therefore, the capacitor can charge fully when CLK is high. The exponential charging produces a narrow positive voltage strike across the resistor. *The input gates are activated at the instant of this positive strike.*

D flip-flop :

A *D* flip-flop is an improvement over the *R-S* flip-flop to *avoid race condition.* It can be *level clocked or edge triggered.* The edge triggered one causes the change in output state at a unique instant.

In a *clocked R-S* flip-flop two input signals are required to drive the flip-flop which is a disadvantage with many digital circuits. In some events, both input signals become *high* which is again an undesirable condition. So these shortcomings/drawbacks of clocked *R-S* flip-flop are overcome in *D* flip-flop.



(b) Circuit diagram (b) Symbol

Input D_n	Output Q_{n+1}
0	0
1	1

(c) Truth table

Fig. 2.136. D flip-flop

Fig. 2.136(*a*), (*b*), (*c*) show the circuit diagram, symbol and truth table of *D* flip-flop respectively. It may be observed that only single data bit, *D* is required to drive the flip-flop.

— When the clock signal is at low level, data bit *D* is prevented to reach at output *Q* until clock signal becomes *high* at next pulse.

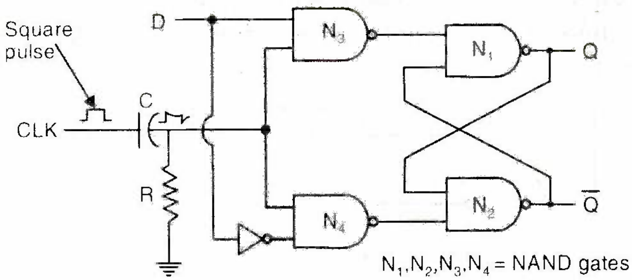
— It may be noted from the truth table that when data bit D_n is *high*, output Q_{n+1} gets at *high* level and when data bit D_n is low, Q_{n+1} gets at low level. Thus *D* flip-flop transfers the data bit *D* to *Q* as it is, and *Q* remains in the same state until the next pulse of the clock arrives.

- The flip-flop is named (*D*) flip-flop since *the transfer of data from the input to output is delayed.*
- The *D*-type flip-flop is *either used as a delay device* or *as a latch to store 1-bit of binary information.*

Edge triggered D flip-flop :

Fig. 2.137.(*a*) shows the circuit diagram and symbols of an edge triggered *D* flip-flop. The *clock provides the square wave signal.* RC circuit *converts* this signal *into strikes* so that triggering occurs at the instant of positive strike. The data bit *D drives one of the inputs.* Because of inverter, the complement $\overline{D}$ *drives the other output.* At the instant of positive strike, input *D* and its complement $\overline{D}$ cause the output *Q* to set or reset. Fig. 2.137(*b*) shows the truth table.

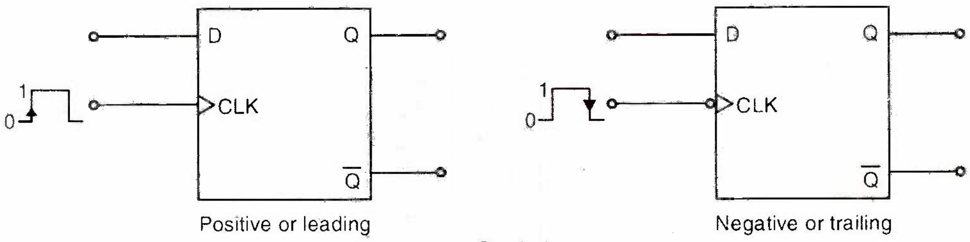
- When CLK is 0 or 1, the *D* input is not ***there*** and there is no change in state of *Q*.
- On the *negative* edge of the clock (marked ↓) the output remains in the same state.
- On the *positive* edge of the clock (marked ↑) *Q* changes to 0 if *D* is 0 and to 1 if *D* is 1.



(a) Circuit diagram

CLK	D	Q
0	×	No change
1	×	No change
↓	×	No change
↑	0	0
↑	1	1

(b) Truth table



Symbols

Fig. 2.137. Edge triggered *D* flip-flop.

Edge triggered J-K flip-flop :

— *J-K* fip-flop is *very versatile* and is perhaps the most *widely used type of flip-flop.*

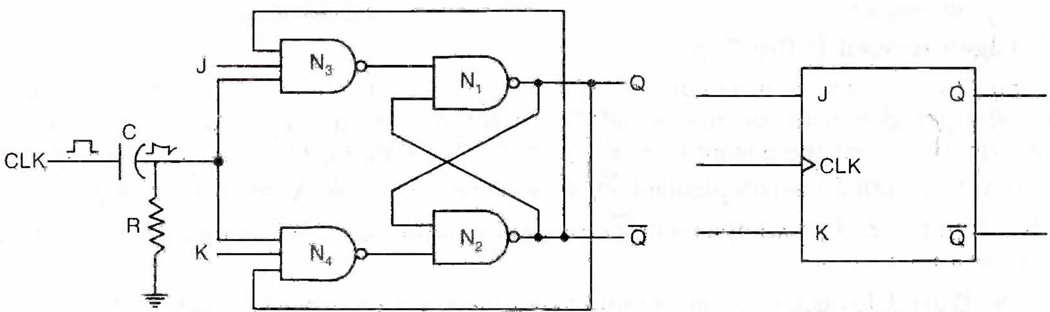
— The *J* and *K* designations for the inputs have no known significance except that they are adjacent letters in the alphabet.

— *J-K* flip-flop *functions identically to R-S flip-flop.*

The difference is that the *J-K* flip-flop has no *invalid state* as does the *R-S* flip-flop.

— It is widely used in digital devices such as *counters, registers, arithmetic logic units, and other digital systems.*

Fig. 2.138(*a*) shows the circuit diagram of a edge triggered *J-K* flip-flop used in *digital counters.* The CLK input is through an *RC* circuit with a short time constant. The *RC* circuit converts the rectangular clock pulse to narrow spikes as shown. Due to double inversion through NAND gates, the circuit is positive edge triggered.



(a) Circuit diagram

(b) Symbol for positive edge triggered J.K. flip-flop

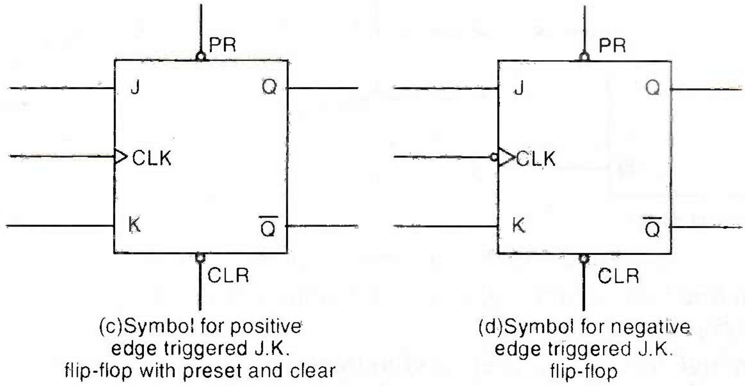


Fig. 2.138. Edge triggered *J-K* flip-flop.

- When both inputs J and K are low, the circuit is inactive at all times irrespective of the presence of CLK pulse.
- When J is low (*i.e.*, 0) and K is high (*i.e.*, 1), the circuit will be reset when positive CLK edge strikes the circuit and $Q = 0$. The flip-flop will remain in reset state if it is already in reset state.
- When $J = 1$ and $K = 0$, the circuit sets at the arrival of next positive clock edge.
- When $J = 1$ and $K = 1$, the flip-flop will toggle (means to switch to opposite state) on the next positive CLK edge. The action is illustrated in the table 2.10 :

Table 2.10. Positive edge triggered J-K flip-flop

CLK	J	K	Q
0	×	×	No change
1	×	×	No change
↓	×	×	No cahnge
X	0	0	No change
↑	0	1	0 (reset)
↑	1	0	1 (set)
↑	1	1	toggle

Using of *RC* circuit for edge triggering is not very convenient for fabrication. Actual circuits use additional NAND gates for edge triggering, such circuits are known as *direct coupled circuit.*

— Fig. 2.138(*b*) shows the symbol for positive edge triggered *J-K* flip-flop.

— Fig. 2.138(*c*) shows a positive edge *J-K* flip-flop with present (*PR*) and clear (CLR).

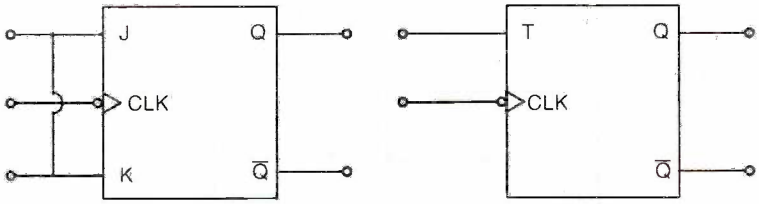
— Fig. 2.138(*d*) shows the symbol for negative edge triggered *J-K* flip-flop with *PR* and CLR.

The *small bubble at CLR indicates negative triggering.*

T flip-flop :

T flip-flop is basically a *J-K* flip-flop, in this circuit input terminals *J* and *K* are connected with each other and this input is named as *T.*

Fig. 2.139(*a*) and (*b*) show the circuit diagram and symbol respectively of a trailing edge triggered *T* flip-flop.



(a) Circuit diagram (b) Symbol

Input T_n	Output T_{n+1}
0	Q_n
1	$\bar{Q}_n$

(c) Truth table

Fig. 2.139. Trailing edge triggered T flip-flop.

— When *low level signal* is applied to the input terminal T, then *initial state of output of flip-flop remains the same.*

— When ***high*** level signal is applied to the input terminal T, output of the flip-flop toggles after arrival of every new clock pulse. So the frequency of output signal is *half* of the clock signal frequency.

This flip-flop can be treated as frequency divider or a device which takes the input frequency at the clock terminal and divide it by *two.*

2.3.17. Counters

— *A **counter** is a sequential circuit that goes through a prescribed sequence of states upon the application of input pulses.*

— The input pulses, called *count pulses,* may be clock pulse or may originate from an external source and may occur at prescribed intervals of true or random.

— The sequence of states in a counter may follow a binary count or any other sequence of states.

— They are used for counting the number of occurrences of an event and are useful for generating time sequences to control operations in a digital system.

Straight binary sequence counter. It is the simple and most straight forward. An n-bit binary counter has n flip-fops and can count in binary from 0 to 2^{n-1}.

Binary ripple couter. It consists of a series connections of T flip-flops without any logic gates. Each flip-flop is triggered by the output of its preceding flip-flop goes from 1 to 0. The signal propagates through the counter in a *Ripple* manner, *i.e.,* the flip-flop essentially changes once at a time in rapid succession. It is the most simplest and most straight forward. It, however, has speed limitations ; an increase in speed can be obtained by the use of a parallel or a synchronous counter.

Synchronous 3-bit binary counter. In this all flip-flops are triggered simultaneously by count pulse. The flip-flop is complemented only if its T input is equal to 1.

Counter-decoder circuits :

— Counters together with decoders are used to generate timing and sequencing signals that control the operation of digital systems.

— The counter-decoder can be designated to give any desired number of repeated timing sequence.

Applications of counters :

The fundamental applications of counters are given below :

1. Measurement of time interval.
2. Direct counting.
3. Measurement of speed.
4. Measurement of frequency.

5. Measurement of distance.
6. Gating a counter.

2.3.18. Registers

A **register** *is a group of memory elements which work together as one unit.* The simple registers only store a binary word. The other registers modify the stored word by shifting its bits to left or right.

The registers can be *classified* as :

(*i*) Accumulator.
(*ii*) General purpose registers.
(*iii*) Special purpose registers.

— A *counter* is a special kind of register to count the number of clock pulses arriving at the input.

2.3.19. Logic Families

The basic building block for digital systems is the *logic gate.* Logic circuits have evolved into *logic families.* Usually a system is fabricated with circuits from one logic famiy. When circuits from more than one family are to be used to implement a given function, it is necessary to ensure that output of one family is *compatible with input of the other.*

The logic famjies are *classified* as follows :

1. *Bipolar families :*
(*i*) DTL (Diode Transistor Logic)
(*ii*) TTL (Transistor Transistor Logic)
(*iii*) ECL (Emitter Coupled Logic)
2. *MOS families :*
(*i*) PMOS (P-channel MOSFET Logic)
(*ii*) NMOS (N-channel MOSFET Logic)
(*iii*) CMOS (Complementary MOSFET Logic)

Note. The PMOS and DTL are now obsolete.

2.3.20. Integrated Circuits

General aspects :

An ***integrated circuit*** *(IC) is a complete electronic circuit in which both the *active (e.g. transistors and FETs) and passive components (e.g. resistance, capacitors and inductors) are fabricated on a tiny single chip of silicon.*

An *IC* is different from a discrete (*i.e.*, distinct or separate) circuit, which is built by *connecting separated devices.* In this case, each device is fabricated separately and then all the devices are assembled together to make an electronic circuit. Discrete circuits have two main *disadvantages :*

(*i*) In a large circuit (*e.g.* T.V. circuit, computer circuit) there may be hundreds of components and consequently discrete assembly would occupy large space,

(*ii*) There will be hundreds of soldered points posing a considerable problem of reliability. To overcome these drawbacks of space conservation and reliability, engineers started a drive for *miniatured circuits.* This led to the development of integrated circuits.

– J.S. Kilby of Texas Instruments was the first person to develop in 1959 an integrated circuit – a single monolithic silicon chip in which active and passive components were fabricated by successive deposition, etching and diffusions. He was soon followed by Robert Noyce of fair-child who successfully fabricated a complete *IC* including the interconnections on a single silicon chip. Since then a lot of progress has been made.

Advantages and disadvantages of Integrated Circuits (ICs) :

As compared to standard printed circuits which use discrete components *ICs* have the following *advantages :*

1. *Exteremely small size* (physical)—Often the size is thousands of time smaller than a discrete circuit.
2. *Very small weight* (owing to miniaturised circuit).
3. *Reduced cost* (since many identical circuits can be built simultaneously on a single wafer).
4. *Extremely high reliability* (IC logic gate has been found to be 100 000 times more reliable than a vacuum tube and 100 times more reliable than a transistor logic gate.
5. *Increased response time and speed.*
6. *Low power consumption* (due to smaller size).
7. *Easy replacement.*
8. *Higher yield* (Because of the batch production, the yield is very high).
9. *Improved functional performance* as more complex circuits can be fabricated for achieving better characteristics.
10. *Greater ability of operating at extreme temperatures.*

Disadvantages :

1. They are *quite delicate* and *cannot withstand rough handling* or *excessive heat.*
2. They function at *fairly low voltages.*
3. They handle only *limited amount of power.*
4. If any component in an *IC* goes out of order, the whole *IC* has to be *replaced* by the new one.
5. It is not possible to produce high power *ICs* (greater than 10 W).
6. Coils or inductors *cannot* be fabricated.
7. There is a *lack of flexibility* in an *IC, i.e.,* it is not generally possible to modify the parameters within which an integrated circuit will operate.
8. In a *IC,* it is neither convenient nor economical to fabricate capacitances exceeding 30 pF. Therefore, for high values of capacitance, discrete components exterior to *IC* chip are connectted.
9. High grade *P-N-P* assembly is not possible.
10. Difficult to produce an *IC* with low noise.
11. Voltage dependence of resistors and capacitors.

Scale of Integration :

The *number of electonic circuits or components, which can be fabricated on a standard size of a silicon chip is called the* **Scale or level of integration.**

The scale of integration is generally classified on two basis : (*i*) The number of *circuits,* and (*ii*) The number of *components.*

The various types of scale of integration are :

1. *Small scale integration* **(SSI) :**

No. of circuits per package .... Less than 12
No. of components .... Less than 50

2. *Medium scale integration* **(MSI) :**
   No. of circuits per package .... Beteween 30 and 100
   No. of components .... Between 50 and 5000
3. *Large scale integration* **(LSI) :**
   No. of circuits per package .... Between 100 and 10 000
   No. of components .... Between 5000 and 100 000
4. *Very large scale integration* **(VLSI) :**
   No. of circuits per package .... Between 10 000 and 100 000
   No. of components .... Between 100 000 and 1 000 000
5. *Ultra large scale integration* **(ULSI) :**
   No. of circuits per package .... Between 100 000 and 1 000 000
   No. of components .... Between 1 000 000 and 10 000 000
6. *Giga scale integration* **(GSI) :**
   No. of circuits per package .... 1 000 000 or more
   No. of components .... over 100 000 000.

Classification of Integrated Circuits :

There are many ways of classifying integrated circuits but the following two classifications are important from subject point of view :

1. *Fabrication or structure :*

(*i*) Monolithic integrated circuits.

(*ii*) Thick and thin film integrated circuits.

(*iii*) Hybrid or multichip integrated circuits.

2. *Application or function :*

(*ii*) Linear (or analog) integrated circuits.

(*ii*) Non-linear (or digital) integrated circits.

Monolithic Integrated Circuits :

The word 'monolithic' means 'single stone' or more appropriately *'a single solid structure'*. In this *IC*, all circuit components (both active and passive) are fabricated variably *within a single continuous piece of silicon crystalline material called* ***water*** (or ***subtrate***). All components are atomically part of the same chip.

Monolithic *ICs* are by *far the most common type of ICs used in practice* because of :

(*i*) *Mass production ;*

(*ii*) *Lower cost ;*

(*iii*) *Higher reliability.*

Commercially available *ICs* of this type can be used is :

— *Amplifiers ;*

— *Voltage regulators ;*

— *A. M. receivers ;*

— *T. V. circuits;*

— *Computer circuits.*

Limitations of monolithic ICs :

(*i*) Low power rating.

(*ii*) Lack of flexibility in circuit design.

(*iii*) Poorer isolation between components.

(*iv*) Small range of values of passive components used in the *ICs*.

(*v*) No possibility of fabrication of inductors.

Thick and thin film Integrated Circuits :

The essential difference between thick-film and thin-film *ICs* is *not their relative thickness* but the *method of depositing the film.* Both have *similar appearance, properties and general characteristics.*

- These devices are larger than monolithic *ICs* but smaller than discretes circuits.
- These *ICs* can be used *when power requirement is comparatively higher.*

MOS Integrated Circuits :

Integrated circuits (*ICs*) based upon the active devices are of the following *two types :*

1. Bipolar *ICs* using *bipolar* active devices such as BJT.
2. Unipolar *ICs* using *unipolar* active devices like FET.

MOS *ICs* based on MOSFET structure find wide applications particularly in *digital field,* because of the following *"advantages"* over bipolar *ICs* :

(*i*) *Fabrication process is simple and cheaper* comparatively.

(*ii*) Occupy less area (the MOS IC typically occupies only 5 percent of the surface required by an expitaxial double-diffused transistor in conventional *IC* ; a MOS resistor occupies 1 percent of the area of a conventional diffused resistor).

(*iii*) *Low power consumption.*

(*iv*) *Less costly* to fabricate.

(*v*) MOS transistor has a *higher bandwidth than bipolar transistor.*

(*vi*) *High packing density.*

Disadvantage :

The major demerit of MOS *ICs* is that their *operating speed is smaller than that of* bipolar *ICs* and as such they are *not suitable for ultra high-speed applications.*

Applications :

MOS *ICs* find wide applications in LSI and VLSI *chips* such as :

— Calculator chips ;

— Memory chips ;

— Micro processors (μP) ;

— Single-chip computers.

IC symbols :

In general, no standard symbol exist for *ICs*. Oftenly, the circuit diagram merely shows a block with numbered terminals. However, sometimes standard symbols are used for operational amplifiers or digital logic gates. Some of the symbols used with *ICs* are shown in Fig. 2.140.

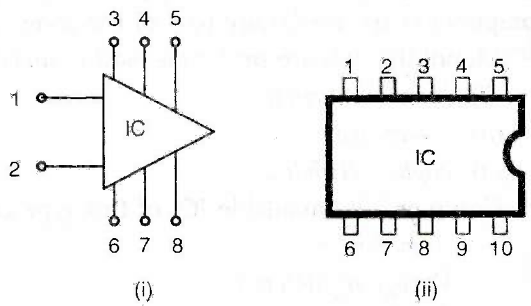
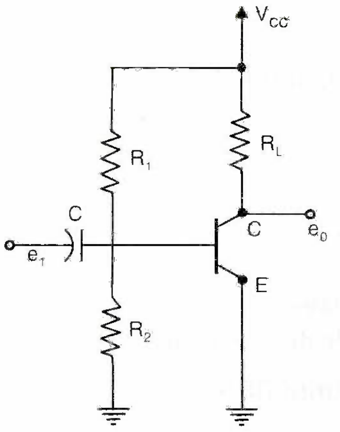


Fig. 2.140. *IC* symbols.

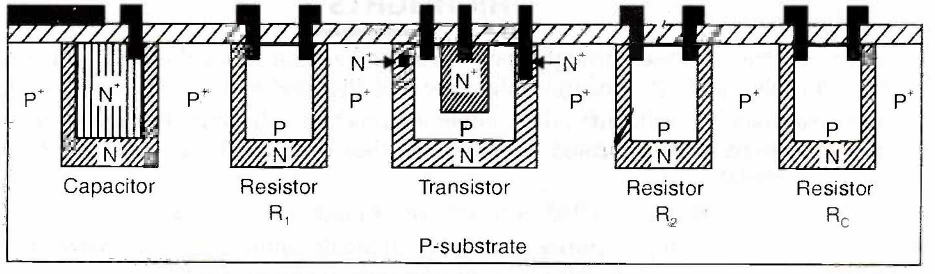
- *IC* symbol does not show the internal circuit.

The Integrated Transistor Amplifier :

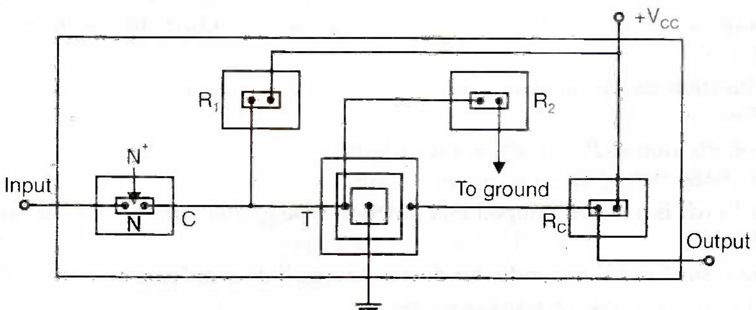
Fig. 2.141. (*i*) shows the schematic diagram of an integrated transistor amplifier ; the cross section view and top view of the interconnections are shown in Fig. 2.141 (*ii*) and (*iii*) respectively.



(*i*) Schematic diagram.



(*ii*) Cross-sectional view showing each element.



(*iii*) Top view showing the interconnections

Fig. 2.141. Integrated transistor amplifier.

The five circuit elements-one capacitor, three resistors and one transistor—and all the interconnections are created by the same masking, etching, and diffusion process. In actual *IC* the circuit elements would not appear in the tandem arrangement shown in Fig. 2.141 (*ii*) and (*iii*) ; rather the *elements would be so placed that as to make optimum use of the available space and to reduce the length of the interconnections to the minimum possible.* The tandem arrangement is used here merely for convenience and classification.

- The total area on the chip covered by this amplifier is only a very small fraction of a square mm.

Applications of ICs :

The popular applications of *ICs* are :

1. *Digital watch ;*

2. *Electronic calculator* ;
3. Pocket *PC* ;
4. Personal digital assistant (PDA) ;
5. MP3 players ;
6. Digital cameras ;
7. Mobile phones ;
8. Digital dictionaries ;
9. Digital translators ;
10. CD (compact disk) player ;
11. DVD (Digital versatile disk) players.

2.3.21. Operational Amplifiers

Refer to Article 4.8.6

HIGHLIGHTS

1. When electricity flows through open space or vacuum as in the case of lightning or vacuum tubes instead of being confined to metallic conductors, it is termed as *electronic*.
2. *Semiconductors* are solid materials, either non-metallic elements or compounds, which allow electrons to pass through them so that they conduct electricity in much the same way as a metal.
3. A pure semiconductor is called *instrinsic semiconductor.*
4. The process of adding impurity (extremely in small amounts, about 1 part in 10^8) to a semiconductor to make it extrinsic (impure) semiconductor is called *doping*.
5. The *N*- and *P*-type materials represent the basic building blocks of semiconductor devices.
6. The outstanding property of *P-N* junction diode to conduct current in one direction only permits it to be used as a *rectifier.*
7. *P-N* junction diodes usually made of germanium or silicon, are commonly used as *power rectifiers*.
8. A properly doped *P-N* junction diode which has a sharp breakdown voltage is known as *Zener diode.*
9. *Tunnel diode* is a heavily doped *P-N* junction type germanium having an extremely narrow junction.
10. A *"transistor"* is a semiconductor device having both rectifying and amplifying properties.
11. The two basic types of transistors are :
    (*i*) Bipolar junction transistor (BJT)
    (*ii*) Field-effect transistor (FET).
12. *Transistor circuit configurations :*
    (*i*) Common-base (CB) configuration
    (*ii*) Common-emitter (CE) configuration
    (*iii*) Common-collector (CC) configuration.
13. A *FET* is a three terminal (namely drain, source, gate) semiconductor device in which current conduction is by only one type of majority carriers (electrons in case of an *N*-channel FET or holes in a *P*-channel *FET*).
14. In a broad sense, following are two main types of FETs :
    (*i*) JFET (*ii*) MOSFET
15. Metal oxide semiconductor FET (MOSFET) is an important semiconductor device and is widely used in many circuit applications. It is also called insulated gate FET (IGFET).

16. The term SCR is often used for the member of the thyristor family which is the most widely used *power-switching* device.
17. A *rectifier* is a circuit which uses one or more diodes to convert A. C. voltage into pulsating D. C. voltage. A rectifier may be half-wave or full-wave.
18. The ratio of D.C. power output to the applied A.C. input power is known as *rectifier efficiency.*
19. The ratio of D.C. power output to the applied A.C. input power is known as *rectifier efficiency.*
20. The A.C. component present in the output is called a *ripple.*
21. The branch of electronics which deals with digital circuits is called *digital electronics.*
22. An electronic circuit that handles only a digital signal is called a *digital circuit.* In digital circuits the following four systems of arithmetic are often used :

    Decimal, Binary, Octal, Hexadecimal.
23. A digital circuit with one or more input signals but only one output signal is called a *logic gate.*

    In the complex circuits, the following *six* different digital electronics gates are used as basic elements :

    (*i*) NOT gate (*ii*) NAND gate
    (*ii*) AND gate (*iv*) OR gate
    (*v*) NOR gate (*vi*) XOR gate.
24. The algebra used to symbolically describe logic functions is called *Boolean algebra.*
25. A *combinational circuit* consists of logic gates whose outputs at any time are determined directly from the combination of inputs without regard for previous input.
26. The *synchronous sequential circuits* are built to operate at *a clocked rate* whereas *asynchronous ones* are *without clocking.*
27. The memory elements used in clocked sequential circuits are called *flip-flops.*
28. A *counter* is a sequential circuit that goes through a prescribed sequence of states upon the application of input pulses.
29. A *integrated circuit* (*IC*) is a complete electronic circuit in which both the active (*e.g.* transistors and FETs) and passive components (e.g. resistors, capacitors and inductors) are fabricated on a tiny single chip of silicon.
30. The number of electronic circuits or components, which can be fabricated on a standard size of a silicon chip is called the *scale or level of integration.*

    The various types of scale of integration are : SSI, MSI, LSI, VLSI, ULSI, GSI.
31. The *ICs* can be classified as follows :

    (*i*) Monolithic integrated circuits.
    (*ii*) Thick and thin-film integrated circuits.
    (*iii*) Hybrid or multichip integrated circuits.

    Or

    (*i*) Linear (or analog) integrated circuits.
    (*ii*) Non-linear (or digital) integrated circuits.

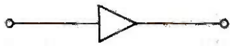
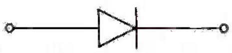
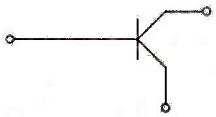
OBJECTIVE TYPE QUESTIONS

Choose the correct answer :

P-N junction diode

1. A *P-N* junction diode has ....

    (*a*) one *P-N* junction (*b*) two *P-N* junctions
    (*c*) three *P-N* junctions (*d*) none of these.

2. A crystal diode has forward resistance of the order of ......
   (a) kΩ (b) Ω
   (c) MΩ (d) none of these.
3. The schematic symbol for a *P–N* junction diode is
   (a)  (b) 
   (c)  (d) 
4. The resistance of a diode is equal to
   (a) ohmic resistance of *P*- and *N*- semiconductors
   (b) junction resistance
   (c) reverse resistance
   (d) algebraic sum of (a) and (b).
5. An ideal crystal diode is one which behaves as a perfect .... when forward biased.
   (a) conductor (b) insulator
   (c) resistance material (d) algebraic sum of (a) and (b)
6. A crystal diode is ....device.
   (a) non-linear (b) bilateral
   (c) linear (d) none of the these.
7. When a diode is reverse biased, it is equivalent to
   (a) an OFF switch (b) an ON switch
   (c) a high resistance (d) none of these.
8. The reverse current in a diode is of the order of ....
   (a) kA (b) mA
   (c) μA (d) A.
9. The conventional current in a P-N junction diode flows
   (a) from positive to negative (b) from negative to positive
   (c) in direction opposite to the electron flow (d) both (a) and (c) above.
10. The leakage current in a crystal diode is due to
    (a) minority carriers (b) majority carriers
    (c) junction capacitance (d) none of the above.
11. The cut in voltage (or knee voltage) of a silicon diode is
    (a) 0.2 V (b) 0.6 V
    (c) 0.8 V (d) 1.0 V.
12. A crystal diode is used as ....
    (a) an amplifier (b) a rectifier
    (c) an oscillator (d) A voltage regulator.
13. If the doping level of a crystal diode is increased, the breakdown voltage ....
    (a) remains the same (b) is increased
    (c) is decreased (d) none of the above.
14. The knee voltage of a crystal diode is approximately equal to ....
    (a) applied voltage (b) breakdown voltage
    (c) forward voltage (d) barrier potential

15. A Zener diode has ....
(a) one P-N junction
(b) twoP-N junctions
(c) three P-N junctions
(d) none of these.

16. A Zener diode is always .... connected.
(a) reverse
(b) forward
(c) either reverse or forward
(d) none of these.

17. A Zener diode is used as ....
(a) an amplifier
(b) a voltage regulator
(c) a rectifier
(d) a multivibrator.

18. In the breakdown region, a Zener diode behaves like a .... source.
(a) constant voltage
(b) constant current
(c) constant resistance
(d) none of these.

19. The doping level in a Zener diode is .... that of a crystal diode.
(a) the same as
(b) less than
(c) more than
(d) none of these.

20. A Zener diode is .... device.
(a) a non-linear
(b) a linear
(c) an amplifying
(d) none of these.

Transistors (BJP, FET, etc.)

21. In P-N-P transistor, base will be of
(a) P material
(b) N material
(c) either of the above
(d) none of these.

22. In a transistor symbol, slant line to bar without any arrow head represents
(a) emitter
(b) base
(c) collector
(d) none of these.

23. In a transistor symbol, slant line to the bar with arrow head represnts
(a) emitter
(b) base
(c) collector
(d) none of these.

24. In a transistor highly doped part is
(a) emitter
(b) base
(c) collector
(d) none of these.

25. In a transistor lightly doped part is
(a) emitter
(b) base
(c) collector
(d) none of these.

26. In a transistor largest dimension is that of
(a) emitter
(b) base
(c) collector
(d) none of these.

27. A dot near the transistor pin denotes
(a) emitter
(b) base
(c) collector
(d) none of these.

28. In a transistor symbol, if slant line arrow head is drawn towards the bar, then the transistor is.
(a) P-N-P
(b) N-P-N
(c) either of these
(d) none of these.

29. A notch or a tab on the transistor cap denotes
(a) emitter pin (b) base pin
(c) collector pin (d) none of these.
30. Resistance across which of the following two pairs of transistor be nearly equal?
(a) Emitter base, emitter collector (b) Emitter base, base collector
(c) Base collector, collector emitter (d) None of these.
31. Which of the following is valid for both P-N-P as well as N-P-N transistors?
(a) The emitter injects holes into the base region
(b) The electrons are the minority carriers in the base region
(c) The EB junction is forward biased for active operation
(d) When biased in the active region, current flows into emitter terminal.
32. In a transistor with normal bias
(a) the emitter junction offers high resistance (b) the emitter junction is reverse biased
(c) the emitter junction has a low resistance (d) none of the above.
33. Which region of a transistor is lightly doped?
(a) Collector (b) Base
(c) Emitter (d) All regions are equally doped.
34. Semiconductor is a material which
(a) allows one type of carriers to pass through it
(b) has conductivity greater than insulator
(c) allows current to flow in one direction but not in the opposite direction
(d) none of these.
35. ..... is the region of a transistor which has highest conductivity.
(a) Base (b) Emitter
(c) Collector (d) Any of the above.
36. Bipolar transistor is a
(a) three terminal semiconductor device (b) three layer semiconductor device
(c) three junction semiconductor device (d) none of these.
37. Silicon controlled rectifier belongs to
(a) diode family (b) triode family
(c) thyristor family (d) non of these.
38. Current flow through a bipolar transistor is by means of
(a) electrons (b) holes
(c) both electrons and holes (d) none of these.
39. Transistor works as an open switch when emitter junction is .....biased and collector junction is ..... biased.
(a) forward, reverse (b) reverse, reverse
(c) reverse, forward (d) forward, forward.
40 Transistor works as a closed switch when emitter junction is ..... biased and collector junction is .....biased.
(a) forward, reverse (b) reverse, reverse
(c) reverse, forward (d) forward, forward.
41. Transistor works as a variable rheostat when emitter junction is ..... biased and collector junction is ..... biased.
(a) forward, reverse (b) reverse, reverse
(c) reverse, forward (d) forward, forward.

42. With two diodes connected back to back with emitter diode forward biased, and collector diode reverse biased
    (*a*) emitter and collector currents are nearly equal and base current is very small
    (*b*) emitter and base currents are nearlyequal and collection current is very small
    (*c*) base and collector currents are nearly equal and emitter current is large
    (*d*) any of the above.
43. Current base part of a transistor behaves like
    (*a*) constant current source
    (*b*) forward biased diode
    (*c*) a resistance
    (*d*) none of the above.
44. Majority carriers emitted by the emitter
    (*a*) mostly recombine in base region
    (*b*) mostly pass through the base region
    (*c*) are stopped by the collector junction barrier
    (*d*) recombine in the collector region.
45. The following relationship between α and β are correct *except*
    (*a*) $1 - \alpha = \dfrac{1}{1+\beta}$
    (*b*) $\alpha = \dfrac{\beta}{1+\beta}$
    (*c*) $\beta = \dfrac{\alpha}{1-\alpha}$
    (*d*) $\alpha = \dfrac{\beta}{1-\beta}$
46. A P-N-P transistor has
    (*a*) only acceptor ions
    (*b*) only donor ions
    (*c*) two-P-regions and one N-region
    (*d*) three P-N junctions.
47. Regarding common emitter configuration which of the following statements is *incorrect*?
    (*a*) Its output resistance is very high.
    (*b*) It is the only circuit which has voltage and current gains higher than unity.
    (*c*) Its power gain is the best.
    (*d*) It is the only configuration which provides inversion.
48. The active region of the output characteristics for a common base transistor is that in which
    (*a*) emitter is forward-biased but collector is reverse-biased
    (*b*) both emitter and collector are forward-biased
    (*c*) collector is forward-biased and emitter is reverse-biased
    (*d*) both emitter and collector are reversed-biased.
49. The set of transistor characteristics that enables α to be determined directly from the slope is ..... characteristics.
    (*a*) common emitter transfer
    (*b*) common emitter output
    (*c*) common base transfer
    (*d*) common base input.
50. When a common emitter transistor is cut off which of the following happens?
    (*a*) Maximum voltage appears across the collector.
    (*b*) Maximum collector current flows.
    (*c*) Minimum voltage appears across the collector.
    (*d*) Miximum voltage appears across the load resistor.
51. In amplifier circuit, biasing of transistor is necessary to
    (*a*) fix the value of current amplification
    (*b*) establish suitable D.C. workig conditions
    (*c*) ensure that transistor is saturated
    (*d*) ensure that transistor is cut off.

52. The configuration in which current gain of transistor amplifier is lowest is
    (*a*) common base (*b*) common collector
    (*c*) common emitter (*d*) any of these.
53. The configuration in which voltage gain of transistor amplifier is lowest is
    (*a*) common base (*b*) common collector
    (*c*) common emitter (*d*) any of these.
54. The configuration in which input impedance of transistor amplifier is lowest is
    (*a*) common base (*b*) common collector
    (*c*) common emitter (*d*) any of these.
55. The configuration in which output impedance of transistor amplifier is highest is
    (*a*) common base (*b*) common collector
    (*c*) common emitter (*d*) any of these.
56. The configuration which provides both high current gain and high voltage gain of transistor amplifier is
    (*a*) common base (*b*) common collector
    (*c*) common emitter (*d*) any of these.
57. A transistor-terminal current is positive when the
    (*a*) current is due to flow of electrons (*b*) current is due to flow of holes
    (*c*) electrons flow into the transistor at the terminal
    (*d*) electrons flow out of the transistor at the terminal.
58. A transistor is said to be in quiescent state when
    (*a*) no signal is applied to the input (*b*) no currents are flowing
    (*c*) it is unbiased
    (*d*) emitter junction and collector-junction biases are equal.
59. The most noticeable effect of a small increase in temperature in the CE transistor is the
    (*a*) increase in the A.C. current gain (*b*) decrease in the A.C. current gain
    (*c*) increases in I_{CEO} (*d*) increases in the output resistance.
60. FETs have similar properties to
    (*a*) thermionic valves (*b*) unijunction transistor
    (*c*) P-N-P transistors (*d*) N-P-N transistors.
61. A JFET can operate in
    (*a*) depletion mode only (*b*) enhancement mode only
    (*c*) depletion and enhancement modes
    (*d*) neither enhancement nor depletion mode.
62. In a JFET ..... is usually the point of reference.
    (*a*) gate (*b*) drain
    (*c*) source (*d*) either (*b*) or (*c*).
63. The primary control on drain current in a JFET, is exerted by which of the following?
    (*a*) Gate reverse bias (*b*) Channel resistance
    (*c*) Voltage drop across channel (*d*) Size of depletion regions.
64. For the operation of enhancement-only N-channel MOSFET, value of gate voltage has to be
    (*a*) zero (*b*) low positive
    (*c*) high positive (*d*) high negative.

65. Which of the following statements is *correct* regarding a JFET operating above pinch-off voltage?
   (*a*) The depletion regions become smaller.
   (*b*) The drain current starts decreasing.
   (*c*) The drain current remains practically constant.
   (*d*) The drain current increases steeply.
66. A FET can be used as a variable
   (*a*) inductor
   (*b*) capacitor
   (*c*) resistor
   (*d*) voltage source
   (*e*) current source.
67. In FET the drain voltage above which there is no increases in the drain current is called ..... voltage
   (*a*) pick off
   (*b*) pinch off
   (*c*) breakdown
   (*d*) critical.
68. The operation of JFET involves a flow of
   (*a*) minority carriers
   (*b*) majority carriers
   (*c*) recombination carriers
   (*d*) any of the above.
69. When the positive voltage on the gate of a P-channel JFET is increased, the drain current will
   (*a*) increase
   (*b*) decrease
   (*c*) remain the same
   (*d*) any of the above.
70. A FET differs from a bipolar transistor as it has
   (*a*) simpler fabrication
   (*b*) negative resistance
   (*c*) high input impedance
   (*d*) any of the above.
71. A FET, for its operation, depends on the variation of
   (*a*) forward-biased junction
   (*b*) reversed-biased junction
   (*c*) magnetic field
   (*d*) The depletion-layer width with reverse voltage.
72. N-channel FETs are superior to P-channel FETs because
   (*a*) they have a higher switching time
   (*b*) they have a higher input impedance
   (*c*) mobility of electrons is greater than that of holes
   (*d*) all of the above.
73. UJT is also called a
   (*a*) transistorized junction
   (*b*) voltage controllable device
   (*c*) current controllable device
   (*d*) relaxation oscillator.
74. The difference between a thyristor and a silicon diode is that the thyristor
   (*a*) conducts when it is triggered in addition to being forward biased
   (*b*) conducts when it is forward biased
   (*c*) blocks when it is reverse biased
   (*d*) none of the above.
75. On stopping the gate pulse to a SCR
   (*a*) it will stop conduction
   (*b*) it will continue conduction in the same direction
   (*c*) it will continue conduction in the opposite direction
   (*d*) none of the bove will happen.

76. SCR is used for current control in
(*a*) D.C. circuit only
(*b*) A.C. circuit only
(*c*) both (*a*) and (*b*)
(*d*) none of these.
77. SCR is
(*a*) three layer two terminal device
(*b*) three layer three terminal device
(*c*) four layer two terminal device
(*d*) four layer three terminal device.
78. A triac is a ..... switch.
(*a*) undirectional
(*b*) bi-directional
(*c*) either of these
(*d*) none of these.
79. A diac is ..... switch.
(*a*) an A.C.
(*b*) D.C.
(*c*) either of these
(*d*) none of these.
80. The normal way to turn on a diac is by
(*a*) gate current
(*b*) breakover voltage
(*c*) either of these
(*d*) none of these.
81. A diac is equivalent to a
(*a*) triac with two gates
(*b*) diode and two resistors
(*c*) pair of SCRs
(*d*) pair of four-layer SCRs.
82. Regarding triac which of the following statements is *incorrect*?
(*a*) It is not particularly suited for A.C. or mains power control.
(*b*) It is a 5-layer bi-directional semiconductor device.
(*c*) Any one of its main terminals can be used either as cathode or as anode.
(*d*) It can be triggered in response to both positive and negative gate terminals.
83. ..... is the best electronic device for fast switching.
(*a*) MOSFET
(*b*) JFET
(*c*) BJT
(*d*) Triode.
84. Which semiconductor device acts like a diode and two resistors ?
(*a*) UJT
(*b*) SCR
(*c*) diac
(*d*) triac.
85. ..... is the device which acts like an N-P-N and P-N-P transistor connected base-to-base and emitter-to-collector.
(*a*) SCR
(*b*) UJT
(*c*) diac
(*d*) triac.
86. An SCR may be considered to be diodes back-to-back consisting of an anode, cathode and .....
(*a*) two, plate
(*b*) three, plate
(*c*) three, gate
(*d*) four, base.
87. A LASCR in just like a conventional SCR except that it
(*a*) has no gate terminal
(*b*) can also be light-triggered
(*c*) cannot carry large current
(*d*) cannot be pulse-triggered.
88. Which semiconductor device behaves like two SCRs?
(*a*) MOSFET
(*b*) JFET
(*c*) UJT
(*d*) Triac.
89. An SCR conducts appreciable current when
(*a*) gate is negative and anode is positive with respect to cathode
(*b*) anode is negative and gate is positive with respect to cathode

(c) anode and gate are both negative with respect to cathode
(d) anode and gate are both positive with respect to cathode.

90. An SCS has which of the following?
(a) One anode, one cathode and two gates (b) Two anodes and two gates
(c) Four layers and three terminals (d) Three layers and four terminals.

91. An SCS may be switched ON by a
(a) positive pulse at its anode gate G_1 (b) positive pulse at its cathode gate G_2
(c) negative pulse at its cathode (d) positive pulse at its anode.

92. Which of the following methods used for protecting MOSFET against damage from stray voltage developing at the gate is *incorrect*?
(a) Only source terminal is earthed during transit.
(b) Back-to-back Zener diodes are formed into the monolithic structure of MOSFET.
(c) Grounding rings are used which are removed only when it is wired securely into the circuit.
(d) It is inserted into conducting sponge during visit.

93. Regarding MOSFET which of the following statements is *incorrect*?
(a) It can operate in depletion mode.
(b) It can operate in enhancement mode.
(c) It can operate in depletion and enhancement modes.
(d) It can operate in depletion-only-mode.
(e) It can operate in enhancement-only mode.

94. The main factor which differentiates a DE MOSFET from an E-only MOSFET is the absence of
(a) P-N junctions (b) insulated gate
(c) electrons (d) channel.

95. The input gate current of a FET is
(a) a few amperes (b) a few milliamperes
(c) a few microamperes (d) negligibly small.

96. Silicon devices are preferred at high temperature operations as compared to germanium because
(a) silicon can dissipate more power
(b) reverse saturation current is less in case of silicon
(c) silicon is more thermally stable
(d) all of the above.

97. Hall effect can be used to measure
(a) carrier concentration (b) electric field intensity
(c) magnetic field intensity (d) none of the above.

98. Which of the following statements is *correct* in case of a properly biased transistor?
(a) The emitter to base depletion region is small and collector to base depletion region is large
(b) The emitter to base depletion region is large and collector to base deplection region is small
(c) both depletion regions are small
(d) both depletion regions are large.

99. Ebers-Moll equations for transistors provide
(a) true terminal currents regardless of junction biases
(b) true terminal voltages dependent on junction biases

(*c*) separate input and output circuits
(*d*) all of the above.

100. In the symbols of P-N-P transistors and N-P-N transistor the arrow on the emitter shows the direction of flow of
(*a*) electrons, electrons (*b*) holes, holes
(*c*) holes, electrons (*d*) electrons, holes.

ANSWERS

1. (*a*)	2. (*b*)	3. (*b*)	4. (*d*)	5. (*a*)	6. (*a*)	7. (*a*)
8. (*c*)	9. (*d*)	10. (*a*)	11. (*b*)	12. (*b*)	13. (*c*)	14. (*d*)
15. (*a*)	16. (*a*)	17. (*b*)	18. (*a*)	19. (*c*)	20. (*a*)	21. (*b*)
22. (*c*)	23. (*a*)	24. (*a*)	25. (*b*)	26. (*c*)	27. (*c*)	28. (*a*)
29. (*a*)	30. (*b*)	31. (*c*)	32. (*c*)	33. (*b*)	34. (*b*)	35. (*b*)
36. (*b*)	37. (*c*)	38. (*c*)	39. (*b*)	40. (*d*)	41. (*a*)	42. (*b*)
43. (*a*)	44. (*b*)	45. (*d*)	46. (*c*)	47. (*a*)	48. (*a*)	49. (*c*)
50. (*a*)	51. (*a*)	52. (*a*)	53. (*b*)	54. (*a*)	55. (*a*)	56. (*c*)
57. (*d*)	58. (*a*)	59. (*c*)	60. (*a*)	61. (*a*)	62. (*c*)	63. (*a*)
64. (*c*)	65. (*c*)	66. (*c*)	67. (*b*)	68. (*b*)	69. (*b*)	70. (*c*)
71. (*d*)	72. (*c*)	73. (*b*)	74. (*a*)	75. (*b*)	76. (*c*)	77. (*d*)
78. (*b*)	79. (*a*)	80. (*b*)	81. (*d*)	82. (*a*)	83. (*c*)	84. (*a*)
85. (*a*)	86 (*c*)	87. (*a*)	88. (*d*)	89. (*d*)	90. (*a*)	91. (*b*)
92. (*a*)	93. (*d*)	94. (*d*)	95. (*d*)	96. (*b*)	97. (*c*)	98. (*a*)
99. (*a*)	100. (*b*).					

THEORETICAL QUESTIONS

1. Define a 'semiconductor'.
2. List the important characteristics of semiconductors.
3. Give examples of semiconducting materials.
4. What is the difference between a semiconductor and an insulator?
5. What is an intrinsic semiconductor?
6. What do you mean by the term *doping*?
7. How does an extrinsic semiconductor differ from an intrinsic semiconductor?
8. Explain the structure of a *P*-type semiconductor with help of neat sketches.
9. Explain briefly about 'atomic binding in semiconductors'.
10. How are holes formed in semiconductors?
11. Derive an expression for electron conductivity of a metal.
12. Derive expressions for conductivity of *N*-type and *P*-type semiconductors.
13. What do you mean by conductivity modulation ?
14. Explain briefly the following :
(*i*) Thermistors and sensitors (*ii*) Photoconductors.
15. List the applications of semiconductor materials.
16. How is germanium prepared?
17. What is a *P-N* junction diode? How its terminals are identified?
18. Draw the *V-I* characteristics of a junction diode when it is (*a*) forward biased and (*b*) reverse biased.

19. Draw the graphical symbol of a crystal diode and explain its significance. How the polarities of function diode are identified ?
20. Draw the equivalent circuit of a crystal diode.
21. What is an ideal diode and a real diode?
22. Explain the following terms :
    (i) Static resistance
    (ii) Bulk resistance
    (iii) Junction resistance
    (iv) A.C. or dynamic resistance
    (v) Reverse resistance of a diode.
23. What are the important applications of a diode?
24. Write a short note on the power and current ratings of a diode.
25. What is a Zener diode ? Draw its equivalent circuit.
26. Explain briefly the applications of a Zener diode.
27. What do you understand by Zener voltage?
28. Explain why Zener diode is always operated in reverse biasing.
29. Explain how a Zener diode can stabilize the voltage across the load.
30. Explain the process of Zener breakdown.
31. Draw and explain a Zener diode voltage regulator.
32. Define the term 'Transistor'.
33. What are the various types of transistors?
34. Explain the function of emitter in the operation of a junction transistor.
35. What is the significance of arrow in the transistor symbol?
36. Why is emitter wider than collector and base?
37. Why is base made thin?
38. Draw *N-P-N* and *P-N-P* transistors.
39. Explain the working of a *P-N-P* transistor.
40. Differentiate between *P-N-P* and *N-P-N* transistors. Why are collector and emitter currents nearly equal in these transistors?
41. Define α and β of a transistor and derive the relationship between them.
42. Draw three basic configurations of *N-P-N* transistor.
43. Draw input and output characteristics of *CB* transistor configuration.
44. Draw the circuits of the various transistor configurations. List their important features. Why *CE* configuration is mainly used?
45. Explain the construction and working of a JFET.
46. What is the difference between a JFET and a BJT?
47. How will you determine the drain characteristics of JFET? What do they indicate?
48. What are the advantages and disadvantages of JFET?
49. What are the applications of FETs?
50. What is the difference between MOSFET and JFET?
51. Define the following terms for a JFET :
    (i) The pinch-off voltage.
    (ii) Channel ohmic resistance.
    (iii) Drain resistance.
52. Draw the *V-I* characteristics of an *N*-channel FET.
53. Discuss briefly, the construction, working, characteristics and applications of SCR.

54. Explain the forward and reverse characteristics of a thyristor.
55. What is the difference between SCR and Triac?
56. List the applications of thyristor?
57. Describe a half-wave rectifier using a crystal diode.
58. Derive and expression for the efficiency of a half-wave rectifier.
59. With neat sketch, explain the working of the following :
    (*i*) Centre-tapped full-wave rectifier.
    (*ii*) Full-wave bridge rectifier.
60. Derive an expression for the efficiency of a full-wave rectifier.
61. What is a ripple factor? What is its value for a half-wave and a full-wave rectifier?
62. What is 'digital electronics'?
63. State the advantages and disadvantages of digital electronics?
64. What is a 'digital circuit'?
65. Why binary system is preferred in 'digital system'?
66. Discuss the importance of 1's and 2's complement numbers.
67. Explain the Gray code and alphanumeric codes.
68. What is meant by a radix (or base of a number system)?
69. Draw the diagram of a clocked R-S flip-flop and give the truth table.
70. Show that a R-S flip-flop results when two NOR gates are cross-coupled.
71. What is a flip-flop? Explain the principle of operation of S-R flip-flop with truth table.
72. With the aid of a neat sketch, explain the operation of J-K flip-flop.
73. Briefly describe J-K, D- and T-type flip-flops.
84. Write short notes on "logic families".

CHAPTER

3 Sensors and Transducers

3.1 Introduction; 3.2 Mechanical detector-transducer elements; 3.3 Definition of transducer; 3.4 Classification of transducers – Transducer sensitivity – Specification for transducers; **3.5 Electromechanical transducers; 3.6 Transducer actuating mechanisms; 3.7 Resistance transducers** – Linear and angular motion potentiometers – Thermistors and resistance thermometers – Wire resistance strain gauges; **3.8 Variable inductance transducers** – Self generating type – Electromagnetic type – Electrodynamic type – Eddy current type – Passive type – Variable reluctance transducer – Mutual inductance transducer – Linear variable–differential transformer (*LVDT*); **3.9 Capacitive transducers** – Capacitive transducers – Using change in area of plates – Capacitive transducers – Using change in distance between the plates; **3.10 Piezoelectric transducers** – Piezoelectric materials – desirable properties of piezoelectric materials – working of a piezoelectric device – advantages and disadvantages of piezoelectric transducers; **3.11 Hall effect transducers** – Hall effect – Hall effect transducers; **3.12 Thermoelectric transducers; 5.13 Photoelectric transducers** – principle of operation – applications – classification – Photoemissive cell – Photovoltaic cell – Photoconductive cell; **3.14 Strain gauges** – Types of strain gauges – Wire wound strain gauges – Foil strain gauges – Semiconductor strain gauges – Capacitive strain gauges – Theory of strain gauges; **3.15 Load cells; 3.16 Proximity sensors; 3.17 Pneumatic sensors; 3.18 Light sensors; 3.19 Digital optical encoder; 3.20 Recent trends** – Smart pressure transmitters; **3.21 Selection of sensors; 3.22 Static and dynamic characteristics of transducers** – Measurement systems – Instruments. Highlights – Objective Type Questions – Theoretical Questions – Unsolved Examples.

3.1 INTRODUCTION

The *primary sensing element* (*sensor*) is the first and foremost requirement for measurement and automatic controls. The *sensors sense the condition, state or value of the process variable and produce on output which reflects this condition, state or value. The* **transducers** *transform the energy of the process variable to an output of some other type of energy which is able to operate some control device.* Sometimes a *secondary transducer* may be employed to transform the output of the primary sensor to still another type of energy.

Examples :

- In the ordinary *dial indicator* the indicating spindle acts as a sensor/detector for displacement. It simply performs the function of sensor/detector and nothing else.
- The function of a *Bourdon tube* of a pressure gauge is *twofold*: Firstly to sense the pressure and secondly to give the resulting effect or output in the form of displacement. Here the tube acts a sensor/detector transducer.

- In a *compressive load cell*, the platform detects the force and gives an output in the form of deflection. This deflection may be further converted into an electrical output by strain gauges (called *secondary transducer*).

For the measurement of particular quantity, different types of sensors and transducers are available and the choice of a suitable unit depends upon the static and dynamic performance characteristics.

3.2 MECHANICAL DETECTOR-TRANSDUCER ELEMENTS

The various mechanical detector-transducer elements may be enumerated and discussed as follows:

1. Elastic members/elements
2. "Mass" sensing elements
3. Thermal detectors
4. Hydro-pneumatic elements.

1. Elastic members/elements:

These elements *work on the principle of direct tension or compression, bending and torsion.* These are invariably used to change force into displacement. The following elastic members/elements are commonly used :

(*i*) *Proving ring (stress ring).* It is a ring of known physical dimension and mechanical properties. An external tensile or compressive force applied across the ring diameter causes distortion which is proportional to that force. The distortion is measured by means of a dial gauge, a sensitive micrometer, or a strain gauge.

The proving rings have been used as *standards for calibrating tensile testing machines and for accurate measurement of large plastic loads.*

(*ii*) *Elastic torsion member.* Several times torque meters make use of elastic torsion members which twist in proportion to applied torque and deformation is used as a measure of torque.

(*iii*) *Springs.* In a spring type indicating scale, unknown weight applied to the free end of spring causes displacement which is indicated by the pointer.

(*iv*) *Bourdon tube, bellows, diaphragm.* Most pressure measuring devices use either a Bourdon tube, bellows or diaphragms. The action of these devices is based on the elastic deformation brought about by the force resulting from pressure summation.

2. "Mass" sensing elements:

- The *inertia of a concentrated mass* provides another basic mechanical detector-transducer element, which is used in the *accelerometers* and *vibration pick-ups* and serves to measure the characteristics of dynamic motion (e.g., displacement, velocity, acceleration, frequency, etc.) through application of Newton's second law of motion.
- Any simple mechanically vibrating member (*e.g.*, a *pendulum*) would sever as a time or frequency transducer, chopping the passage of time into discrete bits.
- Further the *manometer*, used for pressure measurement, also works on the *principle of mass displacement.*

3. Thermal detectors:

These are the device employed to measure the *temperature* of solids, liquids and gases. They sense the temperature by employing one of the following *primary effects:*

(*i*) Change in physical stage; (*ii*) Change in chemical state;
(*iii*) Change in electrical properties; (*iv*) Change in radiating ability.

The following *thermal detectors* are most commonly used

(*i*) Glass thermometers
(*ii*) Pressure gauge thermometers
(*iii*) Bimetallic thermometers
(*iv*) Resistance thermometers
(*v*) thermistors
(*vi*) Pyrometers
(*vii*) Thermocouples.

4. Hydro-pneumatic sensors:

Following are the *common examples* of the hydro-pneumatic sensors.

(*a*) ***Applied to static conditions:***

(*i*) *Simple float.* A simple float converts the fluid level into displacement; it makes no allowance for change in the density of the supporting liquid.

(*ii*) *Hydrometer.* It senses specific gravity and converts it into displacement. It uses the immersion depth as a means for detecting variation in specific gravity of the supplying liquid.

(*b*) ***Applied to dynamic conditions:***

(*i*) *Orifices and venturies.* These are used for flow measurement in pipes and provide information in the form of pressure change as a result of transformation of energy.

(*ii*) *Pitot tube.* It measures the pressure resulting from total-flow rate rather than the change of rate.

(*iii*) *Vanes in the form of air foils or turbine wheels.* These are also used to sense fluid flow.

3.3 DEFINITION OF TRANSDUCER

A broad definition of a transducer is as follows:

"A **transducer** *is a device which converts the energy from one form to another".*

Most of the transducers either convert electrical energy into mechanical displacement and/or convert some non-electrical physical quantity (*e.g.*, force, sound, temperature etc.) to an electrical signal.

A transducer performs the following *functions* in an electronic instrumentation system :

1. *Detects or senses the presence, magnitude and changes in physical quantity being measured.*
2. *Provides a proportional electrical output signal (see Fig. 3.1.)*

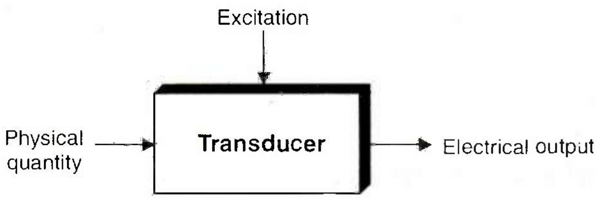


Fig. 3.1. Transducer.

- A **transducer** can be broadly defined as a *device which converts a non-electrical quantity into an electrical quantity.*

An **inverse transducer** *is defined as a device which converts an electrical quantity into a non-electrical quantity.* It is a precision actuator which has an electrical input and a low power non-electrical output. A *piezoelectric* crystal acts as an inverse transducer because when a voltage is applied across its surfaces, it changes its dimensions causing a mechanical displacement.

3.4 CLASSIFICATION OF TRANSDUCERS

A. Transducers are broadly *classified* into two groups as follows:

1. ***Active transducers.*** They are also known as *self-generating type transducers.* These transducers *develop their own voltage or current. The energy required for production of an output signal is obtained from the physical phenomenon being measured.*

   Examples: *Thermocouples and thermopiles, piezoelectric pick-up, photovoltaic cell.*

2. ***Passive transduces.*** *They are known as externally-powered transducer.* These transducers *derive the power required for the energy conversion from an external power source.* However, they *may absorb some energy from the physical phenomenon under study.*

   Examples: *Resistance thermometers and thermistors, potentiometric devices, differential transformer, photoemission cell etc.*

B. Classification based on the type of output :

1. ***Analogue transducers.*** These transducers convert the input physical phenomenon into an analogous output which is a continuous function of time.

   Examples: *Strain gauge, a thermocouple, a thermistor or an LVDT* (linear voltage differential transformer).

2. ***Digital transducers.*** These transducers convert the input physical phenomenon into an electrical output which may be in form of pulse.

C. Classification based on electrical principle involved :

1. ***Variable-resistance type :***
   (*i*) Strain and pressure gauges.
   (*ii*) Thermistors, resistance thermometers.
   (*iii*) Photoconductive cell etc.
2. ***Variable-inductance type :***
   (*i*) Linear voltage differential transformer (LVDT).
   (*ii*) Reluctance pick-up.
   (*iii*) Eddy current gauge.
3. ***Variable-capacitance type :***
   (*i*) Capacitor microphone.
   (*ii*) Pressure gauge.
   (*iii*) Dielectric gauge.
4. ***Voltage-generating type :***
   (*i*) Thermocouple.
   (*ii*) Photovoltaic cell.
   (*iii*) Rotational motion tachometer.
   (*iv*) Piezoelectric pick-up.
5. ***Voltage-divider type :***
   (*i*) Potentiometer position censor.
   (*ii*) Pressure-actuated voltage divider.

Table 3.1. shows the measurements versus transduction methods.

- While *describing a particular transducer* the information must be available about the following aspects :

(*i*) The measurand.

(*ii*) The sensing element which responds directly to the measurand.

(*iii*) The principle of operation of the transducer and where the output of the transducer originates.

(*iv*) The useful range.

Table 3.1. Measurements versus Transduction Methods

S. No.	Quantity to be measured	Type of transducer	S. No.	Quantity to be measured	Type of transducer
1.	*Displacement*	— Resistive — Inductive — Capacitive — Piezoelectric — Magnetoelectric — Radioactive — Electron tube.	7.	*Pressure*	— Resistive — Inductive — Capacitive — Piezoelectric — Thermoelectric — Magnetoelectric — Magnetostrictive — Radioactive — Electron tube.
2.	*Thickness*	— Inductive — Capacitive — Piezoelectric — Photoelectric — Radioactive.	8.	*Flow*	— Resistive — Inductive — Capacitive — Piezoelectric — Magnetoelectric — Radioactive.
3.	*Velocity*	— Resistive — Inductive — Capacitive — Piezoelectric — Photoelectric — Magnetoelectric — Radioactive — Electron tube.	9.	*Level*	— Resistive — Capacitive — Piezoelectric — Photoelectric — Radioactive.
4.	*Acceleration*	— Resistive — Inductive — Capacitive — Piezoelectric — Magnetoelectric — Electron tube.	10.	*Temperature*	— Resistive — Photoelectric — Thermoelectric — Radioactive.
5.	*Mass*	— Inductive — Piezoelectric — Magnetoelectric — Radioactive.	11.	*Humidity*	— Resistive — Capacitive.
6.	*Force*	— Resistive — Inductive — Piezoelectric — Radioactive.	12	*Viscosity*	— Resistive — Capacitive — Piezoelectric — Magnetostrictive.

- While *selecting a dector-transducer element*, the following major consideration need to be looked into:

  (*i*) Mechanical suitability in terms of

  — Physical size, weight and shape;

— Mounting arrangement;
— Ruggedness.

(ii) Electrical suitability in terms of:
— Sensitivity ;
— Frequency response ;
— Ease of signal transmission.

(iii) Environmental suitability in terms of
— Sensitivity to temperature and self-heating effects;
— Magnetic fields;
— Vibration; dust and humidity;
— Supply frequency etc.

(iv) Transducer performance in terms of calibration accuracy.

(v) Desired measurement accuracy and range, power requirements, overload protection and vulnerability to sudden failure.

(vi) Purchase aspects.

3.4.1. Transducer Sensitivity

The relationship between the measurand and the transducer output signal is referred to as "transducer sensitivity".

i.e., $$\text{Transducer sensitivity} = \frac{\text{Output signal increment}}{\text{Measuand increment}}$$

Sensitivity of a transducer should be *usually as high as possible* since then it becomes easier to take the measurements.

3.4.2. Specifications for Transducers

While selecting the proper transducer for any applications, or ordering the transducers, the following specifications should be thoroughly considered:

(i) Ranges available.
(ii) Squaring system.
(iii) Sensitivity.
(iv) Maximum working temperature.
(v) Method of cooling employed.
(vi) Mounting details.
(vii) Maximum depth.
(viii) Linearity and hysteresis
(ix) Output for zero input.
(x) Temperature coefficient of zero drift.
(xi) Natural frequency.

3.5 ELECTRO-MECHANICAL TRANSDUCERS

These days electrical/electronic techniques of measurement are being increasingly applied to the measurements in many fields other than in electrical engineering. These methods claim the following *advantages* :

Advantages :

1. Less power consumption and less loading on the system to be measured.
2. Friction and mass inertia effects minimum.
3. More compact instrumentation.
4. Possibility of non-contact measurements.

5. Good frequency and transient response.
6. Feasibility of remote indication and recording.
7. Amplification greater than that produced by a mechanical contrivance.
8. Possibility of mathematical processing of signals like summation, integration etc.

3.6 TRANSDUCERS ACTUATING MECHANISMS

Transducers are also known as *gauges, pick-ups and signal generators.* Most of the pick-ups have following *two basic elements* :

(*i*) Activating device.
(*ii*) Transducing element.

Fig. 3.2. shows some typical actuating mechanisms

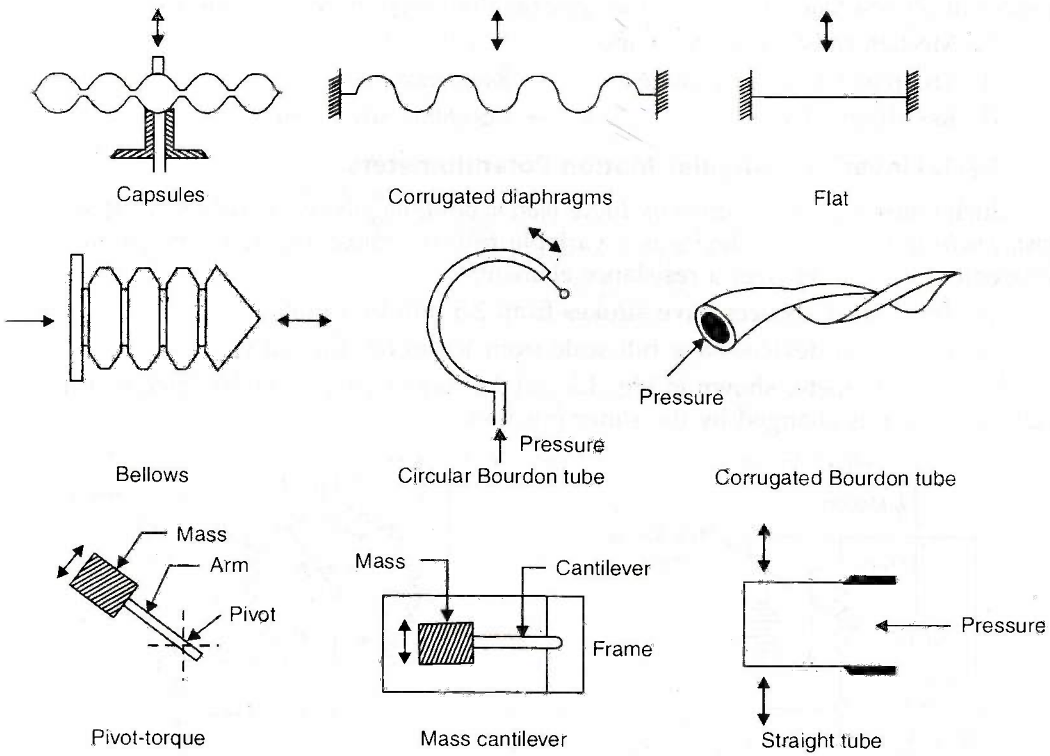


Fig. 3.2. Transducer actuating mechanisms.

3.7 RESISTANCE TRANSDUCERS

The resistance of a metal conductor is expressed by a simple equation that involves a few physical quantities. The relationship is given by $R = \frac{\rho L}{A}$;

where, R = Resistance, Ω,
ρ = Resistivity of conductor materials, Ω-m,
L = Length of conductor, m, and
A = Cross-sectional area of the conductor, m^2.

Any method of varying one of the quantities involved in the above relationship can be the designed basis of an *electrical resistance transducer.* There are a number of ways in which resistance can be changed by a physical phenomenon.

The *translational and rotational "potentiometers"* which work on the basis of change in the value of resistance with change in length of the conductor can be used for measurement of translational or rotary displacements.

*"**Strain gauges**" work on the principle that the resistance of a conductor or a semiconductor changes when strained.* This property can be used for measurement of displacement, force and pressure.

The resistivity of materials changes with the change of temperature thus causing a change of resistance. This property may be used for measurement of *"temperature"*.

In a *resistance transducer* an indication of measured physical quantity is given-by *a change in the resistance.* It may be *classified* (as discussed above) as follows :

1. Mechanically varied resistance — *Potentiometer*
2. Thermal resistance change — *Resistance thermometers*
3. Resistivity change — *Resistance strain gauge.*

3.7.1. Linear and Angular Motion Potentiometers

Such potentiometers *convert the linear motion or the angular motion of a rotating shaft into changes in resistance.* The device is a variable resistor whose resistance is varied by the movement of a slider over a resistance element.

— Translatory devices have strokes from 2.5 mm to 5 mm.

— Rotational devices have full scale from 10° to 60° full turn.

The potentiometer shown in Fig. 3.3 and 3.4 form a part of the bridge circuit whose output voltage is changed by the slider position.

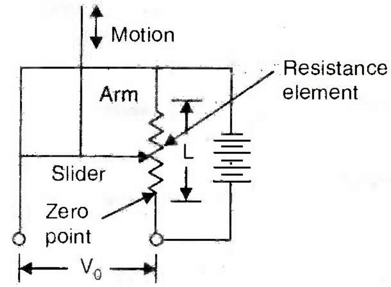


Fig. 3.3. Linear motion potentiometer.

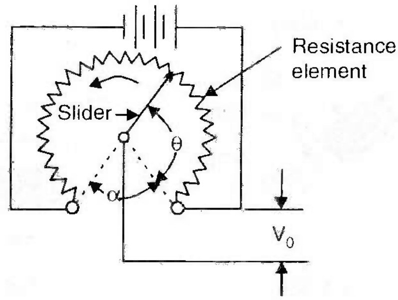


Fig. 3.4. Rotary motion potentiometer.

— The slider is powered by the mechanical part on which the linear displacement or angular measurement are to be made.

— Due to arm movement, the slider moves over the resistance element and thus shorts out a portion of the resistance. *The change in resistance in the potentiometer is then an indication of the amount of motion and the direction of movement is indicated by whether the resistance is increasing or decreasing.* The unbalanced voltage is measured directly or fed into an amplifier and recorded.

The potentiometers are used in many transducers designed to measure :

(*i*) Pressure (*ii*) Force
(*iii*) Acceleration (*iv*) Liquid level.

The potentiometers have the folowign *advantages* and *disadvantages* :

Advantages :

(*i*) High output.
(*ii*) Less expensive.
(*iii*) Available in different sizes, shapes and ranges.
(*iv*) Simple to operate.
(*v*) Their electrical efficiency is very high.
(*vi*) Rugged construction.
(*vii*) Insensitivity towards vibration and temperature.

Disadvantages :

(*i*) Limited life due to early wear of the sliding ram.
(*ii*) The output tends to noisy and erratic in high speed operation or when in high vibration environment.
(*iii*) In wire wound potentiometers the resolution is limited while in cermet and metal film potentiometers, the resolution is infinite.

Power rating of potentiometers :

The potentiometers are designed with a definite power rating which is related directly to their heat dissipating capacity. The manufacturer normally designs a series of potentiometers of single turn with a diameter of 50 mm with a wide range of ohmic values ranging from *100 Ω to 10 kΩ* in *steps of 100 Ω*. These potentiometers are essentially of the *same size* and of the *same mechanical configuration*. They have the *same heat transfer capabilities*. Their rating is typically *5 W at an ambient temperature of 21°C*. This limits their input excitation voltage.

Materials used for potentiometers :

The materials used for potentiometers may be classified as *wire wound* and *non-wire wound* as follows:

1. Wire wound potentiometers :

The materials used are:

- Platinum;
- Nickel chromium;
- Nickel copper;
- Other precious resistance elements.
  - — These potentiometers carry relatively large currents at high temperatures.
  - — Their temperature coefficients of resistance is usually small, of the order of $20 \times 10^{-6}/°C$ or less.
  - — Their resolution is about 0.025 – 0.05 mm and is limited by the number of turns that can be accommodated on the body.
  - — The response is limited to about 5 Hz.
  - — The maximum speed with which a wire wound potentiometer may be turned is about 300 r.p.m.

2. Non-wire wound potentiometers :

These are also called *continuous potentiometers*.

The *materials* used are :

- Cermet;
- Hot moulded carbon;

- Carbon film;
- This metal film.
  - These potentiometers provide *improved resolution and life* (since resolution is no longer limited by the number of turns that can be wrapped onto a body)
  - A continuous potentiometer may be turned at a speed of 2000 r.p.m.
  - These are *more sensitive to temperature changes* and have a *higher wiper contact resistance,* which is variable and can *carry only moderate currents.*

Example 3.1. *Explain briefly the types of errors encountered in a transducer.*

Solution. The types of errors encountered in a transducer are briefly discussed below:

1. Scale errors.
2. Dynamic errors.
3. Noise and drift errors.

1. *Scale errors. Calibration* in general may be defined as the *process for determination, by measurement or comparison with a standard, of the correct value of each scale reading on the measuring instrument.* It is the determination of the settings of a control device that correspond to particular values of voltage, current, frequency, pressure or some other output.
2. *Dynamic errors. Fidelity* of an instrument system is defined as the *degree of closeness with which the system records the signal which is impressed upon it.* It refers to the ability of the system to reproduce the output in the same form as the input. "Dynamic error" is the *difference between the true value of a quantity changing with time and the value indicated by the instrument if no static error is assumed.*
3. *Noise and drift errors. Drift* is undesired change or a gradual variation in output over a period of time that is unrelated to changes in output, operating conditions or load. Drift for a measuring device can either be systematic, random, or some combination of the two. For most devices, drift is measured and specified as a percentage of output span.

Example 3.2. *A linear resistance potentiometer is 50 mm long and is uniformly wound with wire having a resistance of 10000 Ω. Under normal conditions, the slider is at the centre of the potentiometer. Find the linear displacement when the resistance of the potentiometer as measured by a Wheatstone bridge for two cases is (i) 3850 Ω (ii) 7560 Ω.*

Are the two displacements in the same direction? If it is possible to measure a minimum value of 10 Ω resistance with the above arrangements, find the resolution of the potentiometer in mm.

(Anna University)

Solution. Under normal condition, it is given that the slider is at the centre of the potentiometer, hence under normal position the resistance of the potentiometer

$$= \frac{10000}{2} = 5000\ \Omega$$

The resistance of the potentiometer wire per unit length

$$= \frac{10000}{50} = 200\ \Omega/\text{mm}$$

(*i*) Change in resistance of potentiometer from its normal position

$$= 5000 - 3850 = 1150\ \Omega$$

$$\therefore \quad \textbf{Displacement} = \frac{1150}{200} = \textbf{5.75 mm (Ans.)}$$

(*ii*) Change in resistance of potentiometer from its normal position

$$= 7560 - 5000 = 2650\ \Omega \text{ (in the } \textit{opposite direction}\text{)}$$

$$\therefore \quad \textbf{Displacement} = \frac{2560}{200} = \textbf{12.8 mm. (Ans.)}$$

The two displacements are in the *opposite direction.*

Resolution of the potentiometer

$$= \frac{\text{Min. measurable resistance}}{\Omega/\text{mm}} = \frac{10}{200} = \textbf{0.05 mm. (Ans.)}$$

3.7.2. Thermistors and Resistance Thermometers

These transducers are thermally sensitive variable resistors made of certain conducting and ceramic-like semiconducting materials. They are used *as temperature detecting elements used to sense temperature for the purpose of measurements and control.*

Thermistors are essentially semiconductors which behave as resistors with a *high negative temperature coefficient of resistance.* The high sensitivity to temperature changes make the thermistors extremely useful for precision temperature (–60°C to + 15°C) measurements, control and compensation. Their resistance ranges from 0.5 Ω to 0.75 MΩ.

Thermistors are composed of sintered mixture of metallic oxides such as manganese, nickel, cobalt, copper, iron and uranium.

Fig. 3.5. shows the commercial forms of thermistors.

Applications of thermistors :

1. *Measurements of temperature* .... (major application).
2. Temperature compensation in complex electronic equipment, magnetic amplifiers and instrumentation equipment.
3. Measurement of power at high frequencies.
4. Vacuum measurements.
5. Measurements of level, flow and pressure of liquid.
6. Measurement of thermal conductivity.

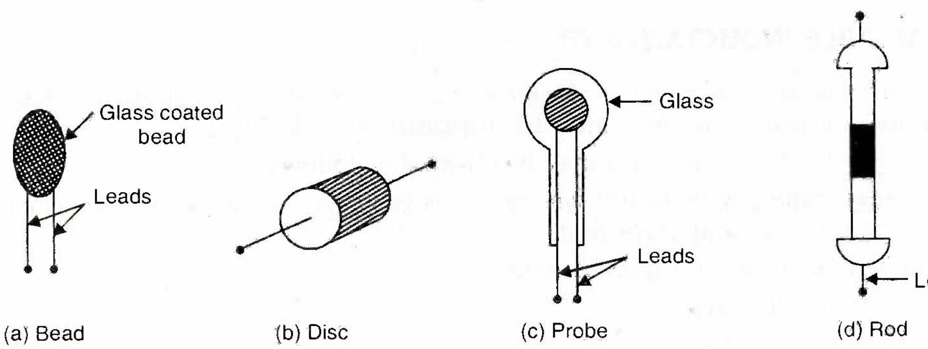


Fig. 3.5. Commercial forms of thermistors.

3.7.3. Wire Resistance Strain Gauges

Refer to article 3.14.

Salient features of thermistors :

1. The thermistors are compact, rugged and inexpensive.

2. They have good stability, when properly aged.
3. Their response time can vary from a fraction of a second to minutes depending on the size of the detecting mass and thermal capacity of the thermistors. It varies *inversely* with dissipation factor.
4. The upper limit of temperature for thermistors is dependent on physical changes in the material or solder used in attaching the electrical connections and is usually 400°C or less.
5. These can be installed at a distance from their associated measuring circuits if elements of high resistances are used such that the resistance of leads is negligible.
6. The measuring current should be maintained to as low a value as possible so that self-heating of thermistors is avoided otherwise errors are introduced on account of change of resistance caused by self-heating.

Example 3.3. *(a) As thermistor has a resistance temperature coefficient of –5% over a temperature range of 25°C to 50°C. If the resistance of the thermistor is 100 W at 25° C, what is the resistance at 35°C?*

(b) Suggest a complete instrumentation scheme in block diagram form to measure the temperature in a closed oven with the help of thermistor.

Solution. (*a*) $R_{35} = R_{25}[1 + \alpha(35 - 25) = 100[1 - 0.05(35 - 25)] = 50\ \Omega$ **(Ans.)**

(*b*) Fig. 3.6. shows the complete instrumentation scheme for the measurement of temperature with the help of a thermistor. Thermistor is mounted in the oven at a place where temperature is to be sensed. With the increase in temperature, resistance of the thermistor decreases causing imbalance in Wheatstone bridge circuit whose output balance voltage is amplified by signal conditioning device, the amplified output when connected to a suitable output device gives the value of the temperature of the even.

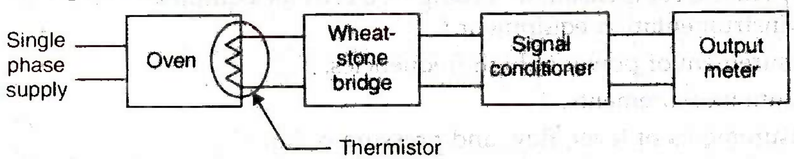


Fig. 3.6.

3.8 VARIABLE INDUCTANCE TRANSDUCERS

These are based on a change in the magnetic characteristic of an electrical circuit in response to a measure and which may be displacement, velocity, acceleration etc.

Variable inductive transducers may be *classified* as follows :

1. Self-generating type. *In this type voltage is generated because of the relative motion between a conductor and a magnetic field.*

These may be further *classified* as follows:

(*i*) Electromagnetic type.

(*ii*) Electrodynamic type.

(*iii*) Eddy current type.

2. Passive type. In this type the *motion of an object results in changes in the inductance of the coils of the transducer.*

These may be further *classified* as follows :

(*i*) Variable reluctance.

(*ii*) Mutual inductance.

(*iii*) Differential transfer type.

3.8.1. Self-generating Type

3.8.1.1. Electromagnetic type

Fig. 3.7 shows an electromagnetic type of self-generating variable inductance transducer.

— It consists of a permanent magnet core on which a coil is directly wound.

— When a plate of iron or other ferromagnetic material is moved with respect to the magnet, the flux field expands or collapses and a voltage is induced in the coil.

— This device is used for *indication of angular speed.* The measurements of speed can be made with great accuracy when the pick-up is placed near the teeth of a rotating gear.

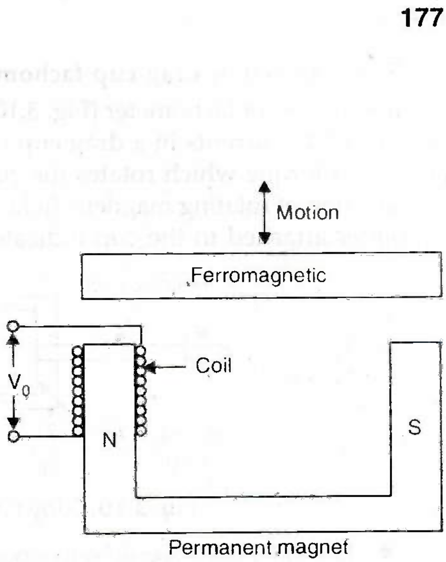


Fig. 3.7. Self-generating variable inductance transducer—Electromagnetic type.

3.8.1.2. Electrodynamic type

This type of transducer (linear and rotational is shown in Fig. 3.8).

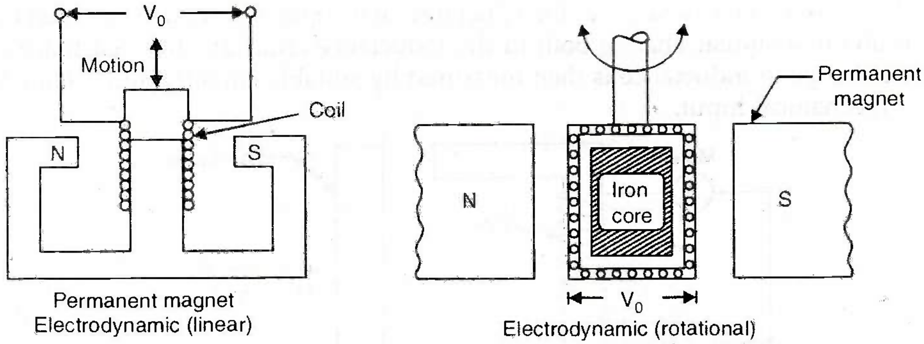


Fig. 3.8. Self-generating variable inductance transducer—Electrodynamic type.

— In this type, coil moves within the field of the magnet. The turns of the coil are perpendicular to the intersecting lines of force.

— When the coil moves it induces a voltage which at any moment is *proportional to the velocity of the coil.*

- The principle of these transducer is used in the *magnetic flow meters.*

3.8.1.3. Eddy current type

Fig. 3.9 shows an eddy current type of self-generating variable inductance transducer.

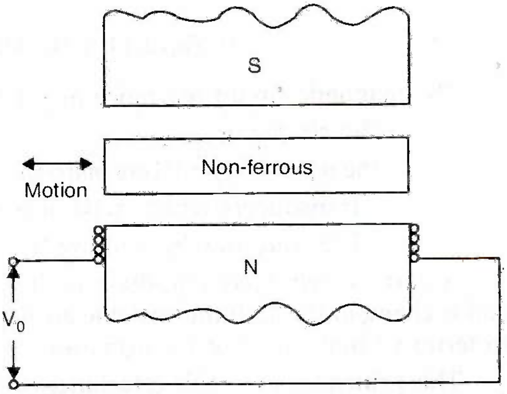


Fig. 3.9. Self-generating variable inductance transducer—Eddy current type.

Eddy current or drag cup tachometer :

In this type of tachometer (Fig. 3.10) the test-shaft rotates a permanent magnet and this induces eddy currents in a drag cup of disc held close to the magnet. The eddy currents produce a torque which rotates the cup against the torque of a spring. The disc turns in the direction of rotating magnetic field until the torque developed equals that of the spring. A pointer attached to the cup indicates the rotational speed on a calibrated scale.

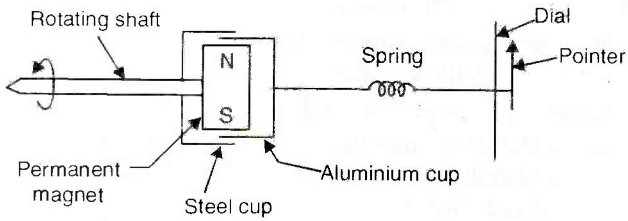


Fig. 3.10. Eddy current or drag type tachometer.

- The *automobile speedometers* operate on this principle.
- These tachometers are used for measuring rotational speeds upto 12000 r.p.m. with an accuracy of ± 3 per cent.

3.8.2. Passive Type

3.8.2.1. Variable reluctance transducer

In these transducers (comprising of a magnetic field and core with a gap between the core and the fixed coils) a change in the reluctance of the magnetic circuit by a mechanical input results in a similar change both in the inductance and inductive reactance of the coils. The change in inductance is then measured by suitable circuitry and related to the value of mechanical input.

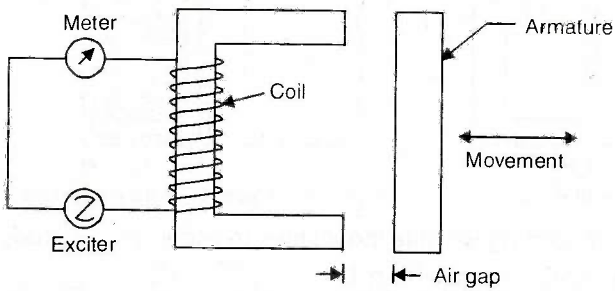


Fig. 3.11. Variable reluctance transducer.

The magnetic circuit reactance may be changed by affecting a change :

(*i*) in the *air gap* or

(*ii*) in the *amount/type of core material.*

— Transducers which make use of *air gap change* are referred as *reluctance type.*

— Transducers which utilize a *variable core* are referred as *permeance type.*

A variable *reluctance transducer* is shown in Fig. 3.12. Here the inductance of a single coil is changed through the variable air gap. The change in inductance may be calibrated in terms of movement of the armature.

This principle of variable reluctance is used for the measurement of dynamic quantities such as :

(i) Pressure (ii) Force
(iii) Displacement (iv) Acceleration
(v) Angular position etc.

Example 3.4. *Fig. 3.12 shows a variable reluctance type proximity inductive transducer in which the coil has inductance of 2 mH when the target made of ferromagnetic material is 1 mm away.*

(i) *Calculate the value of inductance when a displacement of 0.02 mm is applied to the target in a direction moving it towards the core.*

(ii) *Show that the change in inductance is linearly proportional to the displacement. Neglect the reluctance of the iron parts.*

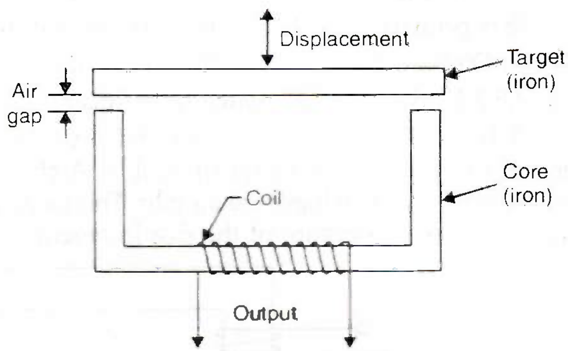


Fig. 3.12. Variable reluctance inductive transducer.

Solution. Inductance with air gap length of 1.00 mm, $L = 2$ mH

(i) **Value of inductance when a displacement of 0.02 mm is applied :**

Length of air gap when a displacement of 0.02 mm is applied towards the core

$$= 1.00 - 0.02 = 0.98 \text{ mm}$$

Now, the inductance is inversely proportional to the length of air gap as the reluctance of flux paths through iron are neglected. *Since the gap length decreases the inductance increases by ΔL.*

$$\therefore \quad L + \Delta L = 2 \times \frac{1}{0.98} = 2.04 \text{ mH (Ans.)}$$

or, $$\Delta L = 2.04 - 2 = 0.04 \text{ mH}$$

(ii) **ΔL α displacement :**

The ratio of change in inductance to the original inductance

$$= \frac{\Delta L}{L} = \frac{0.04}{2} = 0.02$$

Also, the ratio of displacement to original gap length

$$= \frac{0.02}{1} = 0.02$$

Hence the ΔL α **displacement** .... *Proved.*

- This relationship, however, is *true of only very small values of displacement.*

Variable permeance transducer :

Fig. 3.13 show a *variable permeance transducer* in which the inductance of coil is changed by *varying the core material.*

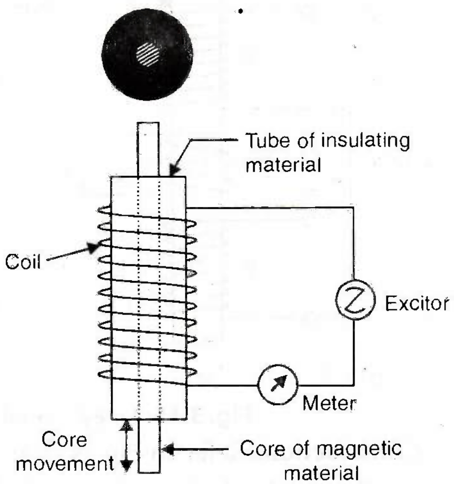


Fig. 3.13. Variable permeance transducer (self inductance arrangements).

— The transducer consists of a coil of many turns of wire wound on a tube or insulating material with a moveable core of magnetic material.

— When the coil is energized and the core enters the solenoid cell, the *inductance of the coil increases in proportion to the amount of metal within the coil.*

It is primarily used for the measurement of :

(*i*) Displacement; (*ii*) Strain; (*iii*) Force.

3.8.2.2. Mutual inductance transducer

A two-coil mutual inductance transducer is illustrated in Fig. 3.14. It consists of an energising coil *X* and a pick-up coil *Y*. A change in the position of the armature by a mechanical input changes the air gap. This cause a change in the output from coil *Y*, which may be used as measure of the displacement of the armature *i.e.*, the mechanical input.

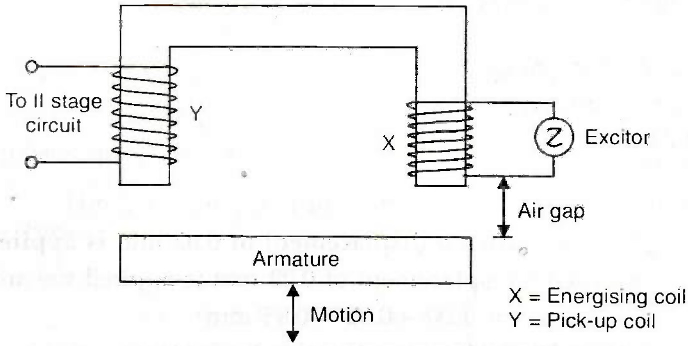


Fig. 3.14. Mutual inductance transducer.

3.8.2.3. Linear-variable-differential transformer (LVDT)

LVDT *is a passive inductive transducer and is commonly employed to measure force (or weight, pressure and acceleration etc. which depend on force) in terms of the amount and direction of displacement of an object.*

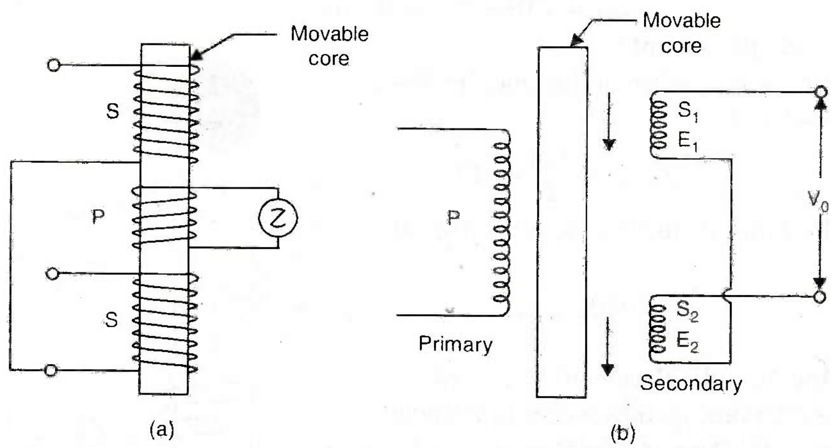


Fig. 3.15. Linear-variable-differential transducer (LVDT).

Construction. Refer to Fig. 3.15(*a*).

— It consists of one primary winding (*P*) and two secondary windings (S_1 and S_2) which are placed on either side of the primary mounted on the same magnetic core. The magnetic core is free to move axially inside the coil assembly and the motion being measured is mechanically coupled to it.

— The two secondaries S_1 and S_2 have equal number of turns but are connected in series opposition so that e.m.fs, (E_1 and E_2) induced in them are 180° out of phase with each other and hence, cancel each other out. [See Fig. 3.15 (*b*)]

— The primary is energised from a suitable A.C. source.

Working :

— When the core is in the centre (*called reference position*) the induced voltages E_1 and E_2 are *equal and opposite*. Hence they cancel out and the output voltages V_0 is zero.

— When the external applied force moves the core towards coil S_2, E_2 is increased but E_1 is *decreased* in magnitude though they are still antiphase with each other. The net voltage available is $(E_2 - E_1)$ and is *in phase with* E_2.

Similarly, when the magnitude core moves towards coil S_1, $E_1 > E_2$ and $V_0 = E_1 - E_2$ and is in *phase with* E_1.

Thus, from above discussion, we find that the magnitude of V_0 is a *function of the distance moved by the core* and its *polarity* or *phase* indicates as to in which direction it has moved.

If core is attached to moving object, the *magnitude* of V_0 *gives the position of that object.*

Fig. 3.15(*c*) shows the pressure measurement by LVDT.

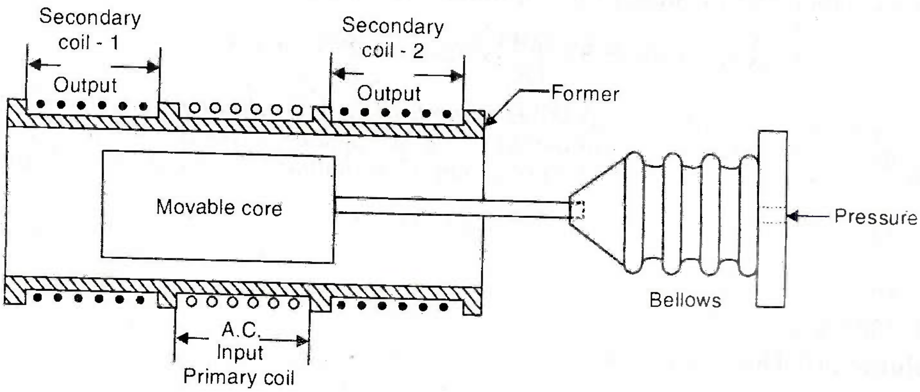


Fig. 3.15 (*c*). Pressure measurement by LVDT.

Advantages :

1. It gives a high output and therefore many a times there is no need for intermediate amplification devices.
2. The transducer possesses a high sensitivity as high as 40 V/mm.
3. It shows a low hysteresis and hence repeatability is excellent under all conditions.
4. Most of the LVDTs consume a power of less than 1 W.
5. Less friction and less noise (due to absence of sliding contacts).
6. These transducers can usually tolerate a high degree of shock and vibration without any adverse effects.
7. It can operate over a temperature range from –265°C to 600°C.
8. It is available in radiation-resistant design for operation in nuclear reactors.

Disadvantages :

1. These transducers are sensitive to stray magnetic fields but shielding is possible. This is done by providing magnetic shields with longitudinal slots.

2. Relatively large displacements are required for appreciable differential output.
3. The receiving instrument must be selected to operate on A.C. signals or demodulator network must be used if a D.C. output is required.
4. Several times, the transducer performance is affected by vibrations.
5. The dynamic response is limited mechanically by the mass of core and electrically by the frequency of applied voltage. The frequency of the carrier should be at least ten times the highest frequency component to be measured.

Applications:

1. Measurement of material thickness in hot strip or slab steel mills.
2. In accelerometers.
3. Jet engine controls in close proximity to exhaust gases.

Note. LVDT *is not suited for fast dynamic measurements on account of mass of the core.*

Example 3.5. *In a linear voltage differential transformer (LVDT) the output voltage is 1.8 V at maximum displacement. At a certain load the deviation from linearity is maximum and it is ± 0.0045 V from a straight line through the origin. Find the linearity at the given load.*

Solution. *Given :* The output voltage of LVDT at maximum displacement = 1.8 V

The deviation from a straight line through the origin = ± 0.0045 V

$$\therefore \quad \text{\%age linearity} = \pm\frac{0.0045}{1.8}\times 100 = \mathbf{\pm 0.25\%\ (Ans.)}$$

Example 3.6. *The output of a LVDT is connected to a 4 V voltmeter through an amplifier whose amplification factor is 500. An output of 1.8 mV appears across the terminals of LVDT when the core moves through a distance of 0.6 mm. If the millivoltmeter scale has 100 divisions and the scale can be read to $\frac{1}{4}$ of a division, calculate:*

(i) The sensitivity of LVDT.

(ii) The resolution of the instrument in mm.

Solution. (*i*) **The sensitivity of LVDT :**

$$\text{The sensitivity of LVDT} = \frac{\text{Output voltage}}{\text{Displacement}} = \frac{1.8}{0.6} = \mathbf{3\ mV/mm\ (Ans.)}$$

(*ii*) **The resolution of the instrument :**

Sensitivity of measurement = Amplification factor × sensitivity of LVDT

= 500 × 3 = 1500 mV/mm

$$1 \text{ scale division} = \frac{4}{100}\ \text{V} = 40\ \text{mV}$$

Minimum voltage that can be read on the voltmeter

$$= \frac{1}{4}\times 40 = 10\ \text{mV}$$

∴ Resolution of the instrument

$$= 10\times\left(\frac{1}{1500}\right) = \mathbf{0.0067\ mm\ (Ans.)}$$

3.9 CAPACITIVE TRANSDUCERS

The principle of operation of capacitive transducers is based upon the familiar equation for capacitance of a parallel plate capacitor :

$$\text{Capacitance, } C = \frac{\epsilon A}{d} = \frac{\epsilon_r \epsilon_0 A}{d} \qquad ...(3.1)$$

where,
$\epsilon = \epsilon_r \epsilon_0$ = Permittivity of medium, F/m,
ϵ_r = Relative permittivity, (for air $\epsilon_r = 1$),
ϵ_0 = Permittivity of free space = 8.85×10^{-12} F/m,
A = Overlapping area of plates, and
d = Distance between the two plates.

Any physical quantity which can cause a change in ϵ, A or d can be measured by the *capacitance gauge.*

The *displacement is measured by measuring the change in capacitance brought about by :*

(*i*) Change in area, or

(*ii*) Change in distance between the plates.

The change in capacitance on account of change in dielectric is used to measure change in liquid and gas levels.

3.9.1. Capacitive Transducers—Using Change in Area of Plates

Figure 3.16(*a*), (*b*) shows the elementary diagrams of the arrangements of a capacitive transducer where capacitance change occurs because of change in the area of plates.

Since capacitance is directly proportional to the effective area of the plates, response of such a system is linear.

Fig. 3.16(*c*) shows variation of the capacitance.

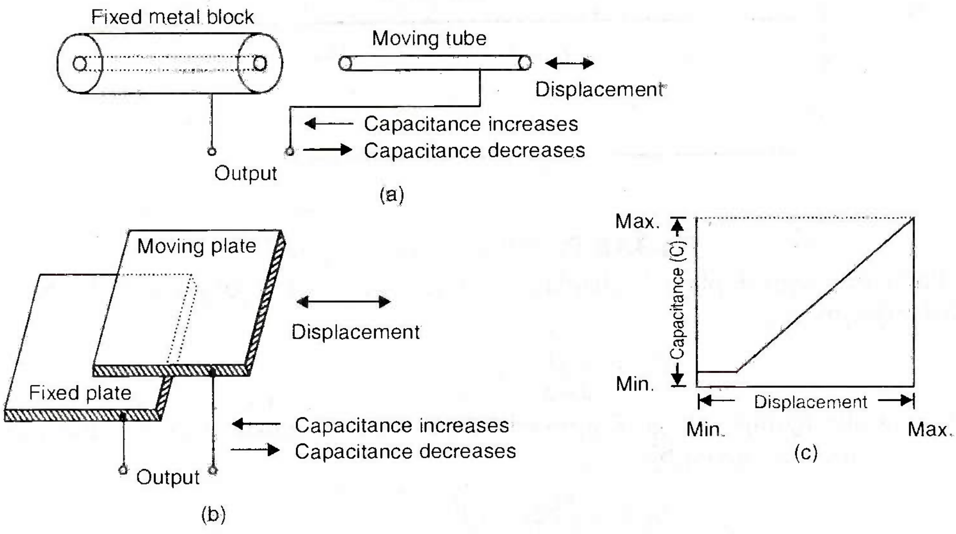


Fig. 3.16. Capacitive transducers working on the principle of change of capacitance with change of area.

3.9.2. Capacitive Transducer–Using Change in Distance Between the Plates

Fig. 3.17 shows the basic form of a capacitive transducer utilizing the effect of *change of capacitance with change in distance between the plates.*

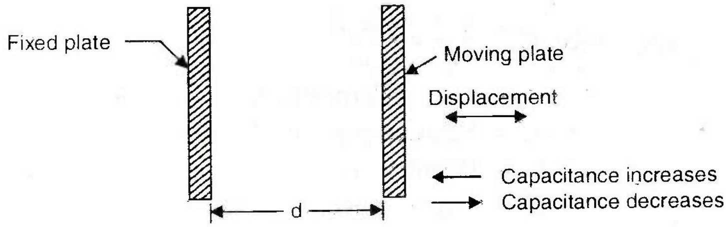


Fig. 3.17. Capacitive transducer.

One is a *fixed plate* and the displacement to be measured is applied to the other plate which is *movable.* Since, the capacitance, C varies inversely as the distance between the plates the response of this transducer is *not linear.*

Differential capacitor system:

In a differential capacitor system, let the normal position of the central plate be represented by a solid lines as shown in Fig. 3.18. The capacitances C_1 and C_2 are then identical.

i.e.,
$$C_1 = C_2 = C = \frac{\in A}{d} \quad ...(3.2)$$

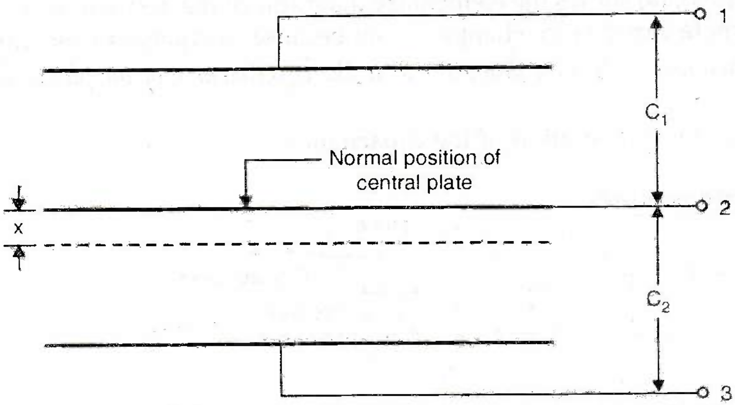


Fig. 3.18. Differential capacitor system.

When the central plate is displaced parallel to itself through a distance x, the capacitances are

$$C_1 = \frac{\in A}{d+x}, C_2 = \frac{\in A}{d-x} \quad ...(3.3)$$

For an alternating voltage E applied between the terminals 1 and 2, the voltages across C_1 and C_2 are given by

$$E_1 = \frac{EC_2}{C_1+C_2} = E\frac{d+x}{2d}$$

and,
$$E_2 = \frac{EC_1}{C_1+C_2} = E\frac{d-x}{2d} \quad ...(3.4)$$

When the differential measurement circuit is fed with output from the terminals pairs 1 and 3, and 2 and 3, the difference voltage would be recorded.

$$E_1 - E_2 = E\frac{x}{d} \quad \text{...(3.5)}$$

The difference voltage is a *linear function of the displacement of the linear plate.*

The differential method can be used for displacement of 10^{-8} mm to 10 mm with an accuracy of 0.1%.

Advantages and disadvantages of capacitive transducers :

Advantages. The *major advantages* of capacitive transducers are:

1. Require extremely small force for operation (hence very useful for use in small systems).
2. Extremely sensitive.
3. Require small power for operation.
4. High input impedance; therefore, loading effects are minimum.
5. Frequency response is good.
6. A resolution of the order of 2.5×10^{-3} mm can be obtained.
7. Can be used for applications where stray magnetic fields render the inductive transducers useless.

Disadvantages. The *principal disadvantages* of capacitive transducers are:

1. The metallic parts must be insulated from each other. The frames must be earthed to reduce the effects of stray capacitances.
2. They show non-linear behaviour several times on account of *edge effects ; guard rings* must be used to eliminate this effect.
3. The cable connecting the transducer to the measuring point is also a source of error. The cable may be source of loading resulting in loss of sensitivity. Also loading makes the low frequency response poor.

Uses of the capacitive transducers. The capacitive transducers are used for the following *purposes* :

1. To measure *both linear and angular displacements.*
2. To *measure force and pressure.*
3. Used as pressure transducers in all those cases where the dielectric constant of a medium changes with pressures.
4. To measure humidity in gases.
5. Used in conjunction with mechanical modifiers for measurement of *volume, density, weight, input level* etc.

Example 3.7. *A parallel plate capacitive transducer uses plates of area 300 mm^2 which are separated by a distance 0.2 mm.*

(i) Determine the value of capacitance when the dielectric is air having a permittivity of 8.85×10^{-12} F/m.

(ii) Determine the change in capacitance if a linear displacement reduces the distance between the plates to 0.18 mm. Also determine the ratio of per unit change of capacitance to per unit change of displacement.

(iii) If a mica sheet 0.01 mm thick is inserted in the gap, calculate the value of original capacitance and change in capacitance for the same displacement. Also calculate the ratio

of per unit change in capacitance to per unit change in displacement. The dielectric constant of mica is 8.

Solution. *Given* : $A = 300\ mm^2 = 300 \times 10^{-6}\ m^2$; $d = 0.2$ mm; $\epsilon_0 = 8.854 \times 10^{-12}$ F/m; ϵ_r (mica) = 8

(*i*) Value of capacitance:

$$\text{Value of capacitance, } C = \frac{\epsilon_0 \epsilon_r A}{d} = \frac{\epsilon_0 A}{d} \qquad (\because\ \epsilon_r = 1)$$

$$= \frac{8.85 \times 10^{-12} \times 300 \times 10^{-6}}{0.2 \times 10^{-3}} \text{F} = \mathbf{13.275\ pF\ (Ans.)}$$

(*ii*) Change in capacitance, ΔC:

Change is displacement Δd= 0.2 – 0.18 = 0.02 mm.

Capacitance after application of displacement,

$$C + \Delta C = \frac{8.85 \times 10^{-12} \times 300 \times 10^{-6}}{0.18 \times 10^{-3}} \text{F} = 14.75 \text{ pF}$$

Change in capacitance, ΔC = 14.75 – 13.275 = **1.475 pF (Ans.)**

Ratio of per unit change of capacitance to per unit change of displacement,

$$\frac{\Delta C / C}{\Delta d / d} = \frac{(1.475/13.275)}{(0.02/0.2)} = \mathbf{1.111\ (Ans.)}$$

(*iii*) *C*, *ΔC*, Ratio (*ΔC/C*)/(*Δd/d*) when mica sheet is inserted:

Initially, the displacement between the plates is 0.2 mm. Since the thickness of mica is 0.01 mm, the length of air between the plates = 0.2 – 0.01 = 0.19 mm.

Initial capacitance of transducer,

$$C = \frac{\epsilon_0 A}{\dfrac{d_1}{\epsilon_{r_1}} + \dfrac{d_2}{\epsilon_{r_2}}}$$

or,

$$C = \frac{8.85 \times 10^{-12} \times 300 \times 10^{-6}}{\left(\dfrac{0.19}{1} + \dfrac{0.01}{8}\right) \times 10^{-3}} \text{F} = \mathbf{13.88\ pF\ (Ans.)}$$

When a displacement of 0.02 mm is applied, the length of air gap is reduced to 0.19 – 0.02 = 0.17 mm.

∴ Capacitance with displacement applied

$$= \frac{8.85 \times 10^{-12} \times 300 \times 10^{-6}}{\left(\dfrac{0.17}{1} + \dfrac{0.01}{8}\right) \times 10^{-3}} \text{F} = 15.5 \text{ pF}$$

Change in capacitance, ΔC = 15.5 – 13.88 = **1.62 pF (Ans.)**

$$\textit{Ratio}\ \frac{\Delta C / C}{\Delta d / d} = \frac{(1.62/13.88)}{(0.02/0.2)} = \mathbf{1.167\ (Ans.)}$$

3.9.3. Capacitive Tachometer

Refer to Fig. 3.19. A capacitive pick-up tachometer consists of a vane attached to one end of the rotating machine shaft. When the shaft rotates between the field capacitive plates, there occurs a change in the capacitance. The capacitor forms a part of an oscillator tank so that *number of frequency changes per unit of time is a measure of the shaft speed.* The pulses thus produced are *amplified* and *squared,* and may then be *fed to frequency measuring unit* or to a digital counter so as to provide a digital analog of the shaft rotation.

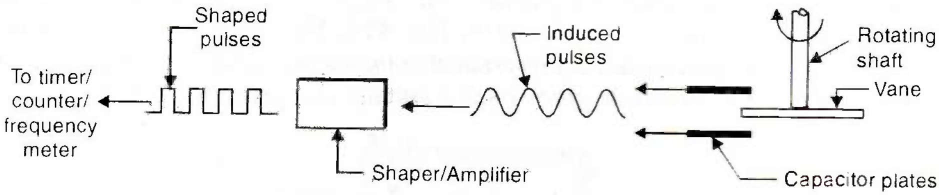


Fig. 3.19. Capacitive pick-up tachometer.

3.10 PIEZOELECTRIC TRANSDUCERS

3.10.1. Piezoelectric Materials

A ***"piezoelectric material"*** *is one in which an electric potential appears across certain surfaces of a crystal if the dimensions of the crystals are changed by the application of a mechanical force.* This potential is produced by the displacement of external charges. The effect is reversible, *i.e.,* conversely, if a varying potential is applied to the proper axis of the crystal, it will change the dimensions of the crystal thereby deforming it. This effect is known as **piezoelectric effect.**

Elements exhibiting piezoelectric qualities are sometimes known as *electro-resistive elements.* Common piezoelectric materials are : *Ammonium dihydrogen phosphate, Rôchelle salts, lithium sulphate, dipotassium tartrate, potassium dihydrogen phosphate, quartz and ceramics A and B.*

There are two main groups of piezoelectric crystals:

1. ***Natural crystals*** ..... such as *quartz and tourmaline.*
2. ***Synthetic crystals*** ..... such as *Rochelle salt, lithium sulphate, dipotassium tartrate etc.*

3.10.2. Desirable Properties of Piezoelectric Materials

The *desirable properties* of piezoelectric materials are :

(*i*) Stability.
(*ii*) High output.
(*iii*) Insensitivity to temperature and humidity.
(*iv*) The ability to be formed into most desirable shape.

Natural crystals entail the following *advantages* ;

(*i*) Higher mechanical and thermal stability.
(*ii*) Ability to withstand higher stresses.
(*iii*) Low leakage.
(*iv*) Good frequency response.

Synthetic materials, in general, have a *higher voltage sensitivity.*

- *"Quartz"* is the most stable piezoelectric material. However, its output is quite small.

- *"Rochelle"* salt provides the highest output but it can be worked over a limited humidity range and has to be protected against moisture. The highest temperature is limited to 45°C.
- *"Barium titanate"* has the advantage that it can be formed into a variety of shapes and sizes since it is polycrystalline. It has also a higher dielectric constant.

3.10.3. Working of a Piezoelectric Device

A typical mode of operation of a piezoelectric device employed for measuring varying force applied to a simple plate is shown in Fig. 3.20. *The magnitude and polarity of the induced charge on the crystal surface is proportional to the magnitude and direction of the applied force.* The charge at the electrode gives rise to voltage (E), given by,

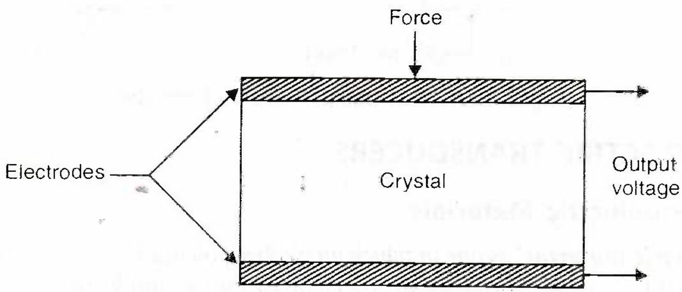


Fig. 3.20. Piezoelectric transducer.

$$E = \frac{gtF}{A} = gtp$$

where, g = Voltage sensitivity in Vm/N, $\left[g = \frac{K}{t},\ k = \text{Piezoelectric constant},\ t = \text{Thickness of the crystal.} \right]$

F = Force in N (newton),

A = Area of the crystal in m^2, and

p = Pressure $\left(= \frac{F}{A}\right)$ in N/m^2.

3.10.4. Advantages and Disadvantages of Piezoelectric Transducers

Advantages:

1. High frequency response.
2. Small size.
3. High output.
4. Rugged construction.
5. Negliible phase shift.

Disadvantages:

1. Output affected by changes in temperature.
2. Cannot measure static conditions.

Applications:

These transducers find the following *fields of application:*

1. Acceleroeters.

2. Pressure cells.
3. Force cells.
4. Ceramic microphones.
5. Phonograph pick-up.
6. Cartridges.
7. Industrial cleansing apparatus.
8. Under-water detection system.

3.10.5. Piezoelectric Accelerometer

Refer to Fig. 3.21. A piezoelectric accelerator is probably the *simplest* and *most commonly used transducer for measuring acceleration*.

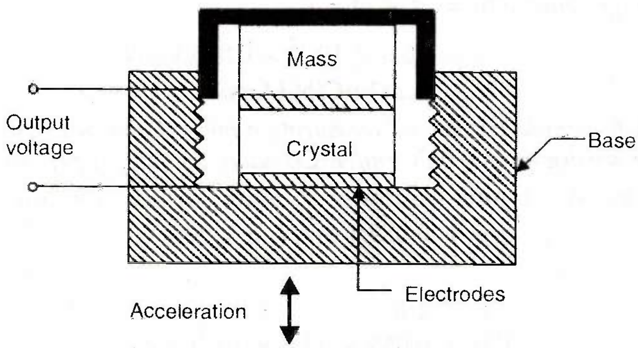


Fig. 3.21. Piezoelectric accelerometer.

Construction. It consists of a *piezoelectric crystal* sandwitched between two *electrodes* and has *mass* placed on it. The unit is fastened to the base whose acceleration characteristics are to be obtained. The *can* threaded to the base acts as a spring and squeezes the mass against the crystal. *Mass exerts a force on the crystal and a certain voltage output is generated.*

Working. When the base is accelerated *downward* inertial reaction force on the base acts upward against the top of the can. This relieves stress on the crystal. According to Newton's second law of motion, *force = mass × acceleration*, since the mass is a fixed quantity, the *decrease in force is proportional to the acceleration*. Similarly, an acceleration in the upward direction would increase the force on the crystal in proportion to the acceleration. The resulting change in the output voltage is recorded and correlated to the acceleration imposed on the base.

Advantages:

1. Small size and a small weight.
2. High output impedance.
3. Can measure acceleration from a fraction of g to thousands of g.
4. High sensitivity.
5. High frequency response (10 Hz to 50 kHz).

Disadvantages:

1. Unsuitable for applications where the input frequency is lower than 10 Hz.
2. Subject to hysteresis errors.
3. Sensitive to temperature changes.

Example 3.8. *A 2.5 mm thick quartz piezoelectric crystal having a voltage intensity of 0.055 Vm/N is subjected to a pressure of 1.4 MN/m^2. If the permittivity of quartz is 40.6×10^{-12} F/m, calculate:*

(i) Voltage output.

(ii) Charge sensitivity of the crystal.

Solution. *Given:* $t = 2.5$ mm or 2.5×10^{-3} m; $g = 0.055$ Vm/N; $p = 1.4$ MN/m^2; $\in (= \in_o \in_r) = 40.6 \times 10^{-12}$ F

(i) Voltage output, *E*:

$$E = gtp \qquad \text{...[Eqn. (3.6)]}$$
$$= 0.055 \times 2.5 \times 10^{-3} \times 1.4 \times 10^{6} = \mathbf{192.5\ V\ (Ans.)}$$

(ii) Charge sensitivity of the crystal:

$$\textit{Charge sensitivity} = \in_o \in_r g = \in_g$$
$$= 40.6 \times 10^{-12} \times 0.055 \text{ C/N}$$
$$= \mathbf{2.233\ pC/N\ (Ans.)}$$

Example 3.9. *A piezoelectric crystal measuring 6 mm × 6 mm × 1.8 mm is used to measure a force. Its voltage sensitivity is 0.055 Vm/N. Calculate the force if voltage developed is 120 V.*

Solution. *Given:* $A = 6 \text{ mm} \times 6 \text{ mm} = 36 \times 10^{-6}$ m^2; $t = 1.8$ mm or 1.8×10^{-3} m;

$$g = 0.055 \text{ Vm/N}; E = 120 \text{ V}$$

Force *F*:

We know that,
$$E = gtp \qquad \text{...[Eqn. (3.6)]}$$
$$120 = 0.055 \times (1.8 \times 10^{-3}) \times p$$

or
$$p = \frac{120}{0.055 \times 1.8 \times 10^{-3}} \text{ N/m}^2 = 1.212 \text{ MN/m}^2$$

∴ Force
$$F = p \times A = 1.212 \times 10^{6} \times 36 \times 10^{-6} = \mathbf{43.63\ N\ (Ans.)}$$

Example 3.10. *The following data relate to a barium titanate pick-up:*

Dimensions	*6 mm × 6 mm × 1.5 mm*
Force acting on the pick-up	*6N*
The charge sensitivity of the crystal	*150 pC/N*
Permittivity	*12.5×10^{-9} F/m*
Modulus of elasticity	*12×10^{6} N/m^2*

Calculate the following:

(i) The strain.

(ii) The charge and capacitance.

Solution. *Given:* $A = 6 \times 6 = 36$ mm^2 or 36×10^{-6} m^2; $t = 1.5$ mm or 1.5×10^{-3} m; $\in = 12.5 \times 10^{-9}$ F/m; $F = 6$ N; d(charge sensitivity) = 150 pC/N; Modulus of elasticity = 12×10^{6} N/m^2.

$$= 150 \text{ pC/N}; \in = 12.5 \times 10^{-9}; E = 12 \times 10^{6} \text{ N/m}^2.$$

(i) The strain, *e*;

$$\text{Pressure, } p = \frac{F}{A} = \frac{6}{36 \times 10^{-6}} \text{ N/m}^2 = 0.167 \text{ MN/m}^2$$

$$\textbf{Strain, } e = \frac{\text{Stress}}{\text{Young's modulus}} = \frac{0.167 \times 10^{6}}{12 \times 10^{6}} = \mathbf{0.0139\ (Ans.)}$$

(*ii*) Charge and capacitance; Q, C:

$$\text{Voltage sensitivity, } g = \frac{d}{\epsilon_0 \epsilon_r} = \frac{d}{\epsilon} = \frac{150 \times 10^{-12}}{12.5 \times 10^{-9}} = 12 \times 10^{-3} \text{ Vm/N}$$

$$\therefore \quad \text{Voltage generated, } E = gtp$$

$$= 12 \times 10^{-3} \times 1.5 \times 10^{-3} \times 0.167 \times 10^{6} = 3\text{V}$$

$$\text{Hence, } charge,\ Q = d \times F = 150 \times 10^{-12} \times 6 \text{ C} = \mathbf{900\ pC\ (Ans.)}$$

$$\text{and,} \quad Capacitance,\ C = \frac{900 \times 10^{-12}}{3} \text{F} = \mathbf{300\ pF\ (Ans.)}$$

3.11 HALL EFFECT TRANSDUCERS

3.11.1. Hall Effect

When a current carrying conductor is placed in a magnetic field, a *transverse effect* is noted. This effect is called Hall effect (discovered by Hall in 1879). Hall found that: "*When a magnetic field is applied at right angles to the direction of electric current an electric field is set up which is perpendicular to both the direction of electric current and the applied magnetic field*".

In other words:

"*When any specimen carrying a current I is placed in the transverse magnetic field B, then an electric field E is induced in the specimen in the direction perpendicular to both I and B. The phenomenon is known as* ***Hall effect***".

In the Fig. 3.22 is shown a specimen bar carrying a current I in the positive-x direction. Let a magnetic field B, be applied in the positive-z direction. Then according to Hall effect, a force gets exerted on the charge carriers (whether electrons or holes) in the negative-y direction. This current I may be due to holes moving in the positive-x or due to *free electrons* moving in the negative-x direction through the *semiconductor specimen.* Hence irrespective of the nature of the charge carriers, whether holes or electrons, these charge carriers get passed *downwards towards face* 1 of the specimen shown in Fig. 3.22.

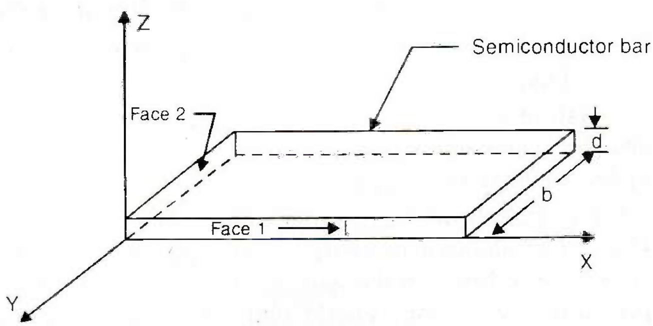


Fig. 3.22. Current carrying semiconductor bar subject to transverse magnetic field.

The current, in an N-type specimen, is carried almost fully by electrons. These electrons, as a result of Hall effect, accumulate on side 1 which surface then gets negatively charged relative to side 2. Consequently, a potential difference develops between surfaces 1 and 2 and is called the '***Hall voltage***' (V_H). This Hall voltage in an N-type semiconductor is positive at surface 2. On the other hand, in a P-type specimen, the Hall voltage is positive at surface 1. These two results have been verified experimentally.

The Polarity of Hall voltage enables us to determine experimentally whether the semiconductor specimen is of N-type or P-type.

The magnitude of Hall voltage (V_H) is given by the expression :

$$V_H = \frac{R_H BI}{b}$$

where, R_H = Hall coefficient,
B = Magnetic field strength,
I = Current carried by the specimen, and
b = Width of the specimen along the magnetic field.

The *Hall effect* may be used for:

1. Determining whether a semiconductor is N-type or P-type.
2. Determining the carrier concentration.
3. Calculating the mobility having measured the conductivity.
4. *Magnetic field meter.* The Hall voltage V_H for a given current is proportional to B. Hence measurement of V_H measures the magnetic field B.
5. *Hall effect multiplier.* The instrument gives an output proportional to the product of two signals. Thus if current I made proportional to one input and if B is proportional to the second input, then *Hall voltage V_H is proportional to the product of the two inputs.*

3.11.2. Hall Effect Transducers

Hall effect transducers are the transducers in which Hall effect is utilised to measure various electrical or non-electrical quantities.

Commercial Hall effect transducers are made from germanium or other semiconductor materials.

The following are the *applications of Hall effect transducers :*

1. Displacement measurement:

Hall effect transducer may be used to measure a linear displacement or to locate a structural element is cases *where it is possible to change the magnetic field strength by variation in the geometry of a magnetic structure.*

Fig. 3.23 shows the arrangement of Hall-effect displacement (linear) transducer . The Hall effect element is located in the gap, adjacent to the permanent magnet. The field strength produced in the gap due to the permanent member is changed by changing the position of a ferromagnetic plate. The voltage output of the Hall effect element is proportional to the field strength in the gap which is function of the position (*i.e.,* displacement) of ferromagnetic plate with respect to the structure.

- With this method the *displacements as small as 0.025 mm can be measured.*

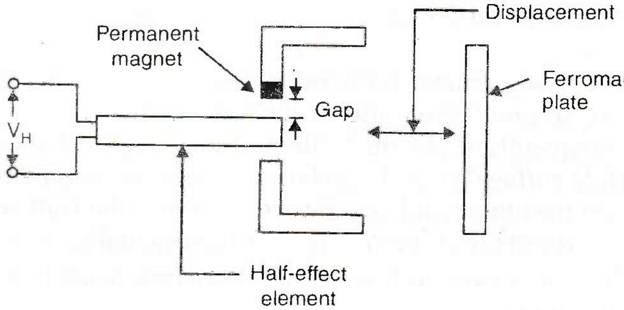


Fig. 3.23. Hall-effect displacement transducer.

2. Current measurement:

Hall effect transducer can be used to measure current in a conductor without interrupting the circuit and without making electrical connection between the conductor circuit and the meter.

When a D.C. or A.C. current flows through the conductor, it sets up a magnetic field around. This *magnetic field is proportional to the current.* A Hall effect transducer is inserted in a slotted ferromagnetic tube which acts as a magnetic concentrator. *The voltage produced at the output terminals is proportional to the magnetic field strength and hence is proportional to the current, flowing through the conductor.*

— The magnetic concentrator can be omitted at high current levels since the magnetic field is fairly strong in the vicinity of the Hall element and thus can cause output voltages which can be detected easily.

- This method can be used *to measure current from less than a mA to thousands of amperes.*

3. Magnetic flux measurement:

The magnetic flux can be measured by using Hall effect transducer. Here, a semiconductor plate is inserted into the magnetic field which is to be measured. The magnetic lines of force are perpendicular to the semiconductor. The transducer gives an output voltage which is *proportional to the magnetic field intensity (B).*

Following are the *advantages* and *disadvantages* of the system:

Advantages:

(*i*) The system requires a very small space in the direction of the magnetic field and thus the Hall effect element can be inserted in narrow gaps for magnetic measurements in air spaces.

(*ii*) The Hall effect element gives out a continuous electric signal in direct response to the magnetic field strength.

Disadvantages:

High sensitivity to temperature variations, and Hall coefficient may vary from plate to plate which may need individual calibration in each case.

4. Fluid level measurement:

Hall effect sensors can be used as position, displacement and proximity sensors if object being sensed with a small permanent magnet.

Such a sensor can be used to determine the level of fuel in an automobile fuel tank. Fig. 3.24 shows a fluid level detector.

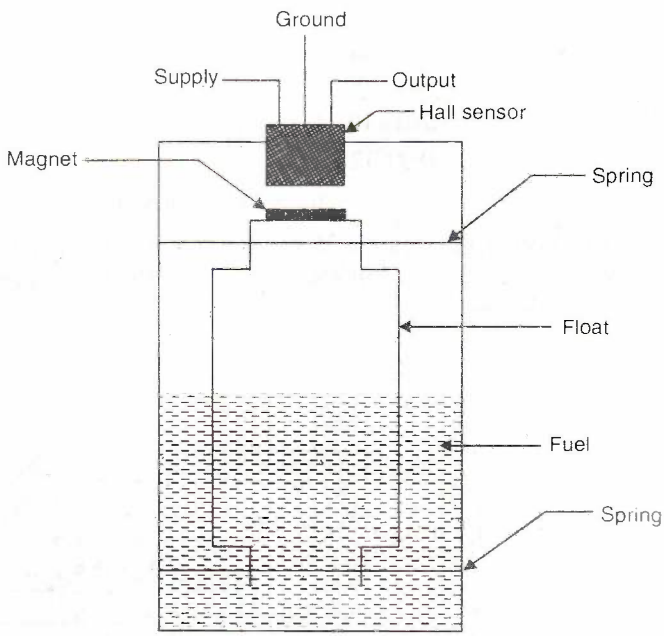


Fig. 3.24. Fluid level detector.

A magnet is attached to a float and as the level of fuel changes and so the float distance from the Hall sensor changes. The result is a Hall voltage output which is a measure of the distance of the float from the sensor and hence the level of the fuel in the tank.

Example 3.11. *The resistivity of semiconductor material was known to be 0.00912 Ω m at room temperature. The flux density in the Hall model was 0.45 Wb/m².*

Calculate the Hall angle for a Hall co-efficient of 3.55 × 10⁻⁴ m³/coloumb.

Solution. Refer to Fig. 3.25.

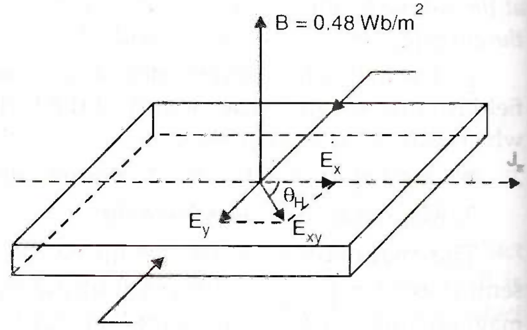


Fig. 3.25

Resistivity of the semiconductor material, $\rho = 0.00912\ \Omega m$

Flux density in the Hall model,

$$B = 0.48 \text{ Wb/m}^2$$

Hall co-efficient,

$$R_H = 3.55 \times 10^{-4} \text{ m}^3/\text{C}$$

Hall angle, θ_H:

Resistivity,
$$\rho = \frac{E_x}{J_x}$$

$$0.00912 = \frac{E_x}{J_x}$$

$$E_x = 0.00912\ J_x$$

Also,
$$R_H = \frac{e_y}{B\ J_x}$$

$$3.55 \times 10^{-4} = \frac{e_y}{0.48 J_x}$$

$$E_y = 3.55 \times 10^{-4} \times 0.48\ J_x = 1.704 \times 10^{-4}\ J_x$$

Now,
$$\tan \theta_H = \frac{E_y}{E_x}$$

$$\frac{1.704 \times 10^{-4} J_x}{0.00912\ J_x} = 0.01868$$

$$\theta_H = \mathbf{1° \ 4' \ (Ans.)}$$

Example 3.12. *Figure 3.26 shows a specimen of silicon doped semiconductor having the Hall co-efficient of 3.55 × 10⁻⁴ m³/coloumb. Calculate the voltage between contacts when a current of 15 mA is flowing.*

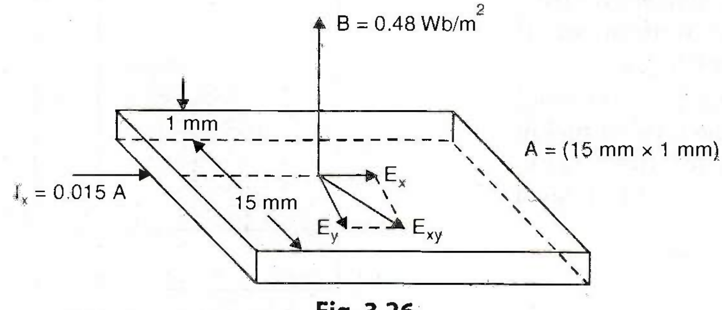


Fig. 3.26

Solution.

$$\text{Hall co-efficient, } R_H = 3.55 \times 10^{-4} \text{ m}^3/\text{C}$$
$$\text{Current, } I = 15 \text{ mA} = 0.015 \text{ A}$$
$$\text{Area, } A = 15 \text{ mm} \times 1 \text{ mm} = 15 \times 10^{-6} \text{ m}^2$$
$$\text{Flux density, } B = 0.48 \text{ Wb/m}^2.$$

Voltage between contacts:

Now, current density, $J_x = \frac{I_x}{A} = \frac{0.015}{15 \times 10^{-6}} = 1000 \text{ A/m}^2$

Hall co-efficient is given by the relation:

$$R_H = \frac{E_y}{BJ_x}$$

$$3.55 \times 10^{-4} = \frac{E_y}{0.48 \times 1000}$$

$$E_y = 3.55 \times 10^{-4} \times 0.48 \times 1000 = 0.1704 \text{ V/m}$$

and, *voltage between contacts*$=0.1704 \times (15 \times 10^{-3}) =$ **0.002556 V (Ans.)**

3.12 THERMOELECTRIC TRANSDUCERS

Two dissimilar metal conductors when joined at the ends and the two junctions kept at different temperatures, then a small e.m.f. is produced in the circuit. *The magnitude of this voltage depends upon the materials of conductors and the temperature difference between the two junctions.* This thermoelectric effect is used in thermocouples for the *measurement of temperature.*

Any number of combination of metals may be used. Two commonly employed combinations are:

1. *Iron and constantan* (an alloy of copper and nickel).
2. *Chromel* (an alloy of chromium and nickel) and *alumel* (an alloy of aluminium and nickel).

3.13 PHOTOELECTRIC TRANSDUCERS

3.13.1. Principle of Operation

The photoelectric transducers *operate on the principle that when light strikes special combination of materials then following may result:*

(*i*) Electrons may flow.

(*ii*) A voltage may be generated.

(*iii*) A resistance change may take place.

3.13.2. Applications

These transducers find the following *fields of application:*

1. Control engineering.
2. Precision measuring devices.
3. Exposure meters used in photography.
4. Solar batteries as sources of electric power for rockets and television, counting machines etc.
5. Satellites used in space research.

3.13.3. Classification

Photoelectric transducers may be *grouped as follows:*

1. Photoemissive cell.
2. Photovoltaic cell.
3. Photoconductive cell.

3.13.4. Photoemissive Cell

This cell is also known as *photo tube*. It is based on the emission of electrons from a metal cathode (or photo-sensitive surface) when it is exposed to light or other radiation.

Refer to Fig. 3.27. It consists of two metallic electrodes (*i.e.*, a cathode and an anode) supported in an evacuated glass bulb fitted with a base like a thermionic valve. The cathode is either semi-cylindrical or V-shaped and is made of a metal coated with an emissive material. The anode is in the form of a thin wire facing the cathode.

When the light falls on the cathode photo-electrons are emitted which are attracted by the positive anode. Subsequently current is produced whose magnitude (for a given cathode) depends on (*i*) intensity of incident radiation and (*ii*) anode cathode voltage.

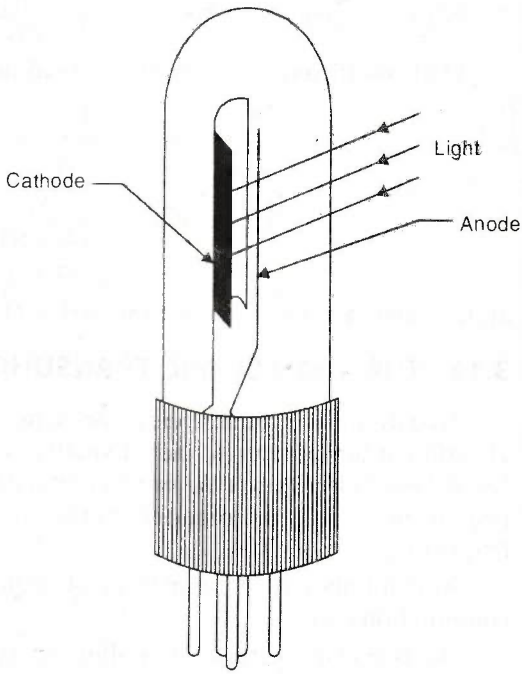


Fig. 3.27. Photoemissive cell.

Photo-emissive cell finds use in: (*i*) field of *photometry and calorimetry*, (*ii*) *sound reproduction* from a *motor-picture film*, (*iii*) *'on and off' circuits* and *other circuits* concerning the counting or sorting of objects on a conveyor belt, automatic opening of a door etc.

3.13.5. Photovoltaic Cell

In this cell sensitive element is a *semiconductor* (not metal) which generates voltage in proportion to the light or any radiant energy incident on it. The most commonly used photo-voltaic cells are barrier layer type like iron-selenium cells or Cu-CuO_2 cells.

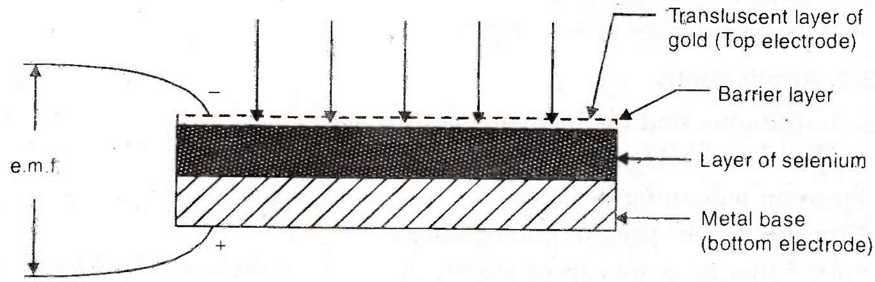


Fig. 3.28. Photovoltaic cell.

Fig. 3.28 shows a typical widely used photo-voltaic cell—"Selenium cell". It consists of a metal electrode on which a layer of selenium is deposited; on the top of this a barrier layer is formed which is coated with a very thin layer of gold. The latter serves as a translucent electrode through which light can impinge on the layer below. Under the influence of this light, a negative charge will build up on the gold electrode and a positive charge on the bottom electrode.

Photo-voltaic cells are widely used in the following *fields:*

(*i*) Automatic control systems.

(*ii*) Television circuits.

(*iii*) Sound motion picture and reproducing equipment.

3.13.6. Photoconductive Cell

"Photoconductive" cell uses a semiconductor material whose resistance changes in accordance with the radiant energy received. The resistivity of semiconductor materials like selenium, cadmium sulphide, lead sulphide and thalmium sulphide is *decreased* when irradiated.

Fig. 3.29 shows the simplest form of such a cell using selenium. There are two electrodes provided with the semiconductor material attached to them. As soon as the cell is illuminated its resistance decreases and current through the circuit becomes large. The shape of the semiconductor material is so made as to obtain a large ratio of 'dark to light' resistance.

A cadmium sulphide cell has two electrodes which are extended in an inter-digital pattern in order to increase the contact area with the sensitive material. It has high 'dark to light' ratio.

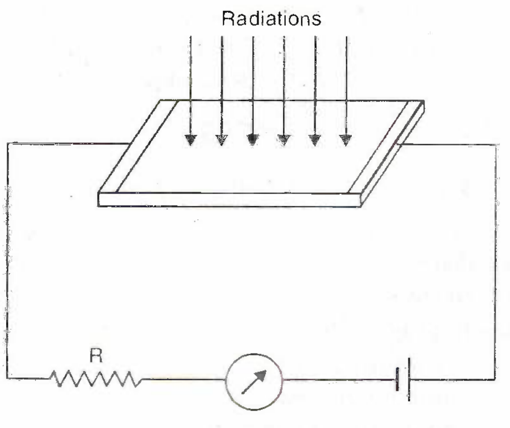


Fig. 3.29. Photoconductive cell.

Photoconductive cells are generally used for *detecting ships and aircrafts by the radiations given out by their exhausts or (funnels) and for telephony by modulated infrared lights.*

3.13.7. Photoelectric Tachometer

Fig. 3.30 shows a photoelectric tachometer.

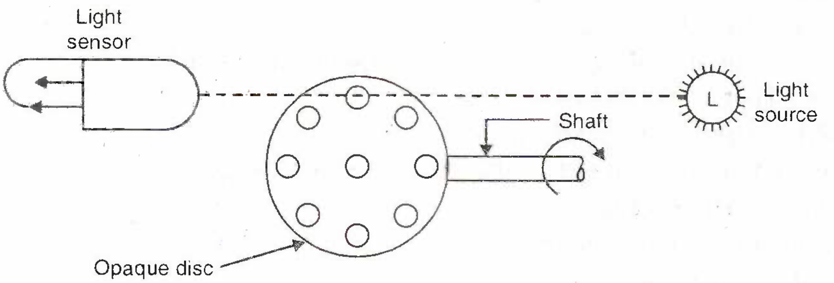


Fig. 3.30. Photoelectric tachometer

- It consists of an *opaque disc* mounted on the shaft whose speed is to be measured. The disc has a number of equivalent holes around the periphery. On one side of the disc there is a source of light (L) while on the other side there is a light sensor (may be a photosensitive device or phototube) in line with it (light-source).

- On the rotation of the disc, holes and opaque portions of the disc come alternatory in between the light source and the light sensor. When a hole comes in between the *two*, light passes through the holes and falls on the light sensor, with the result that an *output pulse is generated.* But when the opaque portion of the disc comes in between, the light from the source is blocked and hence there is no pulse output. Thus *whenever a hole comes in line with the light source and sensor, a pulse is generated.* These pulses are counted/measured through an *electric counter.*

The number of pulses generated depends upon the following *factors :*

(*i*) The number of holes in the disc;

(*ii*) The shaft speed.

Since the number of holes are fixed, therefore, the number of pulses generated depends on the speed of the shaft only. The electronic counter may therefore be calibrated in terms of speed (r.p.m.)

Advantages. It is a *digital instrument.*

Disadvantages. It is required to replace the light source periodically and if the grating period is small then errors might creep in the output.

3.14 STRAIN GAUGES

3.14.1. Introduction

When a metal conductor is stretched or compressed, its resistance changes on account of the fact that both length and diameter of conductor change. The value of resistivity of the conductor also changes. When it is strained its property is called ***piezo-resistance.*** Therefore, resistance strain gauges are also known as ***piezo-resistive gauges.***

The **strain** *gauge is a measurement transducer for measuring strain and associated stress in experimental stress analysis.* Secondly many other detectors, and transducers, notably the load cells, torque meters, diaphragm type pressure gauges, temperature sensors, accelerometers and flow meters, employ *a strain gauge as a secondary transducer.*

3.14.2. Type of Strain Gauges

Four types of strain gauges are:

1. Wire-wound strain gauges.
2. Foil-type strain gauges.
3. Semiconductor strain gauges.
4. Capacitive strain gauges.

(Although these strain gauges have been discussed in chapter 4 they are being dealt with in details again for better understanding by the reader.)

3.14.2.1. Wire-wound strain gauges

There are two main classes of wire-wound strain gauges:

1. Bonded strain gauge.
2. Unbonded strain gauge.

Bonded strain gauge:

It is composed of fine wire, wound and cemented on a resilient insulating support, usually a wafer unit. Such units may be mounted upon or incorporated in mechanical elements or structures whose deformations under stress are to be determined. While there are no limits to the basic values which may be selected for strain-gauge resistances, a typical example may be taken as of the order or 100 to 500 Ω.

Fig. 3.31 shows the commonly used form of resistance wire strain gauges.

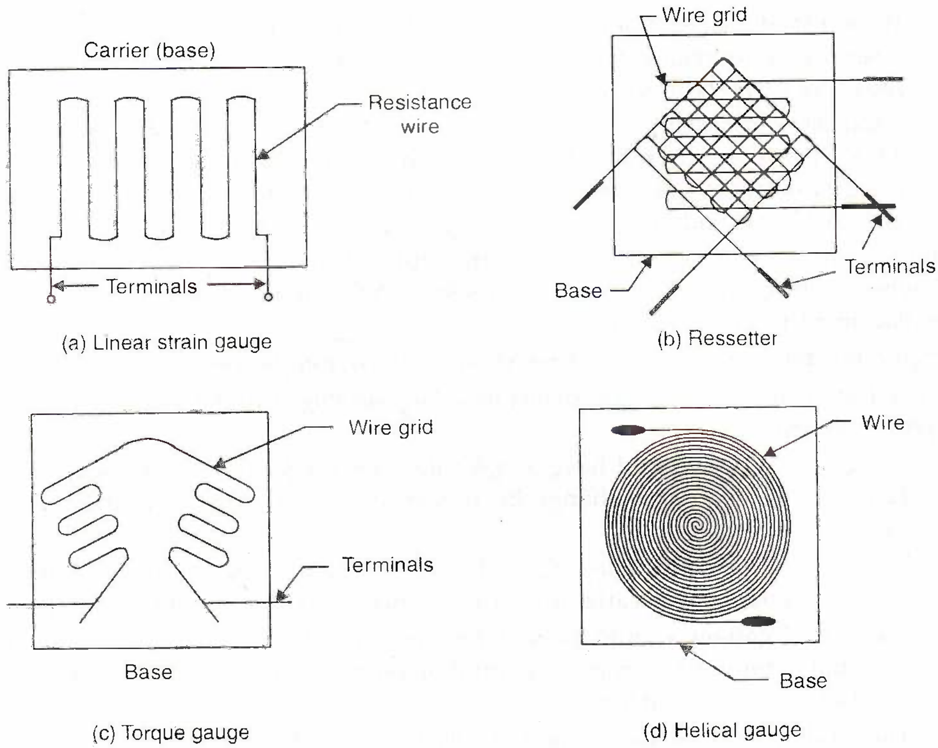


Fig. 3.31. Resistance wire strain gauges.

Unbonded strain gauge:

Figure 3.32 shows an *unbonded strain gauge*. *M* and *N* are attached by rods *m* and *n*, respectively, to points between which *displacement is to be measured*. Pick-up and measuring networks are energized from similar but isolated source. Unbalance originating in pick-up is detected and balanced by servo-actuated measuring network, *providing a reading of strain on graduated scale.*

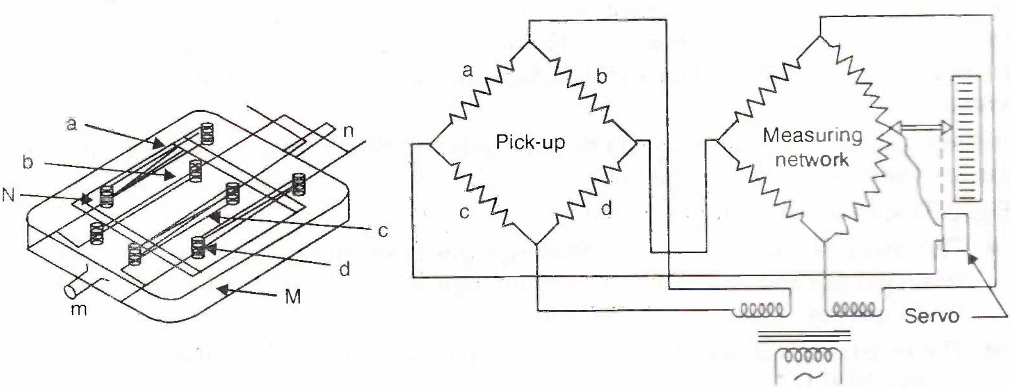


Fig. 3.32. Unbonded strain gauge.

In the unbonded strain gauge the resistance structure comprises of fine wire winding stretched between insulating supports mounted alternately on the two members between which displacement is to be measured (see Fig. 3.32). These wires comprise the four arms of a Wheatstone-bridge network of *which two opposite arms are tightened and the other two slackened by the displacement.*

— *While a bonded gauge tends to respond to the average strain in the surface to which it is cemented, the unbonded form measures displacement between the two points to which the respective supports are attached.*

— *Unbonded wire strain gauges* are usually operated on input potentials ranging upto 35 V direct or alternating current. Under conditions of extreme balance, corresponding to full operating range, the open-circuit e.m.f. may be of the order of 8 to 10 mV and closed circuit current upto 100 μA.

Strain gauges for use on A.C. circuits are supplied in both capacitive and inductive forms, wherein the corresponding characteristics of A.C. circuit components are varied by the displacement to be measured.

Requirements/Characteristics of resistance wire strain gauges:

The resistance wire strain gauges should have the following characteristics to have excellent and reproducible results.

1. The strain gauge should have a *high value of gauge factor.* A high value of gauge factor indicates a large change in resistance for a particular strain resulting in high sensitivity.
2. The *resistance of strain gauge should be as high as possible* since this minimizes the effects of undesirable variations of resistance in the measurement circuit.
3. The strain gauges should have a *low resistance temperature co-efficient.* This is essential to minimize errors on account of temperature variations which affect the accuracy of measurements.
4. The strain gauge *should not have any hysteresis effect in its response.*
5. In order to maintain constancy of calibration over the entire range of strain gauge, it should have *linear characteristics i.e.,* the variations in resistance should be a linear function of the strain.
6. The strain gauges are frequently used for dynamic measurements and hence their frequency response should be good. The linearity should be maintained within accuracy limits over the entire frequency range.

3.14.2.2. Foil strain gauges

In these gauges *the strain is sensed with the help of metal foil.* Foil gauges have a *much greater dissipation capacity as compared with wire wound gauges on account of their greater surface area for the same volume.* Due to this reason they can be employed for *higher operating temperature range.*

In these gauges, the bounding is better due to large surface area of the foil. The *bonded foil* gauges *find* a wider field of action.

Fig. 3.33 shows a typical foil gauge.

- The characteristics of foil type strain gauges are similar to those of wire wound strain gauges and their gauge factors are typically the same as that of wire wound strain gauges.
- The resistance value of foil gauges which are commercially available is between 50 and 1000 Ω.

The *advantage* of foil type strain gauges is that they can be *fabricated economically on a mass scale.*

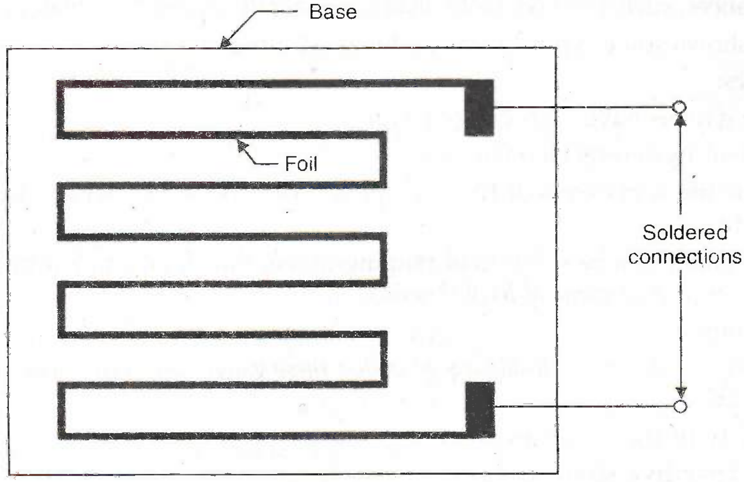


Fig. 3.33. Foil gauge.

3.14.2.3. Semiconductor strain gauges

- *Semiconductor strain gauges depend for their action upon piezo-resistive effect i.e., the change in value of the resistance due to change in resistivity.*
- These gauges are used where a *very high gauge factor and small envelope are required.*

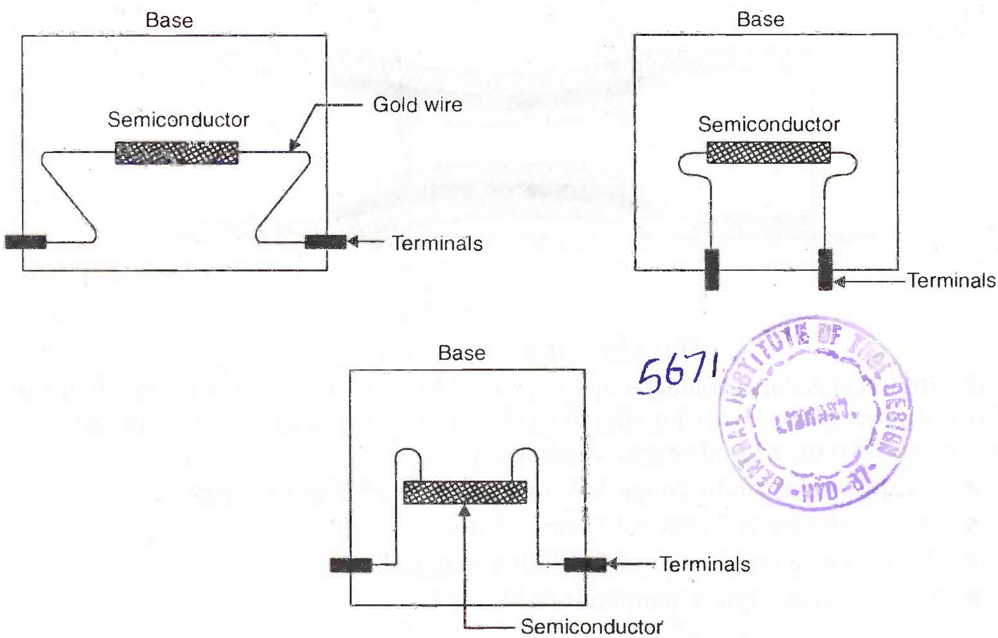


Fig. 3.34. Semiconductor strain gauges.

- For semiconductor strain gauges semiconducting materials such as *silicon* and *germanuium* are used.
- A typical strain gauge consists of a *strain sensitive crystal material and leads that are sandwiched in a protective matrix.* The production of these gauges employs

conventional semiconductor technology using semiconducting wafers or filaments which have a thickness of 0.05 mm and bonding them on suitable insulating substances, such as teflon. Gold, lead are generally applied for making the contacts.

Fig. 3.34 shows some typical semiconductor strain gauges.

Advantages:

1. These gauges have high gauge factor.
2. Excellent hysteresis characteristics.
3. Fatigue life is in excess of 10×10^5 operations and the frequency response is upto 10^{12} Hz.
4. These gauge can be very small ranging in length from 0.7 to 7 mm. *They are very useful for measurement of local strains.*

Disadvantages:

1. The *major and serious disadvantage is that these gauges are very sensitive to change in temperature.*
2. Linearity of these gauges is *poor.*

3.14.2.4. Capacitive strain gauges

Fig. 3.35 shows a capacitive strain gauge. It uses the *principle of variation of capacitance with variation of distance between electrodes.* The electrodes are flexible metal strips of about 0.1 mm thickness. The strain to be measured is applied to the top plate. This changes the distance between the curved electrodes resulting in change of capacitance.

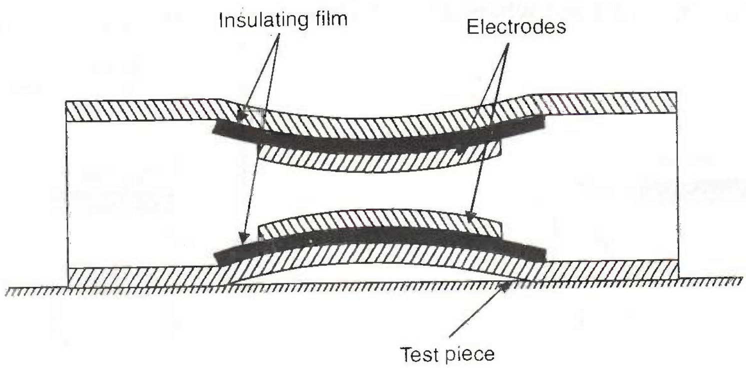


Fig. 3.35. Capacitive strain gauge.

The strain-capacitance relationship, in general, is *not linear* but variations in dimensions and shape allow gauge characteristics to be chosen so as to match the range of capacitance to be measured with a good degree of accuracy.

- A capacitance strain gauge has a capacitance of about 0.5 pF.
- Its overall size is 5 mm × 17 mm × 1 mm.
- It uses a polyamide film of insulating material.
- It can be used upto a temperature of 300°C.

3.14.3. Theory of Strain Gauges

When a strain gauge is subjected to tension (*i.e.*, positive strain) its length *increases* while its cross-sectional area *decreases*. Since the resistance of a conductor is proportional to its length and inversely proportional to its area of cross-section, the resistance of the gauge increases with positive strain. The change in the value of resistance of strained conductor is more than what can be accounted for an increase in resistance due to dimensional changes. The extra change in the value of resistance is attributed to a *change in the value of resistivity of a conductor when strained; this property is known as* ***piezo-resistive effect.***

Strain gauges are most commonly used in wheatstone bridge circuits to measure the change of resistance of grid of wire for calibration proposes; the "**gauge factor**" *is defined as the ratio of per unit change in resistance to per unit change in length.*

i.e., $$\text{Gauge factor } (G_f) = \frac{\Delta R/R}{\Delta L/L} \quad ...(3.8)$$

where, ΔR = Corresponding change in resistance R, and

ΔL = Change in length per unit length L.

The resistance of the wire of strain gauge R is given by :

$$R = \frac{\rho L}{A}$$

where, ρ = Resistivity of the material of wire (of strain gauge),

L = Length of the wire, and

A = Cross-sectional area of the wire,

= KD^2, K and D being a constant and diameter of the wire respectively.

As earlier stated, when the wire is strained its length increases and lateral dimension is reduced as a function of Poisson's ratio (μ); consequently there is an increase in resistance.

Now, $$R = \frac{\rho L}{KD^2}$$

Differentiating it, get we

$$dR = \frac{KD^2(\rho.dL + L.d\rho) - \rho L(2KD.dD)}{(KD^2)^2}$$

$$= \frac{1}{KD^2}\left[(\rho \,.\, dL + L \,.\, d\rho) - 2\rho L \,.\, \frac{dD}{D}\right]$$

or, $$\frac{dR}{R} = \frac{\frac{1}{KD^2}\left[\rho \,.\, dL + L \,.\, d\rho - 2\rho L \,.\, \frac{dD}{D}\right]}{\frac{\rho L}{KD^2}}$$

$$= \frac{dL}{L} + \frac{d\rho}{\rho} - 2\frac{dD}{D}$$

Now, Poisson's ratio, $$\mu = \frac{\text{Lateral strain}}{\text{Longitudinal strain}} = \frac{-dD/D}{dL/L}$$

or, $$\frac{dD}{D} = -\mu \times \frac{dL}{L}$$

For small variations, the above relationship can be written as:

$$\frac{\Delta R}{R} = \frac{\Delta L}{L} + 2\mu\frac{\Delta L}{L} + \frac{\Delta\rho}{\rho} \quad ...(3.9)$$

Gauge factor, $$G_f = \frac{\Delta R/R}{\Delta L/L}$$

or, $$\frac{\Delta R}{R} = G_f \cdot \frac{\Delta L}{L} = G_f \times e \quad ...(3.10)$$

where, e = strain = $\frac{\Delta L}{L}$

The *gauge factor* can be written as:

$$G_f = 1 + 2\mu + \frac{\Delta\rho/\rho}{e} \quad ...(3.11)$$

$$= 1 + 2\mu + \frac{\Delta\rho/\rho}{e}$$

= Resistance change due to change of length + Resistance change due to change in area + Resistance change due to piezo-resistive effect

or, $$G_f = 1 + 2\mu + \frac{\Delta\rho/\rho}{\Delta L/L} \quad ...(3.12)$$

The strain is usually expressed in terms of *microstrain;* 1 micro strain = 1 μm/m

If the change in the value of resistisity of a material when strained is *neglected*, the gauge factor can be rewritten as:

$$G_f = 1 + 2\mu \quad ...(3.13)$$

Eqn. (3.13) is valid only when *piezo-resistive effect* (*i.e.*, change in resistisity due to strain) is *almost negligible.*

- The Poission's ratio for all metals lies between 0 and 0.5. This give G_f as 2 approximately. In case of wire wound strain gauges where the common value for Poisson's ratio is 0.3, the value of G_f amounts to 1.6.
- The value of the gauge factor varies from material to material but it is generally assumed that it remains constant in the working range of strain required. *Its value is determined experimentally.*
- *Knowing the gauge factor (G_f), the strain in the member can be directly found out by the change of resistance.*

Properties of gauge materials:

The grid material for its proper functioning must possess the following desirable *properties:*

1. High resistivity.
2. High gauge factor.
3. High mechanical strength.
4. High electrical stability.
5. Low temperature sensitivity.
6. Low hysteresis.
7. Low thermal e.m.f. when joined with other materials.
8. Good corrosion resistance.
9. Good weldabililty.

Adhesing techniques:

For proper mounting of the strain gauges, the following *steps* should be strictly followed:

1. Before mounting the strain gauge on the surface, the surface must be preferably cleaned by emery cloth and base material exposed.
2. Remove the various traces of grease or oil etc. by using a solvent like acetone.
3. Swab the back of the strain gauge by cotton dipped in acetone once, to ensure that the back is free from grease etc.
4. Apply a generous quantity of cement to the cleaned resistance and then simply place the cleaned gauge on it and excess cement worked out. Make sure that there is no bubble between the surface and the gauge, if any one is there, that should be removed. Avoid using heavy pressure, otherwise cement may puncture the paper and short the grid.
5. Allow the gauge to sit for at least eight or ten hours before using it. If possible a slight weight might be placed by keeping a strong rubber on the gauge.
6. After cement has been fully cured, check the continuity of wire by an ohmmeter and weld the electric leads.

Example 3.13. *The gauge factor of a resistance wire strain gauge using a soft iron wire of small diameter is 4.2. Neglecting the piezo-resistive effect, calculate the Poisson's ratio.*

Solution. *Given:* $G_f = 4.2$

When the *piezo-resistive effect* is neglected, the gauge factor is given by:

$$G_f = 1 + 2\mu \qquad \text{...[Eqn. (3.13)]}$$

or,

$$4.2 = 1 + 2\mu$$

$$\therefore \quad \mu = \frac{4.2-1}{2} = \mathbf{1.6\ (Ans.)}$$

Example 3.14. *A simple electrical strain gauge of resistance 120 Ω and having a gauge factor of 2 is bonded to steel having an elastic limit stress of 400 MN/m² and modulus of elasticity is 200 GN/m². Calculate the change in resistance,*

(i) due to a change in stress equal to $\frac{1}{10}$ *of the elastic range;*

(ii) due to change of temperature of 20°C if the material is advance alloy. The resistance temperature coefficient of advance alloy is 20×10^{-6}/°C.

Solution. *Given:* $R = 120\ \Omega$; $G_f = 2$; Elastic limit stress = 400 MN/m²; Modulus of elasticity = 200 GN/m²; Resistance temperature coefficient, $\alpha_0 = 20 \times 10^{-6}$/°C.

Change in resistance:

(i)
$$\text{Change in stress} = \frac{1}{10} \times 400\ \text{MN/m}^2 = 40 \times 10^6\ \text{N/m}^2$$

$$\text{Modulus of elasticity} = 200\ \text{GN/m}^2 = 200 \times 10^{12}\ \text{N/m}^2$$

$$\text{Strain, } e = \frac{\text{Stress}}{\text{Modulus of elasticity}} = \frac{40 \times 10^6}{200 \times 10^{12}} = \frac{1}{5} \times 10^{-6}$$

$$\text{Gauge factor } G_f = \frac{\text{Per unit change in resistance}}{\text{Per unit change in length}}$$

or

$$G_f = \frac{\Delta R/R}{e} \quad \text{or} \quad \Delta R = R\, G_f e$$

or, $$\Delta R = 120 \times 2 \times \frac{1}{5} \times 10^{-6} = 48 \times 10^{-6}\ \Omega = 45\ \mu\Omega \textbf{ (Ans.)}$$

(*ii*) $$R_{t2} = R_{t1}[1 + \alpha_0(t_2 - t_1)]$$

$\therefore$ Change in resistance $R_{t2} - R_{t1} = R_{t1}\,\alpha_0(t_2 - t_1)$

or, $$\Delta R = R_{t2} - R_{t1} = 120 \times 20 \times 10^{-6} \times (20)$$
$$= 48 \times 10^{-3}\ \Omega = 48\ \text{m}\Omega \textbf{ (Ans.)}$$

Example 3.15. *A strain gauge is bounded to a beam which is 12 cm long and has a cross-sectional area of 3.8 cm². The unstrained resistance and gauge factor of the strain gauge are 220 Ω and 2.2 respectively. On the application of load the resistance of the gauge changes by 0.015 Ω. If the modulus of elasticity for steel is 207 GN/m², calculate:*

(*i*) *The change in length of the steel beam.*

(*ii*) *The amount of force applied to the beam.*

Solution. *Given:* $L = 12$ cm $= 0.12$ m; $A = 3.8$ cm² $= 3.8 \times 10^{-4}$ m²; $R = 220\ \Omega$; $G_f = 2.2$; $\Delta R = 0.015\ \Omega$; $E = 207$ GN/m².

(*i*) The Change in length of steel beam. ΔL:

$$\text{Gauge factor, } G_f = \frac{\Delta R/R}{\Delta L/L}$$

$$\therefore \quad \Delta L = \frac{(\Delta R/R).L}{G_f} = \frac{(0.015/220) \times 0.12}{2.2} = 3.72 \times 10^{-6} \text{ m } \textbf{(Ans.)}$$

(*ii*) The amount of force applied to the beam, *F*:

$$E = \frac{\text{Stress}}{\text{Strain}} = \frac{\sigma}{e}$$

$$\therefore \quad \sigma = E \times e = E \times \frac{\Delta L}{L}$$

$$= (207 \times 10)^9 \times \frac{3.72 \times 10^{-6}}{0.12} = 6.417 \times 10^6 \text{ N/m}^2$$

$$\therefore \quad \textit{Force, } F = \sigma \cdot A = 6.417 \times 10^6 \times 3.8 \times 10^{-4} \text{ N} = \textbf{2.438 kN (Ans.)}$$

3.14.4. Strain-gauge Circuits

The following strain gauge circuits will be discussed :

1. Ballast circuit.
2. Wheatstone bridge circuit.
   (*i*) Balanced (null) condition
   (*ii*) Unbalanced (deflection) condition
   — Quarter bridge
   — Half bridge
   — Full bridge.

3.14.4.1. Ballast-circuit (Voltage-sensitive potentiometric circuit)

Fig. 3.36 shows a ballast circuit-voltage-sensitive potentiometric circuit. Here,

v_i = Input supply voltage,

v_0 = Output voltage,

R_b = Ballast resistance, and

R_g = Resistance of the unstrained resistance gauge.

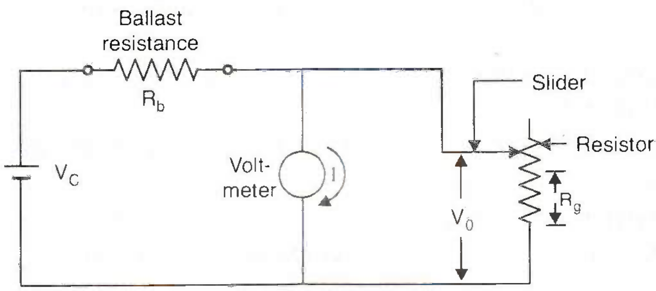


Fig. 3.36. Ballast circuit-voltage-sensitive potentiometric circuit.

The output voltage, when *no stress* is applied to the strain gauge, is given by

$$v_0 = \left(\frac{R_g}{R_g + R_b}\right) v_i \quad \text{...(3.14)}$$

When the gauge is strained, the gauge resistance changes to $(R_g + dR_g)$ and the output voltage becomes,

$$v_0 + dv_0 = \left[\frac{(R_g + dR_g)}{(R_g + dR_g) + R_b}\right] v_i \quad \text{...(3.15)}$$

∴ The *change in the output voltage,*

$$dv_0 = \left[\frac{(R_g + dR_g)}{(R_g + dR_g) + R_b} - \frac{R_g}{R_g + R_b}\right] v_i$$

$$= \left[\frac{(R_g + dR_g)(R_g + R_b) - R_g\{(R_g + dR_g) + R_b\}}{\{(R_g + dR_g) + R_b\}(R_g + R_b)}\right] v_i$$

$$= \left[\frac{\left(R_g^2 + R_g R_b + dR_g R_g + dR_g R_b - R_g^2 - dR_g R_g - R_g R_b\right)}{(R_g + R_b)^2}\right] v_i$$

$$\text{...}(\because \; dR_g > R_g)$$

$$= \left[\frac{dR_g \cdot R_b}{(R_g + R_b)^2}\right] v_i = \frac{R_g \cdot R_b}{(R_g + R_b)^2} \cdot \frac{dR_g}{R_g} v_i \quad \text{...(3.16)}$$

... Multiplying numerator and denominator by R_g.

Also, *condition of maximum sensitivity* is given by : $R_g = R_b$

Hence, $$dv_0 = \frac{v_i}{4} \cdot \frac{dR_g}{R_g}$$

Also, $$\frac{dR_g}{R_g} = G_f \cdot e$$

$$\therefore \qquad dv_0 = \left(\frac{G_f}{4}\right) e\, v_i \quad \text{...where, } G_f \text{ denotes the gauge factor} \qquad \text{...(3.16a)}$$

From eqn. (3.16*a*) it is evident that *change in output-voltage when gauge is strained is directly proportional to strain.*

- The ballast circuit is used for *dynamic strain measurements where static strain components are ignored.*

Limitations of potentiometric circuits:

(*i*) No possibility of compensation for temperature variations.

(*ii*) High sensitivity to A.C. interference giving hum due to ground loops, induction from high current lines and poor connections.

3.14.4.2. Wheatstone bridge circuit

The wheatstone bridge technique can be used in the following *two ways*: (*i*) Null mode; (*ii*) Deflection mode.

1. Null mode:

Refer to Fig. 3.37. The resistance, with *no straining*, are so arranged that $v_B = v_D$ and the galvanometer gives zero deflection.

Then,
$$\frac{R_1}{R_3} = \frac{R_2}{R_4} \qquad \text{...(3.17)}$$

where, $R_1 = R_g$ = Unstrained resistance of the gauge.

In measurement of strains, generally R_1 is the strain gauge, R_2 and R_4 are the fixed resistances and R_3 is a variable resistor.

When the gauge is strained, its resistance R_1 changes by an amount dR_1. This change unbalances the bridge resulting into the deflection of the galvanometer. The balance is then regained by adjusting R_3 by an amount dR_3. The rebalanced conditions gives:

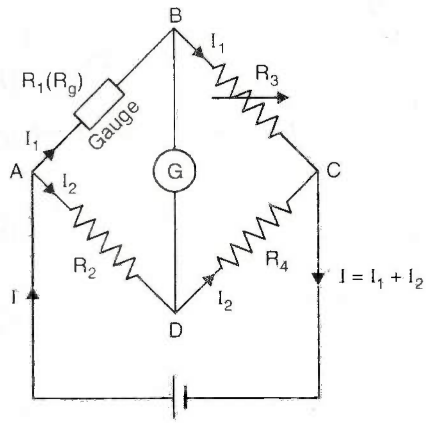


Fig. 3.37. Wheatstone bridge circuit.

$$\frac{R_1 + dR_1}{R_3 + dR_3} = \frac{R_2}{R_4}$$

or,
$$R_1 + dR_1 = (R_3 + dR_3)\,\frac{R_2}{R_4}$$

or,
$$R_1 + dR_1 = R_3 \times \frac{R_2}{R_4} + dR_3 \times \frac{R_2}{R_4}$$

or,
$$R_1 + dR_1 = R_1 + dR_3 \times \frac{R_2}{R_4} \qquad \left[\because R_1 = \frac{R_3}{R_4} \times R_2 \text{ from eqn. (3.17)}\right]$$

or,
$$dR_1 = dR_3 \times \left(\frac{R_2}{R_4}\right) \qquad \text{...(3.18)}$$

If the resistances, of all the limbs of the wheatstone bridge are equal, then

$$R_1 = R_2 = R_3 = R_4 = R_g$$

and, $$dR_1 = dR_3 \quad ...(3.19)$$

The change in resistance dR_1, in terms of strain, is given by

$dR_1 = G_f\, e\, R_g$, where G_f is the gauge factor and e is the strain

$$\therefore \quad dR_3 = G_f\, e\, R_g \quad ...(3.20)$$

Eqn. (3.20) indicates that the *change in the value of resistance R_3 is direct measurement of strain.*

2. Deflection mode:

Initially the bridge resistances are so adjusted that the bridge is in balanced. The equilibrium gets disturbed when the gauges are strained. Then, the voltage v_0 is measured under this unbalanced condition.

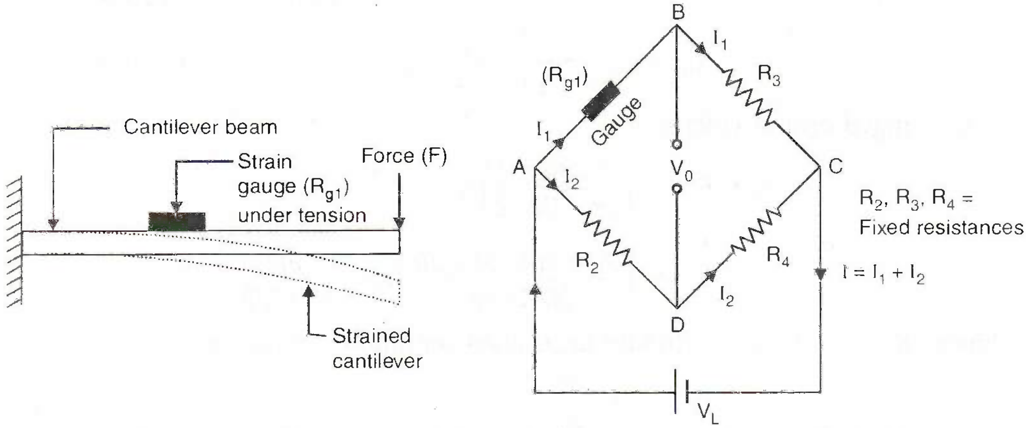


Fig. 3.38. Single gauge used for strain measurement (Quarter-bridge).

(i) Quarter-bridge:

Fig. 3.38 shows single gauge used for strain measurement (quarter-bridge). In this arrangement only one strain gauge is used and the other three elements of the bridge are fixed resistors.

Let us assume that the galvanometer (measuring instrument) has infinite impedance and therefore no current flows through it. Then,

Current flowing through the limbs AB and BC, $I_1 = \dfrac{v_i}{R_{g1} + R_3}$ ...(3.21)

Voltage drop in limb AB (or voltage at terminal B),

$$v_{AB} = I_1 R_{g1} = \frac{R_{g1}}{R_{g1} + R_3} \cdot v_i \quad ...(3.22)$$

Similarly, $$I_2 = \frac{v_i}{R_2 + R_4} \quad ...(3.23)$$

and, $$v_{AD} = \frac{R_2}{R_2 + R_4} \cdot v_i \quad ...(3.24)$$

Initially, $R_{g1} = R_2 = R_3 = R_4 = R$

$$\therefore \quad v_{AB} = v_{AD} = \frac{v_i}{2} \qquad ...(3.25)$$

and, v_0 = Voltage across the terminals B and D

$$= v_{AB} - v_{AD} = 0$$

Obviously, the bridge is balanced under *unstrained conditions.*

When the gauge is strained (see Fig. 3.38), the resistance R_{g1} changes by an amount dR_{g1}. Then,

$$v_{AB} = \left[\frac{R_{g1} + dR_{g1}}{(R_{g1} + dR_{g1}) + R_3}\right] v_i = \left(\frac{R + dR}{2R + dR}\right) v_i$$

$$[\because Rg_1 = R_3 = R \text{ and } dR_{g1} = dR]$$

$$v_{AD} = \left(\frac{R_2}{R_2 + R_4}\right) v_i = \frac{v_i}{2} \qquad (\because R_2 = R_4 = R)$$

The changed output voltage,

$$v_0 + dv_0 = \left(\frac{R + dR}{2R + dR} - \frac{1}{2}\right) v_i$$

$$= \left(\frac{2R + 2dR - 2R - dR}{2(2R + dR)}\right) v_i = \left(\frac{dR}{4R + 2dR}\right) v_i$$

Since $dR << R$ and $v_i = 0$ (under unstrained conditions), therefore

$$dv_0 = \frac{v_i}{4} \cdot \frac{dR}{R} \qquad ...(3.26)$$

or,

$$dv_0 = \left(\frac{G_f}{4}\right) e \cdot v_i \qquad ...(3.27)$$

...in terms of gauge factor G_f and applied strain e

From eqn. (3.27) it is obvious that the *output voltage is directly proportional to the applied strain.*

(ii) Half-bridge:

Fig. 3.39, shows two gauges used for strain measurement (Half-bridge). In this arrangement two of the bridge elements are strain gauges and the other two are fixed resistors. The strain gauge-1 is bonded to the upper surface of the cantilever beam and a second strain gauge-3 is bonded to the lower surface and located precisely underneath the gauge-1. These gauges are connected electrically to form adjacent limbs of the Wheatstone bridge circuit.

The temperature effects are cancelled out by having $R_2 = R_4$ and using *two identical gauges* in the opposite arms of the bridge.

Suppose, $R_{g1} = R_{g3} = R_2 = R_4 = R$

Under *no strain conditions:*

$$v_{AB} = v_{AD} = \frac{v_i}{2};\ v_B = v_D \text{ and } v_0 = 0.$$

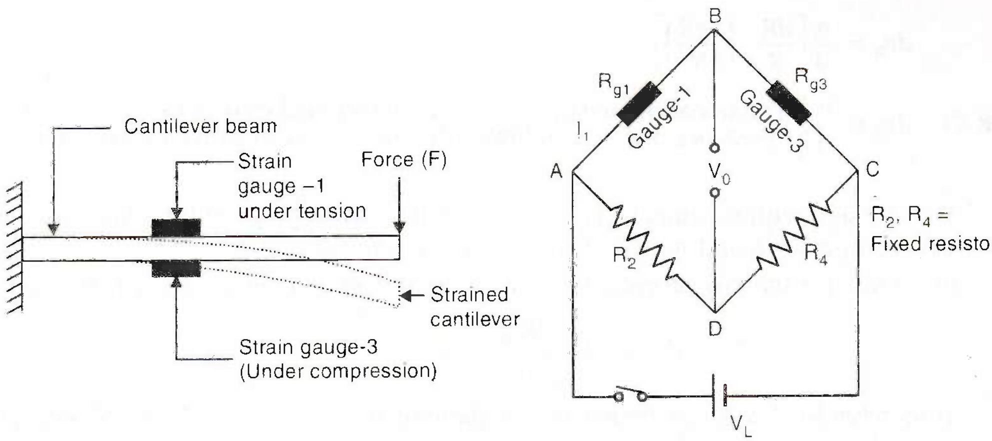


Fig. 3.39. Two gauges used for strain measurement (Half-bridge).

On the application of load to the cantilever beam, the resistance of the gauge R_{g1} *increases* due to *tensile load* whilst R_{g3} *decreased* due to equal *compressive strain* so that,

Resistance of gauge – 1 = $R_{g1} + dR_{g1}$

and, Resistance of gauge – 3 = $R_{g3} - dR_{g3}$

Now,
$$v_{AB} = \frac{R_{g1}}{R_{g1}+R_{g3}} \cdot v_i$$

$$= \frac{R+dR}{(R+dR)+(R-dR)} v_i = \frac{R+dR}{2R} \cdot v_i \qquad \text{...(3.28)}$$

and,
$$v_{AD} = \frac{R_2}{R_2+R_4} v_i = \frac{v_i}{2} \qquad \text{...(3.29)} \quad (\because R_2 = R_4)$$

The changed output voltage,

$$v_0 + dv_0 = \frac{R+dR}{2R} \cdot v_i - \frac{v_i}{2}$$

$$= v_i\left(\frac{R+dR}{2R} - \frac{1}{2}\right) = v_i\left(\frac{2R+2dR-2R}{4R}\right)$$

or,
$$v_0 + dv_0 = \frac{v_i}{2} \cdot \frac{dR}{R} \qquad \text{...(3.30)}$$

Since under unstrained conditions $v_0 = 0$, therefore, change in output voltage due to applied strain becomes,

$$dv_0 = \frac{v_i}{2} \cdot \frac{dR}{R}$$

or,
$$dv_0 = \left(\frac{G_f}{2}\right) e \cdot v_i \qquad \text{...(3.31)}$$

which is *twice* the output of Wheatstone bridge using one gauge only.

The eqn. (3.31) can be rewritten as:

$$dv_0 = \frac{v_i}{4}\left[\frac{dR}{R} - \left(\frac{-dR}{R}\right)\right]$$

or, $$dv_0 = \frac{v_i}{4}\left[\begin{Bmatrix}\text{Fractional change in} \\ \text{resistance of gauge in limb } AB\end{Bmatrix} - \begin{Bmatrix}\text{Fractional change in} \\ \text{resistance of gauge in limb } BC\end{Bmatrix}\right] \quad ...(3.32)$$

The –ve sign with fractional change in resistance of the gauge in limb BC is due to the fact that compressive and tensile strain are of *opposite signs*.

In general, for the *two gauges connected in the adjacent limbs of a bridge circuit*, we have:

$$dv_0 = \frac{v_i}{4}\left(\frac{dR_{g1}}{R} - \frac{dR_{g3}}{R}\right) \quad ...(3.33)$$

Thus, *when both the gauges are mounted on the top of the cantilever beam, the two effects cancel each other and the output voltage is zero.*

(*iii*) *Full-bridge:*

Fig. 3.40, shows four gauges used for stain measurement (Full-bridge). In this arrangement all the four elements of the bridge are strain gauges.

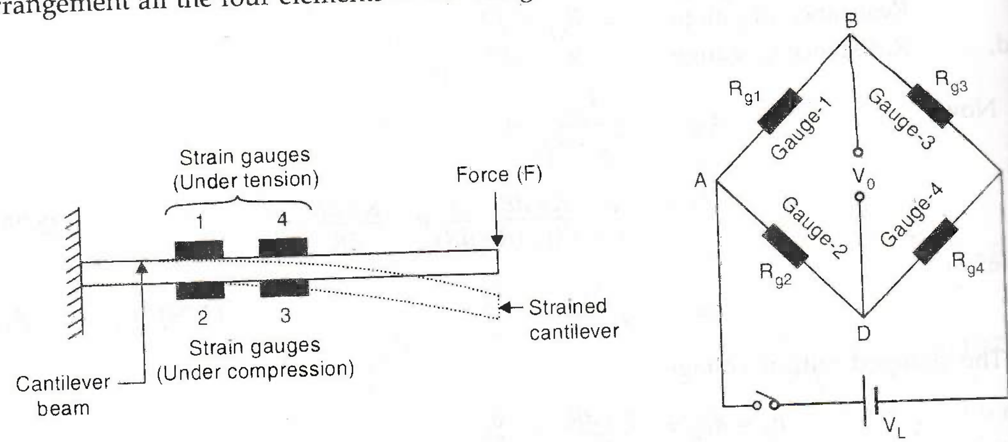


Fig. 3.40. Four gauges used for strain measurement (Full-bridge).

All the four gauges are similar and have equal resistances when unstained, *i.e.*,

$$R_{g1} = R_{g2} = R_{g3} = R_{g4} = R.$$

Under no-strain condition : $v_{AB} = v_{AD} = \frac{v_i}{2}$; $v_B = v_D$ and $v_0 = 0$.

When the load is applied to the cantilever beam, the resistance R_{g1} and R_{g4} *increase* due to *tensile load* whilst resistances R_{g2} and R_{g3} *decrease* due to equal *compressive strain*. When strained the resistances of the various gauges are

$$R_{g1} = R_{g4} = R + dR \text{ (tension)}$$

$$R_{g2} = R_{g3} = R - dR \text{ (compression)}$$

and, $$v_{AB} = \frac{R_{g1}}{R_{g1} + R_{g3}} \cdot v_i$$

or, $$v_{AB} = \frac{R + dR}{(R + dR) + (R - dR)} v_i = \frac{R + dR}{2R} \cdot v_i \quad ...(3.34)$$

and, $$v_{AD} = \frac{R_{g2}}{R_{g2}+R_{g4}} v_i$$

or, $$v_{AD} = \frac{R-dR}{(R-dR)+(R+dR)} \cdot v_i = \frac{R-dR}{2R} \cdot v_i \quad ...(3.35)$$

The changed output voltage,

$$v_0 + dv_0 = \frac{R+dR}{2R} \cdot v_i - \frac{R-dR}{2R} \cdot v_i$$

$$= v_i\left(\frac{R+dR}{2R} - \frac{R-dR}{2R}\right) = v_i \frac{dR}{R}$$

Since the output voltage under unstrained conditions, $v_0 = 0$, therefore, change in output voltage due to applied strain becomes,

$$dv_0 = v_i \frac{dR}{R}$$

i.e., $$dv_0 = G_f \, e.v_i \quad ...(3.36)$$

which is the *four times* the output of Wheatstone bridge using *one gauge only*.

● It may be noted that all the relations derived above are subject to the following conditions : (*i*) the values of the resistances of all the four limbs of the bridge are initially equal, and (*ii*) the galvanometer has infinite impedance and no current flows through it.

Important points – *worth noting* :

1. If there are more than one strain gauge active, the output of the bridge and hence the sensitivity of the system increases. In general, if there are n active strain gauges in the bridges, then the output voltage is given by

$$dv_0 = n \frac{dR}{R} \cdot v_i$$

or, $$dv_0 = n \cdot \left(\frac{G_f}{4}\right) e \cdot v_i \quad ...(3.37)$$

(G_f and e are gauge factor and strain respectively).

The increased bridge output is expressed in terms of "*bridge constant*" (it represents the ratio of the actual bridge output to that if only one gauge were effective). The bridge constants for the three arrangements discussed above are 1, 2 and 4 respectively.

2. *High gauge sensitivity* can be obtained with :

(*i*) *High gauge factor*: It depends upon :

— The gauge material;

— The configuration of the gauge wire;

— The mechanical loading.

● In general the *foil* and *wire gauges* have gauge factor of about 2 and semiconductor gauges have typical values of about – 100 to + 200 (approx).

(*ii*) *Large excitation voltage*. It depends upon current or power rating of the gauge; typical values being 15 mA and 15 mW respectively.

3.15 LOAD CELLS

Load cells *are elastic devices that can be used for measurement of force through indirect methods i.e., through use of secondary transducers.*

Load cells utilize an elastic member as the primary transducer and strain gauges as secondary transducer. When the combination of the *strain gauge–elastic member* is used for weighing, it is called a *"load cell"*.

While designing load cells using strain gauges the following *factors* should be considered :

(*i*) Stiffness of the elastic element.

(*ii*) Optimum positioning of gauges on the element.

(*iii*) Provision for compensation of the temperature.

When large loads are to be measured, the direct tensile-compressive member may be used, whereas, in case of small loads, strain amplification provided by bending may be used with advantage.

3.15.1. Hydraulic Load Cell

Fig. 3.41 shows a *hydraulic load cell.*

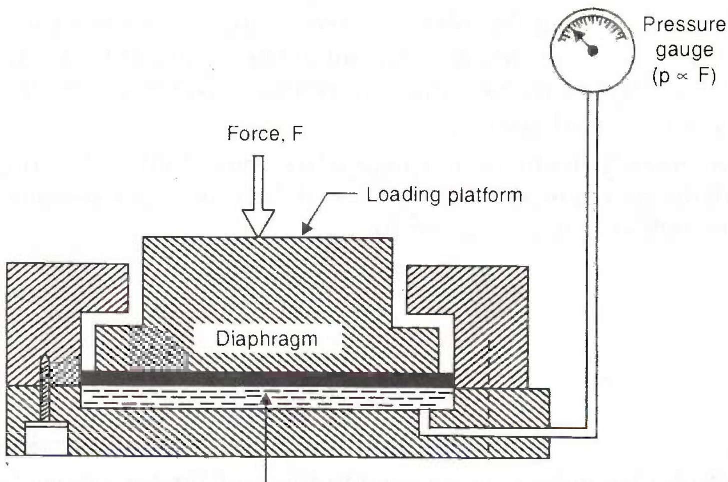


Fig. 3.41. Hydraulic load cell.

Here the force variable is impressed upon a diaphragm which deflects and thereby transmits the force to a liquid. The liquid medium contained in a confined space, has a preload pressure of 2 bar. On the application of the force the liquid pressure increases and *equals the force magnitude divided by the effective area of the diaphragm.* The pressure is transmitted to and read on an accurate pressure gauge calibrated directly on force units.

- These cells have been used to measure loads upto about 25 MN (with an *accuracy of 0.1% of full scale); resolution is about 0.02 per cent.*

3.15.2. Pneumatic Load Cell

This cell operates on the *force-balance principle.* It employs a nozzle-flapper transducer similar to the conventional relay system. For any constant applied force, the system attains equilibrium at a specific nozzle opening and *corresponding pressure is indicated by the height of mercury column in a manometer.*

- The commercially available load cells (operating on this principle) can measure loads upto 25 kN with an accuracy of 0.5% of full scale.

3.15.3. Strain-Gauge Load Cells

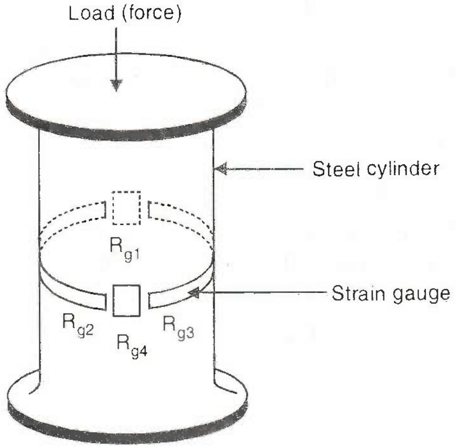
These cells convert weight or force into electrical outputs which are provided by the strain gauges; these outputs can be connected to various measuring instruments for indicating, recording and controlling the weight or force.

Usually the *strain gauges are directly applied to the force-developing device and the device is calibrated against strain-gauge output.*

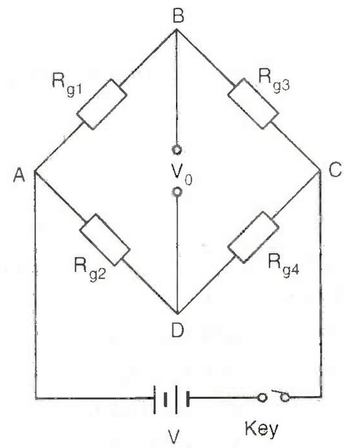
- These are excellent force-measuring devices, particularly for *transient and non-steady forces.*
- These are used in conjunction with CRO (for display purposes) for measurement of *rapidly changing loads.*

Construction and working of the load cell :

Fig. 3.42 shows a simple strain gauge load cell. It consists of a steel cylinder, on which are mounted four identical strain gauges. The gauges R_{g1} and R_{g4} are along the direction of applied load and the gauges R_{g2} and R_{g3} are attached circumferentially to gauges R_{g1} and R_{g4}. All the four gauges are connected electrically to the four limbs of a Wheatstone bridge circuit.



(a) Load cell



(b) Wheatstone bridge circuit

Fig. 3.42. Strain gauge load cell.

When there is *no load on the cell*, all the four gauges have the same resistance (*i.e.*, $R_{g1} = R_{g2} = R_{g3} = R_{g4}$). Obviously the terminals B and D are at the same potential, the bridge is balanced and the output voltage is zero.

i.e., $$v_{AB} = v_{AD} = \frac{v}{2} \quad ...(3.38)$$

and, $$v_{AB} - v_{AD} = v_0 = 0 \quad ...(3.39)$$

On the application of a compressive load to the unit, the *vertical gauges* (R_{g1} and R_{g4}) undergo compression (*i.e.*, negative strain) and, therefore, there is *decrease in resistance*. The *circumferential gauges* R_{g2} and R_{g3}, simultaneously, undergo tension (*i.e.*, positive strain) leading to *increase in resistance*. The two strains are not equal; these are related to each other by a factor, μ, the Poisson's ratio. Thus when strained, the resistances of various gauges are :

$$R_{g1} = R_{g4} = R - dR \quad \text{...(compression)}$$
$$R_{g2} = R_{g3} = R + dR \quad \text{...(tension)}$$

Potential at terminal B, $v_{AB} = \dfrac{R_{g1}}{R_{g1} + R_{g3}} v = \dfrac{R - dR}{(R - dR) + (R + \mu.dR)} \times v$

$$= \frac{R - dR}{2R - dR(1-\mu)} \times v \quad \text{...}(i)$$

Potential at terminal D, $v_{AD} = \dfrac{R_{g2}}{R_{g2} + R_{g4}} v = \dfrac{R + \mu.dR}{(R + \mu.dR) + (R - dR)} \times v$

$$= \frac{R + \mu.dR}{2R - dR(1-\mu)} \times v \quad \text{...}(ii)$$

The changed output voltage,

$$v_0 + dv_0 = \frac{R - dR}{2R - dR(1-\mu)} \times v - \frac{R + \mu.dR}{2R - dR(1-\mu)} \times v$$

...[Using (*i*) and (*ii*)]

$$= \frac{dR(1+\mu)}{2R} = 2(1+\mu)\left(\frac{dR}{R} \cdot \frac{v}{4}\right) \quad \text{...(3.40)}$$

...*in magnitude*

Since the output voltage $v_0 = 0$ under *unloaded* conditions, therefore, change in output voltage due to applied load becomes :

$$dv_0 = 2(1+\mu)\left(\frac{dR}{R} \cdot \frac{v}{4}\right) \quad \text{...(3.41)}$$

Obviously, this *voltage is a measure of the applied load.*

The use of four identical strain gauges in each arm of the bridge *provides full temperature compensation* and also *increases the sensitivity of the bridge 2 (1 + μ) times.*

Uses : The strain gauge load cells find extensive *use* in the following :

(*i*) Road vehicle weighing devices.

(*ii*) Draw bar and tool-force dynamometers.

(*iii*) Crane load monitoring etc.

Example 3.16. *The following data relate to strain gauge load cell arranged with four identical strain gauges as shown in Fig. 3.42.*

Diameter of the steel cylinder = 60 mm; Nominal resistance of each gauge = 120Ω; Gauge factor = 2.0; Supply voltage (v) = 6V; Modulus of elasticity for steel = 200 GN/m²; Poisson's ratio = 0.3.

Calculate the sensitivity of the load cell.

Solution. *Given* : d = 60 mm = 0.06 m; R_g (each gauge) = 120 Ω; G_f = 2.0, v = 6V; E = 200 GN/m²; μ = 0.3.

Sensitivity of the load cell :

Consider a load of 1 kN applied to the load cell.

$$\text{Stress } (\sigma) = \frac{\text{Load}}{\text{Cross-sectional area}} = \frac{1 \times 10^3}{\frac{\pi}{4} \times (0.06)^2} = 0.3537 \times 10^6 \text{N/m}^2$$

$$\text{Strain, } e = \frac{\text{Stress }(\sigma)}{\text{Modulus of elasticity (E)}} = \frac{0.3537 \times 10^{6}}{200 \times 10^{9}} = 1.7685 \times 10^{-6}$$

Fraction change in resistance,

$$\frac{dR}{R} = G_f \times e = 2.0 \times 1.7685 \times 10^{-6} = 3.537 \times 10^{-6}$$

$$\text{Output voltage, } dv_0 = 2(1+\mu)\left(\frac{dR}{R}\cdot\frac{v}{4}\right) \qquad \text{...[Eqn. (3.41)]}$$

$$= 2(1+0.3)\left(3.537 \times 10^{-6} \times \frac{6}{4}\right) = 13.79 \times 10^{-6}\text{V} = 13.79\mu\text{V}$$

Hence, the *sensitivity of the load cell* = **13.79 μV/kN** **(Ans.)**

3.16 PROXIMITY SENSORS

A ***proximity sensor*** consists of an *element that changes either its state or an analog signal when it is close to, but often not actually touching, an object.*

Magnetic, electrical capacitance, inductance, and eddy current methods are particularly suited to the design of a proximity sensor.

- A ***photoemitter-detector pairs*** represents another approach, where *interruption or reflection of a beam of light in used to detect an object in a non-contact manner.* The emitter and detecter are usually a phototransistor and a photodiode.

Common ***applications*** for proximity sensors and limit switches include :

(*i*) Counting moving objects;

(*ii*) Limiting the traverse of a mechanism.

3.16.1. Eddy Current Proximity Sensors

Working principle :

When a coil is supplied with an alternating current an alternating magnetic field is produced. If there is a metal object in close proximity to this attending magnetic field, then eddy currents are induced in it. The eddy currents themselves produce a magnetic field which *distorts* the magnetic field responsible for their production. Consequently, the *impedance of the coil changes and so the amplitude of the alternating current. This change, at some preset level, can be used to trigger a switch.*

Fig. 3.43. shows the basic form of an eddy current proximity sensor. It is *used for the detection of non-magnetic but conductive materials.*

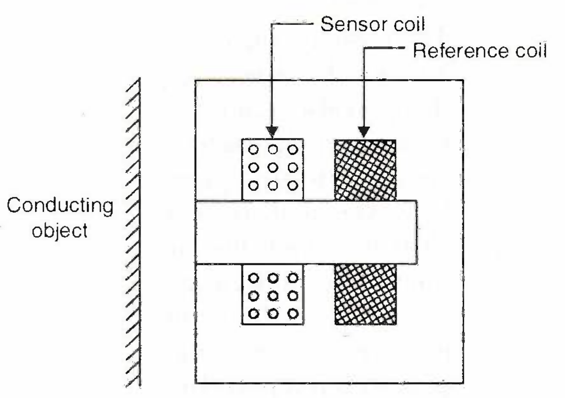


Fig. 3.43. Eddy current proximity sensor.

Advantages :

(*i*) Small in size.

(*ii*) Relatively inexpensive.

(*iii*) High flexibility.

(*iv*) High sensitivity to small displacements.

3.16.2. Capacitive Proximity Sensor

Fig. 3.44. shows a schematic diagram of a capacitance proximity sensor.

— It consists of a simple plate (one of the forms), with the object (*earthed*) acting as the other plate.

— As the object approaches the sensor, separation between the plate of the capacitor and object changes which becomes significant as the object is close to the sensor.

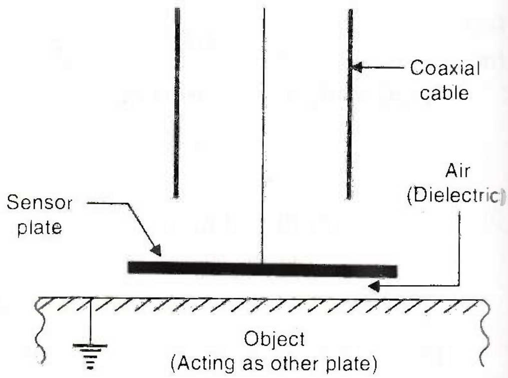


Fig. 3.44. Capacitance proximity sensor.

3.16.3. Inductive Proximity Switch

— An inductive proximity switch consists of a coil wound round a core.

— When the end of the coil is close to a metal object its *inductance changes*. This change can be monitored by its effect on a resonant circuit and the change used to trigger a switch.

- It can only be used *for the detection of metal objects* and is best with ferrous metals.

3.17 PREUMATIC SENSORS

These sensors involve the use of compressed air, displacement or proximity of an object being transformed into a *change in air pressure*.

Fig. 3.45. shown the basic form of a preumatic sensor.

— Low pressure air is allowed to escape through a port in front of the sensor.

— This escaping air, in the absence of any close by object, escapes and in doing so also *reduces the pressure* in the nearby sensor output port. However, if there is a close by object, the air cannot so readily escape and result is that the *pressure increases* in the sensor output port. The output pressure from the sensor thus depends on the proximity of objects.

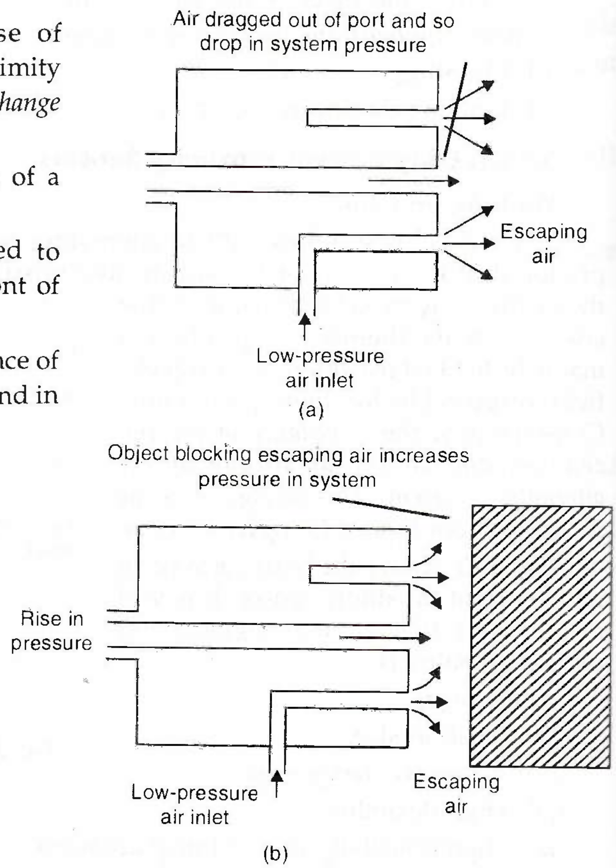


Fig. 3.45. Preumatic proximity sensor.

- Pneumatic sensors are used for the *measurement of the displacements of fractions of millimeters in ranges which typically are about 3 to 12 mm.*

3.18 LIGHT SENSORS

1. **Photodiodes :**

   "Photodiodes" are semiconductor junction diodes which are connected into a circuit in reverse bias, so giving a very high resistance, so that when light, falls on the junction the diode resistance drops and the current in the circuit rises appreciably

   — A photodiode can be used as a variable resistance device controlled by the light incident on it.

   — These diodes have a *very fast response to light.*

2. **Phototransistors :**

   The phototransistors have a light-sensitive collector-base P-N junction. When there is no incident light there is a very small collector-to-emitter current. When light is incident, a base current is produced that is *directly proportional to the light intensity*. This leads to the production of a collector current which is then a *measure of the light intensity.*

   — Phototransistors are often available as integrated packages with the phototransistor connected in a Darlington arrangement with a conventional transistor (Fig. 3.46). Since this arrangement gives a higher current gain, the device gives a *much greater collector current for a given light intensity.*

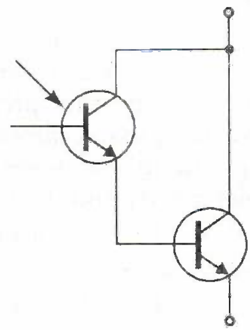


Fig. 3.46. Photo Darlington.

3. **Photoresistor :**

   It has a *resistance* which depends on the intensity of the light falling on it, *decreasing linearly as the intensity increases.*

   — The cadmium sulphide photoresistor is most responsive to light having wavelengths shorter than about 515 nm and the cadmium selinide photoresistor for wavelengths less than about 700 nm.

- An array of light sensors is often required in a small space in order to determine the variations of light intensity across that space.

3.19 DIGITAL OPTICAL ENCODER

A ***digital optical encoder*** *is a device that converts motion into a sequence of digital pulses.* By counting a single bit or decoding a set of bits, the pulses can be converted to relative or absolute position measurements.

- Encoders have both linear and rotary configurations, but the most common type is rotary.
- Rotary encoders are manufactured in two basic forms :
  - (*i*) *Absolute encoder* – Here a unique digital word corresponds to each rotational position of the shaft.
  - (*ii*) *Incremental encoder* – Here digital pulses are produced as the shaft rotates, allowing measurement of relative displacement of the shaft.

Most rotary encoders are composed of a glass or plastic code disc with a photographically deposited radial pattern organised in tracks. As radial lines in each track interrupt the beam between a photoemitter – detector pair, digital pulses are produced. The optical disc of *absolute encoder* is designed to produce a digital word that distinguishes N distinct positions of the shaft. The *incremental encoder*, sometimes called a *relative encodes*, is simpler in design than the absolute encoder.

— ***Incremental encodes** provide more resolution at lower cost than absolute encoders, but they measure only relative motion and do not provide absolute position directly.* However, an incremental encoder can be used in conjunction with a limit switch to define absolute position relative to a reference position defined by the switch.

— ***Absolute encoders*** are chosen in applications where *establishing a reference position is impractical or undesirable.*

3.20 RECENT TRENDS-SMART PRESSURE TRANSMITTERS

The microprocessors are now being used in *transmitters* also; as a consequence of the availability of computing power the transmitters have become more intelligent.

The output in case of smart transmitters is 4–20 mA on 2-wire but with the added capability of digital communication from a hand-held interface connected anywhere on 4–20 mA signal, the remote adjustment of the transmitter data base and acquisition diagnostic information to minimise loop downtime is possible. It has *high rangeability and much better performance.*

The transmitter senses all the *three parameters. 'differential pressure', 'static pressure'* and *temperature.* The meter body is pre-programmed in manufacturing to characterise the unit for linearity, static pressure and temperature effects, and it *computes a highly repeatable and accurate pressure measurement.* These characteristics are held in PROM memory and being specific to one meter are kept with the meter body. The combination of characterised meter body and digital electronics has enabled a quantum leap forward in performance.

- The *rangeability* to smart transmitters is very high (400 : 1). Thus only 3 sensors would be required to cover the entire range of 2.5 millibar to 700 bar.
- The *reliability* is very high due to use of minimum number of components and protection against all foreseeable damping like radio frequency, reverse polarity, overpressure, surge voltage and lightning.

Advantages of digital transmitters :

The major advantages of digital (so called "*smart*") transmitters over their conventional analog counterparts are :

(*i*) Increased rangeability (400 : 1 against 6 : 1 of analog transmitters).
(*ii*) Higher accuracy.
(*iii*) Self-diagnostic facilities.
(*iv*) Almost no drift with time.
(*v*) Reduced cabling cost due to the use of a field bus cuts.
(*vi*) Better noise immunity.
(*vii*) Economical, because of improved overall performance.
(*viii*) Ambient temperature compensation.
(*ix*) Remote adjustability of range, damping, polarity etc. (This makes the commissioning of the entire system simpler).

3.21 SELECTION OF SENSORS

A number of factors need to be considered for selecting of a sensor for a particular application are :

1. The nature of the measurement required *e.g.*,
   — The variable to be measured, its nomial value, the range of values;
   — The accuracy required;
   — The required speed of measurement;
   — The reliability required;
   — 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 conditioning requirements in order to give suitable output signals from the measurement.

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

3.22 STATIC AND DYNAMIC CHARACTERISTICS OF TRANSDUCERS/ MEASUREMENT SYSTEMS–INSTRUMENTS

3.22.1. Introduction

• The ***static characteristics*** *pertain to a system where the quantities to be measured are constant or vary slowly with time. Performance criteria based on dynamic relations (involving rapidly varying quantities) constitute* ***dynamic characteristics***.

— The static characteristics, in a real sense, also influence the quantity of measurement under dyanmic conditions, but these characteristics (static) *show up as non-linear or statistical effects* in otherwise linear differential equations giving the *dynamic characteristics. These effects would make the differential equations analytically unmanageable and so the conventional approach is to treat the two aspects of the problem separately.* Thus, even though these effects influence the dynamic behaviour, the differential equations of dynamic performance generally *neglect* the effects of dry friction, backlash, hysteresis statistical scatter etc.

• The *overall performance of an instrument is judged by a semiquantitative superimposition of the static and dynamic characteristics.*

3.22.2. Performance Terminology

Some important terms used in connection with transducers/measurement systems–instruments are discussed below :

1. **True or actual value.** *The actual magnitude of a signal input to a measuring system which can only be approached and never evaluated is termed as* ***true or actual value***.
2. **Indicated value.** *It is the magnitude of a variable indicated by a measuring instrument.*
3. **Correction.** *The revision applied to the critical value os that the final result obtained improves the worth of the result is called* ***correction***.
4. **Overall error.** *It is the difference of the scale reading and the true value.*
   — When the instrument is properly designed and correctly adjusted the consistent bias in error is very rare.
5. **Range.** *The region between the limits within which an instrument is designed to operate for measuring, indicating or recording a physical quantity is called the* **range** *of the instrument.*
6. **Sensitivity**. *The ratio of output response to a specified change in the input is called* ***sensitivity***.

- The minimum change in the measured variable which produces an effective response of the instrument is called *"Resolution sensitivity"*. It is also called *"discrimination"*.
- The lowest level of measured variable which produces effective response of the instrument is called *"Threshold sensitivity"*.

7. **Scale sensitivity.** *It is defined as the ratio of a change in scale reading to the corresponding change in pointer deflection.*
8. **Scale readability**. The scale readability (in analog instruments) *indicates the closeness with which the scale can be read.*
9. **Repeatability.** It is defined as the *variation of scale reading;* it is random in nature.
   - *It is a measure of closeness with which a given input can be measured over and over again.*
10. **Accuracy.** It may be defined as *conformity with or closeness to an accepted standard value (true value).*
    - Accuracy of an instrument is influenced by factors like static error, dynamic error, reproducibility, dead zone.
11. **Uncertainty**. Uncertainty denotes the *range of error, i.e.,* the region in which one guesses the error to be.
12. **Precision**. It refers to the *degree of agreement* within a group measurements.
    - It is usually expressed in *terms of the deviation in measurement.*
13. **Drift.** *An undesired gradual departure of the instrument output over a period of time that is unrelated to changes in input, operating conditions or lead is called* ***drift****.*
14. **Linearity or non-linearity.** *Deviation of transducer output curve from a specified straight line.* The *"non-linearity"* may be : (*i*) *Terminal linearity* (deviation from a straight line through the end points; (*ii*) *Best-fit linearity* (deviation from the straight line which gives minimum errors, both plus and minus).
15. **Dead zone.** *It is the range within which variable can vary without being detected.*
16. **Dead time.** *It is the time before the instrument begins to respond after the measured quantity has been changed.*
17. **Speed of response.** *The quickness of an instrument to read the measured variable is called* ***speed of response****.*
18. **Reproducibility.** *The degree of closeness with which the same value of a variable may be measured at different times is called* ***reproducibility****.*
19. **Tolerance.** *It is the range of inaccuracy which can be tolerated in measurements.*
20. **Backlash.** It is defined as the *maximum distance or angle through which any part of a mechanical system may be moved in one direction without applying appreciable force or motion to the next part in a mechanical system.*
21. **Stiction (static friction).** *It is the force or torque that is necessary just to initiate motion from rest.*
22. **Noise.** It may be defined *extraneous disturbance generated in a measuring system which convey no meaningful information w.r.t. desired signal.*

3.22.3. Static Characteristics

Measurements of applications in which parameter of interest is more or less constant; or varies very slowly with time are called **static measurements**. *A set of criteria (e.g., "accuracy", "error",* *"reproducibility", "drift", "sensitivity", "dead zone") that provide meaningful description of measurements under static conditions are called* **static characteristics.**

The main *static characteristics* may be summed up as follows :

(*i*) Accuracy
(*ii*) Sensitivity
(*iii*) Reproducibility
(*iv*) Drift
(*v*) Static error
(*vi*) Dead zone.

Range and span :

Range. *The difference between the largest and the smallest reading of the transducer/instrument is called the* ***Range*** *of an instrument.* The range is expressed by stating the lower and upper values.

Span *represents the algebraic difference between the upper and lower range values of the transducer/instrument.*

If the highest point of calibration is S_{max} units while the lowest is S_{min} units and that the calibration is continuous between the points, then we say that the *instrument* ***range*** *is between* S_{min} *and* S_{max}.

The *instrument* ***span*** is given by, $S_{max} - S_{min}$.

The above definitions apply both to analog as well as digital instruments.

Examples : (*i*) *Range* : 2 kN/m^2 to 50 kN/m^2;
Span : $50 - 2 = 48$ kN/m^2
(*ii*) *Range* : –5°C to 90°C;
Span : $90 - (-5) = 95$°C.

Repeatability and reproducibility :

Although the meaning of the terms *repeatability* and *reproducibility* is same but they are applied in different contexts.

Repeatability pertains to the *closeness of output readings when the same input is applied repetitively over a short period of time with the same measurement conditions, same instrument and observer, same location and same conditions of use maintained throughout.*

Reproducibility relates to the *closeness of output readings for the same input when there are changes in the method of measurement, observer, measuring instrument, location, conditions of use and time of measurement.*

Sensitivity :

The ratio of the magnitude of output signal to the input signal or response of measuring system to the quantity being measured is called **sensitivity**.

It is represented by the slope of the calibration curve if the ordinates are expressed in actual units.

Hysteresis :

The maximum differences in output at any measured value within the specified range when approaching the point first with increasing and then with decreasing input may be termed as **hysteresis.**

- It is a phenomenon which shows different output effects when loading and unloading. It is *non-coincidence of loading and unloading curves.*

Fig. 3.47 (*a*) shows output and input curves (loading and unloading) for an instrument which has no friction due to sliding parts. The non-coincidence of loading and unloading curves is on account of *internal friction or hystereses damping.*

Fig. 3.47 (*b*) shows the input-output relations of instruments which do not have internal friction but have external sliding friction.

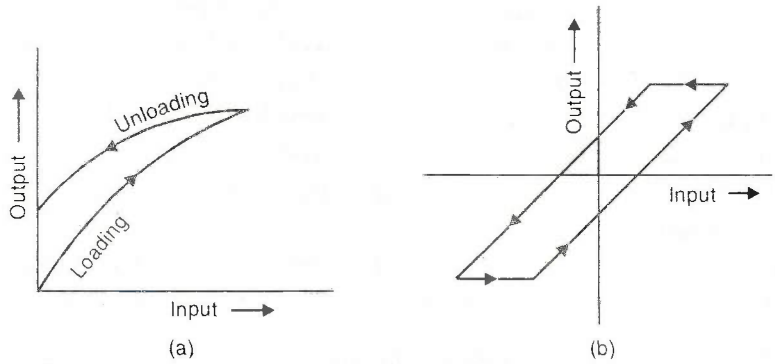


Fig. 3.47. *Hysteresis effects.*

- The numerical value of hysteresis can be specified in terms of either output or input and is usually given as *%age of full scale.*
- Hysteresis results from the *presence of irreversible phenomenon such as :*
  — Mechanical friction;
  — Slack motion in bearings;
  — Magnetic and thermal effects.

Dead band/time :

- The *dead band* or *dead space* of a transducer is the range of input values for which there is no output.
- The *dead time* is the length of time from the application of an input until the output begins to respond and change.

Resolution or Discrimination :

When the input is slowly increased from some arbitrary (non-zero) input value, it is observed that the output does not change at all until a certain increment is exceeded; this increment is called **Resolution or discrimination** of the instrument. Thus *resolution defines the smallest change of input for which there will be a change of output.*

— In case of *analog instruments,* the resolution is determined by the observer's ability to judge the position of a pointer on a scale. Resolution is usually reckoned to be no better than ±0.2 of the smallest division of the scale.

— In case of digital instruments, resolution is determined by the number of neon tubes taken to show the measured value.

- **Threshold** defines the *smallest measurable input* while the **resolution** defines the *smallest measurable input change.*
- *"Threshold"* and *"resolution"* may be *expressed as an actual value or as a fraction or percentage of full scale value.*

3.22.4. Dynamic Responses/Analysis of Measurement Systems

The dynamic behaviour of measurement systems is studied in the following two domains :

1. Time domain analysis.
2. Frequency domain analysis.

1. **Time domain analysis :**

In this the input signal is applied to the measurement system and the behaviour of the system is studied as a *function of time.* The dynamic response of the system to different types of inputs, which are a function of time is analysed at different intervals of time after the application of the input signals. In most cases, the actual input signals vary in random fashion with respect to time and therefore cannot be mathematically defined. Consequently the performance of a system can be analysed (in the time domain analysis) by using the following standard *test signals/inputs* :

(*i*) Step input;
(*ii*) Ramp input;
(*iii*) Parabolic input;
(*iv*) Impulse input.

2. **Frequency domain analysis :**

This type of analysis of a system *pertains to the steady state response of the system to a sinusoidal input*. Here, the system is subjected to a sinusoidal input and the system response is studied with *frequency as the independent variable.*

- *Frequency response.* It is the maximum frequency of the measured variable that an instrument is capable of following without error. The usual requirement is that the frequency of measurand should not exceed 60 per cent of the natural frequency of the measuring instrument.

Standard test signals/inputs :

The most common standard inputs used for dynamic analysis are discussed below :

1. ***Step function*** :

Refer to Fig. 3.68 (*a*). It is a *sudden change from one steady value to another.*

It is mathematically represented by the relationship :

$$x = 0 \text{ at } t < 0$$
$$x = x_c \text{ at } t \geq 0$$

where x_c is a constant value of the input signal x_i.

- The "*transient response*" *indicates the capacity of the system to cope with changes in the input signal.*

2. ***Ramp or linear function*** :

In this case (see Fig. 3.48 (*b*)) the *input varies linearly with time.*

This input is mathematically represented as :

$$x = 0 \text{ at } t < 0$$
$$x = \psi \text{ at } t \geq 0$$

where ψ is the slope of the input versus time relationship.

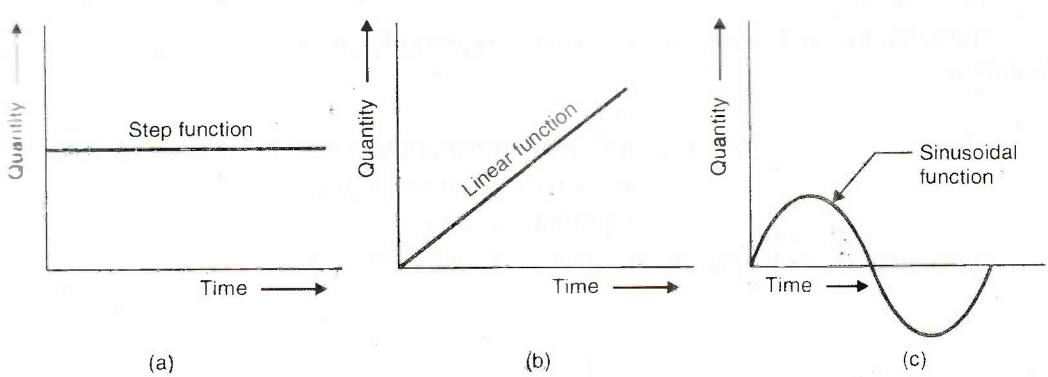


Fig. 3.48. Standard input function.

- *The ramp-response becomes indicative of the steady state error in following the changes in the input signal.*

3. ***Sinusoidal function*** :

In this case (see Fig. 3.68 (*c*)) the *input varies sinusoidally with a constant maximum amplitude.*

It is represented mathematically as follows :

$$I_i = A \sin \omega t,$$

where, A = Amplitude, and

ω = Frequency in rad/s.

- *The frequency or harmonic response is a measure of the capability of the system to respond to inputs of cylic nature.*

A general measurement system can be mathematically described by the following differential equation :

$$(A_n D^n + A_{n-1} D^{n-1} + ... + A_1 D + A_0)\, I_0 = (B_m D^m + B_{m-1} D^{m-1} + ... + B_1 D + B_0) I_i \qquad ...(3.42)$$

where, A's and B's = Constants, depending upon the physical parameters of the system,

D^k = Operative derivative of the order k,

I_0 = The information out of the measurement system, and

I_i = The input information.

The order of the measurement system is generally classified by the value of the power of n.

- *Zero-order system* : $n = 0$ and A_1, A_2, A_3 ..... $A_n = 0$
- *First-order system* : $n = 1$ and A_2, A_3, A_4 ..... $A_n = 0$
- *Second-order system* : $n = 2$ and A_3, A_4, A_5 ..... $A_n = 0$

The above method of classification is used for most of the instruments and systems.

Although general equation can be solved by various methods, we shall be using method of D-operator for getting its solution.

3.22.4.1. Zero, First and Second Order Systems :

1. **Zero order systems :**

Fig. 3.49 shows the block diagram of a 'Zero-order system'. In this case the output of the measuring system (ideal) is directly proportional to input, no matter how the input varies. The *output is faithful reproduction of input without any distortion or time lag.*

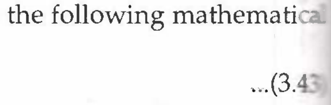


Fig. 3.49. Block diagram for zero-order system.

The behaviour of the zero-order system is represented by the following mathematical solution.

$$I_0 = SI_i \qquad ...(3.43)$$

where, I_0 = Information out of the measuring system,

S = Sensitivity of the system, and

I_i = Input information.

This equation is obtained by putting $n = 0$ in the general equation (3.42)

i.e., $$A_0 I_0 = B_0 I_i$$

or, $$I_0 = \frac{B_0}{A_0} I_i = SI_i \qquad ...(3.44)$$

The zero-order system is characterised only by the static sensitivity (parameter), the value of which is obtained through the process of static calibration.

Examples of zero-order system :

- Mechanical levers;
- Amplifiers;
- Potentiometer (It gives an output voltage which is proportional to wiper's displacement) etc.

2. **First-order systems :**

Fig. 3.50 shows the block diagram of a 'First-order system'.

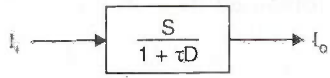


Fig. 3.50. Block diagram for first-order system.

The behaviour of a first-order system is given by following first-order differential equation :

$$A_1 \frac{dI_0}{dt} + A_0 I_0 = B_0 I_i \qquad ...(3.45)$$

(This equation is obtained by inserting $n = 1$ in the general equation).

Eqn. (3.45) may be written in standard form as follows : ...(3.46)

$$\frac{A_1}{A_0}\frac{dI_0}{dt} + I_0 = \frac{B_0}{A_0} I_i$$

or, $$\tau \frac{dI_0}{dt} + I_0 = SI_i \qquad ...(3.47)$$

where, $\tau = \frac{A_1}{A_0} =$ Time constant, and

$S = \frac{B_0}{A_0} =$ Sensitivity.

Using D-operator, we get

$$\left[\text{where}, D = \frac{d}{dt}, \text{and } D^2 = \frac{d^2}{dt^2}\right]$$

$$\tau D I_0 + I_0 = SI_i$$

or, $$I_0 (\tau D + 1) = SI_i$$

or, $$\frac{I_0}{I_i} = \frac{S}{1 + \tau D} \qquad ...(3.48)$$

Equation (3.48) gives the standard form of *transfer operator* for first-order system.

Examples of first-order system :

- Velocity of a true falling mass;
- Air pressure build-up in bellows;
- Measurement of temperature by mercury-in-glass thermometers;
- Thermisters and thermocouples;
- Resistance-capacitance network.

3. **Second-order systems :**

Fig. 3.51., shows the block diagram of 'Second-order system' :

$$I_i \longrightarrow \boxed{\frac{S}{1+(2\gamma/\omega_n)D+(1/\omega_n^2)D^2}} \longrightarrow I_o$$

Fig. 3.51. Block diagram for second-order system.

The behaviour of a second-order system is given by the following differential equation (obtained by putting $n = 2$ in the general equation);

$$A_2 \frac{d^2 I_0}{dt^2} + A_1 \frac{dI_0}{dt} + A_0 I_0 = B_0 I_i \qquad ...(3.49)$$

Dividing the above equation by A_0, we have

$$\frac{A_2}{A_0}\frac{d^2 I_0}{dt^2} + \frac{A_1}{A_0}\frac{dI_0}{dt} + I_0 = \frac{B_0}{A_0} I_i \qquad ...(3.49\ a)$$

Let, $\omega_n = \sqrt{\dfrac{A_0}{A_2}}$ = Undamped natural frequency, rad/s.

$\gamma = \dfrac{A_1}{2\sqrt{A_0 A_2}}$ = Damping ratio, dimensionless, and

$S = \dfrac{B_0}{A_0}$ = Static sensitivity or steady-state gain.

Then, by substituting these values in eqn. (3.49 *a*), we get

$$\frac{1}{\omega_n^2}\cdot\frac{d^2 I_0}{dt^2} + \frac{2\gamma}{\omega_n}\cdot\frac{dI_0}{dt} + I_0 = SI_i \qquad ...(3.50)$$

or, in terms of *D*-operator, we have

$$\left(\frac{D^2}{\omega_n^2} + \frac{2\gamma}{\omega_n} D + 1\right) I_0 = SI_i$$

or,

$$\frac{I_0}{I_i} = \frac{S}{\frac{1}{\omega_n^2} D^2 + \frac{2\gamma}{\omega_n} D + 1} \qquad ...(3.51)$$

Examples of second-order system :

- Piezoelectric pick-up ;
- Spring-mass system (used for acceleration and force measurements)
- Pen control system on X-Y plotters;
- U.V. galvanometer, etc.

Damping ratio :

In the design of instruments a term which is very frequently used is the "*damping ratio*" (γ) defined as the *ratio of the actual value of coefficient of viscous friction in movement and the value required to produce critical damping.*

i.e., $\gamma = \dfrac{A_1}{2\sqrt{A_0 A_2}}$

This dimensions term is very useful because to determine its value, it is not necessary that the values of A_1, A_0 and A_2 may be known. In practice it is not easy to determine accurately the values of A_1 and A_2. Further, even if these values are known, they do not in themselves specify whether the instrument is under, over or critically damped, since a numerical calculation has to be performed with them first. Therefore, designers find "*damping ratio*" as a very convenient measure of the amount of the damping present in the movement.

The terms *damping ratio* (γ) and *underdamped natural frequency* (ω_n) immediately conjure up a physical picture of the response of an instrument and both of the quantities are very easy to measure. Thus g and ω_n easily do away with quantities A_2, A_1 and A_0.

3.22.4.2. First-order System Responses :

The complete solution of an equation which describes the dynamical behaviour of a system consists of the following *two parts* :

(*i*) **Complementary function.** *It corresponds to the short time or transient response.*

(*ii*) **Particular integral.** *It refers to the long time steady state response.*

The transfer operator form of the first-order system is given by :

$$\frac{I_0}{I_i} = \frac{S}{1+\tau D}$$

When S (static sensitivity or steady state gain) equals *unity*, we get

$$(1 + \tau D)\, I_0 = I_i \qquad ...(3.52)$$

Now we shall obtain the solution of this equation for different standard inputs (The solutions are not mathematically rigorous, but are practical).

Transient response (complementary function) :

The transient response from the auxiliary equation is obtained by putting input I_i equal to zero;

i.e., $$(1 + \tau D)\, I_{0,\,t} = 1 \qquad ...(3.53)$$

(subscript t refers to the transient value)

Let the solution be of the form :

$$I_{0,\,t} = A\, e^{mt}$$

where, m is an algebraic variable

or, $$(1 + \tau D)\, A\, e^{mt} = 0$$

or, $$Ae^{mt} + \tau.\frac{d}{dt}(Ae^{mt}) = 0$$

or, $$Ae^{mt} + \tau.Ame^{mt} = 0$$

$$Ae^{mt}(1+\tau.m) = 0$$

$\therefore$ $$m = \frac{1}{\tau}$$

Then, $$I_{0,\,t} = A\, e^{mt} = Ae^{-t/\tau} \qquad ...(3.54)$$

The transient response of a first-order system is same for different standard inputs.

Steady state response (Particular integral) :

The steady state response is given by :

$$(1 + \tau D)\, I_{0,\,s} = I_i \qquad ...(3.55)$$

(Subscript s refers to the steady state value)

or,
$$I_{o,s} = (1 + \tau D)^{-1} I_i$$
$$= (1 - \tau D + \text{terms in } D^2 \text{ and higher}) I_i \quad ...(3.56)$$

1. **Step input :**

Since the input I_i is a step of constant magnitude; its differential equals zero, and subsequently, we get

$$I_{0,s} = (1 + \tau D)^{-1} I_i = I_i \quad ...(3.57)$$

Total response = Transient response + steady state response

or,
$$I_o = Ae^{-t/\tau} + SI_i \quad ...(3.58)$$

The constant A is evaluated from the initial conditions as follows :

At $t = 0, \quad I_o = 0$

$\therefore \quad 0 = A + SI_i \quad$ or, $\quad A = -SI_i$

$$\therefore \quad I_o' = \underbrace{-I_i e^{-t/\tau}}_{\text{Transient}} + \underbrace{I_i}_{\text{steady state}}$$

or,
$$I_o = I_i (1 - e^{-t/\tau}) \quad ...(3.59)$$

or,
$$\frac{I_o}{I_i} = (1 - e^{-t/\tau}) \quad ...(3.60)$$

... in *non-dimensional form.*

Salient features (with step input) :

Following are the salient features of first-order system with step input :

(*i*) The transient response of the first-order system is *time dependent* ; as the time passes, grows its value decreases (Refer to eqn. 3.60) and after a very long time the value becomes zero approximately. Thus magnitude of output (I_o) will be same as input (I_i) when the time is very large.

(*ii*) The speed of response relates to the time constant τ. A large τ indicates that response of the system is slow, whereas a small τ represents a fast system response. Thus in order to get good fidelity (*i.e.,* for accurate dynamic measurements) efforts should be made to minimise the value of τ.

(*iii*) Refer to Fig. 3.52, which shows the time response of a first-order system to a step-input when $t = \tau; \frac{I_o}{I_i} = (1 - e^{-1}) = 0.632.$ Thus, the *time constant* (τ), *for a rising exponential function, is defined as the time to reach* 63.2% of its steady state value. The time constant, for a *decaying function* would correspond to the time taken to fall to 36.8% *of its initial value.*

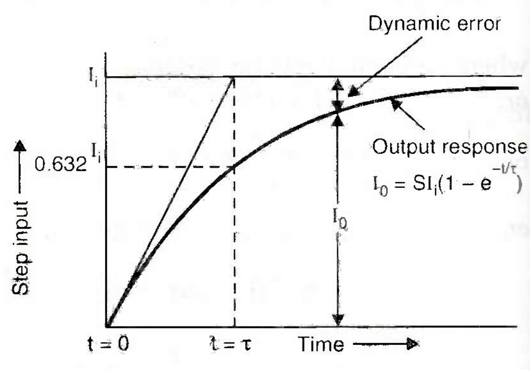


Fig. 3.52. Time response of a first-order system to step input.

(*iv*) Dynamic error (*i.e.,* vertical difference between the input and output response curve).

The value of constant A can be evaluated by applying the initial condition.

At, $t = 0$ $I_0 = 0$

or, $0 = A - \psi\tau \quad \therefore A = \psi\tau$

$$\therefore \quad I_0 = \psi t - \psi\tau + \psi\tau \times e^{-t/\tau} = \psi(t-\tau) + \psi\tau e^{-t/\tau} \quad ...(3.67)$$

or, $$I_o = \psi[t - \tau(1 - e^{-t/\tau})] \quad ...(3.67\ (a))$$

Fig. 3.53 shows the time response of a first-order system to a ramp input.

The dynamic error.

$$E_{dy.} = I_i - I_o$$

$$= \psi t - [\psi t - \psi\tau + \psi\tau e^{-t/\tau}]$$

$$= \underbrace{\psi\tau}_{\text{steady}} - \underbrace{\psi\tau e^{-t/\tau}}_{\text{Transient}} \quad ...(3.68)$$

or, $$\frac{E_{dy.}}{\psi\tau} = 1 - e^{-t/\tau} \quad ...(3.69)$$

...(*in dimensionless form*)

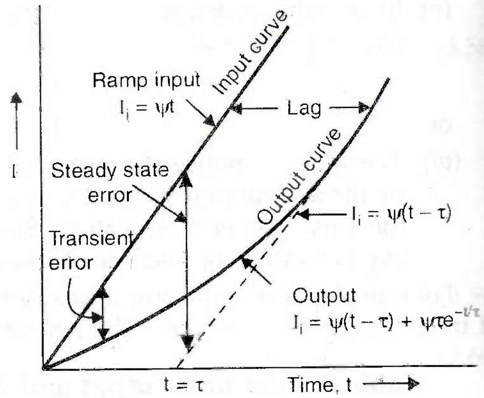


Fig. 3.53. Time response of a first-order system to a ramp input.

Salient features (with ramp input) :

(*i*) The term $\psi\tau$ being independent of time continues to exist and so it is called the steady state error. The term $\psi\tau\, e^{-t/\tau}$ gradually decreases with time and hence is called the *transient error.*

— Since the steady state error is directly proportional to τ (time constant), therefore, the larger the value of τ the larger will be the magnitude of the error.

— When τ is made small the transient error decreases rapidly; this implies, that the system attains the steady state at a faster pace.

(*ii*) The output response curve always lags behind the input curve by a constant amount known as *lag.*

3. **Sinusoidal (Harmonic) input :**

The frequency analysis of a system pertains to the steady state response of the system to a sinusoidal input. In this analysis, the system is subjected to a sinusoidal input and the system response studied with frequency as the independent variable. The sinusoid is a unique input signal and the resulting output signal for a linear system is sinusoidal in the steady state. However, the output signal differs from the input waveform in amplitude and phase.

In order to determine the frequency response of sinusoidal input to a first-order system, let us replace the transfer operator D by a factor $j\omega$ in the input/output relationship; then we get,

$$\frac{I_o}{I_i} = \frac{1}{1 + D\tau} = \frac{1}{1 + j\omega\tau} \quad ...(3.70)$$

where, ω = Input frequency, rad/s, and

$j = \sqrt{(-1)}$

In a frequency response the following two quantities are of interest : Refer to Fig. 3.54.

(*i*) *Amplitude ratio or modulus* $\left(\frac{I_o}{I_i}\right)$. It prescribes the size of the output amplitude relative to the input amplitude.

(*ii*) *Phase shift* of output relative to input.

For the first-order system represented by the equation (3.70),

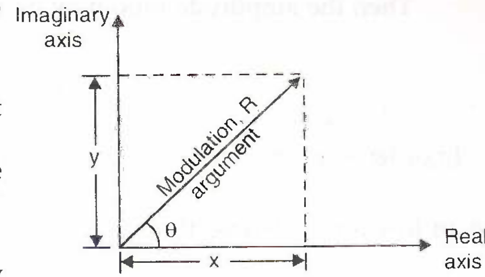


Fig. 3.54.

$$\text{Modulus} = \sqrt{1+(\omega\tau)^2}$$

$$\text{Argument/Phase angle} = \tan^{-1}(\omega\tau) \qquad ...(3.71)$$

Salient features (with sinusoidal input) : Refer to Fig. 3.55.

(*i*) When a system is subjected to a sinusoidal input with frequency ω, its output will also be sinusoidal, but the magnitude of the output amplitude necessarily may not be the same (as the input one). *The ratio of the amplitude* (often called *attenuation*) is given as :

$$\frac{I_o}{I_i} = \frac{1}{\sqrt{1+(\omega\tau)^2}} \qquad ...(3.72)$$

Thus, with the increase in input frequency, the amplitude ratio decreases.

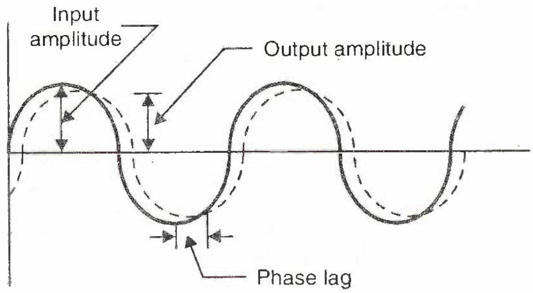


Fig. 3.55. Relationship between an input frequency and corresponding output frequency.

(*ii*) The output from the system may not necessarily be in phase with the input; and the phase difference is given by

$$\phi \text{ (phase angle)} = -\tan^{-1}(\omega\tau) \qquad ...(3.73)$$

–ve indicates that output *lags* behind the input. When $\omega = \frac{1}{\tau}$ the phase lag is $\frac{\pi}{4}$ or 45°.

As the accuracy of an instrument measuring dynamic input depends upon the time constant, therefore, smaller the time constant, greater the accuracy; for phase shift to be small, the time period τ should be small.

(*iii*) When the input and output signals are given by the relations :

$$I_i = A \sin \omega\tau, \text{ and } I_o = B \sin(\omega\tau + \phi) = zA \sin(\omega\tau + \phi),$$

Then the amplitude ratio may be represented as follows :

$$K = \left|\frac{I_o}{I_i}\right| = \frac{z}{\sqrt{1+(\omega\tau)^2}} \qquad ...(3.74)$$

In order to produce amplitude of sinewave without any attenuation ($K = 1$) we must use an instrument whose time constant, $\tau = \frac{\sqrt{z^2-1}}{\omega}$.

3.22.4.3. Second-order System Responses :

In case of typical second-order system having, unit static sensitivity, the homogeneous equation is given by :

$$\frac{I_o}{I_i} = \frac{1}{\left(\frac{1}{\omega_n^2}\right)D^2 + \left(\frac{2\gamma}{\omega_n}\right)D + 1}$$

or,

$$\left[\frac{1}{\omega_n^2}D^2 + \left(\frac{2\gamma}{\omega_n}\right)D + 1\right]I_o = I_i \qquad ...(3.75)$$

(where, γ = damping ratio)

(*a*) **Transient response** (*complimentary function*). It is obtained from the auxiliary equation by replacing D (transfer operator) by an algebraic variable s and putting I_i equal to zero; we get the auxiliary equation as :

$$\frac{1}{\omega_n^2}s^2 + \frac{2\gamma}{\omega_n}s + 1 = 0$$

The roots are

$$s_1, s_2 = \frac{\frac{-2\gamma}{\omega_n} \pm \sqrt{\left(\frac{2\gamma}{\omega_n}\right)^2 - \frac{4}{\omega_n^2}}}{\frac{2}{\omega_n^2}}$$

$$= -\gamma\omega_n \pm \sqrt{\gamma^2\omega_n^2 - \omega_n^2}$$

or,

$$s_1, s_2 = -\gamma\omega_n \pm \omega_n\sqrt{\gamma^2 - 1} \qquad ...(3.76)$$

The *transient solution* has the accepted form,

$$I_{o,\,t} = Ae^{s_1 t} + Be^{s_2 t}$$

where, A and B = Arbitrary constants to be determined from initial conditions, and

s_1, s_2 = Roots of the auxiliary equation (The roots may be real and different, real and equal or imaginary and *that determines the nature of transient response of the system*).

The response of the system is of the following three types depending upon the roots of the characteristic equations :

(*i*) Over-damped systems.

(*ii*) Critically-damped systems.

(*iii*) Under-damped systems.

(*i*) ***Over-damped systems.*** In this case $\gamma > 1$ and the roots are *real and unequal.*

- There is *heavy damping* and the system responds to the final steady-state value without any oscillations but in a *sluggish manner.*
- *No overshoot* in step response and no *"resonance"* (resonance refers to the output signal *greater* in magnitude than the ideal output) in the *frequency response.*
- The overdamped systems, owing to their sluggish response, are usually unsuitable for several control applications.

(*ii*) ***Critically-damped systems.*** In this case $\gamma = 1$ and the roots are *real and equal.*

- The system has a *quick and smooth response to the final steady state value without any oscillations.*
- *No overshoot in the step response and no resonance in the frequency response.*

(*iii*) ***Under-damped systems.*** For an undamped system $\gamma < 1$ and roots of the characteristics equation are a complex conjugate pair; these are given as

$$s_1, s_2 = -\gamma\omega_n \pm j\omega_n\sqrt{(1-\gamma^2)}$$

$$= -\gamma\omega_n \pm j\omega_d \qquad ...(3.77)$$

where, $\omega_d = \omega_n\sqrt{(1-\gamma^2)}$; this quantity is called *"damped natural frequency"* of the system (This is the frequency at which the damped system freely oscillates when disturbed).

- Such systems *take a long time to reach steady state, but have quick initial response.*
- In these systems, there are oscillations in the step response and resonance effects in the frequency response for values of $\gamma < 0.707$.
- Majority of instruments and control systems are generally underdamped (*light damping*).

(*b*) **The steady state response** (*particular integral*). It is given by :

$$\left(\frac{1}{\omega_n^2}D^2 + \frac{2\gamma}{\omega_n}D + 1\right)I_{o,s} = I_i$$

or

$$I_{o,s} = \left(1 + \frac{2\gamma}{\omega_n}D + \frac{1}{\omega_n^2}D^2\right)^{-1} I_i$$

$$= \left(1 - \frac{2\gamma}{\omega_n}D + \text{terms in } D^2 \text{and higher}\right)I_i \qquad ...(3.78)$$

1. **Step input :**

Fig. 3.56. shows the time response of a second-order system to a step input.

Since the input I_i is a step of constant magnitude, its differential equals zero and we have,

$$I_{o,s} = \left(1 - \frac{2\gamma}{\omega_n}D\right)I_i = I_i$$

∴ The complete response

$$I_o = I_{o,s} + I_{o,t}$$

or,
$$I_o = I_i A e^{s_1 t} + B e^{s_2 t} \quad ...(3.79)$$

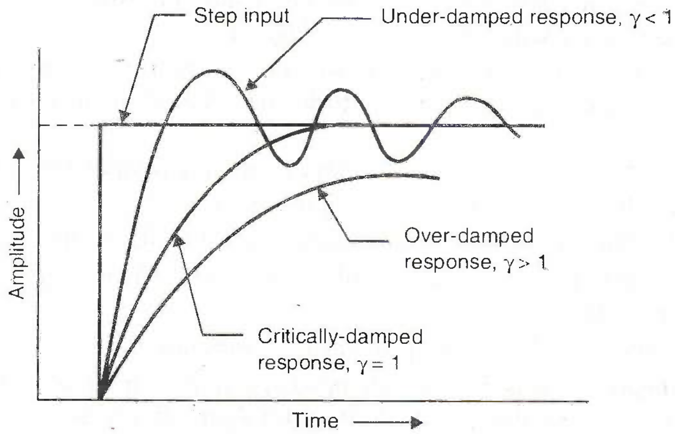


Fig. 3.56. Time response of a second-order system to a step input.

For the under-damped system, the complex conjugate pair of roots are given by

$$s_1, s_2 = -\gamma\omega_n \pm j\omega_d$$

$$\therefore \quad I_o = I_i + Ae^{-(\gamma\omega_n + j\omega_d)t} + Be^{-(\gamma\omega_n - j\omega_d)t} \quad ...(3.80)$$

Replacing the complex exponentials by sines and cosines, we get

$$I_o = I_i + e^{-\gamma\omega_n t}(A\cos\omega_d t + B\sin\omega_d t) \quad ...(3.81)$$

By applying the initial conditions,

At $t = 0$, $\quad I_o = 0$ and $\dfrac{dI_o}{dt} = 0$

we get the values of the constant A and B as :

$$A = -I_i \text{ and}$$

$$B = \frac{-I_i\gamma}{\sqrt{1-\gamma^2}}$$

Inserting these values in equation (3.81), we have

$$I_o = I_i\left[1 - e^{-\gamma\omega_n t}\left\{\cos\omega_d t + \frac{\gamma}{\sqrt{1-\gamma^2}}\sin\omega_d t\right\}\right] \quad ...(3.82)$$

Fig. 3.57. shows the transient response of a second-order system to a unit step function for different values of damping factor γ; *the curves indicate the overshoot and oscillations increase with a reduced damping in the system.*

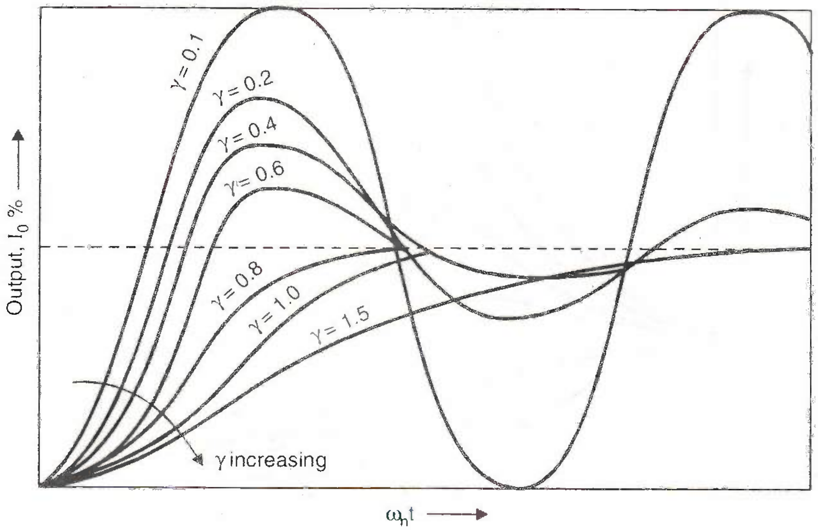


Fig. 3.57. Transient response of a second-order system to a unit step function input for different values of damping factor γ.

2. **Sinusoidal (Harmonic) input :**

When a sinusoidal input is given to the system, its steady state response is determined by replacing *the operator D by jω in the input/output relationship, as given below* :

$$\frac{I_o}{I_i} = \frac{1}{\frac{1}{\omega_n^2}D^2 + \frac{2\gamma}{\omega_n}D + 1} = \frac{\omega_n^2}{D^2 + 2\gamma\omega_n D + \omega_n^2}$$

$$= \frac{\omega_n^2}{(j\omega)^2 + 2\gamma\omega_n(j\omega) + \omega_n^2} = \frac{\omega_n^2}{(\omega_n^2 - \omega^2) + j(2\gamma\omega_n\omega)} \quad ...(3.83)$$

where, ω = Input frequency in rad/s, and $j = \sqrt{-1}$

The denominator is a complex number having :

$$\text{Modulus} = \sqrt{[(\omega_n^2 - \omega^2)^2 + (2\gamma\omega_n\omega)^2]}$$

$$\text{Argument} = \tan^{-1}\left\{\frac{2\gamma\omega_n\omega}{\omega_n^2 - \omega^2}\right\} \quad ...(3.84)$$

∴ The amplitude ratio,

$$\left|\frac{I_o}{I_i}\right| = \frac{\omega_n^2}{\sqrt{[\omega_n^2 - \omega^2)^2 + (2\gamma\omega_n\omega)^2]}} \quad ...(3.85)$$

and, the system phase lag,

$$\phi = \tan^{-1}\left(\frac{2\gamma\omega_n\omega}{\omega_n^2 - \omega^2}\right) \quad ...(3.86)$$

...[using eqn. (3.84)]

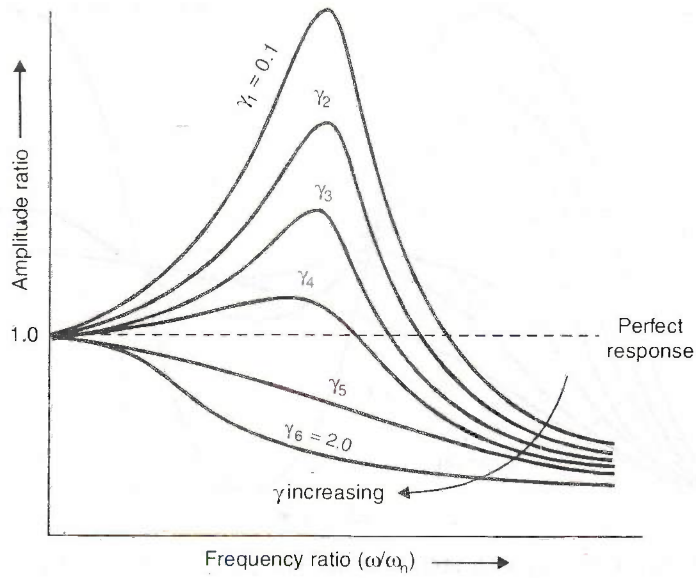


Fig. 3.58. Step response of a second-order system.

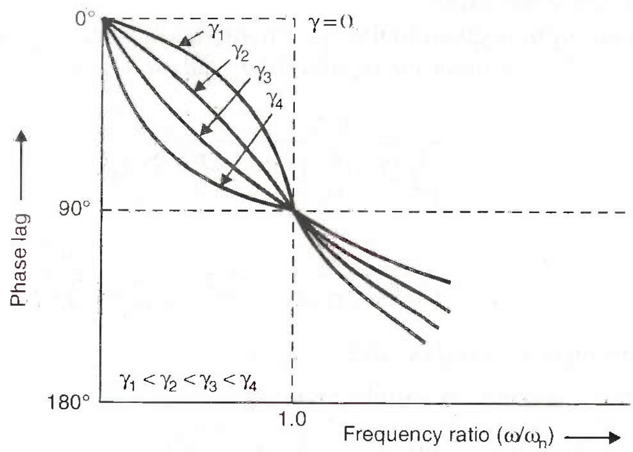


Fig. 3.59. Plot between phase lag and frequency ratio for a second-order system.

Fig. 3.58. and Fig. 3.59. show the variation of amplitude ratio and phase lag verses frequency ratio (ω/ω_n) for various values of damping ratio (γ). From these graphs we observe that the *salient features* of the *steady state response* of a second-order system are :

(*i*) As *frequency ratio* $\to 0$:

Amplitude ratio $\to 1$;

Phase lag $\to 0°$.

(*ii*) As frequency ratio $\to \infty$:

Amplitude ratio $\to 0$;

Phase lag $\to 180°$.

(*iii*) When frequency ratio = 1 :

Amplitude ratio $\to \infty$ in undamped systems ($\gamma = 0$)

Phase lag $\to -\,90°$ in all the systems.

This condition is known as *"resonance"* and can *result in destructive oscillation in lightly damped system.*

(*iv*) *When the amplitude ratio is unity for all frequencies the frequency response is considered to be ideal.* The nearest response to this effect is achieved when the value of γ (damping ratio) lies between 0.6 and 0.7 for both the step and sinusoidal inputs.

WORKED EXAMPLES–FIRST-ORDER SYSTEMS

Example 3.17. *(a) How is the order of the system determined ?*

(b) The following equation characterises the dynamic response of a temperature measuring instrument :

$$\frac{dI_o}{dt} = C\,(I_i - I_o)$$

where, I_o = *Indicated temperature,*

I_i = *Input temperature, and*

C = *A numerical constant.*

(i) Determine the transfer operator form of the equation.

(ii) What is the order of the system ?

Solution. (*b*) Given equation $= \dfrac{dI_o}{dt} = C(I_i - I_o)$

(*i*) **Transfer operator form :**

The given equation can be rewritten as :

$$\frac{1}{C}\cdot\frac{dI_o}{dt} = I_i - I_o$$

or, $$\frac{1}{C}\cdot\frac{dI_o}{dt} + I_o = I_i$$

or, $$I\cdot\frac{dI_o}{dt} + I_o = I_i \qquad \left(\text{where, } \tau = \text{ time constant } = \frac{1}{C}\right)$$

or, $$(\tau D + 1)\,I_o = I_i$$

$\therefore$ The transfer operator form of the equation is given by :

$$\frac{I_o}{I_i} = \frac{1}{(\tau D + 1)} \quad \textbf{(Ans.)}$$

(*ii*) **Order of the system :**

Since the highest differential in the denominator of the transfer operator is *unity*, therefore, the temperature measuring instrument has a *first-order system.* **(Ans.)**

The grouping of the measurement and control systems is done according to the *order of the highest differential in the denominator* of the system transfer operation.

***Examples* :**

- First-order system : $\dfrac{I_o}{I_i} = \dfrac{2}{1 + 3D}$

- Second-order system : $$\frac{I_o}{I_i} = \frac{6}{D^2+3D+4},$$

or, $$\frac{I_o}{I_i} = \frac{14}{(1+0.3D)(1+0.2D)}.$$

Example 3.18. *Formulate the governing equation for a first-order system-temperature measurement by a thermal measuring element (say a thermometer or thermocouple).*

Solution. Refer to Fig. 3.60.

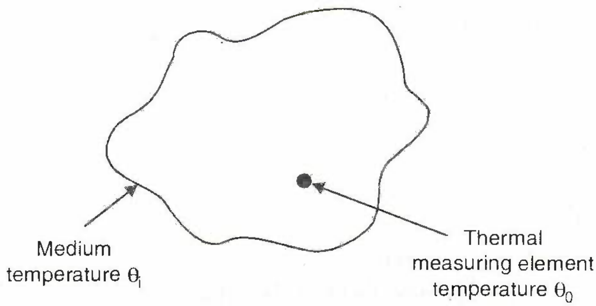


Fig.3.60. Thermal element.

Let, I_i = Temperature of the medium,

I_o = Temperature indicated by the thermal measuring instrument (say a thermometer or thermocouple),

A = Exposed area of the thermal measuring element,

h = Convective heat transfer coefficient,

m = Mass of thermal element, and

c = Specific heat of the element.

Then, the rate of heat flux into the element is,

$$Q = hA\,(\theta_i - \theta_o) \quad \text{...}(i)$$

The rate of enthalpy gain by the element

$$= mc\frac{d\theta_o}{dt} \quad \text{...}(ii)$$

Since the rate of heat flow equals the rate of enthalpy gain by the element, therefore, equating (*i*) and (*ii*) we get :

$$mc\frac{d\theta_o}{dt} = hA(\theta_i - \theta_o)$$

or, $$\frac{mc}{hA}\cdot\frac{d\theta_o}{dt}+\theta_o = \theta_i$$

$$\tau\frac{d\theta_o}{dt}+\theta_o = \theta_i$$

where, $\tau = \dfrac{mc}{hA}$ is known as *time constant* of the system.

In terms of D-operator (where $D = \frac{d}{dt}$), we have

$$(\tau D + 1)\theta_o = \theta_i$$

or, $$\frac{\theta_o}{\theta_i} = \frac{1}{\tau D + 1}$$ ...*Required equation.*

which is an equation of *first-order.* **(Ans.)**

Example 3.19. *A thermometer, idealised as a first-order system with a time constant of 2.2 seconds, is suddenly given an input of 160°C from 0°C.*

(i) What will be reading of the thermometer after 1.2 seconds ?

(ii) Determine its reading if it is initially held at 20°C.

Solution. *Given* : I_i = 160°C; t = 1.2s ; τ = 2.2 s; I_{initial} = 20°C.

(*i*) **Thermometer's reading after 1.2 s :**

We know that, $$I_o = I_i\,(I - e^{-t/\tau})$$ ...[Eqn. 3.59]

$$= 160\,[1 - e^{-(1.2/2.2)}] = \mathbf{67.27°C} \quad \textbf{(Ans.)}$$

(*ii*) **Thermometer's reading if it was initially held at 20°C :**

For a step input from 20°C to 160°C, we have

$$I_o = I_i + (I_{\text{initial}} + I_i)\,e^{-t/\tau}$$ ...[Eqn. 3.62]

$$= 160 + (20 - 160)\,e^{-(1.2/2.2)}$$

$$= 160 + (20 - 160) \times 0.5796 = \mathbf{78.86°C} \quad \textbf{(Ans.)}$$

Example 3.20. *A temperature sensing device can be modelled as a first-order system with a time constant of 5 seconds. It is suddenly subjected to a step input of 30°C to 160°C. Calculate the temperature indicated by the device after 10 seconds after the start of the process.*

Solution. *Given* : τ = 5s ; I_{initial} = 30°C; I_i = 160°C ; t = 10s.

Temperature after 10 seconds is calculated as follows :

$$I_o = I_i + (I_{\text{initial}} - I_i)\,e^{-t/\tau}$$ ...[Eqn. 3.63]

$$= 160 + (30 - 160)\,e^{-10/5}$$

$$= 160 - 130 \times 0.1353 = \mathbf{142.4°C} \quad \textbf{(Ans.)}$$

Example 3.21. *A temperature sensitive transducer when subjected to sudden temperature change takes 9 seconds to reach equilibrium conditions (Three time constants). Calculate the time taken by the transducer to read half of the temperature difference.*

Solution. Time taken to reach equilibrium condition = 3τ = 9s (*Given*).

$\therefore$ Time constant, $$\tau = \frac{9}{3} = 3s$$

Time taken by the transducer to read half of the temperature difference is calculated as follows :

$$I_o = I_i\,(1 - e^{-t/\tau})$$ ...[Eqn. 3.59]

or, $$\frac{I_o}{I_i} = 1 - e^{-t/\tau}$$

or, $$0.5 = 1 - e^{-t/3}$$

or, $$e^{-t/3} = 0.5 \quad \text{or,} \quad e^{t/3} = 2$$

$$t = \mathbf{2.08\ s} \quad \textbf{(Ans.)}$$

SECOND-ORDER SYSTEMS

Example 3.22. *Formulate the governing equation for a second-order system—spring mass system with damping.*

Solution. Refer to Fig. 3.61.

Let, x_i = Input displacement,

x_o = Output displacement,

k = Stiffness of the spring,

C_d = Viscous damping coefficient, and

γ = Damping ratio.

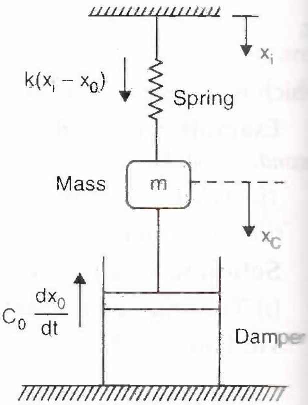


Fig. 3.61. Spring-mass system.

The *forces* acting on the mass are :

(*i*) As both ends of the spring are free to move, therefore

Spring force = Spring stiffness × displacement of one end of the spring relative to other

= $k\,(x_i - x_o)$, acting *downward*.

(*ii*) One end of the dashpot fixed; there is a reaction force acting in the *upward* direction.

Damping force = Damping coefficient × velocity

$$= C_d \times \frac{dx_o}{dt}$$

For translational systems, the Newton's law states that,

$$\Sigma \text{ Force} = \text{Mass} \times \text{acceleration} = m\frac{d^2x_o}{dt^2}$$

or,
$$m\cdot\frac{d^2x_o}{dt^2} = k(x_i - x_o) - C_d\frac{dx_o}{dt}$$

or,
$$mD^2x_o + C_dDx_o + kx_o = kx_i \qquad \left(\text{where, } D = \frac{d}{dt}, \text{ and } D^2 = \frac{d^2}{dt^2}\right)$$

or,
$$x_i = \left(\frac{m}{k}D^2 + \frac{C_d}{k}D + 1\right)x_o \qquad \text{...Required equation}$$

which is an equation of *second-order* type. **(Ans.)**

Comparing the above expression with the standard second-order form, we have :

Undamped natural frequency,

$$\omega_n = \sqrt{\frac{k}{m}}$$

and,
$$\frac{2\gamma}{\omega_n} = \frac{C_d}{k}$$

$$\therefore \quad \text{Damping ratio, } \gamma = \frac{C_d\omega_n}{2k} = \frac{C_d}{2k}\sqrt{\frac{k}{m}}$$

or,
$$\gamma = \frac{C_d}{2\sqrt{km}} \qquad \text{...Required equation}$$

Example 3.23. *(a) Write down the expressions describing the motion of linear and rotational displacement systems of the second-order.*

(b) The pen arrangement of U.V. recorder (second-order system) has a mass of 4.5 g. Calculate the percentage reduction in mass if it is desired to have 15 percent increase in natural frequency of the recorder.

Solution. (*a*) ● The expression for *linear displacement* (spring-mass-damper) *system* is given by :

$$m\frac{d^2I_o}{dt^2}+C_d\frac{dI_o}{dt}+kI_o = kI_i \quad ...(i)$$

where, m = Mass (kg),
C_d = Viscous damping force (Ns/m),
k = Spring stiffness (N/m),
I_o = Output reading, and
I_i = Input reading.

● For the *rotational system*, the expression may be written as :

$$J = \frac{d^2I_o}{dt^2}+C_d\frac{dI_o}{dt}+qI_o = qI_i$$

where, J = Inertia (kg m^2), and
q = Torsional stiffness.

Comparing these expressions with differential equation in the standard form.

$$\frac{1}{\omega_n^2}\cdot\frac{d^2I_o}{dt^2}+\frac{2\gamma}{\omega_n}\cdot\frac{dI_o}{dt}+I_o = I_i, \text{ we get}$$

Natural frequency, $\omega_n = \sqrt{\frac{k}{m}} = \sqrt{\frac{q}{J}}$, as the case may be. **(Ans.)**

(*b*) *Given* : m = 4.5 g; Percentage increase required = 15%.

Using subscripts 1 and 2 for initial and final values respectively, we have

$$\omega_{n1}^2 = \frac{k}{m_1}, \text{and, } \omega_n^2 = \frac{k}{m_2}$$

$$\therefore \quad m_2 = m_1 \times \left(\frac{\omega_{n1}}{\omega_{n2}}\right)^2$$

But, $\omega_{n2} = 1.15\,\omega_{n1}$ ...(*Given*)

$$\therefore \quad m_2 = m_1 \times \left(\frac{\omega_{n1}}{1.15\,\omega_{n1}}\right)^2 = 0.756\,m_1$$

$\therefore$ *Percentage reduction in mass*

$$= \left(\frac{m_1 - m_2}{m}\right) = \left(\frac{1-0.756}{1}\right)\times 100 = 24.4\% \quad \textbf{(Ans.)}$$

Example 3.24. *A second-order system follows the different equation given below :*

$$\frac{d^2I_o}{dt^2}+3\frac{dI_o}{dt}+30I_o = 30I_i$$

where, I_o and I_i are the output and input quantities respectively. Determine the following :

(a) Damping ratio, *(b) Damped natural frequency,*

(c) Static sensitivity, *(d) Time constant.*

Solution. The standard form of the differential equation of a second-order system is given as :

$$\frac{1}{\omega_n^2}\cdot\frac{d^2I_o}{dt^2}+\frac{2\gamma}{\omega_n}\cdot\frac{dI_o}{dt}+I_o = kI_i \quad ...(i)$$

Since the term I_o in eqn. (*i*) has a unit coefficient, therefore to recast the given equation in the standard form, let us divide the given equation throughout by 30; we get

$$\frac{1}{30}\cdot\frac{d^2I_o}{dt^2}+\frac{1}{10}\frac{dI_o}{dt}+I_o = I_i \quad ...(ii)$$

Comparing eqns. (*i*) and (*ii*), we get

$$\omega_n^2 = 30;\ \frac{2\gamma}{\omega_n}=\frac{1}{10}=0.1;\ k=1$$

$\therefore$ Natural frequency, $\omega_n = \sqrt{30} = 5.477\ \text{rad/s}$

(*a*) **Damping ratio,** γ **:**

$$\frac{2\gamma}{\omega_n} = 0.1, \text{ or } \gamma = \frac{\omega_n}{2}\times 0.1$$

or,

$$\gamma = \frac{5.477}{2}\times 0.1 = \mathbf{0.274} \quad \textbf{(Ans.)}$$

(*b*) **Damped natural frequency,** ω_d **:**

$$\omega_d = \omega_n\sqrt{1-\gamma^2} = 5.477\sqrt{1-0.274^2} = \mathbf{5.267\ rad/s} \quad \textbf{(Ans.)}$$

(*c*) **Static sensitivity :**

Static sensitivity, $k = 1$ **(Ans.)**

(*d*) **Time constant** τ **:**

$$\tau = \frac{1}{\omega_n}=\frac{1}{5.477} = \mathbf{0.1826\ s} \quad \textbf{(Ans.)}$$

HIGHLIGHTS

1. The technology of using instruments to measure and control the physical and chemical properties of materials is called *instrumentation.*
2. Modes of measurements are :
   (*i*) Primary measurements (*ii*) Secondary measurements
   (*iii*) Tertiary measurements.
3. A transducer is a device which converts the energy from one from to another.
4. Transducers may be classified as follows :
   A. (*i*) Active transducers (*ii*) Passive transducers
   B. (*i*) Variable-resistance type (*ii*) Digital transducers.

C. (*i*) Variable-resistance type (*ii*) Variable-inductance type
(*iii*) Variable-capacitance type (*iv*) Voltage-generating type
(*v*) Voltage-divider type.

OBJECTIVE TYPE QUESTIONS

Choose the Correct Answer ;

1. LVDT is a
   (*a*) capacitive transducer (*b*) resistive transducer
   (*c*) inductive transducer (*d*) none of them.
2. The size of air-cored transducers in comparison to their iron-cored counter parts is
   (*a*) smaller (*b*) bigger
   (*c*) same (*d*) unpredicatable.
3. LVDT windings are wound on
   (*a*) steel sheets (laminated) (*b*) aluminium
   (*c*) ferrite (*d*) copper.
4. Piezoelectric crystals are used for measurement of .... changes.
   (*a*) static (*b*) dynamic
   (*c*) static and dynamic (*d*) any of these.
5. Piezoelectric crystals produce an e.m.f.
   (*a*) When external mechanical force is applied
   (*b*) when external magnetic field is applied
   (*c*) when radiant energy stimulates the crystal
   (*d*) when the junction of two such crystals is heated.
6. Piezoelectric crystals are used for the measurement of
   (*a*) temperature (*b*) velocity
   (*c*) sound (*d*) none of the above.
7. In the given circuit, how much the voltmeter will read ?
   (*a*) 0 V (*b*) 10 V
   (*c*) 4 V (*d*) 5 V
   (*e*) 3.33 V.

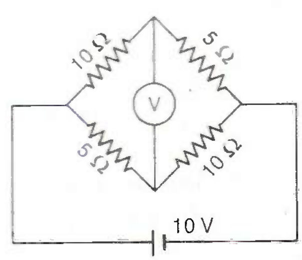


Fig. 3.62

8. Capacitive transducers operate upon the principle (*s*) of
   (*a*) variation of over-lapping area of plates
   (*b*) variation of separation of plates
   (*c*) variation of relative permittivity of dielectric material between two plates
   (*d*) all of the above.
9. Capacitive transducers are normally employed for .... measurements.
   (*a*) static (*b*) dynamic
   (*c*) both static and dynamic (*d*) transient.
10. The thermo-electric effect was first observed by
    (*a*) Seebeck (*b*) Thomas Young
    (*c*) Pirani (*d*) Thermus.

11. Thermocouples are .... transducers.
    (a) active
    (b) passive
    (c) adhesive
    (d) both (a) and (c).
12. Nitro-cellulose cement is used in strain gauges as
    (a) carrier
    (b) base
    (c) adhesive
    (d) lead.
13. A resistance thermometer is basically a/an
    (a) active transducer
    (b) passive transducer
    (c) potentiometer
    (d) none of these.
14. Platinum resistance thermometer can be used upto
    (a) 200°C
    (b) 850°C
    (c) 1200°C
    (d) 1500°C.
15. Which of the following should be incorporated in the RTD to make a temperature sensitive bridge most sensitive to temperature ?
    (a) Platinum
    (b) Nickel
    (c) Thermistor
    (d) Copper.
16. Bourdon tubes have the advantages of
    (a) high accuracy and good dynamic response
    (b) high sensitivity and good repeatability
    (c) not being prone to shock vibrations
    (d) not being susceptible to hysteresis.
17. A transducer is basically a device which converts
    (a) mechanical energy into electrical
    (b) energy or information from one form to another
    (c) mechanical displacement into electrical
    (d) none of these.
18. The gauge factor of a strain gauge is given as
    (a) $G = \dfrac{\Delta R/R}{\Delta l/l}$
    (b) $G = \dfrac{\Delta l/l}{\Delta R/R}$
    (c) $G = \dfrac{\Delta R/R}{\Delta D/D}$
    (d) none of these.
19. The gauge factor G and the Poisson's ratio μ are related as
    (a) $g = 1 + \mu$
    (b) $G = \mu$
    (c) $G = 2 + \mu$
    (d) $G = \dfrac{1+\mu}{2}$.
20. Thermocouples are generally used for accurate temperature measurement upto
    (a) 350°
    (b) 550°C
    (c) 1400°C
    (d) 3500°C.
21. For surface temperature measurement one can use
    (a) strain gauges
    (b) diaphragm
    (c) RTD
    (d) thermocouple.
22. LVDT can be used for
    (a) vibration measurement
    (b) angular velocity measurement
    (c) force measurement in beam
    (d) load measurement on a column.

23. The principle of operation of LVDT is based on variation of
    (*a*) self inductance (*b*) mutual inductance
    (*c*) reluctance (*d*) permeance.
24. An LVDT has an output in the form of
    (*a*) linear displacement of core (*b*) pulse
    (*c*) rotary movement of core (*d*) none of the above.
25. Hall effect transducers have the drawbacks of
    (*a*) high-sensitivity to temperature variation
    (*b*) variation of Hall's coefficient from plate to plate
    (*c*) poor resolution
    (*d*) both (*a*) and (*b*).
26. A Hall's effect pick-up can be used for measuring
    (*a*) pressure (*b*) magnetic flux
    (*c*) relative humidity (*d*) current.
27. Self generating transducers are .... transducers.
    (*a*) active (*b*) passive
    (*c*) secondary (*d*) inverse.
28. The transducer that converts the input signal into the output signal, which is a continuous function of time, is known as .... transducer.
    (*a*) active (*b*) passive
    (*c*) analog (*d*0 digital.
29. A transducer that converts measurand into the form of pulse is called the .... transducer.
    (*a*) active (*b*) analog
    (*c*) digital (*d*) pulse.
30. Certain types of materials generate an electrostatic charge or voltage when mechanical force is applied across them. Such materials are called the
    (*a*) Piezo-electric (*b*) photo-electric
    (*c*) thermo-electric (*d*) none of these.
31. Piezo-electric transdcuers are .... transducers.
    (*a*) active (*b*) passive
    (*c*) inverse (*d*) both (*a*) and (*c*).
32. Piezo-electric transducers work when we apply to it
    (*a*) heat (*b*) mechanical force
    (*c*) vibrations (*d*) illumination.
33. In semiconductor strain gauges, the change in resistance on application of strain is mainly on account of change in
    (*a*) length of wire (*b*) diameter of wire
    (*c*) resistivity (*d*) both (*a*) and (*b*).
34. The drawbacks of semiconductor strain gauges are
    (*a*) low fatigue life (*b*) poor linearity
    (*c*) that they are expensive and brittle (*d*) none of these.
35. In a resistance potentiometer, the non-linearity .... the ratio of potentiometer to load resistance
    (*a*) decreases with the increase in (*b*) increases with the increase in
    (*c*) is independent of (*d*) none of the above.

36. High value pot resistance leads to
    (a) low sensitivity
    (b) high sensitivity
    (c) low non-linearity
    (d) less error.
37. A strain gauge is a passive transducer and is employed for converting
    (a) mechanical displacement into a change of resistance
    (b) pressure into a change of resistance
    (c) force into displacement
    (d) none of the above.
38. Semiconductor strain gauges have
    (a) high gauge factor (–100 to 150)
    (b) very small size (0.7 to 7 mm)
    (c) higher fatigue value
    (d) all of these.
39. The strain gauges should have low
    (a) gauge factor
    (b) resistance
    (c) resistance temperature coefficient
    (d) all of the above.
40. The carrier material employed with strain gauges at room temperature is
    (a) impregnated paper
    (b) bakelite
    (c) aluminium foil
    (d) epoxy.
41. Seismic transducer is used for measurement of
    (a) linear velocity
    (b) angular velocity
    (c) acceleration
    (d) pressure.
42. The accelerometer using LVDT has the advantage of
    (a) high natural frequency
    (b) being contactless device
    (c) better resolution
    (d) all of these.
43. Bonded strain gauges are
    (a) exclusively used for construction of transducers
    (b) exclusively used for stress analysis
    (c) used for both stress analysis and for construction of transducers
    (d) none of the above.
44. A load cell is essentially a
    (a) strain gauge
    (b) thermistor
    (c) resistive potentiometer
    (d) inductive transducer.
45. A load cell is an electro-mechanical device and is widely used for measurement of
    (a) static forces
    (b) dynamic forces
    (c) temperature
    (d0 both (a) and (b).
46. Which of the following can be used for pressure measurement ?
    (a) Thermometer
    (b) Pyrometer
    (c) Bolometer
    (d) Piezoelectric crystal.
47. Radiation pyrometers are used for measuring temperature in the range of
    (a) 500 – 1000°C
    (b) 1000–2000°C
    (c) 1200 – 3500°C
    (d) above 4000°C.
48. The best method of measuring the temperature of hot bodies radiating energy in the visible spectrum is
    (a) thermocouple
    (b) optical pyrometer
    (c) RTD
    (d) thermistor.

49. Pirani gauge is used for measuring .... pressure.
(*a*) very high (*b*) high
(*c*) very low (*d*) atmospheric.

50. Pirani gauge are used for measurement of pressure ranging from
(*a*) 10^{-4} to 1 torr (*b*) 1 to 10 torrs
(*c*) 10 to 100 torrs (*d*) above 100 torrs.

51. The ionization vacuum gauge, in construction, is similar to a
(*a*) vacuum diode (*b*) vacuum triode
(*c*) thyratron (*d*) none of these.

52. The device used for measuring temperatures exceeding 1500°C is
(*a*) radiation pyrometer (*b*) RTD
(*c*) thermocouple (*d*) bimetallic thermometer.

53. The most suitable device for measuring temperature of a furnace is
(*a*) RTD (*b*) thermistor
(*c*) optical pyrometer (*d*) bimetallic thermometer.

54. Which of the following devices cannot be used for measurement of temperature ?
(*a*) RTD (*b*) Thermocouple
(*c*) LVDT (*d*) Pyrometer.

55. Which of the following is not the drawback of radiation pyrometers ?
(*a*) Their initial as well as installation costs are high
(*b*) Poor precision and slow response
(*c*) They need maintenance
(*d*) Each pyrometer needs individual calibration.

56. Pyrometer is used to measure
(*a*) strain (*b*) pressure
(*c*) displacement (*d*) temperature.

57. The device used for measuring low pressure, of the order of 10^{-2} torr, is
(*a*) strain gauge (*b*) Pirani gauge
(*c*) ionization gauge (*d*) any of these.

58. Moving-coil pick-up is used for measuring
(*a*) linear velocity (*b*) vibrations
(*c*) displacement (*d*) pressure.

59. Electronic counters are used for measuring
(*a*) linear velocity (*b*) angular velocity
(*c*) acceleration (*d*) pressure.

60. Angular velocity is measured by
(*a*) strain gauge (*b*) solar cell
(*c*) A.C. tacho-generator (*d*) none of the above.

61. A Wheatstone bridge circuit using strain gauges can be used for measuring
(*a*) static strains (*b*) dynamic strains
(*c*) both (*a*) and (*b*) (*d*) none of these.

62. Dummy strain gauges are used for
(*a*) calibration of strain gauges
(*b*) compensation of temperature variations
(*c*) increasing bridge sensitivity
(*d*) all of the above.

63. In measurements using two strain gauges, the second strain gauge is provided for
 (*a*) temperature compensation (*b*) increasing sensitivity
 (*c*) stability (*d*) linearity
 (*e*) both (*a*) and (*b*).
64. Which of the following additional devices is required for measuring pressure with an LVDT ?
 (*a*) Bellows (*b*) Bourdon tubes
 (*c*) Bolometer (*d*) rotameter
 (*e*) either (*a*) or (*b*).
65. Which of the following devices cannot be used for measurement of pressure ?
 (*a*) LVDT (*b*) RTD
 (*c*) Piezo-electric transducers (*d*) Piezo-resistive transducers.
66. The transducers used for measurement of linear displacement are
 (*a*) strain gauges and resistive potentiometers
 (*b*) LVDTs, capacitive transducers and Hall effect transducers
 (*c*) thermocouples, thermistors and RTDs
 (*d*) both (*a*) and (*b*).
67. Rotational displacement can be measured by
 (*a*) strain gauges (*b*) resistive potentiometers
 (*c*) reluctance transducers (*d*) both (*b*) and (*c*).
68. Temperature compensation, in bridge circuit arrangement, is affected by
 (*a*) using dummy strain gauges
 (*b*) using strain gauges of smaller gauge factor
 (*c*) reversing strain gauges
 (*d*) any of these.
69. Which one of the following devices cannot be used to measure pressure ?
 (*a*) Strain gauge (*b*) LVDT
 (*c*) Piezoelectric crystal (*d*) Pyrometer.
70. Which of the following additional is required for measuring pressure with piezoelectric crystal ?
 (*a*) Bellows (*b*) Strain gauges
 (*c*) Rotameter (*d*) RTD.

ANSWERS

1. (*c*)	2. (*b*)	3. (*c*)	4. (*b*)	5. (*a*)	6. (*a*)	7. (*e*)
8. (*d*)	9. (*c*)	10. (*a*)	11. (*a*)	12. (*e*)	13. (*b*)	14. (*b*)
15. (*c*)	16. (*b*)	17. (*b*)	18. (*a*)	19. (*c*)	20. (*c*)	21. (*d*)
22. (*c*)	23. (*b*)	24. (*a*)	25. (*d*)	26. (*d*)	27. (*a*)	28. (*c*)
29. (*c*)	30. (*a*)	31. (*d*)	32. (*b*)	33. (*c*)	34. (*c*)	35. (*b*)
36. (*b*)	37. (*a*)	38. (*d*)	39. (*c*)	40. (*a*)	41. (*c*)	42. (*d*)
43. (*c*)	44. (*a*)	45. (*d*)	46. (*d*)	47. (*c*)	48. (*b*)	49. (*c*)
50. (*a*)	51. (*b*)	52. (*a*)	53. (*c*)	54. (*c*)	55. (*b*)	56. (*d*)
57. (*b*)	58. (*a*)	59. (*b*)	60. (*c*)	61. (*c*)	62. (*b*)	63. (*e*)
64. (*e*)	65. (*b*)	66. (*d*)	67. (*d*)	68. (*a*)	69. (*d*)	70. (*a*).

THEORETICAL QUESTIONS

1. Define the term "instrumentation".
2. List the various modes of measurement.
3. Enumerate the elements of a measurement system.
4. What is transducer ?
5. What are the functions of a transducer in an electronic instrumentation system ?
6. How are transducers classified ?
7. What are the advantages of electromechanical transducers ?
8. Explain briefly with diagrams important transducer actuating mechanisms.
9. Describe briefly the following :
   (*i*) Thermistors and resistance thermometers.
   (*ii*) Wire resistance strain gauges.
10. Give the classification of variable inductance transducers.
11. Explain briefly any two of the following transducers :
   (*i*) Self-generating variable inductance transducer – Electromagnetic type
   (*ii*) Variable reluctance transducer
   (*iii*) Mutual inductance transducer
   (*iv*) Linear-variable-differential transformer (LVDT).
12. What is the principle on which a capacitive transducer works ?
13. What are the advantages and disadvantages of capacitive transducers ?
14. Give the applications of capacitive transducers.
15. What is a piezo-electric transducer ? List the advantages and disadvantages of piezo-electric transducers.
16. How are photoelectric transducers classified ?
17. Explain briefly the following :
   (*i*) Photoemissive cell (*ii*) Photoconductive cell
   (*iii*) Photovoltaic cell.
18. What is a strain gauge ?
19. Explain briefly with neat diagrams, any two of the following :
   (*i*) Wire-wound strain gauges (*ii*) Foil-type strain gauges
   (*iii*) Semiconductor strain gauges (*iv*) Capacitive strain gauges.

UNSOLVED EXAMPLES

1. A linear resistance potentiometer is 50 mm long and is uniformly wound with a wire having a resistance of 10000 Ω. Under normal conditions, the slider is at the centre of the potentiometer.
   (*i*) Find the linear displacements when the resistances of the potentiometer are measured by a wheatstone bridge for two cases are : (a) 3800 ohms and (b) 7500 ohms.
   (*ii*) If it is possible to measure a minimum value of 12 ohms resistance with the above arrangement, find the resolution of the potentiometer in mm.

   [Ans. (*i*) 6 mm, 12.5 mm; (*ii*) 0.06 mm]
2. In a linear voltage differential transformer the output voltage is 2.0 V at maximum displacement. At a certain load, the deviation from linearity is maximum and it is ± 0.004 V from a straight line through origin. Find the linearity at the given load.

   [Ans. ±2%]

3. The output of a LVDT is connected to a 5V voltmeter through an amplifier whose amplification factor is 250. An output of 2 mV appears across the terminals of LVDT when the core moves through a distance of 0.5 mm. If the multimeter has 100 divisions and the scale can be read to a $\frac{1}{5}$ of a division. Calculate :
   (*i*) The sensitivity of LVDT, and
   (*ii*) The resolution of the instrument in mm, **[Ans.** (*i*) 4 m V/mm, (*ii*) 0.01 mm]
4. A parallel plate capacitive transducer uses plates of area 250 mm^2 which are separated by a distance 0.2 mm.
   (*i*) Calculate the value of capacitance when the dielectric is air having a permittivity of 8.85×10^{-12}F/m.
   (*ii*) Calculate the change in capacitance if a linear displacement reduces the distance between the plates to 0.18 mm. Also calculate the ratio of per unit change of capacitance to per unit change of displacement.
   (*iii*) If a mica sheet 0.01 mm thick is inserted in the gap, calculate the value of original capacitance and change in capacitance for the same displacement. Also calculate the ratio of per unit change of capacitance to per unit change in displacement. The dielectric constant of mica is 8.
   [Ans. (*i*) 11.06 pF; (*ii*) 1.23 pF, 1.11; (*iii*) 11.57 pF, 1.35 pF, 1.167]
5. A capacitive transducer uses two quartz diaphragms of area 675 mm^2 separated by a distance of 3.8 mm. A pressure of 850 kN/m^2 when applied to the top diaphragm produces a deflection of 0.55 mm. The capacitance is 330 pF when no pressure is applied to the diaphragms. Determine the value of capacitance after the application of a pressure of 850 kN/m^2. **[Ans.** 385.8 kN/m^2]
6. A capacitive transducer, used in pressure measuring instrument has a spacing of 4.2 mm between its diaphragms. A pressure of 600 kN/m^2 produces an average deflection of 0.28 mm of the diaphragm of the transducer. A transducer which has a capacitance of 250 pF before the application of pressure is connected in an oscillation circuit having a frequency of 120 kHz. Determine the change in frequency of oscillator after the application of pressure to the transducer.
   [Ans. 4.1 kHz approx]
7. A 2 mm thick quartz piezoelectric crystal having a voltage intensity of 0.055 Vm/N is subjected to a pressure of 1.8 MN/m^2. Calculate the voltage output and charge density of the crystal. Take the permittivity of quartz as 40.6×10^{-12} F/m.
   [Ans. 198 V, 2.23 pC/N]
8. A piezoelectric material measuring 5 mm × 5 mm × 1.5 mm is used to measure a force. Its voltage sensitivity is 0.055 Vm/N. Calculate the force if voltage developed is 110V.
   [Ans. 33N]
9. The following data relate to a barium titrate pick-up :
   Dimensions : 5 mm × 5 mm × 1.25 mm; Force acting on the pick-up = 5N. The charge sensitivity of the crystal = 150 pC/N; Permittivity = 12.5×10^{-9} F/m; Modulus of elasticity = 12×10^6 N/m^2. Calculate strain, charge and capacitance.
   [Ans. 0.0167 : 750 pC; 250 pF]
10. A strain of 5 micro-strain is caused in a structural member when subjected to a compressive force. Two separate strain gauges are attached to the structural member, one is nickel wire strain gauge (gauge factor = –12.1) and other is nichrome wire strain gauge (gauge factor = 2). If the resistance of strain gauges before being strained is 130 Ω, calculate the change in the value of resistance of the gauges after they are strained.
    [Ans. 7.865 mΩ (increase); 1.3 mΩ (decrease)]

11. A strain gauge is bonded to a beam which is 10 cm long and has a cross-sectional area of 4 cm^2. The unstrained resistance and gauge factor of the strain gauge are 220Ω and 2.2 respectively. On the application of load the resistance of the gauge changes by 0.013Ω. If the modulus of elasticity for steel is 207 GN/m^2, calculate : (*i*) the change in length of the steel beam, and (*ii*) the amount of force applied to the beam.

    [**Ans.** (*i*) 2.68×10^{-6} m; (*ii*) 2.219 kN]

12. A single strain gauge having resistance 144 Ω is mounted on a steel cantilever beam at a distance of 0.15 m from the free end. The beam dimensions are 25 cm (length) × 2.0 cm (width) × 0.3 cm (depth). An unknown force applied at the free end produces a deflection of 12.7 mm of the end. If the change in gauge resistance is found to be 0.1824Ω, calculate the gauge factor. Take Young's modulus for steel as 200 GN/m^2. [**Ans.** 2.3]

CHAPTER

4 Signal Conditioning; Data Acquisition, Transmission and Presentation /Display

4.1 INTRODUCTION

4.1.1. General Measurement System Components

Fig. 4.1 shows the general measurement system components :

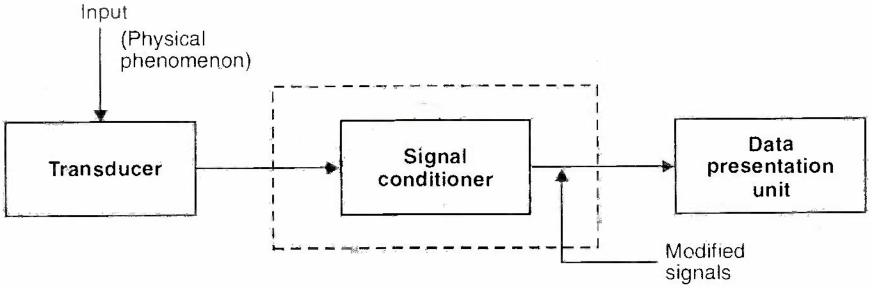


Fig. 4.1. Components of a general measurement system.

— The *"first stage" of the instrumentation or measurement system which detects the measurand (which is basically a physical quantity) is termed as* **detector-transducer stage.** In this stage, in most of the cases, the quantity is detected and is transduced into an electrical form.

— The output from the first stage *needs certain modifications* before it becomes compatible with the data presentation stage. The necessary modification is carried out in the *"intermediate stage"*, more commonly referred to as the **signal conditioning stage**.

— The *"last stage"* of the measurement system may consist of indicating, recording, displaying, data processing elements or may consist of control elements.

- *Measurement of dynamic mechanical quantities* places special requirements on the elements in the signal conditioning stage.

Large amplifications, as well good transient response, are often desired, both of which are difficult to obtain by mechanical hydraulic, or pneumatic methods. Consequently, *electrical or electronic elements are usually required.*

4.1.2. Signal Conditioning and its Necessity

Signal conditioning *may be defined as the process of modifying the output signals from the transducer into usable and satisfactory signal using amplification, attenuation, non-linearisation, linearisation or multiplication by another function.*

The necessity of signal conditioning may be due to following *reasons* :

1. Signals may be too noisy due to electromagnetic interference.
2. Signals may be too small, usually is mV range.
3. Signals may be non-linear and require to be converted into digital form.
4. Signals may be analog one and require to be converted into digital form.
5. Signals may be digital one and need to be converted into analog signals.
6. It may be required to improve the quality of digital signals.

4.1.3. Process Adopted in Signal Conditioning

Following processes are usually adopted in signal conditioning :

1. **Protection.** The range of the output signals from the transducer may be so high that it may damage the next unit or element which needs to be protected.

   Example: If a high voltage/current signals are fed to the microprocessor, it will get damaged. *The microprocessor can be protected by* :

   (*i*) employing a series of current limiting resistors, fuses to break if current is too high;

   (*ii*) using a step down transformer if the voltage is too high;

   (*iii*) employing polarity protection;

   (*iv*) using voltage limitation circuits etc.

2. **Getting right type of signals:**

   — The output signals of a *transducer* is of *analog* type, this needs to be converted to D.C. voltage or current.

   — The output signal of a *microprocessor* is of digital nature, it needs to be converted into analog form to feed it to an actuator for process controlling.

3. **Getting correct level of signals:**

   — The level of the output signal may be too small (to the tune of few mV), this needs to be amplified for feeding it into an analog-to-digital converter. It may also be difficult to measure such low level signals.

   — For amplification operational amplifiers (Op-amp) are widely used.

4. **Elimination of interferences :**

   — Some undesired signals or disturbances (say noise disturbance due to electromagnetic interference) may be associated with the output signals, these need to be eliminated.

   — Such interferences are eliminated by the use of *filters*.

5. **Manipulation of signals:** The output signals may be non-linear in nature, these need to be linearised and vice versa.

4.1.4. Mechanical Amplification and Electrical Signal Conditioning

Limitations/disadvantages of mechanical amplification:

In the field of dynamic measurements, strictly, mechanical systems are much more uncommon than they were in the years past, largely because of several inherent *disadvantages.* Mechanical amplification by the elements such as linkages, gearing, or cams (these elements present design problems of immense magnitudes particularly if dynamic inputs are to be handled) is quite limited because of the following *reasons :*

(*i*) When amplification is required *frictional forces are also amplified, resulting in considerable undesirable signal loading.* These effects, coupled with backlash and elastic deformations, result in *poor response.*

(*ii*) Initial loading results in reduced frequency response and in certain cases, depending on the particular configuration of the system, *phase response is also a problem.*

Advantages of electrical signal conditioning:

In several detector-transducer combinations which provide an output in electrical form, it is convenient to perform further signal conditioning electrically.

- Such conditioning may typically *include:*
  - — Converting resistance changes to voltage changes;
  - — Subtracting offset voltages;
  - — Increasing signal voltages;
  - — Removing unwanted frequency components.
- Electrical methods are also preferred for their ease of *power amplification.*
  - — *Additional power may be fed into the system to provide a greater output power than input by the users of "power amplifiers"*, which have no important mechanical counterpart in most instrumentation. (It is true that hydraulic and pneumatic systems may be set up to increase signal power; however, their use is limited to relatively slow-acting control applications, primarily in the fields of chemical processing and electric power generation). This technology is of a particular value when recording procedures employ stylus-type recorders, mirror galvanometers, or magnetic-disc methods.

4.2 FUNCTIONS OF SIGNAL CONDITIONING EQUIPMENT

The signal conditioning equipment may be required to perform the following *functions on the transduced signal:*

1. Amplification
2. Modification or modulation
3. Impedance matching
4. Data processing
5. Data transmission.

1. **Amplification.** It means *enhancement of the signal level* which is often in the low level range. The amplification system must bring the level of transducer signal to a value adequate enough to make it useful for *conversion, processing, indicating* and *recording.*
2. **Modification or modulation.** It means to *change the form of signal.* The signal may be *smoothened, linearised, filtered or converted into digital form.*
3. **Impedance matching.** The signal conditioning equipment arranges the input, and output impedances of the matching device so as to prevent loading of the transducer and to maintain a high signal level at the recorder.
4. **Data processing.** To carry out mathematical operations (*e.g.*, addition, subtraction, differentiation, integration etc.) before indication or recording of data.
5. **Data transmission.** To transmit signal from one location to another without changing the contents of the information.

The whole task of signal conditioning requires the following:

(*i*) Ingenuity;

(*ii*) Proper selection of components;

(*iii*) Faithful reproduction of signal.

- The elements of signal conditioning are designed in such a fashion as to be *insensitive to all extraneous inputs. The accuracy, range and dynamic response are all designed to be compatible with the detector transducer.*
- In several situations the *"signal conditioning"* or *"data acquisition equipment"* is an *excitation* and *amplification system for passive transducers.* It may be an *amplification system for active transducers.* In both the applications, the transducer output is brought upto adequate level to make it useful for *conversion, processing, indicating* and *recording.*
- In case of *"passive transducers"* (*e.g.,* strain gauges, potentiometer resistance thermometers, inductive and capacitive transducers) excitation is needed because these transducers do not generate their own voltage or current; the excitation is provided from external sources.
- The *"active transducers"* (*e.g.,* technogenerators, thermocouples, inductive pick-ups and piezoelectric crystals) do not require excitation from an external source since they produce their own electrical output. However, these signals have a low voltage level and as such they need to be amplified.

The excitation sources may be:

- D.C. voltage source.
- A.C. voltage source.

Figures 4.2 and 4.3 show D.C. and A.C. signal conditioning systems respectively.

D.C. signal conditioning system:

Refer to Fig. 4.2. The resistance transducers like strain gauges constitute one or more than one arm of a Wheatstone bridge which is excited by an isolated D.C. source. The bridge can be balanced by a potentiometer and can also be calibrated for unbalanced conditions.

Characteristics of a D.C. amplifier:

(*i*) It should have extremely good thermal and long term stability.

(*ii*) It may require balanced differential inputs giving high mode rejection ratio (CMRR); CMRR is a *measure of ratio of desired signal to undesired signal, this value should be as high as possible.*

Advantages:

(*i*) D.C. amplifier is easy to calibrate at low frequencies.

(*ii*) It is able to recover from an overload condition unlike its A.C. counterpart.

Disadvantages:

- The major disadvantage of a D.C. amplifier is that it suffers from the *problem of drift.* As a result, the *low frequency spurious signals come out as data information.* This problem is overcome by the *use of the drift amplifiers.*
- The D.C. amplifier is followed by a lowpass filter which eliminates high frequency components or noise from the data signal.

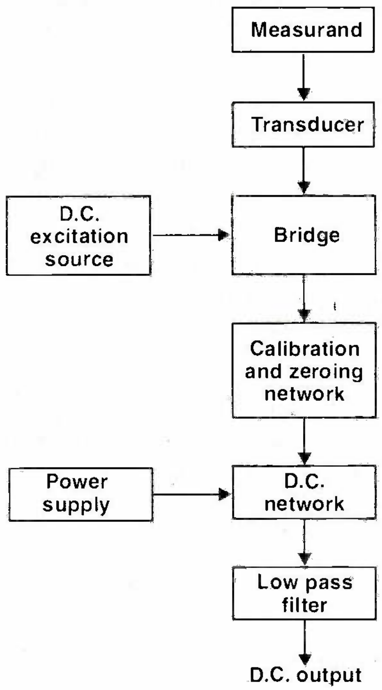


Fig. 4.2. D.C. signal conditioning system.

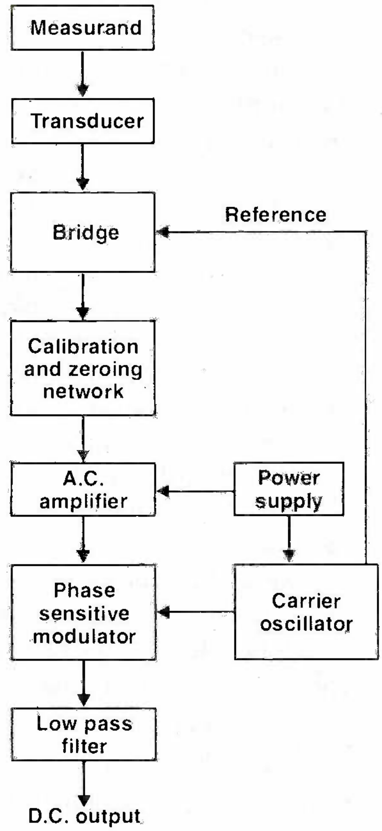


Fig. 4.3. A.C. signal conditioning system.

Uses. D.C. systems are generally used for common resistance transducers such as *potentiometers and resistance strain gauges.*

A.C. signal conditioning system:

Refer to Fig. 4.3. The problems which are encountered in D.C. systems are overcome through *carrier type A.C. signal conditioning system.*

The transducer parameter variations amplitude modulate the carrier frequencies at the bridge output and the waveform is *amplified and demodulated.* The demodulation is phase sensitive so that polarity of D.C. output indicates the direction of the parameter change in the bridge output.

In carrier systems, it is *very easy to obtain very high rejection of mains frequency pick-up.*

— Active filters can be used to reject this frequency and prevent overloading of A.C. amplifier.

— The carrier frequency components of the data signal are filtered out by the phase-sensitive demodulators.

Uses. A.C. systems are used for variable reactance transducers and for systems where signals have to be transmitted long via cables to connect the transducers to the signal conditioning equipment.

— The physical quantities like pressure, temperature, acceleration, strain etc. after having being *transduced* into their analogous electrical form and *amplified* to sufficient current or voltage levels (say 1 to 10 V) are further *processed by electronic circuits.*

— The signal, in some applications, does not need any further processing and the amplified signal may be directly applied to indicating or recording or control instruments.

— Several applications, however, involve further processing of signals which involve linear and non-linear operations.

4.3 AMPLIFICATION

An **amplifier** *is a device which is used to increase or augment the weak signal. It may operate on mechanical (levers, gears etc.) optical, pneumatic and hydraulic, or electrical and electronic principles.*

The *ratio of output signal (I_o) to input signal (I_i) for an amplifier is termed as* **gain, amplification** *or* **magnification**. The gain of amplification (G) is expressed as:

$$G = \frac{I_o}{I_i} \qquad ...(4.1)$$

Since $\frac{I_o}{I_i}$ are in the same units, the gain G is a *dimensionless quantity.*

Invariably, in order to get greater magnification, two or more amplifiers are arranged in series/cascades. The overall gain of the arrangement (assuming that no loading occurs) is given by the product of individual gains of the amplifying units,

i.e.,

$$\frac{I_o}{I_i} = G_1 \cdot G_2 \cdot G_3 \cdots\cdots \qquad ...(4.2)$$

4.4 TYPES OF AMPLIFIERS

The amplifiers, on the basis of principle of working, may be *categorised* as follows:

1. Mechanical amplifiers.
2. Fluid amplifiers.
3. Optical amplifiers.
4. Electrical and electronic amplifiers.

4.5 MECHANICAL AMPLIFIERS

The mechanical amplifiers may be further *classified* as follows:

(*i*) ***Simple and compound levers :*** The compound lever has two or more levers linked together so that output from one lever provides the input to the other.

Example. The *Huggenberger extensometer* is one of the most popular and *accurate mechanical amplifier.* It uses a system of compound levers to give *very high magnification* to the order of *2000 or even more.*

(*ii*) ***Simple and compound gears :*** The simple and compound gear trains are used quite frequently *to provide mechanical amplification of either angular displacement or rotary speed.*

A "*compound gear train*" gives *greater modification with the additional advantage of no change in the direction of input signal.*

Examples. The gear trains are used for the magnification of displacement in the *Bourdon tube pressure gauge* and in the *dial-test indicator* where linear movement is translated into rotation by means of rack and pinion.

Limitations of mechanical amplification:

The mechanical amplification, as earlier stated, usually suffers from the errors caused by the following *factors* :

(*i*) Internal loading;

(*ii*) Friction at the mating parts;

(*iii*) Elastic deformation;

(*iv*) Backlash.

4.6 FLUID AMPLIFIERS

Fluid amplifiers may be *classified* as follows:

(*i*) ***Hydraulic amplifier*** : When a small displacement is applied to a piston operating inside a cylinder containing some liquid, there occurs a large displacement of the liquid in the output tube which has a small diameter.

Example. This principle is employed in the *mercury-in-glass thermometer* and the *single-column manometers.*

(*ii*) ***Pneumatic amplifier*** : Pneumatic methods are extensively used and can be applied to any type of measurement.

4.7 OPTICAL AMPLIFIERS

In optical amplification, a ray of light strikes a mirror with an angle of incidence i and gets reflected with angle of reflection equal to the angle of incidence. When the mirror rotates through an angle θ, the angle of incidence change to $(i + \theta)$. Before rotation of the mirror, the angle between the incident ray and reflected ray is $2i$ and after rotation it is $2(i + \theta)$. Obviously there is angular magnification of 2θ between the incident and reflected rays. In order to get a greater magnification, more number of mirrors surfaces may be used.

Examples. This principle to amplify the input signals is used in the following cases:

— Optical levers;

— U.V. galvanometers;

— Mechanical-pointer galvanometers.

4.8 ELECTRICAL AND ELECTRONIC AMPLIFIERS

The electrical amplifiers are used to *increase the magnitude of weak voltage or current signals resulting from electromechanical transducers.*

4.8.1. Desirable Characteristics of Electronic Amplifiers

The following are the desirable characteristics of electronic amplifiers:

(*i*) *High input impedance* so that its loading effect on the transducer in minimum.

(*ii*) *Low output impedance* so that the amplifier is *not unduly loaded by the display or recording device.*

(*iii*) *Frequency response* should be as *good as that of the transducer.*

4.8.2. Electronic Amplification of Gain

- The following are the several generalities that can be listed for the ideal (but non-existent) electronic amplifier:

(*i*) Infinite gain (lower gain can be obtained by adding attenuation circuits).

(*ii*) Infinite input impedance; no input current, hence no load on the previous stage or device.

(*iii*) Zero output impedance (low noise).

(*iv*) Instant response (wide frequency bandwidth).

(*v*) Zero output for zero input.

(*vi*) Ability to ignore or reject, extraneous inputs.

Of course, none of these aims can be completely realized, it is often possible to approach them, and their assumption simplifies circuit analysis.

- In an electronic amplifier, separate power is provided so that the output power may exceed the input if that is required.

Here, if v_i = Input voltage,

i_i = Input current,

v_o = Output voltage, and

i_o = Output current,

Then ;

$$\text{Gain} = \frac{\text{Power output}}{\text{Power input}} = \frac{v_o i_o}{v_i i_i} \qquad \text{...(4.3)}$$

$$\text{Voltage amplification} = \frac{\text{Voltage output}}{\text{Voltage input}} = \frac{v_o}{v_i} \qquad \text{...(4.4)}$$

$$\text{Current amplification} = \frac{\text{Current output}}{\text{Current input}} = \frac{i_o}{i_i} \qquad \text{...(4.5)}$$

- Another way of expressing *power gain* is through the use of *decibel*.

The common logarithm (log to the base 10) of power gain is known as *bel power gain*.

$$\text{Power gain} = \log_{10}\left(\frac{P_o}{P_i}\right) \text{bel}$$

$$1 \text{ bel} = 10 \, dB$$

$$\therefore \quad \text{Power gain} = 10 \log_{10}\left(\frac{P_o}{P_i}\right) \text{dB} \qquad \text{...(4.6)}$$

If the two powers are developed in the same resistance or equal resistance, then

$$P_i = \frac{V_i^2}{R} = I_i^2 R$$

$$P_o = \frac{V_0^2}{R} = I_o^2 R$$

$$\therefore \quad \textit{Voltage gain} = 10 \log_{10} \frac{V_0^2/R}{V_i^2/R} = 20 \log_{10} \frac{V_o}{V_i} \text{ dB} \qquad \text{...(4.7)}$$

$$\text{Current gain} = 10 \log_{10} \frac{I_o^2 R}{I_i^2 R} = 20 \log_{10} \frac{I_o}{I_i} \, dB \qquad ...(4.8)$$

Example 4.1. *A three-stage amplifier has a first voltage gain of 100, second stage voltage gain of 200 and third stage voltage gain of 400. Find the total voltage gain in dB.*

Solution. First-stage voltage gain in *dB*

$$= 20 \log_{10} 100 = 20 \times 2 = 40$$

Second-stage voltage gain in *dB*

$$= 20 \log_{10} 200 = 20 \times 2.3 = 46$$

Third-stage voltage gain in *dB*

$$= 20 \log_{10} 400 = 20 \times 2.6 = 52$$

Total voltage gain = 40 + 46 + 52 = 138 *dB* **(Ans.)**

Example 4.2. *(i) A multistage amplifier employs five stages each of which has a power gain of 30. What is the total gain of the amplifier in dB?*

(ii) If a negative feedback of 10 dB is employed, find the resultant gain.

Solution. Absolute gain of each stage = 30

Number of stages = 5

(*i*) *Power gain of one stage* = $10 \log_{10} 30$ dB = 14.77

∴ Total *power gain* = 5 × 14.77 = **73.85 dB (Ans.)**

(*ii*) *Resultant power gain with negative feedback*

= 73.85 – 10 = **63.85 dB (Ans.)**

4.8.3. A.C. and D.C. Amplifiers

The instrumentation systems usually employ the following two types of electronic amplifiers.

(*i*) A.C. coupled amplifiers.

(*ii*) D.C. coupled amplifiers.

- For an "*A.C. amplifiers*" bandwidth is the range of frequencies between which gain or amplitude ratio is constant to within – 3dB (3dB down points). This corresponds to the frequencies at which the voltage output amplitude falls by 29.3% to 70.7% of the maximum value.
  - — The "*A.C. amplifiers*" are *only capable of dealing with rapid, repetitive signals but are usually simpler and cheaper* when compared with their D.C. counterparts.
  - — In an "A.C. amplifier system" the amplifier drift and spurious noise are *not significant*; the mains frequency pick-up *rejection is very high.*
- The "*D.C. amplifiers*" are capable of *amplifying static, slowly changing or rapid-repetitive input signals.*
  - — The "D.C. amplifier systems" are *easy to calibrate at low frequencies, and have the ability to recover rapidly from overload conditions.*

4.8.4. Modulated and Unmodulated Signals

The measurands may be "*pure*" in the sense that analog electrical signal contains *nothing more than the real time variation of the measurand information itself.*

On the other hand, the signal may be "*mixed*" with a **carrier** which consists of a *voltage oscillation at some frequency higher than that of the signal.* A common rule of thumb is that the *frequency ratio should be at least 10:1.*

The measurand affects the *carrier* by varying either its *amplitude* or its *frequency:*

— In the *former* case the *carrier frequency is held constant and its amplitude is varied by the measurand. This process is known as* **Amplitude modulation** (or **AM**).

— In the *latter* case the *carrier amplitude is held constant and its frequency is varied by the measurand.* This is known as **Frequency modulation** (or **FM**).

The most familiar use of AM and FM transfer of signals is in AM and FM *radio broadcasting.*

When *"modulation"* is used in instrumentation "amplitude modulation" (AM) is the *more common form.*

— Nearly any mechanical signal from a passive pick-up can be transduced into an analogous AM form. Sensors based on either *inductance* (*e.g.* differential transformer) or *capacitance* (*e.g.* capacitance pickup for liquid level) require an A.C. excitation.

In addition, however, resistance-type sensors may also use an A.C. excitation, as with some strain gauge circuits.

It is required to extract signal information from the modulated carrier.

- This operation, *when AM is used,* may take several forms:
  — The simplest is merely to display the entire signal using an *oscilloscope or oscillograph,* and then to *"read"* the result from the envelope of the carrier.
  — More commonly, the *mixed signal and carrier are* ***"demodulated"*** by *"rectification* and *filtering".*
- FM demodulation is more complex operation and may be accomplished through the use of
  — Frequency discrimination,
  — Ratio detection, or
  — IC phase-locked loops.

4.8.5. Integrated Circuits (ICs)

The integrated circuits (ICs), as the name implies, are groups of circuit elements combined to perform specific purposes. For the most part the elements consist of *transistors, diodes, resistors,* and, to lesser extent, *capacitors,* all connected and packaged in convenient plug-in units.

ICs from the building blocks are used to construct more complex circuits such as :

— Differential amplifiers;

— Mixers (for combining signals);

— Timers;

— Filters;

— Audio preamps;

— Auto-power amplifiers;

— Voltage references;

— Regulators and comparators;

— Several digital devices.

4.8.6. Operational Amplifiers (Op-amp)

An ***operational amplifier (Op-amp)*** *is a linear integrated circuit (IC) that has a very high voltage gain, a high input impedance and a low output impedance.*

It is so called because it can be employed to carry out many different mathematical operations like *"addition". "subtraction", "multiplication", "division", "integration", "differentiation"* etc.

- Operational amplifiers are linear integrated circuits that work on *relatively low supply voltage.*
- They are *reliable and inexpensive.*
- An *deal operational amplifier* is device of infinite voltage gain, infinite bandwidth, infinite input impedance (open) and zero output impedance.
- An Op-amp may contain two dozen transistors, a dozen resistors and one or two capacitors.

  Examples : µA 709, LM 108-LM 208, CA 741 CT and CA741T.

4.8.6.1. Specifications/Characteristics of an Op-amp

While selecting an Op-amp, the following characteristics need to be considered:

1. *Input offset voltage.* It is the voltage that must be applied at the input terminals to make the output voltage zero (This is about 2 mV for a 741 amplifier). The offset voltage changes with temperature.
2. *Input offset current.* It is defined as the net difference in current that must be applied at the input terminals to make the output voltage zero (This is 20 nA for a 741 amplifier).
3. *Input check current.* It is the mean of the two input currents to make the voltage zero.
4. *Slew rate.* It is the maximum rate at which the output can change. It is expressed as volts/microseconds.
5. *Unity gain frequency.* This is the frequency at which the open loop gain of the amplifier becomes unity.
6. *Common mode rejection ratio (CMRR).* It is the ratio of desirable signals to undesirable signals.

- An Op-amp is the *basic building block* for:

— Amplifiers
— Integrators
— Summers
— Differentiators
— Comparators
— A/D and D/A converters
— Active filters
— Sample and hold amplifiers.

4.8.6.2. Op-amp description

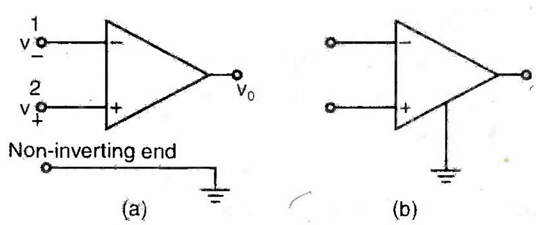


Fig. 4.4. Op-amp symbol

Fig. 4.4(*a*) shows a standard symbol (a triangle having two input labelled differently and a single output) for an Op-amp, the one shown in Fig. 4.4(*b*) is also oftenly used.

One input terminal is designated by *–ve* sign, it is called *inverting end* while other input terminal is designated by a *+ve* sign, it is called *non-inverting end.*

These plus (+) and minus (–) polarities indicate *phase reversal* only. It does not mean that v_1 and v_2 [Fig. 4.4(*a*)] are negative and positive respectively. Additionally, it also *does not imply* that a positive input voltage has to be connected to the plus-marked non-inverting terminal 2 and negative input voltage to negative-marked inverting terminal 1. *In fact amplifier can be used either way up so to speak.* It may also be noted that all input and output voltages are referred to a common reference usually the ground shown in Fig. 4.4(*a*).

Operational amplifier operating with –ve feedback possesses stable closed loop gain and also insensitivity to variation of supply voltage and ambient temperature.

• The voltage at the output terminal v_o, is the product of the amplifier gain G, and the voltage difference :

$$v_o = G(v_+ - v_-) \qquad ...(4.9)$$

The output voltage is roughly limited to the power supply voltages V_{cc} and V_{ee}, as the voltage difference increases; if the voltage difference becomes too large, the output *saturates* near one of these values and remain constant.

• *The differential characteristic of op-amp has great importance in instrumentation because it eliminates offset voltages and noise signals common to both input terminals.* For example, nearby power lines may induce 50-cycle noise in the exterior circuitry leading to the amplifier. Such line noise is often present in identical form at both input terminals. This behaviour is known as *common-mode rejection.* If, instead, an op-amp receives the output of a voltage-sensitive Wheatstone bridge, the common offset voltages of the two voltage dividers are cancelled, and only the desired difference voltage is applied.

Limitations of Op-amp : Most of the Op-amps have a nonideal characteristic according to which they *do not completely satisfy the differential amplifying property.* With *both inputs grounded residual output voltage remains.*

— The multitude of transistors, resistors, and other elements within the Op-amp are *never perfectly matched,* so the amp output actually reaches zero at some small *non-zero* input voltage. To accommodate this input offset voltage, the common op-amp is provided with pins marked "offset null" or "balance" which provides a means for adjusting the unwanted offset voltage towards zero.

Another limitation is that the actual common-mode rejection is finite. If the two input signals each include a common-mode voltage v_{cm}, the Op-amp's actual voltage will be,

$$v_o = G(v_+ - v_-) + G_{cm}\, v_{cm} \qquad ...(4.10)$$

The finite common-mode rejection is characterised by the "*common-mode rejection ratio* (CMRR)" in decibels:

$$\text{CMRR} = 20 \log_{10}\left(\frac{G}{G_{cm}}\right) \text{dB} \qquad ...(4.11)$$

Since typical Op-amps have a CMRR of 60 to 120 dB, therefore, the common-mode gain is typically 10^3 to 10^6 times smaller than the differential gain, hence a *high CMRR is desirable.*

• Further, the performance of Op-amp may be limited by thermal drift. Both internal and external circuit elements may be temperature sensitive, and design of each circuit usually includes compensating features.

A wide variety of Op-amps are available, and their differences largely represent attempts to *improve*:

— Thermal stability;
— CMRR;
— Offset voltages;
— Frequency response .... These refinements, however, increase the cost.

4.8.6.3. Applications of Op-amp

Operational amplifiers may be used as the *basic components of* :

- Linear voltage amplifiers;
- Integrators and differentiators;
- Function generators;
- Impedance transformers;
- Differential amplifiers;
- Voltage comparators;
- Filters;
- Many other devices.

4.8.6.4. Op-amp circuits used in Instrumentation

Some of the commonly used Op-amp circuits are described below:

1. Inverter;
2. Adder;
3. Subtracter;
4. Multiplier and divider;
5. Integrator;
6. Differentiator;
7. Buffer amplifier;
8. Differential amplifier.

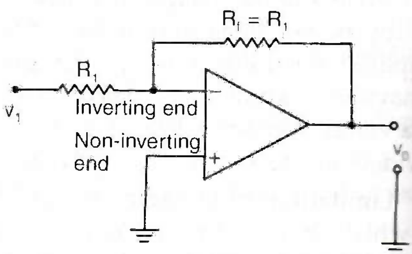


Fig. 4.5. Op-amp as an inverter.

1. Inverter. Fig. 4.5, shows the circuit of an Op-amp used as *inverter.* The feedback resistance R_f is made equal the resistance R_1, connected to the inverting end of the amplifier,

$$\text{Output voltage, } v_o = -\frac{R_f}{R_1} v_1 = -v_1 \qquad \ldots(4.12)$$

$$(\because R_f = R_1)$$

Obviously, the output voltage is 180° out of phase with the input voltage.

2. Adder. Fig. 4.6 shows an Op-amp circuit that performs the signals with amplification (if desired); using superposition theorem, we get

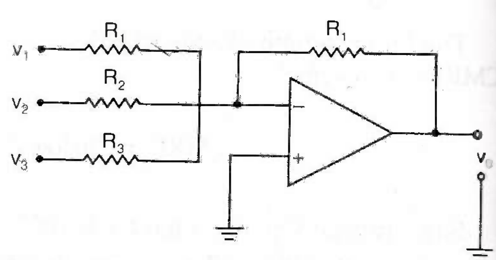


Fig. 4.6. Op-amp as an adder.

Output voltage,

$$v_o = -\left(\frac{R_f}{R_1} v_1 + \frac{R_f}{R_2} v_2 + \frac{R_f}{R_3} v_3\right) \qquad \ldots(4.13)$$

If $R_1 = R_2 = R_3 = R_f$, then

$$v_o = -(v_1 + v_2 + v_3) \qquad \ldots(4.14)$$

i.e., sum of the individual input voltages. The inversion that occurs cannot be avoided.

3. Subtracter. The Op-amp circuit used for subtraction of two input signals is shown in Fig. 4.7. The output of the 2nd Op-amp is given by :

$$v_o = -\left(-v_1 \frac{R_f}{R_1} \cdot \frac{R_f}{R_3} + v_2 \frac{R_f}{R_2}\right) \quad \text{...(4.15)}$$

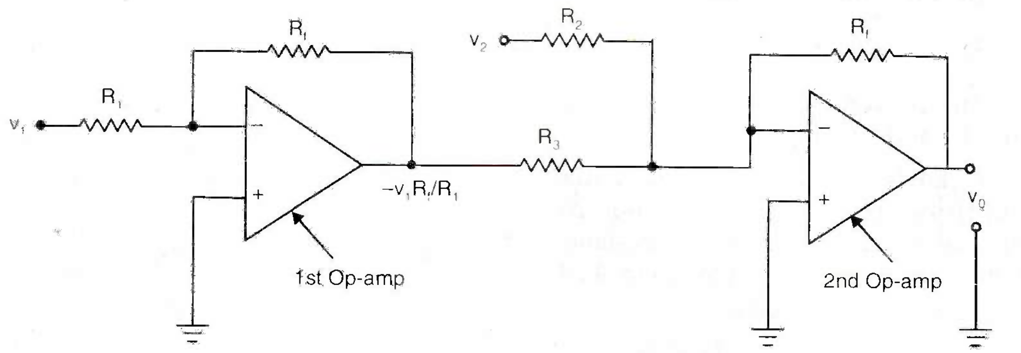


Fig. 4.7. Op-amp as a subtractor.

If $R_f = R_1 = R_2 = R_3$, the circuit acts as a *pure subtractor* and the output, in this case, is given by

$$v_o = v_1 - v_2 \quad \text{...(4.16)}$$

4. Multiplier and divider. The output of an Op-amp in the inverting mode is given by

$$v_o = -\left(\frac{R_f}{R_1}\right) \times v_1$$

In case $R_f > R_1$, the circuit shown in Fig. 4.8 acts as a multiplier and in case $R_f < R_1$ it acts as a divider.

Thus, by choosing the values of R_f and R_1, the multiplier and the divider circuits can be designed for multiplication and division by any number.

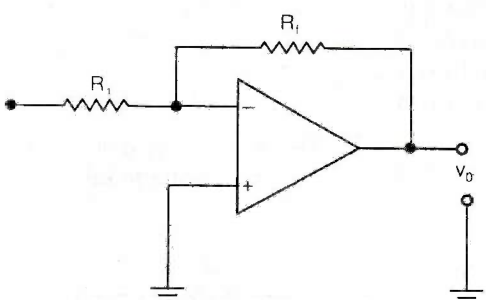


Fig. 4.8. Op-amp as multiplier and divider.

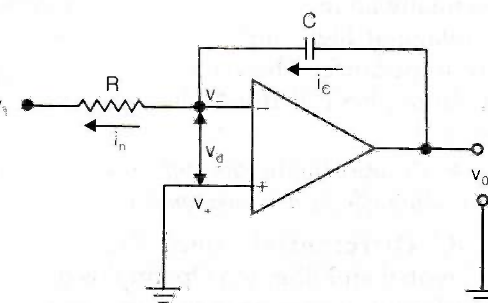


Fig. 4.9. Op-amp as an integrator.

5. Integrator. Fig. 4.9. shows a circuit in which the output voltage is proportional to the integral of the input voltage.

In order to show that the circuit shown in Fig. 4.9. acts as an integrator using KCL (Kirchoff's Current Law) at node v_-, we have,

$$i_n = i_C$$

or,
$$\frac{v_- - v_1}{R} = C\frac{d}{dt}(v_o - v_-)$$

For *infinite differential gains*, $v_- = 0$

$$\therefore \qquad -\frac{v_1}{R} = C\frac{d}{dt}v_o$$

By integration, we have

$$v_o = -\frac{1}{RC}\int v_1\,dt \qquad \text{...(4.17)}$$

The convenient values of R and C are MΩ and μF range respectively.

6. Differentiator. The differential amplifier circuit is obtained by *interchanging* the positions of resistance R and capacitor C as shown in Fig. 4.10.

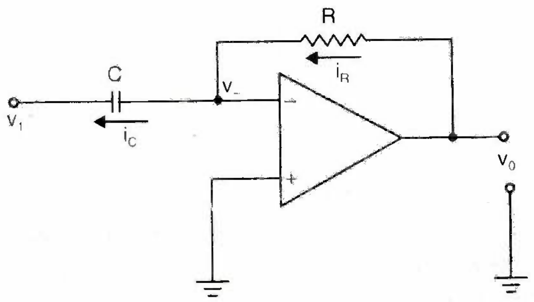


Fig. 4.10. Op-amp as differentiator.

At node v_-, we have :

$$i_C = i_R$$

$$C\frac{d}{dt}(v_- - v_1) = \frac{v_o - v_-}{R}$$

Now, $$v_- = 0$$

$$\therefore \qquad -C\frac{d}{dt}(v_1) = \frac{v_o}{R}$$

or, Output voltage, $$v_o = -RC\frac{d}{dt}(v_1) \qquad \text{...(4.18)}$$

Thus, the output voltage is equal to the differentiated input voltage.

- The *Op-amps are normally used as differentiators as they tend to decrease the signal to noise (S/N) ratio.*

7. Buffer amplifier. The buffer amplifier is essentially an impedance transformer which converts a voltage at high impedance to the same voltage at low impedance. The circuit of a unity gain buffer amplifier also called a *"voltage follower"* is shown in Fig. 4.11.

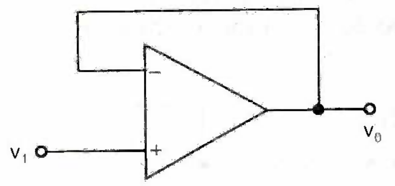


Fig. 4.11. Unity gain buffer amplifier voltage follower.

- *The use of unity gain buffer amplifiers greatly reduces the loading effects in measurements systems.*

8. Differential amplifier. A differential amplifier (an Op-amp) is of significant importance in an instrumentation system.

In its basic form it has two *inputs* and *outputs*. The signals available to the two outputs are identical except that the two are 180° out-of-phase with each other.

The output voltage of the amplifier is proportional to the difference between the two input voltages.

Fig. 4.12, shows an Op-amp used as a differential amplifier.

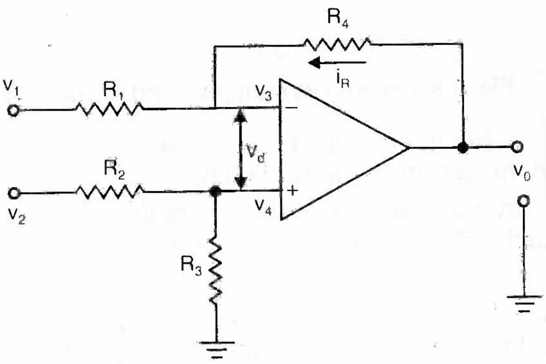


Fig. 4.12. Op-amp used as a differential amplifier.

Here, $v_o = G_d(v_+ - v_-)$

where, G_d = Differential gain.

The signal $v_d = (v_+ - v_-)$ is called *"Difference Mode Signal"* or simply *"Difference Signal"*

$$G_d = \frac{v_o}{v_+ - v_-} = \frac{v_o}{v_d} \qquad ...(4.19)$$

It can be proved that,

$$v_o = G_d(v_2 - v_1) \qquad ...(4.20)$$

When the *two input voltages are equal,* the output voltage is zero. Equal inputs are known as *"Common mode signals"* because the input signal is common to both inputs. However, in actual practice when equal input voltages are applied to the inputs, the *output voltage is not exactly equal to zero (difference is typically of the order of several hundred microvolts) on account of difference in response of the two inputs to common mode signals.*

Common mode gain,

$$G_{cm} = \frac{v_o}{v_{cm}} \qquad ...(4.21)$$

where, G_{cm} = Common mode gain, and

v_{cm} = Common mode input signal.

The *"Common mode rejection ratio (CMRR)"* is defined as

$$CMRR = \frac{G_d}{G_{cm}} \qquad ...(4.22)$$

Also, $$CMRR = 20\log_{10}\left(\frac{G_d}{G_{cm}}\right)\text{dB} \qquad ...(4.22a)$$

... when expressed in *dB*.

Advantages of differential amplifiers :

1. ***Noise immunity:***
   - These amplifiers are extensively used in equipment such as *electronic voltmeters and oscilloscopes.*
2. ***Drift immunity :***
   - The differential amplifier *has inherent capabilities of eliminating problem of drift.*
   - The differential amplifier construction is used for the *early stages of oscilloscope and electronic voltmeter amplifiers, where low drift is extremely important.*

Instrumentation amplifiers. The instrumentation amplifier is a dedicated differential amplifier with *extremely high impedance. The high common mode rejection makes this amplifier very useful in receiving small signals buried in large common-mode offsets and noise.*

These amplifiers consist of *two stages:*

— The *first stage* offers very high input impedance to both input signals and allows to set the gain with a single resistor.

— The *second stage* is a differential amplifier (unity gain) with output, negative feedback and ground connections all throughout.

4.8.7. Attenuators

An ***attenuator*** *is a two-port resistive network and is used to reduce the signal level by a given amount.*

In a number of applications, it is necessary to introduce a specified loss between the source and a matched load without altering the impedance relationship. Attenuators may be used for this purpose.

Attenuator may be *symmetrical* or *asymmetrical*, and can be either *fixed or variable. A fixed attenuator with constant attenuation* is called a **pad**.

- Variable attenuators are used as control volumes in radio broadcasting sections.
- Attenuators are also used in *laboratory to obtain small value of voltage or current for testing circuits.*

The attenuation is expressed in **decibels (dB)** or, in *naper*. The attenuation offered by a network in decibels is given by

$$\text{Attenuation in } dB = 10\log_{10}\left(\frac{P_i}{P_o}\right) \quad ...(4.23)$$

where, P_i is the input power and P_o is the output power.

The attenuators may be of the following types:

1. Resistance attenuator.
2. Symmetrical T-attenuator.
3. L-type attenuator.
4. π-type attenuator.

4.8.8. Filters

Filtering *is the process of attenuating unwanted components of a measurement while permitting the desired component to pass.*

The **filter** *is an electronic circuit which can pass or stop a particular band of frequencies through it.* The filters was first designed by G.A. Campbell and D.Z. Zobel at Bell laboratories.

The band of frequencies which will pass through filter is called the ***pass band*** and the band of all remaining frequencies is called ***attenuation band.*** In case of **ideal filter**, all frequencies of pass band will pass without suffering from any attenuation while the band of all remaining frequencies of attenuation band will be suppressed completely.

Classification of filters:

The filters may be *classified* as follows:

A. On the basis of passing and attenuating of frequencies:

1. **Low pass filters :**
   - These are those filters which pass only low frequencies through them and which reject all high frequencies above the cut-off frequencies.
   - A low pass filter is also called "*lag network*" because it causes a phase lag in the output signal.
   - This type of filter is also called "*integrating network*".
2. **High pass filters :**
   - These are those filters which pass only high frequencies through them and which reject all low frequencies below the cut-off frequency.
   - The high pass filter is a differentiating network and is also called as "*lead network*" because it will cause a phase lead in the output signal.
3. **Band pass filters :**
   - These are those filters which pass a band of frequencies through them and which reject all other frequencies to pass through them.

4. **Band stop filters :**

   - These filters, which are also known as *"band elimination filters"*, are those which reject a band of frequencies to pass through them and which allow the other frequencies to pass though them.

B. On the basis of relation between series and shunt impedances :

1. **Constant filters** (*or prototype filters*). In this filter the series impedance z_1 and stunt impedance z_2 are interrelated by the relation:

$$z_1 z_2 = k^2, \quad \text{...(4.24)}$$

where, k is a constant independent of frequency.

2. ***m*-derived filters.** These filters do not have the product of series and shunt impedances equal to k^2, but have the same characteristic impedance as the corresponding k section, with sharper attenuation characteristic.

   - Fig. 4.13(a) shows some terminology as applied to a low pass filter (Similar terms are applicable to the high pass and notch or band reject filters, respectively) while in the Fig. 4.13(b) are shown the band-pass filter characteristics.

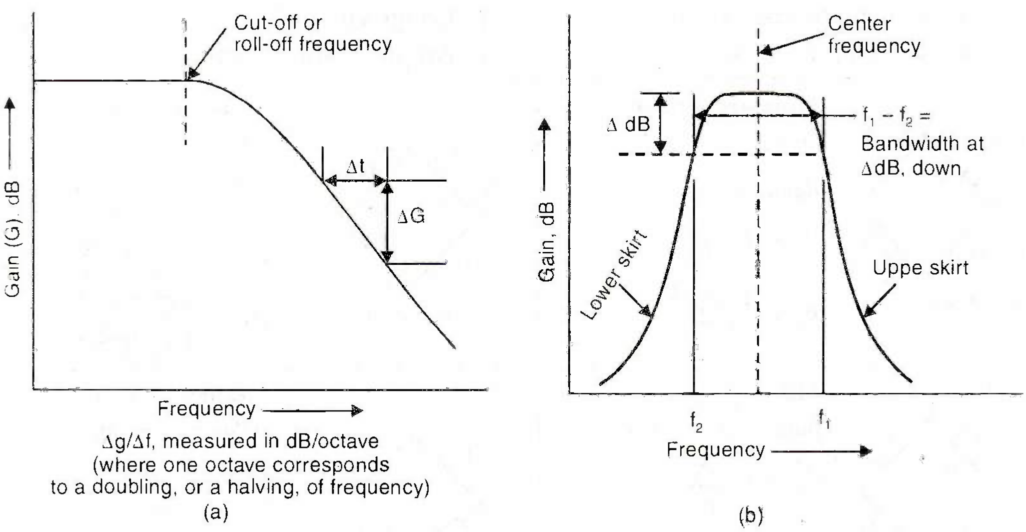
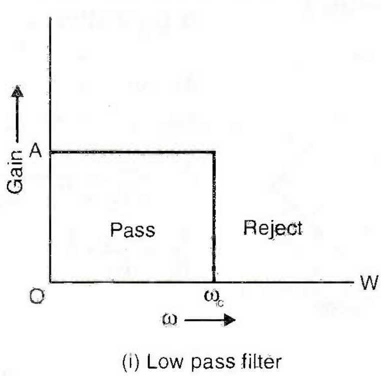
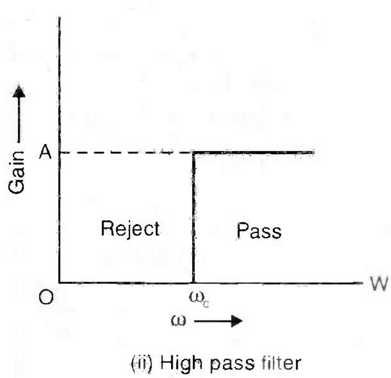


Fig. 4.13

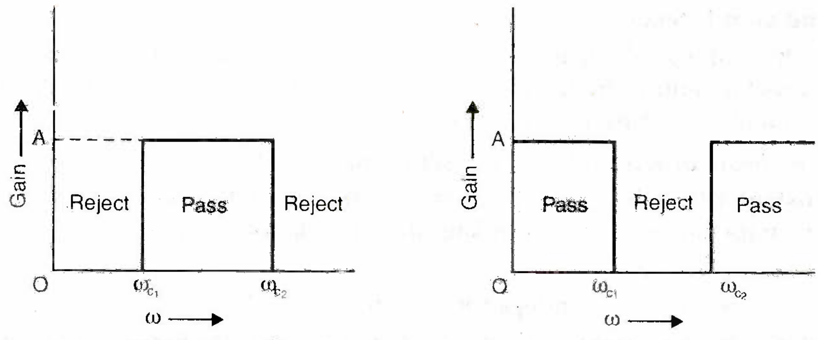
- Fig. 4.14 shows the ideal characteristics of filters.



(i) Low pass filter



(ii) High pass filter



(iii) Band pass filter (iv) Band stop filter

Fig. 4.14. Ideal characteristics of filters.

4.8.9. Input Circuitry

Although not all-inclusive, the following types of input circuits are used for signal conditioning of electrical transducers :

1. Simple current-sensitive circuits.
2. Ballast circuits.
3. Voltage-dividing circuits.
4. Bridge circuits.
5. Resonant circuits.
6. Amplifier input circuits.

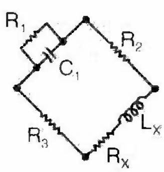
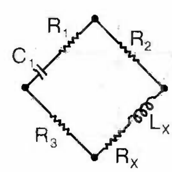
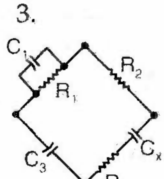
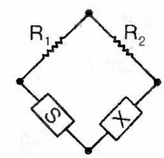
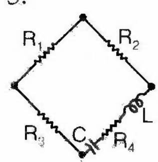
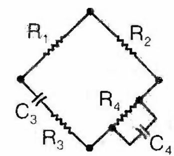
1.  Maxwell circuit	*Measures L (best for $\omega L_x/R_x < 10$)* *Balance equations :* $L_x = R_2R_3C_1$ $R_x = \frac{R_2R_3}{R_1}$	2.  Hay circuit	*Measures L (best of $\omega L_x/R_x > 10$)* *Balance equations :* $L_x = \frac{R_2R_3C_1}{1+\omega^2C_1^2R_1^2}$ $R_x = \frac{\omega^2C_1^2R_1R_2R_3}{1+\omega^2C_1^2R_1^2}$
3.  Schering circuit	*Measures C* *Balance equations :* $C_x = C_3\frac{R_1}{R_2}$ $R_x = R_2\frac{C_1}{C_3}$	4.  Comparison with series constants	*Meausres L or C* *Balance equations :* $R_x = R_s\frac{R_1}{R_2}$ If inductive, $L_x = L_s\frac{R_2}{R_1}$ If capacitive, $C_x = C_s\frac{R_1}{R_2}$
5.  Resonant circuit	*Measures L or C (f known), f (L and C known)* *Balance equations :* $X_L = X_C$ or $L.\ C = \frac{1}{\omega^2}$ $f = \frac{1}{2\pi\sqrt{LC}}$	6.  Wien or RC frequency bridge	*Measures f* *Balance equations :* $f = \frac{1}{2\pi\sqrt{R_3R_4C_3C_4}}$ $\frac{R_1}{R_2} = \frac{R_3}{R_4} + \frac{C_4}{C_3}$

Fig. 4.15. Impedance bridge arrangements.

4.9 DATA ACQUISITION

4.9.1. Introduction

Now-a-days, in mechatronic and measurement systems, microprocessors, microcontrollers, single-board computers, and personal computers are widely used. As such, it is increasingly important for engineers to understand how to directly access information and analog data from the surrounding environment with these devices.

Consider a signal from a sensor as illustrated by the analog signal in Fig. 4.16. In this case there are *two options* :

- Firstly, one could record the signal with an analog device such as chart recorder (which physically plots the signal on the paper) or display *it with an oscilloscope.*
- Secondly, *the data may be stored by using a microprocessor or computer. This process is called computer* ***"data acquisition"*** and entails the following *merits*:
  (*i*) Can result in greater data accuracy;
  (*ii*) Provides more compact storage of the data;
  (*iii*) Enables data processing long after the occurrence of the events;
  (*iv*) Allows use of the data in real time control system.

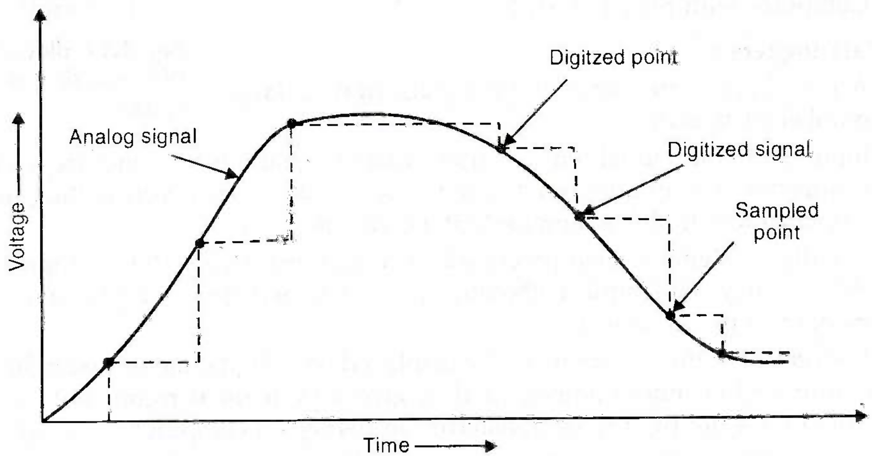


Fig. 4.16. Analog signal and sampled equivalent.

In order to input analog data to a digital circuit or microprocessor, the analog data *must be transformed into digital valves.* The first step is to *numerically evaluate the signal at discrete instants in time. This process is called* ***"sampling"***, and the result is *"digitized signal"* composed of discrete values corresponding to each sample (See Fig. 4.16).

4.9.2. Data Acquisition (DAQ) Systems

Data Acquisition *is the process of using output signals and inputting that into a computer.* The output signal may be one that originates from direct measurement of electrical quantities such as voltage, frequency, resistance etc. or that originates from sensors.

Data acquisition systems are of the following *two* types:

(*i*) Analog data acquisition system.

(*ii*) Digital data acquisition system.

Fig. 4.17, shows the block diagram of elements of analog data acquisition system:

— This system consists of a sensor-transducer the output of which is connected to DAC board (this is a PCB) through a signal conditioning unit. The DAC board is plugged to a computer. The DAC board consists of a *multiplexer, amplifier, ADC , register* and *control circuitry*, the output of control circuitry connected to a computer system.

— A software is employed to control the acquisition of data through DAC. When the program requires input from a particular sensor, it activates the DAC board by sending control word to the control and status register. The control word indicates what type of operation the board has to carry out.

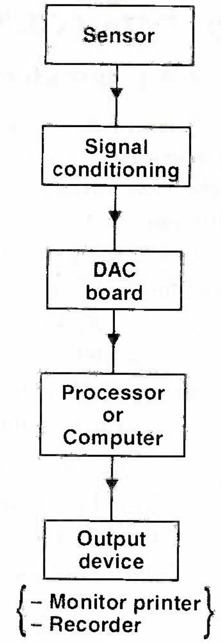


Fig. 4.17. Block diagram of analog data acquisition system.

Automated data acquisition systems may take the following forms:

1. Data loggers;
2. Computer with plug-in boards.

1. Data loggers :

— A *data logger* can monitor the inputs from a large number of sensors.

— Inputs from individual sensors, after suitable signal conditioning, are fed into the *multiplexer*. The multiplexer is used *to select one signal* which is then fed, after amplification, to the analog-to-digital converter.

— The digital signal is then processed by a microprocessor. The microprocessor is able to carry out simple arithmetic operations, perhaps taking the average of a number of measurements.

— The output of the system might be displayed on a digital meter that indicates the output and channel number, used to give a permanent record with a printer, stored on a floppy disc or transferred to perhaps a computer for analysis.

- As data loggers are often used with *thermocouples*, there are often special inputs for thermocouples, these providing cold junction compensation and linearisation. The multiplexer can be switched to each sensor in turn and so the output consists of a sequence of samples. Scanning of the inputs can be selected by programming the microprocessor to switch the multiplexer to just sample a single channel, carry out a single scan of all channels, a continuous scan of all channels, or perhaps carry out a periodic span of all channels.

2. Computer with plug-in boards :

Fig. 4.18, shows the basic elements of a data acquisition system using plug-in boards with a computer.

The signal conditioning prior to the inputs to the board depends on the sensors concerned.

Examples:

(*i*) *Thermocouples* — Amplification, cold junction compensation and linearisation;
(*ii*) *Strain gauges* — Wheatstone bridge, voltage supply for bridge and linearisation;
(*iii*) *RTDs* — Current supply, circuitry and linearisation.

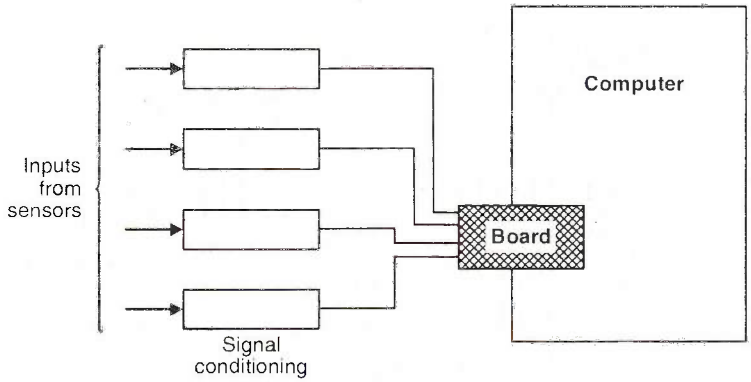


Fig. 4.18. Data acquisition system.

4.9.3. Analog-to-Digital Conversion (ADC)

4.9.3.1. Digital signals

The majority of sensors supply the output which tends to be in analog form. Where a *microprocessor* is used as part of the measurement or control system, the analog output from the sensor *has to be converted into a digital form* before it can be used as an input to the microprocessor. Similarly, most actuators operate with analog inputs and so the digital output from a microprocessor has to be converted into an analog form before it can be used as input by the actuator.

4.9.3.2. ADC process

The *"analog-to-digital conversion" process* changes a sampled analog voltage into digital form. This process, conceptually involves the following *two steps:*

(*i*) ***Quantizing.*** It is defined as the *transformation of a continuous analog input into a set of discrete output states.*

(*ii*) ***Coding.*** It is *assignment of a digital code word or number to each output state.*

Procedure of conversion:

Analog-to-digital conversion involves converting analog signals into binary words. Fig. 4.19 shows the basic elements of analog-to-digital conversion. The procedure of A/D conversion is that a *clock supplies regular time signal pulses to ADC (analog-to-digital converter) and every time it receives a pulse it samples the analog signal.* The types of signals involved at various stages of analog-to-digital conversion are shown in Fig. 4.20:

- Fig. 4.20(*a*), shows the analog signal;
- Fig. 4.20(*b*), shows the clock signals which supply the time signals at which the sampling occurs;
- Fig. 4.20(*c*), shows a series of various pulses which is the result of sampling (sampled signal);
- Fig. 4.20(*d*), shows the sampled and held signal which is obtained by using a *sample and hold unit.* (This unit is necessary because A/D converter requires a finite amount of time, termed the *'conversion time'*, to convert analog signal into a digital one) which holds each sampled value until the next pulse occurs.

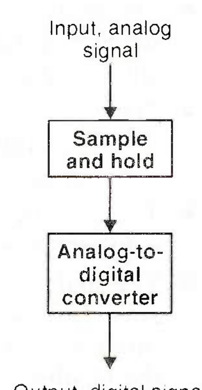


Fig. 4.19. Basic elements of A/D conversion.

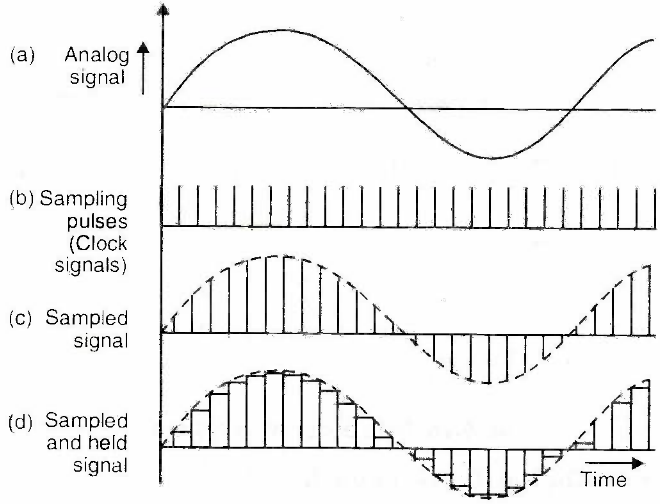


Fig. 4.20. Signals : (*a*) Analog; (*b*) Clock; (*c*) Sampled; (*d*) Sampled and held.

4.9.3.3. Components used in A/D conversion

In order to acquire an analog voltage for digital processing, it is imperative to properly select the following components and apply them this sequence:

(*i*) Buffer amplifier; (*ii*) Low-pass filter;
(*iii*) Sample and hold amplifier; (*iv*) Analog-to-digital (A/D) converter;
(*v*) Computer.

Figure 4.21, shows the components used in A/D conversion:

- The *buffer amplifier* provides a signal in a range close to but not exceeding the full input voltage range of the A/D converter.
- The *low-pass filter* is necessary to remove any undesirable high-frequency components in the signal that could produce aliasing. The out-off frequency of the low-pass filter should not be greater than half the sampling rate.
- The *sample and hold amplifier* maintains a fixed input value (from an instantaneous sample) during the short conversion time of the A/D converter.
- The *A/D converter* should have a resolution and analog quantization size appropriate to the system and signal.
- The *computer* must be properly interfaced to A/D converter system to store and process the data.
- The analog-to-digital conversion process requires a small but finite interval of time that must be taken into consideration when assessing the accuracy of the results.
- The *conversion time depends on the design of the*

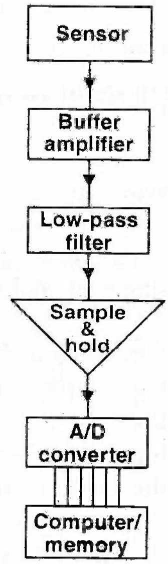


Fig. 4.21. Components used in A/D conversion

converter, the method used for conversion, and the speed of the components used in the electronic design.

— Owing to the continuous change in the analog signals, the uncertainty about when in the sample time window the conversion occurs causes corresponding uncertainty in the digital value. This is of significant importance if there is no sample and hold amplifier on the A/D input. The term *"aperture time"* refers to the duration of the time window and is associated with any error in the digital output due to changes in the input during this time.

— In a signal, sampling at or about Nyquist frequency will yield the correct frequency components. In order to obtain accurate amplitude resolution, we must have an A/D converter with an adequately small aperture time.

4.9.3.4. Analog-to-digital (A/D) converter

An ***analog-to-digital (A/D)*** *converter is an electronic device that converts an analog voltage to a digital code.* The output of the A/D converter can be directly interfaced to digital devices such as microcontroller and computers.

The ***"resolution"*** *of an A/D converter is the number of bits used to digitally approximate the analog value of the input.* The number of possible states N is equal to the number of bit combinations that can be output from the converter : $N = 2^n$

where, n = the number of bits.

The number of analog *"decision points"* that occur in the process of quantizing is $(N - 1)$. The *"analog quantization size"* Q is defined as the *full scale range of the A/D converter by the number of output states.*

Design principles:

Analog-to-Digital (A/D) converters *are designed based on a number of different principles; these are:*

(*i*) Successive approximations.

(*ii*) Flash or parallel encoding.

(*iii*) Single-slope and dual-slope integration.

(*iv*) Switched capacitor.

(*v*) Delta sigma.

Some of these are discussed below:

(i) Successive approximation A/D converter :

The successive approximation A/D converter is very widely used because :

- It is fast in operation;
- It has high resolution;
- It is less expensive.

The various subsystems involved in this type of converter are shown in Fig. 4.22.

— The *"clock"* generates a voltage, emitting a regular sequence of pulses which are counted, in a *binary manner*, and the resulting binary word is converted into an analog voltage by a *"DAC"* (*digital-to-analog converter*). This voltage rises in steps and is *compared with the 'analog input voltage'* from the sensor.

— When the clock-generated voltage passes the input analog voltage the pulses from the clock are stopped from being counted by a *"gate"* being closed. The output from the counter at that time is then a digital representation of the analog voltage.

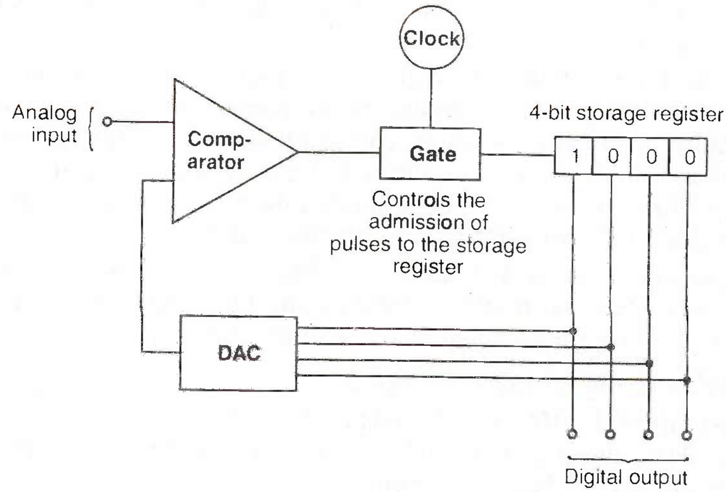


Fig. 4.22. Successive approximations analog-to-digital converter (ADC)

Note: When frequency of the clock is f, the time taken between the pulses is $\frac{1}{f}$; hence the time taken to generate the word, *i.e.*, the conversion time is $\frac{n}{f}$.

(ii) Flash A/D converter :

The fastest type of A/D converter is known as a ***flash converter.***

Fig. 4.23, shows a flash ADC:

— For an n-bit converter, 2^n-1 separate voltage comparators are used in parallel, with each having the analog input voltage as one input.

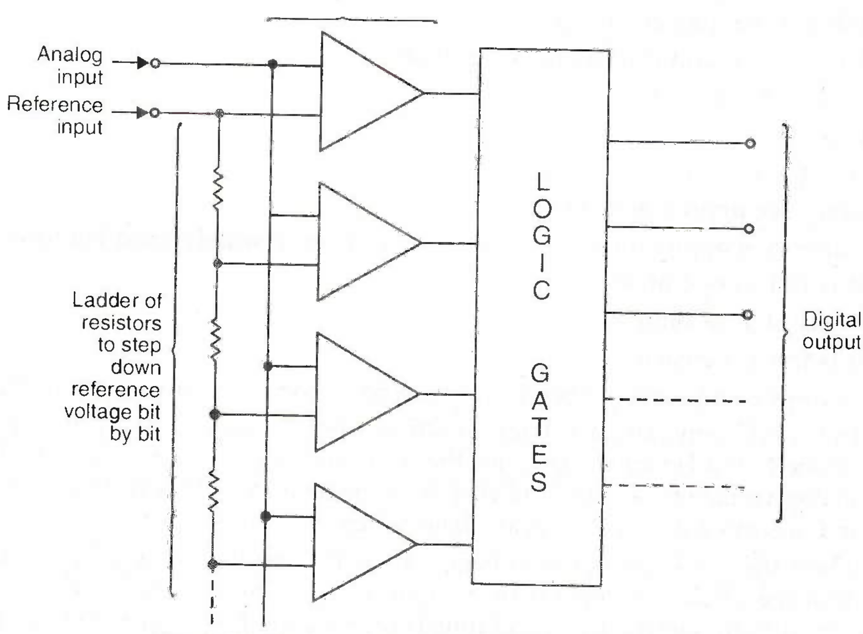


Fig. 4.23. Flash analog-to-digital converter (ADC)

— A *reference voltage* is applied to a ladder of resistors so that the voltage applied as the other input to each comparator is one bit *larger* in size than the voltage applied to the previous comparator in the ladder. Thus when the analog voltage is applied to the ADC, all these comparators for which the analog voltage is greater than the reference voltage of a comparator will give a *high* output and those for which it is less will be *low.*

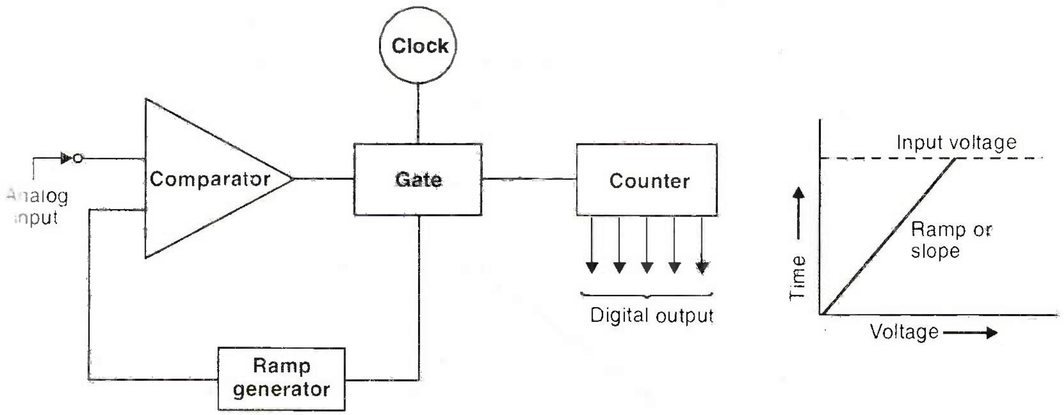
— The resulting outputs are fed in parallel to a *logic system* which translates into a digital word.

(iii) Single-slope and dual-slope integration :

(a) Single-slope or ramp or voltage-to-time A/D converter :

Fig. 4.24 shows the schematic of a ramp ADC:

— A ramp analog-to-digital converter (ADC) involves an analog voltage which is increased at constant rate (and hence called *ramp* voltage) and is applied to a comparator where it is compared with the analog voltage from the sensor. The time consumed by the ramp voltage to increase to the value of the sensor voltage will depend on the size of the sampled analog voltage.



(a) Ramp ADC circuit

(b) Graphical representation

Fig. 4.24. Single slope or ramp ADC.

— When the ramp voltage starts, a gate is opened which starts a binary counter counting the regular pulses from a clock. When the two voltages are equal, the gate closes and the word indicated by the counter is the digital representation of the sampled analog voltage.

(b) Dual-slope integration AD/converter or dual ramp converter :

This type of converter, as compared to single ramp converter, is more commonly used.

Fig. 4.25, shows the dual slope/dual ramp *ADC*. The analog input voltage and the reference input voltage are successively connected to the integrator with the help of a switch. These two voltages (analog input and reference input) *must be of opposite polarities.*

— The fixed voltage is integrated for a fixed sample time.

— The integrated value is then discharged at a fixed rate and the time to do this is measured by a counter. The count is then a measure of the analog input voltage.

- The "*advantage*" of these converters is that they have *excellent noise rejection* because the integral action averages out random negative and positive contributions over the sampling period.
- Their "*limitation*" is that they are *very slow* in operation.

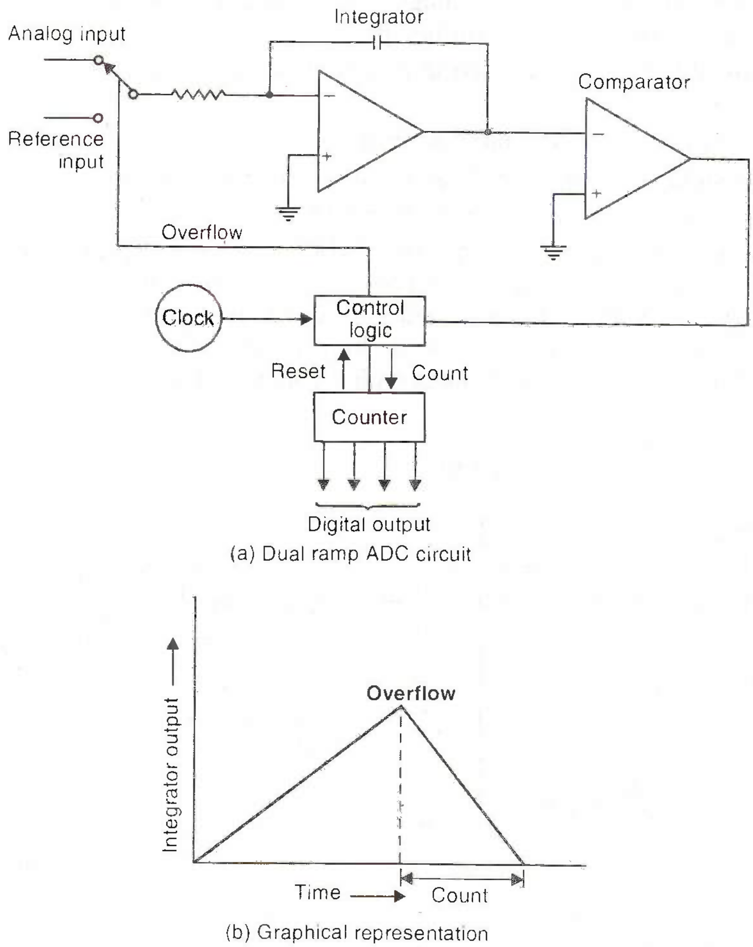


Fig. 4.25. Dual slope/dual ramp A/D converter.

Multiplexers:

*The "**multiplexer**" is essentially an electronic switching device which enables each of the inputs to be sampled in turn.*

A "*multiplexer*" is a circuit that is able to have inputs of data from a number of sources and then, by selecting an input channel, give an output from just one of them.

- In *applications* where there is a need for measurement to be made at a number of different locations, rather than use a separate ADC and microprocessor of each measurement, a "*multiplexer*" can be used *to select each input in turn and switch it through a single ADC and microprocessor* (Fig. 4.26).

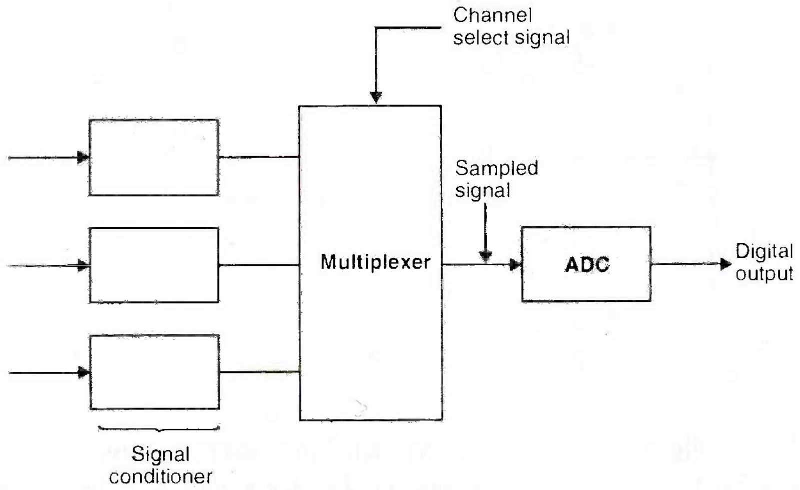


Fig. 4.26. Multiplexer.

Digital multiplexer :

Fig. 4.27 shows a two channel multiplexer. The logic level applied to the select input determines which AND gate is enabled so that its data input passes through the *OR* gate to the output.

A number of forms of multiplexers are available in *IC* packages.

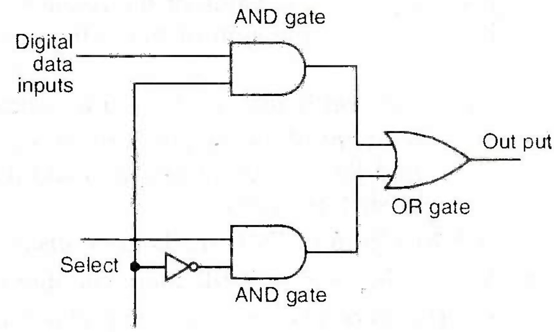


Fig. 4.27. Two channel multiplexer.

- ***"Demultiplexer"*** is similar to multiplexer but with *reversed action*. It accepts a digital signal through its one input and then channelises it to a particular output selected by binary value at the control port.

4.9.4. Digital-to-Analog (D/A) Conversion

Invariably we have to reverse the process of analog-to-digital (A/D) conversion by changing a digital value to an analog value. This is called ***digital-to-analog (D/A) conversion.***

A D/A converter (DAC) *allows a computer or other digital device to interface with external analog circuits and devices.*

- The input to a digital-to-analog converter (DAC) is a *binary word;* the output is an analog signal that represents the weighted sum of the non-zero bits represented by the word.

  Example: An output of 0010 must give an analog output which is twice that given by an input of 0001.

Digital-to-analog converters :

— Figure 4.28, shows a simple form of DAC using a summing amplifier to form the weighted sum of all the non-zero bits in the input word.

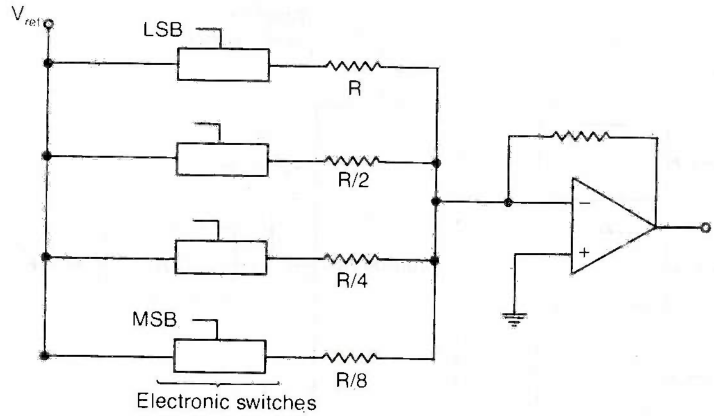


Fig. 4.28. Weighted resistor digital-to-analog converter.

— The reference voltage (V_{ref}) is connected to the resistors by means of electronic switches which respond to binary 1.

— The values of input resistances depend on which bit in the word a switch is responding to, the value of the resistor for successive bits from the LSB being halved. Hence the sum of the voltages is a weighted sum of the digits in the work.

- Such a system is referred to as a *weighted-resistor network.*
- The limitations of the weighted-resistor network is that accurate resistances have to be used for each of the resistors and it is difficult to obtain the required wide range of such resistors.

As such this form of DAC tends to be *limited to 4-bit-conversions.*

- *R-2R ladder network* is the more commonly used version (Fig. 4.29).
  — This version overcomes the problem of obtaining accurate resistances over a wide range of values, *only two values being required.*
  — The output voltage is generated by switching sections of the ladder to either the reference voltage or 0 V according to whether there is a 1 or 0 in the digital input.

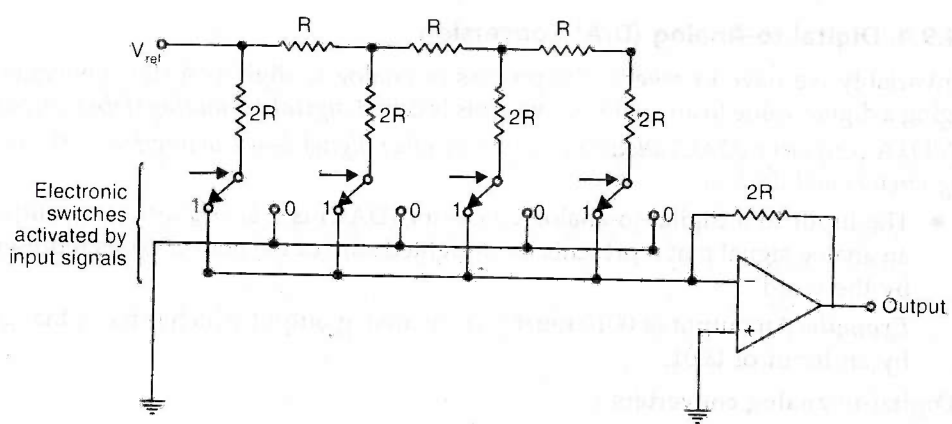


Fig. 4.29. R-2R ladder digital-to-analog converter.

- Fig. 4.30 shows computer control hardware, illustrating the *roles that A/D and D/A converters play in a* ***mechatronic control system.***
  — An analog voltage signal from a *sensor* (*e.g.*, a thermocouple) is converted to a digital value.
  — The *computer* uses this value in a control algorithm, and the computer outputs an analog signal to an *actuator* (*e.g.*, an electric motor) to cause some change in the system being controlled.

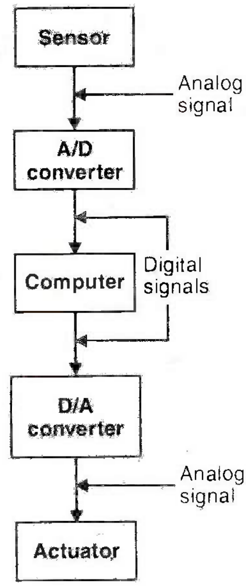


Fig. 4.30. Computer control hardware.

Pulse-Modulation:

While dealing with the transmission of low-level D.C. signals from sensors, a problem that is encountered is that the gain of Op-amp (operational amplifier) used to amplify them may drift causing a drift in the output. This problem can be solved if the signal is a *sequence of pulses rather than a continuous-time signal.* This can be achieved in the following *two* ways:

1. Pulse amplitude modulation (PAM).
2. Pulse width modulation (PWM).

1. Pulse amplitude modulation:

— In this method of conversion, D.C. signal (Fig. 4.31(*a*)] is chopped in the way as shown in Fig. 4.31(*b*). The output from the chopper is a chain of pulses, the heights of which depends on the D.C. level of the input signal. This process is called *"pulse amplitude modulation"*.

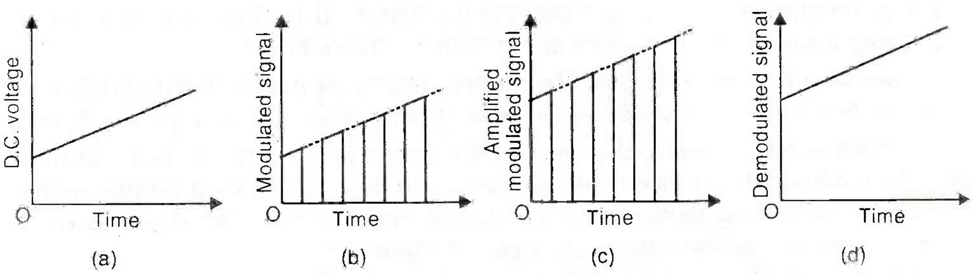


Fig. 4.31. Pulse amplitude modulation.

— After amplification and any other signal conditioning, the modulated signal [Fig. 4.31(*c*)] can be demodulated [Fig. 4.31(*d*)] to give a D.C. output.

2. Pulse width modulation (PWM) :

This type of modulation is used where the width, *i.e., duration of a pulse rather than its amplitude* depends on the size of the voltage, as shown in Fig. 4.32.

— PWM is widely used with control systems as a means of controlling the average value of a D.C. voltage.

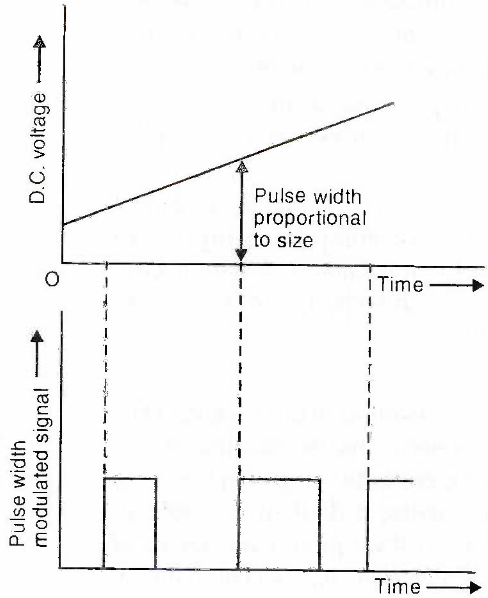


Fig. 4.32. Pulse width modulation (PWM).

4.10 DATA SIGNAL TRANSMISSION

The terms *"measuring devices"* and *"transmitters"* generally go side by side and it is very difficult to make any distinction between them.

A measuring device converts a primary indication into some form of energy that can easily be displayed on a scale; some transmitters also do the same things. In the *stricter sense* **"transmitters"** *could be considered as devices which transmit the value of the primary variable at a considerable distance from the primary element.* If transmission is to be carried over *very long distances*, then devices are known as **"telemeters"**.

The terms **data transmission** *and* **"telemetry"** *refer to the process by which the measurand is transferred to a remote location for the purpose of being processed, recorded and displayed.*

For transmission purposes, the measured variable is converted into a transmittable signal (either pneumatic or electrical), so that it can be received by a remote indicating, recording, or controlling device. The *selection of transmission device* depends upon the *nature of the variable and the distance the signal is required to be sent.*

For data transmission various methods have been developed; the choice of a particular method depends upon :

(*i*) The physical variable;

(*ii*) The distance involved.

- The **hydraulic and pneumatic methods** are employed for transmission over a *short distance.*
  — The pneumatic type transmission devices are generally suitable for transmission upto maximum distance of 200 m.
- The **electrical/electronic methods** are suitable *equally for short as well as long distance transmission.*
  — Generally short transmission is carried out on own communication connections between sending and receiving devices.

— The *telemeters* which are designed for *long-distance transmission* may be designed to transmit over their own wires or over phone wires or by microwave.

4.10.1. Mechanical Transmission

The *"rack and pinion arrangement"* and the *"gear trains"* as used in *Bourdon-tube pressure* gauge and *dial indicator gauge* constitute mechanical transmission. They *amplify the displacement* and also *transmit the signal to a pointer which moves across a calibrated dial.*

4.10.2. Hydraulic Transmission

Fig. 4.33 shows the hydraulic method of transmission, which is commonly used. Here four bellows are employed, two at the transmission end and two at the receiving end. The four bellows are connected by an impulse pipeline and the whole system is filled with liquid. When the actuating link, on the transmission end, is operated by the measurand, then one bellow is expanded and other is contracted. This expansion and contraction is communicated to receiving end, which moves the receiving pointer an equal amount. The purpose of using two bellows on either side is to compensate for changes in ambient temperature.

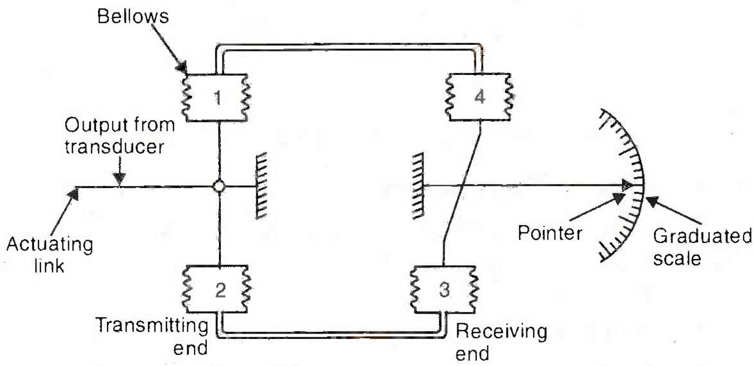


Fig. 4.33. Hydraulic method of transmission.

4.10.3. Pneumatic Transmission

Fig. 4.34 shows the one of the pneumatic methods of transmission **(Flapper nozzle mechanism).**

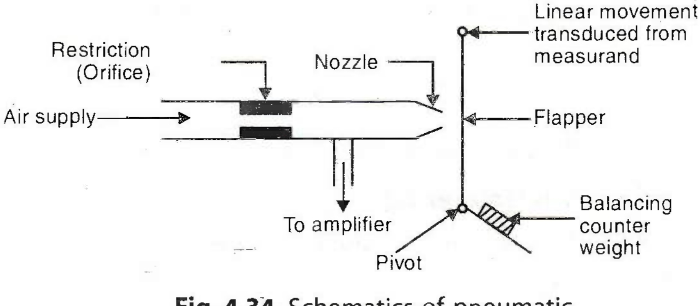


Fig. 4.34. Schematics of pneumatic transmission–Flapper nozzle mechanism.

It consists of an open nozzle which is supplied with air through a restriction/orifice (its diameter being *smaller* than nozzle diameter for proper functioning). In front of the nozzle there is a flapper which is positioned by the measuring element. The force on the flapper is produced by a transducer which converts the measurand into linear displacement. The flapper is pivoted about a point and at the other end, it contains some balancing counter weight.

When the flapper is moved against the nozzle the air cannot escape and maximum air passes to the amplifier, and when flapper is moved away from the nozzle, minimum air passes to the amplifier as most of the air escapes to atmosphere. Thus, the movement of flapper from one extreme position to another serves to control the amplifier, which produces an air pressure proportional to the measurand of adequate strength for transmission over the required distance.

4.10.4. Magnetic Transmission

Fig. 4.35 shows the schematics of magnetic transmission. In this arrangement/device, an armature is attached at the end of the mechanical moving part whose movement is to be transmitted outside the armature moving inside a non-magnetic tube. A magnet is placed around the armature outside the tube. The magnet follows the movement of the armature and repositions a pneumatic transmitter. The magnet movement could also be utilised to operate an electronic transmitter.

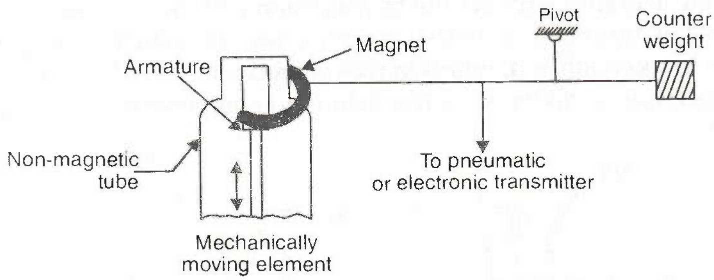


Fig. 4.35. Schematics of magnetic transmission.

4.10.5. Electric Type of Transmitters

Most of the electric type of transmitters employ A.C. bridge circuits in which degree of coupling between inductances is varied by changing the amount of iron core within a coil. The common *examples* are:

1. Wheatstone bridge transmitter.
2. Inductance bridge.
3. Impedance bridge.
4. Differential transformer.
5. Self synchronous motor (Selsyn)
6. Resistance manometers.

4.10.6. Converters

The converters are series of transducers which play an important role in the modern instrumentation, linking electrical (voltage and current based) and pneumatic control systems together.

Following are the most commonly used converters :

1. Current-to-pneumatic converters.
2. Pneumatic-to-current converters.
3. Voltage-to-current converters.
4. Voltage-to pneumatic converters.

4.10.7. Telemetering

According to the primary measurement involved, the telemetering system can be *classified* as follows:

1. Voltage telemetering.
2. Current telemetering.
3. Position or ratio telemetering.
4. Impulse telemetering.
5. Frequency telemetering.

1. Voltage telemetering :

- In these systems the measurand is converted to A.C. or D.C. voltage.

- For such systems, the self-balancing potentiometers are the usual receivers.
- These systems are affected by line resistance, leakage, interfering sources nearly, noise and require higher-quality circuits than current systems, especially for low voltages.

The voltage telemetering system is *limited for transmission upto 300 meters distances.*

2. Current telemetering :

- This system is also *not suitable for long distances* since the current output is varied by means of an adjustable resistance in the line.

Advantages:

(*i*) The current systems can develop higher voltages than most voltage systems and, consequently, it can be made more immune to the effect of thermal and inductance voltages in the interconnecting leads as well as line resistance.

(*ii*) Simple D.C. milliammeters can be used with special calibration for line resistance.

(*iii*) Several receivers can be operated simultaneously.

(*iv*) The received signals can be added or subtracted directly.

(*v*) Changes in line resistance are compensated by basic feedback method.

(*vi*) The response of the system to an input change is almost instantaneous.

(*vii*) The energy level is adequately high to minimise the effects of extraneous voltages.

3. Position or ratio telemetering :

The *synchromotor* (selsyn) *telemetering system* is the most common example of this category. Another example being the *inductance bridge.*

In this system *angular input displacement is converted into relative magnitude of three phase A.C. voltages.*

Advantages:

(*i*) Require no intermediate amplifiers or conversions.

(*ii*) Relatively inexpensive.

(*iii*) Minimum moving parts, so the maintenance is low.

(*iv*) Instantaneous response.

(*v*) Power taken for their operation directly from the line.

Limitation : These systems are *affected by excessive line resistance.*

4. Impulse telemetering :

An impulse telemetering may be *used over extreme distance by operating a carrier or radio transmitter.*

The four typical systems commonly used in impulse telemetering are :

(*i*) Impulse amplitude.

(*ii*) Impulse spacing.

(*iii*) Impulse duration.

(*iv*) Impulse rate.

These systems have the advantages of giving *accuracy independent of supply-voltage variations.*

5. Frequency telemetering :

In frequency telemetering, the *frequency of an A.C. signal is varied in accordance with the measured quantity.*

4.11 DATA PRESENTATION/DISPLAY

4.11.1. General Aspects

The main purpose of any measurement system is to provide information concerning the state and condition of the physical phenomenon being investigated. The measuring systems may be activated either directly from the measuring means (*e.g.* bellows, pressure spiral etc., to which indicating pointer is attached directly through level and leverage system) or by means of a servo-operated system (null-balance system which incorporates a feedback circuit in a closed loop). The last stage of a measurement system is the *data presentation stage*; if the results of the system are meaningful they must be displayed for *instant observation by a display device* or *for storage for observation at a later stage by a recorder* (The data presentation devices may be called *"output devices"*).

The following factors decide about the *choice between the display devices and recorders* :

(*i*) The information content of the output.

(*ii*) The expected use of the output.

The output devices may be *categorized* as follows:

1. Single number output devices.
2. Time domain output devices.

1. Single number output devices:

Such devices indicate the value of some particular quantity under condition such that the value to be measured can be regarded as time variant over the time interval during which measurements are made; thus a *single number will represent measurement.*

"Indicating instruments" and *digital display units* belong to this class.

2. Time domain output devices:

The indicating instruments or the digital display units (suitable only when the *output varies at a very slow rate*) do not serve the purpose *when the values of the quantity are to be taken as a function of time.*

- *For fast changing outputs* (where signal waveform or shape is the desired information) .... *"Cathode ray oscilloscope (CRO)* is used.
- For keeping a *permanent record of the variation of the output with time* .... *"Cathode ray tube photographs"*, *direct writing recorders "strips chart recorders"*, *magnetic taperecorders* etc. are used.

The machine interpretable outputs can be had from:

(*i*) Magnetic tapes;

(*ii*) Punched paper tapes;

(*iii*) Punched cards;

(*iv*) Pulsed signals.

- The information available from an instrument may take the following forms:
  (*i*) *Quantitative information* (*e.g.*, angular spread in r.p.m.; force in newtons).
  (*ii*) *Qualitative information* (*e.g.*, the approximate value or direction of change of some variable, a check reading).
  (*iii*) *Status informations* (*e.g.*, On/Off, in/out).
  (*iv*) *Alphanumeric and symbolic information* (*e.g.*, the labels and instructions; letters *A* to *Z*, the numerals 0 to 9, punctuation marks and various other simple symbols can be generated and interpreted).

- A *good display, functionally, is one which permits the best combination of speed accuracy and sensitivity when transferring* the *necessary information from the instrument to the operator.*

4.11.2. Electrical Indicating Instruments

Qualitatively the electrical indicating instruments are widely used for measurement of current, voltage, resistance and power. These instruments can be *classified* as follows :

1. Analog instruments.
2. Digital instruments.

Analog type of meters use scale and needle (pointer) type arrangement to display the value of the measured parameter. *The measured parameter is converted into electrical signal which further actuates some electro-mechanical device to move the pointer continuously as the parameter changes.*

Digital types meters *indicate the reading in exact numerals.*

Table 4.1 shows the comparison between analog and digital type instruments.

Table 4.1. Comparison between Analog and Digital type Instruments

S. No.	Aspects	Analog type instruments	Digital type instruments
1.	*Information form*	As the position of a pointer against a calibrated scale or dial.	As a number.
2.	*Possibility of human error*	Exists	Does not exist.
3.	*Best possible accuracy*	± 0.25%	± 0.005% or better.
4.	*Resolution*	One in several hundreds.	One part in several hundred thousands.
5.	*Presence of moving parts*	Moving parts involved.	These instruments can be made without moving parts.
6.	*Construction*	Simple in construction and direct reading type; can perform under favourable conditions.	Since these instruments involve electronics, proper environmental conditions are essential.
7.	*Rate of change of parameter*	These instruments enable the operator to judge the rate of change of parameter by seeing the needle movement.	Change of digital reading does not give any knowledge of rate of change of parameter.
8.	*Time required to observe the reading*	If exact reading is required operator takes more time as he has to guess the approximate tenths of small division.	Reading of digital meters is very fast.
9.	*Auxiliary power requirement*	These instruments require no auxiliary source of power for actuation but derive driving power for indicating system from the process.	These instruments require auxiliary source of power.

S. No.	*Aspects*	*Analog type instruments*	*Digital type instruments*
10.	*Mobility* *Examples :*	Can be portable also. Examples of pointer-dial output devices are: ● Micrometer and platform scales; ● Manometers and Bourdon-tube pressure gauge; ● Mercury in glass and filled system thermo-meters; ● Speedometer of an automobile; ● Common voltmeters and ammeters etc.	Usually stationary type. Examples of Digital output devices are: ● Digital ammeters and voltmeters; ● Electronic and mechanical counters; ● Odometers; ● Yes-No light (On or Off); ● Time on a scoreboard etc.

Pointer-scale analog indicators:

In analog instruments the value of the measured parameter is indicated by positioning of the indicating pointer again a calibrated scale. This purpose can be achieved either by moving the pointer with relation to a stationary scale (fixed-scale moving-pointer indicators or the scale may be moved with relation to a fixed reference (moving-scale fixed-pointer indicators).

1. Single-point indicators:

The *fixed-scale and movable-pointer indicators*, available in a variety of forms, are shown in Fig. 4.36. Figure 4.37 shows the fixed-pointer and movable-scale indicators.

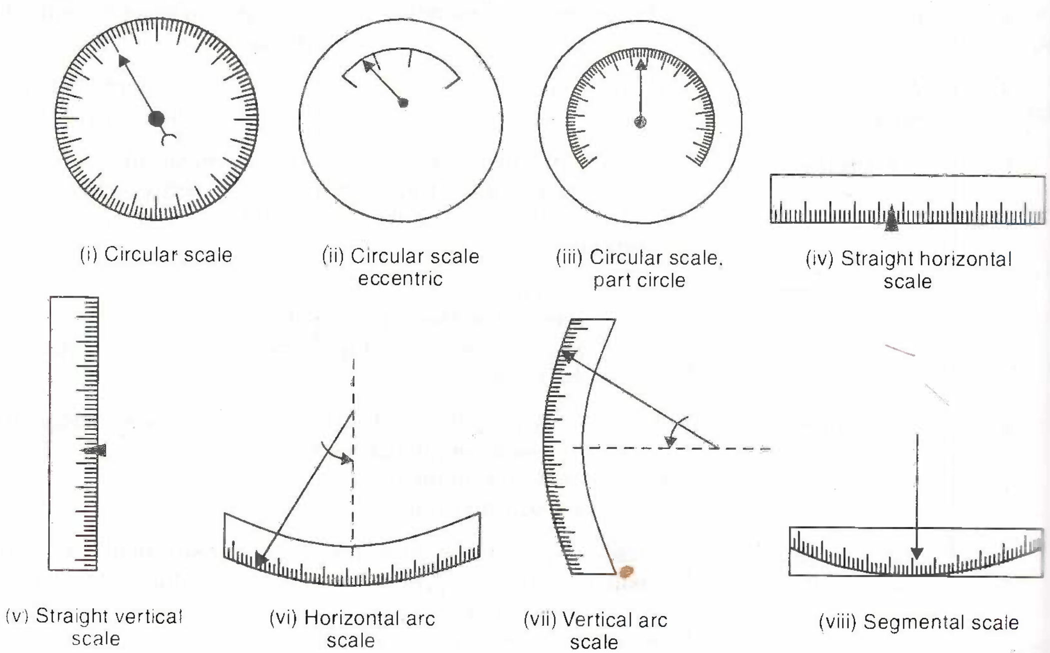


Fig. 4.36. Fixed-scale and movable-pointer indicators.

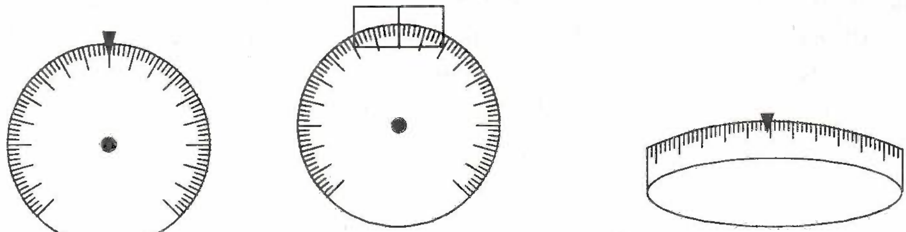


Fig. 4.37. Fixed-pointer and movable-scale indicators.

The readability of graduated dials is influenced by the following *factors*:

- The shape and length of the pointer.
- The number, spacing, length and thickness of scale marking.
- The system of numbering of the scale marks.
- The size and design of the numerals.

2. ***Multi-point, multi-pointer and multi-range indicators:***

Multi-point indicator. In this system the indicator pointer can be connected to a number of inputs, one at a time with the help of a selector switch.

The selector switch may be operated either manually or automatically after a pre-determined time. The observed reading is multiplied by a factor corresponding to the particular measurand.

Generally such systems are confined where measurable variables are of electrical signals as the selection is accomplished by switching electrical circuits. However, gas selector switches also exist which connect one gas pipe at a time to the measuring instrument and are well designed to avoid leakage of gas.

Multi-pointer indicator. This type of indicator contains more than one number of pointers and above each point the identification number of the medium being measured is marked. *Usually this arrangement is used in recorders and not in indicators.*

Multi-range indicators. An instrument with multi-range indicators has *different scales for different ranges;* the choice of a particular scale is made by a selector switch.

Essential features of indicating instruments:

Indicating instruments possess three essential features:

1. *Deflecting device.* ..... Whereby a mechanical force is produced by the electric current, voltage or power.
2. *Controlling device.* ..... Whereby the value of deflection is dependent upon the magnitude of the quantity being measured.
3. *Damping device.* ..... To prevent oscillation of the moving system and enable the latter to reach its final position quickly.

4.11.3. Analog Instruments

Moving-iron instruments (Ammeters and voltmeters):

Moving-iron instruments are commonly used in laboratories and switch board at commercial frequencies because they *are very cheap and can be manufactured with required accuracy.*

Moving-iron instruments can be divided into two types:

1. Attraction type .... in which a sheet of soft iron is *attracted towards a solenoid.*
2. Repulsion type .... in which two parallel rods or strips of soft iron, magnetised inside a solenoid, are regarded as *repelling each other.*

Moving-coil instruments:

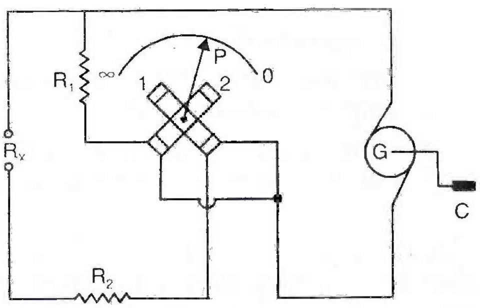
The moving-coil instruments are of the following two types:

1. Permanent-magnet type .... can be used for D.C. only.
2. Dynamometer type .... can be used both for A.C. and D.C.

Megger :

Meggers (or megohmmeters) are instruments which measure the *insulation resistance of electric circuits relative to earth and one another.*

A megger consists of an e.m.f. source and a *voltmeter.* The scale of the voltmeter is calibrated in ohms (kilo-ohms or megohms, as the case may be). In measurements the e.m.f. of the self-contained source must be equal to that of the source used in calibration.



G = Generator
C = Crank
1, 2 = Coils
P = Pointer
1 = Current coil
2 = Pressure/Voltage coil
R_x = Unknown resistance
R_1 = Fixed resistance
R_2 = Safety resistance

Fig. 4.38. Circuit diagram of megger.

Fig. 4.38 shows diagrammatically a megger whose readings are independent of the speed of the self-contained generator. The moving system incorporates two coils 1 (current coil) and 2 (pressure coil) mounted on the same shaft and *placed in the field of a permanent magnet* (not shown) 90° apart. The generator energizes the two coils over separate wires. Connected in series with one coil is a fixed resistance R_1 (or several different resistances in order to extend the range of the instrument). The unknown resistance R_x is connected in series with the other coil. The *currents in the coils interact with the magnetic field and produce opposing torques.*

The deflection of the moving system depends on the ratio of the currents in the coils and is independent of the applied voltage. The unknown resistance is read *directly* from the scale of the instrument. (The accuracy of measurement is unaffected by variations in the speed of the generator between 60 and 180 r.p.m.).

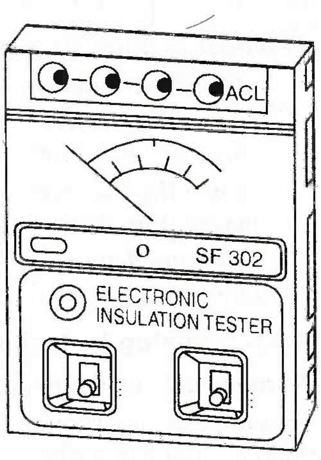


Fig. 4.39. Electronic insulator tester.

Electronic insulation tester:

Fig. 4.39 shows an electronic insulator tester :

- These days electronic tester is used to test the insulation.

- It can also measure low resistance 0 to 2 kΩ, high resistance 0.05 to 100 MΩ and A.C. voltage upto 0 to 100 V.
- It is easy to use.
- It does not require hand rotation.
- It works on six cells of 1.5 V each.

Multimeter (AVO):

Fig. 4.40 shows a *Multimeter* (AVO *meter*). The basic circuit of the multimeter is shown in Fig. 4.41.

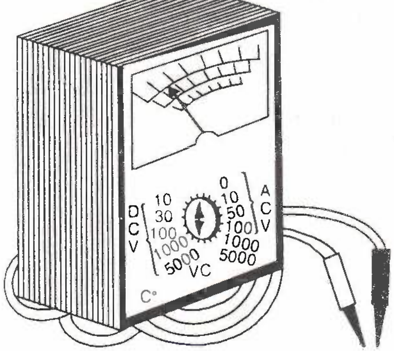


Fig. 4.40. Multimeter.

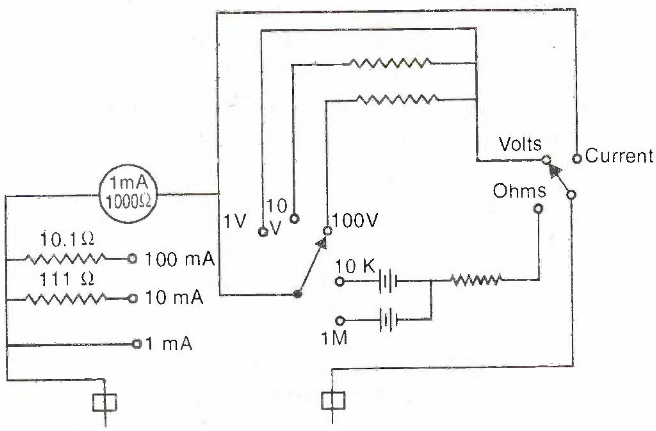


Fig. 4.41. Basic circuit of a multimeter.

Main parts. The following are the main parts of a multimeter :

1. One or two cells 1.5 V.
2. 12 position rotary switch.
3. Moving coil meter.
4. Different types of resistances.
5. Rectifier.
6. Many condensers.
7. Case.

This meter can work as **voltmeter, ammeter** or **ohmmeter.**

Voltmeter. Ten ranges, 5 for D.C. and 5 for A.C.

Ampere meter. Due to several ranges in this meter, we can measure mA (milliampere) also.

Ohmmeter. When using it as ohmmeter 3 ranges are available × 1, × 10, × 100. If we are taking reading on × 10, we are to multiply the reading by 10, *e.g.*, if reading is 5 Ω the actual reading will be 10 × 5 = 50 Ω.

Applications:

The multimeter can be used to accomplish the following jobs:

- *D.C. 0 to 10 V scale.* To test one or two cells voltage or to test radio voltage upto 10 V.
- *D.C. 0 to 30 V scale.* To test 6 coils storage battery or to test hearing aid machine.
- *D.C. 0 to 300 V scale.* To measure supply voltage and to measure radio D.C. voltage.

- *D.C. 0 to 1000 V scale.* To test the voltage of photo-flash battery.
- *A.C. 0 to 10 V scale.* Bell transformer, night lamp voltage testing.
- *A.C. 0 to 30 V scale.* To check bell, or toy train transformer.
- *A.C. 0 to 300 V scale.* To check house meter voltage, radio voltage.

Testing purpose. When using as ohmmeter, this meter can be used to *measure the resistance and to test the continuity of wires etc.*

Electronic voltmeters:

Almost all electronic voltmeters make use of the *rectifying properties of diodes* whether vacuum tubes or metal rectifiers or semiconductor diodes.

- Vacuum tube diode was first used in electronic voltmeters way back in 1895 and is still popular as sensing element of Vacuum Tube Voltmeters (VTVM).
- With the introduction of the transistor and other semiconductor devices vacuum tubes are on their way out. Solid state models with junction field effect transistor (JFET) input stages are known as Transistor Voltmeter (TVM) and Field Effect Transistor Voltmeters (FETVM) are taking their place.

The electronic voltmeters claim the following *advantages:*

1. Detection of low level signals.
2. Low power consumption.
3. High frequency range.

4.11.4. Digital Instruments

The **digital instruments** *indicate the value of the measured in the form of decimal number* (whereas the analog instruments display the quantity to be measured in terms of deflection of a pointer, *i.e.,* an analog displacement or an angle corresponding to the electrical quantity).

The **digital meters** *work on the principle of "qantization".* The analog quantity to be measured is first subdivided or quantized into a number of small intervals upto many decimal places. The objective of the digital instrument is then to determine in which portion of the subdivision the measured can thus be identified as an integral multiple of the smallest unit called the *quantum,* chosen for subdivision. The measuring procedure thus reduces to one of counting the number of quanta present in the measurand.

A **digital instrument** *can be considered as a counter which counts the pulses in a predetermined time.* Digital transducers whose output is in the form of pulses are used to monitor the desired parameter. *Accuracy of digital instrument is dependent on the number of pulses generated by transducer because the fraction of pulse cannot be generated and in counting there can be ambiguity of only one pulse or start/stop.* Hence *more are pulses* corresponding to a measure less the possibility of error corresponding to one pulse and *more the accuracy.*

The *information* in the electronic digital read-out (display) devices is *presented as a series of digits on tubes, screen or printed on a piece of paper.* The relevant characters (letters of alphabet from A to Z, numerals from 0 to 9, punctuation mark and other symbols in common use) can be generated by :

(*i*) Semiconductor light emitting diodes (LED).
(*ii*) Liquid crystal displays (LCD).
(*iii*) Numerical indicators tubes (NIT).
(*iv*) Hot filament or bar tubes.

4.11.5. Recorders

A **recorder** *records electrical and non-electrical quantities as a function of time.* The record may show how one variable varies with respect to another, or how the input signal varies with time.

The record serves the following *objectives* :

(*i*) It preserves the details of measurement at a particular time.

(*ii*) It provides at a glance the overall picture of the performance of unit.

(*iii*) It provides immediate reflection on the action taken by the operator.

Types of recorders :

In an instrumentation system, one of the important considerations is the method by which the data required is recorded. The recording method should be consistent with the type of system. If we are dealing with a wholly analog system, then *analog recording techniques* should be used. While, on the other hand, if the system has a digital output, *digital recording devices* are employed.

Two types of recorders are:

1. ***Analog recorders :***

(*i*) Graphic recorders

  (*a*) Strip chart recorders

   - Galvanometer type
   - Null type
     — Potentiometric recorders
     — Bridge recorders
     — LVDT recorders

  (*b*) X-Y recorders

(*ii*) Oscillographic recorders.

(*iii*) Magnetic tape recorders.

2. ***Digital recorders :***

The above recorders are discussed briefly below :

1. Strip chart recorders : Fig. 4.42 shows the basic constructional features of a strip chart recorder.

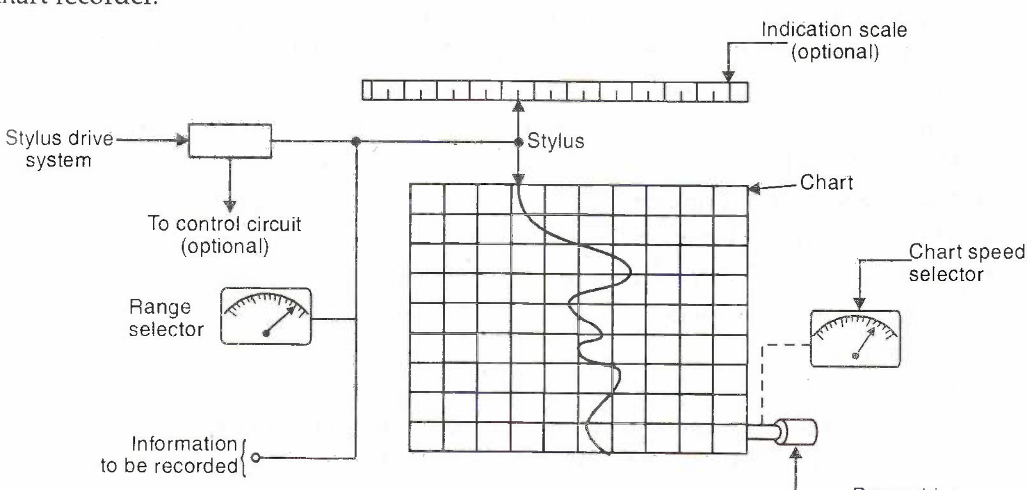


Fig. 4.42. Strip chart recorder.

A strip chart consists of the following:

- A long wall of graph paper moving *vertically*.
- A system for driving the paper at some selected speed.
- A stylus for marking paper on the moving graph paper (Most recorders use a pointer attached to the stylus, which (pointer) moves over a calibrated scale thus showing instantaneous value of the quantity being measured).
- A stylus driving system which moves the stylus in nearby exact replica or analog of the quantity being measured (A spring wound mechanism may be used but in most of the recorders a *synchronous motor* is used for driving the paper).

Marking mechanisms. The most commonly used mechanisms employed for marking marks on the paper are :

(*i*) Marking with ink filled stylus.
(*ii*) Marking with heated stylus.
(*iii*) Chopper bar.
(*iv*) Electric stylus marking.
(*v*) Electrostatic stylus.
(*vi*) Optical marking method.

Tracing systems. For producing graphic representations, the following two types of tracing systems are used:

(*i*) Curvilinear system.
(*ii*) Rectilinear system,

Galvanometer type strip chart recorders :

- This type of recorder operates on the *"deflection principle"*.
- The deflection is produced by a galvanometer, (d'Arsonval) which produces a torque on account of a current passing through its coil. This current is proportional to the quantity being measured.
- These recorders can work on ranges for a few mA to several mA or from a few mV to several mV.
- The moving galvanometer type recorder is *comparatively inexpensive instrument*, having a narrow bandwidth of 0 to 10 Hz. It has a sensitivity of 0.4 mV/mm or from a chart of 100 mm width a full scale deflection of 40 mV is obtained. Linear amplifiers are used for measurement of smaller voltages.
- This type of recorder is not useful for recording fast variations in either current or voltage or power.

Null type strip chart recorder:

- This type of recorder operates on *"comparison basis"*.

The null type strip chart recorders are of the following types :

1. Potentiometric recorders.
2. Bridge recorders.
3. LVDT recorders.

The most common application of **potentiometric recorder** is for *recording and control of process temperatures*. Self balancing potentiometers are unduly used in industry because of the following *reasons:*

(*i*) Their action is automatic and thus eliminates the constant operation of an operator.

(*ii*) They draw a curve of the quantity of being measured with the help of recording mechanism.

(*iii*) They can be mounted on switchboard or panel and thus act as mounting devices for the quantity under measurement.

Single-point and multi-point recorders:

- *Instruments that record changes of only one measured variable are called* **single-point recorders.**
- A **multi-point recorder** may have as many as 24 inputs, with traces displaced in 6 colours.

2. X-Y recorders:

A X-Y recorder is an instrument *which gives a graphic record of the relationship between two variables.* This system has a pen which can be positioned along the two axes with the writing paper remaining stationary. There are *two* amplifier units, one amplifier actuates the pen in the Y-direction as the input signal is applied, while the second amplifier actuates the pen in X-direction. The movements of the pen in X-and Y-directions are automatically controlled by means of a motor, pulleys and a linear potentiometer. Obviously, trace of the marking pen will be due to the combined effects of two signals applied simultaneously. In these recorders, an e.m.f. is plotted as a function of another e.m.f. There are many variations of X-Y recorders. With the help of these *recorders and appropriate transducers a physical quantity may be plotted against another physical quantity.*

A few **examples** in which *use of X-Y recorders is made* are as under:

(*i*) Plotting of stress-strain curves, hysteresis curves and vibrations amplitude against swept frequency.

(*ii*) Pressure-volume diagrams for I.C. engines.

(*iii*) Pressure-flow studies for lungs.

(*iv*) Lift drag wind tunnel tests.

(*v*) Electrical characteristics of materials such as resistance versus temperature and plotting the output from electronic calculators and computers.

(*vi*) Speed-torque characteristics of motors.

(*vii*) Regulation curves of power supplies.

(*viii*) Plotting of characteristics of vacuum tubes, zener diodes, rectifiers and transistors etc.

3. Ultraviolet (U.V.) recorders:

These recorders are basically electro-mechanical oscillographic recorders and modified version of Duddel's oscillographs.

An ultraviolet recorder consists of a number of galvanometer (moving coil) elements mounted in a single magnet block as shown in Fig. 4.43. A paper sensitive to ultraviolet light is used for producing a trace for the purpose of recording. The u.v. light is projected on the paper with the help of mirrors attached to the moving coils.

Working. When a current is passed through the moving (galvanometer) coil, it deflects under the influence of the magnetic field of the permanent magnet. The ultraviolet light falling on the mirrors is deflected and projected on to the u.v. light sensitive paper through a lens and mirror system. The paper is driven past the moving high spot and thus a trace of variation of current with respect to time is produced.

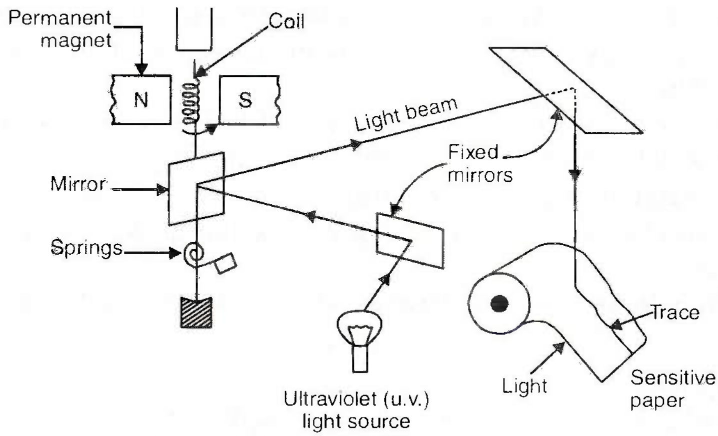


Fig. 4.43. Ultraviolet (u.v.) recorder.

The recorder, in addition to the input currents, may have the following additional traces:

(*i*) Grid lines.

(*ii*) Timing lines.

(*iii*) Trace identification.

- The ultraviolet (u.v.) recorders, compared to the mechanical and pen recorders, have *better frequency and response characteristics;* the typical values are :

  Frequency response = 0 to 300 Hz; 0 to –12 kHz (maximum)

  Response time = 16 μs

  — The maximum frequency that may be recorded depends upon the frequency response of the galvanometer used. When high frequency signals are to be recorded, the marking paper is moved with sufficient speed so as to spread out the trace along the time-axis. *The u.v. recorders have an additional advantage of multi-trace recording.*

- *"Typical applications"* of U.V. recorders are in recording:

  (*i*) Regulation transients of generators.

  (*ii*) Output of transducers.

  (*iii*) Control system performance.

- These recorders are also used for *recording the magnitude of low frequency signals which cannot be measured with analog (pointers) type instruments.*

4. Magnetic tape recorders:

These recorders have response characteristics which enable them to be used at *higher frequencies;* hence they find an *extensive use in Instrumentation systems.*

A magnetic tape recorder consists of the following *basic components*:

1. Recording head.
2. Magnetic tape.
3. Reproducing head.
4. Tape transport mechanism.
5. Conditioning devices.

Advantages:

(*i*) Low distortion.

(*ii*) Wide frequency range from D.C. to several MHz.

(*iii*) Wide dynamic range which exceeds 50 dB.

(*iv*) The recorded signal is immediately available with no time lost in processing. The recorded signal can be played back, or reproduced as many times as desired without loss of signal.

(*v*) Multi-channel recording possible.

5. Cathode Ray Oscilloscope (CRO):

A cathode-ray oscilloscope is an instrument which presents signal waveforms visually. It is also useful for *comparing two signals in phase, frequency or amplitude.*

A CRO can operate upto 50 MHz, can allow viewing of signals within a time span of a few nanoseconds and can provide a number of waveform displays simultaneously on the screen. It also has the ability to hold the displays for a short or long time (for many hours) so that original signal may be compared with one coming on later.

A block diagram of cathode-ray oscilloscope is shown in Fig. 4.44.

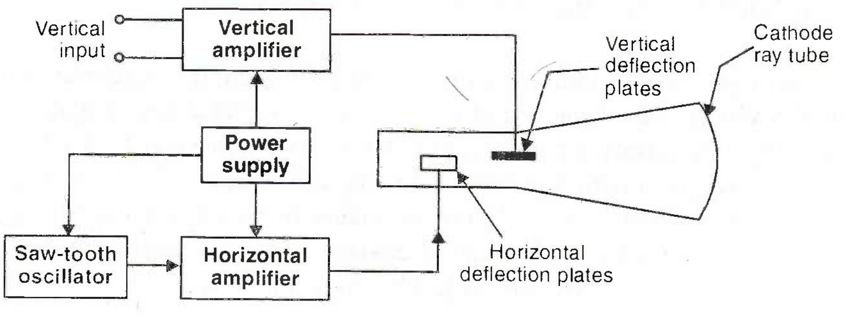


Fig. 4.44. Cathode-ray oscilloscope.

Cathode ray tube (CRT):

A cathode ray tube is the 'heart' of an oscilloscope and is very similar to the picture tube in a television set.

Fig. 4.45 shows the cross-sectional view of a general-purpose electrostatic C.R.T.

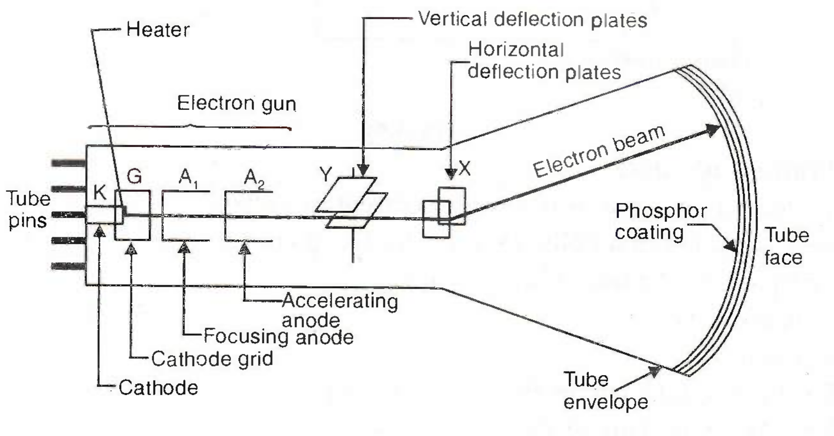


Fig. 4.45. Cathode ray tube.

It has the following four major *components*:

1. ***Electron gun*** .... it produces a stream of electrons.
2. ***Focusing and accelerating anodes*** .... they produce a narrow and sharply-focused beam of electrons.
3. ***Horizontal and vertical deflecting plates*** .... for the path of beam.
4. ***An evacuated glass envelope with a phosphorescent screen*** .... produces a bright spot when struck by a high velocity electron beam.

Working of a C.R.O :

When a signal is to be displayed or viewed on the screen it is applied across the Y-plates of a cathode ray tube. But to see its waveform or pattern, it is essential to spread it to horizontally from left to right. This is achieved by applying a sawtooth voltage wave to X-plate.

Under these conditions, the electron beam would move uniformly from left to right thereby graphing vertical variations of the input signal versus time. Due to repetitive tracing of the viewed waveform, we get a continuous display because of *persistence of vision*.

However, to get a stable stationary display on the screen, it is essential to *synchronize the horizontal sweeping of the beam* (sync) with the input signal across Y-plates. The signal will be properly synced only when its frequency *equals the sweep-generator frequency*. The usual method of synchronizing the input signal is to use a portion of the input signal to trigger the sweep generator so that the frequency of the sweep signal is locked or synchronized to the input signal. It is called internal sync because the synchronization is obtained by internal wiring connections as shown in Fig. 4.46.

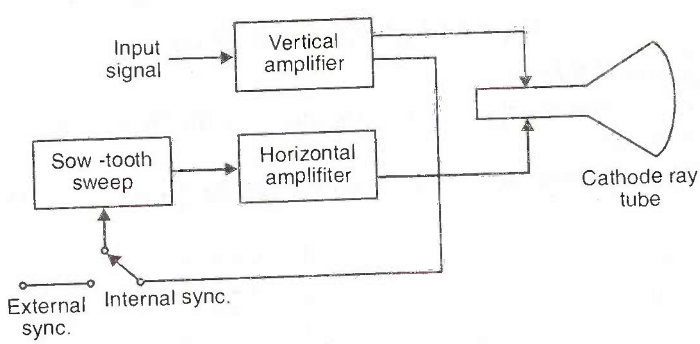


Fig. 4.46.

Applications of C.R.O.:

1. Tracing of an actual waveform of current or voltage.
2. Determination of amplitude of a variable quantity.
3. Comparison of phase and frequency.
4. In televisions.
5. In radar.
6. For finding B.H. curves for hysteresis loop.
7. For engine pressure analysis.
8. For studying the heart beats, nervous reactions etc.
9. For tracing transistor curves.

Example 4.3. *What do you mean by "Direct recording"? State also its advantages and limitations.*

Solution. Direct recording. It is the simplest method of recording and *usually requires one tape track for each channel. The signal to be recorded is amplified, mixed with a high frequency bias and fed directly to the recording head as a varying electric current.*

Advantages:

(*i*) This recording process requires only simple, moderately priced electronic circuitry.

(*ii*) It has a wide frequency response ranging from 50 Hz to about 2 MHz for a tape speed of 3.05m/s. It provides the *greatest bandwidth obtainable from a given recorder.*

(*iii*) It has a good dynamic response and takes overload without increase in distortion.

(*iv*) It can be used for recording voice and multiplexing a number of channels of information into one channel of tape recording.

(*v*) It is used to record signal where information is contained in the relation between frequency and amplitude, such as spectrum analysis of noise.

Limitations :

(*i*) It is *used only when maximum bandwidth is required and when variations in amplitude are acceptable.*

(*ii*) It is *mainly used for recording of speech and music.*

Example 4.4. *What are the advantages and disadvantages of strip chart recorders ?*

Solution. Advantages of strip chart recorders:

(*i*) The rate of movement of the chart can easily be changed to spread out the trace of the variable being observed.

(*ii*) Data conversion is easier when rectangular coordinates are used.

(*iii*) These recorders require the use of servo-mechanisms to position the pointer or pen. Therefore more than adequate power is available, there being no real limitations on the weight of the pen, pressure between pen and paper or length of the pointer.

(*iv*) Relatively large amount of paper can be inserted at one time in the form of a nell.

(*v*) Many more separate variables can be recorded on a strip chart than on circular chart.

Disadvantages :

(*i*) The mechanism is considerably more complicated than is required to drive a circular chart.

(*ii*) Observing behaviour several hours or days back is not as easy as picking out one circular chart which covers the desired period of time.

Example 4.5. *How do "Circular chart recorders" differ from "Strip chart recorders"?*

Solution. The differences between circular chart and strip chart recorders are given, in a tabular form, on page 302:

Table 4.2

S. No.	Aspects	Circular chart recorders	Strip chart recorders
1.	*Handling and storing*	Easy	Very easy
2.	*Shape and size of chart*	Circular, varying in size from 100 mm to 250 mm diameter.	Curvilinear type, available in the form of long strips usually rolled on to a drum.
3.	*Usable recording area*	40 to 50% area of chart is calibrated and rest is the space covered by mechanism involved.	90% or more of the chart width is usable recording area, very small position being taken up by punched holes for guide purposes.
4.	*Amount of information that can be carried.*	Strictly limited amount.	It can be packed with information.
5.	*Exhibition of information*	It shows all the information recorded at a glance.	It needs to be unrolled to see past records.
6.	*Facility to record*	It is possible to simultaneously record on the full-chart range upto four separate variables.	It is possible to record upto 4 to 6 points simultaneously and thus afford saving a lot of panel space.
7.	*Cost*	Low initial cost.	Cost though high is justified, considering its versatility, predictive diagnostic capability, invaluable tool for analysing the overall dynamic response.
8.	*Range of chart speeds*	Usually the circular chart moves at one constant speed and high speed phenomenon cannot be recorded.	It is possible to have wide range of chart speeds and records fast changing phenomenon.
9.	*Chart speed*	The chart speed is limited and as such recording cycle takes longer time for multiple points.	The availability of wide range of chart speeds enables the recording of greater number of points and at a much higher speed than is practical with circular chart instruments.
10.	*Type of operators required*	Less skilled operators can do the job since it is easy to adjust and repair instruments.	Skilled operators are required.

4.11.6. Printers

The *printers provide a record of data on paper.* Such printers are available in the following versions:

1. The dot matrix printer.
2. The ink jet printer.
3. The laser printer.

1. The dot matrix printer :

Fig. 4.47 shows the head mechanism of a dot matrix printer.

— It consists of either 9 or 24 pins in a vertical line.

— Each pin is controlled by an electromagnet which when turned on propels the pin onto the inking ribbon. This transfers a small bob of ink onto the paper behind the ribbon. A character is formed by moving the print head in horizontal lines back-and-forth across the *paper* and *firing* the appropriate pins.

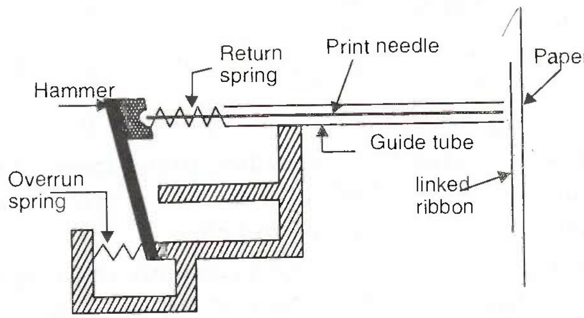


Fig. 4.47. Head mechanism of a dot matrix printer.

2. The ink jet printer :

This type of the printer uses a conductive ink which is forced through a small nozzle to produce a jet of very small drops of ink of constant diameter at a constant frequency.

— In 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. (Fig. 4.48).

— In another form is used a small heater in the print head with vaporized ink in a capillary tube, so producing gas bubbles which push out 'ink drops'.

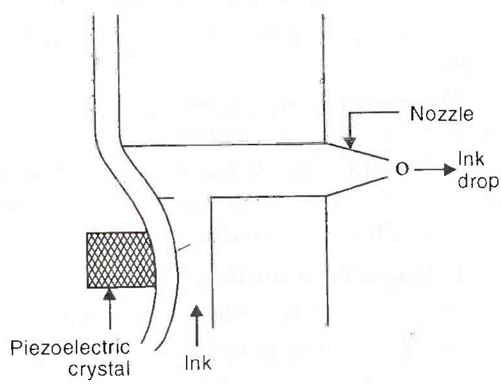


Fig. 4.48. Production of stream of drops.

3. The laser printer :

Figure 4.49 shows the basic elements of a laser printer.

— It has a photosensitive drum which is coated with a selenium-based light sensitive material. The selenium, in the dark, has a high resistance and consequently becomes charged as it passes close to the charging wire; this is a wire at a high voltage and off which charge leaks.

— A light beam is made to scan along the length of the drum by a small rotating eight-sided mirror. 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.

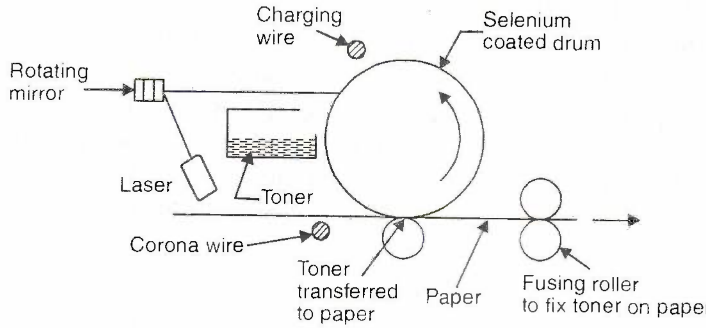


Fig. 4.49. Laser printer's basic elements.

— As the drum passes the toner reservoir, the charged areas attract particles of toner which thus stick to the areas that have not been exposed to light and do not stick on the areas that have been exposed to light.

— The paper is given a charge as it passes another charging wire, the so called corona wire, so that as it 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 finely adhere to paper.

4.11.7. Magnetic Recording

The use of *'magnetic recording'* is restored *to store data on the floppy disc and hard discs of computers.*

The basic principles are that *a recording head, which responds to the input signal, produces corresponding magnetic patterns on a thin layer of magnetic material and a read head gives an output by converting the magnetic patterns on the magnetic material to electrical signals.* Besides these heads the systems require a transport system which moves the magnetic material in a controlled way under the heads.

1. Magnetic recording codes :

- In digital recording, the signals are recorded as a *coded combination of bits.*
- A *bit cell* is the element of the magnetic coating where the magnetism is either completely saturated in one direction or completely saturated in the reverse direction. Saturation is when the magnetising field has been increased to such an extent that the magnetic material has reached its maximum amount of magnetic flux and further increases in magnetising current produce no further change.
- For getting proper flux reversals, some of the commonly used methods (involving encoding) are :
  - (*i*) Phase encoding (PE);
  - (*ii*) Non-return-to-zero (NRZ);
  - (*iii*) Frequency modulation (FM);
  - (*iv*) Modified frequency modulation (MFM);
  - (*v*) Run length limited (RLL).

Optimum code is the one that allows the bits to be packed as close as possible and which can be read without error. The read head can locate reversals quite easily but they must not be too close together.

— The RLL code has the advantage of being *more compact* than the other codes, PE and FM taking up the most space.

— MFM and NRZ take up the same amount of space. *NRZ* has the disadvantage of, unlike the other codes, *not being self locking.*

2. Magnetic discs :

Digital recording is commonly done *on a floppy or hard disc.*

The digital data is stored on the disc surface along concentric circles called *tracks*, a single disc having many such tracks. A single read-write head is used for each disc surface and heads are moved, by means of a mechanical actuators, backwards and forwards to access different tracks. The disc is spum by the drive and the read/write heads *read or write data into a track.*

The 3.5 *"floppy disc"* used in the personal computer can store 1.4 *Mbytes* of data.

"Hard discs". These are sealed units with data stored on the disc surface along concentric circles.

— A hard disc assembly has more than one such disc and the data is stored on magnetic coatings on both sides of the discs.

— The discs are rotated at high speeds and tracks assessed by moving the read-write heads.

— Large amounts of data can be stored on such assemblies of discs; storage of the order of many *G bytes* are now common.

4.11.8. Display Systems

Several display systems use light indicators to indicate on-off status or give alphanumeric displays.

Some of these display systems are enumerated and briefly discussed below:

1. Light indicators.
2. LED displays.
3. A 5 by 7 dot matrix LED display.
4. Liquid crystal displays.

1. Light indicators :

For such displays, the light indicators may be neon lamps, incandescent lamps, light-emitting diodes (LEDs) or liquid crystal displays (LCDs).

- *Neon lamps* need high voltages and low currents and can be powered directly from the mains voltage but can only be used to give a red light.
- *Incandescent lamps* can be used with a wide range of voltages but need a comparatively high current. They emit white light to use lenses to generate any required colour. Their main advantage is their *brightness.*
- LEDs (light-emitting diodes) require low voltages and low currents and are cheap.
  — These diodes when forward biased emit light over a certain band of wavelengths.
  — The most commonly used LEDs can give red, yellow or green colours.
  — *With microprocessor-based systems, LEDs are the most common form of indicators.*

2. LED displays:

— With a LED a current-limiting resistor is generally required in order to limit the current to below the maximum rated current of about 10 to 30 mA.

— Some LEDs are supplied with *built in resistors* so they *can be directly connected to microprocessor systems.*

— LEDs are available as *single light displays, seven-and-sixteen-segment alphanumeric displays,* in dot matrix format and bar graph form.

3. A 5 by 7 dot matrix LED display:

In this type of display the array consists of five column connectors, each connecting the anodes of seven LEDs. Each row connects to the cathodes of five LEDs. To turn on a particular LED, power is applied to its column and its row is grounded.

4. Liquid crystal displays :

Such displays are used in battery-operated devices such as watches and calculators.

— Five by seven dot matrix forms are also available.

HIGHLIGHTS

1. The *signal conditioning equipment* may be required to perform the following functions on the transduced signal :
   (*i*) Amplification (*ii*) Modification or modulation
   (*iii*) Impedance matching (*iv*) Data processing
   (*v*) Data transmission.
2. An *amplifier* is a device which is used to increase or augment the weak signal.
3. An *operational amplifier* (Op-amp) is a linear integrated circuit (IC) that has very high voltage gain, a high input impedance and a low output impedance.
4. *Filtering* is process of attenuating unwanted components of a measurement while permitting the desired component to pass.
5. A *good display,* functionally, is one which permits the best combination of speed, accuracy and sensitivity when transferring the necessary information from the instrument to the operator.
6. The electrical indicating instruments may be classified as:
   (*i*) Analog instruments.
   (*ii*) Digital instruments.
7. Essential features of indicating instruments are :
   (*i*) Deflecting device.
   (*ii*) Controlling device.
   (*iii*) Damping device.
8. The *digital instruments* indicate the value of the measurand in the form of decimal numbers whereas the analog instruments display the quantity to be measured in terms of deflection of a pointer *i.e.,* an analog displacement or an angle corresponding to the electrical quantity.
9. The *digital meters* work on the principle of *"quantization".*
10. A *digital instrument* can be considered as a counter which counts the pulses in a predetermined time.
11. A numerical indicator tube (NIT) consists of a gas filled glass tube having ten cathodes in the form of numbers and an anode.
12. A *recorder* records electrical and non-electrical quantities as a function of time. Two types of recorders are :

(i) Analog recorders.

(ii) Digital recorders.

13. Instruments that record change of only one measured variable are called *single point recorders.* A *multi-point recorder* may have as many as 24 inputs, with traces displaced in 6 colours.
14. A *X-Y recorder* is an instrument which gives a graphic record of the relationship between two variables.
15. *Cathode ray oscilloscope (CRO)* is an instrument which presents signal waveforms visually. It is also useful for comparing two signals in phase, frequency or amplitude.

OBJECTIVE TYPE QUESTIONS

A. Choose the Correct Answer :

1. The closed loop gain of an Op-amp is dependent upon whether the Op-amp is used
   (a) in inverting mode (b) in non-inverting mode
   (c) is independent of the fact whether the input is connected to inverting or non-inverting terminal.
   (d) is dependent upon the fact whether the input is connected to inverting or the non-inverting terminal.
2. A buffer amplifier has gain of
   (a) infinity (b) zero
   (c) unity (d) dependent upon the circuit parameters
3. A.C. amplifiers are best suited for
   (a) steady-state signals (b) low frequency signals
   (c) rapidly varying signals (d) none of these.
4. The amplifier drift and spurious noise signals are not significant in
   (a) a.c. amplifiers (b) d.c. amplifiers
   (c) charge amplifiers (d) none of these.
5. In a carrier system, drift and spurious signals are important
   (a) because they modulate the carrier
   (b) because they do not modulate the carrier
   (c) because it is easier to achieve a stable carrier than a stabilized d.c. source.
   (d) none of the above.
6. When using d.c. signal conditioning system, with a carrier of 3 kHz, the data frequency should be limited to :
   (a) 1 kHz (b) 5 Hz
   (c) 600 Hz (d) 2 MHz.
7. The input and output displacements are of opposite phase in
   (a) simple lever (b) compound lever
   (c) compound gear trains (d) none of these.
8. What is the desirable feature in an electronic amplifier?
   (a) High output impedance (b) Low input impedance
   (c) good frequency response (d) All of these.
9. Charge amplifiers are used in order to amplify the output signals of
   (a) inductive (b) capacitive
   (c) resistive
   (d) piezo-electric and capacitive transducers.

10. Filters that transmit all frequencies below a defined cut-off frequency are known as
    (*a*) low-pass filters (*b*) high-pass filters
    (*c*) band-pass filters (*d*) any of these.
11. Excitation and amplification systems are needed
    (*a*) for active transducers only (*b*) for passive transducers only
    (*c*) for both active and passive transducers
    (*d*) for both passive and output transducers.
12. A d.c. amplifier
    (*a*) needs to have a balanced differential inputs with a high common mode rejection ratio (CMRR) to give very good thermal and long term stability.
    (*b*) easy to calibrate at low frequencies and has ability to recover from overload conditions.
    (*c*) is immune to drift and low frequency spurious signals come out as data information.
    (*d*) is followed by a low pass filter to eliminate high frequency components including noise from the data signal.
    (*e*) all of the above.
13. The output from frequency-modulation system is
    (*a*) a.c. voltage (*b*) d.c. voltage
    (*c*) a.c. and d.c. voltage (*d*) any of these.
14. The data transmission with synchro systems employs .... telemetering to convey the requisite information.
    (*a*) frequency (*b*) position
    (*c*) impulses (*d*) voltage.
15. When using a.c. signal conditioning system for capacitive transducers, the carrier frequencies
    (*a*) range between 50 Hz and 20 kHz (*b*) should be of the order of 0.5 MHz
    (*c*) should be of the order to 20 MHz (*d*) none of the above.
16. An a.c. signal conditioning system is normally used for
    (*a*) resistive transducers like strain gauges (*b*) inductive and capacitive transducers
    (*c*) piezoelectric transducers (*d*) all of the above.
17. The overall gain or amplification of a system of two amplifiers arranged in series is
    (*a*) $G_1 + G_2$ (*b*) $G_1 - G_2$
    (*c*) $G_1 \times G_2$
    (*d*) $\frac{G_1}{G_2}$ where G_1 and G_2 are the two gains expressed as pure numbers.
18. The properties of an ideal Op-amp are :
    (*a*) It should have zero input impedance (*b*) It should have high input impedance
    (*c*) It should have a zero open loop gain (*d*) None of the above.
19. The moving iron voltmeters indicate :
    (*a*) the same value of d.c. and a.c. voltages.
    (*b*) lower values for a.c. voltages than the corresponding values of d.c. voltages.
    (*c*) higher value for a.c. voltages than the corresponding values of d.c. voltages.
    (*d*) none of the above.
20. Which of the following is the visual display unit?
    (*a*) Cathode any oscilloscope (*b*) U.V. recorder
    (*c*) Storage oscilloscope (*d*) Moving coil oscillograph.
21. Which of the following units has a high frequency response but presents difficulty in getting a permanent record?

(a) Servo recorder (b) Moving coil oscillograph
(c) X-Y recorder (d) Cathode ray oscilloscope.

22. The switching time of LEDs is of the order of
(a) 1 s (b) 1 ms
(c) 1 μs (d) 1 ns.

23. LEDs emit light
(a) only in red colour (b) only in yellow colour
(c) only in green colour (d) in red, green yellow and amber colours,

24. The advantages of F.M. magnetic tape recording are :
(a) It can record from d.c. to several kHz (b) It is free from dropout effects
(c) It is independent of amplitude and accurately reproduces the waveform of input signal
(d) All of the above.

25. The source of emission of electrons in a CRT is :
(a) PN junction diode
(b) a barium and strontium oxide coated cathode
(c) accelerating anodes (d) post accelerating anodes.

26. The pointer-scale instruments have a
(a) very high (b) very low
(c) linerar (d) stable frequencies response.

27. The operation of a moving-coil current recording instruments is based on
(a) photo-electric principle (b) D' Arsonval principle
(c) piezo-electric principle (d) thermo-electric principle.

28. The turn on and turn off times of a LCD are of the order of :
(a) 1 s (b) 1 ms
(c) 10 ms (d) 10 ns.

29. The power requirement of an LED is
(a) 40 mW per numeral (b) 40 μW per numeral
(c) 10 W per numeral (d) 10 μW per numeral.

30. A Nixie tube requires
(a) 10 cathodes (b) 12 cathodes
(c) 15 cathodes (d) 20 cathodes.

31. The time bases of an oscilloscope are generated by
(a) horizontal amplifier (b) vertical amplifier
(c) sweep generators (d) storage oscilloscope.

32. Which of the following devices requires a matching network to avoid overloading of the signal source and prevent damage from excessive current?
(a) U.V. recorder (b) X-Y recorder
(c) Storage oscilloscope (d) Servo recorder.

33. X-Y recorders record a quantity
(a) with respect to another quantity (b) on X axis with respect to time on Y axis.
(c) on Y axis with respect to time on X axis (d) any of these.

34. In an electrodynamometer type of wattmeter
(a) the current coil is made fixed (b) the pressure coil is made fixed
(c) any of the two coils can be made fixed (d) both the coils should be movable.

35. An average reading VTVM uses one diode with an external series resistance. A high value of series resistance is used so that the instrument should have

(*a*) a high input impedance (*b*) a linear *v-i* characteristics
(*c*) low power consumption (*d*) all of the above.

36. In a CRT the focusing anode is located
 (*a*) between pre-accelerating and accelerating anode
 (*b*) after accerating anode
 (*c*) before pre-accelerating anode (*d*) none of the above.
37. Post acceleration is needed in a CRO if the frequency of the signals is
 (*a*) less than 1 MHz (*b*) more than 1 MHz
 (*c*) more than 10 MHz (*d*) more than 10 Hz.
38. The slewing speed in an *x-y* recorder refers to
 (*a*) time base
 (*b*) maximum constant velocity that the marking pen achieves
 (*c*) frequency response
 (*d*) relationship between inputs to *x* and *y* channels.
39. Which of the following recorder/display units has the highest frequency response and least response time?
 (*a*) X-Y plotters (*b*) U.V. recorders
 (*c*) Pen recorders (*d*) CRO.
40. An LCD requires a power of approximately
 (*a*) 20 W (*b*) 20 mW
 (*c*) 20 μW (*d*) 20 nW.

ANSWERS

A. Choose the correct answer

1. (*a*)	2. (*c*)	3. (*c*)	4. (*a*)	5. (*a*)	6. (*c*)	7. (*a*)
8. (*c*)	9. (*d*)	10. (*a*)	11. (*d*)	12. (*e*)	13. (*b*)	14. (*b*)
15. (*b*)	16. (*b*)	17. (*c*)	18. (*b*)	19. (*b*)	20. (*a*)	21. (*d*)
22. (*d*)	23. (*d*)	24. (*d*)	25. (*b*)	26. (*b*)	27. (*b*)	28. (*c*)
29. (*a*)	30. (*c*)	31. (*c*)	32. (*a*)	33. (*a*)	34. (*a*)	35. (*d*)
36. (*a*)	37. (*c*)	38. (*b*)	39. (*d*)	40. (*c*).		

B. Fill in the Blanks or say "Yes" or "No"

1. The first stage of the instrumentation or measurement system which detects the measurand is termed as .... stage.
2. Amplification means enhancement of the signal level which is given in the low level range.
3. Modulation means to change the form of signal.
4. .... transmission means to transmit signal from one location to another without changing the contents of the information.
5. D.C. amplifier is difficult to calibrate at low frequencies.
6. The major disadvantage of a D.C. amplifier is that it suffers from the problem of drift.
7. .... is a device which is used to increase or augment the weak signal.
8. The ratio of output signal to input signal for an amplifier is termed as gain or amplification.
9. A "Compound gear train" gives small modification.
10. The D.C. amplifiers are capable of amplifying static, slowly changing or rapid-repetitive input signals.

11. When "modulation" is used in instrumentation, frequency modulation is the more common form.
12. More commonly, the mixed signal and carrier are demodulated by rectification and filtering.
13. An .... is a linear integrated circuit that has a very high voltage gain, a high input impedance and a low voltage output impedance.
14. An-amps are linear integrated circuits that work on relatively .... supply voltage.
15. .... amplifier converts a voltage at high impedance to the same voltage at low impedance.
16. .... is a two-port resistive network and is used to reduce the signal level by a given amount.
17. Variable attenuators are used as control volumes in radio broadcasting stations.
18. .... is an electronic circuit which can pass or stop a particular band of frequencies through it.
19. A low pass filter is also called lag network.
20. Current telemetering is quite suitable for long distances.
21. The last stage of a measurement system is the .... presentation stage.
22. A good display, functionally, is one which permits the best combination of .... and .... when transferring the necessary information from the instrument to the operator.
23. .... type meters indicate the reading in exact numerals.
24. Almost all electronic voltmeters make use of the rectifying properties of ....
25. The .... instruments indicate the value of the measurand in the form of decimal number.
26. The analog meters work on the principle of quantization.
27. An analog instrument can be considered as a counter which counts the pulses in a predetermined time.
28. A P-N junction diode, which emits light when forward biased is known as a light emitting diode (LED).
29. Liquid crystal displays (LED) have extremely low power requirement.
30. A .... indicator tube consists of a gas filled glass tube having ten cathodes in the form of numbers and an anode.
31. Digclampter gives reading in .... form.
32. A .... records electrical and non-electrical quantities as a function of time.
33. .... type strip chart recorder operates on "comparison basis".
34. Instruments that record changes of only one measured variable are called .... recorders.
35. A .... recorder may have as many as 24 inputs, with traces displaced in 6 colours.
36. A X-Y recorder is an instrument, which gives a graphic record of relationship between two variables.
37. Magnetic tape recorders have response characteristics which enable them to be used at .... frequencies.
38. A CRO is an instrument which presents signal wave-forms visually.
39. A CRO cannot be used to compare two signals in phase, frequency or amplitude.
40. A CRO can be used for tracing transistor curves.

ANSWERS

B. Fill in the Blanks or say "Yes" or "No"

1. detector-transducer	2. Yes	3. Yes	4. Data
5. No.	6. Yes	7. Amplifier	8. Yes

9. No	10. Yes	11. No	12. Yes
13. Op-amp	14. low	15. Buffer	16. Attenuator
17. Yes	18. Filter	19. Yes	20. No
21. data	22. speed, accuracy, sensitivity		23. Digital
24. diodes	25. digital	26. No	27. No
28. Yes	29. Yes	30. Numerical	31. digital
32. recorder	33. Null	34. single point	35. multipoint
36. Yes	37. higher	38. Yes	39. No
40. Yes.			

THEORETICAL QUESTIONS

1. What do you mean by the following terms as applied to instrumentation or measurement system?
   (i) Detector-transducer stage.
   (ii) Signal conditioning stage.
2. State the limitations of mechanical amplification.
3. What are the advantages of electrical signal conditioning?
4. Explain briefly the following functions of signal conditioning equipment:
   (i) Amplification (ii) Modification or modulation
   (iii) Impedance matching (iv) Data processing
   (v) Data transmission.
5. Explain briefly the following:
   (i) D.C. signal conditioning system.
   (ii) A.C. signal conditioning systems.
6. Describe briefly the term "Amplification".
7. Explain briefly any two of the following amplifiers:
   (i) Mechanical amplifiers (ii) Fluid amplifiers
   (iii) Electrical and electronic amplifiers
8. State the generalities that can be listed for an ideal electronic amplifier.
9. What are A.C. and D.C. amplifiers? Explain briefly.
10. What do you mean by "Modulated and unmodulated signals"?
11. What is an Op-amp? State its limitations as well.
12. Explain briefly the term "Common-mode rejection ratio (CMRR).
13. State the applications of Op-amp.
14. Enumerate some of the commonly used Op-amp circuits.
15. Explain briefly the following:
    (i) Buffer amplifier. (ii) Differential amplifier.
16. State the advantages of differential amplifiers.
17. What is an attenuator? How are the attenuators classified?
18. What do you mean by the terms "Filtering" and "Filter"?
19. How are filters classified?
20. What do you mean by "Signal transmission"?
21. Explain briefly any three of the following types of transmission ?
    (i) Mechanical transmission (ii) Hydraulic transmission
    (iii) Pneumatic transmission (iv) Magnetic transmission.

22. Give five examples of electric type of transmitters.
23. What do you mean by "Converters"?
24. How are telemetering systems classified?
25. Explain briefly any two of the following types of telemetering systems:
    (*i*) Voltage telemetering (*ii*) Current telemetering
    (*iii*) Impulse telemetering (*iv*) Frequency telemetering.
26. What is the main purpose of any measurement system?
27. What is the function of the display or recording element of a generalised measurement system?
28. How does a display unit differ from a recorder?
29. How are the output devices categorized? Explain briefly.
30. How can we get machine interpretable outputs?
31. List the different forms in which the display is available from an instrument.
32. How are electrical indicating instruments classified?
33. How analog display meters differ from digital type meters?
34. Give a comparison between analog type and digital type instruments.
35. Give four examples each of the analog type and digital type instrumentation.
36. Explain briefly the following:
    (*i*) Single-point indicators.
    (*ii*) Multi-point multi-pointer and multi-range indicators.
37. Explain briefly the essential features of indicating instruments.
38. Describe briefly any two of the following :
    (*i*) Moving-iron instruments; (*ii*) Moving-coil instruments;
    (*iii*) Rectifier instruments.
39. What are advantages of electronic voltmeters?
40. What are digital instruments? State their principle of operation.
41. Explain briefly any two of the following:
    (*i*) Semiconductor light emitting diodes (LED);
    (*ii*) Liquid crystal displays;
    (*iii*) Hot filament or bar tubes;
    (*iv*) Numerical indicator tubes (NIT).
42. What is a recorder?
43. Elaborate the difference between a display unit and a recorder.
44. What is meant by a direct reading instrument?
45. Explain the functioning of a basic type of strip chart recorder. Enumerate the different types of marking mechanisms used in it.
46. Distinguish between single point and multi-point recorders.
47. What is a X-Y recorder? State its applications.
48. Explain the moving of an ultraviolet (U.V) recorder. State its applications.
49. What are the basic components of a magnetic tape recorder for instrumentation applications? List its advantages and disadvantages.
50. Explain with neat diagram the construction and working of a cathode ray oscilloscope (CRO).
51. Describe the different parts of a cathode ray tube (CRT).
52. What are the applications of a CRO?

CHAPTER

5

Microprocessors

5.1 **Computers-***Brief description* – History and development of computers – Definition of a computer – Characteristics of a computer – Classification of computers – Analog computers – Digital computers – Differences between analog and digital computers – Block diagram of a digital computer – Rating of chips – Computer peripherals – Storage devices – Hardware, software and liveware – Translators – Computer languages – Computer programming process for writing programs – Computing elements of analog computers; **5.2 Microprocessors** – Microprocessor – General aspects – Definition and brief description – Characteristics of microprocessors – Important features – Uses of microprocessors – ***Microprocessor systems*** – The microprocessor – Buses – Memory – Input/Output – ***Intel 8085 Microprocessor*** – Brief history – Introduction – Arithmetic and logic unit (ALU) – Timing and control unit – Registers – Data and address – Pin configuration – Opcode and operands – Microprocessor programming – ***Microcontrollers.***

5.1 COMPUTER—BRIEF DESCRIPTION

5.1.1. History and Development of Computers

- Charles Babbage (an English Mathematician) was responsible for conceiving the concept of the Modern computer, and is called "Father of Computers".
- He designed the early computer called *"Difference Engine"* in the year 1822, with which reliable tables could be produced. In 1833 he improved upon the machine and put forth new of idea of *"Analytical Engine"* which could perform the basic arithmetic functions automatically. In this machine punched cards were used as input/output devices for basic input and output.

The concept of use of punched cards was developed further by Horman Hollerith in the year 1889.

- Leonards Torres demonstrated a *digital calculating machine* in Paris in 1920.
- In 1944 Prof. Howard Aiken (Howard University) developed Electromechanical calculators known as Mark-I. This machine could handle about a sequence of 5 arithmetic operations by using memory for previous results.
- On the basis of research done for U.S. army during the World War-II in 1946, the first electronic computer, ENiAC (Electric Numerical Integer and Computer) was designed in 1946. This computer was about 15 metres long and 2 metre high and weighed about 50 tons. It consumed about 200 kW power. This machine did not have any facility for storing program.
- In 1949, the *concept of stored program was adopted.*
- In 1951, was introduced the commercial version of stored program computer UNIVAC-(Universal Automatic Computer)—the first digital computer.

Generations of computers:

First generation .... Developed during the years 1951–1959.

- These computers are *"based" on* ***"Vacuum tubes"***.
- Very slow in operation (10^3 operations/sec.)
- Big in size and unreliable.
- Short span of life.
- Frequent breakdowns.
- High power consumption and great amount of heat generation.
- Small primitive memories and no auxiliary storage.
- Limited programming capabilities.

Examples: UNIVAC-1 and IBM 650.

Second generation. Developed during the years 1960–1965.

- These computers are *based on* **"Transistors"**.
- Faster in operation, comparatively (10^6 operations/sec.)
- Smaller in size.
- More reliable.
- Consume less power.
- Generate less heat than vacuum tubes.
- Auxiliary memory in the form of magnetic tape was introduced.

Examples: UNIVAC 1107, IBM 7090, CDC 1604, Honeywell 800 etc.

Third generation. .... Introduced during 1965–1970, also being used presently.

- These computers are based on ***"Integrated circuits"***, based on silicon technology.
- Much more smaller in size.
- More reliable.
- Faster in operation (10^9 operation/sec).
- Less expensive.
- Employ higher capacity internal storage.
- Wide range of peripheral used.
- Make use of new concepts like *multi-programming, multi-processing, high level languages.*

Examples: IBM-360/370, Honeywell 6000.

Fourth generation .... Introduced in 70s.

- These computers are based on VLSI (Very large scale integration) chips and microprocessors chips.
- Possess high processing power.
- Low maintenance.
- Faster in operation.
- High reliability.
- Very low power consumption.
- Less expensive.
- Small size.

This generation also includes the following :

— Microcomputers;

— Office automation systems;

— Distributed processing systems.

Fifth generation .... Introduced during late 1990s.

- These computers use optic fibre technology to handle *Artificial Intelligence, Expert Systems, Robotics* etc.
- Possess very high processing speeds.
- More reliable.

5.1.2. Definition of a Computer

A **computer** is *a machine that processes data according to set of instructions stored within the machine.*

- It *receives data as input, processes the data, i.e.,* performs arithmetic and logical operations on the same and *produces output in the desired form on output device as per the instructions coded in the program.*
- The processing function of the computer is *directed by the stored program, a set of codes instructions stored in the memory unit, which guides the sequence of steps to be followed during processing.*

5.1.3. Characteristics of a Computer

The following are the *characteristics* which make a computer an indispensable unit :

1. Speed
2. Consistency
3. Accuracy
4. Flexibility
5. Reliability
6. Large storage capacity
7. Automatic operation
8. Diligent
9. No emotional ego and psychological problems.

Limitations of a computer :

A computer entails the following *limitations* :

1. It does not work on itself, a set of instructions is required for its operation.
2. It cannot take decision on its own, it has to be programmed as per requirements.
3. It is not intelligent, it has to be instructed in detail for the performance of each and every task.
4. It cannot learn by experience, as human beings do.

5.1.4. Classification of Computers

The computers may be *classified* as follows:

1. **On the basis of the type of data :**

   (*i*) *Analog computers* (These computers process the data in analog form).

   (*ii*) *Digital computers* (These computers process the data in digital form).

2. **On the basis of the size and capacity :**
   (*i*) *Microcomputers*
   (*ii*) *Mini computers*
   (*iii*) *Main frame*
   (*iv*) *Super computers.*
3. **On the basis of the type of application :**
   (*i*) *Special purpose computers*
   (*ii*) *General purpose computers.*
4. **On the basis of the number of users :**
   (*i*) *Single user computers*
   (*ii*) *Multi-user computers.*
5. **On the basis of the number of processors :**
   (*i*) *Single processor computers.*
   (*ii*) *Multiprocessor computers.*
6. **On the basis of the type of instructions set :**
   (*i*) *Complex Instruction Set Computers* (CISC).
   (*ii*) *Reduced Instruction Set Computers* (RISC).

5.1.5. Analog Computers

- The *principle of operation* of analog computers is *to create a physical analog of mathematical problems.*
- Measure physical variables *continuously.*
- Use signals as input (which may be supplied by devices like barometers, speedometers, thermometers etc.).
- The *result* given by an analog computer is *not very precise, accurate and consistent.*
- These computers find limited applications.

Example: Speedometer of a vehicle (here speed varies continuously).

5.1.6. Digital Computers

- The digital computers *accept digits and alphabets as input.*
- Receive data in the form of discrete signals representing ON (high) or OFF (low) voltage.
- The data input can be represented as sets of O's and 1's representing low and high respectively.
- The digital computers convert data into discrete form before operating on it.
- The *most important characteristic* of a digital computer is that it is general purpose device *capable of being used in a number of different applications.* By changing the stored program, the same machine can be used to implement totally different tasks.

Example: Digital watches.

Digital computers may be further classified based upon : (*i*) Purpose of use (*e.g.*, General purpose, special purpose); (*ii*) Size and capabilities.

On the basis of *size and capabilities*, the digital computers are classified as :

1. Super computers.

2. Mainframe computers.
3. Medium sized computers.
4. Mini computers.
5. Micro computers.

1. **Super computers :**
   - These computers are the fastest (speed of calculations upto 1.2 billion instructions per second) and have very high processing speeds.
   - They are very large in size and most powerful and costliest.
   - Their fields of applications include processing weather data, geological data, genetic engineering etc.
   - Word length : 64 bits and more.
   - These computers can receive input from more than 1000 individual work stations.

   Example: PARAM (a super computer developed in India).

2. **Main frame computers :**
   - These are large scale general purpose computer systems.
   - Possess large storage capacities in several million words.
   - Secondary storage directly accessible—of the order of several billion words.
   - Can support a large number of terminals (upto 100 or more).
   - Faster in operation (100 million instructions/sec. approx).
   - Accept all types of high level languages.
   - Word length—16 or 32 or 64 bits.

3. **Medium sized computers :**
   - Mini versions of mainframe computers.
   - They have smaller power than mainframes.
   - Processing speeds relatively high with support for about 200 remote systems.

4. **Mini computers :**
   - These are general purpose computer systems.
   - Reduced storage capacity and performance (as compared to main frame).
   - CPU speed—few million instructions/sec.
   - Word length—16 or 32 bits.
   - Can accept all types of high level languages.
   - Can support upto about 20 terminals.

   Note: In view of fast development in electronics it is difficult to draw a line of demarcation between small main frame computers and large mini-computers.

5. **Micro-computers :**
   - These are small sizer computers *utilising microprocessors*. These are popularly known as personal computer (PC).
   - CPU is usually contained on one chip.
   - Possess low storage capacity (maximum being 256 K words).
   - Slow in operation (10^5 instructions/sec.).

- Usually provided with *video display unit, floppy drive and printer.* Some micro-computers can support hard disc also.
- Maximum word length is 16 bits; however most of these use 8-bit words.
- Commonly used language—BASIC. However these computers can also accept other high level languages viz., PASCAL, FORTRAN etc.

Note: **A single chip microcomputer* consists of a single chip on which the central processing unit, input/output and memory units are integrated. This is used for *industrial applications* and also in *product calculators.*

*Its *advantage is the reduction in cost and size, increase in performance and reliability.*

5.1.7. Differences between Analog and Digital Computers

The differences between analog and digital computers are given in Table 5.1.

Table 5.1. Differences between Analog and Digital Computers

S. No.	Digital computer	Analog computer
1.	It performs calculations by counting and thus counts directly. It is the most *versatile machine.*	It processes work electronically by *analogy.* It does not produce number but produces its results *in the form of graph.* It is *more efficient in continuous calculations.*
2.	It operates on inputs which are on-off type (being digits 0 or 1) and its output is in the form of signals.	It accepts variable electrical signals (analog values) as inputs, and its output is also in the form of analog electrical signals.
3.	It is based on counting operation.	It operates by measuring analog signals.

These days *digital computers* are being *widely used.*

A *hybrid computer* is combination of both analog and digital computers. It is used for *simulation applications.*

5.1.8. Block Diagram of a Digital Computer

Figure 5.1 shows a block diagram of a typical digital computer.

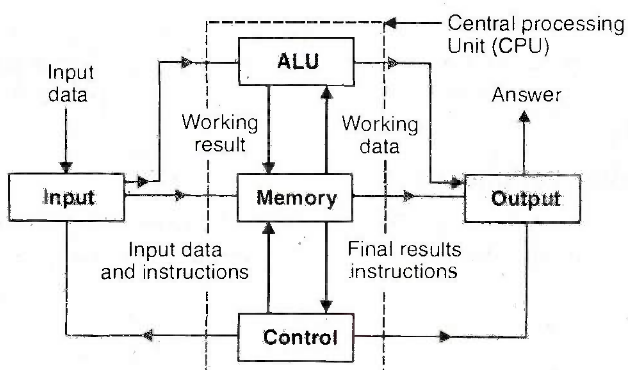


Fig. 5.1. Block diagram of a digital computer.

The following are the *five basics elements* of a computer system :

1. **Input :**
   - The data and instructions are first recorded on a machine readable medium, like punched card, and then fed into the computer via a device that codes them in a manner which is suited to conversion into electrical pulses before entering memory.
   - The input supplies data to the computer in digital (binary) form.
2. **Memory :**
   - The memory section within the computer is where data are stored or memorized.
   - Problem to be solved, inputs for the problem, a program of instruction, working data, intermediate results and final results are *types of memory data.*
   - The memory section holds data between high speed computer operation and slower input and output devices.
3. **Arithmetic Logic Unit (ALU) :**
   - ALU performs necessary arithmetical operations on the data and ensures that instructions are obeyed.
   - It also performs *logical operations.*
   - The ALU combined with control unit is called *Central Processing Unit (CPU).*
4. **Control Unit :**
   - It fetches instructions from main memory, interprets them and issues the necessary signals to the components making up the system.
   - It issues commands for all hardware operations necessary in obeying instructions.
5. **Output :**
   - The output is the path for data out of the computer and may include devices for reading out answers.

5.1.9. Rating of Chips

Chips are rated in terms of their *capacity* and *speed.*

- *Capacity* of a chip refers to the amount of *kilo-bites it can store.*
- *Chip speed* refers to the rate at which the microprocessor can write to the chip. It is usually measured in nano-seconds (ns). As the chip speed increases, its cost also goes up.

5.1.10. Computer Peripherals

A **peripheral** *is any device commonly used with a CPU of a computer for input or output of information or for memory functionally separate from the CPU and electronically detachable.*

Input devices :

1. *Keyboard :*
   - It is the most common and simplest input device.
   - It is merely a collection of momentary switches. The outputs of the key switches are fed to electronic circuitry known as *keyboard encodes* which converts them

into binary coded values. The values are then fed into the computer which interprets the key which was pressed. Thus the function of the key changes with the type of work we are doing.

2. *Mouse :*
   - It is a pointing device and its size is about the size of palm.
   - It is a hand-held device that *controls a pointer on the screen.*
   - It rolls on a small ball. A mouse has one or more buttons on the top. When the user moves the mouse over a flat surface, the screen cursor moves in the direction of the mouse movement.
3. *Digitizer (or Graphic tablet) :*
   - It is similar to light pen.
   - It consists of a glass plate on which digitizing tablet is moved.
   - It is used for fine drawing works and for image manipulation applications such as Auto-cad.
4. *Optical Mark Reader (OMR) :*
   - OMR is being used for reading the answer sheet by means of light. It can read upto 150 documents per minute; when on-line with respect to the computer system, can read upto 2000 documents per minute.
   - OMR can also be used for such applications as *order writing payroll, inventory control, insurance, questionnaires* etc.
5. *Magnetic Ink Character Reader (MICR) :*
   - MICR uses a special ink to print character. These characters can be decoded by special magnetic devices.
   - This system is employed *by banks for processing cheques.*
6. *Scanner :*
   - It is used for getting existing graphical image (like photographs, mats, etc.) into computer.
   - Once the graphical image is scanned and brought into the computer user can include them into documents or can edit them.
7. *Light pen :*
   - It consists of a pen like device and photoelectric cell.
   - It is used to draw pictures on the screen.
   - When light pen is in contact with screen, the electron beam activates the photoelectic cell which in turn sends signals into the computer and ultimately a mark is made on the screen where light pen contacted the screen.
8. *Joy-stick :*
   - It is screen-pointing device.
   - A stick is present with a button at the top. It can be held in the hand and bent in any one of the four directions. As the stick is moved, the action on the screen changes in the appropriate direction.
   - A joy-stick is *widely used for playing computer games.*
9. *Touch screen :*
   - The touch screen technique involves beam and ultrasonic waves.

- By using touch screen we can issue command to the computer by touching the screen.
- Limited amount of data can be entered via a terminal or a microcomputer that has a touch screen.

10. *Compact Disk Read Only Memory (CDROM) :*
    - It is a 120 mm diameter disc with a polycarbonate subtrate, a reflective metalised layer on one side, with a protective lacquer finish.
    - Here a laser beam is used to burn a small hole or pit which represent binary '1'. The absence of pit represents '0'. In this way digital information is stored on the disc in large quantities (in Giga Bytes).
11. *Voice Recognition System or Voice Synthesizer :*
    - Voice recognition techniques, along with several other techniques, are used to convert the voice signals to appropriate words and device the correct meanings of words. There has been a limited success in this area and these days devices are available commercially to recognize and interpret human voices.

Output devices :

1. *Printer :*
    - A printer is *device that produces copies of text and graphics on paper.*
    - The printers are classified/categorised as follows:

      A. *Impact printers :*
        (*i*) Solid font
        (*ii*) Dot matrix.

      B. *Non-impact printers :*
        (*i*) Thermal printer (*ii*) Inkjet printer
        (*iii*) Laser printer (*iv*) Electrographic printer
        (*v*) Electrostatic printer.
2. *Plotters :*
    - Plotters are those devices which *reproduce drawings using pens that are attached to movable arms.*
    - Platting in different colours is possible.
3. *Monitors or Visual Display Unit (VDU) :*
    - A monitor is a television like device, which is used to display information, output and input data.
    - It consists of a cathade ray tube (CRT), on which the information is displayed. When the user processes any key on the keyboard, the keyboard encoder generates code of that key which is depressed. This code is then fed to the computer; from there VDU system takes that code and displays it on the screen.

5.1.11. Storage Devices

The memory devices in a memory unit (which stores the data, instructions and intermediate results) may be of the following types :

1. Internal storage device ..... also known as *main* or *primary* storage device.

The primary storage devices currently in use in computers are :

(*i*) Magnetic core memory device

(*ii*) Thin film memory device

(*iii*) Thin rod memory device

(*iv*) Plated wire memory device.

2. Auxiliary storage device

The popular *secondary memory devices* are :

(*i*) Magnetic tape drive

(*ii*) Magnetic disk drive

(*iii*) Magnetic drum

(*iv*) Floppy disk

(*v*) Winchester disk.

Methods of Input to Backing Stores :

The following methods are generally used :

(*i*) Key-to-tape

(*ii*) Key-to-cassette/cartridge

(*iii*) Key-to-disk/diskette

Memory. The memory is used to store information/data so that it can be retrieved whenever required. There are mainly two types of memories :

1. Primary memory
2. Secondary memory.

1. **Primary memory :**
   - It is also known as core memory, main memory, RAM (Random Access Memory).
   - It is constructed using purely semiconductor devices, data is stored in the form voltages.
   - It is a volatile memory whereas ROM (Read Only Memories) are non-volatile memories.

2. **Secondary memory :**
   - Secondary memory, also known as *auxiliary memory*, is used to store large volumes of data.
   - Data is stored in the form of magnetic energy and can be stored (in the secondary memory) for large periods.

Difference between Read Only Memory (ROM) and Random Access Memory (RAM).

ROM :

- As the name implies ROM is a memory unit that performs the read operation only; it *does not have a write capability*. This implies that the binary *information stored in a ROM is made permanent during the hardware production of the unit and cannot be altered by writing different words into it*. Whereas a *RAM is a general-purpose device whose contents can be altered during the computational process.*

- ROM is a type of memory chip that we can read only and we cannot write on it.
- ROM provides permanent storage for program instructions.
- The most important ROM chip in any computer is ROM BIOS (Basic Input/Output System).
- ROM is most oftenly used in microprocessors that always execute the same program such as boot strap loader.

Disadvantage of ROM :

(*i*) A ROM is prepared by the manufacturer and cannot be altered once the chip has been made.

(*ii*) It is slow.

The *ROM memory* may be **classified** as follows:

(*i*) *Programmable Read Only Memory (PROM).* Here, the information can be altered, but *not as easily as in the ordinary memory.* Once the operations to be performed have been written into a PROM chip, they are permanent and cannot be changed.

(*ii*) *Erasable Programmable Read Only Memory (EPROM).* This type of ROM can be erased and programmed with the help of special equipment. It has a window at its top, which if exposed to ultraviolet light, allows data to be erased.

(*iii*) *Electrically Erasable Programmable ROM (EEPROM).* In order to erase and reprogramme this type of ROM, it is required to be removed from the socket.

(*iv*) *Flash EPROM.* It is the latest type of ROM. A manufacturer can make changes to the flash EPROM while it remains in the PC, by running a special program.

RAM :

- This memory is so named since memory registers can be accessed for information transfer as required.
- RAM chip is made with Metal Oxide Semiconductor (MOS).
- RAM chips may be classified as :
  (*i*) *Dynamic Ram:* It provides volatile storage (*i.e.*, the data stored is lost in the event of a power failure).
  (*ii*) *Static RAM:* These chips are more complicated and take up more space for a given storage capacity than dynamic RMA chips. These chips are also volatile in nature but as long as they are supplied with power, they need not require special regenerator circuits to retain the stored data.
- *Static RAM chips are thus used in specialised applications while Dynamic RAM chips are used in the primary locations.*
- Owing to the volatile nature of these storage elements, a back up *Uninterrupted Power System (UPS) is often installed along with larger computer systems.*

5.1.12. Hardware, Software and Liveware

Hardware :

The set of *physical components, modules and peripherals comprising a computer system* is called ***Hardware.***

Apart from wires and nut bolts, the major hardware components of computer are :

(*i*) Input-output devices

(*ii*) Control unit

(*iii*) Memory

(*iv*) ALU.

Software :

The software is a *set of programs required for data processing activities of the computer.* In other words, the program written in any one of the computer languages, is called *software.*

System software includes the following :

(*i*) Operating systems

(*ii*) Language processors (assembles, compilers, interpreters)

(*iii*) Utility program

(*iv*) Subroutine program.

Liveware :

All persons concerned with computers, *i.e.,* complier, programmer, etc. are called *liveware.*

5.1.13. Translators

A *translator* is a software program which converts statements written in one language into another *e.g.,* converting assembly language to machine code etc. The assembly language program is called *'source program'* and the machine code program is called *'object program'.*

There are three types of *translators* :

1. Assembler
2. Compiler
3. Interpreter.

5.1.14. Computer Languages

1. ***Machine language.*** It is a programming language in which the instructions are in a form which allows the computer to perform them immediately, without any further translation. Instructions in machine language are *in the form of a binary code,* also called machine code and are known as machine *instructions.*
2. ***Low level language.*** Low level languages are machine-oriented languages in which each instruction corresponds or resembles a machine instruction. The low level language must be translated into machine language before use.
3. ***High level language.*** The development of high level language was intended to overcome main limitations of level language. The high level languages have an extensive vocabulary of word, symbols and sentences.

Different types of high level languages are :

(*i*) *Commercial languages.* ... The most well commercial language is COBOL (Commercial Business Oriented Language).

(*ii*) *Scientific language.* ... The most well-known languages among this group are :

(*a*) ALGOL (Arithmetic Oriented Language)

(*b*) FORTRAN (Formula Translation)

(*c*) BASIC (Beginner All Purpose Symbolic Instruction Code).

(*iii*) *Special purpose language.*

(*iv*) *Command language.*

(*v*) *Multipurpose language.*

5.1.15. Computer Programming Process for Writing Programs

The complete computer programming process followed by programmer for writing comprises the following **steps** :

1. Analysis
2. Flow charting
3. Coding
4. Debugging
5. Documentation
6. Production.

5.1.16. Computing Elements of Analog Computers

1. *Attenuators* ...... are used to multiply a variable quantity by a constant.
2. *Summing amplifiers* ...... are used to add or subtract variables as required.
3. *Servo multipliers* ...... are used to multiply two variables.
4. *Function generators* ...... are used to simulate the arbitrary behaviour of variables.
5. *Integrating amplifiers* ...... are used to integrate a variable with respect to time.

5.2 MICROPROCESSORS

5.2.1. Microprocessor—General Aspects

5.2.1.1. Definition and brief description

A ***microprocessor*** *is a large scale integration (LSI) chip that is capable of performing arithmetic and logic functions as defined by a given programme.* This system by itself does not form an operational computer, and additional circuit for memory and input/output must be supplied and interfaced with the system. The software (it is the programme for controlling the operation of the microprocessor itself) is also to be provided.

Or

A state machine on a single IC chip with very large scale integration, capable at a desired instant of working as per programme or an instruction of a programme, and which is driven by a clock of frequency of 1 MHz or more, is called a ***microprocessor***. Such machine is also called a central processing unit (CPU). A CPU forms main part of a computer.

The ***microprocessor*** consists of the following *three segments* (See Fig. 5.2).

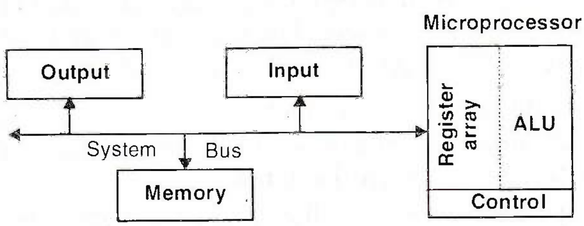


Fig. 5.2. Block diagram of a microcomputer.

1. **Arithmetic/Logic Unit (ALU).** In this area of the microprocessor, computing functions are performed on data. The ALU performs arithmetic operations such as addition and subtraction, and logic operations such as AND, OR and exclusive OR. Results are stored *either* in registers or in memory or sent to output devices.
2. **Register unit.** This area of the microprocessor consists of various registers. The registers are used primarily to store data temporarily during the execution of a program. Some of the registers are accessible to the user through instructions.

3. **Control unit.** The control unit provides the necessary timing and control signals to all the operations in the microcomputer. It controls the flow of data between the microprocessor and peripherals including memory.

In short a *microprocessor performs the following functions* :

- Communicates with all peripherals (memory and I/O) using system.
- Controls timing of information flow.
- Performs the computing tasks specified in a programme.

5.2.1.2. Characteristics of microprocessor

In nearly every type of design, with any complexity at all, microprocessors have potential for *drastically reducing component count and shortening design time.* In fact a microprocessor is considered to represent long-awaited next generation of digital building blocks, and that microprocessor will provide the best single approach to the system-level digital integrated circuit.

Some of the characteristics of a microprocessor are listed below :

1. It *handles shorter words than other computers,* usually 4 to as many as 16 bits.
2. It consists of *integrated circuits* from 1 to 30 in number.
3. It contains arithmetic logic unit (ALU), registers, control, random access memory (RAM), data buses and read only memory (ROM) with programmes.

5.2.1.3. Important features

The important features of the microprocessors are :

1. Low cost
2. Small size
3. Low power consumption
4. Versatile (The versatility of a microprocessor results from its 'stored programme' mode of operation).
5. Extremely reliable.

Note: Probably the term 'micro' in the name of the device can be contributed to its *low cost, small size and low power consumption.* The processing capability of a microprocessor should not, however, be underestimated. Currently available 32-bit microprocessors have a processing power similar to that of the mainframe computer of a few years ago. Even the early 8-bit microprocessors are powerful enough to perform several applications.

5.2.1.4. Uses of microprocessor

The processing power of the 8-bit microprocessors is more than adequate to satisfy the requirements of most of the *instrumentation applications.* By making an instrument microprocessor-based, it can be made *intelligent by incorporating new features like programmability* which cannot be easily provided in its hard-wired counterpart.

Some important *uses* of microprocessors in instrumentation area are listed below :

1. Frequency meters.
2. Function generators.
3. Frequency synthesizers.
4. Spectrum synthesizers.
5. *Intelligent instruments CRT terminals*
6. Digital millimeters.
7. Oscilloscopes.

8. Counters.
9. **Process control**
   — Instrumentation
   — Monitoring and control
   — Data acquisition
   — Logging and processing.
10. **Medical Electronics**
   — Patient-monitoring in intensive care unit
   — Pathological analysis
   — Measurement of parameters like blood pressure and temperature.

Under this heading the following instruments/machines are included:

(*i*) Microprocessor based medical instrument.
(*ii*) Microprocessor based ECG machines.
(*iii*) Microprocessor based EEG machines etc.

Other Applications of microprocessors :

(*i*) High level language computers.
(*ii*) Replacing hard-wired logic by a microprocessor.
(*iii*) Control of automation and continuous processes.
(*iv*) Computer peripheral controllers.
(*v*) Home entertainment and games.
(*vi*) Inventory control system, pay roll banking etc.

5.2.2. Microprocessor Systems

Microprocessor systems consist of the following *three* parts :

1. *Central processing unit (CPU) :*
   — This part uses the microprocessor.
   — It recognises and carries out program instructions.
2. *Input and output interfaces :*
   — These interfaces handle communications between the computer and the outside world.
   — For the interface, the term *port* is used.
3. *Memory :*
   — Its function is to hold the program instructions and data.

- *"Microprocessors" which have memory and various input/output arrangements on the same chip are called* ***microcontrollers.***

5.2.2.1. The microprocessor

The microprocessor (generally referred to as CPU) is that part of the processor system which carries out the following functions :

(*i*) Processes the data ;
(*ii*) Fetches instructions from memory;
(*iii*) Decodes and executes the instructions.

Following are the various parts of a microprocessor :

1. *Arithmetic and logic unit (ALU):* This part of the microprocessor *manipulates the data.*

2. *Registers:* Registers are memory locations within the microprocessor and are employed to store information involved in program execution. The various types of registers used are :

(*i*) *Accumulator register :*

— It is a temporary holding register for data to be operated on by the arithmetic and logic unit and also, after the operation, the register for holding the results.

— It deals with all data transfers associated with the execution of arithmetic and logic operations.

(*ii*) *Status register :*

— The status register (also called flag register or condition code register) contains information concerning the result of the latest process carried out in the ALU.

— It contains individual bits with each bit having special significance; the bits are called flags.

— The status of the latest operation is indicated by each flag with each flag being set or reset to indicate a specific status.

(*iii*) *Program counter register (PC) :*

— This register, also called instruction pointer (IP), contains the address of the memory location that contains the next program instruction.

— This register is *updated* after the execution of each instruction, so that it contains the memory location where the next instruction to be executed is stored.

(*iv*) *Memory address register (MAR) :*

— This register contains the address of data.

(*v*) *Instruction register (IR) :*

— This register stores an instruction.

— The control processing unit (CPU), after fetching an instruction from the memory via the data bus, stores it in the instruction register. The microprocessor, after each such fetch, increments the program counter by one with the result that the program counter points to the next instruction waiting to be fetched. The instruction can then be decoded and used to execute an operation. This sequence is known as *fetch-execute cycle.*

(*vi*) *General purpose registers :*

— These registers serve as temporary storage for data or addresses and *used in operations involving transfers between other registers.*

(*vii*) *Stack pointer register (SP) :*

— The *stack* is a special area of the memory in which program counter values can be stored when a subroutine part of the program is being used.

— The contents of this register form an address which defines the top of the stack in RAM.

The number and form of the registers depends on the microprocessor concerned.

5.2.2.2. Buses

Buses are the paths along which digital signals move from one section to another.

— A bus is just a number of conductors along which electrical signals can be carried. It might be tracks on a printed circuit board or wires in a ribbon cable.

In a microprocessor system there are the following three forms of bus :

1. Data bus;
2. Address bus;
3. Control bus.

1. ***Data bus :***
   — The data bus carries the data associated with the processing function of the CPU.
   — Word lengths used may be 4, 8, 16, 32 or 64.
   — Each wire in the bus carries a binary signal, *i.e.*, a0 or a1.
   — The more wires the data bus has the longer the word length that can be used.

- The earliest microprocessors were 4-bit (word length : $2^4 = 16$) devices and such 4-bit microprocessors are still used in such devices as toys, washing machines etc. They were followed by 8-bit microprocessors (*e.g.*, Motorola 6800, the Intel 8085 A and the Zilog Z80). Now 16-bit, 32-bit and 64-bit microprocessors are available.

2. ***Address bus :***
   — It carries signals which indicate where data is to be found and so the selection of certain memory locations or input or output ports.
   — Each storage location within a memory device has a unique identification, termed its address, so that system is able to select a particular instruction or data item in the memory.
   — Each input/output interface also has an address.
   — When a particular address is selected by its address being placed on the address bus, *only that location is open to the communications from the CPU*. The CPU is thus *able to communicate with just one location, at a time*.
   — A computer with an 8-bit data has typically a 16-bit wide address bus, *i.e.*, 16 wires. This size of address enables 2^{16} locations to be addressed. 2^{16} is 65 536 locations and is usually written as 64 K, where K is equal to 1024.

3. ***Control bus :***
   — This bus *carries the signals relating to control actions*.
   — It is also used *to carry the system clock signals*; these are to synchronise all the actions of the microprocessor system.

5.2.2.3. Memory

In a microprocessor, the memory unit stores binary data and takes the form of one or more integrated circuits (ICs).

— The data may be program instruction codes or numbers being operated on.
— *The size of the memory is determined by the number of wires in the address bus.*

Following are the various forms of memory unit :

1. ROM
2. PROM
3. EPROM
4. EEPROM
5. RAM.

1. ***ROM :***
   — ROM (*Read Only Memory*) is a memory device in which *data is stored permanently*.

remains permanently in the chip until *erased by shining ultraviolet light through a quartz window on the top of the device.*

4. *EEPROM :*

— EEPROM (Electrically erasable PROM) is similar to EPROM; erasure, however, is done by the application a relatively high voltage rather than using ultraviolet light.

5. *RAM :*

— RAM (Random-access memory) is a read/write memory in which data currently being operated on (temporary data) is stored.

— Such a memory can be read or written to.

— When RAM is used for program storage then such programs are referred to as ***software.*** When the system is switched on, software may be loaded into RAM from some other peripheral equipment such as a keyboard or hard disc or floppy disc.

Difference between a software of a computer and a microprocessor :

In computer software is loaded into the computer at the beginning of each computation, software in microprocessor is stored within the computer itself in a ROM chip. The modification of the program is achieved by merely replacing ROM IC with another ROM IC containing a different control program. This as a very notable advantage of software implementation in microprocessor.

5.2.2.4. Input/Output

The transfer of data between the microprocessor and the external world is termed as the ***input/ output operation.***

— The pieces of equipment that exchange data with a microprocessor system are called ***peripheral devices.***

— In *input operations* the input device places the data in the data register of the interface chip; this holds the data until it is read by the microprocessor. In *output operations* the microprocessor places the data in the register until it is read by the peripheral.

5.2.3. Intel 8085 Microprocessor

5.2.3.1. Brief history

— Intel Corporation in early seventies introduced the first microprocessor, *Intel 4004*. This microprocessor was a single chip device which was capable of performing simple arithmetic and logic operations such as addition, subtraction, comparison, AND, and OR. Its control unit could perform various functions such as fetching of an instruction from the memory, decoding it and generating control pulses for executing it. It was a 4-bit microprocessor operating upon 4-bits of data at a time.

— Intel introduced *4040* as modified version of microprocessor 4004.

— Intel, later on, introduced 8-bit microprocessors called 8008 and 8080 which could perform arithmetic and logic operations on 8-bit words.

— These days, modified and better version of 8-bit microprocessor is **Intel 8085** which is most widely used and most popular microprocessor.

Now-a-days 12-bit, 16-bit and 32-bit microprocessors are also available.

Fig. 5.3, *shows the block diagram of Intel 8085 microprocessor.*

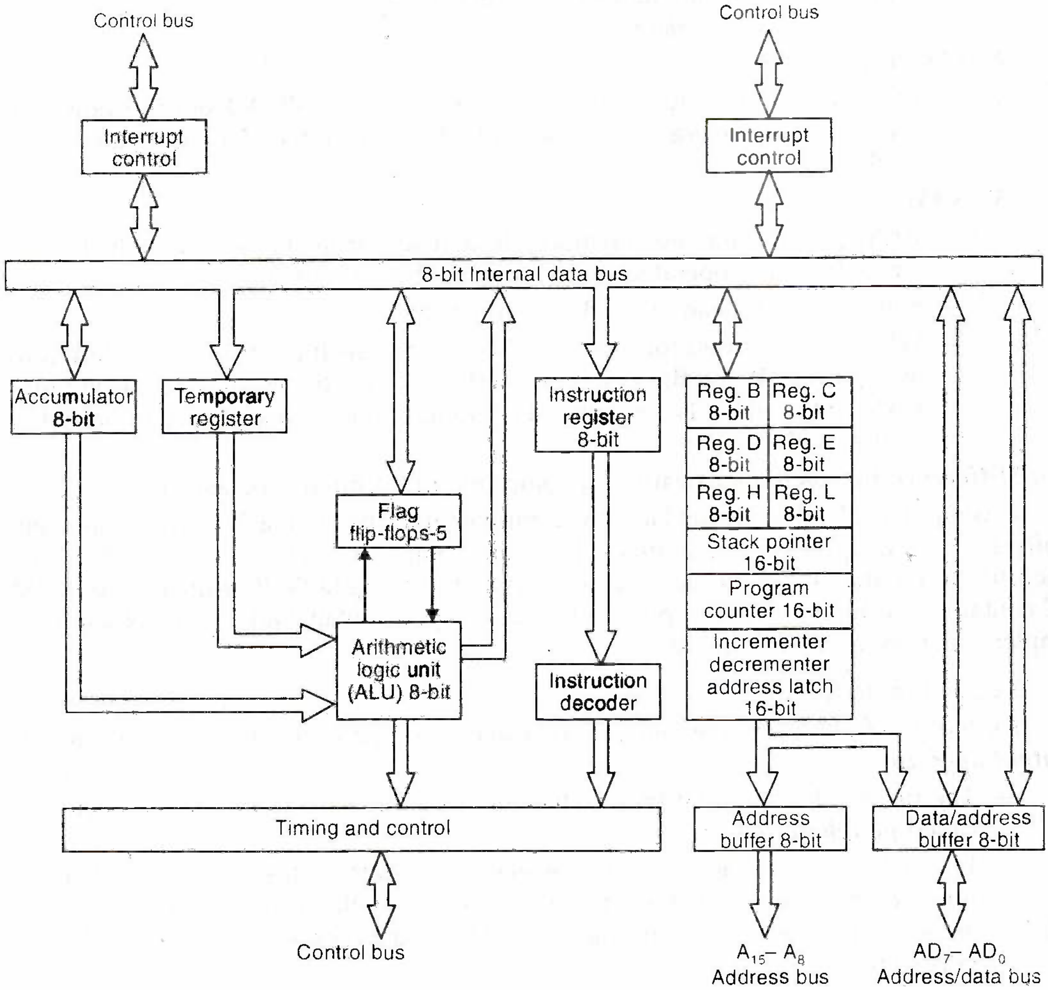


Fig. 5.3. Block diagram of Intel 8085 microprocessor.

5.2.3.2. Introduction

- Intel 8085 is an 8-bit, NMOS microprocessor.
- It is a 40 pin I.C. package fabricated on a single LSI chip.
- It uses a single +5 $V_{D.C.}$ supply for its operation.
- Its clock speed is about 3 MHz. The clock cycle is of 320 ns. The time for the clock cycle of the Intel 8085 AH-Z version is 200 ns. It has *80 basic instructions and 246 op-codes.*

5.2.3.3. Arithmetic and logic unit (ALU)

ALU performs the following arithmetic and logical operations :

1. Addition
2. Subtraction
3. Logical AND
4. Logical OR
5. Logical EXCLUSIVE OR
6. Complement (logical NOT)
7. Increment (add 1)
8. Decrement (subtract 1)
9. Left shift, Rotate left, Rotate right.
10. Clear etc.

5.2.3.4. Timing and control unit

— This unit is a section of CPU.

— It *generates timing and control signals* which are necessary for the execution of instructions.

— It controls data flow between CPU and peripherals (including memory).

— It provides status, control and timing signals which are required for the operation of memory and input/output devices.

— *It controls the entire operations of microprocessor and peripherals connected to it.*

5.2.3.5. Registers

***Registers** are digital devices used by the microprocessor for temporary storage and manipulation of data and instructions.* Data remain in the registers till they are sent to the memory or I/O devices.

- Many registers use the D-type flip-flop although J-K flip-flop is commonly used as well. Both types are readily available as commercial MSI units.
- Operationally, registers exhibit two notable characteristics: they are edge-triggered devices and all switching of flip-flops is synchronised by applying the clock pulse to each flip-flop simultaneously.

Activation of the register itself is achieved by means of an appropriate control signal.

Registers like counters may be either *parallel registers* or *serial (shift) registers*, although combined versions are also possible.

— In *parallel registers* all the binary data that appear at input terminals of the flip-flops are transferred to the output terminals in a single clock pulse. This makes the operation of the register very fast; it is the reason for its preference in digital computers.

— *Serial or shift register* processes each bit of word in succession and, therefore, operation is slow. However the shift register does offer the compensating advantage of requiring less equipment.

Intel 8085 microprocessor has the following registers :

1. One 8-bit accumulator (ACC), *i.e.*, register A.

2. Six 8-bit general purpose registers (B, C, D, E, H and L).
3. One 16-bit stack pointer, SP.
4. One 16-bit program counter, PC.
5. Instruction register.
6. Temporary register.

These registers are described below :

1. *Accumulator (ACC) :*
   - It is an 8-bit special purpose register that is a part of the ALU. It is also identified as register A.
   - In arithmetic and logical operations the accumulator may store the operand, execute an instruction with the help of other registers, and memory and finally store the result of the operation. In the former case it acts as a source, and in the latter a destination.
2. *General purpose registers :*
   - The 8085 microprocessor contains six 8-bit general purpose registers. These are identified as B, C, D, E, H and L as shown in Fig. 5.3.
   - These registers are used in microprocessor for temporary storage of operands or intermediate data in calculations.
   - These registers can be used either simply for storage of 8-bit data or in pairs for storage of 16-bit data. When used in pairs, only selected combination can be used for pairing, *i.e.*, B-C, D-E and H-L. When registers are used in pairs the high order byte resides in the first register and low order byte in the second register.
3. *Stack pointer (SP) :*
   - It is a 16-bit special function register.
   - The *stack* is a sequence of memory locations set aside by a programmer to store/retrieve the contents of accumulator, flags, program counter and general purpose registers during the execution of a program. Any portion of the memory can be used as a stack.
   - In this register, data is stored temporarily on first come and last go basis.
4. *Program counter (PC) :*
   - It is a 16-bit special-purpose register and is used to hold the memory address of the next instruction to be executed.
   - The contents of the PC are automatically updated by the microprocessor during the execution of an instruction so that at the end of execution it points to the address of the next instruction in the memory.
   - The microprocessor uses the PC for sequencing the execution of instructions.
5. *Instruction register :*
   - During the execution of a program, microprocessor addresses some memory which supplies an 8-bit data of instruction code to the data bus which gets stored in the register called the *instruction register.*
   - The instruction register holds the op-code (operation code or instruction code) of the instruction which is being decoded and executed.

6. *Temporary register :*
   - It is an 8-bit register associated with ALU.
   - It holds data during an arithmetic/logical operation.
   - It is used by the microprocessor and is not accessible to programmer.

Flags. There are five-flops in Intel 8085 microprocessor to serve as status flags; these are : (*i*) Carry Flag (CS), (*ii*) Parity Flag (P), (*iii*) Auxiliary Carry Flag (AC); (*iv*) Zero Flag (Z), and (*v*) Sign Flag (S).

— The flip-flops are set or reset according to the conditions which arise during an arithmetic or logical operation.

— If a flip-flop for a particular flag is set, it indicates 1. When it is reset, it indicates 0.

Instruction decoder. Data from the instruction register is sent to the instruction decoder, where microprocessor decodes it and then translates into specific actions.

5.2.3.6. Data and address bus

The data bus of Intel 8085 microprocessor is 8-bit wide and hence, 8 bits of data can be transmitted in parallel from or to the microprocessor. This microprocessor requires a 16-bit wide address bus as the memory addresses are of 16-bits. The 8 *most significant* bits of the address are transmitted by the address bus, A-bus (pins A_8 to A_{15}). The 8 *least significant* bits of the address are transmitted by address/data bus, AD-bus (pins AD_0–AD_7).

The address/data bus transmits data and address at different movements. At particular moment it transmits either data or address. Thus *AD-bus operates in time shared mode. This technique is called* ***multiplexing.***

5.2.3.7. Pin configuration

Fig. 5.4. shows the schematic diagram of Intel 8085 microprocessor.

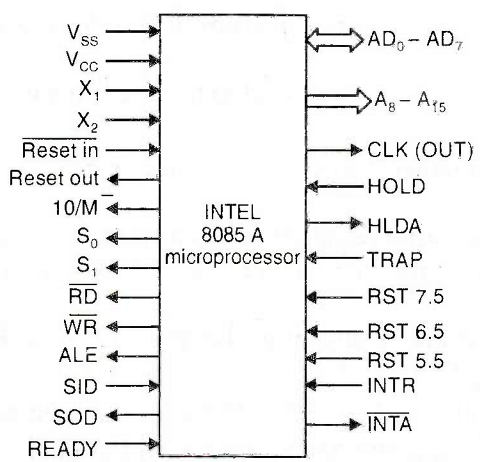


Fig. 5.4. Schematic diagram of Intel 8085 microprocessor.

ADo–AD_7 (Input/Output): They are used for the least significant 8-bits of the memory address of I/O address during the first clock cycle of a machine cycle. Again they are used for data during second and third clock cycles.

A_8–A_{15} (output): These are address bus and are used for the most significant bits of the memory address of 8-bits of I/O address.

$IO/\overline{M}$ (Output): It is a status signal which distinguishes whether the address is for memory or I/O. When it goes high the address on the address bus is for an I/O device, whereas, when it goes low the address on the address bus is for a memory location.

ALE (Output): It is an *address latch enable signal.* It goes high during first clock cycle of a machine cycle and enables the lower bits of the address to be latched either into the memory or external latch.

S_0, S_1 (Output): These are status signals sent by the microprocessor to distinguish the various types of operations given in the table below,

S_1	S_0	*Operations/Machine cycle*
0	0	HALT
0	1	WRITE
1	0	READ
1	1	FETCH

$\overline{RD}$ (Output) :

— It is a signal to control READ operation.

— When it goes low the selected memory or I/O device is read.

$\overline{WR}$ (Output) :

— It is a signal to control WRITE operation.

— When it goes low the data on the data bus is written into the selected memory or I/O location.

READY INPUT :

— It is used by the microprocessor to sense whether a peripheral is ready to transfer data or not.

— If READY is high the peripheral is ready, if it is low the microprocessor waits till it goes high.

— A slow peripheral may be connected to the microprocessor through READY line.

HOLD (Input) :

— It indicates that another device is requesting for the use of the address and data bus.

— The microprocessor after having received a HOLD request relinquishes the use of the buses as soon as the current machine cycle is completed. Internal processing may continue.

— The processor regains the bus after the removal of the HOLD signal.

HLDA (Output) :

— This signal indicates that the HOLD request has been received.

— After the removal of a HOLD request the HLDA goes low. The CPU takes over the buses half clock cycle after the HLDA goes low.

INTR (Input); $\overline{INTA}$:

— INTR signal indicates an interrupt request.

— The INTR line is sampled in the last state of the last machine cycle of an instruction. The microprocessor acknowledges the interrupt signals and issues an $\overline{INTA}$ signal

$\overline{RESET\ IN}$ (Output) :

— When this signal is applied the program counter is set to zero and resets the Interrupt Enable and HLDA flip flops. Except the instruction register no other register or flag is affected.

— The CPU remains in the reset condition as long as reset is applied.

RESET OUT :

— This is an output signal which shows that CPU is being reset. This can be used as a system RESET.

— The signal is synchronised to the microprocessor clock.

X_1, X_2 (Input) :

— X_1 and X_2 are the terminals to be connected to an external crystal oscillator which drives an internal circuitry of the microprocessor to produce a suitable clock for the operation of microprocessor.

CLK (Output) :

— It is a clock output for user, which can be used for other digital ICs.

— Its frequency is same at which processor operates.

SID and SOD line :

— SID is an input line and it is for serial input data. The serial input data at SID can be identified by an instruction called RIM and serial data can be an instruction called SIM.

— In Intel 8085 microprocessor only serial transmission facility is available. The SID and SOD lines pursuit the input and output serial data. The actual transfer of the data is accomplished by software using the RIM and SIM instructions. Both these instructions are single byte and are also used to read or set/reset interrupt masks.

5.2.3.8. Opcode and operands

Each instruction contains the following *two parts* :

(*i*) Operation code (opcode);

(*ii*) Operand.

- ***Opcode*** specifies the task to be performed by the computer.
- ***Operand*** is the data to be operated on.
  - — The operand (or data) given in the instruction may be in various forms such as 8-bit or 16-bit data, 8-bit or 16-bit address, internal registers or a register or memory location.
  - — In some instructions the operand is implicit. When the operand is a register it is understood that data is the content of the register.

5.2.3.9. Instruction cycle

*An **instruction** is a command given to the computer to perform a specified operation on given data.*

In order to perform a particular task a programmer writes a sequence of instructions, called a *program*. Program and data are stored in the memory. The CPU fetches one instruction from the memory at a time and executes it. It executes all instructions of a program one by one to produce the final result.

An ***instruction cycle*** consists of a *fetch cycle* and *execute cycle*. The total time required to execute an instruction is the some of time required to fetch an opcode and the time required to execute it.

— The opcode (the Ist byte of an instruction is its opcode; the instruction may be more than one byte long) fetched from the memory goes to the data register, DR (data/address buffer in Intel 8085 microprocessor) and then to instruction register, IR. From the instruction register it goes to the decode circuitry which *decodes* the instruction. The decoded circuitry is within the microprocessor.

— After the instruction is decoded, *execution begins.* If the operand is in the general purpose registers, execution is immediately performed. The time taken in decoding and execution is one clock cycle. If an instruction contains data or operand address which are still in the memory, the CPU has to perform some read operations to get the desired data. After receiving the data it performs execute operation.

A *read cycle* is similar to a fetch cycle. In case of a read cycle the quantity received from the memory are data or operand address instead of an opcode. In some instructions write operation is performed.

In write cycle data are sent from CPU to the memory or an output device.

In some cases an execute cycle may involve one or more read or write cycles or both.

— The necessary steps carried out to perform a fetch, a read or write operation constitute a *"Machine cycle"*. An instruction cycle consists of several machine cycles.

5.2.3.10. Microprocessor programming

Prior to performing to task, a microprocessor has to be programmed (*Program* is a sequence of instruction that operates the microprocessor on a certain data to achieve desired results).

- The program written for performing a particular task, is stored in the semiconductors memory that is accessible to the microprocessors.
- During the execution of the program, microprocessor fetches one instruction at a time from the memory and executes it.
- Microprocessor understands only instructions written in sequence by using 0s and 1s, and this type of program is known as *machine language program.* These types of programs are very difficult to write. So *first of all programs are written is assembly language using mnemonic operation codes and symbolic addresses.* After that this program is *translated into machine language programme* manually or by using some special translator known as an *assembler.*

5.2.4. Microcontrollers

*The **microcontroller** is the integration of a microprocessor with memory and input/output interfaces, and other peripherals such as timers, on a single chip.* It is basically a *microcomputer on a single IC.*

Fig. 1.13 (Article 1.3) shows the general block diagram of a microcontroller.

Microcontrollers entails the following *"characteristics"*.

(*i*) Low cost;

(*ii*) Versatility;

(*iii*) Ease of programming;

(*iv*) Small size.

- Microcontrollers are attractive in mechatronic system design since their small size and broad functionality allow them to be *physically embedded in a system* to perform all of necessary control functions.

Examples :

(*i*) *8-bit microcontrollers (data path 8-bit wide) :*
- — The Motorola 68 HC11;
- — The Intel 8051;
- — The PIC16 C6 X/7X.

(*ii*) *16-bit microcontroller :*
- — The Motorola 68 HC 16.

(*iii*) *32-bit microcontroller :*
- — The Motorola 68300.

- Microcontrollers have *limited* amounts of ROM and RAM. These are *widely used for embedded control systems.*
- A microprocessor system with separate memory and input/output chips is more suitable for processing information in a computer system.

Applications:

Microcontrollers are used in wide array of applications including :

- Entertainment equipment;
- Home appliances;
- Automobiles;
- Trucks.
- Air planes;
- Toys;
- Office equipment;

All these products involve devices that require some sort of intelligent control based on various inputs; ***Examples*** being :

— In a *microwave oven*, the microcontroller monitors the control panel for user input, updates the graphical displays when necessary, and controls the timing and cooking functions.

— In an *automobile*, there are many microcontrollers to control various subsystems, including cruise control, antilock braking, ignition control, keyless entry, environmental control, and air and fuel flow.

— A *toy* robot dog has various sensors to detect inputs from its environment and an on board microcontroller actuates motors to mimic actual dog behaviour based on their input.

— An *office fax machine* controls actuators to feed papers, use photo sensors to scan a page, sends or receives data on a phone line, and provides a user interface complete with menu-driven controls.

All the above mentioned devices are *controlled by 'microcontrollers' and the 'software' running on them.*

- Typically, *'microcontrollers'* have *less than 1 kilobyte to several tens of kilobytes* of program memory, compared with *'microcomputers'* whose ram memory is measured in *megabytes or gigabytes.* Also, microcontroller clock speeds are *slower than those used for microcomputers.*

— A selected microcontroller, for some applications, may not have enough speed or memory to satisfy the needs of the application. The manufacturers of microcontrollers usually provide a wide range of products to accommodate different applications.

— When more memory or Input/Output capability is required, the functionality of the microcontroller can be expanded with *additional components, e.g.*, RAM or EEPROM chips, external A/D and D/A converters, and other microcontrollers.

Microchip controllers :

— The microchip microcontrollers use a form of architecture termed *Harvard architecture* (Fig. 5.5). With this architecture, instructions are fetched from program memory using accessible variables.

— Harvard architecture enables faster execution speeds to be achieved for a given clock frequency.

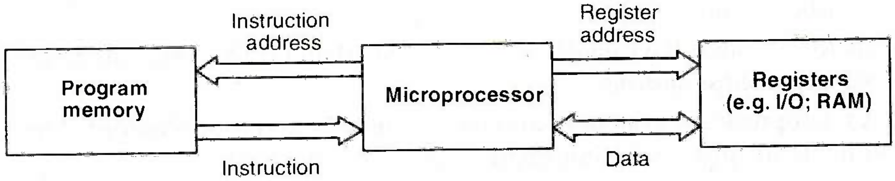


Fig. 5.5. Harvard architecture.

Selection of a microcontroller :

While selecting a microcontroller the following factors should be considered :

1. Number of input/output pins.
2. Interfaces required.
3. Memory requirements.
4. The number of interrupts required.
5. Processing speed required.

HIGHLIGHTS

1. A *computer* is a machine that processes data according to set of instructions stored within the machine.
2. The principle of operation of *analog computers* is to create a physical analog of mathematical problems. The *digital* computers *accept digits and alphabets as inputs.*
3. The complete programming process followed by programmer for writing comprises the following *steps:*

   (*i*) Analysis; (*ii*) Flow charting; (*iii*) Coding; (*iv*) Debugging; (*v*) Documentation; (*vi*) Production.
4. A *"microprocessor"* is a large scale integration (LSI) chip that is capable of performing arithmetic and logic functions as defined by a given programme. A microprocessor consists of: (*i*) ALU; (*ii*) Register unit; (*iii*) Control unit.
5. *Registers* are digital devices used by the microprocessor for temporary storage and manipulation of data and instructions.
6. The *microcontroller* is the integration of a microprocessor with memory and input/output interfaces, and other peripherals such as liners, on single chip.

OBJECTIVE TYPE QUESTIONS

Fill in the blanks or Say 'Yes' or 'No'

1. Charle's Babbage is called "Father of computers".
2. A .... is a machine that processes data according to set of instructions stored within the machine.
3. Speedometer is an example of analog computer.
4. .... are popularly known as personal computer (PC).
5. A hybrid computer is a combination of both analog and digital computers.
6. Chips are rated in terms of their .... and ....
7. RAM chip is made with Metal Oxide Semiconductor (MOS).
8. .... are used to multiply a variable quantity by a constant.
9. Servomultipliers are used to multiply two variables.
10. A .... is a large scale integration (LSI) chip that is capable of performing arithmetic and logic functions as defined by a given programme.
11. Microprocessors which have memory and various input/output arrangements on the same chip are called ....
12. .... are memory locations within the microprocessor.
13. MAR (Memory address register) contains the address of data.
14. General purpose registers serve as .... storage for data or addresses and used in operations involving transfers between other registers.
15. .... are the paths along which digital signals move from one section to another.
16. .... bus carries the signals relating to control actions.
17. The size of the memory is determined by the number of wires in the address bus.
18. .... is a device in which data is stored permanently.
19. RAM is a memory that can be read only.
20. Intel 8085 is an 8-bit, NMOS microprocessor.

ANSWERS

1. Yes	2. Computer	3. No	4. Microcomputers
5. Yes	6. Capacity, speed	7. Yes	8. Attenuators
9. Yes	10. microprocessor	11. microcontroller	12. Registers
13. Yes	14. temporary	15. Buses	16. Control
17. Yes	18. ROM	19. No	20. Yes.

THEORETICAL QUESTIONS

1. What is a 'Computer'? Explain.
2. List the characteristics of a computer.
3. What are the limitations of a computer?
4. How are computers classified?
5. How are digital computers classified on the basis of size and capabilities?
6. Explain briefly the following :
   Super computer; Main frame computers; Minicomputer; Microcomputers.
7. What are the differences between analog and digital computers?
8. Draw the block diagram of a computer and explain briefly its various parts.
9. How are chips rated?

10. What do you mean by the term 'Peripheral'? Explain briefly the following devices:
    (i) Input devices; (ii) Output devices.
11. Explain briefly 'storage devices'.
12. List the steps which are required for computer programming process for writing programs.
13. What is a "Microprocessor"?
14. Draw the block diagram of a microcomputer and explain briefly the three segments (ALU, Register and Control unit) of a microprocessor.
15. What are the characteristics of microprocessor?
16. Mention the important features of the microprocessors.
17. What are the uses of microprocessors?
18. Explain briefly the various parts of a microprocessor system.
19. Explain briefly the following registers:
    Accumulator register; Status register; Program counter register (PC); memory address register; Instruction register; General purpose registers; Stack pointer register.
20. What are 'buses'? Explain briefly the following buses:
    Data bus, Address bus, Control bus.
21. Explain briefly the following forms of memory unit:
    ROM; PROM; EPROM; EEPROM; RAM.
22. Explain briefly Intel 8085 microprocessor with the help of a block diagram.
23. Write a short note on 'Microprocessor programming'.
24. What are 'Microcontroller'? Explain briefly.
25. What are 'Microchip controller'?

CHAPTER

6 System Models and Controllers

6.1 **Basic system models** – Introduction – Mechanical system building blocks – Rotational systems – Building up a mechanical system – Electrical system building blocks – Building up a model for an electrical system – Fluid system building blocks – Building up a model for a fluid system – Thermal system building blocks – Building up a model for a fluid system – Thermal system building blocks; **6.2 System models** – Introduction – Rotational – Translation systems – Electromechanical systems – Hydro-mechanical systems; **6.3 Controllers** – Introduction – Control modes – Two-steps mode – Proportional mode (P) – Derivative mode (D) – PD controllers – Integral mode (I) – PI controllers – PID controllers – Digital controllers – Adaptive control system – Programmable logic controllers – Introduction – Special features – Basic structure – Selection of a PLC – Highlights – Objective Type Questions – Theoretical Questions.

6.1 BASIC SYSTEM MODELS

6.1.1. Introduction

This chapter relating to system models in mainly concerned to determine how systems behave with time when subject to some disturbance. For understanding the behaviour of the systems, ***mathematical models*** are needed :

— The mathematical models are *equations which describe the relation between the input and output of a system.*

— The basis for any mathematical model is provided by the *fundamental physical laws* that govern the system's behaviour.

— These models can be used to enable forecasts to be made of the system's behaviour under specific conditions.

Systems can be made up from a range of *building blocks* (as a child builds houses, cars etc.) from a number of basic building blocks.

Here follows the description of building blocks for mechanical, electrical, fluid and thermal systems.

6.1.2. Mechanical System Building Blocks

The basis building blocks of the models used to represent mechanical systems are :

1. Springs;
2. Dashpots;
3. Masses.

1. Springs. The springs represent the *stiffness* of a system. Figure 6.1 shows a spring subjected to force F. In the case of a linear spring (*i.e.,* where the extension/elongation or compression is proportional to the applied force),

$$F = kx \quad ...(6.1)$$

where, F = Applied force,

k = A constant, and

x = Extension (or compression).

Eqn. (6.1), indicates, that as per Newton's third law, the force F is equal in size and in the opposite direction to the force exerted by the stretched spring (*i.e.*, kx).

The spring when stretched stores energy, the energy being released when the spring springs back to its original length.

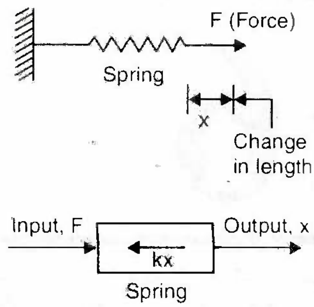


Fig. 6.1. Spring.

- *"Energy stored"* when there is an extension x,

$$E = \frac{1}{2}kx^2 = \frac{1}{2}\frac{F^2}{k} \quad (\because F = kx) \quad ...(6.2)$$

2. Dashpots. The dashpots represent the forces opposing motion, *i.e., frictional or damping effects.* Fig. 6.2 shows a dashpot. Here, the faster the object is pushed greater becomes the opposing forces.

In an ideal case, the damping or resisting force F is proportional to the velocity v of the piston. Thus,

$$F = c.v \quad ...(6.3)$$

where c is a constant.

Further, since velocity is the rate of change of displacement x, therefore,

$$F = c \cdot \frac{dx}{dt} \quad ...(6.4)$$

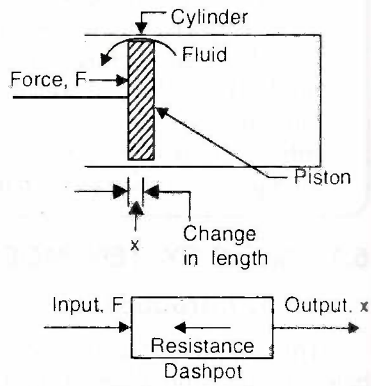


Fig. 6.2. Dashpot.

In a dashpot *no energy is stored.* It does not return to its original position when there is no force input. The dashpot *dissipates* energy rather storing it.

- *"Power dissipated"*, $P = cv^2$ ...(6.5)

3. Masses. The masses represent the *inertia or resistance to acceleration.* Fig. 6.3 shows a mass; the mass building block exhibits the property that the bigger the mass the greater the force required to a specific acceleration. As per Newton's, law:

$$F = ma \quad ...(6.6)$$

or, $$F = m \times \frac{dv}{dt} = m \times \frac{d}{dt}\left(\frac{dx}{dt}\right)$$

or, $$F = m \times \frac{d^2x}{dt^2} \quad ...(6.7)$$

- *"Energy (kinetic energy) stored"* in the mass when it is moving with a velocity v, and released when it stops moving,

$$E = \frac{1}{2}mv^2 \quad ...(6.8)$$

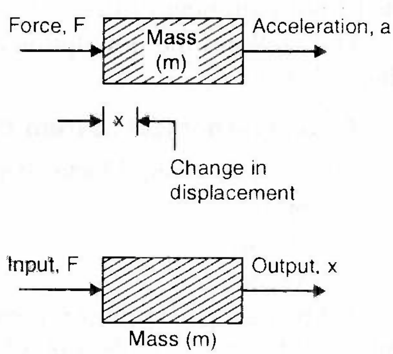


Fig. 6.3. Mass.

6.1.2.1. Rotational systems

In mechanical systems, when *rotation* is involved, the three building blocks are :

1. Torsional spring; 2. Rotary damper; 3. Moment of inertia.

With such building blocks, the inputs are *torque* and the output *angle rotated.*

1. A torsional spring. In a torsional spring the angle rotated (θ) is proportional to the torque (T),

i.e., $$T = k\theta \quad \text{...(6.9)}$$

- *"Energy stored"* by the torsional spring when twisted through an angle θ,

$$E = \frac{1}{2}k\theta^2 = \frac{1}{2}\frac{T^2}{k} \qquad (\because T = k\theta) \quad \text{...(6.10)}$$

2. Rotary damper. In the rotary damper, a disc is rotated in a fluid and the resistive torque (T) is proportional to the angular velocity (ω),

i.e., $$T = c\omega = c \cdot \frac{d\theta}{dt} \quad \text{...(6.11)}$$

(since ω is the rate of change of angular displacement)

- *"Power dissipated"* by the rotary damper when rotating with an angular velocity ω,

$$P = c\omega^2 \quad \text{...(6.12)}$$

3. Moment of inertia (I). The moment of inertia of building block exhibits the property that the greater the moment of inertia I the greater the torque needed to produce an angular acceleration α.

$$T = 1 \cdot \alpha \quad \text{...(6.13)}$$

or, $$T = I \cdot \frac{d\omega}{dt} = I\frac{d(d\theta/dt)}{dt} = I\frac{d^2\theta}{dt^2} \quad \text{...[6.13(a)]}$$

($\because$ angular acceleration is rate of change of angular velocity and angular velocity is the rate of change of angular displacement)

- *"Energy stored"* by the mass rotating with an angular velocity, ω

$$= \frac{1}{2}I\omega^2 \quad \text{...(6.14)}$$

6.1.2.2. Building up a mechanical system

Several systems can be considered to consist of a mass, spring and a dashpot as shown in Fig. 6.4.

Net force applied to the mass (m) to cause the mass to accelerate = $F - kx - cv$.

where, v = The velocity with which the piston in the dashpot, and hence m is moving,

x = The change in length of the spring, and

k = Stiffness of the spring.

Hence, $$F - kx - cv = ma$$

or, $$F - kx - c \cdot \frac{dx}{dt} = m \times \frac{d^2x}{dt} \qquad \left(\because a = \frac{d^2x}{dt^2}\right)$$

On rearranging, we get

$$m \cdot \frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = F \quad ...(6.15)$$

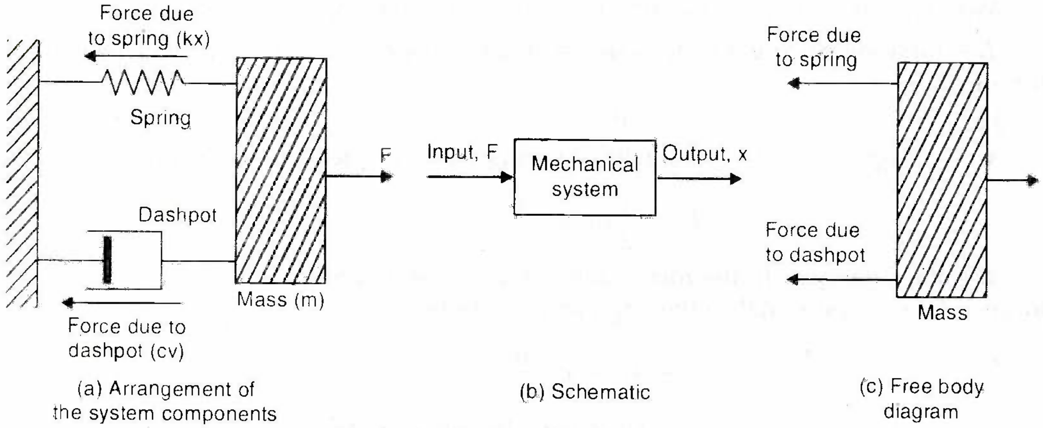
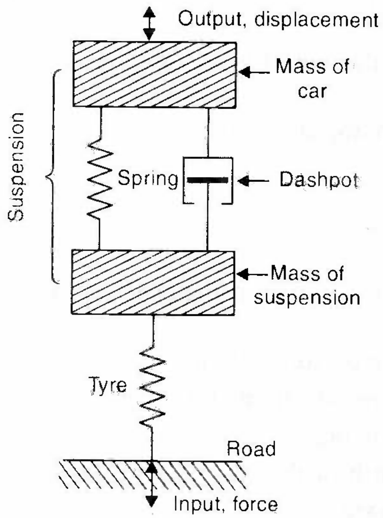


Fig. 6.4. Mechanical system.

It is *second-order differential equation* (because of the term $\frac{d^2x}{dt^2}$)

— Many systems can be built up from suitable combinations of the mass, spring and dashpot building blocks. As an example, Fig. 6.5 shows a mathematical model of a wheel of a car moving along a road. The procedure of analysing such a model is same as discussed above.



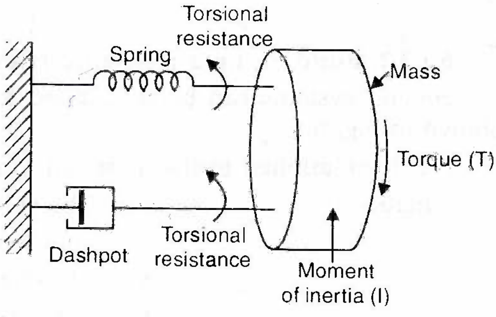


Fig. 6.5. Mathematical model of a car moving on a road.

Fi.g. 6.6. Building block model (rotational).

— Similar models can be constructed for rotating systems; such a model is shown in Fig. 6.6. This is a comparable situation to that analysed above for linear displacements and yields a similar equation given as follows :

$$I\frac{d^2\theta}{dt^2}+c\cdot\frac{d\theta}{dt}+k\theta = T \quad ...(6.16)$$

where, θ = Angular displacement.

6.1.3. Electrical System Building Blocks

For electrical systems, the building blocks are :

1. Resistors; 2. Inductors; 3. Capacitors.

1. Resistors. The potential difference across a resistor at any instant depends on the current I through it.

$$V = I \times R \quad ...(6.17)$$

where R is the resistance.

- *"Power dissipated"* by a resistor,

$$P = I \times V = \frac{V^2}{R} \quad ...(6.18)$$

2. Inductors. The potential difference V across an inductor at any instant depends on the rate of change of current $\left(\frac{dI}{dt}\right)$ through it;

i.e.,
$$V = L\cdot\frac{dI}{dt} \quad ...(6.19)$$

where L is the inductance.

The direction of the potential difference is in the *opposite* direction to the potential difference used to drive the current through the inductor, hence the term back e.m.f.

By rearranging the eqn. (6.19), we have

$$I = \frac{1}{L}\int V\,dt \quad ...(6.20)$$

- *"Energy stored"* by an inductor $= \frac{1}{2}LI^2$ ...(6.21)

3. Capacitors. The potential difference across a capacitor depends on the charge Q on the capacitor plates at the instant concerned.

$$V = \frac{Q}{C} \quad ...(6.22)$$

where C is the capacitance.

Since current i to or from the capacitor is the rate at which charge moves to or from the capacitor plates, *i.e.*,

$$i = \frac{dq}{dt},$$

therefore, total charge Q on the plates is given by

$$Q = \int i\,dt \quad ...(6.23)$$

and
$$V = \frac{1}{C}\int i\,dt \quad \text{... from eqn. (6.19)}$$

- *"Energy stored"* by a capacitor $= \frac{1}{2}CV^2$ ...(6.24)

6.1.3.1. Building up a model for an electrical system

The various electrical building blocks can be combined by using *Kirchhoff's laws;* these are as follows:

1. *Kirchhoff's current law (KCL):* It states as follows:

   "The sum of currents entering a junction is equal to the sum of the currents leaving the junction".

2. *Kirchhoff's voltage law (KVL):* It states as follows:

   "The sum of the e.m.fs. (rises of potential) around any closed loop of a circuit equals the sum of the potential drops in that loop".

The convenient method of using KCL is *"node analysis"* and that of using KVL is *"mesh analysis"*. To illustrate these two methods of analysis, let us consider the circuit shown in Fig. 6.7.

- To illustrate the use of ***"node analysis"*** (all components being resistors) let us pick up a principal node point A on the figure and let the value at this node point be V_A with reference to some other principal node that has been picked up as the reference.

According to Kirchhoff's KCL, we have:

$$I_1 = I_2 + I_3 \qquad ...(i)$$

Now, $$I_1R_1 = V - V_A$$

or, $$I_1 = \frac{V - V_A}{R_1}$$

and, $$I_2R_2 = V_A$$

or, $$I_2 = \frac{V_A}{R_2} \qquad ...(ii)$$

and, $$I_3(R_3 + R_4) = V_A$$

or, $$I_3 = \frac{V_A}{R_3 + R_4}$$

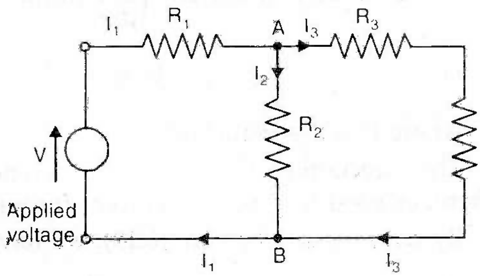


Fig. 6.7. Node analysis.

Now substituting for the currents in eqn. (*i*), we get

$$\frac{V - V_A}{R_1} = \frac{V_A}{R_2} + \frac{V_A}{R_3 + R_4} \qquad ...(6.25)$$

- To illustrate the use of ***"mesh analysis"*** for the circuit in Fig. 6.7 we assume there are currents circulating in each mesh in the way shown in Fig. 6.8. Then by applying KVL to each mesh, we have:

— For the mesh with current I_1 circulating and source of e.m.f. V :

$$V = I_1R_1 + (I_1 - I_2)R_2 \qquad ...(i)$$

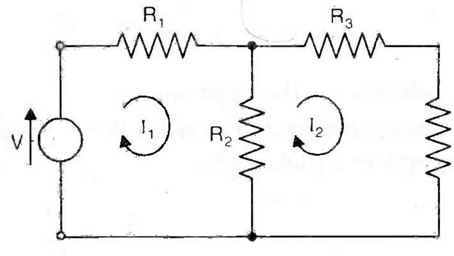


Fig. 6.8. Mesh analysis.

— For the mesh with current I_2 circulating, there being no source of e.m.f.

$$0 = I_2R_3 + I_2R_4 + (I_2 - I_1)R_2 \qquad ...(ii)$$

Now the two mesh currents I_1 and I_2 can be found out from the above two equations.

— In general, it is easier to employ mesh analysis when the number of nodes in a circuit is less than the number of meshes.

1. *Resistor-inductor (R-L) system :*

Fig. 6.9 shows a simple electrical system consisting of a resistor and an inductor in series.

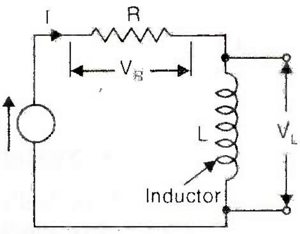


Fig. 6.9. Resistor-inductor system.

Applying KVL to circuit loop gives :

$$V = V_R + V_L$$

where, V_R is potential difference across the resistor R and V_L that across the inductor.

Since, $\qquad V_R = IR$

therefore, $\qquad V = IR + V_L$

Since, $\qquad I = \frac{1}{L}\int V_L dt \qquad$ ...[Eqn. 6.20)]

Then, the relationship between the input and output is

$$V = \frac{R}{L}\int V_L dt + V_L \qquad ...(6.26)$$

2. *Resistor-capacitor system :*

Consider a simple electrical system consisting of a resistor and capacitor in series as shown in Fig. 6.10.

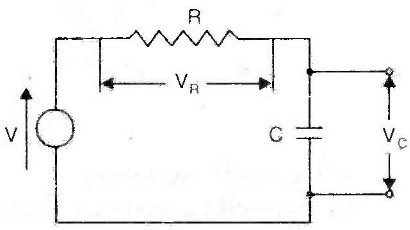


Fig. 6.10. Resistor-capacitor system.

Applying KVL to the circuit loop gives:

$$V = V_R + V_C$$

where V_R is the potential difference across the resistor and V_C that across the capacitor.

Now, $\qquad V_R = IR \text{ and } I = C \cdot \frac{dV_C}{dt}$

$$\therefore \qquad V = RC\frac{dV_C}{dt} + V_C \qquad ...(6.27)$$

This gives the relationship between output V_C and the input and is *first-order* differential equation.

3. *Resistor-inductor-capacitor system :*

Figure 6.11 shows a resistor-inductor-capacitor system.

Applying KVL to the circuit loop, we get :

$$V = V_R + V_L + V_C$$

or, $\qquad V = IR + L.\frac{dI}{dt} + V_C$

$$\left(\because V_L = L\frac{dI}{dt}\right)$$

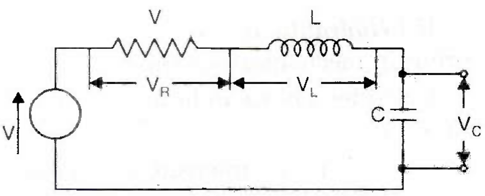


Fig. 6.11. Resistor-inductor-capacitor system.

But, $\qquad I = C \cdot \frac{dV_C}{dt}$

$$\therefore \qquad \frac{dI}{dt} = C\frac{d(dV_C/dt)}{dt} = C\frac{d^2V_C}{dt^2}$$

Hence, $$V = RC\frac{dV_C}{dt} + LC\frac{d^2V_C}{dt^2} + V_C \quad ...(6.28)$$

This is a *second-order* differential equation.

6.1.4. Fluid System Building Blocks

The three basic building blocks of a fluid flow systems (Fig. 6.12), can be considered to be equivalent of electrical resistance, inductance and capacitance.

Fluid system can be considered to fall into *two* categories :

(*i*) *Hydraulic:* Here the fluid is a liquid and is considered to be incompressible.

(*ii*) *Pneumatic:* Here it is a gas which can be compressed and consequently shows a change of density.

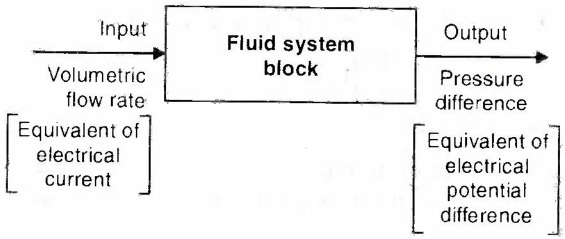


Fig. 6.12. Fluid system

1. Hydraulic systems :

(i) Hydraulic resistance (R_h). It is the resistance to flow which occurs as a result of a liquid flowing through valves or changes in a pipe diameter. The following relation holds good:

$$p_1 - p_2 = R_h \times Q_l \quad ...(6.29)$$

where, $p_1 - p_2$ = Difference of pressure,

R_h = A constant, called *hydraulic resistance*, and

Q_l = Volume rate of flow of liquid.

— Hydraulic linear resistances occur with orderly flow through capillary tubes and porous plugs but non-linear resistances occur with flow through sharp-edged orifices or when the flow is turbulent.

- The *"energy dissipated"*, $P = \frac{1}{R_h}(p_1 - p_2)^2$ ...(6.30)

(ii) Hydraulic intertance (I_h). It is equivalent of inductance in electrical systems or a spring in mechanical systems.

Consider a block of liquid of mass, m, as shown in Fig. 6.13.

Let, p_1 = Intensity of pressure at section-1,

F_1 = Force acting at section-1,

p_2 = Intensity of pressure at section-2,

F_2 = Force acting at section-2,

A = Cross-sectional area, and

L = Length of the block of liquid.

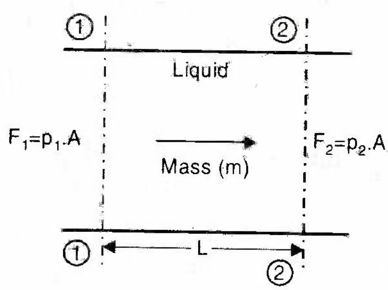


Fig. 6.13. Hydraulic inertance.

Then, net force acting on the liquid is :

$$F_1 - F_2 = p_1 \cdot A - p_2 \cdot A = (p_1 - p_2)A$$

This net force causes the mass to accelerate with an acceleration a, and therefore,

$$(p_1 - p_2)A = ma$$

or, $$(p_1 - p_2)A = m \cdot \frac{dv}{dt} \qquad (\because \text{a is the rate of change of velocity } \frac{dv}{dt})$$

Also, mass of the liquid, $m = AL\rho$

$\therefore$ $$(p_1 - p_2)A = A.L.\rho.\frac{dv}{dt}$$

But, the volume rate of flow, $q = Av$

$\therefore$ $$(p_1 - p_2)A = L\rho\frac{dQ_l}{dt}$$

or, $$p_1 - p_2 = I_h\frac{dQ_l}{dt} \qquad ...(6.31)$$

where the hydraulic intertance I_h is defined as: $I_h = \frac{L\rho}{A}$.

- *"Energy stored"* by intertance, $E = \frac{1}{2} I_h Q_l^2$ ...(6.32)

(iii) Hydraulic capacitance (C_h). This term is used to describe energy storage with a liquid when it is stored in the form of *potential energy (P.E.).*

Consider a container filled with a liquid as shown in Fig. 6.14.

Let, A = Cross-sectional area of the container,

H = Height of liquid in the container,

Q_{l1}, Q_{l2} = The rates of liquid flow at the entrance and exit of the container

V = Volume of liquid in the container,

p = Pressure difference between the input and output.

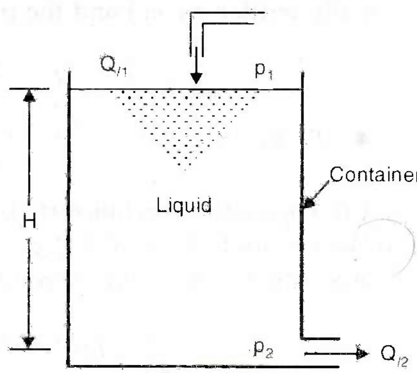


Fig. 6.14. Hydraulic capacitance.

Then, $$Q_{l1} - Q_{l2} = \frac{dV}{dt} \qquad ...(6.33)$$

(where $\frac{dV}{dt}$ = rate of change of volume V in the container)

or, $$Q_{l1} - Q_{l2} = \frac{d(AH)}{dt} = A \cdot \frac{dH}{dt} \qquad (\because V = A \cdot H)$$

Also, $$p = \rho g H,$$

or, $$H = \frac{p}{\rho g}$$

where ρ is the liquid density and g is the acceleration due to gravity.

If the liquid is assumed incompressible then ρ does not change with pressure.

Then, $$Q_{l1} - Q_{l2} = A\frac{d(p/\rho g)}{dt} = \frac{A}{\rho g}\cdot\frac{dp}{dt}$$

The hydraulic capacitance C_h is defined as being :

$$C_h = \frac{A}{\rho g}.$$

Thus, $$Q_{l1} - Q_{l2} = C_h \cdot \frac{dp}{dt} \quad ...(6.34)$$

By integrating this equation we get,

$$p = \frac{1}{C_h}\int(Q_{l1} - Q_{l2})dt \quad ...(6.35)$$

- *"Energy stored"* by the capacitance, $E = \frac{1}{2}C(p_1 - p_2)^2$. ...(6.36)

2. Pneumatic systems :

Like hydraulic systems, pneumatic systems also have three base building blocks: *Resistance, inertance* and *capacitance.*

(i) Pneumatic resistance (R_{pn}). It is defined in terms of the mass rate of flow $\frac{dm}{dt}$ (this normally written as $\dot{m}$) and the pressure difference $(p_1 - p_2)$ as:

$$p_1 - p_2 = R_{pn}\cdot\frac{dm}{dt} = R_{pn}\,\dot{m} \quad ...(6.37)$$

- *"Power dissipated"*, $P = \frac{1}{R_{pn}}(p_1 - p_2)^2$. ...(6.38)

(ii) Pneumatic inertance (I_{pn}). The pneumatic inertance is *due to the pressure drop necessary to accelerate a block of gas.*

According to Newton's second law,

$$(p_1 - p_2)A = ma = m\cdot\frac{dv}{dt} = \frac{d(mv)}{dt} \quad ...(6.39)$$

where, $(p_1 - p_2)$ = Pressure difference,

A = Area of cross-section, and

a = Acceleration of the gas.

Also, $$mv = (\rho LA) \times v = \rho LA \times \frac{Q_{pn}}{A} = \rho L Q_{pn}$$

where, L = Length of the block of gas being accelerated,

v = Velocity of gas, and

Q_{pn} = Volume rate of gas flow.

Thus, $$(p_1 - p_2)A = L\frac{d(\rho Q_{pn})}{dt} \quad \text{...from eqn. (6.39)}$$

But $\dot{m} = \rho Q_{pn}$, therefore

$$(p_1 - p_2) = \frac{L}{A}\cdot\frac{d\dot{m}}{dt}$$

or, $$(p_1 - p_2) = I_{pn} \cdot \frac{d\dot{m}}{dt} \quad \text{...(6.40)}$$

where, I_{pn} (pneumatic inertance) $= \dfrac{L}{A}$

(iii) Pneumatic capacitance (C_{pn}). The pneumatic capacitance is *due to the compressibility of the gas and is comparable to the way in which the compression of the spring stores energy.*

Let us consider a container containing gas.

Let, V = Volume of gas entering the container,

$\dfrac{dm_1}{dt}$ = Mass rate of flow entering the container,

$\dfrac{dm_2}{dt}$ = Mass rate of flow leaving the container, and

ρ = Density of the gas in the container.

Since the gas can be compressed, both p and V can vary with time. Hence,

Rate of change of mass in container

$$= \rho\frac{dV}{dt} + V\frac{d\rho}{dt}.$$

Since, $$\frac{dV}{dt} = \frac{dV}{dp} \times \frac{dp}{dt}$$ and, for an ideal gas,

$$pV = mRT$$

$$\therefore \quad p = \left(\frac{m}{V}\right)RT \text{ or } \rho RT \text{ or } \rho = \frac{p}{RT}$$

and, $$\frac{d\rho}{dt} = \frac{1}{RT}\left(\frac{dp}{dt}\right)$$

Then, rate of change of mass in container

$$= \rho \cdot \frac{dV}{dp} \cdot \frac{dp}{dt} + \frac{V}{RT}\frac{dp}{dT}$$

where, R = The gas constant, and

T = Absolute temperature, (K).

Now, the rate at which the mass in the container is changing is given as :

$$\frac{dm_1}{dt} - \frac{dm_2}{dt} = \left(\underbrace{\rho\frac{dV}{dp}}_{C_{pn1}} + \underbrace{\frac{V}{RT}}_{C_{pn2}}\right)\frac{dp}{dt} \quad \text{...(6.41)}$$

where, $\rho \cdot \dfrac{dV}{dp} = C_{pn1}$ = The pneumatic capacitance due to change of volume of the container, and

$\dfrac{V}{RT} = C_{pn2}$ = The pneumatic capacitance due to the compressibility of the gas.

Hence, $$\frac{dm_1}{dt} - \frac{dm_2}{dt} = (C_{pn1} + C_{pn2})\frac{dp}{dt} \quad ...(6.42)$$

or, $$\dot{m}_1 - \dot{m}_2 = (C_{pn1} + C_{pn2})\frac{dp}{dt}$$

or $$dp = \frac{1}{C_{pn1} + C_{pn2}}(\dot{m}_1 - \dot{m}_2) \cdot dt$$

On integration, we get

$$p_1 - p_2 = \frac{1}{C_{pn1} + C_{pn2}} \int (\dot{m}_1 - \dot{m}_2)\, dt \quad ...(6.43)$$

- *"Energy stored"* by capacitance, $E = \frac{1}{2} C_{pn}(p_1 - p_2)^2$ ...(6.44)

6.1.4.1. Building up a model for a fluid system

Hydraulic system :

Fig. 6.15, shows a simple hydraulic system in which a liquid is entering and leaving a container, such a system can be considered to consist of :

- A *capacitor*—the liquid in the container,
- A *resistor*—the valve;
- *Inertance neglected*—since flow rates change only very slowly.

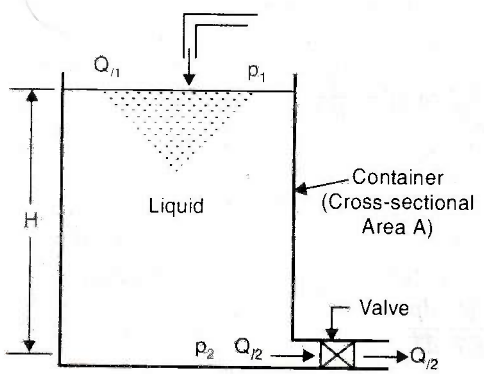


Fig. 6.15. A hydraulic system.

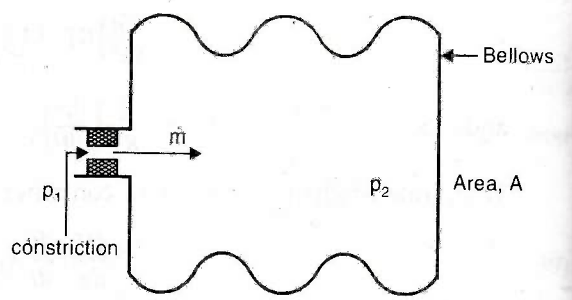


Fig. 6.16. A pneumatic system.

For the *capacitor*, we can write the following equation :

$$Q_{l1} - Q_{l2} = C_h \cdot \frac{dp}{dt} \quad ...(i)$$

For the *resistor*, we have :

$$p_1 - p_2 = R_h Q_{l2}$$

...since the rate at which liquid leaves the container Q_{l2} equals the rate at which it leaves the valve.

Since, $$p_1 - p_2 = p = \rho g H,$$

$$Q_{l2} = \frac{\rho g H}{R_h}$$

Substituting for Q_{l2} in eqn. (i), we get

$$Q_{l1} - \frac{\rho g H}{R_h} = C_h \frac{d(\rho g H)}{dt}$$

But,
$$C_h = \frac{A}{\rho g}$$

$$\therefore \quad Q_{l1} = A \cdot \frac{dH}{dt} + \frac{\rho g H}{R_h} \quad \text{...(6.45)}$$

Eqn. (6.45) conveys that how the liquid height in the container depends on the rate of input of liquid into the container.

Pneumatic system :

The example of a simple pneumatic system is the *bellows* as shown in Fig. 6.16. Such a system can be considered to consist of :

- A *capacitor*—the bellows itself;
- A *resistor*—a constriction which restricts the rate of flow of gas into the bellows;
- *Inertance neglected*—since the flow rate changes only slowly.

The rate of mass flow ($\dot{m}$) into the bellows is given by:

$$p_1 - p_2 = R_{pn}\dot{m} \quad \text{...(6.46)}$$

where,
p_1 = Pressure prior to the constriction,
p_2 = Pressure after constriction, *i.e.*, the pressure in the bellow, and
R_{pn} = Resistance provided by the constriction.

All the gas that flows into the bellows remains in the bellows, there being no exit from the bellows.

The capacitance of the bellows is given by:

$$\dot{m}_1 - \dot{m}_2 = (C_{pn1} + C_{pn2})\frac{dp_2}{dt} \quad \text{...(6.47)}$$

Since the mass flow rate entering the bellows is given by the equation for the resistance and the mass leaving the bellows is zero, therefore,

$$\frac{p_1 - p_2}{R_h} = (C_{pm1} + C_{pn2})\frac{dp_2}{dt}$$

or,
$$p_1 = R_h(C_{pnl} + C_{pm2})\frac{dp_2}{dt} + p_2 \quad \text{...(6.48)}$$

This eqn. conveys that how the pressure in the bellows p_2 varies with time when there is an input of a pressure p_1.

Since bellows are just a form of spring (the bellows expands or contracts due to pressure changes inside it), we can write:

$$F = kx$$

where,
F = The force causing expansion or contraction of the bellows,
x = The resulting displacement (due to force F), and
k = The spring constant for the bellows.

Also, $p_2 = \dfrac{F}{A}$

where, A = Cross-sectional area of the bellows.

Thus, $p_2 A = F = kx$

Hence substituting for p_2 in eqn. (6.48), we get

$$p_1 = R_h(C_{pn1} + C_{pn2})\frac{k}{A}\cdot\frac{dx}{dt} + \frac{k}{A}\cdot x \qquad ...(6.49)$$

Eqn. (6.49) is a *first-order differential equation*, and describes how the value of x (extension or contraction of bellows) changes with time when there is an input of a pressure p_1.

The pneumatic capacitance due to change in volume of the container C_{pn1} is $\rho\dfrac{dV}{dp_2}$ and since $V = Ax$,

$$C_{pn1} = \rho A\frac{dx}{dp_2}.$$

But for the bellows $p_2 A = kx$,

thus
$$C_{pn1} = \rho A\frac{dx}{d(kx/A)} = \frac{\rho A^2}{k} \qquad ...(6.50)$$

C_{pn2}, the pneumatic capacitance due to the compressibility of the air, is given by :

$$C_{pn2} = \frac{V}{RT} = \frac{Ax}{RT} \qquad ...(6.51)$$

6.1.5. Thermal System Building Blocks

For thermal systems, there are only two building blocks:

1. Resistance (R_{th}).
2. Capacitance (C_{th}).

Thermal resistance (R_{th}). If Q_{th} is the rate of flow of heat and $(T_2 - T_1)$ the temperature difference, then

$$Q_{th} = \frac{T_2 - T_1}{R_{th}} \qquad ...(6.52)$$

The value of R_{th} depends on the *mode of heat transfer.*

- In the case of ***conduction*** through a solid, for unidirectional conduction,

$$Q_{th} = kA\frac{(T_1 - T_2)}{L} \qquad ...(6.53)$$

where, k = Thermal conductivity,
A = Cross-sectional area of the material through which the heat is being conducted,
L = The length of the material between the points at which temperatures are T_1 and T_2.

Hence, with this mode of heat transfer,

$$R_{th} = \frac{L}{Ak}$$

- When the mode of heat transfer is ***convection***, as with liquids and gases, then

$$Q_{th} = Ah(T_2 - T_1) \quad ...(6.54)$$

where, A = The surface area across which there is the temperature difference, $(T_2 - T_1)$, and

h = The heat transfer coefficient.

Thus, with this mode of heat transfer

$$R_{th} = \frac{1}{Ah}$$

Thermal capacitance (C_{th}). *"Thermal capacitance" is a measure of the store of internal energy in a system.*

Thus, if the rate of flow of heat into a system is Q_{th1} and the rate of flow out is Q_{th2}, then

$$Q_{th1} - Q_{th2} = mc\frac{dT}{dt} \quad ...(6.55)$$

where, m = Mass,

c = Specific heat capacity, and

$\frac{dT}{dt}$ = Rate of change of temperature.

Eqn. (6.55) can be written as:

$$Q_{th1} - Q_{th2} = C_{th}\frac{dT}{dt} \quad ...(6.56)$$

where C_{th} (thermal capacitance) = mc.

6.1.5.1. Building up a model for a thermal system

Let us consider a thermometer at temperature T which has just been inserted into a liquid at temperature T_l,

Then, $$Q_{th} = \frac{T_l - T}{R_{th}} \quad ...(6.57)$$

where, Q_{th} = Net rate of heat flow from liquid to thermometer, and

R_{th} = Thermal resistance to heat flow from the liquid to the thermometer.

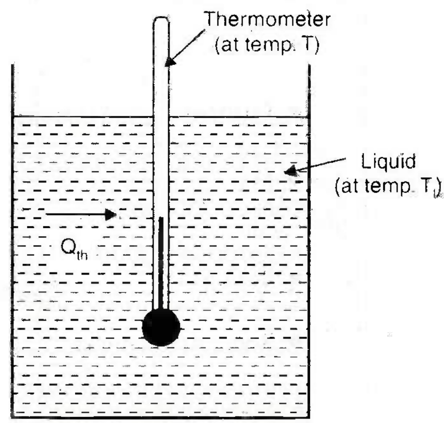


Fig. 6.17. A thermal system.

The thermal capacitance (C_{th}) of the thermometer is given by:

$$Q_{th1} - Q_{th2} = C_{th} \cdot \frac{dT}{dt} \quad ...(6.58)$$

Here, $Q_{th1} = Q_{th}$ (since there is only a net heat flow from the liquid to the thermometer), and

$Q_{th2} = 0$

Thus, $$Q_{th} = C_{th} \cdot \frac{dT}{dt} \qquad ...(6.59)$$

Substituting the value of Q_{th} in eqn. (6.55), we have

$$C_{th} \cdot \frac{dT}{dt} = \frac{T_l - T}{R_{th}} \qquad ...(6.60)$$

By rearranging this equation, we get

$$R_{th}\, C_{th} \frac{dT}{dt} + T = T_l \qquad ...[6.60(a)]$$

This is a *first-order* differential equation and describes how the temperature indicated by the thermometer T will vary with time when thermometer is inserted into a hot liquid.

Note: In the above thermal system, the parameters have been considered to be *lumped* (*i.e.*, the temperatures are only functions of time and not position within the body).

The summary of mechanical, electrical, fluid and thermal systems building blocks is given in Table 6.1.

Table 6.1. Summary of Mechanical, Electrical, Fluid and Thermal Systems Building Blocks

S. No.	Building blocks	Working equation	Energy stored or power dissipated
1.	*Mechanical systems :*		
	(*i*) *Translational :*		
	• Spring	$F = kx$	$E = \frac{1}{2}\frac{F^2}{k}$
	• Dashpot	$F = c\frac{dx}{dt}$	$P = cv^2$
	• Mass	$F = m\frac{d^2x}{dt^2}$	$E = \frac{1}{2}mv^2$
	(*ii*) *Rotational :*		
	• Spring	$T = k\theta$	$E = \frac{1}{2}\frac{T^2}{k}$
	• Rotational damper	$T = c\frac{d\theta}{dt}$	$P = c\omega^2$
	• Moment of inertia	$T = I\frac{d^2\theta}{dt^2}$	$E = \frac{1}{2}\omega^2$
2.	*Electrical systems :*		
	• Resistor	$I = \frac{V}{R}$	$P = \frac{V^2}{R}$
	• Inductor	$I = \frac{1}{L}\int V dt$	$E = \frac{1}{2}LI^2$
	• Capacitor	$I = C\frac{dV}{dt}$	$E = \frac{1}{2}CV^2$

3.	***Fluid systems :***		
	(i) Hydraulic systems:		
	• Resistance	$Q_l = \dfrac{p_1 - p_2}{R_h}$	$P = \dfrac{1}{R_h}(p_1 - p_2)^2$
	• Inertance	$Q_l = \dfrac{1}{I_h}\int(p_1 - p_2)dt$	$E = \dfrac{1}{2}I_h Q_l^2$
	• Capacitance	$Q_l = C_h \dfrac{d(p_1 - p_2)}{dt}$	$E = \dfrac{1}{2}C_h(p_1 - p_2)^2$
	(ii) Pneumatic systems:		
	• Resistance	$\dot{m} = \dfrac{p_1 - p_2}{R_{pn}}$	$P = \dfrac{1}{R_{pn}}(p_1 - p_2)^2$
	• Inertance	$\dot{m} = \dfrac{1}{I_{pn}}\int(p_1 - p_2)dt$	$E = \dfrac{1}{2}I_{pn}\,\dot{m}^2$
	• Capacitance	$\dot{m} = C_{pn}\dfrac{d(p_1 - p_2)}{dt}$	$E = \dfrac{1}{2}C_{pn}(p_1 - p_2)^2$
4.	***Thermal systems :***		
	• Resistance	$Q_{th} = \dfrac{T_1 - T_2}{R_{th}}$	–
	• Capacitance	$Q_{th1} - Q_{th2} = C_{th}\dfrac{dT}{dt}$	$E = C_{th}T$

6.2 SYSTEM MODELS

6.2.1. Introduction

In the previous article we have discussed the basic building blocks for mechanical, electrical and fluid systems separately. However, in engineering many systems encountered involve aspects of more than one of these systems (*e.g.*, electric motor involves electrical as well as mechanical elements). In this article we shall discuss how single-discipline building blocks can be combined to give models for such multi-discipline systems.

Usually ***"linearised" mathematical models*** *are used* because of the following *reasons:*

(*i*) *Most of the techniques of control systems are based on there being linear relationships for the elements of such systems.*

(*ii*) Most control systems are maintaining an output equal to some reference value, the *variations from this value tend to be rather small and so the linearised model is perfectly appropriate.*

6.2.2. Rotational-Translational Systems

In several mechanisms the conversion of rotational motion to translational motion or vice versa is involved (*e.g.*, Rack-and-pinion, shafts with lead screws etc.)

In order to analyse such a system let us consider a rack-and-pinion system as shown in Fig. 6.18. The rotational motion of the pinion is converted into translational motion of the rack.

Let, T_{in} = Input torque,

T_{out} = Torque output,

I = Moment of inertia of the pinion,

r = Radius of the pinion,

ω = Angular velocity of the pinion, and

v = Output velocity of the rack

$= \omega r$

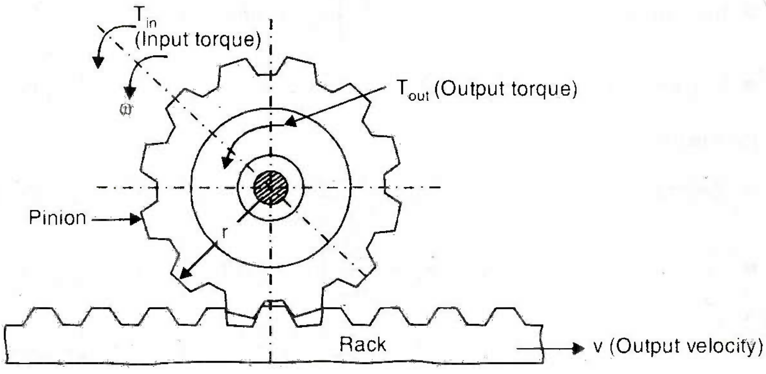


Fig. 6.18. Rack and pinion.

For pinion element :

$$T_{in} - T_{out} = I \cdot \frac{d\omega}{dt}$$ ...assuming negligible damping

(where, $\frac{d\omega}{dt} = \alpha$ = angular acceleration) ...(6.61)

or, $$T_{in} - T_{out} = \frac{I}{r} \cdot \frac{dv}{dt}$$ ...(6.62)

$\left(\because\ v = wr, \text{ and } \frac{dv}{dt} = r \cdot \frac{dw}{dt} \text{ or } \frac{dw}{dt} = \frac{1}{r} \cdot \frac{dv}{dt}\right)$

For rack element :

Due to the movement of the pinion, the rack element will be subjected to a force of $\frac{T}{r}$.
If cv is the frictional force then the net force is :

$$\frac{T_{out}}{r} - cv = m \cdot \frac{dv}{dt}$$ ...(6.63)

($\because$ As per Newton's second law, $F = ma$)

or, $$T_{out} - rcv = r \cdot m \cdot \frac{dv}{dt}$$

or, $$T_{out} = rcv + rm \cdot \frac{dv}{dt}$$

Substituting the value of T_{out} in eqn. (6.62) we get,

$$T_{in} - rcv - rm\frac{dv}{dt} = \frac{I}{r} \cdot \frac{dv}{dt}$$

or, $$T_{in} - rcv = \left(\frac{I}{r} + mr\right)\frac{dv}{dt}$$

or, $$T_{in} - rcv = \left(\frac{I+mr^2}{r}\right)\frac{dv}{dt}$$

or, $$\frac{dv}{dt} = \left(\frac{r}{1+mr^2}\right)(T_{in} - rcv) \quad ...(6.64)$$

This equation is a *first-order differential equation* describing how the output is related to the input.

6.2.3. Electromechanical Systems

The electromechanical devices transform electrical signals to rotation or vice versa:

Examples :

(*i*) An *electric motor* gets an input of a potential difference and gives an output of rotation of a shaft.

(*ii*) A *generator* receives rotation of shaft as input and gives an output of a potential difference.

(*iii*) A *potentiometer* gets an input of a rotation and supplies an output of a potential difference.

Fig. 6.19 shows a *rotary potentiometer* which is potential divider. Thus,

$$\frac{V_{out}}{V} = \frac{\theta}{\theta_{max}} \quad ...(6.63)$$

where, V_{out} = Output voltage for input θ,

V = Potential difference across the full length of the potentiometer track,

θ = Angle swept for V_{out}, and

θ_{max} = The total angle swept out by the slider in being rotated from one end of the track to the other.

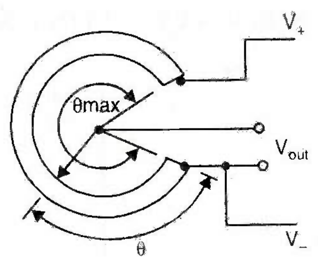


Fig. 6.19. Rotary potentiometer.

6.2.4. Hydraulic-mechanical Systems

These systems *involve the transformation of hydraulic signals to translational or rotational motion, or vice versa.*

Example: The movement of a piston in a cylinder as a result of hydraulic pressure involves the transformation of a hydraulic pressure input to the system to a translational motion output.

6.3 CONTROLLERS

6.3.1. Introduction

Whereas the ***open-loop control*** is essentially just a *switch on-switch off form of control,* but with a ***closed-lop control systems*** *a controller is used to compare the output of a system with the required condition and convert the error into a control action designed to reduce the error.*

- The *digital control* is used when the computer is in the feedback loop and exercising control in this way.
- The term *programmable logic control* (PLC) is used for a simple controller *based on*

a microprocessor and operates by examining the input signals from sensors and carrying out logic instructions which have been programmed into the memory.

Here we shall discuss about closed-loop control.

6.3.2. Control Modes

The various types of control modes (*i.e.*, the ways in which controllers can react to error signals) are:

1. Two-step mode.
2. Proportional mode (P).
3. Derivative mode (D).
4. Integral mode (I).
5. Combinations of modes: PD, PI and PID.

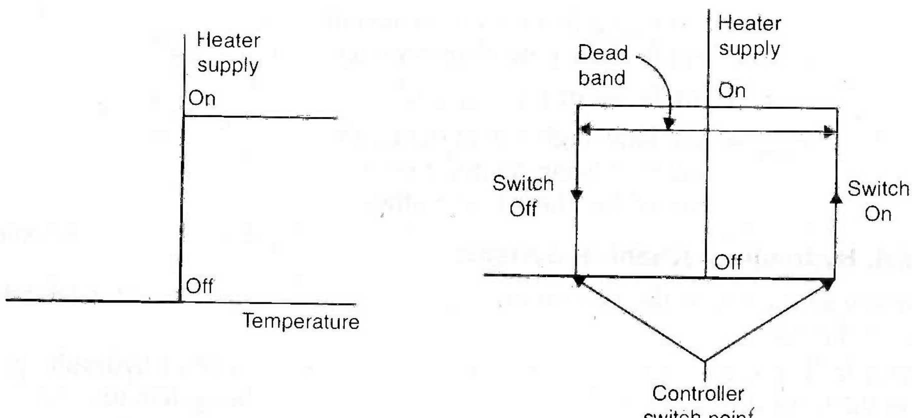
The above modes can be achieved by a controller by means of pneumatic circuits, analogue electronic circuits involving operational amplifiers or by the programming of a microprocessor or computer.

6.3.3. Two-step Mode

In such a mode the controller is essentially just a switch which is activated by the error signal and supplies just as an on-off correcting signal.

Example: The *'bimetallic strip'* that may be used with a simple temperature control system.

- In this type of mode control action is *discontinuous*.



(a) One controller switch point (b) Two controller switch points

Fig. 6.20. Two-stop control with one and two controller switch points.

Fig. 6.20(*a, b*) shows two-step control with one controller switch point and two controller switch points respectively.

6.3.4. Proportional Mode (P)

In a proportional-mode method of control, the size of the controller is *proportional to the size of the error* (whereas in a two-step method of control the control output is either an 'on' or an 'off' signal, irrespective of the magnitude of the error).

Fig. 6.21 shows the output variations of a proportional-mode controller, with the size and sign of error.

Proportional band is the range of errors over which the linear relationship between controller output and error tends to exist. The equation of straight line within the proportional band can be represented by :

Change in output of the controller from set point = $K_P \cdot e$ ...(6.66)

where, I_o = The controller output percentage at zero error,

I_{out} = The controller output percentage at error e,

K_P = A constant, and

e = The error.

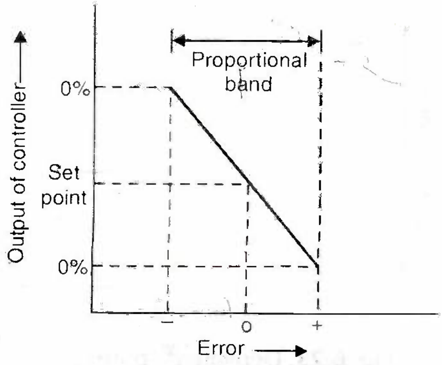


Fig. 6.21. Proportional band.

- Fig. 6.22 shows a summing operational amplifier with an inverter *used as a proportional controller.*

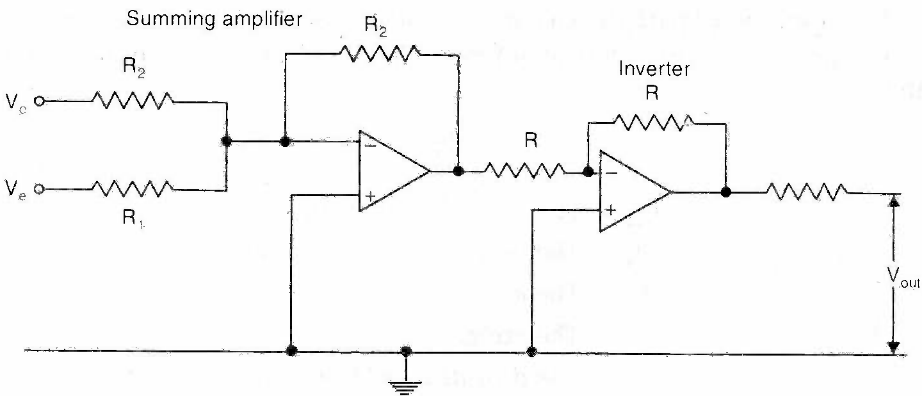


Fig. 6.22. Proportional controller.

6.3.5. Derivative Mode (D)

In this type of control the change in controller output from the set point value is *proportional to the rate of change with time of the error signal.* This can be represented by two equations:

$$I_{out} - I_o = K_D \frac{de}{dt} \quad ...(6.67)$$

where, I_o = The set point output value, and

I_{out} = The output value that will occur when the error e is changing at the rate $\frac{de}{dt}$, and

K_D = Constant of proportionality.

Fig. 6.23, shows the output of controller that results when there is a constant rate of change of error with time.

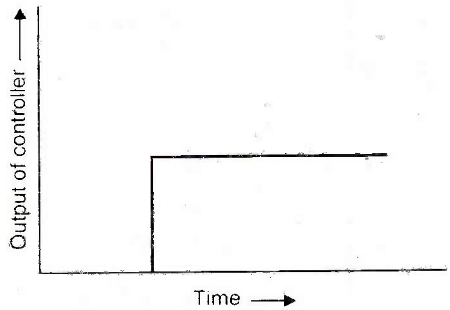


Fig. 6.23. Derivative control.

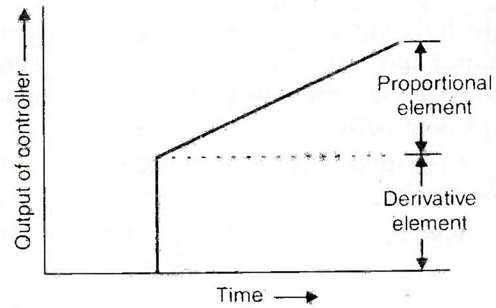


Fig. 6.24. PD control.

6.3.5.1. PD controller

Since derivative controllers do not respond to steady-state error signals (as with these signals the rate of error change with time is zero), *the derivative control is always combined with proportional control.*

— The derivative part *responds to the rate of change;*

— The proportional part gives a *response to all error signals* (including steady signals).

In a PD (proportional plus derivative) controller the change in the output of controller from the set point value is given by:

$$I_{out} - I_o = K_P e + K_D \frac{de}{dt} \quad ...(6.68)$$

where, I_{out} = The output when error is e,

I_o = The output at the set point,

K_P = The *proportionality constant,*

e = The error,

K_D = The *derivative constant,* and

$\frac{de}{dt}$ = The rate of change of error.

Fig. 6.24 shows the variation of the output of controller when the error changes constantly.

6.3.6. Integral Mode (I)

In this type of control the rate of change of the output of the control I is proportional to the input error signal e.

i.e., $$\frac{dI}{dt} = K_I e \quad ...(6.69)$$

where, K_I = The constant of proportionality.

Integrating the above equation we get

$$\int_{I_o}^{I_{out}} dI = \int_0^t K_I e \, dt$$

or, $$I_{out} - I_0 = \int_0^t K_I e \, dt \qquad \text{...(6.70)}$$

where, I_0 = The output of controller at zero time, and

I_{out} = The output of controller at time t.

Fig. 6.25 shows the action of an integral controller when there is a constant error input to the controller:

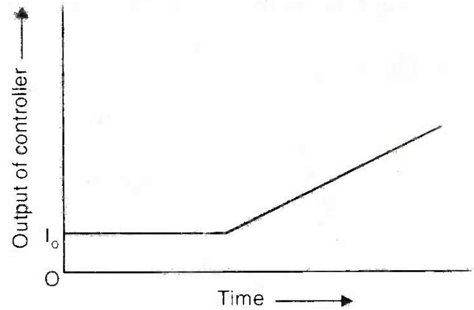


Fig. 6.25. Integral control.

— When the controller output is constant, the error is zero;

— When the controller output varies at a constant rate the error has a constant value.

6.3.6.1. PI controllers

Normally, the integral mode is not used alone but is frequently used in conjunction with the proportional mode. The equation of the PI control system is given as :

$$I_{out} - I_0 = K_P e + \int K_I e \, dt \qquad \text{...(6.71)}$$

Fig. 6.26, shows how the system reacts when there is an abrupt change to a constant error:

— The error gives rise to a proportional controller output which remains constant since there is no change in error;

— On this is then superimposed a steadily increasing controller output due to the integral action.

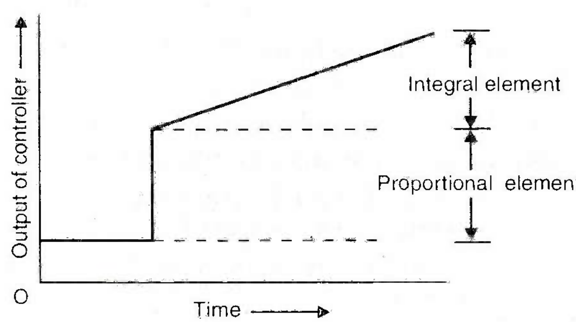


Fig. 6.26. PI control.

6.3.7. PID controllers

PID controller is one in which all the three modes of control, Proportional (P), Integral (I) and Derivative (D) are combined together. In such a controller there *is no offset error and tendency for oscillations is reduced.*

The equation of this controller is written as :

$$I_{out} - I_0 = K_P e + K_I \int e \, dt + K_D \frac{de}{dt} \qquad \text{...(6.72)}$$

where, I_{out} = The output from the controller,

I_0 = The set point output when there is no error,

K_P = The proportionality constant,

e = Error,

K_I = The integral constant, and

K_D = The derivative constant.

Fig. 6.27, shows an operational amplifier PID circuit.

Here, $$K_P = \frac{R_1}{R+R_D};\ K_D = R_D C_D;\ K_I = \frac{1}{R_1 C_1}$$

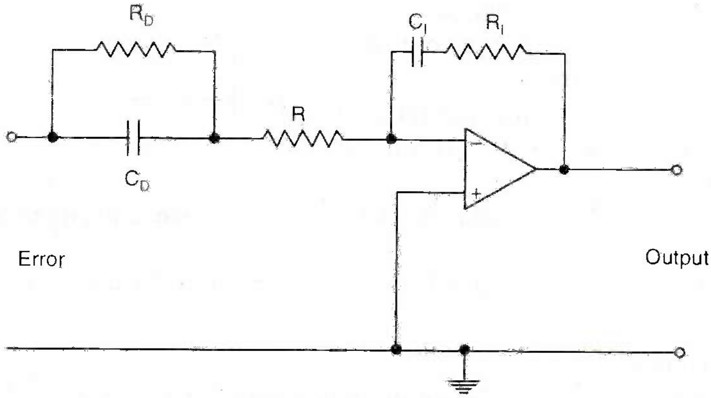


Fig. 6.27. PID circuit.

6.3.8. Digital Controllers

*The *digital controllers require inputs which are digital, process the information in digital form and give an output in digital form.*

The controller performs the following functions:

(*i*) Receives input from *sensors;*

(*ii*) Executes control programs;

(*iii*) Provides the output to the correction elements.

— As several control systems have analogue measurements an analog-to-digital converter (ADC) is used for the inputs.

Fig. 6.28 shows the digital closed-loop control system which can be used with a continuous process:

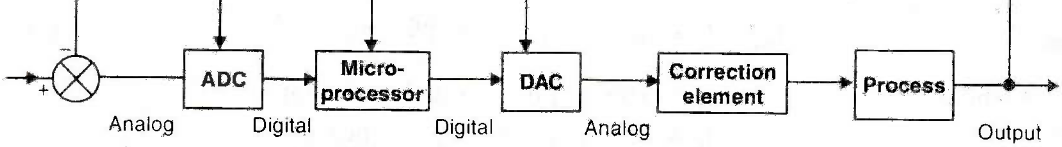


Fig. 6.28. Digital closed-loop control system.

— The *clock* supplies a pulse at regular time intervals and dictates when samples of controlled variables are taken by the *ADC*.

* The term *digital control* is used when the digital controller, basically microprocessor, is in control of the closed-loop control system.

— These samples are then converted to digital signals which are compared by the *microprocessor* with the set point value to give the error signal.
— The error signal is then processed by a *control mode* (initiated by the microprocessor) and digital output is produced.
— The digital output, generally after processing by an ADC since correcting elements generally require analog signals, can be used *to initiate the corrective action.*

Basically, a digital controller carries out the following sequence of operations:

- *Samples* the measured value.
- *Compares* this measured value with the set value and establishes the error.
- *Makes calculations* based on the error value and stored values of previous inputs and outputs to obtain the output signal.
- *Sends* the output signal to the digital-to-analog converter (DAC).
- *Waits* until the next sample time before repeating the cycle.

Advantages of microprocessors as controllers over analog controllers :

The microprocessor, as controllers, claim the following *advantages over analog controllers* :

1. The form of controlling action (*e.g.*, proportional or three mode) can be changed by purely a change in the computer software.
2. No alteration in hardware or electrical wiring is required.
3. Whereas with analog control, separate controllers are required for each process being controlled, however, with a microprocessor many separate processes can be controlled by sampling processes with a *multiplexer.*

• As compared to analog control, *digital control gives better accuracy* because the amplifiers and other components used with analog systems change their characteristics with time and temperature and so show drift, while digital control does not suffer from drift in the same way since it *operates on signals in only the on-off mode.*

6.3.9. Adaptive Control System

An "adaptive control system" is one which adapts to changes and changes its parameters to fit the prevailing circumstances. It is based on the use of a microprocessor as the controller. This system consists of the following *three stages* of operation :

(*i*) To start to operate with controller conditions set on the basis of as assumed condition.
(*ii*) To compare continuously the desired performance with the actual performance of the system.
(*iii*) To adjust automatically and continuously the control system mode and parameters in order to minimise the difference between the desired and actual performance of the system.

Out of the several forms of the adaptive control systems, the following three are commonly used :

1. ***Gain-scheduled control.*** In this type of control preset changes in the controller's parameters are made on the basis of some auxiliary measurement of some process variable.
   - The *advantage* of this control is that the changes in the parameters can be made quickly when the conditions change. However, the *limitation* of this

system is that the control parameters have to be determined for many operating conditions so that the controller can select the one to fit the prevailing conditions.

2. ***Self-timing.*** This system (also being referred to as *auto-timing*) continuously times its own parameters based on monitoring the variable that the system is controlling and the output from the controller.
   - It is often being used in commercial PID controllers.
3. ***Model-reference adaptive systems :***
   - In this system an accurate system model is developed.
   - The set value is then used as input to both the actual and the model systems and the difference between the actual output and the output from the model compared.
   - The difference in the above signals is then used to adjust the parameters of the controller to minimise the difference.

6.3.10. Programmable Logic Controllers (PLCs)

6.3.10.1. Introduction

PLCs are specialised industrial devices for interfacing to and controlling analog and digital devices.

— They are designed with a small instruction set suitable for industrial control applications.

— They are usually programmed with *"ladder logic"*, which is graphical method of laying out the connectivity and logic between system inputs and outputs.

— They are designed with industrial control and industrial environments specifically in mind. Therefore, in addition to being *flexible and easy to program, they are robust and relatively immune to external interference.*

- A programmable logic controller (first conceived in 1968), is shown in Fig. 6.29. *It is a "digital electronic device" that uses a programmable memory to store instructions and to implement functions such as logic sequencing, timing, counting and arithmetic in order to control machines and processes.*

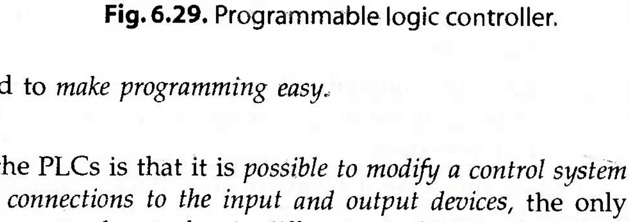


Fig. 6.29. Programmable logic controller.

It has been specifically designed to *make programming easy.*

Advantages :

(*i*) The primary ***advantage*** of the PLCs is that it is *possible to modify a control system without having to rewire the connections to the input and output devices,* the only requirement being that an operator has to *key in different set of instructions.*

(*ii*) PLCs are also *much faster than relay-operated systems.*

Uses. PLCs are widely used and extend from small-contained units for use with perhaps 20 digital inputs/outputs to modular systems which can be used for large numbers of inputs/outputs, handle digital or analog inputs/output, and also carry out PID control modes.

6.3.10.2. Special features

Although PLCs are similar to computers, yet they have the following *specific features* to their use as controllers:

1. The interfacing for inputs and outputs is *inside* the controller.
2. *Easily programmable.* They have an easily understood programming language.
   — Programming is mainly concerned with *logic* and *switching operation.*
3. *Rugged* and designed to withstand vibrations, temperatures, humidity and noise.

6.3.10.3. Architecture basic structure

Fig. 6.30 shows the architecture/internal structure of a programmable logic controller (PLC):

A PLC consists of the following *main components* :

1. Central processing unit (CPU);
2. Memory;
3. Input/Output circuitry.

1. ***Central processing unit (CPU) :***
   - It controls and processes all the operations within the PLC.

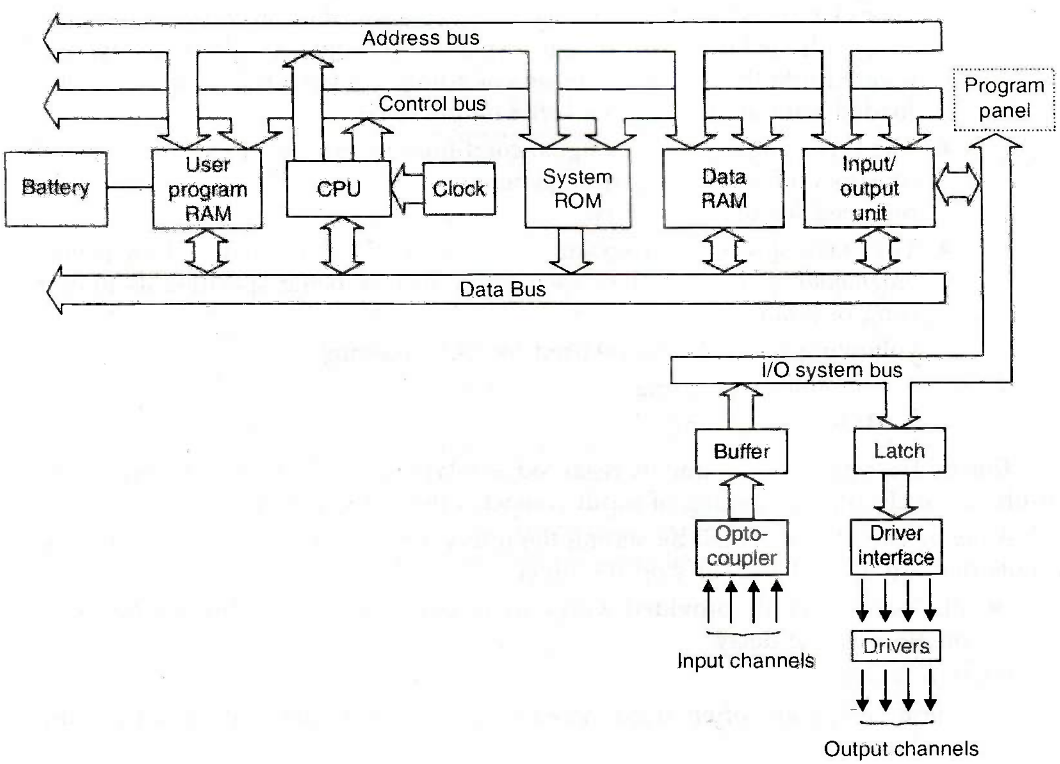


Fig. 6.30. Architecture of a programmable logic controller (PLC).

- It is provided with a *"clock"* with a frequency of typically between 1 and 8 MHz. This frequency determines the operating speed of the PLC and provides the timing and synchronisation for all elements in the system.

- A *"bus system"* carries information and data to and from the CPU, memory and input/output units.

2. *Memory:* The various memory elements available in a PLC are:
   (*i*) A system ROM to give permanent storage for the operating system and fixed data.
   (*ii*) RAM for user's program.
   (*iii*) Temporary buffer stores for input/output channels.
   - The programs in RAM can be changed by the user. However, to prevent the loss of these programs when the supply is switched off a battery is likely to be used in the PLC to maintain the RAM contents for a period of time.
   - Specifications for small PLCs often specify the program memory size in terms of the number of program step (A program step is an instruction for some event to occur) that can be stored. Typically the number of steps that can be handled by a small PLC is of the order of 300 to 1000, which is generally more than adequate for most control situations.

3. *Input/Output (I/O) circuitry:* The I/O unit provides the interface between the system and outside world.
   - Programs are entered into the I/O unit from a panel which can vary from small keyboards with liquid crystal displays to those using a visual display unit with keyboard and screen display. The programs, alternatively, can be entered into the system by means of a link to a personal computer which is loaded with an appropriate software package.
   - The I/O channel provides signal conditioning and isolation functions so that sensors and actuators can be generally directly connected to them without the need for other circuitry.
   - The basic form of programming commonly used with PLCs is *ladder programming*. This involves each program task being specified as though a rung of a ladder.

   Following methods can be used for *I/O processing* :

   1. Continuous updating.
   2. Mass I/O copying.

Timers: The timers are commonly regarded as relays with coils which, when energised, result in the closing or opening of input contacts after some preset time.

A *timing circuit* is specified by stating the interval to be timed and the conditions or events that are to start and/or stop the timer.

- PLCs are generally provided with only *delay-on timers, i.e.,* a timer which comes on after a time delay.

Internal relays:

- These relays are often used when there are programs with multiple input conditions.
- The internal relays are also used for the starting of multiple outputs.

Counters: The use of counters is restored to when there is *a need to count a specified number of contact operations.*

- Counter circuits are supplied as an internal feature of PLCs.

Shift registers: Several internal relays can be grouped together to form a *register* which can provide a storage area for a series sequence of individual bits. Thus a 4-bit and a 8-bit registers would be formed by using four and eight internal registers respectively.

The term *shift register* is used because the bits can be shifted along by one bit when there is a suitable input to the register.

Shift registers have *three inputs :*

- One to load data into the first element of the register (OUT);
- One as the shift command (SFT);
- One for resetting (RST).

6.3.10.4. Selection of a PLC

For selection of a PLC, the following criteria need to be considered:

1. *Types of inputs/outputs* required, such as:
   — Isolation;
   — Out-board power supply for inputs/outputs;
   — Signal conditioning.
2. *Input/Output capacity* required.
3. *Size of memory* required. This is linked to the number of inputs/outputs and the complexity of program used.
4. *Speed and power* required for CPU—This is linked to the number of types of instructions that can be handled by a PLC.

HIGHLIGHTS

1. The *mathematical models* are equations which describe the relation between the input and output of a system.
2. Mechanical system building blocks are: *Springs; dashpots; masses.*
3. Electrical system building blocks are: Resistors; inductors; capacitors.
4. Fluid system building blocks are: *Resistance; inertance, capacitance.*
5. Thermal system building blocks are: *Resistance; capacitance.*
6. Various types of control modes are: Two-step mode; proportional mode (P); derivative mode (D); integral mode (I); combinations of modes.
7. PLCs (Programmable logic controllers) are special industrial devices for interfacing to and controlling analog and digital devices.

OBJECTIVE TYPE QUESTIONS

Fill in the Blanks or Ray 'Yes' or 'No'

1. Systems can be made up from a range of ....
2. The .... models are equations which describe the relation between the input and output of a system.
3. The mechanical system building blocks are: Springs, dashpots and ....
4. In a dashpot no energy is stored.
5. Energy stored by the mass rotating with an angular velocity, $\omega = \frac{1}{2}I\omega^3$.
6. The electrical system building blocks are resistor, inductors and capacitors.
7. The various electrical building blocks can be combined by using .... laws.
8. Hydraulic .... is equivalent of a spring in mechanical systems.

9. The pneumatic inertance is due to the pressure drop necessary to accelerate a block of gas.
10. Open-loop control is just a switch on–switch off form of control.
11. In a two-step mode control action is continuous.
12. In proportional mode (P) method of control the size of the controller is .... to the size of the error.
13. The derivative control is always combined with proportional control.
14. The .... controllers require inputs which are digital, process the information in digital form and give an output in digital form.
15. As compared to digital control, the analog control gives better accuracy.
16. .... are specialised industrial devices for interfacing to and controlling analog and digital devices.
17. PLC consists of CPU, memory and .... circuitry.
18. The .... are commonly regarded as relays with coils which, when energised, result in the closing or opening of input contacts after some preset time.
19. Internal relays are often used when there are programs with multiple input conditions.
20. PLCs are rarely provided with delay-on timers.

ANSWERS

1. building blocks	2. mathematical	3. masses	4. Yes
5. No	6. Yes	7. Kirchhoff's	8. Inertance
9. Yes	10. Yes	11. No	12. proportional
13. Yes	14. digital	15. No	16. PLCs
17. Input/Output	18. timers	19. Yes	20. No

THEORETICAL QUESTIONS

1. What are mathematical models? Explain briefly.
2. Explain briefly the following basic building blocks of a mechanical system:
   (*i*) Springs; (*ii*) Dashpots; (*iii*) Masses.
3. Enumerate and explain briefly the three building blocks of a rotational system.
4. Explain briefly a mathematical model of a car moving on a road.
5. Explain briefly the following building blocks of an electrical system:
   (*i*) Resistors; (*ii*) Inductors; (*iii*) Capacitors.
6. How can Kirchhoff's laws be used for combining building blocks of electrical systems? Explain briefly.
7. Discuss briefly the various fluid systems building blocks.
8. What is a hydraulic inertance? Explain briefly.
9. What is pneumatic inertance? Explain briefly.
10. Explain briefly building up models for the following systems:
    (*i*) Mechanical system.
    (*ii*) Hydraulic system.
    (*iii*) Pneumatic system.
11. Explain briefly the following thermal system building blocks:
    (*i*) Resistance; (*ii*) Capacitance.
12. How is the model for a thermal system built up? Explain.
13. Write a short note on system models.
14. Explain briefly the following:
    (*i*) Rotational-translational systems.

(*ii*) Electromechanical systems.

(*iii*) Hydro-mechanical systems.

15. Explain briefly any two of the control modes:

(*i*) Two-step mode;
(*ii*) Proportional mode (P);
(*iii*) Derivative mode;
(*iv*) Integral mode (I).

16. Discuss briefly the following controllers:

(*i*) PI controllers;
(*ii*) PD controllers;
(*iii*) PID controllers.

17. What are digital controllers? Explain briefly.
18. What the advantages of microprocessors as controllers over analog controllers?
19. Discuss briefly 'Adaptive control system'.
20. What are programmable logic controllers? Explain.
21. State the advantages and uses of PLCs.
22. State the special features of a PLC.
23. Discuss briefly with a neat sketch the architecture of a PLC.
24. Explain briefly the following:
Timers; Counters; Shift registers.
25. What criteria should be considered while selecting a PLC?

CHAPTER

7 Actuators—Mechanical, Electrical, Hydraulic and Pneumatic

7.1 INTRODUCTION

In most mechatronic systems, motion or action of some sort is involved; it is created by a force or torque that results is acceleration and displacement. *This motion or action (which can be applied to anything from a single atom to large articulated structure) is produced by the devices known as* **activators.**

Actuators *produce physical changes such as linear and angular displacement. They also modulate the rate and power associated with these changes.*

The proper selection of the appropriate type of actuator is an important aspect of mechatronic system design.

We shall discuss briefly the following actuators:

1. Mechanical actuators
2. Electrical actuators
3. Hydraulic actuators
4. Pneumatic actuators

7.2 MECHANICAL ACTUATORS

7.2.1. General Aspects

Mechanical actuators or mechanisms *are devices which can be considered to be motion converters in that they transform motion from one form to some other required form.* For example, they might transform linear motion into rotational motion, or motion in one direction into a motion in a direction at right angles, or perhaps a linear reciprocating motion into rotary motion, as in the internal combustion engine where the reciprocating motion of the pistons is converted into rotation of the work and hence the drive shaft.

Mechanical elements include the use of linkages, cams, gears, rack-and-pinion, chains, belt drives etc. For example:

— Cams and linkages can be used to obtain motions which are prescribed to vary in a particular manner.

— Parallel shaft gears might be used to reduce a shaft-speed. Bevel gears might be used for the transmission of rotary motion through 90°.

— Rack-and-pinion can be used to convert rotational motion to linear motion.

— A toothed belt or chain drive might be used to transform rotary motion about one axis to motion about another.

Several actions which were earlier obtained by use of mechanisms are, however, often now days being obtained by the use of "microprocessor system". For example earlier, cams on a rotating shaft were used for domestic washing machines in order to give a timed sequence of actions such as opening a valve to let water into the drum, switching the water off, switching a heater on etc. But now-a-days *modern washing machines employ a microprocess-based system with the microprocessor programmed to switch on outputs in the required sequence.*

However, mechanisms/mechanical actuators still have a role in mechatronic systems to provide such functions as:

(*i*) Change of speed, e.g., that given by gears.

(*ii*) Specific type of motion, e.g., that given by a quick-return mechanism.

(*iii*) Amplification of force, e.g., that given by levers.

(*iv*) Transfer of rotation about one axis to rotation about another, e.g., timing belt.

7.2.2. Machine

"It is an apparatus for applying mechanical power, consisting of a number of interrelated parts, each having a definite function."

Or

"It is a device by means of which available energy can be converted into desired form of useful work".

— *A* ***Machine*** *is the assembly of resistant bodies or links whose relative motions are successfully constrained so that available energy can be converted into useful work.*

— Machines are used *to transmit both motion and force.*

— *A body is said to be resistant if it can transmit the required force with negligible deformation.* These bodies are the parts of the machines which are employed for transmitting motion and forces.

7.2.3. Kinematic Link or Element

Definition and characteristics :

Definition. ***Kinematic element*** *is a resistant body or an assembly of resistant bodies which go to make a part or parts of a machine connecting other parts which have motion 'relative' to it.*

— A kinematic link is assumed to be completely rigid. The machine components which do not fit this assumption of rigidity, such as springs, usually have no effect on the kinematics of device but do play a role in supplying forces. Such members are *not* called links.

Example. Fig. 7.1 shows a reciprocating steam engine. Here,

— Piston, piston rod and cross head ... *one link.*

— Connecting rod with big and small end bearings ... *second link.*

— Crankshaft and flywheel ... *third link.*

— Cylinder, engine frame and main bearings ... *fourth link.*

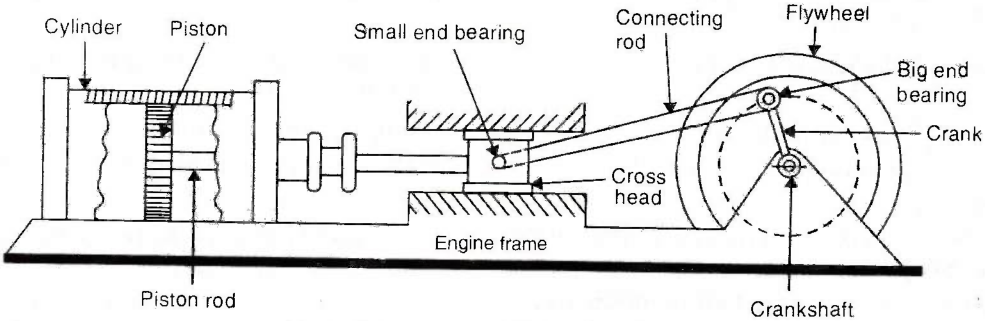


Fig. 7.1. Reciprocating steam engine.

Characteristics of a link. A link should have two *characteristics* :

1. It should have *relative motion.*
2. It must be a *resistant body* (*need not be rigid body*).

Types of links: The various types of links are :

1. **Rigid link.** *A link which does not undergo any deformation while transmitting motion is called a "**rigid link**".*
   — Strictly speaking, rigid links do not exist. However, since the deformation of a connecting rod, crank etc. of a connecting rod, crank etc. of reciprocating steam engine is *not appreciable,* they *can* be considered as rigid links.
2. **Flexible link.** A *flexible link in one which is partly deformed in a manner not to affect the transmission of motion.*
   Example : Belts, ropes, chains and wires (these link transmit tensile forces only).
3. **Fluid link.** A *fluid link is one which is formed by having a fluid in receptacle and the motion is transmitted through the fluid by 'pressure' or 'compression' only*
   Example : Hydraulic presses, jacks and *brakes.*

Difference between Machine and Structure :

Structure is an assemblage of a number of resistant bodies (known as *members*) having *no relative motion* between them and meant for carrying *load having straining action.*

Examples. A railway bridge, a roof truss, machine frames etc.

The differences between a *'machine'* and *'structure'* are given in tabular form below.

Machine	*Structure*
1. Parts of a machine *move* relative to each other.	1. The members of a structure *do not move* relative to one another.
2. It transforms the available energy into some useful work.	2. No energy is transformed into useful work.
3. The links may *transmit both power and motion.*	3. The members of a structure *transmit forces only.*
Examples. Shaper, lathe etc.	*Examples.* Roof truss frame etc.

7.2.4. Kinematic Pair

A **kinematic pair** is *a joint of two links that permits relative motion.*

— The relative motion between the elements or links that form a pair is required to be *completely constrained or successfully constrained.*

— The degree of freedom of a kinematic pair is given by the number of independent coordinates required to completely specify the relative motion. These coordinate are called *"variables"*.

- ***Completely constrained motion.*** When the motion between a pair is *limited to a definite direction* irrespective of the direction of force applied, then the motion is said to be a completely constrained motion.

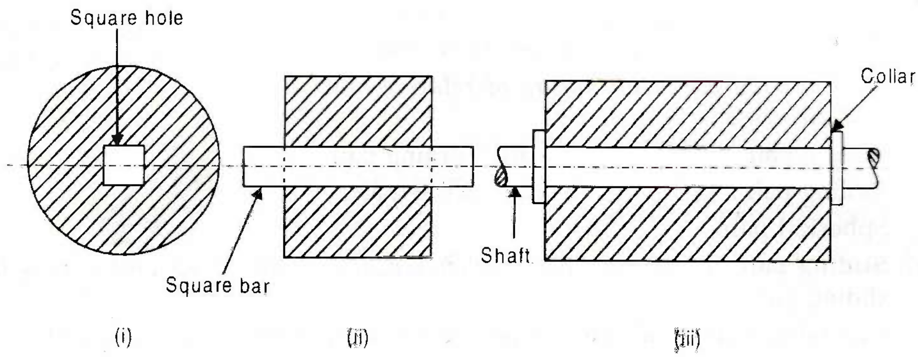


Fig. 7.2. Completely constrained motion.

Examples : The motion of a square bar in a square hole [Fig. 7.2(*i*)], and the motion of a shaft with collars at each end are the examples of the completely constrained motion.

— The motion of the piston and cylinder, (forming a pair) in a steam engine (Fig. 7.1) in which the motion of the piston is limited to a definite direction (*i.e.*, it will only reciprocate) is also an example of completely constrained motion.

- **Incompletely constrained motion.** When the motion between a pair can take place in *more than one direction* then the motion is called an *"incompletely constrained motion"*.

  Examples. A circular bar or shaft in a circular round role, as shown in Fig. 7.3, is an example of incompletely constrained motion as it may either *rotate* or *slide* in a hole.

- **Successfully constrained motion.** The motion is said to be successfully constrained when the motion between the elements, forming a pair, is such that constrained motion is not completed by itself, *but by some means.*

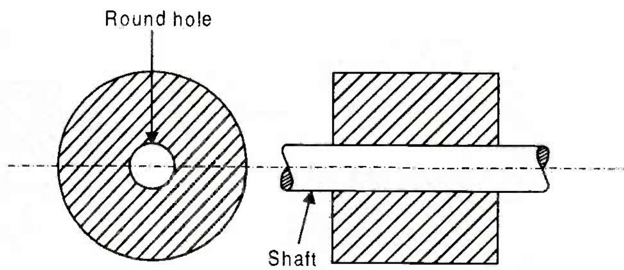


Fig. 7.3. Incompletely constrained motion.

Example. Refer to Fig. 7.4. The shaft may rotate in the bearing or it may move upwards. This is the case of *incompletely* constrained motion. However, if the *load is placed on the shaft to prevent axial upward movement* of the shaft, then the motion of the pair is said to be *successfully* constrained.

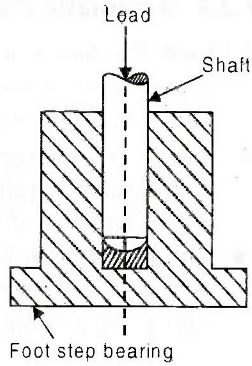


Fig. 7.4. Successfully constrained motion.

Classification of kinematic pairs :

The kinematic pairs may be *classified* on the following considerations :

1. Nature of relative motion between the elements.
2. Nature of contact between the elements.
3. Nature of the mechanical arrangement for complete or successful constraint between the elements.

1. ***Classification based on nature of relative motion between the elements :***

(*i*) Sliding pair (*ii*) Turning pair
(*iii*) Rolling pair (*iv*) Screw pair
(*v*) Spherical pair.

(*i*) **Sliding pair.** If two links have a *sliding motion relative to each other*, they form a sliding pair.

Examples. Piston and cylinder pair, rectangular rod in rectangle hole (Fig. 7.5(*i*)) etc.

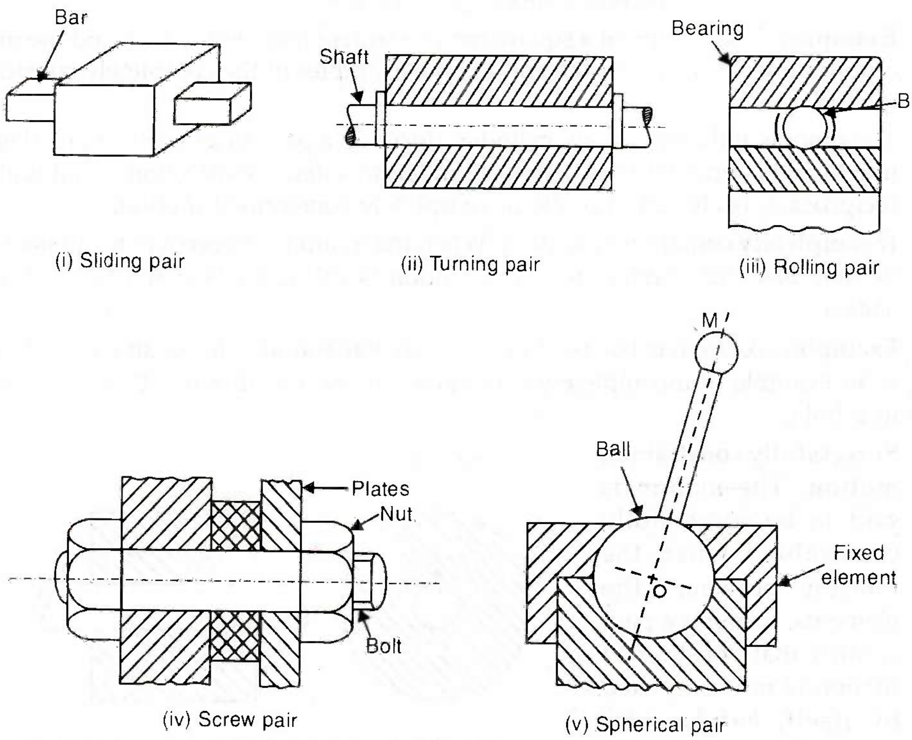


Fig. 7.5

- A sliding pair has a *completely constrained motion.*

(*ii*) **Turning pair.** When one link has *turning or revolving motion relative to the other,* they constitute a turning or revolving pair.

Examples. A shaft rotating in a bearing [Fig. 7.5(*ii*)]. Rotation of a crank in a slider crank mechanism is another turning pair. (It is also known as hinged pair).

(*iii*) **Rolling pair.** *When the links of a pair have a rolling motion relative to each other,* they form a rolling pair.

Examples. Ball and roller bearings. In a ball bearing [Fig. 7.5(*iii*)], the ball and the shaft constitute one rolling pair whereas the ball and the bearing is the second rolling pair.

(*iv*) **Screw (or helical) pair.** When the two elements of a pair are connected in such a way that *one element can turn about the other by screw threads,* the pair is known as *'screw pair'.*

Example. Nut and bolt arrangement [Fig. 7.5(*iv*)].

(*v*) **Spherical pair.** When two elements of a pair are connected in such a way that *one element with spherical shape turns or swivels about the other fixed element* ; the pair formed is called a *'spherical pair'.*

Examples. The ball and socket joint [Fig. 7.5(*v*)]; attachment of a car mirror, pen stand etc.

2. ***Classification based on the nature of contact between elements :***

(*i*) Lower pairs (*ii*) Higher pairs.

(*i*) **Lower pair.** If a pair in motion has a *surface contact* between its elements it is called a *'lower pair'.*

Examples. Shaft rotating in a bearing ; piston moving in a cylinder etc.

(*ii*) **Higher pair.** In a higher pair there is a *line or point* contact between the elements of a pair. The contact surfaces of the two elements are not alike or similar.

Examples. A pair of friction discs, toothed gearing, belt and rope drives, cam and follower, ball and roller bearings etc.

3. ***Classification based on the nature of mechanical constraint :***

(*i*) Closed pairs (*ii*) Unclosed pairs.

(*i*) **Closed pairs.** If the elements of the pair, are *held together mechanically,* they constitute a *'closed pair'.*

Examples. *All lower pairs.*

(*ii*) **Unclosed pairs.** When the two elements are *not held together mechanically,* it forms an *'unclosed pair'.* The two elements are connected *together by gravity or spring force.*

Example. Cam and follower pair (Fig. 7.6).

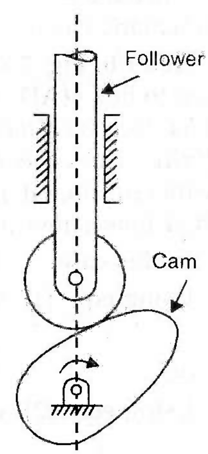


Fig. 7.6. Cam and follower.

7.2.5. Kinematic Chain

When a number of links are connected in space such that, the relative motion of any point on a link with respect to any other point on the other link follows a law the chain is called a **kinematic chain.**

In order to determine whether the assemblage of links and pairs form the kinematic chain or not, the following *two equations* for lower pairs are available :

Eqn. 1. $$l = 2p - 4$$

where, l = Number of *links*, and p = Number of *pairs*.

Eqn. 2. $$l = \frac{2}{3}(j + 2)$$

where, j = Number of *joints*.

If the above *equations are satisfied, the links form a kinematic chain.*

Refer to Fig. 7.7. There are three members and there is *no relative motion* between them. Therefore, it forms a *"structure"* only ; it cannot be chain.

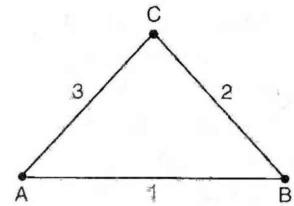


Fig. 7.7

Here, $p = 3; l = 3$

Using Eqn. (1), we have

$$l = 2p - 4 = 2 \times 3 - 4 = 2$$

But $l = 3$, therefore, eqn. 1 is *not satisfied* and hence it is not *a kinematic chain.*

Using Eqn. (2), we have

Here $j = 3$

$$\therefore \quad l = \frac{2}{3}(j + 2) = \frac{2}{3}(3 + 2)$$

or $$l = \frac{10}{3}$$ which is not true.

Since the eqn. (2) is not satisfied, hence it not a kinematic chain.

Refer to Fig. 7.8. If a definite displacement is given to link 4(AD), keeping link 1(AB) fixed, the *resultant motion of the two remaining links is **perfectly definite**.* Thus the relative motion is *completely constrained,* and it is the basis of all machines; dotted lines show the displacement.

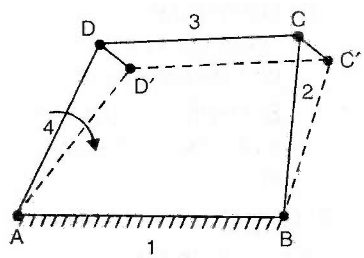


Fig. 7.8

In this case, $l = 4, p = 4, j = 4.$

Using eqn. (1), we have

$$l = 2p - 4$$

or, $$l = 2 \times 4 - 4 = 4$$ which is *true.*

Using eqn. (2), we get

$$l = \frac{2}{3}(j + 2) = \frac{2}{3}(4 + 2) = 4$$

which is again *true.*

Thus, both the eqns. (1 and 2) are satisfied for kinematic chain.

Thus, *the links and pairs from the kinematic chain.*

Refer to Fig. 7.9. In this case, $l = 5, p = 5, j = 5$; with the given data the eqns. (1) and (2) are not satisfied, hence a *kinematic chain is not formed.*

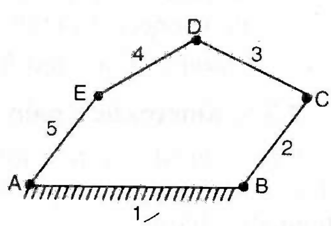


Fig. 7.9

7.2.6. Mechanism

When one of the links of a kinematic chain is fixed, the chain is known as ***mechanism.***

It may be *used for transmitting or transforming motion.*

Examples. Engine *indicators, typewriter etc.*

Mechanisms are of two types:

Simple mechanism. A mechanism with *four links* is known as *simple mechanism.*

Compound mechanism. The mechanism with *more than four links* is known as *compound mechanism.* It may be made by adding two or more simple mechanisms.

- *When a mechanism is required to transmit power or to do some particular type of work, it then becomes a* **machine**.

Difference between mechanism and machine :

Mechanism	*Machine*
1. Transmits and modifies motion.	1. Modifies mechanical work.
2. Skeleton outline of the machine to produce definite motion between various links.	2. May have several mechanism for transmitting mechanical work or power.
3. When kinematic chain is analysed as mechanisms so special consideration need be given to the forms and the cross-sectional proportions of the links except in so far as the assembly locations are involved. *Examples.* Clock work, type writer.	3. As to the machine cross-sectional and proportion requirement to give strength, stiffness, clearance etc. make it necessary to consider links in their details. *Examples.* Shaper etc.

7.2.7. Inversion of Mechanism

As we know that when one of the links in a kinematic chain is fixed, it is called a mechanism, therefore, we can obtain as *many mechanisms as the number of links in a kinematic chain by fixing, in turn different links in a kinematic chain.* This method of *obtaining different mechanisms by fixing different links in a kinematic chain, is known as* ***inversion of the mechanism.***

7.2.8. Types of Kinematic Chains and their Inversions

Important kinematic chains have *four lower pairs, each pair being either a sliding pair or a turning pair;* following are the three important types of kinematic chains.

1. Four bar or quadric cyclic chain — *All four turning pairs.*
2. Single slider crank chain — *Three turning and one sliding pair.*
3. Double slider crank chain — *Two turning and two sliding pairs.*

7.2.8.1. Four bar chain. Refer to Fig. 7.10.

This is also known as *quadric cycle chain.*

— It has *four links* and *four pairs* which are turning in nature.

— Links are of different length.

— One of the rotating links is known as crank or driver and the other link as follower or rocker. The member connecting the crank and the follower is known as connecting rod and fixed link is the frame.

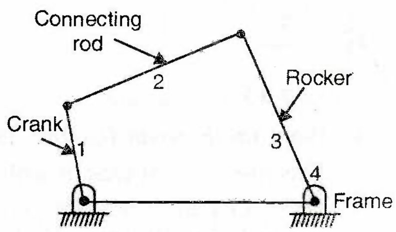


Fig. 7.10. Four bar chain.

Some important **inversions** of *the four bar chain* are :

1. Beam engine
2. Coupled locomotive
3. Pantograph
4. Watt mechanism.

1. ***Beam engine*** (*Crank and lever mechanism*) : Refer to Fig. 7.11.

In this mechanism, when the crank rotates about the fixed centre A, the lever oscillates about a fixed centre D. The end E of the lever CDE is connected to a piston rod which reciprocates due to the rotation of the crank. In other words, the *purpose of this mechanism is to convert rotary motion into reciprocating motion.*

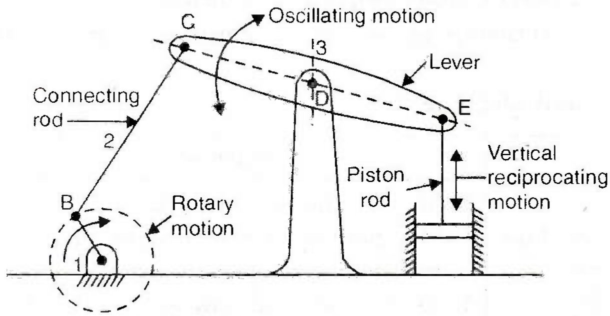


Fig. 7.11. Beam engine.

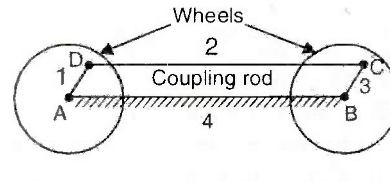


Fig. 7.12. Coupled locomotive.

2. ***Coupled locomotive*** (*Double crank mechanism*) : Refer to Fig. 7.12.

In this mechanism, links AD and BC (having equal length) act as cranks and are connected to the respective wheels. The link CD acts as a coupling rod and the link AB is fixed in order to maintain a constant centre to centre distance between them.

This mechanism is meant for *transmitting rotary motion from one wheel to the other wheel.*

3. ***Pantograph.*** Refer to Fig. 7.13.

- It is a device *used to reproduce a displacement in a reduced* or *an enlarged scale.* It is used for duplicating the drawing maps, plants, etc.
- It is basically a quadric cycle in the form of a parallelogram as shown in Fig. 7.13; all the four pairs are turning in nature.

According to the geometry of the figure: $\dfrac{AC}{AC'} = \dfrac{AE}{AE'}$.

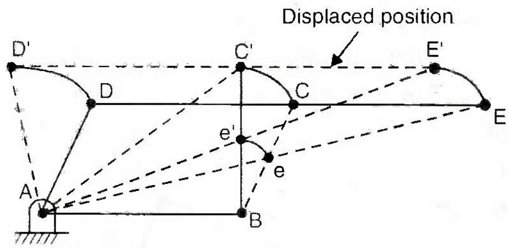


Fig. 7.13. Pantograph.

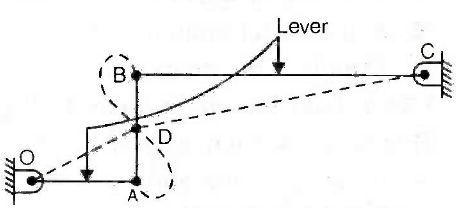


Fig. 7.14. Watt mechanism.

4. ***Watt mechanism*** (*Double lever mechanism*). Refer to Fig. 7.14.

This mechanism was invented by watt for his steam engine to guide the piston rod. Links OA and BC are *parallel in the mean position of the mechanism.* They are connected with links AB. OA and BC are levers, the ends of which are at O and C. For a small displacement of levers OA and BC, D will trace an approximate straight line where point D is located on AB such that $\dfrac{AD}{DB} = \dfrac{BC}{OA}$.

7.2.8.2. Slider crank chain

A single slider crank chain is a modification of the basic four bar chain. It consists of *one sliding pair* and *three turning pairs*. It is, usually, found in *reciprocating steam engine mechanism*. This type of mechanism *converts rotary motion into reciprocating motion and vice versa*.

In a single slider crank chain (Fig. 7.15), the links 1 and 2, links 2 and 3 and links 3 and 4 form three *turning pairs* while links 4 and 1 form a *sliding pair*.

Some important **inversions** of *slider crank chain* are:

1. Pendulum pump.
2. Oscillating cylinder engine.
3. Rotary I.C. engine.
4. Crank and slotted lever quick return motion mechanism.
5. Whitworth quick return motion mechanism.

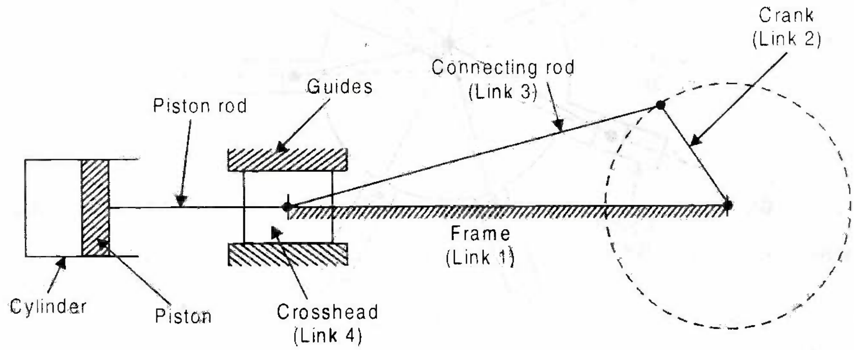


Fig. 7.15. Single slider crank chain.

First three inversion of mechanisms will be discussed here.

1. *Pendulum pump* (*or Bull engine*). Refer to Fig. 7.16.

When crank (link 2) rotates, the connecting rod (link 3) oscillates about a pin pivoted to the fixed link 4 at A and the piston attached to the piston rod (link 1) reciprocates.

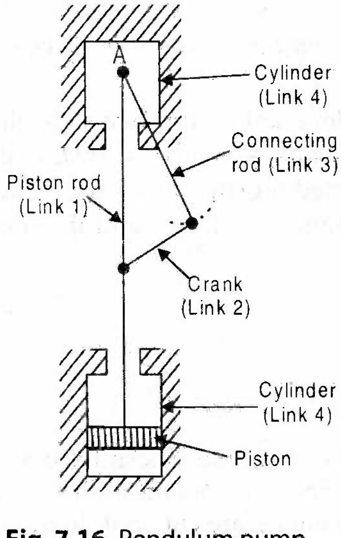


Fig. 7.16. Pendulum pump.

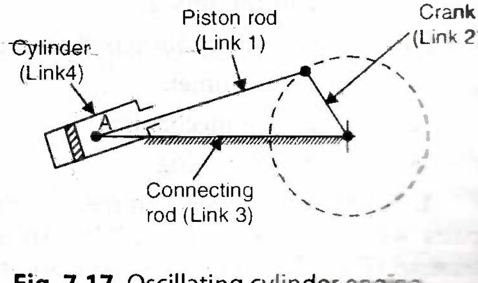


Fig. 7.17. Oscillating cylinder engine.

2. ***Oscillating cylinder engine.*** Refer to Fig. 7.17.

This arrangement is employed to *convert reciprocating motion into rotary motion*. In this mechanism, the link 3 forming the turning pair is fixed (The link 3 corresponds to the connecting rod of a reciprocating steam engine mechanism).

When crank (link 2) rotates, the piston attached to piston rod (link 1) reciprocates and the cylinder (link 4) oscillates about a pin pivoted to the fixed link at *A*.

3. ***Rotary internal combustion engine*** (*or Gnome engine*). Refer to Fig. 7.18.

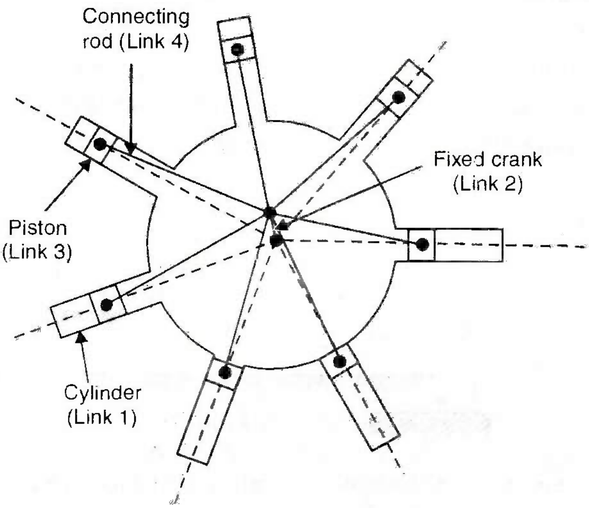


Fig. 7.18. Rotary internal combustion engine.

It consists of seven cylinders in one plane and all revolve about the fixed centre, while the crank (link 2) is fixed.

Here, when the connecting rod (link 4) rotates, the piston (link 3) reciprocates in the cylinders forming link 1.

7.2.8.3. Double slider crank chain.

A kinematic chain which consists of two turning pairs and two sliding pairs is known as ***double slider crank chain.***

- Fig. 7.19 shows the arrangement of a double slider crank chain. Two slide blocks, links 1 and 3, slide along the slots in a frame, link 4, which is *fixed*, and the turning pairs formed at pins *A* and *B* are connected together by a link 2. Each of the slide blocks forms a sliding pair with the frame, *i.e.*, link 4 and the turning pair with the link 2.

Such a kinematic chain has *three inversions* :

1. Elliptical trammel.
2. Scotch yoke mechanism.
3. Oldham's coupling.

1. ***Elliptical trammel.*** Frame, *i.e.*, link 4, is fixed and the slide blocks form sliding pairs with the link 2 in Fig. 7.19. An application of such an inversion is the *Elliptical trammel* (Fig. 7.20). A plate is taken and two slots at right angles are cut on it. In the slots, two sliding blocks are fitted. And these slide blocks are connected by a link. Any point except the mid-points of AB or points A and on the link will trace *"ellipse"*. The points A and B move in *straight line.* The mid-point of AB traces a circle.

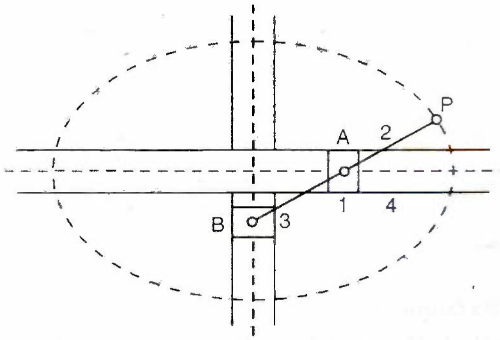


Fig. 7.19. Elliptical trammel.

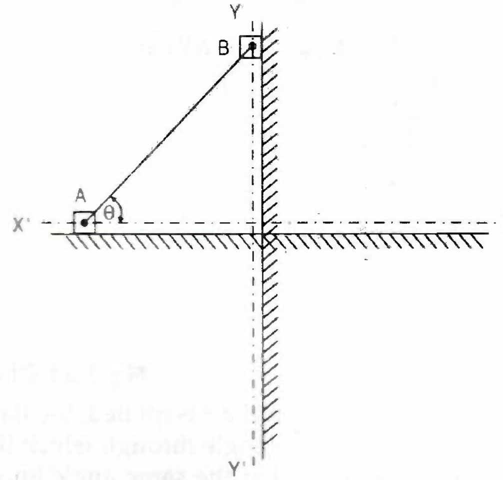


Fig. 7.20

2. ***Scotch yoke mechanism.*** This inversion is used for *converting rotary motion into a sliding or reciprocating motion.*

In Fig. 7.21, link 1 is fixed. In its mechanism when, the link 2 (which corresponds to crank) *rotates* about A as centre, the link 4 (which corresponds to frame) *reciprocates*. The fixed link 1 guides the frame.

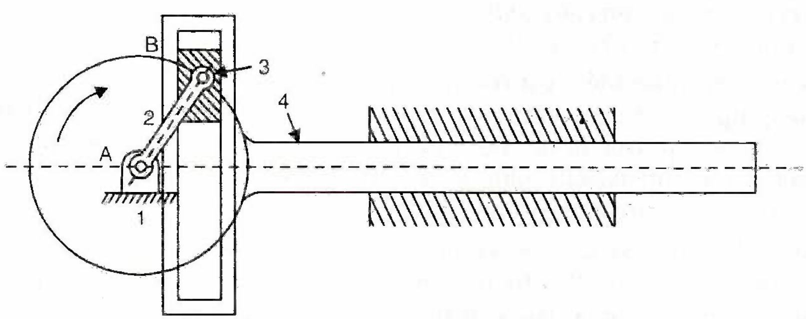


Fig. 7.21. Scotch yoke mechanism.

3. ***Oldham's coupling.*** Refer to Fig. 7.22.

This coupling is used for connecting two parallel shafts when distance between the shafts is small. The shafts are coupled in such a way that if one shaft rotates, the other shaft also rotates at the same speed. This inversion is obtained by fixing the link 2, as shown in Fig. 7.22(*a*).

The two shafts to be connected have flanges rigidly fastened to the shafts, generally forged at the ends. These flanges form links 1 and 3. These links (1 and 3) form turning pairs with link 2. These flanges have diametrical slots cut in their inner faces as shown in Fig. 7.22(*b*). The intermediate piece (link 4) which is a circular disc, having two tongues (*i.e.*, diametrical projections) T_1 and T_2 on each face at right angles to each other, as shown in Fig. 7.22(*c*). The tongues on the link 4 closely fit into the slots in the two flanges (link 1 and link 3). The *link 4 can slide or reciprocate in the slots in the flanges.*

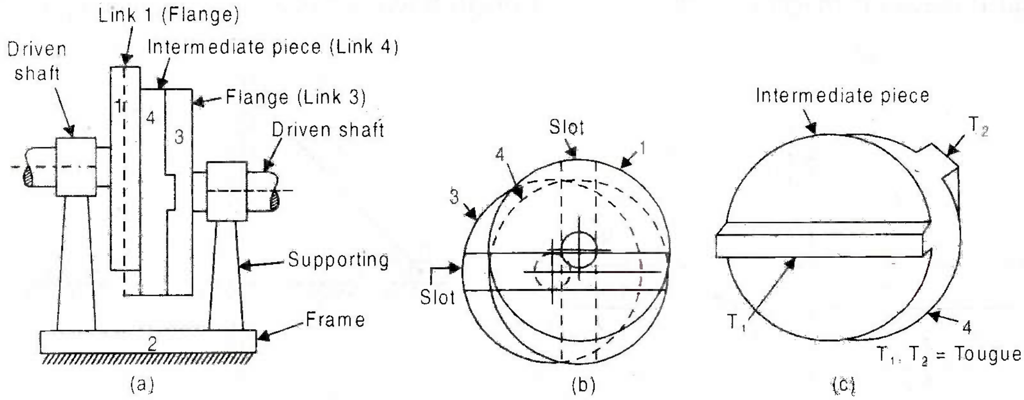


Fig. 7.22. Oldham's coupling.

When the driven shaft is rotated, the flange (link 1) causes the intermediate piece (link 4) to rotate the same angle through which the flange has rotated, and it further rotates the other flange (link 3) at the same angle and thus the driven shaft rotates.

The distance between the axes of the shafts is constant and, therefore, the centre of intermediate piece will follow the path of a circle with diameter equal to the distance between the axes of the shafts. Therefore, *maximum sliding speed of each tongue of intermediate piece in the slot will be given by the peripheral velocity of the centre of the disc along its circular path.*

7.2.9. Gear Drive

Introduction :

- *A* ***gear*** *is a wheel provided with teeth which mesh with the teeth on another wheel, or on to a rack, so as to give a positive transmission of motion from one component to another.*
- Gears constitute the most commonly used device for power transmission or for changing power-speed ratios in a power system. They are used for transmitting motion and power from one shaft to another *when they are not too for apart and when a constant velocity ratio is desired.* Gears also afford a convenient way of changing the direction of motion.
- A number of devices such as *differentials, transmission gear boxes, planetary drives* etc., used in many construction machines employ gears as basic component.

Advantages and disadvantages of toothed gearing :

The following are the *advantages* and *disadvantages* of toothed gearing/gear drive :

Advantages:

1. High efficiency.
2. Long service life.
3. High reliability.
4. More compact.
5. Can operate at high speeds.
6. Can be used where precise timing is required.
7. Large power can be transmitted.
8. Constant speed ratio owing to absence of slipping.
9. Possibility of being applied for a wide range of torques, speeds and speed ratios.

10. The force required to hold the gears in position is much less than in an equivalent friction drive. This results in lower bearing pressure, less wear on the bearing surface and efficiency.

Disadvantages:

1. Special equipment and tools are required to manufacture the gears.
2. When one wheel gets damaged the whole set up is affected.
3. Noisy in operation at considerable speeds.

Definitions:

Refer to Fig. 7.23.

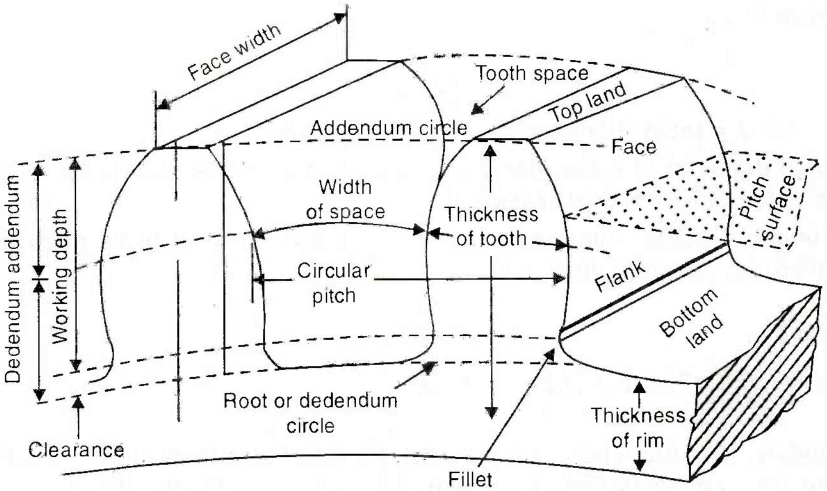


Fig. 7.23. Terms of gears.

1. ***Pitch circle.*** It is an imaginary circle which would transmit the same motion as the actual gear, by *pure rolling action.*

   The diameter of the pitch circle is known as *pitch circle diameter.*
2. ***Addendum circle.*** A circle concentric with the pitch circle and *bounding the outer ends* to the teeth is called an *addendum circle.*

   The diameter of the addendum circle is known as addendum circle diameter.
3. ***Addendum.*** It is the radial distance between the pitch circle and addendum circle.
4. ***Dedendum (Or root) circle.*** It is a circle concentric with the pitch circle and bounding the bottom of the tooth.
5. ***Dedendum.*** It is the radial distance between the pitch circle and the dedendum circle.
6. ***Clearance.*** The difference between the dedendum (of one gear) and addendum (of the mating gear) is called as *clearance.*
7. ***Working depth.*** It is the sum of the addenda of the two mating gears.
8. ***Circular thickness (or Thickness of tooth).*** The length of arc between the sides of a gear tooth, measured on the pitch circle is known as *circular thickness (or thickness of tooth).*
9. ***Tooth space.*** It is the width of the recess between two adjacent teeth measured along pitch circle.

10. ***Backlash.*** It is the difference between the tooth space and the tooth thickness.
11. ***Face.*** It is the action or working surface of the addendum.
12. ***Flank.*** The working face of the dedendum is called the *flank*.
13. ***Top land.*** It is the surface of the top of the tooth.
14. ***Bottom land.*** It is the surface of the bottom of the tooth space.
15. ***Whole depth.*** It is the total depth of the tooth space, equal to addendum plus dedendum; also it is equal to the working depth plus clearance.
16. ***Tooth fillet.*** It is the radius which connects the root circle to the tooth profile.
17. ***Circular pitch.*** The distance measured along the pitch circle from a point on one tooth to the corresponding point on an adjacent tooth is called *circular pitch*. It is represented by p_c.

$$p_c = \frac{\pi D}{T} \qquad \text{...(7.1)}$$

where D = pitch diameter, T = number of teeth.

18. ***Pitch diameter.*** It is the diameter of a pitch circle. It is usually represented by d_p or d_g for pinion and gear respectively.
19. ***Diametral pitch.*** Number of teeth on a wheel per unit of its pitch diameter is called the *diametral pitch*. It is denoted by p_d

$$\therefore \qquad p_d = \frac{T}{D} \qquad \text{...(7.2)}$$

From eqns. (7.1) and (7.2), we have

$$p_c \cdot p_d = \pi \qquad \text{...(7.3)}$$

20. ***Module.*** It is the *reverse of the diametral pitch*. Ratio between the pitch diameter and the number of teeth is known as *module*, it is denoted by m.

$$m = \frac{D}{T} \qquad \text{...(7.4)}$$

Types of gears :

The types of gear are discussed below :

1. Spur gear. A spur gear is a gear wheel or pinion for transmitting motion between two *parallel shafts*. This is the simplest form of geared drive. The teeth are cast or machined *parallel with the axis of rotation of the gear*. Normally the teeth are of involute form. Fig. 7.24 illustrates a spur gear drive, consisting of a pinion and a spur wheel.

The efficiency of power transmission by these gears is very high and may be as much as 99% in case of high-speed gears with good material and workmanship of construction and good lubrication in operation. Under average conditions, efficiency of 96–98% are commonly attainable. The *disadvantages* are that they are liable to be *more noisy* in operation and may *wear out* and develop *backlash* more readily than the other types.

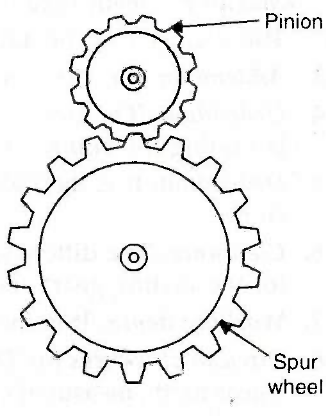


Fig. 7.24. Spur gear.

2. Helical gear. Refer to Fig. 7.25, helical gear is one in which teeth instead, of being parallel with shaft as in ordinary spur gears, are *inclined*. This ensures *smooth action* and *more accurate maintenance of velocity ratio.*

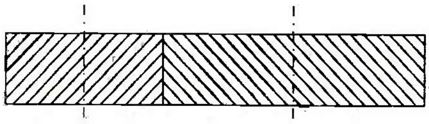


Fig. 7.25. Helical gear.

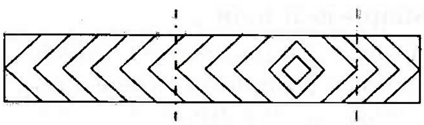


Fig. 7.26. Double helical gear.

A *disadvantage* is that the inclination of the teeth sets up a *lateral thrust*. A method of neutralising this lateral or axial thrust is to use *double-helical gears* (also known as *Herring bone gears*) shown in Fig. 7.26.

3. Bevel gear. Refer to Fig. 7.27. A bevel gear transmits motion between two shafts which *intersect*. If the shafts are at right angles and wheels equal in size, they are called *mitre gears;* if the shafts are not at right angles, they are sometimes called angle bevel gears. Spiral toothed bevel gears are preferred to straight-toothed bevels in certain applications, because they will run more smoothly and make less noise at high speeds.

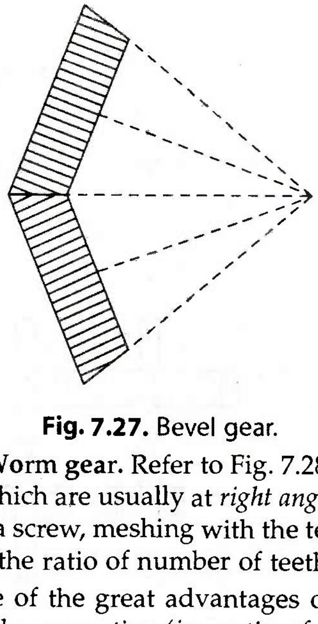


Fig. 7.27. Bevel gear.

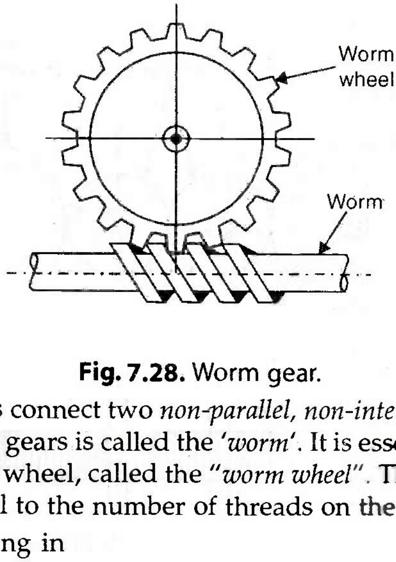


Fig. 7.28. Worm gear.

4. Worm gear. Refer to Fig. 7.28. Worm gears connect two *non-parallel, non-intersecting shafts* which are usually at *right angles*. One of the gears is called the *'worm'*. It is essentially part of a screw, meshing with the teeth on a gear wheel, called the *"worm wheel"*. The gear ratio is the ratio of number of teeth on the wheel to the number of threads on the worm.

One of the great advantages of worm gearing in that high gear ratios (*i.e.*, ratio of rotational speed of worm to that of worm wheel) are easily obtained. Worm gearing is *smooth and quiet*.

5. Rack and pinion. Refer to Fig. 7.29. A rack is a spur gear of infinite diameter, thus it assumes the shape of a straight gear. The rack is generally used with a pinion to *convert rotary motion into rectilinear motion.*

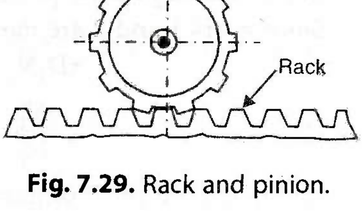


Fig. 7.29. Rack and pinion.

Types of gear trains :

The combination of gear wheels by means of which motion is transmitted from one shaft to another shaft is called a **gear train.**

The gear trains are of the following types:

1. Simple gear train
2. Compound gear train
3. Epicyclic gear train.

Simple gear train:

A *simple gear train* is one in which each shaft carries one wheel only (Fig. 7.30). Simple gear trains are employed where a *small velocity ratio* is required. The gear train in which the driving and the driven shafts are co-axial or coincident is known as the *reverted gear train* (Fig. 7.31).

Refer to Fig. 7.30. 1 is the driving wheel and 4 the driven wheel.

Let, N_1 = Speed of driver in r.p.m.,

N_2 = Speed of the idle gear 2 in r.p.m.,

N_3 = Speed of the idle gear 3 in r.p.m.,

N_4 = Speed of the driven (or follower) in r.p.m.

and T_1, T_2, T_3, and T_4 be the number of teeth on the gears 1, 2, 3 and 4 respectively.

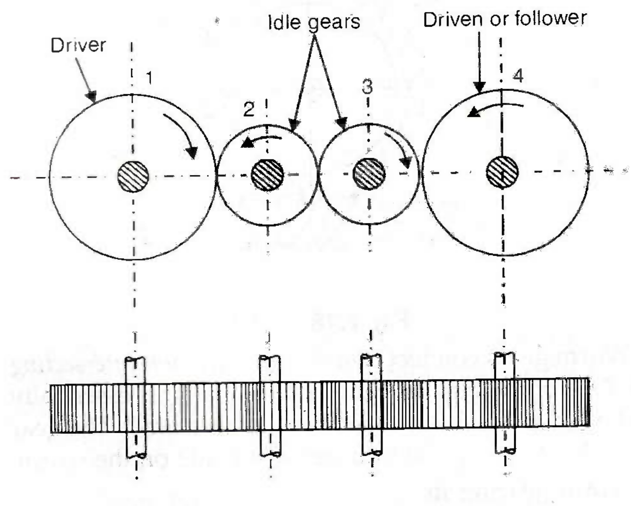


Fig. 7.30. Simple gear train.

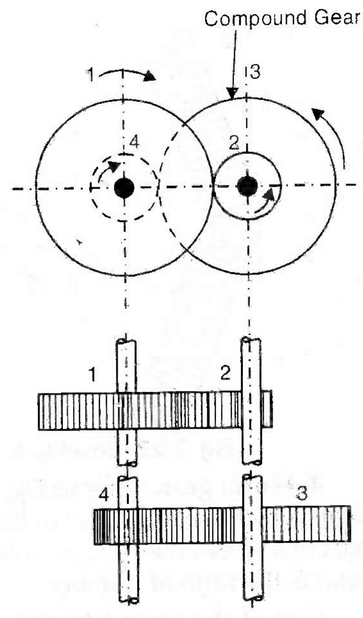


Fig. 7.31. Reverted gear train.

Let D_1 and D_2 be the pitch diameters of wheels 1 and 2.

Since gears 1 and 2 are meshing together, therefore,

$$\pi D_1 N_1 = \pi D_2 N_2$$

$$\frac{N_1}{N_2} = \frac{D_2}{D_1} \quad ...(i)$$

and diametral pitch of gear 1 = diametral pitch of gear 2.

$$\therefore \quad \frac{T_1}{D_1} = \frac{T_2}{D_2} \text{ or } \frac{T_2}{T_1} = \frac{D_2}{D_1} \quad ...(ii)$$

From eqns. (*i*) and (*ii*), we have

$$\frac{N_1}{N_2} = \frac{T_2}{T_1} \quad ...(1)$$

Similarly,

$$\frac{N_2}{N_3} = \frac{T_3}{T_2} \quad ...(2)$$

$$\frac{N_3}{N_4} = \frac{T_4}{T_3} \quad ...(3)$$

Multiplying eqns. (1), (2) and (3), we get

$$\frac{N_1}{N_2} \times \frac{N_2}{N_3} \times \frac{N_3}{N_4} = \frac{T_2}{T_1} \times \frac{T_3}{T_2} \times \frac{T_4}{T_3} \text{ or } \frac{N_1}{N_4} = \frac{T_4}{T_1} \quad ...(7.5)$$

$$\therefore \text{ Speed (or velocity) ratio} = \frac{\text{Speed of the driver}}{\text{Speed of the driven}} = \frac{\text{No. of teeth on driven}}{\text{No. of teeth on driver}}$$

Train value. It is the *reciprocal of velocity ratio*

$$= \frac{\text{Speed of driven}}{\text{Speed of driver}} = \frac{\text{No. of teeth on driver}}{\text{No. of teeth on driven}}$$

Similarly, it can be proved that the above equation holds good even if there are any number of intermediate gears. These intermediate gears are called *idle gears*, as they *do not affect the speed ratio or train value of the system.* In simple train of gears (as seen above) the speed ratio and train value is *independent of the size and number of intermediate/idle gears.*

The idle gears are provided for the following *purposes* :

1. To *bridge the distance* between the driving and driven wheels of moderate sizes instead of providing two wheels (driving and driven) of extra-ordinary big sizes.
2. To help achieving the *required direction of driven wheel.*

Fig. 7.31 shows a reverted gear train. *The reverted gear trains are used in automotive transmissions, lathe back gears, industrial speed reducers etc.*

Compound gear train:

A ***compound gear train*** is one in which *each shaft carries two wheels, one of which acts as the follower and the other acts as a driver to the other shaft* (Fig. 7.32). These gear trains are used for *high velocity ratio* and the same can be obtained with wheels of comparatively small diameter and, moreover, the driver can be had in smaller and limited space and if need arises, can be brought back so that the driving and driven wheels axes are coincident (*i.e.*, in one line). Usually for a *speed reduction in excess of 7 to 1* a compound train or worm gearing is employed (instead of a simple train).

Refer to Fig. 7.32.

The gear 1 is driving gear mounted on shaft L, gears 2 and 3 are compound gears which are mounted on shaft M. The gears 4 and 5 are also compound gear which are mounted on shaft P and gear 6 is the driven gear mounted on shaft Q.

Let,
N_1 = Speed of driving gear 1 in r.p.m.,
T_1 = No. of teeth on driving gear 1,
N_2, N_3, N_4, N_5, N_6 = Speed of respective gears in r.p.m., and
T_2, T_3, T_4, T_5, T_6 = No. of teeth on respective gears.

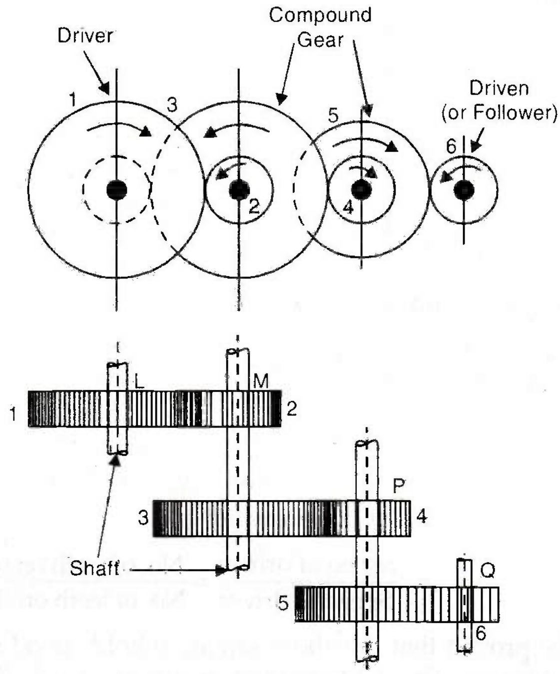


Fig. 7.32. Compound gear train.

Since gear 1 meshes with gear 2, therefore its speed (or velocity) ratio is

$$\frac{N_1}{N_2} = \frac{T_2}{T_1} \qquad ...(i)$$

Similarly, for gears 3 and 4, speed ratio is

$$\frac{N_3}{N_4} = \frac{T_4}{T_3} \qquad ...(ii)$$

and for gears 5 and 6, speed ratio is

$$\frac{N_5}{N_6} = \frac{T_6}{T_5} \qquad ...(iii)$$

The speed ratio of the compound gear train is obtained by multiplying eqns. (*i*), (*ii*), and (*iii*).

$$\therefore \qquad \frac{N_1}{N_2} \times \frac{N_3}{N_4} \times \frac{N_5}{N_6} = \frac{T_2}{T_1} \times \frac{T_4}{T_3} \times \frac{T_6}{T_5}$$

But, $N_2 = N_3$ ($\because$ gears 2 and 3 are mounted on shaft M)

$N_5 = N_4$ ($\because$ gears 5 and 4 are mounted on shaft P)

$$\therefore \qquad \frac{N_1}{N_6} = \frac{T_2}{T_1} \times \frac{T_4}{T_3} \times \frac{T_6}{T_5}$$

i.e., Speed (or velocity) ratio $= \dfrac{\text{Speed of the first driver}}{\text{Speed of the last driven or follower}}$

$= \dfrac{\text{Product or the number of teeth on drivens}}{\text{Product of the numbers teeth on the drivers}}$

$$\text{Train value} \left(= \frac{N_6}{N_1}\right) = \frac{\text{Speed of the last driven or follower}}{\text{Speed of the first driver}}$$

$$= \frac{\text{Product of the number of teeth on the drivers}}{\text{Product of the number of teeth on the drivens}}.$$

Example 7.1. *A toothed gear has 72 teeth and circular pitch of 26 mm, find the following:*

(i) Pitch diameter. *(ii) Diametral pitch.*

(iii) Module of the gear.

Solution. Number of teeth, $T = 72$

Circular pitch, $p_c = 26$ mm

(i) Pitch diameter, D :

$$p_c = \frac{\pi D}{T}$$

$$\therefore \quad 26 = \frac{\pi \times D}{72} \text{ or } \frac{26 \times 72}{\pi} = 595.87 \text{ mm (Ans.)}$$

(ii) Diametral pitch, p_d :

$$p_c \cdot p_d = \pi$$

$$p_d = \frac{\pi}{p_c} = \frac{\pi}{26} = 0.12 \text{ teeth/mm (Ans.)}$$

(iii) Module, m :

$$m = \frac{D}{T} = \frac{595.87}{72} = 8.27 \text{ mm/tooth (Ans.)}$$

Example 7.2. *Fig. 7.33 shows the gearing of a machine tool. The gear A is connected to the motor shaft which rotates at 1000 r.p.m. Find the speed of the gear F mounted on the output shaft L.*

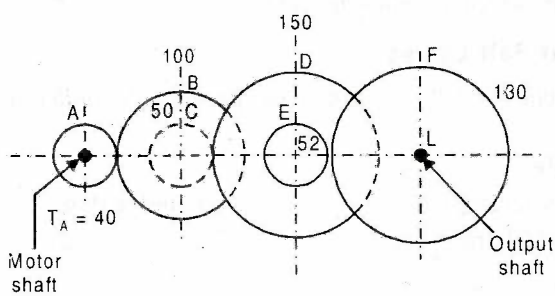


Fig. 7.33.

Solution. Number of teeth of gear A, $T_A = 40$

Number of teeth of gear B, $T_B = 100$

Number of teeth of gear C, $T_C = 50$

Number of teeth of gear D, $T_D = 150$

Number of teeth of gear E, $T_E = 52$

Number of teeth of gear F, $T_F = 130$

Speed of the motor shaft, $N_A = 1000$ r.p.m.

Speed of the output shaft, N_F :

We know that, $\frac{N_F}{N_A} = \frac{T_A \times T_C \times T_E}{T_B \times T_D \times T_F}$ or $\frac{N_F}{1000} = \frac{40 \times 50 \times 52}{100 \times 150 \times 130}$

$$N_F = 53.33 \text{ r.p.m. (Ans.)}$$

Epicyclic (or Planetary) gear train :

So far we have discussed those gear trains in which axes of the wheels remain *fixed* relative to one another. But there is another system of gear train in which there is *relative motion* between two or more of the axes of the wheels (constituting the train); such an arrangement of wheels is known as *"epicyclic gear train"*. The wheels are usually carried on an arm or link pivoted about a fixed centre and itself capable of rotating. For example, in Fig. 7.34 gear rolls around the outside of the stationary gear 2 as the arm *A* revolves. Epicyclic trains are sometimes called as *planetary gear trains* because of the fact that gear 1 goes round and round the gear 2 just like a planet moving round the sun. The motion of the planets around the sun is called planetary motion, so the motion of gear 1 around gear 2 is called planetary motion.

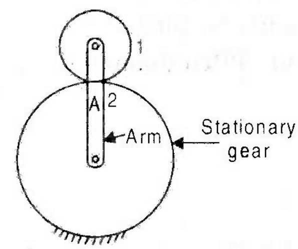


Fig. 7.34

Epicyclic gear trains are also *simple* as well as *compound* exactly in the same manner as explained earlier.

The following points are worth noting :

1. The *epicyclic gear trains are useful for transmitting very high velocity ratios, with gears of moderate size in a comparatively lesser space.*
2. These trains are of great practical importance and find use in almost all kinds of workshop and electrical machines; *e.g. back gear of lathe, differential gears and gear boxes for motor vehicles, cyclometers etc.*

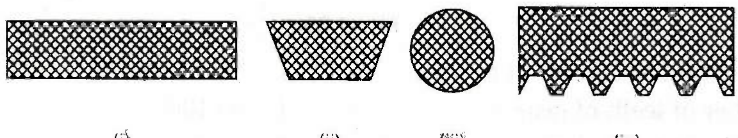
7.2.10. Belts and Belt Drives

A ***belt*** *is a continuous band of flexible material passing over pulleys to transmit motion from one shaft to another.*

Belts are available :

(*i*) with a narrow *rectangular cross-section*—**Flat belts** [Fig. 7.35(*i*)].

(*ii*) with a *trapezoidal cross-section*—**V-belts** [Fig. 7.35(*ii*)] and multiple **V-belts** [Fig. 7.35(*iv*)].



(i) (ii) (iii) (iv)

Fig. 7.35

(*iii*) *Round cross-section*—**Round belts** [Fig. 7.35(*iii*)].

Chiefly used in machinery are *flat and V-belts.*

Flat belts :

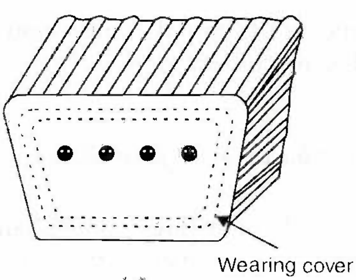
- Flat belts are used for their simplicity and because they are subjected to minimum bending stress on the pulleys. The load capacity of flat belts is varied by varying their width, and only one is used in each drive. They are made of *leather, rubber, textile, balata and steel.*

- *Leather belts* have the best *pulling capacity*. Because of high cost of leather they are used very rarely.
- *Rubber belts* made of rubber on a cotton-duck base are used where the belt is exposed to the weather or steam, as they do not absorb moisture so readily as leather. They get destroyed if kept in contact with oil or grease.
- *Textile belts* are made of cotton and are used for rough and short service.
- *Balata belts* are acid and water proof and cannot withstand temperature higher than 100°C.
- *Steel belts* are claimed to transmit more horse power per cm width, and to remain unaffected by dampness or heat and be immune from stretching and slipping. The pulleys on which they are mounted *do not have camber*. Steel belts are sometimes used, the belt being subjected to considerable initial tension, to maintain the pressure on the pulley, on which the friction depends.

Note: The pulley of the flat belts is made *convex* at centre. This feature of the pulley is called **camber** or *crown* and due to it the lateral displacement of belt is prevented.

V-belts :

A V-belt is a belt of trapezoidal section running on pulleys with grooves cut to match the belt. The normal angle between the sides of the groove is 40 deg. Fig. 7.36(*a, b*).



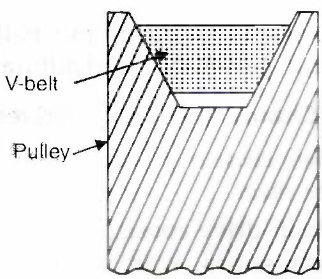


Fig. 7.36. V-belt.

- V-belts are usually made of fabric coated with rubber. They are silent and resilient. They are used when the distance between the shafts is too short for flat-belt drives. Owing to the *wedge action* between the belt and the sides of the groove in the pulley, the V-belt is *less likely to slip,* hence *more power can be transmitted for the same belt tension.*

Round belts:

Round belts are employed to *transmit low power, mainly in instruments, table-type machine tools, machinery of the clothing industry and household appliances.* Round belts are used singly, as a rule. They may be made of *leather, canvas* and *rubber.* The diameter range is from 3 to 12 mm, usually from 4 to 8 mm. The minimum allowable ratio of the diameter of smaller pulley to the belt diameter is about 20, the recommended ratio is 30.

Belt drive :

A belt drive consists of the driving and driven pulleys and the belt which is mounted on the pulleys with a certain amount of tension and transmits peripheral force by friction.

- Belt drives may be:

(*i*) Open belt drive

(*ii*) Crossed belt drive.

Open belt drives [Fig. 7.37(*a*)] are applied, as a rule, between parallel shafts which *rotate in the same direction.* Here the belt is subject to tension and bending.

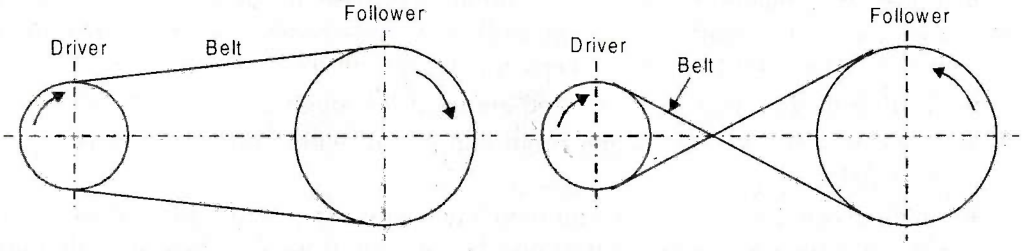


Fig. 7.37. Belt drives.

In *crossed belt drives* [Fig. 7.37(*b*)] the power is transmitted between small shafts *rotating opposite direction.* Since *the angle of contact in this type of drive is more,* it can transmit *more power than open belt drive.* However there is *more wear and tear of the belt in this drive.*

Applications of belt drives :

The main applications of belt drives are:

(*i*) To transmit power from low or medium capacity electric motors to operative machines.

(*ii*) To transmit power from small prime movers (internal combustion engines) to electric generators, agricultural and other machinery.

9.2.11. Chains and Chain Drives

- A *chain consists of links connected by joints which provide for articulation or flexibility of the chain.*
- A chain drive consists of two sprockets and chain (Fig. 7.38). Chain drives, or transmissions, with several driven sprockets are also employed. Besides the enumerated components, chain drives may also include tensioning devices, lubricating devices and guards.

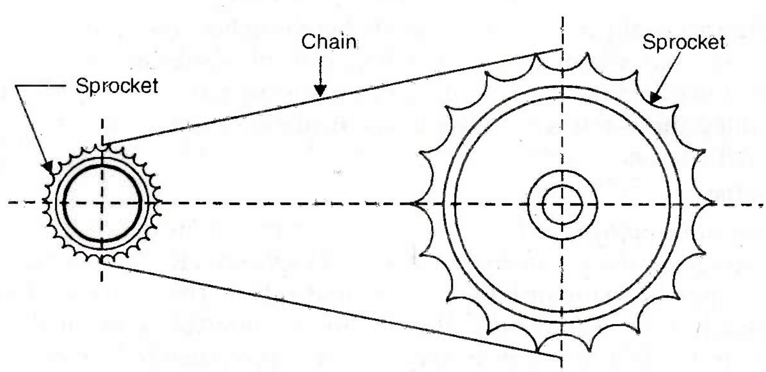


Fig. 7.38. Chain drive.

Chains drives are used for:

(1) *Medium centre to centre distances* which, in the case of a gear drive, would require idle gears, or intermediate stages not necessary to obtain the required speed ratio.

(2) Drives with strict requirements as to overall size or ones *requiring positive transmission without slippage* (preventing the use of V-belts drives).

There are two principal types of chain drives:

(*i*) Roller chain drive, and

(*ii*) Inverted tooth or silent chain drive.

7.2.12. Bearings

Introduction

- A **bearing** *is a device which supports, guides and restrains moving elements.*
- The material used for bearing is commonly *cast-iron for slow speeds, bronze or brass lining being fitted for higher speeds.*

White metal or antifrication metal is used as a lining for the bronze, or it may be held directly in the cast-iron or in the steel of a connecting rod. The value of soft metals such as these is that *they do not roughen the journal, and they are able to flow slightly under pressure if insufficient clearance has been allowed or if the shaft is very slightly out of line.*

7.2.12.1. Classification of Bearings

Bearings may be *classified* as follows :

1. ***Plain bearings :***

(*a*) Journal bearing. (*b*) Pivot bearing.

(*c*) Collar or thrust bearing.

2. ***Ball and roller bearings.***

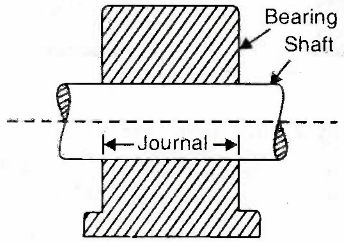


Fig. 7.39. Journal bearing.

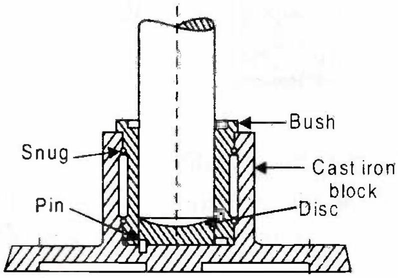


Fig. 7.40. Pivot bearing.

1. **Plain bearings.**

- A ***journal bearing*** (Fig. 7.39) is one in which the bearing pressure is *perpendicular to the axis of the shaft*. The portion of the rotating element which is in contact with the bearing is called *journal.*
- A ***pivot bearing*** is one in which the pressure is *parallel to the axis of the shaft* (Fig. 7.40) and the end of the shaft rests on the bearing surface.
- In ***collar bearing*** (Fig. 7.41) *the pressure is parallel to the axis of the shaft, which is passed and extended through the bearings.*

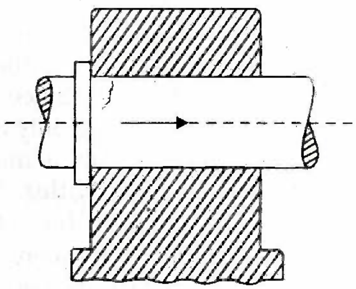


Fig. 7.41. Collar or thrust bearing.

- These bearings are employed to *take up unbalanced axial loads on the horizontal shaft.* If the load is light, a single collar thrust bearing may serve the purpose but in case of large loads the use of multiple collar bearings is restored to.

2. ***Ball and roller bearings :*** Refer to Figs. 7.42 and 7.43

- Ball and roller bearings are also known as *rolling contact bearings or rolling-element bearings* because the bearing elements especially are in a rolling contact. Sometimes these are also referred to as *"antifrication bearings"*, through some friction is always present owing to rolling resistance between the balls/rollers and the races, retainers and contacting parts etc. The starting friction in ball and roller bearings is lower than that in an equivalent journal bearing in which metal-to-metal rubbing takes place at the time of starting. *The ball and roller bearings are also quite suitable at moderate speeds but at high speeds it is found that a properly designed and lubricated journal bearing has less friction.*

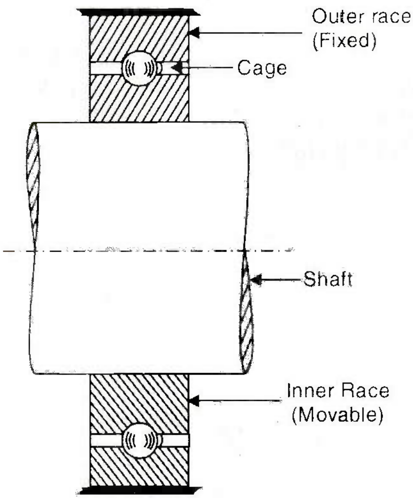


Fig. 7.42. Ball bearing.

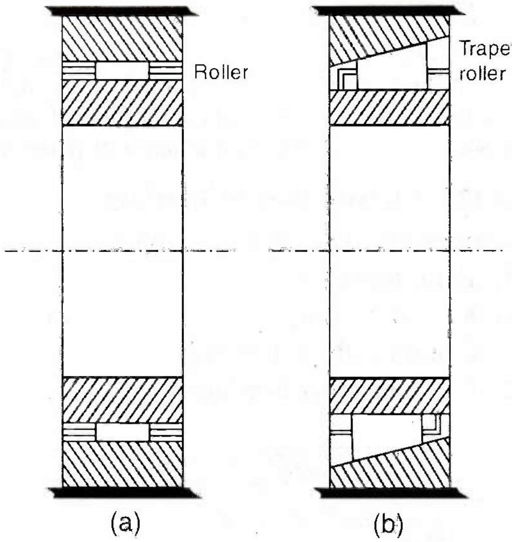


Fig. 7.43. Roller bearings.

- The friction-speed relationship for various cases is shown in Fig. 7.44. It may be noted that *in the case of ball and roller bearings, the coefficient of friction varies little with load and speed, except at extreme values; this property makes the ball and roller bearings extremely suitable for machines that are started and stopped frequently, especially under load.* Since in case of ball bearing only a kinematic *point contact* is made, and in case of roller bearing a kinematic *line contact,* the latter is frequently used for large bearing loads. However, *in case of roller bearings the*

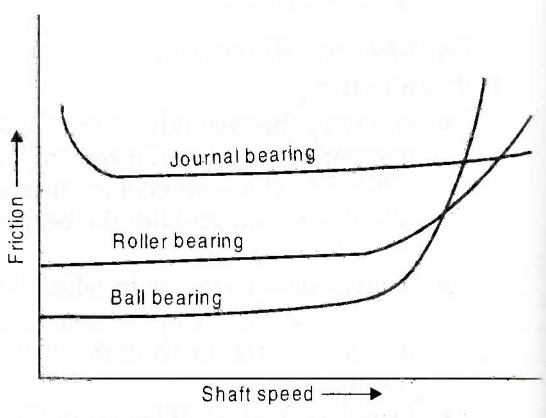


Fig. 7.44. Friction-speed relationship.

inherent disadvantage is the variation of pressure along the band of contact, due to deflection of shaft and mountings.

Uses of bearings :

The uses of bearings in *electrical equipment* are given in tabular form below :

S. No.	Equipment	Bearings
1.	High H.P. motors, generators or alternators whose shafts are horizontal and have no *thrust* (end pressure)	*Journal bearing*
2.	High H.P. electrical machines with horizontal shafts and having end *thrust.*	Thrust or roller bearings
3.	Turbogenerator sets with vertical shafts.	*Foot step or pivot bearings*
4.	Table fans.	*Ball bearings (Radial type)*
5.	Ceiling fans.	*Ball bearings (Thrust type)*
6.	Medium H.P. motors or generators (shafts with horizontal axis without end thrust).	*Roller bearings (Radial type)*
7.	Medium H.P. motors or generators (shafts with vertical axis).	*Roller bearing (Thrust type)*
8.	Medium H.P. motors and generators (shafts with horizontal axis and having end thrust or shafts placed in an inclined position).	*Roller bearing (Tapered Type)*

7.3 ELECTRICAL ACTUATORS

7.3.1. General Aspects

Actuator: *A mechanical device or a system which has motion or movement is called an actuator.*

Actuation system: *A group of elements which is responsible directly or indirectly for imparting motion to an actuator is called an* ***actuation system.***

Electrical actuator: *An actuator receiving electrical energy for motion is called an* ***electrical actuators.***

Electrical actuators systems include the following:

I. Switching devices:

1. Mechanical switches:
   - Solenoids.
   - Relays.
2. Solid state switches:
   - Diodes.
   - Tyristors.
   - Transistors.

Here the *control signal switches on or off some electrical device, perhaps a heater or motor.*

II. Drive systems:

1. D.C. motors.
2. A.C. motors.

7.3.2. Switching Devices

1. Mechanical switches :

Mechanical switches are elements which are often used as sensors to give input to systems *e.g., keyboards.* Here we are concerned with their use as actuators to perhaps switch on *electric motors* or *heating elements* or switch on the current to actuate *solenoid valves controlling hydraulic or pneumatic cylinders.*

***Mechanical switches** are those where in switching action is by the application of force on the switch and during switching action mechanical elements move with the switch.* These switches consists of one or more pair of contacts which are mechanically closed or opened and in doing so make or break electrical circuit.

- Mechanical switches are specified in terms of *number of poles and throws.*
  - — ***Poles (P)** are number of separate circuits that can be completed by the same switching action.*
  - — ***Throws (T)** are number of individual contacts for each pole.*
- There are many designs for limit *"switches"* including push-button and levered microswitches. All switches are used to open or close connections within circuits. As illustrated in Fig. 7.45, switches are characterised by the number of poles throws *and whether connections are "normally open (NO) or "normally closed (NC)".*

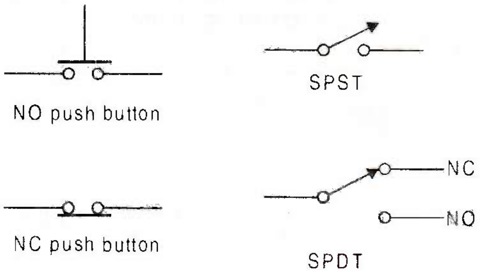


Fig. 7.45. Switches.

— The SPST switch is a single pole (SP), single throw (ST) device that opens or closes a single connection.

— The SPDT switch changes the pole between two different throw positions.

There are many variations on the pole and throw configurations of switches, but their function is easily understood from the basic terminology.

Bouncing and debouncing :

*When mechanically switches are opened or closed, there are brief current oscillations due to mechanical bouncing or electric arcing; this phenonon is called **switch bounce.***

Fig. 7.46, illustrates that the mechanical contact associated with a switch closing results in multiple voltage transitions over a short period of time.

— Bouncing can occur when the switch is opened.

— Generally the bouncing time is about 20 ms.

The problem of bouncing can be solved by using the following *methods* :

(*a*) Specially designed switches.

(*b*) Software solution.

(*c*) Hardware solution.

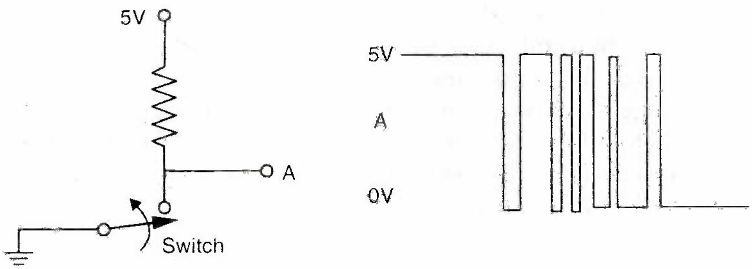


Fig. 7.46

(*a*) ***Specially designed switches:*** Specially designed switches include the following :

- Keys of a keyboard (toggle switch);
- Membrane switch;
- The keys used in calculators, mobile phones and telephones.

(*b*) ***Software solution:*** In this method, the microprocessor is programmed with a software to detect that the switch is closed and then wait for the bouncing period (say 20 ms). After checking that bouncing has ceased, the switch being in the closed position processing of next instruction can take place.

(*c*) ***Hardware solution:*** The hardware solutions to the bouncing problem are :

- *Set reset flip-flop circuit (also called latch circuit);*
- D flip-plop circuits
- Schmitt trigger.

Fig. 7.47, shows the sequential logic circuit which can provide an output that is free from multiple transitions associated with switch bounce.

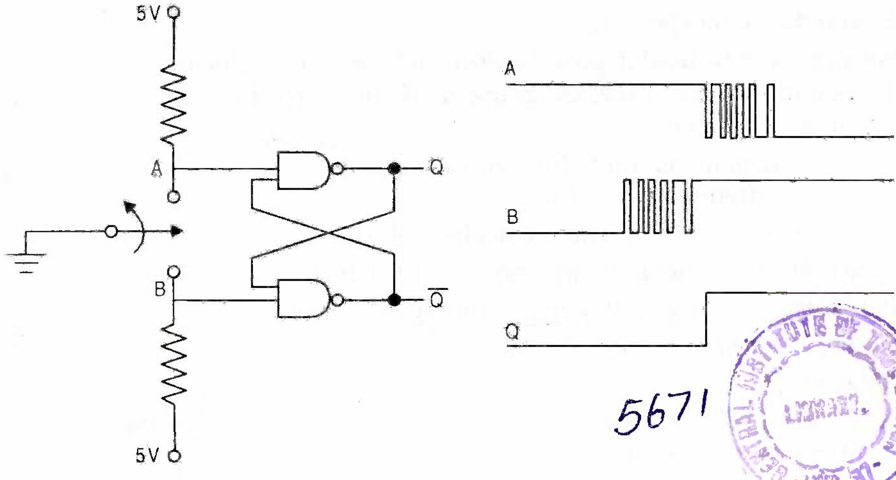


Fig. 7.47. Switch debounce circuits.

— As the switch breaks contact with *B*, single bounce occurs on the *B* line. There is a small delay as the switch moves from contact *B* to *A*, and then single bounce occurs on the *A* line as contact is established with *A*. The output of the debouncer *Q* is a single transition from 0 V to 5 V.

— The circuit functions very much like a flip-flop.

(*i*) **Solenoids :** Refer to Fig. 7.48.

A *"solenoid"* consists of a coil and a movable iron core called the *armature.* When the current is passed through the coil it gets energized and consequently the core moves to increase the flux linkage by closing the air gap between the cores. The movable core is usually spring-loaded to allow the core to retract when the current is switched off. *The force generated is approximately proportional to the square of the current and inversely proportion to the square of the width of the air gap.*

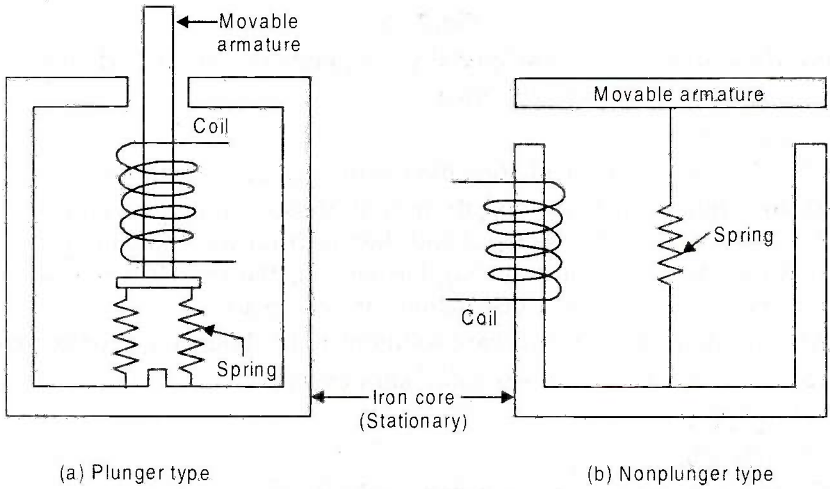


Fig. 7.48. Solenoids.

- Solenoids are inexpensive.
- Solenoids can be used to provide electrically operated actuators. *Solenoid valves* are an example of such devices, being used to control fluid flow in hydraulic or pneumatic systems.

The use of solenoids is limited to on-off applications such as *latching, locking,* and *triggering.* They are frequently used in:

— *Home appliances* (*e.g.*, washing machine valves).

— *Automobiles* (*e.g.*, door latches and starter solenoid)

— *Pinball machines* (*e.g.*, plungers and bumpers).

— *Factory automation.*

(*ii*) **Relays :**

Relays *are electrically operated switches in which changing current in one electrical circuit switches a current on or off in another circuit.*

Relays are often used in control systems; the output from the controller is a relatively small current and a much larger current is needed to switch on or off the final connection element, *e.g.*, the current required by an electric heater in a temperature control system or a motor.

Relays are used in *'power switches'* and *'electromechanical control elements'.*

- *A relay performs a function similar to a power transistor switch circuit but has the capability to switch much larger currents.*

The input circuit of a relay is *electrically isolated* form the output circuit, unlike the common-emitter transistor circuit, where there is a common ground between the input and output. Since the relay is electrically isolated, noise, induced voltages, and ground faults occurring in the output circuit have *minimal impact* on the input circuit.

- The *disadvantage* of the relays is that they have *slower switching times than transistors.*

2. Solid state switches:

Following are the solid-state devices which can be used to electronically switch circuits:

(*i*) Diodes (*ii*) Thyristors and triacs
(*iii*) Bipolar transistors (BPT) (*iv*) Power, MOSFETs.

(*i*) Diodes:

— A diode can be regarded as a 'directional element', only passing a current when forward biased (*i.e.*, with the anode being positive with respect to the cathode). If diode is sufficiently reverse biased, it will breakdown.

— If an alternating voltage is applied across a diode, *it can be regarded as only switching on when the direction of the voltage is such as to forward bias it and being off in the reverse biased direction.*

- Generally diodes are not used as switches, but are used as *rectifiers*. It can also be used for full-wave rectification by forming a bridge using diodes.

(*ii*) Thyristors and triacs:

Thyristors: The thyristor, or silicon-controlled rectifier (SCR) can be regarded as a diode which has a gate controlling the condition under which the diode can be switched on.

- The power-handling capability of a thyristors is high and thus it is widely used for *switching high power applications.*

Thyristors are employed in:

— D.C. controls;

— Electric heaters;

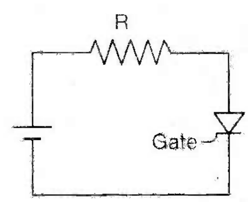
— Electric motors;

— Lamp dimmers etc.

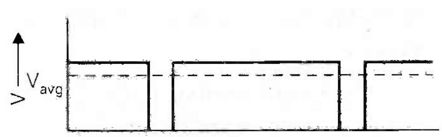
(*a*) *D.C. control :* Fig. 7.49 shows the thyristor D.C. control circuit and output :

— The thyristor is used as a switch to *on* and *off* the device.

— An intermittent voltage is generated by chopping of the supply voltage using an alternate signal to the gate. The average value of the D.C. voltage can thus be varied and hence controlled by the alternating signal at the gate.



(a) D. C. control circuit



(b) Output

Fig. 7.49. Thyristor application.

(*b*) *Thyristor application in lamp dimmer (A.C. circuit) :*

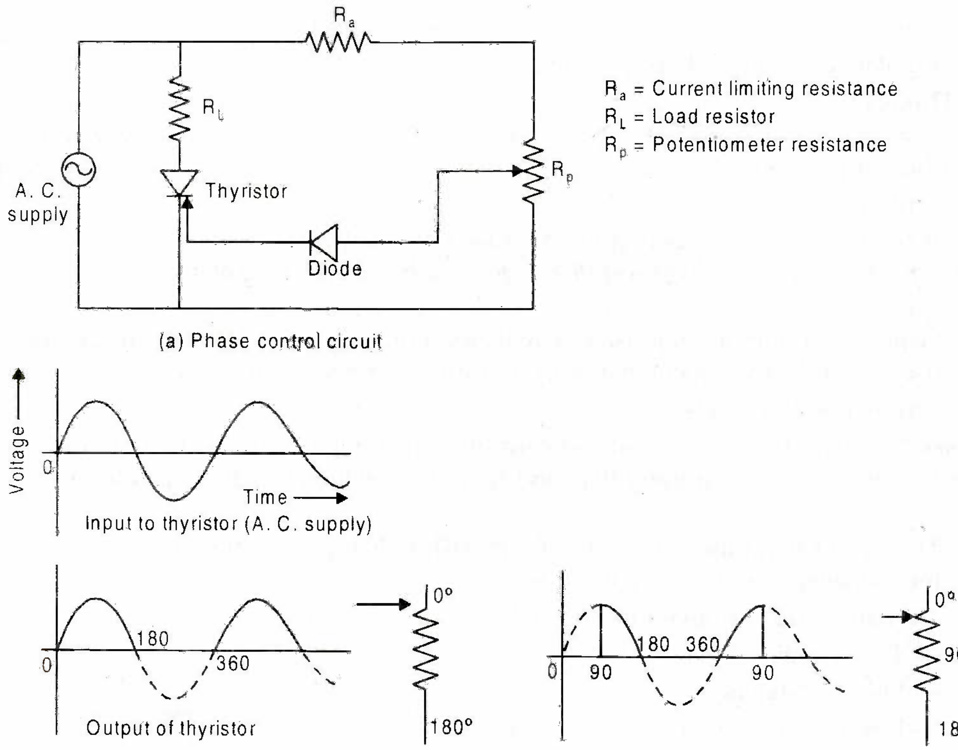
Fig. 7.50, shows the circuit using thyristor for a lamp dimmer (also called phase control circuit) and output of the thyristor.

— A.C. supply is applied across R_L (may be a lamp or an electric heater) in series with a thyristor. R_p is the potentiometer (resistance) which sets the voltage at which the thyristor is triggered.

— Diode prevents the negative part of A.C. supplied to the gate.

— By adjusting the triggering voltage to the thyristor (using potentiometer), it can be made to trigger at any point between 0° and 90° in the +ve half cycle of A.C.

— When the thyristor is triggered at the beginning of the cycle *i.e.,* 0° full power supply is applied to the load and by varying the triggering voltage the supply to the load can be varied.



(a) Phase control circuit

(b) Input to and output of thyristor

Fig. 7.50. Thyristor application in lamp dimmer.

Thus *thyristor can be employed to control the A.C. supply to several devices.*

Triac :

— The triac is similar to the thyristor and is equivalent to a pair of thyristors connected in reverse parallel on the same chip.

— The triac can be turned on in either forward or reverse direction.

— Triacs are simple, relatively inexpensive, methods of controlling A.C. power.

***(iii)* Bipolar transistor (BPT):**

Bipolar transistors come in two forms, the NPN and the PNP. For the NPN transistor, the main current flows in at the collector and out at the emitter, a controlling signal being applied to the base. The PNP transistor has the main current flowing in at the emitter and out at the collector, a controlling signal being applied to the base.

- In using transistor switched actuators with a microprocessor, attention has to be given to the size of the base current required and its direction. The base current required can be too high and so a buffer might be used. The buffer increases the drive current to the required value. It might also be used to invert.
- *Bipolar transistor switching is implemented by base currents and higher frequencies of switching are possible than with thyristors. The power handling capacity is less than that of thyristors.*

(*iv*) MOSFETs:

MOSFETs (Metal-oxide field-effect transistors) are available in two types, the N-channel and the P-channel.

The main difference between the use of a MOSFET for switching and a bipolar transistor is that no current flows into the gate to exercise the control. The gate voltage is the controlling signal. Thus drive circuitry can be simplified in that there is no need to be concerned about the size of the current.

With MOSFETs, very high frequency switching is possible, up to 1 MHz, and interfacing with a microprocessor is simpler than with bipolar transistors.

Control of D.C. motor using MOSFET :

MOSFET can be employed as a control switch for a D.C. motor as *on-off* switch. Here, as compared to BJT for D.C. motor control, a level shifter buffer is used to raise the voltage level to that required for the MOSFET.

Figure 7.51 shows the circuit diagram for D.C. motor control using MOSFET.

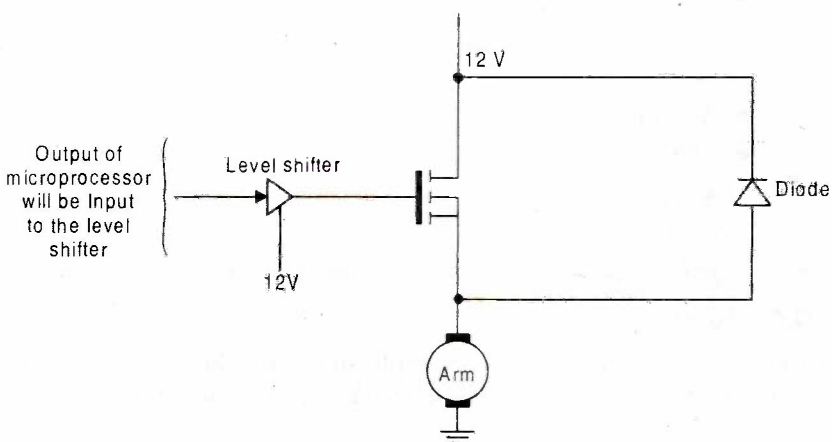


Fig. 7.51. MOSFET (N-channel type) application in D.C. motor control.

7.3.3. Drive Systems—Electric Motors

Electric motors are frequently used as the final control element in positional or speed control systems.

Electric motors for mechatronics applications, can be *classified* by electrical configuration as follows :

1. D.C. Motors:

(*i*) *Permanent magnet.*

(*ii*) *Series wound.*

(*iii*) *Shunt wound.*

(*iv*) *Compound wound.*

2. **A.C. Motors**

(*i*) *Single phase:*

(*a*) Induction:

- Squirrel cage:
  - — Split phase
  - — Capacitor start
  - — Permanent split capacitor
  - — Shaded pole
  - — Two-valve capacitor.
- Wound rotor:
  - — Repulsion
  - — Repulsion start
  - — Repulsion induction.

(*b*) Synchronous:

- Shaded pole
- Hysteresis
- Reluctance
- Permanent magnet.

(*ii*) *Polyphase:*

(*a*) Induction:

- Wound rotor
- Squirrel cage.

(*b*) Synchronous.

(*iii*) Universal motors.

- *In modern control systems* **D.C. motors** *are mostly used.*

7.3.4. D.C. Motors

D.C. (Direct current) motors find wide applications in a large number of mechatronic designs because of the *torque-speed characteristics* achievable with different electrical configurations.

— The speeds of the D.C. motors can be *smoothly controlled* and in most cases are *reversible.*

— These motors *can respond quickly* since they have a *high ratio of torque to rotor-inertia.*

— *'Dynamic braking'* (where motor generated energy is fed to a resistor dissipater) and *'regenerative braking'* (where motor-generated energy is fed back to the D.C. power supply) can be implemented in applications where quick stops and high efficiency are desired.

7.3.4.1. Permanent magnet (PM) D.C. motors

In these motors (See Fig. 7.52) field excitation is obtained by suitably mounting permanent magnets (which require no power source and therefore produce no heating) on the stator. Magnets made from ferrites or rare earth (cobalt samarium) are used.

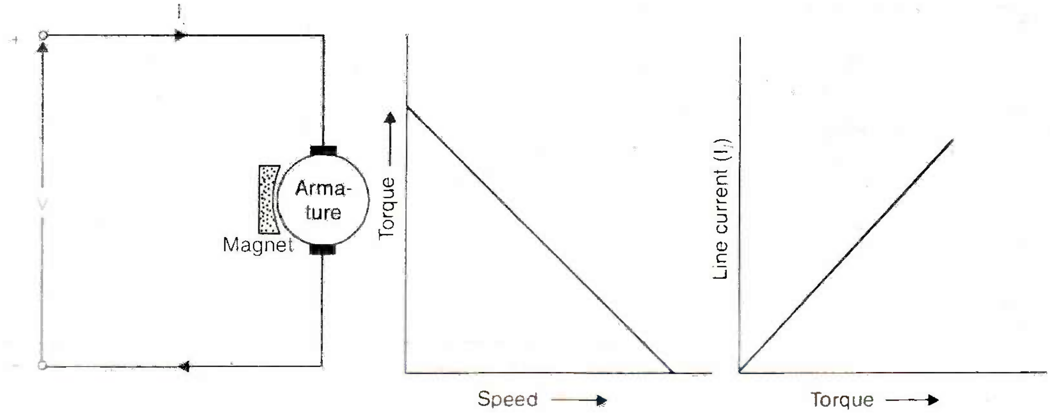


Fig. 7.52. Permanent magnet D.C. motor schematic and torque-speed and current-torque curves.

— A PM motor is *lighter and smaller* than others, equivalent D.C. motors because the field strength of permanent magnets is high.

— The radial width of a typical permanent magnet is roughly one-fourth that of an equivalent field winding.

— *PM motors are easily reversed by switching the direction of the applied voltage, because the current and field change direction only in the rotor.*

— PM motors can be brushed, brushless, or stepper motors.

Applications:

- The PM motor is *ideal in computer control applications* because of the *linearity* of its torque-speed relation.

— When a motor is used in a position or speed control application with sensor feedback to a controller, it is referred to as *"servomotor"*.

- PM motors are used only in *low-power applications* since their rated power is limited to 5 H.P. or less, with fractional horsepower ratings being more common.

Advantages:

As compared to field wound motors, these motors possess the following *advantages*:

(*i*) More efficient.

(*ii*) More reliable.

(*iii*) More study and compact.

(*iv*) The field flux remains constant for all loads giving a more linear speed-torque characteristic.

(*v*) In a separately excited motor, failure of field can lead to runaway condition. This does not happen in PM motors.

Limitation. As the *flux is constant* in these motors, *speed cannot be controlled above base speed.*

7.3.4.2. D.C. Shunt motors:

Refer to Fig. 7.53.

In these motors armature and field windings are connected in *parallel* and powered by the same supply.

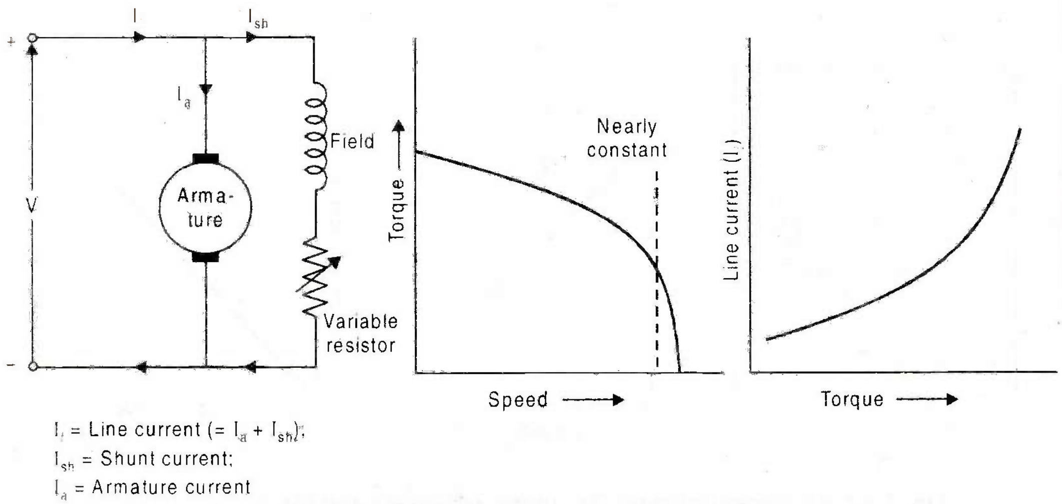


Fig. 7.53. D.C. shunt motor schematic and torque-speed and current-torque curves.

- These motors exhibit nearly *constant speed* over a long range of loading.
- They have starting torques about *1.5 times* the rated operating torque.
- They have *lowest starting torque* of any of D.C. motors.
- They can be economically converted to allow adjustable speed by placing a potentiometer in series with the field windings.

7.3.4.3. D.C. Series motors

In this type of motor (See Fig. 7.54) armature and field windings are connected in *series* so the armature and field currents are *equal.*

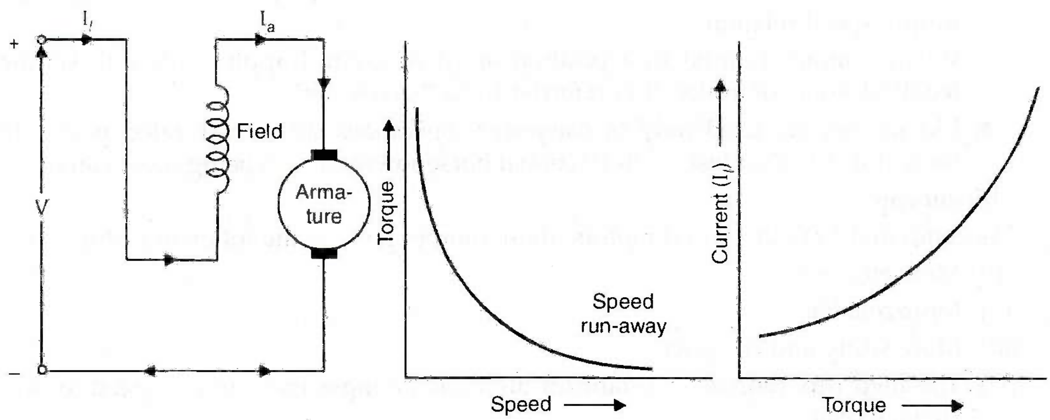


Fig. 7.54. D.C. series motor schematic and torque-speed and current-torque curves.

- *These motors exhibit very high starting torques, highly variable speed depending on load, and very high speed when the load is small.*
  — In fact, large series motors can fail catastrophically when they are suddenly unloaded (*e.g.*, in a belt drive application when the belt fails) due to dynamic forces at high speeds; this is called ***"run-away"***. As long as the motor remains loaded, this poses no problem.

- The torque-speed curve for a series motor is *hyperbolic* in shape, implying an inverse relationship between the torque and speed and *nearly constant power over a wide range.*

7.3.4.4. D.C. Compound motors

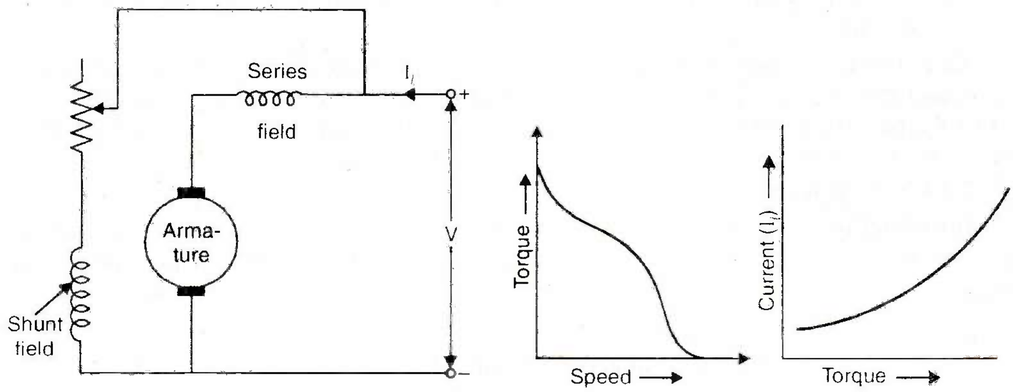


Fig. 7.55. D.C. compound motor schematic and torque-speed and current-torque curves.

Refer to Fig. 7.55. The compound motor has a shunt field winding in addition to the series winding so that the number of magnetic lines of force produced by each of its poles is the resultant of the flux produced by the shunt coil and that due to the series coil. The *flux so produced depends not only on the current and number of turns of each coil, but also on the winding direction of the shunt coil in relation to that of the series coil.* When the two fluxes *assist each other* the machine is a ***cumulative compound motor***, while if they *oppose each other*, it is said to be ***differential compound motor***.

Fig. 7.56 shows the field windings and interpole connections of a *differential compound wound motor.* The shunt coil is made up of *many turns of fine wire*, whilst the series coil comprises relatively *few turns of thickwire.*

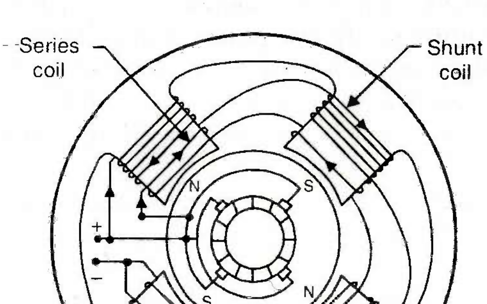


Fig. 7.56. Field windings of a differential compound motor.

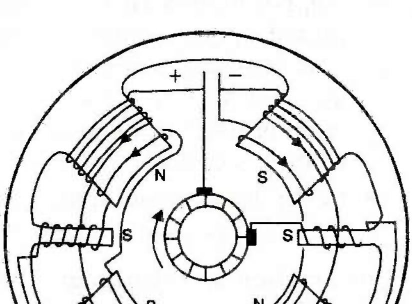


Fig. 7.57. Field windings and interpole connections of a cumulative compound wound motor.

Fig. 7.57, shows the field windings of a *cumulative compound motor.* The flow of currents in the shunt and series coil is worth noting in Fig. 7.56 and Fig. 7.57.

- *The maximum speed of compound motor is limited, unlike a series motor, but-its speed regulation is not as good as with a shunt motor.*
- *The torque produced by compound motors is somewhat lower than that of series motors of similar size.*

Note: Unlike the permanent magnet motor, when voltage polarity for a shunt, series, or compound D.C. motor is changed, the direction of rotation does not change. The reason for this is that the polarity of both the stator and rotor changes, because the field and armature windings are excited by the same source.

7.3.4.5. Stepper motors

Introduction: A ***stepper motor****, a special type of D.C. motor, is an incremental motion machine.* It is a *permanent magnet or variable reluctance D.C. motor* and has the following *characteristics:*

(*i*) It can rotate in both directions.

(*ii*) It can move in precise angular increments.

(*iii*) It can sustain a holding torque at zero speed.

(*iv*) It can be controlled with digital circuits.

— A stepper motor moves in accurate equal angular increments, known as ***steps***, in response to the application of digital pulses to an electric drive circuit. The number and rate of the pulses control the position and speed of the motor shaft.

— Generally, stepper motors are manufactured with steps per revolution of 12, 24, 72, 144, 180, and 200, resulting shaft increments of 30°, 15°, 5°, 2.5°, 2°, and 1.8° per step. Special ***micro-stepping*** circuitry can be designed to allow many more steps per revolution, often 10,000 steps/revolution or more.

— The stepper motor is used in digitally controlled position control system in open loop mode. The input command is in the form of a train of pulses to turn a shaft through a specified angle.

— Stepper motors are either ***bipolar***, requiring two power sources or a switchable polarity power source, or ***unipolar***, requiring only one power source. They are *powered by D.C. sources and require digital circuitry to produce coil energising sequences for rotation of the motor.* Feedback is *not* always required for control, but the use of an encoder or other position sensor can ensure accuracy when exact position control is critical.

- Generally, stepper motors produce *less than 1 H.P.* and are therefore *used only in low-power position control applications.*

Construction and working:

- A stepper motor consists of a *slotted stator having multi-pole, multi-phase winding and a rotor structure carrying no winding.* They typically use three and four phase windings, the number of poles depends upon the required angular change per input pulse.
- The rotors may be of the *permanent magnet or variable reluctance type.*
- Stepper motors operate with an *external drive logic circuit.* When a train of pulse is applied to the input of the drive circuit, the circuit supplies currents to the stator windings of the motor to make the axis of the air-gap field around in coincidence with the input pulses. The rotor follows the axis of the air-gap magnetic

field by virtue of the permanent magnet torque and/or the reluctance torque, depending upon the pulse rate and load torque (including inertia effects).

1. Permanent magnet stepper motor :

In the case of a permanent magnet stepper motor, the stator consists of wound poles, the rotor poles are permanent magnets.

Fig. 7.58, shows the phases or stacks of a 2-phase, 4-pole permanent magnet stepper motor.

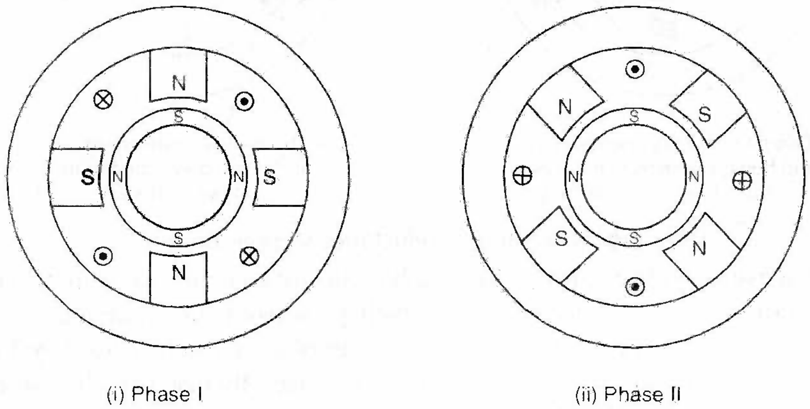


Fig. 7.58. Permanent magnet stepper motor.

— The rotor is made of *ferrite or rate-earth material* which is *permanently magnetised.*

— The stator stack of phase II is staggered from that of phase I by an angle of 90°.

— When the phase 'I' is excited, the rotor is aligned as shown in Fig. 7.58(*i*), If now the phase 'II' is also excited, the effective stator poles shift anti-clockwise by 22.5° [Fig. 7.58(*ii*)] causing the rotor to move accordingly. Now, keeping the phase 'II' still energised, if the phase 'I' is now de-energised, the rotor will move another step of 22.5°. The reversal of phase 'I' winding current will produce a further forward movement of 22.5°, and so on.

It can be easily observed/visualised as to how the direction of movement can be reversed.

— Each phase is provided with double coils to simplify the switching arrangement (which is electronically accomplished).

— This type of motor has the *advantage of small residual holding torque, called* ***detent torque****, even when stator is not energized.*

2. Variable reluctance stepper motor :

— A variable-reluctance stepper motor has no permanent magnet on the rotor and the rotor employed is a ferro-magnetic multi-toothed one.

— *The large differences in magnetic reluctances that exist between the direct and quadrature axes develop the torque.* The stationary field developed by the direct current in some stator coils tends to develop a torque which causes the rotor to move to the position where the reluctance of the flux path is minimum.

Fig. 7.59 shows the basic form of the *variable reluctance stepper motor.*

— With this form the rotor is made of soft-steel and is cylindrical with four poles, *i.e.*, fewer poles than on the stator.

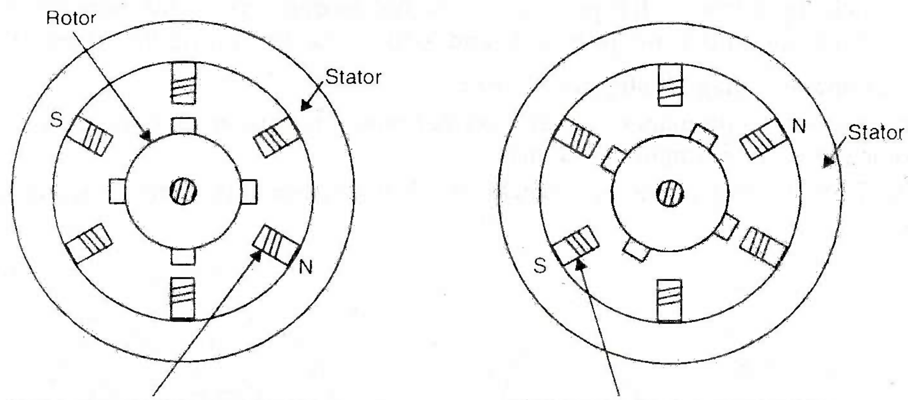


Fig. 7.59. Variable reluctance stepper motor.

— When an opposite pair of windings 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 poles 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 and stator poles line up. This is termed the position of minimum reluctance.

— This form of stepper generally gives step angles of 7.5° or 15°.

- *Stepping angle, irrespective of the type of stepper motor* is given as

$$\alpha = \frac{360°}{\text{Number of phases} \times \text{number of poles}} = \frac{360}{np} \qquad ...(1)$$

3. Hybrid stepper motor :

— This is *infact a permanent magnet stepper motor with constructional features of toothed and stacked rotor adopted from the variable-reluctance motor.*

— The stator has only one set of winding-excited poles which interact with the two rotor stacks.

— The permanent magnet is placed *axially along the rotor in the form of an annular cylinder over the motor shaft* (See Fig. 7.60).

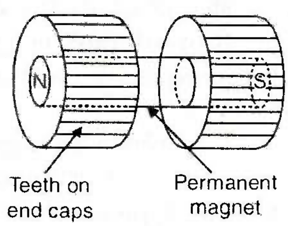


Fig. 7.60. Hybrid motor rotor.

— The stacks at each end of the rotor are toothed. So all the teeth on the stack at one end of the rotor acquire the *same polarity* while the teeth of the stack at the other end of the rotor acquire the *opposite polarity*. The two sets of the teeth are displaced from each other by *one half of the tooth pitch* (also called pole pitch).

— The primary **advantage** of the hybrid motor is that if *stator excitation is removed, the rotor continues to remain locked into the same position, as before removal of excitation.* This is due to the reason that the rotor is prevented to move in either direction by torque because of the permanent excitation.

Torque-speed characteristics of a stepper motor:

Fig. 7.61 shows the torque-speed characteristics of a stepper motor.

— In the "*locked step mode*', the rotor decelerates and may even come to rest between each step. Within this range, the motor *can be instantaneously started, stopped or reversed without losing step integrity.*

— In the "*slewing mode*", the speed is too fast to allow instantaneous starting, stopping, or reversing. The rotor must be gradually accelerated to enter this mode and gradually decelerated to leave the mode. While in slewing mode, the rotor is in synch with the stator field rotation and does not settle between steps.

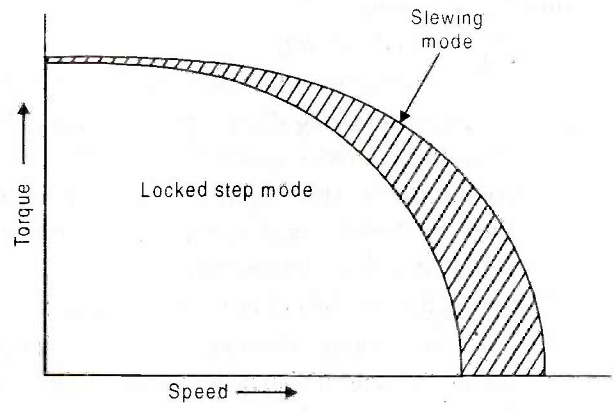


Fig. 7.61. Torque-speed curves of stepper motor.

The curve between the regions in the figure indicates the maximum torques that the stepper can provide at different speeds without slewing. The curve bordering the outside of the slewing mode region represents the absolute maximum torques the stepper motor can provide at different speeds.

Advantages and applications of stepper motor:

Advantages: The stepper motor (*a position control device*) entails the following advantages:

1. Compatibility with digital systems.
2. The angular displacement can be precisely controlled without any feedback arrangement.
3. No sensors are needed for position and speed sensing.
4. It can be readily interfaced with microprocessor (or computer based controller).

Applications: Stepper motors have a wide range of *applications,* mentioned below :

1. Paper feed motors in typewriters and printers.
2. Positioning of print heads.
3. Pens in XY-plotters.
4. Recording heads in computer disc drives.
5. Positioning of worktables and tools in numerically controlled machining equipment.
6. Also employed to perform many other functions such as *metering, mixing, cutting, blending, stirring* etc. in several commercial, military and medical applications.

7.3.4.6. Servomotors

Introduction: The term *servo* or *servo mechanism* refers to a *feedback control system* in which the controlled variable is:

— Mechanical position, or

— Time derivatives *e.g.*, velocity and acceleration.

Following *characteristics* are usually required for a feedback control system:

(*i*) High accuracy.

(*ii*) Remote operation.

(*iii*) Fast-response.

(*iv*) Unattended control.

Following are the *essentials* of a feedback control system :

1. ***An error detecting device.*** It determines when the regulated quantity is *different* from the reference quantity and sends out the error signal to the *amplifier.*
2. ***An amplifier.*** The amplifier receives the error signal and then supplies power to the error-correcting devices, which in turn changes the regulated quantity so that it matches the reference input.

A *servo-motor* should entail the following *characteristics* :

1. *The output torque of the motor should be proportional to the voltage applied* (*i.e.*, the control voltage which is developed by the amplifier in response to an error signal).
2. *The direction of the torque developed by the servo-motor should depend upon the instantaneous polarity of the control voltage.*

Types of servo-motors :

The servo-motors are of the following two types :

1. *D.C. servo-motors.*
2. *A.C. servo-motors.*

1. D.C. servo-motors :

These motors are preferred for very high power systems since they operate more efficiently (as compared to A.C. servo-motors).

These motors may be of the following types :

— Series motors;

— Split series motors ;

— Shunt control motors ;

— Permanent magnet (*fixed excitation*) shunt motor.

(*i*) **Series motors :**

- This motor has a high starting torque.
- It draws large current.
- The speed regulation is poor.
- Reversal can be obtained by reversing field voltage polarity with split series field winding.

(*ii*) **Split series motor :**

- The D.C. series motor with split field (small fraction kW) may be operated as a separately excited field-controlled motor (Fig. 7.62).

  The armature may be supplied from a constant current source.

- A typical torque curve shows

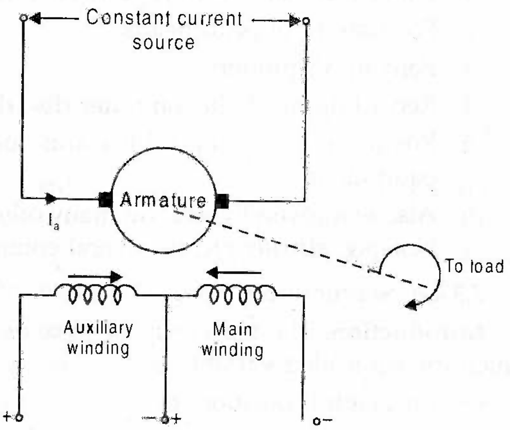


Fig. 7.62. From D.C. amplifier.

the following :

— High stall torque ;

— Rapid reduction in torque with increase in speed.

(*iii*) **Shunt control motor :**

- This type of motor has *two separate windings : Field winding placed on the stator and the armature winding placed on the rotor of the machine.* Both the windings are connected to a D.C. supply source.
- Whereas in a conventional D.C. shunt motor, the two windings are connected in parallel across the D.C. supply mains, but in a *servo-application* in windings are driven by *separate D.C. supplies.*

(*iv*) **Permanent magnet shunt motor :**

- It is a fixed excitation shunt motor where the field is actually supplied by a permanent magnet.
- Its performance is similar to that of armature controlled fixed field motor.

2. A.C. servo motors :

Applications :

- These motors are best suited for *low power applications.*
- Precision servo-motors are used in :

  — Instrument servos ;

  — Computers ;

  — Inertial guidance systems etc.
- The mechanical output power of A.C. servo-motor varies from 2 watts to a few hundred watts.
- An A.C. servo-motor is basically a two-phase induction motor except for certain special design features. The *main important difference* between a standard split-phase motor and an A.C. servo motor is that the *latter* has *thinner conducting bars in the squirrel cage motor, so that the motor resistance is higher.* The torque-speed characteristics should be *linear* as shown by the curve II in Fig. 7.63.

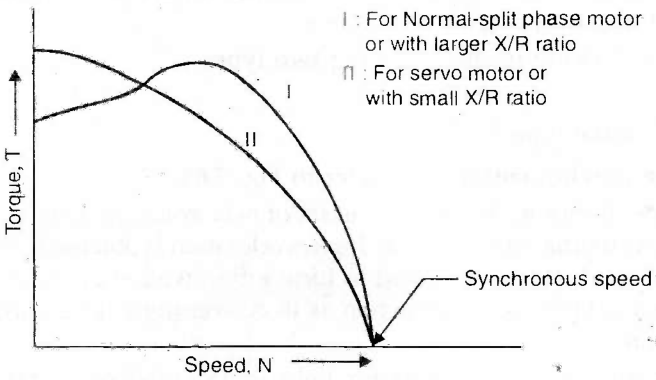


Fig. 7.63. Torque-speed characteristics.

Description of A.C. servo-motors :

1. Drag-cup rotor servo-motor. Refer to Fig. 7.64.

- Drag-cup construction is used for *very low inertia applications.*
- In this type of motor the rotor construction is usually of squirrel cage or drag-cup type; here only a *light cup rotates while the rotor core is stationary* (thus inertia is quite small).

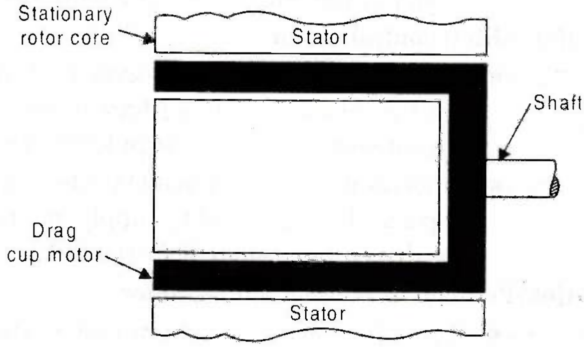


Fig. 7.64. Drag-cup rotor servo-motor.

- The servo-motors contains two windings namely, *main winding* (sometimes called fixed or reference winding) and *control winding.* The voltage applied to the windings are at *right angles* to one another. Usually one winding is excited with a fixed voltage while the other one is excited by the control voltage (which is the output from servo-amplifier).
- *While in operation, the output torque of the motor is roughly proportional to the applied control voltage, and the direction of torque is determined by the polarity of the control voltage.*

2. Shaded-pole type servo-motor :

- This type of motor employs a *phase-sensitive relay to actuate those contacts which produce a short-circuit of the shaded-pole winding to produce rotation in the desired direction.*
- *The main shortcoming of this motor is that it responds only when the amplifier error signal is of adequate magnitude to cause the relay to operate.*

7.3.4.7. Moving coil motors

There are certain applications which require acceleration much higher than what can be achieved in a conventional D.C. servo-motor. The armatures of moving coil D.C. motors have special constructions which allow a substantial reduction in armature inertia and inductance, permitting very high accelerations.

Moving coil motors are of the following two types :

1. Shell type.
2. Disc or Pancake type.

1. Shell type moving coil motor. Refer to Fig. 7.65.

- In this type of motor, the rotor consists of *only armature winding* due to which it has very low inertia ; consequently high acceleration is obtained. Armature winding consists of conductors assembled to form a thin walled cylinder. The *commutator* may have a cylindrical construction as in conventional D.C. motors or disc type construction.

  Low reluctance path for the stator field is provided by a stationary magnetic material cylinder.

  In such a motor the *current is axial and flux is radial.*

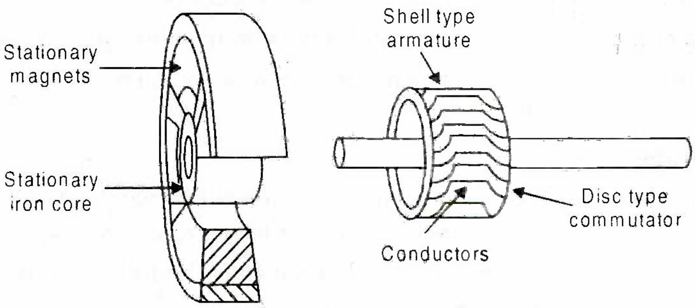


Fig. 7.65. Shell type moving coil motor.

- **Micromotors** (*Tiny motors with diameters around 1 cm*) have armature winding consisting of simply varnished wires arranged in cylindrical form and a disc type commutator.

Such motors find wide applications in *card readers, video systems, cameras* etc.

In *bigger size motors* the armature winding is made by bonding conductors together using polymer resins and fibre glass to provide adequate mechanical strength.

2. Disc or Pancake type moving coil motor. Refer to Fig. 7.66.

In this motor armature is made in disc or pancake form, and armature conductors resemble spokes on a wheel. The armature winding is formed by stamping conductors from a sheet of copper, welding them together and placing them on a light weight disc. Conductor segments are then joined with a *commutator* at the centre of the disc.

Here the direction of *flux is axial* and *armature current is radial* (just opposite to shell type conventional motors).

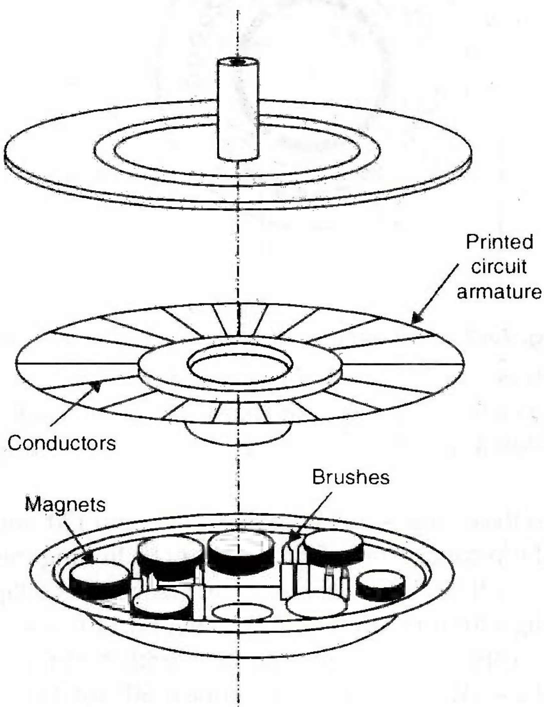


Fig. 7.66. Disc or Pancake type moving coil motor.

The *principle of operation is same as that of a conventional D.C. motor.*

- These motors are more robust and available in sizes *upto few kilowatts.*
- They find *applications* where *axial space is at a premium* such as *machine tools, disc drives* etc.

7.3.4.8. Torque motors

"Torque motors" are *the D.C. motors designed to run for long periods in a stalled or a low speed condition.* Some torque motors are designed to operate at low speeds intermittently.

The torque motor *applications* can be divided into the following three types :

(*i*) Motor is required to *operate in stalled condition.*

— The purpose of the motor is to develop the required tension or pressure on a material, similar to spring.

(*ii*) Motor is required to *move through only a few revolutions* or *degrees of revolution.*

Examples. Opening of valves, switches, clamping devices etc.

(*iii*) This category involves *continuous movement of the motor at low speed.*

Example. Reel drive.

7.3.4.9. Brushless D.C. (or trapezoidal PMAC) motors

Fig. 7.67 shows the cross-section of a 3-phase 2-pole trapezoidal PMAC motor.

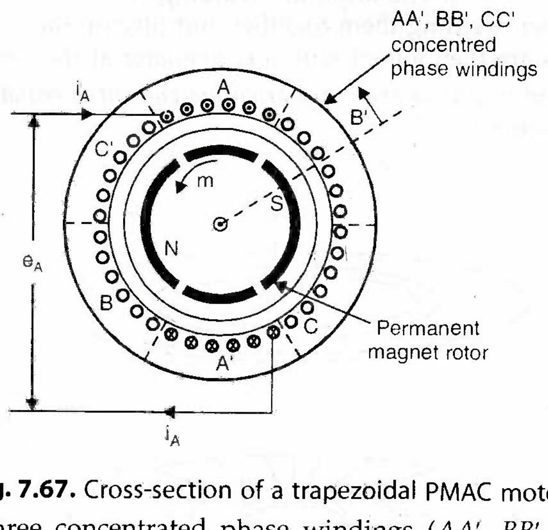


Fig. 7.67. Cross-section of a trapezoidal PMAC motor.

The stator has three concentrated phase windings (*AA'*, *BB'* and *CC'*) which are displaced by 120° and each phase winding spans 60° on each side. The voltages induced in three phases are shown in Fig. 7.68. The reason for getting trapezoidal waveforms is explain below :

When revolving in the counter-clockwise direction, upto 120° rotation from the position shown in Fig. 7.67, all top conductors of phase A will be linking the S-pole and all bottom conductors of phase A will be linking the N-pole. Hence the voltage induced in phase *A* will be the same during 120° rotation (Fig. 7.67). Beyond 120°, some conductors in the top link N-pole and others the S-pole. Same happens with bottom conductors. Hence, the voltage induced in phase *A* linearly reverses in next 60° rotation. Rest of the waveform of phase *A* and waveforms of *B* and *C* can be explained on the same lines.

Fig. 7.68 shows the induced voltage, phase current and torque waveforms of a brushless D.C. motor.

An *inverter fed trapezoidal PMAC motor drive operating in self-controlled mode is called a* **brushless D.C. motor.** This motor is also conceived as *electronically commutated D.C. motor,* because inverter here performs the same function as the brushes and commutator in a D.C. motor *i.e.,* to shift currents between armature conductors to keep the stator and rotor fields stationary, and in quadrature with respect to each other.

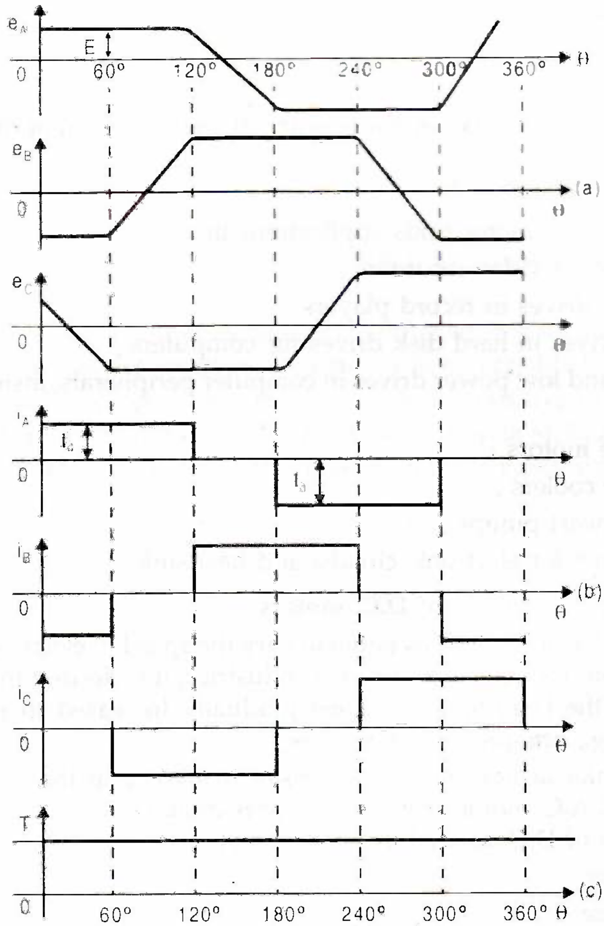


Fig. 7.68. Induced voltage, phase current and torque waveforms of a brushless D.C. motor.

Advantages:

Owing to the absence of brushes and commutator, brushless D.C. motors claim the following *advantages over the conventional D.C. motors :*

(*i*) Long life.

(*ii*) Require practically no maintenance.

(*iii*) High reliability.

(*iv*) Low inertia and friction.

(*v*) Low radio frequency interference and noise.

(*vi*) Because armature windings are on the stator, cooling is much better, *i.e., high specific outputs* can be obtained.

(*vii*) They have a faster acceleration (due to low inertia and friction) and can be run at much higher speeds-upto 100000 r.p.m. and higher are common.

(*viii*) High efficiency, exceeding 75 percent (whereas wound field motors of low power ratings have much lower efficiency).

Disadvantages:

(*i*) High cost.

(*ii*) Low stalling torque.

The size of a brushless D.C. motor is nearly the same as that of the conventional D.C. motor.

Applications:

The brushless D.C. motor finds applications in :

(*i*) Tape drive for video recorders ;

(*ii*) Turn table drives in record players ;

(*iii*) Spindle drives in hard disk drives for computers ;

(*iv*) Low cost and low power drives in computer peripherals, instruments and control systems.

(*v*) Gyroscope motors ;

(*vi*) Cryogenic coolers ;

(*vii*) Artificial heart pumps ;

(*viii*) Cooling fans for electronic circuits and heat sinks.

9.3.4.10. Electronic control of D.C. motors

Introduction : Normally, it is essential to vary the speed of electrical drives in different fields of application. Usually, in all process industries, it is desired that the system be set at slow speed in the beginning and then gradually increased to meet the maximum production rate, *e.g., newspaper printing press.*

One of the major achievements of *thyristor technology* in the field of control is the control of D.C. and A.C. motor drives. *Thyristor controlled schemes* have totally dominated the field of control of D.C. as well as A.C. motors because of the following *advantages* :

(*i*) *Compactness.*

(*ii*) *Fast response.*

(*iii*) *More efficiency.*

(*iv*) *More control capabilities.*

(*v*) *More reliability.*

(*vi*) *Less cost etc.*

Advantages of electronic control systems :

The electronic control system claims the following *advantages over conventional methods* :

1. Very compact and small in size.
2. Consumes very less power.
3. Very fast in response.

4. Much more accurate and efficient than a conventional system.
5. Control ranges are much more than any other systems.
6. High reliability comparatively.
7. Economical, since maintenance cost in minimum.
8. Highly protective.
9. In electronic systems more automation, as required for highly sophisticated machines, is possible.

D.C. Motor speed control :

There are several methods by which the speed of a D.C. shunt motor can be controlled *using thyristors,* some of the commonly used methods are discussed below.

1. Armature voltage control method :

This is also called the *phase control method of speed control.* The complete diagram for this scheme is shown in Fig. 7.69.

- The field of motor is excited by a constant D.C. obtained from the *full-wave rectifier.*
- The armature voltage is *varied by varying the firing angle of the SCRs of the thyristor bridge.* Voltage across the armature terminals will be variable D.C. obtained from the full-wave half-controlled thyristor bridge. In the positive half-cycle SCR_1 and diode D_1 will conduct whereas in the negative half cycle SCR_2 and diode D_2 will conduct. Gates of *SCRs* will be given signal from the triggering circuit (not shown in the Fig. 7.69).

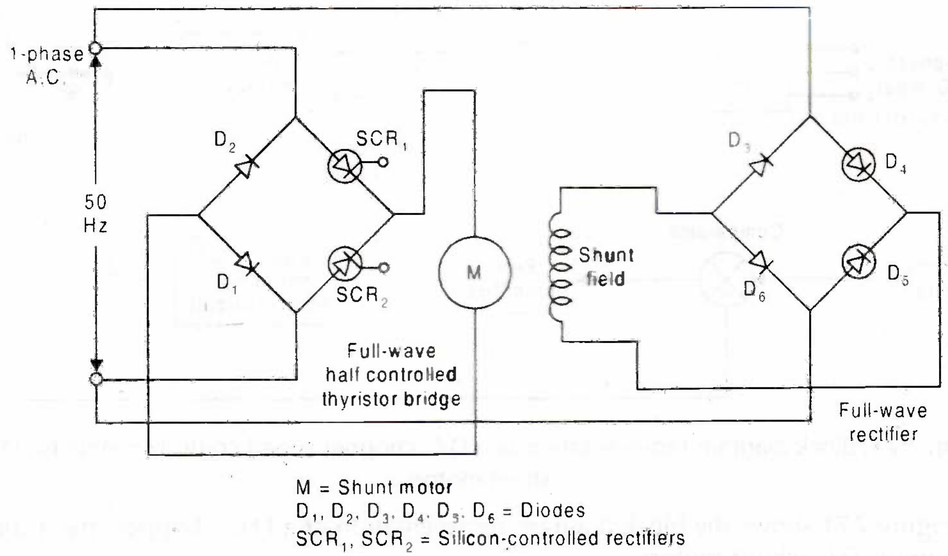


Fig. 7.69. Complete circuit diagram for the armature voltage control method for speed control of D.C. shunt motor.

- The wave shapes for the A.C. input voltage and controlled D.C. armature voltage are shown in Fig. 7.70.

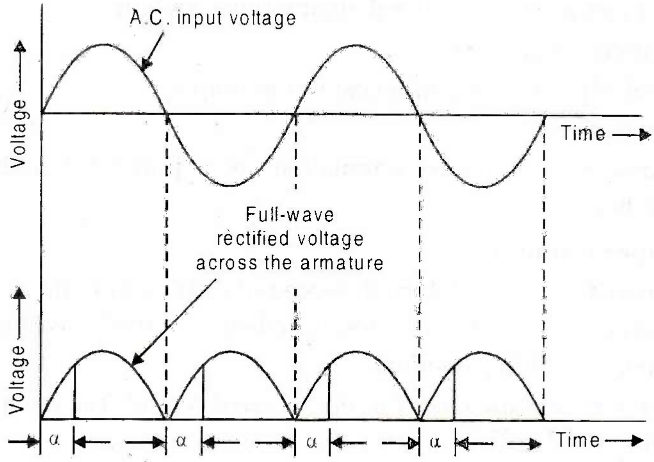


Fig. 7.70. Wave shapes for A.C. input voltage and controlled D.C. armature voltage.

2. *D.C. chopper speed control :*

A *D.C. chopper can give variable D.C. at its output.* This variable can be utilised for the purpose of speed control of D.C. shunt motors. This method of speed control has gained popularity since the introduction of semiconductor devices.

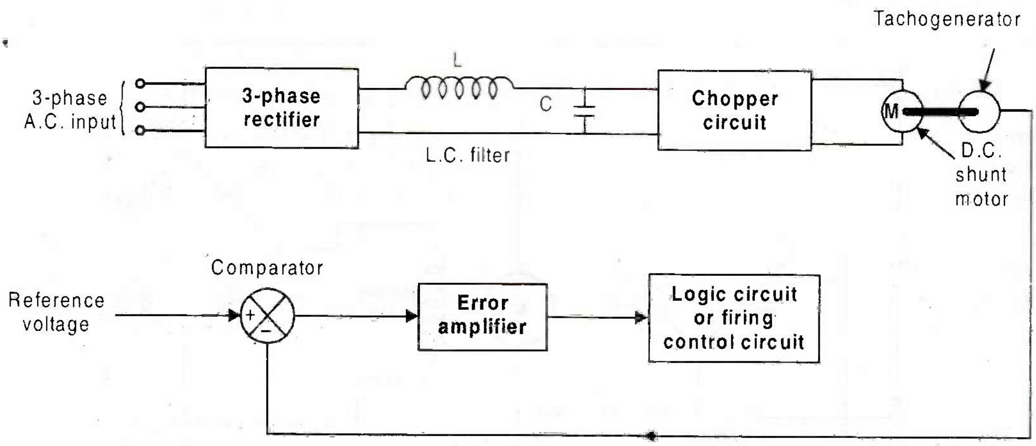


Fig. 7.71. Block diagram representation of a D.C. chopper speed control scheme for D.C. shunt motors.

Figure 7.71 shows the block diagram representation of a D.C. chopper speed control scheme for D.C. shunt motors.

— In this scheme the 3-phase A.C. is rectified into D.C. by means of a *3-phase rectifier.*

— The ripples are minimised with the help of a proper *'LC filter'*.

This filtered rectified D.C. serves as the *input* for the chopper circuit. There is a *'logic circuit'* which decides the firing of the thyristors used in the chopper. The ON, OFF durations for the thyristors used are decided by this unit. The input signal to this logic circuit is obtained from a *'comparator'* through an *'error amplifier'*.

— The speed feedback from the D.C. shunt motor is converted into equivalent voltage signal by means of a *'tachogenerator'*. The speed feedback in the form of voltage signal is given to the comparator where it is compared with the set reference voltage. If there is a difference between the two, it will generate an error signal which is amplified by the error amplifier and sent to the logic circuit to decide the ON, OFF durations of the thyristors connected in the chopper.

— Choppers are built by using one or two SCRs depending upon the type and circuitry used. *This is a very efficient method and is widely used in industries these days because of its fast response.*

— Fig. 7.72 shows a simple circuit diagram for speed control of a D.C. shunt motor. An *L-C* filter is used in the input side of the chopper to *reduce ripples* in the D.C. input. Diode is the freewheeling diode.

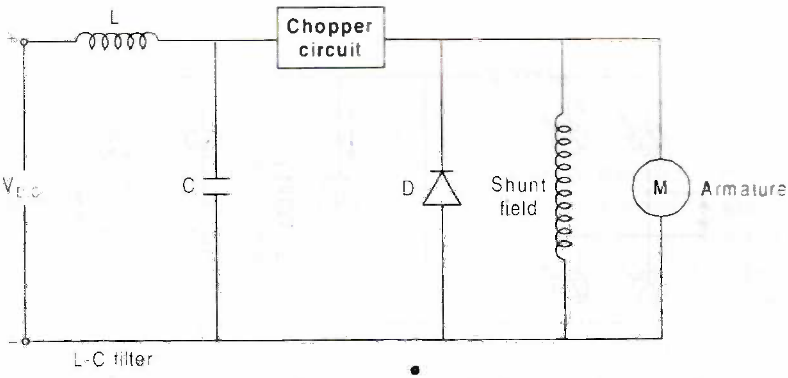


Fig. 7.72. Circuit diagram of a D.C. chopper for speed control of a D.C. shunt motor.

— Fig. 7.73 shows a simple chopper circuit which may be used for controlling the speed of a D.C. series motor.

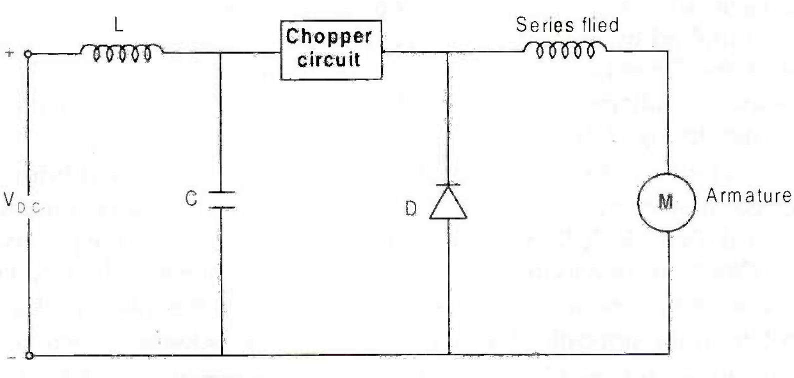


Fig. 7.73. D.C. chopper application for speed control of a D.C. series motor.

— Variation of T_{ON} and T_{OFF} will vary the load voltage at the output of the chopper which will change the speed of the motor accordingly.

— Diode *D* has been used as a *freewheeling diode to provide low resistance path for the* *current which will flow even at the OFF period of the thyristors.* This current flows for a little time due to the stored energy in the winding which is inductive in nature.

— An *L-C* filter has been used in the input side of the chopper to reduce the ripples in the D.C. input voltage.

3. *Speed control by using a dual converter :*

A dual converter, as the name indicates, uses two converters a *rectifier* and an *inverter.* Both the bridges are built by using *SCRs.* A dual converter may be used to obtain the following controls of a D.C. motor :

- Reversible speed control.
- Plugging.
- Regenerative braking.

The above controls are discussed below.

Fig. 7.74 shows the circuit diagram for speed control a D.C. shunt motor using a *dual converter.*

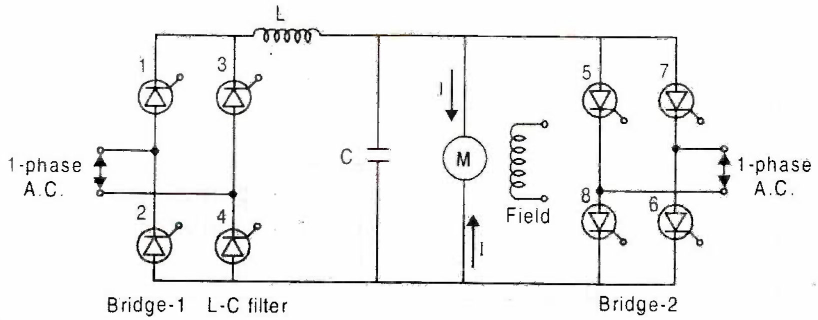


Fig. 7.74. Circuit diagram for speed control of a D.C. shunt motor using a dual converter.

1. *Reversible speed control and plugging :*

- Four *SCRs,* 1, 2, 3 and 4 form the first bridge (Bridge-1) which serves as a 1-phase full-wave fully-controlled bridge and rectifies the 1-phase A.C. into D.C. This D.C. is filtered by an *L.C. filter* to remove the ripples. In the positive half cycle *SCRs* 1 and 2 conduct simultaneously and in the negative half cycle *SCRs,* 3 and 4 conduct simultaneously. The direction of flow of armature current *I* is clockwise as shown in Fig. 7.74.
- For ***reversing*** the direction of rotation of the motor, the second bridge (Bridge-2) is gated after commutating the first bridge. The Bridge-2 is constituted by the *SCRs* 5, 6, 7 and 8. *SCRs* 5 and 6 conduct simultaneously in the positive half cycle and *SCRs* 7 and 8 conduct simultaneously in the negative half cycle. Thus, the direction of flow of armature current is reversed in this case and the motor tries to rotate in the opposite direction *i.e.* in the anticlockwise direction.

Because the motor was originally running in the clockwise direction, the inertia would oppose the torque developed in the anticlockwise direction. When the *two torques become equal, the motor becomes stationary provided bridge-2 is commutated.* This process of stopping the motor is called ***plugging***. If the bridge-2 further continues to conduct, the motor would start-running in the opposite direction resulting in speed reversal. In the opposite direction of rotation of the motor, the speed can be controlled by varying the *firing angle* of the second bridge.

2. Regenerative braking :

In this case, after bridge-1 is commutated and bridge-2 is triggered the counter e.m.f. generated in the armature of the motor acts as input for bridge-2 which is connected in the *inverter mode*. The output of bridge-2 which is 1-phase A.C. may be fed back to the mains supply. Thus, we see that bridge-1 acts as a rectifier and bridge-2 acts as inverter. Therefore, in regenerative braking the K.E. of the motor is converted into electrical energy and fed back to the supply system thereby saving energy.

7.3.5. Single-Phase Motors

7.3.5.1. General aspects

- The number of machines operating from single-phase supplies is greater than all other types taken in total. For the most part, however, they are only used in the smaller sizes, less than 5 kW and mostly in the fractional H.P. range. They operate at *lower power-factors* and are relatively *inefficient when compared with polyphase motors.* Though simplicity might be expected in view of the two-line supply, the *analysis is quite complicated.*
- *Single-phase motors* perform a great variety of useful services in the home, the office, the factory, in business establishments, on the farm and many other places where electricity is available. Since the requirements of the numerous applications differ so widely, the motor-manufacturing industry has developed several types of such machines, each type having operating characteristics that meet definite demands. For example, one type operates satisfactorily on direct current or any frequency upto 60 cycles ; another rotates at absolutely constant speed, regardless of load ; another develops considerable starting torque and still another, although not capable of developing much starting torque, is nevertheless extremely cheap to make and very rugged.

7.3.5.2. Applications and Disadvantages

Applications :

- Single-phase induction motors are in very wide use in industry especially in *fractional horse-power field.*

  They are extensively used for electric drive for *low power constant speed apparatus* such as *machine tools, domestic apparatus* and *agricultural machinery* in circumstances where a three-phase supply is *not* readily available.
- There is a large demand for single-phase induction motors in sizes ranging from a *fraction of horse-power upto about 5 H.P.*

Disadvantages:

Though these machines are useful for small outputs, they are not used for large powers as they suffer from many disadvantages and are never used in cases where three-phase machines can be adopted.

The main *disadvantages* of single-phase induction motors are :

1. Their output is only 50% of the three-phase motor, for a given frame size and temperature rise.
2. They have lower power factor.
3. Lower-efficiency.
4. These motors do not have inherent starting torque.
5. More expensive than three-phase motors of the same output.

7.3.5.3. Construction and working

Construction :

- A single phase induction motor is similar to a 3-ϕ squirrel-cage induction motor in physical appearance. Its rotor is essentially the same as that used in 3-ϕ induction motors. Except for shaded pole motors, the stator is also very similar. There is a uniform air-gap between the stator and rotor but no electrical connection between them. It can be wound for any even number of poles, two, four and six being most common. Adjacent poles have opposite magnetic property and synchronous speed equation, $N_s = \frac{120f}{p}$ also applies.
- The stator windings differ in the following two aspects :
  — ***Firstly*** single phase motors are usually provided with concentric coils.
  — ***Secondly,*** these motors normally have two stator windings. In motors that operate with both windings energised, the winding with the *heaviest wire* is known as the **main winding** and the other is called the **auxiliary winding.** If the motor runs with auxiliary winding open, these windings are usually referred as *running* and *starting.*
  — In *most of motors the main winding is placed at the bottom of the slots and the starting winding on top but shifted 90° from the running winding.*

Working :

When the stator winding of a single phase induction motor is connected to single phase A.C. supply, a *magnetic field* is developed, *whose axis is always along the axis of stator coils.* The magnetic field produced by the stator coils is pulsating, varying sinusoidally with time. Currents are induced in the rotor conductors by transformer action, these currents being in such a direction as to oppose the stator m.m.f. Then the axis of the rotor m.m.f. wave coincides with that of the stator field, the *torque angle* is, therefore, *zero* and *no torque is developed on starting.* However, if the rotor is given a push by hand or by other means in either direction, it will pick-up the speed and continue to rotate in the same direction developing operating torque. Thus a single phase induction motor is not inherently self starting and requires some special means for starting.

The above mentioned behaviour of this type of motor can be explained by any one of the following theories :

1. Double revolving field theory 2. Cross-field theory.

The results given by both the theories are approximately same.

Double revolving field theory is described below :

The magnetic field produced by the stator coils is *pulsating,* varying sinusoidally with time. *Ferrari* pointed out that *such a field can be resolved into two equal fields but rotating in opposite directions with equal angular velocities. The maximum value of each component is equal to half the maximum of the pulsating field.*

If the initial time is such that the rotating vectors of the two component fields are along the *Y*-axis in the positive direction, the two component waves ϕ_1 and ϕ_2 coincide. The resultant of these two is ϕ_{max}. After a short interval of time the two vectors rotate, through an angle θ in their respective directions and the waves are shown to occupy the positions in Fig. 7.75. These waves intersect at *A* on the *Y*-axis and as the waves travel *A* moves along the *Y*-axis. Hence the resultant of these two component waves at any instant is equal to 2*OA*.

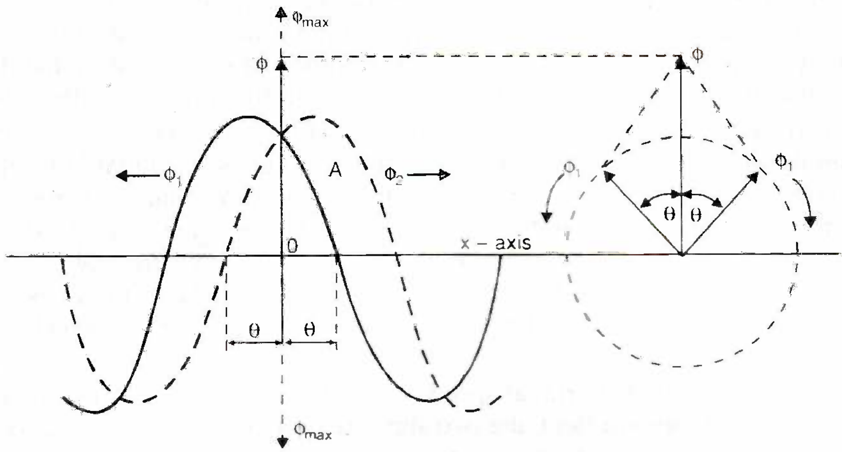


Fig. 7.75.

$$\phi_1 = OA = \phi_{1(max)} \cos(\omega t - \theta) \quad ...(i)$$

$$\phi_2 = OA = \phi_{2(max)} \cos(\omega t + \theta) \quad ...(ii)$$

and, $\phi_{1(max)} = \phi_{2(max)}$

By expanding and adding (*i*) and (*ii*),

$$\phi_1 + \phi_2 = 2\phi_{1(max)} \cos\theta \cos\omega t$$

$$2\,OA = \phi_{(max)} \cos\theta \cos\omega t \quad ...(iii)$$

which is the equation of the pulsating field and *proves "Ferrari's statement"*. Thus a single-phase induction motor is *not inherently self-starting.*

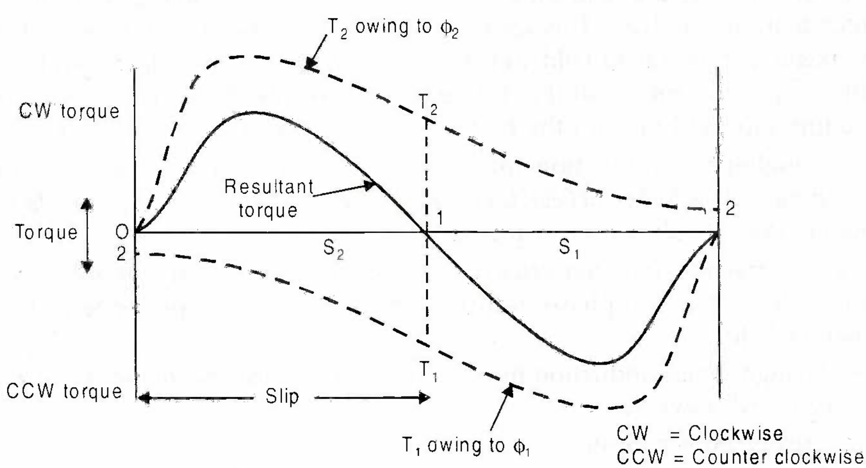


Fig. 7.76. Balanced torque at standstill in squirrel cage rotor excited by a single-phase winding.

The existence of these two fluxes (*forward and backward*) rotating in opposite directions can be verified by supplying a fractional horsepower single-phase induction motor with rated voltage. The motor does not start, but if the shaft is turned by hand, say in clockwise direction, the rotor picks up speed. This means that the rotor conductors are rotating in the direction of that field which rotates in clockwise direction. When the motor is braked and stopped without switching off the supply, the rotor remains at rest. If now the shaft is turned by hand in anti-clockwise direction, the motor picks up speed in that direction. This means that the rotor conductors are now rotating in the direction of the other field.

This behaviour of the motor is *due to the presence of two opposing torques due to the two fields.* When the rotor is at rest, (*i.e.*, slip = 1) the two torques are equal but opposite in direction. Hence the net torque is zero and therefore the rotor remains at rest. Fig. 7.76 shows the torque variations due to the two fields. If the rotor is made to speed up in one direction, say in that in which T_1 increases, T_1 exceeds the opposing torque T_2 and the motor begins to accelerate. T_2 goes on diminishing until at the working speed it is negligibly small. Hence the single-phase induction motor rotates in the direction in which it is made to run.

Thus, if the rotor is made to run at speed N by some external means in any direction, say in the direction of forward field, the two slips are now s and $(2 - s)$, as shown below :

The slip of the rotor w.r.t. the forward rotating field F_f,

$$s_f = \frac{N_s - N}{N_s} = s \qquad \text{...(1)}$$

The slip of the rotor w.r.t. the backward rotating field F_b,

$$s_b = \frac{N_s - (-N)}{N_s} = \frac{2N_s - (N_s - N)}{N_s} = (2 - s) \qquad \text{...(2)}$$

- Under normal running condition $(2 - s) >> s$ and as a consequence the backward field rotor currents are much larger than at standstill and have a low power factor. The corresponding opposing rotor m.m.f., owing to stator impedance, causes the backward field to be greatly reduced in strength. On the other hand, the low-slip forward rotating field induces smaller currents of a high power factor in the rotor than at standstill. This leads to greatly strengthening of forward field. The weakening of backward field and strengthening of forward field depends upon the slip or speed of rotor and the difference increases with the decrease in slip w.r.t. the forward field or with the increase in rotor speed in forward direction.
- In a single-phase induction motor, *the increase in rotor resistance increases the effectiveness of the backward field, which reduces the breakdown torque, lowers the efficiency and increases the slip corresponding to maximum torque.*
- The ***performance characteristics*** of a single phase induction motor are somewhat *inferior* to that of a 3-phase induction motor due to the presence of backward rotating field.
  - — A single-phase induction motor has a *lower breakdown torque at larger slip and increased power losses.*
  - — Greater power input.
  - — The *speed regulation* tends to be *poorer* than that for a polyphase motor.
  - — The *power factor tends to be lower* (since the normal slip of a single-phase induction motor under load conditions is rather greater than that of the corresponding 3-phase motor).
- In view of the above factors, a single phase induction motor has a *larger frame size* than that of 3-phase motor.
- Single-phase motors *tend to be somewhat noiser than 3-phase motors which have no such pulsating torque.*

7.3.6. Three Phase Induction Motors

7.3.6.1. Introduction

An induction motor is simply an *electric transformer* whose magnetic circuit is separated by an air gap into two relatively movable portions, one carrying the primary and the other the secondary winding. Alternating current supplied to the primary winding from an electric power system induces an opposing current in the secondary winding, when *latter is short-circuited or closed through an external impedance.* Relative motion between the primary and secondary structures is produced by the *electromagnetic forces corresponding to the power thus transferred across the air gap by induction.*

The *essential feature which distinguishes the induction machine from other types of electric motors is that the secondary currents are created solely by induction, as in a transformer instead of being supplied by a D.C. exciter or other external power source, as in synchronous and D.C. machines.*

Advantages : Three-phase induction motor is the *most commonly used motor* in industrial applications because of the *advantages* listed below :

1. Simple design.
2. Rugged construction.
3. Reliable operation.
3. Low initial cost.
5. Easy operation and simple maintenance.
6. High efficiency.
7. Simple control gear for starting and speed control.

Applications :

Induction motors are available with torque characteristics suitable for a *wide variety of loads :*

(*i*) The standard motor has a starting torque of about 120 to 150 per cent of full-load torque. Such motors are suitable for most applications.

(*ii*) For starting loads such as small refrigerating machines or plunger pumps operating against full pressure or belt conveyors, *high torque motors with a starting torque of twice normal full-load torque,* or more, are used.

(*iii*) For driving machines that use large flywheels to carry peak loads, such as punch presses and shears, a high-torque motor with a slip at full-load up to 10 per cent is available. *The high slip permits enough change in speed to make possible the proper functioning of the flywheel.*

(*iv*) By the use of a wound-rotor with suitable controller and external resistances connected in series with the rotor winding, it is possible to obtain any value of starting torque up to the maximum breakdown torque. *Such motors are well adapted as constant-speed drives for loads that have large friction loads to overcome at starting.*

7.3.6.2. Constructional details

The Stator :

- The stator frame consists of a symmetrical and substantial casting, having feet cast integral with it. The stator core, consisting of high grade, low loss electrical sheet-steel stampings, is assembled in the frame under hydraulic pressure. The thickness of stampings/laminations is usually from 0.35 to 0.6 mm. The

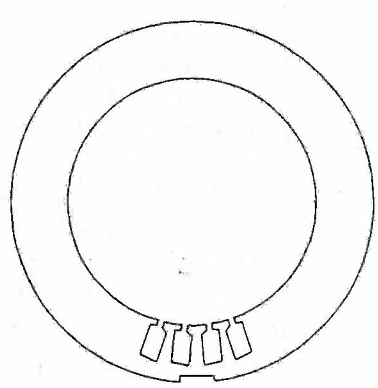


Fig. 7.77. Stator stamping.

stator laminations are punched in one piece for small induction motor (Fig. 7.77). In induction machines of large size the stator core is assembled from a large number of *segmental laminations.*

The slots are *sometimes of the 'open type* (*i.e.*, having parallel walls) for the accommodation of former wound coils. But the usual practice is to have practically *'enclosed slots' in order to reduce the effective* length of air-gap.

- The *stator windings* are given the utmost care to make them mechanically and electrically sound, so as to ensure long life and high efficiency. After the winding is in position it is thoroughly dried out whilst still hot and is completely immersed in a high grade synthetic resin varnish. It is then *acid, alkali, moisture* and *oil proof.*

  For *small motors* working at ordinary voltages, *single layer mush winding* is used. For *medium size machines double layer lap winding with diamond shaped coils* is used. *Single layer concentric windings are used for large motors working at high voltages.*

- *Frames* of electrical machines house the stator core. Frames of small and medium sizes of induction motors have hollow cylindrical form and that of large motors have the shape of a circular box. *In small induction motors,* having a frame diameter of up to about 150 cm, *the frame also supports the end shields.* The frame should be *strong and rigid* as rigidity is very important in the case of induction motors of large dimensions. *This is because of the length of the air gap is very small and if the frame is not rigid, it would create an irregular air gap around the machine resulting in production of unbalanced magnetic pull.* Frames for small machines are made as a *single unit* and are usually cast. The frames of *medium and large sized* machines are *fabricated from rolled steel plates.*

The Rotor : The rotors are of two types :

1. Squirrel-cage ;
2. Wound rotor.

1. ***Squirrel-cage.*** The squirrel-cage rotor is made up of stampings (Fig. 7.78), which are keyed directly to the shaft. The slots are partially closed and the winding consists of embedded copper bars to which the short-circuited rings are brazed. The squirrel cage rotor is so robust that it is almost indestructible.

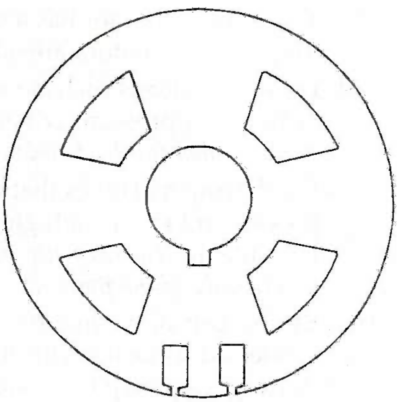


Fig. 7.78. Rotor stamping.

The great majority of present day induction motors are manufactured with squirrel-cage rotors, a *common practice being to employ winding of cast aluminium.* In this construction the assembled rotor laminations are placed in a mould after which molten aluminium is forced in, under pressure, to form bars, end rings and cooling fans as extension of end rings. This is known as *die cast rotor* and has become very popular as there are no joints and thus there is no possibility of high contact resistance.

In this type of rotor, it may be noted that *slots are not made parallel to the shaft but they are 'skewed'* to serve the following purposes :

(*i*) *To make the motor run quietly by reducing the magnetic hum.*

(*ii*) *To reduce the locking tendency of the rotor.*

2. ***Wound rotor.*** The wound rotor has also slotted stampings and the *windings are former wound*. The *wound rotor construction is employed for induction motors requiring speed control or extremely high values of starting torque*. The wound rotor has completely insulated copper windings very much like the stator windings. The winding can be connected in star or delta and the three ends are brought out at the three slip rings. The current is collected from these slip rings with carbon brushes from which it is led to the *resistances for starting purposes*. When the motor is *running, the slip rings are short-circuited by means of a collar, which is pushed along the shaft and connects all the slip rings together on the inside*. Usually the *brushes are provided with a device for lifting them from the slip rings when the motor has started up, thus reducing the wear and the frictional losses.*

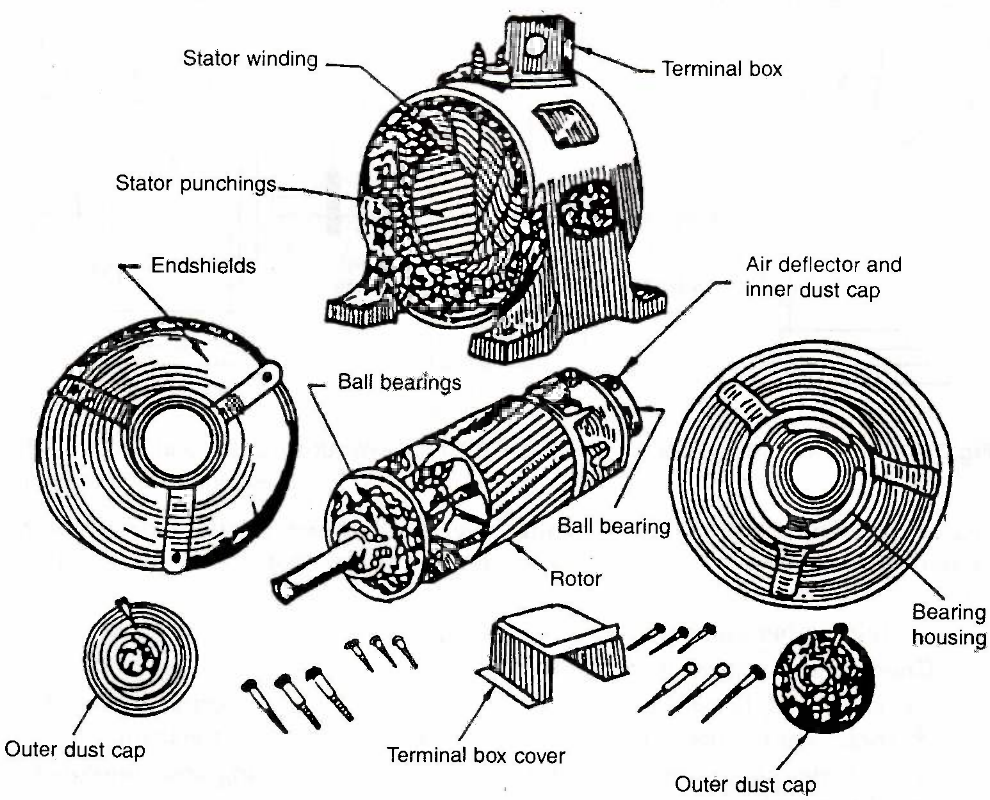


Fig. 7.79. Component parts of a small squirrel-cage induction motor.

The *number of slots in the rotor should never be equal to the number of slots in the stator*. If they are, there would be a *variation of reluctance of the magnetic path* from maximum, when teeth are opposite slots, to minimum when teeth are opposite teeth. The resulting flux pulsations would have a high frequency, since the periodic time would be the interval period for a tooth to occupy similar positions opposite two successive teeth. This will *not only cause extra iron loss but the rotor will tend to lock with the stator* if at the time of starting teeth are opposite teeth. The best plan is to make the number of the stator and the rotor teeth *prime to each other.*

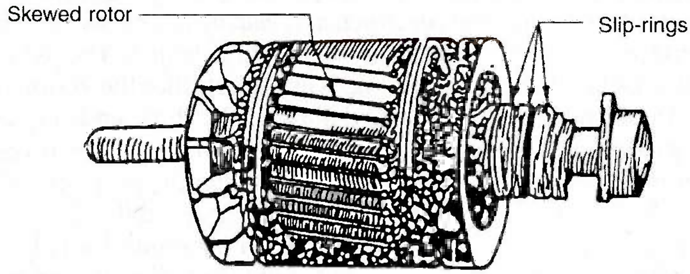


Fig. 7.80. Induction motor with phase-wound rotor, showing the three slip rings on the rotor shaft.

Figs. 7.81 and 7.82 show squirrel-cage and phase-wound induction motors respectively.

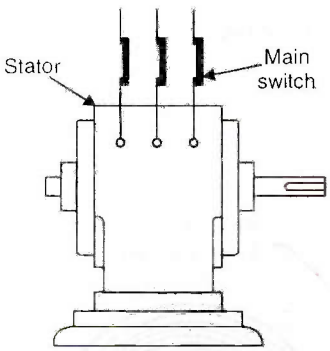


Fig. 7.81. Squirrel-cage motor.

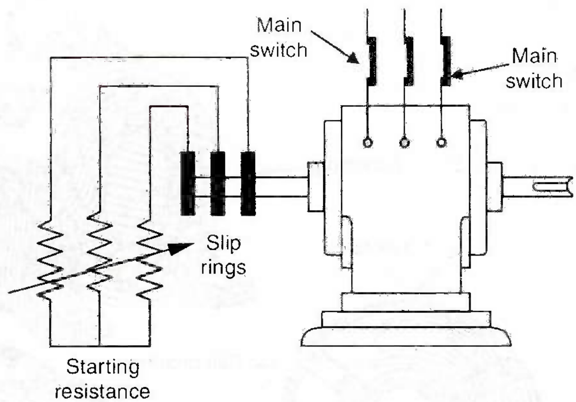


Fig. 7.82. Phase -wound motor connected to a three-phase star-connected starting resistance.

Advantages of a squirrel-cage motor over a phase-wound induction motor. As compared with a wound rotor a squirrel-cage induction motor entails the following *advantages :*

1. Slightly higher efficiency.
2. Cheaper and rugged in construction.
3. No slip rings, brush gear, short-circuiting devices, rotor terminals for starting rheostats are required. The star-delta starter is sufficient for starting.
4. It has better space factor for rotor slots, a shorter overhang and consequently a small copper loss.
5. It has a smaller rotor overhang leakage which gives a better power factor and a greater pull out torque and overload capacity.
6. It has bare end rings, a large space for fans and thus the cooling conditions are better.

The major *'disadvantage' of squirrel-cage motor* is that *it is not possible to insert resistance in the rotor circuit for the purpose of increasing the starting torque. The cage rotor has a smaller starting torque and large starting currents as compared with wound rotor.*

Slip rings. The slip rings for wound-rotor machines are made of *either brass or phosphor bronze.* They are shrunk on to a cast iron sleeve with moulded silica insulation. This assembly is passed on to the rotor shaft. The slip rings are rotated either between the core and the bearing or on the shaft extension. In the latter case the shaft is made hollow to allow the three connections from rotor to slip rings to pass through bearings.

Shaft and bearings. *In an induction motor the air gap is made as small as possible.* Therefore the shaft is made short and stiff in order that the rotor may not have any significant deflection, *as even a small deflection would create large irregularities in the air gap which would lead to production of an unbalanced magnetic pull. There is also a possibility of rotor and stator fouling with each other. Ball and roller bearings are generally used as with their use, accurate centering is much simpler than with journal bearings.* Also the *overall length of machine is reduced. For small motors,* a roller bearing may be used at the driving end and a ball bearing at the non-driving end. *For large and heavy rotors journal bearings are used.*

7.3.6.3. Theory of operation of an induction motor

When a three-phase is given to the stator winding a rotating field is setup. This field sweeps past the rotor (conductors) and by *virtue of relative motion,* an e.m.f. is induced in the conductors which form the rotor winding. Since this winding is in the form of a closed circuit, a current flows, the direction of which is, by Lenz's law, such as to oppose the change causing it.

Now, *the change is the relative motion of the rotating field and the rotor, so that, to oppose this, the rotor runs in the same direction as the field and attempts to catch up with it.* It is clear that torque must be produced to cause rotation, and *this torque is due to the fact that currents flow in the rotor conductors* which are situated in, and at right angles to, a magnetic field.

Fig. 7.83 shows the induction motor action.

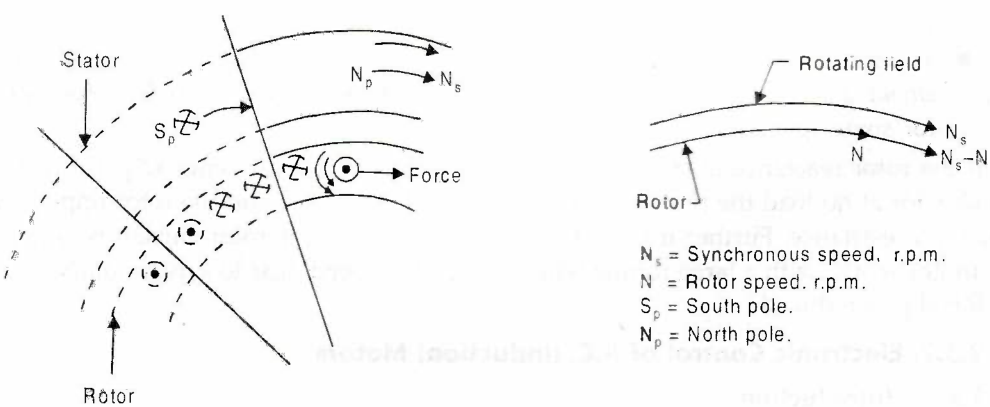


Fig. 7.83. Induction motor action.

- When the motor shaft is *not loaded,* the machine has only to rotate itself against the mechanical losses and the rotor speed is *very close to the synchronous speed.* However, the rotor speed cannot become equal to the synchronous speed because if it does so, *the e.m.f. induced in the rotor winding would become zero and there will be no torque.* Hence the speed remains slightly less than the synchronous speed. If the motor shaft is *loaded,* the rotor will slow down and the relative speed of the rotor with respect to the stator rotating field will increase. The e.m.f. induced in the rotor winding will increase and will produce more rotor current which will increase the electromagnetic torque produced by the motor. Conditions of equilibrium are attained when the rotor speed has adjusted to a new value so that the electromagnetic torque is sufficient to balance the mechanical or load torque applied to the shaft. The speed of the motor *when running under full load conditions is somewhat less than the no-load speed.*

7.3.6.4. Slip

- As earlier stated, *the rotor speed must always remain less than the synchronous speed. The difference between the synchronous speed and the rotor speed is known as* **'slip'**. It is usually expressed as a fraction of the synchronous speed. Thus slip s is

$$s = \frac{N_s - N}{N_s} \quad ...(7.6)$$

or,
$$N = N_s(1 - s) \quad ...(7.6a)$$

where, N_s = Synchronous speed (r.p.m.)

N = Motor speed (r.p.m.)

In practice the value of slip is very small. At no-load, slip is around 1% or so and at full-load it is around 3%. For large efficient machines the slip at full-load may be around 1% only. The induction motor, is therefore, a motor with substantially constant speed and fills the same role as *D.C. shunt motor.*

- When the rotor is stationary (standstill) its speed is zero and $s = 1$. The rotor cannot run at synchronous speed because then their will be no rotor e.m.f. and no rotor current and torque. If the rotor is to run at synchronous speed an external torque is necessary. *If the rotor is driven such that $N > N_s$, the slip becomes negative, the rotor torque opposes the external driving torque and the machine acts as* ***induction generator.***
- The induction motor derives its name from the fact that the *current in the rotor circuit is induced from the stator.* There is no external connection to the rotor except for some special purposes.

If the rotor reactance at standstill is X_2 its value at slip 's' becomes sX_2. *This is very desirable,* for at no-load the reactance becomes almost negligible and the rotor impedance is now all resistance. Further if the rotor resistance is small the rotor current is large, so that motor works with a large torque which brings the speed near to synchronous speed, *i.e.,* the slip is reduced.

7.3.7. Electronic Control of A.C. (Induction) Motors

7.3.7.1. Introduction

The speed of a D.C. motor can be controlled by varying the field current or the armature voltage through a phase controlled rectifier or by a D.C.–D.C. converter if the input supply is D.C. Also, in a D.C. machine the torque is developed due to the interaction of field flux and the D.C. armature flux which remains stationary in space. Whereas in A.C. machine, a 3-phase supply to the stator winding *produces rotating magnetic field of constant magnitude and which reacts with the rotor m.m.f. to develop the torque.* The rotor m.m.f. in case of in induction motor is created by the stator induction effect, whereas in case of synchronous motor the rotor m.m.f. is created by a separate field winding which carries D.C. current.

The speed of an A.C. machine depends upon the stator supply frequency which produces the synchronously rotating magnetic field. If the frequency of the stator supply is increased to increase the speed of the motor, the magnitude of air gap flux is reduced due to increased magnetising reactance and correspondingly the developed torque is reduced. This shows that the *speed and torque of an A.C. motor can not be controlled independently by the conventional methods of speed control.* For this reason, an *A.C. motor requires 'variable voltage variable frequency' power supply for its speed control* A *'D.C. link converter system'* consisting of a rectifier and an inverter or a *'cycloconverter'* can be used as a variable voltage-variable frequency source.

However, it may be noted that the *voltage and current waves obtained by solid state devices are rich in harmonics and cause problem of harmonic heating torque pulsation.*

7.3.7.2. Speed control of a single-phase induction motor

The most common method for speed control of a single-phase induction motor is the *stator voltage control method.*

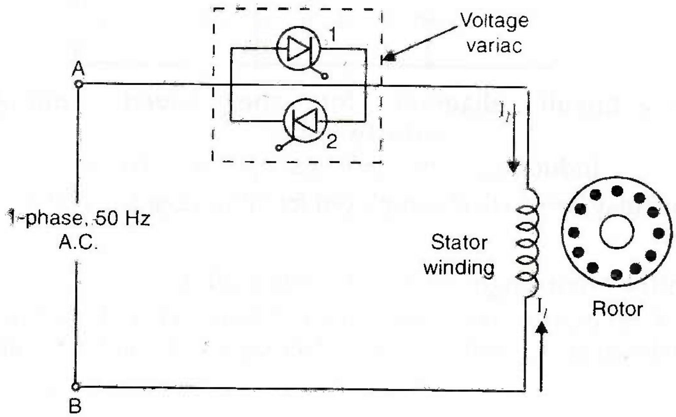


Fig. 7.84. Circuit diagram for speed control of a single-phase induction motor by stator voltage control method.

Figure 7.84, shows the circuit diagram for speed control of a single-phase induction motor by stator voltage control method. The circuit uses two *SCRs* connected in anti parallel. In the +ve half cycle when point *A* is positive and point *B* is negative SCR_1 is triggered. The direction of flow of current in the stator winding is from the top to bottom. In the –ve half-cycle point *A* becomes negative and *B* becomes positive. SCR_1 is turned OFF and SCR_2 is triggered. The direction of flow of current in the stator winding is reversed. In other words, the alternating current supply becomes available across the stator winding of the motor. By varying the firing angles of *SCRs* 1 and 2 the magnitude of the A.C. voltage across the stator winding of the motor can be *varied* ; this in turn will *vary the speed of the motor.*

Different schemes under this method are discussed below.

1. Speed control by using triac :

By using a triac, *very smooth speed control* of a single-phase induction motor can be obtained. A diac is used as a triggering agent for the triac in the circuit. Fig. 7.85 shows the circuit diagram for this arrangement. A diac-triac pair can provide the widest range of control.

There are two *R-C* networks. R_1-C_1 from the triggering circuit whereas R_2-C_2 along with C_1 form the π-network (filter) which would bypass any spike in the A.C. mains supply. The values of R_2 and C_2 are lower than the values of R_1 and C_1. R_2 also works as a current limiting resistance for the diac. The *R-C* triggering gate control process adopted in this circuit provides a very wide and smooth speed control for 1-phase induction motor.

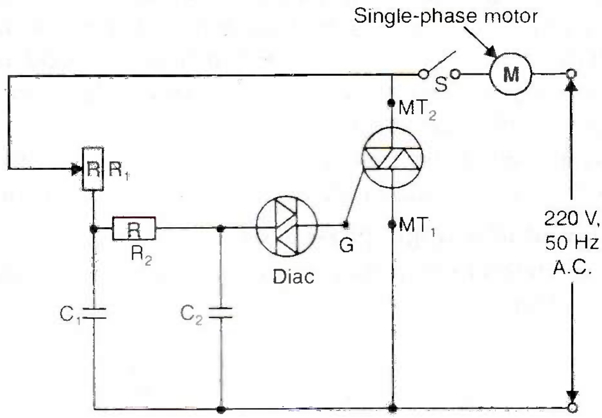


Fig. 7.85. Circuit diagram for the speed control of a single-phase induction motor using a triac.

- This circuit may be effectively employed for fabricating *fan regulators and illumination controllers.*

2. Speed control using single-phase inverter circuit :

With the help of an 'inverter circuit' we can obtain variable voltage fixed frequency A.C. supply which can be fed to the motor for speed control. Fig. 7.86, shows the block diagram representation for this scheme.

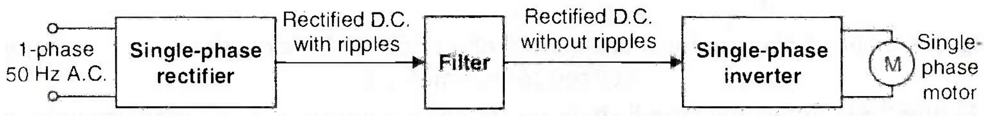


Fig. 7.86. Block diagram representation for the scheme of speed control of a single-phase induction motor using an inverter circuit.

- Single-phase A.C. is rectified with the help of a single-phase full-wave rectifier and then filtered to minimise the ripple content.
- The inverter output, which is a *fixed frequency variable A.C. voltage,* is fed to the motor whose speed is to be controlled. Inverter output (A.C. in nature) is *made variable by changing the firing time (angle) of SCRs.*

This process is known as *fixed frequency variable voltage control.*

3. Speed control by using cycloconverter circuits :

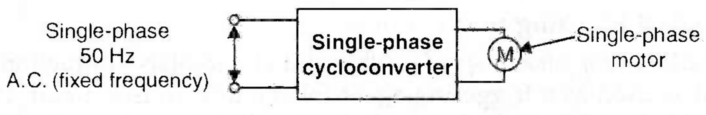


Fig. 7.87. Block diagram representation of the cycloconverter scheme for speed control of a single-phase induction motor.

Basically, this is a *variable frequency method for speed control.* By controlling the firing sequence of the *SCRs* connected in a cycloconverter the frequency of the A.C. input voltage can be changed. The variable frequency A.C. supply available at the point of the cycloconverter circuit may be fed to the motor for speed control. A block diagram representation of such a scheme is shown in Fig. 7.87.

- *"Cycloconverters" are mostly used for speed control of gearless drives.*

7.3.7.3. Speed control of three-phase induction motors

Following methods are used for *controlling the speed of three-phase induction motors :*

1. Stator voltage control or variable voltage constant frequency control.
2. Variable voltage variable frequency control.
3. Variable current variable frequency control.
4. Regulation of slip power.

The basic principles of operation of these methods are given as follows :

1. Stator voltage control :

By using a thyristor A.C. controller circuit, constant frequency variable voltage supply can be generated. Connection diagram for such scheme is shown in Fig. 7.88.

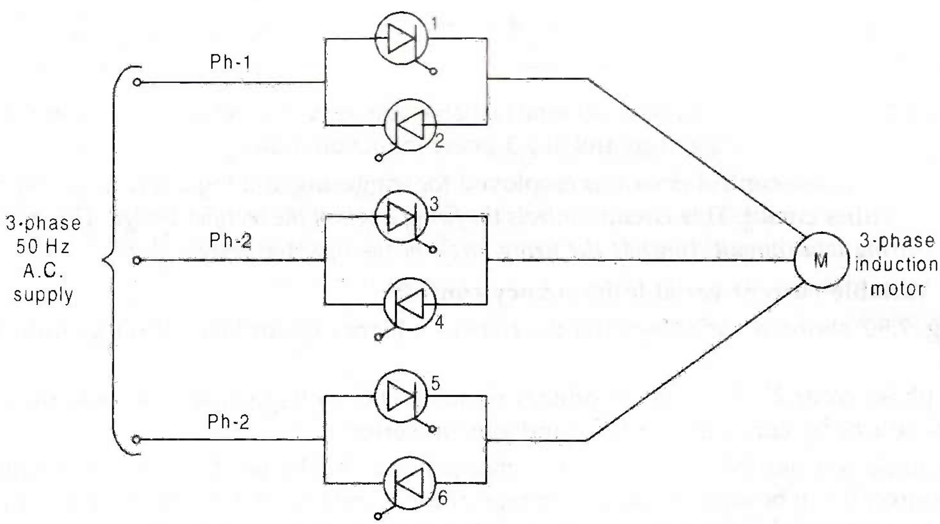


Fig. 7.88. Phase-controlled A.C. supply for three-phase induction motor control.

By changing the applied voltage, air gap flux can be changed so also the slip, and motor speed can be altered. To obtain a reasonable control a full thyristor controller is used. Two *SCRs* connected in antiparallel per phase are used to form three such bridges. *SCRs* 1, 2 form the bridge for phase-1, similarly, 3, 4 form the bridge for phase-2 and 5, 6 for phase-3. The controlled (variable) three-phase voltage, when fed to the 3-phase induction motor, will result in the desired speed control of the motor.

- This arrangement is *quite costly* and *its firing circuit will also be quite complicated.*

2. Variable-voltage variable-frequency control :

Fig. 7.89 shows the basic block diagram for a speed control scheme of a 3-phase induction motor. This is basically a *variable-voltage variable-frequency supply scheme* for the speed control.

- 3-phase A.C. is rectified into D.C. and then filtered to minimise the ripple content. *L-C.* filter is generally used for this purpose. This controlled D.C. is converted into controlled pulses by means of a voltage to frequency converter. These controlled pulses are fed to the inverter bridge for producing the variable-voltage variable-frequency output. This output is fed to the 3-phase induction motor for controlling its speed.

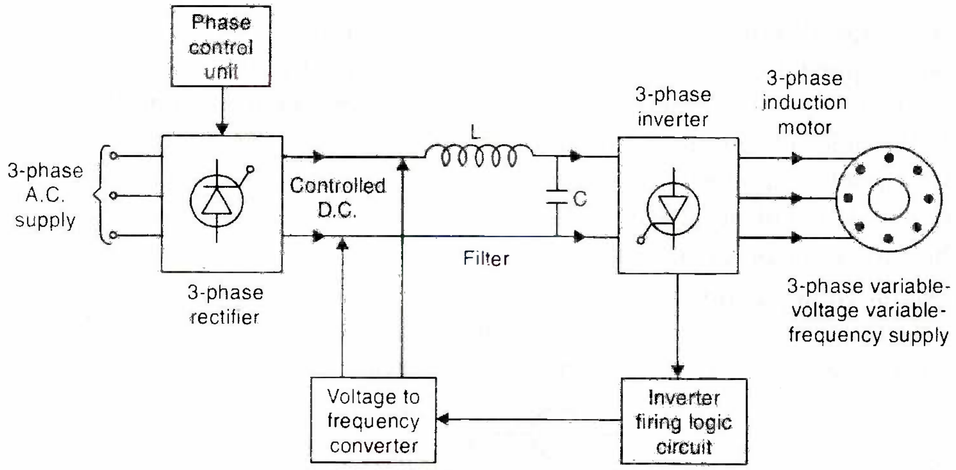


Fig. 7.89. Block diagram for basic scheme (variable-voltage variable-frequency control) for speed control of a 3-phase induction motor.

- The 'phase control circuit' is employed for triggering and logic sensing of 3-phase rectifier circuit. This circuit *controls the firing angle of the rectifier bridge.* The *'inverter firing logic circuit' controls the firing angle of the inverter bridge.*

3. Variable-current variable-frequency control :

Fig. 7.90, shows a variable-current variable-frequency control circuit for an induction motor.

A phase-controlled rectifier produces variable D.C. voltage which is converted to a current source by connecting a large inductor in series.

A diode rectifier followed by a D.C. chopper can also be used as a variable voltage D.C. source. It can be shown that the voltage at the terminals of 3-phase induction motor is almost sinusoidal with superimposed voltage spikes due to commutation.

The *'converter'* used is a *line commutated* whereas the *'inverter'* is *forced commutated* as the induction motor is a lagging p.f. load.

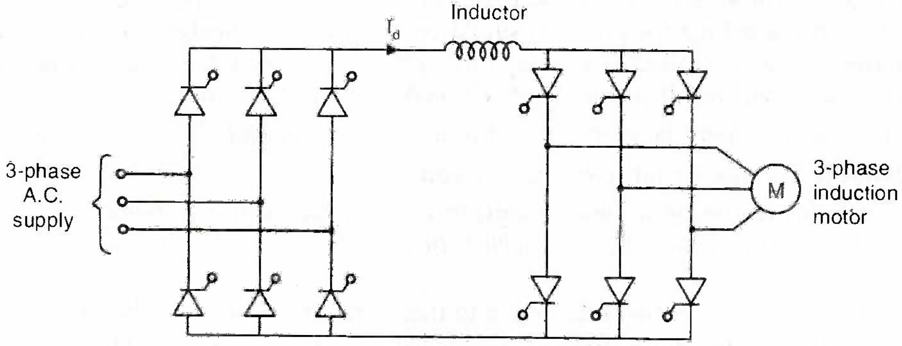


Fig. 7.90. Variable-current variable-frequency control circuit.

The following are the *advantages* and *disadvantages* of this circuit :

Advantages :

(*i*) The control circuit is simpler and more reliable since only six thyristors are to be controlled.

(*ii*) The power circuit is rugged and reliable.

(*iii*) Any fault on the inverter side causes slow rise of fault current which can be cleared by converter grid control.

(*iv*) Less number of components in inverter circuit and less commutation loss.

(*v*) Regenerative process is simple and no additional component is required.

Disadvantages :

(*i*) The inverter is somewhat bulky and expensive (due to the large size of the inductance and commutation capacitors).

(*ii*) The response of the drive is somewhat sluggish.

(*iii*) The frequency range of the inverter is low and it cannot operate under no load condition as some minimum load current is required to commutate the inverter satisfactorily.

4. Slip power recovery method :

When the supply frequency and the voltage are fixed, the speed of an induction motor can be *varied by injecting a counter e.m.f. into the rotor circuit of the motor.* This method is, therefore, *used for a wound-rotor induction motor.* The inefficiency of the drive system because of large slip power dissipation can be overcome by this method.

In this method the slip power of the motor is rectified by a diode rectifier and is then pumped back to the A.C. line through a line commutated inverter. Fig. 7.91 shows the schematic diagram for this method.

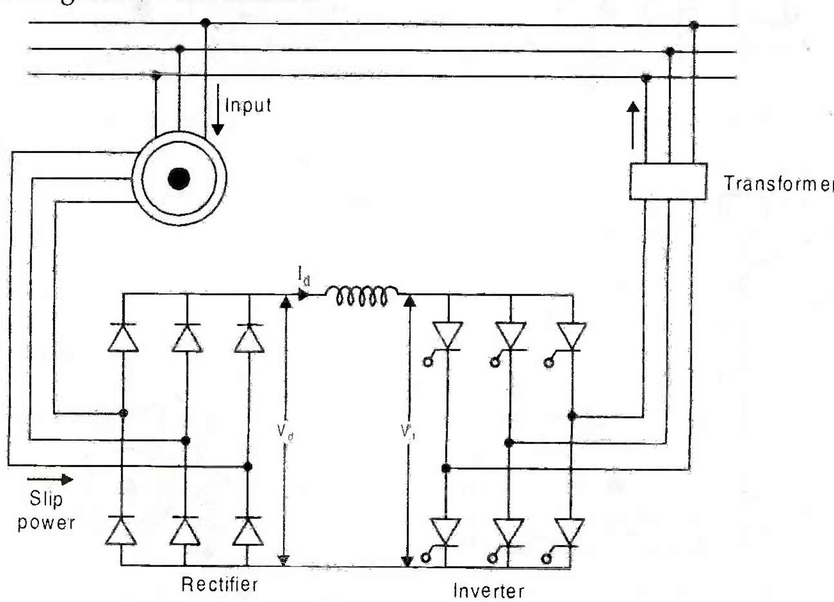


Fig. 7.91. Slip power recovery method.

The required power handling capability of the converter corresponds to the slip power and hence if the range of speed control is small the rating of converter is also small. However, if speed control upto standstill is the required, the converters should be rated for the full machine rating. The initial in-rush current in the converters can be avoided by connecting separate starting resistances in the slip ring circuit.

The torque pulsation and additional heating must be considered while designing the drive system.

*The **disadvantage*** of this system is that *regeneration and speed reversal are not possible in the drive system.*

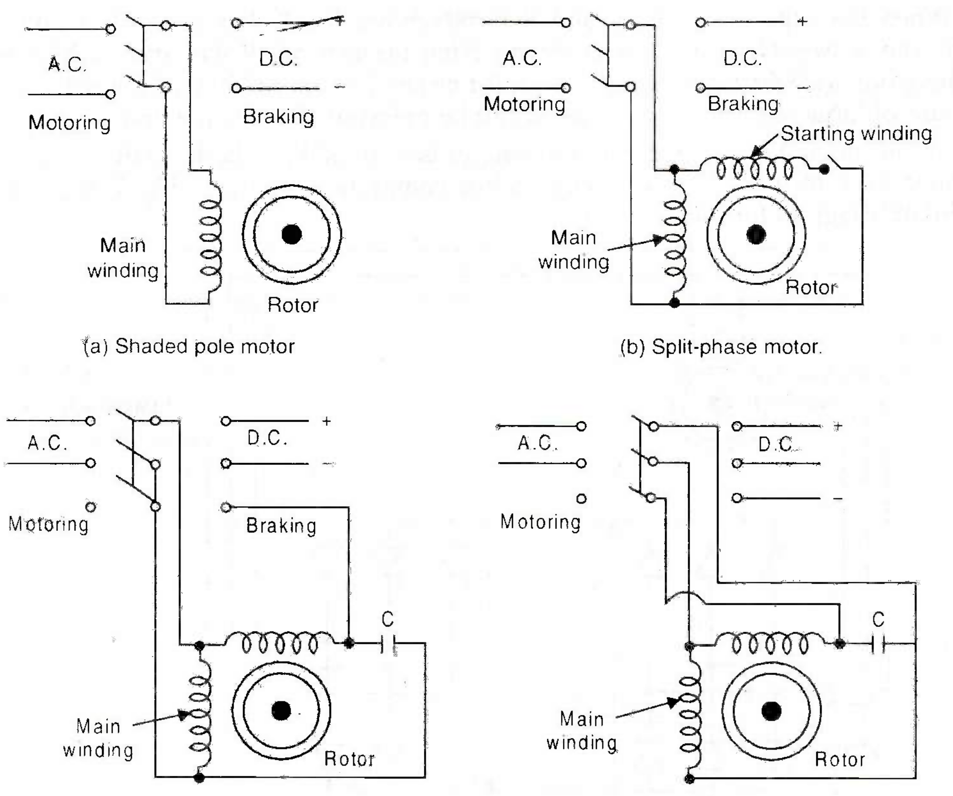
- The drawback of this method that it results in *very poor p.f.*
- This drive system is used in *large H.P. pump and blower type applications where limited range of speed control is required.*

7.3.7.4. Braking of single-phase motors

These motor can be braked by :

(*i*) D.C. dynamic braking (*ii*) Plugging.

(*i*) ***D.C. dynamic braking :*** This method is commonly used for braking of single-phase induction motors. With the help of a double-pole double-throw switch or triple-pole double-throw switch, motor connection is shifted from A.C. (motoring) to D.C. source for braking. These connections for various single-phase induction motors are shown in Fig. 7.92.



(a) Shaded pole motor

(b) Split-phase motor.

(c) Capacitor-run motor, parallel winding connection for braking.

(d) Capacitor-run motor, series winding connection for braking.

Fig. 7.92. D.C. dynamic braking of single-phase induction motors.

- In case of split-phase, capacitor-run, and capacitor-start and capacitor-run motors, either main winding can be connected across the D.C. source [Fig. 7.92(*b*)] or main and auxiliary windings connected in series or parallel [Fig. 7.92(*c*) and (*d*)].
- When in braking connection, D.C. current through the stator winding (or windings) produces a stationary field through which squirrel cage rotor moves. Current induced in rotor bars *interacts with D.C. field to produce braking torque,* as in 3-phase induction motor. Motor *decelerates and stops.* As induced currents are zero at zero speed, the braking torque is also zero. *For braking, supply is obtained by a diode rectifier connected to A.C. mains.* Motor winding can be connected directly across diode rectifier to obtain fast braking. Winding is disconnected from D.C. supply after the motor stops.

(*ii*) ***Plugging and reversal :*** Except in case of shaded pole motor, *plugging and reversal is obtained by changing phase sequence by reversing polarity of one of the windings.*

7.3.7.5. Dynamic braking of a 3-phase induction motor

The speed of an induction motor can be controlled by injecting D.C. voltage in its stator winding. A variable resistance may be used in the rotor (in case of a slip ring induction motor) for dissipating the required amount of power. Now-a-days *thyristor bridges* are used for supplying D.C. which is controllable in nature. With the help of controlled D.C. from a thyristor bridge the dynamic braking can be achieved in a more effective manner. The connection diagram for scheme is shown in Fig. 7.93.

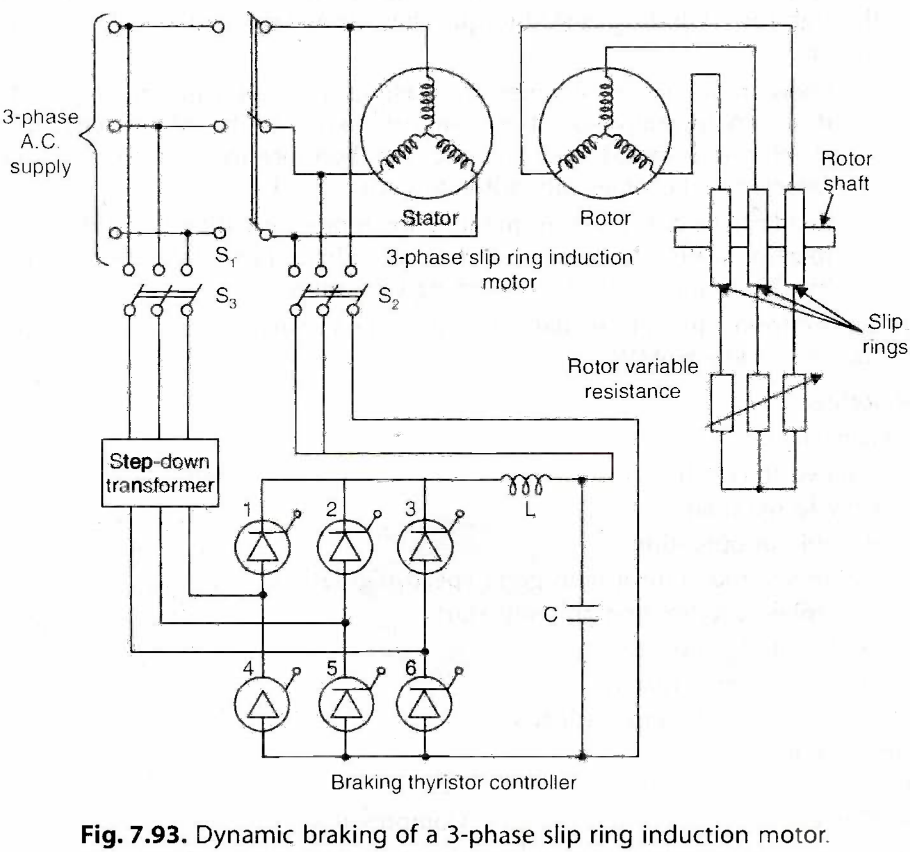


Fig. 7.93. Dynamic braking of a 3-phase slip ring induction motor.

- 3-phase A.C. is stepped down to lower voltage and fed to a 3-phase thyristor bridge which serves as the rectifier.
- This D.C. is filtered by an L.C. filter for minimising the ripples.
- Ripple free D.C. is then fed to the stator winding of the induction motor as shown in Fig. 7.93.

It is to be noted that while feeding D.C. to the stator the *3-phase A.C. input must be disconnected.* A.C. is disconnected with the help of S_1 and D.C. is disconnected with the help of S_2. Since, the input A.C. voltage is stepped down to a lower value, the thyristor converter may be of lower voltage rating.

7.3.7.6. Eddy current drives

- An eddy current drive consists of an *eddy current clutch placed between an induction motor running at a fixed speed and the variable speed load.* Speed is controlled by controlling D.C. excitation to magnetic circuit of the clutch. Since motor itself runs at a fixed speed it can be fed directly from A.C. mains.
- The principle of an eddy-current clutch is identical to an induction motor in which both stator and rotor are allowed to rotate. Stator, which is coupled to driving induction motor, has D.C. winding which produces magnetic field rotating at the speed of rotor. Rotor has a metal drum coupled to the load. Eddy currents are induced in rotor drum by stator magnetic field. Interaction between the stator field and eddy currents produces a torque which causes rotor to move with a slip. Slip, and therefore, the load speed, can be controlled by controlling D.C. current through rotor winding. Speed torque characteristics are identical to an induction motor.
  — Speed reduction is obtained by wasting a power equal to sP_{in} in the rotor drum. Minimum speed is usually restricted to 30 percent below the synchronous speed, because efficiency becomes too low and cooling of the rotor drum becomes difficult below this speed.
  — Load can be decoupled from induction motor by setting D.C. winding current to zero. Motor can now be started on no load. Load can be smoothly started by slowly increasing D.C. winding excitation.
- Eddy current clutch is available in different constructions and sizes ranging from fraction of kW to MW.

Advantages :

The *advantages of eddy current drives* are :

(*i*) Rugged in construction.
(*ii*) Easy to maintain.
(*iii*) Reliable in operation.
(*iv*) Stepless speed control with good speed regulation.
(*v*) Controlled acceleration and soft start.
(*vi*) High starting torque.
(*vii*) High overload capacity.
(*viii*) Ability to handle impact loads.

Applications :

They are widely used in :

— Blowers ; — Compressors ;
— Conveyors ; — Cranes ;

— Dredges ;
— Elevators ;
— Winders ;
— Line shafts and paper machines.

However, due to *poor efficiency and cooling,* they are *rarely* used in new installations.

7.3.8. Synchronous motor—Types, starting, speed control and braking

7.3.8.1. Types of synchronous motors

The following types of synchronous motors are commonly used :

1. Wound field synchronous motors.
2. Permanent magnet (PM) synchronous motors.
3. Synchronous reluctance motors.
4. Hysteresis synchronous motors.
   — All these motors have a stator with a 3-phase winding, which is connected to an A.C. source.
   — Fractional horse power synchronous reluctance and hysteresis motors employ a 1-phase stator.

1. Wound field motors :

Wound field synchronous motor rotor has a D.C. field winding, which is supplied from a D.C. source through slip-rings and brushes. The rotor can have cylindrical or salient pole construction.

— Cylindrical rotors have *higher mechanical strength* and are *employed in high power and high speed applications.*

— Salient pole motors, due to low cost, are preferred for other applications.

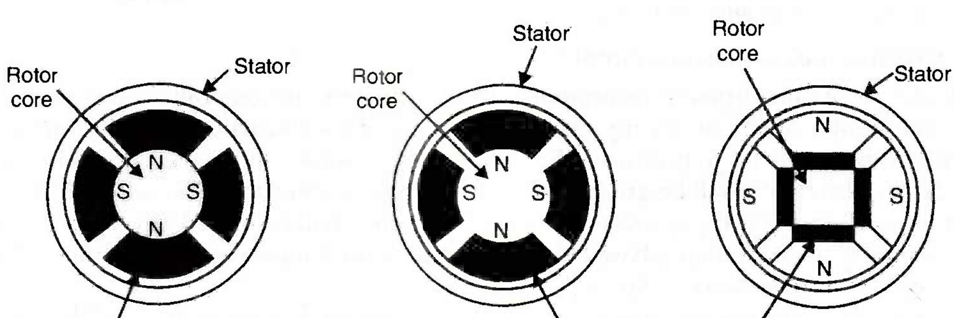
2. Permanent magnet (PM) synchronous motors :

In medium and small size motors, D.C. field can be produced by permanent magnets; thus dispensing with D.C. source, slip rings, brushes and field winding losses. Such motors are known as *permanent magnet (PM) synchronous motors.* Usually ferrite magnets are used. Rare earth (cobalt-samarium) magnets, although very expensive, are sometimes used to *reduce* the volume and weight of the motor.

PM synchronous motors are *classified* as follows :

(i) Surface mounted :

(*a*) *Projecting type.* In such motors magnets project from the surface of the rotor [Fig. 7.94(*i*) (*a*)].



(a) Projecting surface mounting (b) Inset magnet

(i) Surface mounted motor (ii) Interior or buried magnet motors

Fig. 7.94. Types of PM synchronous motors.

(*b*) *Inset type.* In this case, magnets are inserted into the rotor, providing a smooth rotor surface [Fig. 7.94(*i*) (*b*)]

- Epoxy glue is used to fix the magnets to the rotor surface in both.
- While these motors are *easy to construct and are less expensive, they are less robust compared to interior type rotors and are not suitable for high speed applications.*

(*ii*) *Interior (or buried)* : In these motors, magnets are imbedded in the interior of the rotor [Fig. 7.94(*ii*)].

Features of wound field and permanent magnet synchronous motors :

The wound field and PM synchronous motors have a *higher full load efficiency and power factor than an induction motor.*

Wound field motors can be designed for a higher power rating than induction motors. Since the air-gap flux is not produced solely by the magnetising current drawn from the armature, a larger air-gap suiting the mechanical design can be chosen. The ability to control power factor is an important advantage at higher power levels. *Operating at unity power factor minimizes the inverter rating.*

PM synchronous motor, apart from the *robust construction, has low losses and high efficiency.* Because of low losses, it is possible to make motors with very high power density and torque to inertia ratio. These make them *suitable for servo drives requiring fastest possible dynamic response.*

- One *significant difference* between the wound field and PM motors, which are designed to operate with a source of fixed frequency is discussed below :

When a wound field motor is started as induction motor, D.C. field is kept off. In case of a PM motor, the field *cannot* be 'turned off'. When at a speed below synchronous speed, the rotor field induces a voltage in the stator, which has a frequency different than the frequency of stator supply. The current produced by induced voltage interacts with the rotor field to produce a braking torque, which opposes the induction motor torque due to damper winding. The permanent magnet synchronous motor (PMSM) is designed so that the braking torque is very small compared to induction motor torque. Owing to the capability of starting direct on line these motors are called *line start PMSM.*

— PMSM are available in 3-phase and 1-phase construction.

— Although expensive to induction motors, they have advantages of *high efficiency, high power factor and low sensitivity to supply voltage variations.*

— These motors are preferred for industrial applications with *large duty cycles such as pumps, fans and compressors.*

3. Synchronous reluctance motor :

Single-phase salient-pole synchronous-induction motors, are generally called *reluctance motors.* If the rotor of any uniformly distributed single-phase induction motor is altered so that the laminations tend to produce *salient rotor poles,* as shown in Fig. 7.95, the reluctance of the air-gap flux path will be greater where there are no conductors embedded in slots. Such a motor, coming up to speed as an induction motor, will be pulled into synchronism with the pulsating A.C. single-phase field by the reluctance torque developed at the salient iron poles which have lower-reluctance air gaps.

Working of a reluctance motor. In order to understand the working of such a motor the basic fact which must be kept in mind is that *when a piece of magnetic material is located in a magnetic field, a force acts on the material, tending to bring it into the densest portion of the field. The force tends to align the specimen of material in such a way that the reluctance of the magnetic path that passes through the material will be minimum.*

path that passes through the material will be minimum.

When supply is given to the stator winding, the revolving magnetic field will exert reluctance torque on the unsymmetrical rotor tending to align the salient pole axis of the rotor with the axis of the revolving magnetic field (because in this position, the reluctance of the magnetic path would be minimum). If the reluctance torque is sufficient to start the motor and its load, the rotor will pull into step with the revolving field and continue to run at the speed of the revolving field. (Actually the motor starts as an induction motor and after it has reached its maximum speed as an induction motor, the reluctance torque pulls its rotor into step with the revolving field so that the motor now runs as synchronous motor by virtue of its saliency).

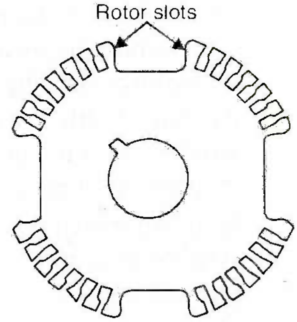


Fig. 7.95. Reluctance-motor lamination.

Reluctance motors have approximately *one-third* the horsepower rating they would have as induction motors with cylindrical rotors, although the ratio may be increased to one-half by proper design of the field windings. *Power factor and efficiency are poorer than for the equivalent induction motor.* Reluctance motors are subject to *'cogging'*, since, the locked-rotor torque varies with the rotor position, but the effect may be *minimized* by *skewing the rotor bars* and by *not* having the number of rotor slots exactly equal to an exact multiple of the number of poles.

Uses. Despite its shortcomings, the reluctance motor is widely used for many *constant speed* applications such as *recording instruments, time devices, control apparatus, regulators,* and *phonograph turntables.*

- *Reversing* is obtained as in any single-phase induction motor.

Speed-torque characteristics. Fig. 7.96 shows speed-torque characteristics of a typical single-phase reluctance motor.

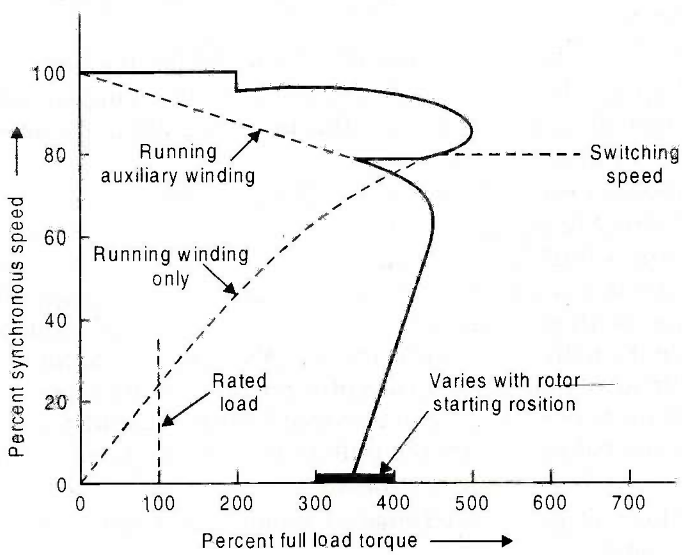


Fig. 7.96. Speed-torque characteristics of a single-phase reluctance motor.

- The motor starts at anywhere from 300 to 400 per cent of its full-load torque (depending on the rotor position of the unsymmetrical rotor with respect to the field windings) as a two-phase motor as a result of the magnetic rotating field created by a starting and running winding (displaced) 90° in both space and time.

- At about 3/4th of the synchronous speed, a centrifugal switch opens the starting winding, and the motor continues to develop a single-phase torque produced by its running winding only.

  As it approaches synchronous speed, the reluctance torque (developed as a synchronous motor) is sufficient to pull the rotor into synchronism with the pulsating single-phase field).

- *The motor operates at a constant speed up to a little over 200% of its full-load torque.* If it is loaded beyond the value of pull-out torque, it will continue to operate as a single-phase induction motor up to 500% of its rated output.

7.3.8.2. Starting of synchronous motor

The purpose of starting method is to bring rotor speed close to synchronous speed. Following methods are used to start synchronous motor :

1. Using damper windings as a squirrel-cage induction motor.
2. Using a low power auxiliary motor.

1. Using damper windings as a squirrel cage induction motor :

One widely used method is to start the synchronous motor as an induction motor with field unexcited and damper winding serving as a squirrel-cage rotor. Regarding this method, following points are worth noting :

(*i*) The currents and starting torque can be reduced and increased respectively, by increasing the damper winding resistance. The motor speed while running as an induction motor, for successful pull-in, must be close to synchronous speed. For this the damper winding resistance must be low. Further, for damping hunting oscillation damper winding resistance must be low. The damper winding resistance is so selected as to strike a compromise between these two contradictory requirements.

(*ii*) D.C. field should be applied only after the motor has reached close to full speed.

(*iii*) When the rotor has salient pole construction, the damper winding can have conductors only over the pole arc. This leads to a dip in the speed-torque curve at half of synchronous speed.

(*iv*) On the application of full supply voltage, the starting current in the motor can be 7 to 10 times of full load value. Except in small size motors, such a high starting current causes fluctuations in supply voltage. In case of large size motors, such a high current may cause a large drop in the terminal voltage, thus reducing the already low starting torque further. Starting current can be reduced by employing any one of the reduced voltage starting methods employed for starting induction motors. Reduction in starting current is obtained at the expense of reduction in starting torque. When started at a reduced voltage, the transition to full voltage can be made before or after the pull-in. *Former is preferred as it improves pull-in performance due to following two reasons :*

— With full voltage the speed attained as induction motor is closer to synchronous speed, and

— The pull-in torque increases in proportion to voltage squared, consequently pull-in can be achieved faster and with large motor loads.

2. Using a low power auxiliary motor :

(*i*) In this method a low power auxiliary motor is coupled to the synchronous motor shaft. With the help of auxiliary motor, the rotor speed is brought near synchronous speed and then D.C. field is switched-in.

(*ii*) This method has a very *low starting torque*.

Note: It is practically impossible to start a synchronous motor with its D.C. field energized. Even when left de-energized, the rapidly rotating magnetic field of the stator will induce extremely high voltages in the many turns of the field winding. It is customary, therefore, *to short-circuit the D.C. field winding during the starting period* ; whatever voltage and current are induced in it may then aid in producing induction motor action. In very large synchronous motors, field sectionalising or field-splitting switches are used which short-circuit individual field windings to prevent cumulative addition of induced voltages from pole to pole.

7.3.8.3. Braking of synchronous motors

- The motor can work in regenerative braking only at synchronous speed. Therefore, *"regenerative braking" cannot be used for stopping or decelerating a load.*
- *"Dynamic braking"* is obtained by *disconnecting stator from the source and connecting it to 3-phase resistor.* Machine works as a synchronous generator and dissipates generated energy in the braking resistor.

7.3.8.4. Speed control of synchronous motors

The *speed* of synchronous motors can be *controlled* as follows :

(*i*) By using current-fed D.C. link.

(*ii*) By using cycloconverter.

1. Speed control by current-fed D.C. link :

Fig. 7.97, shows the circuit diagram for speed control of synchronous motor by current-fed D.C. link.

The typical circuit consists of three converters two of which are connected between the 3-phase source and synchronous motor and the third converter (bridge rectifier) supplies D.C. field excitation for the rotor.

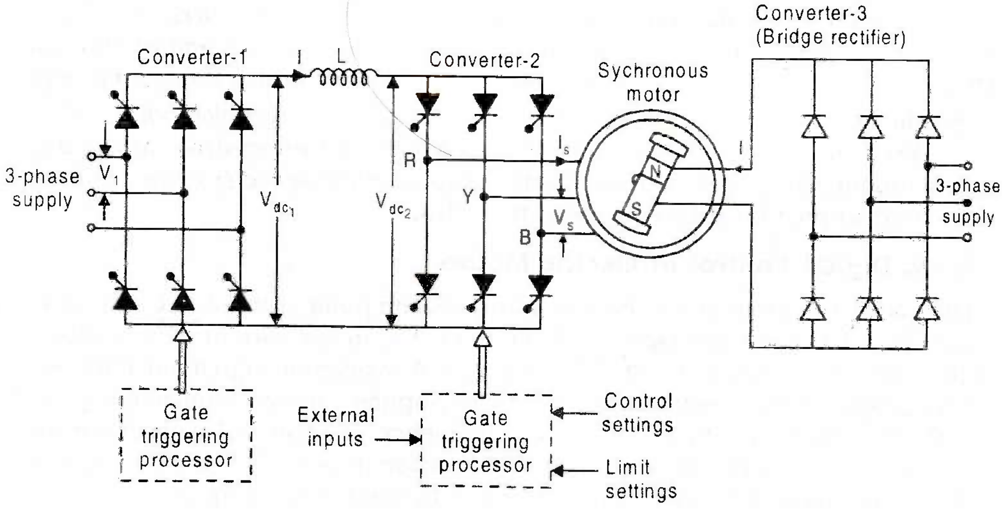


Fig. 7.97. Circuit diagram for speed control of synchronous motor by current-fed D.C. link.

— Converter-1 (C-1) acts as a *controlled rectifier* and feeds D.C. power to converter-2 (C-2). It acts as a current source and controls I.

— Converter-2 (C-2) behaves as a *naturally commutated inverter* whose A.C. voltage and frequency are established by the motor. The converter-2 is naturally commutated by voltage V_S induced across motor terminals by its revolving magnetic flux. The revolving flux which depends on the stator currents and the D.C. field exciting current is usually kept constant. Consequently V_S is proportional to motor speed.

— The function of the smoothing inductor L is to maintain a ripple-free current in the D.C. link between the two converters.

As regards various controls, information picked up from various points is processed in the gate-triggering processors which then send out appropriate gate firing pulses to converters 1 and 2. The processors receive information about the desired rotor speed, its actual speed, instantaneous rotor position, field current, stator voltage and current etc. The processors check whether these inputs represent normal or abnormal conditions and send appropriate gate firing pulses either to correct the situation or meet a specific demand.

Gate triggering of C-1 and C-2 is done at line frequency (50 Hz) and at motor frequency respectively. In fact, gate pulses of C-2 are controlled by rotor position which is pulsed by position transducers mounted at the end of the shaft. The speed of the motor can be increased by increasing either D.C. link I or exciting current I_f.

- This method of speed control is applied to *motors ranging from 1 kW to several MW.*
- *Permanent-magnet synchronous motors used in textile industry and brushless synchronous motors for nuclear reactor pumps are controlled by this method.*

2. Speed control and cycloconverter :

This arrangement consists of *three cycloconverters* connected to the three phases of the synchronous motor and *one controlled rectifier* for supplying field exciting current, I_f to the rotor. Each cycloconverter is composed of two 3-phase bridges and supplies a single-phase output. With a line frequency of 50 Hz, the cycloconverters output frequency can be varied from 0 to 10 Hz (It is well known that a cycloconverter can convert A.C. power at higher frequency to one at a lower frequency).

The cycloconverters and controlled rectifies function as *current sources.* The air-gap flux is kept constant by controlling the magnitude of the stator currents and exciting current I_f. The motor can be made to operate at unity power factor by proper timing of gate pulses.

- The *speed of cycloconverter-driven large slow-speed synchronous motors can be continuously varied from 0 to 15 r.p.m.* Such low speeds permit direct-drive of the ball mill without using a gear reducer. Such high-power low-speed systems are also being used as *propeller drives* on board the ships.

7.3.9. Digital Control of Electric Motors

The speed information can be fed into microcomputer using a D.C. Tacho (speed encoder) and *A/D* converter (speed *I/P* module). The motor current data is usually fed into the computer through a fast *A/D* converter. A synchronizing circuit interface (line synchronizing circuit) is required so that microcomputer can synchronize the generation of the firing pulse data with the supply line frequency. The gate pulse generator receives a firing signal from microcomputer. A set of instruction (Program) is stored in the memory and those are executed by computer for proper functioning of a drive.

Advantages of digital control :

1. High reliability.
2. Easy software control.
3. High precision and accuracy.
4. Better speed regulation.
5. Faster response.
6. Improved performance.
7. Economical.
8. Flexibility.
9. Better time response.
10. *The major advantage of the digital control is that by changing the program, desired control technique can be implemented without any change in the hardware.*

7.3.10. Selection of a motor for mechatronic applications

While selecting a motor for a specific mechatronic applications, the following factors/ specifications should be considered:

(*i*) Speed range.
(*ii*) Torque-speed variations.
(*iii*) Reversibility.
(*iv*) Operating duty cycle.
(*v*) Starting torque.
(*vi*) Power required.

Besides the above factors, the following points should also be considered:

— Will the motor start and will it accelerate fast enough?
— What is the maximum speed the motor can produce?
— What is the operating duty cycle?
— How much power does the load require?
— What is the load inertia?
— Is the load to be driven at constant speed?
— Is accurate position or speed control required?
— Is a transmission or gear box required?

7.4 HYDRAULIC ACTUATORS

7.4.1. General Aspects

An actuator wherein hydraulic energy is used to impart motion is called an **hydraulic actuator**.

A system wherein energy is imparted to oil and this hydraulic energy so imported to the oil is converted into mechanical energy is called an **hydraulic system**.

Advantages and disadvantages of hydraulic system:

Following are the advantages and disadvantages of a hydraulic system :

Advantages:

1. Easy to produce, transmit, store, regulate and control, maintain and transform the hydraulic power.
2. Possible to generate high gain in force and power amplification.
3. Hydraulic systems are uniform and smooth, generate stepless motion and variable speed and force to a greater accuracy.
4. Limiting and balancing of hydraulic forces are easily performed.
5. Weight-to-power ratio of an hydraulic system is comparatively less than that for

an electro-mechanical system.

6. Easy maintenance.
7. Hydraulic systems are cheaper if one considers the high efficiency of power transmission.
8. Hydraulics is mechanically safe, compact and adaptable to other forms of power and can be easily controlled.
9. Division and distribution of hydraulic power is simpler and easier than other forms of energy.
10. Frictional resistance is much less in an hydraulic system as compared to a mechanical movement.

Disadvantages:

1. The manufacturing cost of the system is high since the hydraulic elements have to be machined to a high degree of precision.
2. Hydraulic elements have to be specially treated to protect them against rust, corrosion, dirt, etc.
3. Petroleum based hydraulic oil may pose fire hazards thus limiting the upper level of working temperature.
4. Certain hydraulic systems are exposed to unfriendly climate and dirty atmosphere as in the case of mobile hydraulics like dumpers, loaders, etc.
5. Hydraulic power is not readily available compared to electric power.

7.4.2. Hydraulic Power Supply

Hydraulic systems are designed to move large loads by controlling a high-pressure fluid in distribution lines and pistons with mechanical or electromechanical valves.

7.4.2.1. Basic element of an oil hydraulic system

Fig. 7.98 shows the *element* of an hydraulic system.

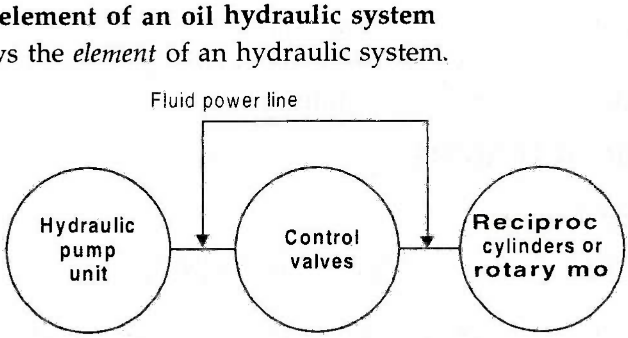


Fig. 7.98. Elements of an hydraulic system.

1. *Hydraulic pump unit :*

- In an actual hydraulic system a pump converts mechanical power into fluid power.
- The intake of the pump is connected to a liquid source usually called a tank or reservoir.

2. ***Control valves :***

- The flow of pressurised liquid discharges by the pump is controlled by valves :
  — *"Pressure control valves"* control the liquid *pressure.*
  — *"Flow control valves"* control the liquid *flow rate.*
  — *Directional control valves* control the *direction* of flow of the liquid.

3. ***Hydraulic motor/cylinder :***

- The liquid discharged by the pump is directed to hydraulic motors or cylinders by control valves.
- *Motors are used where rotary motion is desired and cylinders are used where linear motion is necessary.*

7.4.2.2. Components of an hydraulic system

Fig. 7.99, shows the various *components* of an hydraulic system; these are :

1. ***Pump.*** It delivers high pressure fluid.
2. ***Pressure regulator:*** It limits the pressure in the system.
3. ***Valves:*** These control flow rates and pressures.
4. ***Distribution system:*** It is composed of hoses or pipes.
5. ***Infrastructure:*** It consists of the elements contained in the dashed box in the figure and is typically used to power many hydraulic valve actuator subsystems.

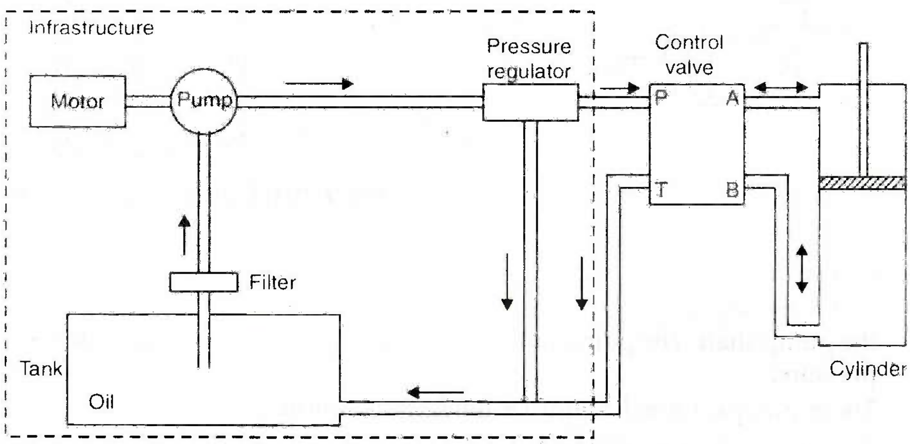


Fig. 7.99. Components of an hydraulic system.

7.4.3. Pumps

- An hydraulic pump is usually driven by an electric motor (*e.g.*, a large AC induction motor) or an internal combustion engine.
- Typical fluid pressures generated by pumps used in heavy equipment (*e.g.*, large industrial machines and construction equipment) are 6.9 MPa to 20.7 MPa range.
- The hydraulic fluid is selected to have the following *characteristics*:
  (*i*) Good lubrication to prevent wear in moving components.
  (*ii*) Corrosion resistance.
  (*iii*) Incompressibility to provide rapid response.

Classification of pumps :

There are two broad classification of pumps as identified by the fluid power industry:

1. ***Hydrodynamic or Non positive displacement pumps :*** Refer to Fig. 7.100.

Examples of these pumps are : Centrifugal and propeller pumps.

— Although these pumps provide smooth continuous flow, their flow output is reduced when the circuit resistance is increased. Since there is a great deal of clearance between the rotating and stationary elements, when the resistance of the external system starts to increase, some of the fluid slips back into the clearance spaces causing a reduction in the discharge flow rate. This slippage is due to the fact that the fluid follows the least resistance path. Thus the pump flow rate depends not only on the rotational speed but also on the resistance of the external system.

— These types of pumps are used for *low pressure, high volume flow applications.*

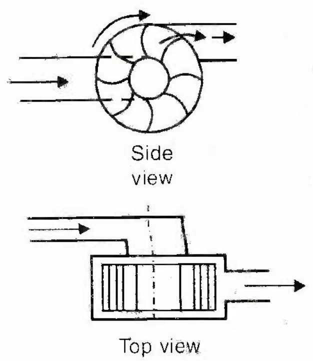


Fig. 7.100. Hydrodynamic or non-positive displacement pump.

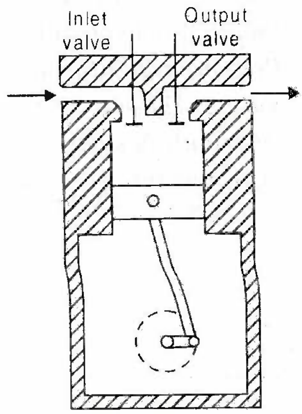


Fig. 7.101. Positive displacement pump.

2. ***Hydrostatic or positive displacement pumps***

— This type of pump *ejects a fixed quantity of fluid* (See Fig. 7.101) per revolution of the pump shaft. The pump outlet flow is constant and is not dependent on system pressure.

— These pumps are *well suited for fluid power system.*

Examples of these pumps are :

(*i*) Gear pumps.

(*ii*) Vane pumps.

(*iii*) Piston pumps.

(*i*) ***Gear pump :*** The gear pump unit (Fig. 7.102) consists of two identical intermeshing spur gears with involute teeth. One of the gears is keyed to the driving shaft of the motor and the other gear revolves *idly*. These gears rotate in opposite directions in a closely fitting stationary *housing*.

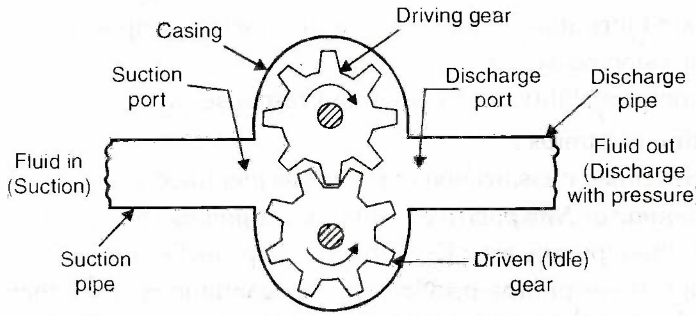


Fig. 7.102. Spur gear pump.

Although the gear pump, which consists of two gears, is a rotating machine, yet its action on the liquid to be pumped is *not dynamic* and it *merely displaces the liquid to be pumped, is continuous and uniform and there is no change of velocity and acceleration under normal stable conditions.*

Working Principle. Referring to Fig. 7.102, the oil coming in at the suction port fills the space between the teeth, is carried around the periphery of the revolving gears and is finally pushed out to the discharge port. The teeth of the gears have a perfect meshing and that serves both to transmit the drive and to maintain a seal between the suction and discharge side. Care is taken to ensure that the oil trapped between the successive lines of contact does not built up pressure.

— For good working it is necessary to have teeth made precise and further they should have hard surface. Above all the whole gear and casing should have good surface and accurate dimensions. The casing should have, in addition, a packing to prevent leakage. For highly viscous liquids the casing is provided with a heating jacket to reduce viscosity.

The direction of flow can be changed by reversing the direction of gear assembly. But the spur gear pump delivers hydraulic fluid always at right angles to the axis of rotation.

Let, A = Area enclosed between two adjacent teeth and casing,

L = Axial length of teeth,

T_g = Number of teeth in each gear wheel, and

N = Speed in r.p.m.

Then, volume of liquid discharged in one revolution,

$$q = 2LAT_g$$

∴ Theoretical volume displaced by the pump per second,

$$Q_{th} = (2LAT_g) \times \frac{N}{60} \quad ...(7.7)$$

Normally the area between the adjacent teeth is larger than the cross-sectional area of the meshing tooth, and that causes some liquid to flow back to the low pressure side. Also, there is some leakage because of clearance between the casing and gear wheel. Thus if η_v is the volumetric efficiency of the pump, then

Actual discharge, $Q_{actual} = Q_{th} \times \eta_v$

or

$$Q_{actual} = (2LAT_g) \times \frac{N}{60} \times \eta_v \quad ...(7.8)$$

If it is not possible to determine easily the area enclosed between the adjacent teeth and casting, then volumetric displacement of the pump per *revolution* is calculated by using the following emperical relation.

$$q = 0.95\,\pi c\,(D - c)t \quad ...(7.9)$$

where, D = Outside diameter of gears, and

c = Centre-to-centre distance between the axes of gears.

High speed pumps can produce a suction of about 7 m. The suction pipe is often connected directly to the casing and avoids stuffing box. This also fixes the direction of rotation of gears.

Normally gear pumps are expected to work against small heads of a few atmospheres. However, pumps and that also in one stage producing pressure upto 100 atm. absolute have been manufactured.

Gears have generally involute teeth but the helical teeth are also used, which provide better grip over the gear width, and also give smooth running and uniform flows. They are, however, comparatively difficult to manufacture and hence are costly. Therefore, their use in limited to high discharge pumps.

Applications. This type of pump is widely used for cooling water and pressure oil to be supplied for *lubrication to motors, turbines, machine tools etc.*

(*ii*) ***Vane pump:*** Fig. 7.103, shows the schematics of a vane pump.

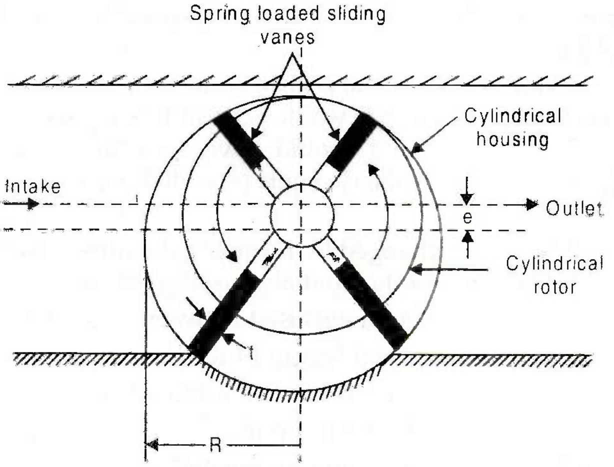


Fig. 7.103. Vane pump.

Construction :

It consists of a cylindrical rotor which is mounted *eccentrically* in relation to a cylindrical housing.

— The motor has radial slots into which are inserted the *sliding vanes*, which are spring loaded. This arrangement provides the required seal between suction and discharge connections.

Working/Operation :

— As the rotor rotates, the vanes undergo free to-and-fro sliding movement in the slots. On suction side, the pocket (space) between the vanes tend to increase in volume and the space gets filled with liquid.

— After the point maximum distance between the rotor and casing has passed, the space opens to delivery and the liquid is discharged. The quantity of liquid pumped and the flow direction can be controlled by affecting a change in the degree of eccentricity.

Let,
b = Width of vane,
t = Thickness of vane,
n = Number of vanes,
R = Inner radius of casing,
e = Eccentricity between the rotor and casing, and
N = Speed in r.p.m.

Then, the volume of liquid discharged in one revolution,

$$q = 2eb\,[2\pi\,(R - e) - nt]$$

∴ Theoretical volume of liquid displaced by the pump per second,

$$Q_{th} = 2eb\,[2\pi(R-e) - nt] \times \frac{N}{60} \qquad ...(7.10)$$

A *single stage* vane pump can develop about 15 to 65 atm. In order to get higher pressures, more than one stage may be used.

(iii) Piston pump: Fig. 7.104 shows a *swash plate piston pump.*

— The cylinder block is rotated by the input shaft, and the piston ends are driven in and out as they side in the fixed *swash plate* slot, which is angled with respect to the shaft axis.

— A piston draws fluid from an inlet manifold over half the swash plate and expels the fluid into the outlet manifold during the other half.

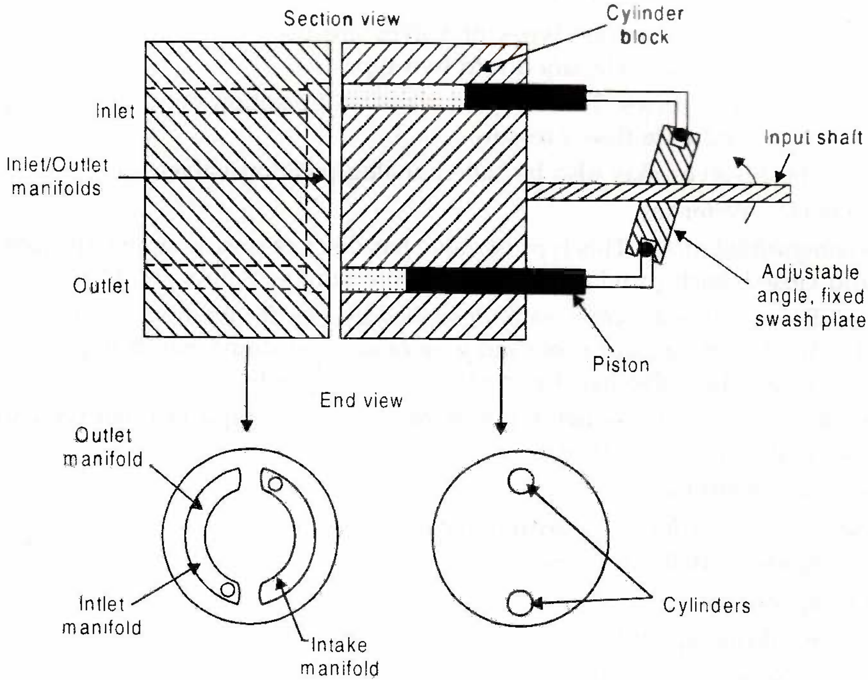


Fig. 7.104. Swash plate piston pump.

— The pump displacement can be altered simply by changing the angle of the fixed swash plate.

— Comparison of characteristics of various pumps are given below :

S. No.	*Aspects*	*Gear pump*	*Vane pump*	*Piston pump*
1.	Displacement	Fixed	Variable	Variable
2.	Cost	Low	Medium	High
3.	Typical pressure	140 MPa	270 MPa	420 MPa

7.4.4. Pressure Regulator

*A **pressure regulator** is a pressure relief valve which is provided in positive displacement pumps (which provide a fixed volumetric flow) to prevent the pressure from exceeding design limits.*

Fig. 7.105, shows the simplest pressure regulator—*spring ball arrangement* :

— When the pressure force exceeds the spring force, fluid is vented back to the tank, preventing a further increase in pressure.

— The *cranking pressure* (threshold pressure) is usually adjusted by altering the compressed length of the spring *i.e.*, its resisting force.

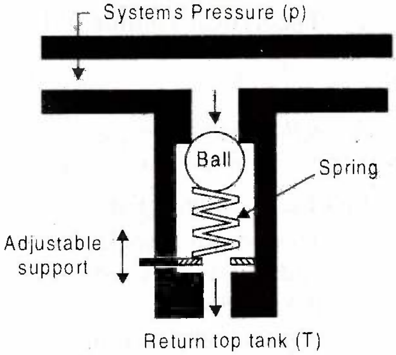


Fig. 7.105. Pressure regulator.

7.4.5. Hydraulic Valves

***Hydraulic valves** are mechanical devices used for controlling hydraulic parameters in hydraulic systems:*

7.4.5.1. Classification of valves

In hydraulic system various types of valves are used to control or regulate flow medium, but can be *broadly* classified into two types:

1. *Infinite positive valves.* These valves allow any position between fully open and closed to modulate flow pressure.

   — These valves may also be called analog position valves.

   Examples: Water tap.

2. *Finite position valves.* This type of valve has discrete positions, usually just opened and closed, each providing a different pressure and flow condition.

   — These valves are commonly described by x/y designation, where x represents number of paths or ports and y number of positions; For example a 4/3 valve means the valve has 4 ports/ways and 3 positions.

An analogy between these two types of valves is the comparison between an *electric light dimmer* and a *simple on/off switch.*

Other classifications :

I. *Based on the method of controlling the flow medium :*

1. Seating—Ball, plate, cone.
2. Spool type :
   - Sliding spool valve
   - Rotary spool valve.

II. *Based on function and hydraulic quantity controlled :*

1. Pressure control valve—To control pressure of flow medium.
2. Flow control valve—To control quantity of flow medium.
3. Direction control valves—To control direction of flow medium.

- In mechatronics and hydro-mechanical systems control of pressure, flow and direction are of significant importance.

7.4.5.2. Graphic valve symbols

A valve is represented by a *square* for each of its switching positions.

Refer to Fig. 7.106.

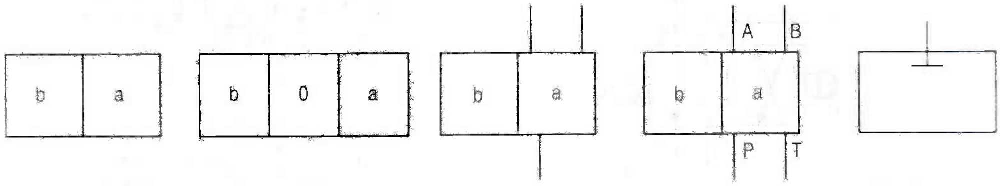
— Fig. 7.106(*i*), shows the symbol of a two position valve.

— Fig. 7.106(*ii*), shows a three position valve. Valve position can be represented by letters *a*, *b*, *c* and so on, with *o* being used for a central neutral position.

— Fig.. 7.106(*iii*), shows two position valve with three ports. Parts of a valve are shown on the outside of the boxes in normal, *unoperated* or *initial position.*

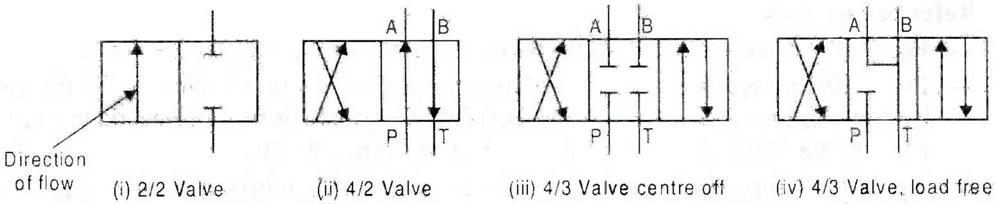
— Fig. 7.106 (*iv*), shows two position valve in the four ports.

— Fig. 7.106(*v*), shows closing of port (s).



(i) Two position valve (ii) Three position valve (ii) 3/2 Valve (iv) 4/2 Valve (v) Closing of port (S)

Fig. 7.106. Basis of graphic symbols.



(i) 2/2 Valve (ii) 4/2 Valve (iii) 4/3 Valve centre off (load isolated) (iv) 4/3 Valve, load free in centre

Fig. 7.107. Valve symbols.

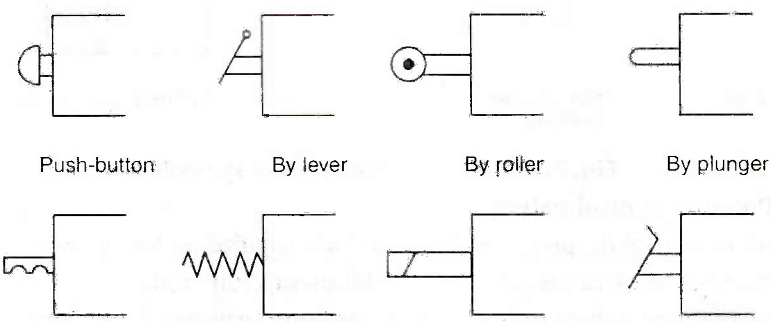
- Fig. 7.107 shows valves symbols for 2/2 valve, 4/2 valve and 4/3 valve centre off (load isolated) and 4/3 valve, load free in centre respectively.

Following ***designations*** are normally given to the ports:

Working links = *A*, *B*, *C* and so on; *Pressure* (*power*) *supply* = *P*;

Exhaust return = *R*, *S*, *T* and so on (*T* normally used for hydraulic systems, *R* and *S* for pneumatic systems).

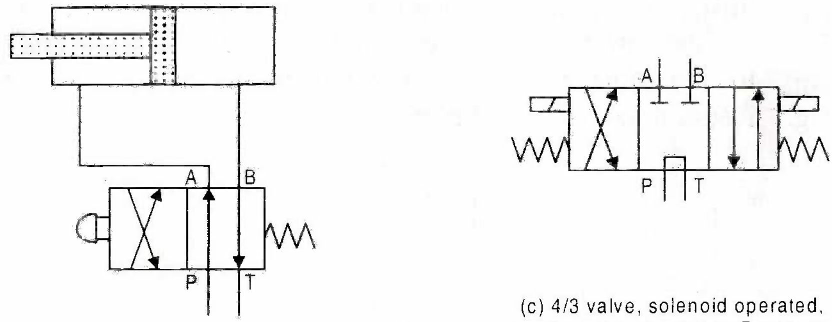
— Arrow-heated lines represent direction of flow. In Fig. 7.107(*ii*), for example, fluid is delivered form port *P* to port *A* and returned from port *B* to port *T*, when valve is in its normal state *a*. In state *b* the flow is reversed. Shut off position are represented by ⊥ as shown by the central position of the valve in Fig. 7.107(*iii*), internal flow paths can be represented as shown in Fig. 7.100(*iv*).



Push-button By lever By roller By plunger

Detente (holds position) By spring By solenoid By pedal

(a) Actuation symbols,



(b) 4/2 valve. Pushbutton extend, spring retract when pushbutton released.

(c) 4/3 valve, solenoid operated, spring return to centre. Pressure line unloads to tank and load locked in centre position.

Fig. 7.108. Complete valve symbols.

Refer to Fig. 7.108.

— Fig. 7.108(*a*) shows the various ways in which valves can be operated.

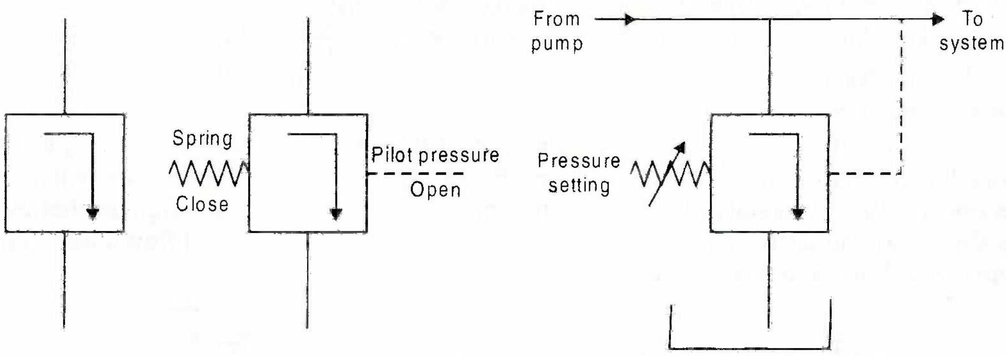
— Fig. 7.108(*b*) represents a 4/2 valve operated by a push button, with the push button depressed the ram extends. With the push button released, the spring pushes the valve to earlier position and the ram retracts.

— Fig. 7.108(*c*) represents a solenoid-operated 4/3 valve, with spring return to centre.

Refer to Fig. 7.109.

This figure shows the "*infinite position valve symbols*".

— A basic valve is represented by a single square, with the valve being shown in normal, or non-operated, position.



(a) Infinite position valve

(b) With actuation symbols

(c) Pressure relief valve

Fig. 7.109. Infinite position valve symbols.

7.4.5.3. Pressure control valves

These valves *control the pressure of flow medium required by the system.*

The pressure control valves perform the following functions :

(*i*) To regulate or reduce oil pressure in certain portions of the circuit.

(*ii*) To unload system pressure.

(*iii*) To limit maximum system pressure as a safety measure.

(*iv*) To assist sequential operation of actuators in a circuit by pressure control.

(*v*) To perform any other pressure related functions by virtue of pressure control.

These valves are classified based on (*i*) the primary functions they have to perform, (*ii*) type of connection; (*iii*) size and pressure operating range.

Based on their primary functions, the pressure control valves are ***classified*** as follows:

1. Pressure relief valve.
2. Pressure sequencing valve.
3. Pressure reducing or regulating valve.
4. Pressure unloading valve.
5. Pressure brake valve.

The operation of a pressure control valve is based mainly on balance between pressure and a mechanical load e.g. a spring force biased against the oil pressure. The valve can assume various positions between fully closed and fully open conditions depending on the flow and pressure differential.

1. Pressure relief valves:

These valves are found in every hydraulic system. It is a normally closed valve connected between the pressure line and the oil reservoir. Its main purpose is to *limit the pressure in a system to a prescribed maximum by diverting some or all of the pump output to the tanks, when the desired set pressure is reached.*

The pressure relief valves are of the following two types : (*i*) Direct acting relief valve; (*ii*) Pilot operated relief valve.

(*i*) ***Direct acting relief valve.*** For details refer to Art. 7.4.4. Fig. 7.110 shows the graphic symbol of this type of valve.

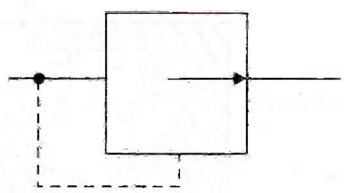


Fig. 7.110. Graphic symbol.

Fig. 7.111 shows the application of pressure relief valve in a hydraulic system containing accumulator as one of the elements.

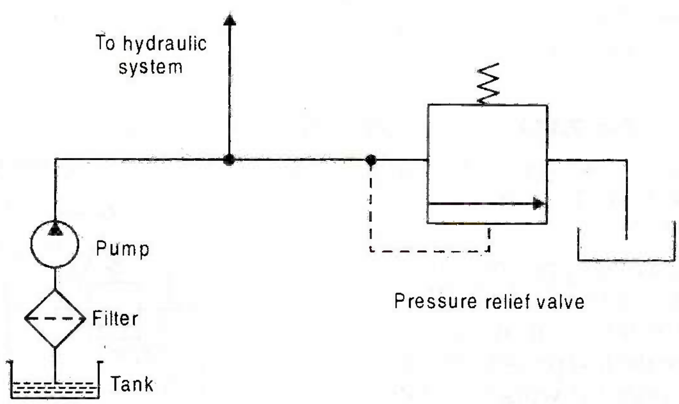


Fig. 7.111. Application of pressure relief valve.

(*ii*) ***Pilot operated relief valve.*** Direct controlled pressure relief valves are used where the flow rate and the system pressure are *reasonably smaller and there is not much variation in system pressure or flow rate.*

The use of pilot operated valve, however, is *most common for a larger flow rate and higher pressure.*

The main *advantage* of such a valve is that here the pilot valve can be kept spatially separated from the main valve such as on a control panel and can introduce suitable D.C. valves in between to set different pressures by pilot valves.

2. Pressure sequencing valve :

Sequencing valves extensively used in hydraulic systems are also pressure control valves.

A pressure sequence valve is used in hydraulic system *to cause various operations in a sequential order, i.e., one after the other.* For example a pressure sequence valve used in *clamping and machining circuit* may permit the clamping operation to take place first and when clamping cylinder is fully extended, the machining cylinder is actuated.

Fig. 7.112 shows the pilot operated sequence valve. Here the required sequential pressure can be adjusted manually.

— In this valve, the fluid flows freely through the primary passage to operate the first phase until the pressure setting of the sequence valve is reached.

— As the spool lifts, flow is diverted to the secondary port to operate the second phase.

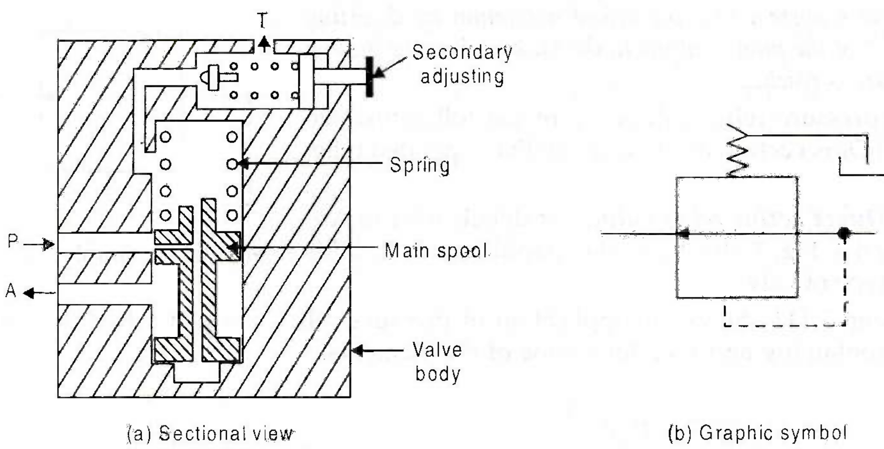


Fig. 7.112. Pilot-operated pressure sequence valve.

- Fig. 7.113 shows a typical example where a workpiece is pushed into position by cylinder-1 and clamped by cylinder-2.

— Sequence valve V_2 is connected to the extend line of cylinder-1. When this cylinder is moving the workpiece the line pressure is low, but rises once the workpiece hits the end stop.

— The sequence valve opens once its inlet pressure rises above a preset value. Cylinder-2 then operates to clamp the workpiece.

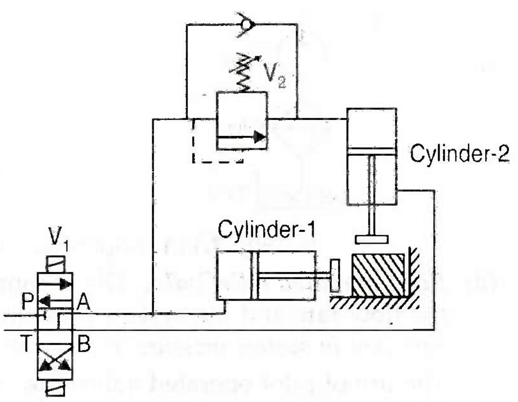


Fig. 7.113. Application of sequence valve.

A check valve across V_2 allows both cylinders to retract together.

3. Pressure reducing valve :

— The pressure reducing valves are normally open pressure control valves used *to maintain reduced pressures in certain portions of the hydraulic system.*

— There are actuated by the pressure sensed in the branch circuit and tend to close as it reaches the pressure of the valve setting preventing farther build-up of pressure.

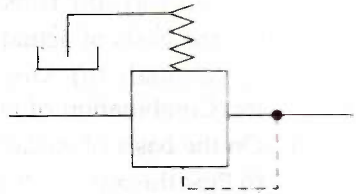


Fig. 7.114. Graphic symbol.

4. Pressure unloading valve:

This type of valve is used to *unload the energy in a system of a lower pressure* (whereas a pressure relief valve requires full system pressure to open thereby causing higher quantity of energy loss due to heat).

— These valves are employed in systems where two pumps provide a large volume of oil at a low pressure and one of them must be unloaded during a specific period requiring only a small volume of oil at a high pressure. This will save the system from undesired heat energy to a great extent.

7.4.5.4. Flow control valves (variable orifice)

— These valves are used to control the speed of hydraulic actuator by controlling the flow rate or discharge.

— The principle of flow valve is based on the flow rate or discharge and the velocity of the actuator.

Needle valves :

— It is the most common hydraulic flow-control device.

— It consists of a needle or pointed threaded stem that can be adjusted manually to control the flow or discharge through the valve. It is made of steel.

— This valve can also be used as a stop valve to prevent the flow of fluid from one part of the hydraulic circuit to another.

Glode valve :

— In this valve the flow area is larger than that of a needle valve.

— These valves are not suitable for throttling but can be used as a stop valves. Normally, they are not used as flow control valves.

7.4.5.5. Direction control valves

As the name implies, *the direction control valves start, stop and control the direction of flow for reversing the direction of motion of the actuator.*

- Direction control valves are employed in a hydraulic system to *determine the direction of the fluid* in the hydraulic circuit. Sometimes they are also used as a *selector switch.*

Direct control (DC) valves may be *classified* as follows:

1. On the basic of *internal valving element :*
   (*i*) Poppet (Ball or piston)
   (*ii*) Spool valve.
2. On the basis of *flow paths :*

(*i*) Two way; (*ii*) Three way; (*iii*) Four way.

3. On the basis of actuation of *internal valving element* :

(*i*) Manual; (*ii*) Mechanical; (*iii*) Electrical; (*iv*) Hydraulic; (*v*) Pneumatic; (*vi*) Combination of any of these.

4. On the basis of *method of connection:*

(*i*) Pipe thread; (*ii*) Straight thread; (*iii*) Flanged or subplot; (*iv*) Manifold mounted.

5. On the basis of *size.*

Here follows the description of finite positioning directional valves which direct the fluid by opening or closing flow paths in definite valve position. Each finite position is represented graphically by a square and the arrows inside showing the path.

1. Check valve :

The check valves only allow flow in one direction and, as such, are similar in operation to *electronic diodes.*

(*i*) *Simple check valve.* It is simplest in construction and consists of a ball and seat arrangement of the valve as shown in Fig. 7.115(*a*). It is commonly used in pneumatic systems.

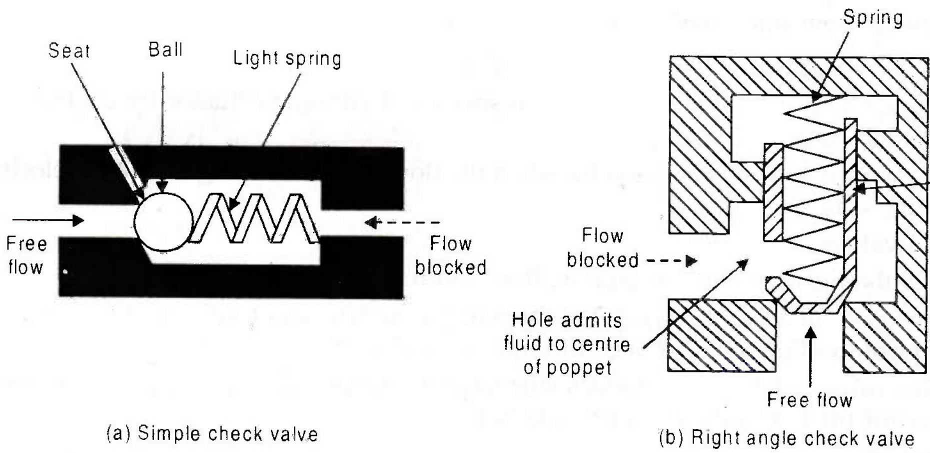


Fig. 7.115. Check valves.

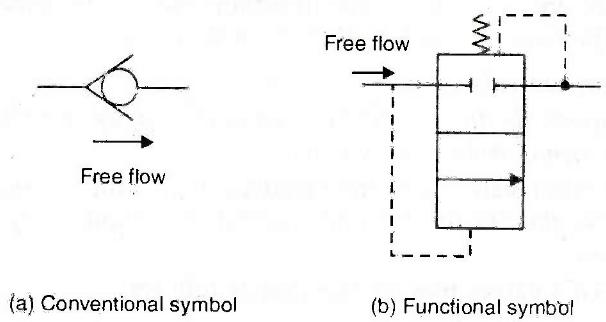


Fig. 7.116. Check valve symbols.

(*ii*) *Right angle check valve.* Its construction is shown in Fig. 7.115(*b*), it is better used to the higher pressures of a hydraulic system.

— A *pilot-operated check valve* is similar to a basic check valve but can be held open permanently by application of an external pilot pressure signal.

Fig. 7.117 shows the *application of a check valve.* In the figure is shown an hydraulic circuit with a pressure storage device called an *accumulator* (An accumulator, besides being a storage tank, supplies oil to the hydraulic actuator at constant pressure depending on the requirement. Here a check valve allow the pump to unload via the pressure regulating valve, while still maintaining system pressure.

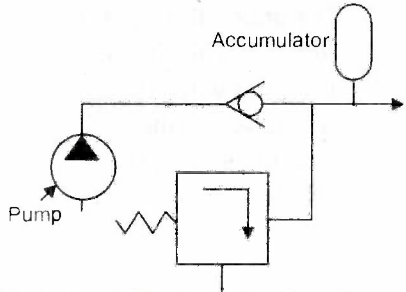


Fig. 7.117. Application of a check valve.

2. Poppet valve :

Fig. 7.118 shows a poppet valve. It is a *check valve that can be forced open to allow reverse flow.*

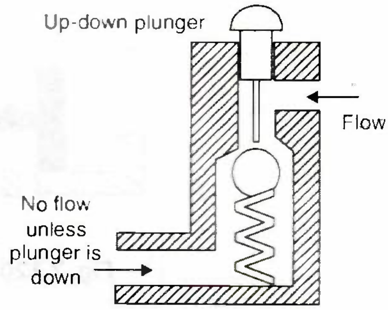


Fig. 7.118. Poppet valve.

3. Spool valves :

Fig. 7.119, shows the schematic of a spool valve:

— It consists of a cylindrical spool with multiple lobes moving within a cylindrical casing containing multiple ports.

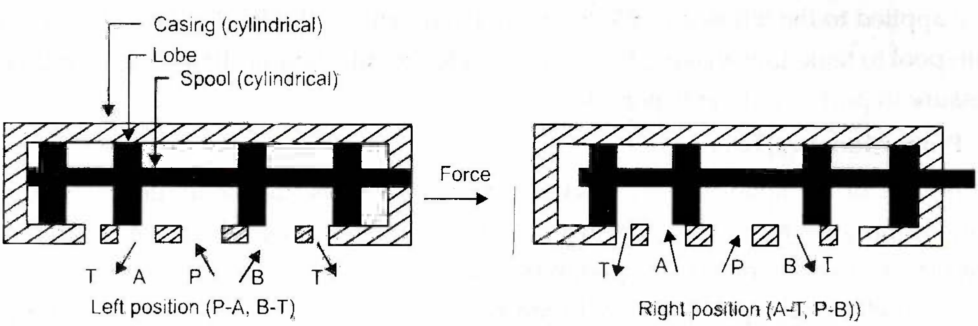


Fig. 7.119. Schematic of spool valve.

— The spool can be moved back and forth to align spaces between the spool lobes with input and output ports in the housing to direct high-pressure flow to different conduits in the system. The static pressure force on the spool is balanced since the pressure is always applied to opposing internal faces of the lobes.

— To move the spool, an axial force (from a solenoid or manual control lever) is required to overcome the hydrodynamic forces associated with changing the momentum of flow.

In the *left position*, port-*A* is pressurised and port-*B* is vented to the tank. A force is required to move the spool from this position to the *right position*, where port-*B* is pressurised and port-*A* is vented.

Pilot-operated spool valve :

Where large hydrodynamic forces occur, a *pilot valve* is added to the design of the spool valve, as shown in Fig. 7.120.

— The pilot valve operates at a *lower pressure*, called *"pilot pressure", and much lower flow rates* and therefore requires *less force to actuate.*

— The pilot valve directs pilot pressure to one side of the main spool, and the force generated by the pressure acting over the main spool lobe force is large enough to actuate the main valve. The effect of the pilot valve is to *amplify* the force provided by the solenoid or mechanical lever acting on the pilot spool.

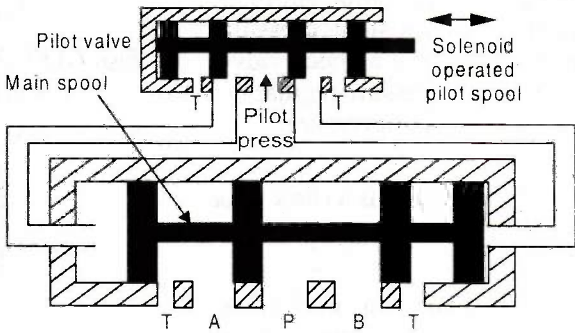


Fig. 7.120. Pilot-operated spool valve.

Fig. 7.120, shows that the pilot spool is in the full left position, causing pilot pressure to be applied to the left side of the main pool and venting fluid from the right side of the main pool to tank, thus driving the main spool to the full right position. This applies main pressure to port-*B* and vents port-*A*.

Proportional valve :

In case of the spool valves, discussed so far, the operation is limited between two positions only *i.e., on* and *off*. However, continuous operation can be achieved by using a *"proportional valve"*, one whose spool moves a distance proportional to a mechanical (say a lever) or electrical input (say an adjustable current solenoid), then changing the flow rate and varying the speed and force of the actuator.

- When the spool position is controlled by solenoids, the proportional valve is called an *"electrohydraulic valve"*. These valves may be used in open control situations with no feedback, but they often include sensors to monitor spool position or actuator output.
- *Proportional valves equipped with sensor and control circuitry* are often called *"servo-valves"*.
- Electrohydraulic valves are often pilot operated where the solenoids drive the pilot spool, which in turn controls the position of the primary spool.

Figs. 7.121 and 7.122, show the block diagrams of mechanical hydraulic servo-valve and electrohydraulic servo-valve.

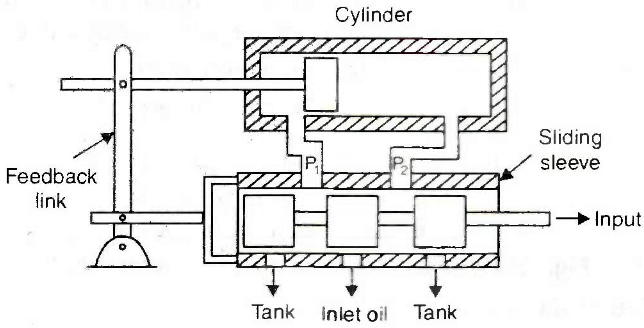


Fig. 7.121. Mechanical hydraulic servo-valve.

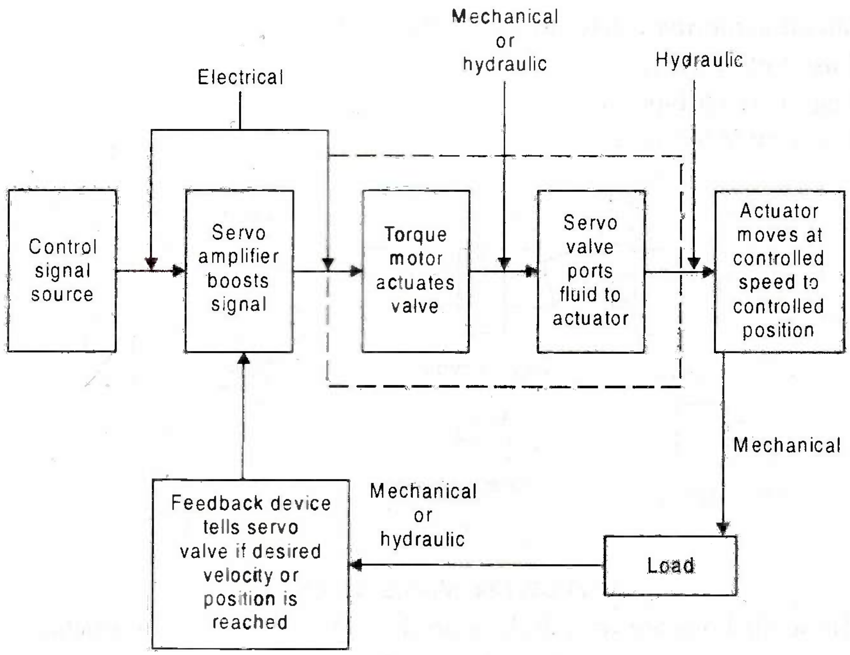


Fig. 7.122. Electrohydraulic servo-valve.

— *Electrohydraulic systems use low power electrical signals (1 W) for controlling the movements of large power hydraulic pistons (7640 W or more).* The typical applications are *aircraft controls and numerical control machines.*

4. Rotary valves :

With a rotary spool the hydraulic fluid is directed through longitudinal grooves machined into a rotatable piston. As a rotational movement is necessary to effect changeover, this type of spool is used predominantly in *manually operated valves.*

A *rotary valve* consists of a rotating spool that is fitted into a circular valve body with a close tolerance. The hydraulic fluid is directed through the longitudinal grooves or passages provided on the rotating spool. These passages connect or block the ports in the valve body and if necessary a centre position can be incorporated.

— Rotary valves can be line, panel or sub-plate mounted.

— These valves are generally *low flow valves.*

— Although rotary valves can be used for reversing cylinders and motors, these are usually used as *pilot valves in pilot* operated directional valve.

Fig. 7.123 shows a rotary four way, three-position valve.

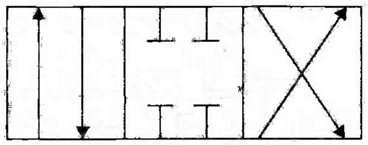


Fig. 7.123. Rotary four way, three-position valve.

Methods of actuation of spools :

The actuating force for providing axial motion to the sliding spool in a sliding spool direction control valve can be imparted by the following five methods :

1. ***Manual actuation :*** Refer to Fig. 7.124(*i, ii, iii*).

(*i*) Push button type;

(*ii*) Leg operated type;

(*iii*) Lever operated type .

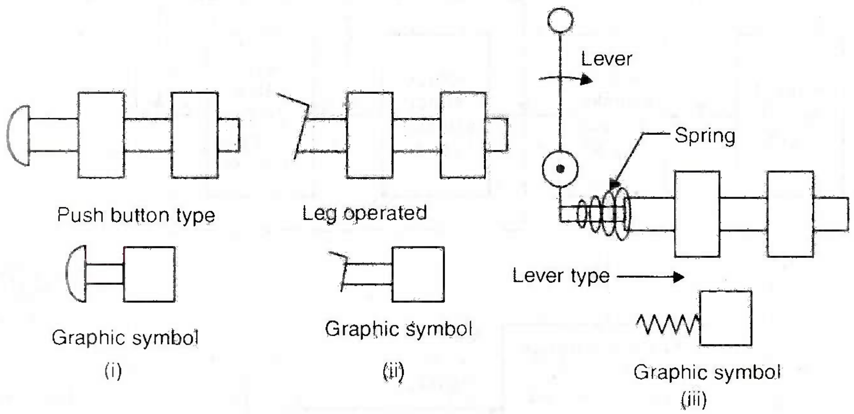


Fig. 7.124. Manual actuators.

- These methods are *generally considered unsafe and hence not in practice.*

2. ***Mechanical actuation :*** Refer to Fig. 7.125(*i, ii*).

(*i*) Plunger type;

(*ii*) Roller wheel type.

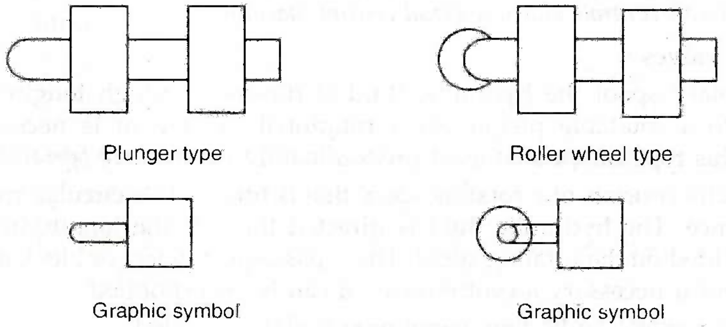


Fig. 7.125. Mechanical actuators.

3. ***Pneumatic actuation :*** Refer to Fig. 7.126.

The pneumatic actuator uses the force of air applied to a piston, which is connected to spool. As the piston moves due to air the spool actuates. In this case as the pressure is quite low the actuator piston must be relatively large to overcome the spring and flow forces.

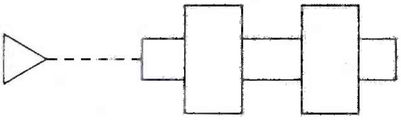


Fig. 7.126. Pneumatic actuation.

4. ***Electrical actuation :*** The most common type of electrical actuator is *"solenoid"*.

— A solenoid is made up of two parts : a coil and an armature as shown in Fig. 7.127. When coil is energised magnetic field is produced which attracts the armature, which is turn pushes the solenoid pin or the spool.

— Solenoid can be operated either on A.C. or D.C. signals. The designs may be different but the operating principle is same.

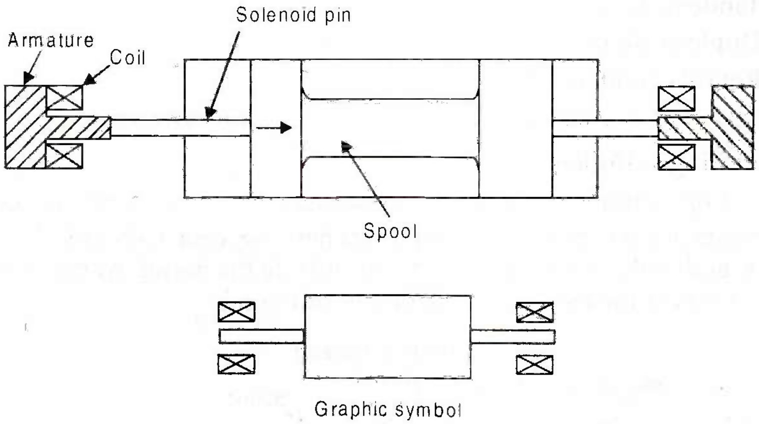


Fig. 7.127. Solenoid actuated valve.

5. ***Hydraulic actuation :*** The fluid discharged by the pump is directed to the 'hydraulic actuator'. The actuator converts the pressure energy of the fluid into mechanical energy. The hydraulic actuators are of the following two types :

(*i*) Linear actuators ;
- Hydraulic cylinder.

(*ii*) Rotary actuators ;
- Hydraulic motor.

7.4.6. Linear Actuators

A **fluid power hydraulic cylinder** *is a linear actuator which is most useful and effective in converting fluid energy to an output force in a linear direction for performing work such as pulling or pushing in a variety of engineering applications such as in machine tools and other industrial machinery, earth moving equipment, construction equipment and space applications.*

A hydraulic cylinder usually consists of a *movable element*, a *piston* and a *piston rod* operating within a cylindrical bore.

7.4.6.1. Types of cylinders

The cylinders may be *classified* as follows:

I. *According to function performed :*

(i) Single acting cylinders.
(ii) Double acting cylinders.

II. *According to construction:*
(i) Tie rod cylinders.
(ii) Mill type cylinders.
(iii) One-piece welded cylinders.
(iv) Threaded head cylinders.

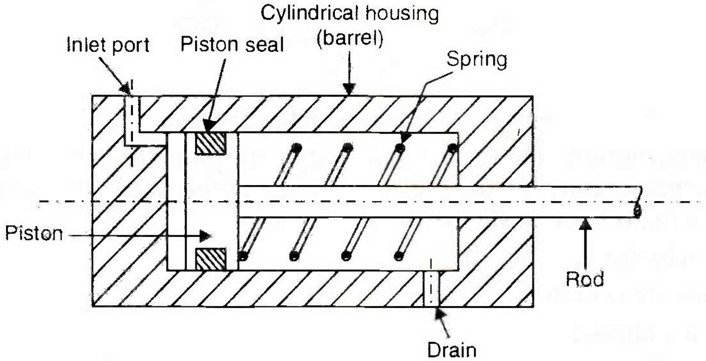
III. *Special types :*
(i) Plunger or ram cylinders.
(ii) Telescoping cylinders.
(iii) Cable cylinders.
(iv) Diaphragm cylinders.
(v) Tandem cylinders.
(vi) Duplex cylinders.
(vii) Rotary cylinders etc.

Some of the commonly used cylinder are discussed below :

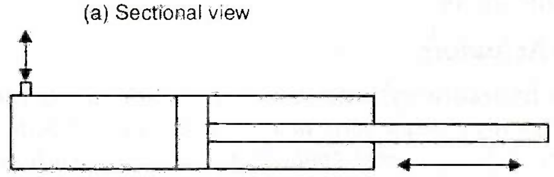
1. Single acting cylinders :

A single acting cylinder (See Fig. 7.128) is designed to apply force in *only one direction.*

— It consists of a piston inside a cylindrical housing, called a barrel. A rod is attached to one end of the piston and it extends outside the barrel. At the other end (blank end) is a port for the entrance and exit of the oil.



(a) Sectional view



(b) Schematic

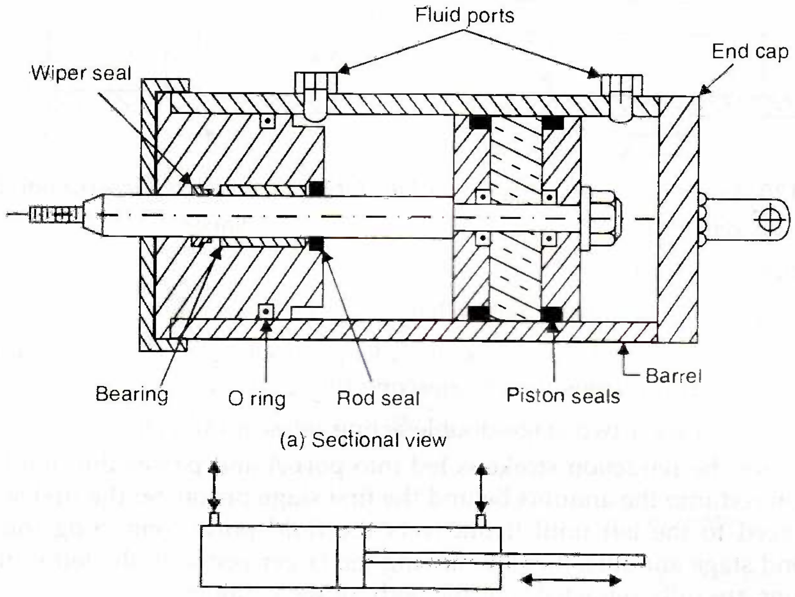
Fig. 7.128. Single acting cylinder.

— A single acting cylinder can exert a force only in the extending direction, as fluid from the pump enters through the blank end of the cylinder.

— These cylinders do not retract hydraulically. Retraction is accomplished by the inclusion of a compression spring or by using gravity.

2. Double acting cylinders :

A double acting cylinder (see Fig. 7.129) is capable of delivering forces in both directions and is most commonly used in industrial hydraulics.



(a) Sectional view

(b) Schematic.

Fig. 7.129. Double acting cylinder.

— The *barrel* is made of seamless steel tubing, honed to a fine finish on the inside surface. The *piston* which is made of ductile iron contains *U* cup packings to seal the leakage between the piston and the barrel. The *ports* are located in the end caps which are secured to the barrel by tie rods. The load of the piston rod at the neck is taken by the bearing, which is generally made of brass or bronze. A rod *wiper* is provided at the end of the neck to prevent foreign particles and dust from entering into the cylinder along with the piston rod.

— As the fluid from pump enters the cylinder through port-1, the piston moves forward and the fluid returns to the reservoir from the cylinder through port-2. During the return stroke the fluid is allowed to enter the cylinder through port-2 and the fluid from the other side of the piston goes back to the reservoir through port-1.

— Double acting cylinders my be either single rod ended, (also called *differential cylinder*) or double rod ended, (also called *non-differential* cylinder) as shown in Figs. 7.130 and 7.131 respectively.

— The single rod cylinders have piston connected to a smaller diameter piston rod. For a given pressure, these cylinders exert greater force when extending than when retracting.

— The double rod ended cylinder is used when it is required to exert *equal forces in both direction.* However, the maximum force of the cylinder for a given tube size is smaller than the single rod end type.

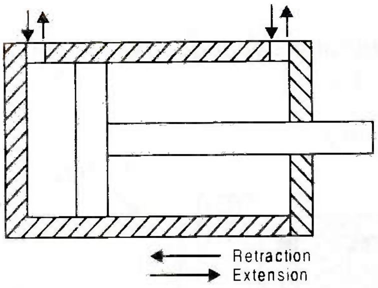


Fig. 7.130. Single rod ended or differential cylinder.

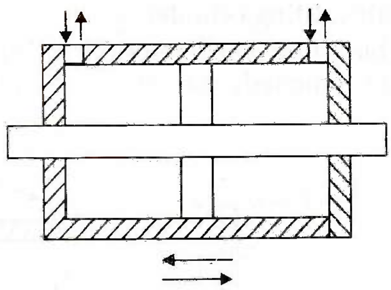


Fig. 7.131. Double rod ended or non-differential cylinder.

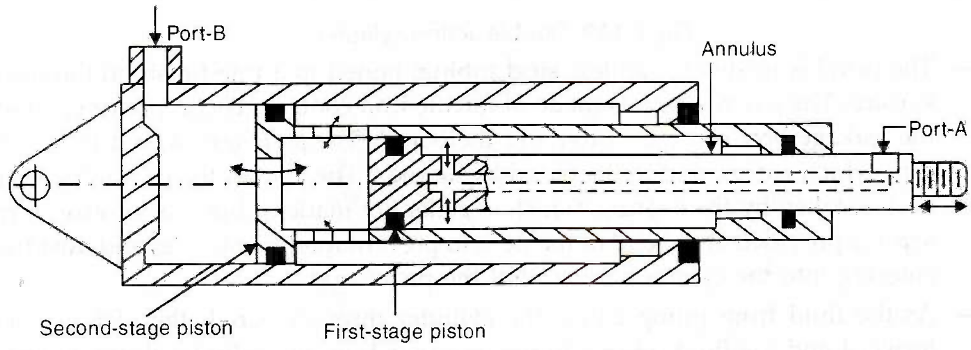
3. Telescoping cylinders :

These cylinders are employed where *long work-strokes are required.*

A telescoping cylinder provides a relatively long working stroke for an overall reduced length by using several pistons which telescope into each other.

Figure 7.132, shows a two-stage double acting telescoping cylinder :

— Fluid for the retraction stroke is fed into port-*A* and passes through the hollow piston rod into the annulus behind the first stage piston. So the first stage piston is forced to the left until it uncovers the fluid ports connecting this with the second stage annulus, thereby moving the larger piston to the left until both the pistons are fully retracted into the body of the cylinder.



Fi.g 7.132. A two-stage double acting telescoping cylinder.

— Fluid for the extension stroke is then fed through port-*B*, forcing both pistons to the right until the cylinder is fully extended.

Cylinder ratings :

The ratings of cylinder are based on its size and pressure capability.

— The principal sizes features are :

(*i*) Piston diameter; (*ii*) Piston rod diameter; (*iii*) Stroke length.

— The pressure rating is established by the manufacturer and it is available in the manufacturer's catalogue.

Hydraulic cylinders applications :

The following are the *applications* of hydraulic cylinders:

1. Hydraulic jacks.

2. Aircraft landing system.
3. Hydraulic shapers and many machine tools.
4. Oceanography applications.
5. Tusk tabler for handling huge logs.
6. Power steering control for off-highway vehicles.
7. Earth excavators.
8. Automobile hoisting.

- As illustrated in Fig. 7.133, the *linear actuator* is very versatile in achieving a variety of motions.

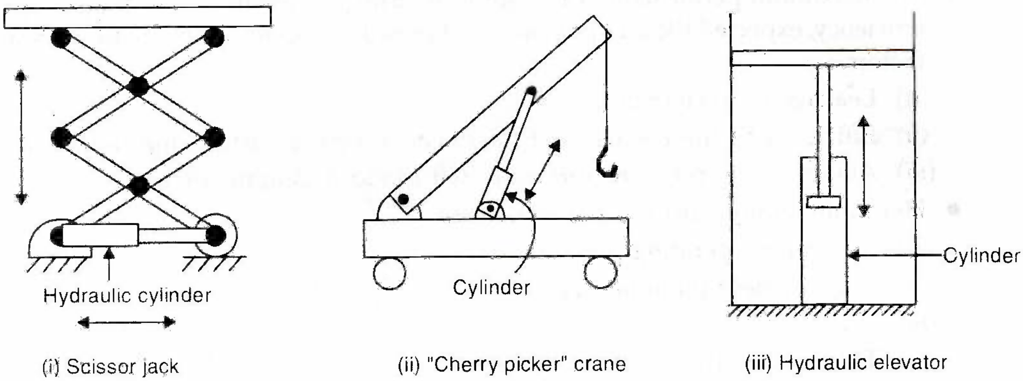


Fig. 7.133. Some mechanisms driven by an hydraulic cylinder.

(*i*) The *scissor jack* converts small linear motion in the horizontal direction to larger linear motion in the vertical direction.

(*ii*) Linear motion of the cylinder in the *crane* results in rotary motion of its pivoted boom.

(*iii*) Cylinder motion in the *hydraulic elevator* drives the elevator directly.

Some common cylinder problems :

Following are some common cylinder problems :

1. Oil leakage at rod end and at other parts.
2. Rod stripping and/or breakage.
3. Sticky, slow start-up cushion.
4. Tubing end leakage.
5. Problems associated with link mechanism and machine members.
6. Premature seal wear out.

7.4.7. Rotary Actuators

Rotary actuators *are the hydraulic pneumatic equivalents of electric motors*. For a given torque, or power, a rotary converter :

— is more compact than an equivalent motor;

— cannot be damaged by an indefinite stall;

— can be safely used in an explosive atmosphere.

For variable speed applications, the complexity and maintenance requirements of a rotary actuator are similar to a thyristor-controlled D.C. drive, but for fixed applications, the A.C. induction motor is simpler to install and maintain.

7.4.7.1. Hydraulic motors

An ***hydraulic motor*** *is a rotary actuator wherein hydraulic energy is converted into mechanical energy in the form of rotary motion and torque which may be used for doing work.* Thus, its function is just opposite to that of an hydraulic pump.

These motors very closely resemble pumps in construction. In an hydraulic motor high pressure oil pushes a freely moving element, thereby developing torque and continuous rotating motion. Since both inlet and outlet ports at times be pressurised, therefore, most hydraulic motors are drained externally.

- The maximum performance of a motor in terms of pressure, flow, torque, speed, efficiency, expected life and physical configuration is determined by the following factors :
  - (*i*) Leakage characteristics;
  - (*ii*) Efficiency of the means used to convert pressure surface and output shaft;
  - (*iii*) Ability of the pressure surface to withstand hydraulic force.
- Hydraulic motors are *specified or rated* by :
  - (*i*) Maximum operating pressure;
  - (*ii*) Displacement (volume size);
  - (*iii*) Speed;
  - (*iv*) Torque capacity.

Rotary actuator symbols :

Fig. 7.134, shows the rotary actuator symbols :

— *Internal leakage always occurs in a 'hydraulic motor'* and a drain line, shown dotted, is used to return the leakage fluid to the tank. If this leakage return is inhibited the motor may pressure lock and cease to rotate or even suffer damage.

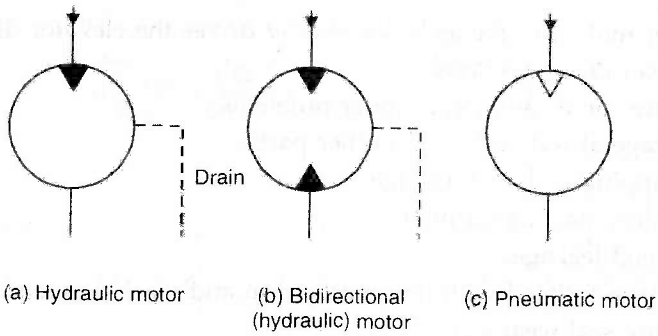


Fig. 7.134. Rotary actuator symbols.

Classification of hydraulic motors :

Hydraulic motors are *classified* as follows :

1. Gear motors.
2. Vane motors.
3. Piston motor
   - (*i*) Radial type;
   - (*ii*) Axial type.

1. *Gear motors :*

Fig. 7.135, shows the schematic of a gear motor.

— Fluid enters at the top and pressurises the top chamber. Pressure is applied to two gear faces at A, and a single gear face at B. *There is, thus, an imbalance of forces on the gears resulting in rotation* as shown.

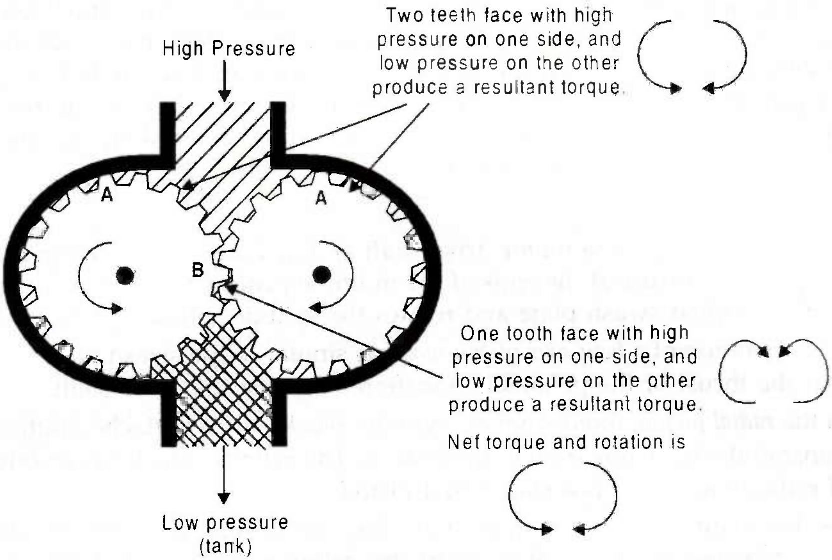


Fig. 7.135. A gear motor.

— Gear motors *suffer from leakage* which is more pronounced at low speed. Thus they *tend to be used in medium speed, low torque applications.*

2. *Vane motors :*

A vane motor is a *positive displacement motor* which develops an output torque at its shaft by allowing hydraulic pressure to act on the vanes which are extended. Basically a vane motor consists of *rotor, vanes, cam ring, port plates* with kidney shaded inlet and outlet ports and shaft.

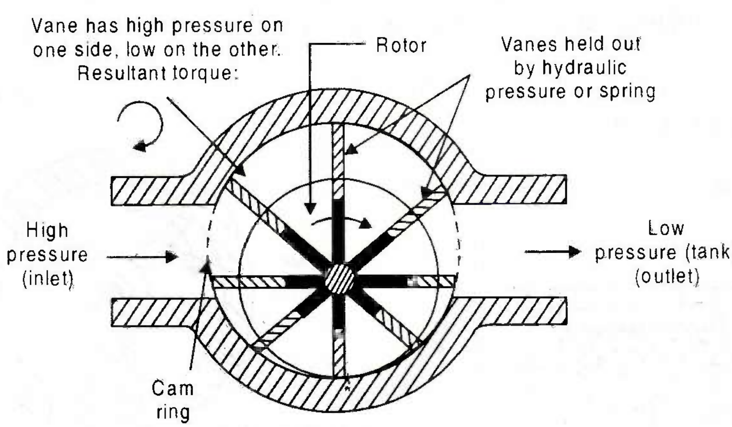


Fig. 7.136. A vane motor.

Fig. 7.136 shows the schematic of a vane motor:

— Vane motors develop torque by the hydraulic pressure that acts on the exposed surfaces of the vanes which slide in and out of the rotor connected to the driver shaft. Larger the exposed surface of the vane or higher the pressure of oil, more will be the torque developed.

— Since no centrifugal force exists until the motor begins to rotate, some means have to be provided to initially hold the vanes against the casting contour. Springs are often used for this purpose.

— In vane motor two different pressures, system pressure and outlet pressure are involved. System pressure will be greater than the outlet pressure resulting in *side loading* in the motor shaft. The side loading in the motor is avoided by using cam-shaped ring, instead of a circular ring. With this arrangement the two pressure quadrants oppose each other and the forces acting on the shaft are *balanced*; such motors are balanced vane motors.

3. ***Piston motors :***

- In *swash plate type*, the motor drive shaft and cylinder block are centered on the same axis. Pressure at the ends of the multiple pistons causes a sequential action against a tilted swash plate and rotates the cylinder block and the motor shaft.
- The operation of a *bent axis piston motor* is similar to the swash plate type, except that the thrust of the pistons is transferred against the drive shaft.
- In the *radial piston*, motors have a cylinder block with an attached output shaft to transmit the force imparted to the pistons. The cylinder block has an odd number of radial bores with precision fitted pistons.
  — When oil enters the cylinder bore, the piston is forced against the thrust ring, imparting a tangential force to the cylinder block and shaft, causing the assembly to rotate. Each piston is pushed inward by the thrust ring once it reaches the outlet port, thus pushing the fluid to the reservoir.

Semi-rotary/Limited motion rotary actuators:

These actuators are used to convert fluid pressure energy into torque which turns through an angle limited by the design of the actuator. With majority of the designs, the angle of rotation is within 360° although it is possible to considerably exceed this when using piston-operated actuators.

Some examples are illustrated in Figs. 7.137(*a*)(*b*).

— In Fig. 7.137(*a*), a double-acting piston is coupled to the output shaft by a rack in piston.

— The actuator shown in Fig. 7.137(*b*) is driven by a single vane coupled to the output shaft.

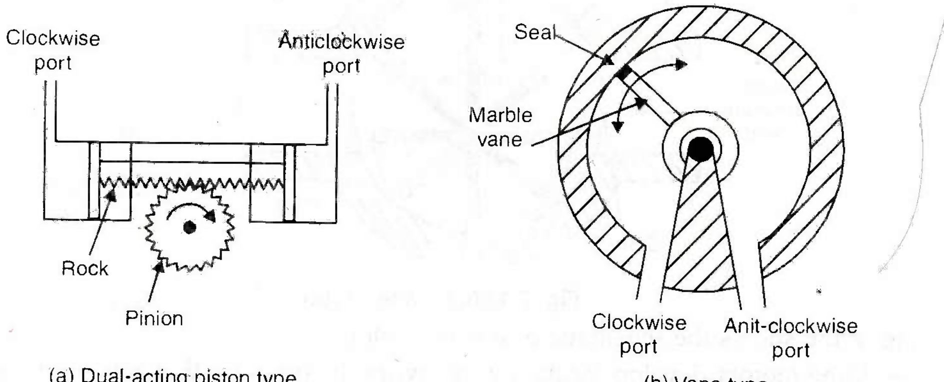


Fig. 7.137. Semi-rotary actuators.

In both the above cases the *shaft angle can be finely controlled by fluid applied to the ports.*

7.4.7.2. Advantages and applications of hydraulic motors

Advantages:

1. These motors are explosion proof. An hydraulic motor requires no electrical connections except only two hydraulic lines.
2. It is easy to achieve frequent stopping, starting and reversing by a single four-way three position direction control valve.
3. Stepless variations in speeds are available from zero to that obtainable from the maximum output of pumping system.
4. The hydraulic motor can be suddenly stopped (by closing the control valve) without any harm to the mechanism.
5. The hydraulic motors can be employed for machines requiring controlled variable torque (*e.g.*, paper winding machines).

Applications:

The fields of applications of hydraulic motors include:

(*i*) Mining equipment;
(*ii*) Machine tools;
(*iii*) Pottery machines;
(*iv*) Drill rigs;
(*v*) Winding machines etc.

Note: ● Hydraulic systems have the ***advantage*** of generating *extremely large forces from very compact actuators.* They also can provide *precise control of low speeds* and have built in travel limits defined by the cylinder stroke. ● The ***drawbacks*** of hydraulic systems include the need for a *large infrastructure* (high-pressure pump, tank, and distribution lines); potential for *fluid leaks,* which are undesirable in a clean environment; *possible hazards,* associated with high pressures (*e.g.*, a ruptured line); *noisy operation, vibration, and maintenance requirements.* Because of *these disadvantages, electric motor drives are often the preferred choice.* However, in *large systems, which require extremely large forces, hydraulics often provide the only alternative.*

7.5 PNEUMATIC ACTUATORS

7.5.1. Introduction

Pneumatic systems *use pressurised air to transmit and control power.*

Air is used as the fluid because :

(*i*) It is safe.
(*ii*) It is less expensive and readily available.
(*iii*) It can be inducted and exhausted directly to the atmosphere and a return line is not necessary as with hydraulics.

Comparison between pneumatic systems and hydraulic systems :

The fluid generally found in pneumatic systems is *air,* in hydraulic systems it is *oil.* And it is primarily the different properties of the fluids involved that characterise the differences between the two systems. These differences can be listed as follows:

1. Pneumatic systems are fire-and-explosion proof, whereas hydraulic systems are not, unless non flammable liquid is used.
2. In pneumatic systems no return pipes are used when air is used, whereas they are always needed in hydraulic systems.
3. The normal operating pressure of pneumatic systems is very much lower than that

of hydraulic systems.

4. Pneumatic systems are insensitive to temperature changes, in contrast to hydraulic systems, in which fluid friction due to viscosity depends greatly on temperature. Normally operating temperatures for pneumatic and hydraulic systems are 5°–60°C and 20°–70° respectively.
5. The normal operating pressure of pneumatic systems is very much lower than that of hydraulic systems.
6. Accuracy of pneumatic actuators is poor at low velocities whereas accuracy of hydraulic actuators may be made satisfactory at all velocities.
7. Output powers of pneumatic systems are considerably less than those of hydraulic systems.
8. In pneumatic systems, external leakage is permissible to a certain extent, but internal leakage must be avoided because the effective pressure difference is rather small. In hydraulic systems internal leakage is permissible to a certain extent, but external leakage must be avoided.

7.5.2. Components of a Pneumatic System

Fig. 7.138, shows the various components of a pneumatic system; these are :

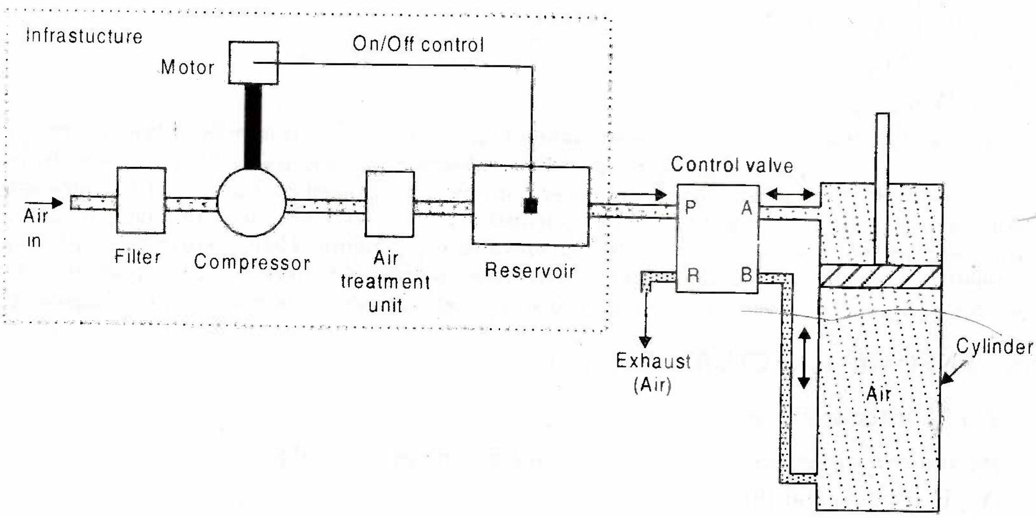


Fig. 7.138. Components of a pneumatic system.

1. **Compressor.** It is used to provide pressurised air, usually on the order of 500 kPa to 1.0 mPa, which is much lower than hydraulic system pressures.

   As a result of the lower operating pressures, pneumatic actuators *generate much lower forces* than hydraulic actuators.
2. **Air treatment unit.** After the inlet air is compressed, excess moisture and heat are removed from the air with an air treatment unit.
3. **Reservoir.** Unlike hydraulic pumps, which provide positive displacement of fluid at high pressure on demand, compressors cannot provide high volume of pressurised air responsively; therefore a large volume of compressed air is stored in a reservoir or tank.

   The reservoir is equipped with a *pressure sensitive switch* that activates the compressor when the pressure starts to fall below the desired level.

4. **Control valve and actuator.** *Control valves* and *actuators* act in much the same way as in hydraulic systems, but instead of returning fluid to tank, the air is simply returned (exhausted) to the atmosphere.

Following points are worth noting :

- Pneumatic systems are *open systems*, always processing new air (whereas hydraulic systems are closed systems, always returning the same oil).
- Since air is compressible, pneumatic cylinders are not typically used for applications requiring accurate motion between the end points, especially in the presence of a varying load.

7.5.3. Pneumatic Valves

In order to control pneumatic actuators, the air energy has to be regulated, controlled and reversed with a predetermined sequence. This is achieved with the help of pneumatic valves; these are enumerated and discussed as follows :

1. Direction control valves :
   (*i*) Two-way valve
   (*ii*) Three-way valve
   (*iii*) Four-way valve
   (*iv*) Five-way valve.
2. Pneumatic check valve.
3. Flow control valve.
4. Pneumatic shuttle valve or 'OR' type valve.
5. 'AND' type or Two-pressure valve.
6. Quick exhaust valve.
7. Time delay valve.

- These valves are *used mainly to direct the flow of the pressure fluid in the desired directions.* The main functions of these valves are to *start, stop and regulate the direction of air flow and help distribution of air in the desired line.*
  — They can be actuated to assume different positions by various actuating mediums, viz electrical, mechanical, pneumatic, or other modes of control. This results in corresponding connection or description of flow between various port openings. The various types of these valves are discussed below :

1. Directional control (D.C.) valves :

(*i*) *Two way valve :* Fig. 7.139 shows the symbolic representation of the valve. This is an *on-off* type of device.

— It is usually provided with two external ports, a supply port and an exhaust port.

— A normally open two-way valve permits the flow in its normal or in its rest position and blocks flow when actuated.

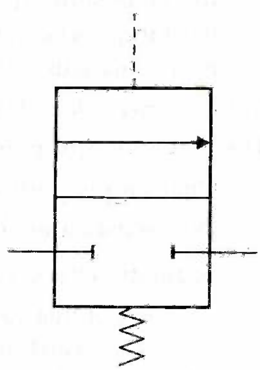


Fig. 7.139. Two-way direction control valve valve-poppet type.

The normally closed valve blocks flow in its normal position and permits the flow when actuated.

(*ii*) *Three-way valve:* Refer to Fig. 7.140.

In this type of valve, one flow port is connected to either of the two ports. It may be used alternatively to pressurise one port and exhaust the other port.

These valves may be used :

- As a pilot relay to operate the other valves;
- To control single acting cylinder or in pairs to control double acting cylinders.

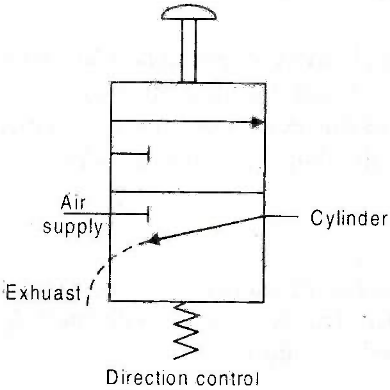


Fig. 7.140. Three-way valve-poppet type.

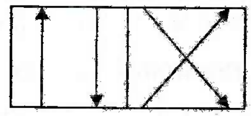
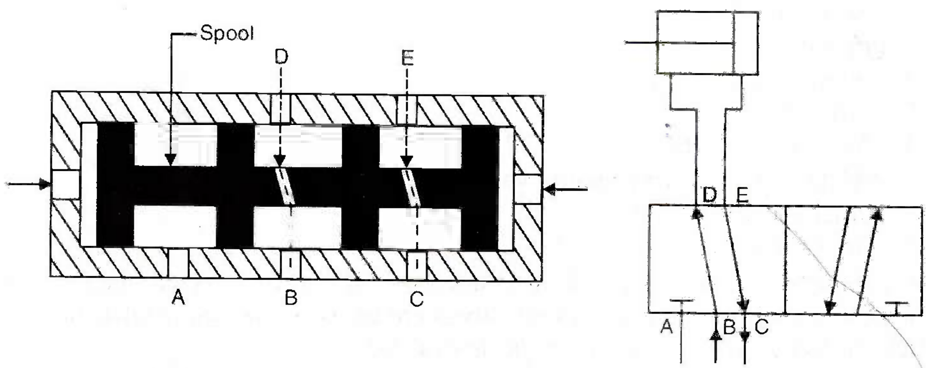


Fig. 7.141. Four-way valve-seat type



(a) Sectoral view (b) Symbol

Fig. 7.142. Five-way valve–spool type.

(*iii*) *Four-way valve:* Refer to Fig. 7.141.

— A four-way valve has two working ports, a supply port and an exhaust port.

— In one position the valve allows air to flow from the supply source to one of the working ports. Simultaneously air is permitted to flow from the other working port to exhaust. The flow paths are reversed when the valve is shifted.

(*iv*) *Five-way valve:* Refer to Fig. 7.142.

This valve design permits the use of either dual supply or dual exhaust.

— Dual supply ports permit the use of different pressures for the cylinder movement.

— Dual exhaust enables easy exhaustion of the valve.

2. Pneumatic check valve :

The function of this valve is to *shut off* against reverse flow and *open* at a low working pressure in the forward direction.

Such valves having metal or light weight plastic body designs are available.

3. Flow control valve :

— This type of valve has a spring loaded disc which allows a *free flow in one direction and an adjustable or controlled flow in the opposite direction.*

— Flow adjustment is performed by a tapped brass stem that controls the flow through the cross hole in the disc.

4. Pneumatic shuttle valve or 'OR' type valve:

— This valve *automatically selects the higher of the two input pressures and connects that pressure to the output port while blocking the lower pressure.*

— A pneumatic shuttle valve delivers an output when one input is present or when both are present.

5. 'AND' type or two-pressure valve : Refer to Fig. 7.143.

In this type of valve, *an output is produced if both the input signals are fed.*

- This has two inlets (X and Y) and one outlet (A). When signal is fed first to X, the valve spool moves towards Y, closing the air passage from X to A. The reverse takes place if air is fed first to Y. If air is fed simultaneously to X and Y, then spool remains in its acquired position and air may pass to A from both X and Y. If different pressures are present, the low pressure is switched to outlet A.

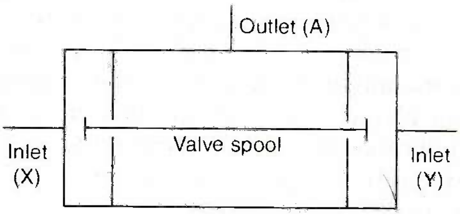


Fig. 7.143. Two-pressure valve.

6. Quick exhaust valve : Refer to Fig. 7.144.

With the use of flow control valve in a pneumatic circuit, the actuator speed is controlled, which means that the speed of the actuator may be reduced over its normal speed to suit a particular need of the system design. But this valve is *used to induce a higher speed in a cylinder by allowing the exhaust air to pass through the direction control valve from the cylinder, so that air energy can act quickly.*

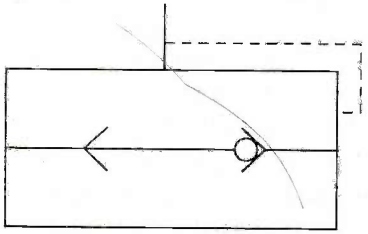


Fig. 7.144. Quick exhaust valve.

When air is fed to the piston side of the cylinder, air in the rod end of the cylinder exhausts to atmosphere quickly by using this valve. Here, the air flowing to the cylinder from the direct control valve will pass to the *port of the quick exhaust valve* and from here to the *port of the valve* and then to cylinder. But the return air from the cylinder will exhaust to the atmosphere without travelling through the exhaust valve port and thus avoids direct control valve. So the resistance to piston movement is eliminated to some extent and speed of the cylinder is *accelerated proportionally to the amount of reduced resistance.*

7. Time delay valve : This valve is used in the pneumatic system to *initiate a delayed signal.*

7.5.4. Linear and Rotary Actuators

7.5.4.1. Linear actuators—Pneumatic cylinders

The pneumatic cylinders *convert the pneumatic power into straight-line reciprocating motions.*

Pneumatic cylinder construction makes extensive use of aluminium and other non-ferrous alloy materials to reduce the weight and the corrosive effects of air and to improve heat transfer capabilities.

Air cylinders, according to the operating principle, are ***classified*** and described as follows : Refer to Fig. 7.145.

1. Single-acting cylinder.
2. Double-acting cylinder.
3. Tandem cylinder.
4. Three position cylinder.
5. Through rod cylinder.
6. Adjustable stroke cylinder.
7. Telescoping cylinder.

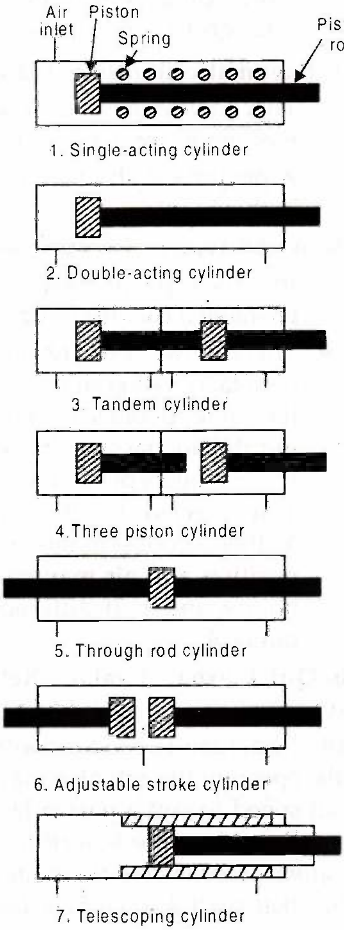


Fig. 7.145. Basic cylinder types.

1. Single-acting cylinder :

In a single-acting cylinder, the compressed air is fed only in one side. Hence, this cylinder can produce work only in one direction. The return movement of the piston is effected by a built in spring or by application of an external force. The spring is designed to return the piston to its initial position with a sufficiently high speed.

— The *advantage* of a single-acting cylinder lies in its reduced air consumption, since air is not wasted while retracting the piston.

2. Double-acting cylinder :

In this type of cylinder, the force exerted by the compressed air moves the piston in *two directions*. This cylinder produces less force during retraction, because the piston rod's cross-sectional area is subtracted from the piston area under pressure.

In principle, the stroke length is unlimited, although buckling and bending must be considered before we select a particular size of piston diameter, rod length and stroke length.

— These are used particularly when the piston is required to perform work not only on the advance movement but also on the return.

3. Tandem cylinder :

Here two cylinders are arranged in series so that the force obtained from the cylinder is almost double.

— Since the available force is doubled, this design is useful when *larger forces are required*, but a single cylinder with a larger diameter cannot be accommodated.

4. Three position cylinder :

A three position cylinder is quite similar to the tandem cylinder, except that the left-piston rod is *not connected* to the right piston and the left cylinder is *shorter* than the right one.

With the left piston extended, the retraction of the right piston is limited to an intermediate position which is determined by the ability of the right-piston to retract fully.

5. Through rod cylinder : Here the piston rod is extended on both ends of the piston. This will *ensure equal force and speed on both sides of the cylinder.*

6. Adjustable stroke cylinder :

— The cylinder stroke can be adjusted by screwing the left hand piston in or out.

— By using the shortest possible stroke needed for a given job, better rapid cycling is achieved and *air consumption is reduced.*

7. Telescoping cylinder :

When pressure is applied to the left side, the inner cylinder acts as a piston and extends. Once it reaches the end of its stroke, the inner most piston begins to extend. The available stroke is almost double when compared to a normal cylinder having the same retracted length.

7.5.4.2. Rotary actuators—Air motors

An **air motor** *is used to generate rotational motion in a pneumatic system.*

- The air motors have been found to provide very high rotational speeds, which may sometimes go upto 10,000 rpm. These motors are manufactured with fractional kW as low as 0.05 kW, while the higher range is upto 20 kW.

Types of air motors : The various types of air motors are :

1. Piston type motors.
2. Vane motors.
3. Turbine motors.
4. Gerotor type motors.

1. Piston type motors :

These motors may be of *axial* or *radial* type design :

— The operation of an *"axial"* piston air motors is similar to the piston type hydraulic motor. As pistons reciprocate in sequence, they actuate a wobble plate and this in turn imparts a rotary motion to the output shaft through a gear train.

— Axial piston motors are low power (2.5 kW) motors while *radial* piston motors give upto 18 kW.

— *"Radial" piston motors are low speed motors.*

— Piston type motors may have 4, 5, or 6 cylinders. The power developed by these motors is dependent on the *inlet pressure, number of pistons, the area of the pistons, the stroke of the pistons* and *the speed.*

— The 5-cylinder design provides an *even torque* at any given operating speed due to the overlap of the five power impulses occurring in the stroke revolution.

2. Vane motor : Fig. 7.146 shows a vane motor.

— An eccentric rotor has slots in which vanes are forced outwards against the walls of the cylinder by rotation. The vanes divide the chamber into separate compartments which increase in size from the inlet port round to the exhaust port. The air entering such a compartment exerts a force on a vane and causes a rotor to rotate.

— The motor can be made to *reverse* its direction of rotation by using a *different inlet port.*

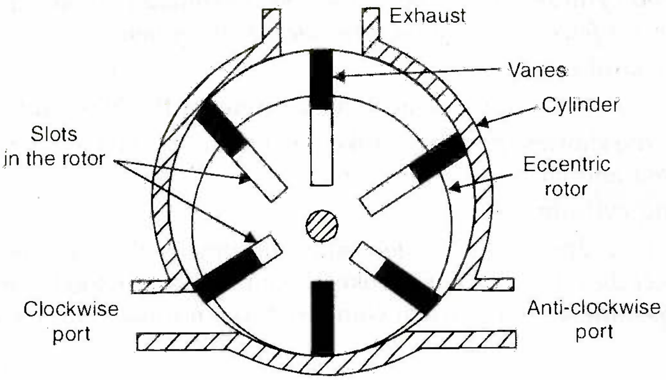


Fig. 7.146. Vane motor.

3. Turbine motors :

These motors *convert low velocity high pressure air to high velocity low pressure air by passing it through metering nozzles.* The advantage of this arrangement is that there is no rubbing or sliding contact between the rotating parts and the body cavity. This *reduces wear and lubricated air is not required to seal and lubricate parts.*

- These are *high speed low torque motors* for the same volume of air than piston vane type.

4. Gerotor type motor : A gerotor type motor is shown in Fig. 7.147.

- These air motors are mostly used for *low r.p.m.* (such as 20 to 30 r.p.m.) pressure, hence they *may not be suitable for high torque application.*

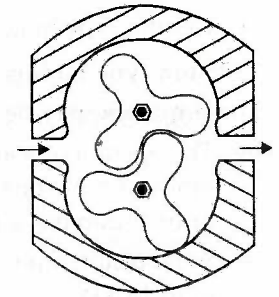


Fig. 7.147. Gerotor motor.

Applications of air motors :

The air motors may be used in conjuction with hydraulic power units :

— Conveyor belts;

— Agitators and mixers;

— Pipe threaders;

— Tool devices;

— Bench grinders etc.

7.5.5. Special Features of Pneumatic Actuators

Pneumatic actuators should possess the following *special features* :

1. Better heat transfer capability.
2. Higher fatigue life.
3. Simple in construction.
4. High reliability against failure.
5. Should be made of anticorrosive materials.
6. Light weight (so that they are easier to manipulate).

7.5.6. Example of Fluid Control System

Fig. 7.148 shows the essential features of a system for the control of a variable such as the level of a liquid in a container by controlling the rate at which liquid enters it.

The output from the liquid level sensor, after signal conditioning, is transmitted to the *current to pressure converter* as a current of 4 to 20 mA. It is then converted into a gauge pressure of 20 to 100 kPa which then actuates a pneumatic control valve and so controls the rate at which liquid is allowed to flow into the container.

- The basic form of a *current to pressure converter* is shown in Fig. 7.149.

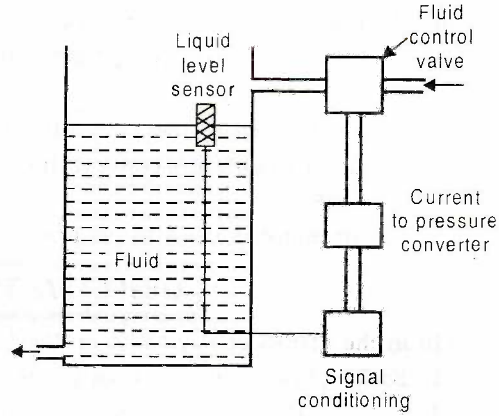


Fig. 7.148.

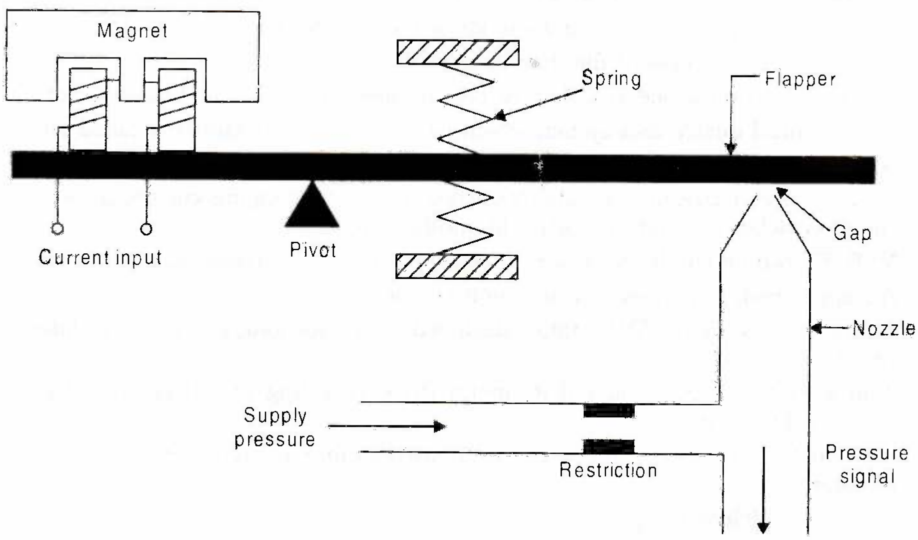


Fig. 7.149. Current to pressure converter.

— The input current passes through the coils mounted on the core which is attracted towards a magnet, the extent of the attraction depending on the size of the current.

— The movement of the core causes movement of a flapper above the nozzle. The position of the flapper in relation to the nozzle determines the rate at which air can escape from the system and hence the air pressure in the system.

Springs on the flapper are used to adjust the sensitivity of the converter so that currents of 4 to 20 mA produce gauge pressures of 20 to 100 kPa. These are the standard values that are generally used in such systems.

HIGHLIGHTS

1. *Actuators* produce physical changes such as linear and angular displacement. They also modulate the rate and power associated with these changes.
2. When one of the links of a kinematic chain is fixed, the chain is known as *mechanism*.

3. *Electrical actuators* may be : Switching devices and Drive systems.
4. *A stepper motor,* a special type of D.C. motor, is an incremental motion machine.
5. An actuator wherein hydraulic energy is used to impart motion is called an *hydraulic actuator.*
6. *Pneumatic systems* use pressurised air to transmit and control power.
7. Pneumatic cylinders convert the pneumatic power into straight-line reciprocating motions.
8. An *air motor* is used to generate rotational motion in a pneumatic system.

OBJECTIVE TYPE QUESTIONS

Fill in the blanks or Say "Yes" or "No"

1. Rack-and-pinion can be used to convert rotational motion to linear motion.
2. .... is a device by means of which available energy can be converted into desired form of useful work.
3. A kinematic pair is a joint of two links that permits relative motion.
4. A mechanism with .... links is known as simple mechanism.
5. .... is the reverse of the diametral pitch.
6. A journal bearing is one in which the bearing pressure is parallel to the axis of the shaft.
7. A mechanical device or a system which has motion or movement is called an ....
8. Solenoids can be used to provide electrically operated actuators.
9. .... are electrically operated switches in which changing current in one electrical circuit switches a current on or off in another circuit.
10. MOSFET can be employed as a control switch for a D.C. motor as on-off switch.
11. A stepper motor is an incremental motion machine.
12. Torque motors are the D.C. motors designed to run for long periods in a stalled or low speed condition.
13. A inverter fed trapezoidal PMAC motor drive operating in self controlled is called a .... D.C. motor.
14. An actuator wherein hydraulic energy is used to import motion is called an hydraulic actuator
15. A .... delivers high pressure fluid.
16. A value is represented by a .... for each of its switching positions.
17. The .... valves start, stop and control the direction of flow for reversing the direction of motion of the actuator.
18. Proportional valves equipped with sensor and control circuitry are often called .... valves.
19. The most common type of electrical actuator is ....
20. .... cylinder are used where long work-strokes are required.

ANSWERS

1. Yes	2. Machine	3. Yes	4. four
5. Module	6. No	7. actuator	8. Yes
9. Relays	10. Yes	11. Yes	12. Yes
13. brusheless	14. Yes	15. pump	16. square
17. direction control	18. servo	19. solenoid	20. Telescoping.

THEORETICAL QUESTIONS

1. What is an 'actuator'? List the various types of actuators.
2. Define the following terms : Machine; Kinematic link; Kinematic pair; Kinematic chain; Inversion of mechanism.
3. What are the advantages and disadvantages of toothed gearing?
4. What is a 'bearing'? How are bearings classified?
5. What is an 'electrical actuator?
6. What is meant by 'electrical actuation system'? What are the devices used in such systems.
7. What are the mechanical switches? Explain.
8. Explain the terms 'bouncing' and 'debouncing' as applied to mechanical switches.
9. What are the various methods which can be used to tackle the problem of 'bouncing' in mechanical switches.
10. Explain briefly the working of following mechanical switching devices:
    (*i*) Solenoids; (*ii*) Relays.
11. Explain briefly any two of the following solid-state devices which can be used to electronically switch circuits:
    (*i*) Diodes; (*ii*) Thyristors;
    (*iii*) Bipolar transistors; (*iv*) Power MOSFETs.
12. Explain briefly D.C. motor control by using MOSFETs.
13. Give brief classification of electric motors.
14. Explain briefly the following electric motors:
    (*i*) Permanent magnet D.C. motors;
    (*ii*) Stepper motors.
15. State the advantages and applications of stepper motor.
16. What are servo-motors? Explain briefly.
17. Explain briefly the following :
    (*i*) Moving coil motor. (*ii*) Brushless D.C. motors.
18. What are the advantages of electronic control systems?
19. Explain briefly the various methods by which speed of D.C. motor can be controlled.
20. Discuss briefly the following electric motors:
    (*i*) Single phase motors.
    (*ii*) Three phase induction motor.
    (*iii*) Synchronous motor.
21. Discuss briefly the electronic control of A.C. (induction) motors.
22. Discuss briefly "Digital control of electric motors".
23. What factors/specifications should be considered while selecting a motor for mechatronic applications?
24. What is an 'hydraulic actuator'?
25. What are the advantages and disadvantages of hydraulic system.
26. What are the basic elements of an oil hydraulic system?
27. What are the components of an hydraulic system? Explain briefly?
28. What is the function of an hydraulic pump?
29. How are hydraulic pumps classified?
30. Explain briefly the following :
    (*i*) Centrifugal pump. (*ii*) Gear pump.
    (*iii*) Vane pump.

31. Explain briefly with the help of a sketch, the swash plate piston pump.
32. Describe briefly a pressure regulator.
33. What are hydraulic valves? How are these classified.
34. What are the functions of pressure control valves?
35. Give the classification of pressure control valves.
36. Explain briefly the following valves:
    (*i*) Pressure relief valves.
    (*ii*) Pressure sequencing valves.
    (*iii*) Pressure reducing valves.
37. Discuss briefly flow control valves.
38. What is the function of a 'direct control valve'?
39. How are direct control valves classified?
40. Explain briefly any two of the following valves:
    (*i*) Check valve. (*ii*) Spool valves.
    (*iii*) Proportional valve. (*iv*) Rotary valves.
41. What is fluid power/hydraulic cylinder? Explain briefly?
42. How are hydraulic cylinders classified?
43. Explain briefly the following :
    (*i*) Double acting cylinders.
    (*ii*) Telescopic cylinders.
44. What are the applications of hydraulic cylinders?
45. List some common cylinder problem.
46. What is an 'hydraulic motor'?
47. How are hydraulic motors specified or rated?
48. Explain briefly the following hydraulic motors :
    (*i*) Gear motors; (*ii*) Vane motors.
49. State the advantages and applications of hydraulic motors.
50. Give the comparison between pneumatic systems and hydraulic systems.
51. Explain with the help of a neat diagram the components of a pneumatic system.
52. Enumerate various types of pneumatic valves and explain briefly and two of them.
53. Explain briefly the following :
    (*i*) Pneumatic check valve.
    (*ii*) Pneumatic shuttle valve.
    (*iii*) Quick exhaust valve.
54. What is the function of a following pneumatic cylinder?
55. How are pneumatic cylinders classified?
56. Explain briefly any two of the pneumatic cylinders:
    (*i*) Double-acting cylinder (*ii*) Tandom cylinder
    (*iii*) Three position cylinder (*iv*) Telescoping cylinder.
57. What is the function of an air motor?
58. Name the various types of motors.
59. Explain briefly any two of the following air motors?
    (*i*) Piston type motors (*ii*) Vane motors
    (*iii*) Turbine motors (*iv*) Gerotor type motors.
60. What are the special features which pneumatic actuators should possess?
61. List the applications of air motors.

CHAPTER

8 Mechatronic Systems

8.1 GENERAL ASPECTS

Earlier we developed the foundations for the integration of mechanical devices, sensors, and signal and power electronics into mechatronic systems. For obtaining completeness in the integration of mechanical devices, sensors, and signal and power electronics in the most advanced mechatronic system, it is essential to include microprocessor–based control systems. Some other *"control architectures"* also useful is mechatronic systems are discussed below:

1. **Analog circuits:** Several mechatronic designs need a specific actuator output based on an analog input signals. In order to effect the desired control, in some cases, analog signal processing circuits consisting of *operational amplifiers* and *transistors* can be used.
   - Analog controllers are often simple to design and easy to implement and can be less expensive than microprocessor-based systems.
2. **Digital circuits:**
   - When the input signals are *digital* or can be converted to a finite set of states, then *combinational or sequential logic controllers* may be easy to implement in mechatronic design.
   - In case of the simplest designs, a few digital chips are employed to create a digital controller. For generating complex Boolean functions on a single IC (integrated chip), specialised digital devices such as PAL (programmable array logic) controllers and PLAs (programmable logic arrays) can be used to reduce the complexity of design. Sometimes, it may be economically feasible to design an ASIC (application-specific integrated circuit) that provides unique digital functionality on a single IC. An ASIC solution, in high-volume manufacturing applications, can be cheaper, smaller and require less power.
3. **Programmable logic controller (PLC):**
   - PLCs are specialised industrial devices for interfacing to and controlling analog and digital devices.
   - They are usually provided with *"ladder logic"* which is a graphical method of laying out the connectivity and logic between system inputs and outputs.

- Besides being *flexible and easy to program, they are robust and relatively immune to external interference.*

4. **Microcontroller:**
   - The *microcontroller* (a microcomputer on a single IC) provides a small, flexible control platform that can be easily *embedded in a mechatronic system.*
   - It can be programmed to perform a wide range of control tasks.
5. **Single-board computer:**
   - A single-board computer is considered to be a good alternative when an application requires more features or resources than can be found on a typical microcontroller and size is not a major concern.
   - These computers are easily interfaced to a personal computer; this is useful in the testing and debugging stages of design development and for downloading software into the memory of the single-board computer.
6. **Personal computer:**
   - A desktop or laptop personal computer (PC) may serve as an appropriate control platform in large sophisticated mechatronic systems.
   - The personal computer can be easily interfaced to sensors and actuators using commercially available plug-in-data acquisition cards.
   - PC-controlled mechatronic systems are especially common in research and development testing and product development laboratories.

8.2 DESIGN PROCESS

For any system, the design process involves the following *stages*:

1. The need.
2. Analysis of the problem.
3. Preparation of a specification.
4. Generation of possible solutions.
5. Selection of a suitable solution.
6. Production of a detailed design.
7. Production of working drawings.

8.3 TRADITIONAL AND MECHATRONIC DESIGNS

The engineering design (a complicated process) involves interactions between many stills and disciplines. The mechatronic approach is based on the inclusion of the disciplines of *electronics, computer technology* and *control engineering*.

Let us consider an **example** of *temperature control for a domestic central heating system:*

— The *"traditional design"* for such a system has been a *bimetallic thermostat* in a closed-loop control system. As the temperature changes the bimetallic strip operates an on/off switch for the heating system.

The bimetallic thermostat system has the following *limitations/disadvantages*:

(*i*) The bimetallic thermostat is comparatively crude.

(*ii*) To devise a method to have different temperatures at different times of the day is complex and not easily achieved.

The *"mechatronic solution"* for the above system may involve the use of a *microprocessor-controlled system using perhaps a thermodiode as the sensor*. Such a system has the following *advantages over the traditional design*:

(*i*) The microprocessor-controlled system can cope easily with giving precision and programmed control;

(*ii*) The system is much more flexible.

8.4 EMBEDDED SYSTEMS

The term *embedded system* is used for a *microprocessor-based system* that is designed to control a function or range of functions and is not designed to be programmed by the system user. The programming is done by the manufacturer and is burnt into the memory system and *cannot be changed by the system user.*

Example: A modern domestic washing machine has an embedded microcontroller which is programmed with different washing programs; here the operator does not have to program the microcontroller. The operator has only to select the required washing program by means of a switch and the required program is implemented.

- In an embedded system the manufacturer makes a ROM containing the program. This is only economical, if there is a need for a larger number of these chips. Alternatively, for prototyping or low volume applications, a program could be loaded into the EPROM/EEPROM of the application hardware.

8.5 DESCRIPTION OF SOME MECHATRONIC SYSTEMS

The following mechatronic systems are briefly described below :

1. Engine management system.
2. Automatic camera.
3. Automatic washing machine.

8.5.1. Engine Management System

The ignition and fuelling requirements of a car engine are fulfilled by the car's engine management system. In a four stroke internal combustion (I.C.) engine there are many cylinders, each of which has a piston connected to a common crankshaft and each of which carries out the four strokes namely, suction stroke, compression stroke, power stroke and exhaust stroke. Fig. 8.1 shows the working of a single-cylinder, four-stroke petrol engine.

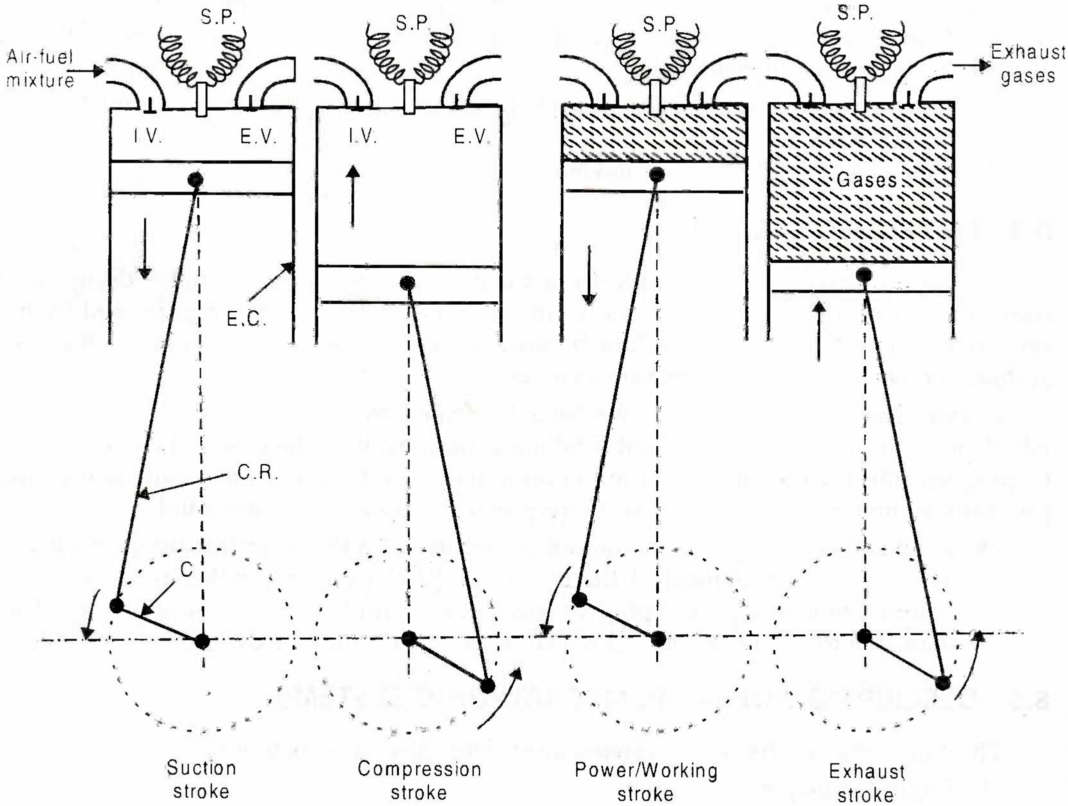
— During the *suction stroke*, when the piston moves down the inlet-valve (I.V.) opens and the air-fuel mixture is drawn into the cylinder; the exhaust valve (E.V.) however remains closed.

— During the *compression stroke* the piston moves up and the air-fuel mixture is compressed; both the inlet and exhaust valves do not open during any part of this stroke. When the piston is near the top of the cylinder the spark plug ignites the mixture with a resulting expansion of the hot gases.

— During *power/working stroke* the hot gases expand, thus doing work on the piston.

— During *exhaust stroke*, the piston moves up, forcing the exhaust gases to escape to the atmosphere through the exhaust valve.

The piston of each cylinder are connected to a common crankshaft and their power strokes occur at different times so that there is continues power for rotating the crankshaft.



I.V=Intel valve, E.V. = Exhaust valve, E.C. = Engine cylinder, C.R. = Connecting eod, C = Crank, S.P. = Spark plug

Fig. 8.1. Four-stroke sequence of an I.C. engine.

The power and speed of the engine are controlled by varying:

— Ignition timing;

— Air-fuel mixture.

In a modern car the above mentioned operations are carried out through a *microprocessor*. Fig. 8.2 shows the *basic elements of an engine management system*:

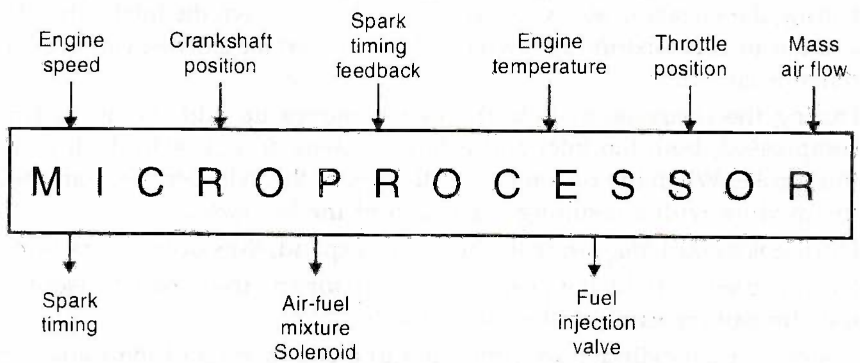


Fig. 8.2. Basic elements of an engine management system.

— To control *ignition timing, the* crankshaft drives a distributor which makes electrical contacts for each spark plug in turn and a timing wheel; this timing wheel generates pulses to indicate the crankshaft position. The *microprocessor then adjusts the timing at which high voltage pulses are sent to the distributor* so they occur at the 'right' moments of time.

— For controlling the amount of *air-fuel mixture* entering a cylinder during the suction strokes, the microprocessor varies the time for which a solenoid is activated to open the inlet valve on the basis of inputs received of the engine temperature and throttle position.

The amount of fuel to be injected into the air stream can be determined by an input from a sensor of the mass rate of air flow, or compute from other measurements, and the *microprocessor then gives an output to control a fuel injection valve.*

8.5.2. Automatic Camera

The modern automatic camera has an automatic focussing and exposure. Fig. 8.3, shows the basic elements of the control system for an automatic camera. The *working of the system* is as follows:

— For activating the system when the switch is operated, and the camera pointed at the object being photographed, the microprocessor takes in the output from the "*range sensor*" and sends an output to the lens position drive to move the lens to achieve focussing.

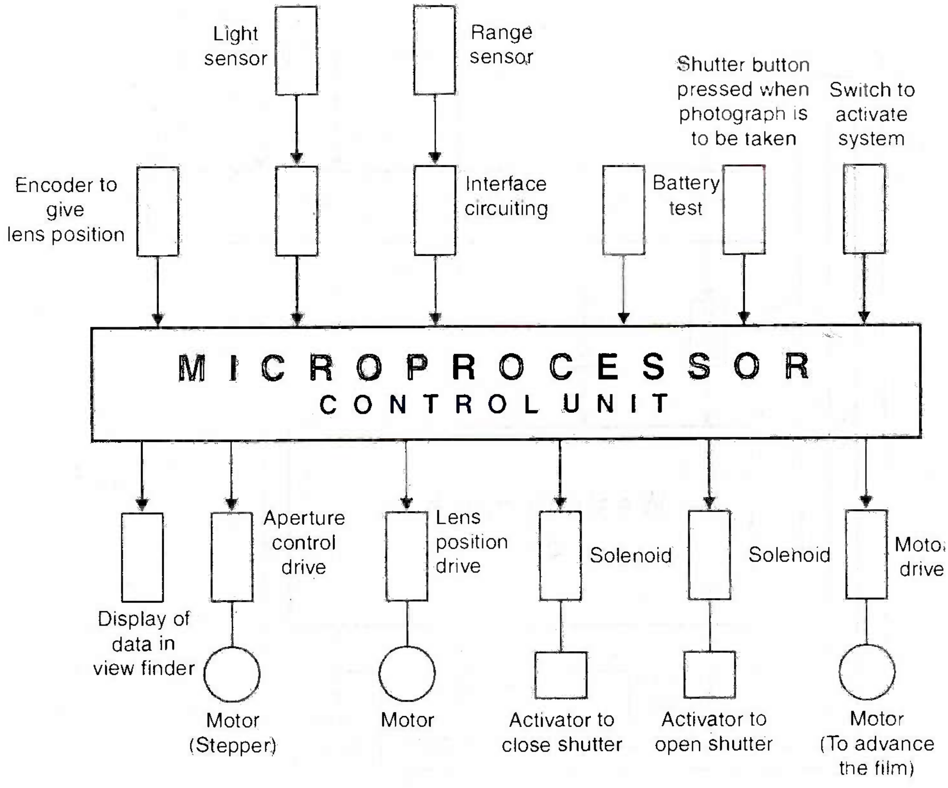


Fig. 8.3. Schematic of the control system for an automatic camera.

— The lens position is fed back to the microprocessor so that the feedback signal can be used to *modify* the lens position according to the input from the range sensor.

— The *"light sensor"* gives an input to the microprocessor which then gives an output to determine, if the photographer has selected the shutter controlled mode rather than aperture controlled mode; the time for which the shutter will be opened.

— After the photograph has been taken, the microprocessor gives an output to the motor drive to advance the film ready for the next photograph to be taken.

The cameras used in the past were adjusted for light, focussing and time or duration of aperture opening based on the sensitivity of the film including winding all being carried out manually.

- These days digital camera are flooding the market, indicating an era of digital technology. In such cameras the image of the object taken by the cameras is converted into *digital images and stored in memory* housed in the camera. Depending on the *memory size,* a large number of photographs can be shot. The photos stored in the memory can be seen on the monitor of a computer system and selection can be made. Handy cams of magnetic tape, and digital types with separate memory chips for still photographs are available in the market.

8.5.3. Automatic Washing Machine

Fig. 8.4 shows the constituent elements of a sequential controlled domestic washing machine in which control action is executed one after another operation.

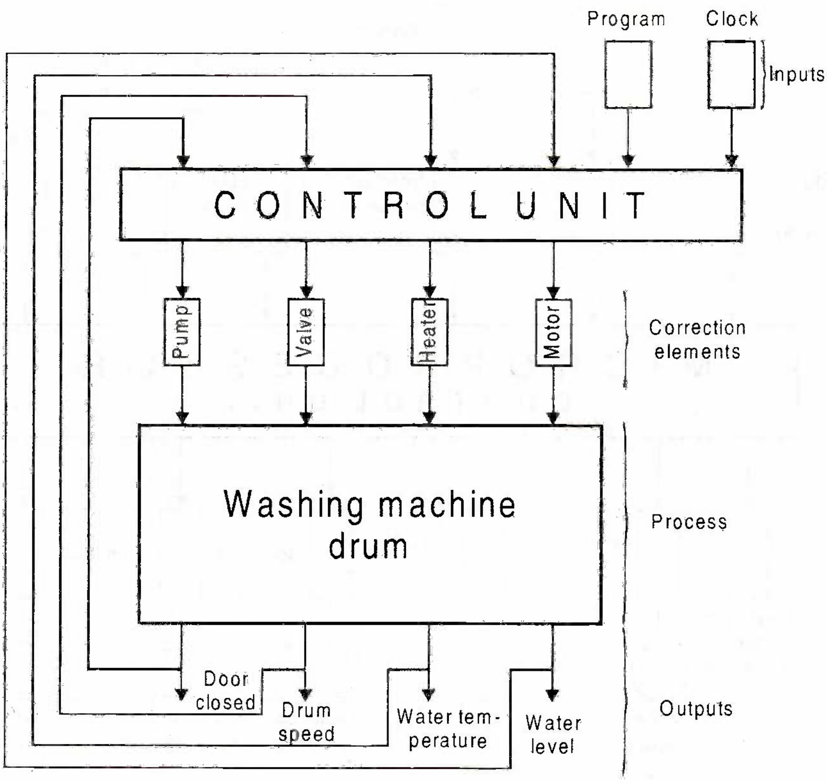


Fig. 8.4. Schematic of washing machine.

In this machine, the *following operations have to be carried out in the correct sequence:*

(*i*) A *prewash cycle* wherein the clothes in the drum are given a wash in cold water;

(*ii*) A *main wash cycle* (after the prewash cycle) when the clothes are washed in hot water;

(*iii*) A *rinse cycle* when the clothes are rinsed with cold water a number of times, followed by *spinning* to remove water from the clothes.

Earlier, all the above operations were controlled with the help of mechanical system which involved a set of *"cam-operated switches"; the contour of the cam operating different switch being proportional to time.* The sequence of instruction used was a function of set of cams used.

In modern machines the controller is a *"microprocessor"* and the program is not supplied by the mechanical arrangement of cams but by *"software program"*.

The working of a modern washing machine is as follows:

(*i*) Prewash cycle: In this cycle an electrically operated valve is opened when a current is supplied and switched off when it ceases. This valve permits cold water into the drum for a period of time determined by the output from the microprocessor used to operate its switch. In order to check the entry of water to the tank, a *sensor* is used to give signal when the water level has reached the preset level and give an output from the microprocessor which is used to switch off the current to the valve.

(*ii*) Main wash cycle:

— When the pre-wash part of the program is completed, the microprocessor gives an output for the main wash cycle; it switches a current into the circuit to open a valve to allow cold water into the drum. This level is sensed and the water shut off when the required level is reached.

— The microprocessor then supplies a current to activate a switch which applies a larger current to an electric heater, to heat the water. A temperature sensor is used to switch off the current when the water temperature reaches a preset valve.

— The microprocessor then switches on the drum motor to *rotate the drum*. This will continue for the time determined by the the microprocessor before switching off. Then the microprocessor switches on the current to a discharge pump to empty the water from the drum.

(*iii*) Rinse cycle: The rinse part of operation is now switched as a sequence of signals to open valves which allow cold water into the machine, switch it off, operate the motor to rotate the drum, operate a pump to empty the water from the drum, and repeat this sequence a number of time. The microprocessor, then finally switches on the motor, at a higher speed than for the rinsing, to *'spin the clothes'*.

8.5.4. List of Some Other Mechatronic Systems

A list of some other mechatronic systems is as follows:

- Automatic ice makers and freezers;
- Dishwashers;
- Welding robots;
- Automatic guided vehicles (AGVs);
- Variable speed drills;
- Video game and virtual reality input control systems;

- Fax machines, document scanners and other semiautomatic office equipment;
- Antilock brake systems, remote automatic door locks and other automobile systems;
- NC lathes;
- Ultrasonic probes, and other medical diagnostic equipment.
- VCRs, video and CD players, camcorders and other sophisticated consumer electronic products;
- Various systems on airplanes;
- Material testing machines;
- Manual and semiautomatic controllers for hydraulic cranes and other construction equipments;
- Laser printers, hard drive head positioning systems and other computer peripherals;
- Factor automation systems etc.

THEORETICAL QUESTIONS

1. Discuss briefly the 'control architectures' useful in mechatronic systems.
2. Enumerate the various stages involved in the design of a system.
3. Explain briefly the following:
   (*i*) Traditional and mechatronics designs.
   (*ii*) Embedded systems.
4. Discuss briefly the following mechatronics systems:
   (*i*) Engine management system.
   (*ii*) Automatic camera.
   (*iii*) Automatic washing machine.

CHAPTER

9 Elements of CNC Machines

9.1 INTRODUCTION TO NUMERICAL CONTROL OF MACHINES AND CAD/CAM

9.1.1. Modern Machine Tools

Newer machine tools have been built to absorb newer machining technologies to cope with newer and tougher materials. New technologies include Ultrasonic Machining (USM). Electro-Chemical Machining (ECM), Laser Beam Machining (LBM) etc. Besides this the advancement in electronics and application of computer in the machine tools have brought in a significant and revolutionary change in the machine tool control concept. This has given birth to an entirely new generation of machine tools. Numerically Controlled (NC) machine tools are highly flexible and are economical for producing a single or a large number of parts. **Numerical Control, NC** can be defined simply as *control by numbers.* A machine tool having a *dedicated computer to help prepare the program and control some or all of the operations of the machine tool is* called **Computer Numerical Control (CNC)** *machine tool.*

9.1.2. NC Machines

Introduction :

NC machines assimilate a method of automation, where automation of medium and small volume production is done by some controls under the instructions of a program. The definition of NC (Numerical Control) as given by EIA (Electronic Industries Association) is as under.

"A system in which actions are controlled by the direct insertion of numerical data at some point. The system must automatically interpret at least some portion of this data."

In NC machines, the input information for controlling the machine tool motion is provided by means of punched tape or magnetic tapes in a coded language.

Working of NC machine tool :

Fig. 9.1., shows the working sequence of a NC machine tool viz-a-viz operator controlled machine tool

- The first two steps, component drawing and process planning are similar in both operator controlled and NC machine tools.

- In the operator controlled machine tools, the operator controls the cutter position during manufacture and also makes necessary adjustments and corrections to produce the desired component.
- However, in NC machine tool the *operator is replaced by the data processing part of the system and the control unit.*
- In the data processing unit, the co-ordinate information regarding the component is *recorded on a tape by means of a teleprinter.*
- Tape is fed to the control unit which sends the *position command signals to slideway transmission elements of the machine.* At the same time, the command signal is constantly *compared* with the actual position achieved, with the help of position *feedback signal derived from automatic monitoring of the machine tool slide position.* The difference in two signals, if any, is corrected until the desired component is produced.

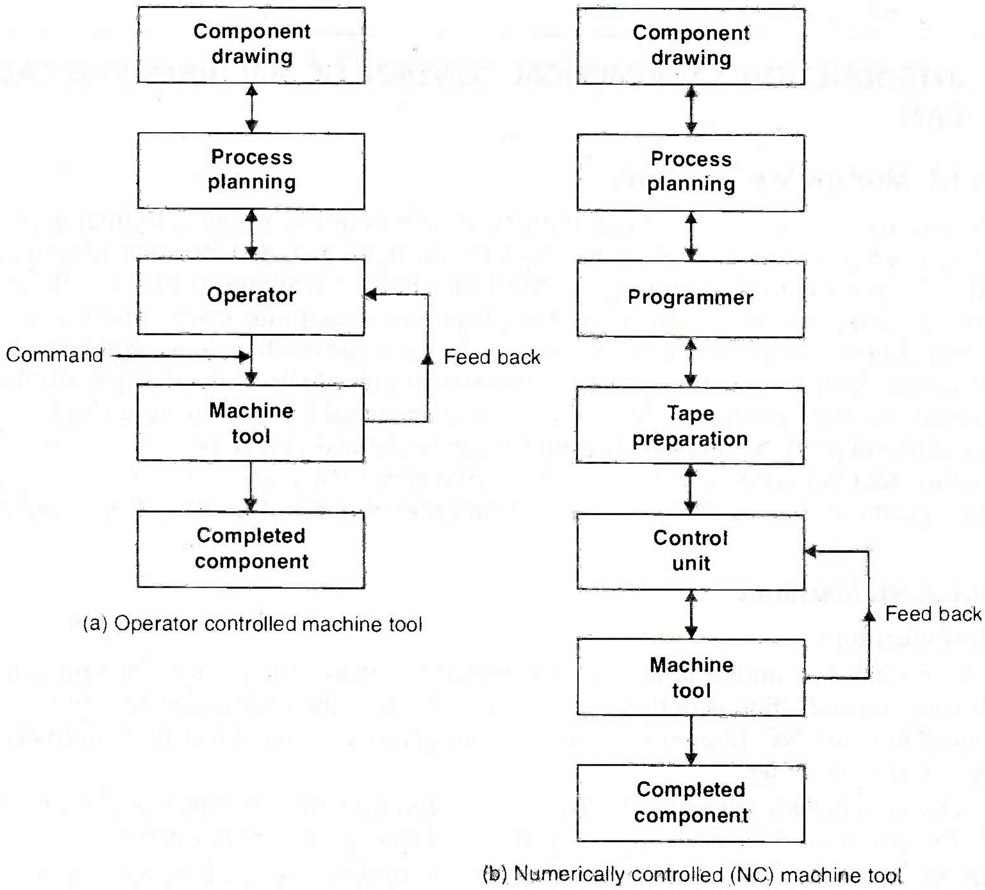


Fig. 9.1.

Main elements of a NC machine tool :

Refer to Fig. 9.2. The *main elements* of a NC machine tool are :

1. The control unit (also known as NC console or Director).
2. The drive units.

3. The position feedback package.
4. Magnetic box.
5. Manual control.

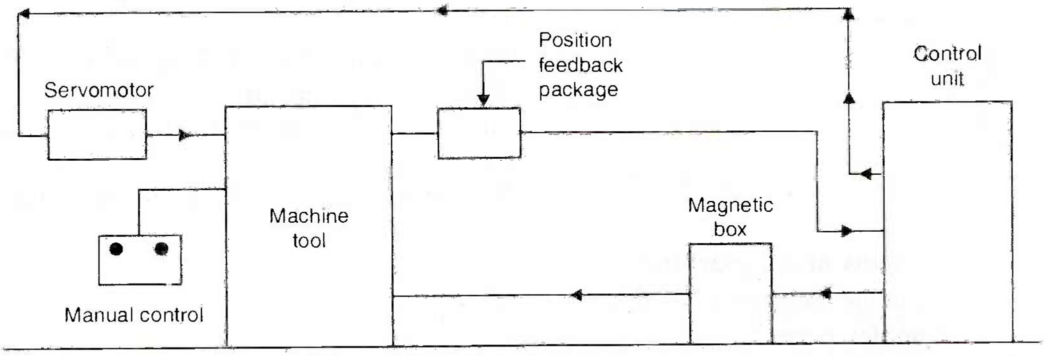


Fig. 9.2. Main elements of a NC machine.

- In the **control unit**, a tape recorder reads the instructions (written in a coded language) for manufacturing the component.
- The instructions under electronic processing and the control unit sends command signals to the **drive units** of the machine tool and also to the **magnetic box** (Electrical control cabinet). Command signals sent to the *drive units* of the machine tool, *control the length of travel and feed rates*, while the command signals sent to the *magnetic box* control other functions such as *spindle motor starting and stopping, selecting spindle speeds, actuation of tool change, coolant supply etc.*
- A **feedback transducer** provided in the machine tool checks whether the required lengths of travel have been obtained. It sends the information of the actual position achieved to the control unit. In case there is any difference between the input command signal and the actual position achieved, the drive unit is actuated by suitable amplifier from the error signal.
- **Manual control** provided in the machine tool assists the operator to perform some functions manually such as motor start-stop, speed change, feed change, axes movements, coolant supply etc.

Classification of NC machines :

NC machines may be *classified* as follows :

A. *According to control system:*

1. Point-to-point system .... The machining is done at specific positions.
   Example : *Drilling machine* operation.
2. Straight line system .... It is an extension of point to point system.
   Example : *Stepped turning on lathe, pocket milling* etc.
3. Contour system .... There are continuous, simultaneous and co-ordinated motions of the tool and workpiece along different coordinate axes.
   Example : Machining of profiles, contours and curved surfaces.

B. *According to feedback :*

1. Open loop system	....	There is no *'feedback'* and no return signal to indicate whether the tool has reached the correct position at the end of operation or not. **Example :** *Co-ordinate drilling machine.*
2. Closed loop system	....	A *feedback* is built into the system, which automatically monitors the position of the tool. It is more expensive than an open loop system.

Applications of NC machines :

The *major applications* of NC machines are :

1. Complex parts.
2. Parts which are frequently subjected to design changes.
3. Repetitive and precision quality parts which are to be produced in low to medium batch quantity.
4. To cut down lead time in manufacture.
5. In situations where the investment on tooling and fixture inventory will be high if parts are made on conventional machines tools.

Advantages of NC machines :

Following are the *advantages* of NC machines :

1. Accuracy achieved is of high order.
2. Reduced production cost per piece.
3. Less scap.
4. High production rates.
5. Less operator skill required.
6. Excellent reliability.
7. Tooling cost low.
8. Less cycle time and increased tool life.
9. Increased flexibility.
10. Production of complex parts.
11. Reduced set-up time.
12. Elimination of special jigs and fixtures.
13. Reduced inspection.
14. Lower labour cost.
15. Reduced floor space.
16. Easy and effective production planning.

9.1.3. CNC Machines

In a CNC machine, a *minicomputer is used to control machine tool functions from stored in information or punched tape input* or *computer terminal input.*

The definition CNC (Computer Numerical Control) as given by EIA is as under :

"The numerical control system where a dedicated, stored program computer is used to perform

some or all of the basic numerical control functions in accordance with control programmes stored in read/write memory (RAM) of the computer".

CNC may also be defined as : *"An NC system with a microcomputer or microprocessor using software to implement control algorithms."*

Fig. 9.3, shows the control unit and panel of a CNC. The following *points* about CNC machines are worth noting :

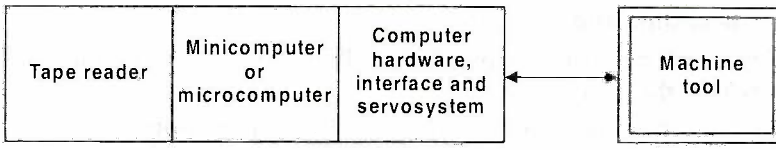


Fig. 9.3. Computer Numerical System (CNC).

- The control unit and a panel of CNC differs from that of NC controls in that, it works in *ON-line mode* whereas NC works in *batch processing mode.*
- A typical CNC may need only the drawing specifications of a part to be manufactured and the computer automatically generates the part program for the loaded part.
- The part program once entered into the computer memory can be used again and again.
- The input information can be reduced to a great extent with the use of special sub-programs developed for repetitive machining sequences.
- The CNC machines have the facility for proving the part program without actually running it on the machine tool.
- CNC control unit allows compensation for any changes in the dimensions of the cutting tool.
- With CNC control systems, it is possible to obtain information on machine utilisation which is useful to the managements.

Functions of CNC :

The principal functions of CNC are :

1. Machine tool control.
2. In-process compensation.
3. Improved programming and operating features.
4. Diagnostics.

Advantages of CNC machines :

CNC machines offer the following *advantages* in manufacturing :

1. Greater flexibility.
2. Reduced data reading error.
3. Increased productivity.
4. Consistent quality.
5. Automatic material handling.
6. Elimination of operator errors.
7. Reduced operator activity.
8. Lower labour cost.
9. Smaller batches.

10. Longer tool life.
11. Just-in-time (JIT) manufacture.
12. Reliable operation.
13. Elimination of special jigs and fixtures.
14. Reduced inspection.
15. Less scrap.
16. Accurate costing and scheduling.
17. CNC machine can diagnose program and can detect the machining malfunctioning even before the part is produced.
18. Conversion of units – possible within computer memory.

Disadvantages of CNC machines :

1. Higher investment cost.
2. Higher maintenance cost.
3. Costlier CNC personnel.
4. Airconditioned places are required for the installation of the machines.
5. Unsuitable for long run applications.
6. Planned support facilities.

Applications of CNC :

CNC is being used in the following machines/areas :

- Drilling machines.
- Turning machines.
- Boring machines.
- Milling machines.
- Grinding machines.
- Pipe bending machines.
- Coil winding machines.
- Flame cutting machines.
- Welding, wire cut EDM and several other areas.

9.1.4. CAD/CAM

CAD/CAM (Computer-Aided Design/Computer-Aided Manufacture) technology was initiated in the aerospace industry but presently it is spreading at a rapid pace in all industries.

It can be *defined* most simply as the *use of computers to translate a product's specific requirements into the final physical product.*

Following points are worth noting about CAD/CAM technology :

- With *this system, a product is designed, produced and inspected in one automatic process.*
- It plays a key *role in areas* such as *design analysis, production planning, detailing, documentation, N/C part programming, tooling fabrication, assembly, jig and fixture design, quality control, and testing.*
- Whenever any deviation is noted, a programmable controller takes automatic corrective action to compensate for the deviation. Thus a *closed loop system* is formed which produces *consistent quality products, reduces wastes* and *improves productivity.*

- CAD/CAM system is *ideally* suited for *designing and manufacturing mechanical components of free form complex with three dimensional shapes.*

9.1.4.1. CAD

Definition :

In the modern sense, CAD (Computer Aided Design) is defined as :

"A design process using sophisticated computer graphics techniques, backed up with computer software packages to aid in the analytical, development, costing and ergonomic problems associated with design work".

Advantages :

The following are the *advantages* of CAD :

1. Drawings can be produced at a faster rate.
2. Drawings produced by CAD systems are more accurate and neat.
3. In this system there is no repetition of the drawings.
4. CAD systems assimilate several special draughting techniques which are not available with conventional means.
5. Design calculations and analysis can be carried out quickly.
6. With CAD systems superior design forms can be produced.
7. CAD simulation and analysis techniques can drastically cut the time and money spent on prototype testing and development – often the costliest stage in the design process.
8. Using CAD systems design can be integrated with other disciplines.

9.1.4.2. CAM

General aspects :

CAM (Computer-Aided Manufacture) concerns any automatic manufacturing process which is controlled by computers.

The most important elements of CAM are :

1. CNC manufacturing and programming techniques.
2. Computer controlled robotics manufacture and assembly.
3. Flexible Manufacturing Systems (FMS).
4. Computer Aided Inspection (CAI) techniques.
5. Computer Aided Testing (CAT) techniques.

Advantages :

CAM entails the following *advantages* :

1. Product obtained is superior in quality.
2. The manufactured form has a greater versatility.
3. Higher production rates with lower work-forces.
4. There is less likelihood of human error.
5. As a result of increased manufacturing efficiency cost savings are materialised.
6. The production processes can be repeated via storage of data.

9.1.4.3. Software and hardware for CAD/CAM

The functions of CAD/CAM systems are mainly determined by the *software.*

Software usually consists of a *number of separate application packages to perform the desired function.* The size of computer depends on the number and sizes of packages and number of work stations.

Hardware is responsible for the reliability and speed of response of the system.

A wide range of standard software is available and generally it is not worth developing users own software. Though a system can be built up from standard software packages from different sources and standard hardware, it is often costly because of the considerable programming effort required to interface the packages to a common data base to provide user friendly software to adapt the system to the user's requirements. It is thus *advisable to adopt turn key system for turn key suppliers.*

9.1.4.4. Functioning of CAD/CAM system

- CAD/CAM is *an interactive computer graphic tool that enhances design and manufacturing functions to create a highly profitable product.* This technique is being applied by big industries for improving overall manufacturing performance.
- It is *not* a standard tool which can be fitted into any company but has to be tailored to suit the needs of the company. It is rather complex technology and has wide potential for immediate benefits.
- Usually this *tool consists of a dedicated computer, which is connected to a number of work-stations.* The system is used to assist in the design and manufacturing, through the use of an *expandable set of linked software modules.* A designer can define dimensions and display views of 2 dimensions, $2\frac{1}{2}$ dimensions and 3 dimensions parts on modules. It is possible to generate the families of part directly by a parametric processor either by direct scaling or using a catalogue of subprograms. From the geometric definition a solid model can be constructed, to assist in visualisation. It is possible to store complete details of designs on numerical control types for subsequent use on demand. Bench making tests are carried out to ensure system's capability.

9.1.4.5. Features and characteristics of CAD/CAM systems

The following are the *features and characteristics* of CAD/CAM system :

1. A major portion of the output of the engineering sector involves batch production and CAD/CAM offers immense cost and quality benefits for such requirements.
2. The work-in-progress, in batch production, is reduced considerably.
3. It is possible to produce at random all the variants and series of a product planned to be manufactured by a firm.
4. Such a system has inherent flexibility to cater to new models of the product in pipeline without major modification.
5. In such a system, several machining centres are arranged one after the other with robots and proper automatic materials handling equipment. Software is developed to integrate the machine CNC control and the handling system. Each machining centre is equipped with several tool magazines. All the tools required to complete each operation on each model of the product can be stored in the magazine.
6. All the part programs for the different models are stored in the memory. System has only to identify the model of the product presented to a machine in order to complete the machining operations. Thus it is possible to have totally random mixes of models of a product proceeding down the line at any one time.
7. The system can be conceived in multiplies of 15-20 minutes operations. If certain operations take longer, then multiples of similar machines can be installed in the line. Sometimes identical machines are introduced for each operation so that production can continue even if one machine goes down.

8. - The components are loaded on to a pallet. Means are provided to identify the exact model.
   - Loaded pallets enter the line and wait at the start of the line until a signal that one of the first operation machines is vacant is obtained.
   - The handling system automatically directs the pallet to the first vacant machine for first operation.
   - The pallets are loaded on a fixture. The fixture is designed so that it permits access to all four sides and end faces and wherever machining operation is required. The pallets are designed to have windows where access for machining is required.
   - As the pallet enters the machining area, air blast clears both the fixture and pallet locations. The fixture is then properly clamped and supported. Touch trigger probes are used to check its location in the pallet.
   - Probes also identify the exact model of the component and signals from the probes active master calling program which selects the appropriate past program and sub-routines from the control memory.
   - An overhead cascade coolant wash is provided to clear away swarf before the pallet is located. All coolant and swarf is carried away via underground ducts to a central separation and coolant filteration plant.

     Some systems can show metal being removed dynamically.
   - It is possible to store libraries of standard tools and tool holders, thus carrying out process planning.
   - By calling up and manipulating standard fixturing components, like studs, stops, clamps, bushes, location devices, fixtures etc., it is possible to design a fixture for a component already designed on the CAD/CAM system.
   - It also allows sheet metal development (unfolding), taking account of the material for the bends. It is also possible to layout sheet metal components on a standard sheet to reduce the waste (nesting). Factory layout process planning and robot programming have also been attempted.
   - Exploded views, schematics and diagrams, 3-D colour shades like photographic views of the parts can be produced.
   - Tenders and estimates can be quickly produced to high quality.

9.1.4.6. Application areas for CAD/CAM

The potential application areas for CAD/CAM are :

1. Design and design analysis :

- CAD system would be best suited for *drawing offices where frequent modifications are required on drawing and several parts repeat.*
- It must be remembered that it is very easy with computer to make modifications and very fast to draw part profile once its details are fed into computer.
- Once a drawing is entered in the CAD system, later modifications can be done quickly, and detail drawings can be prepared quickly from a general arrangement drawing.
- NC tapes can be produced.
- Storing of the drawing is very convenient, easy, occupies very less space and symbols for electrical, hydraulic, control and instrumentation circuits can be called up quickly and positioned on the schematic drawing.

- Standard components can be stored permanently in the data base and called up and positioned on the drawing, resulting in saving of time and enforcement of standards. It is possible to associate nongraphical information like past number, supplier, material etc., for any component assembly.
- It is very convenient to calculate properties like weight, centre of gravity, moment of inertia, etc., because 3-D models can be easily produced.
- It is also possible to carry out finite element analysis by producing meshing for analysis.

2. **Manufacture :**

- With CAD/CAM system the complete NC part programming process can be carried out interactively, including post processing and production of NC tape. Source programs in languages such as APT can be produced. Systems can verify tapes by producing tool centre path plots.

9.2 ELEMENTS OF CNC MACHINES

9.2.1. Introduction

A *Computer numerically controlled* ***(CNC) machine*** *(Fig. 9.3) is a mechatronic system since the machine tool which is a mechanical system is incorporated or integrated with the electronic controls for its different drives and computer system for interfacing the software with the mechanical and electronic system.*

Hardware or electronic circuits control the motions of various drives.

The design and construction of CNC machines differs greatly from that of conventional machine tools. This difference arises from the requirement of higher performance levels. The CNC machines often employ the various *mechatronic elements* that have been developed over the years. However, the quality and reliability of these machines depends on the various *elements* and *subsystems* of the machines.

The following are some of the important constituent parts, and aspects of CNC machines to be considered in their designing :

1. Machine structure.
2. Guideways/Slideways.
3. Drives.
4. Spindle and spindle bearings.
5. Measuring systems.
6. Controls.
7. Gauging
8. Tool monitoring.
9. Swarf removal.
10. Safety.

9.2.2. Machine Structure

The *"machine structure" is the load carrying and supporting member of the machine tool.* The design and construction of CNC machine should be such that it meets the main *"objectives" (i) High precision and repeatability, (ii) reliability; (iii) Efficiency.* In order to meet these requirements, the numerically controlled machine tools should have a structure with the following *characteristics* :

1. It does *not deform or vibrate beyond the permissible limits* under the action of static and dynamic forces, to which it is subjected.

- ***Static load*** of a machine tool results from the weights of slides and the workpiece, and the forces due to cutting.
- ***Dynamic load*** is a term used for the constantly changing forces acting on the structure while the movement is taking place. These forces cause the whole machine to vibrate and the origin of these vibrations may be due to *unbalanced rotating parts, improper meshing of gears, bearings irregularities, and interrupted cuts while machining* (as in milling). These revibrations can be reduced by : (*i*) *Improving the damping properties*, (*ii*) *Reducing the mass of structure and increasing the stiffness of the structure.*

2. Its design should be such that the thermal distortion is minimum. The machine tool should be protected from external and internal heat sources; some of these heat sources are : *Electric motor; friction in mechanical drives, gear boxes, bearings and guideways; machining process; temperature of surrounding objects.*

- Thermal deformation due to ***thermal load*** may be *reduced* by :
  - (*i*) Designing the structure thermo-symmetrically.
  - (*ii*) External mounting of drives.
  - (*iii*) Using a proper lubrication system for removing frictional heat from bearings and guideways.
  - (*iv*) Removing the coolant and swarf efficiently for the dissipation of heat generated from the machining process.

3. The machine structure design should be such that the removal of swarf is easy and the chips etc., do not fall on the slideways.

9.2.3. Guidways/Slidways

9.2.3.1. Introduction

In machine tools the guideways are used to serve the following *purposes* :

(*i*) To control the direction or line of action of the carriage or the table on which a tool or a workpiece is held.

(*ii*) To absorb all static and dynamic loads.

The guideways may be an integral part of the machine structure or may be mounted separately on the structure. These guidways may be horizontal, vertical or inclined. However *vertical and inclined guideways are preferred so that chips produced during the cutting operation do not get collected on the quickways.*

The shape and size of the work produced depends on the accuracy of the movement and kinematic accuracy of the guideway. Kinematic accuracy depends on the straightness, flatness and parallelism errors in the guideway.

- In a CNC *machine* the design of guideway/slideway should :

(*i*) Reduce friction;

(*ii*) Reduce wear;

(*iii*) Satisfy the requirements of movement of the slides;

(*iv*) Improve smoothness of the drive

9.2.3.2. Factors influencing the design of guideways

The following *factors* should be considered while designing guideways :

1. Geometric and kinematic accuracy.
2. Position in relation to work area.

3. Provision for adjustment of play.
4. Rigidity.
5. Damping capability.
6. Velocity slide.
7. Friction characteristics.
8. Wear resistance.
9. Protection against swarf and damage.
10. Protective guards to safeguard the guideways against accidental damages.
11. Freedom from unnecessary restraints.
12. Effective lubrication and efficient lubrication systems.

9.2.3.3. Types of guideways

Guideways are broadly *classified* as follows :

1. Friction guideways.
   (*i*) Vee guideways.
   (*ii*) Flat guideways
   (*iii*) Dovetail guideways.
   (*iv*) Cylindrical guideways
2. Antifriction linear motion (LM) guideways.
3. Frictionless guideways :
   (*i*) Hydrostatic guideways.
   (*ii*) Aerostatic guideways.

- Other types of guideways employed in machine tools are :
  1. Hydrostatic guideways.
  2. Aerostatic guideways.

9.2.3.4. Friction guideways

- These guideways find wide application in conventional machine tools due to their *low manufacturing cost and good damping properties.*
- They *operate under conditions of sliding friction and do not have a constant coefficient of friction.* The frictional coefficient varies with the sliding velocity as shown in Fig. 9.4.

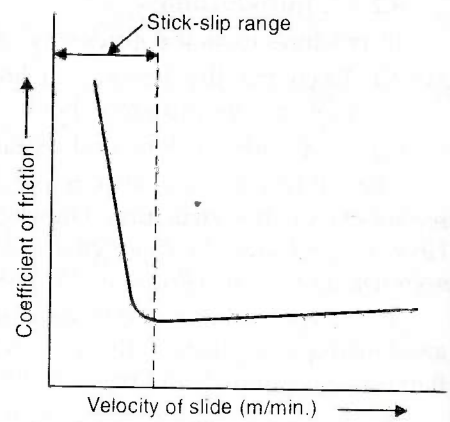


Fig. 9.4. Coefficient of friction v/s velocity of slide graph for friction guideways.

— At the commencement of the movement, the coefficient of friction is very high, but as the velocity increases it falls rapidly and beyond a certain critical velocity it remains almost constant. Thus, to start motion/movement, the force to overcome friction has to be correspondingly high. This force results in the drive mechanism, such as a screw, being elastically deformed.

— With the increase in speed, the friction decreases and a greater amount of movement than that intended for the slide takes place; this may lead ultimately to a jerky motion. This phenomenon is known as *"stick-slip phenomenon".*

The possibility of this phenomenon can be reduced by using materials such as *PTFE* (Poly tetra fluoro ethylene) and *turcite* at the quideways interface; these materials have a low and constant coefficient (of the order of 0.1).

(i) Vee guideways :

Fig. 9.5, shows Vee and inverted Vee guideways.

— In case of Vee guideways with *apex upwards*, there is no chip falling or accumulation. In this case lubrication is difficult. In case of inverted Vee guideways there is a possibility of falling and accumulation of chips; however lubrication is easier.

- The Vee guideways are widely used on machine tools, especially on lathe beds.
- One of the advantages of Vee guideways is that the parallel alignment of the guideway with the spindle axis is not affected by wear.
- These guideways wear away rapidly due to lack of bearing surface. These are difficult to manufacture.

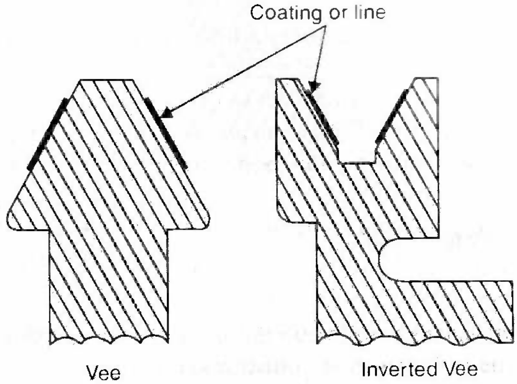


Fig. 9.5. Vee guideways.

(ii) Flat guideways :

Fig. 9.6, shows a flat form of guideways.

— These guideways have better load bearing capabilities than other guideways.

— These are easier to manufacture.

— In such guidways the chip accumulation and lubrication problems are serious.

— These do not wear uniformly.

— Jibs are used to ensure accurate fitting of the slide on the flat surface.

- These guideways are suitable for *heavy load transmission.*

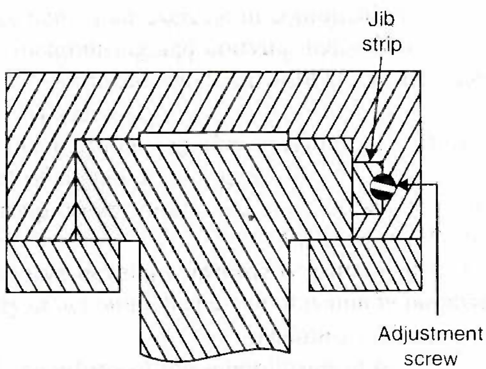


Fig. 9.6. Flat guideways.

(iii) Dovetail guideways :

Fig. 9.7, shows the dovetail form of guideways.

— These guideways have large load carrying capacity and tend to check the overturning tendency under eccentric loading.

— They are preferred when both horizontal and vertical locations of moving parts are considered essential.

— Jibs are used to ensure accurate fitting of the slide on the dovetail surface. The jibs are tapered and can be adjusted to reduce excessive clearance caused by wear.

- Although the Vee type guideways have certain advantages, it is the *flat or dovetail forms which are used on CNC machine tools.*

- The majority of lathes have a combination of Vee and flat guideways *to prevent twisting of the slide*. Provision has also to be made to prevent the carriage from lifting off the guideway.

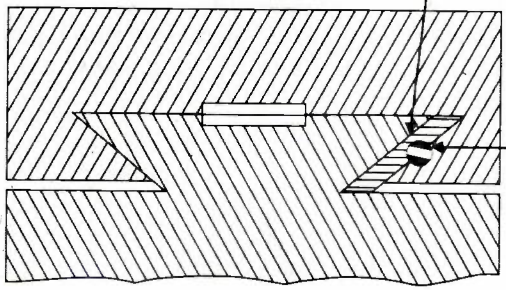


Fig. 9.7 Dovetail guideways.

Note: When the guideways are an integral part of the castings and get worn out after a period of time, it is necessary to dismantle the machine to remachine the guideways so that their accuracy is restored. *To overcome this difficulty, pre-machined hardened steel guideways are fastened to the main castings which can be replaced if they are worn out or damaged.*

(iv) Cylindrical guideways :

Fig. 9.8 shows a cylindrical form of guideways; in this case the bore in the carriage housing provides support all around the guideways.

— These guideways are *very efficient for relatively short traverses and light loads.*

— Their use for long traverses and heavy loads is not suitable because the guideways may sag or bend in the centre of the span under a load.

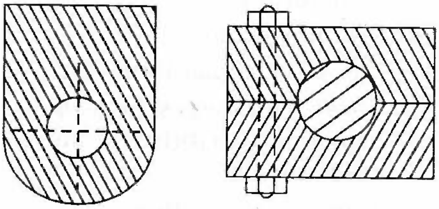


Fig. 9.8 Cylindrical or circular type guideways.

9.2.3.5. Antifriction linear motion (LM) guideways

These guideways are used on *CNC machine tools to reduce amount of wear, friction, heat generation and improve smoothness of the movement.*

- The antifriction guideways are employed to overcome the relatively high coefficient of friction in metal-to-metal contacts and the resulting limitations addressed above.
- They use *rolling elements in between the moving and stationary elements of the machine.*

Advantages : The antifriction guideways claim the following *advantages* over the friction guides :

1. High load carrying capacity.
2. Heavier preloading possibility.
3. High traverse speeds.
4. Low frictional resistance.
5. No stick-slip.
6. Ease of assembly.
7. Commercially available in ready-to-fit condition.

Disadvantage : Their main disadvantage is *'lower damping capacity'*.

- Although the rolling element bearings have less damping characteristics than friction guideways, LM guideways have become common in machine tools on account of their *rapid traverse rates.*

Types of antifriction guideways :

Although several types of antifriction guideways are put to use, yet the most commonly used in CNC machines are :

1. Linear bearing with balls.
2. Linear bearing with rollers.

1. *Linear bearing with balls :*

A linear ball bush shown in Fig. 9.9., uses *recirculating balls* within a bush type of bearing. These are designed to run along precision ground shafts and offer *frictionless movement over varying strokes of length with high linear precision.*

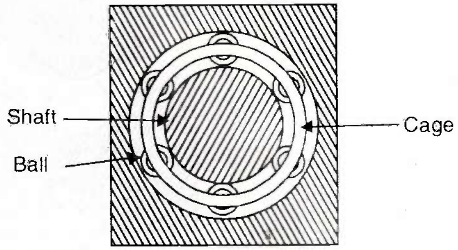


Fig. 9.9. Linear ball bushing.

2. *Linear bearing with rollers :*

The recirculating linear roller bearings are used for *movement along a flat plane.*

Their main characteristic feature is that there is *continuous roller circulation which allows unlimited linear movement.*

Fig. 9.10. shows a linear roller bearing (also called a *"tychoway"*) :

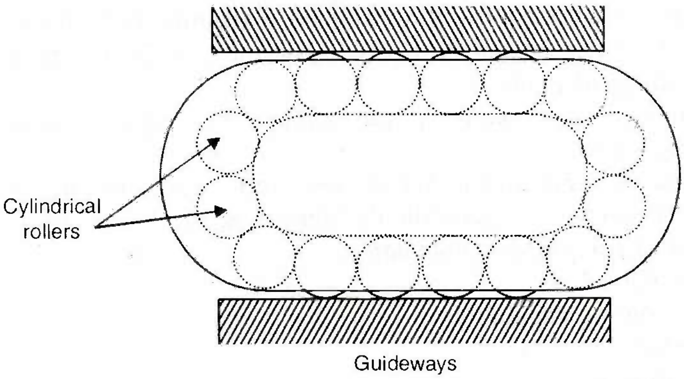


Fig. 9.10. Linear roller bearing.

— It consists of hardened and precision ground supporting elements and a number of cylindrical rollers. As in the case of roller bearings, the rollers are guided between shoulders of the supporting elements with very close tolerances.

— The grinding element prevents the rollers from falling out and sliding against each other. Also the guiding element assists in smooth return of the rollers to the loading zone.

— The rollers are in constact with guideway machined on the bed of the machine. This arrangement provides smooth and easy movement but the machine bed has to be machined to an accurate form. Also the machine bed surfaces coming in contact with the rollers have to be hardened.

— These bearings can be mounted horizontally for load carrying applications such as machine tool table or they can be mounted vertically to provide support, guidance and motion for the vertical elements of the machine tool.

- Vee and flat roller arrangement shown in Fig. 9.11., can also be used to provide frictionless linear movement.

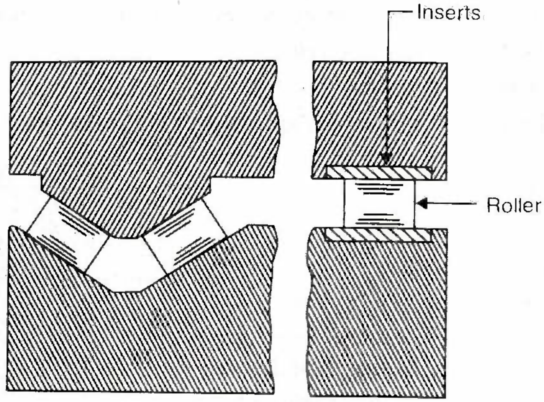


Fig. 9.11. Vee and flat roller.

9.2.3.6. Frictionless guideways

(*i*) ***Hydrostatic guideways ;***

— In these guideways the surface of the slide is separated from the guideway by a very thin film of fluid supplied at pressures as high as 300 bar.

— In hydrostatic guideways *frictional wear and stick slip are entirely eliminated.*

— In such guideways a *high degree of dynamic stiffness and damping is obtained,* both the characteristics contributing to good machining capabilities.

— Owing to high cost and difficulty in assembly, their application is limited.

(*ii*) ***Aerostatic guideways :***

In these guideways, the *slide is raised in a cushion of compressed air which entirely separates the slide and guidway surfaces.*

- Their major limitation is *low stiffness* and this limits their use to *positioning applications only e.g.,* a Coordinate Measuring Machine (CMM).

Advantages of frictionless guideways :

1. Longer life.
2. Large damping capability.
3. Frictionless.
4. High stiffness.
5. No stick-slip.
6. Less thermal distortion due to better heat dissipation.

Disadvantages :

1. Difficulty in assembling the guideways.
2. High cost.
3. Leakage problems.

Selection of guideways :

The selection of guides for a particular application basically depends upon the requirements of :

(*i*) The load carrying capacity;
(*ii*) Denaturing;
(*iii*) The traverse speed.

For getting the maximum benefit, most of the machine tool manufacturers make use of a *combination of antifriction and friction guideways with PTFE/turcite lining.* In such a combination antifriction guideways improve the load carrying capacity while friction guideways improve damping property.

9.2.4. Drives

Drives *are devices which impart motion to mechanical elements.*

- In a CNC machine tool there are three major group of elements:
  - (*i*) Control and electronics.
  - (*ii*) Electric drives (electromechanical drives)
  - (*iii*) Mechanical elements (table, slide, tool holder etc.)

In addition, there can be hydraulic and pneumatic systems which are integrated with CNC machine tool.

The *primary function of the drive is to cause motion of the controlled machine tool member (spindle, slide etc.) to conform as closely as possible to the motion commands issued by the CNC system.*

— In order to ensure a *high degree of consistency in production, variable speed drives are essential.*

— Most of the drives used in machine tools are electrical.

Depending on their characteristics, machine tool drives can be ***classified*** as follows :

1. **Spindle drives** ... (*constant power*)

(*i*) *D.C. spindle drives :*

- Separately excited D.C. shunt motor.
- Controller:
  — Thyristor (SCR) amplifier, or
  — Microprocessor based self-tuned thyristor amplifier.
- Speed control:
  — Armature and field control.

(*ii*) *A.C. spindle drives:*

- Squirrel cage induction motor.
- Controller:
  — Microprocessor based pulse width modulated (PWM) inverter
- Speed control:
  — Frequency, vector control

2. **Feed drives** ... (*constant torque*)

(*i*) *D.C. servo-drive:*

- Motor – permanent magnet.
- Controller:
  — Thyristor D.C. amplifier
  — Transistor PWM D.C. chopper
- Speed control:
  — Armature voltage

(*ii*) *A.C. servo-drive :*

- Motor – Synchronous three phase A.C. motor with permanent magnet rotor.
- Controller:
  — Transistor for PWM frequency inverter; analog drive amplifier
  — Transistor PWM frequency inverter; digital drive amplifier.
- Speed control:
  — Frequency control.

9.2.4.1. Spindle drives

The following motors are used in spindle drives :

(*i*) D.C. shunt motor (separately excited).

(*ii*) Three-phase A.C. induction motor.

The *requirements of a spindle drive motor are* :

1. Compactness.
2. High overload capacity.
3. Large speed range of at least 1 : 1000.
4. Maximum speed upto 9000 – 20000 r.p.m.
5. High rotational accuracy.
6. Range of rated output from 3.7-50 kW.
7. Wide constant power band.
8. Fast dynamic response.
9. Excellent running smoothness.

In CNC machines the D.C. spindle drives are commonly used (say for stepless speed variation of spindles). However, with the advent of microprocessor based A.C. frequency inverter, of late, the A.C. drives are being referred to D.C. drives as they offer several advantages (*e.g., more reliable, easily maintainable and less costly*).

— The *main advantage* of microprocessor-based frequency converter is the possibility of using the spindle motor for C-axis applications for speed control in the range of 1 : 10^6 with positioning.

9.2.4.2. Feed drives

The main components of a feed drive are : (*i*) A feed servomotor; (*ii*) Mechanical transmission system.

A *"feed motor"*, unlike a spindle motor, has special characteristics like *constant torque and positioning*.

In continuing operations where a prescribed path has to be followed continuously, several feed drives have to operate simultaneously; this requires a *sufficiently damped servo system with high band width, i.e.,* fast response and matched dynamic characteristics for different axes.

Following are the *requirements of CNC feed drive* :

1. High torque-to-weight ratio.
2. Integral mounting feedback devices.
3. During machining, the required constant torque for overcoming frictional and working forces must be provided.
4. Low electrical and mechanical constants.
5. Low armature or motor inertia.
6. Permanent magnet construction.
7. Total enclosed non-ventilated design.
8. Maximum speed upto 3000 r.p.m.
9. The drive should be infinitely variable with a speed range of at least 1 : 20,000.
10. Positioning of smallest position increments like 1-2 μm should be possible.
11. Four quadrant operation – quick response characteristics.
12. High peak torque for quick responses.

For CNC machines the commonly used feed drives are D.C. and A.C. servomotors. Although earlier D.C. servomotors, because of their excellent speed regulation, high torque and efficiency, were used most commonly on CNC machine, *but now A.C. servomotors have become more popular for machine tool applications because of the following "characteristics"* :

(*i*) Higher reliability as composed to D.C. servomotors.
(*ii*) Provide a constant torque over their entire speed range.
(*iii*) Require less maintenance due to brushless operation.
(*iv*) Provide a better response and dynamic stiffness.
(*v*) Excellent temperature resistance.
(*vi*) Fast response.
(*vii*) Increased power density.
(*viii*) Low rotor inertia.

- All the axes in a CNC machine are controlled by ***servomotors***. The movement along the different axes is required either to move the cutting tool or the work material to the designed positions.
- In order to accomplish accurate control of position and velocity, ***stepper motors*** are used for axis drive. The use of stepper motor considerably simplifies the system as feedback devices are not used. The cost of machine tool is also less. *The steppers motors are suitable only for light-duty machines due to low power output.*

Mechanical transmission systems :

The mechanical transmission system of a feed drive consists of the following elements :

1. Elements to convert the rotary motion to a linear motion (*Recirculating ball screw-nut or rack-and-pinion system*).
2. Elements to transmit torque (*gear box or timing belt and couplings*).

To keep the *transmission error to a minimum* is the primary requirement in the design of a mechanical transmission system. To achieve this, the following requirements are essential :

(*i*) Low friction. (*ii*) High Stiffness.
(*iii*) Sufficient damping. (*iv*) Backlash free operation.
(*v*) High natural frequency.

I. *Recirculating ballscrew and nut :*

In ballscrews, the sliding friction encountered in conventional screws and nuts is replaced by rolling friction in a manner analogous to the replacement of simple journal bearing by ball bearing.

Fig. 9.12, shows the recirculating ballscrew and nut arrangement.

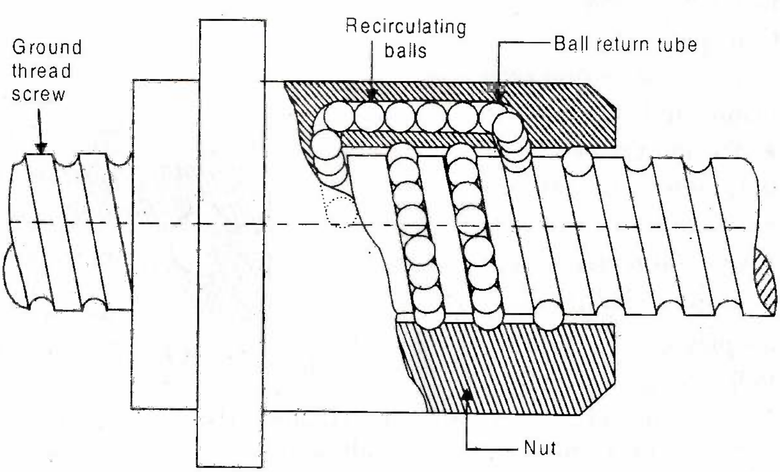


Fig. 9.12. Recirculating ballscrew and nut arrangement.

— The mounting arrangement of a ballscrew depends on its required ~~speed,~~ length and size. The position of the ballscrew should be near the line of the resultant force arising from cutting, frictional and inertial forces.

— The efficiency of a recirculating ballscrew is of the order of 90 percent and is obtained by the balls providing a rolling motion between the screw and the nut.

— In a ballscrew system, attention should be paid to the selection of end bearings to minimise the positioning inaccuracies.

— The ballscrews used on CNC machines are usually of precision grade.

Advantages :

The recirculating ballscrews are widely used on CNC machines because of the following *advantages* :

(*i*) High efficiency.

(*ii*) No stick-slip effect.

(*iii*) Low frictional resistance.

(*iv*) Low drive power requirement.

(*v*) High traverse speed.

(*vi*) Less wear and hence longer life.

(*vii*) Little temperature rise.

Preloading of nuts :

One of the primary requirements of screw and nut mechanism in CNC machines employed for motion transmission is that there should not be any backlash and if any should be minimum between the screw and nut. *Backlash free motion results in the slide travelling without any positioning error.*

As backlash cannot be completely eliminated but can be reduced; *preloading* concept is often used to achieve bare minimum backlash. ***Preloading*** *is the process of applying initial load to the nut which will cause elastic deformation of the screw threads in the axial direction, thereby increasing the axial rigidity of the ballscrew nut.*

Preloading of ballscrews nut may be *classified* as follows :

Preload ballscrews :

1. *Constant pressure preload ballscrews :*
   (*i*) Belleville spring.
   (*ii*) Coil spring.
2. *Constant position preload ballscrews :*
   (*i*) Double nut type preload:
   - Tension preload type.
   - Compression preload type.

   (*ii*) Single nut type preload:
   - Integral preload type.
   - Oversize ball preload type.

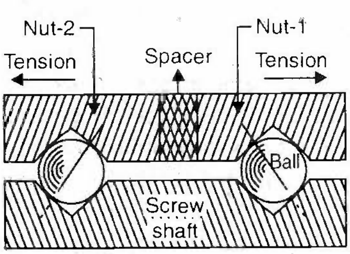


Fig. 9.13. Tension preload.

(*i*) *Tension preload :*

- Refer to Fig. 9.13 :

— Tension preload provides the required amount of preload by the insertion of a spacer of specified width (depending upon the desired amount of preload) between the two nuts.

— Each nut exerts pressure on its respective ball, thus forcing the balls away from each other.

(ii) *Compression preload :*

- Refer to Fig. 9.14 :

— In this case also, the preload is achieved through the insertion of a spacer, between the nuts but the pressure is applied (to the nuts) in the *opposite direction*, squeezing them together and forcing the balls against threads of the screw shafts.

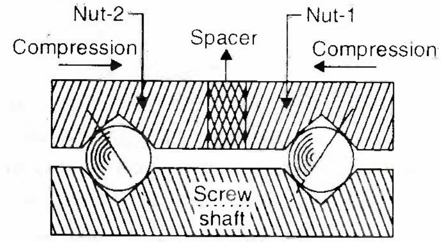


Fig. 9.14. Compression preload.

— The spacer thickness, as in tensile preload also depends on the desired amount of preload.

(iii) *Oversize preload :*

- Refer to Fig. 9.15 :

— For single nut ball screws, one type of preload is accomplished by using balls which are just *slightly larger* than the space provided between the nut and screw shaft. This method is best suited to provide *comparatively light preload, to the extent of eliminating axial clearance.*

— In oversize ball preload, the *balls have 4-point contact, thereby increasing the operational efficiency;* the standard non-loaded spacer balls, however, are used in a ratio of 1 : 1.

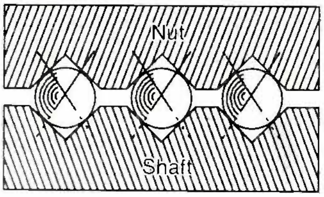


Fig. 9.15. Oversize ball preload.

(iv) *Integral preload :*

- Refer to Fig. 9.16 :

— The integral preload ballscrews, in outward appearance, appear the same as single nut oversize ball or non-preloaded ball screws; however there is a *dimensional allowance* in the nut internal rotating element which provides the proper amounts of preload. Compared with the double nut type, the *nut dimensions can be made shorter, an advantage of integral preload ball screws.*

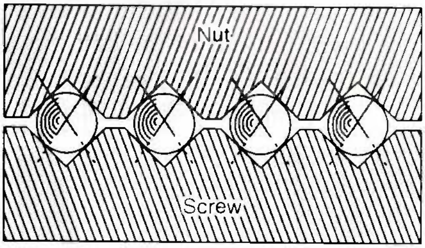


Fig. 9.16. Integral preload.

— These ballscrews are *best suited for light and moderate preload.*

(v) *Constant pressure preload :*

- Refer to Fig. 9.17 :

— In this method of preloading a coil spring or belleville spring is inserted between the two nuts to provide a constant pressure preload.

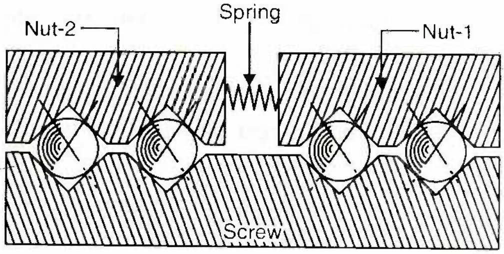


Fig. 9.17. Constant pressure preload.

— Overloading should be avoided otherwise the balls may get jammed resulting in stoppage of the motion.

— By this method transmission efficiency as high as 90% can be achieved.

Backlash can also be eliminated by using "elliptical screw and two-ways ball nuts".

Mounting of ballscrews :

The mounting arrangement of a ballscrew depends on its *required speed, length and size*. The position of the ballscrew should be near the line of the resultant force arising from cutting, frictional and inertial forces.

The commonly used methods of mounting of ballscrews are :

1. Ballscrew fixed at both ends.
2. One end fixed other end supported.
3. Both ends supported.
4. One end fixed other end free.

- In order to minimise the positioning inaccuracies in a ballscrew system, attention should be paid to the selection of *end bearings*. In a ballscrew the function of the bearings is to locate the screw radially and resist the axial thrust force. These bearings should have :

— high load capacity;

— high axial stiffness;

— low axial run-outs (of the order of 2 μm)

Following are the *commonly used ballscrew end bearings* :

(*i*) Sets of angular contact ball bearings.

(*ii*) Set of thrust and radial roller bearings.

(*iii*) Precision deep groove ball bearings.

Dismantling of ballscrew system :

- The ball nut should never be removed from the ballscrew by the user as the balls will fall out of the ball nut.
- The dismantling of the ball nut from the ballscrew should be carried out by the *following very special method* :

A tube whose outside diameter is equal to the root diameter of ballscrew is brought close to the end of ballscrew threads and the nut is driven onto this tube so that the ball in the nut is supported by the outside surface of the tube.

Classification of ballscrews :

The ballscrews, depending upon the accuracy, are *classified* as follows :

1. *"Commercial grades"* ... The threads are invariably *rolled*.
2. *"Precision grades"* ... The threads are *cut and ground* to obtained the required accuracy.

The ballscrew used on CNC *machines* are usually of the *"precision grade"*.

A ballscrew's accuracy can be specified as :

(*i*) Cumulative lead accuracy over a specified length.

(*ii*) Total cumulative lead accuracy.

(*iii*) Fluctuations of cumulative lead accuracy over one revolution.

In consonance with the above accuracies, the ballscrews are *classified* into the following seven grades :

C0, C1, C2, C3, C4, C5 and C7 .... having grade-wise applications for various machine tools.

II. *Roller screws* :

A *roller screw* has a *grooved roller elements* which make physical constant with the threads on the nut and screw. The rollers are not plain cylindrical but have the *circumferential grooves* which may be threads matching with the threads on the screw and nut or grooves of the thread form of the screw and nut.

The two types of roller screws generally used are :

1. Planetary roller screw.
2. Recirculating roller screw.

The rollers in both types of screw are positioned with between the nut and the screw and engage with the thread from inside the nut and on the outside of the screw.

- These roller screws provide *backlash-free movement* and their *efficiency is of the same order as of ballscrews.*
- An *advantage* of rollerscrews is that because the pitch of the screw is *smaller than the minimum pitch of the ballscrew, the less complex electronic circuitry will provide more accurate position control.*
- The roller screws are *much costlier than the ballscrews.*

1. *Planetary roller screw* :

Fig. 9.18, shows the elements of a planetary roller screw.

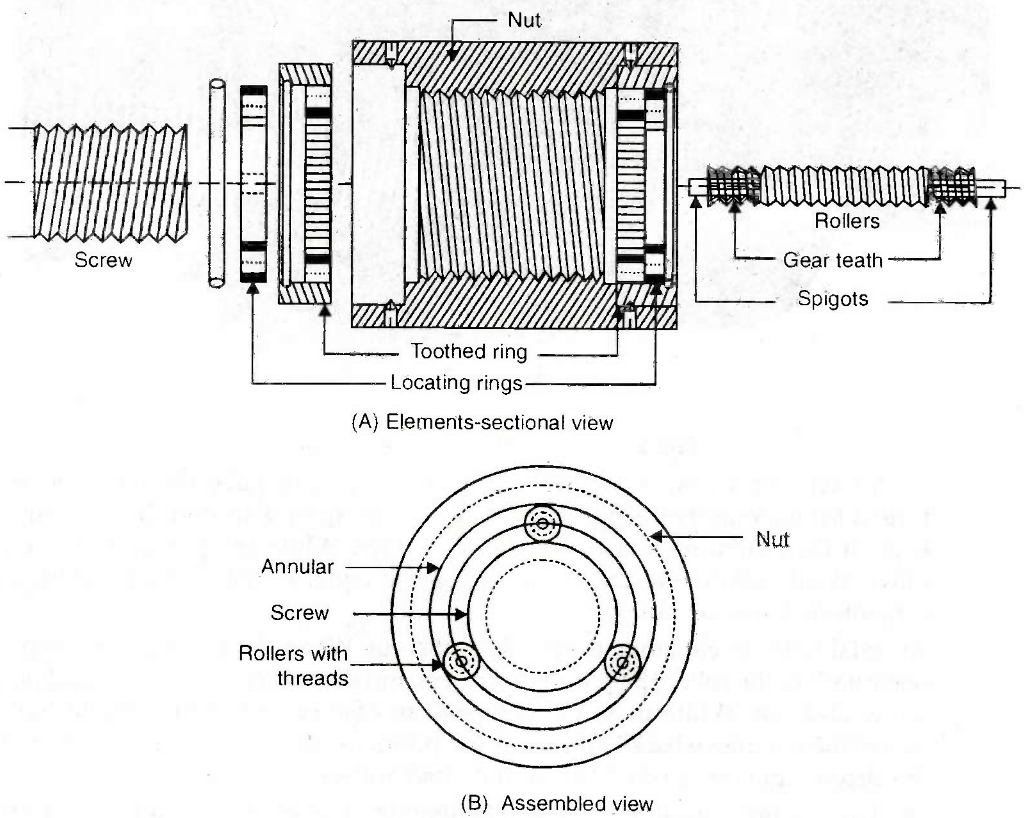


Fig. 9.18. Planetary roller screws.

— A planetary roller screw consists of *rollers* with grooves cut on them.

— At each end of the rollers, *gear teeth* are cut. The gear teeth mesh with an internally *toothed ring* on the *nut*, which drives the rollers to provide rolling motion between the nut and the *screw*.

— The rollers are equally spaced around the shaft and are retained in their circumferential positions by *spigots* which engage themselves in the *locating rings* at each end of the nut. There is no axial movement of the rollers relative to the nut.

- These screws are *capable of transmitting high loads at fast speeds.*

2. *Recirculating roller screw :*

Fig. 9.19 shows a recirculating roller screw :

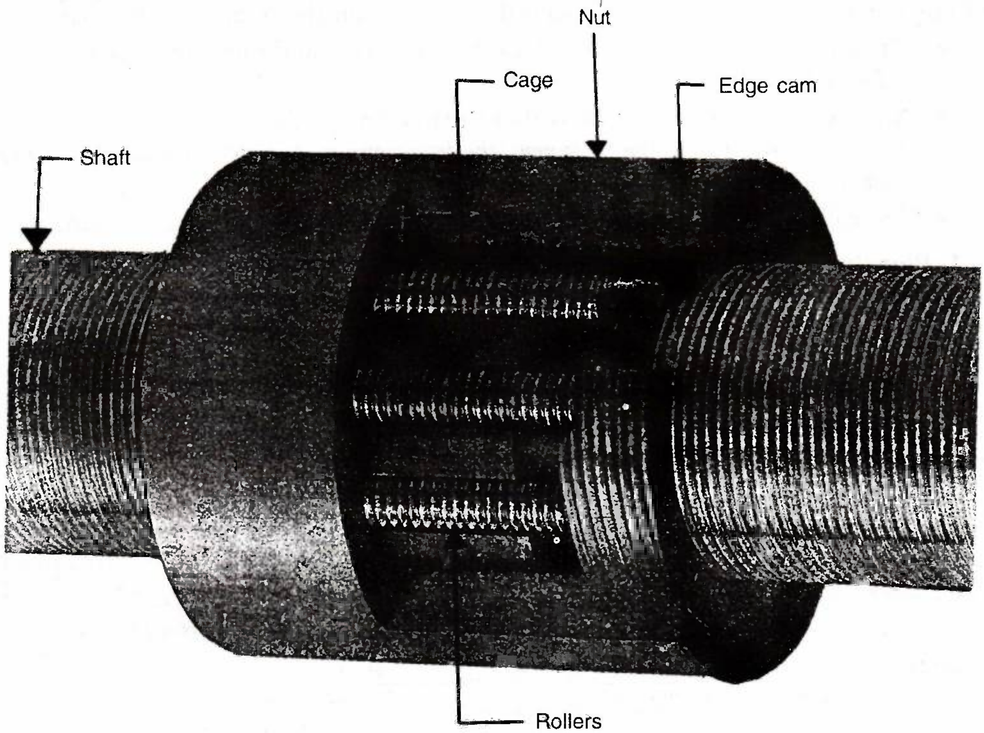


Fig. 9.19. Recirenlating roller screw.

— In this type of screw, the *rollers* are not threaded but have circular grooves of thread form along their length. They are equally spaced around the *shaft* and are kept in their circumferential position by a *cage*. While in operation, the rollers move axially relative to the nut at a distance equal to the screw pitch for each rotation of screw or nut.

— An axial recess is cut along the inside of the nut. When the drive screw completes one rotation, the rollers pass into this recess and disengage from the thread on the screw and nut. While they are in recess, an *edge cam* on a ring inside the nut causes them to move back to their starting positions. While one roller is disengaged, the driving power is provided by the other rollers.

- These screws are *slower in operation* than the planetary type, but are *capable of taking high loads with greater accuracy.*

Differences between ball and roller screws :

Following are the differences between ball and roller screws :

S.No.	Aspects	Ballscrews	Roller screws
1.	*Physical contact*	Balls in the intermediate element which make physical contact with the threads on the screw and nut	Rollers with circumferential grooves make physical contact with the threads on the screw and the nut.
2.	*Accurate position control*	Difficult	Possible
3.	*Backlash*	Backlash exists and is reduced by preloading using two nuts.	Backlash free
4.	*Type of contact*	Contact between ball and screw is *point contact.*	Contact between the roller and screw is *surface contact.*
5.	*Number of components*	Less	Relatively more
6.	*Cost*	Low	High

III. *Rack and pinion:*

In case of machines with *longer strokes,* the use of ballscrews for such machines is restricted due to the following reasons :

(*i*) A ballscrew, for longer strokes needs to the supported at intermediate points to minimise deflection due to its own weight over the length and a large diameter has to be used to reduce torsional deflection.

(*ii*) The drive cannot be run at higher speed due to the lower critical speed of the ballscrew.

The above problems can be tackled by using rack-and-pinion drive which is particularly suitable for transmission of motion over a *longer length.*

Advantages :

(*i*) A slide operated by a rack-and-pinion drive has the advantage that the stiffness of the drive is independent of the stroke length.

(*ii*) This system is cheaper as compared to the ballscrew system.

- Special pinions are available which *provide a minimum backlash.* These pinions are in two sections across the width :

— Teeth on one side of the pinion mesh with one side of the rack teeth;

— Teeth on the other side of the pinion mesh with the other side of the rack teeth.

IV. *Elements of torque transmission :*

The torque from a prime-mover shaft is transmitted to an output shaft (may be a pinion or a ballscrew) through the elements of torque transmission.

Several elements used on CNC machines for transmission of torque are :

(*i*) Gears; (*ii*) Timing belts; (*iii*) Flexible couplings etc.

1. ***Gear box*** : A gear box is employed for torque transmission in the following cases :

(*i*) To reduce torques on the prime-mover shaft.

(*ii*) To reduce load inertia on the prime-mover shaft.

(*iii*) To reduce the high motor speed to a speed suitable for the feed drive.

(*iv*) To provide reduction between the shafts which are not coaxial or parallel.

2. ***Timing belts*** : Fig. 9.20 shows the constructional details of a *timing belt.*

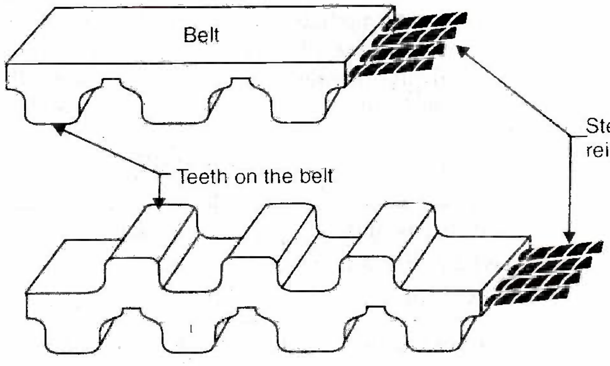


Fig. 9.20. Constructional details of timing belt.

— The timing belts are *endless toothed belts* similar to conventional belt drive.

— The teeth engage with a *timing pulley having teeth on its periphery.* The teeth profiles on the belt and the pulley are compatible with each other.

— Timing belt with tooth profile are commercially available with steel wire reinforcement (Fig. 9.20). Several manufacturers have their own teeth profile and suppliers catalogues give complete information regarding the selection of pitch of teeth, length and width of belt, initial tension, torque transmission, speed, power and application. *Initial tension increases the efficiency of torque transmission of the belt.*

Advantages : *Timing belts claim the following advantages :*

(*i*) Higher efficiency.	(*ii*) Less noise.
(*iii*) Low cost.	(*iv*) Elimination of lubrication.
(*v*) Less maintenance.	(*vi*) Slip free, being a positive drive.

Flexible couplings :

This method of transmission is recommended when the driving and driven shaft are coaxial and the distance between the shafts being small. However, usually it is very difficult to maintain the co-axiality of the shafts (Further, heat and elastic deformation cause additional misalignments between the two coaxial shafts). In such situations, flexible couplings are used which take care of slight misalignment in the axes of the shafts.

The ballscrew and servomotor are coupled directly using this type of coupling (Fig. 9.21).

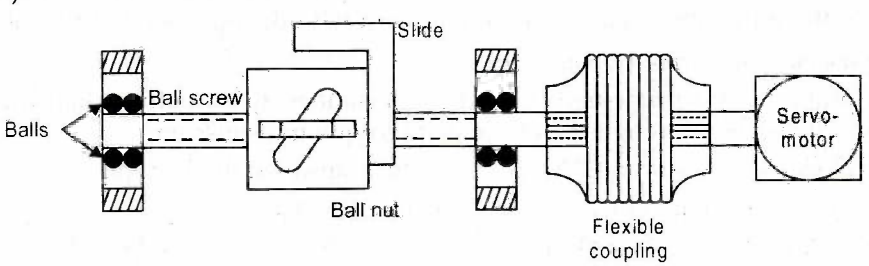


Fig. 9.21. Use of a flexible coupling for coupling ballscrew and servomotor.

- With the use of the flexible couplings the following three kinds of *errors can be compensated* :
  (*i*) Radial misalignment.
  (*ii*) Angular misalignment.
  (*iii*) Axial shift.

9.2.5. Spindle and Spindle Bearings

9.2.5.1. Spindle

The spindle carrying the workpiece or tool when subjected to high cutting speeds and high material removal rates, experience deflection and thrust forces. To ensure increased stability and minimise torsional strain, the machine spindle is *designed to be short and stiff and the final drive to the spindle is located as near to the front bearing as possible.*

The rotational accuracy of the spindle is dependent on the quality and design of bearings used. The *ball or roller bearings are suitable for high speeds and high loads because of low friction, lower wear rate and lesser liability to incorrect adjustment and ease of replacement when necessary.*

9.2.5.2. Spindle bearings

In modern machine tools, which employ high performance cutting tool materials, the designed characteristics of spindles used are :

(*i*) Minimum deflection under varying loads.
(*ii*) Long service life.
(*iii*) Stiffness.
(*iv*) Thermal stability.
(*v*) Good running accuracy both in radial and axial directions.
(*vi*) Axial load carrying capacity.
(*vii*) High speed of operation, without chatter, vibration.

On these characteristics do the accuracy and quality of the jobs produced depend. This can be achieved by using proper spindle bearing.

The various types of spindle bearings used in the design of a spindle for machine tools are :

1. Antifriction bearings.
2. Hydrostatic bearings. $\rbrace$ .....*Fluid bearings*
3. Hydrodynamic bearings.

1. *Antifriction bearings* :

The antifriction bearings are *suitable for high speeds and high loads.*

These are often preferred to hydrodynamic bearings because the following *reasons* :

- High reliability.
- Ease of replacement.
- Low friction.
- Moderate dimensions.
- Lesser liability to suffer from wear or incorrect adjustment.

On CNC machines, the following types of *ball and roller bearings* are used :

(*i*) *Ball bearings* :
  (*a*) Deep groove ball bearings (Fig. 9.22)
  (*b*) Angular contact ball bearings (Fig. 9.23)

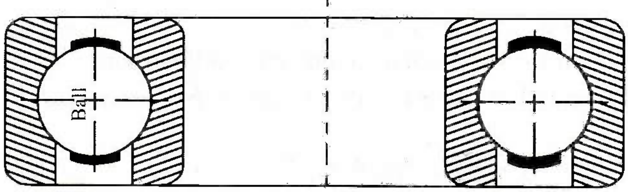


Fig. 9.22. Deep groove ball bearing

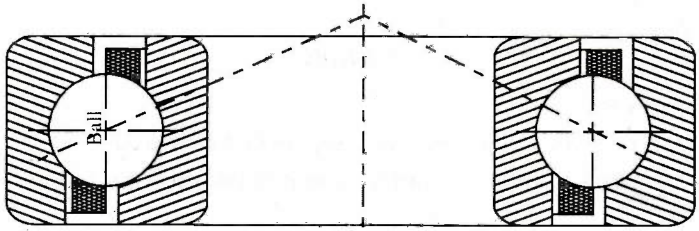


Fig. 9.23. Angular contact ball bearing

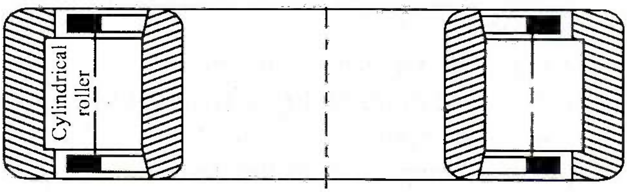


Fig. 9.24. Cylindrical roller bearing

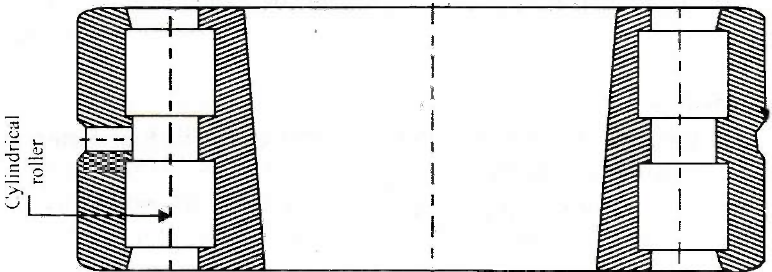


Fig. 9.25. Cylindrical roller bearing with tapered bore.

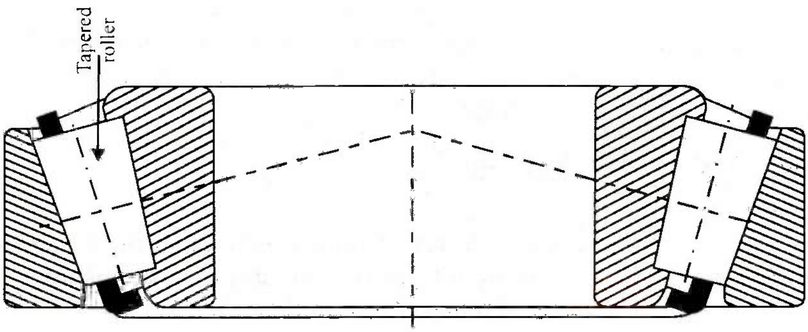


Fig. 9.26. Tapered roller bearing.

(*ii*) *Roller bearings* :

(*a*) Cylindrical roller bearings (Fig. 9.24).

(*b*) Cylindrical roller bearings (double row) with tapered bore (Fig. 9.25).

(*c*) Tapered roller bearings (Fig. 9.26).

- The ball and roller bearings are called antifriction bearings because the contact of support of rolling element is *point contact* in case of ball bearing and *line contact* in case of roller bearing. It is of paramount importance that these bearings are manufactured with *highest accuracy* otherwise any error in any of the elements will severely affect the quality of job produced.
- The selection of a particular type of bearing for the spindle depends on the following requirements of the particular machine :

(*i*) Spindle stiffness.

(*ii*) Spindle accuracy.

(*iii*) Speeds of operation.

Preloading of bearings :

There are some amounts of radial and axial clearances in the ball and roller bearings. When a main spindle is mounted on bearings there should be neither an axial nor radial play in the main spindle assembly. This is achieved by *preloading*.

— In case of tapered roller bearings and angular contact ball bearings, the axial and radial clearances can be taken up simultaneously by *preloading*.

— Cylindrical roller bearings (double row) with tapered bores are radially preloaded by pushing the inner race against the taper on the spindle.

Advantages :

(*i*) Preloading increases the radial and axial rigidity of the bearings.

(*ii*) Improves damping characteristics of bearings.

(*iii*) Prevents rolling elements from disengaging themselves from the raceways.

- *Preloading causes elastic deformation of the bearings. Excess of elastic deformation causes metal-to-metal contact producing noise.*

2. *Hydrostatic bearings* :

Fig. 9.27 shows the principle of hydrostatic bearings :

— Here the spindle is supported by a relatively thick film of oil (called *hydrostatic pockets*) supplied under pressure; the oil in the pockets being stationary. The oil is supplied to the bearing through a throttling system to control pressure and volume. Lubricating seals are used to prevent the leakage of oil. There is no mechanical contact.

— The load carrying capacity of this type of bearing is independent of the speed of rotation.

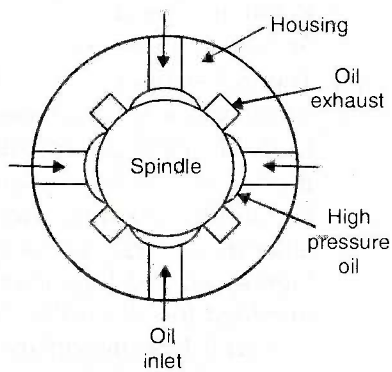


Fig. 9.27. Principle of hydrostatic bearings.

They have the following *merits* :

(*i*) High wear resistance.

(*ii*) High damping properties.

(*iii*) High running accuracy.

- These bearings are used in grinding and boring machines etc. (where temperature effects cause problems in the part accuracy).

3. *Hydrodynamic bearings* :

Fig. 9.28 shows the principle of hydrodynamic bearings :

— The pressure of oil within the bearing is created by the rotation of the spindle. As the spindle rotates, the oil in contact with the spindle is carried into wedge-shape cavities between the spindle and the bearing due to centrifugal action. As the oil is forced through the small clearances between the bearing and spindle, the oil pressure is increased.

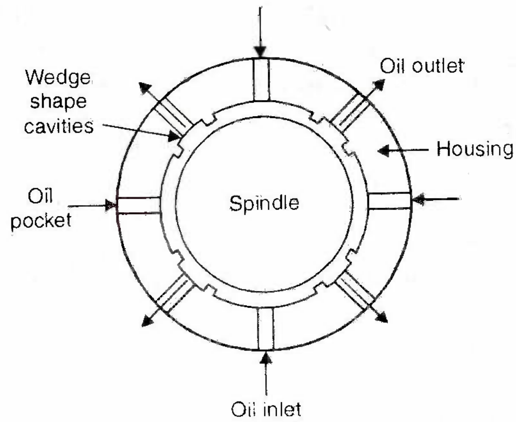


Fig. 9.28. Principle of hydrodynamic bearings.

— In this type of bearing there is a constant flow of oil round the spindle, maintaining a thick oil film.

The *essential features* of these bearings are :

(*i*) Good running accuracy.

(*ii*) Simplicity.

(*iii*) Good damping properties.

The *main limitation* of this type of bearing is that a definite clearance must be provided for the oil film to be maintained between the bearing and the spindle; the clearances normally provided vary from 50 μm to 200 μm depending upon the journal diameter.

- These bearings are used where the load carrying capacities are low and frequent starting and stopping of the spindle is not required as in the case of *grinding machines.*

***Selection of spindle bearing* :**

The selection of spindle bearing depends on the following factors :

(*i*) Type of load – axial, radial, or combination.

(*ii*) Load intensity.

(*iii*) Rotational speed.

(*iv*) Spindle stiffness.

(*v*) Thermal stability.

The *accuracy of a spindle* depends on : (*i*) Radial runout; (*ii*) Axial runout.

— In *radial runout* the spindle shifts radially in any of 360° directions.

— In *axial run out* the spindle moves in the axial direction.

For an ideal condition both the radial runout and axial runout should be zero.

— Since the accuracy of the spindle also depends on thermal stability especially for high speed and high load carrying spindles, a proper provision should also be provided for lubricating the spindle bearings.

- In ***recent development*** the metal balls and rollers are replaced by ceramic balls and rollers because the latter offer the following *advantages* :

(*i*) Low coefficient of friction.

(*ii*) Greater thermal stability.

(*iii*) High hardness.

(*iv*) High wear resistance.

The ceramic bearings can be employed for spindle speed in range of 10,000 to 20,000 r.p.m.

9.2.6. Measuring Systems

Measuring systems are used on all the CNC machines to perform the following *functions* :

1. To monitor the position of a slide on a slideway.
2. To orient the spindle/table.
3. To measure the spindle speed.

- The following two terms are associated with measuring systems :
  (*i*) *Accuracy* : It is the smallest unit of movement that the system can consistently and repeatedly discriminate.
  (*ii*) *Resolution* : It is the smallest unit that can be measured by the measuring system.
- In a measuring system the measuring devices used are *classified* as :
  (*i*) *Rotary measuring devices* :
  Example : Incremental rotary encoder – widely used on CNC machine.
  (*ii*) *Linear measuring devices* :
  Example : A linear scale — used very oftenly.

Methods adopted for measurements :

As the ideal monitoring system for slide position is not yet available (due to reversal problems such as interference caused by the presence of swarf and cutting fluid), the following methods are generally used :

1. Direct measurement system.
2. Indirect measurement system.
3. Incremental rotary encoders.
4. Linear scale.

1. *Direct measurement system* : In this system the linear displacement is measured directly at the slide. The measuring device is fixed onto the moving machine element which detects the actual distance travelled by the machine slide.

Examples : (*i*) Linear scales; (*ii*) Inductosyn.

- By this measurement system a high degree of accuracy is obtained *because backlash errors in axis drive elements do not affect the measuring process.*

2. *Indirect measurement system* : In this system the slide position is determined by the rotation of the ballscrew/pinion or the drive motor.

- In comparison to direct measurement system, this system is *more convenient and less costly*. But, in such systems, there is a possibility of additional sources of error like backlash and torsional deformation on the drive system creeping in. However, these errors can be reduced by using various types of error compensating devices available in CNC machines. In these systems some of the feedback devices used are : *Encoders; resolvers.*

3. *Incremental rotary encoders* : The incremental measurement means *measurement by* *counting i.e.*, the output signals of increment rotary encoders are fed to an *electronic counter* from which the measured value is obtained by *counting the individual increments.*

— The encoder is connected mechanically to the ballscrew or any rotating shaft through a flexible coupling. It must be ensured that the axis of an encoder and that of a connecting shaft are aligned both angularly and radially within the permissible limits imposed by the flexible couplings; working beyond these limits may lead to undue mechanical load on the encoder's bearings.

4. *Linear scale* :

- The linear scale consists of a *glass scale with gratings and a reading head;* one of these two elements is mounted on a fixed member and another on the moving slide.
  - — The glass scale has gratings;
  - — The reading head contains the light source, a condenser, lens for collimating the light beam, the scanning reticle with index gratings, and cells,
- As and when the scale is moved relative to the scanning unit, the lines and spaces of the scale coincides with those of the index grating alternately. The corresponding light fluctuations are sensed by the cells which generate the signals; these signals are further processed for measurements (as is done in case of rotary encoders).

9.2.7. Controls

- For CNC machines, CNC controls are of significant importance. Earlier, CNC controls were developed for simple applications in turning, machining centres and grinding, but these days CNC systems have been developed to meet with the increased machine tools requirements of higher spindle speeds, higher rapid traverses and more number of axes. The new generation computer numerical controls allow *simultaneous control of more axes, interpolate positions faster, and use more data points for precise control.*
- The new controllers offer the following :
  - — Advanced graphic interfaces;
  - — Program simulation;
  - — Some cutter selecting capabilities.

9.2.8. Gauging

The use of hi-tech CNC machines leads to *better workpiece quality.* The quality can be maintained by eliminating the effect of parameter like tool wear and thermal growth, with the use of automatic gauging system.

The gauging on a machine tool may be used for the following *purposes* :

(*i*) To inspect workpiece.

(*ii*) To detect tool breakage.

(*iii*) To define tool offsets.

(*iv*) To automatically align the workpiece.

(*v*) To detect the stock variation.

For measurement and inspection, a *touch trigger probe* is normally employed. These probes are very sensitive switches with a long spring loaded stylus. Once the contact is made these probes are capable of detecting small deflections of the stylus from the home position. The signal is transmitted to the machine system which processes the data according to the gauging requirements. These probes are mounted on the turret, table or headstock housing or can be transferred from the tool magazine to the spindle for use, depending on the machine and function.

9.2.9. Tool Monitoring System

The tools wear out or even break during machining. If tool wear and breakage is not properly monitored, the productivity of the machine and the quality of the component produced are affected. Now-a-days established monitoring sensors and systems are available commercially which can be integrated with CNC machines.

Following are the two ways of monitoring tool wear and breakage :

1. *Direct monitoring* : In this type of monitoring a touch probe is directly used to monitor the tool condition by checking the tool edge position and checking for the existence of a tool edge.
2. *Indirect monitoring* : Here, the tool condition is checked indirectly by monitoring the change in certain perameter whose value when affected reflects the tool condition. Following parameters are used to monitor tool condition :
   (*i*) Cutting forces. (*ii*) Tool life.
   (*iii*) Workpiece dimensions. (*iv*) Emission of noise during cutting.
   (*v*) Power of the spindle or a feed drive or a driven tool.

9.2.10. Swarf Removal

In CNC machines the cutting time is much more and as such the volume of swarf generated is also more.

— Unless the swarf is quickly and efficiently removed from the cutting zone, it can affect the cutting process and quality of the finished product.

— Also the swarf cannot be allowed to accumulate at the machine tool because it may hamper the access to the machine tool.

— In addition some auxiliary functions like automatic component loading or automatic tool change may also be affected by accumulation of swarf.

To overcome all above problems it is necessary to provide an efficient swarf control system with the CNC machine tools with some mechanism to remove the swarf from the cutter and cutting zone and for the disposal of swarf from the machine tool area itself.

— The swarf removal from the cutting zone is generally taken case of by the design configuration of the machine. *Continuously operating linear or rotatory converters are used for removing the swarf from the machine tool.*

9.2.11. Safety

As the CNC machines are under continuous automatic operation, there is a need to protect the machine guideways and to ensure operators safety since the machines run at high speeds with automatic auxiliary operations.

— In order to have efficient working and long life of the machine it is essential to protect machine guideways, drive screws and transducers etc. These elements are protected by the use of various types of collapsible guards and covers. All the sliding elements are fitted with wipers and drive screws are normally protected by using telescopic covers. Jets of cutting fluids are used to wash away swarf and clear the tool work area.

— Operator's safety is very important aspect which cannot be overlooked. To ensure safe working conditions the CNC machine tools are provided with metallic or plastic guards. Where it is not possible to provide effective guards, proximity protection is provided by pressure mats or light barriers.

HIGHLIGHTS

1. *CNC* may be defined as an NC system with a microcomputer or microprocessor using software to implement control algorithms. With *CNC control systems* it is possible to obtaine information on machine utilisation which is useful to the managements.
2. *Elements of CNC machines* : The following are some of the important constituent parts, and aspects of CNC machines to be considered in their designing :

   (*i*) Machine structure; (*ii*) Guideways (slideways); (*iii*) Drives; (*iv*) Spindle and spindle bearings; (*v*) Measuring systems; (*vi*) Controls; (*vii*) Gauging; (*viii*) Tool monitoring; (*ix*) Swarf removal; (*x*) Safety.

OBJECTIVE TYPE QUESTIONS

Fill in the Blanks or Say "Yes" or "No".

1. A machine tool having a dedicated computer to help prepare the program and control some or all of the operations of the machine tool is called .... machine tool.
2. CNC control unit allows compensation for any changes in the dimensions of the cutting tool.
3. .... is defined as a design process using sophisticated computer graphics techniques, backed up with computer software packages to aid in the analytical, development, costing and ergonomic problems associated with design work.
4. .... concerns any automatic manufacturing process which is controlled by computers.
5. The machine structure is the load carrying and supporting member of the machine tool.
6. .... guideways find wide application in conventional machine tools due to their low manufacturing cost and good damping properties.
7. Friction guideways operate under conditions of sliding friction and do not have a constant coefficient of friction.
8. .... guideways wear away rapidly due to lack of bearing surface.
9. In flat guideways the chip accumulation and lubrication problems are not serious.
10. Flat guideways wear uniformly.
11. Dovetail guideways have large load carrying capacity and tend to check the overturning tendency under eccentric loading.
12. Cylindrical guideways are not suitable for relatively short traverses and light loads.
13. Antifriction linear motion (LM) guideways are used on CNC machine tools to reduce amount of wear, friction, heat generation and improve smoothness of the movement.
14. The recirculating linear roller bearings are used for movement along a .... plane.
15. In .... guideways the slide is raised on a cushion of compressed air which entirely separates the slide and guideway surfaces.
16. .... are the devices which impart motion to mechanical elements.
17. Constant torque and positioning are the special characteristics of a feed motor.
18. The motors are suitable only for light duty machines due to low power output.
19. The ballscrews used on CNC machines are usually of .... grade.

20. .... is the process of applying initial load to the nut which will cause elastic deformation of the screw threads in the axial direction, thereby increasing the axial rigidity of the ballscrew nut.
21. Oversize preload method is best suited to provide comparatively light preload, to the extent of eliminating axial clearance.
22. A .... screw has a grooved roller elements which make physical contact with the threads on the nut and screw.
23. The timing belts are toothed belts.
24. The incremental measurement means measurements by counting.
25. For measurement and inspection a touch trigger probe is normally employed.

ANSWERS

1. CNC	2. Yes	3. CAD	4. CAM	5. Yes
6. Friction	7. Yes	8. Vee	9. No	10. No
11. Yes	12. No	13. Yes	14. Flat	15. Aerostatic
16. Drives	17. Yes	18. stepper	19. precision	20. Preloading
21. Yes	22. roller	23. endless	24. Yes	25. Yes

THEORETICAL QUESTIONS

1. What do you mean by "Numerical control" ?
2. What are the areas where "Numerical control" can be used ?
3. Describe briefly the working of NC machine tool.
4. Explain with a neat diagram the main elements of a NC machine tool.
5. How are NC machines classified ?
6. Enumerate various application of NC machines.
7. List the advantages of NC machines.
8. Define CNC.
9. What are the functions of CNC ?
10. State the advantages of CNC machines over NC machines.
11. What are the disadvantages of CNC machines ?
12. What are the applications of CNC machines ?
13. Explain briefly the functioning of CAD/CAM system.
14. What is a CNC machine ?
15. Enumerate important constituent parts, and aspects of CNC machines which should be considered in their designing.
16. What is "Machine structure" ? Explain briefly.
17. Explain briefly static load, dynamic load and thermal load in relations to a machine structure.
18. What are the functions of guideways ?
19. What are the factors which influence the design of guideways ?
20. How are guideways classified ?
21. Discuss briefly the friction guideways.
22. What is 'Stick-slip phenomenon'? Explain.

23. Explain briefly the following :
    (i) Vee guideways; (ii) Flat guideways;
    (iii) Dovetail guideways; (iv) Cylindrical guideways
24. What are antifriction linear motion (LM) guideways ? What are their advantages and disadvantages ?
25. Explain briefly various types of antifriction guideways.
26. Discuss briefly with neat sketches the following :
    (i) Linear bearing will balls.
    (ii) Linear bearing with rollers.
27. What are frictionless guideways ? Explain.
28. Explain briefly the following :
    (i) Hydrostatic guideways.
    (ii) Aerostatic guideways
29. List the advantages and disadvantages of frictionless guideways.
30. What is a 'Drive' ?
31. How are machine tool drives classified ?
32. Explain briefly "Spindle drives".
33. What are the main components of a feed drive ? Explain.
34. What are the requirements of CNC feed drive ?
35. A.C. servomotors have become more popular than D.C. servomotors for machine tool applications. Explain why ?
36. What are the advantages of A.C. servomotors over D.C. servomotors ?
37. Explain briefly "Recirculating ballscrew and nut".
38. What do you mean by preloading of nuts ? Explain.
39. How is preloading of ballscrews-nut classified ?
40. Explain briefly the following :
    (i) Tension preload; (ii) Compression preload;
    (iii) Oversize preload; (iv) Integral preload.
41. What are the commonly used methods of mounting of ballscrews ?
42. How are ballscrews classified ?
43. How can ballscrews accuracy be specified ?
44. What is a roller screw ? How does it differ from ballscrew ?
45. Explain briefly the following :
    (i) Planetary roller screw.
    (ii) Recirculating roller screw.
46. Give the differences between ballscrews and roller screws.
47. What is "rack-and-pinion arrangement" of power transmission ? Explain.
48. List the elements which are used on CNC machines for torque transmission.
49. Explain briefly any two of the following elements of torque transmission :
    (i) Gear box.
    (ii) Timing belts.
    (iii) Flexible couplings.
50. What are the advantages of timing belts ?
51. What are the desired characteristics of spindles used in modern machine tools ?
52. Explain briefly antifriction spindle bearings used in CNC machines.
53. What do you mean by "Preloading of bearings"? Explain.
54. What are the advantages of preloading of bearings ?

55. Explain briefly principle of hydrostatic bearings.
56. What is an hydrodynamic bearing ? Explain.
57. What are the essential features of hydrodynamic bearings ?
58. What are the factors which influence the selection of spindle bearing ?
59. Explain briefly the measurement systems used on CNC machines.
60. Explain briefly the following :
    (*i*) Incremental rotary encoders.
    (*ii*) Linear scale.
61. Write a short note on 'controls' used on CNC machines.
62. Explain briefly the following as applied to CNC machines :
    (*i*) Gauging;
    (*ii*) Tool monitoring system;
    (*iii*) Swarf removal;
    (*iv*) Safety.

APPENDIX – A

Basic Mechanical Concepts

A.1 Engineering materials – Classification of materials – Classification of electrical engineering materials – Biomaterials – Advanced materials – Materials of future – "Smart materials" – Nanotechnology – Mechanical propoerties of metals – Selection of materials; **A.2 Force, moments and friction** – Force – Moments – Friction; **A.3 Stresses and Strains** – Classification of loads – Stress – Simple stress – Strain – Importance of mechanical tests; **A.4 Bending of beams; A.5 Shafts** – Torsion of shafts – Torsion equation – Power transmitted by the shaft; **A.6 Bending moments and shearing forces** – Some basic definitions – Classification of beams – Shearing force (S.F.) and bending moment (B.M.) – General relation between the load, the shearing force and the bending moment; **A.7 Metrology** – Standards of measurement – Limits, fits and tolerance – Classification of measuring equipment – Surface finish; **A.8 Machining processes** – Machining – Classification of machining processes-cutting tools – Orthogonal and oblique cutting – Types of chips – Forces of a single-point tool – Machine tools; **A.9 Heat treatment** – Definition – Objects – Constituents of iron and steel – Heat treatment processes – Highlights – Objective Type Questions – Theoretical Questions.

A.1 ENGINEERING MATERIALS

A.1.1. Classification of Materials

The engineering materials may be *classified* as follows:

1. **Metals** (*e.g.*, iron, aluminium, copper, zinc, lead etc.)
2. **Non-metals** (*e.g.*, leather, rubber, plastics, asbestos, carbon etc.)

Metals may be further subdivided as :

(*i*) *Ferrous metals* (*e.g.*, cast iron, wrought iron and steel) and *alloys* (*e.g.*, silicon steel, high speed steel, spring steel etc.)

(*ii*) *Non ferrous metals* (*e.g.*, copper, aluminium, zinc, lead etc.) and *alloys* (brass, bronze, duralumin etc.)

Engineering materials may also be **classified** as follows :

1. Metals and alloys
2. Ceramics
3. Organic polymers.

1. Metals and alloys :

Metals are polycrystalline bodies consisting of a great number of fine crystals (10^{-1} to 10^{-4} cm size) differently oriented with respect to one another. Depending upon the mode of crystallization, these crystals may be of various irregular shapes, and, in contrast to crystals of regular shape, are called *crystallites or grains* of the *metal*. Metals in the solid state and, to some extent, in the liquid state possess *high thermal and electrical conductivity, and a positive temperature coefficient of electrical resistivity.* The general resistance of pure metals *increases with the temperature.* Many metals display superconductivity; at temperatures near absolute zero, their electrical resistance drops abruptly to extremely low values. Besides, all metals are capable of *thermionic emission, i.e.,* the emission of electrons upon being heated; they are good reflectors of light and lend themselves well to plastic deformation.

Pure metals *are of low strength and in many cases, do not possess the required physiochemical and technological properties for some definite purpose.* Consequently they are *soldom* used in engineering. The overwhelming majority of metals used are *alloys.*

Alloys are produced by *melting or sintering two or more metals,* or metals and a non-metal, together. Alloys possess typical properties inherent in the metallic state, the substances that make up the alloy are called its *components.* An alloy can consist of two or more components.

Examples of metals and alloys: *Steels, copper, aluminium, brasses, bronze, invar, superalloys etc.*

2. Ceramic materials :

These materials are non-metallic solids made of inorganic compounds such as oxides, nitrides, borides, silicides and carbides. They are fabricated by first shaping the powder with or without the application of pressure into a compact which is subsequently subjected to a hight temperature treatment, called *sintering. Traditional ceramics* were made from crude naturally occurring mixtures of materials having inconsistent purity. These have been used essentially in the manufacture of pottery, porcelain, cement and silicate glasses. *New ceramics* possess exceptional electrical, magnetic, chemical, structural and thermal properties. Such ceramics are now extensively used in the *electronic control devices, computers, nuclear engineering and aerospace fields.*

Examples of ceramics : *MgO, CdS, ZnO, SiC,* BaT_iO_2, *silica, sodalime, glass, concrete, cement, ferrites, garnets, etc.*

3. Organic materials :

These materials are derived *directly from carbon.* They usually consist of carbon chemically combined with hydrogen, oxygen or other non-metallic substances. In many instances their structures are fairly complex.

Common organic materials are : *Plastics and Synthetic rubbers.* These are termed *"polymers"* because they are formed by polymerization reaction in which relatively simple molecules are chemically combined into massive long-chain molecules or "three dimensional" structures.

Examples of organic materials : *Plastics: PVC, PTFE, polythene; Fibers, terylene, nylon, cotton; Natural and synthetic rubbers, leather, etc.*

Examples of Composites :

1. **Metals and alloys and ceramics**
   (*i*) Steel reinforced concrete.
   (*ii*) Dispersion hardened alloys.
2. **Metals and alloys and organic polymers**
   (*i*) Vinyl-coated steel.
   (*ii*) Whisker-reinforced plastics.

3. **Ceramics and organic polymers**
   (*i*) Fibre-reinforced plastics.
   (*ii*) Carbon-reinforced rubber.

A.1.2. Classification of Electrical Engineering Materials

The electrical engineering materials may be classified into the following *four* types :

1. Conductors.
2. Semiconductors.
3. Insulators (or dielectrics).
4. Magnetic materials.

1. Conductors :

- ***Conductors*** may be defined, *as the materials which have free valence electrons in plenty for electric conduction*. The commonly used conductors are *copper, aluminium, tungsten, iron and steel, lead, nickle, tin etc.* In this case the valance and conduction bands *overlap*. Since there is no physical distinction between the two bands, therefore, a large number of free electrons (conduction) are available.
- The conductors are used in *electric devices, instruments and all kinds of electrical machine windings*. They are also employed in manufacturing of *cables and wires*, for the *distribution of electrical energy* over long distances and *telephone and telegraph circuits*.

2. Semiconductors :

Semiconductors *are solid materials, either non-metallic elements or compounds which allow electrons to pass through them so that they conduct electricity in much the same way as the metals.* They occupy an intermediate position between conductors and insulators. In this case, the *valance band is almost filled but conduction band is almost empty;* they are *separated by a small energy gap*. The valence band is completely filled at 0°K and no electron is available for conduction. But as the temperature is increased the width of energy gap decreases and some of the electrons are liberated into the conduction band. In other words, the conductivity of semiconductors increases with temperature. Semiconductors usually have *high resistivity, negative temperature coefficient of resistance* and are *generally hard and brittle.*

The main difference between a conductor and semiconductor relates to the dependence of their conductivity on the degree of purity of metals. The conductivity of a good conductor *increases* with purification whereas that of semiconductor generally *decreases* with purification.

Examples of elements which are semiconductors are : Boron (B), Carbon (C), Silicon (Si), Germanium (Ge), Phosphorus (P), Arsenic (As), Antinomy (Sb), Sulphur (S), Selenium (Se), Iodine (I). A number of semiconducting compounds in the form of oxides, alloys, sulphides, halides and solenoids are also available.

Semiconductors are ***used*** in different fields of electrical engineering, *e.g.*, *telecommunication and radio communication, electronics and power engineering. They also render their services as amplifiers, rectifiers, photocells, special sources of electric current etc.*

3. Insulators :

Insulators are *those materials in which valence electrons are very tightly bound to their parent atoms thus requiring very large electric field to remove them from attraction of nuclei.* They are not governed by electrodynamic phenomena involving the direction flow of number of electric charges by the electrostatic phenomena associated with the presence of an electric field. They have (*i*) a full valence band, (*ii*) an empty conduction, and (*iii*) a large energy gap between them; for conduction to take place, electrons must be given sufficient energy to jump from valence band to conduction band. At ordinry temperature the probability of electrons from full valence band gaining sufficient energy so as to surmount energy gap and becoming available for conduction in conduction band is *slight. But increase in temperature enables electrons to go to conduction band.*

In *electric circuits and devices the insulators insulate one current-carrying part from another.*

The insulating materials may be of **Three** types :

1. *Solid* : Mica, micanite, porcelain, asbestos, slate, marble, bakelite, rubber, PVC, polythene, paper, glass, cotton, silk, wood, valcanised fibre, ceramic, aluminium oxide.
2. *Liquid* : Natural resin varnishes, bituminous varnishes, phenolic varnishes, shellac varnishes, etc.
3. *Gaseous* : Air, nitrogen freon.

4. Magnetic materials :

- **Magnetic materials** are *those materials in which a state of magnetisation can be induced.* In accordance with the value of relative permeability the materials may be classified in the following three ways :
  (*i*) **Ferromagnetic materials.** The relative permeability of these materials is *much greater than unity* and is dependent on the field strength. The principal ferromagnetic elements are: *Iron, cobalt and nickel.* Gadolinium, however, also comes under this classification. They have *high susceptibility.*
  (*ii*) **Paramagnetic materials.** They have relative permeability *slightly greater than unity* and are magnetised slightly. *Aluminium, platinum and oxygen* belong to this category.
  (*iii*) **Diamagnetic materials.** The relative permeability of these materials is *slightly less than unity.* The examples are *bismuth, silver, copper and hydrogen.*
- *The magnetic properties of materials arise from the spin of electrons and orbital motion of electrons around the atomic nuclei.* In several atoms the opposite spin neutralises one another, but when there is an excess of electrons spinning in one direction, a magnetic field is produced. All substances, except ferromagnetic materials which can from permanent magnets, exhibit magnetic effects only when subjected to an external electromagnetic field.
- Since magnetic materials strengthen the magnetic field in which they are placed and possess high magnetic permeability, they claim wide field of applications *in the form of magnetic waves, magnetic screens and permanent magnets.*

A.1.3. Biomaterials

- *Biomaterials* are *employed in components implanted into the human body for replacement of diseased or damaged body parts.*
- These materials must not produce toxic substances and must be compatible with body tissues (*i.e.*, must not cause adverse biological reactions).
- All of the above materials–metals, ceramics, polymers, composites and semi-conductors may be used as biomaterials.

A.1.4. Advanced Materials

- *Materials that are utilised in high-technology (or high-tech) applications are sometimes called* ***Advanced materials***. By high technology we mean a device or product that operates or functions using relatively intricate and sophisticated principles: *Examples* include *electronic equipment* (VCRs, CD players etc.) *computers, fiberoptic systems, spacecraft, aircraft, and military rocketry.*

- These advanced materials are typically *either traditional materials whose properties have been enhanced or newly developed, high-performance materials.* Furthermore, they may be of all materials types (*e.g.*, metals, ceramics, polymers) and are normally *relatively expensive.*

A.1.5. Materials of Future—"Smart Materials"

- *Smart (or intelligent) materials* are a group of new and state-of-the-art materials now being developed that will have a significant influence on many of our technologies. The adjective **"smart"** *implies that these materials are able to sense changes in their environments and then respond to these changes in predetermined manners–traits that are also found in living organisms.* In addition, this *"smart"* concept is *being extended to rather sophisticated systems that consist of both smart and traditional materials.*
- Components of a smart material (or system) include some types of *sensor* (that detects an input signal), and an *actuator* (that performs a responsive and adaptive function). Actuators may be called upon to change shape, position, frequency, or mechanical characteristics in response to changes in temperature, electric fields, and/or magnetic fields.

Following *four types of materials* are commonly used for **actuators** :

(*i*) *Shape memory alloys.* These are metals that, after having been deformed, revert back to their original shapes when temperature is changed.

(*ii*) *Piezoelectric ceramics.* These expand and contract in response to an applied electric field (or voltage); conversely they also generate an electric field when their dimensions are altered.

(*iii*) *Magnetostrictive materials.* The behaviour of these materials is analogous to that of piezoelectrics, except that they are responsive to magnetic fields.

(*iv*) *Electrorheological / magnetorheological.* These are liquids that experience dramatic changes in viscosity upon the application of electric and magnetic fields, respectively.

Materials / devices employed as **sensors** include the following :

(*i*) Optical fibers.

(*ii*) Piezoelectric materials (including some polymers).

(*iii*) Microelectromechanical devices.

Example: One type of smart system is used in helicopters to reduce aerodynamic cockpit noise that is created by the rotating rotor blades. *Piezoelectric sensors* inserted into the blades, monitor blade stresses and deformations; feedback signals from these sensors are fed into a computer-controlled adaptive device, which generates noise-cancelling antinoise.

A.1.6. Nanotechnology

- The general procedure utilised by scientists to understand the chemistry and physics of materials, until recent times, has been to begin by studying large and complex structures, and then to investigate the fundamental building blocks of these structures that are smaller and simpler. This approach is sometimes termed ***"top-down"*** *science.*

- However, with the advent to scanning probe microscopes, which permit observation of individual atoms and molecules, it has become possible *to manipulate and move atoms and molecules to form new structures, and, thus design new materials that are built from atomic level constituents, (i.e., materials by design).* This ability to carefully arrange atoms provides opportunities to develop mechanical, electrical, magnetic and other properties that are not otherwise possible. This is termed as *"**botton-up**" approach and the study of the properties of these materials is termed* "**nanotechnology**"; the "nano" prefix denotes that the dimensions of these structural entities are on the order of a nanometer (10^{-9} m) as a rule, less than 100 nanometres (equivalent to approximately 500 atom diameters).
  — One *example* of a material of this type is the ***carbon nanotube***.

A.1.7. Mechanical Properties of Metals

1. **Strength:** *The strength of metal is its ability to withstand various forces to which it is subjected during a test or in service.* It is usually defined as tensile strength, compressive strength, proof stress, shear strength, etc. *Strength of materials is a general expression for the measure of capacity of resistance* possessed by solid masses or pieces of various kinds to any cause tending to produce in them a permanent and disabling change of form or positive fracture. Materials of all kinds owe their strength to the action of the forces residing in and about the molecules of the bodies (the molecular forces) but mainly to that ones of these known as *cohesion;* certain modified results of cohesion as toughness or tenacity, hardness, stiffness and elasticity are also important elements, and strength is in relation of the toughness and stiffness combined.
2. **Elasticity:** A material is said to be *perfectly elastic* if the whole of the stress produced by a load *disappears completely* on the removal of the load, the modulus of elasticity of Young's modulus (E) is the proportionally constant between stress and strain for elastic materials. *Young's modulus is the indicative of the property called **stiffness**: small values of E indicate flexible materials and large value of E reflect stiffness and rigidity.* The property of spring back is a function of modulus of elasticity and refers to the extent to which metal springs back when an elastic deforming load is removed. In metal cutting, modulus of elasticity of the cutting tools and tool holder affects their rigidity.
3. **Plasticity :**
   - *"Plasticity" is the property that enables the formation of permanent deformation is a material.* It is reverse of elasticity; a plastic material will retain exactly the shape it takes under load, even after the load is removed. Gold and lead are the highly plastic materials. *Plasticity is used in stamping images on coins and ornamental work.*
   - During plastic deformation there is the displacement of atoms within metallic grains and consequently the shapes of the metallic components change. It is because of this property that certain synthetic materials are given the name "plastics". These materials can be changed into required shape easily.
4. **Ductility:** *It is the ability of a metal to withstand elongation or bending.* Due to this property, wires are made by drawing out through a hole. The material shows a considerable amount of plasticity during the ductile extension. This is a valuable

property in chains, ropes etc., because they do not snap off, while in service, without giving sufficient warming by elongation.

5. **Malleability:** *This is the property by virtue of which a material may be hammered or rolled into thin sheets without rupture.* This property generally *increases with the increase of temperature.*

6. **Toughness (or Tenacity):** *Toughness (or tenacity) is the strength with which the material opposes rupture.* It is due to the attraction which the molecules have for each other; giving them power to resist tearing apart.

   The area under the stress-strain curve indicates the toughness (*i.e.*, energy which can be absorbed by the material upto the point of rupture). Although the engineering stress-strain curve is often used for this computation, a more realistic result is obtained from a *true-stress curve.* Toughness is expressed as energy absorbed (Nm) per unit volume of material participating in absorption (m^3) or Nm/m^3. This result is obtained by multiplying the ordinate by the abscissa (in appropriate units) of stress-strain plot.

7. **Brittleness:** *Lack of ductility is brittleness.* When a body breaks easily when subjected to shocks it is said to be *brittle.*

8. **Hardness :**

   - *"Hardness" is usually defined as resistance of material to penetration.* Hard materials resist scratches or being worn out by friction with another body.
   - Hardness is primarily a function of the elastic limit (*i.e.*, yield strength) of the material and to a lesser extent a function of the work hardening co-efficient. The modulus of elasticity also exerts a slight effect on hardness.
   - In the most generally accepted test, an indentor is pressed into the surface of the material by slowly applied known load, and the extent of the resulting impression is measured mechanically or optically. A large impression for a given load and indentor indicates soft material, and the opposite is true for small impression.
   - The converse of hardness is known as *softness.*

9. **Fatigue :**

   - When subjected to flucturating or repeating loads (or stresses), materials tend to develop a characteristic behaviour which is different from that (or materials) under *steady loads. Fatigue is the phenomenon that leads to fracture under such conditions.* Fracture takes place under repeated or fluctuating *stresses whose maximum value is less than the tensile strength of the material* (under steady load). Fatigue fracture is progressive, beginning as minute cracks that grow under the action of the fluctuating stress.
   - *Fatigue fracture starts at the point of highest stress.* This point may be determined by the shape of the part; for instant, by stress concentration in a groove. It can also be caused by surface finish, such as tool marks or scratches, and *by internal voids such as shrinking cracks and cooling in castings and weldments and defects introduced during mechanical working and by defects, stresses introduced by electroplating.* It must be remembered that surface and internal defects are stress raisers, and the point of highest actual stress may occur at these rather than at the minimum cross-section of highest normal stress. Thus processing methods are extremely important as they affect fatigue behaviour.

10. **Creep :**

- *Creep is the slow plastic deformation of metals under constant stress or under prolonged loading usually at high temperature. It can take place and lead to fracture at static stresses much smaller than those which will break the specimen by loading it quickly.* Creep is specially taken care of while designing I.C. engines, boilers and turbines.
- The creep at a room temperature is known as *low temperature creep* and occurs in load pipes, roofings, glass as well as in white metal bearings. The creep at high temperatures is known as *high temperature creep*. It mainly depends upon metal, service temperature to be encountered and the *stress involved*. For studying its effects, the specimens are put under a constant load; the creep is measured during various time intervals and results then plotted to get a *creep curve*.

A.1.8. Selection of Materials

General considerations for selection of materials are enumerated below :

1. Mechanical Strength.
2. Ductility.
3. Design.
4. Stability.
5. Availability.
6. Fabricability.
7. Corrosion resistance.
8. Cost.

A.2 FORCE, MOMENTS AND FRICTION

A.2.1. Force

Force is *something which changes or tends to chai ge the state of rest or of uniform motion of a body in a straight line.* Force is the direct or indirect action of one body on another. The bodies may be in direct contact with each other causing direct motion or separated by distance but subjected to gravitational effects.

There are different kinds of forces such as gravitational, frictional, magnetic, inertia or those caused by mass and acceleration. A static force is the one which is caused without relative acceleration of the bodies in question.

The force has a magnitude and direction, therefore, it is **vector**. While the directions of the force is measured in absolute terms of angle relative to a co-ordinate system, the magnitude is measured in different units depending on the situation.

When a force acts on a body, the following effects may be produced in that body : (*i*) *It may bring a change in the motion of the body, i.e., the motion may be accelerated or retarded;* (*ii*) *It may balance the force already acting on the body thus bringing the body to a state of rest or of equilibrium, and* (*iii*) *It may change the size or shape of the body, i.e., the body may be twisted, bent, stretched, compressed or otherwise distorted by the action of the force.*

Characteristics of a force :

The characteristics or elements of the force are the quantities by which a force is fully represented. These are :

1. Magnitude (*i.e.*, 5 kgf, 10 kgf, 50 N, 100 N, etc.)

2. Direction or line of action (angle relative to a co-ordinate system).
3. Sense or nature (push or pull).
4. Point of application.

Representation of forces :

Forces may be represented in the following two ways:

1. Vector representation; 2. Bow's rotation.

Vector representation. A force can be represented graphically by a vector.

Bow's notation. It is a method of designating a force by writing two capital letters one on either side of the force as shown in Fig. A.1, where force P_1 (20 N) is represented by AB and force P_2 (10 N) by CD.

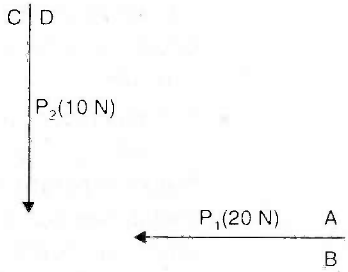


Fig. A.1

Classification of forces :

There are several ways in which forces can be classified. Some of the important classifications are given as under :

1. **According to the effect produced by the force :**
   (*i*) **External force.** *When a force is applied external to a body it is called* ***external force***.
   (*ii*) **Internal force.** *The resistance to deformation, or change of shape, exerted by the material of a body is called an* ***internal force***.
   (*iii*) **Active force.** An ***active force*** *is one which causes a body to move or change its shape.*
   (*iv*) **Passive force.** *A force which prevents the motion, deformation of a body is called a* ***passive force***.
2. **According to nature of the force :**
   (*i*) **Action and reaction.** Whenever there are two bodies in contact, each exerts a force on the other. Out of these forces one is called *action* and other is called *reaction. Action and reaction are equal and opposite.*
   (*ii*) **Attraction and repulsion.** These are actually non-contacting forces exerted by one body or another without any visible medium transmission such as magnetic forces.
   (*iii*) **Tension and thrust.** When a body is dragged with a string the force communicated to the body by the string is called *tension* while, if we push the body with a rod, the force exerted on the body is called a *thrust*.
3. **According to whether the force acts at a point or is distributed over a large area:**
   (*i*) **Concentrated force.** *The force whose point of application is so small that it may be considered as a point is called a* ***concentrated force***.
   (*ii*) **Distributed force.** A ***distributed force*** *is one whose place of application is area.*
4. **According to whether the force acts at a distance or by contact :**
   (*i*) ***Non-contacting forces or forces at a distance.*** Magnetic, electrical and gravitational forces are examples of non-contacting forces or forces at a distance.
   (*ii*) ***Contacting forces or forces by contact.*** The pressure of steam in a cylinder and that of the wheels of a locomotive on the supporting rails are examples of contacting forces.

Force systems :

A *force system is a collection of forces acting on a body in one or more planes.*

According to the relative positions of the lines of action of the forces, the forces may be *classified* as follows:

1. **Coplanar concurrent collinear force system.** It is the simplest force system and includes those forces whose vectors lie along the same straight line (See Fig. A.2).

Fig. A.2. Collinear forces.

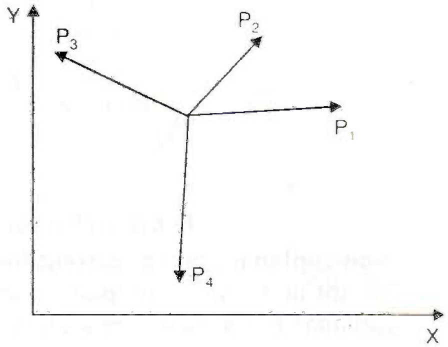


Fig. A.3. Coplanar, concurrent non-parallel forces.

2. **Coplanar concurrent non-parallel force system.** *Forces whose lines of action pass through a common point are called* ***concurrent forces***. In this system lines of action of all the forces meet at a point but have different directions in the same plane as shown in Fig. A.3.
3. **Coplanar non-concurrent parallel force system.** In this system, the lines of action of all the forces lie in the same plane and are parallel to each other but may not have same direction as shown in Fig. A.4.

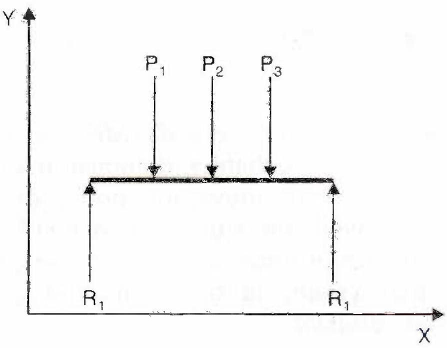


Fig. A.4. Coplanar non-concurrent parallel forces.

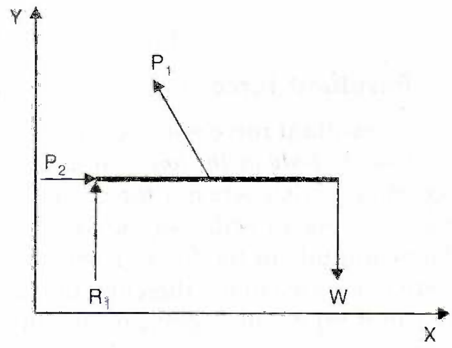


Fig. A.5. Coplanar non-concurrent, non-parallel forces.

4. **Coplanar non-concurrent non-parallel force system.** Such a system exists where the lines of action of all forces lie in the same plane but do not pass through a common point. Fig. A.5 shows such a force system.
5. **Non-coplanar concurrent force system.** This system is evident where the lines of action of all forces do not lie in the same plane but do pass through a common point. An example of this force system is the *forces in the legs of tripod support for camera* (Fig. A.6)

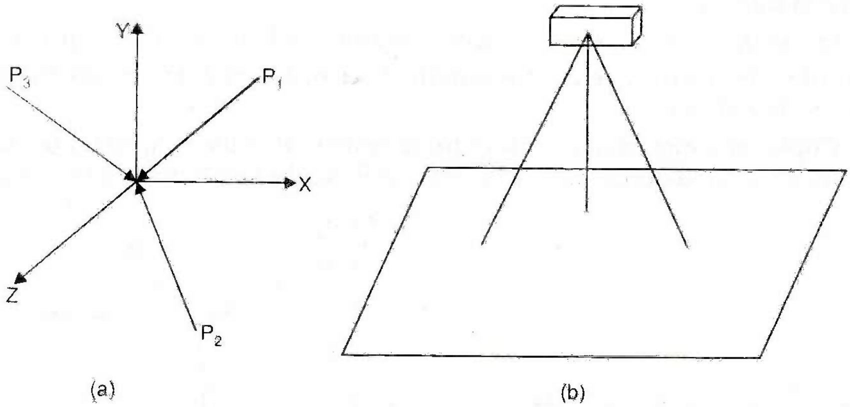


Fig. A.6. Non-coplanar, concurrent forces.

6. **Non-coplanar non-concurrent force system.** Where the lines of action of all forces do not lie in the same plane and do not pass through a common point, a non-coplanar non-concurrent system is present. (Fig. A.7)

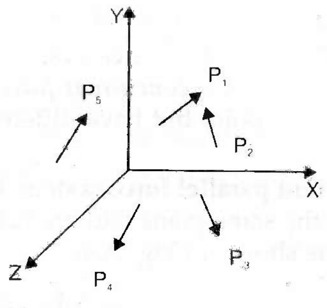


Fig. A.7. Non-coplanar, non-concurrent forces.

Resultant force

A **resultant force** *is a single force which can replace two or more forces and produce the same effect on the body as the forces.* It is fundamental principle of mechanics, demonstrated by experiment, that when a force acts on a body which is free to move, the motion of the body is in the direction of the force, and the distance travelled in a unit time depends on the magnitude of the force. Then for a system of concurrent forces acting on a body, the body will move in the direction of the resultant of that system, and the distance travelled in a unit time will depend on the magnitude of the *resultant.*

Equilibrium conditions for coplanar concurrent forces :

When several forces act on a particle, the particle is said to be in *equilibrium* if there *is no unbalanced* forces acting on it, *i.e.,* the resultant of all the forces acting on the particle is *zero.*

Analytical and graphical conditions of equilibrium of coplanar concurrent forces are given as under :

Analytical conditions :

1. The algebraic sum of components of all the forces in any direction which may be taken as horizontal, in their plane must be zero. Mathematically, $\Sigma H = 0$.
2. The algebraic sum of components of all the forces in a direction perpendicular to

the first direction, which may be taken as *vertical,* in their plane, must be zero. Mathematically, $\Sigma V = 0$.

Graphical conditions. The force polygon, *i.e.,* force or vector diagram *must close.*

Lami's Theorem :

It states as under :

"If three coplanar forces acting on a point in a body keep it in equilibrium, then each force is proportional to the sine of the angle between the other two forces."

Figure A.8, shows three forces P, Q and R acting at a point O. Let the angle between P and Q be γ, between Q and R be α and between R and P be β. If these forces are in equilibrium then according to Lami's theorem :

$$\frac{P}{\sin \alpha} = \frac{Q}{\sin \beta} = \frac{R}{\sin \gamma}$$

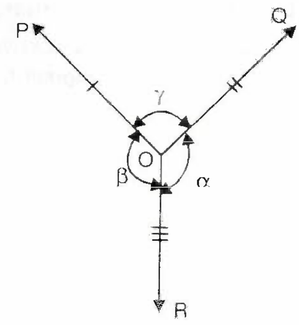


Fig. A.8.

A.2.2. Moments

The tendency of forces is not only to move the body but also to rotate the body. *This rotational tendency of a force is called* ***moment****. The force multiplied by the perpendicular distance from the point to the line of action of the force is called moment about that point.* Unit of moment is equal to the force unit multiplied by the distance unit. It can be in kgfm or Nm etc.

Consider, a finite rigid body capable of rotation about point O as shown in Fig. A.9. The diagram shows the section of the body in the plane of the paper. The axis of rotation is the line perpendicular to the paper and passing through the point O. Let us apply a force P directed along the paper and acting on the body at the point A. The direction PA is the line of action of the force which is perpendicular to OA. Then the moment (or torque) of the force P about the point O is given by the product of force P and the distance OA, *i.e.,*

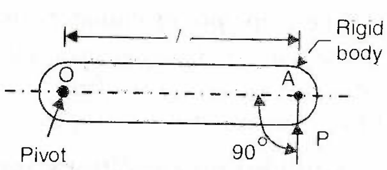


Fig. A.9

$$\text{Moment of force} = \text{Force} \times \text{perpendicular distance}$$
$$= P \times OA = P \times l$$

Moment of a force is vector quantity as it has a magnitude as well as a direction.

It may be noted that the moment of force varies directly with its distance from the pivot. For example, it is much easier to turn a revolving door by pushing at the outer edge of the door, as in Fig. A.10, than by pushing in the centre, as in Fig. A.11.

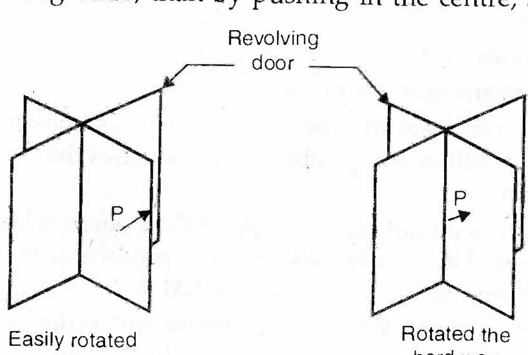


Fig. A.10 **Fig. A.11**

Clockwise and anti-clockwise moments :

If a force P is applied to a body in such a way that it tends to rotate the body in the clockwise sense as, shown in Fig. A.12(a), then the moment is said to be *clockwise.* If, on the other hand, the force P tends to rotate the body in the anti-clockwise sense, as shown in Fig. A.12(b), the moment is said to be *anti-clockwise.*

Conventionally, *clockwise moments* are taken as *negative moments* and *anti-clockwise moments as positive moments.*

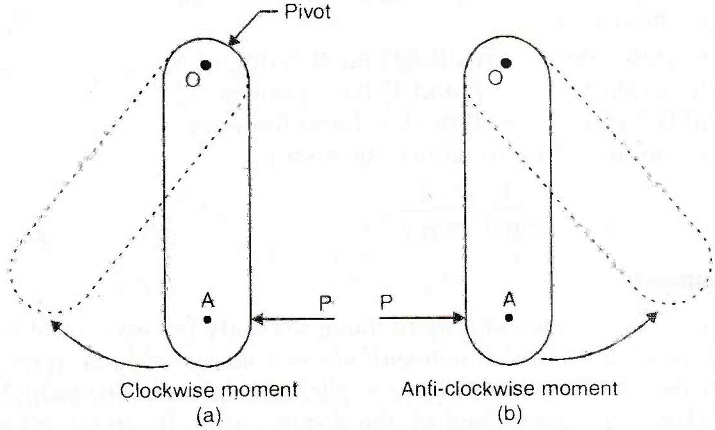


Fig. A.12.

Principle of moments :

The ***principle of moments*** may be stated as follows :

"When a body acted upon by several forces is in rotational equilibrium, the sum of the clockwise moments of the forces about any point is equal to sum of the anti-clockwise moments of the forces about the same point."

Equilibrium conditions for bodies under co-planar non-concurrent forces :

When a body is under the action of a co-planar non-concurrent force system it may rotate due to resultant moment of the force system or it may set in a horizontal or vertical motion due to horizontal and vertical components of forces. The *body, thus can only be in equilibrium if the algebraic sum of all the external forces and their moments about any point in their plane is zero.*

Mathematically, the *conditions of equilibrium* may be expressed as follows :

1. $\Sigma H = 0$ (ΣH means sum of all the horizontal forces)
2. $\Sigma V = 0$ (ΣV means sum of all the vertical forces)
3. $\Sigma M = 0$ (ΣM means sum of all the moments).

When co-planar forces meet in a point the system is known as *co-planar concurrent force system.* This system will be in equilibrium if it satisfies the conditions of equilibrium, *viz.,* $\Sigma H = 0$ and $\Sigma V = 0$.

When co-planar forces do not meet in a point the system is known as ***co-planar non-concurrent force system.*** This system will be in equilibrium if it satisfies all the three conditions of equilibrium, *viz,.* $\Sigma H = 0$, $\Sigma V = 0$, $\Sigma M = 0$.

The conditions $\Sigma H = 0$ and $\Sigma V = 0$ *ensure that the system does not reduce to a single force and condition* $\Sigma M = 0$ *ensures that it does not reduce to a couple.*

In the case of co-planar non-concurrent force system ΣM may be equal to zero but the system may not still be called in equilibrium because the point where the moments are taken about may be lying on the line of action of the resultant. Hence in this case, all the three conditions of equilibrium have to be fulfilled.

Resultant of a coplanar, non-concurrent non-parallel force system :

(*i*) The magnitude, direction and position of the resultant of a given coplanar, non-concurrent, non-parallel force system are found analytically as follows :

$$R = \sqrt{(\Sigma H)^2 + (\Sigma V)^2}$$

where, ΣH = Algebraic sum of the horizontal components of all the forces, and

ΣV = Algebraic sum of vertical components of all the forces.

(*ii*) The *direction* of the resultant is determined by using the relation,

$$\tan \alpha = \frac{\Sigma V}{\Sigma H}.$$

(*iii*) The *position* of the resultant is determined by taking moments of all the rectangular components of forces about a point in their plane and equating the algebraic sum of moments of all the forces to that of the resultant by using the relation,

Moments of resultant '*R*' about the point

= Algebraic sum of rectangular components of all the forces.

A.2.3. Friction

Concept of friction :

It has been observed that surfaces of bodies, however smooth they may be, are not perfect and possess some irregularities and roughness. Therefore, if a block of one substance is placed over the level surface of another, a certain degree of interlocking of minutely projecting particles takes place. This interlocking properties of projecting particles oppose any tendency of the body to move. *The resisting force acts in the direction opposite to that of the motion of the upper block and is called* ***friction***. Thus, *wherever there is a relative motion between two parts, a force of friction comes into play, and hence to overcome friction some energy is wasted.*

Hence, *force of friction or frictional force* may be defined as the *opposing force which is called into play in between the surfaces of contact of two bodies, when one body moves over the surface of another body.* (See Figs. A.13 and A.14).

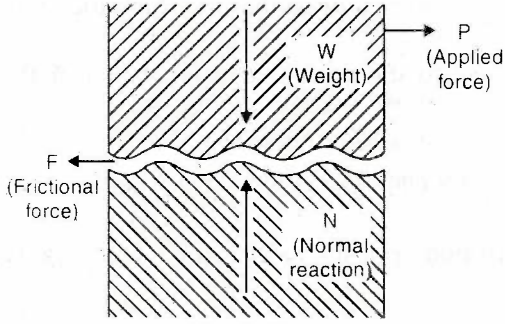


Fig. A.13

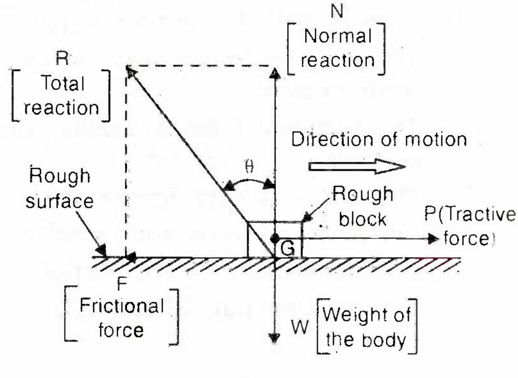


Fig. A.14

In engineering applications friction is both desirable and undesirable.

There are appliances and devices known as *friction* devices such as belts and ropes, friction clutches, jib and cotter joints, brakes, nuts and blots, in which friction is *desirable* and efforts are made to maximise it. On the contrary, the friction is very *undesirable* in *moving parts of machines*. It causes the loss of energy which manifests itself in the forms of heat energy. Due to friction a more force is required to cause motion of the parts. To improve the efficiency of the machines the frictional force is reduced to the minimum possible by *lubrication*.

Static and Dynamic friction :

Static friction. *The static friction is the friction offered by the surfaces subjected to external forces until there is no motion between them.*

Dynamic friction. The dynamic friction is the *friction experienced by a body* when it is in motion. It is also known *as kinetic friction and is always less then static friction* (the kinetic friction is about 40 to 75 per cent of the limiting static friction).

Limiting friction:

Figure A.15, shows a graph between the applied force and the friction. During static condition as the applied force is increased from zero value the frictional force increases in direct proportion to the applied force. A certain stage is reached when the applied force is just sufficient to overcome friction and motion of the body takes place. After this the friction *suddenly decreases* to a magnitude which remains constant throughout the motion period as shown in Fig. A.15.

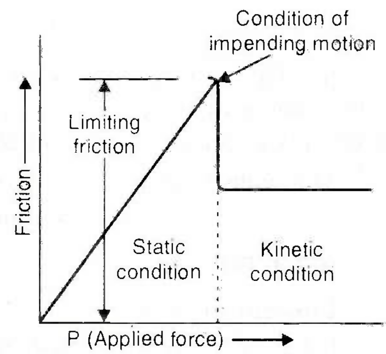


Fig. A.15

When the motion is just to commence, maximum friction is encountered. This condition is known as *limiting equilibrium. The friction acting at this stage is termed as limiting friction.*

Hence, **limiting force of friction** *may be defined as the maximum value of friction force which exists when a body just begins to slide over the surface of the other body.* When the applied force or tractive force P is less than the limiting friction, the body remains at rest, and the friction is called *static friction*, which may have any value between zero and limiting friction.

Laws of friction :

Laws of static friction :

The laws of static friction are as follows:

1. The frictional force always acts in a direction *opposite* to that in which the body tends to move.
2. The frictional force is *directly proportional* to the normal reaction between the surfaces.
3. The frictional force *depends upon the nature of surfaces in contact.*
4. The frictional force is *independent of the area and shape of the contacting surfaces.*

Laws of dynamic or kinetic friction :

1. The frictional force always acts in a direction opposite to that in which the body moves.

2. The frictional force is directly proportional to the normal reaction between the two contacting surfaces.
3. The magnitude of force of dynamic friction bears a constant ratio to the normal reaction between two surfaces but the ratio is slightly *less* than that in case of limiting friction.
4. The frictional force remains *constant for moderate speeds but it decreases slightly with the increase of speed.*

It may be noted that :

(*i*) For extremely low pressure and for very high pressures sufficient to produce excessive deformation, the co-efficient of static friction, *somewhat increases.*

(*ii*) For extremely low relative velocities, the co-efficient of kinetic friction increases and apparently becomes equal to the co-efficient of static friction.

(*iii*) For very high velocities co-efficient of kinetic friction decreases appreciably.

(*iv*) Ordinary changes in temperatures do not materially affect co-efficient of friction.

A.3 STRESSES AND STRAINS

A.3.1. Classification of Loads

A **load** *may be defined as the combined effect of external forces acting on a body.* The loads may be *classified* as: (*i*) deal loads, (*ii*) live or fluctuating loads, (*iii*) inertia loads or forces and (*iv*) centrifugal loads or forces.

The other way of *classification* is (*i*) tensile loads, (*ii*) compressive loads, (*iii*) torsional or twisting loads, (*iv*) bending loads and (*v*) shearing loads.

The load may be a 'point' (or concentrated) or 'distributed'.

Point load. A point load or concentrated load is one which is considered to act at a point. In actual practice, the load has to be distributed over a small area, because, such small knife-edge contacts are generally neither possible, nor desirable.

Distributed load. A distributed load is one which is distributed or spread in some manner over the length of the beam. If the spread is uniform, (*i.e.*, at the uniform rate, say w kN or N/metre run) it is said to be uniformly distributed load and is abbreviated as u.d.l. If the spread is not at uniform rate, it is said to be non-uniformly distributed load. *Triangular* and *trapezoidal* distributed loads fall under this category.

A.3.2. Stress

When a body is acted upon by some load or external force, it undergoes deformation (*i.e.*, change in shape or dimensions) which increases gradually. During deformation, the material of the body resists the tendency of the load to deform the body, and when the load influence is taken over by the internal resistance of the material of the body, it becomes stable. This *internal resistance which the body offers to meet with the load is called* **stress.**

Stress can be considered either as total stress or unit stress. Total stress represents the total resistance to an external effect and is expressed in N, kN or MN. Unit stress represents the resistance developed by a unit area of cross-section, and is expressed in KN/m^2 or MN/m^2 or N/mm^2. For the remainder of this text, the word stress will be used to signify unit stress.

The various types of stresses may be *classified* as :

1. **Simple or direct stress :**

(*i*) Tension; (*ii*) Compression; (*iii*) Shear.

2. **Indirect stress :**
   (*i*) Bending; (*ii*) Torsion.
3. **Combined stress.** Any possible combination of types 1 and 2.

A.3.3. Simple Stress

Simple stress is often called *direct stress* because it develops under direct loading conditions. That is, simple tension and simple compression occur when the applied force, called load, is in line with the axis of the member (axial loading) (Figs. A.16 and A.17), and simple shear occurs, when equal, parallel, and opposite forces tend to cause a surface to slide relative to the adjacent surface (Fig. A.18).

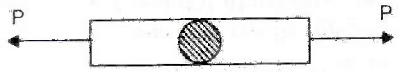


Fig. A.16. Tensile stress.

Fig. A.17. Compressive stress.

In certain loading situations, the stresses that develop are not simple stresses. For example, referring to Fig. A.19, the member is subjected to a load which is perpendicular to the axis of the member (transverse loading) (Fig. A.20). This will cause the member to bend, resulting in deformation of the material and stresses being developed internally to resist the deformation. All three types of stresses—tension, compression and shear—will develop, but they will not be simple stresses, since they were not caused by direct loading.

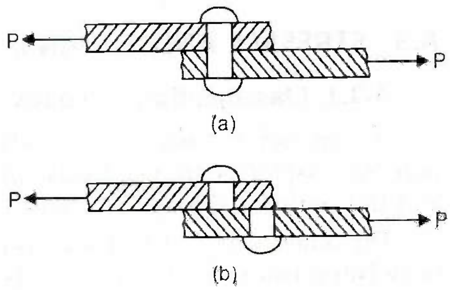


Fig. A.18. (a) Rivet resisting shear (b) Rivet failure due to shear.

When any type of simple stress σ (sigma) develops, we can calculate the magnitude of the stress by,

$$\sigma = \frac{P}{A}$$

where σ = Stress, kN/m^2 or N/mm^2

P = Load $\left[\begin{array}{l}\text{external force causing} \\ \text{stress to develop}\end{array}\right]$, kN or N

A = Area over which stress develops, m^2 or mm^2.

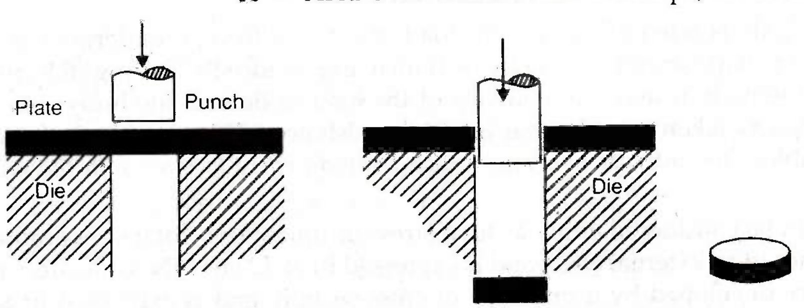


Fig. A.19. (*a*) Punch approaching plate; (*b*) Punch shearing plate; (*c*) Slug showing sheared area.

It may be noted that in cases of either simple tension or simple compression, the areas which resist the load are perpendicular to the direction of forces. When a member is subjected to simple shear, the resisting area is parallel to the direction of the force. Common situations causing shear stresses are shown in Figs. A.18 and A.19.

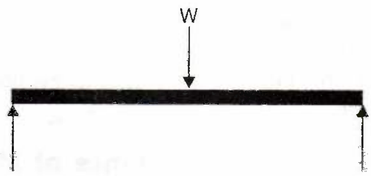


Fig. A.20. Simply supported beam (Transverse loading).

A.3.4. Strain

Any element in a material subjected to stress is said to be strained. *The strain (e) is the deformation produced by stress.* The various types of strains are explained below :

Tensile strain :

A piece of material, with uniform cross-section, subjected to a uniform axial tensile stress, will increase its length from l to $(l + \delta l)$ (Fig. A.21) and the increment of length δl is the actual deformation of the material. The fractional deformation or the tensile strain is given by

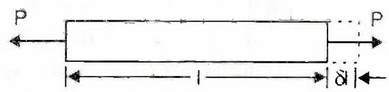


Fig. A.21.

$$e_t = \frac{\delta l}{l}$$

Compression strain :

Under compressive forces, a similar piece of material would be reduced in length (Fig. A.22) from l to $(l - \delta l)$.

The fractional deformation again gives the strain e_c.

Fig. A.22.

where, $$e_c = \frac{\delta l}{l}.$$

Shear strain :

In case of shearing load, a shear strain will be produced which is *measured by the angle through which the body distorts.*

In Fig. A.23 is shown a rectangular block $LMNP$ fixed at one face and subjected to force F. After application of force, it distorts through an angle ϕ and occupies new position $LM'\,N'P$. The shear strain (e_s) is given by

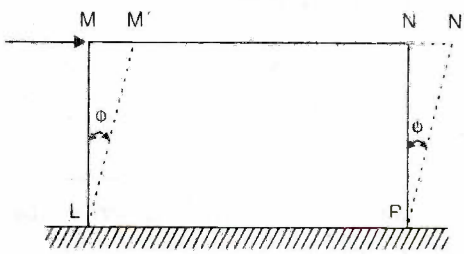


Fig. A.23.

$$e_s = \frac{NN'}{NP} = \tan\phi$$

$$= \phi \text{ (radians) ..... since } \phi \text{ is very small.}$$

The above result has been obtained by assuming NN' equal to arc (as NN' is small) drawn with centre P and radius PN.

Volumetric strain :

It is defined as the *ratio between change is volume and original volume of the body,* and is denoted by e_v,

$$e_v = \frac{\text{Change in volume}}{\text{Original volume}} = \frac{dV}{V}$$

The strains which disappear with the removal of load are termed as *elastic strains* and the body which regains its original position on the removal of force is called an *elastic body*. The body is said to be *plastic* if the *strains exist even after the removal of external force.* There is always a limiting.

A.3.5. Importance of Mechanical Tests

- Structures, machines and products of various kinds are usually subjected to load and deformation. Therefore, the properties of materials under the action of load and deformation so produced under various environments become an important engineering consideration. *The microscopic properties of materials under applied forces or loads are broadly classed as mechanical properties.* They are a measure of the strength and lasting characteristic of a material in service and are of great importance particular to the design engineer. Unfortunately these properties cannot be desired from the structural or bonding considerations alone since most of them are structure-sensitive, are much more *affected by crystal imperfections* and other factors such as composition, grain size, heat treatment etc. Therefore, mechanical properties do not depend on them in all situations. A great number of mechanical properties, are, therefore, best evaluated by mechanical testing of the materials like metals and alloys.
- The following important mechanical tests give valuable information about metals and alloys as given below :

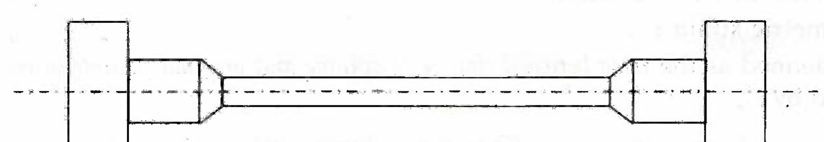
S. No.	Name of test	Information supplied about ...
1.	*Tensile test*	Tensile strength, yield point, elastic limit. Young's modulus, ductility, toughness etc.
2.	*Impact test*	Toughness of a material under shock loading condition.
3.	*Hardness test*	Wear resistance, indentation resistance, scratch resistance or cutting ability of material.
4.	*Fatigue test*	Behaviour of a material under repeatedly applied stress and its endurance limit.
5.	*Creep test*	Behaviour of a material under a steady load over a long period of time and creep limit of a material.

Tensile test (only) is described below :

Tensile test :

- The tensile test is one of the most widely used of the mechanical tests. There are many variations of this test to accommodate the widely differing character of materials such as metals, elastomers, plastics and glasses. The tensile test on a mild steel test piece is described below :

  Fig. A.24 shows standard specimens used for the test.



(a)

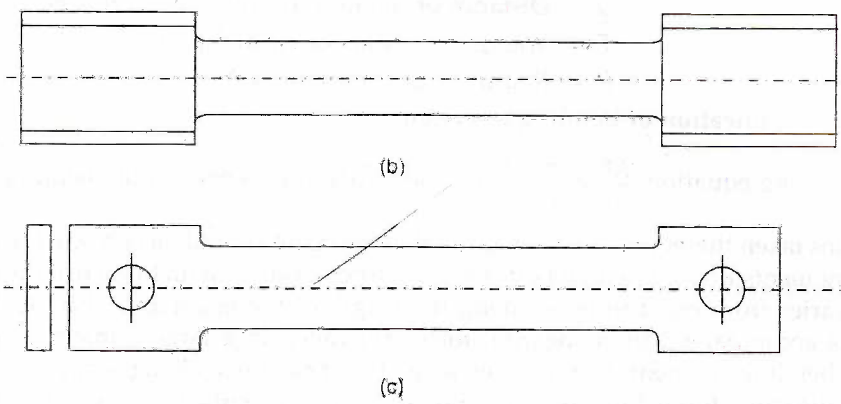


Fig. A.24. Tensile test specimens.

- The tensile test is carried out on a bar of uniform cross-section throughout the gauge length. The specimen is mounted in the jaws of a testing machine with which a gradually increasing load can be applied. The extension or elongation of the gauge length is recorded continuously and finally a graph is drawn between the loads and extensions or between the stress and strain; which is of the type shown in Fig. A.25

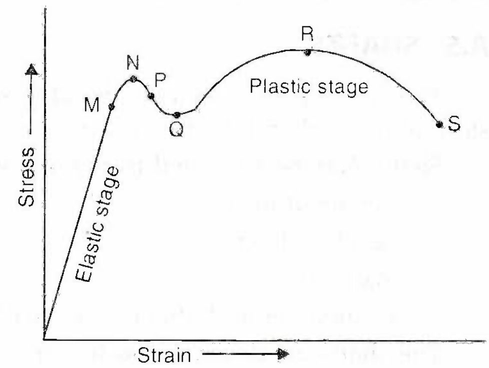


Fig. A.25. Stress-strain curve.

Upto the point M Hooke's law holds good and this point is known as *"limit of proportionality"*. *Beyond* the point M Hooke's law is not obeyed although the material remains elastic *i.e.*, strain completely disappears after the removal of load. At the point *N elastic limit is reached.* If the material is loaded or stressed upto this point the material will regain its original shape on the removal of the load. Up to the point P strain increases more quickly than stress, at this point the metal yields. In the mild steel yielding commences immediately and two point P and Q, the *upper and lower yield points* respectively are obtained. On further increasing the load slightly, the *strain increases rapidly* till R when *neck or waist* is formed. When this point (R) is reached the deformation or extension continues even with lesser load and ultimately *feature occurs.*

A.4 BENDING OF BEAMS

Bending equation is given by:

$$\frac{M}{I} = \frac{\sigma}{y} = \frac{E}{R}$$

where,

M = Moment of resistance,

I = Moment of inertia of the section about neutral axis (N.A.)

σ = Bending stress,

y = Distance of the fibre from the neutral axis,

E = Young's modulus of elasticity, and

R = Radius of curvature of N.A.

Practical application of bending equation :

The bending equation $\frac{M}{I} = \frac{\sigma}{y} = \frac{E}{R}$ is based upon the theory of pure bending and the assumptions taken thereupon, which require that the beam should be subjected to constant bending moments unaccompanied by shearing forces, but in actual practice the bending moment varies from point to point along the length of the beam and also, the bending moment is accompanied by a shearing force. However, in a large number of practical cases, the bending moment is maximum when the shear force changes sign, *i.e.*, crosses the zero shearing force line. In this way the requirements to simple bending are approximately satisfied at the point of maximum bending moment and therefore, it seems justifiable to apply the bending equation at that point only.

A.5 SHAFTS

The shafts are *usually* cylindrical in section, solid or hollow. They are made of mild steel, alloy steel and copper alloys.

Shafts may be subjected to the following loads :

1. Torsional load.
2. Bending load.
3. Axial load.
4. Combination of above three loads.

The shafts are designed on the *basis of strength and rigidity.*

The following values are usually adopted for the design of shaft :

σ = 112 MN/m^2, the maximum permissible tensile or compressive stress.

τ = 56 MN/m^2, the maximum permissible shear stress.

The ultimate tensile stress for commercial steel shafting may be 315 MN/m^2 for hot rolled and turned low carbon steel and 490 MN/m^2 for cold finished low carbon steel, corresponding stresses at the elastic limit would be about 160 MN/m^2 and 315 MN/m^2 respectively. In shafts with key ways the allowable stresses are 75% of the values given.

A.5.1. Torsion of Shafts

To transmit energy by rotation it is necessary to apply a turning force. In case of a shaft if the force is applied tangentially and in the plane of transverse cross-section the torque or twisting moment may be calculated by multiplying the force with the radius of the shaft. If the shaft is subjected to two opposite turning moments it is said to be in *pure torsion* and it will exhibit the tendency of shearing off at every cross-section which is perpendicular to longitudinal axis.

A.5.2. Torsion Equation

Torsion equation is given by : Refer to Fig. A.26.

$$\frac{T}{I_p} = \frac{C\theta}{l} = \frac{\tau}{R}$$

where, T = Maximum twisting torque,

R = Radius of the shaft,
I_p = Polar moment of inertia,
τ = Shear stress,
C = Modulus of rigidity
θ = The angle of twist (radians), and
l = length of the shaft.

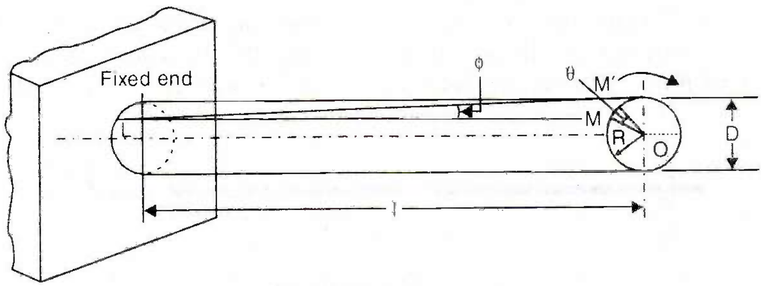


Fig. A.26

A.5.3. Power Transmitted by the Shaft

Consider a force F newtons acting tangentially on the shaft of radius R. If the shaft due to this turning moment ($F \times R$) starts rotating at N r.p.m. then

Work supplied to the shaft/sec.

$$= F \times \text{distance moved/sec.}$$
$$= F \times 2\pi RN/60 \text{ Nm/s}$$

or, $$P = \frac{F \times 2\pi\, RN}{60} \text{ watts} = \frac{T \times 2\pi N}{60 \times 1000} \text{ kW} \qquad (\because T = F \times R)$$

Hence, $$P = \frac{2\pi\, NT}{60 \times 1000} \text{ kW} \qquad ...(i)$$

where T is the mean/average torque in Nm.

A.6 BENDING MOMENTS AND SHEARING FORCES

A.6.1. Some Basic Definitions

Beam. Beam is a structural member which is acted upon by a system of external loads at right angles to the axis.

Bending. Bending implies deformation of a bar produced by loads perpendicular to its axis as well as force-couples acting in a plane passing through the axis of the bar.

Plane bending. If the plane of loading passes through one of the principal centroidal axes of the cross-section of the beam, the bending is said to be *plane* (or *direct*).

Oblique bending. If the plane of loading does *not* pass through one of the principal centroidal axes of the cross-section of the beam, the bending is said to be *oblique.*

Point load. A *point load* or *concentrated load is one which is considered to act at a point.* In actual practice, the load has to be distributed over a small area because such small knife-edge contacts are generally neither possible nor desirable.

Distributed load. A *distributed load is one which is distributed or spread in some manner over the length of the beam.* If the spread is uniform (*i.e.*, at the uniform rate, say *w* kN/metre run) it is said to be uniformly distributed load and is abbreviated as U.D.L. If the spread is not a uniform rate, it is said to be non-uniformly distributed load. *Triangular* and *Trapezium* distributed loads fall under this category.

A.6.2. Classification of Beams

Depending upon the type of supports beams are classified as follows:

1. **Cantilever.** A cantilever is a beam whose one end is fixed and the other end free. Fig. A.27 shows a cantilever with end *A* rigidly fixed into its support and the other end *B* free. The length between *A* and *B* is known as the *length of cantilever.*

Fig. A.27. Cantilever.

2. **Simply (or freely) supported beam.** A simply supported beam is one whose ends freely rest on wall or columns or knife edges (Fig. A.28). In all such cases the reactions are *always upwards.*

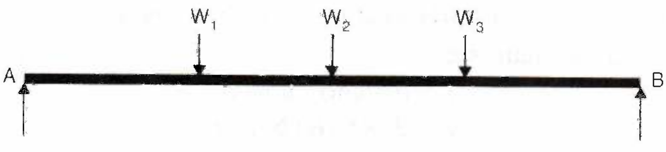


Fig. A.28. Simply supported beam.

3. **Overhanging beam.** An overhanging beam is one in which the supports are not situated at the ends *i.e.*, one or both the ends project *beyond the supports.* In Fig. A.29 *C* and *D* are two supports and both the ends *A* and *B* of the beam are overhanging beyond the supports *C* and *D* respectively.

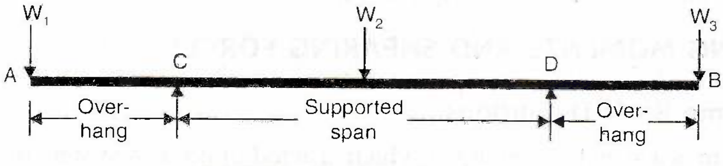


Fig. A.29. Overhanging beam.

4. **Fixed beam.** A fixed beam is one whose both ends are rigidly fixed or built into its supporting walls or columns (Fig. A.30).

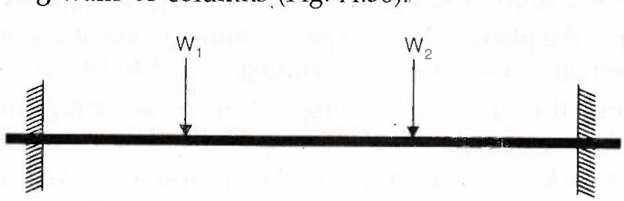


Fig. A.30. Fixed beam.

5. **Continuous beam.** A continuous beam is one which has *more than two supports* (Fig. A.31). The supports at the extreme left and right are called the *end supports* and all the other supports, except the extreme, are called *intermediate supports.*

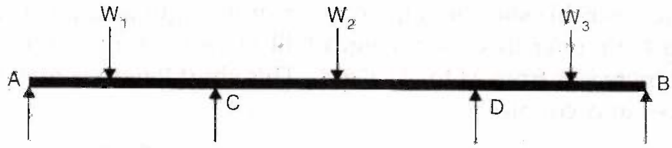


Fig. A.31. Continuous beam.

It may be noted that the first three types of beams (*i.e.*, cantilevers, simply supported beams and overhanging beams) are known as *Statically Determinate Beams* as the *reactions of these beams at their supports can be determined by the use of equations of static equilibrium and the reactions are independent of the deformation of beams.* The last two types of beams (*i.e.*, fixed beams and continuous beams) are known as *Statically Indeterminate Beams as their reactions at supports cannot be determined by the use of equations of static equilibrium.*

A.6.3. Shearing Force (S.F.) and Bending Moment (B.M.)

When a beam, which is in equilibrium under a series of forces, is cut in some section X, and the beam to the left to the section remains in equilibrium (Fig. A.32), then some force must act at the section. Prior to cutting, this force would be provided by the adjacent material, and would act tangentially to the section. Hence there will be a shearing force at the section. Numerically this shearing force will be given by the algebraic sum of the forces to the left or to the right of the section. As a convention, an upward force to the left of a section is counted as producing *negative* shearing force. Similarly an upward force to the right of the section will produce *positive* shearing force.

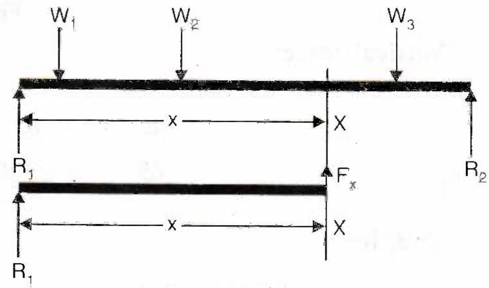


Fig. A.32.

Considering further the equilibrium of the material to the left of the section X (Fig. A.32), it follows that there can be no resultant moment to the left of the section. Hence, any moment produced by the forces acting on the beam must be balanced by an equal and opposite moment produced by the internal forces acting in the beam at the section. This is the bending moment at the section.

The bending moment is the *algebraic sum of moments to the left or right of the section.* In each case, by considering equilibrium, either for forces or moments, the resultant, caused by the applied forces to one side of the section is balanced by the bending moment and shearing force acting at the section. The *sign convention for bending moments* is that a beam in ***"hogging"*** condition is subject to **negative** bending moment, and one in a ***"sagging"*** condition to **positive** bending moment (Fig. A.33).

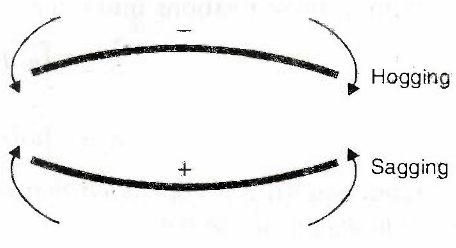


Fig. A.33.

A.6.4. General Relation between the Load, the Shearing Force and the Bending Moment

Refer to Fig. A.34. Consider a short length δx of a beam at a distance x from some origin. Let the load over this short length be w per unit length acting vertical downwards; then the shearing force over this short length will increase from S to $(S + \delta S)$ while the bending moment increases from M to $(M + \delta M)$. This short length is in equilibrium under both vertical forces and couples.

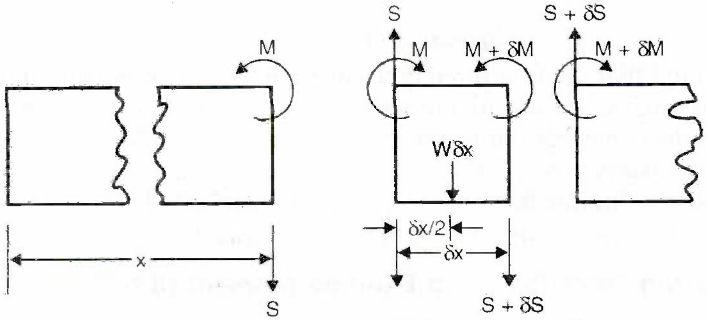


Fig. A.34.

Vertical forces :

$$(S + \delta S) - S = w\,\delta x$$

$$\therefore \quad \delta S = w \, . \, \delta x$$

or,
$$\frac{\delta S}{\delta x} = w \text{ or } \frac{dS}{dx} = w \qquad ...(i)$$

Couples :

$$M - (M + \delta M) = -S\delta x + w\delta x\left(\frac{\delta x}{2}\right)$$

or,
$$-\delta M = -S \cdot \delta x + \frac{w}{2}(\delta x)^2$$

Since $(\delta x)^2$ is a small quantity of the second order, it may be taken as zero.

$$\therefore \quad -\delta M = -S\delta x \quad \text{or} \quad S = \frac{\delta M}{\delta x}$$

and in the
$$\underset{\delta x \to 0}{\text{Lt}}\ S = \frac{dM}{dx} \qquad ...(ii)$$

Putting these relations into integral form we have,

$$S = \int w\,dx \qquad ...(iii)$$

$$M = \int s\,dx \qquad ...(iv)$$

From eqn. (*i*) it is concluded that *the rate of change of S.F. at any section represents the rate of loading at the section.*

From equation (*ii*) it is concluded that the *rate of change of B.M. at any section represents the S.F. at that section.*

The *B.M. M shall be maximum or minimum when* $\frac{dM}{dx}$ *= 0, i.e., S* = 0. Thus at the *sections* *where S.F. is zero or changes sign* (because then it passes through zero) *the B.M. is either maximum or minimum.*

A.7 METROLOGY

Metrology, in literary sense, *means the pure science of measurements.* But for engineering purposes, it is restricted to measurements of length and angles and other quantities which are expressed in linear or angular measurements.

A.7.1. Standards of Measurements

These days only *two standard systems of linear measurement, English* (yard) and *Metric* (metre) are in general use throughout the world.

The *metric system* was *originated in France* and is now being used in many countries in the world.

The British system of linear measurement is based on one arbitrarily unit known as *yard.*

For linear measurements the various standards known are :

1. Line standard.
2. End standard.
3. Wavelength standard.

Line standard :

***A yard** or **metre** is defined as the distance between scribed lines on a bar of metal under certain conditions of temperature and support. These* are legal standards and Act of Parliament authorises their use.

- The ***Metre*** is defined as 1650763.73 wavelengths of the orange radiation in vacuum of krypton-86 isotype.
- The ***Yard*** is defined as 0.9144 metre. This is equivalent to 1509458.35 wavelengths of the same radiation.

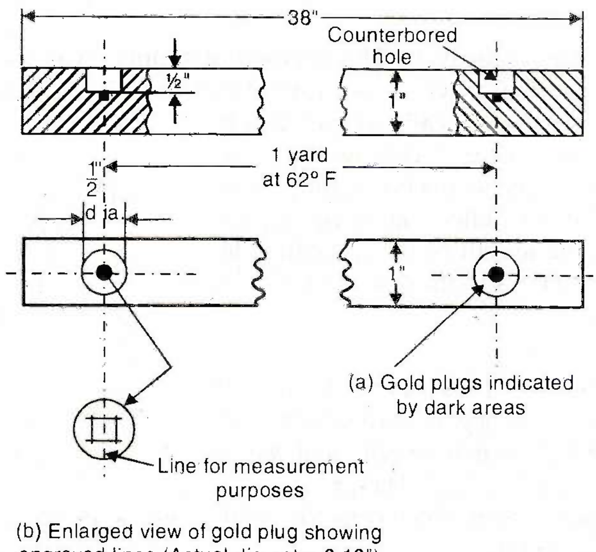


Fig. A.35. Imperial standard yard.

Yard :

A Yard was formerly known as the Imperial Standard Yard. Fig. A.35 shows a diagram indicating its essential features. It consists of a bronze bar made from an alloy known as Baily's metal, consisting of 16 parts copper, $2\frac{1}{2}$ parts tin and 1 part zinc. The bar, 1 sq. in. in cross-section has an overall length of 36". Two counter bored holes, $\frac{1}{2}$" diameter by $\frac{1}{2}$" deep, at 36" centres (1" from each end of the bar) provide sighting holes for two gold plugs inserted in the holes at the base of each counter bore. The faces of the gold plugs are flush with bases of counter bores and, therefore, lie in the neutral plane of the bar where bending effects are minimised when the bar is resting on supports. These plugs are 0.01 in. diameter, and five lines are ruled on the upper polished face of each; three lines at right angles to the length of the bar and two parallel to the bar as shown in Fig. A.35(*b*). *The length of the yard is defined as the distance between the two central transverse lines on the plugs when the temperature of the bar is constant at 62°F, and when the bar is supported on rollers, in a specified manner, to prevent flexure.*

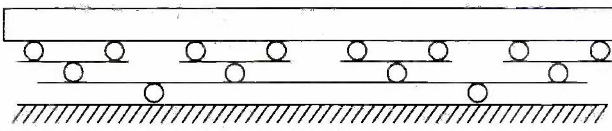


Fig. A.36. 8-point supporting system for imperial yard till 1922.

- The original procedure, when intercomparisons were made between the standard and its copies, was to float the bars in the mercury; but proof that the bars could be effectively supported on rollers, while maintaining the previous accuracy, was provided by Airy. His method is shown in Fig. A.36. The standard was directly supported on eight equally spaced rollers in conjunction with a special frame. The distance between the rollers was proved by Airy to be equal to $\frac{l}{\sqrt{n^2-1}}$ where, n and l represent number of rollers and length of bar respectively. This method of support was used for the purpose of inter-comparisons until 1922 when two supports at the Airy points were introduced. When a bar is supported specifically at two points symmetrically about its centre, a condition can be produced when the bar ends lie in a horizontal plane. With this condition the bar deflects at its centre, but the effective error in the length of bar is negligible. Applying Airy's formula the specific distance between the supports is equal to $0.577\ l$.

Metre :

The *length of the metre is defined as the distance, at 0°C between the centre portions of pure platinum–irridium alloy (10% irridium) of 102 cm total length—and having a cross-section as shown in Fig. A.37.* The graduations are on the upper surface of web which coincides with the neutral axis of the section.

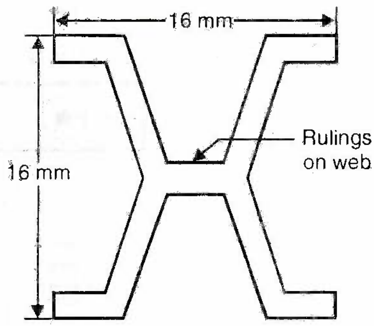


Fig. A.37. International prototype metre (Cross-section).

Sub-division of standards :

The imperial standard yard and international prototype metre defined previously are considered to be perfect or master standards and cannot usually be used for general purposes. Thus depending upon the importance of accuracy required for the work, standards are subdivided into *four grades.*

1. Primary standards. To ensure that standard unit of length, *i.e.,* yard or metre does not change its value and it is strictly followed and precisely defined there should be one and only one material standard preserved under most careful conditions. This has no direct application to a measuring problem encountered in engineering. *These are used only at rare intervals and solely for comparison with secondary standards.*

2. Secondary standards. Secondary standards are made as nearly as possible to the primary standards with which they are compared at intervals. Any error existing in these bars is recorded by comparison with primary standards after long intervals. *These standards are distributed to a number of places for safe custody and used in their turn for occasional comparison with tertiary standards. These standards also act as safeguard against the loss or destruction of primary standards.*

Materials for secondary standards :

(*i*) Invar—an alloy of nickel and steel.

(*ii*) Fuse silica.

(*iii*) Elinvar—an alloy of nickel and chromium.

All the above materials have *usually very low coefficient of linear expansion.*

3. Tertiary standards. Tertiary standards are the *first standards to be used for reference purposes in laboratories and workshops. These* should also be maintained as a *reference for comparison at intervals with working standards.*

4. Working standards. These standards are necessary for use in metrology laboratories and similar institutions. These are derived from fundamental standards.

Sometimes standards are classified as :

- ***Reference standards*** (used for reference purposes)
- ***Calibration standards*** (used for calibration of inspection and working standards)
- ***Inspection standards*** (used by inspectors)
- ***Working standards*** (used by operators).

Characteristics of line standards :

The characteristics of line standard are given below :

1. Accurate engraving on the scales can be done but it is difficult to take full advantage of this accuracy. For example, a steel rule can be read to about ±0.2 mm of true dimension.
2. It is easier and quicker to use a scale over a wide range.
3. The scale markings are not subject to wear although significant wear on leading end *leads to undersizing.*
4. There is no *'built in'* datum in a scale which would allow easy scale alignment with the axis of measurement, this again *leads to undersizing.*
5. Scales are subjected to the parallax effect, a source of both positive and negative reading errors.
6. For close tolerance length measurement (except in conjunction with microscopes) scales are not convenient to be used.

End standard :

End standards, in the form of the bars and slip gauges, are in general use in precision engineering as well as in standard laboratories such as the N.P.L. (National Physical Laboratory). Except for applications where microscopes can be used, scales are not generally convenient for the direct measurement of engineering products, whereas slip gauges are in everyday use in tool-rooms, workshops, and inspection departments throughout the world.

A *modern end standard consists fundamentally of a block or bar of steel generally hardened whose end faces are lapped flat and parallel to within a few millionth of a cm.* By the process of lapping, its size too can be controlled very accurately. Although, from time to time, various types of end bar have been constructed, some having flat and some spherical faces, *the flat, parallel faced bar is firmly established as the most practical method of end measurement.*

Characteristics of end standards :

1. Highly accurate and well suited to close tolerance measurements.
2. Time-consuming in use.
3. Dimensional tolerance as small as 0.0005 mm can be obtained.
4. Subjected to wear on their measuring faces.
5. To provide a given size, the groups of blocks are *"wrung"* together. Faulty wringing leads to damage.
6. There is a *"built-in"* datum in end standards, because their measuring faces are flat and parallel and can be positively located on a datum surface.
7. As their use depends on *"feel"* they are not subject to the parallax effect.

End bars. Primary end standards usually consist of *bars of carbon steel* about 20 mm in diameter and made in sizes varying from 10 mm to 1200 mm. These are hardened at the ends only. They are used for the measurement of work of larger sizes.

Slip gauges. *Slip gauges* are used as standards of measurement in practically *every precision engineering* works in the world. These were invented by C.E. Johansom of Sweden early in the present century. These are made of *high-grade cast steel and are hardened throughout.* With the set of slip gauges, combination of slip gauge enables measurements to be made in the range of 0.0025 to 100 mm but in combinations with end/length bars measurement range upto 1200 mm is possible.

Note: The accuracy of line and end standards is affected by temperature changes and both are originally calibrated at 20 ± ½°C. Also care is taken in manufacture to ensure that change of shape with time is reduced to negligible proportions.

Wavelength standard :

In 1827, Jacqnes Babinet, a French philosopher, suggested that wavelengths of monochromatic light might be used as natural and invariable units of length. It was nearly a century later that the Seventh General Conference of Weights and Measures in Paris approved the definition of a standard of length relative to the metre in terms of the wavelength of the red radiations of cadmium. Although this was not the establishment of a new legal standard of length, it set the seal on work which kept on going for a number of years.

- *Material standards are liable to destruction and their dimensions change slightly with time. But with the monochromatic light we have the advantage of constant wavelength and since the wavelength is not a physical one, it need not be preserved. This is reproducible standard of length, and the error of reproduction can be of the order of 1 part in 100* *millions. It is because of this reason that International standard measures the metre in terms of wavelength of krypton 86 (Kr 86).*

- Light wavelength standard, for sometime, had to be objected because of the impossibility of producing pure monochromatic light as wavelength depends upon the amount of isotope impurity in the elements. But now with the rapid development in atomic energy industry, pure isotopes of natural elements have been produced. *Krypton* 85, *Mercury* 198 *and Cadmium* 114 are possible sources of radiation of wavelength suitable as natural standard of length.

Advantages of wavelength standards :

The following are the *advantages* of using wavelength standard as basic unit to define primary standards :

1. It is not influenced by effects of variation of environmental temperature, pressure, humidity and ageing because it is not a material standard.
2. There is no need to store it under security and thus there is no fear of its being destroyed as in the case of yard and metre.
3. It is easily available to all standardising houses, laboratories and industries.
4. It can be easily transferred to other standards.
5. This standard can be used for making comparative statement of a much higher accuracy.
6. It is easily reproducible.

Classification of standards :

To maintain accuracy and interchangeability it is necessary that the standards be traceable to a single source, usually the National Standards of the country, which are further linked to International Standards. The accuracy of National Standards is transferred to working standards through a chain of intermediate standards in a manner given below :

National Standards
↓
National Reference Standards
↓
Working Standards
↓
Plant Laboratory Reference Standards
↓
Plant Laboratory Working Standards
↓
Shop Floor Standards

Evidently, there is degradation of accuracy in passing from the defining standards to the shop floor standards. The *accuracy of a particular standard depends on a combination of the number of times it has been compared with a standard in a higher echelon, the frequency of such comparisons, the care with which it was done, and the stability of the particular standard itself.*

Relative characteristics of line and end standards :

The relative characteristics of *line* and *end standards* are given in the tabular form below :

S. No.	Aspects	Line standard	End standard
1.	*Manufacture and cost of equipment*	Simple and low.	Complex process and high.
2.	*Accuracy in measurement*	Limited to ±0.2 mm. In order to achieve high accuracy, scales have to be used in conjunction with microscopes.	Very accurate for measurement of close tolerances upto ±0.001 mm.
3.	*Time of measurement*	Quick and easy.	Time consuming.
4.	*Effect of use*	Scale markings not subject to wear but the end of scale is worn. Thus it may be difficult to assume zero of scale as datum.	Measuring faces get worn out. To take care of this end pieces can be hardened, protecting type. Built-in datum is provided.
5.	*Other errors*	There can be parallax error.	Errors may get introduced due to improper wringing of slip gauges. Some errors may be caused due to change in laboratory temperature.

A.7.2. Limits, Fits and Tolerance

General aspects :

- In the design and manufacture of engineering products a great deal of attention has to be paid to the *mating, assembly and fitting of various components.* In the early days of mechanical engineering during the nineteenth century, the majority of such components were actually mated together, their dimensions being adjusted until the required type of fit was obtained. These methods demanded craftsmanship of a high order and a great deal of very fine work was produced.
- Present day standards of quantity production, interchangeability, and continuous assembly of many complex compounds, could not exist under such a system, neither could many of the exacting design requirements of modern machines be fulfilled without the knowledge that certain dimensions can be reproduced with precision on any number of components.
- *Modern mechanical production engineering is based on a system of limits and fits, which, while not only itself ensuring the necessary accuracies of manufacture, forms a schedule or specifications to which manufacturers can adhere.*

In order that a system of limits and fits may be successful, following *conditions* must be fulfilled :

(*i*) The range of sizes covered by the system must be sufficient for most purposes.

(*ii*) It must be based on some standards so that everybody understands alike and a given dimension has the same meaning at all places.

(*iii*) For any basic size it must be possible to select from a carefully designed range of fit the most suitable one for a given application.

(*iv*) Each basic size of hole and shaft must have a range of tolerance values for each of the different fits.

(*v*) The system must provide for both unilateral and bilateral methods of applying the tolerance.

(*vi*) It must be possible for a manufacturer to use the system to apply either a hole-based or a shaft-based system as his manufacturing requirements may need.

(*vii*) The system should cover work from high class tool and gauge work where very wide limits of sizes are permissible.

Nominal size and basic dimensions:

Nominal size. A *'nominal size' is the size which is used for purpose of general identification.* Thus the nominal size of a hole and shaft assembly is 60 mm, even though the basic size of the hole may be 60 mm and the basic size of the shaft 59.5 mm.

Basic dimension. A *'basic dimension' is the dimension, as worked out by purely design considerations.* Since the ideal conditions of producing basic dimension, do not exist, the basic dimension can be treated as theoretical or nominal size, and it has only to be approximated. A study of function of machine part would reveal that it is unnecessary to attain perfection because some variations in dimension, however small, can be tolerated on size of various parts. It is, thus, *general practice to specify a basic dimension and indicate by tolerances as to how much variation in the basic dimension can be tolerated without affecting the functioning of the assembly into which this part will be used.*

Definitions :

The definitions given below are based on those given in IS : 919 *Recommendation for Limits and Fits for Engineering,* which is in line with the ISO recommendation.

Shaft. The term *shaft* refers not only to diameter of a circular shaft but to *any external dimension on a component.*

Hole. This term refers *not only* to the diameter of a circular hole but to *any internal dimension on a component.*

Actual size of the shaft. This is the measured dimensions of the part.

Basic size. Refer to Fig. A.38. The *basic size is the standard size for the part and is the same for both the hole and its shaft.* ***Example :*** A 60 mm diameter hole and shaft.

Zero line. Refer to Fig. A.38. *This is the line which represents the basic size so that the deviation from the basic size is zero.*

Limits of size. These are *maximum and minimum permissible sizes of the part.*

Minimum limit of size. Refer to Fig. A.39. The *minimum size permitted for the part.*

Maximum limit of size. Refer to Fig. A.38. The *maximum size permitted for the part.*

Tolerance. Refer to Fig. A.39. *The difference between the maximum and minimum limits of size.*

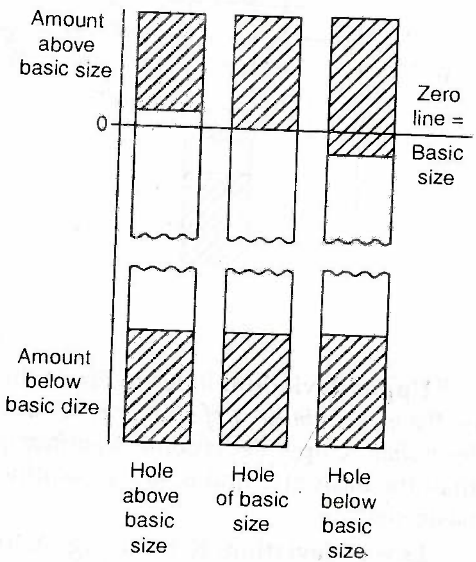


Fig. A.38

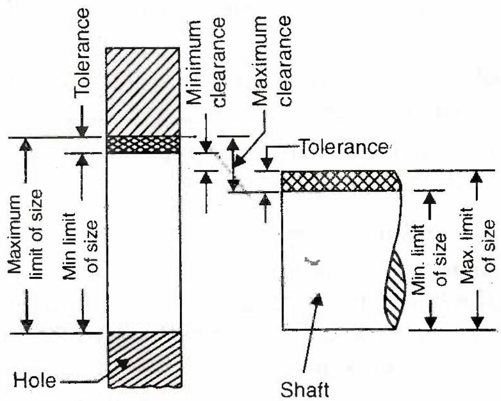


Fig. A.39

Tolerance size. This is the difference between the two limits of size.

Grade of tolerance. The tolerance grade *is an indication of the degree of accuracy of manufacture.* It is designated by the letter IT followed by a number. Tolerance grades are ITO1, ITO2, upto IT16. The *larger the number the larger the tolerance.*

Standard tolerance unit. This is the unit used to calculate the various grades of tolerance for a given basic size.

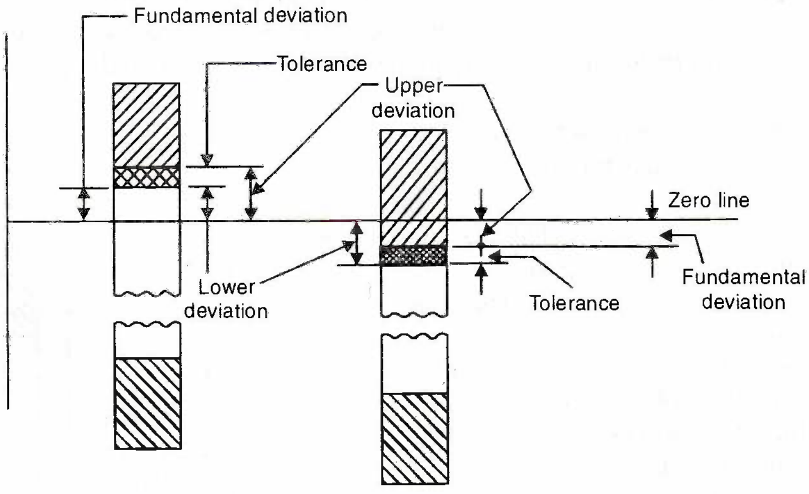


Fig. A.40

Upper deviation. Refer to Fig. A.40. This is the *amount from the basic zero or zero line, on the maximum limit of size for either a hole or a shaft.* It is designated ES for a *hole* and *es* for a *shaft*. Upper deviation is a *positive quantity* when the maximum limit of size is *greater* than the basic size and *negative quantity* when the maximum limit of size is less than the basic size.

Lower deviation. Refer to Fig. A.40. This is the *amount from basic size, or zero line,* to the *minimum limit of size.* It is designated EI for a *hole* and *ei* for a *shaft*. Lower deviation is a *positive quantity* when the minimum limits of *sizes is greater* than the basic size and a *negative quantity* when the minimum limit of size is *less* than the basic size.

Fundamental deviation. Refer to Fig. A.40. This is the deviation, either the upper or lower deviation, which is the *nearest one to the zero line* for either a hole or a shaft. It fixes the position of the tolerance zone in relation to zero line.

Fit. Fit *means a degree of tightness or looseness between two mating parts to perform a definite function.*

Clearance. The *difference between the sizes of a hole and a shaft* which are to be assembled together when the *shaft is smaller than the hole.*

Interference. The *difference between the sizes of a hole and a shaft* which are to be assembled together when the *shaft is larger than the hole.*

Classes of fit. Refer to Fig. A.41.

1. **Clearance fit:** A *clearance fit could be obtained by making the lower limit on the hole equal to or larger than the upper limit on the shaft.* Any hole and any shaft made to these tolerances would assemble with a clearance fit with certainty.
2. **Interference fit:** *An interference fit would be obtained with equal certainty by making the lower limit on the shaft equal to or larger than the upper limit on the hole.*

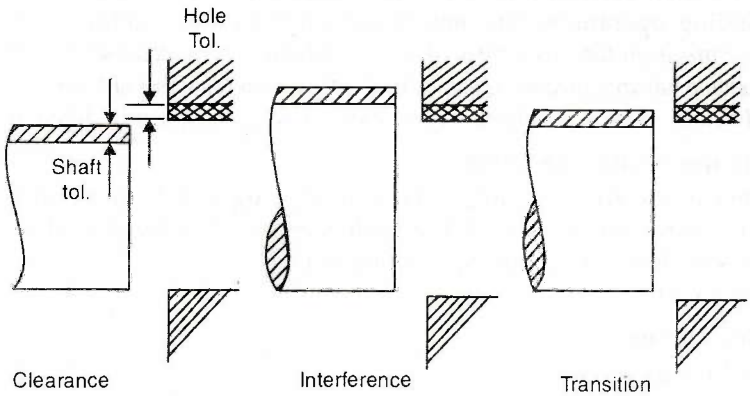


Fig. A.41. Classes of fit

3. **Transition fit:** Between these two conditions lies a *range of fits known as transition fit.* These are obtained when the upper limit on the shaft is larger than the lower limit on the hole, and the lower limit on the shaft is smaller than the upper limit on the hole. It must be realised that transition fits exist only as a class; any actual hole and shaft must assemble with either a clearance or interference fit.

Allowance: The *difference between the maximum shaft and minimum hole is known as allowance.* In a *clearance fit,* this is the *minimum clearance* and is a *positive allowance.* In an *interference fit,* it is the *maximum interference and is a negative allowance.*

Basis of fit (or limit) system:

A fit or limit system consists of a series of tolerances arranged to suit a specific range of sizes and functions, so that limits of size may be selected and given to mating components to ensure specific classes of fit. This system may be arranged on the following basis :

1. Hole basis system.
2. Shaft basis system.

Hole basis system: Refer to Fig. A.42. ***'Hole basis system'*** *is one in which the limits on the hole are kept constant and the variations necessary to obtain the classes of fit are arranged by varying those on the shaft.*

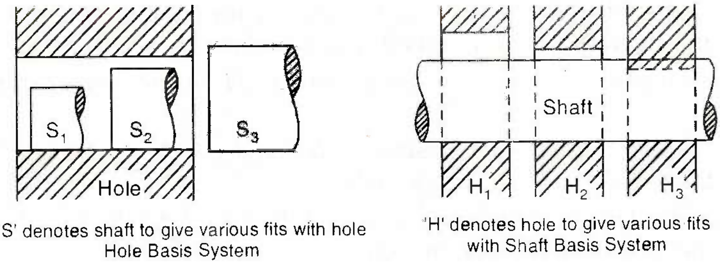


Fig. A.42. Hole and shaft basis systems.

Shaft basis system: Refer to Fig. A.42, ***'Shaft basis system'*** *is one in which the limits on the shafts are kept constant and the variations necessary to obtain the classes of fit are arranged by varying the limits on the holes.*

In present day industrial practice *hole basis system is used because a great many holes are produced by standard tooling*, for example, reamers drills, etc., whose size is not adjustable. Subsequently the shaft sizes are more readily variable about the basic size by means of turning or grinding operations. The *hole basis system results in considerable reduction in reamers and other precision tools as compared to a shaft basis system because in shaft basis system due to non-adjustable nature of reamers, drills etc. great variety (of sizes) of these tools are required for producing different classes of holes for one class of shaft for obtaining different fits.*

Systems of specifying tolerances :

The tolerance or the error permitted in manufacturing a particular dimension may be allowed to vary either on *one side* of the basic size or on *either side* of the basic size. Accordingly two systems of specifying tolerances exit.

Refer to Fig. A.43.

1. Unilateral system.
2. Bilateral system.

In the ***unilateral system***, tolerance is applied only in *one direction*.

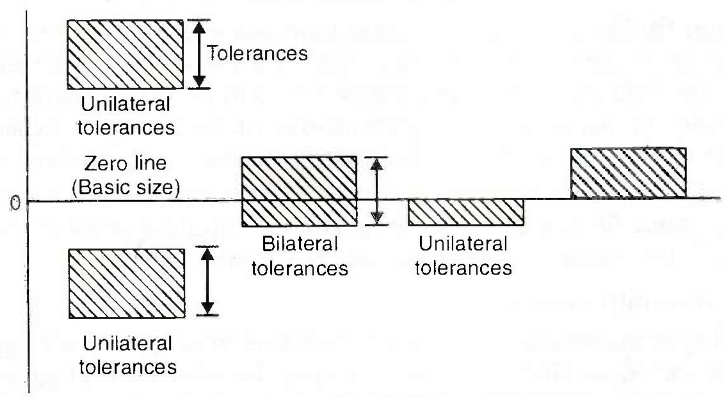


Fig. A.43. Unilateral and bilateral tolerances.

Example: $40.0^{+0.04}_{+0.02}$ or $40.0^{-0.02}_{-0.04}$

In the ***bilateral system*** of writing tolerances, a dimension is permitted to vary in *two directions*.

Example: $40.0^{+0.02}_{-0.04}$

- ***Unilateral system*** is more satisfactorily and realistically applied to certain machining processes where it is common knowledge that dimensions will most likely deviate in one direction. Further, in this system the tolerance can be revised without affecting the allowance or clearance conditions between mating parts, *i.e.*, without changing the type of fit. This system is *most commonly used in interchangeable manufacture especially where precision fits are required.*
- It is not possible, in ***bilateral system***, to retain the same fit when tolerance is varied. The basic size dimension of one or both of the mating parts will also have to be changed. This system clearly points out the theoretically desired size and indicates the possible and probable deviations that can be expected on each side of basic size. *Bilateral tolerances help in machine setting and are used in large scale manufacture.*

Designation of holes, shafts and fits :

A hole or shaft is completely described if the basic size, followed by the appropriate letter and by the number of the tolerance grade, is given.

Example :

- A 25 mm H-hole with the tolerance grade IT8 is given as :
  25 mm H8 or simply 25 H8.
- A 25 mm *f*-shaft with the tolerance grade IT7 is given as :
  25 mm *f*7 or simply 25 *f*7.

A *'fit'* is indicated by combining the designations for both the hole and shaft with the hole designation written first, regardless of system (*i.e.*, hole-basis or shaft-basis).

Example : 26 H8—*f*7 or

25 H8–*f*7 or

$25\frac{H8}{f7}$

Commonly used holes and shafts :

- In several engineering applications the fits required can be met by a quite small selection from the range available in the standards. The holes and shafts commonly used are as follows :

Holes (*commonly used*): H6, H7, H8, H9, H11.

Shafts (*commonly used*): *c*11; *d*10, *e*9, *f*7, *g*6, *h*6, *k*6, *n*6, *p*6, *s*6.

IS : 919 gives the most commonly used holes and shafts upto 500 mm for the purpose of general engineering.

The Newall system :

- The Newall system is the first standard evolved in Great Britain to standardise limits and fits and is still used to a certain extent although all the fits provided by this system can be obtained with approximately the same values by selection from 1916. This system provides a range of clearance, transition and interference fits for size upto 12". *It is a hole basis system*, which stipulates two grades of holes, specified with bilateral tolerances, together with 6 grades of shaft tolerances.
- This system is extremely simple and is earliest of all the systems. It specifies too few fits and those listed do not enforce to modern ideas as regards their basic deviations. Though this served a useful purpose in the past but is *not considered suitable for modern production.*

- Since it is based on hole basis system, therefore, *in this system provision is made in the size of the hole for error in workmanship, and the variation to obtain the quality of fit required is allowed for on the size of the shaft which has to enter the hole.*

ISO system of limits and fits :

ISO system has presently been universally adopted and as a matter of fact IS : 919 is almost in line with this system. While ISO specifies 28 classes of holes designated A, B, C, CD, E, EF, F, FG, G, H, I, J, JS, K, M, N, P, R, S, T, U, V, X, Y, Z, ZA, ZB, ZC and 18 grades of tolerance exactly matching with those of ISO systems. Similarly ISO has 28 classes of shafts while IS : 919 specifies only 25 classes of shaft. Other characteristics such as fundamental deviation and tolerance unit etc. are same in both the systems.

Types of fits :

Some important types of fits are discussed below :

1. **Selective fit.** A *selective fit* may be a *transition* or an *interference fit.* This type of fit is required where the object is to make a shaft and hole with a finite and not a permissible range on it. It is customarily used for tight or interference fits where it is desired to avoid the extremes of maximum tightness or looseness. The ideal selective fit for the tightest class of fit would stress the hole just to its elastic limit, thereby giving the maximum holding power without overstressing or distorting the grain structure.
2. **Push fit.** A *push fit* is a *transition fit.* It is also known as 'sung fit' and represents the closest fit that permits assembling parts by hand. With a push fit, there should be no perceptible play between the mating parts.
3. **Driving fit.** A *driving fit* is an *interference fit.* When a plug or shaft is made slightly larger than the hole into which it is to be inserted and the allowance is such that the parts can be assembled by driving, this is known as a driving fit. *Such fits are employed when the parts are to remain in a fixed position relative to each other.* Before assembling parts with a driving fit, the bearing surfaces should be oiled. A hydraulic press is usually preferable for assembling.
4. **Forced or pressed fit.** A forced or pressed fit is an *interference fit.* It is the term used when a pin, shaft or other cylindrical part is forced into a hole of slightly smaller diameter, ordinarily by the use of hydraulic press or some other type of press capable of exerting a considerable pressure. *A force fit has a larger allowance than a driving fit, and therefore requires greater pressure for assembling.* Forced fits are restricted to parts of *small* and *medium size, e.g., crankpins, car wheel axles,* and *similar parts* (which must be held very securely).
5. **Shrinkage fit.** A *shrinkage fit* is an *interference fit.* It is obtained by making the internal member slightly larger than the hole in the external member. In this type of fit, the pressure is not required for assembly but instead the external member is heated and expanded sufficiently to permit inserting the internal member easily. Then as the external part cools or is cooled by applying water or dry ice, it shrinks tightly around the internal part. In general practice, a smaller allowance for shrinkage fit is favoured.
6. **Freeze fit.** In a *freeze fit,* instead of heating the female member, the male member may be *contracted by cooling and subsequently allowed to expand into the female.* This process uses an industrial refrigerator giving a temperature of about –50°C, or tc obtain lower temperatures the component is cooled in liquid air (boiling point –190°C). Examples of this process are in the insertion of exhaust valve seats

inserts in engine cylinder heads or blocks, or in the insertion of brass bushes in various assemblies.

This process is convenient only for small parts as otherwise the size of the refrigerator equipment is prohibitive or consumption of liquid air is excessive. However, on suitable parts the process is very convenient, since the temperature is controlled and is unlikely to damage the structure of the material in any way. A combination of freezing and heating is also used as this enables reasonable maximum temperatures to be used. A *convenient method of reaching –50/–60°C without expensive equipment, and suitable for occasional use is to cool the part in alcohol to which solid frozen dioxide (known as 'dry ice' or 'dry cold') is added.*

Concept of interchangeability :

'Interchangeability' *refers to assembling a number of unit components taken at random from stock so as to build up a complete assembly without fitting or adjustment.*

A modern motor-car, for example, consists of many hundreds of separate components each of which is manufactured in large numbers. For complete interchangeability it should be possible simply to collect at random the constituent parts then to assemble the whole without the use of any cutting tools and for the assembly to function satisfactorily.

The contacts between the various parts constitute what are termed *fits.*

For correct functioning of parts the fits must be good within certain limits of accuracy. It would be possible so to choose these limits as to ensure absolute interchangeability, and this should always be done in the case of less-important fits such as bolts fitting in bolt-holes and so on. Experience shows, however, that to do this in the case where the tolerance on the fit is very small may call for such fine limits that the cost is excessive.

In cases like this a process known as *selective assembly* is used. Thus if we have a shaft required to run in a close fitting bearing we can arrange, during inspection, to sort the shafts into, say, three grades, those near the upper limit, those near the middle and those nearly at the bottom limit. The same selection is made with the bearings. By arranging to mate the top-limit shafts with the top-limit holes, for example, we shall ensure a much better assembly than if the parts were chosen at random. We can, in fact, increase the limits on the components and thereby very much reduce the cost of production.

One of the objects of interchangeability is to make it possible to replace a work part, such as a complete ball bearing, without making any adjustment to the old or new parts. Here, of course, selective assembly is difficult except by actual manufacturers of the components, and in such cases it is necessary that absolute interchangeability should be possible.

A.7.3. Classification of Measuring Equipment

The measuring equipment can be *classified* as follows :

1. Measuring instruments.
2. Limit gauges.
3. Measuring devices.
4. Measuring machines.

1. Measuring instruments. These are the instruments by means of which a *direct reading of a dimensions or property can be taken without use of any extra attachment.*

Examples :

- Vernier calipers
- Micrometers
- Slip gauges
- Dial gauges
- Speedometers
- Thermometers

- Voltmeters
- Ohm meters
- Ampere meters.

2. Limit gauges. These are gauges by means of which *a certain dimension or a certain form can be checked for which the gauges are designed or adjusted.*

Examples :

- Go and Not-go gauges.
- Thread gauges.
- Taper gauges.

3. Measuring devices. These are the means of measurement by which the *measured value is indicated on a measuring head.* Sometimes it is possible to register the readings on a recording device. These are measuring devices with installed standards and other without installed standards. The latter type can only be used for comparison.

The measuring devices can be *classified* according to the type of the measuring head used.

(*i*) **Mechanical measuring devices.** In these devices the magnification of the reading is done by pure mechanical means such as levers and gears.

Examples : *Dial indicators, Passimeter, Mikrokator and Grapho test.*

(*ii*) **Optical measuring devices.** In these devices optical means are used in measuring process. This can be an optical enlargement just for reading the standards (Loope Microscope or Projector) or it can be an optical magnification of the measured value. Usually a combination of these are used.

(*iii*) **Electrical measuring devices.** These are devices in which electrical energy is used in the measuring process. The measurement is generally done by mechanical means (tracer, stylus, or plunger) and the movement is then converted to electric current, voltage or impulse. These can be amplified, magnified and then indicated or recorded or converted to signals or movements for the purpose of control. These electrical measuring devices compared with *mechanical and the optical devices have the advantage that the results of the measurements can be indicated or registered in a place far away from that where measurement is carried out.* This *facilitates the control of the testing and the production machines.*

Examples : *Perthometer tester* and *Electrical comparator.*

(*iv*) **Pneumatic measuring devices.** In these devices *pneumatic means are used in the measuring process.*

Example : *Solex pneumatic comparator.*

Combination of two or more of the above mentioned principles is also possible.

- *Other measuring devices in which X-rays or γ-rays* (radio-active elements) or *Ultrasonic waves are used.*

Example : *Exatest.*

4. Measuring machines. The measuring machines are employed for universal use in the field of metallurgy. They have a compact construction (column or bed) and contain their own standards of measurement in the form of scales, micrometers or other instruments.

Examples :

- Matrix machine.
- Universal measuring microscope.
- Zeiss universal length measuring machine.

A.7.4. Surface Finish

Introduction:

Each of the methods employed for producing a surface imparts some characteristic quality to the structure of the surface layer of metal and it will be readily appreciated that a chip-producing information method will have a different effect from such treatments as planishing, rolling or drawing where the surface layer is caused to slide over the material beneath. Certain qualities in the surface layer are important for various classes of service; for example the conditions imposed on the races of a ball or roller bearing where heavy loads must be carried by very small areas, whilst under the action of rolling friction demand that the surfaces must possess a high degree of homogeneity, elasticity and hardness. The finish and texture of lifting surfaces have an important effect on bearing friction, rate of wear, initial toterances etc. Every surface, more or less, is composed of minute hills and valleys and when the conditions of service require that such surfaces shall operate one against the other the early stages of action will result in the levelling down of the "hills" on each member. If the combination has a bearing this initial wear will result in an increased clearance, so that the eventual conditions, after the bearing has "settled down", will depend on both the initial fit and quality of the finish. The reliability of a press or force fit and its approach to theoretical conditions is largely dependent on surface quality and finish. Even for surfaces which do not fit or serve as bearing smoothness is often important since in highly stressed parts fatigue cracks originate from surface blemishes and under corrosive influence a high class surface may be more durable than one less finished.

Surface texture :

The surface texture may be defined as :

"The characteristic quality of an actual surface due to small departures from its general geometrical form which, occurring at regular or irregular intervals, tend to form a pattern or texture on the surface".

Surface textures vary according to the machining processes used in producing it. This is certainly true in case of metal machining. These differences are apparent by visual examination and can be felt readily by passing a finger nail over the surface. The texture of surfaces may be regular or irregular in character and may lie in a particular direction or be non-directional in character. Additional factors producing surface irregularities may include faulty tools, inherent imperfections in the machine tools used and, of course, errors due to the personal element. The problem of the measurement of surface texture is complex, owing to the number of possible variables. Therefore, in practice, for the normal control of manufactured components, a complete analysis is not possible, or even desirable. The difficulty is overcome mainly by specifying the finishing process (such as grinding, lapping, etc.) to be used and by stating, in standard units, the quality of finish required.

- The *problem of the measurement of surface texture is basically geometrical.* Although, fundamentally, the problem is three dimensional in character, in practice it is conveniently reduced to one of two-dimensional geometry. This is accomplished by limiting individual measurements to the examination of profiles of plane sections taken through the surface being measured. It is important to make the measurement in the *"correct"* plane, which is usually in a plane approximately at right angles to the "lay" (or direction of the predominant markings) of the surface and, generally, it furnishes the most efficient results. However, other planes may have to be developed in special cases.

Primary and secondary texture :

Any material being machined by chip removal process can't be finished perfectly due to some departures from ideal conditions as envisaged by the designer. Due to conditions not being ideal, the surface produced will have some irregularities; and these geometrical irregularities could be classified into the following four categories :

First order. *Irregularities arising out of inaccuracies in the machine tool itself* (*e.g.* lack of straightness of guideways on which tool post is moving).

Surface irregularities arising due to deformation of work under the action of cutting forces and the height of the material itself are also induced under this head.

Second order. *Irregularities caused due to vibrations of any kind such as chatter marks.*

Third order. *Irregularities caused by a machining itself due to characteristics of the process.* This also includes the feed marks of the cutting tool.

Fourth order. Irregularities arising from the rupture of the material during the separation of the chip.

These irregularities of four orders can be grouped into the following *two* groups :

1. Primary texture (or Roughness)
2. Secondary texture.

Primary texture (or roughness). In this group are included *irregularities of small wavelength caused by direct action of the cutting element on the material or by some other disturbance such as friction, wear, or corrosion.* These errors are chiefly caused due to *tool chatter i.e.*, it includes irregularities of third and fourth order and constitutes the *microgeometrical errors.*

Secondary texture. In this group are included *irregularities of considerable wavelength of a periodic character resulting from mechanical disturbances in the generating set up.* These errors are termed as *macrogeometrical errors* and include irregularities of first and second order and are mainly due to *misalignment of centres, lack of straightness of guideways and non-linear feed motion.*

Any surface could be considered to be combination of two forms of wavelengths (*large wavelength for waviness and smaller wavelength for roughness*) superimposed upon each other.

One of the problems in measuring surfaces finish is to *separate the waviness from the roughness.*

Fig. A.44 shows the various terms used in connection with surface finish.

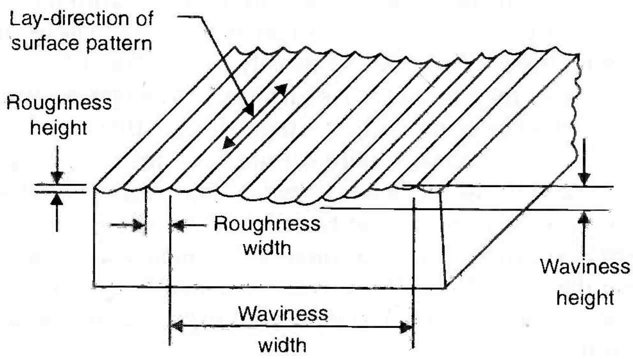


Fig. A.44

Methods of measuring surface finish :

The surface finish of machined part can be measured by the following *two* methods:

1. Surface inspection by comparison methods.
2. Direct instrument measurements.

Surface inspection by comparison methods :

In comparative methods, the surface texture is assessed by observation of the surface. But these methods are not reliable as they can be misleading if comparison is not made with surface produced by same techniques. The various methods available are :

1. Touch inspection.
2. Visual inspection.
3. Scratch inspection.
4. Microscopic inspection.
5. Surface photographs.
6. Micro-interferometer.
7. Wallace surface dynamometer.
8. Reflected light intensity.
9. Comparison with standard specimens.

Indication of surface roughness symbols used :

- The basic symbol, consists of two legs of unequal length inclined at approximately 60° to the line representing the surface under consideration as shown in Fig. A.45.

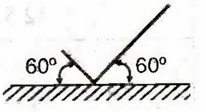


Fig. A.45

- If the removal of material by machining is required, a bar is added to the basic symbol, as shown in Fig. A.46.
- If the removal of material is not permitted, a circle is added to the basic symbol, as shown in Fig. A.47.

  The symbol in Fig. A.47 may also be used in a drawing relating to a production process to indicate that a surface is to be left in the state relating from a preceding manufacturing process, whether this state was achieved by removal of material or otherwise.

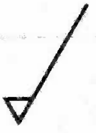


Fig. A.46

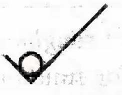


Fig. A.47

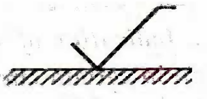


Fig. A.48

- When special surface characteristics have to be indicated a line is added to the longer leg of any of the above symbols, as shown in Fig. A.48.

I. *Indication of surface roughness :*

(*a*) The value or values defining the principle criterion of roughness are added to the symbols as shown in Fig. A.49.

(*b*) *Surface roughness specified :*

- As in Fig. A.49(*i*), may be obtained by any production method.
- As in Fig. A.49(*ii*), shall be obtained by removal of material by machining.
- As in Fig. A.49(*iii*), shall be obtained without removal of material.

(*c*) When only one value is specified it represents the maximum permissible value of surface roughness.

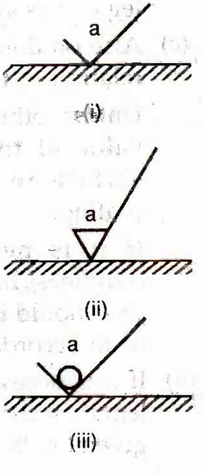


Fig. A.49

(*d*) If it is necessary to impose maximum and minimum limits of the principal criterion of surface roughness, both values should be shown as in Fig. A.50, with the maximum limit a_1 above the minimum limit a_2.

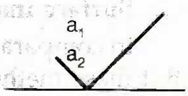


Fig. A.50

(*e*) The principal criterion of roughness R_a may be indicated by the corresponding roughness grade symbol as shown below :

Roughness Value Ra, mm	*Roughness Grade Number*	*Roughness Symbol*
50	N12	~
25 12.5	N11 N10	∇
6.3 3.2 1.6	N9 N8 N7	∇∇
0.8 0.4 0.2	N6 N5 N4	∇∇∇
0.1 0.05 0.025	N3 N2 N1	∇∇∇∇

II. *Indication of special surface roughness characteristics :*

(*a*) In certain circumstances, for fundamental reasons, it may be necessary to specify additional special requirements concerning surface roughness.

(*b*) If it is required that the final surface roughness be produced by one particular production method, this method should be indicated in plain language on an extension of the longer leg of this symbol given in Fig. A.48, as shown in Fig. A.51.

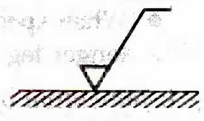


Fig. A.51

(*c*) Also on this extension line should be given any indications relating to treatment or coating.

Unless otherwise stated, the numerical value of the roughness applies to the surface roughness after treatment of coating.

If it is necessary to define surface roughness both before and after treatment, this should be explained in a suitable note or in accordance with Fig. A.52.

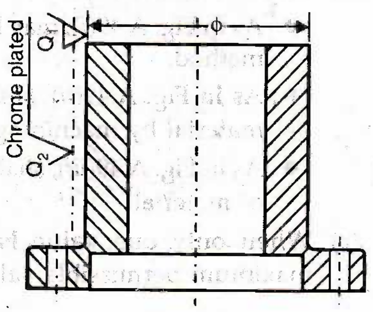


Fig. A.52

(*d*) If it is necessary to indicate the sampling length, it should be selected from the series given in IS : 3073–1967 'Assessment of surface roughness, and then stated adjacent to the symbol, as shown in Fig. A.53.

(e) If it is necessary to control the direction of lay, it is specified by a symbol (see Table A.1) added to the surface roughness symbol as shown in Fig. A.54.

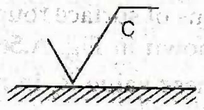


Fig. A.53

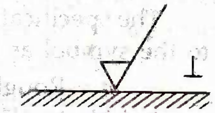


Fig. A.54

III. *Symbols for the direction of lay :*

The series of symbols for the common directions of lay are specified in Table A.1.

IV. *Indication of machining allowance :*

Where it is necessary to specify the value of the machining allowance, this should be indicated on the left of the symbol as shown in Fig. A.55. This value should be expressed in millimetres, according to the general system used for dimensioning the drawing.

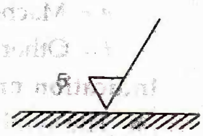
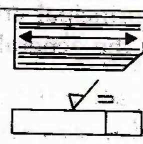
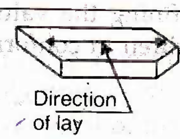
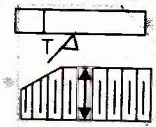
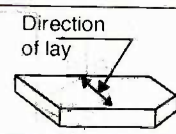
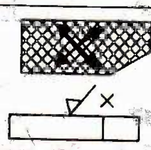
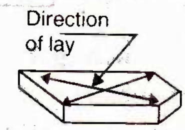
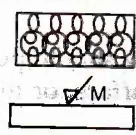
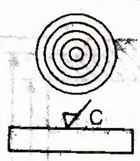
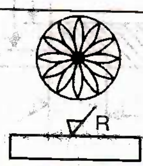


Fig. A.55

Table A.1. Symbols for direction of lay

Symbol	*Interpretation*		
=	Parallel to the plane of projection of the view in which the symbol is used.		
⊥	Perpendicular to the plane of projection of the view in which the symbol is used.		
X	Crossed in two slant directions relative to the plane of projection of the view in which the symbol is used.		
M	Multi-directional.		—
C	Approximately circular relative to the centre of the surface to which the symbol is applied.		—
R	Approximately radial relative to the centre of the surface to which the symbol is applied.		—

Note: It should be necessary to specify a direction of lay not clearly defined by these symbols, then this shall be achieved by a suitable note on the drawing.

V. Position of the specification of surface roughness in the symbol :

The specifications of surface roughness should be placed relative to the symbol as shown in Fig. A.56.

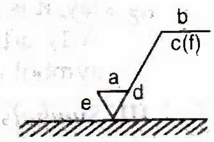


Fig. A.56

a = Roughness value R_a in millimetres or Roughness grade symbol N1 to N12;

b = Production method, treatment or coating;

c = Sampling length;

d = Direction of lay;

e = Machining allowance;

f = Other roughness values (in brackets).

Indication on drawings :

- The symbol, as well as the inscriptions, should be oriented such that they may by read from the bottom or the right-hand side of the drawing (Fig. A.57).

If it is not practicable to adopt this general rule, the symbol may be drawn in any position, but only provided that it does not carry any indication of special surface roughness characteristics or of machining allowances. Nevertheless, in such cases the inscription defining the value of the principal criterion of roughness (if present) shall always be written in conformity with the general rule (Fig. A.58).

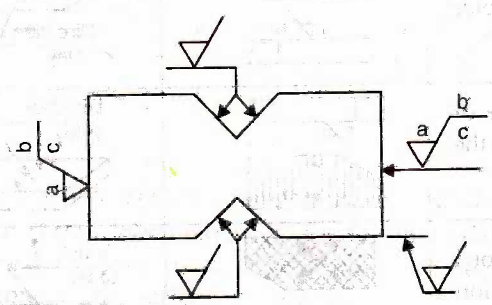


Fig. A.57

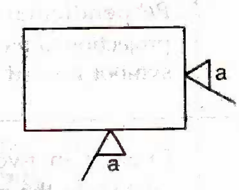


Fig. A.58

If necessary, the symbol may be connected to the surface by a leader line terminating in an arrow.

The symbol or the arrow should point from outside the material of the part, either to the line representing the surface, or to an extension of it (Fig. A.57).

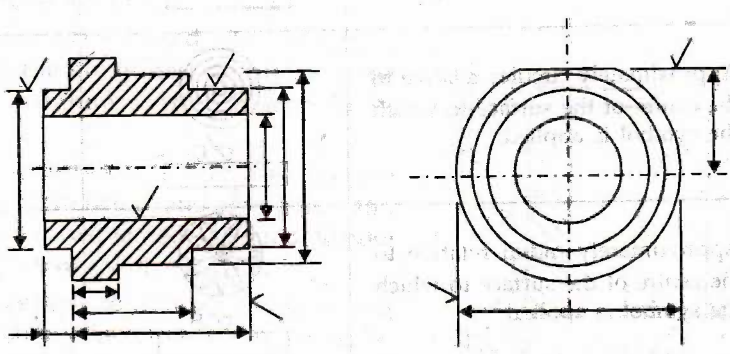


Fig. A.59

- In accordance with the general principles of dimensioning, the symbol is only used once for a given surface and, if possible, on the view which carries the dimension defining the size or position of the surface (Fig. A.59).
- If the same roughness is required on all the surfaces of a part, it is specified:
  (*a*) Either by a note near view of the part (Fig. A.60), near the title block or in the space devoted to general notes; or
  (*b*) Following the part number on the drawing (Fig. A.61)

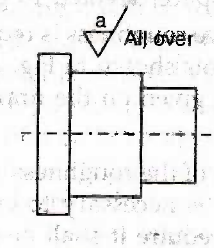


Fig. A.60

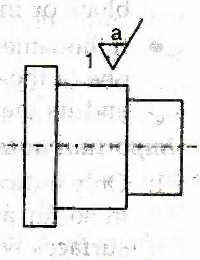


Fig. A.61

- If the same surface roughness is required on the majority of the surfaces of a part, it is specified with the addition of
  (*a*) the notation except where otherwise stated (Fig. A.62).
  (*b*) a basic symbol (in brackets) without any other indication (Fig. A.63), or

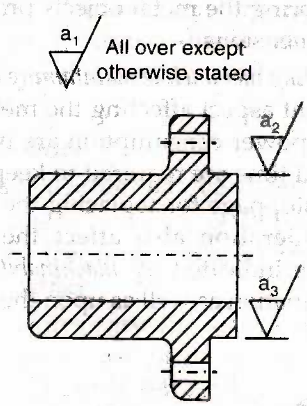


Fig. A.62

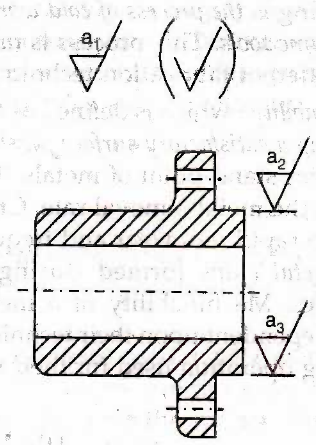


Fig. A.63

(*c*) the symbol or symbols (in brackets) of the special surface roughness or roughness (Fig. A.64). The symbols for the surface roughness which are exceptions to the general symbol are indicated on the corresponding surfaces.

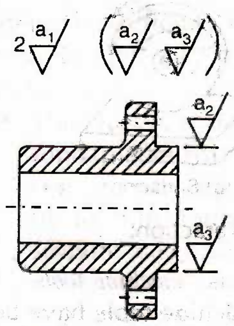


Fig. A.64.

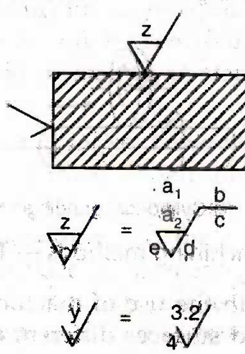


Fig. A.65.

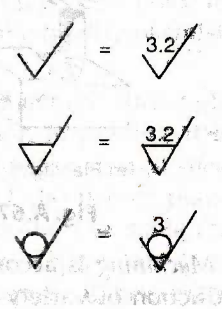


Fig. A.66.

- To avoid the necessity of repeating a complicated specification a number of times, or where space is limited a simplified specification may be used on the surface, provided that its meaning is explained near the drawing of the part, near the title block or in the space devoted to general notes (Fig. A.65).
- If the same surface roughness is required on a large number of surfaces of the part, one of the symbols shown in Fig. A.47 may be used on the appropriate surfaces and its meaning given on the drawing, for example, as shown in Fig. A.66.

Important notes :

1. Only indications of the roughness, method of production or machining allowance in so far as this is necessary to ensure fitness for purposes and only for those surfaces which require it shall be given.
2. The specification of surface's roughness is unnecessary whenever the ordinary manufacturing processes by themselves ensure an acceptable surface finish.

A.8 MACHINING PROCESSES

A.8.1. Machining

Machining *is the process of cold working the metals into different shapes by using different types of machine tools.* This process is mainly used to bring the metal objects produced by means of different fabrication techniques to final dimensions.

"Machinability" which is defined as the *ease of removing metal while maintaining dimensions and developing a satisfactory surface finish* is an important aspect affecting the metallurgical and properties stand-point of metals. Tool wear and power consumption are two factors which affect the metal removal rate. Greater effort and time are required to keep the tools sharp due to rapid tool wear and frequent machine stoppage for replacing the full tools. Types of *metal chips* formed during machining operation also affect the different characteristics. Machinability of a metal is generally indicated by *machinability ratings* (which are dependent upon their techniques of determination as well as upon the particular metal cutting operation used for their measurement).

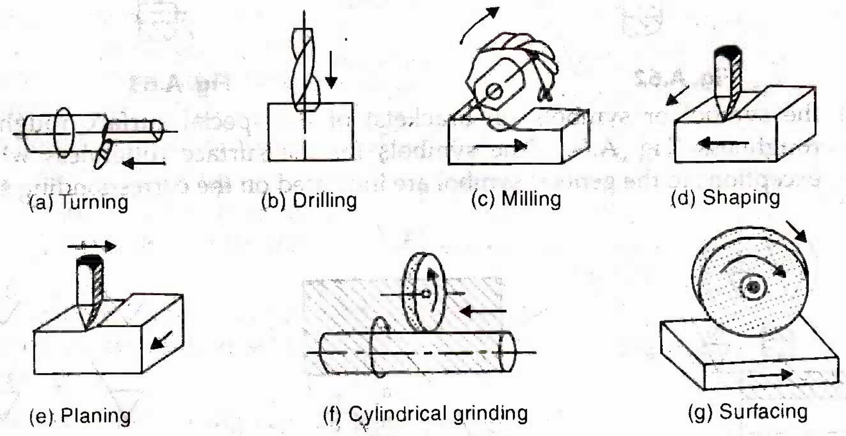


Fig. A.67. Principal machining methods — Tool work interaction.

Machining is accomplished with the use of machines known as *"machine tools"*. For production of variety of machined surfaces different types of machine tools have been developed. *The kind of surface produced depends upon the shape of cutting, the path of the tool* *as it passes through the material or both.* Depending on them metal cutting processes are called either turning or planing or boring or other operations performed by machine tools like lathe, shaper, planer, drill, miller, grinder, etc., as illustrated schematically in Fig. A.67.

A.8.2. Classification of Machining Processes

Machining processes *are material removing operations in which the desired shape, size and surface finish on the finished product are obtained by removing surplus material.*

The machining processes are *classified* as follows:

1. *Metal Cutting:*
   (*i*) *Single point cutting :*
   - Turning
   - Boring
   - Shaping
   - Planing.

   (*ii*) *Multi-point cutting :*
   - Milling
   - Drilling
   - Tapping
   - Hobbing
   - Broaching.
2. *Grinding and finishing :*
   (*i*) *Grinding :*
   - Surface grinding
   - Cylindrical grinding
   - Centreless grinding.

   (*ii*) *Finishing :*
   - Lapping
   - Honing
   - Superfinishing.
3. *Unconventional Machining :*
   - Ultrasonic machining
   - Electrodischarge machining
   - Electro-chemical machining
   - Laser beam machining.

— The **metal cutting** (machining, a generic term, refers to all material removal processes) *refers to only those processes where material removal is affected by the relative motion between tool made of harder material and the workpiece.* The tool would be single-point cutting tool as used in operations like turning or shaping, or a multi-point tool as used in milling or drilling operation.

— **Grinding and finishing** processes are those where metal is *removed by a large number of hard abrasive particles or grains which may be bonded as in grinding wheels, or be in loose form as in lapping.*

— **Unconventional machining processes** are those *which use electrical, chemical and other means of material removal for shaping high strength materials and for producing complicated shapes.*

A.8.3. Cutting Tools

General characteristics of a metal cutting tool :

A typical cutting tool in simplified form is shown in Fig. A.68; in this figure are shown the general characteristics of a metal cutting tool.

(*i*) **Rake angle.** *It is the angle between the face of the tool called the rake face and normal to the machining direction.* This angle specifies the ease with which a metal is cut. *Higher the rake angle, better is the cutting and less are the cutting forces.* There is a maximum limit to the rake angle and this is generally of the order of 15° for high speed steel tools cutting mild steel (increase in the rake angle reduces the strength of the tool tip as well as the heat dissipation).

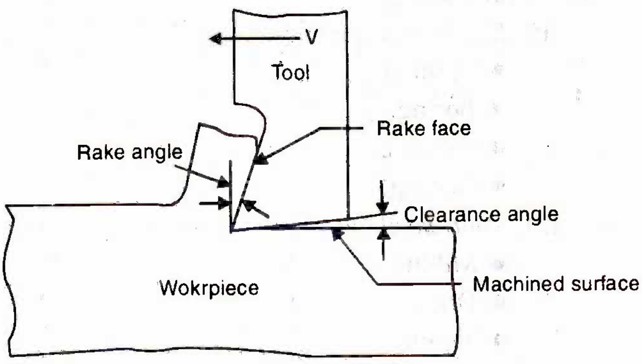


Fig. A.68. General characteristics of a metal cutting tool.

It is possible to have rake angle as *zero* or *negative.* These are generally used in the case of highly brittle tool materials such as carbides or diamonds for giving extra strength to the tool tip.

(*ii*) **Clearance angle.** *This is the angle between the machined surface and underside of the tool called the flank face.* The clearance angle is provided such that the tool will not rub the machined surface thus spoiling the surface and increasing the cutting forces. A very large clearance angle reduces the strength of the tool tip, and hence normally an angle of the order of 5–6° is used.

- The *conditions* which have an important influence on metal cutting are: (*i*) *Work material,* (*ii*) *cutting tool material,* (*iii*) *cutting tool geometry,* (*iv*) *cutting speed,* (*v*) *feed rate,* and (*vi*) *depth of cut and cutting fluid used.*
- The *cutting speed* (V) is the speed with which the tool moves through the work material. This is generally expressed in metres per second (m/s).
- *Feed rate* (f) may be defined as the small relative movement per cycle (per revolution or per stroke) of the cutting tool in a direction usually normal to the cutting speed direction.
- *Depth, of cut* (*d*), is the normal distance between the unmachined surface and the machined surface.

Classification of cutting tools :

Cutting tools are *classified* as follows :

1. Single point cutting tools.
2. Multi-point cutting tools.
   (*i*) Solid tool

(*ii*) Brazed tool

(*iii*) Inserted bit tool.

The various angles of a single point tool are shown in Fig. A.69.

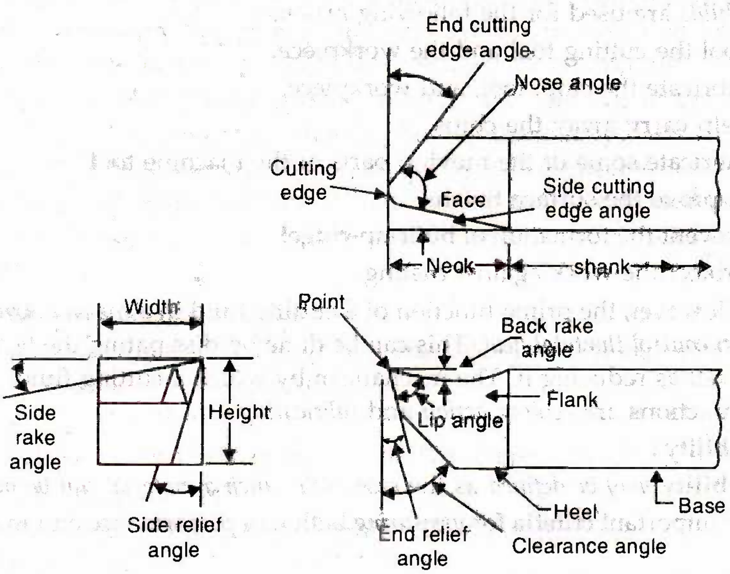


Fig. A.69. Various angles of a single point tool.

Characteristics of an ideal cutting-tool material :

An ideal cutting-tool must possess the following *characteristics :*

1. The material must *remain harder than work material at elevated temperature* (*Hot hardness*)
2. The material must *withstand excessive wear* even through the relative hardness of the tool-work materials changes. (*Wear resistance*)
3. The material must have sufficient *strength and ductility* to withstand shocks and vibrations and to prevent breakage. (*Toughness*)
4. The *coefficient of friction* at the chip tool interface must remain low for minimum wear and reasonable surface finish.
5. The *cost and easeness of fabrication* should be within reasonable limits.

Types of tools materials :

- While selecting proper tool material the *type of service* to which the tool will be subjected should be given primary consideration. No one material is superior in all respects, but rather each has certain characteristics which limits its field of application.
- The *principal carbon tool materials* are :
  1. Carbon steels.
  2. Medium alloy steels.
  3. High speed steels.
  4. Stellilies.
  5. Cemented carbides.
  6. Ceramics.

- ***Unilateral system*** is more satisfactorily and realistically applied to certain machining processes where it is common knowledge that dimensions will most likely deviate in one direction. Further, in this system the tolerance can be revised without affecting the allowance or clearance conditions between mating parts, *i.e.*, without changing the type of fit. This system is *most commonly used in interchangeable manufacture especially where precision fits are required.*
- It is not possible, in ***bilateral system***, to retain the same fit when tolerance is varied. The basic size dimension of one or both of the mating parts will also have to be changed. This system clearly points out the theoretically desired size and indicates the possible and probable deviations that can be expected on each side of basic size. *Bilateral tolerances help in machine setting and are used in large scale manufacture.*

Designation of holes, shafts and fits :

A *hole or shaft is completely described if the basic size, followed by the appropriate letter and by the number of the tolerance grade, is given.*

Example :

- A 25 mm H-hole with the tolerance grade IT8 is given as : 25 mm H8 or simply 25 H8.
- A 25 mm *f*-shaft with the tolerance grade IT7 is given as : 25 mm *f* 7 or simply 25 *f* 7.

A *'fit'* is indicated by combining the designations for both the hole and shaft with the hole designation written first, regardless of system (*i.e.*, hole-basis or shaft-basis).

Example : 26 H8—*f*7 or

25 H8-*f*7 or

$$25\frac{H8}{f7}$$

Commonly used holes and shafts :

- In several engineering applications the fits required can be met by a quite small selection from the range available in the standards. The holes and shafts commonly used are as follows :

Holes (*commonly used*): H6, H7, H8, H9, H11.

Shafts (*commonly used*): c11, *d*10, *e*9, *f*7, *g*6, *h*6, *k*6, *n*6, *p*6, *s*6.

IS : 919 gives the most commonly used holes and shafts upto 500 mm for the purpose of general engineering.

The Newall system :

- The Newall system is the first standard evolved in Great Britain to standardise limits and fits and is still used to a certain extent although all the fits provided by this system can be obtained with approximately the same values by selection from 1916. This system provides a range of clearance, transition and interference fits for size upto 12". *It is a hole basis system*, which stipulates two grades of holes, specified with bilateral tolerances, together with 6 grades of shaft tolerances.
- This system is extremely simple and is earliest of all the systems. It specifies too few fits and those listed do not enforce to modern ideas as regards their basic deviations. Though this served a useful purpose in the past but is *not considered suitable for modern production.*

Tool life of a cutting tool may be calculated by using the following relation :

$$VT^n = C$$

where, V = Cutting speed in m/min.,

T = Tool life in min.,

C = A constant (which is numerically equal to cutting speed that gives the tool life of one min.), and

n = Another constant (depending upon finish, workpiece material and tool material)

= 0.1 for H.S.S. steel tools; 0.2 to 0.25 for carbide tools and 0.4 to 0.55 for ceramic tools.

The following are some of the ***possible tool failure criteria*** *that could be used for limiting tool life :*

Based on tool wear :

(*i*) Wear land size.

(*ii*) Crater depth, width or other parameters.

(*iii*) A combination of the above two.

(*iv*) Chipping or fine cracks developing at the cutting edge.

(*v*) Volume or weight of materials worn off the tool.

(*vi*) Total destruction of the tool.

Based on consequences of worn tool :

(*i*) Limiting value of change in component size.

(*ii*) Limiting value of surface finish.

(*iii*) Fixed increase in cutting force or power required to perform a cut.

A.8.4. Orthogonal and Oblique Cutting

Orthogonal cutting: Refer to Fig. A.70.

- When the tool is pushed into the workpiece, a layer of material is removed from the workpiece and it slides over the front face of the tool called rake face. *When the cutting edge of wedge is perpendicular to the cutting velocity, the process is called* ***orthogonal cutting.***

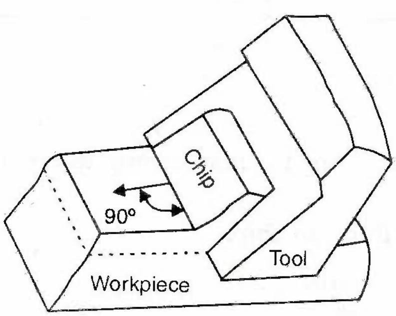


Fig. A.70. Orthogonal cutting.

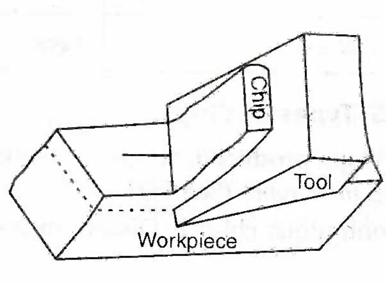


Fig. A.71. Oblique cutting.

- In this case, the material gets deformed under plane strain conditions; the *chip slides directly up the tool face.*

Oblique cutting: Refer to Fig. A.71.

- In most practical metal-cutting processes, the cutting edge of the tool is not perpendicular to the cutting velocity but set at angle with the normal to the cutting velocity.

- *Cutting in this case takes place in three-dimensions (turning or milling) and represents the general case of oblique cutting.*
- In oblique cutting a *lateral direction of chip movement is obtained.*

Comparison between 'Orthogonal cutting' and 'Oblique cutting'

S.No.	Aspects	Orthogonal cutting	Oblique cutting
1.	*Inclination of the cutting edge of the tool.*	Perpendicular to the direction of tool travel.	Inclined at an angle with the normal to the direction of tool travel.
2.	*Clearance of the work-piece width by the cutting edge.*	The cutting edge clears the width of the work-piece on either ends.	The cutting edge may or may not clear the width of the workpiece.
3.	*The chip movement*	The chip flows over the tool face and direction of chip flow velocity is normal to the cutting edge. The chip coils in a tight flat spiral.	The chip flows on the tool face making an angle with the normal on the cutting edge. The chip flows side-ways in a long curl.
4.	*Number of components of cutting force acting on the tool.*	Only two components of the cutting force act on the tool. These two components are perpendicular to each other and can be represented in a plane.	Three components of the forces (mutually perpendicular) act at the cutting edge.
5.	*Maximum chip thickness occurrence.*	Maximum chip thickness occurs at its middle.	The maximum chip thickness may not occur at middle.
6.	*Tool Life.*	Less	More.

A.8.5. Types of Chips

The chips produced, whatever the cutting conditions be, may belong to one of the following *three types* (See Fig. A.72).

1. Continuous chip; 2. Discontinuous chip; 3. Built-up chip.

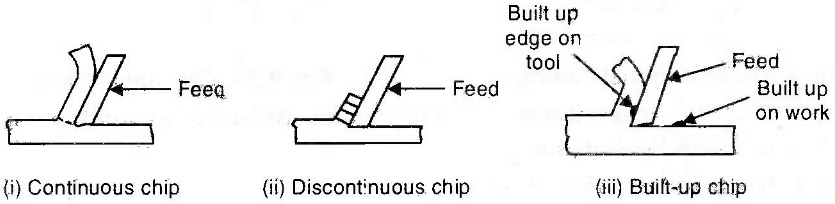


Fig. A.72. Types of chips.

1. Continuous chip: Refer to Fig. A.72(*i*).

- These chips are produced while machining *more ductile materials.* This type of chip is *most desirable.*
- The continuous chip which is like a ribbon flows along the rake face. Production of continuous chips is possible because of ductility of metal.
- Some ideal conditions that promote continuous chips in metal cutting are :
  — Small chip thickness (fine feed);
  — Small cutting edge;
  — Large rake angle;
  — High cutting speed;
  — Less friction between the chip-tool interface through efficient lubrication.
  — Ductile work materials.
- These chips are *most useful chips* since the *surface finish obtained is good and the cutting is smooth.* It also helps in *having higher tool life and lower power consumption.*

However, because of the large coils of chips, *chip disposal is a problem.* For this purpose various forms of *chip breakers* have been developed which are in the form of a step or groove in the tool rake face. The chip breakers allow the chips to be broken into small pieces so that they can be easily disposed off.

2. Discontinuous chip : Refer to Fig. A.72(*ii*).

- These chips are usually produced while cutting *more brittle materials* like grey cast-iron, bronze and hard brass.
- In this type the chip produced is in the form of *discontinuous segments* (deformed material instead of flowing continuously) gets ruptured periodically.
- Discontinuous chips are easier from the view point of chip disposal. However, the cutting force becomes unstable with the variation coinciding with the fracturing cycle. Also they generally provide better surface finish. However, in case of ductile materials they cause poor surface finish and low tool life.
- Discontinuous chips are likely to be produced under the following conditions :
  — Low cutting speeds;
  — Small rake angles;
  — Higher depths of cut (large chip thickness).

3. Built-up chip : Refer Fig. A.72(*iii*).

When machining ductile materials, conditions of high local temperature and extreme pressure in the cutting zone and also high friction in the tool-chip interface may cause the work material to *adhere or weld to the cutting edge of the tool forming the built-up edge (BUE).* This *causes the finished surface to be rough.* However, since the cutting is being carried by the BUE and not the actual tool tip, the *life of the cutting tool increases* while cutting with BUE. That way BUE is not harmful while rough machining.

A.8.6. Forces of a Single-point Tool

Refer to Fig. A.73.

Orthogonal cutting : Resultant, $R = \sqrt{F_x^2 + F_z^2}$

Oblique cutting : Resultant, $R = \sqrt{F_x^2 + F_y^2 + F_z^2}$

Torque to be developed on the workpiece,

$$T = \frac{F_z \times D}{2 \times 1000} \text{ Nm (neglecting the components } F_x \text{ and } F_y)$$

(where, D = diameter of the workpiece in mm)

$$\textit{Heat produced} \text{ (= work done in cutting metal)} = \frac{F_z \times v}{60 \times 1000} \text{ kN m/s or kJ/s or kW}$$

where, v = Cutting speed in m/min.

$$\text{Power required} = \frac{F_z \times v}{60 \times 1000 \times \eta} \text{ kW}$$

where, η = efficiency of the machine)

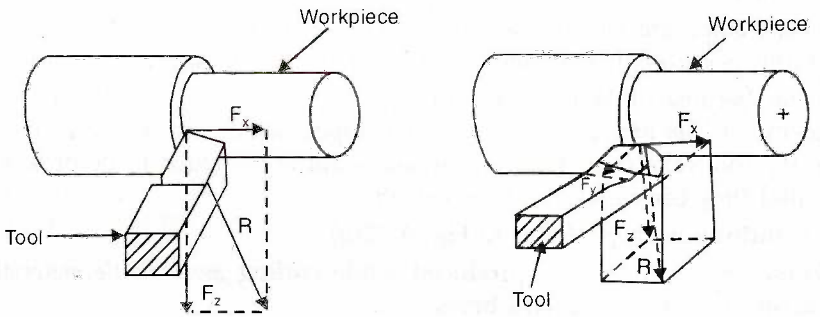


Fig. A.73. Forces on a cutting tool.

The approximate values of efficiencies of the different machines when working at full loads are :

1. Lathes .... 80 to 90%
2. Drilling machines .... 85 to 90%
3. Milling machines .... 80 to 90%
4. Shapers and planers .... 65 to 75%
5. Grinding machines .... 80 to 85%.

A.8.7. Machine Tools

Machine tools are used for machining. They employ cutting tools to remove excess material from the given job. The machine tools can be *classified* as follows :

1. **General purpose :**
   (*i*) Lathe. (*ii*) Drilling machine.
   (*iii*) Shaping machine. (*iv*) Planing machine.
   (*v*) Milling machine. (*vi*) Sawing machine.
2. **Special purpose :**
   (*i*) Special lathes like capstan, turret and copying lathes.
   (*ii*) Boring machine. (*iii*) Broaching machine.
   (*iv*) Production milling machine. (*v*) Production drilling machine.
3. **Automatic machine tools:** These machine tools, also called Automatic screw cutting machines (or simply auto-mats), are used for mass production of essentially small parts using a set of pre-designed and job-specific cams.

4. **Computer Numerical Control (CNC) machine tools:** Under CNC machine tools, we have *CNC turning centre,* which does all the work of lathe and *CNC machining centre* which does milling, drilling etc., with provision for automatic tool changing and tool wear correction built into it.

A.9 HEAT TREATMENT

A.9.1. Definition

Heat treatment *is defined as an operation or combination of operations, involving heating and cooling of a metal or alloy in its solid state with the object of changing the characteristics of the material.*

A.9.2. Objects

Heat treatment is generally employed for following purposes :

1. *To improve machinability.*
2. *To change or refine grain size.*
3. *To relieve the stress of the metal induced during cold or hot working.*
4. *To improve mechanical properties, e.g., tensile strength, hardness, ductility, shock resistance, resistance to corrosion etc.*
5. *To improve magnetic and electric properties.*
6. *To increase resistance to wear, heat and corrosion.*
7. *To produce a hard surface on a ductile interior.*

A.9.3. Constituents of Iron and Steel

The different microscopic constituents of iron and steel which commonly occur are :

1. Ferrite,
2. Cementite,
3. Pearlite,
4. Austenite,
5. Troostite,
6. Sorbite, and
7. Martensite.

The *other constituents comprise the three allotropic forms of nearly pure iron, graphite and slag.*

1. **Ferrite.** Iron which contains little or no carbon is called *ferrite.* It is *very soft and ductile* and is known as *alpha* iron by the metallurgists. Ferrite is present to some extent in a great range of steels, particularly those low in carbon content, and it is also present, in soft cast iron. *Ferrite does not harden when cooled rapidly.* It forms smaller crystals when cooled from a bright red heat at a rapid rate.
2. **Cementite.** This is a definite carbide of iron (Fe_3C) which is *extremely hard,* being harder than ordinary hardened steel or glass. Cementite increases generally with the proportion of carbon present, and the hardness and also the brittleness of cast iron is believed to be due to this substance.

   It contains *6.6 percent carbon* and occurs either in the form of a network or in globular or massive form, depending on the analysis of the steel and the heat treatment to which it is subjected. *It is magnetic below 25°C. Its pressure in iron or steel decreases the tensile strength but increases the hardness and cutting qualities.*
3. **Pearlite.** Pearlite is the name given to a mixture of about 87.5 percent ferrite and 12.5 percent cementite. It comprises of *alternate layers* of ferrite and cementite in

steel. Under high magnification the ferrite and cementite can be seen to be arranged in alternate laminations or plates. When seen in the microscope the surface has appearance like *mother of pearl*, hence the name *pearlite*. The thickness of alternate plates and the distance between them is governed by the rate of cooling, slow cooling produces a coarser structure than rapid cooling. *Pearlite is eutectoid of steel.*

It has been found that the proportion of pearlite increases from nothing in the case of pure carbonless iron upto 100%, or saturation, for steel containing 0.90% of carbon, thus a 0.3 percent carbon steel will consist of about 33 percent perlite and rest ferrite. It is the characteristic of *soft steels* that they contain ferrite and pearlite, and the *hardness increases with the proportion of pearlite. Hard steels are mixtures of pearlite and cementite.*

4. **Martensite.** It is hard brittle mass of fibrous or needle like structures and is the chief constituent of hardened steel. The vickers pyramid numeral is anything upto 900 for an original carbon content of 0.9 percent. It has been found that martensite is produced by the rapid quenching of high carbon steel from a slightly higher temperature than the maximum temperature of critical interval. *It is not as tough as austenite. It differs from austenite in being magnetic.*

5. **Austenite.** It is a solid solution of iron-carbon which is stable only within a particular range of composition and temperature, and is *non-magnetic.* On cooling below 700°C it is completely transformed into ferrite which is magnetic and cementite to form the *eutectoid pearlite*, together with free ferrite or free cementite, depending on whether the carbon content is less or greater than 0.87 percent respectively.

   It is formed when carbon steel with more than 1.1 percent carbon is quenched rapidly from about 1000°C. The amount of austenite increases with the proportion of carbon, 0 upto 1.1 percent carbon, upto 70 percent for 1.6 to 1.8 percent carbon. *Austenitic steels cannot be hardened by usual heat treatment methods and are non-magnetic.*

6. **Troostite.** It is a structure in steel (consisting of very finely divided iron carbide in what is known at *"alpha-iron"*) produced either by tempering a martensitic steel at between 250 and 450°C or by quenching steel at a speed *insufficient* to suppress the thermal change point fully. The structure produced by the latter method should be more accurately termed very fine pearlite.

7. **Sorbite.** It is a structure which consists of evenly distributed carbide or iron particles in a mass of ferrite, formed when a fully hardened steel is tempered at between 550 and 650°C. *A sorbitic structure is characterised by strength and a high degree of toughness.*

A.9.4. Heat Treatment Processes

Refer to Fig. A.74. The various heat treatment processes are enumerated below :

1. Annealing.
2. Normalising.
3. Hardening.
4. Tempering.
5. Surface hardening :
   (*i*) Case hardening (by carburising)

(*ii*) Nitriding

(*iii*) Cyaniding

(*iv*) Flame hardening.

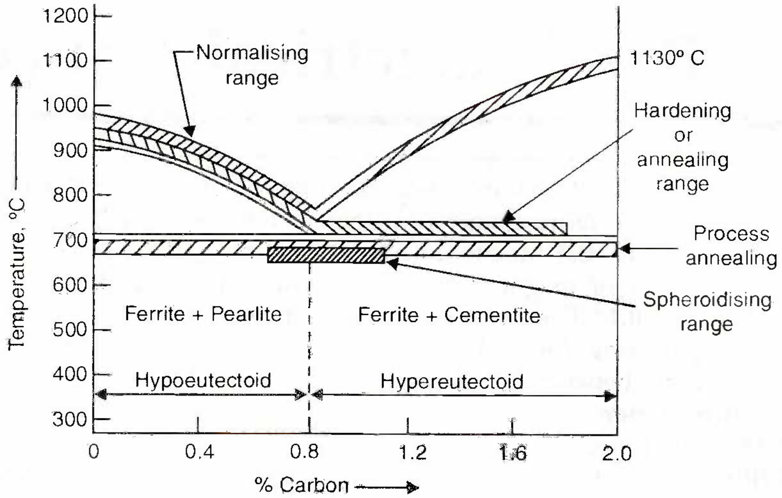


Fig. A.74. Temperature range for heat treatment processes.

❑ ❑ ❑

APPENDIX – B

Basic Electrical Concepts

B.1 ATOMIC STRUCTURE

— An **element** is defined as *a substance which cannot be decomposed into other substances. The smallest particle of an element which takes part in chemical reaction is known as* **atom.**

— All matter is composed of *atoms* which are *infinitesimally small.*

— All atoms are made of *electrons, protons* and *neutrons.* Most solid materials are classed, from the standpoint of electrical conductivity, as *conductors, semiconductors* or *insulators.* To be ***conductor,*** the substance must contain some *mobile electrons*— so that they can move freely between atoms. These free electrons come only from the *valence* (outer) *orbit* of the atom. *Conductivity depends on the number of electrons in the valence orbit.*

"The energy level of an electron increases as its distance from the nucleus increases. Thus an electron in the second orbit possesses more energy than electron in the first orbit, electrons in the third orbit have higher energy than in the second orbit and so on. It follows, therefore, that electrons in the last orbit will possess very high energy. These high energy electrons are less bound to the nucleus and hence they are more mobile. It is the mobility of last orbit electrons that they acquire the property of combining with other atoms. Further due to this combining power of last orbit electrons of an atom they are called valence electrons".

- Atoms with *fewer than four valence electrons* are good *conductors.*
- Atoms with *more than four valence electrons* are *poor conductors.*
- Atoms *with four valence electrons* are *semiconductors.*

Important data of an atom :

(*i*) Electron

Mass of an electron $= 9.11 \times 10^{-31}$ kg

$= \dfrac{1}{1840}$ mass of proton

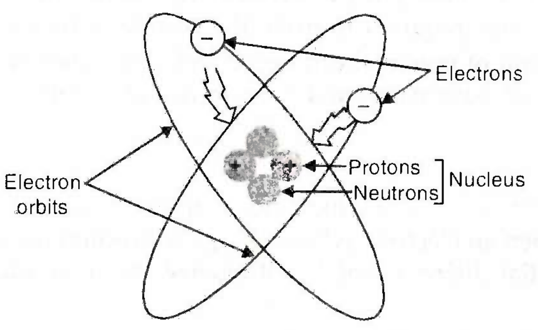


Fig. B.1. Atomic structure: Electron, proton and neutron.

Charge of electron $= -1.602 \times 10^{-19}$ coulomb

Diameter of an electron $= 10^{-15}$ m

(*ii*) Proton

Mass of proton $= 1.67 \times 10^{-27}$ kg

Charge on proton $= +1.602 \times 10^{19}$ coulomb.

(*iii*) Neutron

Mass of neutron = Mass of proton ($= 1.67 \times 10^{-27}$ kg)

Charge of neutron = Nil

Diameter of nucleus ..... of the order of 10^{-14} m

Diameter of orbits $= 10^4$ times the dia of the molecule.

— Normally, the atoms are electrically neutral, that, the number of electrons and protons are the same, cancelling each other's electrical force. Atoms *"stay together"* because *unlike charges attract each other.* The electrical force of the protons hold the electrons in their orbits. *Like electrical charges repel each other* so negatively charged electrons will not collide with each other.

Positive and negative ions :

When an electron is removed from a neutral atom, this atom becomes positively charged and is called *positive ion.* However, if an electron is added to a neutral atom, it becomes negatively charged and is called a *negative ion.* Thus, an atom becomes an ion by the gain or loss of electron.

B.2 ELECTRIC CURRENT

The controlled movement of electrons (or drift) through a substance is called ***current.***

— Current is the *rate at which electrons move.* One *ampere* (unit of current) represents 6.28×10^{18} electrons passing a point each second (1 coulomb past a point in one second).

$$\begin{bmatrix} \text{Ampere = coulomb/second} \\ \text{One coulomb = charge of } 6.28 \times 10^{18} \text{ electrons} \end{bmatrix}$$

— *When electricity flows through open space or vacuum as in the case of lightning or vacuum tubes instead of being confined to metallic conductors, it is termed as* ***electronic.***

B.3 ELECTROMOTIVE FORCE

— *Electromotive force* (e.m.f.) *is the force that causes a current of electricity to flow.*

— The potential difference (p.d.) V, between two points in a circuit is the electrical pressure or voltage required to drive the current between them.

— The ***volt*** is a unit of potential difference and electromotive force. It is defined as the difference of potential across a resistance of 1 ohm carrying a current of 1 ampere.

Electron volt :

Electron volt is a unit in terms of which the energies of atomic particles are expressed. It is the work done when an electron, whose charge is e coulombs, is moved in an electric field through a potential difference of 1 volt against the force (newtons) acting on the charge.

Thus 1 electron volt = e joules

B.4 RESISTANCE

The *opposition to flow of electrons* (due to bonds between protons and electrons, as well as to collisions) is called a *electrical resistance* (R).

— Resistance may also be defined as *"The property of the electric circuit which opposes the flow of current".*

The practical unit of electric resistance is ohm (Ω). It (ohm) is defined as the *resistance in which a constant current of 1 ampere generates heat at the rate of 1 watt. One volt* applied across *1 ohm* will produce *1 ampere.*

1 Mega-ohm (MΩ) = 10^6 Ω
1 kilo-ohm (kΩ) = 10^3 Ω
1 milli-ohm (mΩ) = 10^{-3} Ω
1 micro-ohm (μΩ) = 10^{-6} Ω

Laws of resistance :

The resistance of a conductor, such as a wire, of uniform cross-section depends on the following *factors* :

(*i*) Length (l): varies *directly* as its length l.
(*ii*) Cross-section (A): varies inversely as the cross-section A, of the conductor.
(*iii*) Nature of the material (ρ).
(*iv*) Temperature of the conductor: It almost varies directly with the temperature.

$$\therefore \quad R = \rho \frac{l}{A} \quad \text{...(B.1)}$$

where ρ is known as specific resistance or resistivity.

Specific resistance or resistivity of a material may be defined as *"The resistance between the opposite faces of a metre cube of that material".*

The unit of resistivity is ohm-metre (Ω-m).

Conductance (G):

Conductance (G) is the reciprocal of resistance $\left(G = \frac{1}{R} = \frac{A}{\rho l}\right)$.

Conductivity (σ):

The *reciprocal of specific resistance* $\left(\sigma = \frac{1}{\rho}\right)$ of a material is called its *conductivity.*

The unit of conductivity $\left(\sigma = G\frac{l}{A}\right)$ is mho/metre.

Temperature co-efficient of resistance :

Temperature co-efficient of resistance at 0°C may be defined as follows :

"The change in resistance per ohm for change in temperature of 1°C from 0°C".

Over *large temperature range* the simple formula

$$R_t = R_0(1 + \alpha t) \quad ...(B.2)$$

does not completely fit, but a formula of the type

$$R_t = R_0(1 + \alpha t + \beta t^2) \quad [B.2(a)]$$

(where β is a smaller co-efficient)

applies,

Also $\rho_t = \rho_0(1 + \alpha_0 \cdot t)$

where ρ_t, and ρ_0 are the resistivities at t° and 0°C respectively.

The effect of temperature on resistance

The following points are worth noting :

(*i*) The resistance of *metal conductors 'increases'* (α, *i.e.*, temperature co-efficient of resistance being *positive*) with rise of temperature; the rate of increase is very considerable for most pure metals, being as much as about $\frac{1}{1.50}$ of the total resistance for each centigrade rise in the case of iron; the effect is smaller in case of alloys, and very small indeed for materials such as manganin and constantan which are therefore very suitable for making *standard resistances.*

(*ii*) The resistance of semiconductors such as *carbon*, and all *electrolytes* 'decreases' as the temperature rises (α being *negative*).

Ohm's Law:

— Ohm's law can be stated as follows:

"For a fixed metal conductor, the temperature and other conditions remaining constant the current (I) through it is proportional to the potential difference (V) between its ends".

In other words, $\frac{V}{I}$ = constant or $\frac{V}{I} = R$

where R is the resistance of the conductor between the two points considered :

— The linear relationship ($I \propto V$) does not apply to all non-metallic conductors. For example, for silicone carbide, the relationship is given by:

$V = KI^x$ where K and x are constants and x is *less than unity.*

The following relations hold good :

(i) $P = VI = I^2R = \frac{V^2}{R}$

(ii) $I = \frac{P}{V} = \sqrt{\frac{P}{R}}$

(iii) $R = \frac{P}{I^2} = \frac{V^2}{P}$

(iv) $V = \frac{P}{I} = \sqrt{PR}$

[where, P = power in watts, V = voltage in volts, I = current in amperes, and R = Resistance in ohms]

Power is expressed in terms of kW (kilowatt = 1000 W) or MW (megawatt = 1000 kW or 10^6 W).

Electrical energy is expressed in terms of kWh (kilowatt hours)

$$1 \text{ kWh} = 1 \text{ kW} \times 1 \text{ hour} = 1000 \text{ watt-hours}$$
$$(= 1000 \times 60 \times 60 \text{ watt-sec.})$$

Linear and non-linear resistors :

— A *linear resistor is one which obeys Ohm's law.* A circuit which contains only linear components is called a *linear circuit.*

— Such elements in which the *V/I* (volt-ampere) plots are not *straight lines but curves* are called *non-linear resistors* or *non-linear elements.*

Limitations of Ohm's law :

Ohm's law does not apply under the following conditions :

1. Electrolytes where enormous gases are produced on either electrode.
2. Non-linear resistors like vacuum radio valves, semiconductors, gas filled tubes etc.
3. Arc lamps.
4. Metals which get heated up due to flow of current through them.
5. Appliances like metal rectifiers, crystal detectors, etc. in which operation depends on the direction of current.

Resistances in series :

Figure B.2 shows three resistances connected in series. Obviously *current flowing through each resistance will be same* but *voltage drop* across each of them will *vary* as per value of individual resistance.

Also the sum of all the voltage drops ($V_1 + V_2 + V_3$) is equal to the applied voltage (V).

i.e., $V = V_1 + V_2 + V_3$

$IR = IR_1 + IR_2 + IR_3$ [using Ohm's law: $V = IR$]

i.e., $R = R_1 + R_2 + R_3$ ...(B.3)

where R is the *equivalent resistance* of series combination.

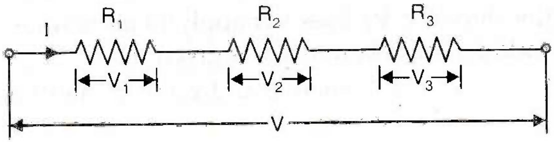


Fig. B.2. Resistances in series.

Resistances in parallel:

Refer to Fig. B.3. In this case *voltage across each resistance will be same but current will be different* depending upon the value of the individual resistance.

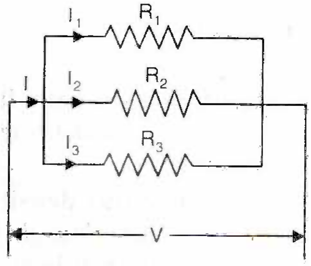


Fig. B.3. Resistances in parallel.

i.e., $I = I_1 + I_2 + I_3$

$$\frac{V}{R} = \frac{V}{R_1} = \frac{V}{R_2} = \frac{V}{R_3}$$

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \quad \text{...(B.4)}$$

where R is the *equivalent resistance* of the parallel combination.

$$R = \frac{R_1R_2R_3}{R_1R_2 + R_2R_3 + R_3R_1} \quad \text{...[B.4(a)]}$$

$$G = G_1 + G_2 + G_3 \quad \text{...(B.5)}$$

Superconductivity :

Equation $R_t = R[1 + \alpha(t - 20)]$ holds good for temperature below 20°C. But at *very low temperature, some metals acquire zero electrical resistance and zero magnetic induction; the property known as* **superconductivity.**

Superconducting elements: Zinc, cadmium, mercury, lead.

Typical superconducting compounds and alloys: PbAu, $PbT1_2$, SnSb, CuS, NbN, NbB, ZrC.

The superconductivity will disappear if

(*i*) *The temperature of the material is raised above its critical temperature.*

OR

(*ii*) *A sufficiently strong magnetic field or current density is employed.*

B.5 MAGNETIC FIELD

Magnetism. It is defined as the *property which certain materials have that permits them to produce or conduct magnetic lines of force.*

Magnet. It is an *object about which a magnetic field exists* and is either natural or man-made. The latter type can be either temporary or permanent.

— Each magnet has a magnetic field around it just as the earth does. The magnet field is *strongest at the end of the magnet.* In the *centre* of the magnet the strength is *negligible.*

— Magnetic lines of force (also called **magnetic flux**) have direction similar to the motion of electric charges. A magnet has a north pole and a south pole just as electric charges are either negative or positive.

— Like poles of magnets repel whereas unlike poles attract.

— Magnetism can be induced in a magnetic material by placing it in a magnetic field.

— The *lines of force tend to spread away from each other because of the mutual repulsion between the lines.* Thus a magnetic field extends outward from the magnet, and the lines are wider spaced (less energy) as the distance from the magnet increases.

B.6 TERMS CONNECTED WITH MAGNETIC MATERIALS

1. **Magnetic force.** It is the force exerted by one magnet on another to attract it or repel it.
2. **Unit pole strength.** It is defined as the *strength of that pole which when placed in a vacuum at a distance of one metre from a similar and equal pole, repels it with a force of one newton.*
3. **Magnetic flux density (B).** It is defined as the flux (ϕ) *or lines of force passing per unit area (A) through any substance through a plane at right angles to the direction of magnetic flux;* it is measured in Wb/m^2 (or *T*, *i.e.*, Telsa).

   Mathematically, $$B = \frac{\phi}{A} \qquad \text{...(B.6)}$$
4. **Magnetic field strength.** It may be defined in the following two ways:
   (*i*) Field strength at any point within a magnetic field is the *number of lines of force passing through a unit area round the point considered and held perpendicular to the lines.*
   (*ii*) Field strength at any point within a magnetic field *is the force exerted by a unit north pole at that point.*
5. **Relative permeability (μ_r).** *It is the ratio of flux density (B) produced in that material to the flux density produced in vacuum by the same magnetising force (H).* It is denoted by μ_r ($\mu_r = 1+\frac{K}{\mu_0}$, where *K* is susceptibility).
6. **Absolute permeability (μ).** It is the *ratio of flux density in that material to the magnetising force producing that flux density* and is denoted by μ ; $\mu = \mu_0\mu_r$; where μ_0 is the permeability of free space having a value of $4\pi \times 10^{-7}$ H/m.
7. **Magnetic potential.** The magnetic potential at any point within a magnetic field is *measured by the work done in carrying a unit north pole from infinity to that point against the force of magnetic field.*
8. **Intensity of magnetisation (I).** It is defined *as the pole strength per unit area of the bar or magnetic moment per unit volume of the bar.* It is denoted by letter *I*.
9. **Susceptibility (K).** It is defined as the ratio of intensity of magnetisation (*I*) to magnetising force (*H*).

   Mathematically, $$K = \frac{I}{H} \qquad \text{...(B.7)}$$
10. **Magnetomotive force (m.m.f.).** *It is that force which drives or tends to drive the flux through a magnetic circuit.* In short it is written as m.m.f. It is the product of number of turns (*N*) and current (*I*) in amperes in those turns, *i.e.*, m.m.f. = *NI*.
11. **Magnetic reluctance.** It is that property of the material *which opposes the production of magnetic flux in it.*
12. **Co-ercive force.** It may be defined as the *demagnetising force which is necessary to neutralise completely the magnetism in an electromagnet after the value of magnetising force becomes zero.*
13. **Remanance.** It is defined as the *magnetic flux density which still persists in magnetic material even when the magnetising force is completely removed.* It is expressed in Wb/m^2 (or *T*).
14. **Retentivity.** *It is that properly of magnetic material which is measured by its maximum value of the residual induction.*

B.7 CLASSIFICATION OF MAGNETIC MATERIALS

In accordance with the value of *relative permeability* the magnetic materials may be classified in the following three ways:

1. **Ferromagnetic materials.** The relative permeabilities of these materials are *much greater than unity* and dependent on the field strengths.
   Examples. *Iron, cobalt* and *nickel.*
   Gadolinium, however, also comes under this classification. These materials have *high susceptibility.*
2. **Paramagnetic materials.** These have relative permeability *slightly greater than unity* and are *magnetised slightly.*
   Examples. *Aluminium, platinum* and *oxygen.*
3. **Diamagnetic materials.** The relative permeability of these materials is *slightly less than unity.* They repel the lines of force slightly.
   Examples. *Bismuth, silver, copper* and *hydrogen.*

B.8 MAGNETICALLY SOFT MATERIALS

The magnetically soft materials (*suitable for making electromagnets*) are characterised as follows :

1. They have high permeability.
2. The magnetic energy stored is not high.
3. They have negligible co-ervice force (due to which these are *not* suitable for making permanent magnets).
4. They have low remanance.

Examples. *Pure or ingot iron, manganese and nickel steels, cast iron, silicon steels, carbon steels, mumetal, perminvar, permalloy.*

B.9 MAGNETICALLY HARD MATERIALS

These are suitable for making *permanent magnets* and have the following characteristics:

1. They possess high value of *BH* product.
2. High retentivity.
3. High co-ercivity.
4. Strong magnetic reluctance.
5. Hysteresis loop is more rectangular in shape.

Examples. *Tungsten steel, cobalt steel, chromium steel, alnico, cunife, hypernic.*

B.10 LAWS OF MAGNETIC FORCE

Coulomb, through his experiments found that the force between two magnetic poles placed in a medium is

(i) directly proportional to their pole strengths (m_1, m_2),
(ii) inversely proportional to the square of the distance (d) between them, and
(iii) inversely proportional to the absolute permeability (μ) of the surrounding medium.

i.e.,
$$F \propto \frac{m_1 m_2}{d^2}$$

or
$$F = k\frac{m_1 m_2}{\mu d^2}$$

where k = constant)

In the S.I. system, the value of $k = \frac{1}{4\pi}$

$$\therefore \quad F = \frac{m_1 m_2}{4\pi \mu d^2} = \frac{m_1 m_2}{4\pi \mu_0 \mu_r d^2} \quad \text{...(B.8)}$$

$$(\because \mu = \mu_0 \mu_r)$$

B.11 MAGNETIC FIELD DUE TO A CURRENT CARRYING CONDUCTOR

— When an electric current flows through a wire, a magnetic field is built up around the wire itself. This can be seen using a card-board, iron filings, and a current-carrying wire. When the wire passes through the cardboard and the current flowing, iron filings are sprinkled on to the cardboard. They can be seen arranging themselves in a magnetic field (Fig. B.4). The magnetic lines of force are referred to as flux. Just as in a natural magnet, the field is strongest near the wire and diminishes as the distance from the wire increases.

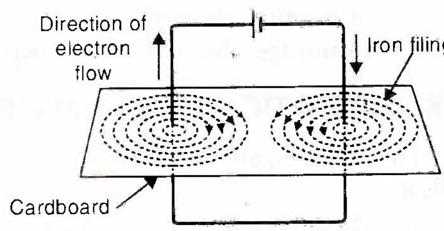


Fig. B.4. Magnetic field around a wire that is carrying electric current.

— Flux around a wire does have direction. Flux direction is determined by the direction of electron flow within the wire. As shown in Fig. B.5, the *North pole of the compass needle indicates the direction of flux or magnetic field around the wire.* The **dot** *in the centre of the wire on the left indicates the point of the current-direction arrow coming* **toward the observer** ; the **X** at the right represents the tail of the current arrow **pointing away from the observer**. If the direction of electron flow within the wire is reversed, the compass needles will reverse themselves, indicating a change in flux direction.

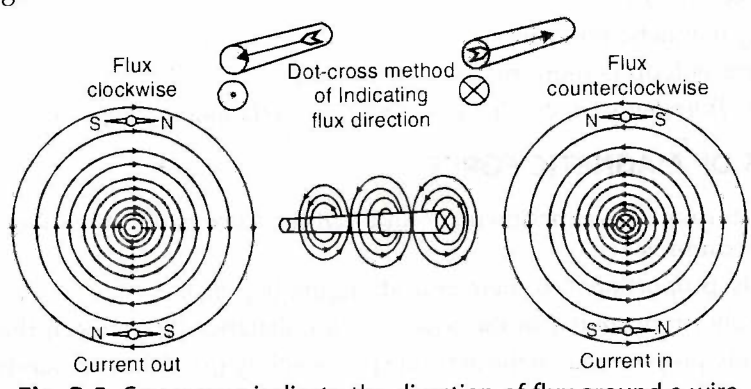


Fig. B.5. Compasses indicate the direction of flux around a wire.

Right hand rule (or right hand screw rule)

The direction of the magnetic field can be found by using *rigint hand rule or the right hand screw rule.* The right hand rule states as follows :

"Grasp the wire in the right hand, with the thumb pointing in the direction of the current. The fingers will curl around the wire in the direction of the magnetic field".

Figure B.6 illustrates this rule.

The right hand screw rule can be explained as follows :

As a wood screw is turned clockwise it moves (or progresses) into the wood. The *horizontal direction of the screw is analogous to the direction of current in a conductor. The circular motion of the screw shows the direction of the magnetic flux around the conductor.*

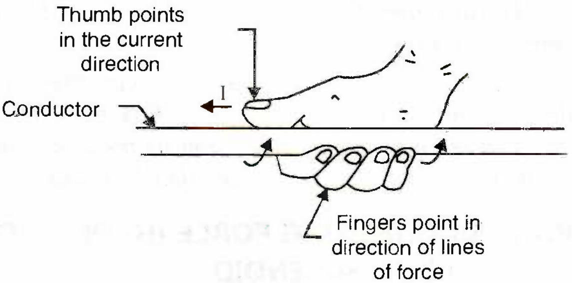


Fig. B.6. Right hand rule (or right hand screw rule).

B.12 FORCE ON A CURRENT-CARRYING CONDUCTOR LYING IN A MAGNETIC FIELD

Refer to Fig. B.7. It has been found that *whenever a current-carrying conductor is placed in a magnetic field, it experiences a force which acts in a direction perpendicular both to the direction of the current and the field.*

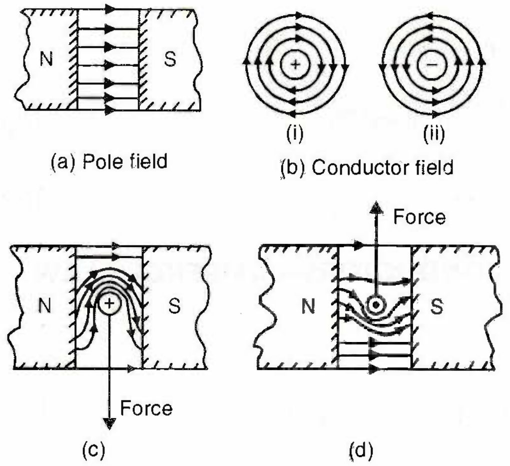


Fig. B.7. Force on a current carrying conductor lying in a magnetic field.

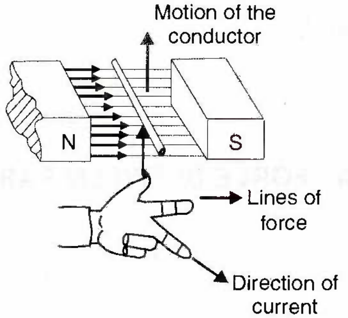


Fig. B.8. Fleming's left hand rule.

The force developed in the conductor is given by the relation :

$$F = BIl \text{ newtons} \qquad \text{...(B.9)}$$

$$(= \mu_0\mu_r\, HIl \text{ newtons})$$

where, F = Force developed in the conductor,

B = Flux density, T (Wb/m^2),

I = Current in the conductor, A, and

l = Exposed length of the conductor, m.

$$\begin{bmatrix} \mu_0 = \text{Absolute permeability;} \\ \mu_r = \text{Relative permeabililty;} \\ H = \text{Magnetising force.} \end{bmatrix}$$

The *direction of this force may be easily found by Fleming's left hand rule* (See Fig. B.8) which states as follows :

"Hold your left hand with index finger, middle finger and thumb at right angles. If the index finger points in the direction of the flux from north to south and middle finger points in the direction of the imposed voltage and its resulting conventional current flow, then the thumb will point in the direction of the force that is developed".

B.13 MAGNETISING FORCE (H) OF A LONG STRAIGHT CONDUCTOR AND A LONG SOLENOID

Long straight conductor :

$$H = \frac{NI}{2\pi r} AT/m \qquad ...(B.10)$$

and,

$$B = \frac{\mu_0 \mu_r NI}{2\pi r} \text{ Wb/m}^2 \text{ (or } T) \text{ ... in a medium} \qquad ...[B.11(a)]$$

$$= \frac{\mu_0 NI}{2\pi r} \text{ Wb/m}^2 \text{ (or } T) \text{ ... in air} \qquad ...[B.11(b)]$$

where, r = Distance of the point from the centre of the conductor.

Long solenoid :

$$H = \frac{NI}{l} AT/m \qquad ...(B.12)$$

and,

$$B = \frac{\mu_0 \mu_r NI}{l} \text{ ... in a medium} \qquad ...[B.13(a)]$$

$$= \frac{\mu_0 NI}{l} \text{ ... in air} \qquad ...[B.13(b)]$$

B.14 FORCE BETWEEN PARALLEL CONDUCTORS—AMPERES'S LAW

$$F = \frac{\mu_0 I_1 I_2 l}{2\pi d} \text{ newtons} \qquad ...(B.14)$$

where, F = Force between two parallel conductors,

I_1, I_2 = Currents flowing through two parallel conductors,

l = Length of each conductor, and

d = Distance between the conductors.

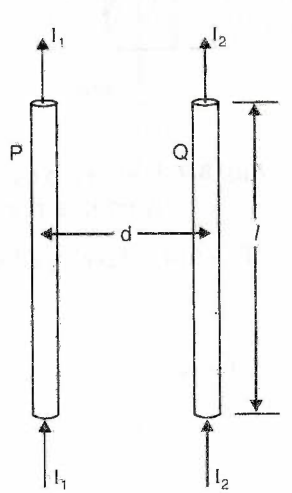


Fig. B.9. Force between two parallel conductors.

Eqn. (B.14) is known as Ampere's law and is used to define the ampere in S.I. units.

If $l = d = 1$ m ; $I_1 = 1$ A, then $F = 2 \times 10^{-7}$ N

Hence, one ampere is defined as follows:

"An ampere is that current when flowing in each of the two infinitely long parallel conductors situated in vacuum and separated 1 metre between centres, produces on each conductor a force of 2×10^{-7} N per metre length".

B.15 FARADAY'S LAWS OF ELECTROMAGNETIC INDUCTION

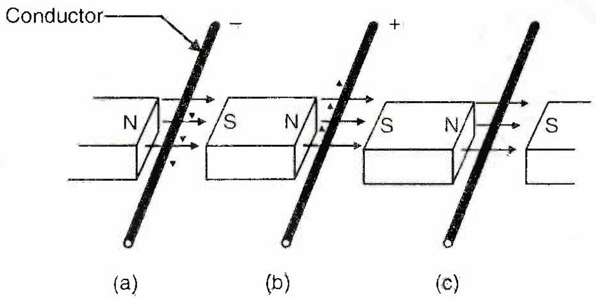
Refer to Fig. B.10. *"The phenomenon whereby an e.m.f. and hence current is induced in any conductor which is cut across or is cut by a magnetic flux is known as* ***electromagnetic induction****"*.

First law. It states as follows :

"Whenever the magnetic flux linked with a circuit changes, an e.m.f. is always induced in it".

OR

"Whenever a conductor cuts magnetic flux, an e.m.f. is induced in that conductor".



(*a*) *Voltage induced across a wire moving downward.* (*b*) *Voltage induced across a wire moving upward.* (*c*) *No voltge is induced in a wire moving paralled to the field.*

Fig. B.10. When a conductor is moved across a magnetic field a voltage is induced in the conductor.

Second Law. It states as follows :

"The magnitude of induced e.m.f. is equal to the rate of change of flux-linkages".

Mathematically, $$e = -\frac{Nd\phi}{dt} \text{ volts} \qquad \text{...(B.15)}$$

where e = Induced e.m.f.,

$\frac{d\phi}{dt}$ = Rate of change of flux, and

N = Number of turns of the coil.

[Usually, a *minus* sign is given to the right-hand side expression to signify the fact that the induced e.m.f. sets up current in such a direction that magnetic effect produced by it opposes the very cause producing it.]

Direction of induced e.m.f. and current :

The direction of the induced current may be found easily by applying either Fleming's Right-hand Rule (Fig. B.11) or Lenz's Law. *Fleming's rule is used where induced e.m.f. is due to flux-cutting (i.e., dynamically induced e.m.f.) and Lenz's Law when it is due to change by flux-linkage (i.e., statically induced e.m.f.).*

Lenz's Law. Figure B.12 shows induction of an e.m.f. in a simple circuit. The direction of the induced e.m.f. is determined by Lenz's law, *which states that the current produced by the induced e.m.f. opposes the change in flux.*

Lenz's may also be stated as follows:

"In all cases of electromagnetic induction, an induced voltage will cause a current to flow in

a closed circuit in such a direction that the magnetic field which is caused by that current will oppose the change that produced the current".

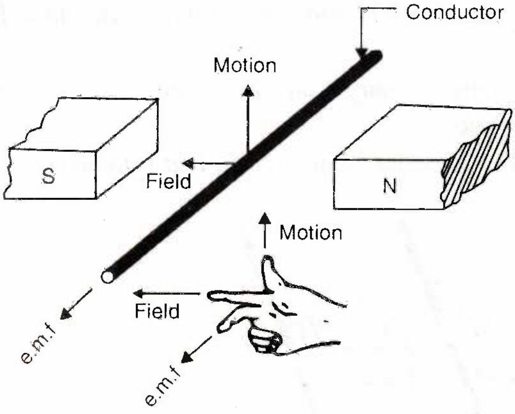


Fig. B.11. Fleming's right hand rule.

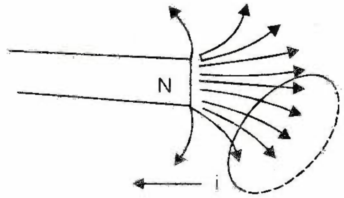


Fig. B.12. Induction of e.m.f. in a simple circuit.

B.16 INDUCED E.M.F.

Induced e.m.f. may be of the following two types :

1. Dynamically induced e.m.f.
2. Statically induced e.m.f.

1. Dynamically induced e.m.f. :

Refer to Fig. B.10. The e.m.f. induced (e) in the conductor is given by :

$$e = Blv \text{ volt} \quad \text{...(B.16)}$$

where,
B = Flux density of the magnetic field in *tesla,*
l = Length of the conductor is *metres,* and
v = Velocity of the conductor in *m/s.*

If the conductor moves at an angle θ with the direction of flux then the induced e.m.f.

$$e = Blv \sin \theta \text{ volts} \quad \text{...(B.17)}$$

The direction of the induced e.m.f. is given by Fleming's Right hand rule.

2. Statically induced e.m.f. :

The e.m.f. induced by *variation of flux is termed as "statically induced e.m.f.".*

Statically induced e.m.f. can be further subdivided as follows :

(*i*) Self-induced e.m.f.
(*ii*) Mutually induced e.m.f.

(*i*) Self-induced e.m.f. :

Self-induced e.m.f. is the *e.m.f. induced in a coil due to the change of its own flux linked with it. If the current through the coil (Fig. B.13) is changed* then the flux linked with its own turns will also change which will produce in it, what its called *self-induced e.m.f.* $\left(e = -N\dfrac{d\phi}{dt}\right)$. The direction of

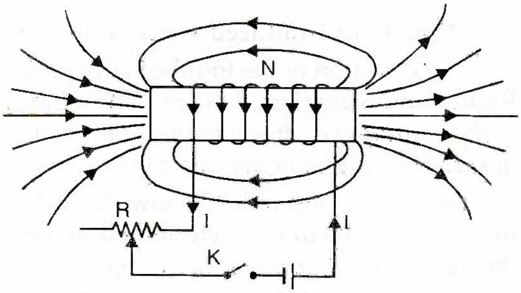


Fig. B.13. Self-induced e.m.f. and self-inductance.

this e.m.f. is given by Lenz's law (and would be such as to oppose any change of flux which is, in fact, the very cause of its production).

Self-inductance. The *property of the coil due to which it opposes any increase or decrease of current or flux through it, is known as* ***self-inductance***. It is measured in terms of self-induction L (in henry).

Self-induction is sometimes analogously called *electromagnetic or electrical inertia.*

Co-efficient of self-induction (L) may be found by the following relations :

1. $$L = \frac{N\phi}{l} \text{ henry} \quad \text{...(B.18)}$$

2. $$L = \frac{\mu_0 \mu_r A N^2}{l} \text{ henry} \quad \text{...(B.19)}$$

3. $$L = \frac{e_L}{\frac{dI}{dt}} \text{ henry} \quad \text{...(B.20)}$$

where, N = Number of turns of the solenoid,

A = Area of cross-section,

e_L = Induced e.m.f., and

$\frac{dI}{dt}$ = Rate of change of current.

Energy in electromagnetic field :

The energy in electromagnetic field is given by :

$$W = \frac{1}{2} L I^2 \text{ joules} \quad \text{...(B.21)}$$

Eqn. (B.21) gives an expression *for stored energy in the magnetic field when the current is increased from zero and the same amount of energy is released when the current is reduced to zero.*

(*ii*) Mutually induced e.m.f.

Refer to Fig. B.14. *Production of e.m.f. in coil B due to change in current A is called* ***mutually induced e.m.f.***

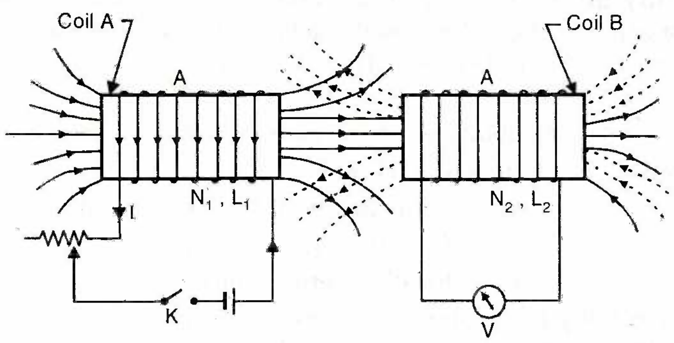


Fig. B.14. Mutually induced e.m.f. and mutual inductance.

Mutual inductance. It is defined as *"The phenomenon by which one circuit causes an e.m.f. induced in the adjacent circuit by induction when flux produced by it is changed".*

Co-efficient of mutual inductance (M) may be found by the following relations :

1. $$M = \frac{N_2\phi_1}{I_1} \text{ henry} \quad \text{...(B.22)}$$

2. $$M = \frac{\mu_0\mu_r A N_1 N_2}{l} \text{ henry} \quad \text{...(B.23)}$$

3. $$M = -\frac{e_M}{\frac{dI_1}{dt}} \quad \text{...(B.24)}$$

Co-efficient of coupling (k)

It is defined as the *ratio of mutual inductance between the coils and the square root of product of self-inductance of each coil.*

In other words, $$k = \frac{M}{\sqrt{L_1L_2}} \quad \text{...(B.25)}$$

B.17 INDUCTANCES IN SERIES

In general we have,

$$L = L_1 + L_2 + 2M \quad \text{...... if } \textit{m.m.fs. are additive} \quad \text{...(B.26)}$$

$$L = L_1 + L_2 - 2M \quad \text{...... if } \textit{m.m.fs. are subtractive} \quad \text{...(B.27)}$$

B.18 INDUCTANCES IN PARALLEL

In general we have,

$$L = \frac{L_1L_2 - M^2}{L_1 + L_2 - 2M} \text{ when mutual field } \textit{assists} \text{ the separate fields.} \quad \text{...(B.28)}$$

$$L = \frac{L_1L_2 - M^2}{L_1 + L_2 + 2M} \text{ when two fields } \textit{oppose} \text{ each other.} \quad \text{...(B.29)}$$

B.19 TERMS CONNECTED WITH MAGNETIC CIRCUIT

A ***magnetic circuit*** *is defined as the route* or *path which is followed by a magnetic flux.*

1: Permeability (μ). *Permeability* of any material is *a measure of ease with which the atoms can be arranged.* It is also defined *as the ability of a material to concentrate magnetic flux and offer little opposition to the flux lines.* The symbol for permeability is the Greek letter mu (μ).

The S.I. unity for permeability is henry/metre (H/m)

Mathematically, $$\mu = \mu_0\mu_r \text{ H/m} \quad \text{...(B.30)}$$

where, μ_0 = Permeability of free space, and

$= 4\pi \times 10^{-7}$ H/m (S.I. units)

μ_r = Relative permeability.

Relative permeability (μ_r) is simply a numeric which expresses the degree to which the material is a better conductor of magnetic flux as compared to free space.

μ_r for *air* (and non-magnetic materials) = 1

μ_r for *diagmagnetic* materials = slightly less than one

μ_r for *paramagnetic* materials = slightly higher than one

μ_r for *ferromagnetic* materials = in the hundreds or thousands.

2. Magnetomotive force (m.m.f.) Magnetomotive force *drives or tends to drive flux through magnetic circuit.* It is equal to the work done in joules in carrying a unit magnetic pole once through the entire magnetic circuit ; m.m.f. is measured in *ampere-turns* (*AT*).

$$AT = NI$$

where, N = Number of turns of a magnetic circuit, and

I = Current in ampere in those turns.

3. Reluctance (S). *Reluctance is a measure of opposition offered by a magnetic circuit to the setting up of flux.*

The reluctance (S) of a magnetic circuit is given by :

$$S = \frac{l}{\mu A} = \frac{l}{\mu_0 \mu_r A} \qquad \text{...(B.31)}$$

where, l = Length of the magnetic circuit,

A = Cross-sectional area of the magnetic circuit,

μ_0 = Absolute permeability, and

μ_r = Relative permeability.

Reluctance of a magnetic circuit is the *ratio of m.m.f. and flux*

i.e., $$\text{Reluctance} = \frac{\text{M.m.f.}}{\text{Flux}}$$

or, $$S = \frac{NI}{\phi} \qquad \text{...(B.32)}$$

The unit of reluctance is AT/Wb. Since 1 AT/Wb = 1/henry, the unit of reluctance is *"reciprocal henry"*.

4. Permeance. The *reciprocal of reluctance* is known as *permeance.*

i.e., $$\text{Permeance} = \frac{1}{\text{Reluctance}} = \frac{1}{S}$$

It is measured in Wb/AT or henry.

5. Reluctivity. It is the *specific reluctance* and corresponds to resistivity which is specific resistance.

Relation between flux density (B) and magnetic field strength (H) :

Comparing Eqns. (B.31) and (B.32), we get

$$\frac{l}{\mu_0 \mu_r A} = \frac{NI}{\phi}$$

or, $$\frac{\phi}{\mu_0 \mu_r A} = \frac{NI}{l} \qquad \text{(By rearranging)}$$

But, $$\frac{\phi}{A} = B$$

and, $$\frac{NI}{l} = H$$

$\therefore$ $$\frac{B}{\mu_0 \mu_r} = \text{H}$$

or, $$B = \mu_0 \mu_r H \qquad \text{...(B.33)}$$

or, $B = \mu H$

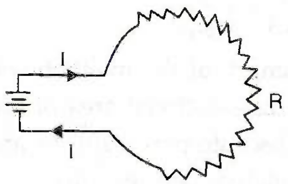
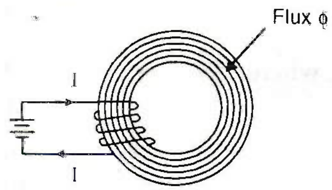
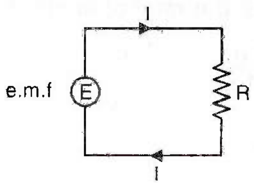
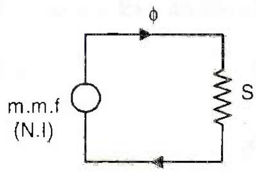
(where μ (permeability) $= \mu_0 \mu_r$]

Thus, *permeability is the ratio of flux density to magnetic field strength.*

B.20 COMPARISON OF ELECTRIC AND MAGNETIC CIRCUITS

The analogy between electric and magnetic circuits is given in table B.1.

Table B.1. Analogy between electric and magnetic circuits (Similarities)

Aspects	Electric circuit	Magnetic circuit
		
1. *Equivalent circuit*		
2. *Exciting force*	Battery voltage (E)	Ampere-turns (AT)
3. *Response*	Current (I)	Flux (ϕ)
4. *Ohm's law*	$I = \frac{E}{R}$	$\phi = \frac{\text{m.m.f.}}{\text{reluctance}} = \frac{NI}{S}$
5. *By dimensions*	$R = \rho \frac{l}{a}$ (Conductance $= 1/R$)	$S = \frac{l}{\mu_0 \mu_r A}$ (Permeance $= 1/S$)
6. *Proportionality*	ρ	$\frac{1}{\mu} \left(= \frac{1}{\mu_0 \mu_r} \right)$
7. *Field intensity*	Electric field intensity $= \frac{E}{l}$ V/m	Magnetic field intensity $= \frac{NI}{l}$ AT/m
8. *Density*	Current density (A/m^2)	Flux density (Wb/m^2)

B.21 ALTERNATING VOLTAGE AND CURRENT

Modern alternators produce an e.m.f. which is for all practical purposes sinusoidal (*i.e.*, a sine curve), the equation between the e.m.f. and time being :

$$e = E_{max} \sin \omega t \quad \text{...(B.34)}$$

where, e = Instantaneous voltage ; E_{max} = Maximum voltage

ωt = Angle through which the armature has turned from neutral.

Taking the frequency as f hertz (cycles per second), the value of ω will be $2\pi f$, so that the equation reads

$$e = E_{max} \sin (2\pi f)t.$$

The graph of the voltage will be as shown in Fig. B.15.

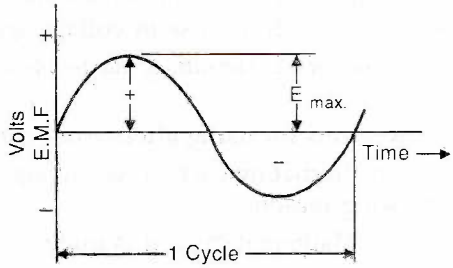


Fig. B.15. The graph of the sinusoidal voltage.

Cycle. One complete set of positive and negative values of an alternating quantity is known as a *cycle.* A cycle may also sometimes be specified in terms of angular measure. In that case, one complete cycle is said to spread over 360° or 2π radians.

Amplitude. The maximum value, positive or negative, of an alternating quantity, is known as its *amplitude.*

Frequency (f). The number of cycles/second is called the frequency of the alternating quantity.

Its unit is *hertz* (Hz).

Time period (T). The time taken by an alternating quantity to complete the cycle is called its *time period.* For example, a 50 hertz (Hz) alternating current has a time period of $\frac{1}{50}$ second.

Time period is reciprocal of frequency,

i.e., $$T = \frac{1}{f}\left(\text{or } f = \frac{1}{T}\right). \quad \text{...(B.35)}$$

Root mean square (R.M.S.) value. The r.m.s. value of an alternating current is given by that steady (D.C.) current *which when flowing through a given circuit for a given time produces the same* **heat** *as produced by the alternating current when flowing through the same circuit for the same time.*

R.M.S. value is the value which is taken for power purposes of any description. This value is obtained by finding the square root of the mean value of the squared ordinates for a cycle or half-cycle (See Fig. B.15).

$$F_{r.m.s.} = E_{max} \times \frac{1}{\sqrt{2}} = 0.707\, E_{max.} \quad \text{...(B.36)}$$

This is the value which is used for all power, lighting and heating purposes, as in these cases the power is proportional to the square of the voltage.

Average or mean value. The average value of an alternating current is expressed by *that steady current which transfers any circuit the same* **charge** *as is transferred by that alternating current during the same time.*

The average value of the voltage will be found to be 0.636 of the maximum value for a perfect sine wave, giving the equation

$$E_{av.} = 0.636\, E_{max.} \quad \text{...(B.37)}$$

The *mean value is only of use in connection with processes where the results depend on the current only, irrespective of the voltage, such as electroplating or battery charging.*

B.22 FORM FACTOR AND PEAK FACTOR

Form factor. *The ratio of r.m.s. (or effective) value to average value is the form factor* (K_f) *of the wave form.* It has use in voltage generation and instrument correction factors.

Peak factor. *The ratio of maximum value to the r.m.s. value is the peak factor* (K_p) of the wave form.

Reasons for using alternating current (or voltage) of sinusoidal form :

An alternating current (or voltage) of sinusoidal form is normally used because of the following reasons :

1. Mathematically, it is quite simple.
2. Its integrals and differentials both are sinusoidal.
3. It lends itself to vector representation.
4. A complex wave form can be analysed into a series of sine waves of various frequencies, and each such component can be dealt with separately.
5. This waveform is desirable for power generation, transmission and utilisation.

B.23 A.C. THROUGH PURE OHMIC RESISTANCE ALONE

Refer to Fig. B.16. Where a sinusoidal e.m.f. is placed across a pure resistance the current will be *in phase with the e.m.f.*, and if shown graphically will be in phase with the e.m.f. curve.

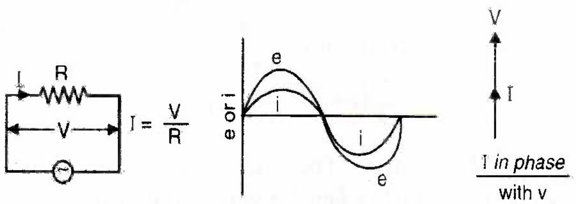


Fig. B.16. Purely resistive circuit.

The current $I = \dfrac{V}{R}$

where, V = R.m.s. value of the applied e.m.f. or voltage, and

R = Resistance in ohms.

(The value of I will be the r.m.s. value)

The *power* (P) in a purely resistive circuit is given by the product of the r.m.s. voltage and the r.m.s. current, *i.e.*,

$$P = VI.$$

B.24 A.C. THROUGH PURE INDUCTANCE ALONE

Refer to Fig. B.17. If a sinusoidal e.m.f. is placed across a pure inductance the current will be found to be,

$$I = \frac{V}{2\pi fL}$$

where, V = Voltage (r.m.s. value),

f = Frequency, and

L = The inductance in henries (H).

(The value of I being the r.m.s. value)

— The current will *lag behind the voltage* and the graphs will as shown in Fig. B.17, the *phase difference being 90°*.

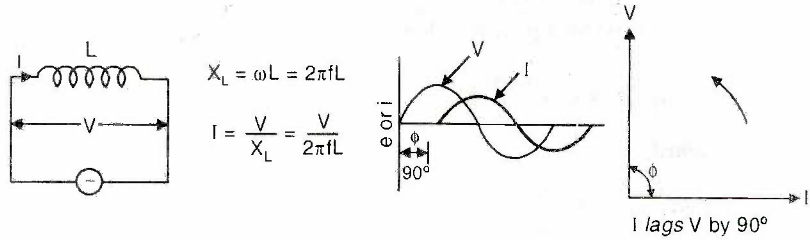


Fig. B.17. Purely inductive circuit.

— The expression $2\pi fL$ (or ωL) is termed as *inductive reactance* (X_L); $I = \dfrac{V}{X_L} = \dfrac{V}{2\pi fL}$.

— *Power consumed is zero.*

B.25 A.C. THROUGH PURE CAPACITANCE ALONE

Refer to Fig. B.18. If a sinusoidal e.m.f. is placed across a capacitor the current will be,

$$I = (2\pi f) \cdot CV$$

where
C = capacitance in farads (F);
f = frequency; and
V = voltage (r.m.s. value)

— In this case the current *leads* the voltage by 90°, as shown in Fig. B.18.

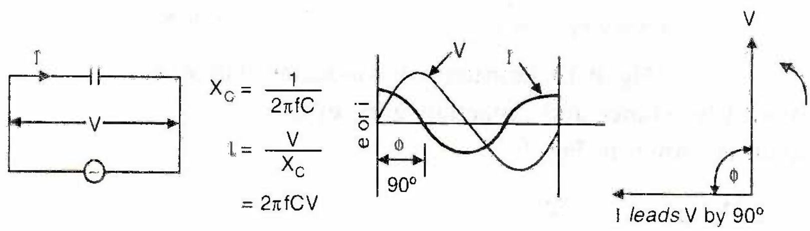


Fig. B.18. Purely capacitive circuit.

— The expression $\dfrac{1}{2\pi fC}\left(\text{or } \dfrac{1}{\omega C}\right)$ is termed the capacitive reactance (X_c) and the current is given by :

$$I = \frac{V}{X_c} = 2\pi f\, CV$$

— *Power consumed is zero.*

B.26 A.C. SERIES CIRCUITS

R-L Circuit (Resistance and inductance in series) :

R-L circuit is shown in the Fig. B.19.

Important formulae :

1. *Impedance*, $Z = \sqrt{R^2 + X_L^2}$ (where, $X_L = 2\pi fL\ \Omega$).

2. *Current*, $I = \dfrac{V}{Z}$.

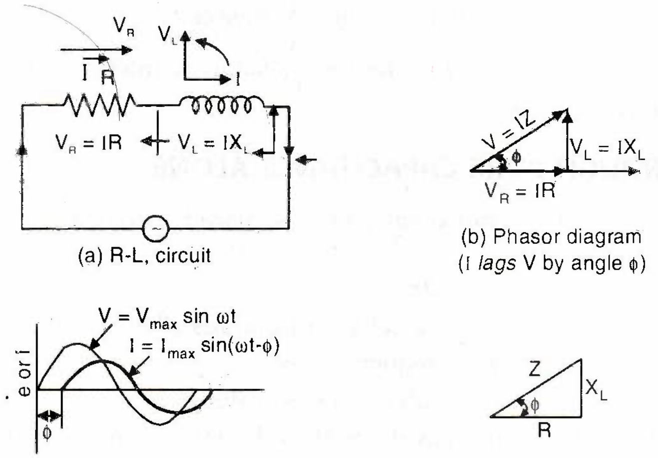
3. *Power factor,*

$$\cos\phi = \frac{R}{Z}\left(=\frac{\text{true power}}{\text{apparent power}}=\frac{W}{VA}\right)$$

(or angle of lag, $\phi = \cos^{-1}\frac{R}{Z}$).

4. *Power consumed,*

$$P = VI\cos\phi\left(=IZ\times I\times\frac{R}{Z}=I^2R\right).$$



(a) R-L, circuit

(b) Phasor diagram
(I *lags* V by angle ϕ)

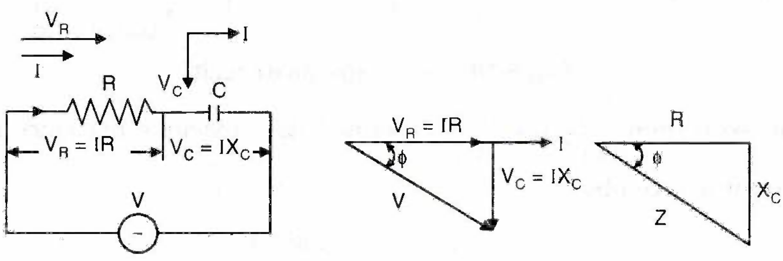
(b) I lags V by angle ϕ

(d) Impedance triangle

Fig. B.19. Resistance and inductance in series.

R-C circuit (Resistance and capacitance in series):

R-C circuit is shown in Fig. B.20.



(a) R-C circuit

(b) Phasor diagram
(I *leads* V by angle ϕ)

(c) Impedance triangle

Fig. B.20. Resistance and capacitance in series.

Important formulae :

1. *Impedance,* $Z = \sqrt{R^2+X_C^2}$ $\left(\text{where, } X_C = \frac{1}{2\pi fC}\Omega, \text{C being in farad}\right)$.

2. *Current,* $I = \frac{V}{Z}$.

3. *Power factor*, $\cos \phi = \dfrac{R}{Z}$

(or angle of *lead*, $\phi = \cos^{-1}\dfrac{R}{Z}$)

4. *Power consumed* = $VI \cos \phi \; (= I^2 R)$

R-L-C Circuit (Resistance, inductance and capacitance in series) :

Figure B.21 shows a *R-L-C* circuit.

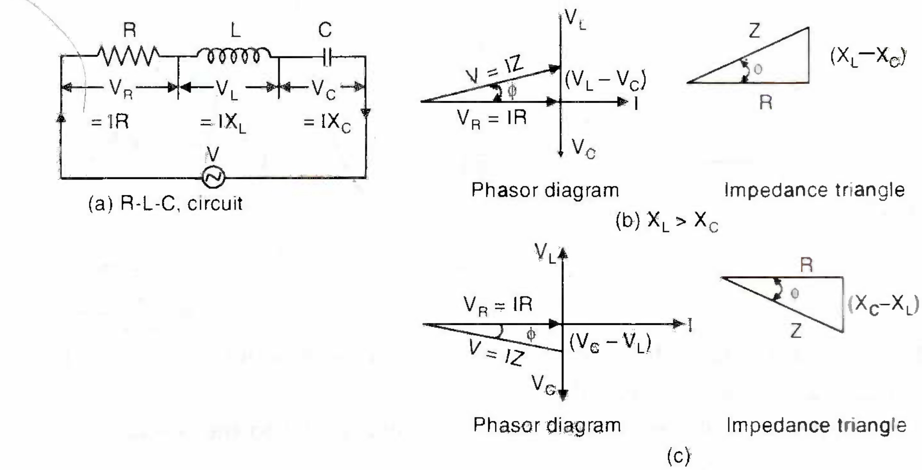


Fig. B.21. Resistance, inductance and capacitance in series.

Important formulae :

1. *Impedance*, $Z = \sqrt{R^2 + (X_L - X_C)^2}$ $\left[\begin{array}{l}\text{where } X_L = 2\pi f L, L \text{ in } henries \\ \text{and } X_L = \dfrac{1}{2\pi f C}, C \text{ in farads}\end{array}\right]$

2. *Current*, $I = \dfrac{V}{Z}$

3. *Power factor*, $\cos \phi = \dfrac{R}{Z}$

$\left[\text{angle of } lag \text{ (when } X_L > X_C) \text{ or } lead \text{ (when } X_C > X_L), \text{ f} = \cos^{?\,1}\dfrac{R}{Z}\right]$

4. *Power consumed* = $VI \cos \phi \; (= I^2 R)$

Resonance in R-L-C Circuits

Refer to Fig. B.22(*a*).

*The frequency of the voltage which gives the maximum value of the current in the circuit is called **resonant frequency**, and the circuit is said to be **resonant**.*

At resonance, $X_L = X_C$

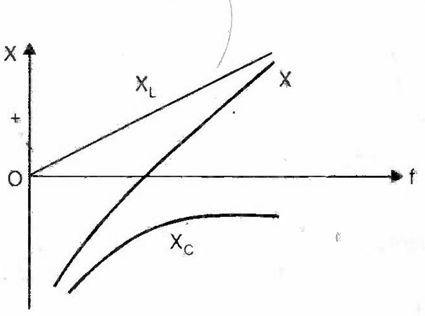
i.e., $$2\pi f_r L = \frac{L}{2\pi f_r C}$$

$$f_r = \frac{1}{2\pi \sqrt{LC}} \qquad \text{...(B.38)}$$

where,
f_r = Resonance frequency in Hz,
L = Inductance in henries, and
C = Capacitance in farads.

Figure B.22, shows variation of X_L, X_C, and X (total reactance = $X_L - X_C$) with variation of frequency f.

Figure B.23 shows the variation of current (I) with frequency (f).



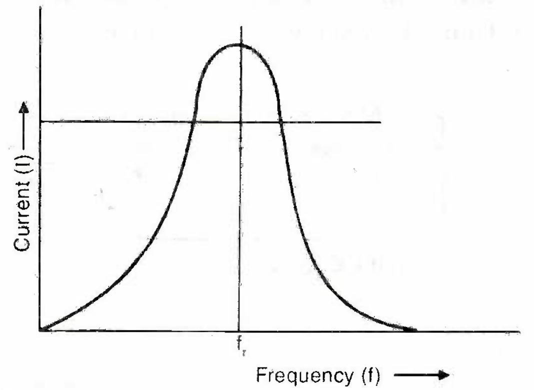


Fig. B.22. Reactance (X) v/s frequency (f). **Fig. B.23.** Current in R-L-C circuit v/s frequency.

At **series resonance**, it is seen that :

1. The impedance of the circuit is minimum and equal to the resistance (R) of the circuit $\left(i.e., I = \frac{V}{R}\right)$.
2. The current drawn is maximum ($i.e., I = I_{max}$).
3. The phase angle between the current and voltage is zero ; the *power factor is unity.*
4. The resonant frequency is given by $f_r = \frac{1}{2\pi\sqrt{LC}}$; if the *frequency is below* the resonant frequency, the net reactance in the circuit is *capacitive* and if the frequency is above the resonant frequency, the net reactance in the circuit is *inductive.*

Q-factor of a series circuit :

In the case of R.L.C. circuit it is defined as *equal to the voltage magnification in the circuit at resonance.*

$$\text{Q-factor} = \frac{1}{R}\sqrt{\frac{L}{C}} \quad \text{...(B.39)}$$

where,
R = Resistance in Ω,
L = Inductance in H, and
C = Capacitance in F.

In the case of series resonance, the higher quality factor, *i.e.*, Q factor means *not only higher voltage magnification but also a higher selectivity of the tuning coil.*

B.27 A.C. PARALLEL CIRCUITS

In A.C. parallel circuit $\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3}$

(In series A.C. circuit : $Z = Z_1 = Z_2 + Z_3$)

The term $\frac{1}{Z}$ written as Y is called the *admittance*. The unit of admittance is *mho*.

Also, $Y = \sqrt{G^2 + B^2}$

where, G = Conductance (always positive), and

and, B = Susceptance (+ve for inductive reactance and negative for capacitive reactance)

The units of conductance and susceptance are **mho.**

$$\text{Also, power factor} = \frac{B}{G}$$

B.28 RESONANCE IN PARALLEL CIRCUITS

At **parallel resonance**, it is seen that :

(*i*) The admittance of the circuit is *minimum* and is *equal* to the conductance of the circuit.

(*ii*) The *current drawn is minimum.*

(*iii*) The phase angle between the current and voltage is zero, the *power factor is unity.*

(*iv*) The resonant frequency is given by $f_r = \frac{1}{2\pi\sqrt{LC}}$ if the *resistance* in the inductance and capacitance branches is *negligible.*

B.29 COMPARISON OF SERIES AND PARALLEL RESONANT CIRCUITS

S. No.	Aspects	Series circuit (R-L-C)	Parallel circuit (R-L and C)
1.	*Impedance at resonance*	Minimum	Maximum
2.	*Current at resonance*	Maximum = $\frac{V}{R}$	Minimum = $V/(L/CR)$
3.	*Effective impedance*	R	L/CR
4.	*Power factor at resonance*	Unity	Unity
5.	*Resonant frequency*	$\frac{1}{2\pi\sqrt{LC}}$	$\frac{1}{2\pi}\sqrt{\left(\frac{1}{LC} - \frac{R^2}{L^2}\right)}$
6.	*It magnifies*	Voltage	Current
7.	*Magnification is*	$\frac{\omega L}{R}$	$\frac{\omega L}{R}$

B.30 Q-FACTOR OF A PARALLEL CIRCUIT

It is defined as the *ratio of the current circulating between its two branches to the line current drawn from the supply or simply, as the current magnification.*

$$\text{Q-factor} = \frac{1}{R}\sqrt{\frac{L}{C}} \quad \text{...(B.40)}$$

B.31 TRANSFORMERS

General Aspects

Function. *The function of a transformer, as the name implies, is to transform alternating current energy from one voltage into another voltage.* The transformer has no rotating parts, hence it is often called a *static transformer.*

When energy is transformed into a higher voltage the transformer is called a *step-up transformer* but when the case is otherwise it is called a *step-down transformer.* Most power transformers operate at constant voltage, *i.e.,* if the power varies the current varies while the voltage remains fairly constant.

Applications. A transformer performs many important functions in prominent areas of electrical engineering.

— In *electrical power engineering* the transformer makes it possible to convert electric power from a generated voltage of about 11 kV (as determined by generator design limitations) to higher values of 132 kV, 220 kV, 400 kV, 500 kV and 765 kV thus permitting transmission of huge amounts of power along long distances to appropriate distribution points at tremendous savings in the cost of transmission lines as well as in power losses.

— *At distribution points* transformers are used to reduce these high voltages to a safe level of 400/230 volts for use in homes, offices etc.

— In *electric communication circuits* transformers are used for a variety of purposes *e.g.*, as an impedance transformation device to allow maximum transfer of power from the input circuit to the output device.

— In *radio and television circuits* input transformers, interstage transformers and output transformers are widely used.

— Transformers are also used in *telephone circuits, instrumentation circuits* and *control circuits.*

Working Principle of a Transformer :

A transformer operates on the principle of *mutual inductance,* between two (and sometimes more) inductively coupled coils. It consists of two windings in close proximity as shown in Fig. B.24. *The two windings are coupled by magnetic induction.* (There is no conductive connection between the windings). One of the windings called *primary* is energised by a sinusoidal voltage. The second winding, called *secondary* feeds the load. The alternating current in the primary winding sets up an alternating flux (ϕ) in the core. The secondary winding is linked by most of this flux and e.m.fs. are induced in the two windings. The e.m.f. induced in the secondary winding drives a current through the load connected to the winding. Energy is transferred from the primary circuit to the secondary circuit through the medium of the magnetic field.

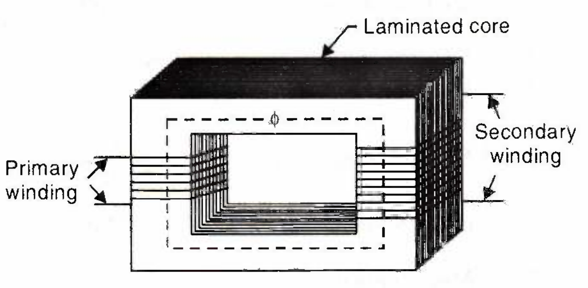


Fig. B.24. Two windings transformer.

In brief, a transformer is a device that :

(*i*) transfers electric power from one circuit to another ;

(*ii*) it does so without change of frequency; and (*iii*) it accomplishes this by electromagnetic induction (or mutual inductance).

Types of Transformers :

S. No.	Type/Kind	Applications/Uses
1.	*Power transformers*	Transmission and distribution of electric power.
2.	*Auto-transformers*	Converting voltages within relatively small limits to connect power systems of different voltages, to start A.C. motors etc.
3.	*Transformer for feeding installation with static converters.* (mercury arc rectifiers, ignitrons, semiconductor valves, etc.)	Converting A.C. into D.C. (rectifying) and converting D.C. into A.C. (inverting).
4.	*Testing transformers*	Conducting tests at high and ultra high voltages.
5.	*Power transformers for special applications.*	Furnace, welding etc.
6.	*Radio transformers*	Radio engineering etc.

Note: Distribution transformers should be designed to have *maximum efficiency* at a load *much lower than full-load* **(about 50%).**

Power transformers should be designed to have *maximum efficiency* **at or near full-load.**

Transformer Construction

All transformers have the following essential elements :

1. Two or more *electrical windings* insulated from each other and from the core (except in auto-transformers).
2. A *core,* which in case of a single-phase distribution transformers usually comprises *cold rolled silicon steel strip* instead of an assembly of punched silicon-steel laminations as are used in the large power-transformer cores. *The flux path in the assembled core is parallel to the directions of steel's grain or 'orientation'.* This results in a *reduction in core losses* for a given flux density and frequency, or *it permits the use of high core densities and reduced size of transformers for given core losses.*

Other necessary parts are:

— A *suitable container* for the assembled core and windings.

— A *suitable medium* for insulating the core and its windings from each other and from the container.

— Suitable *bushings* for insulating and bringing the terminals of the windings out the case.

The two basic types of transformer construction are :

1. **The core type.** The copper windings virtually surround the *iron core.*
2. **The shell type.** The *iron surrounds the copper windings.*

Note: The core stepping (in core type transformers) not only gives high space factor but also results in reduced length of the mean turn and the consequent I^2R loss.

Transformer Windings. The most important requirements of transformer windings are:

1. The winding should be economical both as regards initial cost, with a view to the market availability of copper, and the efficiency of the transformer in service.
2. The heating conditions of the windings should meet standard requirements, since

departure from these requirements towards allowing higher temperature will drastically shorten the service life of the transformer.

3. The winding should be mechanically stable in respect to the forces appearing when sudden short circuit of the transformer occurs.
4. The winding should have the necessary electrical strength in respect to over voltages.

The different types of winding are classified and briefly discussed below :

1. *Concentric windings :*

   (*i*) Cross-over; (*ii*) Helical; and (*iii*) Disc.
2. *Sandwich windings.*

E.M.F. Equation of a Transformer :

$$E_1 = 4.44\, f\, \phi_{max}\, N_1 \quad \text{...(B.41)}$$

$$E_2 = 4.44\, f\, \phi_{max}\, N_2 \quad \text{...(B.42)}$$

In ideal transformer on no-load

$$V_1 = E_1 \text{ and } V_2 = E_2$$

Voltage Transformation Ratio (K)

It is defined as the ratio of the secondary voltage to primary voltage,

i.e.,

$$\frac{E_2}{E_1} = \frac{N_2}{N_1} = K \quad \text{...(B.43)}$$

— If $N_2 > N_1$, *i.e.,* $K > 1$, then transformer is called *step-up transformer.*

— If $N_2 < N_1$, *i.e.,* $K < 1$, then transformer is called *step-down transformer.*

For an *ideal transformer,*

$$\frac{I_2}{I_1} = \frac{E_1}{E_2} = \frac{N_1}{N_2} = \frac{1}{K}.$$

❑❑❑

INDEX

A

B

C

D

E

F

H

I

L

M

N

P

R

S

T

MECHANICAL ENGINEERING

HEAT AND MASS TRANSFER
(IN SI UNITS)

R.K. Rajput

CONTENTS: ● BASIC CONCEPTS ● **PART I: HEAT TRANSFER BY CONDUCTION** ● CONDUCTION ÅSTEADY - STATE ONE DIMENSION ● General Heat Conduction Equation in Spherical Coordinates, ● Heat Conduction Through Plane and Composite Walls, ● Heat Conduction Through Hollow and Composite Cylinders, ● Heat Conduction Through Hollow and Composite Spheres, ● Critical Thickness of Insulation ● Heat conduction with Internal Heat Generation ● Heat Transfer from Extended Surfaces (Fins) ● CONDUCTIONÅSTEADY-STATE TWO DIMENSIONS AND THREE DIMENSIONS ● CONDUCTION ÄUNSTEADY - STATE (TRANSIENT) **PART II: HEAT TRANSFER BY CONVECTION** ● INTRODUCTION TO HYDRODYNAMICS ● DIMEN-SIONAL ANALYSIS ● Characteristic Length or Equivalent Diameter, ● FORCED CONVECTION A. LAMINAR FLOW ● Laminar Flow over a Flat Plate ● Laminar Tube Flow ● Intro-duction ●Turbulent Tube Flow ● Empirical Correlations ● FREE CONVECTION ● Simplified Free Convection Relations for Air, ● Combined Free and Forced Convection, ● BOILING AND CONDENSATION ● Intro-duction ● Boiling Heat Transfer, ● Condensation Heat Transfer ● HEAT EXCHANGERS ● Heat Exchanger Effectiveness and Number of Transfer Units (NTU) **PART III: HEAT TRANSFER BY RADIATION** ● THERMAL RADIATIONÄBASIC RELATIONS ● RADIATION EXCHANGE BETWEEN SURFACES **PART IV: MASS TRANSFER** ● MASS TRANSFER **PART V: OBJECTIVE TYPE QUESTION BANK** ● Index

Code : 10 314 **ISBN : 81-219-1777-8**

ELEMENTS OF MECHANICAL ENGINEERING

Sadhu Singh

CONTENTS : ● Basic Concepts of Thermodynamics ● First Law of Thermodynamics ● Applications of First Law of Theromdynamics ● Second Law of Themodynamics ● Gas Power Cycles ● Mechanisms ● Simple Machines ● Mechanics of Solids ● Explanatory Notes ● Subject ● Index

Code : 10 319 **ISBN : 81-219-2646-7**

TOOLING DATA

P.H. Joshi

CONTENTS: ● Jigs and Fixtures ● Press Tools ● Cutting Tools ● Appendices

AHW 1720 **ISBN : 81-7544-172-0**

A TEXTBOOK OF FLUID MECHANICS AND HYDRAULIC MACHINES

R.K. Rajput

CONTENTS : Part I Fluid Mechanics ● Properties of Fluids ● Pressure Measurement ● Hydrostatic Forces on Surfaces ● Buoyancy and Floatation ● Fluid Kinematics ● Fluid Dynamics ● Dimensional and Model Analysis ● Flow Throuch Orifices and Mouthpieces ● Flow over Notches and Weirs ● Laminar Flow ● Turbulent Flow in Pipes ● Flow Through Pipesi Boundary Layer Theory ● Flow Around Submerged Bodiesadrag and Lift ● Compressible Flow ● Flow in Open Channels ● Part II-Hydraulic Machines ● Impact of Free Jets ● Hydraulic Turbines ● Centrifugal Pumps ● Reciprocating Pumps● Miscellaneous Hydraulic Machines ● Water Power Development ● Objective Type Test Questions ● Laborotary Practicals ● Index

Code : 10 185 **ISBN : 81-219-1666-6**

ENGLISH FOR ENGINEERING AND MANAGEMENT

(Professional Communication in English)

Sutapa Banerjee

CONTENTS: Introduction ● Communication as Sharing ● Context of Communication ● Medium of Communication ● Barriers to Communication ● **Written Communication** ● Grammar: Correction of Sentences ● Transformation of Sentences ● Word Formation ● Single Word for a Group of Words ● Fill in the Blank ● Active/Passive Voice ● Direct and Indirect Narration ● Proposal ● Report Writing F Business Correspondence ● Vocabulary ● Essays ● Comprehension ● Antonyms ● Synonyms ● Idioms ● Group discussion ● Interviews ● Conclusion

Code : 10 311 **ISBN : 81-219-2603-3**

ENGINEERING MATERIALS & METALLURGY

R. K. Rajput

CONTENTS: Review-Basic Concepts • Structure of Atoms and Molecules • Miller Indices F Imperfections (Defects) in Crystals • Grain Size Determination • Constitution of Alloys and Phase Diagrams • Heat Treatment • Ferrous and Non-ferrous Metals • Non-metallic Materials • Mechanical Properties and Testing • Highlights • Objective Type Questions • Theoretical Questions •Objective Type Test Questions (With Answers) • Index

Code : 10318 **ISBN : 81-219-2709-9**

INDUSTRIAL ENGINEERING AND PRODUCTION MANAGEMENT

Martand Telsang

CONTENTS: Part I : Industrial Engineering • Introduction to Industrial Engineering • Productivity • Work-Study • Method Study • Work Measurement • Value Engineering • Plant Location • Plant Layout • Material Handing • Job Evaluation and Merit Rating • Wages and Incentives • Ergonomics • Part II: Production Management • Introduction to Production/Operations Management • New Product Design • Demand Forecasting • Production Planning and Control • Capacity Planning • Material Requirement Planning (MRP) • Process Planning • Project Scheduling with CPM and Pert • Production Control • Inventory Control • Production Cost Concepts and Break-even Analysis • Maintenance Management • Make or Buy Decisions • Planning and Control of Batch Production • Part III: Advanced Topics in Production Management • Application of Linear Programming Technique in Production Management • TQM - Concept and Philosophy • Business Process Reengineering • Group Technology • Just in Time (JIT) Manufacturing • Operations Strategy • Materials Management • Project Management • Service Management • Product and Service Reliability • Theory of Constraints (Toc) • Advanced Manufacturing Technologies and Systems • Supplements • Additional Solved Problems • Appendices

Code : 10 197 **ISBN : 81-219-1773-5**

A TEXTBOOK OF MACHINE DRAWING

(In First Angle Projections)

R.K. Dhawan

CONTENTS: Section I: Introduction & Drawing Instruments • Layout of Drawing Sheet • Conventions • Lettering • Dimensioning • Scales • Section II: Theory of Projection & Orthographic Projection • Orthographic Reading or Interpretation of Views • Indentification of Surfaces • Missing Lines & Views • Sectional Views • Isometric Projections • Auxiliary Views • Freehand Sketching • Section III: Detail & Assembly Drawings • Limits, Fits & Machining Symbols • Rivets and Riveted Joints • Welding • Screw Threads • Fastenings • Keys Cotters and Joints • Shaft Couplings • Bearings • Brackets • Pulleys • Pipe Joints • Steam Engine Parts • I.C. Engine Parts • Valves • Gears • Cams • Jigs & Fixtures • Miscellaneous Drawings

Code : 10 148 **ISBN : 81-219-0824-8**

Win Prizes !

Attention: Students

We request you, for your frank assessment, regarding some of the aspects of the book, given as under:

10 343 MECHATRONICS
R.K. Rajput **First Edition 2007**

Please fill up the given space in neat capital letters. Add additional sheet(s) if the space provided is not sufficient, and if so required.

(i) What topic(s) of your syllabus that are important from your examination point of view are not covered in the book ?

....

....

....

....

(ii) What are the chapters and/or topics, wherein the treatment of the subject-matter is not systematic or organised or updated?

....

....

....

....

....

(iii) Have you come across misprints/mistakes/factual inaccuracies in the book? Please specify the chapters, topics and the page numbers.

....

....

....

....

....

(iv) Name top three books on the same subject (in order of your preference - 1, 2, 3) that you have found/heard better than the present book? Please specify in terms of quality (in all aspects).

1 ....

....

2 ....

....

3 ....

....

(v) Further suggestions and comments for the improvement of the book:

....

....

....

....

....

Other Details:

(i) Who recommended you the book? (Please tick in the box near the option relevant to you.)

☐ Teacher ☐ Friends ☐ Bookseller

(ii) Name of the recommending teacher, his designation and address:

....

....

....

(iii) Name and address of the bookseller you purchased the book from:

....

....

....

(iv) Name and address of your institution (Please mention the University or Board, as the case may be)

....

....

....

(v) Your name and complete postal address:

....

....

....

(vi) Write your preferences of our publications (1, 2, 3) you would like to have

....

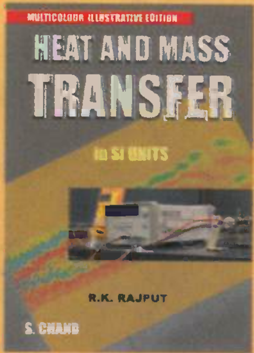
....

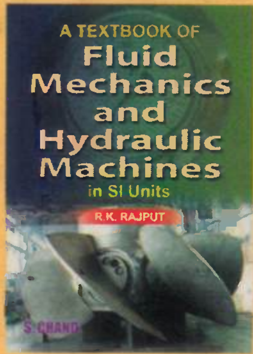
The best assessment will be awarded half-yearly. The award will be in the form of our publications, as decided by the Editorial Board, amounting to Rs. 300 (total).

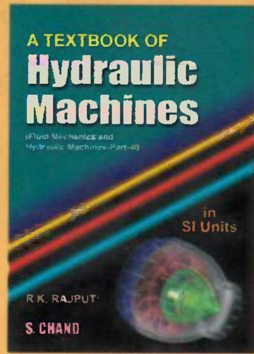
Please mail the filled up coupon at your earliest to:

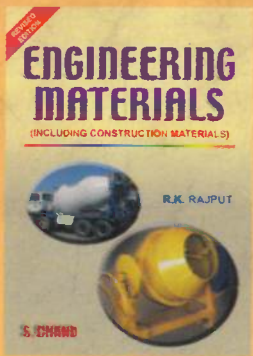
Editorial Department
S. CHAND & COMPANY LTD.,
Post Box No. 5733, Ram Nagar,
New Delhi 110 055

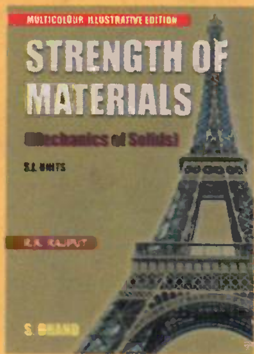
OTHER IMPORTANT BOOKS

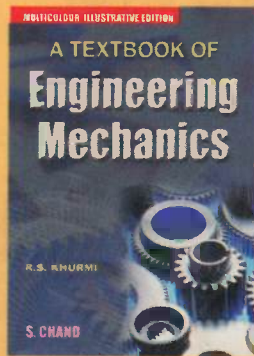


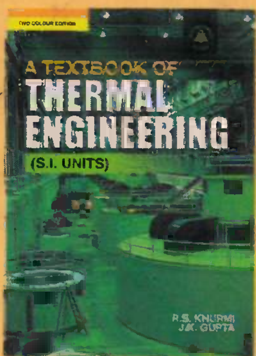


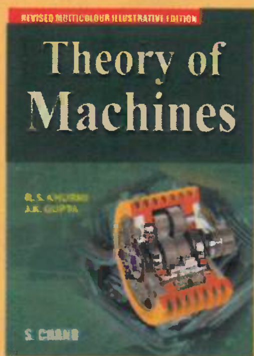


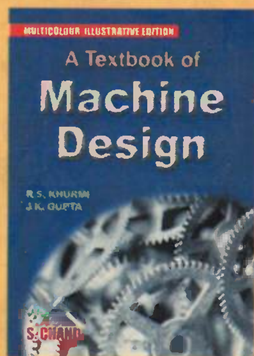












S. CHAND & COMPANY LTD.
(AN ISO 9001 : 2000 COMPANY)
7361, RAM NAGAR, NEW DELHI-110055

Online Shopping Mall:
http://www.schandgroup.com
E-mail : info@schandgroup.com

ISBN 81-219-2859-1
9 788121 928595
10 343 A
Rs. 295.00