SCHOOL OF AERONAUTICS (NEEMRANA)

FACULTY NAME : RAVIRAJ SRIKRISHNA

B.Tech AERONAUTICAL

SUBJECT CODE : EE101

SEMESTER : II

SUBJECT NAME: BASIC ELECTRICAL & ELECTRONICS

UNIT - 1

CHAPTER - 1

BASIC CONCEPTS OF ELECTRICAL ENGINEERING

NATURE OF ELECTRICITY

According to Modern electron theory of matter, all matter whether solid, liquid or gas is composed of very small particles called molecules. A molecule is in turn made up of atoms. An **atom** consists of a central part called nucleus and around the nucleus (called extra-nucleus), there are a number of **electrons** revolving in different paths or orbits. The size of the nucleus is very small as compared to the size of the atom. The nucleus contains **protons** and **neutrons.** A proton is a positively charged particle having mass 1837 times that of an electron. A neutron has the same mass as proton but no charge. Clearly, the nucleus of an atom bears a positive charge. An electron is a negatively charged particle having negative charge equal to the positive charge on a proton. Under normal conditions, the number of electrons is equal to the number of protons in an atom. Therefore, an atom is neutral as whole, the negative charge on electrons cancelling the positive charge on protons.

The above discussion shows that matter is electrical in nature i.e. it contains particles of electricity viz protons and electrons. Wheter a given body exhibits electricity (i.e. chare) or nto depends upon the relative number of these particles of electricity.

i. If the number of protons is equal to the number of electrons in a body, the resultant charge is zero and the body will be electrically neutral. Thus the paper of this book is electrically neutral (i.e. paper exhibits no charge) because it has the same number of protons and electrons.
ii. If from a neutral body, some electrons are removed, there occurs a deficit of electrons in the body. Consequently, the body attains a positive charge. Hence a positively charged body has deficit of electrons from the normal due share.
iii. If a neutral body is supplied with electrons, there occurs an excess of electrons. Consequently, the body attains a negative charge. Hence a negatively charged body has an excess of electrons from the normal due share.

Unit of Charge

The charge on an electron is so small that it is not convenient to select it as the unit of charge. In practice, coulomb is used as the unit of charge. One coulomb of charge is equal to the charge on 625×10^{16} electrons i.e.

$$1 \text{ Coulomb} = \text{Charge on } 625 \times 10^{16} \text{ electrons}$$

Thus when we say that a body has a positive charge of 1 coulomb (1 C), it mans that it has a deficit of 625×10^{16} electrons from the normal due share.

Free Electrons

We know electrons move round the nucleus of an atom in different orbits. The electrons in the last orbit are called valence electrons. In certain substances, especially metals (e.g. copper, aluminium etc), the valence electrons are so weakly attached to their nuclei that they can be easily removed or detached. Such electrons are called free electrons. It may be noted here that all valence electrons in a metal are not free electrons. It has been found that one atom of metal can provide at the msot one of free electrons in metals. For example, 1 cm^3 of copper has about 8.5×10^{22} free electrons at room temperature.

ELECTRIC CURRENT

The flow of free electrons (or charge) in a definite directions is called electric current. The flow of electric current is shown in Fig. 1.1. The copper strip has a large number of free electrons. For simplicity, only the valence orbits are shown because only the valence electrons can take apart in the flow of current. When electric pressure or voltage is applied, the free electrons being negatively charged start moving towards the positive terminal round the circuit as shown in Fig. 1.1. This directed flow of electrons is called electric current.

Conventionally, the direction of electric current is taken along the direction of motion of positive charges. When current is caused by electrons (e.g. in metals), the direction of current is opposite to the direction of electron flow.

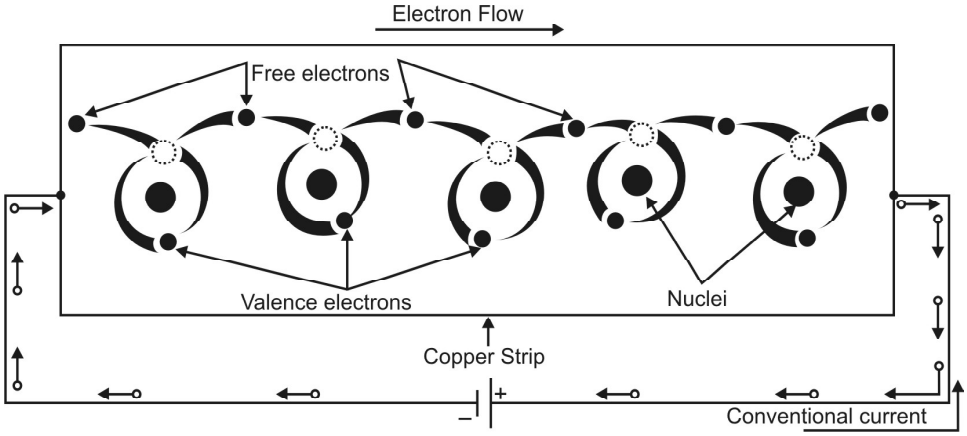


Fig. 1.1

Note: It is important to note that none of the practical consequences nor any of the results of computations performed in the study of electricity and electronics are in any way affected by the direction of current flow that one assumes. In this book, the direction of conventional current will be assumed.

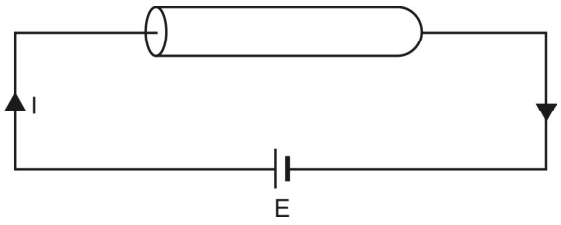


Fig. 1.2

Measurement of Current

The flow of charge in a definite direction is called electric current. It is measured by the time rate of flow of charge through the conductor. If q is the charge flowing through any cross-section of the conductor in time t, then,

Electric current, $$I = \frac{q}{t}$$

If the rate of flow of charge varies with time, then current at any time (instantaneous current) is given by;

$$i = \frac{dq}{dt}$$

where dq is the small charge passing through any cross-section of the conductor in small time dt. The SI unit of electric current is ampere. If q = 1 C and t = 1s, then I = 1/1 = 1 ampere.

One Ampere of current is said to flow through a wire if at any section one coulomb of charge flows in one second.

If n electrons are passing through any cross-section of the wire in time t, then,

$$I = \frac{q}{t} = \frac{ne}{t}$$ where $e = 1.6 \times 10^{-19}$C

Electric Potential

Just as a body raised above the ground has gravitational potential energy, similarly, a charged body has electric potential energy. When a body is charged, work is done in charging the body. This work done is stored in the body in the form of electric potential energy. The charged body has the capacity to do work by moving other charges either by attraction or repulsion. Quantitatively, electric potential is defined as under :

The **electric potential** at a point is the electric potential energy per unit charge.

Electric potential, $$V = \frac{\text{Electric potential energy}}{\text{Charge}} = \frac{W}{Q}$$

The SI unit of energy or work is 1 J and that of charge is 1 C so that SI unit of electric potential is 1 J/C which is also called 1 volt.

Thus when we say that electric potential at a point is 10V, it means that if we place a charge of 1C at that point, the charge will have electric potential energy of 10J. Similarly, if we place a charge of 2C at that point, the charge will have electric potential energy of 20J. Note that potential energy per unit charge (i.e. electric potential) is 10V.

Potential Difference

The difference in the potentials of two charged bodies is called potential difference (p.d.). Consider two bodies A and B having potentials of +5V and +3V respectively as shown in Fig. 1.3 (i). Each coulomb of charge on body A has an energy of 5 Joules while each coulomb of charge on body B has an energy of 3 Joules. Clearly, the body A is at higher potential than body B.

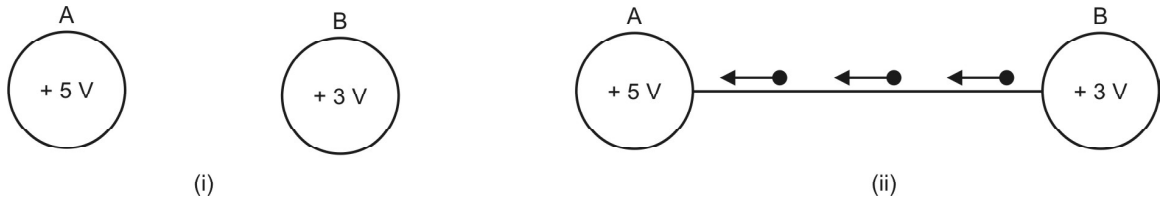


Fig. 1.3

If the two bodies are joined through a conductor [See Fig. 1.3 (ii)], then electrons will flow from body B to body A. When the two bodies attain the same potential, the flow of current stops. Therefore, we arrive at a very important conclusion that current will flow in a circuit if potential difference exists. No potential difference, no current flow. It may be noted that potential difference is sometime called voltage.

EMF, Voltage, Potential Difference

Absolute potential of a point is the work done in moving a unit positive test charge from infinity to that point. The potential difference between two points is the work done in moving a unit positive test charge from one point to the other. If we have two points A and B and it requires work to move a unti positive charge from B to A than A is said to be at higher potential with respect to B or the potential difference between points A and B is positive. When we move from point B to point A, we experience a rise in potential. Conversely, from point A to B there is a fall or potential.

The SI unit of potential difference is volt (symbol V). It is defined as the potential difference across a resistance of 1 ohm when a current of 1A is flowign in the resistance.

Since the potential difference between two points A and B may be positive or negative, it is more appropriate to write is as, say V_{AB} which means potential of point A with respect to that of point B. If A is at higher potential than B, V_{AB} is positive. Then V_{BA} (i.e. potential of point B with respect to that of A) is negative. Thus $V_{BA} = – V_{AB}$. The terms potential difference and voltage are synonymous.

The term emf (electromotive force) is also used instead of voltage. Strictly speaking emf is the total voltage of a source (e.g. a battery or a generarator). There would always be some voltage drop in the source itself and the voltage at the terminals of the source would be a bit less than the source emf. The voltage at the terminals is known as output voltage or terminal voltage.

Electric Potential or Voltage

A continuous path is needed before a continuous flow of electrons will occur. There is a need for some means to push these electrons around the circuit. With electrons, the force for movement is produced by an imbalance of electric charge. In order to move electrons along a conductor, some amount of work is required.

When electrons are present in the static condition, the energy stored is called "potential energy". This potential energy, stored in the form of an electric charge is imbalanced and capable of providing electrons to flow through a conductor, is expressed as "Voltage", which is a measure of potential energy per unit charge on electron. Voltage is the measure of work required to move a unit charge from one location to another, against the force which tries to keep electric charge balanced.

Why Current Flows in the Circuit

Let potential difference exists between two terminals of the resistance.

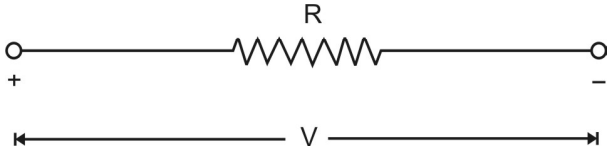


Fig. 1.4. Resistance and Potential

Due to this potential difference, one terminal is positive and other terminal is negative. Now due to positive and negative charges threre is the production of free electrons. Due to poitive and negative charges, electrons start to flow from negative side to positive side because the charge on electron is negative and negative charges repels and negative-positive charges attract each other.

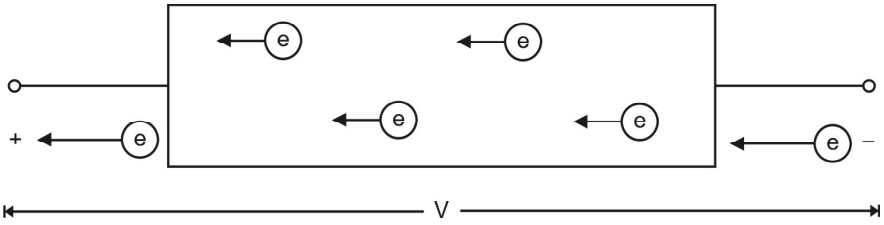


Fig. 1.5. Flow of electrons on application of voltage

As the electric current is due to the flow of electrons, the current to flow in the circuits as shown in Fig. 1.5. The direction of current is opposite to flow of electrons i.e. from positive side to negative side.

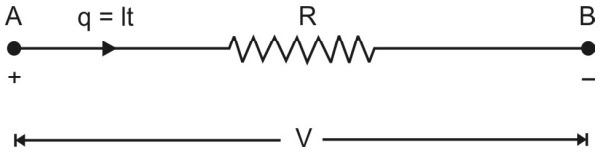


Fig. 1.6. Current through resistance

In other words, if the conductor is connected across positive and negative terminals, due to the potential difference between the terminals, the current starts to flow in conductor and direction of current is from positive terminal to negative terminal.

ELECTROMAGNETIC FORCE

The force exerted by one particle on another by virtue of electric charge on the particle is known as electromagnetic force. An electromagnetic force is a physics concept that refers to a particular force or influence, that effect charged particles. These particles may be positively or negatively charged. The electromagnetic force of interaction that exists between certain elementary particles is regarded as a force between electric charges. Electromagnetic force acts between two electrically charged particles e.g. a negatively charged electron and a positively charged prororon attracts each other with a force which is poroportional to electric charge and inversely proportional to the square of distance between them. The presence of charged particle produces an electric field and when moving it produces a magnetic field. This field manifests itself in a force between chared particles. The difference between these forces is that for a magnetic force to act the charge must be moving, but the electric force is independent of the motion.

Fig. 1.7 shows a charged particle in an electric field. If the charged q is at point 'a' where the electric field is E, the electric force F_{e1} on the charge 1 is

$$F_{e1} = q.E$$

The direction of the electrid field is parallel to the electric field if charge q is positive and antiparallel if it is negative.

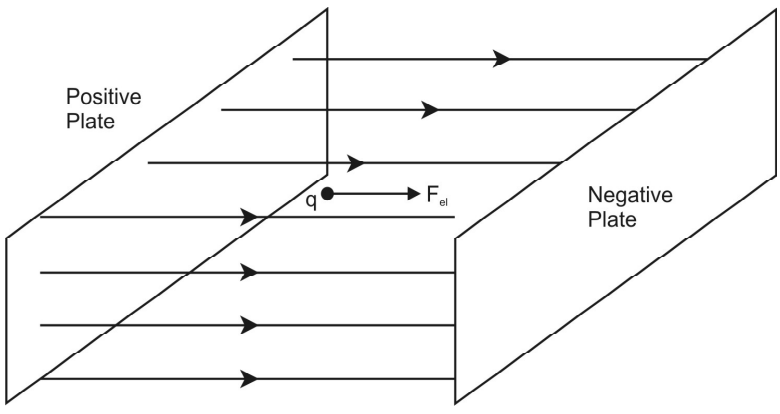


Fig. 1.7. Electric force F_{e1} on charge q in electric field

The magnetic force F_{mag} on charge q moving with speed v due to magnetic field B is

$$F_{mag} = qv \times B$$

The total electromagnetic force on charge 1, arising from the combined electric and magnetic force is:

$$F = F_{e1} + F_{mag} = q (E + v \times B)$$

ELECTRIC POWER

The power of an electric applicance is the rate at which electrical energy is converted into other forms of energy (e.g. heat, etc.). For example, a 60 W bulb converts 60J of electrical energy into heat light each second.

Thus referring to Fig. 1.8, as the charge q (= I t) moves from point A to B, it loses electric potential energy = qV. In other words, qV joules of electrical energy is converted into heat in t seconds.

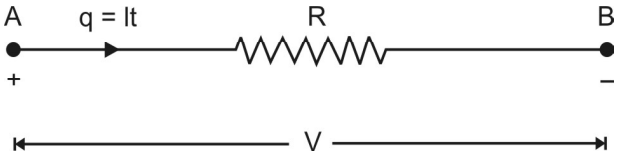


Fig. 1.8

$$\therefore \quad \text{Electric Power} = \frac{qV}{t} = \frac{(It)V}{t} = VI \text{ J/s or watts}$$

$$\therefore \quad \text{Electric Power} = V I \text{ watts} \quad ...(i)$$

$$= I^2 R \text{ watts} \quad ...(ii)$$

$$= V^2/R \text{ watts} \quad ...(iii)$$

Any one of the three formulas can be used to calculate electric power, depending upon the problem in hand.

Unit of Electric Power

$$P = V I$$

The SI unit of p.d. is 1 V and that of current is 1 A so that SI unit of power = 1 V x 1 A = 1 V A or 1 watt (1W).

Hence electric power of a circuit or device is one watt if a current of 1A flows through it when a p.d. of 1V is maintained across it.

The bigger units of electric power are kilowatt (kW) and megawatt (MW).

$$1 \text{ kW} = 1000 \text{ W} ; \qquad 1 \text{ MW} = 10^3 \text{ kW} = 10^6 \text{ W}.$$

Note. Electric appliances are rated in terms of electric power. The faster the appliances converts electrical energy into some other from of energy, the greater the electric` power it has. Thus, in 1 second, a 100 W bulb coverts more electrical energy into heat and light than a 60 W bulb.

ELECTRICAL ENERGY

The loss of electrical potential energy in maintaining current in a circuit is called electrical energy consumed in the circuit.

Thus in Fig. 1.8 above, as the charge q (= I t) moves from point A to B, it loses electric potential energy = q V = V I t joules. This loss of electric potential is converted into heat.

We say that electrical energy consumed in t second is VIt joules.

$$\therefore \quad \text{Electrical energy consumed, } W = VIt = I^2 Rt = \frac{V^2}{R} t \text{ joules}$$

Unit of Electrical Energy

$$W = V I t = \text{power} \times \text{time}$$

The SI unit of power is 1W and that of time is 1s so that SI unit of electrical energy = 1W x 1s = 1Ws or 1J.

I J (or 1Ws) energy is consumed when a device (e.g., bulb, heater, etc.) converts electrical energy to other forms at a rate of 1W for a time of 1 second.

Commercial Unit. In practice, electrical energy is measured in kilowatt-hour (kWh).

1 kWh energy is consumed when a device converts electrical energy to other forms at a rate of 1 kW for a time of 1 hour.

Electrical energy in kWh = Power in kW x Time in hours.

The electricity bills are made on the basis of total electrical energy consumed by the consumer. The unit for billing of electrical energy is 1k Wh. Thus when we say that a consumer has consumed 100 units, it means that electrical energy consumption is 100 kWh.

Use of Power and Energy Formulas

It has already been discussed that electric power as well as electrical energy consumed can be expressed by three formulas. While using these formulas, the following points may be kept in mind :

i. Electric Power, $P = I^2R = \frac{V^2}{R}$ watts

Electrical energy consumed, $W = I^2Rt = \frac{V^2}{R}t$ joules

The above formulas apply only to resistors and to devices (e.g.,. electric bulb, heater, electric kettle etc.) where all electrical energy, consumed is converted into heat.

ii. Electric power, $P = V I$ watts

Electrical energy consumed, $W = V I t$ joules

Electrical Materials

The materials used in electricity and electronics can be broadly divided into three major types viz

1. Conductors 2. Semiconductors 3. Insulators

Conductors (e.g. copper, aluminium etc.) conduct current very easily while insulators (e.g. glass, mica, paper) practically conduct no current. In other words, conductors have small resistivity and insulators have high value of resistivity. The resistivity of semiconductors (e.g. germanium, silicon etc.) lies between conductors and insulators.

Conductors

i. Conductors are formed by metallic bonds. These bonds are based on a structure of positive metal ions surrounded by a cloud of electrons.

ii. Conductors have positive termperature coefficient of resistance i.e. their resistance increases with the rise in temperature and vice-versa [see Fig. 1.9].

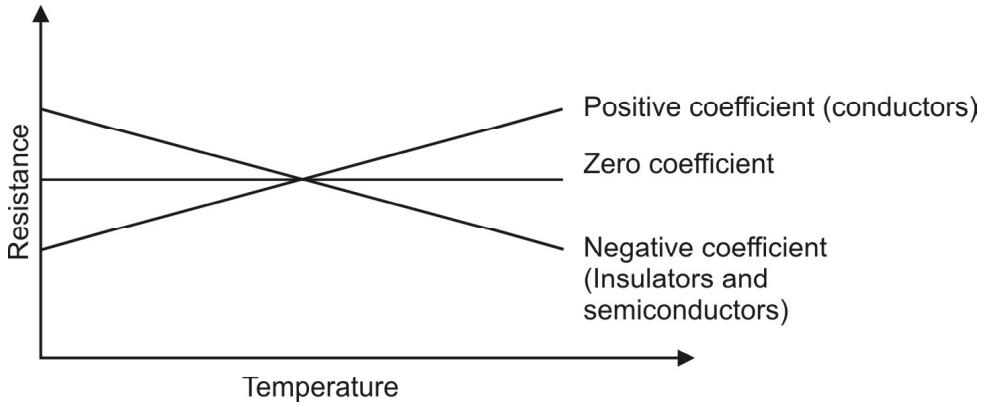


Fig. 1.9

iii. Conductors are used to carry current in electric circuits.

Power and Energy in Resistance

The resistor is an element, which dissipate the energy in the form of heat.

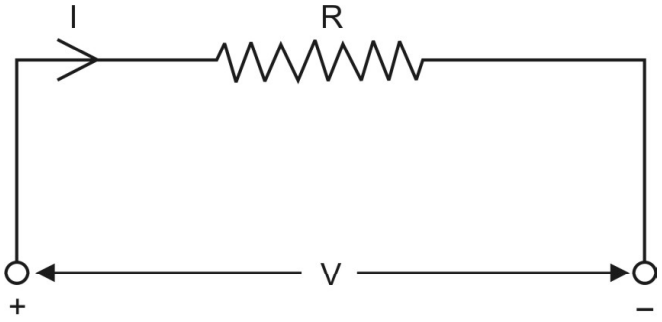


Fig. 1.10. Resistive circuit

By Ohm's law V = IR

Power at any instant $p = VI = I^2 R$ watt

Energy at any instane $e = \int p\,dt = I^2Rt$ joules

At steady state $p = VI = I^2 Rt$ joules

$E = I^2 Rt$ joules

Power and Energy in Inductor

The indicator is an element, which stores the energy by virtue of current flowing through it.

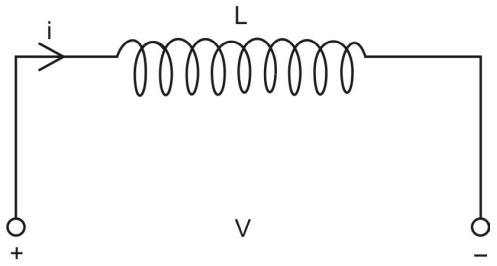


Fig. 1.11. Inductive circuit

The voltage across the capacitor is $V = \frac{1}{C}q$

The current through the capacitor is $i = C\frac{dV}{dt}$

Power in capacitor

$$p = V.i = V.C\frac{dV}{dt}$$

$$p = CV = \frac{dV}{dt}$$

Energy in capacitance

$$E = C\int V\frac{dV}{dt}.dt = C\int V\,dc$$

$$E = \frac{1}{2}CV^2 \text{ joules}$$

The energy is stored in the form of electrostatic field of capacitor and capacitor also does not dissipate any power.

Element	Voltage	Current	Power	Energy
R	I.R.	$\frac{V}{R}$	I^2R	I^2Rt
L	$L\frac{di}{dt}$	$\frac{1}{L}\int V\cdot dt$	$Li\frac{di}{dt}$	$\frac{1}{2}Li^2$
C	$\frac{1}{C}\int i\cdot dt$	$C\frac{dV}{dt}$	$CV\frac{dV}{dt}$	$\frac{1}{2}CV^2$

Table 1.1 : Properties of Circuit Elements

OHM'S LAW

This law applies to electric to electric conduction through good conductros and may be state as follows :

The ratio of potential difference (V) between any two points on a conductor to the current (I) flowing between them, is constant, provided the temperature of the conductors does not change.

In other words, $\frac{V}{I} = \text{constant}$ or $\frac{V}{I} = R$

where R is the resistance of the ocnductor between the two points considered.

Put in another way, it simply means that provided R is kept constant, current is directly proportional to the potential difference across the ends of the conductor. However, this linear relationship between V and I does not apply to all non-metallic conductors. for example, for silicon carbide, the relationship is given by $V = KI^m$ where K and m are constants and m is less than unity. It also does not apply to non-linear devices such as Zener diodes and voltage-regulator (VR) tubes.

Resistance

It may be defined as the property of a substance due to which it opposes (or restricts) the flow of electricity (i.e., electrons) through it.

Metals (as a class), acids and salts solutins are good conductors of electricity. Amongst pure metals, silver, copper and aluminium are very good conductors in the given order. This, as discussed earliear, is due to the presence of a large number of free or loosely-attached electrons in their atoms. These vagrant electrons assume a directed motion on the application of an electric potential difference. These electrons while flowing pass through the molecules or the atoms of the conductor, collide and other atoms and electrons, thereby producing heat.

Those substances which offer relatively greater difficulty or hindrance to the passage of these electrons are said to be relatively poor conductors of electricity like bakelite, mica, glass, rubber, p.v.c. (polyvinyle chloride) and dry wood etc. Amongst good insulators can be included fibrous sunstances such as paper and cotton when dry, mineral oils free from acids and water, ceramics like hard porcelain and asbestos and many other plastics besides p.v.c. It is helpful to remember that electric friction is similar to friction in Mechanics.

The Units of Resistance

The practical unit of resitance is ohm. A conductor is said to have a resistance of one ohm if it permits one ampere current to flow through it when one volt is impressed across its terminals.

For insulators whose resistances are very high, a much bigger unit is used i.e., mega-ohm = 10^6 ohm (the prefix 'mega' or mego meaning a million) or kilo-ohm = 10^3 ohm (kilo means thousand). In the case of very small resistances, smaller units like milli-ohm = 10^{-3} ohm or micro-ohm = 10^{-6} ohm are used. The symbol for ohm is Ω.

Table 1.2. Multiples and Sub-multiples of Ohm

Prefix	Its meaning	Abbreviation	Equal to
Mega-	One million	M Ω	10^6
Koli-	One thousand	k Ω	10^3
Centi-	One hundredth	–	–
Milli-	One thousandth	m Ω	10^{-3}
Micro-	One millionth	$\mu\ \Omega$	10^{-6}

Laws of Resistance

The resistance R offered by a conductor depends on the following factors:

i. It varies directly as its length, l

ii. It varies inversely as the cross-section A of the conductor.

iii. It depends on the nature of the material.

iv. It also depends on the temperature of the conductor.

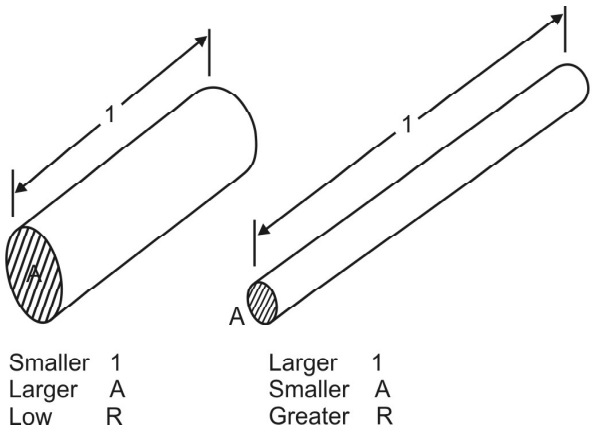


Fig. 1.12

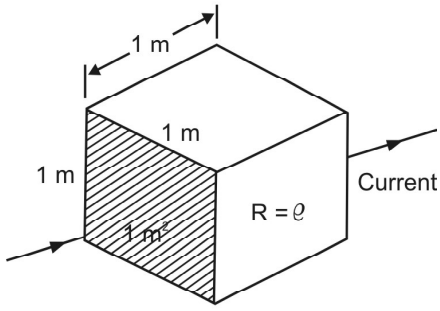


Fig. 1.13

Neglecting the last factor for the time being, we an say that

$$R \propto \frac{l}{A} \text{ or } R = \rho\frac{l}{A}$$

wehre ρ is a constant depending on the nature of the material of the conductor and is known as its specific resistance or resistivity.

If in Eq. (i), we put

l = 1 metre and A = 1 metre^2, then $R = \rho$ (Fig. 1.13)

Hence, specific resistance of a material may be defined as the resistance between the opposite faces of a metre cube of that material.

Units of Resistivity

From Eq. (i), we have $\rho = \frac{AR}{l}$

$$\rho = \frac{A\,\text{metre}^2 \times R\,\text{ohm}}{l\,\text{metre}} = \frac{AR}{l}\ \text{ohm} - \text{metre}$$

Hence, the unit of resistivity is ohm-metre (Ω-m).

Inductance

The inductance is the property of a circuit element by virtue of which it opposes the change of current through it and is capable to store electric energy in the form of magentic field. Inductance has no meaning unless the current through it changes with respect to time.

Inductors react against change in current by dropping voltage in the polarity necessary to oppose the change. When an inductor is faced with an increasing current, it have drop in voltage (load) as it absorbs energy. When an inductor is faced with a decreasing current, it acts as a source, which creates voltage as it release stored energy.

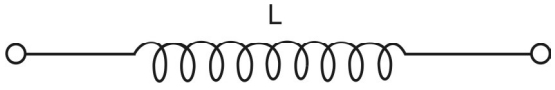


Fig. 1.14

The inductance parameter depends upon the geometry, physical dimension and property of magnetic medium. The unit of inductance is Henry (H). The ability of an inductor is to store energy in the form of magnetic field, is called inductance.

$$L = \frac{N^2 \mu A}{l}$$

When N is number of turns, l is mean length of the core, A is cross-sectional area of the acore and μ is permeability of the material of core. The inductor has no resistance. The energy is stored by the inductor in its magenetic field. Resistance parameter dissipate electrical energy but the inductance parameter does not dissipate energy but it stores the same.

Capacitance

The capacitance is the propterty of circuit element which appears only when time varying potential is applied across its terminals. A capacitor consists of two conducting surfaces separated by a layer of an insulating medium called dielectric. The use of capacitor is to store electric energy by means of electrostatic stress in dielectric.

Capacitors react against changes in voltage by supplying current in the direction necessary to oppose the change. When capacitor is faced with an increasing voltage, it acts as a load and draw current as it absorbs energy. When capacitor is faced with decreasing voltage, it acts as a source and supply current as it releases stored energy.

The capacitance is the amount of charge required to create a unit potential difference between its plates.

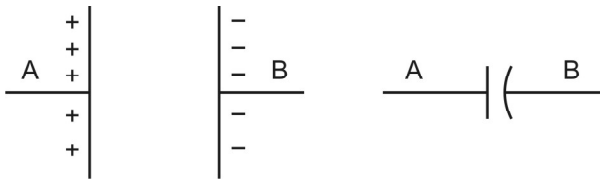


Fig. 1.15. Capacitor and its symbol

Let Q coulomb is the charge on one of two plates of capacitor and if potential difference of V volt's between the plates, then capacitance is given as

$$C = \frac{Q}{V} = \frac{\text{Charge on one plate}}{\text{Potential difference between two plates}}$$

The ability of a capacitor is to store energy in the form of an electric field, is called capacitance. The unit of capacitance is Farad (F). One Farad is defined as the capacitance which requires a charge of one coulomb to establish a potential difference of one volt between its plates.

The capacitane depends on the area, distance between the two plates and permitivity of the medium between the plates.

$$C = \frac{\in A}{d} = \text{Farad}$$

Where $\in$ is permittivity of the medium between the plates, A is area of plates and d is the distance between two plates. The capacitance is the property which delay any change of voltage across it i.e. oppose the change of voltage.

Classification of Circuit Elements

The resistor, inductor, capacitor, etc. are called circuit elements. The circuit elements are classified as follows:

Active Elements

The elements which are source of energy and always supply energy to the network are called active elements. The energy source can be supplying voltage or current. An element which can increase the power level of the circuit, is known as active element. A transistor is an active element as it can amplify the power level in the circuit. Transformer is not an active element as it can not modify the power level.

Examples : Batteries, cell, alternators, etc.

Passive Elements

The elements which either dissipate or store the energy are called passive elements. These elements have the property of absorbing/dissipating or storing energy. These are able to return the energy previously stored in them. These elements are not able to return energy more than that stored in them.

Examples : Resistance, inductance, capacitance, etc.

Linear Elements

Linear elements are elements, whose output is directly proportional to the input. Consider an element A. Let for X_1 input, output is Y_1, for X_2 input, output is Y_2. Then if the element is linear and $X_1 + X_2$ is input, then output must be $Y_1 + Y_2$. This is known as the superposition principle and the elements which follow superposition principle are called linear elements. In electrical term, the elements whose value do not change with the change in current or voltage are called linear elements.

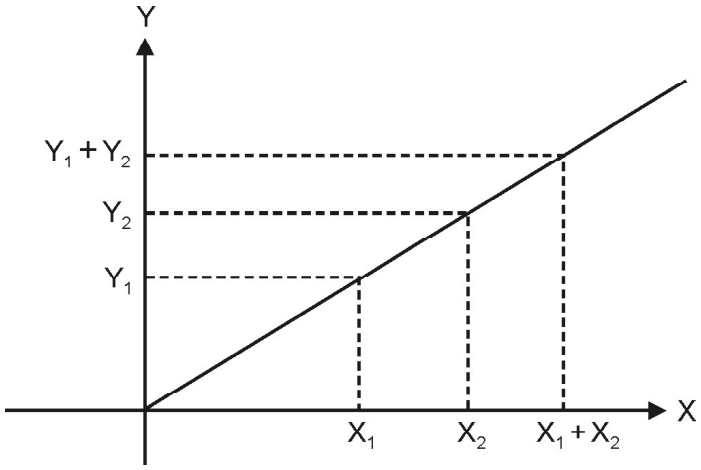


Fig. 1.16. Characteristics of Linear Parameter

Examples : Pure R, L, C, etc.

Non-linear Elements
The elements which do not follow superposition principle are non-linear elements. In the non-linear elements the output is not directly proportional to the input.

Examples : Diodes, choke, etc.

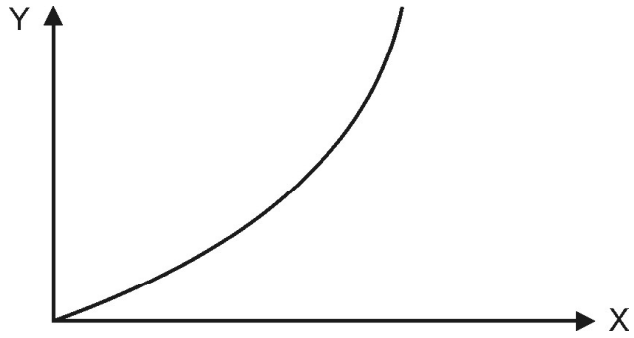


Fig. 1.17. Characteristics of Non-Linear System

Lumped Parameters
If any parameter can be concentrated (lumped) at one point irrespective of their physical size and properties, without affecting the electrical properties ofthe parameter, then it is called as lumped parameter.

Examples : Resistance, inductance, capacitance, etc.

Distributed Parameter
If the parameter cannot be assumed to be lumped at one point and it is distributed all over the circuit, is called 'Distributed Parameter'. Sometimes when we are interested in the intermediate values or point to point values of the electrical signal, then the element is said to be distributed.

Examples : Transmission line, etc.

Bilateral Elements
The elements are assumed to be bilateral, in which voltage and current relationship are same irrespective of the direction of flow of current i.e. properties of elements does not depend on direction of flow of current. The behaviour of the element is equal in either direction. The V-I curve will be same in both the direction for bilateral element.

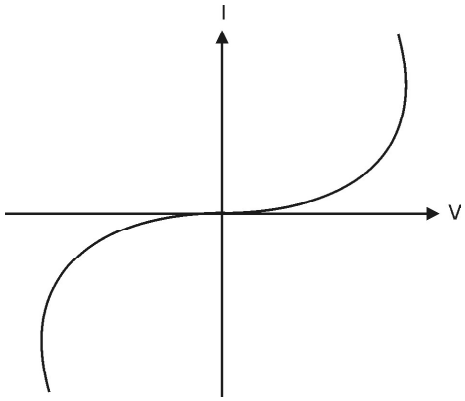


Fig. 1.18. V-I Characteristic of Bilateral Elements

Examples : Resistance (because the V/I ratio is alway R irrespective of direction of flow of current), etc.

Unilateral Elements

The elements are assumed to be unilateral, in which voltage and current relationship are not same if the direction of current is changed i.e. the voltage and current are different for two possible directions of flow of current. The behaviour of the element is unequal in either directions. The V-I curve will be different in both directions for unilateral elements (I and III quadrant).

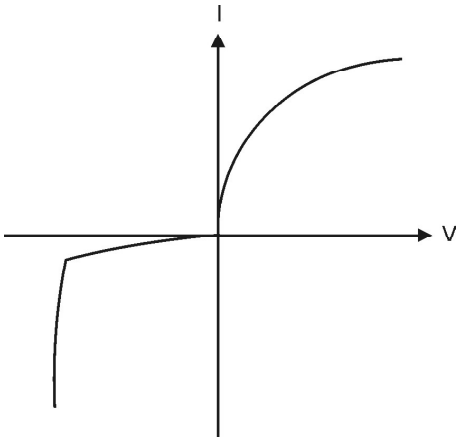


Fig. 1.19. V-I Characteristic of Unilateral Elements

Examples : Diode, etc.

Time-Invariant Elements

The elements which do not change their values with respect to tiem are called time-invariant elements. In these elements, the response with remain same to certain input irrespective of time of application of input i.e. the value of that element is same (constant) at all time.

Example : Inductance, etc.

Time-Varying Elements

The elements whose value change with respect to time are called time-varying elements. In these the response to certain input depend on time of application of input. The value of the element is different at different time.

Example : Resistance, etc.

ELECTRICAL POWER AND ENERGY

When a potential difference (V volts) is applied across a resistance, a current (I amperes) flows through it for a particular time period (t seconds). A work is said to be done for moving electrons and this work done is called electrical energy.

The total amount of work done in an electric circuit is called electrical energy.

$$V = \frac{\text{Work done}}{Q}$$

i.e. work done or electrical energy = V.Q

as $I = \frac{Q}{t}$

so work done = $V.I.t = \frac{V^2}{R}t$

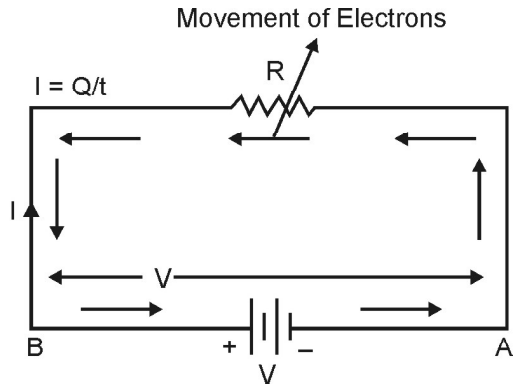


Fig. 1.20. Electrical Energy

The basic unit of electrical energy is joule or watt-second.

If V = 1 volt, I = 1 amp and t = 1 sec

electrical energy = 1 joule

The energy is said to be joule if one ampere current flows through the circuit for one second when a potential difference of 1 volt is applied across it. The other unit of electrical energy is kilowatt-hour (kW-h).

1kWh = 1000 x 60 x watt-second = 36 x 10^5 Watt-sec or joules.

DIRECT CURRENT

The current that always flows in one direction is called direct current (d.c.). The current supplied by a cell/ battery or d.c. generator is direct current. Thus in Fig. 1.21, the battery supplies direct current to the bulb. The direction of current is along ABCDA and always flows in this direction. Note that direct current means steady direct current unless stated otherwise.

D.C. Circuit

The closed path followed by direct current is called a D.C. circuit. A D.C. circuit essentially consists of a source of direct voltage (e.g. battery), the conductors used to carry current and the load. Fig. 1.21 shows a torch bulb (i.e. load) connected to a battey through conducting wires. The direct current starts from the po;sitive terminal of the battery and comes back to the starting point via the load. The direct current follows the closed path ABCDA and hence ABCDA is a D.C. circuit. The load for a D.C. circuit is usually a resistance. In a D.C. circuit, loads (i.e. resistances) may be connected in series or parallel or series-parallel.

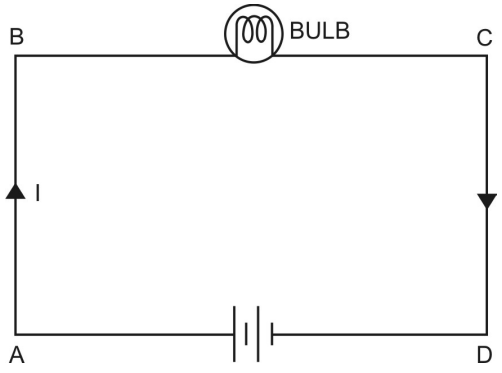


Fig. 1.21

Resistors in Series

A number of resistors are said to be connected in series if the same current flows through each resistor and there is only one path for the current flow throughout. Consider three resistors of resistances R_1, R_2 and R_3 connected in series across a battery of E volts as shown in Fig. 1.22 (i). Then total resistance R_T is given by,

$$R_T = R_1 + R_2 + R_3$$

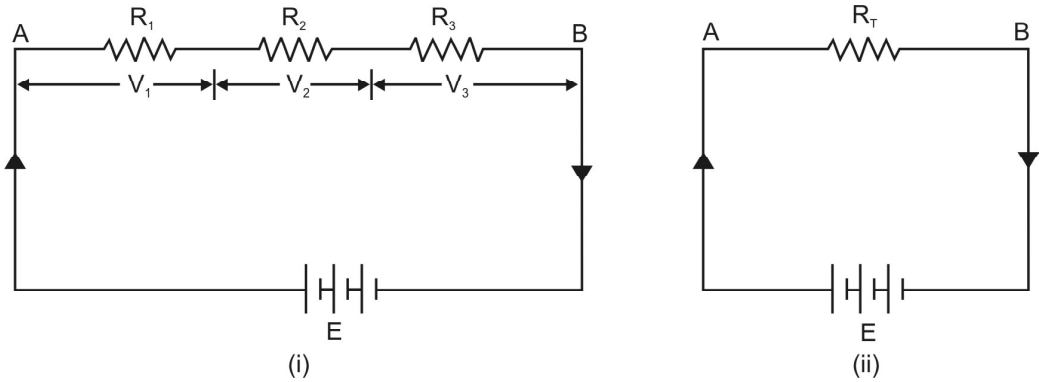


Fig. 1.22

Hence when a number of resistances are connected in series, the total or equivalent resistance is equal to the sum of the individual resistances. Thus we can replace the series connected resistors shown in Fig. 1.22 (i) by a single resistor R_T (= $R_1 + R_2 + R_3$) as shown in Fig. 1.22 (ii). This wil enable us to calculate the circuit current easily ($I = E/R_T$).

i. When resistors are connected in series, the total circuit resistance increases.

ii. $$R_T = R_1 + R_2 + R_3$$

or $$\frac{R_T}{V^2} = \frac{R_1}{V^2} + \frac{R_2}{V^2} + \frac{R_3}{V^2}$$

or $$\frac{1}{P_T} = \frac{1}{P_1} + \frac{1}{P_2} + \frac{1}{P_3}$$

where P_T is the total power dissipated by the series circuit and P_1, P_2 and P_3 are the powers dissipated by individual resistors.

Resistors in Parallel

A number of resistors are said to be connected in parallel if voltage across each resistor is the same and there are as many paths for current as the number of resistors. Consider three resistors of resistances R_1, R_2 and R_3 connected in parallel across a battery of E volts as shown in Fig. 1.23 (i). Then total resistance R_T is given by;

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

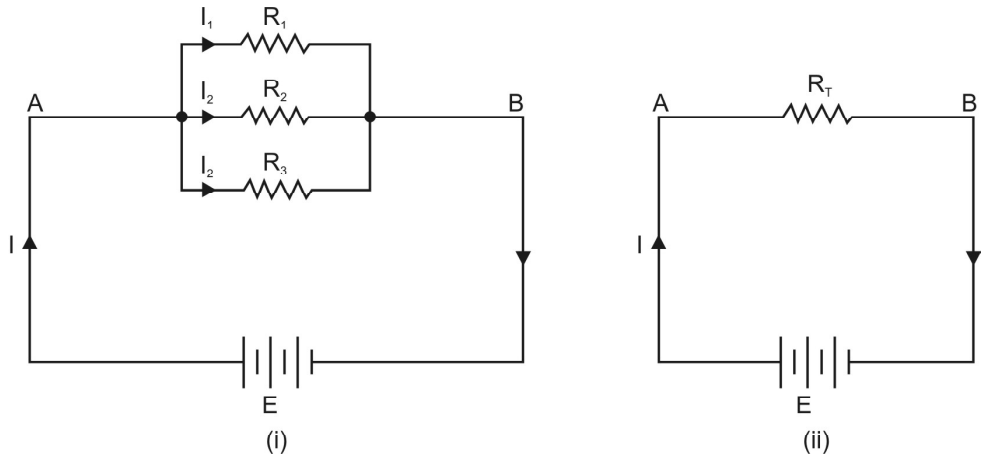


Fig. 1.23

Hence when a number of resistnaces are connected in parallel, the reciporocal of the total resistance is equal to the sum of reciprocals of individual resistances. Again, we can replace the parallel connected resistors shown in Fig. 1.23 (i) by a single resistor R_T shown in Fig. 1.23 (ii).

i. When resistors are connected in parallel, the total circuit resistance decreases.

ii. The total resistance of a parallel circuit is always less than the smallest of the resistances. For example, if three resistors of 1Ω, 3 Ω and 4 Ω are connected in parallel, the total resistance will be less than 1 Ω.

iii. If n resistors, each or resistance R, are connected in parallel, then total resistance $R_T = R/n$.

iv. $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

or $$\frac{V^2}{R_T} = \frac{V^2}{R_1} + \frac{V^2}{R_2} + \frac{V^2}{R_3}$$

or $$R_T = R_1 + R_2 + R_3$$

where P_T is the total power dissipated by the parallel circuit and P_1, P_2 and P_3 are the powers dissipated by individual resistors.

Two Resistors in Parallel

A frequent special case of parallel resistors is a circuit that contains two resistors in parallel as shown in Fig. 1.24. The total circuit current I divides into two parts; I_1 flowing through R_1 and I_2 flowing through R_2.

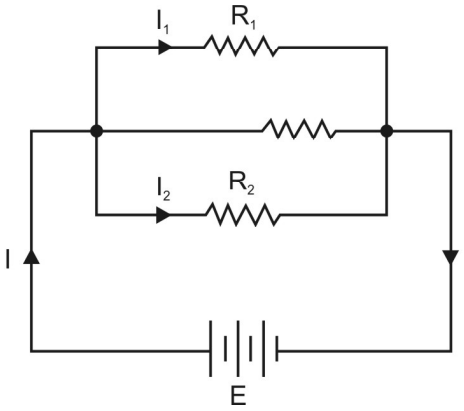


Fig. 1.24

i. **Total Resistance:** $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_2 + R_1}{R_1 R_2}$$

or $$R_T = \frac{R_1 R_2}{R_1 + R_2} \text{ i.e., } \frac{\text{Product}}{\text{Sum}}$$

Thus, if two resistances of 3Ω and 6Ω are connected in parallel, then their total or equivalent resistance R is

$$R = \frac{3 \times 6}{3+6} = \frac{18}{9} = 2\Omega$$

ii. **Branch Currents :** $$E = IR_T = I\frac{R_1 R_2}{R_1 + R_2}$$

Now $$I_1 = \frac{E}{R_1} = \left(I\frac{R_1 R_2}{R_1 + R_2}\right)\frac{1}{R_1} = I\frac{R_2}{R_1 + R_2}$$

$\therefore$ $$I_1 = I \times \frac{R_2}{R_1 + R_2}$$

Similarly, $$I_2 = I \times \frac{R_1}{R_1 + R_2}$$

i.e., current in any of the two branches = Total current x $\frac{\text{Other resistance}}{\text{Sum of the two resistances}}$

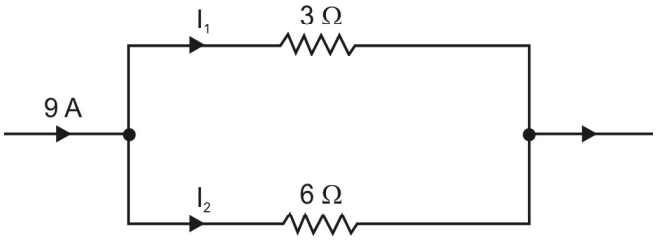


Fig. 1.25

Thus referring to Fig. 1.25, the currents in the two branches are :

$$I_1 = 9 \text{ x } \frac{6}{3+6} = 6\,A$$

$$I_2 = 9 \text{ x } \frac{3}{3+6} = 3\,A$$

Advantages of Parallel Circuits

The most useful property of a parallel circuit is the fact that potential differences has the same value between the terminals of each branch of parallel circuit. This feature of the parallel circuit offers the following advantages:

i. The appliances rated for the same voltage but different powers can be connected in parallel without disturbing each other's performance. Thus a 230V, 230W TV receiver can be operated independently in parallel with a 230V, 40W lamp.

ii. If a break occurs in any one of the branch circuits, it will have no effect on other branch cirucits.

Due to above advantages, electrical appliances in homes are connected in parallel. We can switch on or off any light or appliances without affecting other lights or appliances.

Maxwell,,s Loop Current Method

This method which is particularly well-suited to coupled circuit solutions employs a system of loop or mesh currents instead of branch currents (as in Kirchhoff's laws). Here, the currents in different meshes are assigned continuous paths so that they do not split at a junction into branch currents. This method eliminates a great deal of tedious work involved in the branch-current method and is best suited when energy sources are voltage sources rather than current sources. Basically, this method consists of writing loop voltage eqautions by Kirchhoff's voltage law in terms of unknown loop currents. As will be seen later, the number of independent equations to be solved reduces from b by Kirchhoff's laws to b– (j – 1) for the loop current method where b is the number of branches and j is the number of junctions in a given network.

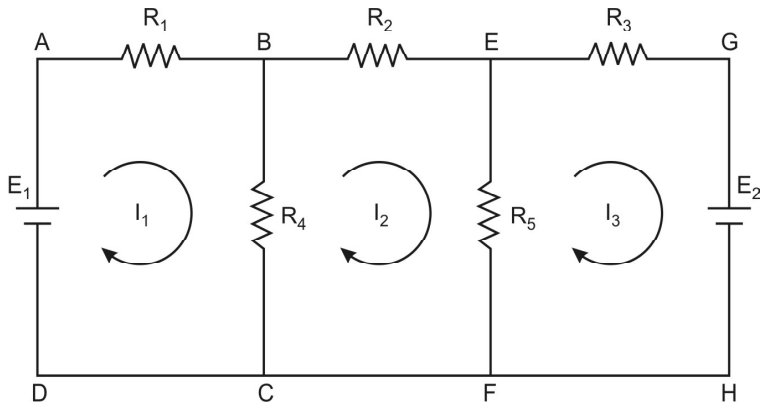


Fig. 1.26

Fig. 1.26 shows two batteries E_1 and E_2 connected in a network consisting of five resistors. Let the loop currents for the three meshes be I_1, I_2 and I_3. It is obvious that current through R_4 (when considered as a part of

the first loop) is $(I_1 - I_2)$ and that through R_S is $(I_2 - I_3)$. However, when R_4 is considered part of the second loop, current through it is $(I_2 - I_1)$. Similarly, when R_5 is considered part of the third loop, current through it is $(I_3 - I_2)$. Applying Kirchhoff's voltage law to the three loops, we get,

$E_1 - I_1R_1 - R_4 (I_1 - I_2) = 0$ or $I_1 (R_1 + R_4) - I_2 R_2 - E_1 = 0$ ...loop 1

Similarly, $- I_2R_2 - R_5)I_2 - I_3) - I_2 - I_1) = 0$

or $I_2R_4 - I_2 (R_2 + R_4 + R_5) + I_3R_5 = 0$ ... loop 2

Also $- I_3R_3 - E_2 - R_5 (I_3 - I_2) = 0$ or $I_2R_5 - I_3 (R_3 + R_5) - E_2 = 0$ ... loop 3

The above three equations can be solved not only to find loop currents but branch currents as well.

Mesh Analysis Using Matrix Form

Consider the network of Fig. 1.27, which contains resistances and independent voltage sources and has three meshes. Let the three mesh currents be designated as I_1, I_2 and I_3 and all the three may be assumed to flow in the clockwise direction for obtaining symmetry in mesh equations.

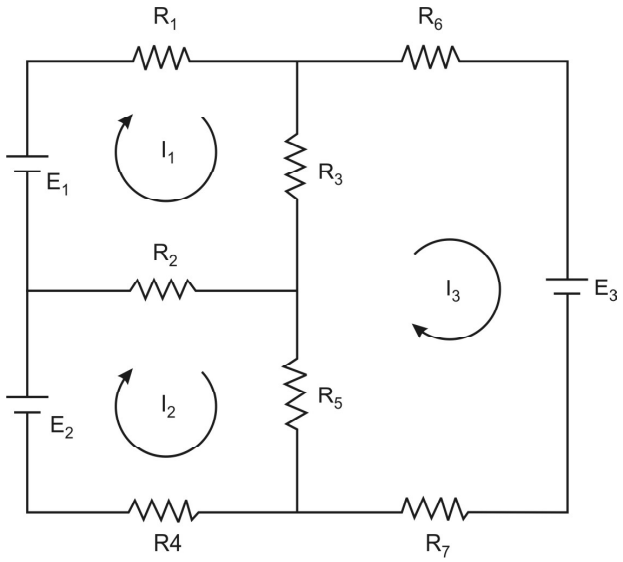


Fig. 1.27

Applying KVL to mesh (i), we have

$E_1 - I_1R_1 - R_3 (I_1 - R_2I_2 - R_3I_3 = E_1$

or $(R_1 + R_2 + R_3) I_1 - R_2I_2 - R_3I_3 = E_1$

Similarly, from mesh (ii) we have

$E_2 - R_2 (I_2 + I_1) - R_5 (I_2 - I_3) - I_2R_4 = 0$

or $R_2I_1 + (R_2 + R_4 + R_5) I_2 - R_5I_3 = E_2$

Applying KVL to mesh (iii), we have

$E_3 - I_3R_7 - R_5 (I_3 - I_2) - R_3 (I_3 - I_1) - I_3 R_6 = 0$

or $R_3I_1 - R_5I_2 + (R_3 + R_5 + R_6 + R_7) I_3 = E_3$

It should be noted that signs of different items in the above three equations have been so changed as to make the items containing self resistances positive (please see further).

The matrix equivalent of the above three equations is

$$\begin{bmatrix} +(R_1+R_2+R_3) & -R_2 & -R_3 \\ -R_2 & +(R_2+R_4+R_5) & -R_5 \\ -R_3 & R_5 & +(R_3+R_5+R_6+R_7) \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}$$

It would be seen that the first item is the firs row i.e. $(R_1 + R_2 + R_3)$ represents the self resistance of mesh (i) which equals the sum of all resistance in mesh (i). Similarly, the second item in the first row represents the

mutual resistance between meshes (i) and (ii) i.e. the sum of the resistance common to mesh (i) and (ii). Similarly, the third item in the first row represents the mutual-resistance of the mesh (i) and mesh (ii).

The item E_1, in general, represents the algebraic sum of the voltage of all the above sources acting around mesh (i). Similar is the case with E_2 and E_3. The sign of the e.m.f.'s is the same as discussed in Art. 2.3 i.e. while going along the current, if we pass from negative to the positive terminal of a battery, then its e.m.f. is taken positive. If it is the other way around, then battery e.m.f. is taken negative.

In general, let

R_{11} = self-resistance of mesh (i)

R_{22} = self-resistance of mesh (ii) i.e. sum of all resistance in mesh (ii)

R_{33} = Self-resistance of mesh (iii) i.e. sum of all resistance in mesh (iii)

$R_{12} = R_{21}$ = – [Sum of all the resistances common to meshes (i) and (ii)]

$R_{23} = R_{32}$ = – [Sum of all the resistances common to meshes (ii) and (iii)]

$R_{31} = R_{13}$ = – [Sum of all the resistances common to meshes (i) and (iii)

Using these symbols, the generalized form of the above matrix equivalent can be written as

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{31} & R_{33} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}$$

If there are m independent meshes in any liner network, then the mesh equations can be written in the matrix form as under :

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} & \ldots & R_{1m} \\ R_{21} & R_{22} & R_{23} & \ldots & R_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ R_{31} & R_{31} & R_{33} & \ldots & R_{3m} \end{bmatrix} \begin{bmatrix} I_2 \\ I_2 \\ \ldots \\ \ldots \\ I_m \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ \ldots \\ \ldots \\ E_m \end{bmatrix}$$

The above equations can be written in a more compact form as $[R_m]\,[I_m] = [E_m]$. It is known as Ohm's law in matrix form.

In the end, it may be pointed out that the directions of mesh currents can be selected arbitrarily. If we assume each mesh current to flow in the clockwise direction, then

i. All self-resistances will always be positive and (ii) all mutual resistance will always be negative. We will adapt this sign convention in the solved examples to follow.

The abovemain advantage of the generalized form of all mesh equations is that they can be easily remembered because of their symmetry. Moreover, for any given network, these can be written byinspection and then solved by the use of determinants. It eliminates the tedium of deriving simultaneous equations.

NODAL ANALYSIS WITH SOURCES

The node-equation method is based directly on Kirchhoff's current law unlike loop-current method which is based on Kirchhoff's voltage law. However, like loop current method, nodal method also has the advantage that a minimum number of equations need be written to determine the unknown quantities. Moreover, it is particularly suited for networks having many parallel circuits with common ground connected such as electronic circuits.

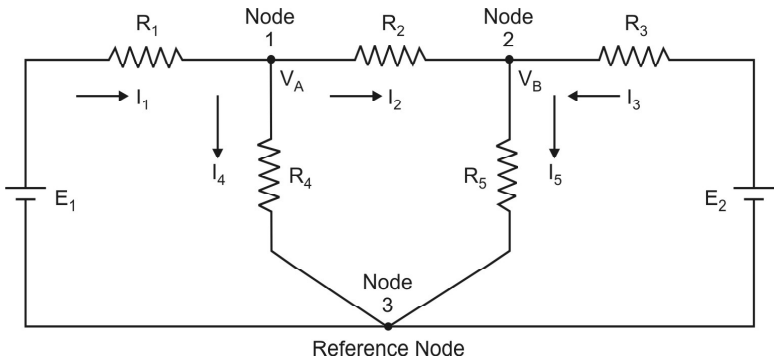


Fig. 1.28

For the application of this method, every junction in the network where three or more branches meet is regarded a node. One of these is regarded as the reference node or datum node or zero-potential node. Hence the number of simultaneous equations to be solved becomes (n – 1) where n is the number of independent nodes. These node equations often become simplified if all voltage sources are converted into current sources.

First Case

Consider the circuit of Fig. 1.28 which has three nodes. One of these i.e. node 3 has been taken in as the reference node. V_A represents the potential of node 1 with reference to the datum node 3. Similarly, V_B is the potential difference between node 2 and node 3. Let the current directions which have been chosen arbitrary be as shown.

For node 1, the following current equation can be written with the help of KCL.

$$I_1 = I_4 + I_2$$

Now $$I_1R_1 = E_1 - V_A \quad \therefore \quad I_1 = (E_1 - V_A)/R_1 \qquad ...(i)$$

Obviously, $$I_4 = V_A/R_4 \quad \text{Also, } I_2R_2 = V_A - V_B \quad (V_A > V_B)$$

$$I_2 = (V_A - V_B)/R_2$$

Substituting these values in Eq. (i) above, we get,

$$\frac{E1\text{-}V_A}{R_1} = \frac{V_A}{R_4} + \frac{V_A - V_B}{R_2}$$

Simplifying the above, we have

$$V_A\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_4}\right) - \frac{V_B}{R_2} - \frac{E_1}{R_1} = 0 \qquad ...(ii)$$

The current equation for node 2 is I5 = I2 + I3

or $$\frac{V_B}{R_5} = \frac{V_A - V_B}{R_2} + \frac{E_2 - V_B}{R_3} \qquad ...(iii)$$

or $$V_B\left(\frac{1}{R_2} + \frac{1}{R_3} + \frac{1}{R_5}\right) - \frac{V_A}{R_2} - \frac{E_2}{R_3} = 0 \qquad ...(iv)$$

Though the above nodal equations (ii) and (iii) seem to be complicated, they employ a very simple and systematic arrangement of terms which can be written simply by inspection. Eq. (ii) at node 1 is represented by

1. The product of node potential V_A and $(1/R_1 + 1/R_4)$ i.e. the sum of the reciprocals of the branch resistance connected to this node.
2. Minus the ratio of adjacent potential V_B and the interconnecting resistance R_2.
3. Minus ratio of adjacent battery (or generator) voltage E_1 and interconnecting resistance R_1.
4. All the above set to zero.

Same is the case with Eq. (iii) which applies to node 2.

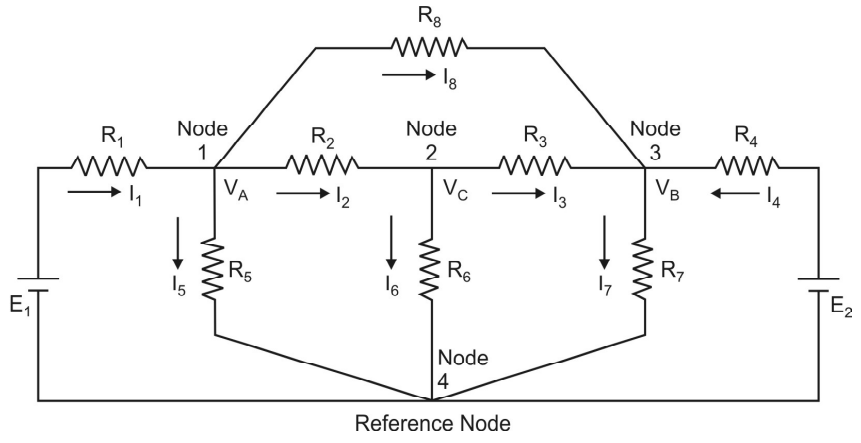


Fig. 1.29

Using conductances instead of resistances, the above two equations may be written as

$$V_A (G_1 + G_2 + G_4) - V_B G_2 - E_1 G_1 = 0$$
$$V_B (G_2 + G_3 + G_5) - V_A G_2 - E_2 G_3 = 0$$

To emphasize the procedure given above, consider the circuit of Fig. 1.29.

The three node equations are

$$V_A\left(\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_5}+\frac{1}{R_8}\right)-\frac{V_C}{R_2}-\frac{V_B}{R_8}-\frac{E_1}{R_1}=0 \quad \text{(nodel 1)}$$

$$V_C\left(\frac{1}{R_1}+\frac{1}{R_3}+\frac{1}{R_6}\right)-\frac{V_A}{R_2}-\frac{V_B}{R_3}=0 \quad \text{(nodel 2)}$$

$$V_B\left(\frac{1}{R_3}+\frac{1}{R_4}+\frac{1}{R_7}+\frac{1}{R_8}\right)-\frac{V_C}{R_3}-\frac{V_A}{R_8}-\frac{E_4}{R_4}=0 \quad \text{(nodel 3)}$$

After finding different node voltages, various currents can be calculated by using Ohm's law.

Second Case

Now, consider the case when a third battery of e.m.f. E_3 is connected between nodes 1 and 2 as shown in Fig. 1.30.

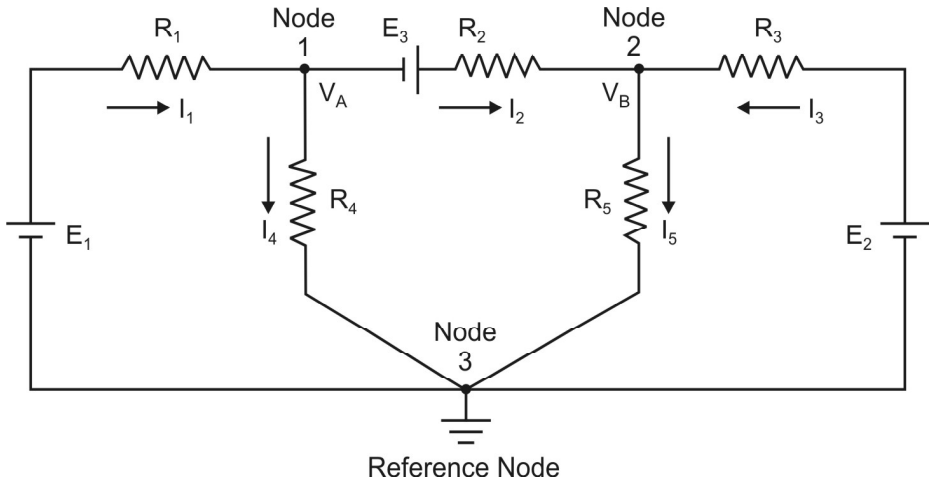


Fig. 1.30

It must be noted that as we travel from node 1 to node 2, we go from the –ve terminal of E_3 to its +ve terminal. Hence, according to the sign convention given in Art.2.3, E_3 must be taken as positive. However, if we travel from node 2 to node 1, we go from the +ve to the –ve terminal of E_3. Hence, when viewed from node 2, E_3 is taken negative.

For Node 1

$$I_1 - I_4 - I_2 = 0 \text{ or } I_1 = I_4 + I_1 \text{ – as per KCL}$$

Now,
$$I_1=\frac{E_1-V_A}{R_1};\ I_2=\frac{V_A+E_3-V_B}{R_2};\ I_4=\frac{V_A}{R_4}$$

$\therefore$
$$\frac{E_1-V_A}{R_1}=\frac{V_A}{R_4}+\frac{V_A+E_3-V_B}{R_2}$$

or
$$V_A\left(\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3}\right)-\frac{E_1}{R_1}-\frac{V_B}{R_2}+\frac{E_2}{R_2}=0$$

It is exactly the same expression as given under the First Case discussed above except for the additional term involving E_3. This additional term is taken as $+E_3/R_2$ (and not as $-E_3/R_2$) because this third battery is so connected that when viewed from mode1, it represents a rise in voltage. Had it been connected the other way around, the additional term would have been taken as $-E_3/R_2$.

For Node 2

$$I_2 + I_3 - I_5 = 0 \quad \text{or} \quad I_2 + I_3 = I_5 \quad \text{– as per KCL}$$

Now, as before, $$I_2 = \frac{V_A + E_3 - V_B}{R_2}, \; I_3 = \frac{E_2 - V_B}{R_3} = \frac{V_B}{R_5}$$

$\therefore$ $$\frac{V_A + E_3 - V_B}{R_2} + \frac{E_2 - V_B}{R_3} = \frac{V_B}{R_5}$$

On simplifying, we get $$V_B\left(\frac{1}{R_2} + \frac{1}{R_3} + \frac{1}{R_5}\right) - \frac{E_2}{R_3} - \frac{V_A}{R_2} + \frac{E_3}{R_2} = 0$$

As seen, the additional terms is $-E_3/R_2$ (and not $+E_3/R_2$) because as viewed from this node, E_3 represents a fall in potential.

It is worth repeating that the additional term in the above Eq. (i) and (ii) can be either $+E_3/R_2$ or $-E_3/R_2$ depending on whether it represents a rise or fall of potential when viewed from the node under consideration.

SOURCE CONVERSION

A given voltage with a series resistnace can be converted into (or replaced by) an equivalent current source with a parallel resistance. Conversely, a current source with a parallel resistance can be converted into a voltage source with a series resistance. Suppose, we want to covnert the voltage source of Fig. 1.31 (a) into an equivalent current source. First, we will find the value of current supplied by the source when a 'short' is put across in terminals A and B as shown in Fig. 1.31 (b). This current is I = V/R.

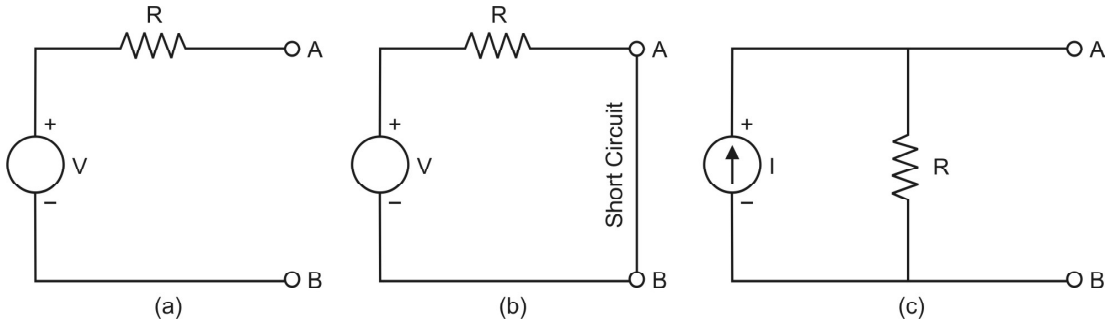


Fig. 1.31

A current source supplying this current I and having the same resistance R connected in parallel with it represents the equivalent source. It is shown in Fig. 1.31 (c). Similarly, a current source of I and a parallel resistance R can be converted into a voltage source of voltage V = IR and a resistance R in series with it. It should be kept in mind that a voltage source-series resistance combination is equivalent to (or replaceable by) a current source-parallel resistance combination if, and only if their

1. respetive open-circuit voltages are equal, and
2. respective short-circuit currents are equal.

For example, in Fig. 1.31(a), voltage across terminals A and B when they are open (i.e. open-circuit voltage V_{OC}) is V itself because there is no drop across R. Short-circuit current across AB = I = V/R.

Now, take the circuit of Fig. 1.31(c). the open-circuit voltage across AB = drop across R = IR = V. If a short is placed across AB, whole of I passes through it because R is completely shorted out.

Ideal Constant-Current Source

It is that voltage whose internal resistance is infinity. In practice, it is approachaed by a source which posses very high resistance as compared to that of the external load resistance. Let the 6-V battery or voltage source have an internal resisance of 1MΩ and let the load resistance vary from 20K to 200K. The current supplied by the source varies from 6.1/1.02 = 5.7 μA to 6/1.2 = 5 μA. As seen, even when load resistance increases 10 times current decreases by 0.9 μA. Hence, the source can be considered, for all practical purposes, to be a constant-current source.

Superposition Theorem

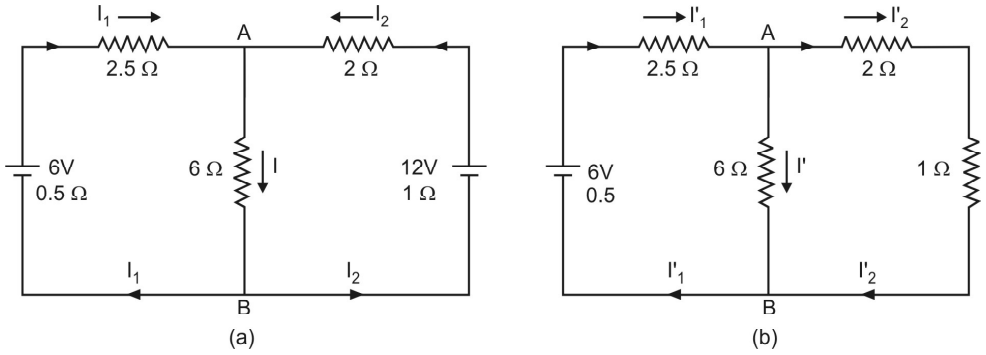


Fig. 1.32

According to this theorem if there are a number of e.m.fs. acting simultaneously in any linear bilateral network, then each e.m.f. acts independently of the others i.e. as if the other e.m.fs. did not exist. The value of current in any conductor is the algebraic sum of the currents due to each e.m.f. Similarly, voltage across any conductor isthe algebraic sum of the voltages which each e.m.f. would have produce while acting singly. In other words, current in or voltage across, any conductor of the network is obtained by sumperimposing the currents and voltages due to each e.m.f. in the network. It is important to keep in mind that this theorem is applicable only to linear networks where current is linearly related to voltage as per Ohm's law.

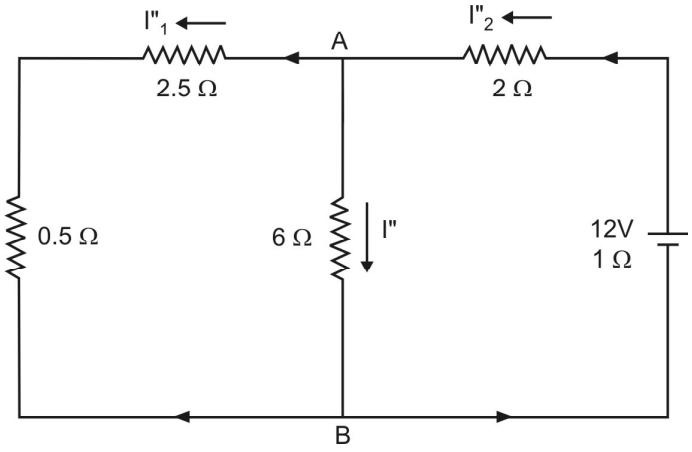


Fig. 1.33

Hence, this theorem may be stated as follows:

In a network of linear resistances containing more than one generator (or source of e.m.f.), the current which flow at any point is the sum of all the currents which would flow at that point if each generator where considered separately and all the other generators replaced for the time being byresistances equal to their internal resistances.

Explanation

In Fig. 1.32 (a) I_1, I_2 and I_3 represent the values of currents which are due to the simultaneous action of the two sources of e.m.f. in the network. In Fig. 1.32 (b) are shown the current values which would have been obtained if left-hand side battery had acted alone. Similarly, Fig. 1.33 represents conditions obtained when right-hand side battery acts alone. By combining the current values of Fig. 1.32 (b) and 1.33 the actual values of Fig. 1.32 (a) can be ontained.

Obviously, $I_1 = I_1' - I_1''$, $I_2 = I_2'' - I_2'$, $I = I' + I''$

THEVENIN THEOREM

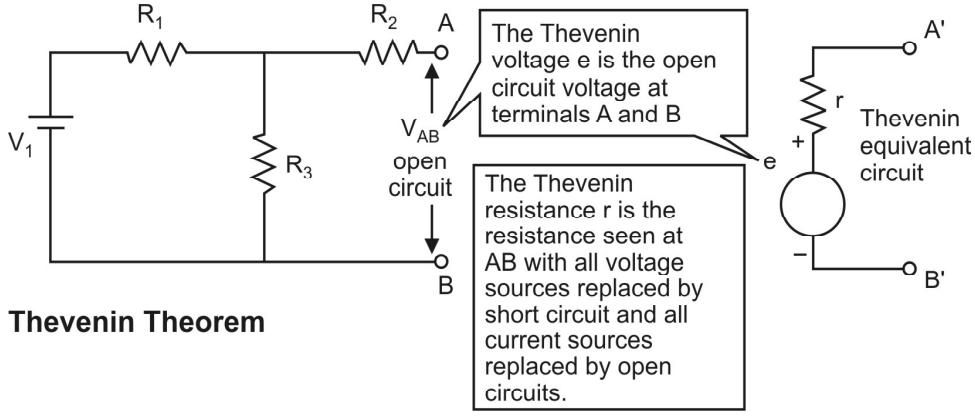


Fig. 1.34

It provides a mathematical technique for replacing a given network, as viewed from two output terminals, by a single voltage source with a series resistance. It makes the solution of complicated networks (particularly, electronic networks) quite quick and easy. The application of this extremely useful theorem will be explained with the help of the following simple example.

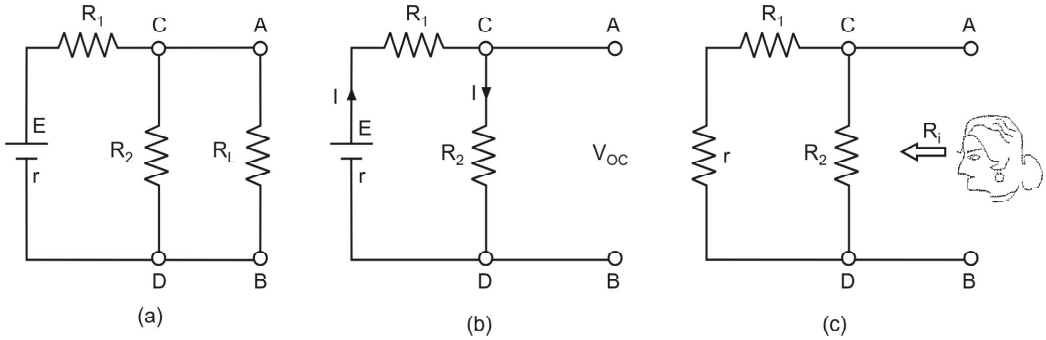


Fig. 1.35

Suppose, it is required to find current flowing through load resistance R_L, as shown in Fig. 1.35 (a). We will proceed as under :

1. Remove R_L from the circuit terminals A and B and redraw the circuit as shown in Fig. 1.35(b). Obviously, the terminals have become open-circuited.
2. Calculate the open-circuit voltage V_{OC} which appears across terminals A and B when they are open i.e. when R_L is removed.
   As seen, V_{OC} = drop across R_2 = IR_2 where I is the circuit current when A and B are open.

   $$I = \frac{E}{R_1 + R_2 + r} \quad \therefore \quad V_{oc} = IR_2 = \frac{ER_2}{R_1 + R_2 + r}$$ [r is the terminal resitance or battery]

   It is also called 'Thevenin voltage ' V_{th}.
3. Now, imaging the battery to be removed from the circuit, leaving its internal resistance r behind and redraw the circuit, as shown in Fig. 1.35 (c). When viewed inwards from terminals A and B, the circuit consists of two parallel paths : one containing R_2 and the other containing (R_1 + r). The equivalent resistance of the network, as viewed from these terminals is given as

   $$R = R_2 \,||\, (R_1 + r) = \frac{R_2(R_1 + r)}{R_2 + (R_1 + r)}$$

   This resistance is also called,* Thevenin resistance R_{sh} (though, it is also sometimes written as R_i or R_0).
   Consequently, as viewed from terminals A and B, the whole network (excluding R_1) can be reduced to a single source (called Thevenin's source) whose e.m.f. equals V_{∞} (or V_{sh}) and whose internal resistance equals R_{sh} (or R_i) as shown in Fig. 1.36.

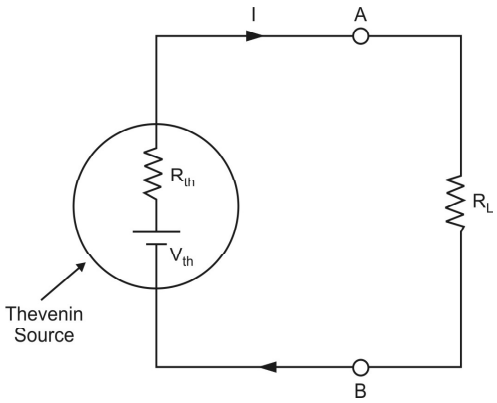


Fig. 1.36

4. R_L is now connected back across terminals A and B from where it was temporarily removed earlier. Current flowing through RL is given by

$$I = \frac{V_{th}}{R_{th} + R_L}$$

It is clear from above that any network of resistors and voltage sources (and current sources as well) when viewed from any points A and B in the network, can be replaced by a single voltage source and a single resistance** in series with the voltage source.

After this replacement of the network by a single voltage source with a series resistance has been accomplished, it is easy to find current in any load resistance joined across terminals A and B. This theorem is valid even for those linear networks which have a nonlinear load.

Hence, Thevenin's theorem, as applied to d.c. circuits, may be stated as under :

The current flowing through a load resitance R_L connected across any two terminals A and B of a linear, avtive bilateral network is given by $V_{OC} \| (R_i + R_L)$ where V_{OC} is the open-circuit voltage (i.e. voltage across the two terminals when R_L is removed) and R_i is the internal resistance of the network as viewed back into the open-circuited network from terminals A and B with all voltage source replaced by their internal resistance (if any) and current sources by infinite resistance.

How to Thevenize a Given Circuit

1. Temporarily remove the resistance (called load resistance R_L) whose current is required.
2. Find the open-circuit voltage V_{OC} which appears across the two terminals from where resistance has been removed. It is also called Thevenin voltage V_{th}.
3. Compute the resistance of the whose network as looked into from these two terminals after all voltage sources have been removed leaving behind their internal resistances (if any) and current sources have been replaced by open-circuit i.e. infinite resisance. It is also called Thevenin resistance R_{th} or T_i.
4. Replace the entire network by a single Thevenin source, whose voltage is V_{th} or V_{OC} and whose internal resistance is R_{th} or R_i.
5. Connect R_L back to its terminals from where it was previously removed.
6. Finally, calculate the current flowing through R_L by using the equation,

$$I = V_{th}/(R_{th} + R_L) \quad \text{or} \quad I = V_{OC}/(R_i + R_L)$$

NORTON'S THEOREM

This theorem is an alternative to the Thevenin's theorem. In fact, it is the dual of Thevenin's theorem. Whereas Thevenin's theorem reduces a two-terminal active network of linear resistances and generators to an equivalent constant-voltage source and series resistance, Norton's theorem replaces the network by an equivalent constant-current source and a parallel resistance.

This theorem may be stated as follows:

i. Any two-terminal active network containing voltage sources and resistance when viewed from its output terminals, is equivalent to a constant-current source and a parallel resistance. The constant current is equal to the current which would flow in tha short-circuit placed acrossthe terminals and parallel resistance is the resistance of the network when viewed from these open-circuited terminals after all voltage and current sources have been removed and replaced by their internal resistances.

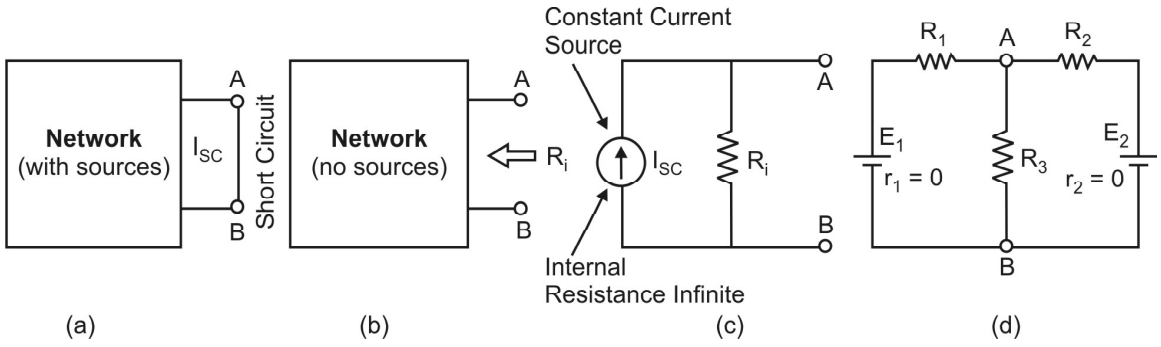


Fig. 1.37

Exaplanation

As seen from Fig. 1.37(a), a short is placed across the terminals A and B of the network with all its energy sources present. The short-circuit current I_{SC} gives the value of constant-current source.

For finding R_i, all sources have been removed as shown in Fig. 1.37(b). The resistance of the network when looked into from terminals A and B gives R_i.

The Norton's equivalent circuit is shown in Fig. 1.37(c). It consists of an ideal constant-current source of infinite internal resistance having a resistance of R_i connected in parallel with it.

ii. Another useful generalized form of this theorem is as follows :

The voltage between any two points in a network is equal to I_{SC}. R_i where I_{SC} isthe short-circuit current between the two points and R_i is the resistance of the network as viewed from these points will all voltage sources being replaced by their internal resistances (if any) and current sources replaced by open-circuits.

Suppose, it is required to findthe voltage across resistance R_3 and hence current through it [Fig. 1.37(d)]. If short-circuit is placed between A and B, then current in it due to battery of e.m.f. E_1 is E_1/R_1 and due to the other battery is E_2/R_2.

$$\therefore \qquad I_{SC} = \frac{E_1}{R_1} + \frac{E_2}{R_2} = E_1 G_1 + E_2 G_2$$

where G_1 and G_2 are branch conductances.

Now, the internal resistnace of the network as viewed from A and B simply consists of three resistances R_1, R_2 and R_3 connected in parallel between A and B. Please note that here load resistance R_3 has not been removed. In the first method given above, it has to be removed.

$$\therefore \qquad \frac{1}{R_i} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} = G_1 + G_2 + G_3$$

$$\therefore \qquad R_i = \frac{1}{G_1 + G_2 + G_3} \quad \therefore \quad V_{AB} = I_{SC}.R_i = \frac{E_1 G_1 + E_2 G_2}{G_1 + G_2 + G_3}$$

Current through R_2 is $I_3 = V_{AB}/R_3$.

Solved example No. 2.96 illustrates this approach.

How To Nortonize a Given Circuit?

This procedure is based on the first statement of the theorem given above.

1. Remove the resistance (if any) across the two given terminals and put a short-circuit across them.
2. Compute the short-circuit current I_{SC}.
3. Remove all voltage sources but retain their internal resisances, if any. Similarly, remove all current sources and replace them by open-circuits i.e. by infinite resistance.
4. Next, find the resistance R_1 (also called R_N) of the network as looked into from the given terminals. It is exactly the same as R_{th}.
5. The current source (I_{SC}) joined in parallel across R_i between the two terminals gives Norton's equivalent circuit.

Maximum Power Transfer Theorem

Although applicable to all branches of electrical engineering , this theorem is particularly useful for analysing communication networks. The overall efficiency of a network supplying maximum power to any branch is 50 per cent. For this reason, the application of this theorem to power transmission and distribution networks is limited because, in their case, the goal is high efficiency and not maximum power transfer.

However, in the case of electronic and communication networks, very often, the goal is either to receive or transmit maximum power (through at reduced efficiency) specially when power involved is only a few milliwatts or microwatts. Frequently, the problem of maximum power transfer is of crucial significance in the operation of transmission lines and antennas.

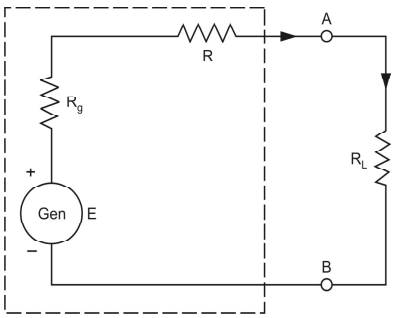


Fig. 1.38

As applied to d.c. networks, this theorem may be stated as follows:

A resistive load will abstract maximum power from a network when the load resistance is equal to the resistance of the network as viewed from the output terminals, with all energy source removed leaving behind their internal resistances.

In Fig. 1.38, a load resistance of R_L is connected across the terminals A and B of a network which consits of a generator of e.m.f. E and internal resistance R_g and a serias resistance R which, in fact, represents the lumped resistance ofthe connecting wires. Let $R_i = R_g + R$ = internal resistance of the network as viewed from a and B.

According to this theorem, R_L will abstract maximum power from the network when $R_L = R_i$.

Proof. Circuit current $$I = \frac{E}{R_L + R_i}$$

Power consumed by the load is

$$P_L = I^2 R_L = \frac{E^2 R_L}{(R_L + R_i)^2}$$

For P_L to be maximum, $\frac{dP_L}{dR_L} = 0$.

Differentiating Eq. (i) above, we have

$$\frac{dP_L}{dR_L} = E^2\left[\frac{1}{(R_L+R_i)^2} + R_L\left(\frac{-2}{(R_L+R_i)^3}\right)\right] = E^2\left[\frac{1}{(R_L+R_i)^2} - \frac{2R_L}{(R_L+R_i)^3}\right]$$

$$0 = E^2\left[\frac{1}{(R_L+R_i)^2} - \frac{2R_L}{(R_L+R_i)^3}\right] \text{ or } 2R_L = R_L + R_i \text{ or } R_L = R_i$$

It is worth noting that under these conditins, the voltage across the load is hold the open-circuit voltage at the terminals A and B.

$\therefore$ Max power is $P_{Lmax} = \frac{E^2 R_L}{4R_L^2} = \frac{E^2}{4R_L} = \frac{E^2}{4R_i}$

Let us consider an a.c. source of internal impedance $(R_1 + j X_1)$ supplying power to a load impedance $(R_L + j X_L)$. It can be proved that maximum powr transfer will take place when the modules of the load impedance is equal to the modulus of the source impedance i.e. $| Z_L | = | Z_1 |$.

Where there is a completely free choice about the load, the maximum power transfer is obtained when load impedance is the complex conjugate of the source impedance. For example, if source impedance is $(R_1 + jX_1)$, then maximum transfer power occurs, when load impedance is $(R_1 - jX_1)$. It can be shown that under this condition, the load power is = $E_2/4R_1$.

■■■

UNIT - 2

CHAPTER - 2
Alternating Currents

ALTERNATING VOLTAGE AND CURRENT

If the polarity of voltage change with time, it is know as an alternating voltage. The current that such a voltage causes to flow repeatedly changes its direction and is called alternating current.

i. **Sinusoidal alternating voltage:** The most commonly encountered type of alternating voltage varies sinusoidally with time. A sinusoidal alternating voltage can be produced by rotating a coil with constant angular ω velocity (sayrad/sec) in a uniform magnetic field. The sinusoidal alternating voltage can be expressed by the equation:

$$\upsilon = V_m \sin \omega t$$

where
ω = value of alternating voltage at time t (called instantaneous value)
V_m = maximum value of alternating voltage
ω = angular velocity of the coil

Fig. 12.1(i) shows the waveform of sinusoidal alternating voltage. Note that voltage varies from zero to a positive peak ($+V_m$), then back via zero to a negative peak (–Vm) and so on. In time period T, the wave completes one cycle.

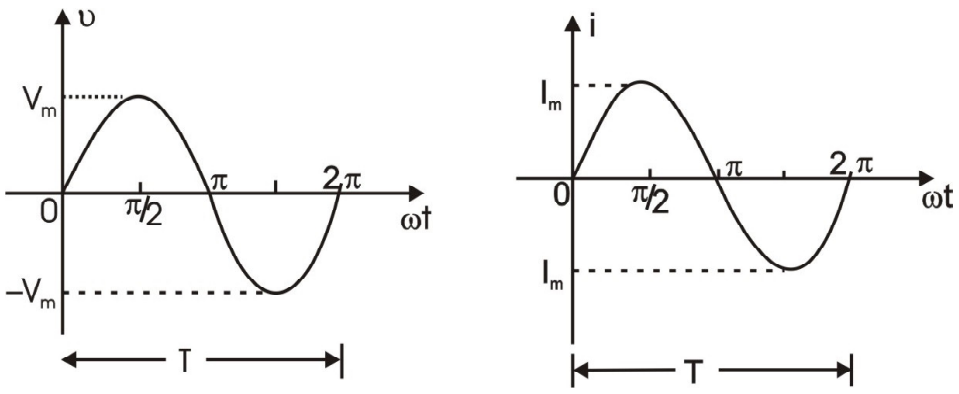


Fig. 2.1

ii. Sinusoidal Alternating current : A sinusoidal alternating voltage applied in a circuit results in a sinusoidal alternating current. Therefore a sinusoidal alternating current can be represented in the same way as voltage i.e.,

$$i = I_m \sin \omega t$$

where
i = value of alternating current at time t (called instantaneous value)
I_m = maximum value of alternating current.

Fig. 2.1(ii) shows the waveform of sinusoidal alternating current. Note that sinusoidal voltage or current not only changes direction at regular intervals but the magnitude is also changing continuously.

Note: Alternating voltage and current mean sinusoidal alternating voltage and current unless stated otherwise. Therefore, we shall omit the word 'sinusoidal' in our further discussion.

FLOW OF ALTERNATING CURRENT

Fig. 2.2 (i) shows an alternating voltage source connected to a resistor R. In Fig. 2.2(i), the upper terminal of alternating voltage source is positive and the lower terminal negative so that current flows in the circuit as shown. After time equal to T/2 (where T is the time period of alternating voltage), the polarities of the voltage source are reversed [See Fig. 2.2(ii)] so that current now flows in the opposite direction. This is called alternating current because the current flows in alternate directions in the circuit.

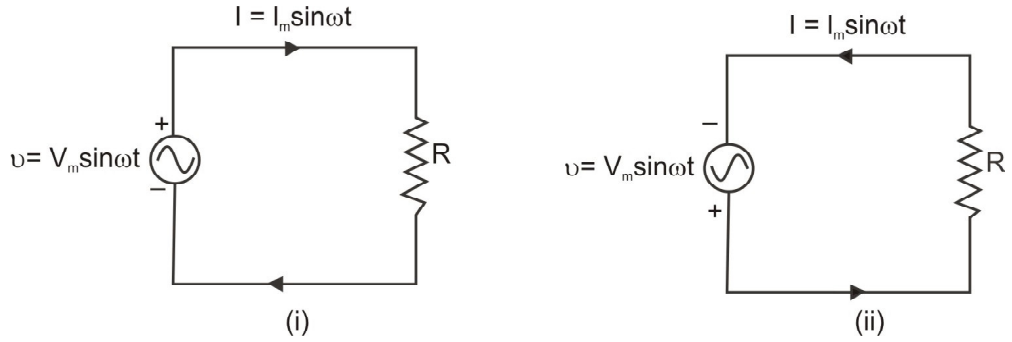


Fig. 2.2

Note that alternating current has the same frequency ($\omega = 2\mu f$) as the alternating voltage that produces it. Therefore, when an alternating voltage is applied in a circuit, the resulting current has the same frequency as that of applied voltage.

IMPORTANT A.C. TERMINOLOGY

The important a.c. terminology is defined below:

i. **Waveform:** The shape of the curve obtained by plotting the instantaneous values of voltage or current as ordinate against time as abcissa is called its waveform or waveshape. Fig.11.3 shows the waveform of an alternating voltage varying sinusoidally.

ii. **Instantaneous value**: The value of an alternating quantity at any instant is called its instantaneous value. The instantaneous values of alternating voltage and current are represented by υ and i respectively. As example, the instantaneous values of voltage (See Fig. 2.3) at 0°, 90° and 270° are 0, + V_m and –V_m respectively.

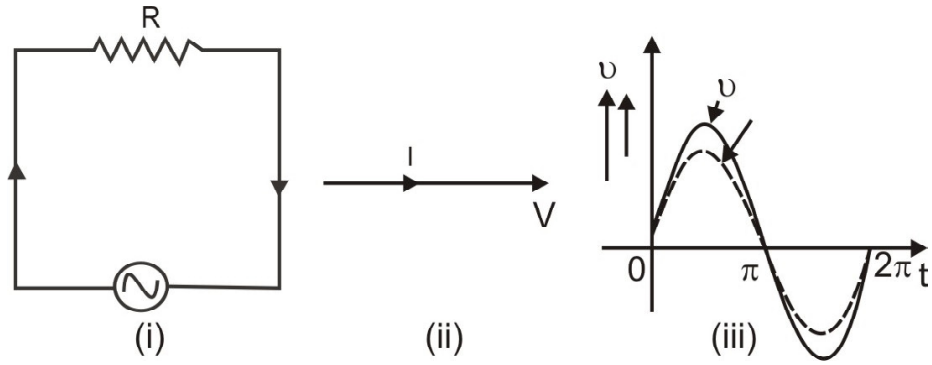


Fig. 2.3

iii. **Cycle :** One complete set of positive and negative values of an alternating quantity is known as a cycle Fig. 2.3 shows one cycle of an alternating voltage.

A cycle can also be defined in terms of angular measure. One cycle corresponds to 360° electrical or 2π radians. The voltage or current generated in a conductor will span 360°electrical (or complete one cycle) when the conductor moves past a north and south pole.

iv. **Alternation.** one-half cycle of an alternating quantity is called an alternation. An alternation spans 180° electrical. Thus in Fig. 2.3, the positive or negative half of alternating voltage is the alternation.

v. **Time period.** The time taken in seconds to complete one cycle of an alternating quantity is called its time period. It is generally represented by T.

vi. **Frequency.** The number of cycles that occur in one second is called the frequency (f) of the alternating quantity. It is measured in cycles/sec (C/s) or Hertz (Hz). One Hertz is equal to 1C/s.

The frequency of power system is low; the most common being 50 C/s or 50 hz. It means that alternating voltage or current completes 50 cycles in one second. The 50 Hz frequency is the most popular because it gives the best results when used for operating both lights and machinery.

vii. **Amplitude.** The maximum value (positive or negative) attained by an alternating quantity is called its amplitude or peak value. The amplitude of an alternating voltage or current is designed by $V_m (or\ E_m) or\ I_m$.

IMPORTANT RELATIONS

Having become familiar with a.c. terminology, we shall now establish some important relations.

i. **Time period and frequency :** Consider an attenuating quantity having a frequency of f C/s and time period T second.

Time taken to complete f cycles = 1 second

Time taken to compete 1 cycle = 1/f second

But the time taken to complete one cycle is the time period T (by definition).

$$\therefore \qquad T = \frac{1}{f} \text{ or } \qquad f = \frac{1}{T}$$

ii. **Angular velocity and frequency :** Suppose the coil is rotating with an angular velocity of ω rad/sec in a uniform magnetic field. In one revolution of the coil, the angle turned is 2π radians and the voltage wave completes 1 cycle. The time taken to complete one cycle is the time period T of the alternating voltage.

$$\therefore \qquad \text{Angular velocity, } \omega = \frac{Angle\ turned}{Time\ taken} = \frac{2\pi}{T}$$

$$\text{or} \qquad \omega = 2\pi f \qquad (\because f = 1/T)$$

iii. **Frequency and speed.** Consider a coil rotating at a speed of N r.p.m. in the field of p poles. As the coil moves past a north and south pole, one complete cycle is generated. Obviously, in one revolution of the coil, P/2 cycles will be generated.

Now, Frequency, f = No of cycles/sec

= (No. of cycles/revolution) × (No. of revolutions/sec)

$$= \left(\frac{P}{2}\right) \times \left(\frac{N}{60}\right) = \frac{PN}{120}$$

$$\therefore \qquad f = \frac{NP}{120}$$

For example, an a.c. generator having 10 poles and running at 600 r.p.m. will generate alternating voltage whose frequency is:

$$f = \frac{NP}{120} = \frac{10 \times 600}{120} = 50Hz.$$

GENERATION OF ALTERNATING VOLTAGE AND CURRENTS

Alternating voltage may be generated by rotating a coil in a magnetic field, as shown in Fig.2.4(a) or by rotating a magnetic field within a stationary coil, as shown in Fig.2.4(b).

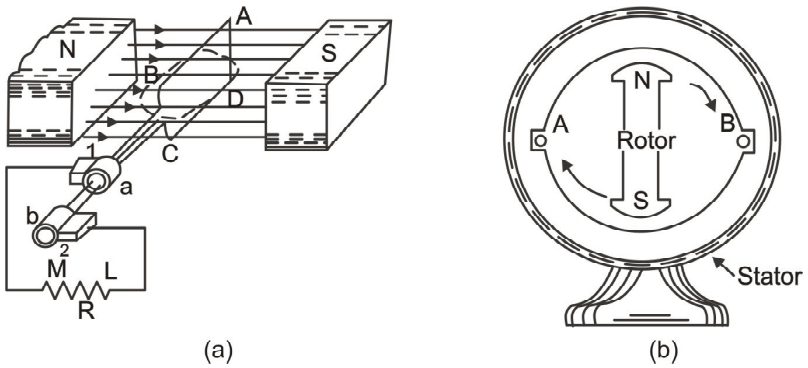


Fig. 2.4

The value of the voltage generated depends, in each case, upon the number of turns in the coil, strength of the field and the speed at which the coil or magnetic field rotates. Alternating voltage may be generated in either of the two ways shown above, but rotating-field method is the one which is mostly used in practice.

Equations of the Alternating Voltages and currents

Consider a rectangular coil, having N turns and rotating in a uniform magnetic field, with an angular velocity of ω radian/second, as shown in Fig.2.5. Let time be measured from the X-axis. Maximum flux Φ_m is linked with the coil, when its plane coincides with the X-axis. In time *t* seconds, this coil rotates through an angle $\theta = \omega t$.

In this deflected position, the component of the flux which is perpendicular to the plane of the coil, is $\Phi = \Phi_m \cos \omega t$. Hence, flux linkages of the coil at any time are $N\Phi = N\Phi_m \cos \omega t$.

According to Faraday's Laws of Electromagnetic Induction, the e.m.f. induced in the coil if given by the rate of change of flux-linkages of the coil. Hence, the value of the induced e.m.f. at this instant (i.e. when $\theta = \omega t$) or the instantaneous value of the induced e.m.f. is

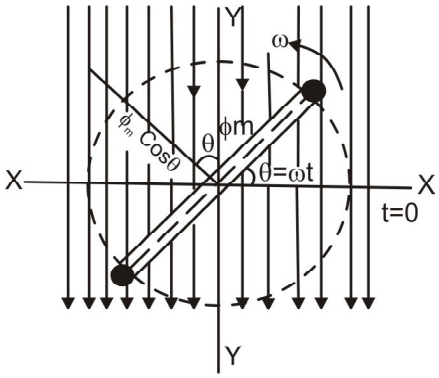


Fig. 2.5

$$e = -\frac{d}{dt}(N\Phi) \text{ volt} = -N.\frac{d}{dt}(\Phi_m \cos \omega t) \text{ volt} = -N\Phi_m \omega(-\sin \omega t) \text{ volt}$$

$$= \omega N\Phi_m \sin \omega t \text{ volt} = \omega N\Phi_m \sin \theta \text{ volt} \quad ...(i)$$

When the coil has turned through 90° i.e. when $\theta = 90°$, then sin $\theta = 1$, hence e has maximum value, says E_m. Therefore, from Eq.(i) we get

$$E_m = \omega n\Phi_m = \omega NB_m A = 2\pi fNB_m A \text{ volt} \quad ...(ii)$$

where B_m = maximum flux density in Wb/m^2; A = area of the coil in m^2

f = frequency of rotation of the coil in rev/second

Substituting this value of E_m in Eq. (i), we get $e = E_m \sin \theta = E_m \sin \omega t$ ...(iii)

Similarly, the equation of induced alternating current is $i = I_m \sin \omega t$ ...(iv)

provided the coil circuit has been closed through a resistive load.

Since $\omega = 2\pi f$, where if is the frequency of rotation of the coil, the above equations if the voltage and current can be written as

$$e = E_m \sin 2\pi ft = E_m \sin\left(\frac{2\pi}{T}\right)t \text{ and } i = I_m \sin 2\pi ft = I_m \sin\left(\frac{2\pi}{T}\right)t$$

where T = time-period of the alternating voltage or current = 1/f

It is seen that the induced e.m.f. varies as sine function of the time angle ωt and when e.m.f. is plotted against time, a curve similar to the one shown in Fig.2.6 is obtained. This curve is known as sine curve and the e.m.f. which varies in this manner is known as sinusoidal e.m.f. Such a sine curve can be conveniently drawn, as shown in Fig.2.7. A vector, equal in length to E_m is drawn. It rotates in the counter-clockwise direction with a velocity of ω radian/second, making one revolution while the generated e.m.f. makes two loops or one cycle. The projection of this vector on Y-axis gives the instantaneous value e of the induced e.m.f. i.e. $E_m \sin \omega t$.

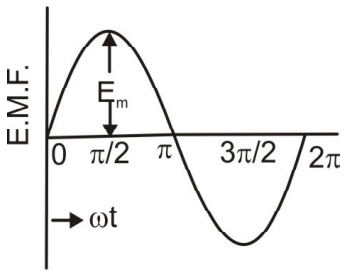


Fig. 2.6

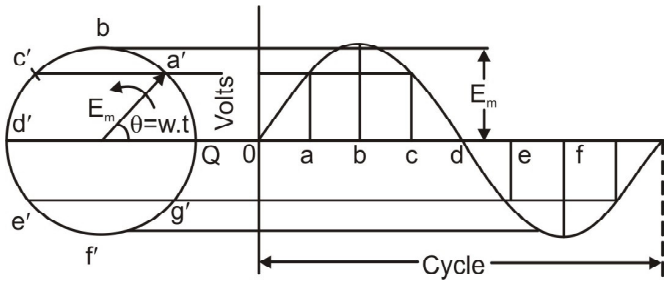


Fig. 2.7

To construct curve, lay off along X-axis equal angular distance oa, ab, bc, cd etc. corresponding to suitable angular displacement of the rotating vector. Now, erect coordinates at the points a, b, c and d etc. (Fig.2.7) and then project the free ends of the vector E_m at the corresponding position a', b', c', etc to meet these ordinates. Next draw a curve passing through these intersecting points. The curve so obtained is the graphic representation of equation (iii) above.

Alternate Method for the Equations of Alternating Voltages and Currents

In Fig.2.8 is shown a rectangular coil AC having N turns and rotating in a magnetic field of flux density B Wb/m^2. Let the length of each of its sides A and C be/meters and their peripheral velocity v metre/second. Let angle be measured from the horizontal position i.e. from the X-axis. When in horizontal position, the two sides A and C move parallel to the lines of the magnetic flux. Hence, no flux is cut and so no e.m.f. is generated in the coil.

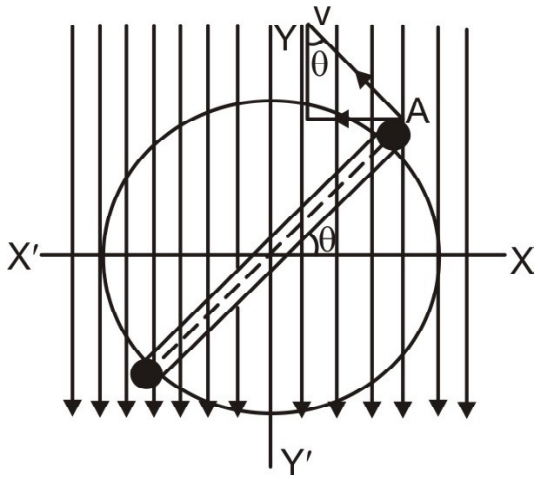


Fig. 2.8

When the coil has turned through angle θ, its velocity can be resolved into two mutually perpendicular components (i) v cos θ component-parallel to the direction of the magnetic flux and (ii) v sin θ component-perpendicular to the direction of the magnetic flux. The e.m.f. is generated due entirely to the perpendicular component i.e. v sinθ.

Hence, the e.m.f. generated in one side of the coil which contains N conductors, is given by, e = N × Bl v sin θ.

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.

It is also known as the effective or virtual value of the alternating current, the former term being used more extensively. For computing the r.m.s. value of symmetrical sinusoidal alternating currents, either mid-ordinate method or analytical method may be used, although for symmetrical but non-sinusoidal waves, the mid-ordinate method would be found more convenient.

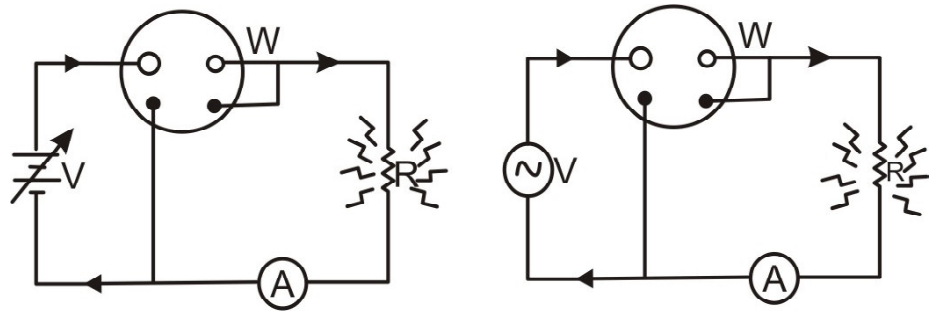


Fig. 2.9

A simple experimental arrangement for measuring the equivalent d.c. value of a sinusoidal current is shown in Fig.2.9. The two circuits have identical resistance but one is connected to battery and the other to a sinusoidal generator. Wattmeters are used to measure heat power in each circuit. The voltage applied to each circuit is so adjusted that heat power production in each circuit is the same. In that case, the direct current will equal $I_m/\sqrt{2}$ which is called r.m.s. value of the sinusoidal current.

Mid-ordinate Method

In Fig.2.10 are shown the positive half cycles for both symmetrical sinusoidal and non-sinusoidal alternating current. Divide time base 't' into n equal intervals of time each of duration t/n seconds. Let the average values of instantaneous currents during these intervals be respectively i_1, i_2, i_3, ...i_n (i.e. mid-ordinates in Fig.2.10). Suppose that this alternating current is passed through a circuit of resistance R ohms. Then,

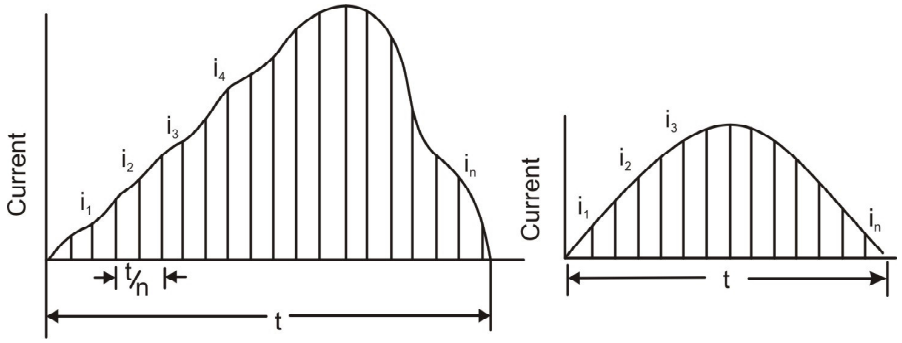


Fig. 2.10

Heat produced in 1st interval = $0.24 \times 10^{-3}\ i_1^2$ Rt/n kcal ($\because$ 1/J = 1/4200 = 0.24×10^{-3})

Heat produced in 2nd interval = $0.24 \times 10^{-3}\ i_2^2$ Rt/n kcal

: : : : : :
: : : : : :
: : : : : :

Heat produced in nth interval = $0.24 \times 10^{-3}\ i_n^2$ Rt/n kcal

Total heat produced in t seconds is = $0.24 \times 10^{-3}\ Rt\left(\frac{i_1^2 + i_2^2 + ... + i_n^2}{n}\right) kcal$

Now, suppose that a direct current of value I produces the same heat through the same resistance during the same time t. Heat produced by it is = $0.24 \times 10^{-3}\ I^2Rt$ kcal. By definition, the two amounts of heat produced should be equal.

$$\therefore\ 0.24 \times 10^{-3} I^2 Rt = 0.24 \times 10^{-3}\ Rt\left(\frac{i_1^2 + i_2^2 + ... + i_n^2}{n}\right)$$

$$\therefore \qquad I^2 = \frac{i_1^2 + i_2^2 + ... i_n^2}{n} \quad \therefore\ I = \sqrt{\left(\frac{i_1^2 + i_2^2 + ... + i_n^2}{n}\right)}$$

= squire root of the mean of the square of the instantaneous currents

Similarly, the r.m.s. value of alternating voltage is given by the expression

$$V = \sqrt{\left(\frac{v_1^2 + v_2^2 + ... v_n^2}{n}\right)}$$

Analytical Method

The standard form of a sinusoidal alternating current is $i = I_m \sin \omega t = I_m \sin \theta$.

The mean of the squares of the instantaneous values of current over one complete cycle is (even the value over half a cycle will do).

$$= \int_0^{2\pi} \frac{i^2 d\theta}{(2\pi - 0)}$$

The square root of this value is $= \sqrt{\int_0^{2\pi} \frac{i^2 d\theta}{2\pi}}$

Hence, the r.m.s. value of the alternating current is

$$I = \sqrt{\left(\int_0^{2\pi} \frac{i^2 d\theta}{2\pi}\right)} = \sqrt{\left(\frac{I_m^2}{2\pi}\int_0^{2\pi} \sin^2 \theta d\theta\right)} \qquad (\text{put } i = I_m \sin\theta)$$

Now, $\cos 2\theta = 1 - 2\sin^2\theta \quad \therefore\ \sin^2\theta = \frac{1 - \cos 2\theta}{2}$

$$\therefore \qquad I = \sqrt{\left(\frac{I_m^2}{4\pi} \times 2\pi\right)} = \sqrt{\left(\frac{I_m^2}{2}\right)} \qquad \therefore\ I = \frac{I_m}{\sqrt{2}} = 0.707\ I_m$$

Hence, we find that for a symmetrical sinusoidal current

r.m.s value of current = 0.707 × max. value of current

The r.m.s. value of an alternating current is of considerable importance in practice, because the ammeters and voltmeters record the r.m.s. value of alternating current and voltage respectively. In electrical engineering work, unless indicated otherwise, the values of the given current and voltage are always the r.m.s. values.

It should be noted that the average heating effect produced during one cycle is

$$= I^2R = \left(I_m/\sqrt{2}\right)^2 R = \frac{1}{2}I_m^2R$$

R.M.S. Value of a Complex Wave

In their case also, either the mid-ordinate method (when equation of the wave is not known) or analytical method (when equation of the wave is known) may be used. Suppose a current having the equation i = 12 sin ωt + 6 sin (3ωt – π/6) + 4 sin (5ωt – π/3) flows through a resistor of R ohm. Then, in the time period T second of the wave, the effect due to each component is as follows:

Fundamental .....$(12/\sqrt{2})^2$ RT watt

3rd harmonic.....$(6/\sqrt{2})^2$ RT watt

5th harmonic .....$\left(4/\sqrt{2}\right)^2$ RT watt

$\therefore$ Total heating effect = RT $\left[\left(12/\sqrt{2}\right)^2+\left(6/\sqrt{2}\right)^2+\left(4/\sqrt{2}\right)^2\right]$

If I is the r.m.s. value of the complex wave, then equivalent heating effect is I^2RT

$\therefore$ $$I^2RT = RT\left[\left(12/\sqrt{2}\right)^2+\left(6/\sqrt{2}\right)^2+\left(4/\sqrt{2}\right)^2\right]$$

$\therefore$ $$I = \sqrt{\left[\left(12/\sqrt{2}\right)^2+\left(6/\sqrt{2}\right)^2+\left(4/\sqrt{2}\right)^2\right]} = 9.74A$$

Had there been a direct current of (say) 5 amperes flowing in the circuit also*, then the r.m.s. value would have been

$$= \sqrt{\left(12/\sqrt{2}\right)^2+\left(6/\sqrt{2}\right)^2+\left(4/\sqrt{2}\right)^2+5^2} = 10.93A$$

Hence, for complex waves the rule is as follows: The r.m.s value of a complex current wave is equal to the square root of the sum of the squares of the r.m.s. values of its individual components.

Average Value

The average value I_a of an alternating current is expressed by that steady current which transfers across any circuit the same charge as is transferred by that alternating current during the same time.

In the case of a symmetrical alternating current (i.e. one whose two half-cycle are exactly similar, whether sinusoidal or non-sinusoidal), the average value over a complete cycle is zero. Hence, in their case, the average value is obtained by adding or integrating the instantaneous values of current over one half-cycle only. But in the case of an unsymmetrical alternating current (like half-wave rectified current) the average value must always be taken over the whole cycle.

i. **Mid-ordinate Method**

With reference to Fig.2.16, $I_{av} = \frac{i_1+i_2+....+i_n}{n}$

This method may be used both for sinusoidal and non-sinusoidal waves, although it is specially convenient for the latter.

ii. **Analytical Method**

The standard equation of an alternating current is, $i = I_m \sin\theta$

$$I_{av} = \int_0^{\pi}\frac{id\theta}{(\pi-0)} = \frac{I_m}{\pi}\int_0^{\pi}\sin\theta d\theta \qquad \text{(putting value of i)}$$

$$= \frac{I_m}{\pi}\left|-\cos\theta\right|_0^{\pi} = \frac{I_m}{\pi}\left|+1-(-1)\right| = \frac{2I_m}{\pi} = \frac{I_m}{\pi/2} = \frac{\text{twice the maximum current}}{\pi}$$

$\therefore$ $I_{av} = I_m/\frac{1}{2}\pi = 0.637I_m$ $\therefore$ average value of current = 0.637 × maximum value

Note: R.m.S. value is always greater than average value except in the case of a rectangular wave when both are equal.

Form Factor

It is defined as the ratio, $K_f = \frac{r.m.s.value}{average\,value} = \frac{0.707 I_m}{0.637 I_m} = 1.1$ (for sinusoidal alternating currents only).

In the case of sinusoidal alternating voltage also, $K_f = \frac{0.707 E_m}{0.637 E_m} 1.11$

As is clear, the knowledge of form factor will enable the r.m.s value to be found from the arithmetic mean value and vice-versa.

Crest or Peak or Amplitude Factor

It is defined as the ratio $K_a = \frac{\text{maximum value}}{\text{r.m.s value}} = \frac{I_m}{I_m/\sqrt{2}} = \sqrt{2} = 1.414$ (for sinusoidal a.c. only)

For sinusoidal alternating voltage also, $K_a = \frac{E_m}{E_m/\sqrt{2}} = 1.414$

Knowledge of this factor is of importance in dielectric insulation testing, because the dielectric stress to which the insulation is subjected, is proportional to the maximum or peak value of the applied voltage. The knowledge is also necessary when measuring iron losses, because the iron loss depends on the value of maximum flux.

A.C. Through Resistance, Inductance and Capacitance

We will now consider the phase angle introduced between an alternating voltage and current when the circuit contains resistance only, inductance only and capacitance only. In each case, we will assume that we are given the alternating voltage of equation $e = E_m \sin \omega t$ and will proceed to find the equation and the phase of the alternating current produced in each case.

A.C. Through Pure Ohmic Resistance Alone

The circuit is shown in Fig.2.56 Let the applied voltage be given by the equation

$v = V_m \sin \theta = V_m \sin \omega t$ ...(i)

Let R = ohmic resistance; i = instantaneous current

Obviously, the applied voltage has to supply ohmic voltage drop only. Hence for equilibrium

$v = iR;$

Putting the value of 'v' from above, we get $V_m \sin \omega t = iR;\ i = \frac{V_m}{R} \sin \omega t$ ...(ii)

Current 'i' is maximum when $\sin \omega t$ is unity $\therefore\ I_m = V_m / R$ Hence, equation (ii) becomes, $i = I_m \sin \omega t$

Comparing (i) and (ii), we find that the alternating voltage and current are in phase with each other as shown in Fig.2.12. It is also shown vectorially by vectors V_R and I in Fig. 2.11.

Power: Instantaneous power, $p = vi = V_m I_m \sin^2 \omega t$ ... (Fig.2.11)

$$= \frac{V_m I_m}{2}(1 - \cos 2\omega t) = \frac{V_m I_m}{2} - \frac{V_m I_m}{2} \cos 2\omega t$$

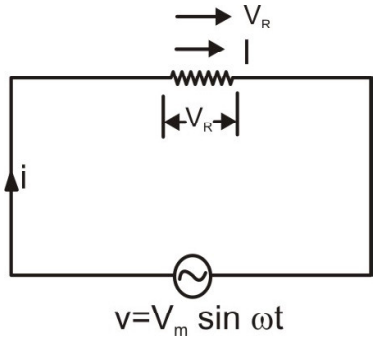


Fig. 2.11

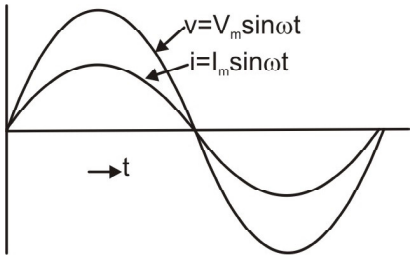


Fig. 2.12

Power consists of a constant part $\frac{V_m I_m}{2}$ and a fluctuating part $\frac{V_m I_m}{2}$ cos 2ωt of frequency double that of voltage and current waves. For a complete cycle. The average value of $\frac{V_m I_m}{2}$ cos $2\omega t$ is zero.

Hence, power for the whole cycle is

$$P = \frac{V_m I_m}{2} = \frac{V_m}{\sqrt{2}} \times \frac{I_m}{\sqrt{2}}$$

or $P = V \times I$ watt

where V = r.m.s. value of applied voltage.

I = r.m.s. value of the current .

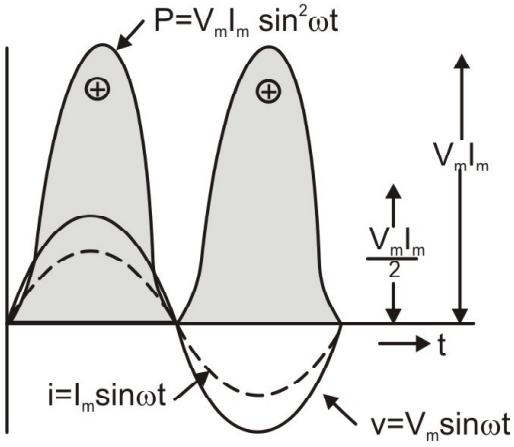


Fig. 2.13

It is seen from Fig.2.13 that no part of the power cycle becomes negative at any time. In other words, in a purely resistive circuit, power is never zero. This is so because the instantaneous values of voltage and current are always either both positive or negative and hence the product is always positive.

A.C. Through Pure Inductance Alone

Whenever an alternating voltage is applied to a purely inductive coil, a back e.m.f. is produced due to the self-inductance of the coil. The back e.m.f., at every step, opposes the rise of fall of current through the coil. As there is no ohmic voltage drop, the applied voltage has to overcome this self-induced e.m.f. only. So at every step

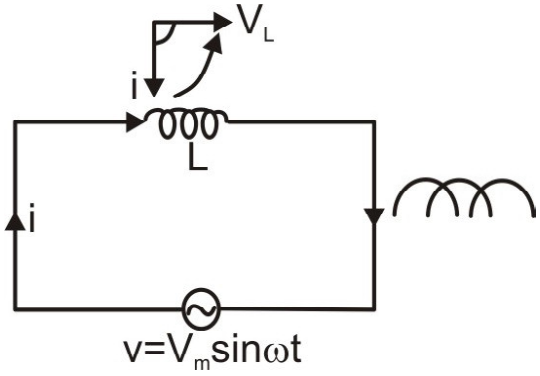


Fig. 2.14

$$v = L\frac{di}{dt}$$

Now $v = V_m \sin \omega t$

$$\therefore \quad V_m \sin \omega t = L\frac{di}{dt} \therefore di = \frac{V_m}{L} \sin \omega t \, dt$$

Integrating both sides, we get i = $\frac{V_m}{L}\int \sin\omega t\ dt$

$= \frac{V_m}{\omega L}(-\cos\omega t)$ ... (constant of integration=0)

$\therefore \quad = \frac{V_m}{\omega L}\sin\left(\omega t - \frac{\pi}{2}\right) = \frac{V_m}{X_L}\sin(\omega t - \pi/2)$

Max, value of *i* is $I_m = \frac{V_m}{\omega L}$ when $\sin\left(\omega t - \frac{\pi}{2}\right)$ is unity.

Hence, the equation of the current becomes i = $I_m \sin(\omega t - \pi/2)$.

So, we find that if applied voltage is represented by v = V_m sinωt, then current flowing in a purely inductive circuit is given by $i = I_m \sin\left(\omega t - \frac{\pi}{2}\right)$

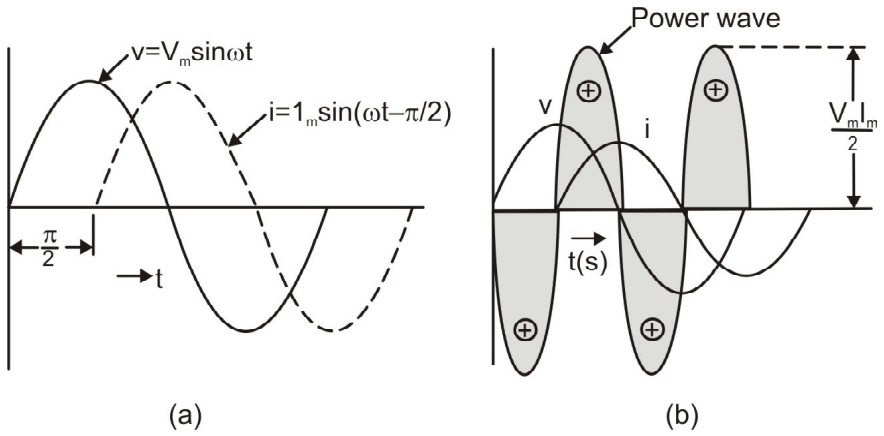


Fig. 2.15

Clearly, the current lags behind the applied voltage by a quarter cycle (Fig.2.15) or the phase difference between the two is π/2 with voltage leading. Vectors are shown in Fig. 2.14 where voltage has been taken along the reference axis. We have seen that $I_m = V_m/\omega L = V_m/X_L$. Here 'ωL' plays the part of 'resistance'. It is called the (inductive) reactance X_L of the coil and is given in ohms if L is in henry and ω is in radian/second.

Now, $X_L = \omega L = 2\pi fL$ ohm. It is seen that X_L depends directly on frequency of the voltage. Higher the value of f, greater the reactance offered and vice-versa.

Power

Instantaneous power = $v_i = V_m I_m \sin\omega t \sin\left(\omega t - \frac{\pi}{2}\right) = -V_m I_m \sin\omega t .\cos\omega t = -\frac{V_m I_m}{2}\sin 2\omega t$

Power for whole cycle is P = $-\frac{V_m I_m}{2}\int_0^{2\pi} \sin 2\omega t\ dt = 0$

It is also clear from Fig.2.15(b) that the average demand of power from the supply for a complete cycle is zero. Here again it is seen that power wave is a sine wave of frequency double that of the voltage and current waves. The maximum value of the instantaneous power is $V_m I_m/2$.

A.C. Through Pure Capacitance Alone

When an alternating voltage is applied to the plates of a capacitor, the capacitor is charged first in one direction and then in the opposite direction. When reference to Fig.2.16, let

v = p.d. developed between plates at any instant

q = Charge on plates at that instant.

Then q = Cv ...where C is the capacitance

= C V_m sin ωt ...putting the value of v,

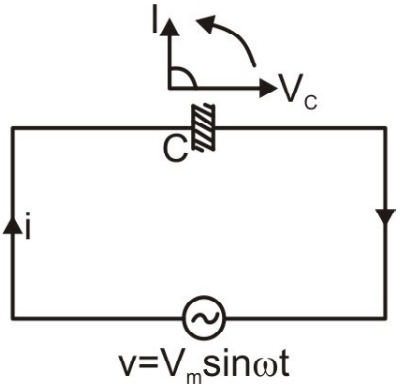


Fig. 2.16

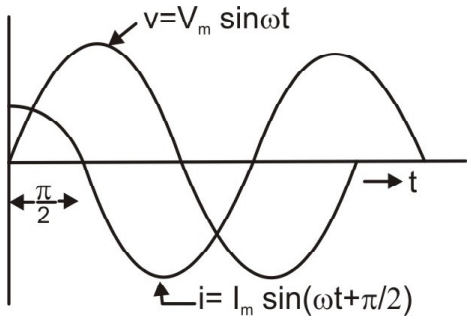


Fig. 2.17

Now, current i is given by the rate of flow of charge.

$$\therefore\ i = \frac{dq}{dt} = \frac{d}{dt}(CV_m \sin\omega t) = \omega C V_m \cos\omega t \quad \text{or } i = \frac{V_m}{1/\omega C}\cos\omega t = \frac{V_m}{1/\omega C}\sin\left(\omega t + \frac{\pi}{2}\right)$$

Obviously, $I_m = \frac{V_m}{1/\omega C} = \frac{V_m}{X_C} \therefore i = I_m \sin\left(\omega t + \frac{\pi}{2}\right)$

The denominator $X_C = 1/\omega C$ is known as capacitive reactance and is in ohms if C is in farad and ω in radian/second. It is seen that if the applied voltage is given by $v = V_m \sin \omega t$, then the current is given by $i = I_m \sin (\omega t + \pi/2)$.

Hence, we find that the current in a pure capacitor leads its voltage by a quarter cycle as shown in fig.11.17 or phase difference between its voltage and current is $\pi/2$ with the current leading. Vector representation is given in Fig.2.17. Note that V_c is taken along the reference axis.

Power. Instantaneous power

$$p = vi = V_m \sin \omega t.\ I_m \sin (\omega t + 90°)$$

$$= V_m I_m \sin\omega t \cos\omega t = \frac{1}{2} V_m I_m \sin 2\omega t$$

Power for the whole cycle

$$= \frac{1}{2} V_m I_m \int_0^{2\pi} \sin 2\omega t\, dt = 0$$

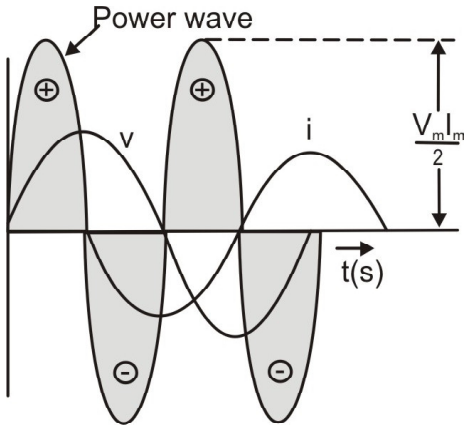


Fig. 2.18

The fact is graphically illustrated in Fig.2.18. We find that in a purely capacitive circuit, the average demand of power from supply is zero (as in a purely inductive circuit). Again, it is seen that power wave is a sine wave of frequency double that of the voltage and current waves. The maximum value of the instantaneous power is $V_m I_m/2$.

A.C. Through Resistance and Inductance

A pure resistance R and a pure inductive coil of inductance L are shown connected in series in Fig.2.19.

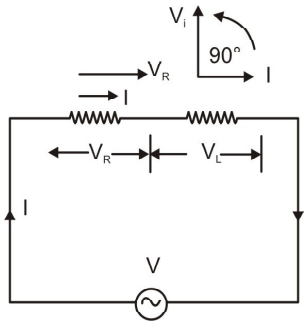


Fig. 2.19

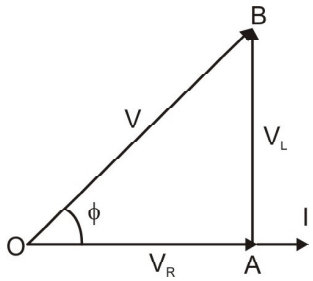


Fig. 2.20

Let V = r.m.s. value of the applied voltage, I = r.m.s. value of the resultant current V_R = IR – voltage drop across R (in phase with I), $V_L = I.X_L$ –voltage drop across coil (ahead of I by 90°)

These voltage drops are shown in voltage triangle OAB in Fig.2.20. Vector OA represents ohmic drop V_R and AB represents inductive drop V_L. The applied voltage V is the vector sum of the two i.e. OB.

$$\therefore \quad V = \sqrt{\left(V_R^2 + V_L^2\right)} = \sqrt{\left[\left(IR\right)^2 + \left(I.X_L\right)^2\right]} = I\sqrt{R^2 + X_L^2}\ , \quad \frac{V}{\sqrt{\left(R^2 + X_L^2\right)}} = I$$

The quantity $\sqrt{\left(R^2 + X_L^2\right)}$ is known as the impedance (Z) of the circuit. As seen from the impedance triangle ABC (Fig.2.20) $Z^2 = R^2 + X_L^2$.

i.e. (Impedance)^2 = (resistance)^2+(reactance)^2.

From Fig.11.20 it is clear that the applied voltage V leads the current I by an angle ϕ such that

$$\tan\phi = \frac{V_L}{V_R} = \frac{I.X_L}{I.R} = \frac{X_L}{R} = \frac{\omega L}{R} = \frac{reactance}{reactance} \quad \therefore \quad \phi = \tan^{-1}\frac{x_L}{R}$$

The same fact is illustrated graphically in Fig.2.22.

In other words, current I lags behind the applied voltage V by an angle ϕ.

Hence, if applied voltage is given by $v = V_m \sin\omega t$, then current equation is

$i = I_m \sin(\omega t - \phi)$ where $I_m = V_m/Z$

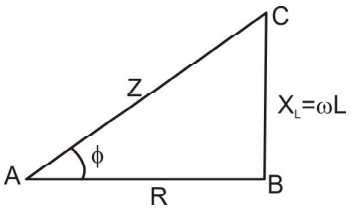


Fig. 2.21

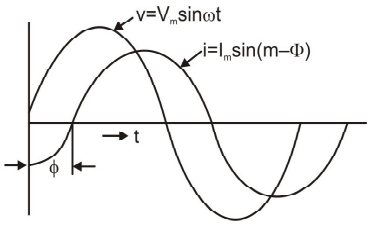


Fig. 2.22

In Fig. 2.23. It has been resolved into its two mutually perpendicular components, I cos ϕ along the applied voltage V and I sin ϕ in quadrature (i.e. perpendicular) with V.

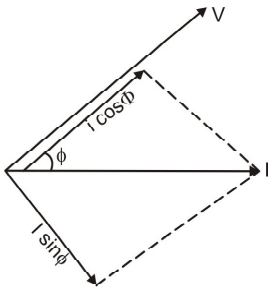


Fig. 2.23

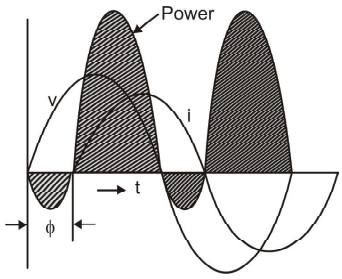


Fig. 2.24

The mean power consumed by the circuit is given by the product of V and that a component of the current i which is in phase with V.

So P = V × I cosϕ = r.m.s. voltage × r.m.s. current × cosϕ

The term 'cosϕ' is called the power factor of the circuit.

Remember that in an a.c. circuit, the product of r.m.s. volts and r.m.s. amperes gives volt amperes (VA) and not true power in watts. True power (W) = volt-amperes (V.A) × power factor.

or Watts = VA × cosϕ

It should be noted that power consumed is due to ohmic resistance only because pure inductance does not consume any power.

Now P = VI cos ϕ = VI × (R/Z) = (V/Z) × I.R = I^2 *R (∵ cosϕ = R/Zzz) or P = I^2R watt*

Graphical representing of the power consumed is shown in Fig.11.24.

let us calculate power in terms of instantaneous values.

Instantaneous power is = iυ = V_m sin ωt x I_m sin (ωt – ϕ) = $V_m I_m$ sin ωt sin (ωt – ϕ)

$$= \frac{1}{2} V_m I_m [\cos\phi - \cos(2\omega t - \Phi)]$$

Obviously, this power consists of two parts (Fig.2.25)

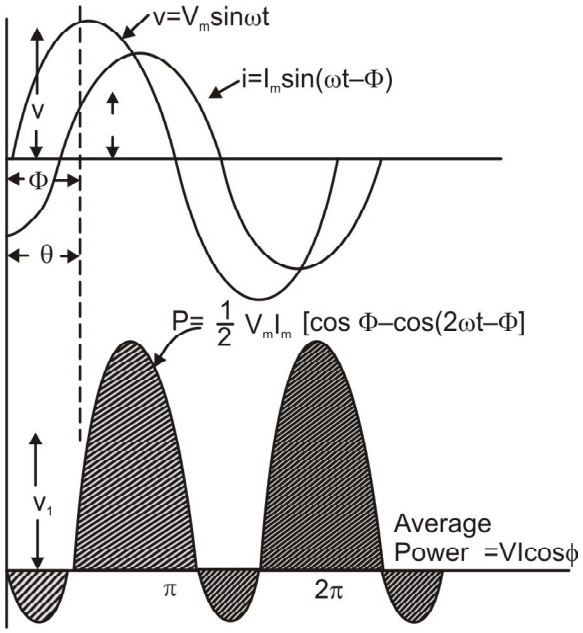


Fig. 2.25

i. a constant part $\frac{1}{2} V_m I_m \cos\phi$ which contributes to real power.

ii. a pulsating component $\frac{1}{2} V_m I_m \cos(2\omega t - \phi)$ which has a frequency twice that of the voltage and current. It does not contribute to actual power since its average value over a complete cycle is zero.

Hence, average power consumed = $\frac{1}{2} V_m I_m \cos\phi = \frac{V_m}{\sqrt{2}} . \frac{I_m}{\sqrt{2}} \cos\phi = VI\cos\phi$, where V and *i* represent the r.m.s. values.

Symbolic Notation, Z = R + jX_L

Impedance vector has numerical value of $\sqrt{(R^2 + X_L^2)}$

It is phase angle with the reference axis is ϕ = $\tan^{-1}(X_L / R)$

it may also be expressed in the polar form as Z = Z $\angle\phi°$

i. Assuming V = V $\angle 0°$; $I = \frac{V}{Z} = \frac{V\angle 0°}{Z\angle\phi°} = \frac{V}{Z}\angle -\phi°$ (Fig.2.26)

It shows that current vector is lagging behind the voltage vector by $\phi°$. The numerical value of current is V/Z.

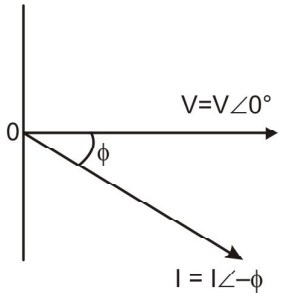


Fig. 2.26

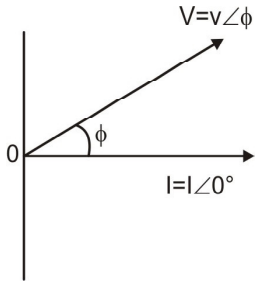


Fig. 2.27

ii. However, if we assumed that

$I = I\angle 0$, then

$V = IZ = I\angle 0° \times Z\angle \phi°$

$= IZ\angle \phi°$

It shows that voltage vector is $\phi°$ ahead of current vector in ccw direction as shown in Fig.2.27.

A.C. Through Resistance and Capacitance

The circuit is shown in Fig.2.28(a). Here V_R = IR = drop across R in phase with I.

$V_C = IX_C$ = drop across capacitor-lagging I by $\pi/2$.

As capacitive reactance X_C is taken negative, V_C is shown along negative direction of Y-axis in the voltage triangle [Fig.2.28(b)]

Now $V = \sqrt{V_R^2 + (-V_C)^2} = \sqrt{(IR)^2 + (-IX_C)^2} = I\sqrt{R^2 + X_C^2}$ or $I = \dfrac{V}{\sqrt{R^2 + X_C^2}} = \dfrac{V}{Z}$

The denominator is called the impedance of the circuit, So, $Z = \sqrt{R^2 + X_C^2}$

From Fig.2.28(b) it is found that I leads V by angle ϕ such that $\tan \phi = -X_C/R$

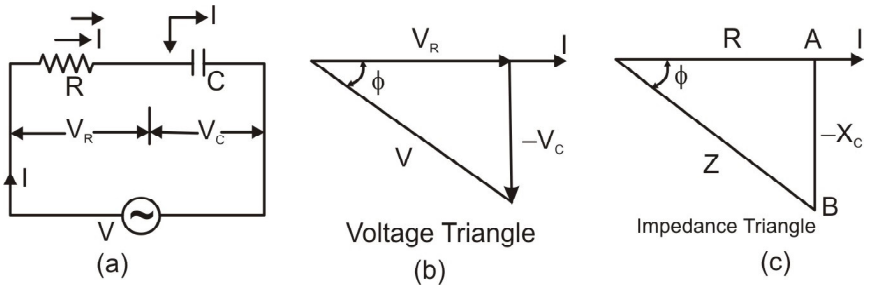


Fig. 2.28

Hence, it means that if the equation of the applied alternating voltage is $v = V_m \sin \omega t$, the equation of the resultant current in the R-C circuit is $i = I_m \sin(\omega t + \phi)$ so that current leads the applied voltage by an angle ϕ. This fact is shown graphically in Fig.2.29.

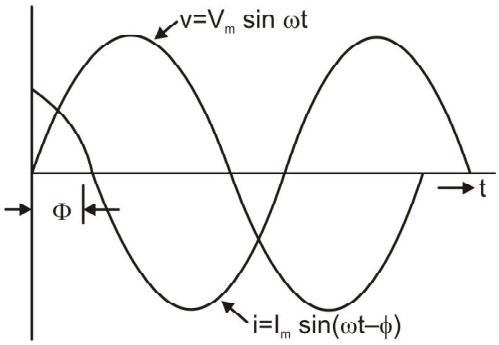


Fig. 2.29

Resistance, Inductance and Capacitance in Series

The three are shown in Fig.2.30(a) joined in series across an ac supply of r.m.s. voltage V.

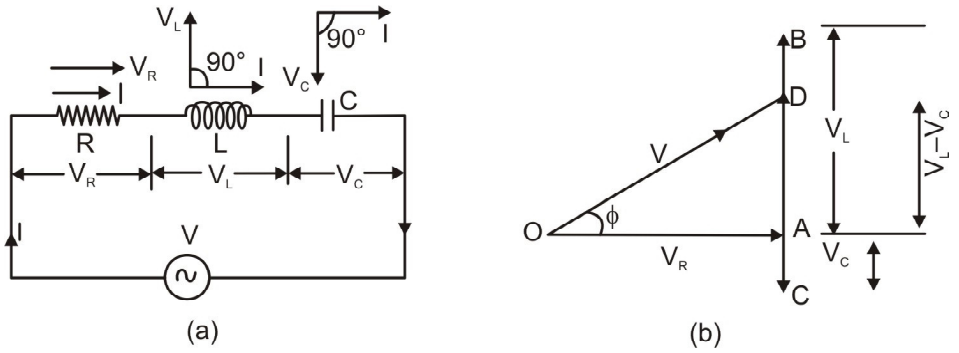


Fig. 2.30

Let $V_R = IR$ = voltage drop across R – in phase with i

$V_L = I.X_L$ = voltage drop across L – leading I by $\pi/2$

$V_C = I.X_C$ = voltage drop across C – lagging I by $\pi/2$

In voltage triangle of Fig.2.33(b), OA represents V_R, AB and AC represent the inductive and capacitive drops respectively. It will be seen that V_L and V_C are 180° out of phase with each other i.e. they are in direct opposition to each other.

Subtracting BD (=AC) from AB, we get the net reactive drop AD = $I(X_L - X_C)$

The applied voltage V is represented by OD and is the vector sum of OA and AD.

$\therefore$ OD $= \sqrt{OA^2 + AD^2}$ or $V = \sqrt{(IR)^2 + (IX_L - IX_C)^2} = I\sqrt{R^2 + (X_L - X_C)^2}$

or $I = \dfrac{V}{\sqrt{R^2 + (X_L - X_C)^2}} = \dfrac{V}{\sqrt{R^2 + X^2}} = \dfrac{V}{Z}$

The term $\sqrt{R^2 + (X_L - X_C)^2}$ is known as the impedance of the circuit. Obviously,

(impedance)^2 = (resistance)^2 + (net reactance)^2

or $Z^2 = R^2 + (X_L - X_C)^2 = R^2 + X^2$

where X is the net reactance (Fig.2.30 and 2.31).

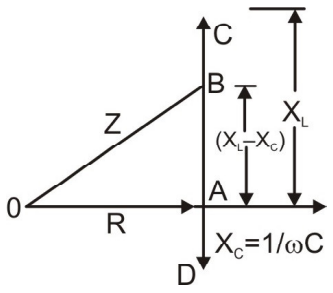


Fig. 2.31

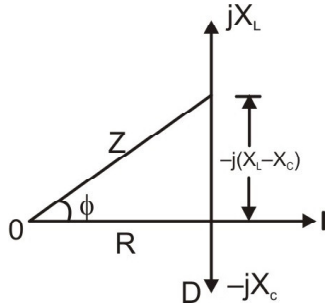


Fig. 2.32

Phase angle ϕ is given by $\tan\phi = (X_L - X_C)/R = X/R$ = net reactance/resistance.

Power factor is $\cos\phi = \dfrac{R}{Z} = \dfrac{R}{\sqrt{R^2 + (X_L - X_C)^2}} = \dfrac{R}{\sqrt{R^2 + X^2}}$

Hence, it is seen that if the equation of the applied voltage is $v = V_m \sin\omega t$, then equation of the resulting current in an R-L-C circuit is given by $i = I_m \sin(\omega t + \phi)$ $(\omega t \pm \phi)$

The +ve sign is to be used when current leads i.e. $X_C > X_L$.

The –ve sign is to be used when current lags i.e. when $X_L > X_C$.

In general, the current lags or leads the supply voltage by an angle ϕ such that $\tan\phi = X/R$.

Using symbolic notation, we have (Fig.2.35), $Z = R + j(X_L - X_C)$

Numerical value of impedance $Z = \sqrt{R^2 + (X_L - X_C)^2}$

Its phase angle is $\Phi = \tan^{-1} [X_L - X_C/R]$

$Z = Z\angle \tan^{-1}\left[(X_L - X_C)/R\right] = Z\angle \tan^{-1}(X/R)$

If $V = V\ \angle 0$, then $I = V/Z$.

Summary of Results of Series AC Circuits

Type of impedance	Value of Impedance	Phase angle for current	power factor
Resistance only	R	0°	1
Inductance only	ωL	90° lag	0
Capacitance only	$1/\omega C$	90° lead	0
Resistance and Inductance	$\sqrt{\left[R^2 + (\omega L)^2\right]}$	0< ϕ < 90° lag	1 > p.f. > 0 lag
Resistance and Capacitance	$\sqrt{\left[R^2 + (-1/\omega C)^2\right]}$	0< ϕ < 90° lead	1 > p.f > 0 lead
R-L-C	$\sqrt{\left[R^2 + (\omega L \sim 1/\omega C)^2\right]}$	between 0° and 90° lag or lead	between 0 and unity lag or lead

Resonance in R-L-C Circuits

We have seen from Art.13.9 that net reactance in an R-L circuit of Fig.2.40(a) is

$$X = X_L - X_C \text{ and } Z = \sqrt{\left[R^2 + (X_L - X_C)^2\right]} = \sqrt{R^2 + X^2}$$

Let such a circuit be connected across an a.c. source of constant voltage V but of frequency varying from zero to infinity. There would be a certain frequency of the applied voltage which would make X_L equal to X_C in magnitude. In that case, X = 0 and Z = R as shown in Fig.2.40(c). Under this condition, the circuit is said to be in electrical resonance.

As shown in fig.2.40(c), $V_L = I.\ X_L$ and $V_C = I.\ X_C$ and the two are equal in magnitude but opposite in phase. Hence, they cancel each other out. The two reactances taken together act as a short-circuit since no voltage develops across them. Whole of the applied voltage drops across R so that $V = X_C$. The circuit impedance Z = R. The phasor diagram for series resonance is shown in Fig.2.40(d).

Calculation of Resonant Frequency

The frequency at which the reactance of the series circuit is zero is called the resonant frequency f_0. Its value can be found as under : $X_L - X_C = 0$ or $X_L = X_C$ or $\omega_0 L = 1/\omega_0 C$.

or $$\omega_0^2 = \frac{1}{LC} \text{ or } (2\pi f_0)^2 = \frac{1}{LC} \text{ or } f_0 = \frac{1}{2\pi\sqrt{LC}}$$

If L is in henry and C in farad, then f_0 is given in Hz.

When a series R-L-C circuit is in resonance, it possesses minimum impedance z = R. Hence, circuit current is maximum, it being limited by value of R alone. The current $I_0 = V/r$ and is in phase with V.

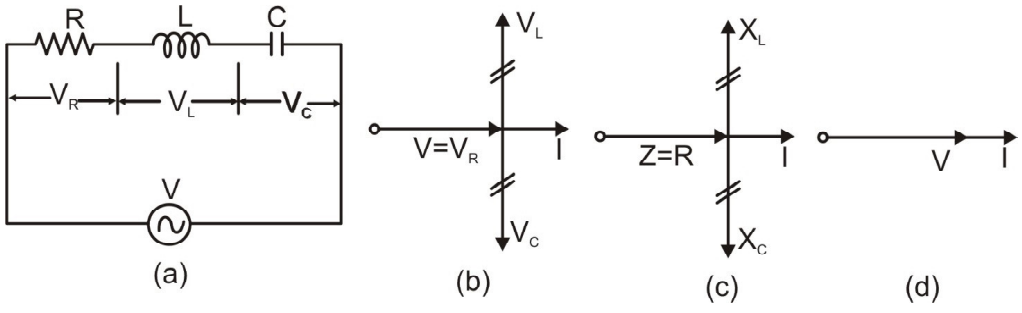


Fig. 2.33

Since circuit current is maximum, it produces large voltage drops across L and C. but these drops being equal and opposite, cancel each other out. Taken together, L and C from part of a circuit across which no voltage develops, however, large the current following. if it were not for the presence of R, such a resonant circuit would act like short-circuit to currents of the frequency to which it resonates. Hence, a series resonant circuit is sometimes called acceptor circuit and the series resonance is often referred to as voltage resonance.

In fact, at resonance the series RLC circuit is reduced to a purely resistive circuit, as shown in Fig.2.33.

Incidentally, it may be noted that if X_L and X_C are shown at any frequency f, that the value of the resonant frequency of such a circuit can be found by the relation $f_0 = f\sqrt{X_C / X_L}$.

When an R-L-C circuit is in resonance

1. net reactance of the circuit is zero i.e. $(X_L - X_C) = 0$ or $X = 0$.
2. circuit impedance is minimum i.e. Z=R. Consequently, circuit admittance is maximum.
3. circuit current is maximum and is given by $I_0 = V/Z_0 = V/R$.
4. power dissipated is maximum i.e. $P_0 = I_0^2 R = V^2 / R$.
5. circuit power factor angle $\theta = 0$. Hence, power factor $\cos\theta = 1$.
6. although $V_L = V_C$ yet V_C is greater than V_C because of its resistance.
7. at resonance, $\omega LC = 1$.
8. $Q = \tan\theta = \tan 0° = 0°$.

Graphical Representation of Resonance

Suppose an alternating voltage of constant magnitude, but of varying frequency is applied to an R-L-C circuit. The variations of resistance, inductive reactance X_L and capacitive reactance X_C with frequency are shown in Fig.2.45(a).

i. **Resistance :** It is independent of f, hence, it is represented by a straight line.

ii. **Inductive Reactance** : It is given by $X_L = \omega L = 2\pi fL$. As seen, X_L is directly proportional to f i.e. X_L increases linearly with f. Hence, its graph is a straight line passing through the origin.

iii. **Capacitive Reactance:** It is given by $X_C = 1/\omega C = 1/2\pi fC$. Obviously, it is inversely proportional to f. Its graph is a rectangular hyperbola which is drawn in the fourth quadrant because X_C is regarded negative. It is asymptotic to the horizontal axis at high frequencies and to the vertical axis at low frequencies.

iv. **Net Reactance:** It is given by $X = X_L \sim X_C$. Its graph is a hyperbola (not rectangular) and crosses the X-axis at point A which represents resonant frequency f_0.

v. **Circuit Impedance:** It is given by $Z = \sqrt{\left[R^2 + \left(X_L \sim X_C\right)^2\right]} = \sqrt{R^2 + X^2}$

At low frequencies Z is large because X_C is large. Since $X_C > X_L$, the net circuit reactance X is capacitive and the p.f. is leading [Fig.2.34(b)]. At high frequencies. Z is again large (because X_L is large) but is inductive because $X_L > X_C$. Circuit impedance has minimum values at f_0 iven by Z = R because X = 0.

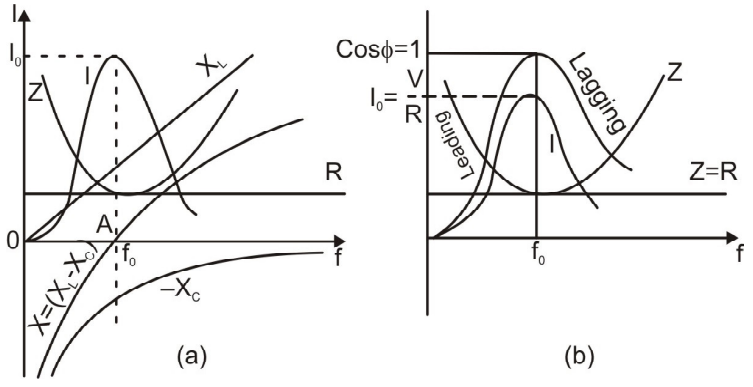


Fig. 2.34

vi. **Current I_0 :** It is the reciprocal of the circuit impedance. When Z is low. I_0 is high and vice versa. As seen, I_0 has low value on both sides of f_0 (because Z is large there) but has maximum value of I_0 = V/R at resonance. Hence, maximum power is dissipated by the series circuit under resonant conditions. At frequencies below and above resonance, current decreases as shown in Fig.11.34 (b). Now, I_0 = V/R and I = V/Z = $VI\sqrt{\left(R^2 + X^2\right)}$. Hence I / I_0 = R/Z = $VI\sqrt{\left(R^2 + X^2\right)}$ where X is the net circuit reactance at any frequency f.

vii. **Power Factor**

As pointed out earlier, X is capacitive below f_0. Hence, current leads the applied voltage. However, at frequencies above f_0. X is inductive. Hence, the current lags the applied voltage as shown in Fig. 2.34. The power factor has maximum value of unity at f_0.

Resonance Curve

The curve, between circuit current and frequency of the applied voltage, is known as resonance curve. The shapes of such a curve, for different values of R are shown in Fig.2.35. For smaller values of R, the resonance curve is sharply peaked and such a circuit is said to be sharply resonant or highly selective. However, for larger values of R, resonance curve is flat and is said to have poor selectivity. The ability of a resonant circuit to discriminate between one particular frequency and all others is called its selectivity. The selectivities of different resonant circuits are compared in terms of their half-power bandwidths.

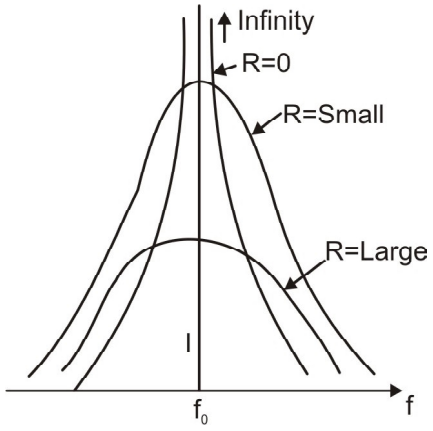


Fig. 2.35

Solving Parallel Circuits

When impedances are joined in parallel, there are three methods available to solve such circuits.
(a) Vector or phasor Method (b) Admittance Method and (c) Vector Algebra

Vector or Phasor Method

Consider the circuits shown in Fig.2.36. Here, two reactors A and B have been joined in parallel across and r.m.s. supply of V volts. The voltage across two parallel branches A and B is the same, but currents through then are different.

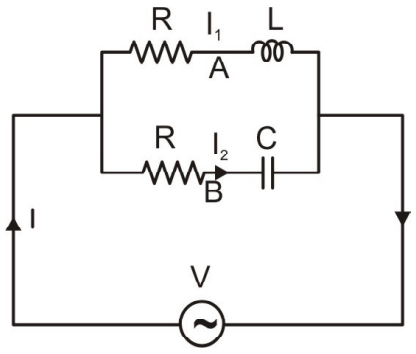


Fig. 2.36

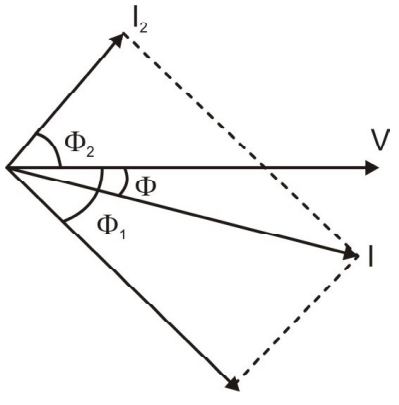


Fig. 2.37

For Branch A, $Z_1 = \sqrt{(R_1^2 + X_L^2)}$; I_1=V/Z_1; cos ϕ_1= R_1/Z_1 or ϕ_1 = $\cos^{-1}(R_1/Z_1)$

Current I_1 lags behind the applied voltage by ϕ_1 (Fig.14.2).

For Branch B, $Z_2 = \sqrt{(R_2^2 + X_c^2)}$; $I_2 = V/Z_2$; cos ϕ_2= (R_2/Z_2)

Current I_2 leads V by ϕ_2 (Fig.2.2).

Resultant current I

The resultant circuit current I is the vector sum of the branch currents I_1 and I_2 and can be found by (i) using parallelogram law of vectors, as shown in Fig.2.37. or (ii) resolving I_2 into their X-and y-components (or active and reactive components respectively) and then by combining these components, as shown in Fig.2.38. Method (ii) is preferable, as it is quick and convenient.

With reference to Fig.2.38(a) we have

Sum of the active components of I_1 and I_2

$$= I_1 \cos \phi_1 + I_2 \cos \phi_2$$

Sum of the reactive components of I_1 and I_2 $\sin \phi_2 - I_1 \sin \phi_1$

If I si the resultant current and ϕ, its phase, then its active and reactive components must be equal to these X-and Y-components respectively [Fig.2.38(b)].

$$I \cos\phi = I_1 \cos\phi_1 + I_2 \cos\phi_2 \text{ and } I \sin\phi = I_2 \sin\phi_2 - I_1 \sin\phi_1$$

$$\therefore \quad I = \sqrt{\left[(I_1 \cos\phi_1 + I_2 \cos\phi_2)^2 + (I_2 \sin g\phi_2 - I_1 \sin\phi_1)^2\right]}$$

$$\text{and} \quad \tan\phi = \frac{I_2 \sin\phi_2 - I_1 \sin\phi_1}{I_1 \cos\phi_1 + I_2 \cos\phi_2} = \frac{Y - \text{component}}{X - \text{component}}$$

If tan ϕ is positive, then current leads and if tan ϕ is negative, then current lags behind the applied voltage V. Power factor for the whole circuit is given by

$$\cos\phi = \frac{I_1 \cos\phi_1 + I_2 \cos\phi_2}{I} = \frac{X - \text{comp.}}{I}$$

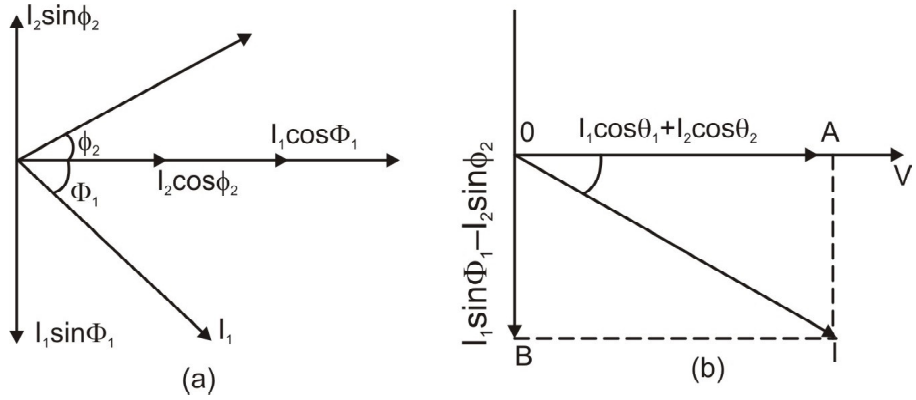


Fig. 2.38

Admittance Method

Admittance of a circuit is defined as the reciprocal of its impedance. It symbol is Y.

$$\therefore \quad Y = \frac{1}{Z} = \frac{1}{V} \text{ or } Y = \frac{\text{r.m.s. amperes}}{\text{r.m.s volts}}$$

Its unit is Sciemens (S). A circuit having an impedance of one ohm has an admittance of one Siemens. The old unit was mho (ohm spelled backwards).

As the impedance Z of a circuit has two components X and R (Fig. 2.39), similarly, admittance Y also has two components as shown in Fig. 2.40. The X-component is known as conductance and Y-component as susceptance.

Obviously, conductance $g = Y \cos \phi$

or $g = \frac{1}{Z} \cdot \frac{R}{Z}$ (from Fig. 2.39)

$$\therefore \quad g = \frac{R}{Z^2} = \frac{R}{R^2 + X^2}$$

Similarly, susceptance $b = Y \sin \phi = \frac{1}{2}, \frac{X}{Z}$ $\quad \therefore\ b = X/Z^2 = X/(R^2+X^2)$ (from Fig. 2.40)

The admittance $Y = \sqrt{(g^2 + b^2)}$ just as $Z = \sqrt{(R^2 + X^2)}$

The unit of g, b and Y is Siemens. We will regard the capacitive suscepance as positive and inductive susceptance as negative.

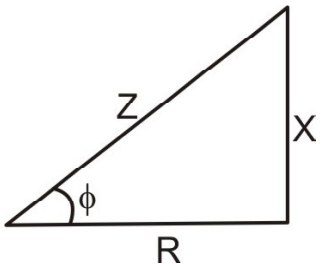


Fig. 2.39

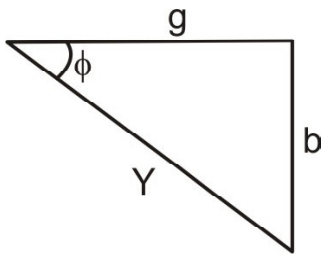


Fig. 2.40

Application of Admittance Method

Consider the 3-branched circuit of Fig.2.41. Total conductance is found by merely adding the conductances of three branches. Similarly, total susceptance is found by algebraically adding the individual susceptances of different branches.

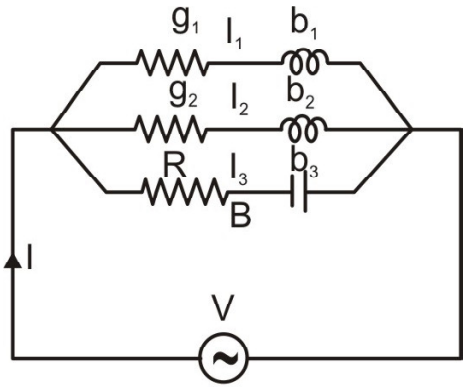


Fig. 2.41

Total conductance $G = g_1 + g_2 + g_3$ .....

Total susceptance $B = (-b_1) + (-b_2) + b_3$ ..... (algebraic sum)

$\therefore$ total admittance $Y = \sqrt{\left(G^2 + B^2\right)}$

Total current $I = VY$; Power factor $\cos \phi = G/Y$.

Complex or Phasor Algebra

Consider the parallel circuit shown in Fig.2.42. The two impedances, Z_1 and Z_2, being in parallel, have the same p.d. across them.

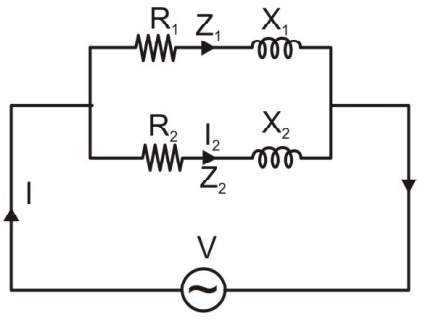


Fig. 2.42

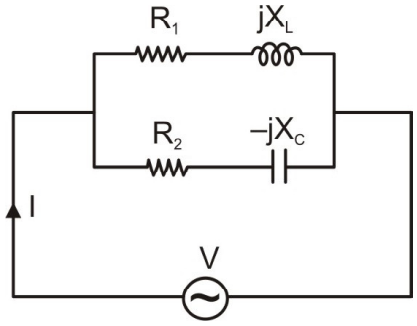


Fig. 2.43

Now $I_1 = \frac{V}{Z_1}$ and $I_2 = \frac{V}{Z_2}$

Total current $I = I_1 + I_2 = \frac{V}{Z_1} + \frac{V}{Z_2} = V\left(\frac{1}{Z_1} + \frac{1}{Z_2}\right) V\left(Y_1 + Y_2\right) = VY$

where Y = total admittance = $Y_1 + Y_2$

It should be noted that admittances are added for parallel branches. Whereas for branches in series, it is the impedances which are added. However, it is important to remember that since both admittances and impedances are complex quantities, all additions must be in complex form. Simple arithmetic additions must not be attempted.

Considering the two parallel branches of Fig.2.43, we have

$$Y_1 = \frac{1}{Z_1} = \frac{1}{R_1 + jX_L} = \frac{(R_1 - jX_L)}{(R_1 + jX_L)(R_1 - jX_L)}$$

$$= \frac{R_1 - jX_L}{R_1^2 + X_L^2} = \frac{R_1}{R_1^2 + X_L^2} - j\frac{X_L}{R_1^2 + X_L^2} = g_1 - jb_1$$

where $g_1 = \frac{R_1}{R_1^2 + X_L^2}$ – conductance of upper branch

$b_1 = -\frac{X_L}{R_1^2 + X_L^2}$ – susceptance of upper branch

Similarly, $Y_2 = \frac{1}{Z_2} = \frac{1}{R_2 - jX_C}$.

$$= \frac{R_2 + jX_C}{(R_2 - jX_C)(R_2 + jX_C)} = \frac{R_2 + jX_C}{R_2^2 + X_C^2} + \frac{R_2}{R_2^2 + X_c^2} + j\frac{X_C}{R_2^2 + X_C^2} = g_2 + jb_2$$

Total admittance $Y = Y_1 + Y_2 = (g_1 - jb_1) + (g_2 + jb_2) = (g_1 + g_2) - j(b_1 - b_2) = G - jB$

$$Y = \sqrt{\left[(g_1 + g_2)^2 + (b_1 - b_2)^2\right]}; \phi = tna^{-1}\left(\frac{b_1 - b_2}{g_1 + g_2}\right)$$

The polar form for admittance is $Y = Y \angle\phi°$ where ϕ is as given above.

$$Y = \sqrt{G^2 + B^2} \angle \tan^{-1}(B/G)$$

Total current $I = VY$; $I_1 = VY_1$ and $I_2 = VY_2$

If $V = V\angle 0°$ and $Y = Y \angle\phi$ then $I = VY = V \angle 0° \times Y \angle\phi = VY \angle\phi$

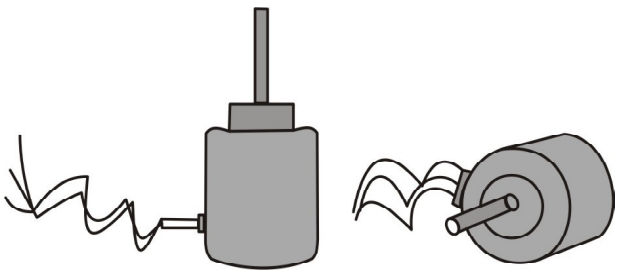
In general, if $V = V \angle\alpha$ and $Y = Y\angle\beta$, then $I = VY = V \angle\alpha \times Y\angle\beta = VY\angle\alpha + \beta$

Hence, it should be noted that when vector voltage is multiplied by admittance either in complex (rectangular) or polar form, the result is vector current in its proper phase relationship with respect to the voltage, regardless of the axis to which the voltage may have been referred to.

Generation of Polyphase Voltage

The kind of alternating currents and voltages discussed in chapter 12 to 15 are known as single-phase voltage and current, because they consist of a single alternating current and voltage wave. A single-phase alternator was diagrammatically depicted in Fig.2.1(b) and it was shown to have one armature winding only. But if the number of armature windings is increased, then it becomes polyphase alternator and it produces as many independent voltage waves as the number of windings or phases. These windings are displaced from one another by equal angles, the values of these angles being determined by the number of phases or windings. In fact, the word 'polyphase' mean poly (i.e. many or numerous) and phases (i.e. winding or circuit).

In a two-phase alternator, the armature windings are displaced 90 electrical degrees apart. A 3-phase alternator, as the name shows, has three independent armature windings which are 120 electrical degrees apart. Hence, the voltages induced in the three windings are 120° apart in time-phase. With the exception of two-phase windings, it can be stated that, in general, the electrical displacement between different phases is 360/n where n is the number of phases or windings.



The rotary Phase Converter

Three-phase systems are the most common, although, for certain special jobs, greater number of phases is also used. For example almost all mercury-arc rectifiers for power purposes are either six-phase or twelve-phase and most of the rotary converters in use are six-phase. All modern generators are practically three-phase. For transmitting large amounts of power, three-phase is invariably used. The reasons for the immense popularity of three-phase apparatus are that (i) it is more efficient (ii) it uses less material for a given capacity and (iii) it costs less than single-phase apparatus etc.

In Fig.2.44 is shown a two-pole, stationary-armature, rotating-field type three-phase alternator. It has three armature coils aa', bb' and cc' displaced 120° apart from one another. With the position and clockwise rotation of the poles as indicated in Fig.2.44, it is found that the e.m.f. induced in conductor 'a' for coil aa' is maximum and its direction is away from the reader. The e.m.f. in conductor 'b' of coil bb' would be maximum and away from the reader when the N-pole has turned through 120° i.e. when N-S axis lies along bb'. it is clear that the induced e.m.f. in conductor 'b' reaches its maximum value 120° later than the maximum value in conductor 'a'. In the like manner, the maximum e.m.f. induced (in the direction away from the reader) in conductor 'c' would occur 120° later than that in 'b' or 240° later than that in 'a'.

Thus the three coils have three e.m.fs. induced in them which are similar in all respects except that they are 120° out of time phase with one another as pictured in Fig.2.46. Each voltage wave is assumed to be sinusoidal and having maximum value of E_m.

In practice, the space on the armature is completely covered and there are many slots per phase per pole.

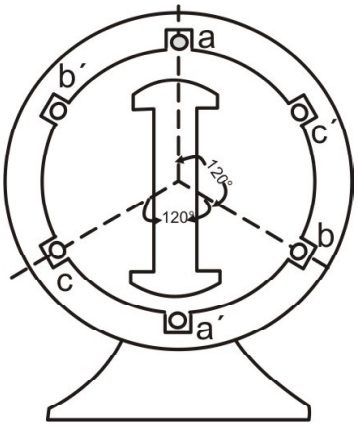


Fig. 2.44

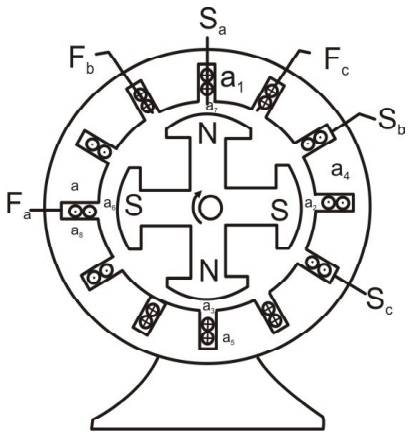


Fig. 2.45

Fig.2.45 illustrates the relative positions of the windings of a 3-phase, 4-pole alternator and Fig.2.47 shows the developed diagram of its armature windings. Assuming full-pitched winding and the direction of rotation as shown, phase 'a' occupies the position under the centres of N and S-poles. It starts and S_a and ends or finishes at F_a.

The second phase 'b' start at S_b which is 120 electrical degrees apart from the start of phase 'a', progresses round the armature clockwise (as does 'a') and finishes at F_b. Similarly, phase 'c' starts at S_c, which is 120 electrical degrees away from S_b, progresses round the armature and finishes at F_C. As the three circuits are exactly similar but are 120 electrical degrees apart, the e.m.f. waves generated in them (when the field rotates) are displaced from each other by 120°. Assuming these waves to be sinusoidal and counting the time from the instant when the e.m.f. in phase 'a' is zero, the instantaneous values of the three e.m.fs. will be given by curves of Fig.193.

Their equations are :

$$e_a = Em \sin \omega t \quad \text{....(i)}$$

$$e_b = Em \sin(\omega t - 120°) \quad \text{....(ii)}$$

$$e_c = Em \sin(\omega t - 240°) \quad \text{....(iii)}$$

As shown, alternating voltages may be represented by revolving vectors which indicate their maximum values (or r.m.s. values if desired). The actual values of these voltages vary from peak positive to zero and to peak negative values in one revolution of the vectors. In Fig. 2.48 are shown the three vectors representing the r.m.s. voltages of the three phases E_o, E_b, and E_c (in the present case $E_a = E_b = E_c = E$, say)

It can be shown that the sum of the three phase e.m.fs. is zero in the following three ways:

i. The sum of the above three equations (i), (ii) and (iii) is zero as shown below:

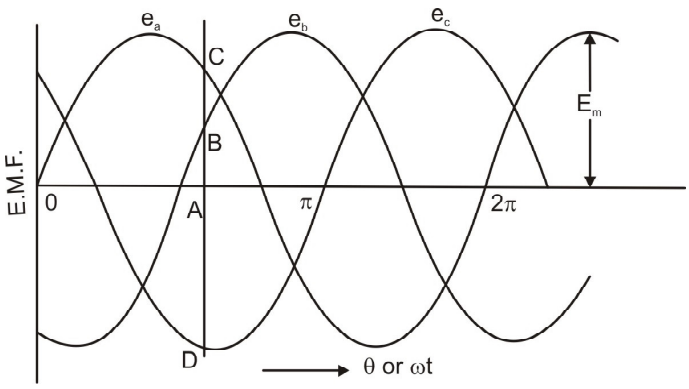


Fig. 2.46

Resultant instantaneous e.m.f. = $e_a + e_b + e_c$

$$= E_m \sin\omega t + E_m \sin(\omega t - 120°) + E_m(\omega t - 240°)$$

$$= E_m\left[\sin\omega t + 2\sin(\omega t - 180°)\cos 60°\right]$$

$$= E_m\left[\sin\omega t - 2\sin\omega t\cos 60°\right] = 0$$

ii. The sum of ordinates of three e.m.f. curves of Fig.2.46 is zero. For example, taking ordinates AB and AC as positive and AD as negative, it can be shown by actual measurement that

AB + AC + (–AD) = 0

iii. If we add the three vectors of Fig.11.48 either vectorially or by calculation, the result is zero.

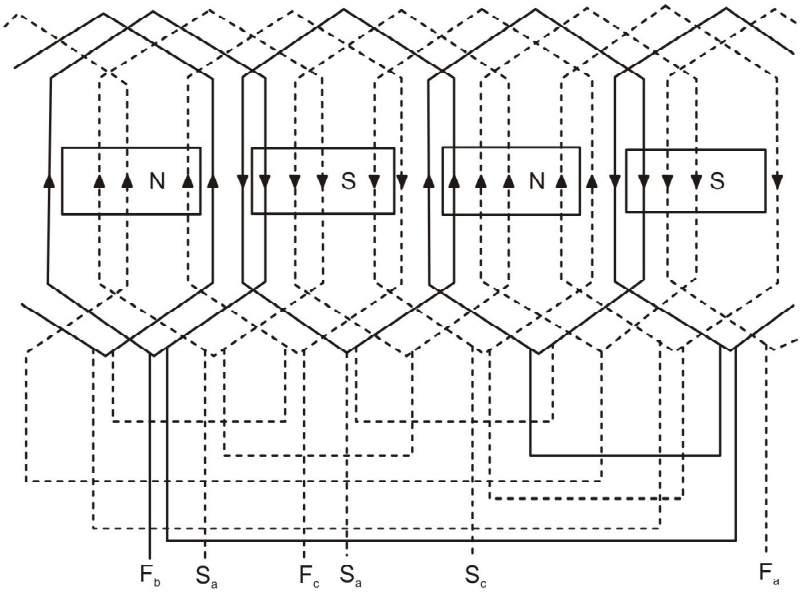


Fig. 2.47

Vector Addition

As shown in Fig.2.49, the resultant of E_a and E_b is E_r and its magnitude is 2E cos 60° = E where $E_a = E_b = E_c = E$.

This resultant E_r is equal and opposite to E_c. Hence, their resultant is zero.

By Calculation

Let us take E_a as reference voltage and assuming clockwise phase sequence

E_a = $E\angle 0° = E + j0$

E_b = $E\angle -240° = E\angle 120° = E(-0.05 + j0.866)$

E_c = $E\angle -240° = E\angle 120° = E(-0.05 + j0.866)$

$$\therefore\ E_a + E_b + E_c = (E + j0) + E(-0.5 - 0.866) + E(-0.05 + j0.866) = 0$$

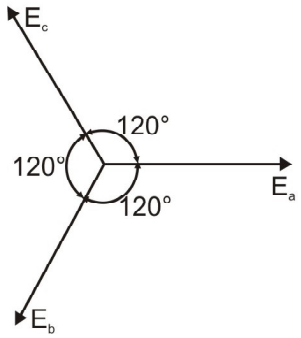


Fig. 2.48

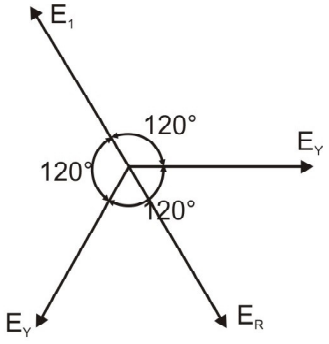


Fig. 2.49

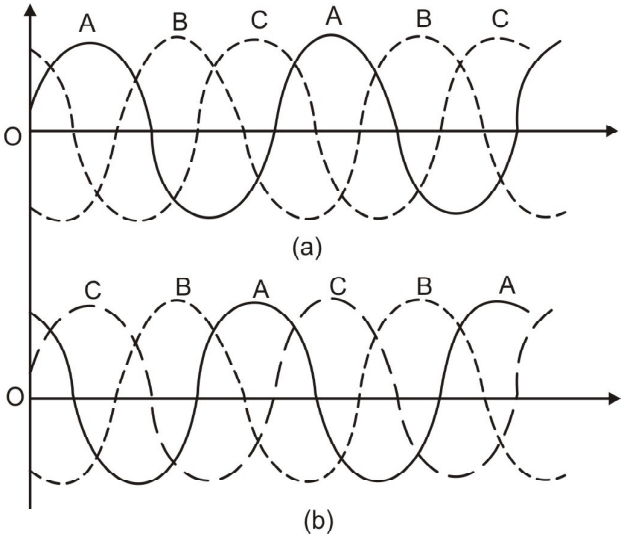


Fig. 2.50

Phase Sequence

By phase sequence is meant the order in which the three phases attain their peak or maximum values. In the development of the three-phase e.m.fs in Fig. 2.50, clockwise rotation of the field system in Fig. 2.44 was assumed. This assumption made the e.m.fs. of phase 'b' lag behind that of 'a' by 120° and in a similar way, made that of 'c' lag behind that of 'b' by 120° (or that of 'a' by 240°). hence, the order in which the e.m.fs. of phases a, b and c attain their maximum values is abc. It is called the phase order or phase sequence $a \rightarrow b \rightarrow c$ as illustrated in Fig. 2.50(a).

If, now, the rotation of the field structure of Fig. 2.50 is reversed i.e. made anticlockwise, then the order in which the three phases would attain their corresponding maximum voltages would also be reversed. The phase sequence would become $a \rightarrow b \rightarrow c$. This means that e.m.f. of phase 'c' would now lag behind that of phase 'a' by 120° instead of 240° as in the previous case as shown in Fig. 2.50(b). By repeating the letters, this phase sequence can be written as acbacba which is the same thing as cba. Obviously, a three-phase system has only two possible sequences : abc and cba (i.e. abc read in the reverse direction).

■■■

UNIT - 3

CHAPTER - 3
Rotating Electrical Machines

Working

Imagine the coil to be rotating in clock-wise direction (Fig. 3.1). As the coil assumes successive positions in the field, the flux linked with it changes. Hence, an e.m.f. is induced in it which is proportional to the rate of change of flux linkages ($e = Nd\Phi dt$). When the plane of the coil is at right angles to lines of flux i.e. when it is in position, 1, then flux linked with the coil is maximum but rate of change of flux linkages is minimum.

It is so because in this position, the coil sides AB and CD do not cut or shear the flux, rather they slide along them i.e. they move parallel to them. Hence, there is no induced e.m.f. in the coil. Let us take this no-e.m.f. or vertical position of the coil as the starting position. The angle of rotation or time will be measured from this position.

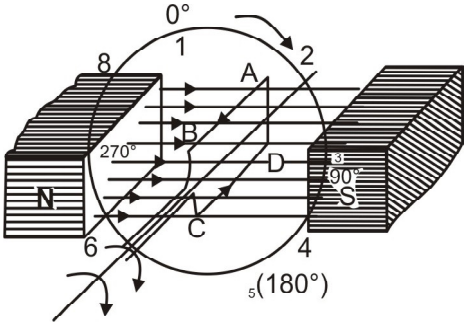


Fig. 3.1

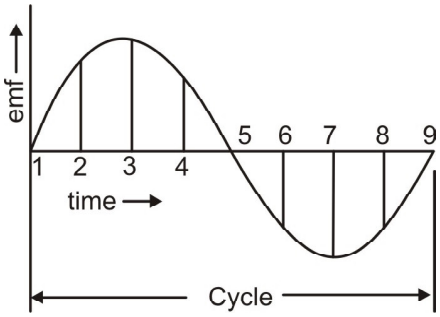


Fig. 3.2

As the coil continues rotating further, the rate of change of flux linkages (and hence induced e.m.f. in it) increases, till position 3 is reached where $\theta = 90°$. Here, the coil plane is horizontal i.e. parallel to the lines of flux. As seen, the flux linked with the coil is minimum but rate of change of flux linkages is maximum. Hence, maximum e.m.f. is induced in the coil when in this position (Fig. 3.2).

In the next quarter revolution i.e. from 90° to 180°, the flux linked with the coil gradually increases but the rate of change of flux linkages decreases. Hence, the induced e.m.f. decreases gradually till in position 5 of the coil, it is reduced to zero value.

So, we find that in the first half revolution of the coil, no (or minimum) e.m.f. is induced in it when in position I, maximum when in position 3 and no e.m.f. when in position 5. The direction of this induced e.m.f. can be found by applying Flaming's Right-hand rule which gives its direction from A to B and C to D. The current through the load resistance R flows from M to L during the first half revolution of the coil.

In the next half revolution i.e. from 180° to 360°, the variations in the magnitude of e.m.f. are similar to those in the first half revolution. Its value is maximum when coil is in position 7 and minimum when in position 1. Hence, the path of current flow is along DCLMBA which is just the reverse of the previous direction of flow.

Therefore, we find that the current which we obtain from such a simple generator reverse its direction after every half revolution. Such a current undergoing periodic reversals is known as alternating current. It is, obviously, different from a direct current which continuously flows in one and the same direction. It should be noted that alternating current not only reverses its direction, it does not even keep its magnitude constant while flowing in any one direction. The two half-cycles may be called positive and negative half-cycles respectively (Fig. 3.2).

For making the flow of current unidirectional in the external circuit, the slip-rings are replaced by split-rings (Fig. 3.3). The split-rings are made out of a conducting cylinder which is cut into two halves or segments insulated from each other by a thin sheet of mica or some other insulating material (Fig. 3.4).

As before, the coil ends are joined to these segments on which rest the carbon or copper brushes. It is seen [Fig. 3.5a)] that in the first half revolution current flows along (ABMNLCD) i.e. the brush No.1 in contact with segment 'n' acts as the positive end of the supply and 'b' as the negative end. In the next half revolution [Fig. 3.5(b)], the direction of the induced current in the coil has reversed. But at the same time, the

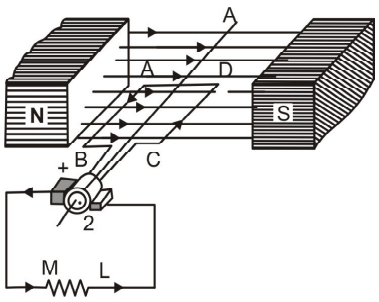


Fig. 3.3

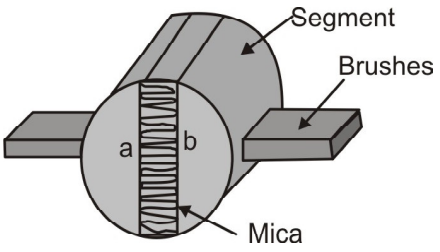


Fig. 3.4

positions of segments 'a' and 'b' have also reversed with the result that brush no.1 comes in touch with the segment which is positive i.e. segment 'b' in this case. hence, current tin the load resistance again flows from M to L. The waveform of the current through the external circuit is as shown in Fig.3.6. The current is unidirectional but not continuous like pure direct current.

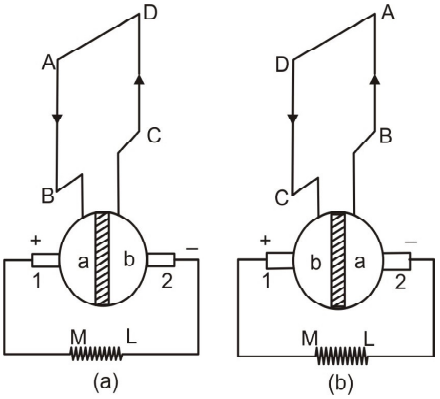


Fig. 3.5

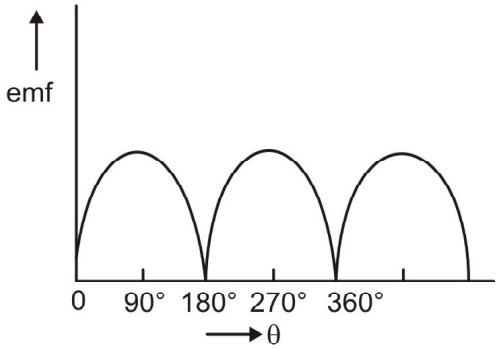


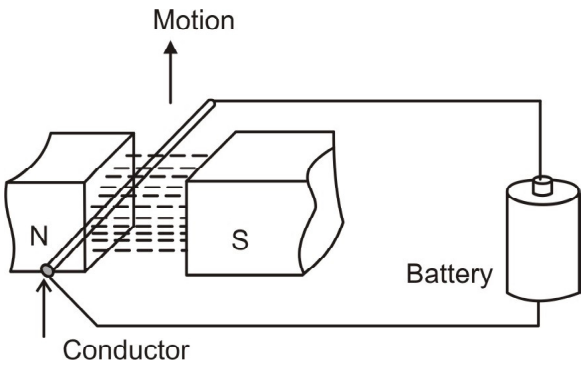
Fig. 3.6

It should be noted that the position of brushes is so arranged that the change over of segments 'a' and 'b' from one brush to the other takes place when the plane of the rotating coil is at right angles to the plane of the lines of flux. It is so because in that position, the induced e.m.f. in the coil is zero.

Another important point worth remembering is that even now the current induced in the coil sides is alternating as before. It is only due to the rectifying action of the split-rings (also called commutator) that it becomes unidirectional in the external circuit. Hence, it should be clearly understood that even in the armature of a d.c. generator, the induced voltage is alternating.

Motor Principle

An Electric motor is a machine which converts electric energy into mechanical energy. Its action is based on the principle that when a current-carrying conductor is placed in a magnetic field is experiences a mechanical force whose direction is given by Fleming's Left-hand Rule and whose magnitude is given by F = BIl Newton.



Principle of Motor

Constructionally, there is no basic difference between a d.c. generator and a d.c. motor. In fact, the same d.c. machine can be used interchangeably as a generator or as a motor. D.C. motors are also like generators, shunt-wound or series-wound or compound-wound.

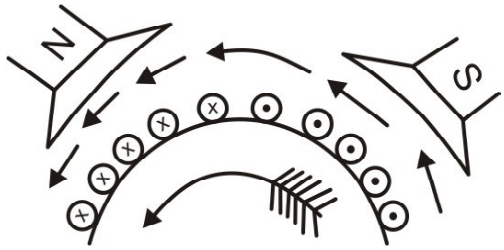


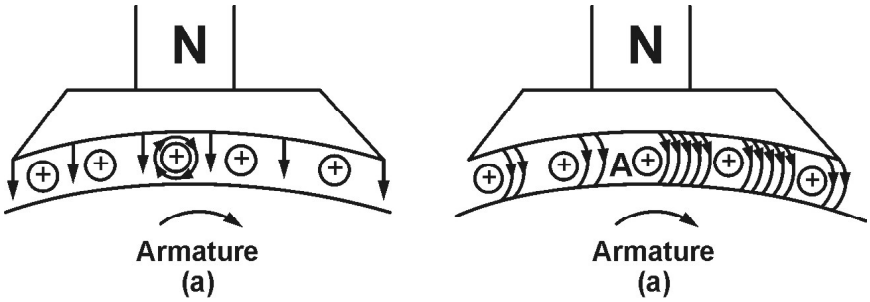
Fig. 3.7

In Fig.3.7 a part of multipolar d.c. motor is shown. When its field magnet are excited and its armature conduction are supplied with current from the supply mains, they experience a force tending to rotate the armature. Armature conductors under N-pole are assumed to carry current downwards (crosses) and those under S-poles, to carry current upwards (dots). By applying fleming's Left-hand Rule, the direction of the force on each conductor can be found. It is shown by small arrows placed above each conductor. It will be seen that each conductor can be found. It will be seen that each conductor experiences a force F which tends to rotate the armature in anticlockwise direction. These forces collectively produce & driving torque which sets the armature rotating.

It should be noted that the function of a commutator in the motor is the same as in a generator. By reversing current in each conductor as it passes from one pole to another, it helps to develop a continuous and unidirectional torque.

Comparison of Generator and Motor Action

As said above, the same d.c. machine can be used, at least theoretically, interchangeably as a generator or as a motor. When operating as a generator, it si driven by a mechanical machine and it develops, voltage which in turn produces current flow in an electric circuit. When operating as a motor, it is supplied by electric current and it develops torque which in turn produces mechanical rotation.



Let us first consider its operation as a generator and see how exactly and through which agency, mechanical power is converted into electric power.

Significance of the Back e.m.f.

As explained when the motor armature rotates, the conductors also rotate and hence cut the flux. In accordance with the laws of electromagnetic induction, e.m.f. is induced in them whose direction, as found by Fleming's Right-hand Rule, is in opposition to the applied voltage (Fig.3.9). Because of its opposing direction, it is referred to as counter e.m.f. or back e.m.f. E_b. The equivalent circuit of a motor is shown in Fig.3.10. The rotating armature generating the back e.m.f. E_b is like a battery of e.m.f. E_b put across a supply mains of V volts. Obviously, V has to drive I_a against the opposition of E_b. The power required to overcome this opposition is $E_b I_a$.

In the case of a cell, this power over an interval of time is converted into chemical energy, but in the present case, it is converted into mechanical energy.

It will be seen that $I_a = \frac{\text{Net voltage}}{\text{Resistance}} = \frac{V - V_b}{R_a}$

where R_a is the resistance of the armature circuit. As pointed out above.

$E_b = \Phi ZN \times (P/A)$ volt where N is in r.p.s.

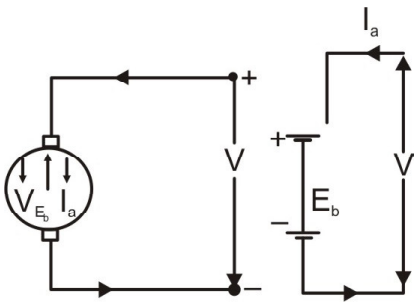


Fig. 3.9

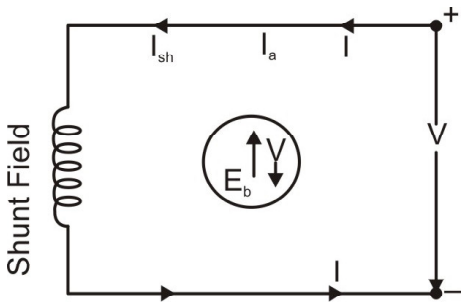


Fig. 3.10

Back e.m.f. depends, among other factors, upon the armature speed. If speed is high, E_b is large hence armature current I_a, seen from the above equation, is small. If the speed is less, then E_b is less, hence more current flows which develops motor torque. So, we find that E_b acts like a governor i.e., it makes a motor self-regulating so that it draws as much current as is just necessary.

Voltage Equation of a Motor

The voltage V applied across the motor armature has to

i. overcome the back e.m.f. E_b and

ii. supply the armature ohmic drop I_aR_a.

$\therefore \quad V = E_b + I_aR_a$

This is known as voltage equation of a motor. Now, multiplying both sides by I_a, we get

$$VI_a = E_bI_a + I_a^2R_a$$

As shown in Fig. 3.10,

VI_a = Electrical input to the armature.

E_bI_a = Electrical equivalent of mechanical power developed in the armature

$I_a^2R_a$ = Cu loss in the armature.

Hence, out of the armature input, some is wasted in I^2R loss and the rest is converted into mechanical power within the armature.

It may also be noted that motor efficiency is given by the ratio of power developed by the armature to its input i.e., $E_bI_a/VI_a = E_b/V$. Obviously, higher the value of E_b as compared to V, higher the motor efficiency.

Condition for Maximum Power

The gross mechanical power developed by a motor is $P_m = VI_a - I_a^2R_a$.

Differentiating both sides with respect to I_a and equating the result to zero, we get

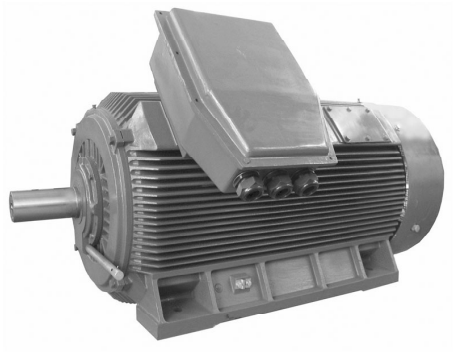
$$dP_m/dI_a = V - 2I_aR_a = 0 \quad \therefore \; I_aR_a = V/2$$

As $V = E_b + I_aR_a$ and $I_aR_a = V/2 \quad \therefore \quad I_aR_a = V/2 \quad \therefore E_b = \frac{V}{2}$

Thus gross mechanical power developed by a motor is maximum when back e.m.f. is equal to half the applied voltage. This condition is, however, not realized in practice, because in that case current would be much beyond the normal current of the motor. Moreover, half the input would be wasted in the form of heat and taking other losses (mechanical and magnetic) into consideration, the motor efficiency will be well below 50 percent.

Classification of A.C. Motors

With the almost universal adoption of a.c. system of distribution of electric energy for light and power, the field of application of a.c. motors has widened considerably during recent years. As a result, motor manufactures have tried, over the last few decades, to perfect various types of a.c. motors suitable for all classes of industrial drives and for both single and three-phase a.c. supply. This has given rise to bewildering multiplicity of types whose proper classification often offers considerable difficulty. Different a.c. motors may, however, be classified and divided into various groups from the following different points of view:



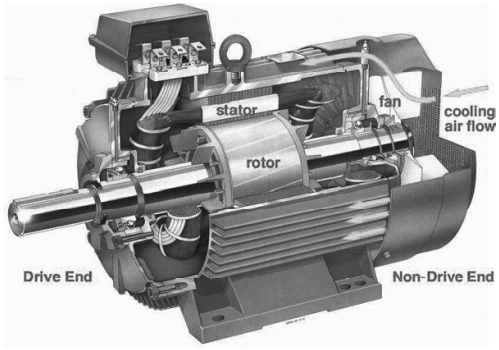
Three phase high voltage asynchronous motors.

1. **As Regards their principle of operation**
   - **(a) Synchronous motors**
     - i. plain and ii. super
   - **(b) Asynchronous motors**
     - a. Induction motors
       - i. Squirrel cage single double
       - ii. Slip-ring (external resistance)
     - b. Commutator motors
       - i. Series { single phase universal}
       - ii. Compensated [conductively inductively
       - iii. shunt { simple compensated
       - iv. repulsion { straight compensated
       - v. repulsion-start induction
       - vi. repulsion induction

2. **As Regards the Type of Current**
   - i. single phase ii. three phase
3. **As Regards Their Speed**
   - i. constant speed ii. variable speed iii. adjustable speed
4. **As Regards Their Structural Features**
   - i. open ii. enclosed iii. semi-enclosed
   - iv. ventilated v. pipe-ventilated vi. reverted frame eye etc.

Induction Motor : General Principle

As a general rule, conversion of electrical power into mechanical power takes place in the rotating part of an electric motor. In d.c. motors, the electric power is conducted directly to the armature (i.e rotating part) through brushes and commutator. Hence, in this sense, a d.c. motor can be called a conduction motor. However, in a.c. motors, the rotor does not receive electric power by conduction but by induction in exactly the same way as the secondary of a 2-winding transformer receives its power from the primary. That is why motors are known as induction motors. In fact, an induction motor can be treated as a rotating transformer i.e. one in which primary winding is stationary but the secondary is free to rotate.



Squirrel cage AC induction motor opened to show the stator and rotor construction, the shaft with bearings, and the cooling fan.

Of all the a.c. motors, the polyphase induction motor is the one which is extensively used for various kinds of industrial drives. It has the following main advantages and also some dis-advantages:

Advantages;

1. It has very simple and extremely rugged, almost unbreakable construction (especially squirrel-cage type).
2. Its cost is low and it is very reliable.
3. It has sufficiently high efficiency. In normal running condition, no brushes are needed, hence frictional losses are reduced. It has a reasonably good power factor.
4. It requires minimum of maintenance.
5. It starts up from rest and needs no extra starting motor and has not to be synchronized. Its starting arrangement is simple-especially for squirrel-cage type motor.

Disadvantages

1. Its speed cannot be varied without sacrificing some of its efficiency.
2. Just like a d.c. shunt motor, its speed decreases with increase in load.
3. Its starting torque is somewhat inferior to that of a d.c. shunt motor.

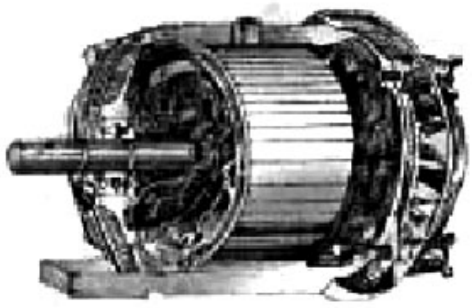
Construction

An induction motor consists essentially of two main parts:

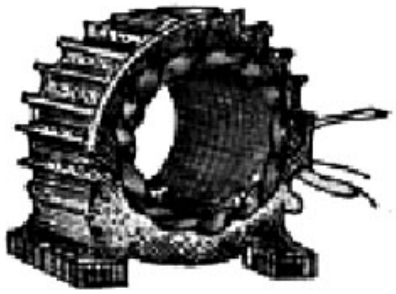
a. a stator and b. a rotor.

a. Stator

The stator of an induction motor is, in principle, the same as that of a synchronous motor or generator. It is made up of a number of stamping, which are slotted to receive the windings [Fig. 3.11(a)]. The stator carries a 3-phase winding [Fig.3.11(b)] and is fed from a 3-phase supply. It is wound for a definite number if poles, the exact number of poles being determined by the requirements of speed. Greater the number of poles, lesser the speed and vice versa. It will be shown in Art.34.6 that the stator windings, when supplied with 3-phase currents, produce a magnetic flux, which is of constant magnitude but which revolves (or rotates) at synchronous speed (given by N_s=120 f/P). This revolving magnetic flux induces an e.em.f. in the motor by mutual induction.



3.11. a. Unwound stator with semi-closed slots. laminations are of high-quality low-loss silicon steel



3.11. b. Completely wound stator for an induction motor.

b. **Rotor**

i. Squirrel-cage rotor: Motors employing this type of rotor are known as squirrel-cage induction motors.
ii. Phase-wound or wound rotor: Motor employing this type of rotor are variously known in 'phase-wound' motors or 'wound' motors or as 'slip-ring' motors.

Squirrel-cage Rotor

Almost 90 per cent of induction motors are squirrel-cage type, because this type of rotor has the simplest and most rugged construction imaginable and is almost indestructible. The rotor consists of a cylindrical laminated core with parallel slots for carrying the rotor conductors which, it should be noted clearly, are not wires but consist of heavy bars of copper, aluminium or alloys. One bar is placed in each slot, rather the bars are inserted from the end when semi-closed slots are used. The rotor bars are brazed or electrically welded or bolted to two heavy and stout short-circuiting end-rings, thus giving us, what is so picturesquely called, a squirrel-case construction (Fig. 3.12).

It should be noted that the rotor bars are permanently short-circuited on themselves, hence it is not possible to add any external resistance in series with the rotor circuit for starting purposes.

The rotor slots are usually not quite parallel to the shaft but are purposely given a slight skew (Fig. 3.13). This is useful in two ways:

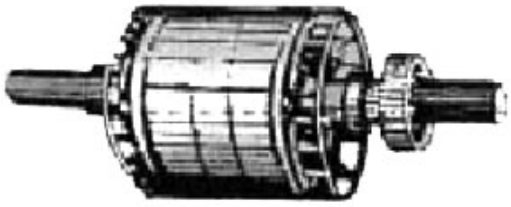


Fig. 3.12a. Squirrel-cage rotor with copper bars and alloy brazed end-rings

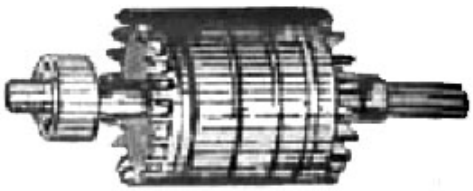


Fig. 3.12b. Rotor with shaft and brings

i. it helps to make the motor run quietly by reducing the magnetic hum and
ii. it helps in reducing the locking tendency of the rotor i.e. tendency of the rotor teeth to remain under the stator teeth due to direct magnetic attraction between the two.

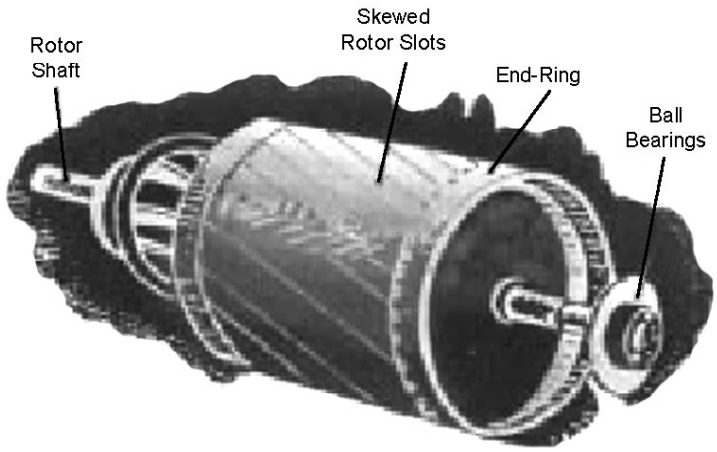
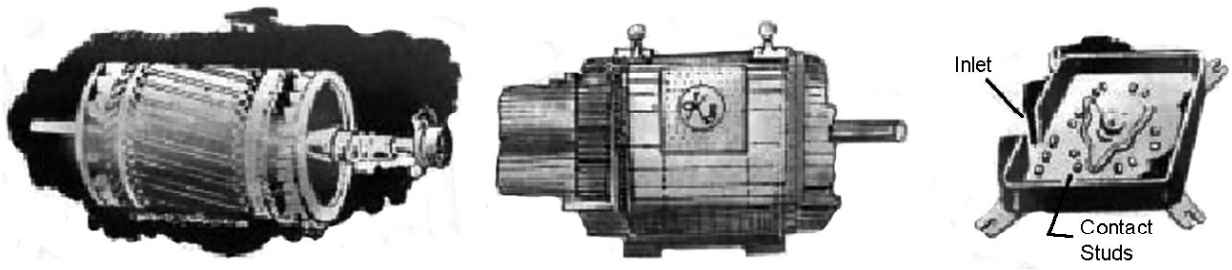


Fig. 3.13

In small motors, another method of construction is used. It consists of placing the entire rotor core in a mould and casting all the bars and end-rings in one piece. The meal commonly used is an aluminium alloy.

Another form of rotor consists of a solid cylinder of steel without any conductors or slots at all. The motor operation depends upon the production of eddy currents in the steel rotor.

This type of rotor is provided with 3-phase, double-layer, distributed winding consisting of coils as used in alternators. The rotor is wound for as many poles as the number of stator poles and is always wound 3-phase even when the stator is wound two-phase.



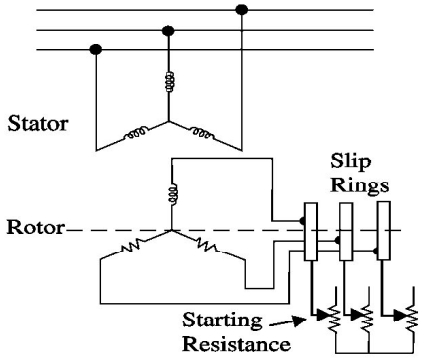
(a) **(b)** **(c)**

Fig. 3.14b. Slip-ring motor with slip-rings brushed and short-circuiting devices

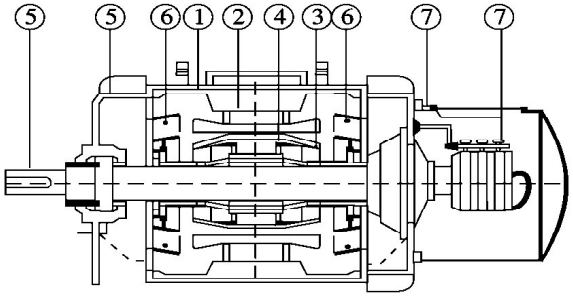
The three phases are starred internally. The other three winding terminals are brought out and connected to three insulated slip-rings on them [Fig. 3.14(b)]. These three brushes are further externally connected to a 3-phase star-connected the rheostat [Fig. 3.14(c)]. This makes possible the introduction of additional resistance

in the rotor circuit during the starting period for increasing the starting torque of the motor, as shown in Fig.3.14(a) (Ex.34.7 and 34.10 and for changing is speed-torque/current characteristics. When running under normal conditions, the slip-rings are automatically short-circuited by means of a metal collar, which is pushed along the shaft and connects all the rings together. Next, the brushes are automatically lifted from the slip-rings to reduce the frictional losses and the wear and tear. Hence, it is seen that under normal running conditions, the wound rotor is short-circuited on itself just like the squirrel-case rotor.

Fig. 3.15(b) shows the longitudinal section of a slip-ring motor, whose structural details are as under:



(a)



(b)

Fig. 3.16(a)

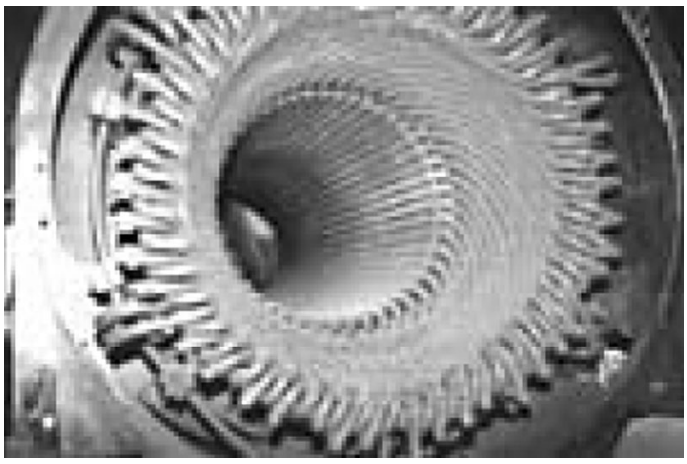
Fig. 3.16 (b) longitudinal section of a Jyoti splash-proof slip-ring motor

1. **Frame.** Made of close-grained alloy cast iron.
2. **Stator and Rotor Core.** Built from high-quality low-loss silicon steel laminations and flash enamelled on both sides.
3. **Stator and Rotor windings.** Have moisture proof tropical insulation embodying mica and high quality varnishes. Are carefully spaced for most effective air circulation and are rigidly braced to withstand centrifugal forces and any short-circuit stresses.
4. **Air-gap.** The stator rabbets and bore are machined carefully to ensure uniformity of air-gap.
5. **Shafts and bearings.** Ball and roller bearings are used to suit heavy duty, trouble-free running and for enhanced service life.
6. **Fans.** Light aluminium fans are used for adequate circulation of cooling air and are securely keyed onto the rotor shaft.
7. **Slip-rings and Slip-ring Enclosures.** *Slip-rings are made of high quality phosphor-bronze and are of moulded construction.*

Why Does the Rotor Rotate

The reason why the rotor of an induction motor is set into rotation is as follow:

When the 3-phase stator windings, are fed by a 3-phase magnitude, but rotating at synchronous speed, is set up. The flux passes through the air-gap, sweeps past the rotor surface and so cuts the rotor conductors which, as yet, are stationary. Due to the relative speed between the rotating flux and the stationary conductors, an e.m.f. is induced in the latter, according to Faraday's laws of electro-magnetic induction. The frequency of the induced e.m.f. is the same as the supply frequency. Its magnitude is proportional to the relative velocity between the flux and the conductors and its direction si given by Fleming's Right-hand rule.



Winding of induction electric motor

Since the rotor bars or conductors form a closed circuit, rotor current is produced whose direction, as given at Lenz's law, is such as to oppose the very cause producing it. In this case, the cause which produces the rotor current is the relative velocity between the rotating flux of the stator and the stationary rotor conductors. Hence, to reduce the relative speed, the rotor starts running in the same direction as that of the flux and tries to catch up with the rotating flux.

The setting up of the torque for rotating the rotor is explained below:

In Fig.3.16(a) is shown the stator field which is assumed to be rotating clockwise. The relative motion of the rotor with respect to the stator is anticlockwise. By applying right-hand rule, the direction of the induction e.m.f. in the rotor is found to be outwards. Hence, the direction of the flux due to rotor current alone, is as shown in Fig.3.16(b). Now, by applying the Left-hand rule, or by the effect of combined field [Fig.3.16(c)] it is clear that the rotor conductors experience a force tending to rotate them in clockwise direction. Hence, the rotor is set into rotation in the same direction as that of the stator flux (or field).

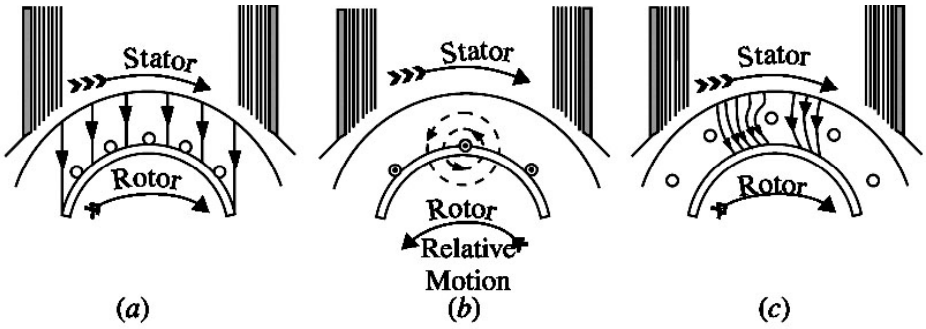


Fig. 3.16

Slip

In practice, the rotor never succeeds in 'catching up' with the stator field. If it really did so, then there would be no relative speed between the two, hence rotor e.m.f., no roto current and so no torque to maintain rotation. That is why the rotor runs at a speed which is always less than the speed of the stator field. The difference in speeds depends upon the load on the motor.

The difference between the synchronous speed N_s and the actual speed n of the rotor is known as slip. Though it may be expressed in so many revolutions/second, yet it is usual to express it as a percentage of the synchronous speed. Actually, the term 'slip' is descriptive of the way in which the rotor 'slips back' from synchronism.

$$\% \text{ slip} \quad s = \frac{N_s - N}{N_s} \times 100$$

Sometimes, $N_s - N$ is called the **slip speed.**

Obviously, rotor (or motor) speed is $N = N_S (1–s)$

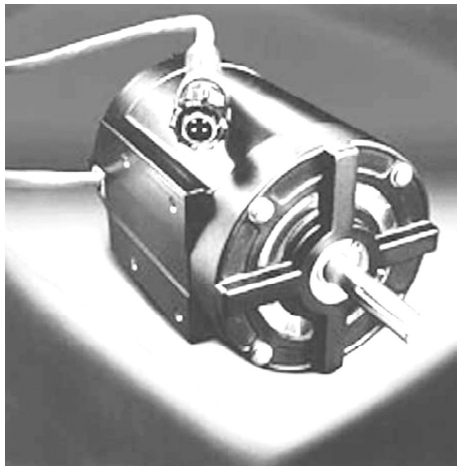
It may be kept in mind that revolving flux is rotating synchronously, relative to the stator (i.e. stationary space) but at slip speed relative to the rotor.

Synchronous Motor - General

A synchronous motor (Fig.3.17) is electrically identical with an alternator or a.c. generator. In fact, a given synchronous machine may be used, at least theoretically, as an alternator, when driven mechanically or as a motor, when driven electrically, just as in the case of d.c. machines. Most synchronous motors are rated between 150 kW and 15 MW and run at speeds ranging from 150 to 1800 r.p.m.

Some characteristic features of a synchronous motor are worth nothing:

1. It runs either at synchronous speed or not at all i.e. while running it maintains a constant speed. The only way to change its speed is to vary the supply frequency (because Ns=120f/P).
2. It is not inherently self-starting. It has to be run upto synchronous (or near synchronous) speed by some means, before it can be synchronized to the supply.



Synchronous motor

3. It is capable of being operated under a wide range of power factors, both lagging and leading. Hence, it can be used for power correction purposes, in addition to supplying torque to drive loads.

Principle of Operation

As shown, when a 3-ϕ winding is fed by a 3-ϕ supply, then a magnetic flux of constant magnitude but rotating at synchronous speed, is produced. Consider a two-pole stator of Fig.3.18, in which are shown two stator poles (marked N_s and S_s) rotating at synchronous speed, say, in clockwise direction. With the rotor position as shown, suppose the stator poles are at that instant situated at points A and B. The two similar poles, N (of rotor) and N_s (or stator) as well as S and S_s will repel each other, with the result that the rotor tends to rotate in the anticlockwise direction.

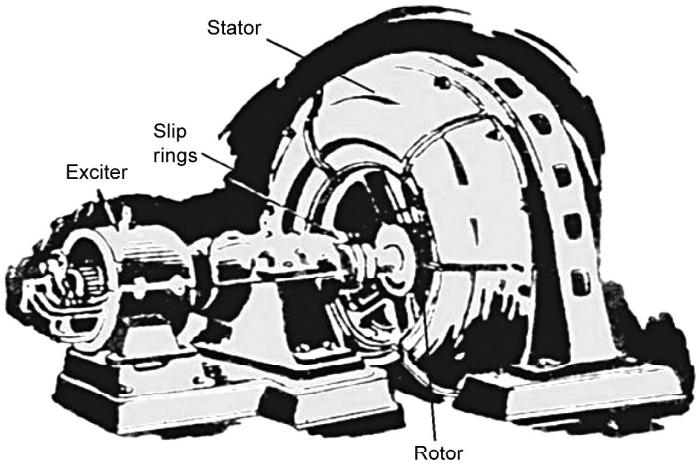


Fig. 3.17

But half a period later, stator poles, having rotated around, interchange their positions i.e. N_s is at point B and S_s at point A. Under these conditions, N_s attracts S and S_s attracts N. Hence, rotor tends to rotate clockwise (which is just the reverse of the first direction). Hence, we find that due to continuous and rapid rotation of stator poles, the rotor is subjected to a torque which is rapidly reversing i.e., in quick succession, the rotor si subjected to torque which tends to move it first in one direction and then in the opposite direction. owing to its large inertia, the rotor cannot instantaneously respond to such quickly-reversing torque, with the result that it remains stationary.

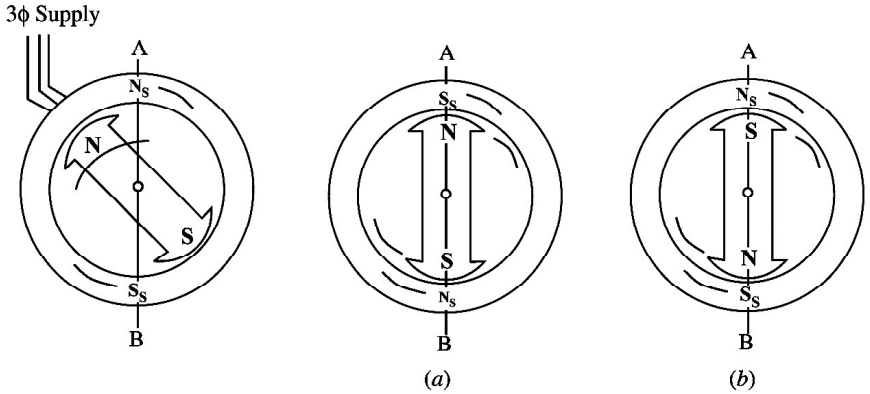
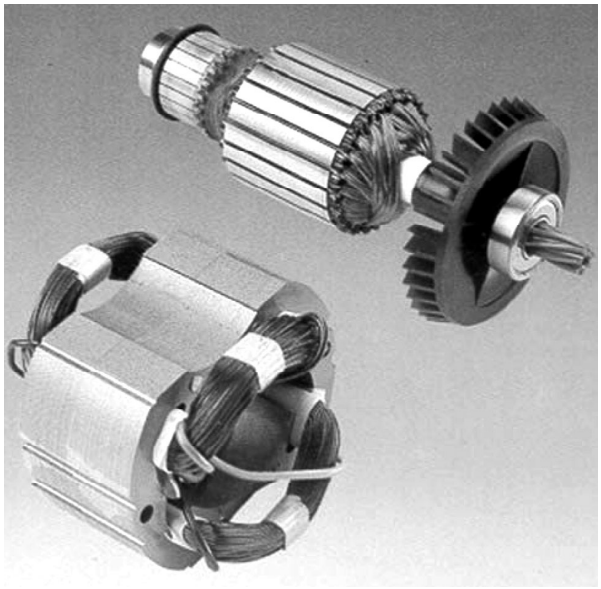


Fig. 3.19 **Fig. 3.19** **Fig. 3.19**

Now, consider the condition shown in Fig.3.19(a). The stator and rotor poles are attracting each other. Suppose that the rotor is not stationary, but is rotating clockwise, with such a speed that it turns through one pole-pitch by the time the stator poles interchange their positions, as shown in Fig.3.19(b). Here , again the stator and rotor poles attract each other. It means that if the rotor poles also shift their positions along with the stator poles, then they will continuously experience a unidirectional torque i.e., clockwise torque, as shown in Fig. 3.19.

Method of Starting

The rotor (which is as yet unexcited) is speeded up to synchronous/near synchronous speed by some arrangement and then excited by the d.c. source. The moment this (near) synchronously rotating rotor is excited, it is magnetically locked into position with the stator i.e., the rotor poles are engaged with the stator poles and both run synchronously in the same direction. It is because of this interlocking of stator and rotor poles that the motor has either to run synchronously or not at all. The synchronous speed is given by the usual relation $N_s = 120\ f/p$.



The rotor and the stator parts of motor

However, it is important to understand that the arrangement between the stator and rotor poles is not an absolutely rigid one. As the load on the motor is increased, the rotor progressively tends to fall back in phase (but not in speed as in d.c. motors) by some angle (Fig. 3.20) but it still continues to run synchronously. The value of this load angle or coupling angle (as it is called) depends on the amount of load to be met by the motor. In other words, the torque developed by the motor depends on this angle, say, α.

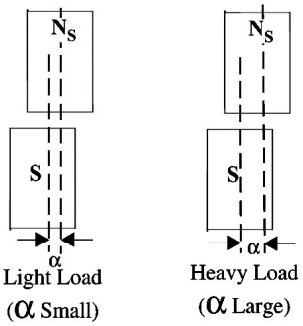


Fig. 3.20

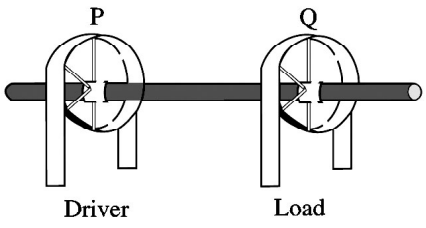


Fig. 3.21

The working of a synchronous motor is, in many ways, similar to the transmission of mechanical power by a shaft. In Fig. 3.21 are shown two pulleys P and Q transmitting power from the driver to the load. The two pulleys are assumed to be keyed together (just as stator and rotor poles are interlocked) hence they run at exactly the same (average) speed. When Q is loaded, it slightly falls behind owing to the twist in the shaft (twist angle corresponds to α in motor), the angle of twist, in fact, being a measure of the torque transmitted. It is clear that unless Q is so heavily loaded as to break the coupling, both pulleys must run at exactly the same (average) speed.

■■■

UNIT - 4

..

CHAPTER - 4
CONDUCTION IN SEMINCONDUCTORS

DEFINITION OF SEMICONDUCTOR

Semiconductors

Semiconductors (e.g. germanium, silicon etc.) are those substances whose electrical conductivity lies in-between conductors and insulators. In terms of energy band, the valence band is almost empty. Further, the energy gap between valence and conduction bands is very small as shows in Fig. 4.1. Therefore, comparatively smaller electric field (smaller than insulators but much greater than conductors) is required to push the electrons from the valence band to the conduction band. In short, a semiconductor has:

(a) almost full valence band
(b) almost empty conduction band
(c) Small energy gap (~ 1 eV) between valence band and conduction bands.

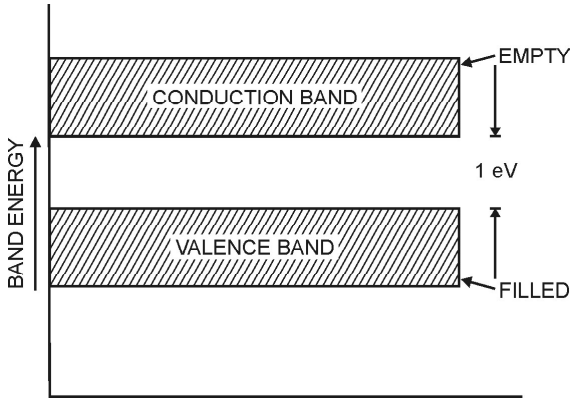


Fig. 4.1

Energy Band Description of Semiconductors

It has already been discussed that a seminconductor is a substance whose resistivity lies between conductors and insultors. The resistivity is of the order of 10^{-4} to 0.5 ohm metre. However, a semiconductor can be defined much more comprehensively on the basis of energy bands as under:

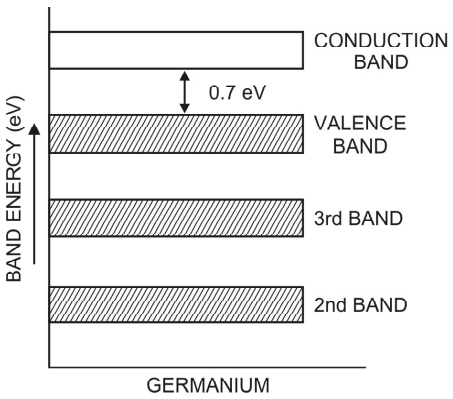


Fig. 4.2

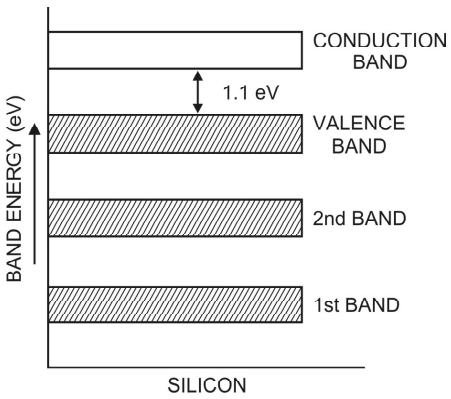


Fig. 4.3

A semiconductor is a substance which has almost filled valence band and nearly empty conduction band with a very small energy gap (~ 1 eV) separating the two.

Fig. 4.2 and 4.3 shows the energy band diagrams of germanium and silicon respectively. It may be seen that forbidden energy gap is very small; being 1.1 eV for silicon and 0.7 eV for germanium. Therefore, relatively small energy is needed by thrie calwnce electrons to cross over to the conduction band. Even at room temperature, some of the valence electrons may acquire sufficient energy to enter into the conduction band

and thus become free electrons. However, at this temperature, the number of free electrons available is very small. Therefore, at room temperature, a piece of germanium or silicon in neither a good conductor nor an insulator. For this reason, such substances are called semiconductors.

The energy band description is extremely helpful in undrstanding the current flow through a semiconductor. Therefore, we shall frequently use this concept in our further discussion.

Majority and Minority Carriers

It has already been discussed that due to the effect of impurity, n-type material has a large number of free electrons whereas p-type material has a large number of holes. However, it may be recalled that even at room temperature, some of the co-valent bonds break, thus releasing an equal number of free electrons and holes. An n-type material has its share of electron-hole pairs (released due to breaking of bonds at room temperature) but in addition has a much larger quantity of free electrons due to the effect of impurity. These impurity-caused free electrons are not associated with holes. Consequently, an n-type material has a large number of free electrons and a small number of holes as shown in Fig. 4.4 (i). The free electrons in this case are considered majority carriers – since the majority portion of current in n-type material is by the flow of free electrons – and the holes are the minority carriers.

Similarly, in a p-type material, holes outnumber the free electrons as shown in Fig. 4.4 (ii). Therefore, holes are the majority carriers and free electrons are the minority carriers.

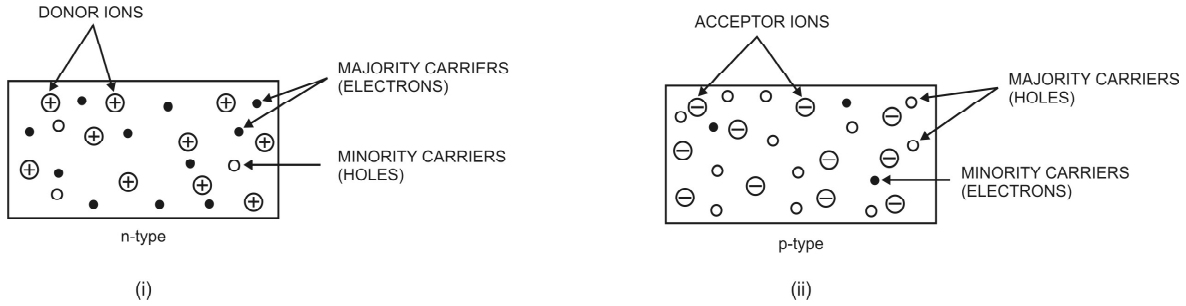


Fig. 4.4

CONDUCTION PROPERTIES OF SEMICONDUCTORS

TYEPS OF SEMICONDUCTORS

Instrinc Semiconductor

A semiconductor in an extremely pure form is known as an instrict semiconductor.

In an intrinsic semiconductor, even at room temperature, hole-electron pairs are created. When electric field is applied across an intrinsic semiconductor, the current conduction takes place by two process, namely; by free electrons and holes as shown in Fig. 4.5. The free electrons are produced due to the breaking up of some covalent bonds by thermal energy. At the same time, holes are created in the covalent bonds. Under the influecnce of electric field, conduction through the semiconductor is by both free electrons and holes. Therefore, the total current inside the semiconductor is the sum of currents due to free electrons and holes.

It may be noted that current in the external wires is fully electronic i.e. by electrons. What about the holes? Referring to Fig. 4.5, holes being positively charged move towards the negative terminal of supply. As the holes reach the negative terminal B, electrons enter the semiconductor crystal near the terminal and combine with holes, thus cancelling them. At the same time, the loosely held electrons near the positive terminal A are attracted away from their atoms into the positive terminal. This creates new holes near the positive terminal which again drift towards the negative terminal.

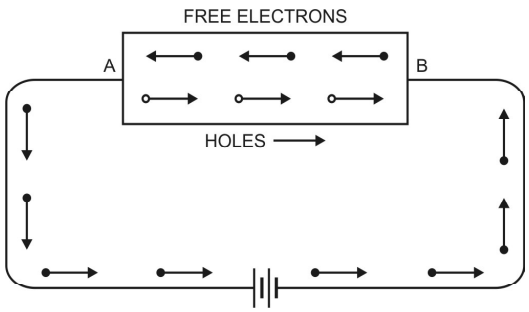


Fig. 4.5

Extrinsic Semiconductor

The intrinsic semiconductor has little current conduction capability at room temperature. To be useful in electronic devices, the pure semiconductor must be altered so as to significantly increase its conducting properties. This is achieved by adding a small amount of suitable impurity to a semiconductor. It is then called impurity or extrinsic semiconductor. The process of adding impurities to a semiconductor is known as doping. The amount and type of such impurities have to be closely controlled during the preparation of extrinsic semiconductor. Generally, for 10^8 atoms of semiconductor, one impurity atom is added.

The purpose of adding impurity is to increase either the number of free electrons or holes in the semiconductor crystal. As we shall see, if a pentavalent impurity (having 5 valence electrons) is added to the semiconductor, a large number of free electrons are produced in the semiconductor. On the other hand, addition of trivalent impurity (having 3 valence electrons(creates a large number of holes in the semiconductor crystal. Depending upon the type of impurity added, extrinsic semiconductors are classified into:

i. n-type semiconductor ii. p-type semiconductor

n-type Semiconductor

When a small amount of pentavalent impurity is added to a pure seminconductor, it is known as **n-type semiconductor.**

The addition of pentavalent impurity provides a large number of free electrons in the semiconductor crystal. Typical examples of petavalent impurities are arsenic (At. No. 33) and antimony (At. No. 51). Such impurities which produce n-type semiconductor are known as donor impurities because they donate or provide free electrons to the semiconductor crystal.

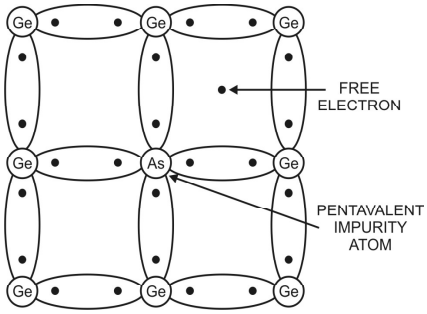


Fig. 4.6

To explain the formation of n-type semiconductor, consider a pure germanium crystal. We know that germanium atom has four valence electrons. When a small amount of pentavalent impurity like arsenic is added to germanium crystal, a large number of free electrons become available in the crystal. The reason is simple. Arsenic is pentavalent i.e. its atom has five valence electrons. an arsenic atom fits in the germanium crystal in such a way that its four valence electrons form covalent bonds with four germanium atoms. The fifth valence electron of arsenic atom finds no place in co-valent bonds and is thus free as shown in Fig. 4.6. Therefore, for each arsenic atom added, one free electron will be available in the germanium crystal. Though each arsenic atom provides one free electron, yet an extremely small amount of arsenic impurity provides enough atoms to supply millions of free electrons.

Fig. 4.7 shows the energy band description of n-type semiconductor. The addition of pentavalent impurity has produced a number of conduction band electrons i.e., free electrons. The four valence electrons of pentavalent atom form covalent bonds with four neighbouring germanium atoms. The fifth left over valence electron of the pentavalent atom cannot be accommodated in the valence band travels to the conduction band. The following points may be noted carefully:

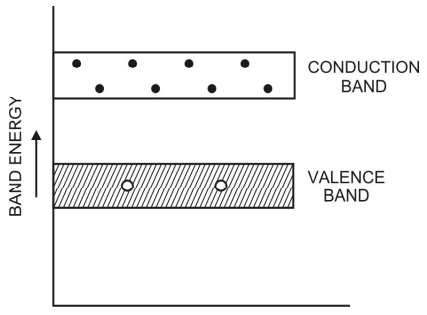


Fig. 4.7

i. Many new free electrons are produced by the addition of pentavalent impurity.

ii. Thermal energy of room temperature still generates a few hole-electron pairs. However, the number of free electrons provided by the pentavalent impurity far exceeds the number of holes. It is due to this predominance of electrons over holes that it is called n-type semiconductor (n stands for negative).

n-type conductivity. The current conduction in an n-type semiconductor is predominatly by free electrons i.e. negative charges and is called n-type or electron type conductivity. To understand n-type conductivity, refer to Fig. 4.8. When p.d. is applied across the n-type semiconductor, the free electrons (donated by impurity) in the crystal will be directed towards the positive terminal, constituting electric current. As the current flow through the crystal is by free electrons which are carriers of negative charge, therefore, this type of conductivity is called negative or n-type conductivity. It may be noted that conduction is just as in ordinary metals like copper.

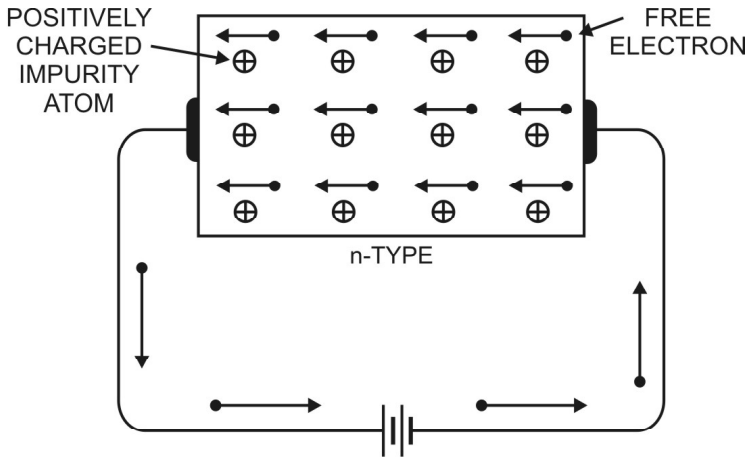


Fig. 4.8

p-type Semiconductor

When a small amount of trivalent impurity is added to a pure semiconductor, it is called p-type semiconductor.

The addition of trivalent impurity provides a large number of holes in the semiconductor. Typical examples or trivalent impurities are gallium (At. No. 31) and indium (At. No. 49). Such impurities which produce p-type semiconductor are known as acceptor impurities because the holes created can accept the electrons.

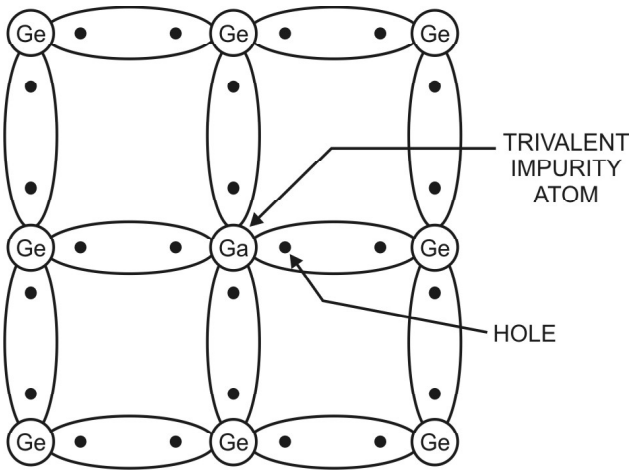


Fig. 4.9

To explain the formation of p-type semiconductor, consider a pure germanium crystal. When a small amount of trivalent impurity like gallium is added to germanium crystal, there exists a large number of holes in the crystal. The reason is simple. Gallium is trivalent i.e. its atom has hree valence electrons. Each atom of gallium fits into the germanium crystal but now only three co-valent bonds can be formed. It is because three valence electrons of gallium atom can form only three single co-valent bonds with three germanium atoms as shown in Fig. 4.9. In the fourth co-valent bond, only germanium atom gallium has no valence eletron to contribute

as all its three valence electrons are already engaged in the co-valent bonds with neighbouring germanium atoms. In other words, fourth bond is incomplete; being short of one electron. This missing electron is called a hole. Therefore, for each gallium atom added, one hole is created. A small amount of gallium provides millions of holes.

Fig. 4.10 shows the energy band description of the p-type semiconductor. The addition of trivalent impurity has produced a large number of holes. However, there are a few conduction band electrons due to thermal energy associated with room temperature. But the holes far outnumber the conduction band electrons. It is due to the predominance of holes over free electrons that it is called p-type semiconductor (p stands for prositive).

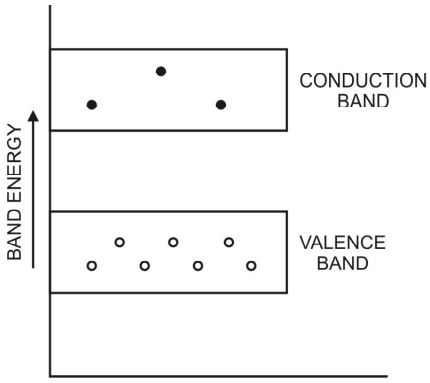


Fig. 4.10

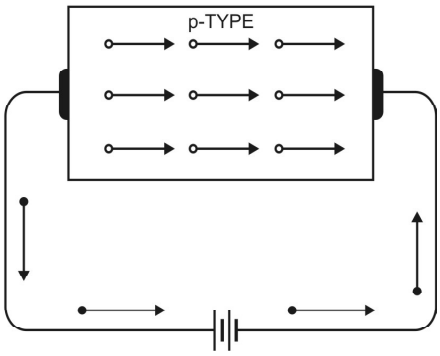


Fig. 4.11

p-type conductivity. The current conduction in p-type semiconductor is predominantly by holes i.e. positive charges and is called p-type or hole-type conductivity. To understand p-type conductivity, refer to Fig. 4.11. When p.d. is applied to the p-type semiconductor, the holes (donated by the impurity) are shifted from one co-valent bond to another, constituting what is known as hole current. It may be noted that in p-type conductivity, the valence electrons move from one co-valent bond to another unlike the n-type where current conduction is by free electrons.

Properties of Semiconductors

i. The resistivity of a semiconductor is less than an indulator but more than a conductor.
ii. Semiconductors have negative temperature co-efficient of resistance i.e. the resistance of a semiconductor decreases with the increase in temperature and vice-versa. For example, germanium is actually an insulator at low temperatures but it becomes a good conductor at high temperatures.
iii. When suitable metallic impurity (e.g. arsenic, gallium etc.) is added to a semiconductor, its current conducting properties change appreciably. This property is most important and is discussed later detail.

Effect of Temperature on Semiconductors

The electrical conductivity of a semiconductor changes appreciably with temperature variations. This is a very important point to keep in mind.

i. At absolute zero. At absolute zero temperature, all the electrons are tightly held by the semiconductor atoms. The inner orbit electorns are bound whereas the valence electrons are engaged in co-valent bonding. At this temperature, the co-valent bonds are very strong and there are no free electrons. Therefore, the semiconductor crystal behaves as a perfect insulator [See Fig. 4.12 (i)].
   In terms of energy band description, the valence band is filled and there is a large energy gap between valence band and conduction band. Therefore, no valence electron can reach the conduction band to become free electron. It is due to the non-availability of free electrons that a semiconductor behaves as an insulator.
ii. Above absolute zero. When the temperature is raised, some of the covalent bonds in the semiconductor break due to the thermal energy supplied. The breaking of bonds sets those electrons free which are engaged in the formation of these bonds. The result is that a few free electrons exist in the semiconductor. These free electrons can constitute a tiny electric current if potential difference is applied across the semiconductor crystal [See Fig. 4.12 (i)]. This shows that the resistance of a semiconductor decreases with the rise in temperature i.e. it has negative temperature coefficient of resistance. It may be added that at room temperature, current through a semiconductor is too small to be of any practical value.

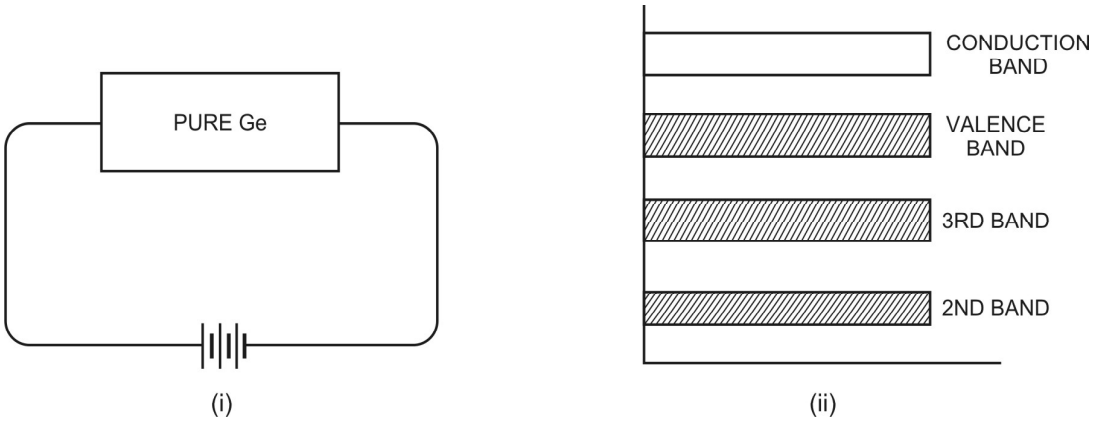


Fig. 4.12

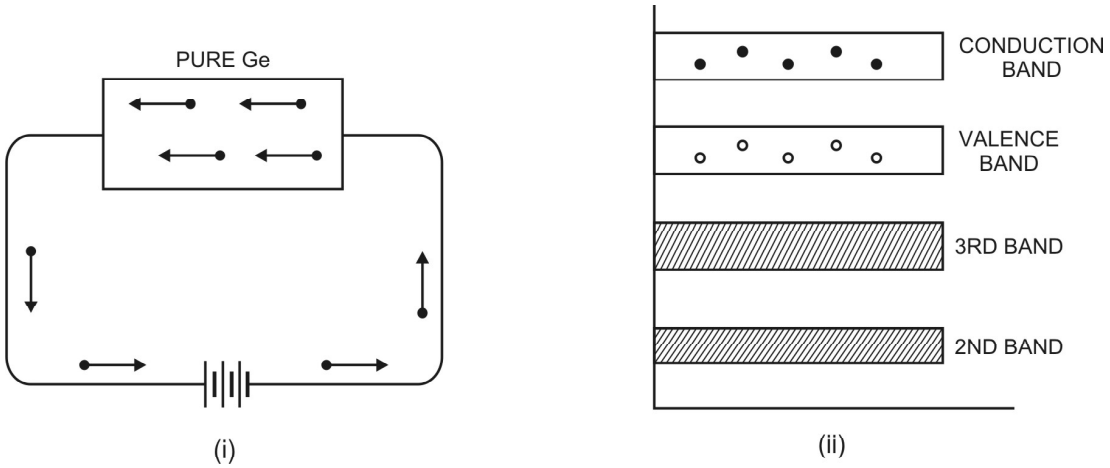


Fig. 4.13

Fig. 4.13 (ii) shows the energy band diagram. As the temperature is raised, some of the valence electrons acquire sufficient energy to enter into the conduction band and thus become free electrons. Under the influence of electric field, these free electrons will constitute electric current. It may be noted that each time a valence electron enters into the conduction band, a hole is created inthe valence band. As we shall see in the next article, holes also contribute to current. In fact, hole current isthe most significant concept in semiconductors.

PN JUNCTION DIODE

Definition

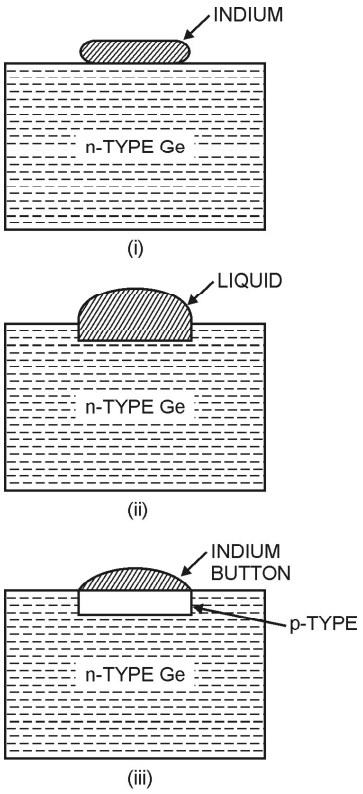
When a p-type semiconductor is suitably joined to n-type semiconductor, the contact surface is called pn junction.

Most semiconductor devices contain one or more pn junctions. The pn junction is of great importance because it is in effect, the control element for semiconductor devices. A thorough knowledge of the formation and properties of pn junction can enable the reader to understand the semiconductor devices.

Formation of pn junction

In actual practice, the characteristic properties of pn junction will not be apparent if a p-type block is just brought in contact with n-type block. In fact, pn junction is fabricated by special techniques. One common method of making pn junction is called alloying. In this method, a small block of indium (trivalent impurity) is placed on an n-type germanium slab as shown in Fig. 4.14.

i. The system is then heated to a temperature of about 500C. The indium and some of the germanium melt to from a small puddle of molten germanium-indium mixture as shown in Fig. 4.14.
ii. The temperature is than lowered and puddle begins to solidify. Under proper conditions, the atoms of indium impurity will be suitably adjusted in the germanium slab to form a single crystal. The addition of



indium overcomes the excess of electrons in the n-type germanium to such an extent that it creates a p-type region.

As the proces goes on, the remaining molten mixture becoms increasingly rich in indium. When all germanium has been redeposited, the remaining material appears as indium button which is frozen on the outer surface of the crystalised portion as shown in Fig. 4.14.

iii. This button serves as a suitable base for soldering on leads.

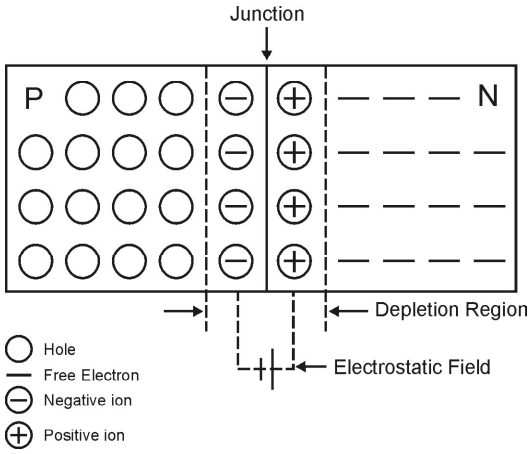


Fig. 4.14

Properties of *pn* Junction

At the instant of pn-junction formation, the free electrons near the junction in the n region begin to diffuse across the junction into the p region where they combine with holes near the junction. The result is that n region loses free electrons as they diffuse into the junction. This creates a layer of positive charges (pentavalent ions) near the junction. As the electrons move across the junction, the p region loses holes as the electrons and holes combine. The result is that there is a layer of negative charges (trivalent ions) near the junction. These two layers of positive and negative charges from the depletion region (or depletion layer). The term depletion is due to the fact that near the junction, the region is depleted (i.e. emptied) of charge carries (free electrons and holes) due to diffusion across the junction. It may be noted that depletion layer is formed very quickly and is very thin compared to the n region and the p region and the p region. For clarity, the width of the depletion layer is shown exaggerated.

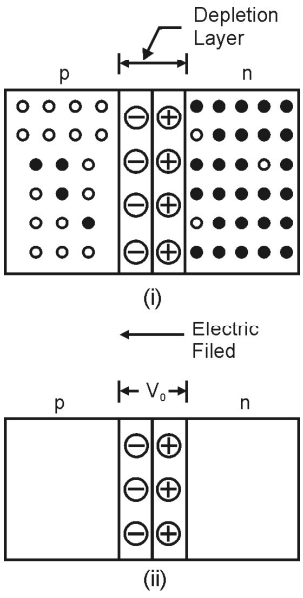


Fig. 4.15

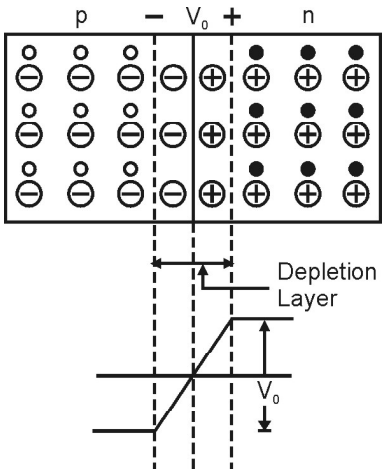


Fig. 4.16

Once pn junction is formed and depletion layer created, the diffusion of free electrons stops. In other words, the depletion region acts as a barrier to the further movement of free electrons across the junction. The positive and negative charges set up an electric field. This is shown by a black arrow in Fig. 4.15 (i). The electric field is a barrier to the free electrons in the n-region. There exists a potential difference across the depletion layer and is called barrier potential (V_0). The barrier potential of a pn junction depends upon several factors including the type of semiconductor material, the amount of doping and temperature. The typical barrier potential is approximately:

For silicon, V_0 = 0.7 V; For germanium, V_0 = 0.3 V

Fig. 4.16 shows the potential (V_0) distribution curve.

BEHAVIOUR OF PN JUNCTION

Applying D.C.Voltage Across pn Junction or Biasing a pn Junction

In electronics, the term bias refers to the use of d.c. voltage to establish certain operating conditions for an electronic device. In relation to a pn junction, there are following two bias conditions

1. Forward biasing **2. Reverse biasing**

Forward biasing

When external d.c. voltage applied to the junction is in such a direction that it cancels the potential barrier, thus permitting current flow, it is called forward biasing.

To apply forward bias, connect positive terminal of the battery to the battery to p-type and negative terminal to n-type as shown in Fig. 4.17. The applied forward potential establishes an electric field which acts against the field due to potential barrier. Therefore, the resultant field is weakened and the barrier height is reduced at the junction as shown in Fig. 4.17. As potential barrier voltage is very small (0.1 to 0.3 V), therefore, a small forward voltage is sufficient to completely eliminate the barrier. Once the potential barrier is eliminated by the forward voltage, junction resistance becomes almost zero and a low resistance path is established for the entire circuit. Therefore, current flows in the circuit. This is called forward current. With forward bias to pn junction, the following points are worth noting:

i. The potential barrier is reduced and at some forward voltage (o.1 to 0.3), it is eliminated altogether.
ii. The junction offers low resistance (called forward resistance, R_f) to current flow.
iii. Current flows in the circuit due to the establishment of low resistance path. The magnitude of current depends upon the applied forward voltage.

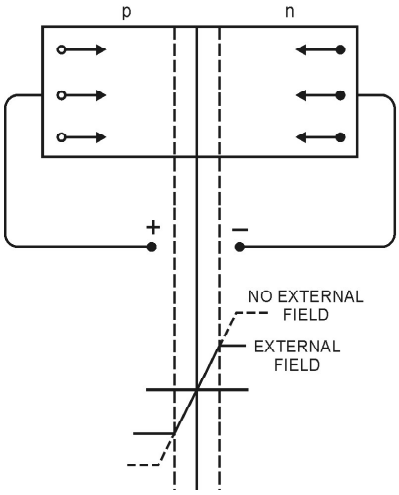


Fig. 4.17

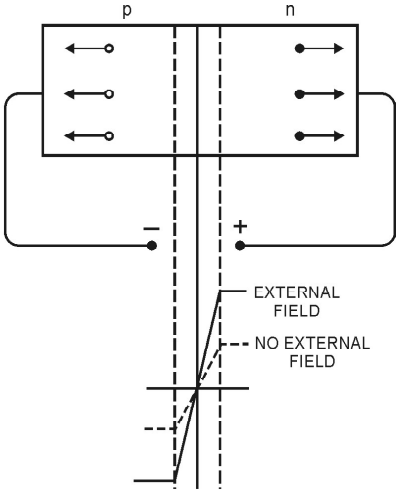


Fig. 4.18

Reverse Biasing

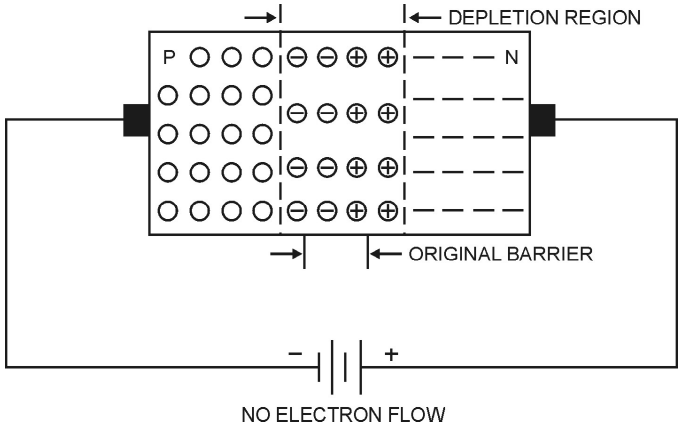
When the external d.c. voltage applied to the junction is in such direction that potential barrier is increased, it is called reverse biasing.

To apply reverse bias, connect negative terminal of the battery to p-type and positive terminals n-type as shown in Fig. 4.18. It is clear that applied reverse voltage establishes an electric field which acts in the same direction as the filed due to potential barrier. Terefore, the resultant field at the junction is strengthened and the barrier height is increased as shown in Fig. 4.18. The increases potential barrier prevents the flow of charge

carriers across the junction. Thus, a high resistance part is established for the entire circuit and hence the current does not flow. With reverse bias to junction, the following points are worth noting:

i. The potential barrier is increased.
ii. The junction offers very high resistance (called reverse resistance, R_r) to current flow.
iii. No current flows in the circuit due to the establishment of high resistance path.

Conclusion. From the above discussion, it follow that with reverse biasto the junction, a high resistance path id established and hence no current flow occurs. On the other hand, with forward bias to the junction, a low resistance path is set up and hence current flows in the circuit.



V.I. CHARACTERISTICS OF ZENER DIODE

It has already discussed that when the reverse bias as a crystal diode is increased, a critical voltage, called breakdown voltage is reached where the reverse current increases sharply to a high value. The breakdown region is the knee of the reverse characteristic as shown in Fig. 4.19. The satisfactory explanation ofthis breakdown of the junction was first given by the American scientist C. Zener. Therefore, the breakdown voltage is sometimes called zener voltage and the sudden increase in current is known as zener current.

The breakdown or zener voltage depends upon the amount of doping. If the diode is heavily doped, depletion layer will be thin and consequently the breakdown of the junction will occur at a lower reverse voltage. On the other hand, a lightly doped diode has higher breakdown voltage. When an ordinary crystal diode is properly doped so that it has a sharp breakdown voltage, is called a zener diode.

A properly doped crystal diode which has a sharp breakdown voltage is known as a zener diode.

Fig. 6.53 shows the symbol of a zener diode. It may be seen that it is just like an ordinary diode except that the bar is turned into z-shape. The following points may be noted about the zener diode:

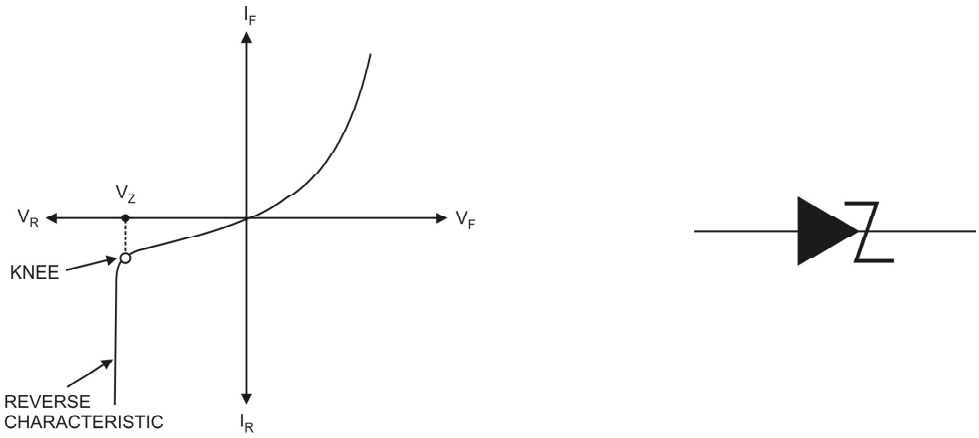


Fig. 4.19

i. A zener diode is like an ordinary diode except that it is properly doped so as to have a sharp breakdown voltage.
ii. A zener diode is always reverse connected i.e. it is always reverse biased.
iii. A zener diode has sharp breakdown voltage, called zener voltage Vz.

iv. When forward biased, its characteristics are just those of ordinary diode.
v. The zener diode is not immediately burnt just because it has entered the breakdown region. As long as the external circuit connected to the diode limits the diode current to less than burn out value, the diode will not burn out.

Equivalent Circuit of Zener Diode

The analysis of circuits using zener diodes can be made quite by replacing the zener diode by its equivalent circuit.

i. **"ON" state.** When reverse voltage across a zener diode is equal to or more than break down voltage V_Z, the current increases very sharply. In this region, the 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 can be represented by a battery of voltage V_Z as shown in Fig. 4.20 (ii). Under such conditions, the zener diode is said to be in the "ON" state.

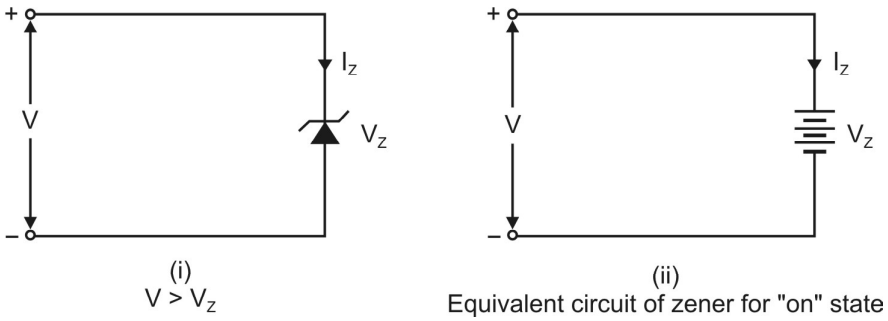


Fig. 4.20

ii. **"OFF" state.** When the reverse voltage across the zener diode is less than V_Z but greater than 0V, the zener diode is in "OFF" state. Under such conditions, the zener diode can be represented by an open-circuit as shown in Fig. 4.21 (ii).

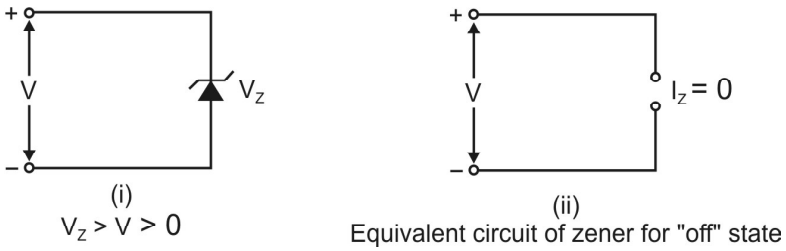


Fig. 4.21

Zener Diode as Voltage Stabiliser

A zener diode can be used as a voltage regulator to provide a constant voltage from a source whose voltage may vary over sufficient range. The circuit arrangement is shown in Fig. 4.22 (i). The zener diode of zener diode of zener voltage V_Z is reverse connected across the load R_L across which constant output is desired. The series resistance R absorbs the output voltage fluctuations so as to maintain constant voltage across the load. It may be noted that the zener will maintain a constant voltage V_Z (= E_0) across the load so long as the input voltage does not fall below V_Z.

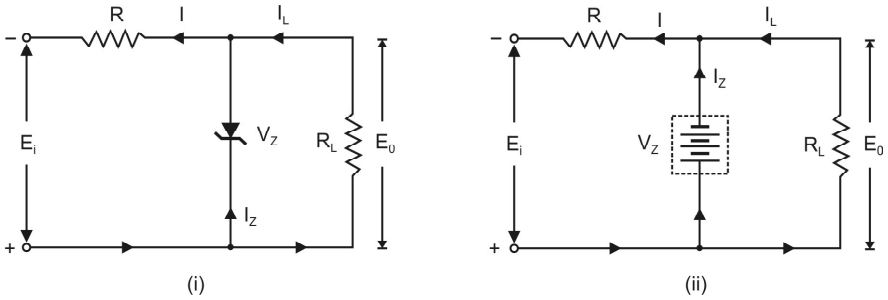


Fig. 4.22

When the circuit is properly designed, the load voltage E_0 remains essentially constant (equal to V_Z) even though the input voltage E_i and load resistance R_L may vary over a wide range.

i. Suppose the input voltage increases. Since the zener is in the breakdown region, the zener diode is equivalent to a battery V_Z as shown in Fig. 4.22 (ii). It is clear that output voltage remains constant at $V_Z(= E_0)$. The excess voltage is dropped across the series resistance R. This will cause an increase in the value of total current I. The zener will conduct the increase of current in I while the load current remains constant. Hence, output voltage E_0 remains constant irrespective of the changes in the input voltage E_i.

ii. Now suppose that input voltage is constant but the load resistance RL decreases. This will cause an increase in load current. The extra current come from the source because drop in R (and hence source current I) will not change as the zener is within its regulating range. The additional load current will come from a decrease in zener current IZ. Consequently, the output voltage stays at constant value.

Voltage drop across $R = E_i - E_0$

Current through R, $I = I_Z + I_L$

Applying Ohm's law, we have,

$$R = \frac{E_i - E_0}{I_Z + I_L}$$

Solving Zener Diode Circuits

The analysis of zener diode circuits is quite similar to that applied to the analysis of semiconductor diodes. The first step is to determine the state of zener diode i.e., whether the zener is in the "on" state of "off" state. Next, the zener is replaced by its appropriate model. Finally, the unknown quantities are determined from the resulting circuit.

E_i and R_L fixed. This is the simpelst case and is shown in Fig. 4.23 (i). Here the applied voltage E_i as well as load R_L is fixed. The first step is to find the state of zener diode. This can be determined by removing the zener from the circuit an calculating the voltage V across the resulting open-circuit as shown in Fig. 4.23 (ii).

If $V > V_Z$, the zener diode is in the "on" state and its equivalent model can be substituted as shown in Fig. 4.24 (i). If $V < V_Z$, the diode is in the "off" state as shown in Fig. 4.24 (ii).

$$V = E_0 \frac{R_L E_i}{R + R_L}$$

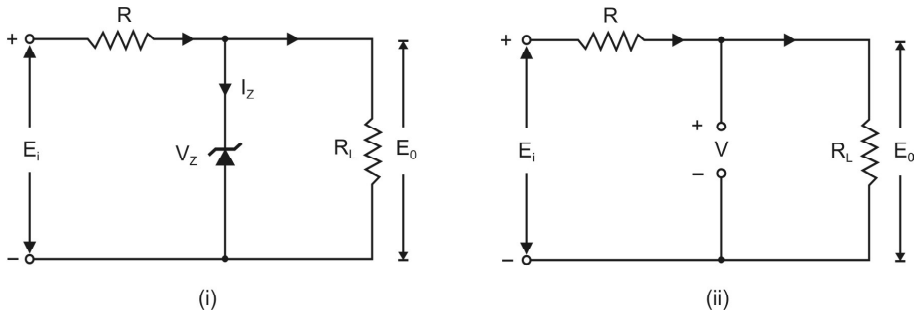


Fig. 4.23

If $V > V_Z$, the zener diode is in the "on" state and its equivalent model can be substituted as shown in Fig. 4.24 (i). If $V < V_Z$, the diode is in the "off" state as shown in Fig. 4.24 (ii).

i. **On State.** Referring to circuit shown in Fig. 4.24 (i),

$E_0 = V_Z$

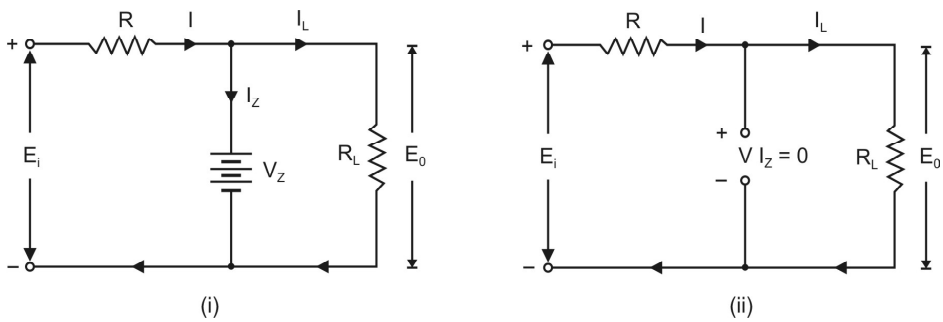


Fig. 4.24

$I_Z = I - I_L$ where $I_L = \frac{E_0}{R_L}$ and $I = \frac{E_i - E_0}{R}$

Power dissipated in zener, $P_Z = V_Z I_Z$

ii. **Off state.** Referring to the circuit shown in Fog. 4.24 (ii)

$I = I_L$ and $I_Z = 0$

$V_R = E_i - E_0$ and $V = E_0$ $(V < V_Z)$

$\therefore$ $P_Z = V I_Z = V(0) = 0$

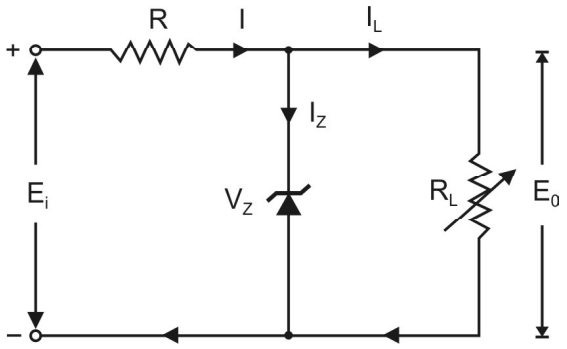


Fig. 4.25

Fixed E_i and Variable R_L. This case is shown in Fig. 4.25. Here the applied voltage (E_i) is fixed while load resistance R_L (and hence load current I_L) changes. Note that there is a definite range of R_L values (and hence I_L values) which will ensure the zener diode to be in "on" state. Let us calculate that range of values.

i. **R_{Lmin} and I_{Lmax}.** Once the zener is in the "on" state, load voltage E_0 (= V_Z) is constant. As a result, when load resistance is minimum (i.e., R_{Lmin}), load current will be maximum ($I_L = E_0/R_L$). In order to find the minimum load resistance that will turn the zener on, we simply calculate the value of R_L that will result in $E_0 = V_Z$ i.e.,

$$E_0 = V_Z = \frac{R_L E_i}{R + R_L}$$

$\therefore$ $$R_{L\min} = \frac{R V_Z}{E_i + V_Z}$$

This is the minimum value of load resistance that will ensure that zener is in the "on" state. Any value of load resistance less than this value will result in a voltage E_0 across the load less than V_Z and the zener will be in the "off" state.

Clearly ; $$I_{L\max} = \frac{E_0}{R_{L\min}} = \frac{V_Z}{R_{L\min}}$$

ii. **I_{Lmin} and R_{Lmax}.** It is easy to see that when load resistance is maximum, load current is minimum.

Now, Zener current, $I_Z = I - I_L$

When the zener is in the "on" state, I remains fixed. This means that when I_L is maximum, I_Z will be minimum. On the other hand, when I_L is minimum, I_Z is maximum. If the maximum current that a zener can carry is I_{ZM}, then,

$$I_{L\min} = I - I_{ZM}$$

and $$I_{L\max} = \frac{E_0}{I_{L\min}} = \frac{V_Z}{I_{L\min}}$$

If the load resistance exceeds this limiting value, the current through zener will exceed I_{ZM} and the device may burn out.

Fixed R_L and Variable E_i. This case is shown in Fig. 4.26. Here the load resistance R_L is fixed while the applied voltage (E_i) changes. Note that there is a definite range of E_i values that will ensure that zener diode is in the "on" state. Let us calculate that range of values.

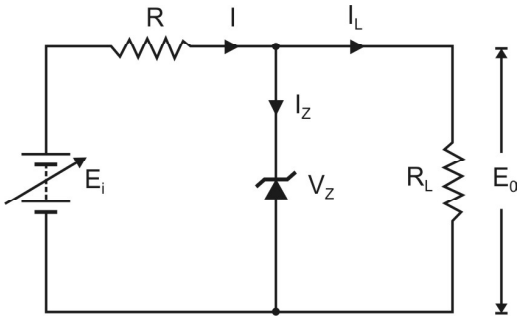


Fig. 4.26

i. **E_i (min).** To dtermine the minimum applied voltage that will turn the zener on, simply calculate the value of E_i that will result in load voltage $E_0 = V_Z$ i.e.,

$$E_0 = V_Z \frac{R_L E_i}{R + R_L}$$

$$\therefore \quad E_{i(min)} = \frac{(R + R_L) V_Z}{R_L}$$

ii. **E_i (max)**

Now, current through R, $I = I_Z + I_L$

Since I_L (= $E_0/R_L = V_Z/R_L$) is fixed, the value of I will be maximum when zener current is maximum i.e.,

$$I_{max} = I_{ZM} + I_L$$

Now $E_i = I R + E_0$

Since E_0 (= V_Z) is constant, the input voltage will be maximum when I is maximum.

$$E_{i(max)} = I_{max} R + V_Z$$

Example 1. For the circuit shown in Fig. 4.27 (i), find :

i. the output voltage ii. the voltage drop across series resistance

iii. the current through zener diode.

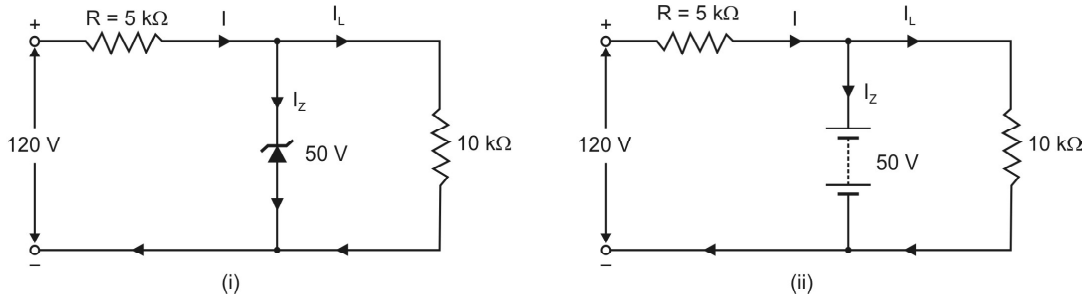


Fig. 4.27

Solution. If you remove the zener diode in Fig. 4.27(i), the voltage V across the open-circuit is given by:

$$V = \frac{R_L E_i}{R + R_L} = \frac{10 \times 120}{5 + 10} = 80\,V$$

Since voltage across zener diode is greater than V_Z (= 50 V), the zener is in the "on" state. It can therefore, be represented by a battery of 50V as shown in Fig. 4.27 (ii).

i. Referring to Fig. 4.27 (ii),

Output voltage = V_Z = 50 V

ii. Voltage drop across R = Input voltage – V_Z = 120 – 50 = **70 V**

iii. Load current, $I_L = V_Z/R_L$ = 50 V/10kΩ = 5 mA

Current through R, $I = \frac{70\,V}{5\,k\Omega} = 14\,mA$

Applying Krichhoff's first law, $I = I_L + I_Z$

$\therefore$ Zener current, $I_Z = I - I_L = 14 - 5 =$ **9 mA**

Example 2. For the circuit shown in Fig. 4.28 (i), find the maximum and mininmum values of zener diode current.

Solution. The first step to determine the state of the zener diode. It is easy to see that for the given range of voltages (80 – 120V), the voltage across the zener is greater than VZ (= 50 V). Hence the zener diode will be in the "on" state for this range of applied voltages. Consequently, it can be replaced by a battery of 50 V as shown in Fig. 4.28 (ii).

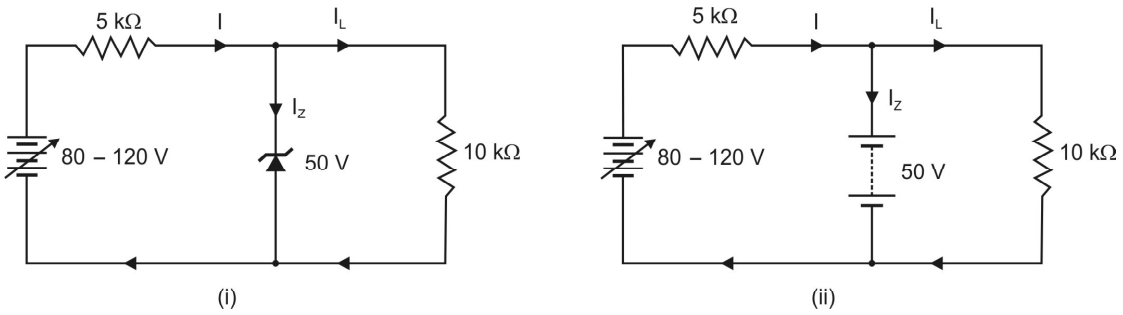


Fig. 4.28

Maximum zener current. The zener will conduct maximum current when the input voltage is maximum i.e. 120V. Under such conditions :

Output voltage 5 kΩ = 120 – 50 = 70 V

Voltage through 5 kΩ, $I = \frac{70V}{5\ k\Omega} = 14\ mA$

Load current, $I_L = \frac{50V}{10\ k\Omega} = 5\ mA$

Applying Krichhoff's first law, $I = I_L + I_Z$

∴ Zener current, $I_Z = I - I_L = 14 - 5 =$ **9 mA**

Minimum Zener current. The zener will conduct minimum current when the input voltage is minimum i.e. 80 V. Under such conditions, we have,

Voltage across 5 kΩ = 80 – 50 = 30 V

Current through 5 Ωk, $I = \frac{30V}{5\ k\Omega} = 6\ mA$

Load current, I_L = 5 mA

∴ Zener current, $I_Z = I - I_L = 6 - 5 =$ **1mA**

Example 3. A 7.2 V zener is used in the circuit shown in Fig. 4.29 and the load current is to vary from 12 to 100 mA. Find the value of series resistance R to maintain a voltage of 7.2 V across the load. The input voltage is constant at 12V and the minimum zener current is 10 mA.

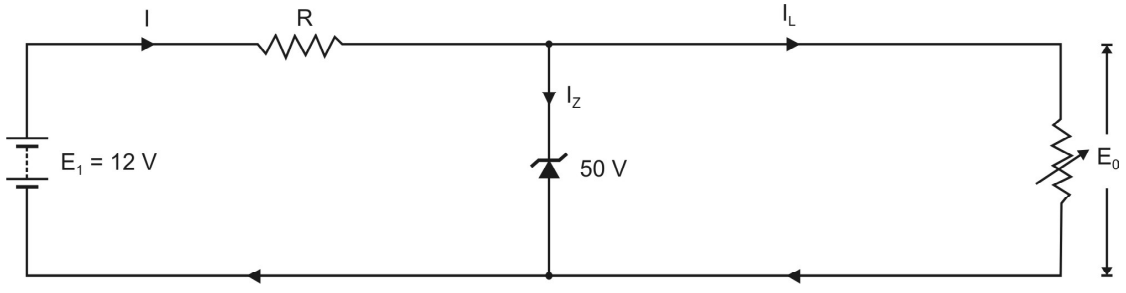


Fig. 4.29

Solution. $E_i = 12\ V;\quad V_Z = 7.2\ V$

$$R = \frac{E_i - E_0}{I_Z + I_L}$$

The voltage across R is to remain constant at 12 – 7.2 = 4.8 V as the load current chanes from 12 to 100 mA. The minimum.

$$\therefore \qquad R = \frac{E_i - E_0}{(I_Z)_{\min} + (I_L)_{\max}} = \frac{12\,V - 7.2\,V}{(10 + 100)\,mA} = \frac{4.8}{110\ mA} = 43.5\Omega$$

If R = 43.5Ω is inserted in the circuit, the output voltage will remain constant over the regulating range. As the load current IL decreases, the zener current I_Z will increase to such a value that $I_Z + I_L = 110$ mA. Note that if load resistance is open-circuited, then $I_L = 0$ and zener current becomes 110 mA.

Example 4. The zener diode shown in Fig. 4.30 has $V_Z = 18V$. The voltage across the load stays at 18 V as long as I_Z is maintained between 200 mA and 2 A. Find the value of series resistance R so that E_0 remains 18 V while input voltage E_i is free to vary between 22V to 28V.

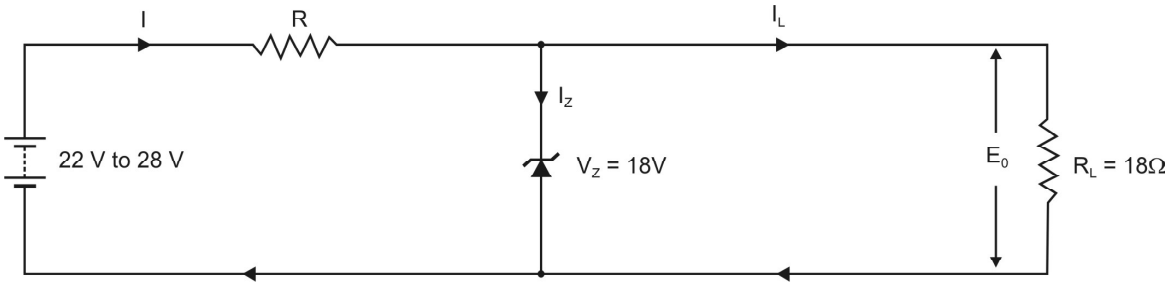


Fig. 4.30

Solution. The zener current will be minimum (i.e. 200 mA) when the input voltage is minimum (i.e. 22 V). The load current stays at constant value $I_L = V_Z/R_L = 18\ V/18\Omega = 1\ A = 1000$ mA.

$$\therefore \qquad R = \frac{E_i - E_0}{(I_Z)_{\min} + (I_L)_{\max}} = \frac{(22-18)V}{(200+1000)\,mA} = \frac{4V}{1200\,mA} = 3.33\Omega$$

Example 5. A 10-V zener diode is used to regulate the voltage across a variable load resistor. The input voltage varies between 13V and 16V and the load current varies between 10 mA and 85 mA. The minimum zener current is 15 mA. Calculate the value of series resistance R.

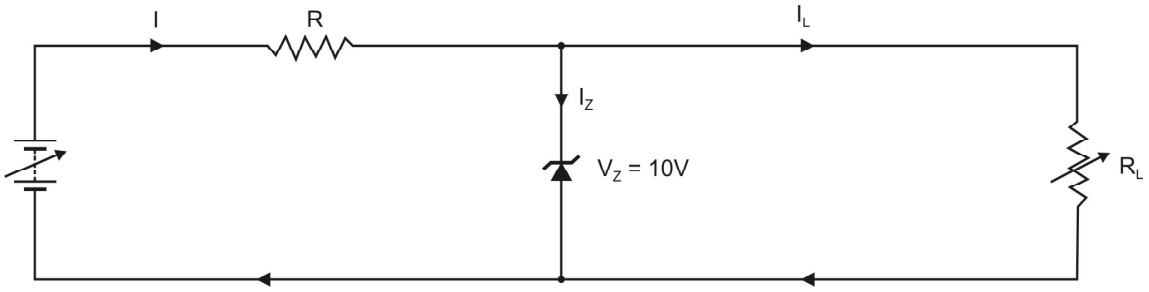


Fig. 1.1

Solution. The zener will conduct minimum current (i.e. 15 mA) when input voltage minimum (i.e. 13V).

$$\therefore \qquad R = \frac{E_i - E_0}{(I_Z)_{\min} + (I_L)_{\max}} = \frac{(13-10)V}{(15+85)\,mA} = \frac{3V}{100\,mA} = 30\Omega$$

Example 6. The circuit of Fig. 4.31 uses two zener diodes, each rated at 15 V, 200 mA. If the cirucit is conencted to a 45-volt unregulated supply, determine :

i. The regulated output voltage. ii. The value of series resistance R

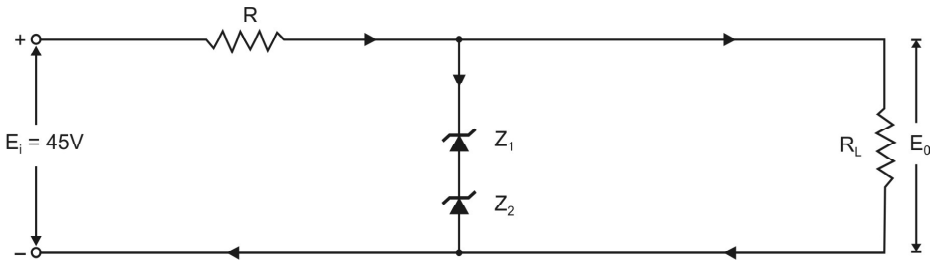


Fig. 4.31

Solution. When the desired regulated output voltage is higher than the rated voltage of the zener, two or more zeners are connected in series as shown in Fig. 4.31. However, in such circuits care must be taken to select those zeners that have the same current rating.

Current rating of each zener, I_Z = 200 mA

Voltage rating of each zener, V_Z = 15 V

Input voltage, E_i = 45 V

i. Regulated output voltage, E_0 = 15 + 15 = 30 V

ii. Series resistance, R = $R = \frac{E_i - E_0}{I_Z} = \frac{45-30}{200\,mA} = \frac{15V}{200\,mA} = 75\Omega$

Example 7. What value of series resistance is required when three 10-watt, 10-volt, 1000 mA zener diodes are connected in series to obtain a 30-volt regulated output from a 45 volt d.c. power source?

Solution. Fig. 4.32 shows the desired circuit. The worst case is at no load because then zeners carry the maximum current.

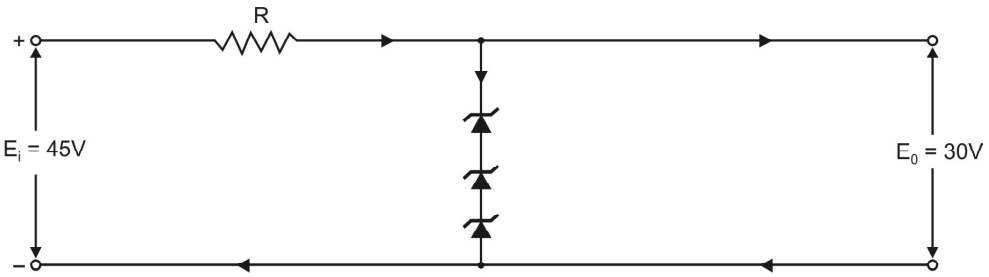


Fig. 4.32

Voltage rating of each zener, V_Z = 10 V

Current rating of each zener, I_Z = 1000 mA

Input unregulated voltage, E_i = 45 V

Regulated output voltage, E_0 = 10 + 10 + 10 = 30 V

Let R ohms be the required series resistance.

Voltage across R = $E_i - E_0$ = 45 – 30 = 15 V

$\therefore$ $R = \frac{E_i - E_0}{I_Z} = \frac{15V}{1000\,mA}\,15\Omega$

Example 8. Over what range of input voltage will the zener circuit shown in Fig. 4.33 maintain 30 V across 2000Ω load, assuming that series resistance R = 200Ω and zener current rating is 25 mA?

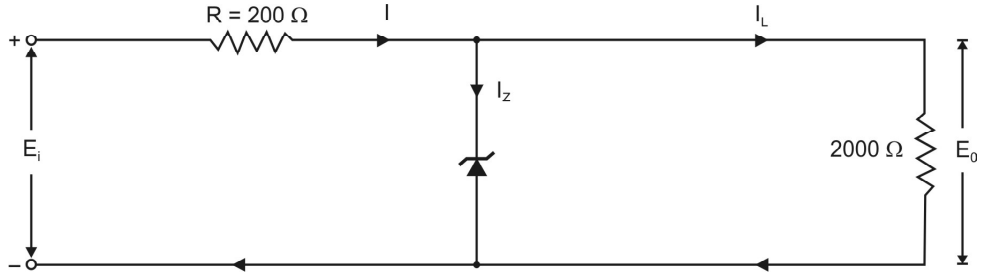


Fig. 4.33

Solution. The minimum input voltage required will be when $I_Z = 0$. Under this condition.

$$I_L = I = \frac{30V}{2000\,\Omega} = 15\,mA$$

$\therefore$ Minimum input voltage = 30 + I R = 30 + 15 mA x 200 Ω
= 30 + 3 = **33 V**

The maximum input voltage required will be when I_Z = 25 mA. Under this condition.

I = I_L + I_Z = 15 + 25 = 40 mA

$\therefore$ Max. input voltage = 30 + I R
= 30 + 40 mA x 200 Ω
= 30 + 8 = **38 V**

Therefore, the input voltage range over which the circuit will maintain 30 V across the load is 33 V to 38V.

Example 9. In the circuit shown in Fig. 4.34, the voltage across the load is to be maintained at 12 V as load current varies from 0 to 200 mA. Desing the regulator. Also find the maximum voltage rating of zener diode.

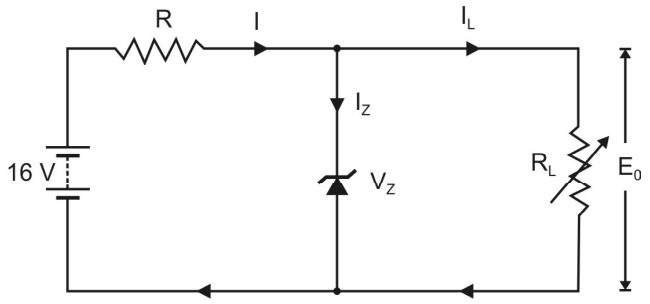


Fig. 4.34

Solution. By designing the regulator here means to find the values of VZ and R. Since the load voltage is to be maintained at 12 V, we will use a zener diode of zener voltage 12 V i.e.,

VZ = **12 V**

The voltage across R is to remain constant at 16 – 12 = 4 V as the load current changes from 0 to 200 mA. The minimum zener current will when the load current is maximum.

$\therefore$ $$R = \frac{E_i - E_0}{(I_Z)_{\min} + (I_L)_{\max}} = \frac{16-12}{(0+200)\,mA} = \frac{4V}{200\,mA} = 20\,\Omega$$

Maximum power rating of zener is

P_{ZM} = $V_Z I_{ZM}$ = (12 V) (200 mA) = 2.4 W

Example. 10. Fig. 4.35 shows the basic zener diode circuits. What will be the circuit behaviour if the zener is (i) working property (ii) shorted (iii) open-circuited?

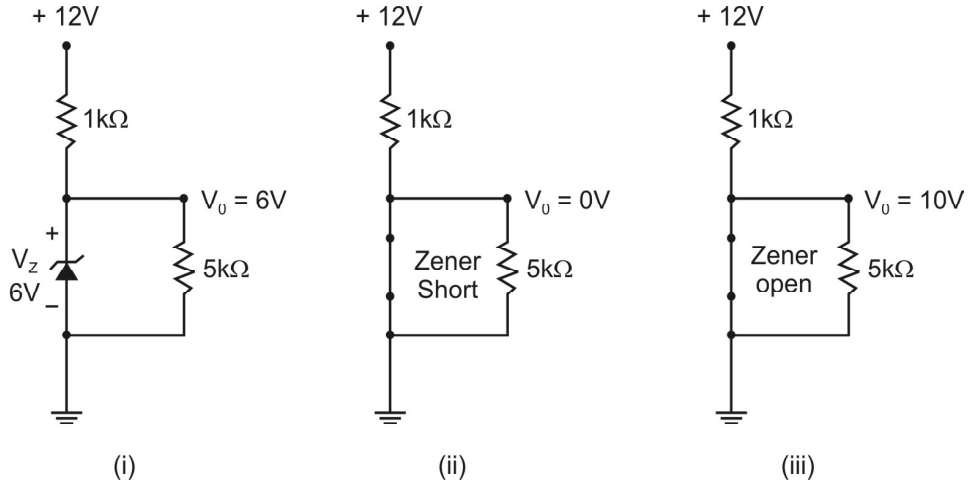


Fig. 4.35

Solution. Zener diodes cannot be tested individually with a multimeter. It is because multimeter usually do not have enough input voltage to put the zener into breakdown region.

i. If the zener diode is working properly, the voltage V_0 across the load (= 5kΩ) will be nearly 6V.

ii. If the zener diode is short [See Fig. 4.35 (ii)], you will measure V_0 as 0V. The same problem could also be caused by a shorted load resistor (=5kΩ) or an opened source resistor (=1 kΩ). The only way to tell which device has failed is to remove the resistors and chekc them with an ohmmeter. If the resistors are good, then zener diode is bad.

iii. If the zener diode is short [See Fig. 4.35 (ii)], your will measure V_0 as 0V. The same problem could also be caused by a shorted load resistor (= 5kΩ) or an opened source resistor (= 1 kΩ). The only way to tell which device has failed is to remove the resistors and check them with an ohmmeter. If the resistors are good, then zener diode is bad.

iii. If the zener diode is open-circuited, the voltage V_0 across the load (= 5kΩ) will be 10V.

Example. 11. Fig. 4.36 shows regulated power supply using a zener diode. What will be the circuit behaviour if (i) filter capacitor shorts (ii) filter capacitor opens?

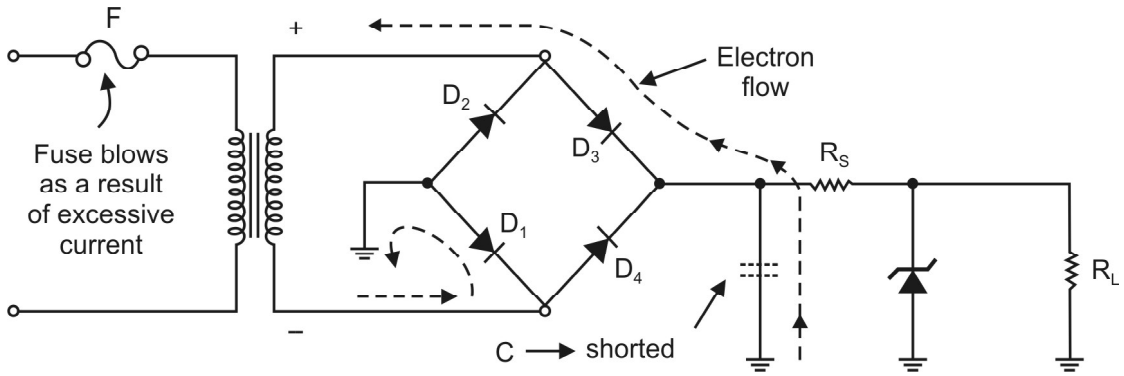


Fig. 4.36

Solutions. The common faults in a zener voltage regulator are shorted filter capacitor or opens filter capacitor.

i. **When filter capacitor shorts.** When the filter capacitor shorts, the primary fuse will blow. The reason for this is illustrated in Fig. 6.71. When the filter capacitor shorts, it shorts out the load resistance R_L. This has the same effect as wiring the two sides of the bridge together.

PHOTO VOLTAIC (SOLAR) CELL

The photo voltaic cell generates a voltage when it is subjected to optical radiation. The generated electric potential called **Photo Voltaic Potential** is proportional to incident light intensity. Since this device is used to generate electricity from sunlight, it is popularity known as solar cell. The cicuit symbol generally used for solar is shown in Fig. 4.37.

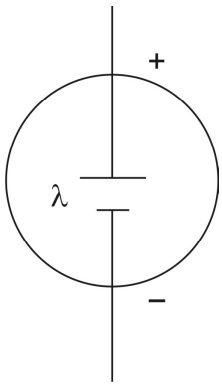


Fig. 4.37. Symbol of Solar Cell

Construction and Operating Principle of Solar Cell

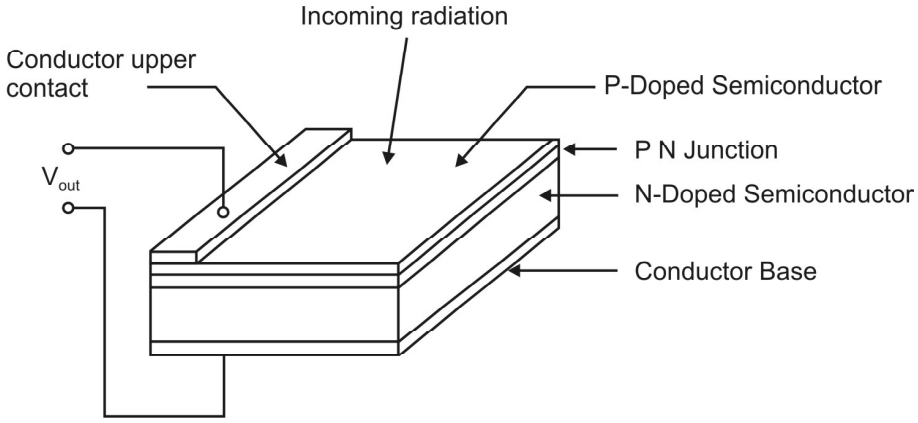


Fig. 4.38. Structure of Solar Cell

Fig. 4.38 shows structure of solar cell. It shows that cell is actually a P-N junction diode with appropriately doped semiconductors. The top P-type semiconductor layer is made very thin so as to permit solar radiation to fall onto junction. Doping level is very high, so the resulting junction will be extremely narrow.

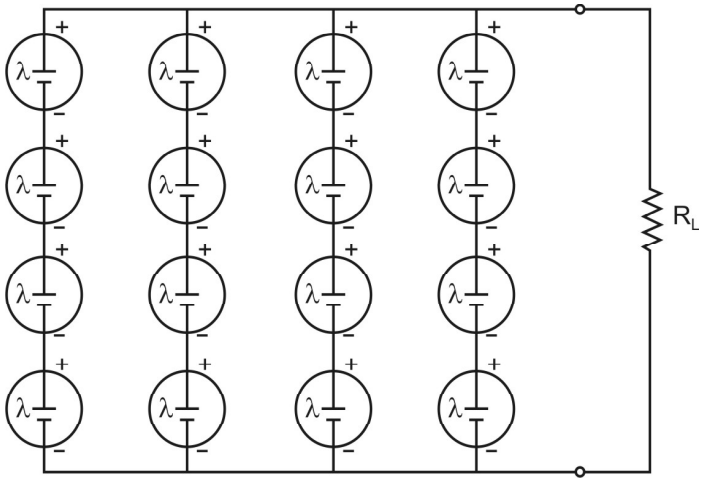


Fig. 1.1. A Solar Cell Array

To permit large amount of light input into the device, solar cells are designed to have large surface area. When photons strike on the thin P-doped upper layer; they are absorbed by the electrons in the N-layer; which caused formation of conduction electrons and holes. These conduction electrons and holes are separated by depletion across the cell, the depletion region potential causes the photo current to flow through the load.

Typical solar cells generates 0.4 V or less with currents ranging from the uA range to mA range, depending on the external load. Usually a large number of solar cells are arranged in an array in order to achieve higher voltage and currents as shown in Fig. 5.31.

In order to efficiently utilize the sun as a source of power, solar cell is an important device. It has become an important component for producing space craft power.

ADVANTAGES AND DISADVANTAGES

Advantages

i. Solar cell do not fuel.
ii. Once installed, they can work for years without any expenditure on maintenance.
iii. They do not pollute atmosphere.
iv. Solar cell do not produce noise.
v. They can produce enough voltage and current required for operation of equipments in rocket and satellites.

Disadvanteage

i. Slow response compared to photodiode.
ii. To get large output voltage, solar panel require large space.
iii. Solar panels are needed to be cleaned regularly to get better result.
iv. Must be supported by storage batteries to get continous operation in night.

HALF-WAVE RECTIFIER

In half-wave rectification, the rectifier conducts current only during the positive half-cycles of input a.c. supply. The negative half-cycles of a.c. supply are suppressed i.e. during negative half-cycles, no current is conducted and hence no voltage appears across the load. Therefore, current always flows in one direction (i.e. d.c.) through the load though after every half-cycle.

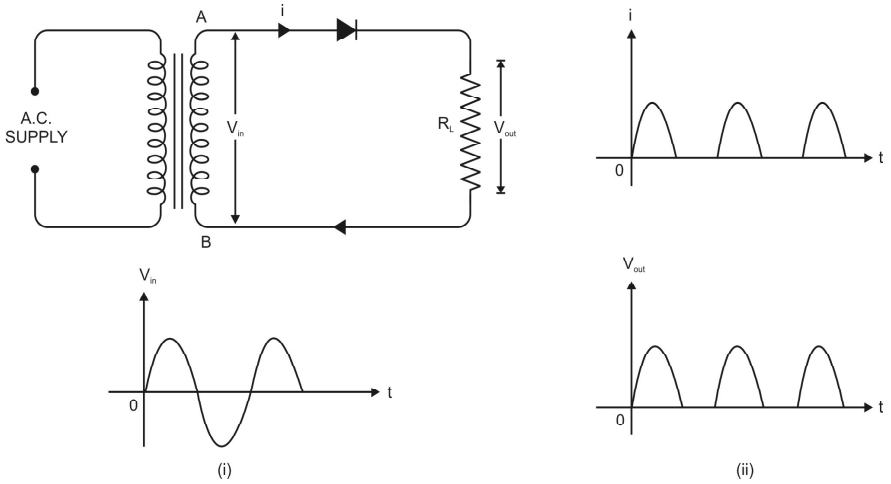


Fig. 2.23

Circuit Details. Fig. 2.23 shows the circuit where a single crystal diode acts as a half-wave rectifier. The a.c. supply to be rectified is applied in series with the diode and load resistance R_L. Generally, a.c. supply is given through a transformer. The use of transformer permits two advantages. Firstly, it allows us to step up or step down the a.c. input voltage as the situation demands. Secondly, the transformer isolates the rectifier circuit from power line and thus reduces the risk of electric shock.

Operation. The a.c. voltage across the secondary windings AB changes polarities after every half-cycle. During the positive half-cycle of input a.c. voltage, end A becomes positive w.r.t. end B. This makes the diode forward biased and hence it conducts current. During the negative half-cycle, end A is negative w.r.t. end B. Under this condition, the diode is reverse biased and it conducts no current. Therefore, current flows through the diode during positive half-cycles of input a.c. voltage only; it is blocked during the negative half-cycles [See Fig. 2.23 (ii)]. In this way, current flows through load R_L always in the same direction. Hende d.c. output is obtained across R_L. It may be noted that output across the load is pulsating d.c. These pulsations in the output are further smoothened with the help of filter circuits discussed later.

Disadvantages : The main disadvantages of a half-wave rectifier are :

i. 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.
ii. The a.c. supply delivers power only half the time. Therefore, the output is low.

OUTPUT FREQUENCY OF HALF-WAVE RECTIFIER

The output frequency of a half-wave rectifier is equal to the input frequency (50 Hz). Recall how a compolete cycle is defined. a waveform has a complete cycle is defined. A waveform has complete cycle when it repeats the same wave pattern over a given time. Thus in Fig. 2.24, the a.c. input voltage repeats the same wave pattern over 0° – 360°, 360° – 720° and so on. This means that when input a.c. completes one cycle, the output half-wave rectified wave also completes one cycle. In other words, the output frequency is equal to the input frequency i.e.

$$f_{out} = f_{in}$$

For example, if the input frequency of sine wave applied to a half-wave rectifier is 100 Hz, then frequency of the output wave will also be 100 Hz.

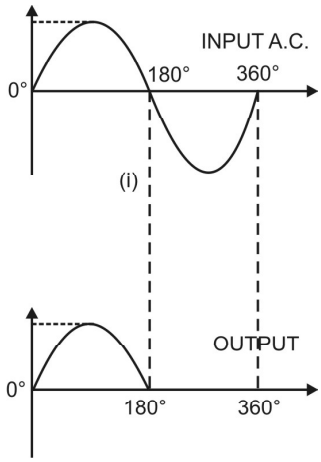


Fig. 2.24

EFFICIENCY OF HALF-WAVE RECTIFIER

The ratio of d.c. power output to the applied input a.c. power is known as rectifier efficiency i.e.

$$\text{Rectifier efficinecy}, \eta = \frac{d.c.\ power\ output}{Input\ a.c.\ power}$$

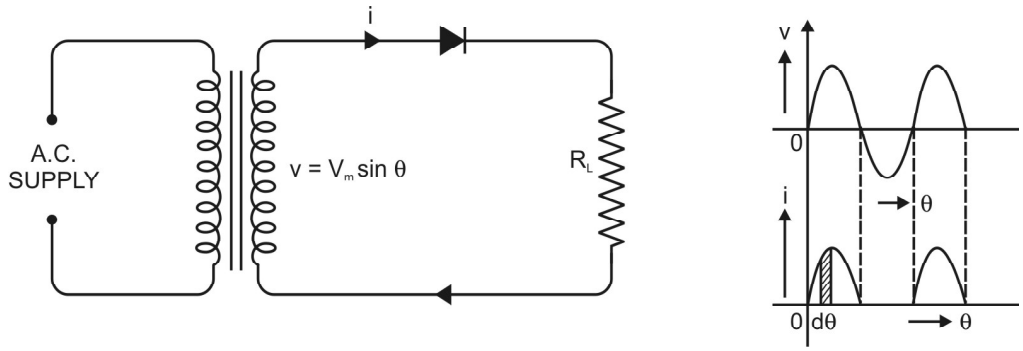


Fig. 1.1

Consider a half-wave rectifier shown in Fig. 6.22. Let $v = V_m \sin \theta$ the alternating voltage that appears across the secondary winding. Let r_f and R_L be the diode resistance and load resistance and load resistance respectively. The diode conducts during positive half-cycle of a.c. supply while no current conduction takes place during negative half-cycles.

D.C. Power. The output current is pulsating direct current. Therefore, in order to find d.c. power, average current has to be found out.

$$I_{av} = I_{dc} = \frac{1}{2\pi}\int_0^{\pi} i\,d\theta = \frac{1}{2\pi}\int_0^{\pi} \frac{V_m \sin\theta}{r_f + R_L}\,d\theta$$

$$= \frac{V_m}{2\pi(r_f + R_L)}\int_0^{\pi} \sin\theta\,d\theta = \frac{V_m}{2\pi\pi(+R_L)}[-\cos\theta]_0^{\pi}$$

$$= \frac{V_m}{2\pi(r_f + R_L)} x\,2 = \frac{V_m}{(r_f + R_L)} x \frac{1}{\pi}$$

$$= \frac{I_m}{\pi} \qquad \left[\because I_m = \frac{V_m}{(r_f + R_L)}\right]$$

$$\therefore \quad \text{d.c. power}, P_{dc} = I_{dc}^2\, x R_L = \left(\frac{I_m}{\pi}\right) x\, R_L \qquad ...(i)$$

A.C. Power Input : The a.c. power input is given by :

$$P_{ac} = I_{rms}^2 (r_f + R_L)$$

For a half-wave rectified wave, $I_{rms} = I_m/2$

$$\therefore \quad P_{ac} = \left(\frac{I_m}{2}\right)^2 (r_f + R_L) \qquad ...(ii)$$

$$\therefore \quad \text{Rectifier efficiency} = \frac{\text{d.c. output power}}{\text{a.c. input power}} = \frac{(I_m/\pi)^2 \, x R_L}{(I_m/2)^2 (r_f + R_L)}$$

$$= \frac{0.406 R_L}{r_f + R_L} = \frac{0.406}{1 + \frac{r_f}{R_L}}$$

The efficiency will be maximum if r_f is negligible as compared to R_L.

$\therefore$ Max. rectifier efficiency = 40.6%

This shows that in half-wave rectification, a maximum of 40.6% of a.c. power is converted into d.c. power.

Example 12. The applied input a.c. power to half-wave rectifier is 100 watts. The d.c. output power obtained is 40 watts.

i. What is the rectification efficiency?

ii. What happens to remaining 60 watts?

Solution.

i. $\text{Rectificat ion efficiency} = \frac{\text{d.c. output power}}{\text{a.c. input power}} = \frac{40}{100} = 0.4 = \mathbf{40\%}$

ii. 40% efficiency of rectification does not mean that 60% of power is lost in the rectifier circuit. In fact, a crystal diode consumes little power due to small internal resistance. The 100 W a.c. power is contained as 50 watts in positive half-cycles and 50 watts in negative half-cycles. The 50 watts in the negative half-cycles not supplied at all. Only 50 watts in the positive half-cycles are converted into 40 watts.

$$\therefore \quad \text{Powerefficiency} = \frac{40}{50} \text{ x } 100 = 80\%$$

Although 100 watts of a.c. power was of a.c. power was supplied, the half-wave rectifier accepted only 50 watts and converted it into 40 watts d.c. power. Therefore, it is appropriate to say that efficiency of rectification is 40% and not 80% which is power efficiency.

Example 13. An a.c. supply of 230 V is applied to a half-wave rectifier circuit through a transformer of turn ratio 10: 1. Find (i) the output d.c. voltage and (ii) the peak inverse voltage Assume the diode to be ideal.

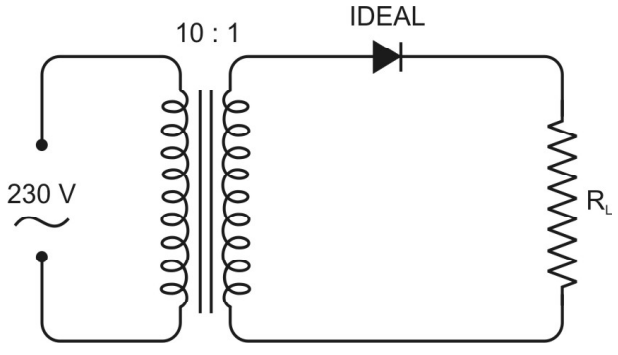


Fig. 1.1

Solution.

Primary to secondary turns is

$$\frac{N_1}{N_2} = 10$$

R.M.S. primary voltage

= 230 V

$\therefore$ Max. primary voltage is

$$V_{pm} = \left(\sqrt{2}\right) \text{x r.m.s. primary voltage}$$
$$= \left(\sqrt{2}\right) \text{x } 230 = 325.3\,\text{V}$$

Max. secondary voltage is

$$V_{sm} = V_{pm} \text{x} \frac{N_2}{N_1} = 325.3 \text{ x } \frac{1}{10} = 32.53\,\text{V}$$

i. $$I_{dc} = \frac{I_m}{\pi}$$

$\therefore$ $$V_{dc} = \frac{I_m}{\pi} \text{x} R_L = \frac{V_{sm}}{\pi} = \frac{32.53}{\pi} = \mathbf{10.36\,V}$$

ii. During the negative half-cycle of a.c. supply, the diode is reverse biased and hence conducts no current. Therefore, the maximum secondary voltage appears across the diode.

$\therefore$ Peak inverse voltage = **32.53 V**

Example 14. A crystal diode having internal resistance $r_f = 20\Omega$ is used for half-wave rectification. If the applied voltage v = 50 sin ωt and load resistance $R_L = 800\Omega$, find:

i. Im, Idc, Irms
ii. a.c. power input and d.c. power output
iii. d.c. output voltage
iv. efficiency of rectification

Solution.

v = 50 sin w t

Maximum voltage, V_m = 50 V

i. $$I_m = \frac{V_m}{r_f + R_L} = \frac{50}{20 + 800} = 0.061\text{A} = \mathbf{61mA}$$

I_{dc} = I_m/π = 61/π = **19.4 mA**

I_{rms} = $I_m/2$ = 61/2 = **30.5 mA**

ii. $$\text{a.c. power input} = (I_{rms})^2 \text{x} (r_f + R_L) = \left(\frac{30.5}{1000}\right)^2 \text{x } (20 + 800) = \mathbf{0.763\,watt}$$

$$\text{d.c. power output} = I^2_{dc} \text{x} R_L = \left(\frac{19.4}{1000}\right)^2 \text{x} 800 = \mathbf{0.31\,watt}$$

iii. $$\text{d.c. output voltage} = I_{dc} R_L = 19.4\,\text{mA x } 800\,\Omega = \mathbf{15.52\,volt}$$

iv. $$\text{Efficiency of rectification} = \frac{0.301}{0.763} \text{ x } 100 = \mathbf{39.5\%}$$

Example 15. A half-wave rectifier is used to supply 50V d.c. to a resistive load of 800Ω. The diode has a resistance of 25Ω. Calculate a.c. voltage required.

Solution.

Output d.c. voltage, V_{dc} = 50 V

Diode resistance, r_f = 25 Ω

Load resistance, R_L = 800 Ω

Let V_m be the maximum vlaue of a.c. voltage required.

$\therefore$ $$V_{dc} = I_{dc} \text{ x } R_L$$
$$= \frac{I_m}{\pi} \text{x} R_L = \frac{V_m}{\pi (r_f + R_L)} \text{ x } R_L \qquad \left[\because I_m = \frac{V_m}{r_f + R_L}\right]$$

or $$50 = \frac{V_m}{\pi (25 + 800)} \text{x} 800$$

$\therefore$ $$V_m = \frac{\pi \text{x} 825 \text{x} 50}{800} = 162\,\text{V}$$

Hence a.c. voltage of maximum value 162 is required.

FULL-WAVE RECTIFIER

In full-wave rectification, current flows through the load in the same direction for both half-cycles of input a.c. voltage. 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 diode does so; current being always in the same direction through the load. Therefore, a full-wave rectifier utilises both half-cycles of input a.c. voltage to produce the d.c. output. The following two circuits are commonly used for full-wave rectification:

i. Centre-tap full-wave rectifier ii. Full-wave bridge rectifier

CENTRE-TAP FULL-WAVE RECTIFIER

The circuit employes two diodes D_1 and D_2 as shown in Fig. 2.25. A centre tapped secondary winding AB is used with two diodes connected so that each uses one half-cycle of input a.c. voltage. In other words, diode D_1 utilises the a.c. voltage appearing across the upper half (OA) of secondary winding for rectification while diode D_2 uses the lower half winding OB.

Operation. During the positive half-cycle of secondary voltage, the end A of the secondary winding becomes positive and end B negaitve. This makes the diode D_1 forward biased and diode D_2 reverse biased. Therefore, diode D_1 conducts while diode D_2 does not. The conventinal current flow is through diode D_1, load resistor R_L and the upper half of secondary winding as shown by the dotted arrow. During the negative half-cycle, end A of the secondary winding becomes negative and end B positive. Therefore, diode D_2 conducts while diode D_1 does not. The conventional current flow through diode D_2, load R_L and lower half winding as shown by solid arrows. Referring to Fig. 2.25, it may be seen that current in the load R_L is in the same direction for both half-cycles of input a.c. voltage. Therefore, d.c. is obtained across the load R_L. Also the polarities of the d.c. output across the load should be noted.

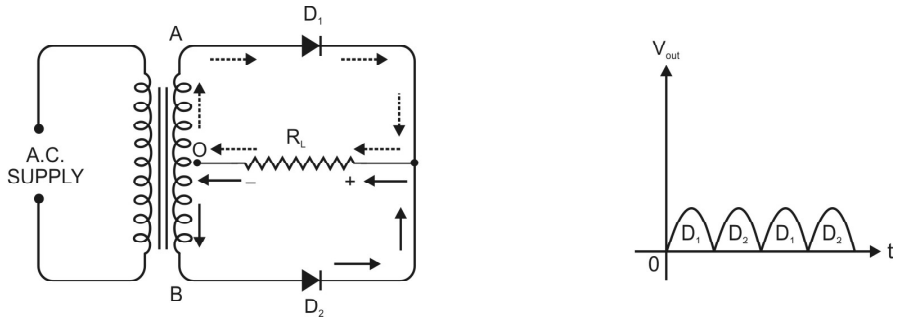


Fig. 2.25

Peak inverse voltage. Suppose V_m is the maximum voltage across the half secondary winding. Fig. 2.26 shows the circuit at the instant secondary voltage reaches it maximum value in the positive direction. At his instant, diode D_1 is conducting while diode D_2 is non-conducting. Therefore, whole of the secondary voltage appears across the non-conducting diode. Consequently, the peak inverse voltage is twice the maximum voltage across the half-secondary winding i.e.

$$PIV = 2\,V_m$$

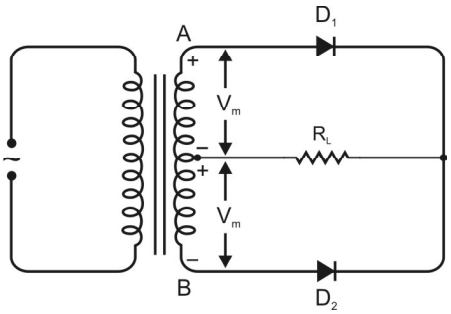


Fig. 2.26

Disadvantages

i. It is difficult to locate the centre tap on the secondary winding.
ii. The d.c. output is small as each diode utilises only one-half of the transformer secondary voltage.
iii. The diodes used must have high peak inverse voltage.

FULL-WAVE BRIDGE RECTIFIER

The need for a centre tapped power transformer is eliminated in the bridge rectifier. It contains four diodes D_1, D_2, D_3 and D_4 connected to form bridge as shown in Fig. 2.23. The a.c. supply to be rectified is applied 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.

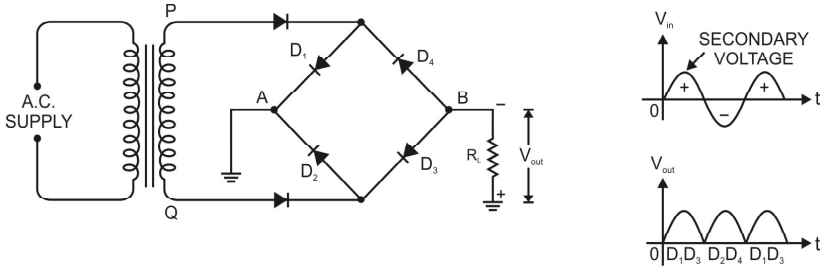


Fig. 2.23

Operation. During the positive half-cycle of secondary voltage, the end P of the secondary winding becomes positive and end Q negative. This makes diodes D_1 and D_3 forward biased while diodes D_2 and D_4 are reverse biased. Tehrefore, only diodes D_1 and D_3 conduct. These two diodes will be in series through the load R_L as shown in Fig. 2.24 (i). The conventional current flow is shown by dotted arrows. It may be seen that current flows from A to B through the load R_L.

During the negative half-cycle of secondary voltage, end P becomes negative and end Q positive. This makes diodes D_2 and D_4 forward biased whereas dioces D_1 and D_3 are reverse biased. Therefore, only diodes D_2 and D_4 conduct. These two diodes will be in series through the load R_L as shown in Fig. 2.24 (ii). The current flow is shown by the solid aorrws. It may be seen that again current flows from A to B through the load i.e. in the sme direction as for the positive half-cycle. Therefore, d.c. output is obtained across load R_L.

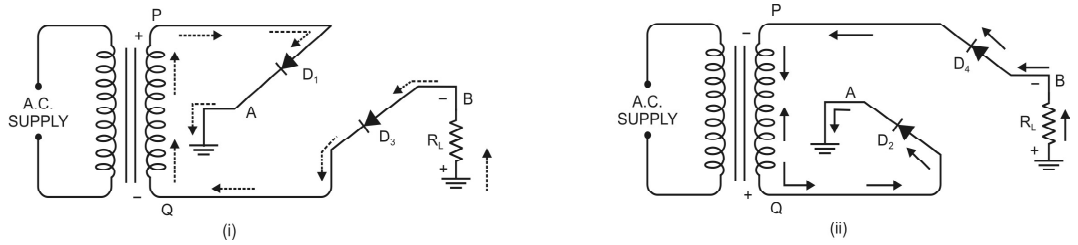


Fig. 2.24

Peak Inverse Voltage. The peak inverse voltage (PIV) of each diode is equal to the maximum secondary voltage of transformer. Suppose during positive half cycle of input a.c., end *P* of secondary is positive and end *Q* negative. Under such conditions, diodes D_1 and D_3 are forward biased while diodes D_2 and D_4 are reverse biased. Since the diodes are considered ideal, diodes D_1 and D_3 can be replaced by wires as shown in Fig. 2.25 (i). This circuit is the same as shown in Fig. 2.25 (ii).

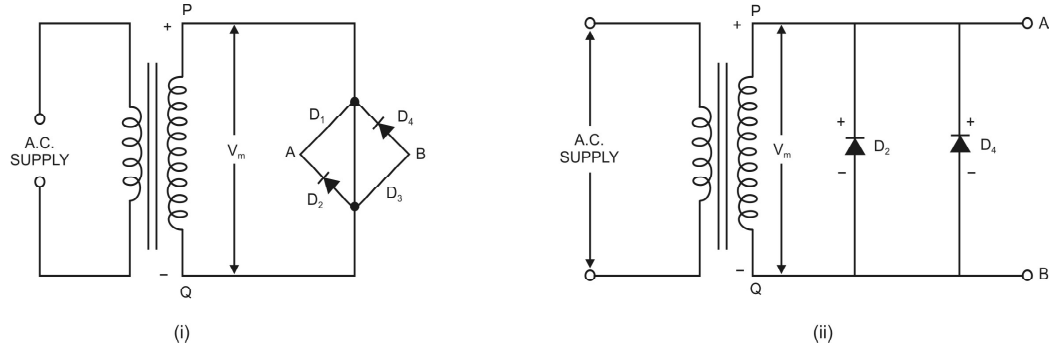


Fig. 2.25

Referring to Fig. 2.25 (ii), it is clear that two reverse biased diodes (i.e. D_2 and D_4) and the secondary of transformer are in parallel. Hence PIV of each diode (D_2 and D_4) is equal to the maximum voltage (V_m) across the secondary. Similarly, during the next half cycle, D_2 and D_4 are forward biased while D_1 and D_3 will be reverse biased. It is easy to see that reverse voltage across D_1 and D_3 is equal to V_m.

Advantages

i. The need for centre-tapped transformer is eliminated.
ii. The output is twice that of the centre-tap circuit for the same secondary voltage.
iii. The PIV is one-half that of the centre-tap circuit (for same d.c. output).

Disadvantages

i. It requires four diodes.
ii. As 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 th centre tap circuit. This is objectionable when secondary voltage is small.

OUTPUT FREQUENCY OF FULL-WAVE RECTIFIER

The output frequency of a full-wave rectifier is double the input frequency. Remember that a wave has a complete cycle when it repeats the same pattern. In Fig. 2.26 (i), the input a.c. competes one cycle from 0° – 360°. However, the full-wave rectified wave completes 2 cycles in this period [See Fig. 2.26 (ii)]. Therefore, output frequency is twice the input frequency i.e.

$$f_{out} = 2f_{in}$$

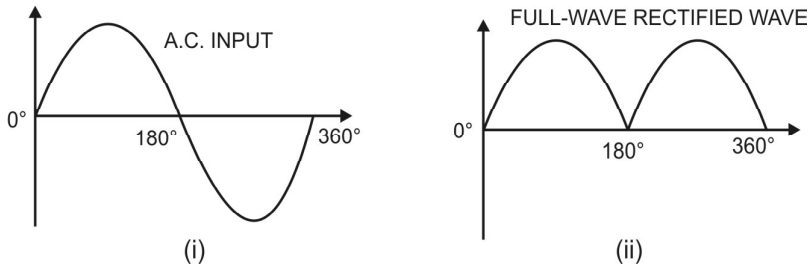


Fig. 2.26

EFFICIENCY OF FULL-WAVE RECTIFIER

Fig. 6.30 shows the process of full-wave rectification. Let v = Vm sin 0 be the a.c. voltage to be rectified. Let r_f and R_L be the diode resistance and load resistance respectively. Obviously, the rectifier will conduct current through the load in the same direction for both half-cycles of input a.c. voltage. The instantaneous current i is given by :

$$i = \frac{v}{r_f + R_L} = \frac{V_m \sin\theta}{r_f + R_L}$$

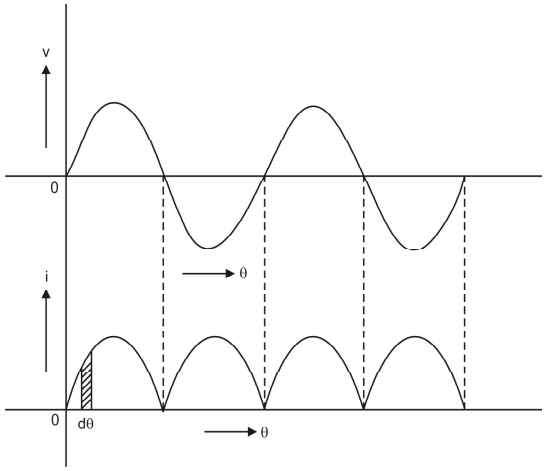


Fig. 1.1

D.C. output power. The output current is pulsating direct current. Therefore, in order to find the d.c. power, average current has to be found out. From the elementary knowledge of electrical engineering.

$$I_{dc} = \frac{2I_m}{\pi}$$

$\therefore$ d.c.powerout, $P_{dc} = I_{dc}^2 \, x \, R_L = \left(\frac{2 I_m}{\pi} \right)^2 x \, R_L$

A.C. input power. The a.c. input power is given by :

$$P_{ac} = I_{rms}^2 (r_f + R_L)$$

For a full-wave rectified wave, we have,

$$I_{rms} = I_m / \sqrt{2}$$

$\therefore$ $$P_{ac} = \left(\frac{I_m}{\sqrt{2}} \right)^2 (r_f + R_L)$$

$\therefore$ Full-wave rectification efficiency is

$$\eta = \frac{P_{dc}}{P_{ac}} = \frac{(2 I_m / \pi)^2 R_L}{\left(\frac{I_m}{\sqrt{2}} \right)^2 (r_f + R_L)}$$

$$= \frac{8}{\pi^2} x \frac{R_L}{(r_f + R_L)} = \frac{0.812 \, R_L}{r_f + R_L} = \frac{0.812}{1 + \frac{r_f}{R_L}}$$

The efficiency will be maximum if r_f is negligible as compared to R_L.

$\therefore$ Maximum efficiency = 81.2%

This is double the efficiency due to half-wave rectifier. Therefore, a full-rectifier is twice as effective as a half-wave rectifier.

Example 16. A full-wave rectifier uses two diodes, the internal resistance of each diode may be assumed constant at 20Ω. The transformer r.m.s. secondary voltage from centre tap to each end of secondary is 50V and load resistance is 980Ω. Find:

i. the mean and current ii. the r.m.s. value of load current

Solution

$$r_f = 20\Omega, \; R_L = 980\,\Omega$$

Max. a.c. voltage, $V_m = 50 \, x \sqrt{2} = 70.7\,\text{V}$

Max. load current, $I_m = \frac{V_m}{r_f + R_L} = \frac{70.7\,\text{V}}{(20+980)\,\Omega} = 70.7\,\text{mA}$

i. Mean load current, $I_{dc} = \frac{2 I_m}{\pi} = \frac{2 \, x \, 70.7}{\pi} = 45\,\text{mA}$

ii. R.M.S. value of load current is

$$I_{rms} = \frac{I_m}{\sqrt{2}} = \frac{70.7}{\sqrt{2}} = 50\,\text{mA}$$

Example 17. In the centre-tap circuit shown in Fig. 6.13, the diodes are assumed to be ideal i.e. having zero internal resistance. Find :

i. d.c. output voltage ii. peak inverse voltage iii. rectification efficiency.

Solution.

Primary to secondary turns, $N_1 / N_2 = 5$

R.M.S. primary voltage = 230 V

$\therefore$ R.M.S. secondary voltage

= 230 x (1/5) = 46 V

Maximum voltage across secondary

$= 46 \text{ x } \sqrt{2} = 65\text{V}$

Maximum voltage across half secondary winding is

V_m = 65/2 = 32.5 V

i. Average current, I_{dc} =

$$\frac{2V_m}{\pi R_L} = \frac{2 \times 32.5}{\pi \times 100} = 0.207 \text{ A}$$

∴ d.c. output voltage, $V_{dc} = I_{dc} = R_L$ = 0.207 x 100 = 20.7 V

ii. The peak inverse voltage is equal to the maximum secondary voltage, i.e.

PIV = 65 V

iii. Rectification efficiency = $\dfrac{0.812}{1+\dfrac{r_f}{R_L}}$

Since r_f = 0

∴ Rectification efficiency = 81.2%

Example 18. In the bridge type circuit shown in Fig. 6.32, the diodes are assumed to be ideal. Find: i. d.c. output voltage ii. peak inverse voltage iii. output frequency. Assume primary to secondary turns to be 4.

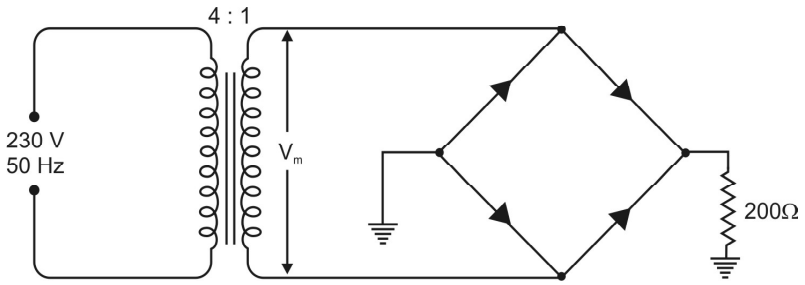


Fig. 1.1

Solution.

Primary/secondary turns, N1/N2 = 4

R.M.S. primary voltage = 230V

∴ R.M.S. secondary voltage = 230 (N2/N1) = 230 x (1/4) = 57.5 V

Maximum voltage across secondary is

$$V_m = 57.5 \, x \sqrt{2} = 81.3 V$$

i. Average current, $I_{dc} = \dfrac{2V_m}{\pi R_L} = \dfrac{2 \, x \, 81.3}{\pi \, x \, 200} = 0.26 \, A$

∴ d.c. output voltage, $V_{dc} = I_{dc} \, x \, R_L = 0.26 \, x \, 200 = 52$ V

ii. The peak inverse voltage is equal to the maximum secondary voltage i.e.

PIV = **81.3 V**

iii. In full-wave rectification, there are two output pulses for each complete cycle of the input a.c. voltage. Therefore, the output frequency is twice that of the a.c. supply frequency i.e.

fout = 2 x fin = 2 x 50 = **100 Hz**

Example 19. Fig. 6.33 (i) and Fig. 6.33 (ii) show the centre-tap and bridge type circuits having the same load resistance and transformer turn ratio. The primary of each is connected to 230V, 50Hz supply.

i. Find the d.c. voltage in each case.

ii. PIV for each case for the same d.c. output. Assume the diodes to be ideal.

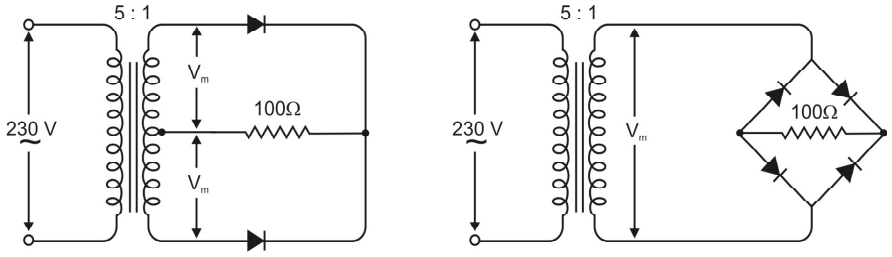


Fig. 1.1

Solution.

i. **D.C. output voltage**

Centre-tap circuit

R.M.S. secondary voltage = 230 x 1/5 = 46 V

Max. voltage across secondary $= 46 \text{ x } \sqrt{2} = 65 \text{ V}$

Max. voltage appearing across half secondary winding is

V_m = 65/2 = 32.5 V

Average current, Idc $= \dfrac{2V_m}{\pi R_L}$

D.C. output voltage, $V_{dc} = I_{dc}\, x\, R_L = \dfrac{2V_m}{\pi R_L} x\, R_L$

$$= \frac{2V_m}{\pi} = \frac{2\,x\,32.5}{\pi} = \mathbf{20.7\ V}$$

Bridge Circuit

Max. voltage across secondary, V_m = 65 V

D.C. output voltage, $V_{dc} = I_{dc} R_L = \dfrac{2V_m}{\pi R_L} x\, R_L = \dfrac{2V_m}{\pi} = \dfrac{2\,x\,65}{\pi} = \mathbf{41.4\ V}$

This shows that for the same secodary, the d.c. output voltage of bridge circuit is twice that of the centre-tap circuit.

ii. **PIV for same d.c. output voltage**

The d.c. output voltage of the two circuits will be the same if Vm (i.e. max. voltage utilised by each circuit for conversion into d.c.) is the same. For this to happer, the turn ratio of the transformers should be as shown in Fig. 6.34.

Centre-tap circuit

R.M.S. secondary voltage = 230 x 1/5 = 46 V

Max. voltage across secondary = $46\,x\,\sqrt{2} = 65V$

Max. voltage across half secondary winding is

V_m = 65/2 = 32.5 V

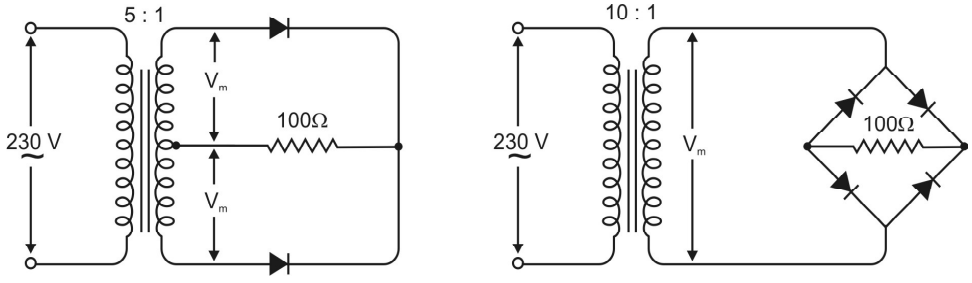


Fig. 1.1

$\therefore$ PIV = $2V_m$ = 2 x 32.5 = **65 V**

Bridge type circuit

R.M.S. secondary voltage = 230 x 1/10 = 23 V

Max. voltage across secondary, $V_m = 23 x \sqrt{2} = 32.5 V$

$\therefore$ PIV = V_m = **32.5 V**

This shows that for the same d.c. output voltage, PIV of bridge circuit is half that of centre-tap circuit. This is a distinct advantage of bridge circuit.

Example 20. The four diodes used in a bridge rectifier have forward resistances which may be considered constant at 1Ω and infinite reverse resistance. The alternating supply voltage is 240 V r.m.s. and load resistance is 480Ω. Calculate (i) mean load current and (ii) power dissipated in each diode.

Solution.

Max. a.c. voltage, $V_m = 240 x \sqrt{2} V$

i. At any instant in the bridge rectifier, two diodes in series are conducting. Therefore, total circuit resistance = 2 r_f + R_L.

Max. load current, $I_m = \frac{V_m}{2 r_f + R_L} = \frac{240 x \sqrt{2}}{2 x 1 + 480} = 0.7 A$

$\therefore$ Mean load current, $I_{dc} = \frac{2 I_m}{\pi} = \frac{2 x 0.7}{\pi} = 0.45 \text{ A}$

ii. Since each diode conducts only half a cycle, diode r.m.s. current is :

$I_{r.m.s.}$ = I_m/2 = 0.7/2 = 0.35 A

Power dissipated in each diode = $I^2_{r.m.s.}$ x r_f = $(0.35)^2$ x 1 = **0.123 W**

Example 21. The bridge rectifier shown in Fig. 6.35 uses silicon diodes, Find (i) output voltage (ii) d.c. output current. Use simplified model for the diodes.

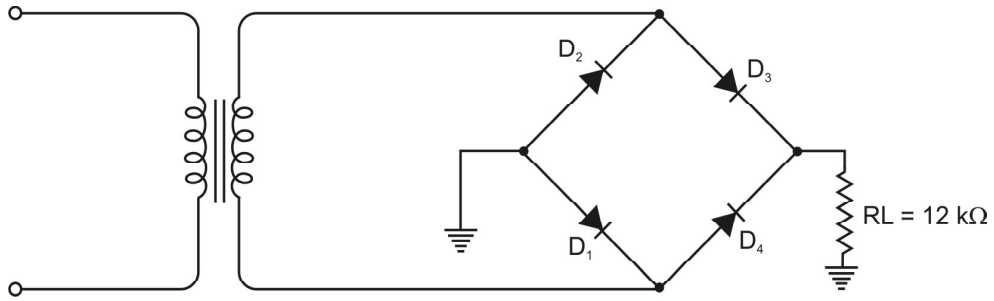


Fig. 2.30

Solution. The conditions of the problem suggest that the a.c. voltage across transformer secondary is 12V r.m.s.

$\therefore$ Peak secondary voltage is

$V_{s(pk)} = 12 x \sqrt{2} = 16.97 \text{ V}$

i. At any instant in the bridge rectifier, two diodes in series are conducting.

$\therefore$ Peak output voltage is

$V_{out\ (pk)}$ = 16.97 – 2 (0.7) = 15.57 V

$\therefore$ Average (or d.c.) output voltage is

$V_{av} = V_{dc} = \frac{2 V_{out(pk)}}{\pi} = \frac{2 x 15.57}{\pi} = 9.91 \text{ V}$

ii. Average (or d.c.) output current is

$I_{av} = \frac{V_{av}}{R_L} = \frac{9.91 V}{12 k\Omega} = \mathbf{825.8\ \mu A}$

RIPPLE FACTOR

The output of a rectifier consists of a d.c. component and an a.c. component (also known as ripple). The a.c. component is undesirable and accounts for the pulsations in the rectifier output. The effectiveness of a rectifier depends upon the magnitude of a.c. component in the output; the smaller this component, the more effective isthe rectifier.

The ratio of r.m.s. value of a.c. component to the d.c. component in the rectifier output is known as **ripple factor** i.e.

$$\text{Ripple factor} = \frac{\text{r.m.s value of a.c. component}}{\text{value of d.c. component}} = \frac{I_{ac}}{I_{dc}}$$

Therefore, ripple factor is very important in deciding the effectiveness of a rectifier. The smaller the ripple factor, the lesser the effective a.c. component and hence more effective is the rectifier.

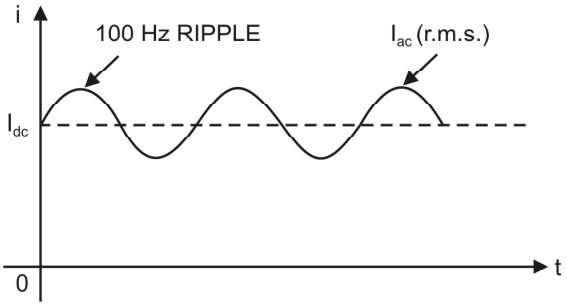


Fig. 1.1

Mathematical Analysis. The output current of a rectifier contains d.c. as well as a.c. component. The undesired a.c. component has a frequency of 100Hz (i.e. double the supply frequency 50 Hz) and is called the ripple (See Fig. 6.39). It is a fluctuation superimposed on the d.c. component.

By definition, the effective (i.e. r.m.s.) value of total load current is given by :

$$I_{rms} = \sqrt{I_{dc}^2 + I_{ac}^2}$$

or
$$I_{ac} = \sqrt{I_{rms}^2 - I_{dc}^2}$$

Dividing throughout by I_{dc}, we get,

$$\frac{I_{ac}}{I_{dc}} = \frac{1}{I_{dc}}\sqrt{I_{rms}^2 - I_{dc}^2}$$

But I_{ac}/I_{dc} is the ripple factor.

$$\therefore \quad \text{Ripple factor} = \frac{1}{I_{dc}}\sqrt{I_{rms}^2 - I_{dc}^2} = \sqrt{\left(\frac{I_{rms}}{I_{dc}}\right)^2 - 1}$$

i. **For half-wave rectification.** In half-wave rectification,

$$I_{rms} = I_m/2 \quad ; \quad I_{dc} = I_m/\pi$$

$$\therefore \quad \text{Ripple factor} = \sqrt{\left(\frac{I_m/2}{I_m/\pi}\right)^2 - 1} = 1.21$$

It is clear that a.c. component exceeds the d.c. component in the output of a half-wave rectifier. This result in greater pulsations in the output. Therefore, half-wave rectifier is ineffective for conversion of a.c. into d.c.

ii. **For full-wave rectification.** In full-wave rectification.

$$I_{rms} = \frac{I_m}{\sqrt{2}} \quad ; \quad I_{dc} = \frac{2I_m}{\pi}$$

$$\therefore \quad \text{Ripplefactor} = \sqrt{\left(\frac{I_m/\sqrt{2}}{2I_m/\pi}\right)^2 - 1} = 0.48$$

i.e. $$\frac{\text{effective a.c. component}}{\text{d.c. component}} = 0.48$$

This shows that in the output of a full-wave rectifier, the d.c. component is more than the a.c. component. Consequently, the pulsations in the output will be less than in half-wave rectifier. For this reason, full-wave rectification is invariably used for conversion of a.c. into d.c.

Example 22. A power supply A delivers 10V dc with a ripple of 0.5 V r.m.s. while the power supply B delivers 25V dc with a ripple of 1 mV r.m.s. Which is better power supply?

Splution. The lower the ripple factor of a power supply, the better it is.
For power supply A

$$\text{Ripple factor} = \frac{V_{ac(r.m.s.)}}{V_{dc}} = \frac{0.5}{10} \text{ x } 100 = 5\%$$

For power supply B

$$\text{Ripple factor} = \frac{V_{ac(r.m.s.)}}{V_{dc}} = \frac{0.001}{25} \text{ x } 100 = 0.004\%$$

Clearly, power supply B is better.

COMPARISON OF RECTIFIERS

S.No.	Particulars	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}	$2 f_{in}$	$2 f_{in}$
6	Peak inverse voltage	V_m	$2 V_m$	V_m

A comparison among the three rectifier circuits must be made very judiciously. Although bridge circuit has some disadvantages, it is the best circuit from the viewpoint of overall performance. When cost of the transformer is the main consideration in a rectifier assembly, we invariably use the bridge circuit. This is particularly true for large rectifiers which have a low-voltage and a high-current rating.

FILTER CIRCUITS

Generally, a rectifier is required to produce pure d.c. supply for using at various places in the electronic circuits. However, the output of a rectifier has pulsating character i.e. it contains a.c. and d.c. components. The a.c. component is undesirable and must be kept from the load. To do so, a filter circuit is used which removes (or filters out) the a.c. component and allows only the d.c. component reach the load.

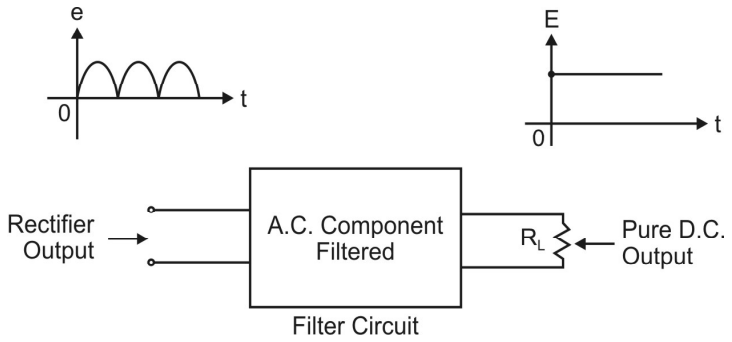


Fig. 1.1

A **filter circuit** is a device which removes the a.c. component of rectifier output but allows the d.c. component to reach the load.

Obviously, a filter circuit should be installed between the rectifier and the load as shown in Fig. 6.40. A filter circuit is generallya combination of inductors (L) and capacitors (C). The filtering action of L and C depends

upon the basic electrical principles. A capacitor passes a.c. readily but does not pass d.c. at all. On the other hand, an inductor opposes a.c. but allows d.c. to pass through it. It then becomes clear that suitable network of L and C can effectively remove the a.c. component, allowing the d.c. component to reach the load.

Types of Filter Circuits

The most commonly used filter circuits are capacitor filter, choke input filter and capacitor input filter on π-filter. We shall discuss these filters in turn.

i. **Capacitor filter.** Fig. 6.41 (ii) shows a typical capacitor filter circuit. It consists of a capacitor C placed across the rectifier output in parallel with load R_L. The pulsating direct voltage of the rectifier is applied across the capacitor. As the rectifier voltage increases, it charges the capacitor and also supplies current to the load, At the end of quarter cycle [Point A in Fig. 6.41 (iii)], the capacitor is charged to the peak value Vm of the rectifier voltage. Now, the rectifier voltage starts to decrease. As this occurs, the capacitor discharges through the load nad voltage across it (i.e. across parallel combination of R-C) decreases as shown by the line AB in Fig. 6.41 (iii). The voltage across load will decrease only slightly because immediately the next voltage peak comes and reacharges the capacitor. This process is repeated again and again and he output voltage waveform becomes ABCDEFG. It may be seen that very little ripple is left in the output. Moreover, output voltage is higher as it remains substantially near the peak value of rectifier output voltage.

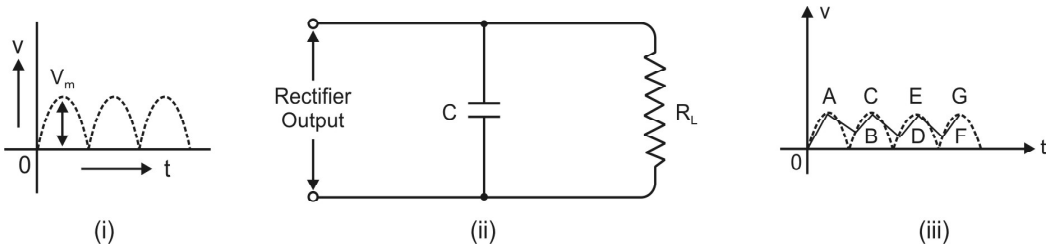


Fig. 1.1

The capacitor filter circuit is extremely popular because of its low cost, small size, little weight and good characteristics. For small load currents (say upto 50 mA), this type of filter is preferred. It is commonly used in transistor radio battery eliminators.

ii. **Choke input filter.** Fig. 6.42 shows a typical choke input filter circuit. It consists of a choke L connected in series with the rectifier output and a filter capacitor C across the load. Only a signle filter section is shown, but several identical sections are often used to reduce the pulsations as effectively as possible.

The pulsating output of the rectifier is applied across terminals 1 and 2 of the filter circuit. As discussed before, the pulsating output of rectifier contains a.c. and d.c. components. The choke offers high opposition to the passage of a.c. component but negligible opposition to the d.c. component. The result is that most of the a.c. component appears across the choke while whole of d.c. component passes through the choke on its way to load. This result in the reduced pulsations at terminal 3.

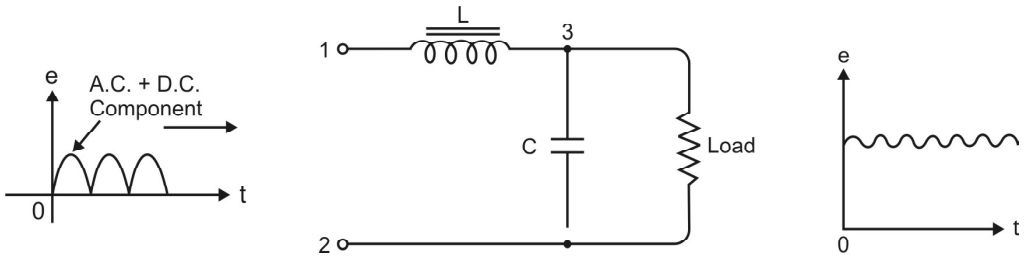


Fig. 1.1

A terminal 3, the rectifier output contains d.c. component and the remaining part of a.c. component which has managed to pass through the choke. Now, the low reactance of filter capacitor by passes the a.c. component but prevents the d.c. component to flow through it. Therefore, only d.c. component reaches the load. In this way, the filter circuit has filtered out the a.c. component from the rectifier output, allowing d.c. component to reach the load.

iii. **Capacitor input filter or π-filter.** Fig. 6.43 shows a typical capacitor input filter or π-filter. It consists of a filter capacitor C_1 connected across the rectifier output, a choke L in series and another filter capacitor C_2 connected across the load. Only one filter section is shown but several identical sections are often used to improve the smoothing action.

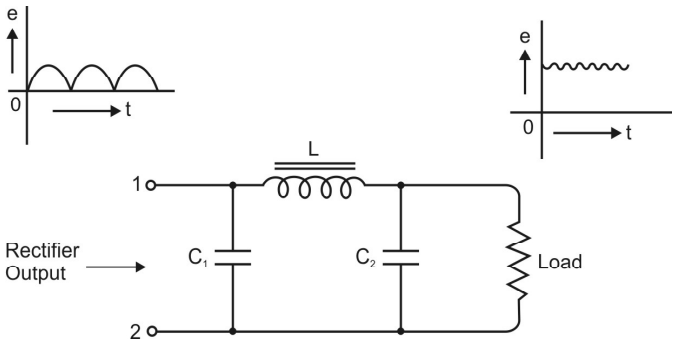


Fig. 1.1

The pulsating output from the rectifier is applied across the input terminals (i.e. terminals 1 and 2) of the filter. The filtering action of the three components viz C_1, L and C_2 of this filter is described below:

a. The *filter capacitor* C_1 offers low rectagnce to a.c. component of rectifier output while it offers infinite reactance to the d.c. component. Therefore, capacitor C_1 bypasses an appreciable amount of a.c. component while the d.c. component continues its journey to the choke L.

b. The *choke L* offers high reactance to the a.c. component but it offers almost zero reactance to the d.c. component. Therefore, it allows the d.c. component to flow through it, while the unbypassed a.c. component is blocked.

c. The *filter capacitor* C_2 bypassed the a.c. component which the choke has failed to block. Therefore, only d.c. component appears across the load and that is what we desire.

Example 23. For the circuit shown in Fig. 6.44, find the output d.c. voltage.

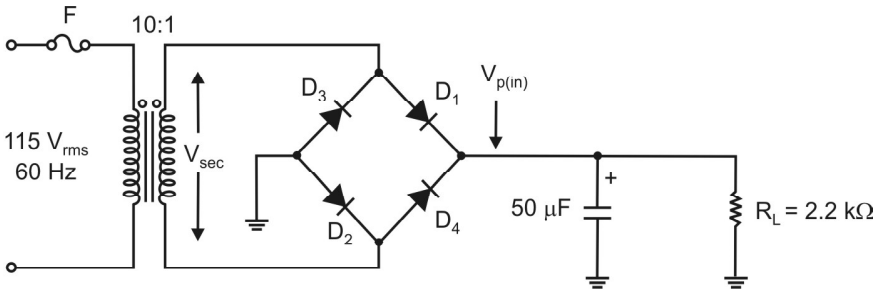


Fig. 1.1

Solution. It can be proved that output d.c. voltage is given by :

$$V_{dc} = V_{p(in)}\left(1-\frac{1}{2f R_L C}\right)$$

Here $V_{p(in)}$ = Peak rectified full-wave voltage applied to the filter

f = Output frequency

Peak primary voltage, $V_{p(prim)} = \sqrt{2}\,x115 = 163\,V$

Peak secondary voltage, $V_{p(sec)} = \left(\frac{1}{10}\right) x\ 163 = 16.3$

Peak full-wave rectified voltage at the filter input is

$$V_{p(in)} = V_{p(sec)} - 20 \times 0.7 = 16.3 - 1.4 = 14.9V$$

For full-wave rectification, $f = 2 f_{in} = 2 \times 60 = 120$ Hz

Now $$\frac{1}{2f R_L C} = \frac{1}{2x120x(2.2x10^3)x(50x10^{-6})} = 0.038$$

$\therefore$ $$V_{dc} = V_{p(in)}\left(1-\frac{1}{2f R_L C}\right) = 14.9(1-0.038) = 14.3V$$

Example 24. The choke of Fig. 6.45 has a d.c. resistance of 25Ω. What is the d.c. voltage if the full-wave signal into the choke has a peak value of 25.7V?

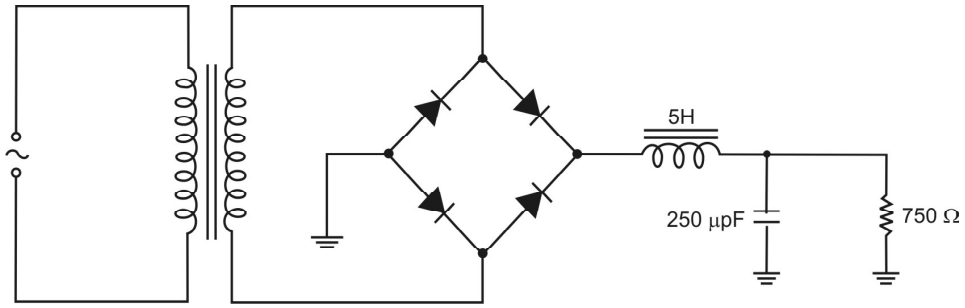


Fig. 1.1

Solution. The output of a full-wave rectifier has a d.c. component and an a.c. component. Due to the presence of a.c. component, the rectifier output has a pulsating character as shown in Fig. 6.46. The maximum value of the pulsating output is V_m and d.c. component is $V'_{dc} = 2\,V_m/\pi$.

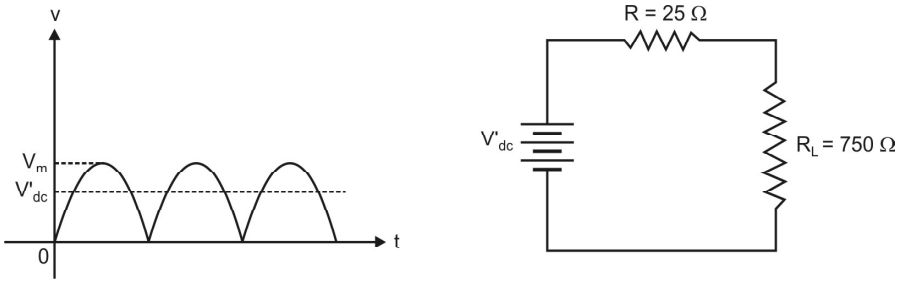


Fig. 1.1

For d.c. component V'dc, the choke resistance is in series with the load as shown in Fig. 6.47.

$\therefore$ Voltage across load, $V_{dc} = \dfrac{V'_{dc}}{R+R_L} x R_L$

In our example, $V'_{dc} = \dfrac{2V_m}{\pi} = \dfrac{2x25.7}{\pi} = 16.4\,V$

$\therefore$ Voltage across load, $V_{dc} = \dfrac{V'_{dc}}{R+R_L} x R_L = \dfrac{16.4}{25+750} x 750 = \mathbf{15.9\,V}$

The voltage across the load is 15.9V d.c. plus a small ripple.

BIPOLAR JUNCTION TRANSISTOR

Introduction to BJT

When a third doped element is added to a crystal diode in such a way that two pn junctions are formed, the resulting device is known as a transistor. The transistor – an entirely new type of electronic device – is capable of achieving amplification of weak signals in a fashion comparable and often superior to that realised by vacuum tubes. Transistors are far smaller than vacuum tubes, have no heating power and may be operated in any position. They are mechanically strong, have practically unlimited life and can do some jobs better than vacuum tubes.

Invented in 1948 by J. Bardeen and W.H. Brattain of Bell Telephone Laboratories, U.S.A. transistor has now become the heart of most electronic applications. Though transistor is only slightly more than 58 years old, yet it is fast replacing vacuum tubes in almost all applications. In this chapter, we shall focus our attention on the various aspects of transistors and their increasing applications in the fast developing electronics industry.

Types of BJT

A transistor consists of two pn junction formed by sandwiching either p-type or n-type semiconductor between a pair of opposite types. Accordingly; there are two types of transistors, namely;

i. n-p-n transistor ii. p-n-p transistor

An n-p-n transistor is composed of two n-type semiconductors separated by a thin section of p-type as shown in Fig. 8.1 (i). However, a p-n-p transistor is formed by two p-sections separated by a thin section of n-type as shown in Fig. 8.1.

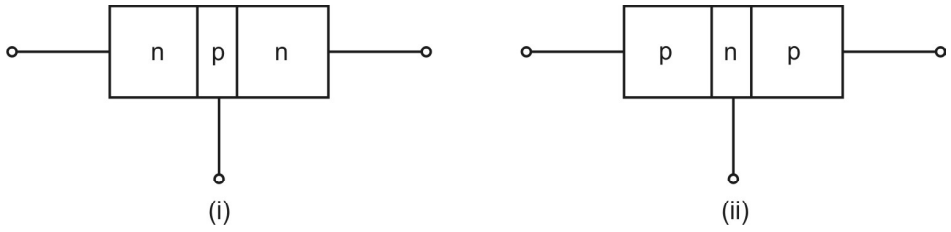


Fig. 1.1.

In the type of transistor, the following points may be noted:

i. These are two pn junctions. Therefore, a transistor may be regarded as a combination of two diodes connected back to back.

ii. There are three terminals, one taken from each type of semiconductor.

iii. The middle section is a very thin layer. This is the most important factor in the function of a transistor.

Origin of the name "Transistor". When new devices are invented, scientists often try to devise a name that will appropriately describe the device. A transistor has two pn junctions. As discussed later, one junction is forward biased and the other is reverse biased. The forward biased junction has a low resistance path whereas a reverse biased junction has a high resistance path. The weak signal is introduced in the low resistance circuit and output is taken from the high resistance circuit. Therefore, a transistor transfers a signal from a low resistance to high resistance. The prefix 'trans' means the signal transfer property of the device while 'istor' classifies it as a solid element in the same general family with resistors.

Naming the Transistor Terminals

A transistor (pnp or npn) has three sections of doped semiconductors. The section on one side is the emitter and the section on the opposite side is the collector. The middle section is called the bas and forms two junctions between the emitter and collector.

i. **Emitter.** The section on one side that supplies charge carries (electrons or holes) is called the emitter. The emitter is always forward biased w.r.t. base so that it can supply a large number of majority carriers. In Fig. 8.2 (i), the emitter (p-type) or pnp transistor is forward biased and supplies hole charges to tis junction with the base. Similarly, in Fig. 8.2 (ii), the emitter (n-type) of npn transistor has a forward bias and supplies free electrons to its junction with the base.

ii. **Collector.** The section on the other side that collects the charges is called collector. The collector is always reverse biased. Its function is to remove charges from its junction with the base. In Fig. 8.2 (i), the collector (p-type) of pnp transistor has a reverse bias and receives hole charges that flow in the output circuit. Similarly, in Fig. 8.2 (ii), the collector (n-type) of npn transistor has reverse biase and receives electrons.

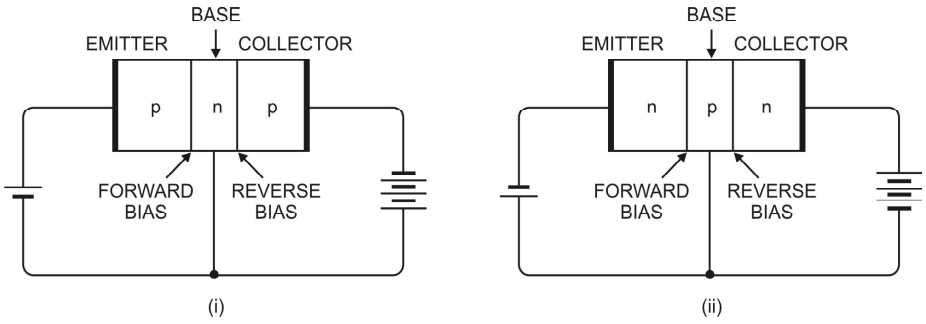


Fig. 1.1.

iii. **Base.** The middle section which forms two pn-junction between the emitter and collector is called the base. The base-emitter junction is forward biased, allowing low resistance for the emitter circuit. The bas-collector junction is reverse biased and provides high resistance in the collector circuit.

Some Facts about the Transistor

Before discussing transistor action, it is important that the reader may keep in mind the following facts about the transistor.

i. The transistor has three regions, namely; emitter, base and collector. The base is much thinner than the emitter while collector is wider than both as shown in Fig. 8.3. However, for the sake of convenience, it is customary to show emitter and collector to be of equal size.
ii. The emitter is heavily droped so that it can inject a large number of charge carriers (electrons or holes) into the base. is lightly doped and very thin; it passes most of the emitter unjected charge carriers to be collector. The collector is moderately doped.
iii. The transistor hastwo pn juncitons i.e. it its like two diodes. The junction between emitter and base may be called emitter-base diode or simply the emitter diode. The junction between the base and collector may be called collector-base diode or simply collector diode.
iv. The emitter diode is always forward biased whereas colelctor diode is always reverse biased.
v. The resistance of emitter diode (forward biased) is very small as compared to collector diode (reverse biased). Therefore, forward bias applied to the emitter diode is generally very small whereas reverse bias on the collector diode much higher.

Transistor Action

The emitter-base junction of a transistor is forward biased whereas collector-base junction is reverse biased. If for a moment, we ignore the presence of emitter-base junction, the practically no current would flow in the collector circuit because of the reverse bias. However, if the emitter-base junction is also present, then forward bias on it causes the emitter current to flow. It is seen that this emitter current almost entirely flows in the collector circuit. Therefore, the current in the collector circuit depends upon the emitter current. If the emitter current is zero, then collector current is nearly zero. However, if the emitter current is 1mA, then collector current is also about mA. This is precisely what happens in a transistor. We shall now discuss this transistor action for npn and pnp transistors.

Working of npn Transistor

Fig. 8.4 shows the npn transistor with forward bias to emitter-base junction and reverse bias to collector-base junction. The forward bias causes the electrons in the n-type emitter to flow towards the base. This constitutes the emitter current I_E. As these electrons flow through the p-type base, they tend to combine with holes. As the base is lightly doped and very thin, therefore, only a few electrons (less than 5%) combine with holes to constitute base current I_B. The remainder (more than 95%) cross over into the collector region to constitute collector current IC. In this way, almost the entire emitter current flows in the collector circuit. It is clear that emitter current is the sum of collector and base currents i.e.

$$I_E = I_B + I_C$$

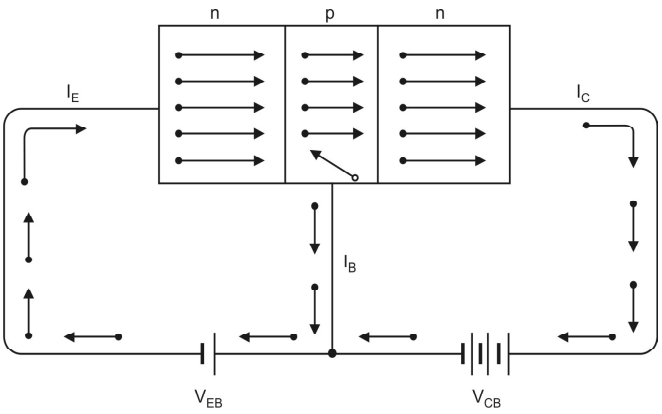


Fig. 2.30. Basic connection of npn transistor

Working of pnp transistor

Fig 8.5 shows the basic connection of a pnp transistor. The forward bias causes the holes in the p-type emitter to flow towards the base. This constitutes the emitter current I_E. As these holes cross into n-type base, they tend to combine with the electrons. As the base is lightly doped and very thin, therefore, only a few holes (less than 5%) combine with the electrons. The remainder (more than 95%) cross into the collector region to constitute collector current I_C. In this way, almost the entire emitter current flows in the collector circuit. It may be noted that current conduction within pnp transistor is by holes. However, in the external connecting wires, the current is still by electrons.

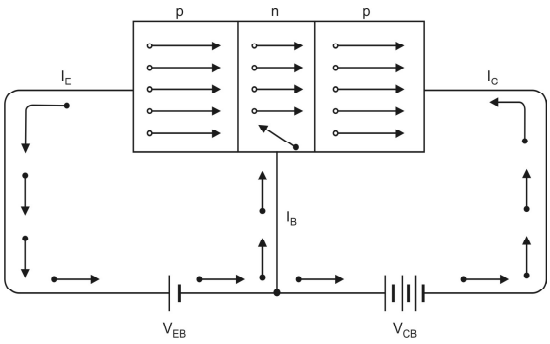


Fig. 2.30. Basic connection of pnp transistor

Importance of Transistor Action. The input circuit (i.e. emitter-base junction) has low resistance because of forward bias whereas output circuit (i.e. collector-base junction) has high resistance due to reverse bias. As we have seen, the input emitter current almost entirely flows in the collector circuit. Therefore, a transistor transfers the input signal current from a low-resistance circuit to a high-resistance circuit. This is the key factor responsible for the amplifying capability of the transistor. We shall discuss the amplifying property of transistor later in this chapter.

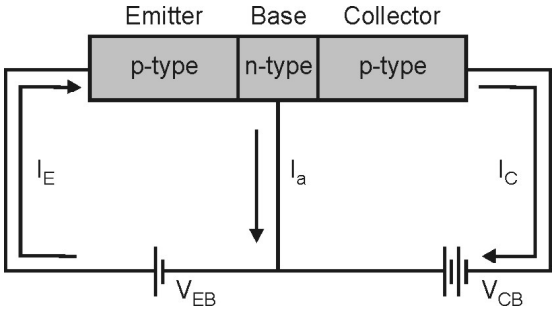


Fig. Conventional Current

Note. There are two basic transistor types : the **bipolar junction transistor** (BJT) and **field effect transistor** (FET). As we shall see, these two transistor types differ in both their operating characteristics and their internal construction. **Note that when we use term transistor, it means bipolar junction transistor** (BJT). The term comes from the fact that in a bipolar transistor, there are two types of charge carriers (viz. electrons and holes) that play part in conductions. Note that bimeans two and polar refers to pilarities. The field-effect transistor is simply referred to as FET.

Transistor Symbols

In the earlier diagrams, the transistors have been shown in diagrammatic form. However, for the sake of convenience, the transistors are represented by schematic diagrams. The symbols used for npn and pnp transistors are shown in Fig. 8.6.

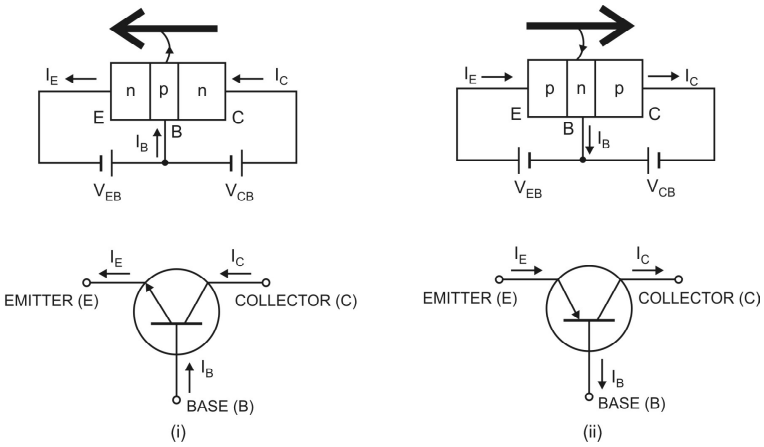


Fig. 1.1.

Note that emitter is shown by an arrow which indicates the direction of conventional current flow with forward bias. For npn connection, it is clear that conventional current flows out of the emitter as indicated by the outgoing arrow in Fig. 8.6 (i). Similarly, for pnp connection, the conventional current flows into the emitter as indicated by inward arrow in Fig. 8.6 (ii).

Transistor Circuit as an Amplifier

A transistor raises the strength of a weak signal and thus acts as an amplifier. Fig. 8.7 shows the basic circuit of a transistor amplifier. The weak signal is applied between emitter-base junction and output is taken across the load R_C connected in the collector circuit. In order to achieve faithful amplification, the input circuit should always remain forward biased. To do so, a d.c. voltage V_{EE} applied in the input circuit in addition to the signal as shown. This d.c. voltage is known as bias voltage and its magnitude is such that it always keeps the input circuit forward biased regardless of the polarity of the signal.

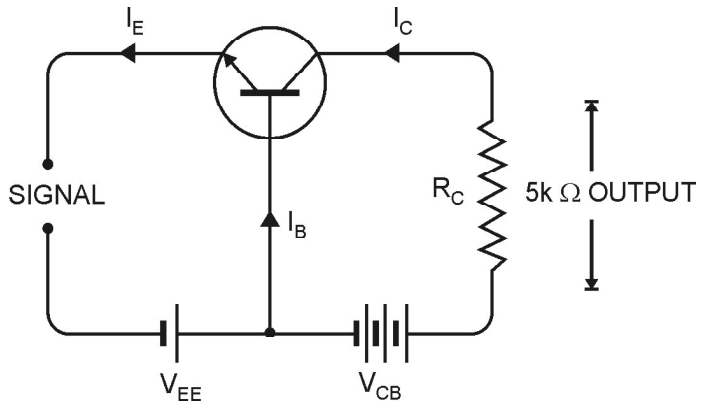
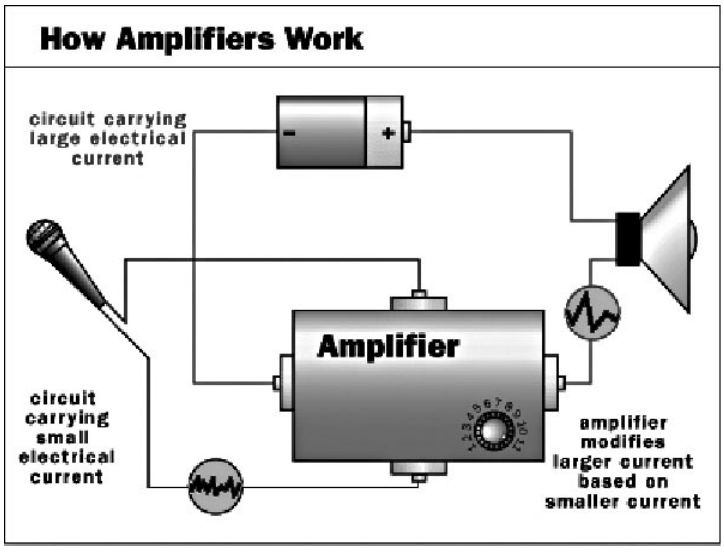


Fig. 1.1

As the input circuit has low resistance, therefore, a small change in signal voltage causes an appreciable change in emitter current. This causes almost the same change in collector current due to transistor action. The collector current flowing through a high load resistance R_C produces a large voltage across it. Thus, a weak signal applied in the input circuit appears in the amplifier form in the collector circuit. It is in this way that a transistor acts as an amplifier.

Illustration. The action of a transistor as an amplifier can be made more illustrative if we consider typical circuit values. Suppose collector load resistance R_C = 5 kΩ. Let us further assuem that a change of 0.1V in signal voltage aproduces a change of 1 mA in emitter current. Obviously, the change in collector current would also be approximately 1 mA. This collector current flowing through collector load R_C would produce a voltage = 5 kΩ x 1mA = 5V. Thus, a change of 0.1V in the signal has caused a change of 5V in the output circuit. In other words, the transistor has been able to raise the voltage level of the signal from 0.1V to 5V i.e. voltage amplification is 50.



Example 25. A common base transistor amplifier has an input resistance of 20Ω and output resistance of 100kΩ. The colelctor load is 1kΩ. If a signal of 500 mV is applied between emitter and base, find the voltage amplification. Assume α_{ac} to be nearly one.

Solution. Fig. 8.8 shows the conditions of the problem. Note that output resistance is very high as compared to input resistance. This is not surprising because input junction (base to emitter) of the transistor is forward biased while the output junction (base to collector) is reverse biased.

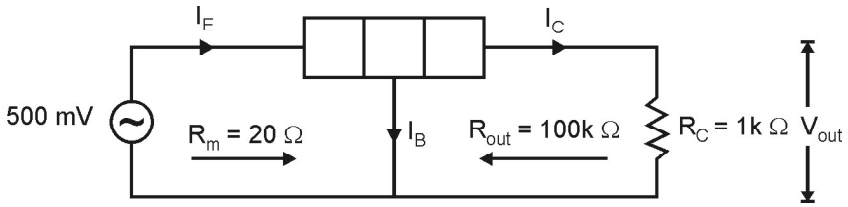


Fig. 8.8

$$\text{Input current, } I_E = \frac{\text{Signal}}{R_{in}} = \frac{500\,\text{mV}}{20\,\Omega} = 25\,\text{mA}.$$ Since α_{ac} is nearly 1, output curren, $I_C = I_E = 25\,\text{mA}$.

Output voltage, $V_{out} = I_C R_C$ = 25 mA x 1 kΩ = 25 V

$$\text{Voltage amplification, } A_v = \frac{V_{out}}{\text{signal}} = \frac{25\,\text{V}}{500\,\text{mV}} = \mathbf{50}$$

Comments. The reader may note that basic amplifying action is produced by transferring current from a low-resistance to a high-resistance circuit. Consequently, the name transistor is given to the device by combining the two terms given in below:

Transfer + Resistor ⟶ Transistor

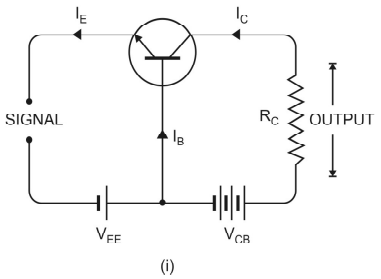
Transistor Connections

There are three leads in a transistor viz., emitter, base and collector terminals. However, when a transistor is to be connected in a circuit, we require four terminals; two for the output. This difficulty is overcome by making one terminal of the transistor common to both input and output terminals. The input is fed between this common terminal and one of the other two terminals. The output is obtained between the common terminal and the remaining terminal. Accordingly; a transistor can be connected in a circuit in the following three ways:

i. common base connection
ii. common emitter connection
iii. common collector connection

Common Base Connection

In this circuit arrangement, 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 connection. In Fig. 8.9 (i), a common base npn transistor circuit is shown whereas Fig. 8.9 (ii) shows the common base pnp transistor circuit.



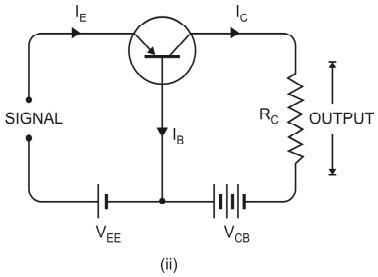


Fig. 1.1.

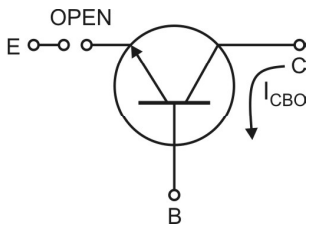


Fig. 1.1.

Current amplification factor (α). It is the ratio of output current to input current. In a common base connection, the input current isthe 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.

It is clear that current amplification factor is less than unitry. This value can be increased (but not more than unitry) by decreasing the base current. This is achieved by making the base thin and doping it lightly. Practical values of α in commercial transistors range from 0.9 to 0.99.

Expression for Collector Current. The whole of emitter current does not reach the collector. It is because a small percentage of it, as a result of electron-hole combinations occurring in base area, gives rise to base current. Moreover, as the collector-base junction is reverse biased, therefore, some leakage current flows due to minority carriers. It follows, therefore, that total collector current consists of :

i. That part of emitter current which reaches the collector current terminal i.e. αI_E.

ii. The leakage current $I_{leakage}$. This current is due to the movement of minority carriers across base-collector junction on account of it being reverse biased. This is generally much smaller than αI_E.

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

It is clear that if $I_E = 0$ (i.e., emitter circuit is open), a small leakage current still flows in the collector circuit. This $I_{leakage}$ is abbreviated as I_{CBO}, meaning collector-base current with emitter open. The I_{CBO} is indicated in Fig. 8.10.

$\therefore$ $I_C = \alpha I_E + I_{CBO}$

Now $I_E = I_C + I_B$

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

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

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

Relation (i) or (ii) can be to find I_C. It is further clear from relations that the collector current of a transistor can be controlled by either the emitter or base current.

Fig. 8.11 shows the concept of I_{CBO}. In CB configuration, a small collector current flows even when the emitter current is zero. This is the leakage collector current (i.e. the collector current when emitter is open) and is denoted by I_{CBO}. When the emitter voltage V_{EE} is also applied, the various currents are as shown in Fig. 8.11 (ii).

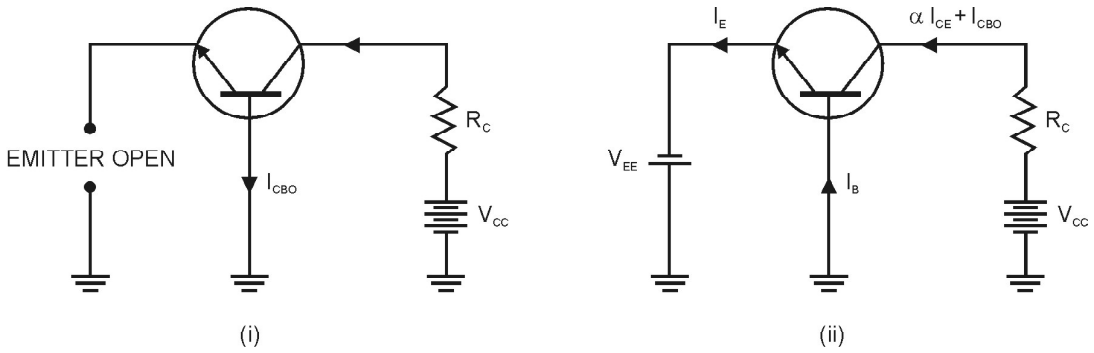


Fig. 8.11

Note. Owing to 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. However, for high power applications, it will appear in microampere range. Further, I_{CBO} is very much temperature dependent; it increases rapidly with the increase in temperature. Therefore, at higher temperature, I_{CBO} plays an important role and must be taken care of in calculations.

Example 26. In a common base connection, $I_E = 1$mA, $I_C = 0.95$mA. Calculate the value of I_B.

Solution. Using the relation, $I_E = I_B + I_C$

or $1 = I_B + 0.95$

$\therefore$ $I_B = 1 - 0.95 =$ **0.05 mA**

Example 27. In a common base connection, current amplification factor is 0.9. If the emitter current is 1mA, determine the value of base current.

Solution. Here, $\alpha = 0.9$, IE = 1 mA

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

or $I_C = \alpha I_E = 0.9 \times 1 = 0.9$ mA

Also $I_E = I_B + I_C$

$\therefore$ Base current, $I_B = I_E - I_C = 1 - 0.9=$ **0.1 mA**

Example 28. In a common base connection, I_C = 0.95 mA and I_B = 0.05 mA. Find the value of α.

Solution. We know $I_E = I_B + I_C = 0.05 + 0.95 = 1$ mA

$\therefore$ Current amplification factor, $\alpha = \frac{I_C}{I_E} = \frac{0.95}{1}$ = **0.95**

Example 29. In a common base connection, the emitter current is 1mA. If the emitter circuit is open, the collector current is 50 μA. Find the total collector current. Given that α = 0.92

Solution. Here, I_E = 1 mA, α = 0.92, I_{CBO} = 50 μA

$\therefore$ Total collector current, $I_C = \alpha I_E + I_{CBO} = 0.92 \times 1 + 50 \times 10^{-3}$

$= 0.92 + 0.05 =$ **0.97 mA**

Example 30. In a common base connection, α = 0.95. The voltage drop across 2kΩ resistance which is connected in the collector is 2V. Find the base current.

Solution. Fig. 8.12 shows the required common base connection. The voltage drop across R_C (= 2kΩ) is 2V.

$\therefore$ I_C = 2 V/2 kΩ = 1 mA

Now $\alpha = I_C/I_E$

$\therefore$ $I_E = \frac{I_C}{\alpha} = \frac{1}{0.95} = 1.05\,\text{mA}$

Using the relation, $I_E = I_B + I_C$

$I_B = I_E - I_C = 1.05 - 1$

$=$ **0.05 mA**

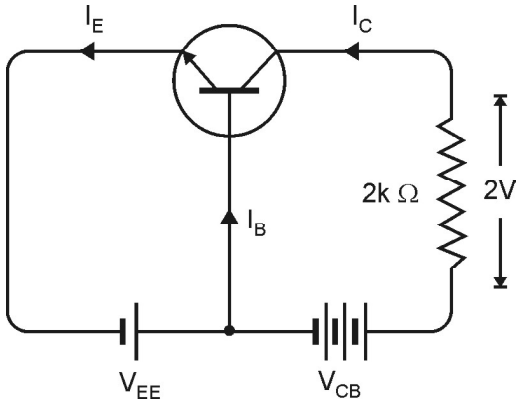


Fig. 1.12

Example 31. For the common base circuit shown in Fig. 8.13, determine IC and VCB. Assume the transistor to be of silicon.

Solution. Since the transistor is of silicon, VBE = 0.7V. Applying Kirchhoff's voltage law to the emitter-side loop, we get,

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

or $I_E = \frac{V_{EE} - V_{BE}}{R_E}$

$= \frac{8\text{V} - 0.7\text{V}}{1.5\,\text{k}\Omega} = 4.87\,\text{mA}$

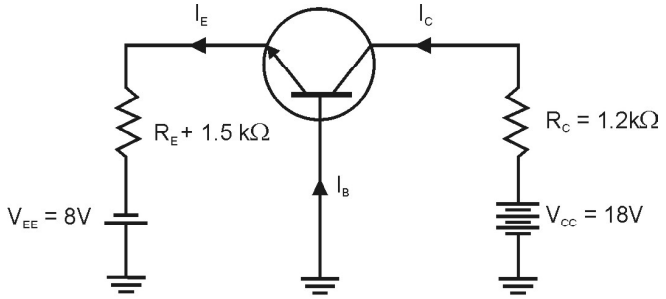
$\therefore$ $I_C \simeq I_E =$ **4.87 mA**

Applying Kirchoff's voltage law to the collector-side loop, we have,

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

$\therefore$ $V_{CB} = V_{CC} - I_C R_C$

$=$ 18 V – 4.87 mA x 1.2 kΩ = **12.16 V**



(i)

Fig. 1.13

Characteristics of Common Base Connection

The complete electrical behaviour of a transistor can be described by stating the interrelation of the various currents and voltages. These relationships can be conveniently displayed graphically and the curves thus obtained are known as the characteristics of transistor. The most important characteristics of common base connection are input characteristics and output characteristics.

Input Characteristics. It is the curve between emitter current I_E and emitter-base voltage V_{EB} at constant collector-base voltage V_{CB}. The emitter current is generally taken along y-axis and emitter-base voltage along x-axis. Fig. 8.14 shows the input characteristics of a typical transistor in CB arrangement. The following points may be noted from these characteristics :

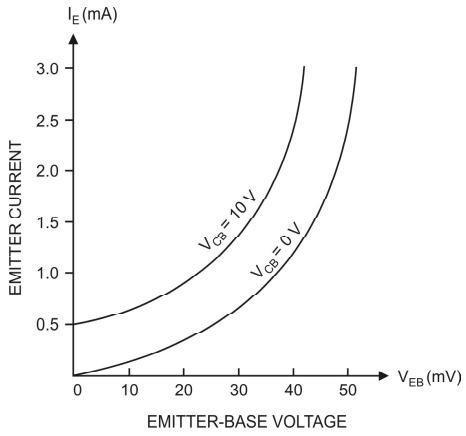


Fig. 1.1.

i. The emitter current I_E increases reapidly with small increase in emitter-base voltage V_{EB}. It means that input resistance is very small.

ii. The emitter current is almost independent of collector-base voltage V_{CB}. This leads to the conclusion that emitter current (and hence collector current) is almost independent of collector voltage.

Input resistance. It is the ratio of change in emitter-base voltage (ΔV_{EB}) to the resulting change in emitter current (I_E) at constant collector-base voltage (V_{CB}) i.e.

$$\text{Input resistance, } r_i \quad \frac{\Delta V_{BE}}{\Delta I_E} \quad \text{at constnat } V_{CB}$$

In fact, input resistance is the opposition offered to the signal current. As a very small V_{EE} sufficient to produce a large flow of emitter current I_E, therefore, input resistance is quite small of the order of a few ohms.

Output characteristic. It isthe curve between collector current I_C and collector-base voltage V_{CB} at constant emitter current I_E. Generally, collector current is taken along y-axis and collector-base voltage along x-axis. Fig. 8.15 shows the output characteristics of a typical transistor in CB arrangment.

The following points may be noted from the characteristics :

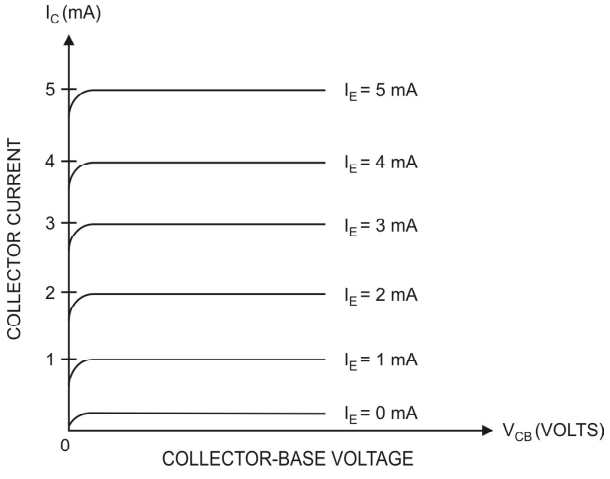


Fig. 1.1.

i. The collector current I_C varies with V_{CB} only at very low voltage (<1V). The transistor is never operated in this region.
ii. When the value of V_{CB} is raised above 1 – 2V, the collector current becomes constant as indicated by straight horizontal curves. It means that now I_C is independent of V_{CB} and depends upon I_E only. This is consistent with the theory that the emitter current flows almost entirely to the collector terminal. The transistor is always operated in this region.
iii. A very large change in collector-base voltage produces only a tiny change in collector current. This means that output resistance is very high.

Output resistance. It is the ratio of change in collector-base voltage (ΔV_{CB}) to the resulting change in collector current (ΔI_C) at constant emitter current i.e.

Output resistance, $r_o = \dfrac{\Delta V_{BE}}{\Delta I_B}$ at constnat I_E

The output resistance of CB circuit is very high, of the order of several tens of kilo-ohms. This is not surprising because the collector current changes very slightly with the change in V_{CB}.

Common Emitter Connection

In this circuit arrangement, 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 output circuits and hence the name common emitter connection. Fig. 8.16 (i) shows common emitter npn transistor circuit whereas Fig. 8.16 (ii) shows common emitter pnp transistor circuit.

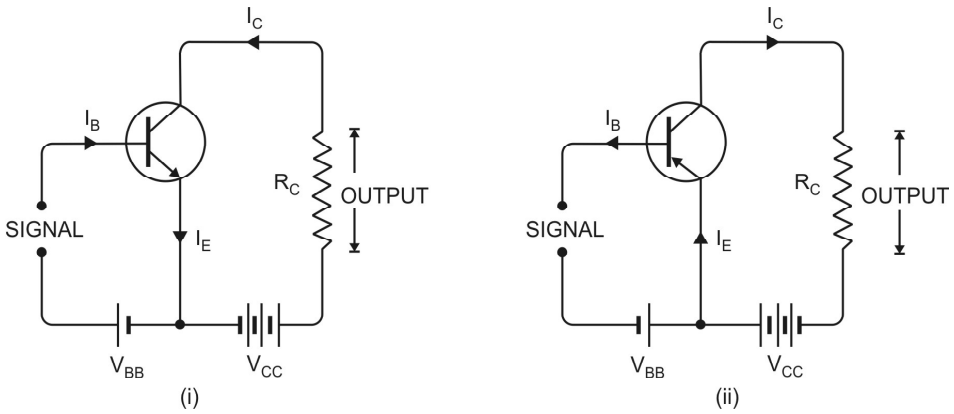


Fig. 1.1.

Base current amplification factor (β)

In common emitter connection, 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.

$$\beta = \frac{\Delta I_C}{\Delta I_B}$$

In almost any transistor, less than 5% of emitter current flows as the base current. Therefore, the value of β is generallay greater than 20. Usually, its value ranges from 20 to 500. This type of connection is frequently used as it gives appreciable current gain as well as voltage gain.

Relation between β and α. A simple relation exists between β and α. This can be derived as follows:

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

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

Now $I_E = I_B + I_C$

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

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

Substituting the value of ΔI_B in exp. (i), we get,

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

Dividing the numerator and denominator of R.H.S. of exp. (iii) by IE. we get,

$$\beta = \frac{\Delta I_C / \Delta I_E}{\frac{\Delta I_E}{\Delta I_E} - \frac{\Delta I_C}{\Delta I_E}} = \frac{\alpha}{1-\alpha} \qquad \left[\because \alpha = \frac{\Delta I_C}{\Delta I_E}\right]$$

$$\therefore \qquad \beta = \frac{\alpha}{1-\alpha}$$

It is clear that as α approaches unity, β approaches infinity. In other words, the current gain in common emitter connection is very high. It is due to this reason that this circuit arrangement is used in about 90 to 95 percent of all transistor applications.

Expression for collector current. In common emitter circuit IB is the input current and is the output current.

We know $I_E = I_B + I_C$

and $I_C = \alpha I_E + I_{CBO}$

From exp. (ii), we get, $I_C = \alpha I_E + I_{CBO} = \alpha (I_B + I_C) + I_{CBO}$

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

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

From exp. (iii), it is apparent that if IB = 0 (i.e. base circuit is open), the collector current will be the current to the emitter. This is abbreviated as ICEO, meaning collector-emitter current with base open.

$$\therefore \qquad I_{CEO} = \frac{1}{1-\alpha} I_{CEO}$$

Substitude the value of $\frac{1}{1-\alpha} I_{CEO} = I_{CEO}$ in exp. (iii) we get,

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

$$I_C = \beta I_B + I_{CEO} \qquad \left(\because \beta = \frac{\alpha}{1-\alpha}\right)$$

Concept of I_{CEO}. In CE configuration, a small collector current flows even when the base current is zero [See Fig. 8.17 (i)]. This is the collector cut off current (i.e. the collector current that flows when base is open) and is denoted by I_{CEO}. The value of I_{CEO} is much larger than I_{CBO}.

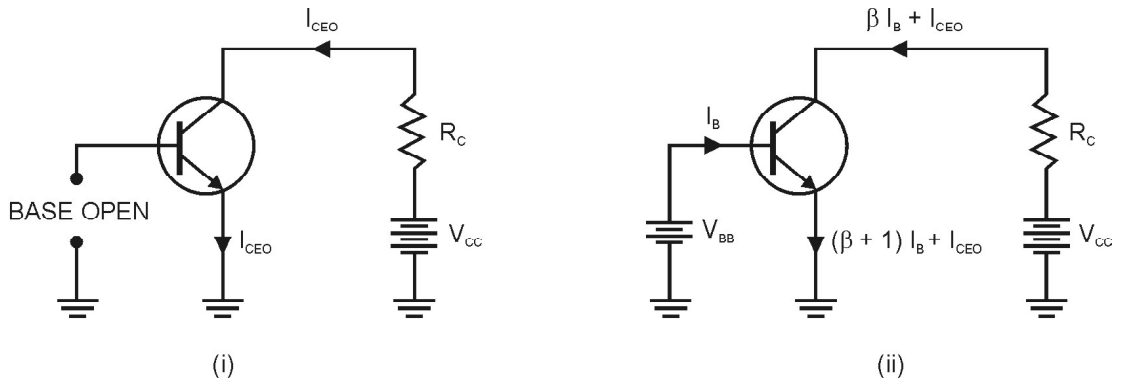


Fig. 1.17

When the base voltage is applied as shown in Fig. 8.17 (ii), then the various currents are :

Base current = I_B

Collector current = $\beta I_B + I_{CEO}$

Emitter current = Collector current + Base current

= $(\beta I_B + I_{CEO}) + I_B = (\beta + 1) I_B + I_{CEO}$

It may be noted here that :

$$I_{CEO} = \frac{I}{1-\alpha} I_{CBO} = (\beta+1) I_{CBO} \qquad \left[\because \frac{1}{1-\alpha} = \beta + 1\right]$$

Measurement fo Leakage Current

A very small leakage current flows in all transistor circuits. However, in most cases, it is quite small and can be neglected.

i. **Circuit for I_{CEO} test.** Fig. 8.18 shows the circuit for measuring I_{CEO}. Since base is open ($I_B = 0$), the transistor is in cut off. Ideally, $I_C = 0$ but actually there is a small current from collector to emitter due to minority carriers. It is called I_{CEO} (collector-to-emitter current with base open). This current is usually in the nA range for silicon. A faulty transistor will often have excessive leakage current.

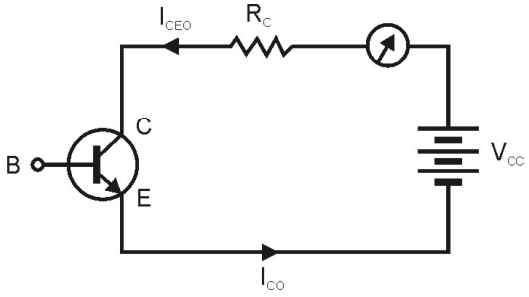


Fig. 1.18

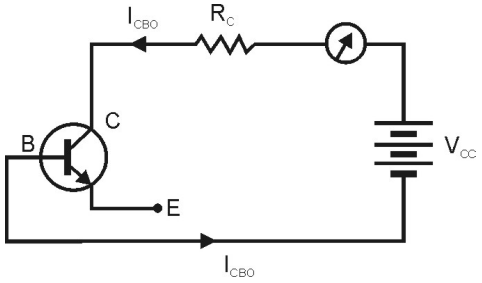


Fig. 1.19

ii. **Circuit for I_{CBO} test.** Fig. 8.19 shows the circuit for measuring I_{CBO}. Since the emitter is open ($I_E = 0$), there is a small current from collector to base. This is called I_{CBO} (collector-to-base current with emitter open). This current is due to the movement of minority carriers across base-collector junction. The value of I_{CBO} is also small. If in measurement, I_{CBO} is excessive, then there is a possibility that collector base is shorted.

Example 8.8. Find the value of β if (i) $\alpha = 0.9$ (ii) $\alpha = 0.98$ (iii) $\alpha = 0.99$.

Solution. i. $$\beta = \frac{\alpha}{1-\alpha} = \frac{0.9}{1-0.9} = \mathbf{9}$$

ii. $$\beta = \frac{\alpha}{1-\alpha} = \frac{0.98}{1-0.98} = \mathbf{49}$$

iii. $$\beta = \frac{\alpha}{1-\alpha} = \frac{0.99}{1-0.99} = \mathbf{99}$$

Example 8.9. Calculate I_E in a transistor for which $\beta = 50$ and $I_B = 20\ \mu A$.

Solution. Here $\beta = 50$, $I_B = 20\mu A = 0.02$ mA

Now $$\beta = \frac{I_C}{I_B}$$

$I_C = \beta I_B = 50 \times 0.02 = 1$ mA

Using the relation, $I_E = I_B + I_C = 0.02 + 1 = 1.02$ **mA**

Example 8.10. Find the α rating of the transistor shown in Fig. 8.20. Here determine the value of I_C using both α and β rating of the transistor.

Solution. Fig. 8.20 shows the conditions of the problem.

$$\alpha = \frac{\beta}{1+\beta} = \frac{49}{1+49} = \mathbf{0.98}$$

The value of I_C can be found by using either α or β rating as under:

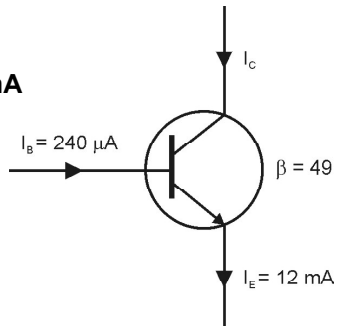


Fig. 8.20

$I_C = \alpha I_E = 0.98\ (12\ mA) = \mathbf{11.76\ mA}$

Also $I_C = \beta I_B = 49\ (240\ uA) = \mathbf{11.76\ mA}$

Example 8.11. For a transistor, $\beta = 45$ and voltage drop across IkΩ which is connected in the collector circuit is 1 volt. Find the base current for common emitter connection.

Solution. Fig. 8.21 shows the required common emitter connection. The voltage drop across R_C (= 1kΩ) is 1 volt.

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

Now $$\beta = \frac{I_C}{I_B}$$

$$\therefore \quad I_B = \frac{I_C}{\beta} = \frac{1}{45} = \mathbf{0.022\,mA}$$

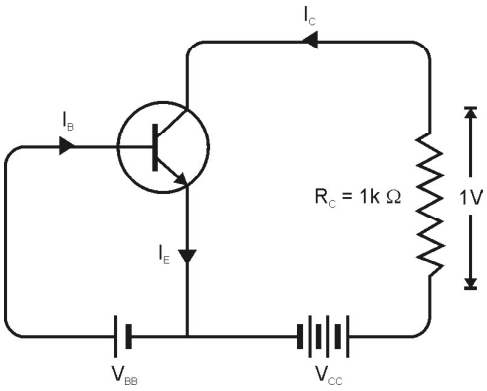


Fig. 1.21

Example 8.12. A transistor is connected in common emitter (CE) configuration in which collector supply is 8V and the voltage drop across resistance R_C connected in the collector circuit is 0.5V. The value of R_C = 800Ω. If $\alpha = 0.96$, determine :

i. collector-emitter voltage

ii. base current

Solution. Fig. 8.22 shows the required common emitter connection with various values.

i. Collector-emitter voltage,

$$V_{CE} = V_{CC} - 0.5 = 8 - 0.5 = \mathbf{7.5\ V}$$

ii. The voltage drop across R_C (= 800Ω) is 0.5 V.

$$\therefore \quad I_C = \frac{0.5}{800\,\Omega} = \frac{5}{8}\,mA = 0.625\,mA$$

Now $$\beta = \frac{\alpha}{1-\alpha} = \frac{0.96}{1-0.96} = 24$$

$$\therefore \text{ Base current, } I_B = \frac{I_C}{\beta} = \frac{0.625}{24} = \mathbf{0.026\,mA}$$

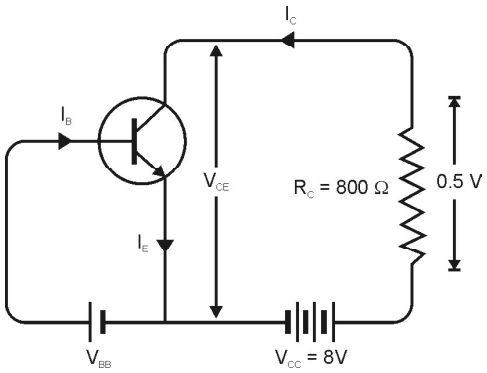


Fig. 1.22

Example 8.13. An n-p-n transistor at room temperature has its emitter disconnected. A voltage of 5V is applied between collector and base. With collector positive, a current of 0.2 μA flows. When the base is disconnected and the same voltage is applied between collector and emitter, the current found to be 20 μA. Find α, I_E and I_B when collector current is 1mA.

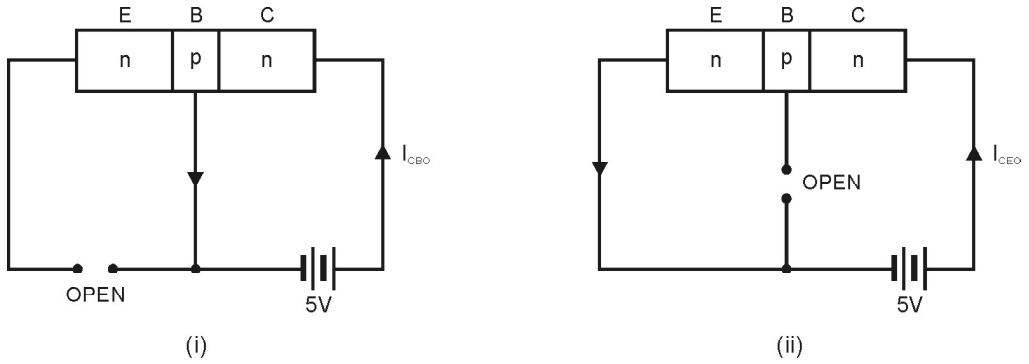


Fig. 1.23

Solution. When the emitter circuit is open [See Fig. 8.23 (i)], the collector-base junction is reverse biased. A small leakage current I_{CBO} flows due to minority carriers.

$\therefore \quad I_{CBO} = 0.2\ \mu A$ ...given

When base is open [See Fig. 8.23 (ii)], a small leakage current I_{CEO} flows due to minority carriers,

$\therefore \quad I_{CEO} = 20\ \mu A$

We know $\quad I_{CEO} = \dfrac{I_{CBO}}{1-\alpha}$

or $\quad 20 = \dfrac{0.2}{1-\alpha}$

$\therefore \quad \alpha = 0.99$

Now $\quad I_C = \alpha I_E + I_{CBO}$

Here $\quad I_C = 1mA = 1000\ \mu A\ ;\ \alpha = 0.99;\ I_{CBO} = 0.2\ \mu A$

$\therefore \quad 1000 = 0.99\ \alpha\ I_E + 0.2$

or $\quad I_E = \dfrac{1000-0.2}{0.99} = \mathbf{1010\,\mu A}$

and $\quad I_B = I_E - I_C = 1010 - 1000 = \mathbf{10\ \mu A}$

Example 8.14. The collector leakage current in a transistor is 300 μA in CE arrangement. If now the transistor is connected in CB arrangement, what will be the leakage current? Given that $\beta = 120$.

Solution. $\quad I_{CEO} = 300\ \mu A$

$$\beta = 120 \quad ; \quad \alpha = \frac{\beta}{\beta+1} = \frac{120}{120+1} = 0.992$$

Now, $\quad I_{CEO} = \dfrac{I_{CBO}}{1-\alpha}$

$\therefore \quad I_{CBO} = (1-\alpha)\ I_{CEO} = (1 - 0.992)\ \alpha\ 300 = \mathbf{2.4\ \mu A}$

Note that leakage current in CE arrangement (i.e. I_{CEO}) is much more than in CB arrangement (i.e. I_{CBO}).

Example 8.15. For a certain transistor, $I_B = 20\ \mu A$; $I_C = 2$ mA and $\beta = 80$. Calculate I_{CBO}.

Solution.

$$I_C = \beta I_B + I_{CEO}$$

or $\quad 2 = 80 \times 0.02 + I_{CEO}$

$\therefore \quad I_{CEO} = 2 - 80 \times 0.02 = 0.4$ mA

Now $\quad \alpha = \dfrac{\beta}{\beta+1} = \dfrac{80}{80+1} = 0.988$

$\therefore \quad I_{CBO} = (1-\alpha)\ I_{CEO} = (1 - 0.988)\ \alpha\ 0.4 = 0.0048$ mA

Example 8.16. Using diagrams, explain the correctness of the relation $I_{CEO} = (\beta + 1)\ I_{CBO}$.

Solution. The leakage current I_{CBO} is the current that flows through the base-collector junction when emitter is open as shown is Fig. 8.24. When th transistor is in CE arrangement, the base current (i.e. I_{CBO}) is multiplied by β in the collector as shown in Fig. 8.25.

$\therefore \quad I_{CEO} = I_{CBO} + \beta I_{CBO} = (\beta + 1)\ I_{CBO}$

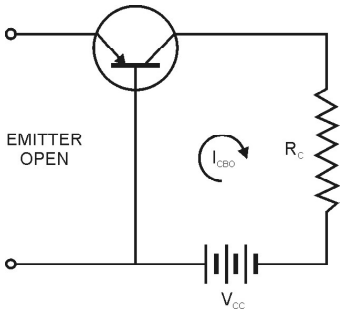


Fig. 1.24

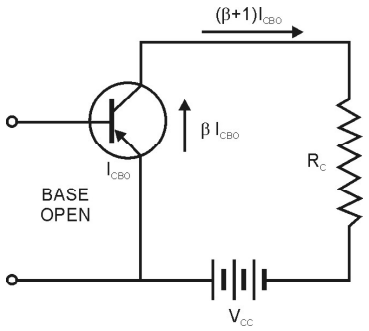


Fig. 1.25

Example 8.17. Determine V_{CB} in the transistor circuit shown in Fig. 8.26 (i). The transtor of silicon and has $\beta = 150$.

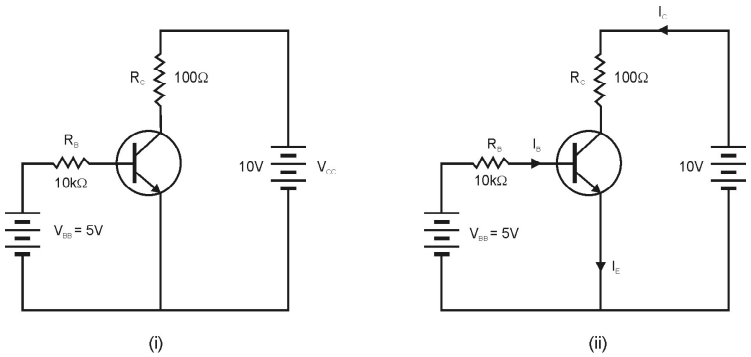


Fig. 1.26

Solution. Fig. 8.26 (i) shows the transistor circuit while Fig. 8.26 (ii) shows the various currents and voltages along with polarities.

Applying Kirchhoff's voltage law to base-emitter loop, we have,

$$V_{BB} = I_B R_B - V_{BE} = 0$$

or $$I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{5V - 0.7V}{10\,k\Omega} = 430\,\mu A$$

$\therefore$ $$I_C = \beta I_B = (150)(430\ \mu A) = 64.5\ mA$$

Now $$V_{CE} = V_{CC} - I_C R_C$$
$$= 10V\ (64.5\ mA)\ (1000\Omega) = 10V - 6.45V = 3.55V$$

We know that : $$V_{CE} = V_{CB} + V_{BE}$$

$\therefore$ $$V_{CB} = V_{CE} - V_{BE} = 3.55 - 0.7 = \mathbf{2.85V}$$

Example 8.18. In a transistor, I_B = 68 μA, I_E = 30 mA and β = 440. Determine the α rating of the transistor. Then determine the value of I_C using both the α rating and β rating of the transistor.

Solution.

$$\alpha = \frac{\beta}{\beta + 1} = \frac{440}{440 + 1} = \mathbf{0.9977}$$

$$I_C = \alpha\, I_E = (0.9977)\ (30\ mA) = \mathbf{29.93\ mA}$$

Also $$I_C = B\, I_B = (440)\ (68\ \mu A) = \mathbf{29.93\ mA}$$

Example 8.19. A transistor has the following rating : $I_{C\,(max)}$ = 500 mA and β_{max} = 300. Determine the maximum allowable value of I_B for the device.

Solution.

$$I_{B(max)} = \frac{I_{C(max)}}{\beta_{max}} = \frac{500\,mA}{300} = \mathbf{1.67\,mA}$$

For this transistor, if the base current is allowed to exceed 1.67 mA, the collector current will exceed its maximum rating of 500 mA and the transistor will probably be destroyed.

Example 8.20. Fig. 8.27 shows the open circuit failures in a transistor. What will be the circuit behaviour in each case?

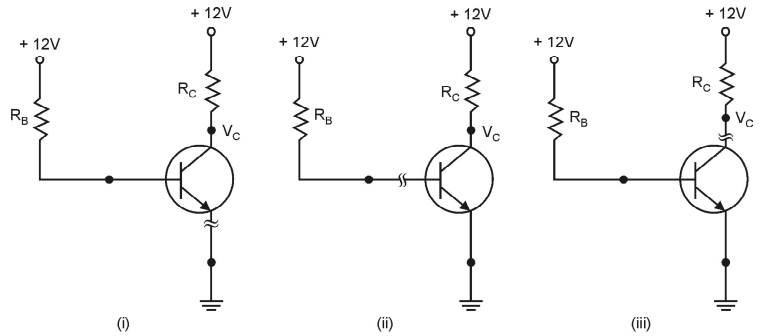


Fig. 1.27

Solution. Fig. 8.27 shows the open circuit failures in a transistor. We shall discuss the circuit behaviour in each case.

i. **Open emitter.** Fig. 8.27 (i) shows an open emitter failure in a transistor. Since the collector diode is not forward baised, it is OFF and there can be neither collector current nor base current. Therefore, there will be no voltage drops across either resistor and the voltage at the base and at the collector leads of the transistor will be 12V.

ii. **Open-base.** Fig. 8.27 (ii) shows an oepn base failure in a transistor. Since the base is open, there can be no base current so that the transistor is in cut-off. Therefore, all the transistor currents are 0A. In this case, the base and collector voltages will both be at 12V.

Note. It may be noted that an open failure at either the base or emitter will produce similar results.

iii. **Open collector.** Fig. 8.27 (iii) shows an open collector failure in a transistor. In this case, the emitter diode is still ON, so we expect to see 0.7V at the base. However, we will see 12V at the collector because there is no collector current.

Example 8.21. Fig. 8.28 shows the short circuit failures in a transistor. What will be the circuit behaviour in each case?

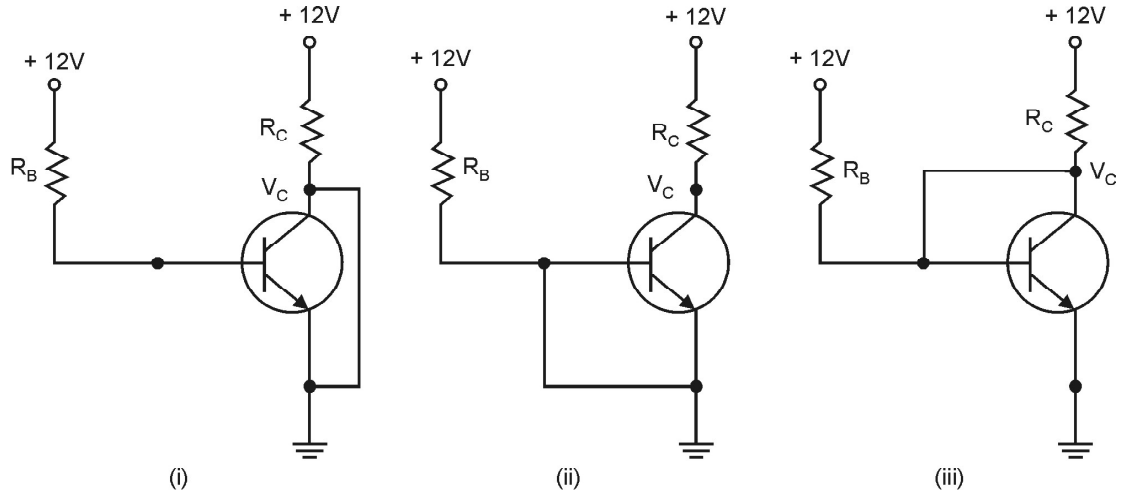


Fig. 1.22

Solution. Fig. 8.28 shows the shor circuit failures in a transistor. We shall discuss the circuit behaviour in each case.

i. **Collector-emitter short.** Fig. 8.28 (i) shows a short between collector and emitter. The emitter diode is still forward biased, so we expect to see 0.7V at the base. Since the collector is shorted to the emitter, $V_C = V_E = 0V$.

ii. **Base-emitter short.** Fig. 8.28 (ii) shows a short between base and emitter. Since the base is now directly connected to ground, $V_B = 0$. Therefore, the current through R_B will be diverted to ground and there is no current to forward bias the emitter diode. As a result, the transistor will be cut off and there is no collector current. So we will expect the collector voltage to be 12V.

iii. **Collector-base short.** Fig. 8.28 (iii) shows a short between the collector and the base. In this case, the emitter diode is still forward biased so $V_B = 0.7V$. Now, however, because the collector is shorted to the base, $V_C = V_B = 0.7V$.

Note. The collector-emitter short is probably the most common type of fault in a transistor. It is because the collector current (I_C) and collector-emitter voltage (V_{CE}) are responsible for the major part of the power dissipation in the transistor. As we shall see (See Art. 8.23), the power dissipation in a transistor is mainly dur to I_C and V_{CE} (i.e. $P_D = V_{CE} I_C$). Therefore, the transistor chip between the collector and the emitter is most likely to melt first.

Characteristics of Common Emitter Connection

The important characteristics of this circuit arrangement are the input characteristics and output characteristics.

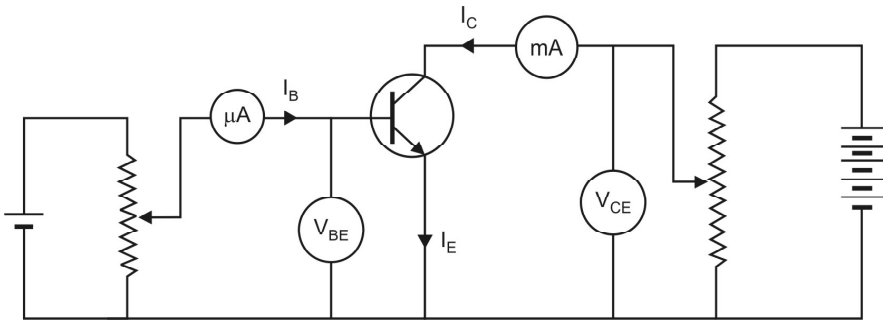


Fig. 8.29

Input Characteristic. It is the curve between base current I_B and base-emitter voltage V_{BE} at constant collector-emitter voltage V_{CE}.

The input characteristics of a CE connection can be determined by the circuit shown in Fig. 8.29. Keeping V_{CE} constant (say at 10V), note the base current I_B for various values of V_{BE}. Then plot the readings obtained on the graph, taking IB along y-axis and VBE along x-axis. This gives the input characteristic at V_{CE} = 10V as shown in Fig. 8.30. Following a similar procedure, a family of input characteristics can be drawn. The following points may be noted from the characteristics :

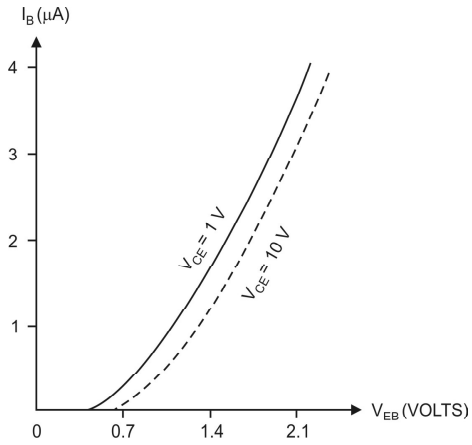


Fig. 1.1.

i. The characteristic resembles that of a forward biased diode curve. This is expected since the base-emitter section of transistor is a diode and it is forward biased.

ii. As compared to CB arrangement, I_B increases less repidly with V_{BE}. Therefore, input resistance of a CE circuit is higher than that of CB circuit.

Input resistance. It is the ratio of change in base-emitter voltage (ΔV_{BE}) to the change in base current (ΔI_B) at constant V_{CE} i.e.

$$\text{Input resistance, } r_i = \frac{\Delta V_{BE}}{\Delta I_B} \text{ at constant } V_{CE}$$

The value of input resistance for a CE circuit is of the order of a few hundred ohms.

Output Characteristic. It is the curve between collector current I_C and collector-emitter voltage V_{CE} at constant base current I_B.

The output characteristics of a CE circuit can be drawn with the help of the circuit shown in Fig. 8.29. Keeping the base current IB fixed at some value say, 5 uA, note the collector current IC for various values of VCE. Then plot the reasings on a graph, taking IC along y-axis and VCE along x-axis. This gives the output characteristics at IB = 5 uA as shown Fig. 8.31 (i). The test can be repeated for IB = 10 uA to obtain the new output characteristic as shown in Fig. 8.31 (ii). Following similar procedure, a family of output characteristics can be drawn as shown in Fig. 8.31 (iii).

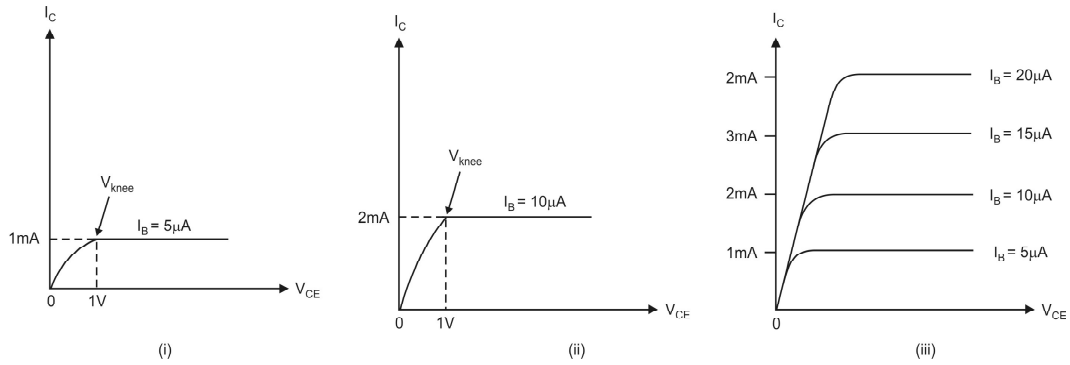


Fig. 1.1.

The following points may be noted from the characteristics:

i. The collector current I_C varies with V_{CE} for V_{CE} between 0 and 1V only. After this, collector current becomes almost constant and independent of V_{CE}. This value of V_{CE} upto which collector current I_C changes with VCE is called the knee voltage (V_{knee}). The transistors are always operated in the region above knee voltage.

ii. Above knee voltage, I_C is almost constant. However, a small increase in I_C with increasing V_{CE} is caused by the collector depletion layer getting wider and capturing a few more majority carriers before electron-hole combinations occur in the base area.

iii. For any value of V_{CE} above knee voltage, the collector current I_C is approximately equal to β x I_B.

Output resistance. It is the ratio of change in collector-emitter voltage (V_{CE}) to the change in collector current (I_C) at constant I_B i.e.

$$\text{Output resistance, } r_o = \frac{\Delta V_{CE}}{\Delta I_C} \text{ at constnat } I_B$$

It may be noted that whereas the output characteristics of C_B circuit are horizontal, they have moticeable slope for the C_E circuit. Therefore, the output resistance of a C_E circuit is less than that of C_B circuit. Its value is the order of 50 kΩ.

Common Collector Connection

In this circuit arrangement, 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. 8.32 (i) shows common collector npn transistor circuit whereas Fig. 8.32 shows common collector pnp circuit.

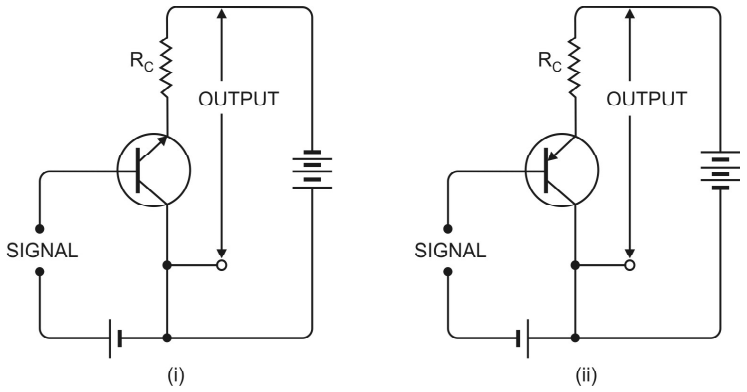


Fig. 1.1.

Current amplification factor γ

In common collector circuit, input current is the base current I_B and output current is the emitter current I_E. Therefore, current amplification in this circuit arrangement can be defined as under :

The ratio of change in emitter current (ΔI_E) to the change in base current (ΔI_B) is known as current amplification factor in common collector (CC) arrangement i.e.

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

This circuit provides about the same current gain as the common emitter as $\Delta I_E \simeq \Delta I_C$. However, its voltage gain is always less than 1.

Relation between γ and α

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

$$\alpha = \frac{\Delta I_C}{\Delta I_E}$$

Now $I_E = I_B + I_C$

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

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

Substituting the value of IB in exp. (i), we get,

$$\gamma = \frac{\Delta I_E}{\Delta I_E - \Delta I_C}$$

Dividing the numberator and denominator of R.H.S. by IE, we get,

$$\gamma = \frac{\frac{\Delta I_E}{\Delta I_E}}{\frac{\Delta I_E}{\Delta I_E} - \frac{\Delta I_C}{\Delta I_E}} = \frac{1}{1-\alpha} \qquad \left(\because \alpha = \frac{\Delta I_C}{\Delta I_E}\right)$$

$\therefore$ $$\gamma = \frac{1}{1-\alpha}$$

Expression for collector current

We know $I_C = \alpha I_E + I_{CBO}$

Also $I_E = I_B + I_C = I_B + (\alpha I_E + I_{CBO})$

$\therefore$ $I_E (1 - \alpha) = I_B + I_{CBO}$

or $$I_E = \frac{I_B}{1-\alpha} + \frac{I_{CBO}}{1-\alpha}$$

or $I_C = I_E = (\beta + 1) I_B + (\beta + 1) I_{CBO}$

iii. **Applications.** The common collector circuit has very high input resistance (about 750 kΩ) and very low output resistance (about 25Ω). Due to this reason, the voltage gain provided by this circuit is always less than 1. Therefore, this circuit arrangement is seldom used for amplification. However, due to relatively high input resistance and low output resitance, this circuit is primarily used for impedance matching i.e. for driving a low impedance load from a high impedance source.

Comparison of Transistor Connections

The comparison of various characteristics of the three connections is given below in the tabular form.

S.No.	Characteristic	Common Base	Common Emitter	Collector Collector
1.	Input resistance	Low (about 100Ω)	Low (about 750 Ω)	Very high (about 750 kΩ)
2.	Output resistance	Very high (about 450 kOΩ)	High (about 45 kΩ)	Low (about 50 Ω)
3.	Voltage gain	About 150	About 500	Less than 1
4.	Applications	For high frequency applications	For audio frequency application	For impedance matching
5.	Current gain	No (less than 1)	High (B)	Appreciable

The following points are worth noting about transistor arrangements :

i. **CB Circuit.** The input resistance (r_i) of CB circuit is low because I_E is high. The output resistance (r_o) is high because of reverse voltage at the collector. It has no current gain ($\alpha < 1$) but voltage gain can be high. The CB circuit is seldom used. The only advantage of CB circuit is that it provides good stability against increase in temperature.

ii. **CE Circuit.** The input resistance (r_i) of a CE circuit is high because of small I_B. Therefore, ri for a CE circuit is much higher than that of CB circuit. The output resistance (ro) of CE circuit is smaller than that of CB circuit. The current gain of CE circuit is large because I_C is much larger than I_B. The voltage gain of CE circuit is larger than that of CB circuit. The CE circuit is generally used because it has the best combination of voltage gain and current gain. The disadvantage of CE circuit is that the leakage current is amplifier in the circuit, but bias stabilisation methods can be used.

iii. **CC Circuit.** The input resistance (r_i) and output resistance (r_o) of CC circuit are respectively high and low as compared to other circuits. There is no voltage gain ($A_v < 1$) in a CC circuit. This circuit is often used for impedance matching.

Commonly Used Transistor Connection

Out of the three transistor connections, the common emitter circuit is the most efficient. It is used in about 90 to 95 per cent of all transistor applications. The main reasons for the widespread use of this circuit arrangement are :

i. **High Current Gain.** In a common emitter connection, I_C is the output current and I_B is the input current. In this circuit arrangement, collector current is given by :

$$I_C = \beta I_B + I_{CEO}$$

As the value of β is very large, therefore, the output current I_C is much more than the input current I_B. Hence, the current gain in CE arrangement is very high. It may range from 20 to 500.

ii. **High voltage and power gain.** Due to high current gain, the common emitter circuit has the highest voltage and power gain of three transistor connections. This is the major reason for using the transistor in this circuit arrangment.

iii. **Moderate output to input impedance ratio.** In a common emitter circuit, the ratio of output impedance to input imperance is small (about 50). This makes this circuit arrangement an ideal one for coupling between various transistor stages. However, in other connections, the ratio of output impedance to input impedance is very large and hence coupling becomes highly inefficient due to gross mismatching.

Transistor as an Amplifier in CE Arrangement

Fig. 8.33 shows the common emitter npn amplifier circuit. Note that a battery V_{BB} is connected in the input circuit in addition to the signal voltage. This a.c. voltage is known as bias voltage and its magnitude is such that it always keeps the emitter-junction forward biased regardless of the polarity of the signal source.

Operation. During the positive half-cycle of the signal, the forward bias across the emitter-base junction is increased. Therefore, more electrons flow from the emitter to the collector via the base. This causes an increase in collector current. The increased collector current produces a greater voltage drop across the collector load resistance R_C. However, during the negative half-cycle of the signal, the forward bias across emitter-base junction is decreased. Therefore, collector current decreases. This results in the decreased output voltage (in the opposite direction). Hence, an amplified output is obtained across the load.

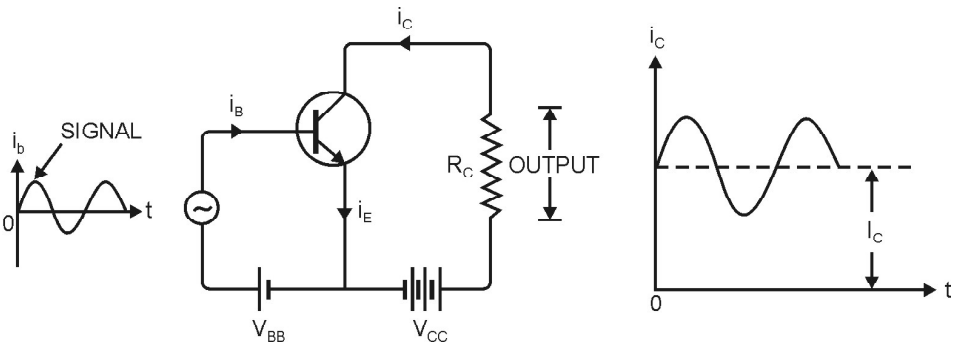


Fig. 1.1

Analysis of collector currents. When no signal is applied, the input circuit is forward biased by the battery V_{BB}. Therefore, a d.c. collector current I_C flows in the collector circuit. This is called zero signal collector current. When the signal voltage is applied, the forward bias on the emitter-bias junction icreases or decreases depending upon whether the signal is positive or negative. During the positive half-cycle of the signal, the

forward bias on emitter-base junction is increased, causing total collector current i_C to increase. Reverse will happen for the negative half-cycle of the signal.

Fig. 8.34 shows the graph of total collector current i_C versus time. From the graph, it is clear that total collector current consists of two components, namely ;

i. The d.c. collector current I_C (zero signal collector current) due to bias battery V_{BB}. This is the current that flows in the collector in the absence of signal.

ii. the a.c. collector current i_C due to signal.

$\therefore$ Total collector current, $i_C = i_C + I_C$

The useful output is the voltage drop across collector load R_C due to the a.c. component i_C. The purpose of zero signal collector current is to ensure that the emitter-base junction is forward biased at all times. The table below gives the symbols usually employed for currents and voltages in transistor applications.

S.No.	Particular	Instantaneous a.c.	d.c.	Total
1.	Emitter current	i_e	I_E	i_E
2.	Collector current	i_c	I_C	i_C
3.	Base current	i_b	I_B	i_B
4.	Collector-emitter voltage	v_{ce}	V_{CE}	v_{CE}
5.	Emitter-base voltage	v_{eb}	V_{EB}	v_{EB}

Transistor Load Line Analysis

In the transistor circuit analysis, it is generally required to determine the collector current for various collector-emitter voltages. One of the methods can be used to plot the output characteristics and determine the collector current at any desired collector-emitter voltage. However, a more convenient method, known as load line method can be used to solve such problems. As explained later in this section, this method is quite easy and is frequently used in the analysis of transistor applications.

D.C. load line. Consider a common emitter npn transistor circuit shown in Fig. 8.35 (i) when no signal is applied. Therefore, d.c. conditions prevail in the circuit. The output characteristics to this circuit are shown in Fig. 8.35 (ii).

The vlaue of collector-emitter voltage V_{CE} at any time is given by ;

$$V_{CE} = V_{CC} - I_C R_C$$

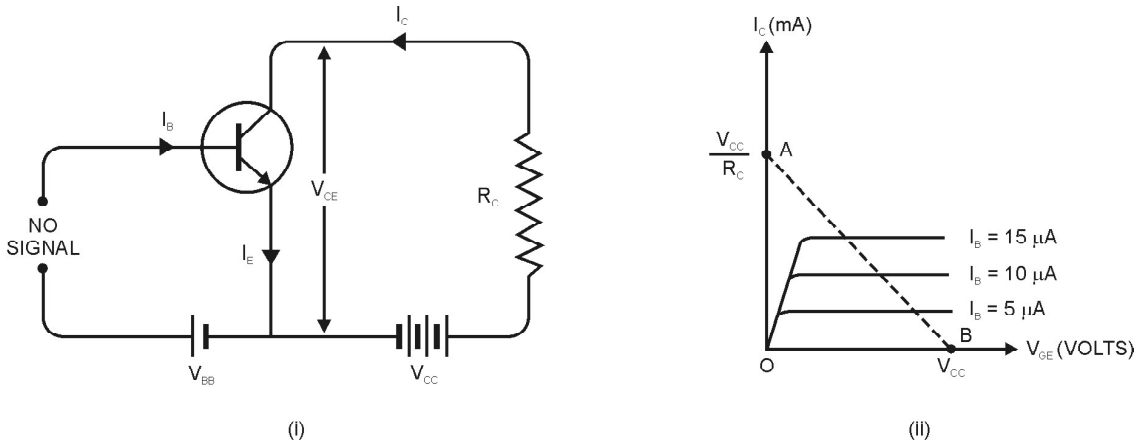


Fig. 8.35

As V_{CC} and R_C are fixed values, therefore, it is a first degree equation and can be represented by a straight line on the output characteristics. This is known as d.c. load line and determines the locus of $V_{CE} - I_C$ points for any given value of R_C. To add load line, we need two end points of the straight line. These two points can be loacted as under :

i. When the collector current $I_C = 0$, then collector-emitter voltage is maximum and is equal to V_{CC} i.e.

$$\text{Max. } V_{CE} = V_{CC} - I_C R_C$$
$$= V_{CC} \qquad (\because I_C = 0)$$

This gives the first point B (OB = V_{CC}) on the collector-emitter voltage axis as shown in Fig. 8.35 (ii).

ii. When collector-emitter voltage V_{CE} – 0, the collector current is maximum and is equal to V_{CC}/R_C i.e.

$$V_{CE} = V_{CC} - I_C R_C$$

or $$0 = V_{CC} - I_C R_C$$

$\therefore$ $$\text{Max. } I_C = V_{CC}/R_C$$

This gives the second point A (OA = V_{CC}/R_C) on the collector current axis as shown in Fig. 8.35 (ii). By joining these two points, d.c. load line AB is constructed.

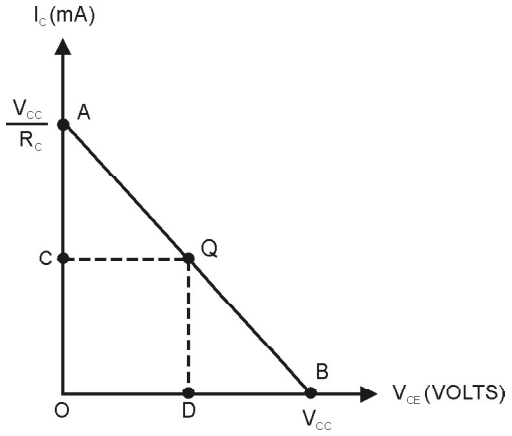


Fig. 8.36

Importance. The current (I_C) and voltage (V_{CE}) conditions in the transistor circuit are represented by some point on the output characteristics. The same information can be obtained from the load line. Thus when I_C is maximum (= V_{CC}/R_C), then V_{CE} = 0 as shown in Fig. 8.36. If I_C = 0, then V_{CE} is maximum and is equal to V_{CC}. For any other value of collector current say OC, the collector-emitter voltage V_{CE} = OD. It follows, therefore, that load line gives a far more convenient and direct solution to the problem.

Operating Point

The zero signal values of I_C and V_{CE} are known as the **operating point.**

It is called operating point because the variations of I_C and V_{CE} take place about this point when signal is applied. It is also called quiescent (silent) point or Q-point because it is the point on $I_C - V_{CE}$ characteristic when the transistor is silent i.e. in the absence of the signal.

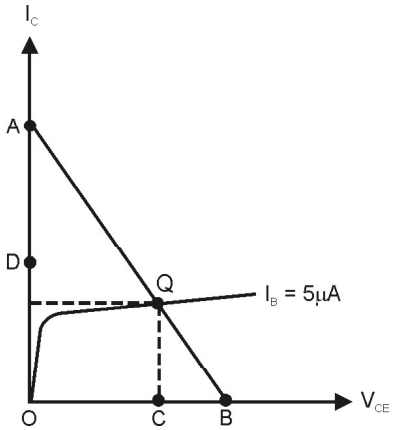


Fig. 8.37

Suppose in the absence of signal, the base current is 5 μA. Then I_C and V_{CE} conditions in the circuit must be represented by some point on I_B = 5 μA characteristic. But I_C and V_{CE} conditions in the circuit should also be represented by some point on the d.c. load line AB. The point Q where the load line and the characteristic intersect is the only point which satisfies both these conditions. Therefore, the point Q describes the actual state of affairs in the circuit in the zero signal conditions and is called the operating point. Referring to Fig. 8.37, for I_B = 5 μA, the zero signal values are :

$$V_{CE} = OC \text{ volts}$$
$$I_C = OD \text{ mA}$$

It follows, therefore, that the zero signal values of I_C and V_{CE} (i.e. operating point) are determined by the point where d.c. load line intersects the proper base current curve.

Example 8.22. For the circuit shown in Fig. 8.38 (i), draw the d.c. load line.

Solution. The collector-emitter voltage VCE is given by;

$$V_{CE} = V_{CC} - I_C R_C$$

When $I_C = 0$, then,

$$V_{CE} = V_{CC} = 12.5 \text{ V}$$

This locates the point B of the load line on the collector-emitter voltage axis.

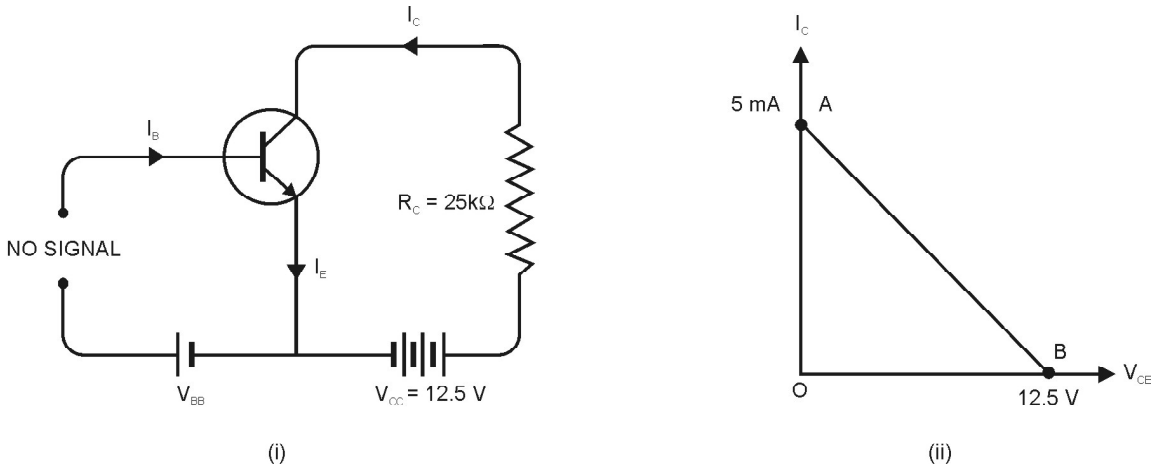


Fig. 8.38

When $V_{CE} = 0$, then,

$$I_C = V_{CC}/R_C = 12.5 \text{ V}/2.5 \text{ k}\Omega = \mathbf{5\ mA}$$

This locates the point A of the load line on the collector current axis. By joining these two points, we get the d.c. load line AB as shown in Fig. 8.38 (ii).

Example 8.23. In the circuit diagram shown in Fig. 8.39 (i), If V_{CC} = 12V and R_C = 6 kΩ, draw the d.c. load line. What will be the Q point if zero signal base current is 20 μA and B = 50?

Solution. The collector-emitter voltage V_{CE} is given by :

$$V_{CE} = V_{CC} - I_C R_C$$

When $I_C = 0$, $V_{CE} = V_{CC} = 12$V. This locates the point B of the load line. When $V_{CE} = 0$, $I_C = V_{CC}/R_C = 12\text{V}/6\text{k}\Omega = 2$ mA. This locates the point A of the load line. By joining these two points, load line AB is constructed as shown in Fig. 8.39 (ii).

Zero signal base current, $I_B = 20\ \mu\text{A} = 0.02$ mA

Current amplification factor, $\beta = 50$

$\therefore$ Zero signal collector current, $I_C = \beta I_B = 50 \times 0.02 = 1$mA

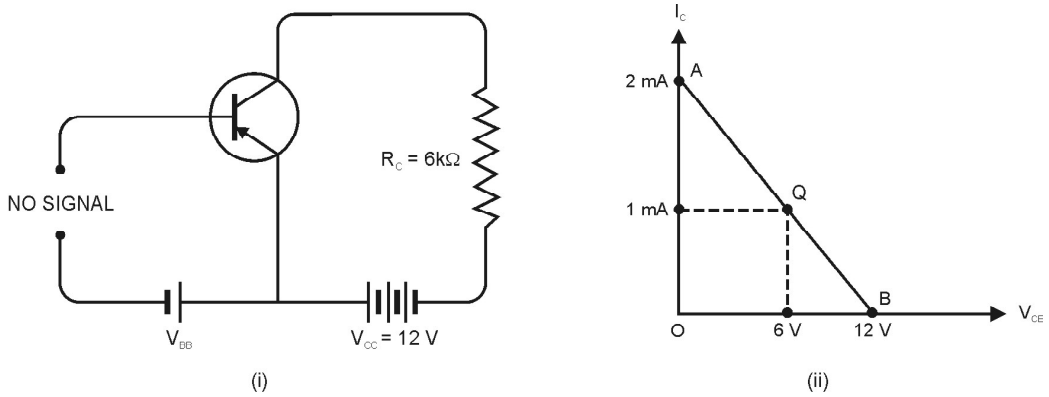


Fig. 8.39

Zero signal collector-emitter voltage is

$$V_{CE} = V_{CC} - I_C R_C = 12 - 1 \text{ mA} \times 6k\Omega = 6V$$

$\therefore$ Operating point is 6V, 1 mA.

Fig. 8.39 (ii) shows the Q point. Its co-ordinates are I_C = 1 mA and V_{CE} = 6 V.

Example 8.24. In a transistor circuit, collector load is 4 kΩ whereas quiescent current (zero signal collector current) is 1mA.

i. What is the operating point if V_{CC} = 10V?

ii. What will be the operating point if R_C = 5 kΩ?

Solution. V_{CC} = 10V, I_C = 1 mA

i. When collector load R_C = 4 kΩ, then,

$$V_{CE} = V_{CC} - I_C R_C = 10 - 1 \text{ mA} \times 4 \text{ k}\Omega = 10 - 4 = 6V$$

$\therefore$ Operating point is **6 V, 1 mA**.

ii. When collector load R_C = 5 kΩ, then,

$$V_{CE} = V_{CC} - I_C R_C = 10 - 1\text{mA} \times 5 \text{ k}\Omega = 10 - 5 = 5V$$

$\therefore$ Operating point is **5 V, 1 mA.**

Example 8.25. Determine the Q point of the transistor circuit shown in Fig. 8.40. Also draw the d.c. load line. Given β = 200 and V_{BE} = 0.7V.

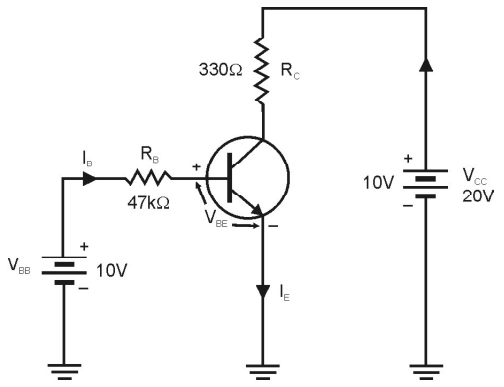


Fig. 8.40

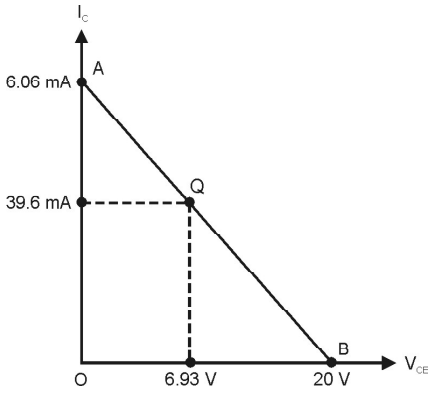


Fig. 8.41

Solution. The presence of resistor R_B in the base circuit should not disturb you because we can apply Kirchhoff's voltage law to find the value of I_B and hence I_C (= βI_B). Referring to Fig. 8.40 and applying Kirchhoff's voltage law to base-emitter loop, we have.

$$V_{BB} - I_B R_B - V_{BE} = 0$$

$$\therefore \quad I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{10V - 0.7V}{47k\Omega} 198\,\mu A$$

Now $I_C = \beta I_B$ = (200) (198 μA) = 39.6 mA

Also $V_{CE} = V_{CC} - I_C R_C$ = 20V – (39.6 mA) (330 Ω) = 20v – 13.07V = 6.93V

Therefore, the Q-point is I_C = **39.6 mA** and V_{CE} = **6.93V.**

D.C. Load Line. In order to draw the d.c. load line, we need two end points.

$$V_{CE} = V_{CC} - I_C R_C$$

When I_C = 0, $V_{CE} = V_{CC}$ = 20V. This locates the point B of the load line on the collector-emitter voltage axis as shown in Fig. 8.41. When V_{CE} = 0, $I_C = V_{CC}/R_C$ = 20V/330Ω = 60.6 mA. This locates the point A of the load line on the collector current exis. By joining these two points, d.c. load line AB is constructed as shown in Fig. 8.41.

Example 8.26. Determine the Q point of the transistor circuit shown in Fig. 8.42. Also draw the d.c. load line. Given β = 100 and V_{BE} = 0.7V.

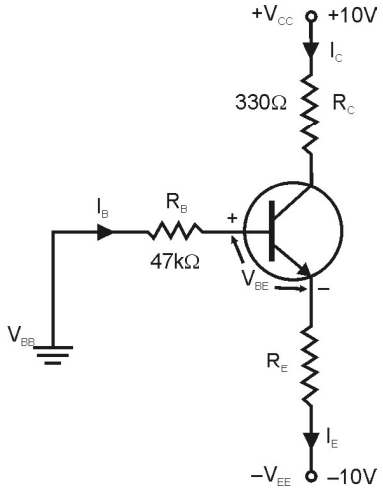


Fig. 8.42

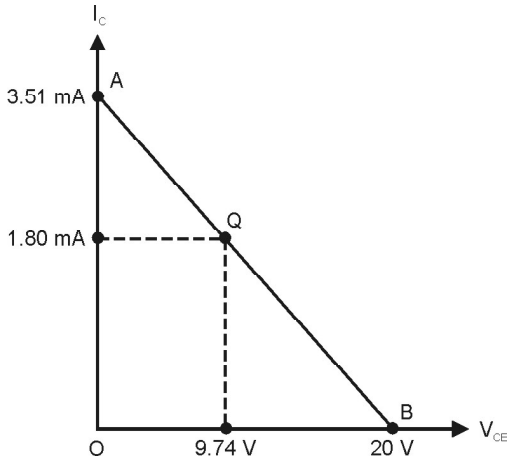


Fig. 8.43

CUT OFF AND SATURATION POINTS

Fig. 8.49 (i) shows CE transistor circuit while Fig. 8.49 (ii) shows the output characteristics along with the d.c. load line.

i. **Cut Off.** The point where the load line intersect the $I_B = 0$ curve is known as cut off. At this point, $I_B = 0$ and only small collector current (i.e. collector leakage current I_{CEO}) exists. At cut off, the base-emitter junction no longer remains forward biased and normal transistor action is lost. The collector-emitter voltage is nearly equal to V_{CC}. i.e.

$$V_{CE}(\text{cut off}) = V_{CC}$$

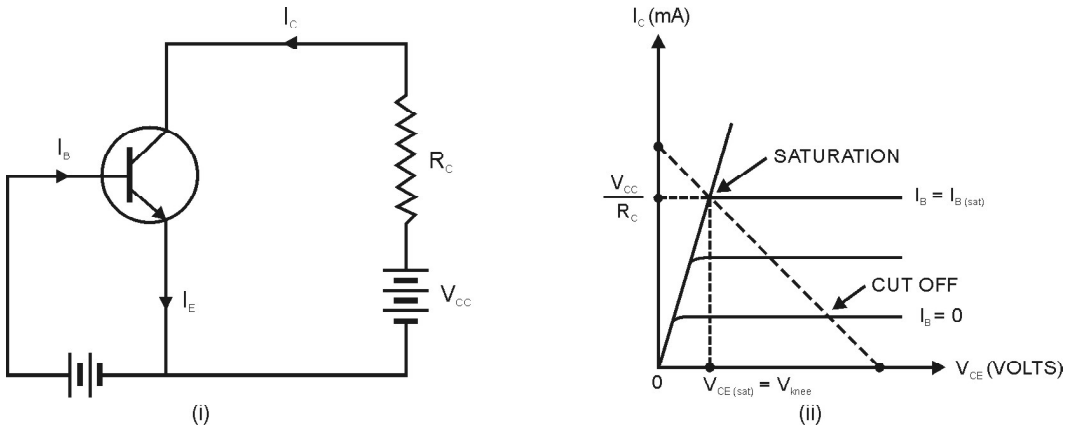


Fig. 8.49

ii. **Saturation.** The point where the load line intersect the $I_B = I_{B(sat)}$ curve is called saturation. At this point, the base current is maximum and so is the collector current. At saturation, collector-base junction no longer remains reverse biased and normal transistor action is lost.

$$I_{C(sat)} = \frac{V_{CC}}{R_C};\ V_{CE} = V_{CE(sat)} = V_{knee}$$

If base current is greater than $I_{B(sat)}$, then collector current cannot increase because collector-base junction is no longer reverse-biased.

iii. **Active reagion.** The region between cut off and saturation is known as active region. In the active region, collector-base junction remains reverse biased while base-emitter junction remains forward biased. Consequetnly, the transistor will function normally in this region.

Note. We provide biasing to the transistor to ensure that it operates in the active region. The reader may find the detailed discussion on transistor biasing in the next chapter.

Summary. A transistor has two pn junctions i.e., it is like two diodes. The junction between base and emitter may be called emitter diode. The junction between base and collector may be called collector diode.

We have seen above that transistor can act in one of the three states : **cut-off, saturated and actived.** The state of a transistor is entirely determined by the states of the emitter diode and collector diode [See Fig. 8.50]. The relations between the diode states and the transistor states are :

CUT-OFF : Emitter diode and collector diode are **OFF.**
ACTIVE : Emitter diode is **ON** and collector diode is **OFF.**
SATURATED : Emitter diode and collector diode are **ON.**

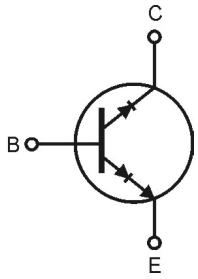


Fig. 8.50

In the **active state,** collector current [See fig. 8.51(i)] is β times times the base current (i.e. $I_C = \beta I_B$). If the transistor is **cut-off,** there is no base current, so there is no collector or emitter current. That is collector emitter pathway is open [See Fig. 8.51 (ii)]. In **saturation,** the collector and emitter are, in effect, shorted together. That is the transistor behaves as though a switch has been closed between the collector and emitter [See Fig. 8.51 (iii)].

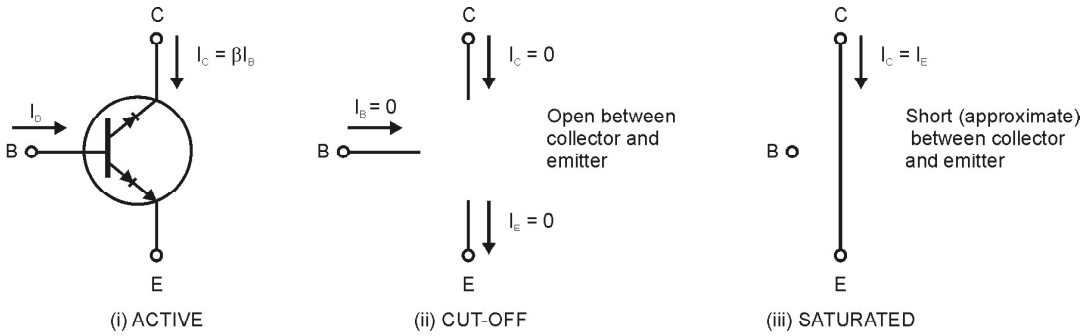


Fig. 8.51

Note. When the transistor is in the active state, $I_C = \beta I_B$. Therefore, a transistor acts as an amplifier when operating in the active state. Amplification means linear amplification. In fact, small signal amplifiers are the most common linear devices.

Example 8.31. Find $I_{C(sat)}$ and $V_{CE(cut\ off)}$ for the circuit shown in Fig. 8.52 (i).

Solution. As we decreases R_B, base current and hence collector current increases. The increased collector current causes a greater voltage drop across RC; this decreases the collector-emitter voltage. Eventually at some value of R_B. VC_E decreases to V_{knee}. At this point, collector-base junction is no longer reverse biased and transistor action is lost. Consequently, further increase in collector current is not possible. The transistor conducts maximum collector current; we say the transistor is saturated.

$$I_{C(sat)} = \frac{V_{CC} - V_{knee}}{R_C} = \frac{V_{CC}}{R_C} = \frac{20\,V}{1k\Omega} = \mathbf{20\,mA}$$

As we increase R_B, base current and hence collector current decreases. This decreases the age drop across R_C. This increases the collector-emitter voltage. Eventually, when $I_B = 0$, the emitter-base junction is no longer forward biased and transistor action is lost. Consequently, further increase in V_{CE} is not possible. In fact, V_{CE} now equals to V_{CC}.

$$V_{CE(cut\text{-}off)} = V_{CC} = 20V$$

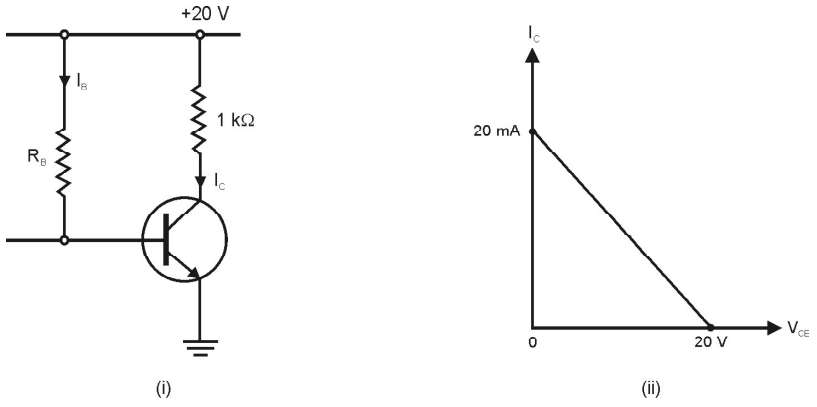


Fig. 8.52

Fig. 8.52 (ii) showsthe saturation and cut off points. Incidentally, they are end points of the d.c. load line.

Note. The exact value of $V_{CE(cut\text{-}off)} = V_{CC} - I_{CEO} R_C$. Since the collector leakage current I_{CEO} is very small, we can neglect $I_{CEO} R_C$ as compared to V_{CC}.

Example 8.32. determine the values of $V_{CE(off)}$ and $I_{C(sat)}$ for the circuit shown in Fig. 8.53.

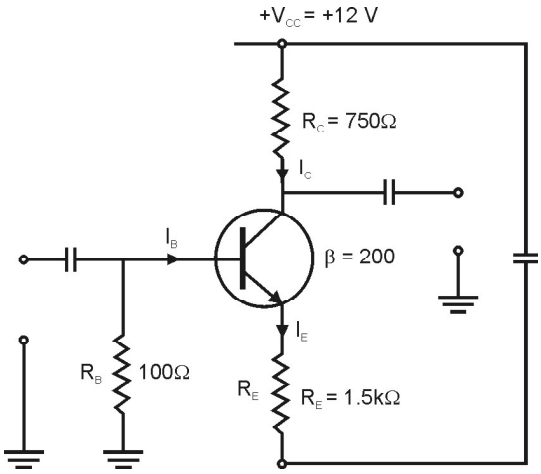


Fig. 8.53

Solution. Applying Kirchhoff's voltage law to the collector side of the circuit in Fig. 8.53, we have,

$$V_{CC} - I_C R_C - V_{CE} - I_C R_E + V_{EE} = 0$$

or $$V_{CE} = V_{CC} + V_{EE} - I_C (R_C + R_E)$$

We have $V_{CE(off)}$ when $I_C = 0$. Therefore, putting $I_C = 0$ in eq. (i), we have,

$$V_{CE(off)} = V_{CC} + V_{EE} = 12 + 12 = \mathbf{24V}$$

We have $I_{C(sat)}$ when $V_{CE} = 0$.

$$I_{C(sat)} = \frac{V_{CC} + V_{EE}}{R_C + R_E} = \frac{(12+12)\,V}{(750+1500)\,\Omega} = \mathbf{10.67\,mA}$$

Example 8.33. Determine whether or not the transistor in Fig. 8.54 is in stauration. Assume $V_{knee} = 0.2V$.

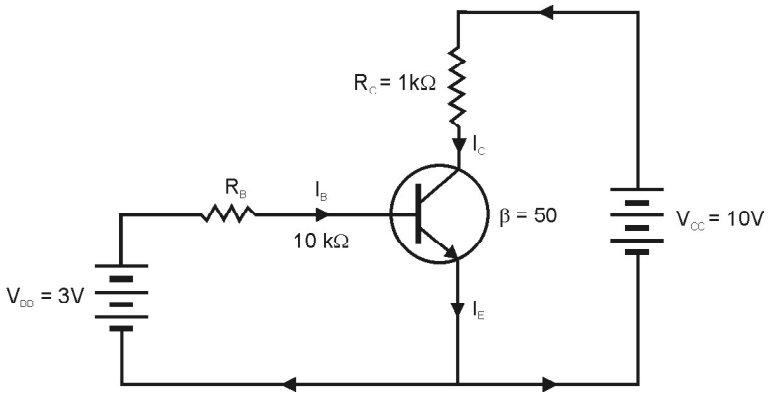


Fig. 8.54

Solution.

$$I_{C(sat)} = \frac{V_{CC} - V_{knee}}{R_C} = \frac{10\,V - 0.2\,V}{1k\Omega} = \frac{9.8\,V}{1k\Omega} = 9.8\,mA$$

Now we shall see if I_B is large enough to produce $I_{C(sat)}$.

Now $$I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{3V - 0.7\,V}{10\,k\Omega} = \frac{2.3\,V}{10\,k\Omega} = 0.23\,mA$$

$I_C = \beta I_B = 50 \times 0.23 = 11.5$ mA

This shows that with specified β, this base current (= 0.23 mA) is capable of producing I_C greater than $I_{C(sat)}$. Therefore, the transistor is **saturated.** In fact, the collector current value of 11.5 mA is never reached. If the base current value corresponding to $I_{C(sat)}$ is increased, the collector current remains at the saturated value (= 9.8 mA).

Example 8.34. Is the transistor in Fig. 8.55 operating in saturated state?

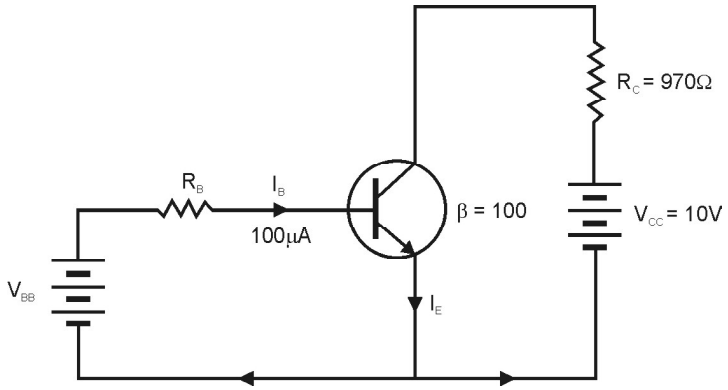


Fig. 8.55

Solution.

$$I_C = \beta I_B = (100)\,(100\ \mu A) = 10\ mA$$
$$V_{CE} = V_{CC} - I_C R_C$$
$$= 10V - (10\ mA)\,(970\Omega) = 0.3V$$

Let us relate the values found to the transistor shown in Fig. 8.56. As you can see, the value of V_{BE} is 0.95V and the value of V_{CE} = 0.3V. This leaves V_{CB} of 0.65V (Note that $V_{CE} = V_{CB} + V_{BE}$). In this case, collector-base junction (i.e. collector diode) is forward biased as is the emitter-base junction (i.e. emitter diode). Therefore, the transistor is operating in the **saturation region.**

Note. When the transistor is in the saturated state, the base current and collector current are independent of each other. The base current is still (and always is) found only from the base circuit. The collector current is found approximately by closing the imaginary switch between the collector and the emitter in the collector circuit.

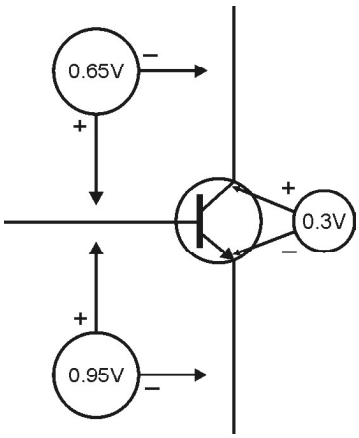


Fig. 8.56

■■■

CHAPTER - 5
FIELD EFFECT TRANSISTOR

INTRODUCTION

In the previous chapters, we have discussed the circuit applications of an ordinary transistor. In this type of transistor, both holes and electrons play part in the conduction process. For this reason, it is sometimes called a bipolar transistor. The ordinary or bipolar transistor has two principal disadvantages. First, it has a low input impedance because of forward biased emitter junction. Secondly, it has considerable noise level. Although low input impedance problem may be improved by careful design and use of more than one transistor, yet it is difficult to achieve input impedance more than a few megaohms. The field effect transistor (FET) has, by virtur of its construction and biasing, larger input impedance which may be more than 100 megaohms. The FET is generally much less noisy than the ordinary or bipolar transistor. The rapidly expanding FET market has led many semiconductor marketing managers to believe that this device will soon become the most important electronic device, primarily because of its integrated-circuit applications. In this chapter, we shall focus our attention on the construction, working and circuit applications of field effect transistors.

Types of Field Effect Transistor

A bipolar junction transistor (BJT) is a current controlled device i.e., output characteristics of the device are controlled by base current and not by base voltage. However, in a field effect transistor (FET), the output characteristics are controlled by input voltage (i.e. electric field) and not by input current. This is probably the biggest difference between BJT and FET. There are two basic types of field effect transistors:

i. Junction field effect transistor (JFET)
ii. Metal oxide semiconductor field effect transistor (MOSFET)

To begin with, we shall study about JFET and then improved form of JFET, namely; MOSFET.

JUNCTION FIELD EFFECT TRANSISTOR (JFET)

Junction field effect transistor is a three terminal semiconductor device in which current conduction is by one type of carrier i.e., electrons or holes.

The JFET was developed about the same time as the transistor but it came into general use only in the late 1960s. In a JFET, the current conduction is either by electrons or holes and is controlled by means of an electric field between the gate electrode and the conducting channel of the device. The JFET has high input impedance and low noise level.

Constructional Details

A JFET consists of a p-type or n-type silicon bar containing two pn junctions at the sides as shown in Fig. 19.1. The bar forms the conducting channel for the charge carriers. If the bar is of n-type, it is called n-channel JFET as shown in Fig. 19.1 (i) and if the bar is of p-type, it is called a p-channel JFET as shown in Fig. 19.1 (ii). The two pn junctions forming diodes are connected internally and a common terminal called gate is taken out. Other terminals are source and drain taken out from the bar as shown. Thus a JFET has essentially three terminals viz., gate (G), source (S) and drain (D).

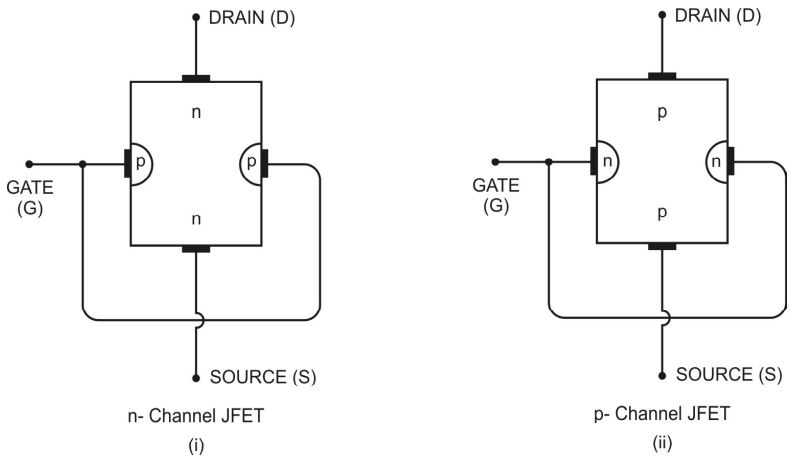


Fig. 5.1

JFET Polarities

Fig. 19.2 (i) shows n-channel JFET polarities whereas Fig. 19.2 (ii) shows the p-channel JFET polarities. Note that in each case, the voltage between the 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 i.e., either end can be used as source and the other end as drain.

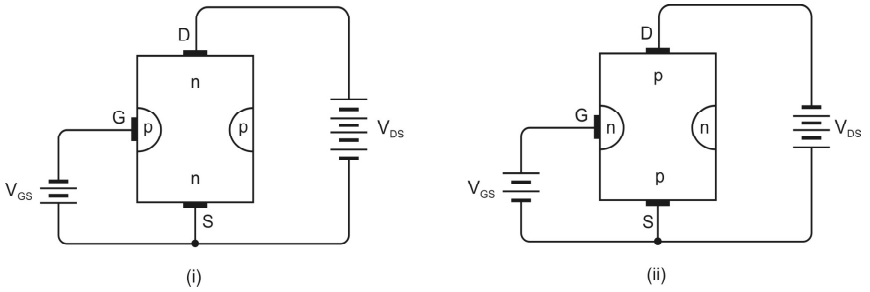


Fig. 5.2

The following points may be noted :

i. The input circuit (i.e. gate to source) of a JFET is reverse biased. This means that the device has high input impedance.

ii. The drain is so biased w.r.t. source that drain current I_D flows from the source to drain.

iii. In all JFETs, source current I_S is equal to the drain current i.e. $I_S = I_D$.

Principle and Working Of JFET

Fig. 19.3 shows the circuit of n-channel JFET with normal polarities. Note that the gate is reverse biased.

Principle

The two pn junctions at the sides form two depletion layers. The current conduction by charge carriers (i.e. free electrons in this case) is through th channel between the two depletion layers and out of the drain. The width and hence resistance of this channel can be controlled by changing the input voltage V_{GS}. The greater the reverse voltage V_{GS}, the wider will be the depletion layers and narrower will be the conducting channel. The narrower channel means greater resistance and hence source to drain current decreases. Reverse will happen should V_{GS} decrease. Thus JFET operates on the principle that width and hence resistance of the conducting channel can be varied by changing the reverse voltage V_{GS}. In other words, the magnitude of drain current (ID) can be changed by altering V_{GS}.

Working

The working of JFET is as under :

i. When a voltage VDS is applied between drain and source terminals and voltage on the gate is zero (See Fig. 19.3 i), the two pn junctions at the sides of the bar establish depletion layers. The electrons will flow from source to drain through a channel between the depletion layers. Teh size of these layers determines the width of the channel and hence the current conduction through the bar.

ii. When a reverse voltage VGS is applied between the gate and source [See Fig. 19.3 ii], 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 source to drain is decreased. On the other hand, if the reverse voltage on the gate is decreased the width of the deplietion layers also decreases. This incerases the width of the conducting channel and hence source to drain current.

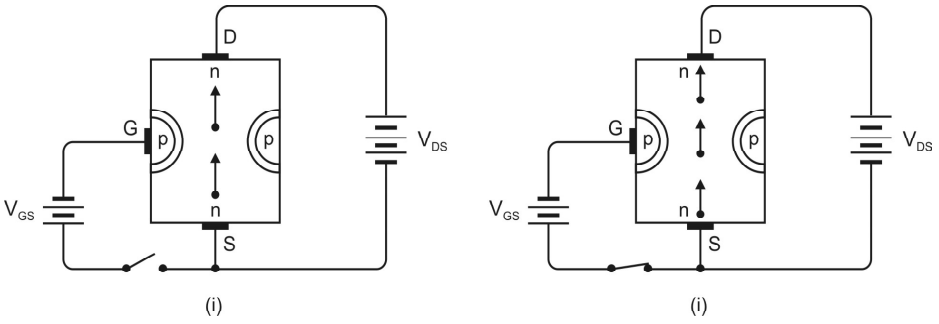


Fig. 5.3

It is clear from the above discussion that current from source to drain can be controlled by the application of potential (i.e. electric field) on the gate. For this reason, the device is called field effect transistor. It may be noted 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.

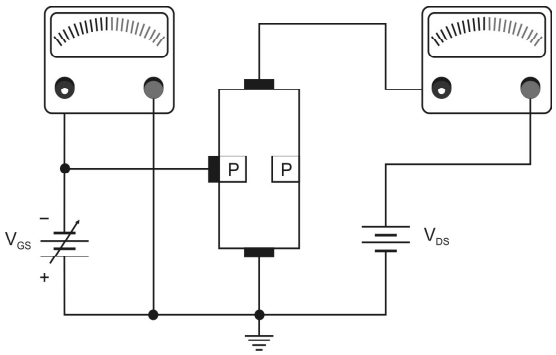


Fig. 5.4

Note. If the reverse voltage VGS on the gate is continuously increased, a state is reached when the two depletion layers touch each other and the channel is cut off. Under such conditions, the channel becomes a non-conductor.

Schematic Symbol of JFET

Fig. 19.4 shows the schematic symbol of JFET. The vertical line in the symbol may be thought as channel and source (S) and drain (D) connected to this line. If the channel is n-type, the arrow on the gate points towards the channel as shown in Fig. 19.4 (i). However, for p-type channel, the arrow on the gate points from channel to gate [See Fig. 19.4 (ii)].

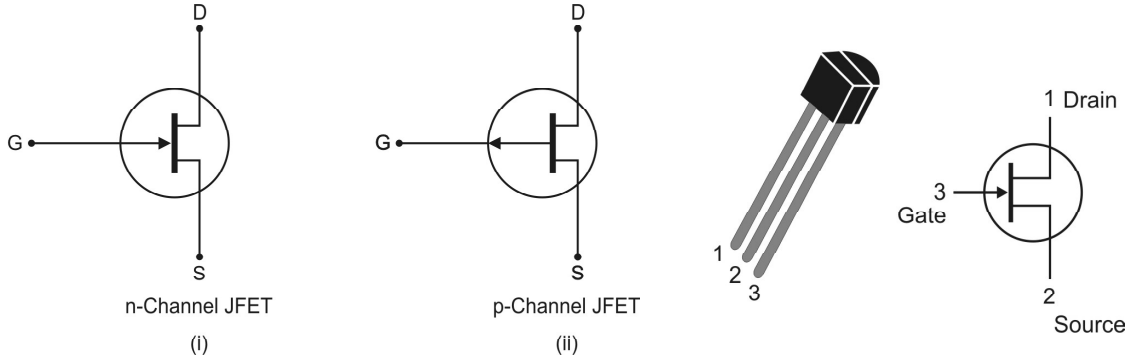


Fig. 5.5

Importance of JFET

A JFET acts like a voltage controlled device i.e. input voltage (V_{GS}) controls the output current. This is different from ordinary transistor (or bipolar transistor) where input current controls the output current. Thus JFET is a semiconductor device acting like a vacuum tube. The need for JFET arose because as modern electronic equipment became increasingly were transistorised, it became apparent that there were many functions in which bipolar transistor were unable to replace vacuum tubes. Owing to their extremely high input impedance, JFET devices are more like vacuum tubes than are the bipolar transistors and hence are able to take over many vacuum-tube functions. Thus, because of JFET, electronic equipment is closer today to being completely solid state.

The JFET devices have not only taken over the functions of vacuum tubes but they now also threaten to depose the bipolar transistors as the most widely used semiconductor devices. As an amplifier, the JFET has higher input impedance than that of a conventional transistor, generates less noise and has greater resistance to nuclear radiations.

Difference Between JFET and Bipolar Transistor

The JFET differs from an ordinary or bipolar transistor in the following ways:

i. In a JFET, there is only one type of carrier, holes in p-type channel and electronis in n-type channel. For this reason, it is also called a unipolar transistor. However, in an ordinary transistor, both holes and

electrons play part in conduction. Therefore, an ordinary transistor is cometimes called a bipolar transistor.

ii. As the input circuit (i.e., gate to source) of a JFET is reverse biased, therefore, the device has high input impedance. However, the input circuit of an ordinary transistor is forward biased and hence has low input impedance.

iii. The primary functional different between the JFET and the BJT is that no current (actually, a very, very small current) enters the gate of JFET (i.e. I_G = 0A). However, typical BJT base current might be a few μA while JFET gate current a thousand times smaller (See Fig. 19.5].

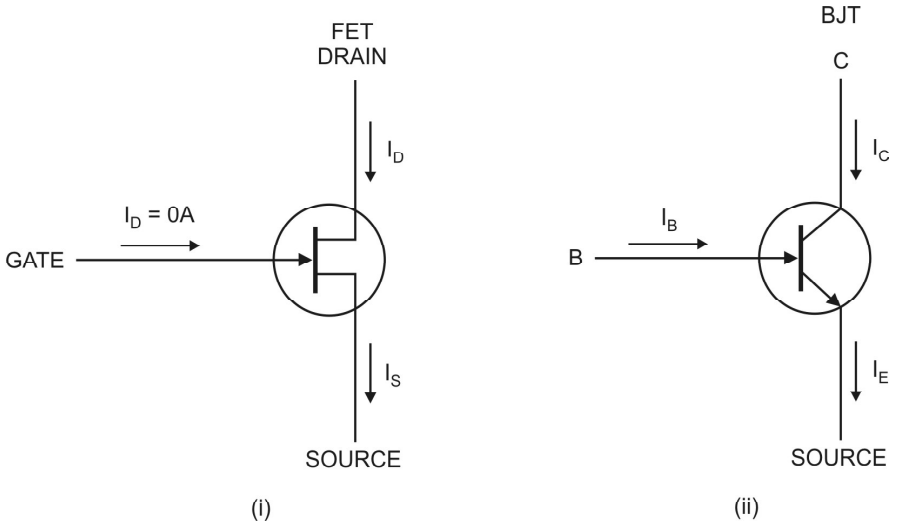


Fig. 5.6

iv. A bipolar transistor uses a current into its base to control a large current between collector and emitter whereas a JFET uses voltage on the 'gate' (= base) terminal to control the current between drain (= collector) and source (= emitter). Thus a bipolar transistor gain is characterised by current gain whereas the JFET gain is characterised as a transconductance i.e., the ratio of change in output current (drain current) to the input (gate) voltage.

v. In JFET, there are no junctions as in an ordinary transistor. The conduction is through an n-type or p-type semi-conductor material. For this reason, noise level in JFET is very small.

JFET as an Amplifier

Fig. 19.6 shows JFET amplifier circuit. The weak signal is applied between gate and source and amplified output is obtained in the drain-source circuit. For the proper operation of JFET, the gate must be negative w.r.t. source i.e., input circuit should always be reverse biased. This is achieved either by inserting a battery V_{GG} in the gate circuit or by a circuit known as biasing circuit. In the present case, we are providing biasing by the battery V_{GG}.

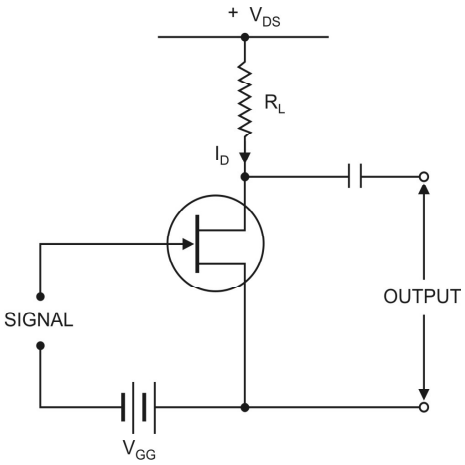


Fig. 5.7

A small change in the reverse bias on the gate produces a large change in drain current. This fact makes JFET capable of raising the strength of a weak signal. During the positive half of signal, the reverse bias on the gate decreases. This increases the channel width and hence the drain current. During the negative half-cycle of the signal, the reverse voltage on the gate increases. Consequently, the drain current decreases. The result is that a small change in voltage at the gate produces a large change in drain current. These large variations in drain current produce large output across the load R_L. In this way, JFET acts as an amplifier.

Output Characteristics of JFET

The curve between drain current (I_D) and drain-source voltage (V_{DS}) of a JFET at constant gate-source voltage ($V_{GS)}$ is known as output characteristics of JFET. Fig. 19.7 shows the circuit for determining the output characteristics of JFET. Keeping V_{GS} fixed at some value, say 1V, the drain-source voltage is changed in steps. Corresponding to each value of V_{DS}, the drain current I_D is noted. A plot of these values gives the output characteristic of JFET at V_{GS} = 1V. Repeating similar procedure, output characteristics at other gate-source voltages can be drawn. Fig. 19.8 shows a family of output characteristics.

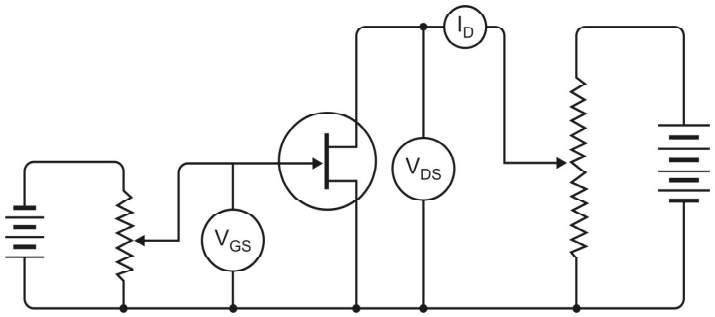


Fig. 5.8

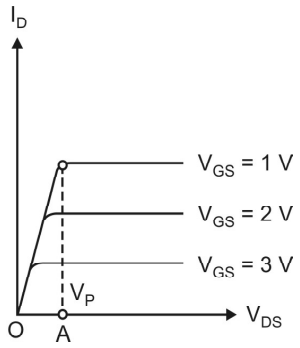


Fig. 5.9

The following points may be noted from the characteristics:

i. At first, the drain current ID rises rapidly with drain-source voltage V_{DS} but then becomes constant. The drain-source voltage above which drain current becomes constant is known as pinch off voltage. Thus in Fig. 19.8 OA is the pinch off voltage V_P.

ii. After pinch off voltage, the channel width become so narrow that depletion layers almost touch each other. The drain current passes through the small passage between these layers. Therefore, increase in drain current isvery small with V_{DS} above pinch off voltage. Consequently, drain current remains constant.

iii. The characteristics resemble that of a pentode valve.

Salient Features of JFET

The following are some salient features of JFET.

i. A JFET is a three-terminal voltage-controlled semiconductor device i.e. input voltage controls the output characteristics of JFET.

ii. The JFET is always operated with gate-source pn junction reverse biased.

iii. In a JFET, the gate current is zero i.e. I_G = 0A.

iv. Since there is no gate current $I_D = I_S$.

v. The JFET must be operated between V_{GS} = 0V and $V_{GS(off)}$. For this range of gate-source votages, I_D will vary from a maximum of I_{DSS} to a minimum of almost zero. [See Fig. 19.12].

vi. Because the two gates are at the same potential, both depletion layers widen or narrow down by an equal amount.

vii. The JFET is not subjected to thermal runway when the temperature of the device increases.

viii. The drain current I_D is controlled by canging the channel width.

ix. Since JFET has no gate current, there is no β rating of the device. We can find drain current I_D by using the eq. mentioned in Art. 19.11.

Important Terms

In the analysis of a JFET circuit, the following important terms are often used:

1. Shorted-gate drain current (I_{DSS})
2. Pinch off voltage (V_P)
3. Gate-source cut off voltage [$V_{GS(off}$]

Shorted-gate Drain Current (I_{DSS})

It is the drain current with source short-circuited to gate (i.e. $V_{GS} = 0$) and drain voltage (V_{DS}) equal to pinch off voltage. It is sometimes called zero-bias current.

Fig. 19.9 shows the JFET circuit with $V_{GS} = 0$ i.e., source shorted-circuited to gate. This is normally called shorted-gate condition. Fig. 19.10 shows the graph between I_D and V_{DS} for the shorted gate condition. The drain current rises rapidly at first and then levels off at pinch off voltage V_P. The drain current has now reached the maximum value I_{DSS}. When V_{DS} is increased beyond V_P, the depletion layers expand at the top of the channel. The channel now acts as a current limiter and holds drain current contant at I_{DSS}.

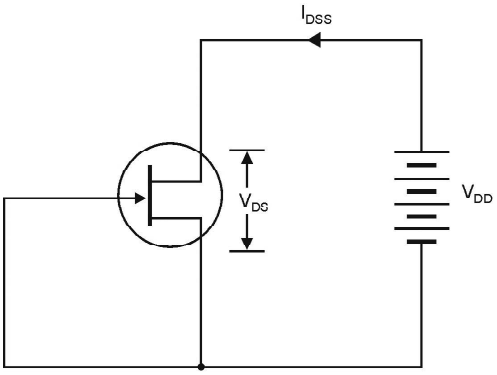


Fig. 5.10

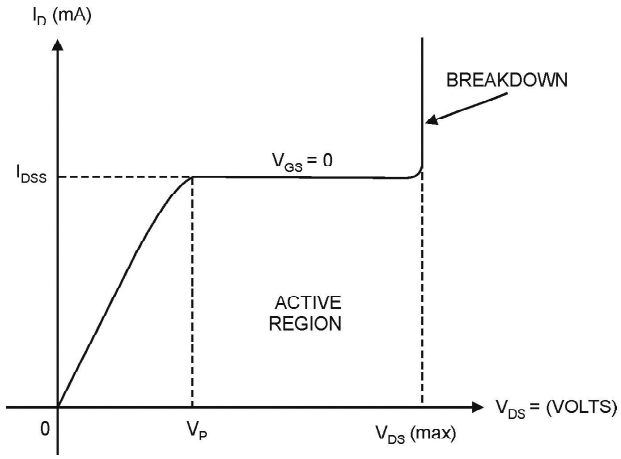


Fig. 5.11

The following points may be noted carefully:

i. Since IDSS is measured under shorted gate conditions, it is the maximum drain current that you can get with normal operation of JFET.

ii. There is a maximum drain voltage [VDS (max)] that can be applied to a JFET. If the drain voltage exceeds VDS (max), JFET would breakdown as shown in Fig. 19.10.

iii. The region between VP and VDS (max) (breakdown voltage) is called constant-current region or active region. As long as VDS is kept within this range, ID will remain constant for a constant value of VGS. In other words, in the active region, JFET behaves as a constant-current device. For proper working of JFET, it must be operated in the active region.

Pinch off Voltage (V_P)

It is the minimum drain-source voltage at which the drain current essentially becomes constant.

Fig. 19.11 shows the drain curves of a JFET. Note that pinch off voltage is V_P. The highest curve is for V_{GS} = 0V, the shorted-gate condition. For values of V_{DS} greater than V_P, the drain current is almost constant. It is because when V_{DS} equals V_P, the channel is effectively closed and does not allow further increase in drain current. It may be noted that for proper function of JFET, it is always operated for $V_{DS} > V_P$. However, V_{DS} should not exceed $V_{DS\,(max)}$ otherwise JFET may breakdown.

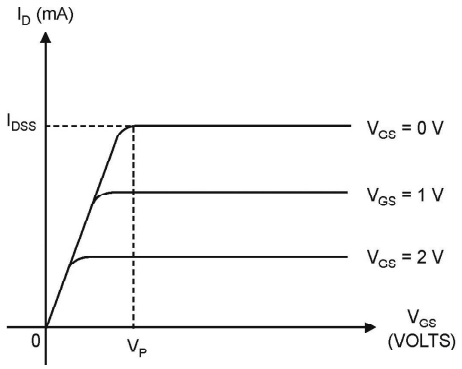


Fig. 5.12

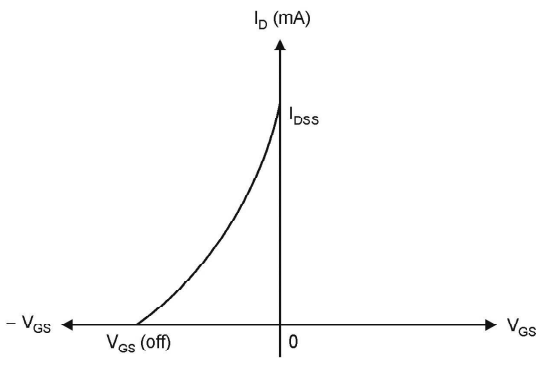


Fig. 5.13

Gate-source cut off Voltage VGS (off)

It is the gate-source voltage where the channel is completely cut off and the drain becomes zero.

The idea of gate-source cut off voltage can be easily understood if we refer to the transfer characteristic of a JFET shown in Fig. 19.12. As the reverse gate-source voltage is increased, the cross-sectional area of the channel decreases. This in turn decreases the drain current. At some reverse gate-source voltage, the depletion layers extend completely across the channel. In this condition, the channel is cut off and the drain current reduces to zero. The gate voltage at which the channel is cut off (i.e. channel becomes non-conducting) is called gate-source cut off voltage $V_{GS\,(off)}$.

Notes i. It is interesting to note that $V_{GS(off)}$ will always have the same magnitude value a V_P. For example if V_P = 6V, then $V_{GS\,(off)}$ = – 6V. Since these two values are always equal and opposite, only one is listed on the specification sheet for a given JFET.

ii. There is a distinct difference between V_P and $V_{GS\,(off)}$. Note that V_P is the value of V_{DS} that causes the JFET to become a constant current device. It is measured at V_{GS} = 0V and will have a constant drain current = I_{DSS}. However, $V_{GS\,(off)}$ is the value of V_{GS} that causes I_D to drop to nearly zero.

Expression for Drain Current (I_D)

The relation between I_{DSS} and V_P is shown in Fig. 19.13. We note that gate-source cut off voltage [i.e. $V_{GS(off)}$] on the transfer characteristic is equal to pinch off voltage V_P on the drain characteristic i.e.

$$V_P = \left|V_{GS(off)}\right|$$

For example, if a JFET has $V_{GS\,(off)}$ = – 4V, then V_P = 4V.

The transfer characteristic of JFET shown in Fig. 19.13 is part of a parabola. A rather complex mathematical analysis yields the following expression for drain current :

$$I_D = I_{DDS}\left[1 - \frac{V_{GS}}{V_{GS}}\right]^2$$

where

ID = drain current at given VGS
IDSS = shorted – gate drain current
VGS = gate–source voltage
VGS(off) = gate–source cut off voltage

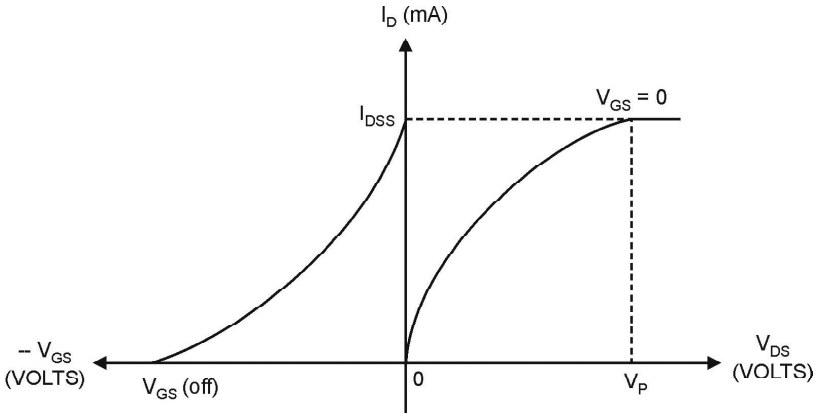


Fig. 5.14

Example 1. Fig. 19.14 shows the transfer characteristic curve of a JFET. Write the equation for drain current.

Solution. Referring to the transfer characteristic curve in Fig. 19.14, we have,

I_{DSS} = 12 mA
$V_{GS(off)}$ = 5V

$$\therefore \quad I_D = I_{DDS}\left[1 - \frac{V_{GS}}{V_{GS(off)}}\right]^2$$

or

$$I_D = 12\left[1 + \frac{V_{GS}}{5}\right]^2 \text{ mA} \quad \textbf{Ans.}$$

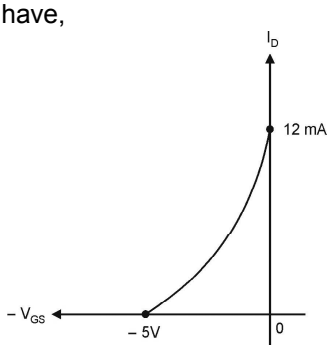


Fig. 5.15

Example 2. A JFET has the following parameters: I_{DSS} = 32 mA; VGS(off) = – 8V ; VGS = – 4.5 V. Find the value of drain current.

Solution.

$$I_D = I_{DSS}\left[1 - \frac{V_{GS}}{V_{GS(off)}}\right]^2$$

$$= 32\left[1 - \frac{(-4.5)}{-8}\right]^2 \text{ mA}$$

= 6.12 mA

Example 3. A JFET has a drain current of 5 mA. If I_{DSS} = 10 mA and $V_{GS(off)}$ = – 6V, find the value of (i) V_{GS} and (ii) V_P.

Solution.

$$I_D = I_{DSS}\left[1 - \frac{V_{GS}}{V_{GS(off)}}\right]^2$$

or

$$5 = 10\left[1 + \frac{V_{GS}}{6}\right]^2$$

or

$$1 + \frac{V_{GS}}{6} = \sqrt{5/10} = 0.707$$

i. ∴ V_{GS} = **1.76 V**

ii. and $V_P = -V_{GS(off)}$ = **6 V**

Example 4. For the JFET in Fig. 19.15, $V_{GS(off)}$ = –4V I_{DSS} = 12 mA. Determine the minimum value of V_{DD} required to put the device in the constant-current region of operation.

Solution. Since $V_{GS(off)}$ = – 4V, V_P = 4V. The minimum value of V_{DS} for the JFET to be in constant-current region is

$$V_{DS} = V_P = 4V$$

In the constant current region with V_{GS} = 0V,

$$I_D = I_{DSS} = 12 \text{ mA}$$

Applying Kirchhoff's voltage law around the drain circuit, we have,

$$V_{DD} = V_{DS} + V_{RD} = V_{DS} + I_D R_D$$

$$= 4V + (12 \text{ mA})\ (560\Omega) = 4V + 6.72V = \mathbf{10.72V}$$

This is the value of V_{DD} to make $V_{DS} = V_P$ and put the device in the constant-current region.

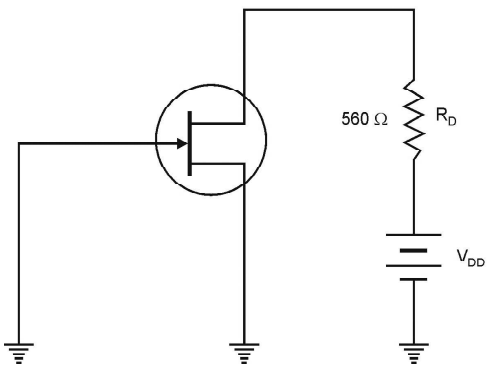


Fig. 5.16

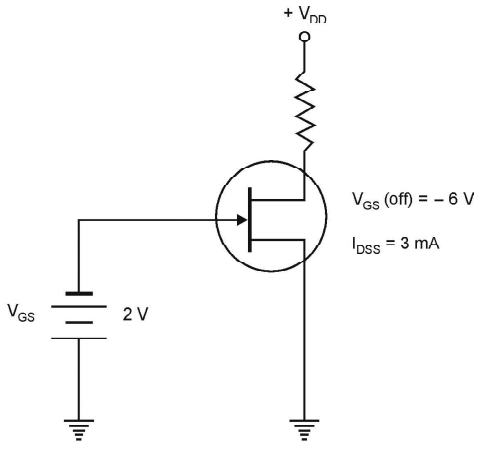


Fig. 5.17

Example 5. Determine the value of drain current for the circuit shown in Fig. 19.16.

Solution . It is clear from Fig. 19.16 that V_{GS} = – 2V. The drain current for the circuit is given by;

$$I_D = I_{DSS}\left(1-\frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$= 3mA\left(1-\frac{-2V}{-6V}\right)^2$$

= (3 mA) (0.444) = 1.33 mA

Example 6. A particular p-channel JFET has a $V_{GS(off)}$ = + 4V. What is I_D when V_{GS} = + 6V?

Solution. The p-channel JFET requires a positive gate-to-source voltage to pass drain current I_D. The more the positive voltage, the less the drain current. When V_{GS} = 4V, I_D = 0 and JFET is cut off. Any further increase in V_{GS} keeps the JFET cut off. Therefore, at V_{GS} = +6V, I_D = **0A.**

Advantages of JFET

A JFET is a voltage controlled, constant current device (similar to a vacuum pentode) in which variations in input voltage control the output current. It combines the many advantages of both bipolar transistor and vacuum pentode. Some of the advantages of JFET are:

i. It has a very high input impedance (of the order of 100 MΩ). This permits high degree of isolation between the input and output circuits.
ii. The operation of a JFET depends upon the bulk material current carriers that do not cross junctions. Therefore, the inherent noise of tubes (due to high-temperature operation) and those of transistors (due to junction transitions) are not present in a JFET.
iii. A JFET has a negative temperature co-efficient of resistance. This avoids the risk of thermal runaway.
iv. A JFET has a very high power gain. This eliminates the necessity of using driver stages.
v. A JFET has a smaller size, longer life and high efficiency.

Parameters of JFET

Like vacuum tubes, a JFET has certain parameters which determine its performance in a circuit. The main parameters of a JFET are (i) a.c. drain resistance (ii) transconductance (iii) amplification factor.

i. **a.c. drain resistance (r_d).** Corresponding to the a.c. plate resistance, we have a.c. drain resistance in a JFET. It may be defined as follows:

It is the ratio of change in drain-source voltage (ΔV_{DS}) to the change in drain current (ΔI_D) at constant gate-source voltage i.e.

$$\text{a.c. drain resistance, } r_d = \frac{\Delta V_{DS}}{\Delta I_D} \text{ at constant } V_{GS}$$

For instance, if a change in drain voltage of 2 V produces a change in drain current of 0.02 mA, then,

$$\text{a.c. drain resistance, } r_d = \frac{2V}{0.02\,mA} = 100\ k\Omega$$

Referring to the output characteristic of a JFET in Fig. 19.8, it is clear that above the pinch off votage, the change in I_D is small for a change in V_{DS} because the curve is almost flat. Therefore, drain resistance of a JFET has a large value, ranging from 10 kΩ to 1MΩ.

ii. **Transconductance (g_{fs}).** The control that the gate voltage has over the drain current is measured by transconductance g_{fs} and is similar to the transconductance gm of the tube. It may be defined as follows:

It is the ratio of change in drain current (ΔI_D) to the change in gate-source voltage (ΔV_{GS}) at constant drain-source voltage i.e.

$$\text{Transconductance, } g_{fs} = \frac{\Delta I_D}{\Delta V_{GS}} \text{ at constant } V_{DS}$$

The transconductance of a JFET is usually expressed either in mA/volt or micromho. As an example, if a change in gate voltage of 0.1 V causes a change in drain current of 0.3 mA, then,

$$\text{Transconductance, } g_{fs} = \frac{0.3\,mA}{0.1V} = 3 \text{ mA/V} = 3 \times 10^{-3} \text{ A/V or mho or S (siemens)}$$

$$= 3 \times 10^{-3} \times 106\ \mu \text{ mho} = 3000\ \mu \text{ mho (or } \mu\text{S)}$$

iii. Amplification factor (μ). It is the ratio of change in drain-source voltage (ΔV_{DS}) to the change in gate-source voltage (ΔV_{GS}) at constant drain current i.e.

$$\text{Amplification factor, } \mu = \frac{\Delta V_{DS}}{\Delta V_{GS}} \text{ at constant ID}$$

Amplification factor of a JFET indicates how much more control the gate voltage has over drain current than has the drain voltage. For instance, if the amplification factor of a JFET is 50, it means that gate voltage is 50 times as effective as the drain voltage in controlling the drain current.

Relation Among JFET Parameters

The relationship among JFET parameters can be established as under:

$$\text{We know} \qquad \mu = \frac{\Delta V_{DS}}{\Delta V_{GS}}$$

Multiplying the numerator and denominator on R.H.S. by ΔI_D, we get,

$$\mu = \frac{\Delta V_{DS}}{\Delta V_{GS}} \times \frac{\Delta I_D}{\Delta I_D} = \frac{\Delta V_{DS}}{\Delta I_D} \times \frac{\Delta I_D}{\Delta V_{GS}}$$

$\therefore \qquad \mu = r_d \times g_{fs}$

i.e. amplification factor = a.c. drain resistance x transconductance

Example 19.7. When a reverse gate voltage of 15V is applied to a JFET, the gate current is 10^{-3} μA. Find the resistance between gate and source.

Solution. $V_{GS} = 15V$; $I_G = 10^{-3}\ \mu A = 10^{-9}$ A

$\therefore$ Gate to source resistance $= \dfrac{V_{GS}}{I_G} = \dfrac{15\,V}{10^{-9}\,A} = 15 \times 10^9\ \Omega = \mathbf{15{,}000\ M\Omega}$

This example shows the major difference between a JFET and a bipolar transistor. Whereas the input impedance of a JFET is several hundred MΩ, the input impedance of a bipolar transistor is only hundreds or thousands of ohms. The large input impedance of a JFET permits high degree of isolation between the input and output.

Example 19.8. When VGS of a JFET changes from –3V, the drain current changes from 1 mA to 1.3 mA. What is the value of transconductance?

Solution. $\Delta V_{GS} = 3.1 - 3 = 0.1$ V ...magnitude

$\Delta I_D = 1.3 - 1 = 0.3$ mA

$\therefore$ Transconductance, gfs $= \dfrac{\Delta I_D}{\Delta V_{GS}} = \dfrac{0.3\,mA}{0.1V} = 3\,mA/V = \mathbf{3000\ \mu mho}$

Example 19.9. The following readings were obtained experimentally from a JFET :

V_{GS}	0 V	0 V	– 0.2 V
V_{DS}	7 V	15 V	15 V
I_D	10 mA	10.25 mA	9.65 mA

Determine (i) a.c. drain resistance (ii) transcondctance and (iii) amplification factor.

Solution. i. With V_{GS} constant at 0V, the increase in V_{DS} from 7V to 15V increases the drain current from 10 mA to 10.25 mA i.e.

Change in drain-source voltage, ΔV_{DS} = 15 – 7 = 8V

Change in drain current, ΔI_D = 10.25 – 10 = 0.25 mA

$\therefore$ a.c. drain resistance, $r_d = \dfrac{\Delta V_{DS}}{\Delta I_D} = \dfrac{8\,V}{0.25\,mA} = \mathbf{32k\Omega}$

ii. With V_{DS} constant at 15V, drain current changes from 10.25 mA to 9.65 mA as V_{GS} is changed from 0 V to – 0.2 V.

$$V_{GS} = 0.2 - 0 = 0.2\ V$$

$$I_D = 10.25 - 9.65 = 0.6\ mA$$

$$\text{Transconductance, } g_{fs} = \frac{\Delta I_D}{\Delta V_{GS}} = \frac{0.6\,mA}{0.2\,V} = 3\,mA/V = \mathbf{3000\,\mu mho}$$

iii. Amplification factor, $\mu = r_d \times g_{fs} = (32 \times 10^3) \times (3000 \times 10^{-6}) = \mathbf{96}$

Variation of Transconductance (g_m or g_{fs}) of JFET

We have seen that transconductance gm of a jfet is the ratio of a change in drain current (ΔI_D) to a change in gate-source voltage (ΔV_{GS}) at constant V_{DS} i.e.

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

The transconductance g_m of a JFET is an important parameter because it is a major factor in determining the voltage gain of JFET amplifiers. However, the transfer characteristic curve for a JFET is nonlinear so that the value of g_m depends upon the location on the curve. Thus the value of g_m at point A in Fig.19.17 will be different from that at point B. Luckily there is following equation to determine the value of g_m at a specified value of V_{GS}.

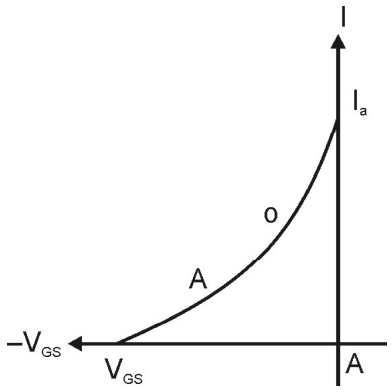


Fig. 5.18

$$g_m = g_{mn}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)$$

where g_m = value of transconductance at any point on the transfer characteristic curve

g_{mo} = value of transconductance (maximum) at $V_{GS} = 0$

Normally, the data sheet provides the value of g_{mo}. When the value of g_{mo} is not available, you can approximately calculate g_{mo} using the following relation:

$$g_{mo} = \frac{2I_{DSS}}{\left|V_{GS(off)}\right|}$$

Example 19.10. A JFET has a value of g_{mo} = 4000 μS . Determine the value of g_m at V_{GS} = –3V. Given that $V_{GS(off)}$ = –8V.

Solution.

$$g_m = g_{mo}\left(1-\frac{V_{GS}}{V_{GS(off)}}\right)$$

$$= 4000\ \mu S\left(1-\frac{-3V}{-8V}\right)$$

$$= 4000\ \mu S\ (0.625) = 2500\ \mu S$$

Example 19.11. The data sheet of a JFET gives the following information : I_{DSS} = 3mA, $V_{GS(off)}$ = –6V and $g_{m(max)}$ = 5000 μS. Determine the transconductance for V_{GS} = –4V and find drain current I_D at this point.

Solution. At V_{GS} = 0, the value of g_m is maximum i.e. g_{mo}.

$$\therefore \quad g_{mo} = 5000\ \mu S$$

Now

$$g_m = g_{ma}\left(1-\frac{V_{GS}}{V_{GS(off)}}\right)$$

$$= 5000\ \mu S\left(1-\frac{-4V}{-6V}\right)$$

$$= 5000\ \mu S\ (1/3) = 1667\ \mu S$$

Also

$$I_D = I_{DSS}\left(1-\frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$= 3\ mA\left(1-\frac{-4}{-6}\right)^2 = 333\ \mu A$$

Types of MOSFETs

There are two basic types of MOSFETs viz,

1. Depletion-type MOSFET or D-MOSFET can be operated in both the depletion-mode and the enhancement-mode. For this reason, a D-MOSFET is sometimes called depletion/enhancement MOSFET.
2. Enhancement-type MOSFET or E-MOSFET. The E-MOSFET can be operated only in enhancement-mode.

The manner in which a MOSFET is constructed determines whether it is D-MOSFET or E-MOSFET.

1. **D-MOSFET.** Fig.19.43 shows the constructional details of n-channel D-MOSFET. It is similar to n-channel JFET except with the following modifications/remarks.

   i. The n-channel D-MOSFET is a piece of n-type material with a p-type region (called substrate) on the right and an insulated gate on the left as shown in Fig.19.43. The free electrons (∵ it is n-channel) flowing from source to drain must pass through the narrow channel between the gate and the p-type region (i.e. substrate).

   ii. Note carefully the gate construction of D-MOSFET. A thin layer of metal oxide (usually silicon dioxide, SiO_2 is an insulator, therefore, gate is insulated from the channel. Note that the arrangement forms a capacitor. One plate of this capacitor is the gate and the other plate is the channel with SiO_2 as the dielectric. Recall that we have a gate diode in a JFET.

   iii. It is a usual practice to connect the substrate to the source (S) internally so that a MOSFET has three terminals viz source (S), gate (G) and drain (D).

   iv. Since the gate is insulated from the channel, we can apply either negative or positive voltage to the gate. Therefore, D-MOSFET can be operated in both depletion-mode and enhancement-mode. However, JFET can be operated only in depletion-mode.

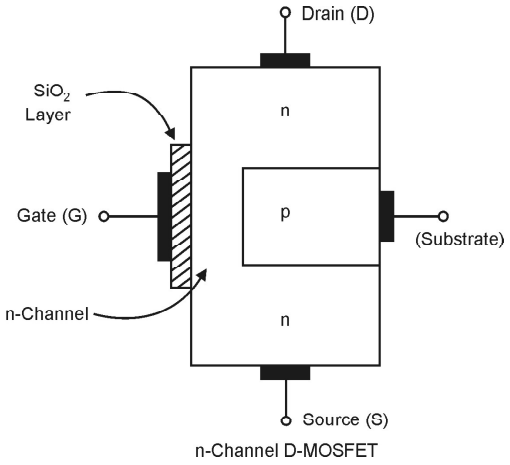


Fig. 5.19

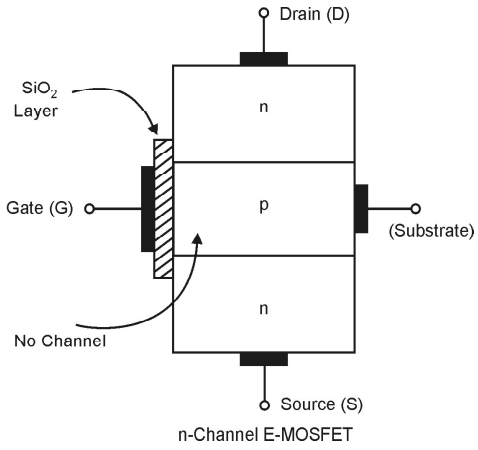


Fig. 5.20

2. **E-MOSFET.** Fig. 19.44 shows the constructional details of n-channel E-MOSFET. Its gate construction is similar to that of D-MOSFET. The E-MOSFET has no channel between source and drain unlike the D-MOSFET. Note that the substrate extends completely to the SiO_2 layer so that no channel exists. The E-MOSFET requires a proper gate voltage to form a channel (called induced channel). It is reminded that E-MOSFET can be operated only in enhancement mode. In short, the construction of E-MOSFET is quite similar to that of the D-MOSFET except for the absence of a channel between the drain and source terminals.

Why the name MOSFET? The reader may wonder why is the device called MOSFET? The answer is simple. The SiO2 layer is an insulator. The gate terminal is made of a metal conductor. Thus, going from gate to substrate, you have a metal oxide semiconductor and hence the name MOSFET. Since the gate is insulated from the channel, the MOSFET is sometimes called insulated-gate FET (IGFET). However, this term is rarely used in place of the term MOSFET.

Symbol for D-MOSFET

There are two types of D-MOSFETs viz (i) n-channel D-MOSFET and (ii) p-channel D-MOSFET.

i. **n-channel D-MOSFET.** Fig. 19.45 (i) show the various parts n-channel D-MOSFET. The p-type substrate constricts the channel between the source and drain so that only a small passage remains at the left side.

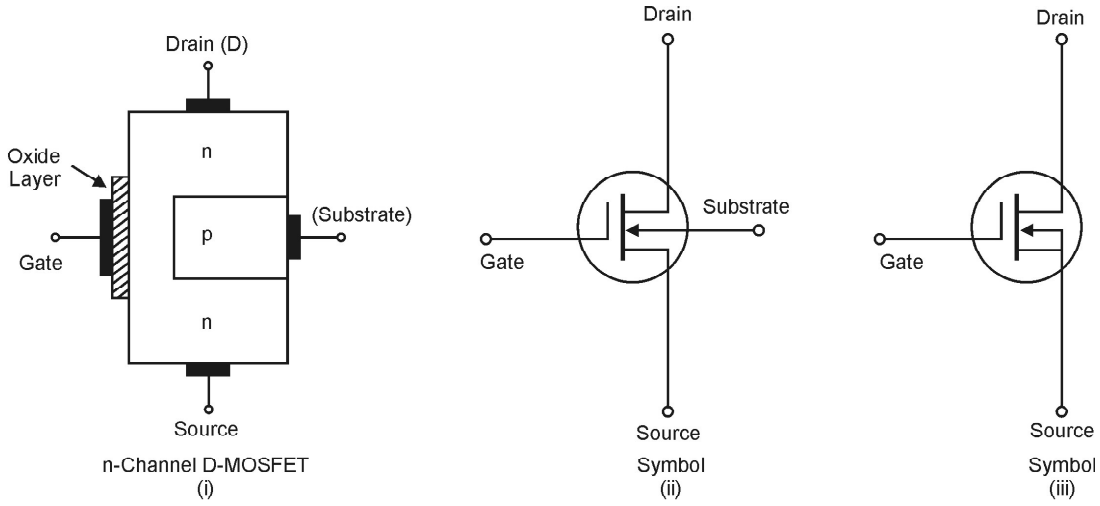


Fig. 5.21

Electrons flowing from source (when drain is positive w.r.t. source) must pass through this narrow channel. The symbol for n-channel D-MOSFET is shown in Fig. 19.45 (ii). The gate appears like a capacitor plate. Just to the right of the gate is a thick vertical line representing the channel. The drain lead comes out of the top of the channel and the source lead connects to the bottom. The arrow is on the substrate and points to the n-material, therefore we have n-channle D-MOSFET. It is a usual practice to connect the substrate to source internally as shown in Fig. 19.45 (iii). This gives rise to a three-terminal device.

ii. **p-channel D-MOSFET.** Fig. 19.46(i) shows the various parts of p-channel D-MOSFET. The n-type substrate constricts the channel between the source and drain so that only a small passage remains at the left side. The conduction takes place by the flow of holes from source to drain through this narrow channel. The symbol for p-channel D-MOSFET is shown in Fig. 19.46(ii). It is a usual practice to connect the substrate to source internally. This results in a three-terminal device whose schematic symbol is shown in Fig. 19.46(iii).

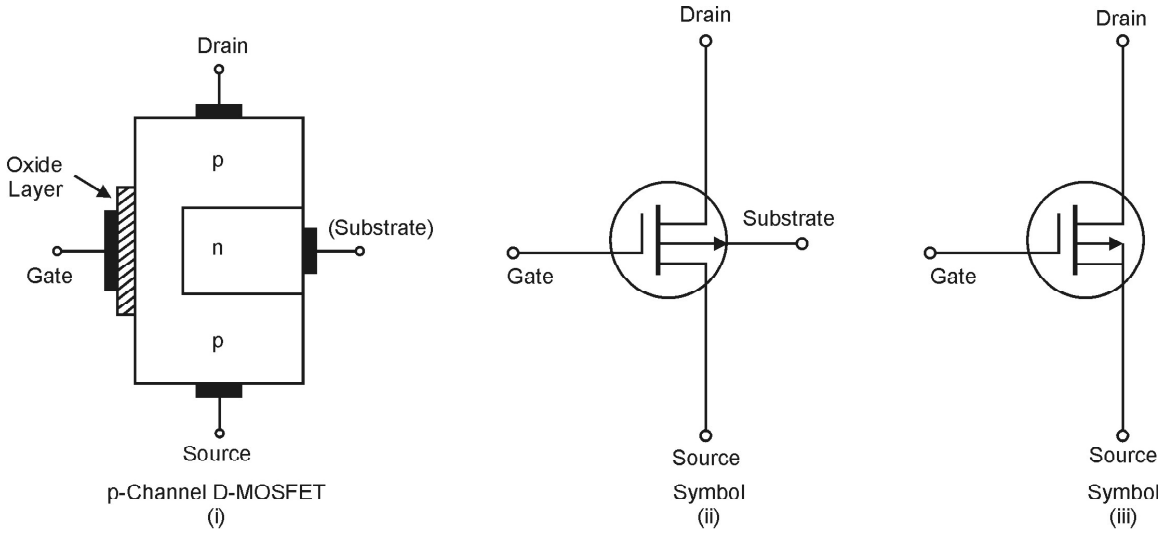


Fig. 5.22

Circuit Operation of D-MOSFET

Fig.19.47 (i) shows that circuit of n-channel D-MOSFET. The gate forms a small capacitor. One plate of this capacitor is the gate and the other plate is the channel with metal oxide layer as the dielectric. When gate voltage is changed, the electric field of the capacitor changes which in turn changes the resistance of the n-channel. Since the gate is insulated from the channel, we can apply either negative or positive voltage to the gate. The negative-gate operation is called depletion mode whereas positive-gate operation is known as enhancement mode.

i. **Depletion mode**. Fig.19.47 (i) shows depletion-mode operation of n-channel D-MOSFET. Since gate is negative, it means electrons are on the gate as shown in Fig.19.47 (ii). These electrons repel the free electrons in the n-channel, leaving a layer of positive ions in a part of the channel as shown in fig.19.47(ii). In other words, we have depleted (i.e. emptied) the n-channel of some of its free electrons. Therefore, lesser number of free electrons are made available for current conduction through the n-channel. This is the same thing as if the resistance of the channel is increased. The greater the negative voltage on the gate, the lesser is the current from source of drain.

Thus by changing the negative voltage on the gate, we can vary the resistance of the n-channel and hence the current from source do drain. Note that with negative voltage to the gate, the action of D-MOSFET is similar to JFET. Because the action with negative gate depends upon depleting (i.e. emptying the channel of free electrons, the negative-gate operation is called depletion mode.

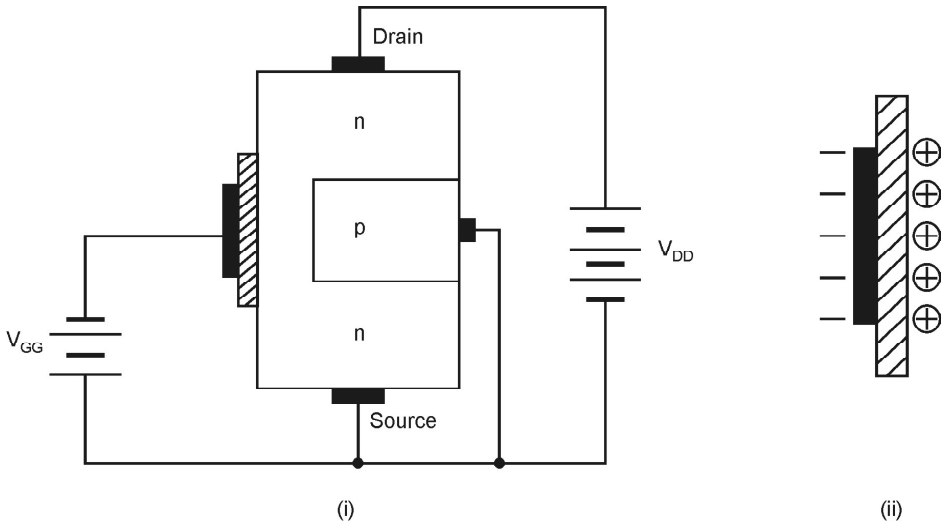


Fig. 5.23

ii. **Enhancement mode.** Fig.19.48(i) shows enhancement-mode operation of n-channel D-MOSFET. Again, the gate acts like a capacitor. Since the gate is positive, it induces negative charges in the n-channel as shown in Fig.19.48(ii). These negative charges are the free electrons drawn the channel. Because these free electrons are added to those already in the channel, the total number of free electrons in the channel is increased. Thus a positive gate voltage enhances or increases conductivity of the channel. The greater the positive voltage on the gate, greater the conduction source to drain.

Thus by changing the positive voltage on the gate, we can change the conductivity of the nel. The main difference between D-MOSFET and JFET is that we can apply positive gate voltage D-MOSFET and still have essentially zero current. Because the action with a positive gate depends upon enhancing the conductivity of the channel, the positive gate operation is called enhancement mode.

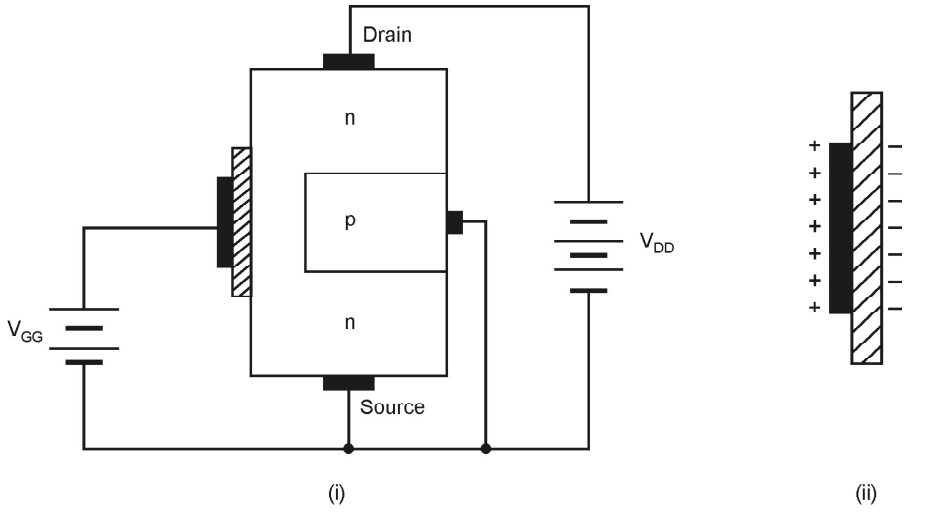


Fig. 5.24

The following points may be noted about D-MOSFET operation:

i. In a D-MOSFET, the source to drain current is controlled by the electric field of capacitor formed at the gate.

ii. The gate of JFET behaves as a reverse-biased diode whereas the gate of a D-MOSFET like a capacitor. For this reason, it is possible to operate D-MOSFET with positive or negative gate voltage.

iii. As the gate of D-MOSFET forms a capacitor, therefore, negligible gate current flows whether positive or negative voltage is applied to the gate. For this reason, the input impedance of D-MOSFET is very high, ranging from 10,000 MΩ to 10,000,00 MΩ.

iv. The externally small dimensions of the oxide layer under the gate terminal result in a very low capacitance and the D-MOSFET has, therefore, a very low input capacitance. This characteristic makes the D-MOSFET useful in high-frequency applications.

D-MOSFET Transfer Characteristic

Fig.19.49 shows the transfer characteristic curve (or transconductance curve) for n-channel D-mosfet. The behaviour of this device can be beautifully explained with the help of this curve as under:

i. The point on the curve where $V_{GS} = 0$, $I_D = I_{DSS}$. It is expected because I_{DSS} is the value of I_D when gate and source terminals are shorted i.e. $V_{GS} = 0$.

ii. As V_{GS} goes negative, I_D decreases below the value of I_{DSS} till I_D reaches zero when $V_{GS} = V_{GS(off)}$ just as with JFET.

iii. When V_{GS} is positive, I_D increases above the value of I_{DSS}. The maximum allowable value of I_D is given on the data sheet of D-MOSFET.

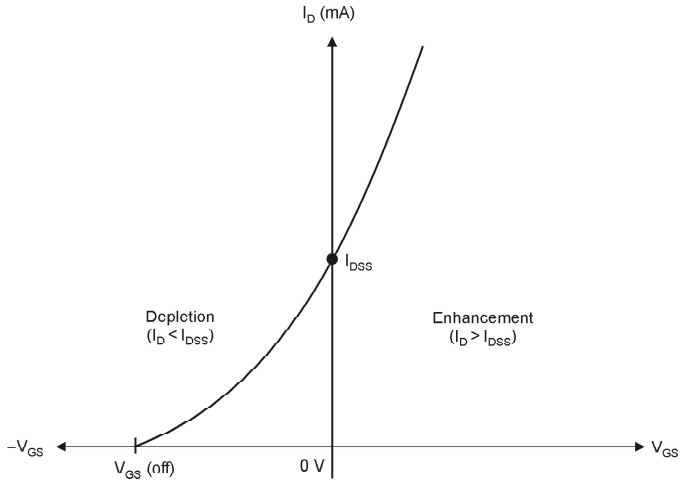


Fig. 5.25

Note that the transconductance curve for the D-MOSFET is very similar to the curve for a JFET. Because of this similarity, the JFET and the D-MOSFET have the same transconductance equation viz.

$$I_D = I_{DSS}\left(I - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

Example 19.30. For a certain D-MOSFET, I_{DSS} = 10 mA and $V_{GS(off)} = -\delta V$.

i. Is this an n-channel or a p-channel?

ii. Calculate I_D at $V_{GS} = -3V$.

iii. Calculate I_D at $V_{GS} = +3V$.

Solution.

i. The device has a negative $V_{GS(off)}$. Therefore, it is n-channel D-MOSFET.

ii. $$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$= 10\text{ mA}\left(1 - \frac{-3}{-8}\right)^2 = 3.91\text{ mA}$$

iii. $$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$= 10\text{ mA}\left(1 - \frac{+3v}{-8V}\right)^2 = 18.9\text{ mA}$$

Example 19.31. A D-MOSFET has parameters of $V_{GS(off)}$ =–6V and I_{DSS} = 1 mA. How will you plot the transconductance curve for the device?

Solution. When V_{GS} = 0 V, $I_D = I_{DSS}$ = 1 mA and when $V_{GS} = V_{GS(off)}$, I_D = 0A. This locates two points viz I_{DSS} and $V_{GS(off)}$ on the transconductance curve. We can locate more points of the curve changing V_{GS} values.

When V_{GS} = –3V ; $I_D = 1\text{ mA}\left(1-\frac{-3V}{-6V}\right)^2 = 0.25\text{ mA}$

When V_{GS} = –1V ; $I_D = 1\text{ mA}\left(1-\frac{-1V}{-6V}\right)^2 = 0.694\text{mA}$

When V_{GS} = +1V ; $I_D = 1\text{ mA}\left(1-\frac{+1V}{-6V}\right)^2 = 1.36\text{ mA}$

When V_{GS} = +3V ; $I_D = 1\text{ mA}\left(1-\frac{+3V}{-6V}\right)^2 = 2.25\text{ mA}$

Thus we have a number of $V_{GS} - I_D$ readings so that transconductance curve for the device can be readily plotted.

Transconductance and Input Impedance of D-MOSFET

There are important parameters of a D-MOSFET and a brief discussion on them is desirable.

i. D-MOSFET Transconductance (g_m). The value of g_m is found for a D-MOSFET in the same way that it is for the JFET i.e.

$$g_m = g_{mo}\left(1-\frac{V_{GS}}{V_{GS(off)}}\right)$$

ii. D-MOSFET Input Impedance. The gate impedance of a D-MOSFET is extremely high. For example, a typical D-MOSFET may have a maximum gate current of 10 pA when V_{GS} = 35V

$$\therefore \text{ Input impedance} = \frac{35V}{10pA} = \frac{35V}{10\times10^{-12}A} = 3.5\times10^{12}\Omega$$

With an input impedance in this range, D-MOSFET would present virtually no load to a source circuit.

D-MOSFET Biasing

The following methods may be used for D-MOSFET biasing:

i. Gate bias ii. Self-bias
iii. Voltage-divider bias iv. Zero bias

The first three methods are exactly the same as those used for jfets and are not discussed here. However, the last method of zero-bias is widely used in D-MOSFET circuits.

Zero bias. Since a D-MOSFET can be operated with either positive or negative values of V_{GS}, we can set its Q-point at V_{GS} = 0V as shown in fig.19.50. Then an input a.c. signal to the gate can produce variations above and below the Q-point.

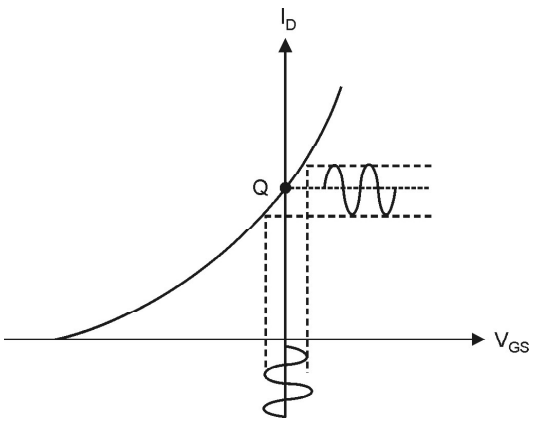


Fig. 5.26

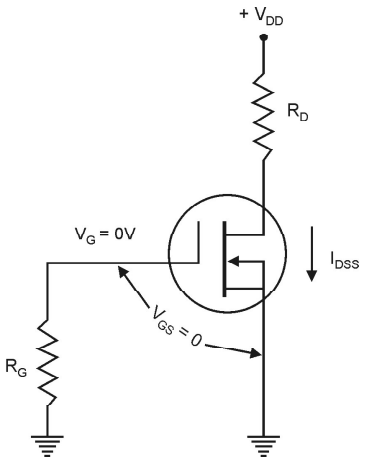


Fig. 5.27

We can use the simple circuit of Fig.19.51 to provide zero bias. This circuit has V_{GS} = 0V and $I_D = I_{DSS}$. We can find V_{DS} as under:

$$V_{DS} = V_{DD} - I_{DSS}R_D$$

Note that for the D-MOSFET zero bias circuit, the source resistor (R_S) is not necessary. With no source resistor, the value of V_S is 0V. This gives us a value of V_{GS} = 0V. This biases the circuit at $I_D = I_{DSS}$ and V_{GS} = 0V. For mid-point biasing, the value of R_D is so selected that $V_{DS} = V_{DD}/2$.

Example 19.32 Determine the drain-to-source voltage (V_{DS}) in the circuit shown in fig.19.51 above if V_{DD} = +18V and R_D = 620Ω. The MOSFET data sheet gives $V_{GS(off)}$ = –8V and I_{DSS} = 12 mA.

Solution. Since $I_D = I_{DSS}$ = 12 mA, the V_{DS} is given by;

$$V_{DS} = V_{DD} - I_{DSS}\, R_D$$
$$= 18V - (12mA)\,(0.62k\Omega) = 10.6V$$

Common-Source D-MOSFET Amplifier

Fig.19.52 shows a common-source amplifier using n-channel D-MOSFET. Since the source terminal is common to the input and output terminals, the circuit is called common-source amplifier. The circuit is zero biased with an a.c. source coupled to the gate through the coupling capacitor C_1. The gate is at approximately 0V d.c. and the source terminal is grounded, thus making V_{GS} = 0V.

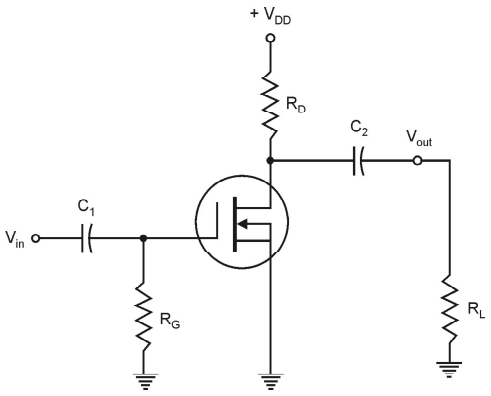


Fig. 5.28

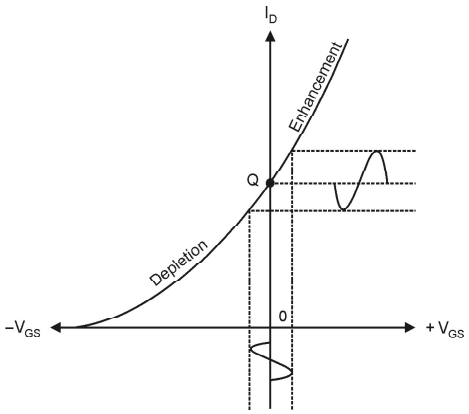


Fig. 5.29

Operation. The input signal (V_{in}) is capacitively coupled to the gate terminal. In the absence of the signal d.c., value of V_{GS} = 0V. When signal (V_{in}) is applied, V_{GS} swings above and below value (∵ d.c. value of V_{GS} = 0V), producing a swing in drain current I_d.

i. A small change in gate voltage produces a large change in drain current as in a JFET. This fact makes MOSFET capable of raising the strength of a weak signal; thus acting as an amplifier.

ii. During the positive half-cycle of the signal, the positive voltage on the gate increases produces the enhancement-mode. This increases the channel conductivity and hence the drain rent.

iii. During the negative half-cycle of the signal, the positive voltage on the gate decreases produces depletion-mode. This decreases the conductivity and hence the drain current.

The result of above action is that a small change in gate voltage produces a large change in the drain current. This large variation in drain current produces a large a.c. output voltage across drain resistance R_D. In this way, D-MOSFET acts as an amplifier. Fig.19.53 shows the amplifying actions D-MOSFET on transconductance curve.

Voltage gain. The a.c. analysis of D-MOSFET is similar to that of the JFET. Therefore, voltage gain expressions derived for JFET are also applicable to D-MOSFET.

Voltage gain, $A_v = g_m R_D$ .... for unloaded D-MOSFET amplifier

$= g_m R_{AC}$ .... for loaded D-MOSFET amplifier.

Note the total a.c. drain resistance $R_{AC} = R_D \,||R_D$.

Example 19.33. The D-MOSFET used in the amplifier of Fig.19.54 has an I_{DSS} =12 mA and g_m = 3.2 mS. Determine (i) d.c. drain-to-source voltage V_{DS} and (ii) a.c. output voltage. Given V_{in} = 500 mV.

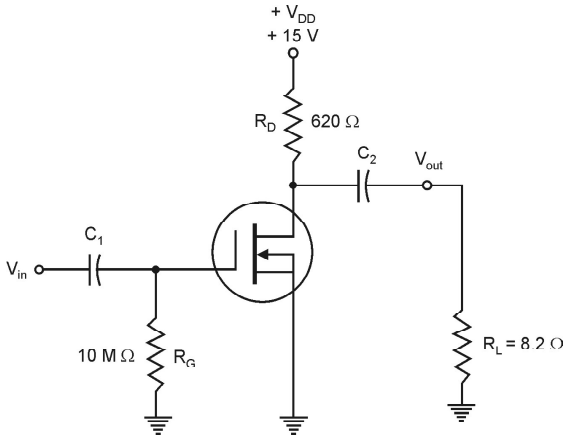


Fig. 5.30

Solution.

i. Since the amplifier is zero biased. $I_D = I_{DSS}$ = 12 mA.

$$\therefore \quad V_{DS} = V_{DD} - I_{DSS}R_D$$
$$= 15V - (12\ mA)\ (0.62\ k\Omega) = 7.56V$$

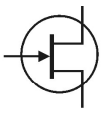
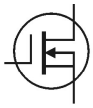
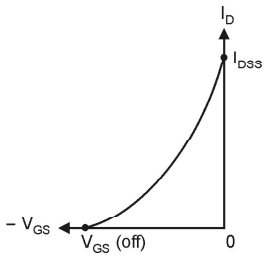
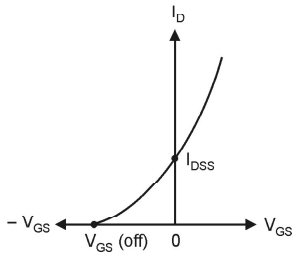
ii. Total a.c. drain resistance R_{AC} of the circuit is

$$R_{AC} = R_D \| R_L = 620\Omega \| 8.2k\Omega = 576\Omega$$

$$\therefore \quad V_{ou} = A_v \times V_{in} = (g_m R_{AC})\ (V_{in})$$
$$= (3.2 \times 10^{-3S} \times 576\Omega)\ (500mV) = 922mV$$

D-MOSFETs Versus JFETs

Table below summarises many of the characteristics of JFETs and D-MOSFETs.

Devices	JFETs	D-MOSFETs
Schematic Symbol:		
Transconductance curve		
Modes of operation	Depletion only	Depletion and enhancement
Commonly used bias circuits	Gate bias Self bias Voltage-divider bias	Gate bias Self bias Voltage-divider bias Zero bias
Advantages	Extremely high input impedance.	Higher input impedance than a comparable JFET. Can operate in both modes (depletion and enhancement)
Disadvantages	Bias instability. Can operate only in the depletion mode.	Bias instability. More sensitive to changes in temperature than the JFET.

E-MOSFET

Two things are worth noting about E-MOSFET. First E-MOSFET operates only in the enhancement mode and has no depletion. Secondly, the E-MOSFET has no physical channel from source to drain because the substrate extends completely to the SiO_2 layer [See Fig.19.55(i)]. It is only by the application of V_{GS} (gate-to-source voltage) of proper magnitude and polarity that the device starts conducting. The minimum, value of V_{GS} of proper polarity that turns on the E-MOSFET is called Threshold voltage [$V_{GS(th)}$]. The n-channel device requires positive V_{GS} ($\geq V_{GS(th)}$) and the p-channel device requires negative V_{GS} ($\geq V_{GS(th)}$) and the p-channel device requires negative($\geq V_{GS(th)}$).

Operation. Fig.19.55(i) shows the circuit of n-channel E-MOSFET. The circuit action is as under:

i. When V_{GS} = 0V [See Fig. 19.55(i)]. there is no channel connecting the source and drain. The p substrate has only a few thermally produced free electrons (minority carriers) so that drain current is essentially zero. For this reason, E-MOSFET is normally OFF when V_{GS} = 0V. Note that this behaviour of E-MOSFET is quite different from JFET or D-MOSFET.

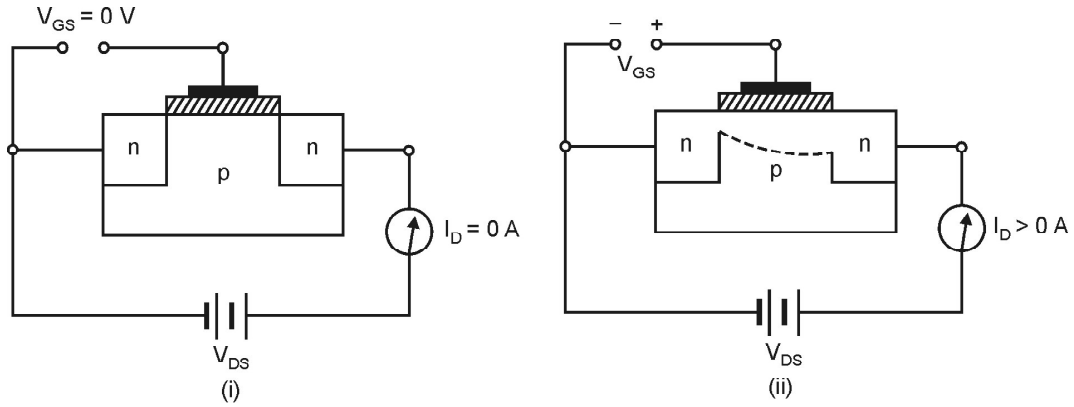


Fig. 5.31

ii. When gate is made positive (i.e. V_{GS} is positive) as shown in Fig.19.55(ii), it attracts free electrons into the p-region. The free electrons combine with the holes next to the SiO_2 layer. If V_{GS} is positive enough, all the holes touching the SiO_2 layer are filled and free electrons begin to flow from the source to drain. The effect is the same as creating a thin layer of n-type material (i.e. inducing a thin n-channel) adjacent to the SiO_2 layer. Thus the E-MOSFET is turned ON and drain current I_D starts flowing form the source to the drain. The minimum value of V_{GS} that turns the E-MOSFET ON is called threshold voltage [$V_{GS(th)}$].

iii. When V_{GS} is less than $V_{GS(th)}$, there is no induced channel and the drain current I_D is zero. When V_{GS} is equal to $V_{GS(th)}$, the E-MOSFET is turned ON and the induced channel conducts drain current from the source of the drain. Beyond $V_{GS(th)}$, if the value of V_{GS} is increased, the newly formed channel becomes wider, causing I_D to increase. If the value of V_{GS} decreases [not less than $V_{GS(th)}$], the channel becomes narrower and I_D will decrease. This fact is revealed by the transconductance curve of n-channel E-MOSFET shown in Fig.19.56. As you can see, I_D = 0 when V_{GS} = 0. Therefore the value of I_{DSS} of the E-MOSFET is zero. Note also that there is no drain current until V_{GS} reaches $V_{GS(th)}$.

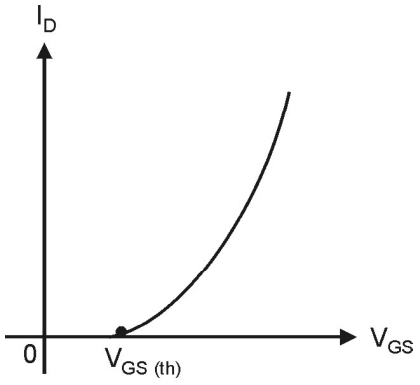


Fig. 5.32

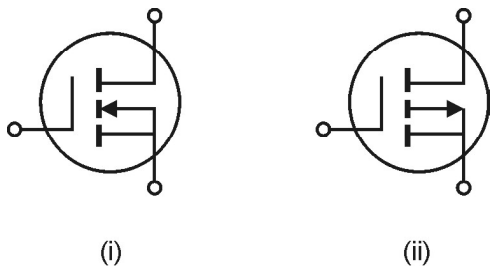


Fig. 5.33

Schematic Symbols. Fig.19.57(i) shows the schematic symbols for n-channel E-MOSFET whereas Fig.19.57 (ii) shows the schematic symbol for p-channel E-MOSFET. When $V_{GS} = 0$, the E-MOSFET is OFF because there is no conducting channel between source and drain. The broken channel line in the symbols indicates the normally OFF condition.

Equation for Transconductance Curve. Fig.19.58 shows the transconductance curve for n-channel E-MOSFET, note that this curve is different from the transconductance curve for n-channel JFET or n-channel D-MOSFET. It is because it starts at $V_{GS(th)}$ rather than $V_{GS(off)}$ on the horizontal axis and never intersects the vertical axis. The equation for the E-MOSFET transconductance curve for ($V_{GS} > V_{GS(th)}$) is

$$I_D = K\left(V_{GS} - V_{GS(th)}\right)^2$$

The constant K depends on the particular E-MOSFET and its value is determined from the following equation:

$$K = \frac{I_{D(on)}}{(V_{GS(on)} - V_{GS(th)})^2}$$

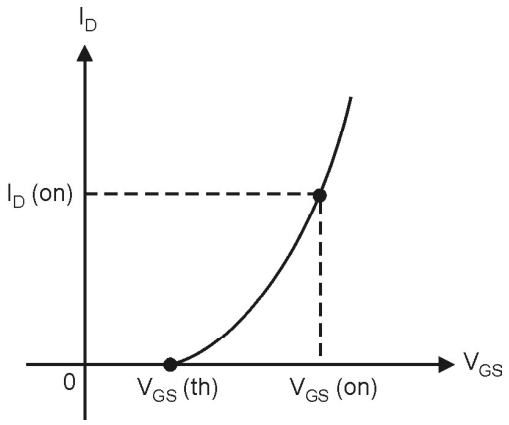


Fig. 5.34

Any data sheet for an E-mosfet will include the current $I_{D(on)}$ and the voltage $V_{GS(on)}$ for one point well above the threshold voltage as shown in Fig.19.58.

Example 19.34. The data sheet for an E-MOSFET gives $I_{D(on)}$ = 500 mA at V_{GS} = 10V and $V_{GS(th)}$ = 1V. Determine the drain current for V_{GS} = 5V.

Solution. Here $V_{GS(on)}$ = 10V.

$$I_D = K\left(V_{GS} - V_{GS(th)}\right)^2$$

Here $$K = \frac{I_{D(on)}}{\left(V_{GS(on)} - V_{GS(th)}\right)^2} = \frac{500mA}{(10V - 1V)^2} = 6.17mA/V^2$$

Putting the various values in eq.(i), we have,

$$I_D = 6.17\ (5V - 1V)^2 = 98.7\ mA$$

Example 19.35. The data sheet for an E-MOSFET gives $I_{D(on)}$ = 3 mA at V_{GS} = 10V and $V_{GS(th)}$ = 3V. Determine the resulting value of K for the device. How will you plot the transconductance curve for this MOSFET?

Solution. The value of K can be determine from the following equation:

$$K = \frac{I_{D(on)}}{\left(V_{GS(on)} - V_{GS(th)}\right)^2}$$

Here $I_{D(on)} = 3\ mA;\ V_{GS(on)} = 10V : V_{GS(th)} = 3V$

$\therefore$ $$K = \frac{3mA}{(10V - 3V)^2} = \frac{3mA}{(7V)^2} = 0.061 \times 10^{-3} A/V^2$$

Now $$I_D = K\left(V_{GS} - V_{GS(th)}\right)^2$$

In order to plot the transconductance curve for the device, we shall determine a few points for the curve by changing the value of V_{GS} and noting the corresponding values of I_D.

For V_{GS} = 5V ; $I_D = 0.061 \times 10^{-3} (5V - 3V)^2 = 0.244$ mA

For V_{GS} = 8V ; $I_D = 0.061 \times 10^{-3} (8V - 3V)^2 = 1.525$ mA

For V_{GS} = 10V ; $I_D = 0.061 \times 10^{-3} (10V - 3V)^2 = 3$ mA

For V_{GS} = 12V ; $I_D = 0.061 \times 10^{-3} (12V - 3V)^2 = 4.94$ mA

Thus we can plot the transconductance curve for the E-MOSFET from these V_{GS}/I_D points.

Boolean Algebra

Mathematician George Boole invented a new kind of algebra – the algebra of logic in the year 1854, popularly known as **Boolean algebra** or Switching algebra. He states that symbols can be used to represent the structure of logical thought. Boolean algebra differs significantly from conventional algebra. This algebra deals with the rules by which the logical operations are carried out. Here a digital circuit is represented by a set of input and output symbols and the circuit function expressed as a set of Boolean relationships between the symbols

Boolean algebra is an algebra for the manipulation of binary variables that can take on only two values 0 or 1.

Basic Laws of Boolean Algebra

Logic operation can be expressed and minimized mathematically using the rules, laws and theorems of Boolean algebra. It is a convenient and systematic method of expressing and analyzing the operation of digital circuits and systems.

Boolean Addition and Multiplication

Addition and multiplication by the Boolean methods involves variables having values of either a binary 1 or a 0. The basic rules of Boolean addition and multiplication are given below:

Boolean Addition	Boolean Multiplication
0 + 0 = 0	0 – 0 = 0
0 + 1 = 1	0 – 1 = 0
1 + 0 = 1	1 – 0 = 0
1 + 1 = 1	1 – 1 = 1

Table 6.11 : Rules for Boolean addition and Boolean Multiplication

Boolean addition is same as the logical OR operation, while boolean multiplication is same as the logical AND operation.

Boolean Algebraic Theorems

binary variables can be represented by a letter symbol such as A, B, X, Y,.... The variable can have only one of the two possible values at any time. viz. 0 or 1. The boolean algebraic theorems are given in the following table:

Theorem No.	Theorem
1	A + 0 = A
2	A – 1 = A
3	A + 1 = A
4	A + 1 = 1
5	A + A = A
6	A – A = A
7	$A + \bar{A} = 1$
8	$A . \bar{A} = 0$

Theorem No.	Theorem
9	A . (B + C) = A. B + A .C
10	A + BC = (A + B) (A + C)
11	A + AB = A
12	A. (A + B) = A
13	$A + \bar{A}B = A + B$
14	$A(\bar{A} + B) = AB$
15	$AB + A\bar{B} = A$
16	$(A + B)(A + \bar{B}) = A$
17	$AB + \bar{A}C = (A + C)(\bar{A} + B)$
18	$(A + B)(\bar{A} + C) = AC + \bar{A}B$
19	$AB + \bar{A}C + BC = AB + \bar{A}C$
20	$(A + B)(\bar{A} + C)(B+C) = (A+B)(\bar{A} + C)$

Table 6.12 Boolean Algebraic Theorems

From these theorems we observe that the even numbered theorems can be obtained from their proceeding odd numbered theorems by interchanging (+) and (.) and by interchanging 1 and 0 by **Duality concept**.

The theorem can be proved by making a truth table. For example.

Theorem 10; A + BC = (A + B) (A + C)

A	B	C	B.C	A + BC	A + B	A + C	(A+B) (A+C)
0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0
0	1	0	0	0	1	0	0
0	1	1	1	1	1	1	1
1	0	0	0	1	1	1	1
1	0	1	0	1	1	1	1
1	1	0	0	1	1	1	1
1	1	1	1	1	1	1	1

Column 5 and 8 proves the theorem true.

Boolean algebra has the commutative, associative, distributive, absorption, consensus and idempotency property which are described as following:

i. **Commutative property:**

$$\boxed{A + B = B + A} \quad ...6.1(a)$$

$$\boxed{A . B = B . A} \quad ...6.1(b)$$

According to commutative property of boolean algebra the order of the AND and OR operation conducted on the variables makes no difference.

ii. **Associative property :**

$$\boxed{A + (B + C) = (A + B) + C} \quad \text{...6.2 (a)}$$

$$\boxed{A \cdot (B \cdot C) = (A \cdot B) \cdot C} \quad \text{...6.2 (b)}$$

According to this law, it makes no difference in what order the variables are grouped during the AND and OR operation of several variables.

iii. **Distributive property :**

$$\boxed{A + BC = (A + B)(A + C)} \quad \text{....6.3(a)}$$

The AND operation of several variables and then the OR operation of the result with a single variable is equivalent to the OR operation of the single variable with each of the several variables and then the AND operation of the sums.

Similarly, $\boxed{A \cdot (B + C) = A \cdot B + A \cdot C}$ ....6.3(b)

The OR operation of several variables and then the AND operation of the result with single variable is equivalent to the AND operation of the single variable with each of the several variable and then the OR operation of the products.

iv. **Absorption property :**

1. $\boxed{A + AB = A}$ ....6.4(a)

Proof:

$$\begin{aligned} A + AB &= A \cdot 1 + AB \\ &= A(1 + B) \\ &= A \cdot 1 = A \qquad [\because 1 \text{ OR } B = 1] \\ A + AB &= A \end{aligned}$$

2. $\boxed{A \cdot (A + B) = A}$ ...6.4(b)

Proof

$$\begin{aligned} A \cdot (A + B) &= A \cdot A + A \cdot B \\ &= A + AB \\ &= A(I + B) \\ &= A \cdot I = A \\ A(A + B) &= A \end{aligned}$$

3. $\boxed{A + \bar{A}B = A + B}$ ....6.5(a)

Proof

$$\begin{aligned} A + \bar{A}B &= (A + \bar{A})(A + B) \\ &= 1 \cdot (A + B) \\ &= (A + B) \\ A + \bar{A}B &= A + B \end{aligned}$$

4. $\boxed{A \cdot (\bar{A} + B) = AB}$ ....6.5(b)

Proof

$$\begin{aligned} A \cdot (\bar{A} + B) &= A\bar{A} + AB \\ &= 0 + AB \\ &= AB \\ A \cdot (\bar{A} + B) &= AB \end{aligned}$$

v. **Consensus Law :**

1. $\boxed{AB + \bar{A}C + BC = AB + \bar{A}C}$ ....6.6(a)

Proof

$$\begin{aligned} AB + \bar{A}C + BC &= AB + \bar{A}C + BC \cdot 1 \\ &= AB + \bar{A}C + BC(A + \bar{A}) \qquad (\because A + \bar{A} = 1) \\ &= AB + \bar{A}C + ABC + \bar{A}BC \\ &= AB(1 + C) + \bar{A}C(1 + B) \\ &= AB \cdot 1 + \bar{A}C \cdot 1 \qquad (\because 1 + C = 1 = 1 + B) \\ &= AB + \bar{A}C \\ AB + \bar{A}C + BC &= AB + \bar{A}C \end{aligned}$$

2. $\boxed{(A + B)(\bar{A} + C)(B + C) = (A + B)(\bar{A} + C)}$ ...6.6 (b)

Proof $(A + B)(\bar{A} + C)(B + C) = (A + B)(\bar{A} + C)(B + C + 0)$

$= (A + B)(\bar{A} + C)(B + C + A\bar{A})$

$= (A + B)(\bar{A} + C)(B + C + A)(B + C + \bar{A})$

$[\because A + BC = (A + B)(A + C)]$

$= (A + B)(A + B + C)(\bar{A}+C)(\bar{A}+ C + B)$

$= (A + B)(\bar{A} + C)$ $[\because A(A + B) = A]$

$(A + B)(\bar{A} + C)(B + C) = (A + B)(\bar{A} + C)$

vi. Principle of Duality

The above properties and laws of boolean algebra are grouped in pairs as (a) and (b). One expression can be obtained from other in each pair by replacing every 0 and 1, ever 1 with 0, every (+) with (.) and every (.) with (+). Any pair of expression satisfying this rule are called dual expression and the rule is called principle of duality.

De-Morgan's Theorem

Two theorems that are an important part of Boolean algebra were proposed by De-Morgan. The first theorem states that the complement of a product is equal to the sum of the complement, i.e., if the variable are A and B, then -

$$\boxed{\overline{AB} = \bar{A} + \bar{B}}$$

The second theorem states that, the complement of a sum is equal to the product of the complements. In equation form, this can be written as –

$$\boxed{\overline{A+B} = \bar{A}.\bar{B}}$$

Proof of De- Morgan's Theorem :

1. **By Truth Table (Perfect Induction Method)**

1 A	2 B	3 $\bar{A}$	4 $\bar{B}$	5 A + B	6 $\overline{A+B}$	7 $\bar{A}.\bar{B}$	8 A · B	9 $\overline{A \cdot B}$	10 $\bar{A}+\bar{B}$
0	0	1	1	0	1	1	0	1	1
0	1	1	0	1	0	0	0	1	1
1	0	0	1	1	0	0	0	1	1
1	1	0	0	1	0	0	1	0	0

Table 6.13. Prof for DE-Morgan's Theorem

To construct the above table, firstly, take all possible combinations of inputs A and B in the first two column. Take complements of A and B to generate columns 3 and 4 respectively. During A and B gives us 5th column whose complement is the 6th column. The 7th and 8th column show logical AND of complemented and true inputs respectively. Column 9 is the complement of column 8 and column 10 gives logical OR of the complemented inputs.

From the above truth table, we observe that column 6 and 7 are equal. Therefore,

$$\boxed{\overline{A+B} = \bar{A}+\bar{B}}$$

Similarly, column 9 and 10 are equal. Therefore,

$$\boxed{\overline{A.B} = \bar{A}+\bar{B}}$$

2. Algebraic method ;

According to the first theorem, $(\bar{A} + \bar{B})$ is the complement of AB and we know that $A + \bar{A} = 1$ and $A \cdot \bar{A} = 0$. So,

Substituting AB for A and $(\bar{A} + \bar{B})$ for $\bar{A}$ in the above expression

$$AB + \bar{A} + \bar{B} = 1 \quad \text{and} \quad AB(\bar{A} + \bar{B}) = 0$$
$$\bar{A} + B + \bar{B} = 1 \quad \text{and} \quad \bar{A}BA + AB\bar{B} = 0$$
$$\bar{A} + 1 = 1 \quad \text{and} \quad 0 + 0 = 0$$
$$1 = 1 \quad \text{and} \quad 0 = 0$$

Thus, De-Morgan's first theorem is proved algebraically. Similarly according to De-Morgan's second theorem, $\bar{A} \cdot \bar{B}$ is the complement of (A + B). So

$$A + B + \bar{A} \cdot \bar{B} = 1 \quad \text{and} \quad (A + B)(\bar{A} \cdot \bar{B}) = 0$$
$$A + B + \bar{B} = 1 \quad \text{and} \quad A\bar{A}\bar{B} + B\bar{A}\bar{B} = 0$$
$$A + 1 = 1 \quad \text{and} \quad 0 + 0 = 0$$
$$1 = 1 \quad \text{and} \quad 0 = 0$$

Thus, De-Morgan's second theorem is proved algebraically.

Minimization (Simplification) of Boolean Expression using Algebraic Method

The switching or Boolean expressions can be simplified by applying properties, laws and theorems of Boolean algebra. The simplification of different Boolean expressions are demonstrated in the following examples;

Example 12 : Prove that $AB + BC + \bar{B}C = AB + C$.

Solution.
$$AB + BC + \bar{B}C = AB + C(B + \bar{B})$$
$$= AB + C \bullet 1 \qquad [\because B + \bar{B} = 1 : \text{Theorem 7}]$$
$$= AB + C \qquad [\because C \bullet 1 = C : \text{Theorem 2}]$$

Example 13 : Simplify the expression $\bar{A} \cdot B + A \cdot B + \bar{A} \cdot \bar{B}$.

Solution :
$$\bar{A} \cdot B + A \cdot B + \bar{A} \cdot \bar{B} = B(\bar{A} + A) + \bar{A} \cdot \bar{B}$$
$$= B \bullet 1 + \bar{A} \cdot \bar{B} \quad [\because \bar{A} + A = 1 : \text{Theorem 7}]$$
$$= B + \bar{A} \cdot \bar{B} \qquad [\because B \cdot 1 = B : \text{Theorem 2}]$$
$$= B + \bar{A} \qquad [\because A + \bar{A}B = A + B : \text{Theorem 13}]$$

Example 14 : Simplify the expression $A + A \cdot B + \bar{A} \cdot B$

Solution.
$$A + A \cdot \bar{B} + \bar{A} \cdot B = A(1 + \bar{B}) + \bar{A} \cdot B$$
$$= A \cdot 1 + \bar{A}B \qquad [1 + A = 1 : \text{Theorem 3}]$$
$$= A + \bar{A}B$$
$$= A + B \qquad [\because A + \bar{A}B = A + B : \text{Theorem 13}]$$

Example 15 : Complement of the expression $\bar{A}B + C\bar{D}$.

Solution.
$$\overline{\bar{A}B + C\bar{D}} = \overline{(\bar{A}B)} \cdot \overline{(C\bar{D})} \qquad [\text{from De-Morgan's theorem}]$$
$$= \left(\left(\bar{\bar{A}}\right) + \bar{B}\right) \cdot \left(\bar{C} + \left(\bar{\bar{D}}\right)\right)$$
$$= \left(A + \bar{B}\right)\left(\bar{C} + D\right) \qquad [\because \bar{\bar{A}} = A]$$

Example 16 : Simplify the following boolean expression:

i. $\overline{YZ} + \bar{W}\overline{XZ} + \bar{W}XY\bar{Z} + WY\bar{Z}$

ii. $\overline{A + B\bar{C}}$

Solution:

i. $\overline{YZ} + \bar{W}\overline{XZ} + \bar{W}XY\bar{Z} + WY\bar{Z}$

$= \overline{YZ} + \bar{W}\bar{X}\bar{Z} + \bar{W}XY\bar{Z} + WY\bar{Z}$

$= \overline{YZ} + \bar{W}\bar{X}\bar{Z} + Y\bar{Z}(\bar{W}X + W)$ $[\because A + \bar{A}B = A + B]$

$= \overline{YZ} + \bar{W}\bar{X}\bar{Z} + Y\bar{Z}(X + W)$

$= \overline{YZ} + \bar{W}\bar{X}\bar{Z} + XY\bar{Z} + WY\bar{Z}$

$= \overline{YZ} + XY\bar{Z} + \overline{WXZ} + WY\bar{Z}$

$= \bar{Z}(\bar{Y} + XY) + \bar{W}\overline{XZ} + WY\bar{Z}$ $[\because A + \bar{A}B = A + B]$

$= \bar{Z} + (\bar{Y} + X) + \overline{WXZ} + WY\bar{Z}$

$= \overline{YZ} + X\bar{Z} + \overline{WXZ} + WY\bar{Z}$

$= \overline{YZ} + WY\bar{Z} + X\bar{Z} + \overline{WXZ}$

$= \bar{Z}\left(\bar{Y} + WY\right) + \bar{Z}\left(X + \overline{WX}\right)$ $[\because A + \bar{A}B = A + B]$

$= \bar{Z}\left(\bar{Y} + W\right) + \bar{Z}\left(X + \bar{W}\right)$

$= \bar{Z}(\bar{Y} + W + X + \bar{W})$

$= \bar{Z}(\bar{Y} + X + 1)$ $[\because A + \bar{A} = 1]$

$= \bar{Z}(1)$

$= \bar{Z}$

ii. $\overline{A + B\bar{C}}$

Using De-Morgan's law

$\overline{A + B\bar{C}} = \left(\bar{A}\right) \cdot \left(\overline{B\bar{C}}\right)$

$= \bar{A} \cdot \left(\bar{B} + \bar{\bar{C}}\right)$

$= \bar{A} \cdot \left(\bar{B} + C\right)$

$= \bar{A}\bar{B} + \bar{A}C$

Example 17 : Simplify the expression $AB + \overline{AC} + A\bar{B}C\,(AB + C)$

Solution. $AB + \overline{AC} + A\bar{B}C\,(AB + C)$

$= AB + \overline{AC} + A\bar{B}CAB + A\bar{B}C \cdot C$

$= AB + \overline{AC} + A \cdot A \cdot C \cdot \bar{B} \cdot B + A \cdot \bar{B}C$ $[\because C \cdot C = C]$

$= AB + \overline{AC} + 0 + A \cdot \bar{B} \cdot C$ $[\because \bar{B} \cdot B = 0]$

$= AB + \overline{AC} + A\bar{B}C$

$= A(B + \bar{B}C) + \bar{A} + \bar{C}$ $[\because \overline{AC} = \bar{A} + \bar{C}]$

$= A\,(B + C) + \bar{A} + \bar{C}$ $[\because A + \bar{A}B = A + B$

$= AB + AC + \bar{A} + \bar{C}$

$= \bar{A} + AB + \bar{C} + AC$

$= \bar{A} + B + \bar{C} + A$ $[\because A + \bar{A}B = A + B]$

$= 1 + B + \bar{C}$ $[\because A + \bar{A} = 1]$

$= 1 + \bar{C}$ $[\because A + 1 = 1]$

$= 1$

Example 18 : Simplify the expression Y = $(\bar{A}+B)(A+B)$

Solution : Y = $(\bar{A}+B)(A+B)$ $(\bar{A}+B)(A+B)$

= $\bar{A}\cdot A+\bar{A}\cdot B+A\cdot B+B\cdot B$

= $0+\bar{A}\cdot B+A\cdot B+B$ $[\because A\cdot\bar{A}=0]$

= $\bar{A}\cdot B+B[A+1]$

= $\bar{A}\cdot B+B\cdot 1$ $[\because A+1=1]$

= $\bar{A}\cdot B+B$

= $B(\bar{A}+1)$

= $B\cdot 1$

= B

Example 19 : If $\bar{A}B+CD=0$, then prove that

$AB+\bar{C}(\bar{A}+\bar{D})=AB+\bar{B}\bar{D}+\bar{A}\bar{C}D$

Solution. : LHS : $AB+\bar{C}(\bar{A}+\bar{D})$ = $AB+\bar{C}(\bar{A}+\bar{D})+0$

= $AB+\bar{A}\bar{C}+\bar{C}\bar{D}+\bar{A}B+C\bar{D}$ $[\because$ given that $\bar{A}B+C\bar{D}=0]$

= $AB+\bar{A}B+\overline{CD}+C\bar{D}+\overline{AC}$

= $B(A+\bar{A})+\bar{D}(C+\bar{C})+\overline{AC}$

= $B+\bar{D}+\bar{A}\bar{C}$

RHS : $AB+BD+\overline{BD}+\bar{A}\bar{C}D+0=AB+BD+\bar{B}\bar{D}+\bar{A}\bar{C}D+\bar{A}B+C\bar{D}$

$[\because$ given that $\bar{A}B+C\bar{D}=0]$

= $AB+\bar{A}B+BD+\overline{BD}+\overline{AC}D+C\bar{D}$

= $B(A+\bar{A})+BD+\bar{B}\bar{D}+\bar{A}\bar{C}D+C\bar{D}$

= $B(1+D)+\bar{B}\bar{D}+\bar{A}\bar{C}D+C\bar{D}$

= $B+\bar{B}\bar{D}+\bar{A}\bar{C}D+C\bar{D}$

= $B+\bar{D}+\bar{A}\bar{C}D+C\bar{D}$ $[\because A+\bar{A}B=A+B]$

= $B+\bar{D}(1+C)+\bar{A}\bar{C}D$

= $B+\bar{D}+D\bar{A}\bar{C}$

= $B+\bar{D}+\bar{A}\bar{C}$ $[\because A+\bar{A}B=A+B]$

Hence LHS = RHS

Example 20 : Simplify = $A\bar{B}+(\bar{A}+B)C$

Solution. Y = $A\bar{B}+(\bar{A}+B)C$

= $A\bar{B}+(\overline{A\bar{B}})C$ $[\because \overline{A\bar{B}}=\bar{A}+B]$

= $A\bar{B}+(\overline{A\bar{B}})C$ $[\because A+\bar{A}B=A+B,$ Here $A=A\bar{B}, B=C]$

= $A\bar{B}+C$

Example 21 : Simplify Y = A + $\bar{A}B+\bar{A}\bar{B}C+\bar{A}\bar{B}\bar{C}D$

Solution. Y = A + $\bar{A}B+\bar{A}\bar{B}C+\bar{A}\bar{B}\bar{C}D$

= $A+B+\bar{A}\bar{B}(C+\bar{C}D)$

= $A+B+\bar{A}\bar{B}(C+D)$ $[\because x+\bar{x}y=x+y]$

= $A+\bar{A}\bar{B}(C+D)+B$

= $A+\bar{B}(C+D)+B$

= $A+B+\bar{B}(C+D)$

= A + B + C + D

Example 22 : If $A\bar{B} + \bar{A}B = C$ show that $A\bar{C} + \bar{A}C = B$

Solution:

$Y = A\bar{C} + \bar{A}C$

$= A(\overline{A\bar{B} + \bar{A}B}) + \bar{A}(A\bar{B} + \bar{A}B)$

$= A[(\overline{A\bar{B}}).(\overline{\bar{A}B})] + \bar{A}(A\bar{B} + \bar{A}B)$ $[\because \bar{A}A = 0]$

$= A(\bar{A} + B)(A + \bar{B}) + \bar{A}A\bar{B} + \bar{A}\bar{A}B$

$= (A\bar{A} + AB)(A + \bar{B}) + \bar{A}B$ $[\because \bar{A}A = 0]$

$= AB(A + \bar{B}) + \bar{A}B$

$= AAB + AB\bar{B} + \bar{A}B$

$= AB + \bar{A}B$ $[\because B\bar{B} = 0]$

$= B(A + \bar{A})$

$= B \cdot 1$

$= B$

Example 23 : Simplify the following as much as possible $A[B + C(\overline{AB + AC})]$

Solution. Using De-Morgans law on inner brackets, we get

$A[B + C(\overline{AB + AC})] = A[B + C(\overline{AB} \cdot \overline{AC})]$

$= A[B + C\{(\bar{A} + \bar{B}) \cdot (\bar{A} + \bar{C})\}]$ [Using De-Morgan's law]

$= A[B + C\{(\bar{A} \cdot \bar{A} + \bar{A}\bar{C} + \bar{A}\bar{B} + \bar{B}\bar{C}\}]$

$= A[B + C\{\bar{A}(1 + \bar{C} + \bar{B}) + \bar{B}\bar{C}]$

$= A[B + C\{\bar{A} + \bar{B}\bar{C}\}]$ $[\because 1 + A = 1]$

$= A[B + \bar{A}C + C\bar{B}\bar{C}]$

$= A[B + \bar{A}C]$ $[\because A\bar{A} = 0]$

$= AB + A\bar{A}C$

$= AB$ $[\because A\bar{A} = 0]$

Example 24 : Using boolean algebra simply.

$$(\overline{\overline{\bar{A} + B}}) + (\overline{\overline{A + \bar{B}}}) + (\overline{\overline{\bar{A}B}})(\overline{\overline{A\bar{B}}})$$

Solution. Using De-morgan's theorem

$= (\overline{\overline{\bar{A} + B}}) + (\overline{\overline{A + \bar{B}}}) + (\overline{\overline{\bar{A}B}})(\overline{\overline{A\bar{B}}})$

$= (\bar{A} + B)(A + \bar{B}) + \bar{A}B + A\bar{B}$

$= A\bar{A} + \bar{A}\bar{B} + AB + B\bar{B} + \bar{A}B + A\bar{B}$

$= \bar{A}\bar{B} + AB + \bar{A}B + A\bar{B}$ $[\because A\bar{A} = 0]$

$= A(B + \bar{B}) + A(B + \bar{B})$

$= \bar{A} + A$ $[\because B + \bar{B} = 1]$

$= 1$ $[\because B + \bar{B} = 1]$

Binary Number System

A number system is a code that uses symbols to count the number of items. The most common and familiar number system is the decimal number system. The decimal number system uses the symbols 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. Thus, the decimal system uses 10 digits for counting the items. A binary system uses only two digits (0 and 1) for counting the items. The reader may wonder how to count the items in a binary system. Let us see how it is done.

Counting in Decimal and Binary systems. Figures 26.3 shows the counting of stones in decimal as well as binary system. As you will see, the counting in the binary number system is performed much the same way as in the decimal number system.

Stones	Decimal	Binary
○	0	0
○ ○	1	1
○ ○ ○	2	10
○ ○ ○ ○	3	11
○ ○ ○ ○ ○	4	100
○ ○ ○ ○ ○ ○	5	101
○ ○ ○ ○ ○ ○ ○	6	110
○ ○ ○ ○ ○ ○ ○ ○	7	111
○ ○ ○ ○ ○ ○ ○ ○ ○	8	1000
○ ○ ○ ○ ○ ○ ○ ○ ○ ○	9	1001

Fig. 5.35

i. Let us first see how items are counted in decimal system. In this system, the count starts as 0, 1, ..., 9. After 9, we are to write the next number. To do so, we use the second digit of the decimal system (i.e., 1) followed by the first digit (i.e,0). So after 9, the next number is 10. The count again continues as 10, 11, 12....., 19. After 19, we use the third digit of the system (i.e.,2) followed by the first digit (i.e.,0) and the count continues as 20, 21, ...etc. In this way, we get the number upto 99. In order to represent a number next to 99, we use three decimal digits (100). That is to say second digit of the decimal system (i.e.,1) followed by two first digits (i.e. two zeros).

ii. Let us now turn to binary system. Note that 0 and 1 count in the binary system is the same as in the decimal counting. To represent 2 stones, we use the second binary digit (i.e.,1) followed by the first (i.e., 0). This gives binary number 10 (read as one-zero and not ten) as an equivalent of 2 in the decimal system. Likewise, 3 in the decimal system can be represented by the binary number 11 (read as one-one and not eleven). After this, the two binary digits are exhausted. We shall use three digits to represent the next binary number. Thus, to represent 4 (four), we use the second binary digit followed by two first binary digits. This gives the binary 100 (read as one-zero-zero) as equivalent to 4 in the decimal system. Here is a simple way to find binary equivalents. Each time the two digits 1 and 0 in one position are exhausted (counted as high as they will go), a 1 is added at the left, all digits to the right are made 0, and the count continues. The reader may apply this simple rule to find next binary numbers.

Notes :

i. Each binary digit (0 or 1) is referred to as a bit. A string of four bits is called as a nibble and eight bits make a byte. Thus, 1001 is a nibble and 10010110 is a binary byte.

ii. The binary number system is the most useful in digital circuits because there are only two digits (0 and 1).

Place Value

Consider the decimal number 642. This can be expressed as:

$$642 = 600 + 40 + 2$$

Note that in a multi digit decimal number (i.e., 642 in the present case), each position has a value that is 10 times the value of the next position to its immediate right. In other words, every position can be expressed as:

$$642 = 6 \times 10^2 + 4 \times 10^1 + 2 \times 10^0$$

Thus, we find that values of various positions in a decimal number system are powers of 10 i.e., equal to the number of digits used in the system. This number is called base or radix of the system. Thus, the decimal system has base of 10 (ten).

For the decimals, the digit to the extreme right is referred to as the least significant digit (LSD) because its positional value or weight is the lowest. For the decimal number 642, is the LSD. The left-most digit in the decimal number is the most significant digit (MSD) because its positional value or weight is the highest. For the decimal number 642, 6 is the MSD with a value of 600.

Binary number system. In the binary number system, only two digits (0 and 1) are used. Therefore, the base of this system is 2. In the binary number , each position has a value that is 2 times the 2 raised to some power. We know that binary number 1001 is equal to the decimal number 9. This can be readily shown as under:

$$1001 = 1\times 2^3 + 0\times 2^3 + 0\times 2^1 + 1\times 2^0 = 9$$

For binary numbers, the digit at the extreme right is referred to as least significant bit (LSB). In the binary number 1001, the 1 at the right is the LSB. The left-most digit is called the most significant bit (MSB). In the binary number 1001, the 1 at the right is the LSB. The left-most digit is called the most significant bit (MSB). In the binary number 1001, at the left is the MSB with the value of 8 in decimal terms.

Decimal to Binary Conversion

There are many methods to perform this conversion. The method described here is called double -dabble because it requires successive divisions by 2. This method can be summarised as under:

Divide progressively the decimal number by 2 and write down the remainder after each division. Continue this process till you get a quotient of 0 and remainder of 1, the conversion is now complete. The remainders, taken in reverse order; form the binary number [See Fig.26.4].

Note that 13 is first divided by 2, giving a quotient of 6 with a remainder of 1. This remainder becomes the 2^0 position in the binary number. The 6 is then divided by 2, giving a quotient of 3 with a remainder of 0. This remainder becomes the2^1 position in the binary number.

Continuing this procedure, the equivalent binary number is 1101.

Decimal number

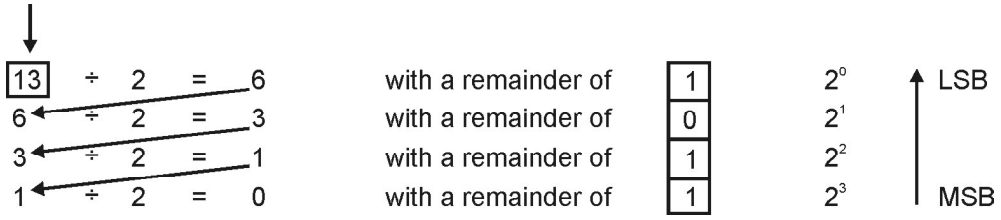


Fig. 5.36

Example 26.1. Convert the decimal number 37 to its equivalent binary number.

2	37
2	18 – 1
2	9 – 0
2	4 – 1
2	2 – 0
2	1 – 0
	0 – 1

Solution. Using double-dabble method, we find that the equivalent binary number is 100101. It is a usual practice to mention the base of the number system.

The decimal system has a base of 10 while binary system has a base of 2.

$\therefore \quad (37)_{10} = (100101)_2$

Note. This notation avoids the confusion that may arise because decimal number also involves the digits 0 and 1. Thus, $(101)_{10}$ denotes the decimal number hundred one while the binary number $(101)_2$ is equivalent to decimal number 5.

Example 26.2. Convert the decimal number 23 to its equivalent binary number.

2	23
2	11 – 1
2	5 – 1
2	2 – 1
2	1 – 0
	0 – 1

Solution. Using double-dabble method, we find that the equivalent binary number is 10111.

$\therefore \quad (23)_{10} = (10111)_2$

Note that binary number 10111 has five bits.

Binary to Decimal Conversion

Binary numbers can be converted to equivalent decimal numbers quite easily. Suppose you are given the binary number 110011. Its conversion to equivalent decimal number involves the following two steps:

i. Place the decimal value of each position of the binary number.

1	1	0	0	1	1
2^5	2^4	2^3	2^2	2^1	2^0

ii. Add all the decimal values to get the decimal number.

Thus, $(110011)_2 = 1\times 2^5 + 1\times 2^4 + 0\times 2^3 + 0\times 2^2 + 1\times 2^1 + 1\times 2^0$

$= 32 + 16 + 0 + 0 + 2 + 1 = 51$

$\therefore \quad (110011)_2 = (51)_{10}$

Note. In binary to decimal conversion, all positions containing 0 can be ignored. Only add the decimal values of the positions where 1 appears. Thus, in case of the above binary number,

$$(110011)_2 = 1\times 2^5 + 1\times 2^4 + 1\times 2^1 + 1\times 2^0$$
$$= 32 + 16 + 2 + 1 = 51$$

Example 26.3. Convert the binary number 110001 to its equivalent decimal number.

Solution. The binary number along with its decimal values of various positions is shown.

$\therefore$ $(110001)_2 = 1\times 2^5 + 1\times 2^4 + 1\times 2^0$

$= 32 + 16 + 1 = 49$

1	1	0	0	0	1
2^5	2^4	2^3	2^2	2^1	2^0

or $(110001)_2 = (49)_{10}$

Octal Number System

The octal number system has a radix of eight so that it uses eight digits : 0, 1, 2, 3, 4, 5, 6 and 7. The position weights in the system are powers of eight. The digit positions of first six powers of eight are :

$8^0 = 1$; $8^1 = 8$; $8^2 = 64$

$8^3 = 512$; $8^4 = 4096$; $8^5 = 32768$

The octal number system is frequently used in digital circuits due to two principal reasons. First, it can be easily converted to binary. Secondly, there are significantly fewer digits in any given octal number than in the corresponding binary number so that it is much easier to work with shorter octal numbers.

1. Decimal-to-Octal Conversion. To convert a decimal number to octal, we employ the same repeated-division method that we used in decimal-to-binary conversion. However, here the division factor is 8 instead of two. The following examples illustrate decimal-to-octal conversion.

i. To convert decimal number 91 to octal number, the procedure is as under:

Division	Remainder	
$91 \div 8 = 11$	3	(LSB)
$11 \div 8 = 1$	3	
$1 \div 8 = 0$	1	(MSB)

$\therefore$ $(91)_{10} = (133)_8$

ii. As another example, consider the conversion of decimal number 266 to octal number.

Division	Remainder	
$266 \div 8 = 33$	2	(LSB)
$4 \div 8 = 0$	4	(MSB)

$\therefore$ $(266)_{10} =$

2. Octal-to-Decimal Conversion. An octal-to-decimal conversion can be done in the same manner as a binary-to-decimal conversion i.e. simply add up the position weights to obtain the decimal number. The following examples illustrate octal-to-decimal conversion.

i. To convert octal number $(133)_8$ to decimal number, the procedure is as under:

Position weights	8^2	8^1	8^0
Octal number	1	3	3

$\therefore$ $(133)_8 = (8^2\times 1) + (8^1\times 3) + (8^0\times 3)$

$= 64 + 24 + 3 = 91$

$\therefore$ $(133)_8 = (91)_{10}$

ii. As another example, consider the conversion of octal number $(372)_8$ to decimal number.

Position weights	8^2	8^1	8^0
Octal number	3	7	2

$\therefore$ $(372)_8 = (8^2\times 3) + (8^1\times 7) + (8^0\times 2)$

$= 192 + 56 + 2 = 250$

$\therefore$ $(372)_8 = (250)_{10}$

3. Octal-to-Binary Conversion. The advantage of octal number system is the case with which an octal number can be converted to a binary number and vice-versa. It is because eight is the third power of two, providing a direct correlation between three-bit groups in a binary number and the octal digits i.e. each three-bit group of binary bits can be represented by one octal digit. Therefore, conversion from octal to binary is performed by converting each octal digit to its 3-bit binary equivalent. The eight digits are converted as shown in the adjoining table.

Octal and Binary Equivalents	
Octal Digit	Binary Bias
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

i. The conversion of octal number $(472)_8$ to binary number is done as under:

4	7	2
↓	↓	↓
100	111	010

Therefore, octal 472 is equivalent to binary 100111010 i.e.

$$(472)_8 = (100111010)_2$$

ii. As another example, consider the conversion of octal number $(5431)_8$ to binary number.

5	4	3	1
101	100	011	001

Therefore, octal 5431 is equivalent to binary 101100011001 i.e.

$$(5431)_8 = (101100011001)_2$$

4. Binary-to-Octal Conversion. The conversion of binary number to octal number is simply the reverse of the above process. The bits of the binary number are grouped into groups of three bits starting at the LSB. Then each group is converted to its octal equivalent. To illustrate this method, consider the conversion of binary number $(100111010)_2$ to octal number. The procedure is as under:

$\underbrace{100}$	$\underbrace{111}$	$\underbrace{010}$
↓	↓	↓
4	7	2

$\therefore \quad (10111010)_2 = (472)_8$

Note that there are fewer digits in the octal number than in the corresponding binary number. Therefore, it is much easier to work with shorter octal numbers.

Sometimes the binary number will not have even groups of 3 bits. In that case, we can add one or two 0s to the left of the MSB of the binary number to fill the last group. This point is illustrated below for the binary number 11010110.

$\underbrace{011}$	$\underbrace{010}$	$\underbrace{110}$
↓	↓	↓
3	2	6

Note that a 0 is placed to the left of the MSB to produce even groups of 3 bits.

Example 26.4. Convert the following decimal numbers to octal equivalent.

i. 76 ii. 255 iii. 372

Solution.

i.

Division	Remainder	
$76 \div 8 = 9$	4	(LSB)
$9 \div 8 = 1$	1	
$1 \div 8 = 0$	1	(MSB)

$\therefore \quad (76)_{10} = (114)_8$

ii.

Division	Remainder	
255 ÷ 8 = 31	7	(LSB)
31 ÷ 8 = 3	7	
3 ÷ 8 = 0	3	(MSB)

$\therefore \quad (255)_{10} = (377)_8$

iii.

Division	Remainder	
372 ÷ 8 = 46	4	(LSB)
46 ÷ 8 = 5	6	
5 ÷ 8 = 0	5	(MSB)

$\therefore \quad (372)_{10} = (564)_8$

Example 26.5. Converted octal number $(24.6)_8$ to the equivalent decimal number.

Solution.

$$(24.6)_8 = (2 \times 8^1) + (4 \times 8^0) + (6 \times 8^{-1})$$
$$= 16 + 4 + 0.75 = 20.75$$
$$\therefore \quad (24.6)_8 = (20.75)_{10}$$

Example 26.6. Convert $(177)_{10}$ to its 8-bit binary equivalent by first converting to octal.

Solution. We shall first convert $(177)_{10}$ to its equivalent octal number as under:

Division	Remainder	
177 ÷ 8 = 22	1	(LSB)
22 ÷ 8 = 2	6	
2 ÷ 8 = 0	2	(MSB)

$\therefore \quad (177)_{10} = (261)_8$

2	6	1
↓	↓	↓
010	110	001

Therefore, the binary equivalent is 010110001. We remove the leading zero to express the result as 8 bits.

$\therefore \quad (177)_{10} = (10110001)_2$

Hexadecimal Number System

The hexadecimal system uses a redix of 16. Therefore, it has 16 possible digit symbols. The first ten digits in the hexadecimal system are represented by the numbers 0 through 9 (0, 1, 2, 3, 4, 5, 6, 7, 8 and 9) and the letters A though F are used to represent the numbers 10, 11, 12, 13, 14 and 15 respectively. The adjoining table shows the relationships among hexadecimal, decimal and binary. Note that each hexadecimal digit represents a group of four binary digits.

As is true for binary and decimal numbers, each digit in the hexadecimal system has a positional value or weight. For the right most digit of a hex (abbreviation for hexadecimal) number, the positional weight is 16^0 (=1), the next digit to the left has a positional weight of 16^1 (=16) and so on. The positional weight distribution of a hex number system is given below:

	16^3	16^2	16^1	16^0
etc.	4096	256	16	1

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

1. Decimal-to-Hex Conversion. To convert a decimal number to hex number, the technique is the same as used for decimal-to-binary conversion or decimal-to-octal conversion. Recall that we did decimal-to-binary conversion using repeated division by 2 and decimal-to-octal conversion using repeated division by 8. Linkewise,

decimal-to-hex conversion is done using repeated division by 16. Let us illustrate the decimal-to-hex conversion procedure. Suppose we are to convert the decimal number 423 to hex number.

Division	Remainder	
423 ÷ 16 = 26	7	(LSB)
26 ÷ 16 = 1	10	
1 ÷ 16 = 0	1	(MSB)

$\therefore \quad (423)_{10} = (1\,A7)_{16}$

Note that 10 is represented by the letter A.

2. Hex-to-Decimal Conversion. In order to convert a hex number to its decimal equivalent , simply and up the position weight of each digit in the hex number. The following example illustrates this conversion.

$$(356)_{16} = (3\times16^2)+(5\times16^1)+(6\times16^0)$$
$$= 768 + 80 + 6 = 854$$
$$\therefore \quad (356)_{16} = (854)_{10}$$

3. Hex-to-Binary Conversion. The conversion from hex to binary is performed by Converting each hex digit to its 4-bit binary equivalent (See above table). The following example illustrates this point. Hence we shall convert hex number (9 F2)16 to its binary equivalent.

9	F	2
↓	↓	↓
1001	1111	0010

$\therefore \quad (9F2)_{16} = (100111110010)_2$

4. Binary-to-Hex Conversion. The conversion from binary to hex is just the reverse of the above process. The binary number is grouped into groups of four bits and each group is converted to its equivalent hex digit. The following example illustrates this point. Here, we shall convert binary number $(1110100110)_2$ to its equivalent hex number.

$\underbrace{0011}$	$\underbrace{1010}$	$\underbrace{0110}$
3	A	6

$\therefore \quad (1110100110)_2 = (3A6)_{16}$

Example 26.7. Convert decimal number 541 to hexadecimal.
Solution.

Division	Remainder	
541 ÷ 16 = 33	13	(LSB)
33 ÷ 16 = 2	1	
2 ÷ 16 = 0	2	(MSB)

$\therefore \quad (541)_{10} = (21D)_{16}$

Example 26.8. Convert decimal number 378 to a 16-bit number by first converting to hexa-decimal.
Solution.

Division	Remainder	
378 ÷ 16 = 23	10	(LSB)
23 ÷ 16 = 1	7	
1 ÷ 16 = 0	1	(MSB)

Thus $(378)_{10} = (17A)_{16}$. We can easily convert this hex number to binary 000101111010. Therefore, we can express $(378)_{10}$ as a 16-bit binary number by adding four leading 0s.

$$(378)_{10} = (0000000101111010)_2$$

Example 26.9. Convert $(B2F)_{16}$ to octal.
Solution. It is easier to first convert hex to binary and then to octal.

$(B2F)_{16}$	=	1011	0010	1111		.... conversion to binary
	=	101	100	101	111	... 3-bit groupings
	=	5	4	5	7	

$\therefore \quad (B2F)_{16} = (5457)_8$

Binary-Coded Decimal Code (BCD Code)

Circuit and machines can deal readily with binary numbers, but people are used to working with decimal numbers. Moreover, there are considerably fewer decimal digits required to represent a number than there are binary. It is much easier to remember just a few digits than it is to remember many. Thus whenever there is an interface between digital circuits and people, the interface data usually takes the decimal from. As a result, the digital circuit must utilise some binary code to conveniently represent the decimal numbers. The code used for this purpose is called BCD code. In a BCD code.

4	8	9
0100	1000	1001

Note that the highest BCD value that a 4-bit binary number could represent is 9 which would be (1001)2 in binary. Clearly, only the 4-bit binary numbers from 0000 through 1001 are used.

BCD Code code	Decimal digit
0000	0
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
1000	8
1001	9

The adjoining table shows the BCD code. Each of the decimal digits (0 through 9) is represented by its binary equivalent. Since a decimal digit can be as large as 9, four bits are required to code each decimal digit (the bianry code for 9 is 1001).

Note that each decimal digit is assigned a 4-bit binary number even though the binary equivalent binary require fewer than four binary places. This way, circuits which use BCD always handle the string of binary bits in four- place groups. When using BCD code, remember that all zeros must be retained, unlike when it is necessary to transfer decimal information into and out of a digital machine. Example of digital machines include the digital clocks, calculators, digital voltmeters and frequency counters.

Example 26.10. What decimal number is represented by the BCD string given below?

0100 0000 0010

Solution. Divide the BCD number into 4-bit groups and convert each to decimal.

0100	0000	0010
4	0	2

Therefore, the equivalent decimal number is $(402)_{10}$.

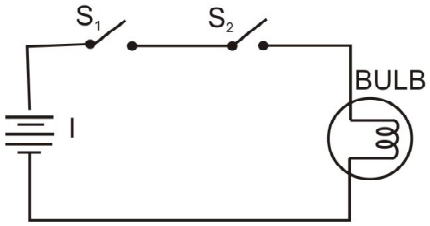
Note. To avoid confusion between BCD and true binary, a BCD string is often separated into groups of 4 binary bits or a subscript BCD is sometimes attached to the string as illustrated under:

0100 0000 0010 or 010000000010_{BCD}

Logic Gates

A digital circuit with one or more input signals but only one output signal is called a logic gate.

Since a logic gate is a switching circuit (i.e. a digital circuit), its output can have only one of the two possible states viz., either a high voltage (1) or a low voltage (0) – it is either ON or OFF. Whether the output voltage of a logic gate is high (1) or low (0) will depend upon the conditions at its input. Fig.26.5 shows the basic idea of a logic gate using switches.



(i)

S_1	S_2	BULB
open	open	OFF
open	open	OFF
closed	open	OFF
closed	closed	ON

(ii)

Truth Table

S_1	S_2	Output
0	0	0
0	1	0
1	0	0
1	1	1

(iii)

Fig. 5.37

i. When S_1 and S_2 are open, the bulb is OFF.
ii. When S_1 is open and S_2 closed, the bulb is OFF.
iii. When S_2 is open and S_1 closed, the bulb is OFF.
iv. When both S_1 and S_2 are closed, the bulb is ON.

Note that output (OFF or ON) depends upon the conditions at the input.

The four possible combinations of switches S_1 and S_2 are shown in the table on the previous page. It is clear that when either of the switches (S_1 or S_2) or both are open, the bulb is OFF. In binary language, when either of the inputs or both the inputs are low (0), the output is low. When both switches are closed, the bulb is ON. In terms of binary language, when both the inputs are high (1), the output is high. It is usual practice to show the conditions at the input and output of a logic gate in the binary form as shown in the table on the previous page. Such a table is called truth table.

The term "logic" is usually used to refer to a decision-making process. A logic gate makes logical decisions regarding the existence of output depending upon the nature of the input. Hence, such circuits are called logic circuits.

Three Basic Logic Gates

A logic gate is a circuit that has one or more input signals but only one output signal. All logic gates can be analysed by constructing a truth table. A truth table lists all input possibilities and the corresponding output for each input. The three basic logic gates that make up all digital circuits are (i) OR gate (ii) AND gate and (iii) NOT gate. We shall first discuss these three basic logic gates and then the combination of these gates. The following points may be noted about logic gates"

i. A binary 0 represents 0 V and binary 1 represents +5V. It is common to refer to binary 0 as LOW input or output and binary 1 as HIGH input or output.

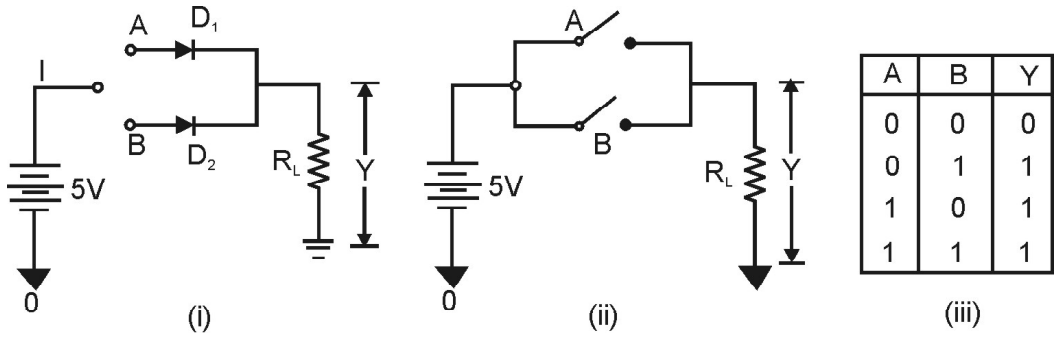
ii. A logic gate has only one output signal. The output will upon the input signal/signals and the type of gate.

iii. The operation of a logic gate may be described either by truth table or Boolean algebra.

OR Gate

An OR gate is a logic gate that has two or more inputs but only one output. However, the output y of an OR gate is LOW when all inputs are LOW. The ouput Y of an OR gate is HIGH if any or all the inputs are HIGH. It is called OR gate because the ouput is high if any or all the inputs are high. For the same reason, an OR gate is sometimes called "any or all gate". For example, consider a 2-input OR gate. The output Y will be high if either or both inputs are high.

OR gate operation, Fig.26.6 (i) shows one way to build a 2-input OR gate while Fig.26.6 (ii) shows its simplified schematic diagram. The input voltages are labeled as A and B while the ouput voltage is Y. Note that negative terminal of the battery is grounded and corresponds to 0 state (LOW level). The positive terminal of the battery (+5V) corresponds to 1 state (HIGH level). There are only four input-output possibilities.



A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(iii)

Fig. 5.38

i. When both A and B are connected to ground, both diodes are non-conducting. Hence, the ouput voltage is ideally zero (low voltage). In terms of binary, when A = 0 and B = 0, then Y = 0 as shown in the truth table in fig.26.6 (iii).

ii. When A is connected to ground and b connected to the positive terminal of the battery, diode D_2 is forward biased and diode D_1 is non-conducting. Therefore, diode D_2 conducts and the output voltage is ideally +5V. In terms of binary, when A = 0 and B = 1, then Y = 1 [See Fig.26.6 (iii)].

iii. When A is connected to the positive terminal of the battery and B to the ground, diode D_1 is on and diode D_2 is off. A gain the ouput voltage is +5V. In binary terms, when A = 1 and B = 0, then Y = 1 [See Fi.g26.6(iii)].

iv. When both A and B are connected to the positive terminal of the battery, both diodes are on. Since the diodes are in parallel, the output voltage is +5V. In binary terms, when A = 1 and B = 0, then Y = 1 [See Fig.26.6(iii)].

It is clear from the truth table that for OR gate, the output is high if any or all of the inputs are high. The only way to get a low output is by having all inputs low. Fig.26.7 shows the logic symbol of OR gate. Note that the symbol has curved line at the input.

A+B	=	Y
0+ 0	=	0
0+ 1	=	1
1+ 0	=	1
1+ 1	=	1

Fig. 5.39

Boolean expression. The algebra used to symbolically describe logic functions is called Boolean algebra. The "+" sign in Boolean algebra refers to the logical OR function. The Boolean expression for OR function is

A + B = Y
↑
OR symbol

The adjoining table shows possibilities for the inputs. According to this table, when 0 is ORed with 0, the result equals 0. Also any variable ORed with 1 equals 1. The OR function can be summed up as under

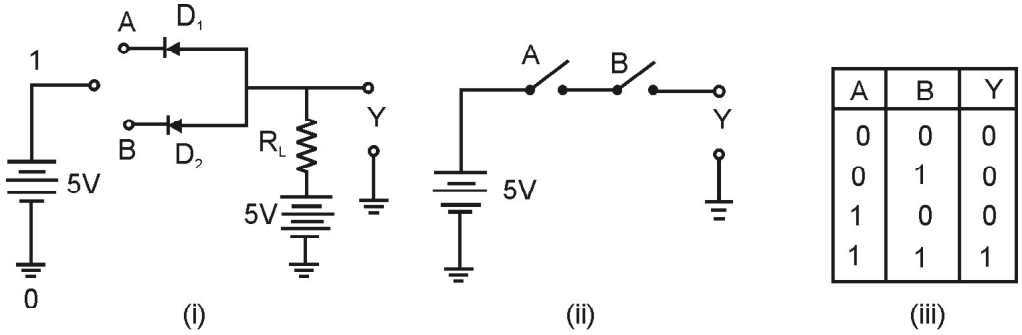
0	ORed with 0 equals	0
0	ORed with 1 equals	1
1	ORed with 1 equals	1

AND Gate

The AND gate is a logic gate that has two or more inputs but only one output. The output Y of AND gate is HIGH when all inputs are HIGH. However, the output Y of AND gate is LOW if any or all inputs are LOW.

It is called AND gate because output is HIGH only when all the inputs are HIGH. For this reason, the AND gate is sometimes called "all or nothing gate". For example, consider a 2-input AND gate. The output will be HIGH when both the inputs are HIGH.

AND gate operation. Fig.26.8(i) shows one way to build a 2-input AND gate while Fig.26.8 (ii) shows its simplified schematic diagram. There are only four input-output possibilities.



A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(iii)

Fig. 5.40

i. When both A and B are connected to ground, both the diodes (D_1 and D_2) are forward biased and hence they conduct current. Consequently, the two diodes are grounded and output voltage is zero. In terms of binary, when A = 0 and B = 0, then y = 0 as shown in truth table in Fig.26.8 (iii).

ii. When A is connected to the ground and B connected to the positive terminal of the battery, diode D_1 is forward biased while diode D_2 will not conduct. Therefore, diode D_1 conducts and is grounded. Again output voltage will be zero. In binary terms, when A = 0 and B = 1, then Y = 0. This fact is shown in the truth table.

iii. When B is connected to the ground and A connected to the positive terminal of the battery, the roles of diodes are interchanged. Now diode D_2 will conduct while diode D_1 does not conduct. B = 0, then Y = 0. This fact is indicated in the truth table.

iv. When both A and B are connected to the positive terminal of the battery, both the diodes do not conduct. Now, the output voltage is +5V because there is no current though R_L.

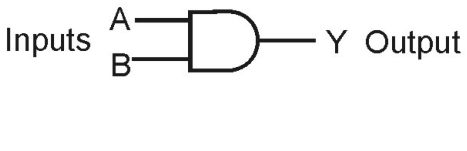
It is clear from the truth table that for AND gate, the output is high if all the inputs are high. However, the output is low if any or all inputs are low. Fig.26.9 shows the logic symbol of AND gate. This is the symbol you should memorise and use from now on for AND gates.

Boolean expression. The Boolean expression for AND function is

$$A \cdot B = Y$$

↑

AND symbol



A · B	=	Y
0 · 0	=	0
0 · 1	=	1
1 · 0	=	1
1 · 1	=	1

Fig. 5.41

where the multiplication dot stands for the AND operation. The adjoining table shows the possibilities for the inputs. Table tells us that 0 ANDed with any variable equals 0. Also, 1 ANDed with 1 equals one. The AND function can be summed up as under:

0 ANDed with 0 equals 0

0 ANDed with 1 equals 0

1 ANDed with 1 equals 1

NOT Gate or Inverter

The NOT gate or inverter is the simplest of all logic gates. It has only one input and output, where the output is opposite of the input. The NOT gate is often called inverter because it inverts the input.

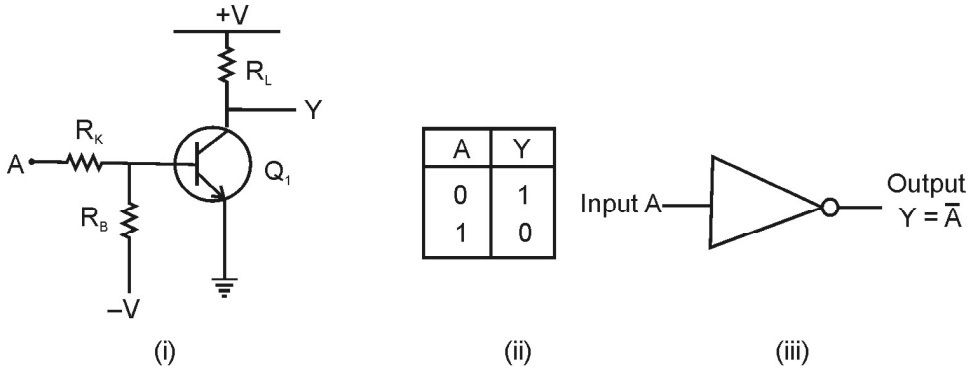


Fig. 5.42

Figure 26.10(i) shows a typical inverter circuit. When A is connected to ground, the base of transistor Q_1 will become negative . This negative potential causes the transistor to cut off and collector current is zero and ouput is +V volts. In binary terms, when A = 0, Y = 1. If sufficiently large positive voltage is applied at A, the base of the transistor will become positive, causing the transistor to conduct heavily. Therefore, the output voltage is zero. In binary terms, when A = 1, Y = 0. Fig. 26.10(ii) shows truth table for an inverter. It is clear from the truth table that whatever the input to the inverter, the output assumes opposite polarity. If the input is 0, the output will the output will be 1; if the input is 1, the output will be 0.

Figure 26.10 (iii) shows the logic symbol for NOT gate or inverter. Note that small bubble on the inverter symbol represents inversion. The Boolean expression for NOT function is

$$Y = \bar{A}$$

Note that bar above the input A represents inversion.

If A = 0, then $Y = \bar{0}$ or Y =1.

If A = 1, then $Y = \bar{1}$ or Y = 0.

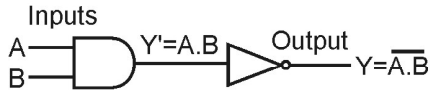
Combination of Basic Logic Gates

The OR, AND and NOT gates are the three basic circuits that make up all digital details circuits. We shall discuss a few combinations of these basic circuits.

i. **NAND gate**. It is a combination of AND gate and NOT gate. In other words, ouput of AND gate is connected to the input of a not gate as shown in Fig.26.11(i). Clearly, the ouput of a NAND gate is opposite to the AND gate. This is illustrated in the truth table for NAND gate. Note that truth table for NAND gate is developed by inverting the outputs of the AND gate.
The Boolean expression for NAND function is

$$Y = \overline{A \cdot B}$$

This Boolean expression can be read as Y = not A.B. To perform the Boolean algebra operation, first the inputs must be ANDed and then the inversion is performed. Note that output from a NAND gate is always I except when all of the inputs are 1. Fig.26.11(iii) shows the logic symbols for a NAND gate. The title bubble (small circle) on the right end of the symbol means to invert the AND.



(i)

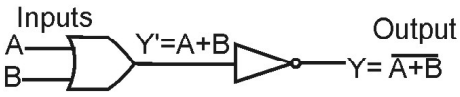
Input		Output	
A	B	AND (Y')	NAND (Y)
0	0	0	1
1	0	0	1
0	1	0	1
1	1	1	0

(ii)

(iii)

Fig. 5.43

ii. **NOR gate**. It is a combination of OR gate and NOT gate. In other words, ouput of OR gate is connected to the input of a NOT gate as shown in Fig.26.12(i). Note that ouput of OR gate is inverted to form NOT gate. This is illustrated in the truth table for NOR gate. It is clear that truth table for NOR gate is developed by inverting the outputs of the OR gate.



Inputs		Output	
A	B	OR(Y')	NOR(Y)
0	0	0	1
1	0	1	0
0	1	1	0
1	1	1	0

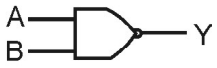


Fig. 5.44

The Boolean expression for NOR function is

$$Y = \overline{A + B}$$

This Boolean expression can be read as Y = not A or B. To perform the Boolean algebra operation, first the inputs must be ORed and then the inversion is performed. Note that output from a NOR gate is high (1) only when all the inputs are low (0). If any of the inputs is high (1), the output is low (0). Fig.26.12 (iii) shows the logic symbols for a NOR gate. The bubble (small circle at the Y output indicate inversion.

NAND Gate as a Universal Gate

The NAND gate is universal gate because its repeated use can produce other logic gates. The table below shows how NAND gates can be connected to produce inverter (i.e., NOT gate), AND gate and OR gate.

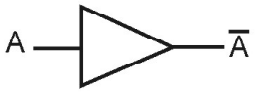
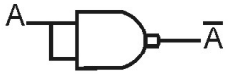
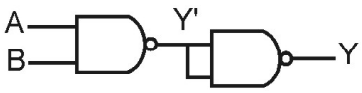
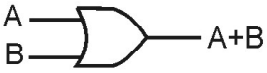
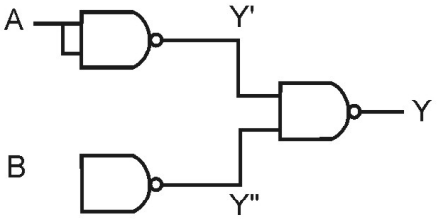
Little Function	Symbol	Circuit using NAND gate only
Inverter		
AND		
OR		

Fig. 5.45

i. NOT gate from NAND gate. When two inputs of NAND gate are joined together so that it has one input, the resulting circuit is NOT gate. The truth table also shows this fact.

A	B (=A)	Y
0	0	1
1	1	0

A	B	Y'	Y"
0	0	1	0
1	0	1	0
0	1	1	0
1	1	0	1

ii. AND gate from NAND gates. For this purpose, we use two NAND gates in a manner as shown above. The ouput of first NAND gate is given to the second NAND gate acting as inverter (i.e., inputs of NAND gate joined). The resulting circuit is the AND gate. The output Y of first NAND gate (AND gate followed by NOT gate) is inverted output of AND gate.

The second NAND gate acting as inverter further inverts it so that the final output Y is that of AND gate. The truth table also shows this fact.

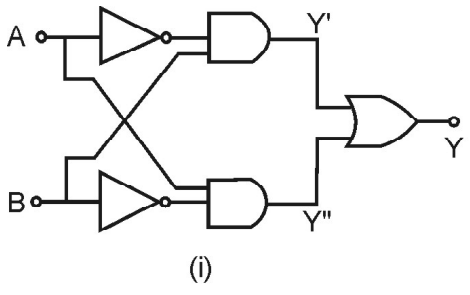
A	B	Y' = $\overline{A}$	Y" = $\overline{B}$	Y
0	0	1	1	0
1	0	0	1	1
0	1	1	0	1
1	1	0	0	1

iii. OR gate from NAND gates. For this purpose, we use three NAND gates in a manner as shown above. The first two NAND gates are operated as NOT gates and their outputs are fed to the third. The resulting circuit is OR gate. This fact is also indicated by the truth table.

Exclusive OR Gate

The name exclusive OR gate is usually shortened to XOR gate. The XOR gate can be obtained by using OR, AND and NOT gates as shown in Fig.26.14(i).

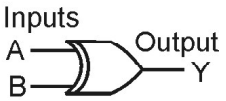
Fig. 26.14 (ii) shows the truth table for XOR gate. The table shows that the output is HIGH (1) if any but not all of the inputs are HIGH (1). This exclusive feature climinates the similarity to the OR gate. The OR gate truth table is also given so that you can compare the OR gate truth table with XOR gate truth table. The logic symbol for XOR gate is shown in Fig.26.14(iii). Note that the symbol is similar that of OR gate except for the additional curved line at the input side.



(i)

Inputs		Output	
A	B	OR	XOR
0	0	0	0
1	0	1	1
0	1	1	1
1	1	1	0

(ii)



(iii)

Fig. 5.46

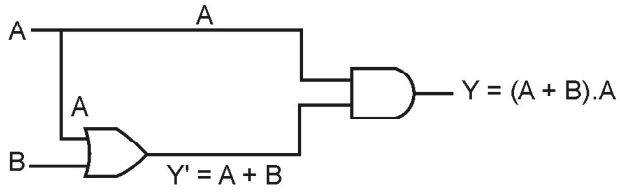
The logic operations in the circuit are as under:

A	B	$\bar{A}$	$\bar{B}$	$\bar{A}\cdot B = Y'$	$A\cdot\bar{B} = Y''$	$Y = Y' + Y''$
0	0	1	1	0	0	0
1	0	0	1	0	1	1
0	1	1	0	1	0	1
1	1	0	0	0	0	0

Note that 0 ANDed with 1 is 0 and 1 ANDed with 1 is 1.

Example 26.11. Obtain the truth table for the circuit shown in fig.26.15 (i).

Solution. Figure 26.15(ii) shows the truth table for the circuit. The truth table can be obtained very easily if the reader remembers the following simple Boolean operations.



(i)

A	B	Y'=A+B	Y=Y'.A
0	0	0	0
1	0	1	1
0	1	1	0
1	1	1	1

(ii)

Fig. 5.47

i. 0 ORed with 0 = 0; 1 ORed with 1 = 1 ; 1 ORed with 0 = 1
ii. 0 ANDed with 0 = 0 ; 0 ANDed with 1 = 0 ; 1 ANDed with 1 = 1

Thus, when A = 0 and B = 0, then A ORed with B = 0 i.e., Y' = 0. When Y' (=0) is ANDed with A (=0), the result is 0. Again when A = 1 and B = 0, then A ORed with B is 1 i.e., Y' = 1. Now Y' (=1) ANDed with A (=1), the result is 1.

Example 26.12. Obtain the truth table for the circuit shown in Fig.26.16.

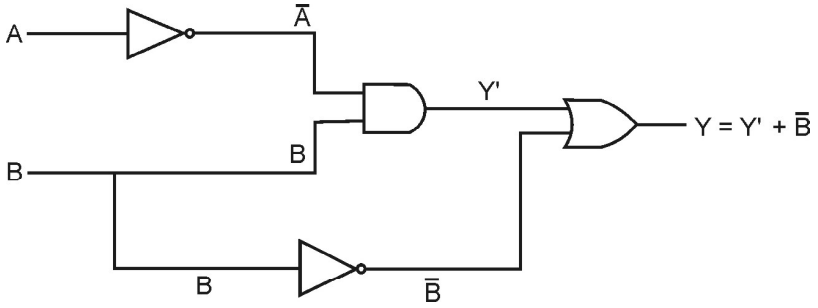


Fig. 26.16

Solution. The truth table for the circuit is shown below:

A	B	$\bar{A}$	$Y' = \bar{A} \cdot B$	$\bar{B}$	$Y = Y' + \bar{B}$
0	0	1	0	1	1
1	0	0	0	1	1
0	1	1	1	0	1
1	1	0	0	0	0

i. When A = 0 and B = 0, then $\bar{A}$ =1. Now Y' is equal to $\bar{A}$ (=1) ANDed with B (=0). The result is 0. Then Y' (=0) ORed with $\bar{B}$ (=1) is 1 i.e. Y = 1.

ii. When A = 1 and B = 0, then $\bar{A}$ = 0. Now Y' is equal to $\bar{A}$ (=0) ANDed with B (=0) and the result is 0 i.e, Y' (=0) ORed with $\bar{B}$ (=1) is 1 i.e., Y = 1.

The reader can proceed in a similar way to find the other output values.

■■■

UNIT - 5

..

CHAPTER - 6
COMMUNICATION SYSTEMS

BLOCK DIAGRAM OF COMMUNICATION SYSTEM

A communication system may be described by the block diagram as shown in Fig. 6.1. The basic elements of a communication system are information source, input transducer, transmitter, transmission channel, receiver and destination.

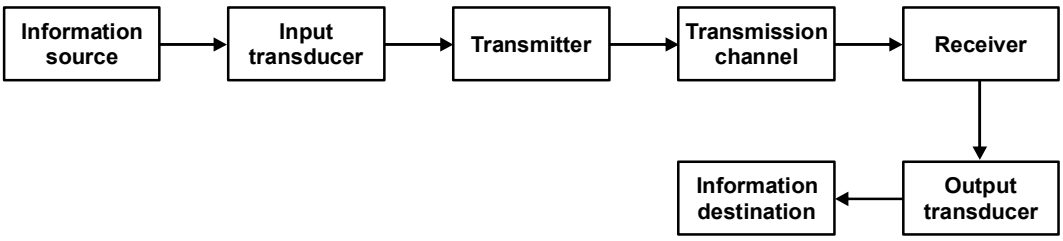


Fig. 6.1. Block Diagram of a Communication System.

Information Source

Originates the messages or information in the form of audio, image, text, data, codes etc. However out of these messages, only the desired message is selected and communicated through the transmission.

Input Transducer

converts source information into electrical signal (incase if the information source is not in the form of electrical signal). It may use microphone, camera, keyboard etc.

Transmitter

It processes the incoming information so as to make it suitable for onward transmission and subsequent reception. The transmitter consists of a number of electronic components e.g. oscillator modulator, amplifier, antenna etc.

Transmission Channel

It is the electrical medium through which the information (message) travels from the transmitter to the receiver. It provides a path (physical connection) or free space for modulated wave to travel between transmitter and receiver. The transmission can be point-to-point or in a broadcast mode. The medium can be one of the following types:

- Wired Lines (twisted pair, co-axial cable, waveguide, fibre optic cable, LASER beam etc.) type
- Wireless / Radio (Microwave, satellite, etc.) type

'Notice' is an unwanted but unavoidable electrical signal which may introduce at any point in the communication system. It should be kept at minimum. It is most noticeable when it occurs in the channel or at the input of the receiver. When the level of noise is too high it makes the information useless and reduces the quality of the system. A good system should have a high signal to noise ratio.

Receiver

It receives the transmitted signal and performs a process which is reverse of transmitter stage. The transmitted signal is received through antenna and then processed to get the original signal. The electronic components of receiver are antenna, filter, amplifier, demodulator etc.

Output Transducer

It converts the electrical signal at its input into the form desired by the system user. It may use loud speaker, personal computer (PC), tape recorder etc.

Information Destination

At last the original information is received at this destination without any noise or distortion.

Figure 6.1 represents one-way (simplex) transmission. Two way communication of course requires a transmitter and receiver at each end. A full duplex system has a channel that allows simultaneous transmission in both directions. A half-duplex system allows transmission in either direction but not at the same time.

IEEE SPECTRUM FOR COMMUNICATION SYSTEMS

Radio waves are a type of Electromagnetic (EM) radiation with wavelengths in the electromagnetic spectrum longer than infrared light. Radio waves have frequencies from 3 KHz to 300 GHz, which corresponds to the wavelength from 100 km to 1 mm. Like all other electromagnetic waves, they travel at the speed of light. Radio waves are used for mobile radio communication, broadcasting, RADAR and other navigational systems, satellite communication, computer network and many other applications. Different frequencies of radio waves have different propagation characteristic in the earth's atmosphere. Radio spectrum is typically government regulated in developed countries and in some cases sold or licensed to operators of private radio transmission systems for example mobile phones and broadcast television stations.

A frequency band is a small section of the spectrum of radio communication frequencies, in which channels are usually used or set aside for the same purpose. Above 300 GHz the absorption of EM radiation by earth's atmosphere is so great that the atmosphere is effectively opaque. To convert from frequency (f) to wavelength (λ) & vice versa, recall that $f = c/\lambda$ or $\lambda = c/f$, where c = speed of light.

$$\lambda_{meter} = \frac{3\times10^8}{f_{Hz}} = \frac{3\times10^5}{f_{KHz}} = \frac{300}{f_{MHz}} = \frac{0.3}{f_{GHz}}$$

or

$$f_{Hz} = \frac{3\times10^8}{\lambda_{meter}},\ f_{KHz} = \frac{3\times10^5}{\lambda_{meter}},\ f_{MHz} = \frac{300}{\lambda_{meter}},\ f_{GHz} = \frac{0.3}{\lambda_{meter}}$$

For example : at 10 GHz, the wavelength = 30/10 = 3 cm.

Military RADAR band nomenclature (L, S, C, X, Ku, K and K_a bands) originated during word war-II as a secret code so scientists and engineers could talk about frequencies without divulging them. After the war the codes were declassified, millimeter (mm) was added and the designation were eventually were eventually adopted by the IEEE. (Institute of Electrical and Electronic Engineers).

The IEEE adopted nomenclature for various bands is shown in table 6.1. The same is used in military RADAR, satellite and terrestrial communications.

Table 6.1 : IEEE RADAR Band Designations

Band	Frequency Range	Origin of Name
HF	3 to 30 MHz	High Frequency
VHF	30 to 300 MHz	Very High Frequency
UHF	300 to 1000 MHz	Ultra High Frequency
L	1 to 2 GHz	Long wave
S	2 to 4 GHz	Short wave
C	4 to 8 GHz	Compromise between S and X
X	8 to 12 GHz	X for cross (used in world war-II for fire control)
Ku	12 to 18 GHz	Kurz - under
K	18 to 27 GHz	German Kurz (short)
Ka	27 to 40 GHz	Kurz - above
V	40 to 75 GHz	-
W	75 to 110 GHz	W follows V in the alphabet
mm	110 to 300 GHz	-

The IEEE system in widely used, but lacks some fine granularity. In that it covers maximum of RADAR frequencies bands i.e. starting from 1 GHz. The designations below for the lower frequencies come from ITU (International Telecommunication Union) frequency bands which are more useful for radio applications. These are shown in Table 6.2.

Table 6.2 : ITU Radio Band Designations

Band	Frequency Range	Origin of Name
VLF	3 to 30 KHz	Very Low Frequency
LF	30 to 300 KHz	Low Frequency
MF	300 to 3000 KHz	Medium Frequency

TYPES OF COMMUNICATION

Based on the type of modulation scheme and the nature of the output of the information sources, the communication systems are divided into two categories :

1. Analog communication, and
2. Digital communication

Other ways of categorizing communication systems include the classification based on the frequency of the carrier and the nature of the communication channel.

Analog Communication

Analog communication systems are designed to transmit analog information using analog modulation methods. The block diagram of an analog communication system is shown in Fig. 6.2.

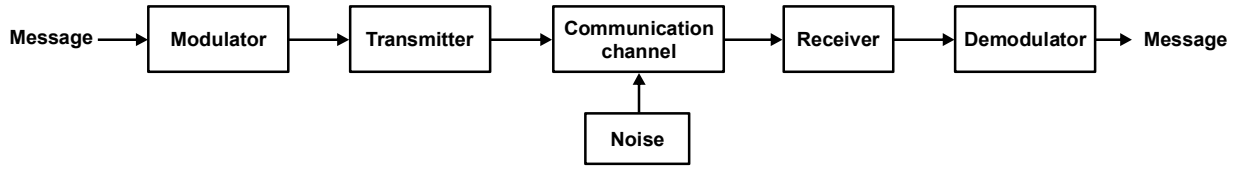


Fig. 6.2. Block Diagram of Analog Communication System.

These include the modulator which produces a signal that is transmitted over the communication channel through the transmitter and on the other side the receiver receives the signal and the demodulator reconstructs the original analog signal. The noise is an unavoidable random signal, which is entirely undesirable in nature. This electrical energy invariably enters any communication system through the electronic components present in the transmitter and receiver circuit. They are also present in the medium itself. Noise is one of the most serious problems to be handled by any communication system and is very crucial in designing the system. Methods are devised to minimize noise but it can not be eliminated fully.

Advantages of Analog Communications

1. Transmitter and receivers are simple
2. Requires smaller bandwidth
3. Frequency Division Multiplexing (FDM) can be used
4. Synchronization problem is relatively easier.

Disadvantages of Analog Communications

1. Repeaters can not be used between transmitters and receivers
2. Not suitable for secret informations
3. Coding is not possible
4. Noise affects the quality of signal
5. No error correction capability
6. Noise and signals are inseparable

Applications

1. Radio broadcasting (AM, FM)
2. T.V. broadcasting
3. Telephones

Digital Communication

Digital communication systems are designed to transmit digital information using digital modulation schemes. Figure 6.3 shows the block diagram of a digital communication system.

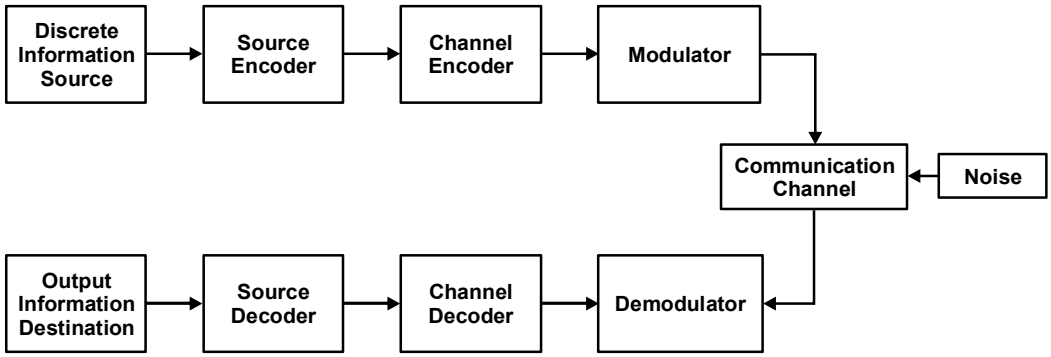


Fig. 6.3. Block Diagram of Digital Communication System

The main purpose of the system is to transmit the message coming out of a source to the destination at a high rate and accuracy as possible. The source and destination points are physically separated in space and communication channel of some sort connects the source to the destination point. The main function of the encoder, modulator, demodulator and the decoder is to combat the degrading effects of the channel on the signal and maximize the information rate and accuracy.

Desecrate information source are characterized by the parameters viz. source alphabet (letters/symbols), symbol rate, source alphabet probabilities and probabilistic dependance of symbols in a sequence. The source encoder converts the symbol sequence into a binary sequence of 0's and 1's by assigning code words to the symbols in the input sequence. Modulation is a reversible process. The extraction of the message from the information bearing waveform produced by the modulator is accomplished by the demodulator.

Advantages of Digital Communications

1. Repeaters can be used between transmitter and receiver to regenerate the digital signal.
2. Digital circuits are simpler and in expensive.
3. Error detection and correction is possible by coding.
4. Can merge different types of data (Audio, Video, Text).
5. Time Division Multiplexing (TDM) is possible.
6. Privacy is maintained.
7. Advanced data processing techniques can be used like digital signal processing, image processing and data compression etc.
8. Better noise immunity.

Disadvantages of Digital Communications

1. The bit rates are very high.
2. Larger bandwidth.
3. Synchronization problem is complicated.

Applications

1. Long range ground to space communication
2. Military communication for coding advantage.
3. Telephones lines
4. Satellite communication
5. Computer and data communications

MODULATION AND DEMODULATION

Modulation

Modulation is the process of altering some characteristics of the carrier wave (such as amplitude, frequency or phase) in accordance with the instantaneous value of the modulating wave. The process is more clear in Fig. 6.4.

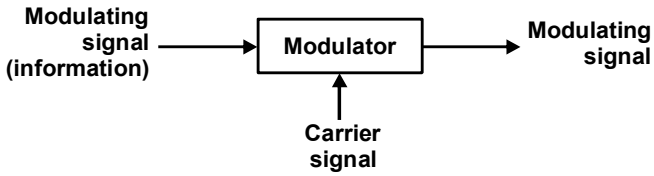


Fig. 6.4. Modulation Process.

In the modulation process two signals are used namely the modulating signal and the carrier signal. The modulating signal is the based and signal which contains the information or intelligence, while the carrier is a high frequency, constant amplitude, constant frequency and non-interrupted wave generated by radio-frequency oscillators. The carrier wave infect acts as a carrier which carries the information signal (modulated signal) from transmitter to the receiver. The modulator combines the two signals to from the modulated signal.

Demodulation

Demodulation or detection is the process of extracting the original information signal from the modulated carrier wave. This is a reverse process of modulations. It is carried out for the following reasons

1. The transmitted modulated wave consists of RF component mixed with AF signal. If it is fed directly to the loud speaker it will not produce any sound as the RF consists of vary high frequency and the loud speaker can not respond to such frequencies due to large inertia of their vibrating discs etc.
2. Such RF wave will also not produce any effect on human ears as its frequency is much higher than the audible range of human ears i.e., 20 Hz to 20 KHz.

   Therefore it is essential to separate the AF signal from RF modulated carrier wave.

NEED FOR MODULATION

Low frequency signals can not be transmitted over a long range directly without modulation. The main factors responsible for the need of modulation are :

1. **Small Operating Range**

   The energy of any wave depends on its frequency, the larger the frequency, the greater the energy contained by it. An audio signal having small frequency and hence a small power can not travel a long distance if radiated directly into the space. Therefore a modulated wave is needed to send the signal over a longer range.
2. **Huge Antenna size**

   The size of the antenna required for an efficient transmission (or reception) is at least $\lambda/4$ i.e., one quarter wavelength.

   i.e. $\ell = \frac{\lambda}{4} = \frac{1}{4} \cdot \frac{(c)}{(f)}$ where c = 3 x 10^8 m/sec. and f = Frequency in Hz

   Let the frequency of transmitted signal is 15 KHz, then an antenna of length 15 km will be required which is practically not possible. On the other hand if the transmitted signal is of frequency 15 MHz then the size of the antenna required will only be about 15 meters, which is quite easily constructed.
3. **Poor Radiation Efficiency**

   The radiation efficiency at audio frequencies is very poor and the transmission is not practicable. But at high frequencies (above 20 KHz) the energy can be radiated efficiently thus is the need of modulation.
4. **Mutual Interference**

   The audio frequency range is from 20 Hz to 20 KHz. If this signal is transmitted directly from different sources then all of these will mix up and will not serve the purpose. Therefore different signals can be transmitted at different carrier frequency through modulation to get the transmission without interference.

   Hence, the only solution is to transmit the low-frequency signal with the help of a high frequency carrier wave using modulation, to avoid the factors mentioned above.

TYPES OF MODULATION

Modulation is a process of mixing a signal with a sinusoid (sine wave) to produce a new signal. This new signal has many advantages over un-modulated signal during transmission. A general function for a sinusoid :

$$f(t) = A \cdot \sin(\omega t + \phi)$$

Here we can see that this sinusoid has 3 parameters, that can be altered to affect the shape of the graph. These parameters are :

- A, the amplitude or magnitude of the sinusoid
- ω, the angular frequency (in radian per second)
- ϕ, the phase angle

Depending upon these three parameters we have three types of modulations i.e., Amplitude Modulation (AM), Frequency Modulation (FM) and Phase Modulation (PM).

i. In **Amplitude Modulation** the amplitude of the carrier signal changed (modulated) in proportion to the message signal while the frequency and phase are kept constant.
ii. In **Frequency Modulation** the frequency of the carrier signal is changed (modulated) in proportion to the message signal while the amplitude and phase are kept constant.
iii. In **Phase Modulation** the phase of the carrier signal is varied in accordance to the low frequency of the message signal keeping the amplitude and frequency unchanged.

In India radio broadcasting using AM as well as FM while for TV transmission AM is used for video signal and FM is used for audio signals. Therefore our discussion in this chapter will be limited to AM and FM only.

Another way of classification of modulation is according to the nature of carrier wave used. As per this it can be either continuous wave modulation, and pulse modulation.

i. **Continuous Wave Modulation**
When the carrier wave is continuous in nature then this modulation is called continuous **wave modulation** or **analog modulation**. AM, FM, PM fall under this category.

ii. **Pulse Modulation**
When the carrier wave is a pulse type waveform then this modulation is called pulse modulation. Here the carrier wave is a periodic sequence of rectangular or square pulses. Pulse modulation can be of two types analog and digital.
- In analog pulse modulation the amplitudes, duration or position of the pulse varies in accordance with the message signal. Here it is of three types, i.e., Pulse Amplitude Modulation (PAM), Pulse Duration Modulation (PDM) and Pulse Position Modulation (PPM).
- In digital type of pulse modulation we have Pulse Code Modulation (PCM).

AMPLITUDE MODULATION (AM)

Amplitude Modulation

Amplitude Modulation is the process of changing amplitude of the high frequency carrier wave in accordance with the instantaneous value of modulating signal, keeping the frequency and phase unchanged. In this process the modulating signal (information) is superimposed upon the RF carrier waves with the help of modulator. Fig. 6.5 shows the AM waveform for sine wave and Fig. 6.6 shows the details of amplitude modulated carrier wave. Here the amplitudes of both positive and negative half cycles are varied according to the information signal. The RF carrier wave consists of sine wave where amplitudes follow the amplitude variations of the modulating wave. The carrier is contained in an envelope formed by the modulating wave. Thus the amplitude variations of the carrier wave are at the signal frequency and the frequency of AM is equal to the frequency of carrier wave.

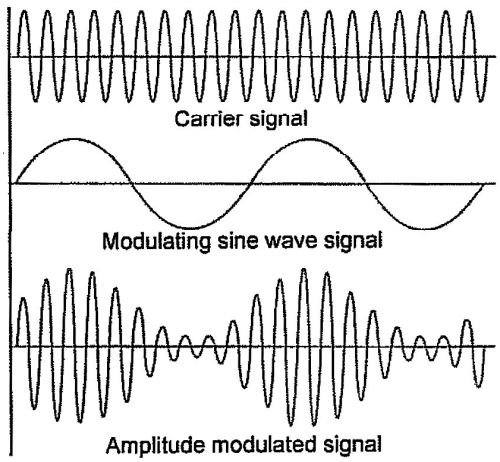


Fig. 6.5. AM waveform for sine wave.

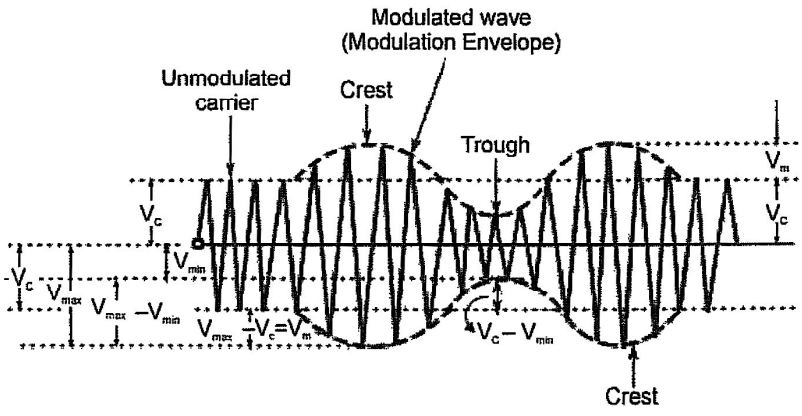


Fig. 6.6. Amplitude modulated carrier wave.

Analysis and Frequency Spectrum of Amplitude Modulated Carrier Wave

Let the carrier and modulating wave be represented by

$$c(t) = V_c \cos \omega_c t \quad ...(i)$$

and $$x(t) = V_m \cos \omega_m t \quad ...(ii)$$

We know that general expression for AM signal is

$$s(t) = [V_c + x(t)] \cos \omega_c t$$

$$= V_c \cos \omega_c t + x(t) \cos \omega_c t \quad ...(iii)$$

By putting the value of x(t) from equation (ii) we get

$$s(t) = V_c \cos \omega_c t + V_m \cos \omega_m t \cos \omega_c t$$

$$= V_c \cos \omega_c t + 0 V_m \cos \omega_c t \cos \omega_m t$$

$$= V_c \cos \omega_c t \left[1 + \frac{V_m}{V_c} \cos \omega_m t \right] \quad ...(iv)$$

We know that the modulation index for AM is given by

$$m_a = \frac{|x(t)|_{max}}{V_c} \quad ...(v)$$

Where $|x(t)|_{max}$ is the maximum amplitude of modulating signal and V_c is the maximum amplitude of carrier signal.

Hence $$|x(t)|_{max} = V_m$$

$$m_a = \frac{V_m}{V_c}$$

Putting this value in equation (4) we get

$$s(t) = V_c \cos \omega_c t \ [1 + m_a \cos \omega_m t] \quad ...(vi)$$

$$= V_c \cos \omega_c t + V_c m_a \cos \omega_c t \cdot \cos \omega_m t$$

$$= V_c \cos \omega_c t + \frac{V_c \cdot m_a}{2} [2 \cos \omega_c t \cdot \cos \omega_m t]$$

$$= V_c \cos \omega_c t + \frac{V_c \cdot m_a}{2} [\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$$

$$= V_c \cos \omega_c t + \frac{V_c \cdot m_a}{2} \cos(\omega_c + \omega_m)t + \frac{V_c \cdot m_a}{2} \cos(\omega_c - \omega_m)t] \quad ...(vii)$$

Equation (vii) reveals that the AM signal has three frequency components as given below:

i. Carrier frequency ω_c having amplitude V_c.

ii. Upper sideband $(\omega_c + \omega_m)$ having amplitude $\frac{m_a \cdot V_c}{2}$.

iii. Lower side band $(\omega_c - \omega_m)$ having amplitude $\frac{m_a \cdot V_c}{2}$.

The frequency spectrum of AM is shown in Fig. 6.7.

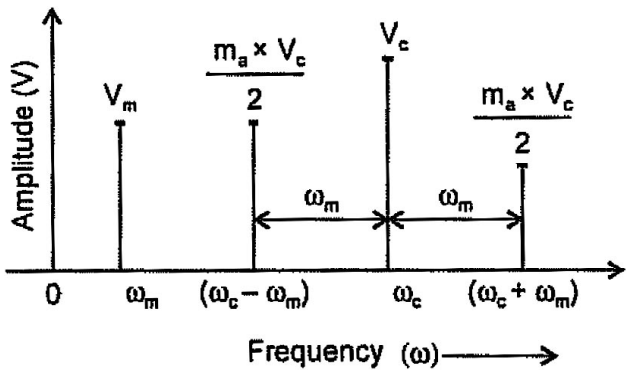


Fig. 6.7. Frequency Spectrum of AM.

Modulation Index

Modulation index is defined as the ratio of change in amplitude of carrier wave to the amplitude of normal carrier wave. Hence

$$m_a = \frac{V_m(\text{amplitude change of carrier wave})}{V_c(\text{amplitude of normal, unmodulated carrier wave})}$$

The modulation index is a number between 0 and 0.8. It can also be represented in percentage by multiplying by 100 :

$$\%\ \text{modulation} = \frac{V_m}{V_c} \times 100$$

It may also be defined in terms of values referred to the modulated carrier wave (refer Fig. 6.6).

$$m = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

Where V_{max} and V_{min} are the maximum and minimum values of the modulated carrier wave.

The Various Cases For

Amplitude modulated wave for different values of 'm' are discussed as: (Assuming in V_m as the amplitude of signal).

Case 1. When $V_m = 0$, it means $m = 0$
This is the case of unmodulated carrier wave.

Case 2. When $V_m = \frac{V_c}{2}$
In this case $m = 0.5$ or 50%
The carrier wave is said to be 50% modulated.

Case 3. When $V_m = V_c$
Then $m = 1$ or 100%
The carrier wave is 100% modulated.

Case 4. When $V_m > V_c$
Then $m >> 1$
The modulation is called over-modulation.

Advantages and Disadvantages of AM

Advantages

Amplitude Modulation has the following advantages :

i. Simple to implement
ii. Inexpensive
iii. Can be demodulated using only a few components

Disadvantages

i. Low efficiency in terms of power usage
ii. Limited range of operation
iii. Noisy reception
iv. Poor recovery of original signal
v. Lower band width

Applications of AM

i. Radio Broadcasting
ii. Inexpensive
iii. Personal (Walkie-Talkie)
iv. Military : to control weapons
v. Business : Conference calls

FREQUENCY MODULATION (FM)

Frequency Modulation

Frequency modulation is the process in which the frequency of carrier wave is varied instantaneously in proportion with the amplitude of the modulating signal, keeping amplitude and phase constant. Thus the information is conveyed via frequency change. Fig. 6.8 shows the FM waveform for sine wave.

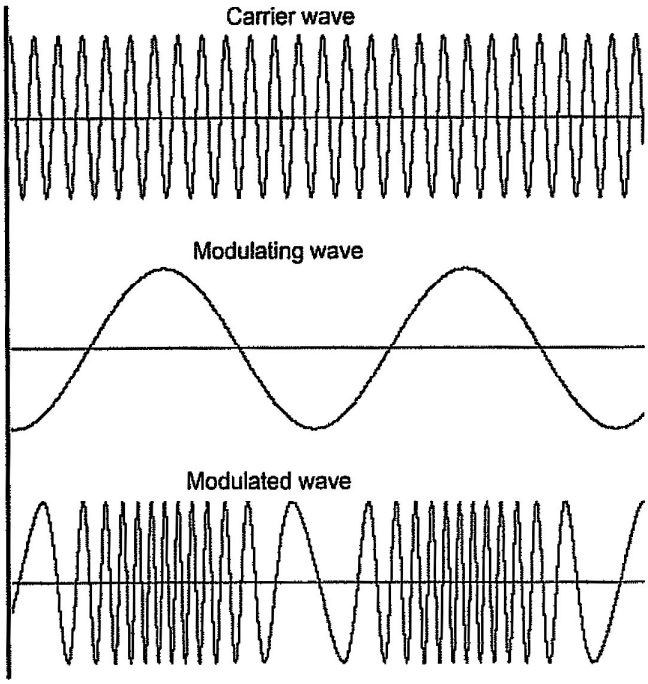


Fig. 6.8. FM waveform for sine wave.

In an FM receiver, it is the amount of frequency shift that is produced in the modulated waveform that determines the volume (audio intensity) heard on the speaker of the receiver. Thus we can say ; as the amplitude of the modulating signal increases or decreases, the frequency of the carrier wave also increases or decreases. However the amplitude of the FM modulated carrier remains constant and it does not change. The problem associated in AM are overcome in FM.

Analysis and Frequency Modulated Carrier Wave

Let the carrier and modulating wave be represented by

$$\nu_c = V_c \text{ Cos } (\omega_c t + \theta) \quad ...(i)$$

and $$\nu_m = V_m \cos \omega_m t \quad ...(ii)$$

Where ν_c, V_c, ω_c and θ are instantaneous values, peak value, angular velocity and the initial phase angle of the carrier ; and ν_m, V_m and ω_m are the instantaneous value, peak value and the angular velocity of the modulating wave.

Let $$\phi_c(t) = \omega_c t + \theta \quad ...(iii)$$

Where $\phi_c(t)$ is the total instantaneous phase angle of the carrier wave at time t. Hence from equation (i), we get

$$\nu_c = V_c \cos \phi_c(t) \quad ...(iv)$$

The ω_c is defined as the instantaneous rate of increase of instantaneous phase (or angular displacement), is related to phase angle ϕ_c as below :

$$\omega_c = \frac{d\phi_c}{dt} \quad ...(v)$$

In FM, f_c (frequency of carrier wave) varies with time according to the instantaneous value of the modulating voltage. Thus the frequency of the carrier after FM is given by.

$$\omega = \omega_c + K_f \cdot v_m = \omega_c + K_f \cdot V_m \cos \omega_m t \quad ...(vi)$$

Where K_f is a constant of probability and represents the frequency conversion factor, whose value depends on the modulation system. The total instantaneous phase of the FM wave is obtained by integrating equation (vi).

$$\phi(t) = \int \omega \cdot dt = \int (\omega_c + K_f \cdot V_m \cos \omega_m t \cdot dt$$

$$= \omega_c t + K_f \cdot V_m \frac{1}{\omega_m} \sin \omega_m t + \theta_l$$

Where θ_t is the constant of integration and represents the initial phase. It may be neglected as it is insignificant in the modulation process. Thus:

$$\phi(t) = \omega_c t + K_f \frac{V_m}{\omega_m} \sin \omega_m t$$

Hence the equation of FM wave is given by

$$v_{FM} = V_c \sin \phi(t)$$

or

$$v_{FM} = V_c \sin \left[\omega_c t + K_f \cdot \frac{V_m}{\omega_m} \sin \omega_m t \right] \quad ...(vii)$$

The instantaneous frequency of FM wave can be drawn from equation (vi). Thus

$$f = \frac{\omega}{2\pi} = \frac{\omega_c}{2\pi} + K_f \frac{V_m}{2\pi} \cos \omega_m t$$

or

$$f = f_c + K_f \frac{V_m}{2\pi} \cos \omega_m t \quad ...(viii)$$

From this equation the maximum and minimum value of frequency are given as:

$$f_{max} = f_c + K_f \frac{V_m}{2\pi}$$

and

$$f_{min} = f_c - K_f \frac{V_m}{2\pi}$$ (as cos 180 = -1)

Hence the **frequency deviation**, which is defined as the maximum change in frequency from mean value f_c, is given by

$$f_d = f_{max} - f_c = f_c - f_{min} = K_f \frac{V_m}{2\pi} \quad ...(ix)$$

Also the **carrier swing**, which is the total variation in frequency from minimum to maximum value i.e., $(f_{max} - f_{min})$ is :

$$CS = 2f_d = K_f \cdot \frac{V_m}{\pi} \quad ...(x)$$

and the **frequency modulation index** (m_f) which is defined as the ratio of frequency deviation to modulation frequency is given by:

$$m_f = \frac{f_d}{f_m} = \frac{\omega_d}{\omega_m} = \frac{K_f \cdot V_m}{\omega_m} \quad ...(xi)$$

Thus the equation for the FM wave equation (vii) in terms of m_f becomes :

$$v_{FM} = V_c \sin[\omega_c t + m_f \sin \omega_m t]$$

or

$$v_{FM} = V_c [\sin \omega_c t \cos(m_f \sin \omega_m t) + \cos \omega_m t \cdot \sin(m_f \sin \omega_m t)] \quad ...(xii)$$

Modulation Index

In FM, the modulation index is normally used instead of percentage of modulation. It is defined fined as the ratio of the frequency deviation to the frequency of the modulating signal. It is expressed in decimal and does not have any unit. For a constant frequency deviation the modulation index drops as the frequency of the modulating signal increases. It varies during the transmission of FM signal as the modulation frequency changes. This term is required for the calculation of bandwidth of an FM signal.

$$m_f = \frac{\Delta f}{f_m} = \frac{K_f \cdot V_m}{\omega_m}$$ [From equation (x)]

Here we can say that for a given frequency (Δf) deviation or for a constant amplitude modulating voltage, varies inversely to the f_m.

Advantages and Disadvantages of FM

Advantages

i. Improved signal to noise ratio
ii. Better immune to noise (Less distortion)
iii. Less interference from adjacent channels
iv. Less radiated power
v. Greater efficiency

Disadvantages

i. Suffer more attenuation than AM signal
ii. Larger band width
iii. More complicated receiver and transmitter

Applications of FM

i. Radio Broadcasting
ii. TV Transmission (For Audio)
iii. Mobile Communication (Earlier)
iv. Police, Fire Department etc.
v. Point to point microwave links used by telecom companies

DIFFERENCE BETWEEN AM AND FM

Table 6.3 gives the main highlights of the difference between AM and FM

Table 6.3 : Difference between AM and FM

S.No.	Parameter	AM	FM
1.	Principle of operation	By changing amplitude of carrier wave	By changing frequency of carrier wave
2.	Amplitude	Depends on modulating index	Constant
3.	Transmitter	Simple	Complex
4.	Transmitted power	Fully utilized	A fraction is utilized
5.	Area of reception	Large	Small
6.	Band width required	Small	Large
7.	Cost	Low	High
8.	Modulating index	Maximum value = 1	No restriction
9.	Prone to noise	More	Less
10.	Broadcast type (Radio)	Can only transmit mono for talk only	Can transmit stereo suitable for music

SOLVED EXAMPLES

Example 6.1. A 1 MHz carrier is amplitude modulated by a 40 KHz modulated signal to a depth of 50%. The unmodulated carrier is having a power of 1kW. Calculate the power of amplitude modulated signal and side band frequencies.

Solution : Given, carrier power,

$$P_c = 1 \text{ kW}$$

Modulation index, m = 50% = 0.5

Power of AM signal, $P_{total} = P_{carrier}\left(1+\frac{m^2}{2}\right) = 1\times\left(1+\frac{0.5^2}{2}\right) = 1.125 \text{ kW}$ **Ans.**

Lower side band frequency, $f_{LSB} = f_c - f_m = 1000 - 40 = 960$ KHz **Ans.**

Upper side band frequency, $f_{USB} = f_c + f_m = 1000 + 40 = 1{,}040$ KHz **Ans.**

Example 6.2. A certain transmitter radiates 9 kW with the carrier unmodulated and 10.125 kW, when the carrier is sinusoidal modulated. Calculate the modulation index. If another sinewave corresponding to 40% modulation is transmitted simultaneously, determine the total radiated power.

Solution : Given, carrier power,

$$P_{carrier} = 9 \text{ kW}$$

Total power of modulated wave,

$$P_{total} = 10.125 \text{ kW}$$

$\therefore$ Modulation index, $m = \sqrt{2\left(\frac{P_{total}}{P_{carrier}} - 1\right)}$ $\because \; P_{total} = P_{carrier}\left(1 + \frac{m^2}{2}\right)$

$$= \sqrt{2\left(\frac{10.125}{9} - 1\right)} = 0.5 = 50\%$$ **Ans.**

Example 6.3. An AM broadcast radio radiates 10 kW of power if modulation percentage is 60. Calculate how much of this is the carrier power.

Solution : We know $P_{total} = P_{carrier}\left(1 + \frac{m^2}{2}\right)$

or $$P_{carrier} = \frac{P_{total}}{\left[1 + \frac{m^2}{2}\right]} = \frac{10}{\left[1 + \frac{(0.6)^2}{2}\right]}$$

$$= \frac{10}{1.18} = 8.47 \text{ kW}$$ **Ans.**

Example 6.4. The r.m.s. antenna current of a radio transmitter is 10A, when unmodulated and rising to 12 A when the carrier is sinusoidally modulated. Calculate the modulation index.

Solution : Given, carrier current

$$I_c = 10 \text{ A}$$

Total modulated current $I_{total} = 12$ A

We know, modulation index, $m = \sqrt{\left[\left(\frac{I_{total}}{I_{carrier}}\right)^2 - 1\right] \times 2}$

$$= \sqrt{\left[\left(\frac{12}{10}\right)^2 - 1\right] \times 2}$$

= 0.938 **Ans.**

Example 6.5. Determine the modulation index of an FM carrie having a frequency deviation of 25 KHz and modulation signal of 5 KHz. Also determine the carrier swing.

Solution : Given, frequency deviation

$$\Delta f = 25 \text{ KHz}$$

Frequency of modulation signal,

$$f_m = 5 \text{ KHz}$$

Modulated index $m_f = \frac{\Delta f}{f_m} = \frac{25}{5} = 5$ **Ans.**

Carrier swing $= 2 \cdot \Delta f = 2 \times 25 \text{ KHz} = 50 \text{ KHz}$ **Ans.**

Example 7.6. An FM transmission has a frequency deviation of 18.75 KHz. Calculate percent modulation if it is broadcasted (i) in the 88 - 108 MHz band (ii) as a portion of a TV broadcast.

Solution : Given frequency deviation,

$$\Delta f = 18.75 \text{ KHz}$$

A maximum frequency deviation of 75 KHz is allowed for commercial FM broadcast :

i.e., $(\Delta f)_{max} = 75 \text{ KHz}$

$$\% \text{ modulation, } m = \frac{\Delta f}{(\Delta f)_{max}} \times 100$$

$$= \frac{18.75}{75} \times 100 = 25\%$$ **Ans.**

A maximum frequency deviation of 25 KHz is allowed for sound portion of the TV broadcast.

i.e. $(\Delta f)_{max} = 25 \text{ KHz}$

$\therefore$ $$\% \text{ modulation, } m = \frac{18.75}{25} \times 100$$

= 75% **Ans.**

SUMMARY

1. **Communication:** Means the process of exchanging information. Radio communication can be used for long range such as from one country to other or from earth to space.

2. **Modulation:** In this process some parameter of the carrier wave such as amplitude, frequency or phase is varied in accordance with the modulating signal.

3. **Demodulation:** or detection is the process of recovering the original modulated signal (information) from the modulated carrier wave.

4. **Modulation index:** or modulation depth is described by how much the modulated variable of the carrier wave signal varies around its unmodulated level. It is defined differently in each modulation scheme i.e. AM, FM and PM.

5. **Spectrum:** A spectrum represents the relative amount of different frequency components in any signal.

6. **Efficiency:** $\text{efficiency} = \frac{\text{Power in all the side bands}}{\text{Total transmitted power}}$

7. **Frequency Deviation:** The change of shift above or below the mean or centre frequency is called frequency deviation.

8. **FM Index:** The ratio of frequency deviation to modulation frequency is called the FM index.

9. **Bandwidth of AM Wave:** The difference between the two extreme frequencies is equal to the bandwidth of AM wave.

   $$\text{B.W.} = (\omega_c + \omega_m) - (\omega_c - \omega_m)$$

   or $$\text{B.W.} = 2\omega_m$$

IMPORTANT FORMULAE

1. **Size of quarter wave antenna**

$$l = \frac{\lambda}{4} = \frac{3 \times 10^8}{4 \times f} \quad \text{meter}$$

where f is frequency in Hz.

2. **General equation of sinusoid signal**

$$f(t) = A \cdot \sin(\omega t + \phi)$$

3. **Modulation index for AM**

$$m_a = \frac{V_m}{V_c} \frac{\text{(Amplitude change of carrier wave)}}{\text{(Amplitude of normal, unmodulated carroer wave)}}$$

or $$\% \text{ Modulation} = \frac{V_m}{V_c} \times 100$$

4. **Modulation index for FM :**

$$m_f = \frac{\Delta f}{f_m}$$

5. **Power of AM signal :**

$$P_{total} = P_{carrier}\left(1 + \frac{m^2}{2}\right)$$

6. **Efficiency :**

$$\eta = \frac{\text{Power in all the side bands}}{\text{Total transmitted power}}$$

REVIEW QUESTIONS

Q.1. What is communication system. Explain with block diagram.

Q.2. Write short note on analog and digital communications.

Q.3. What do you mean by modulation index.

Q.4. Derive an expression for the AM index.

Q.5. What is modulation and why it is necessary.

Q.6. Compare AM and FM.

Q.7. Define AM and FM, use sketches to explain these definitions. Where are these employed.

Q.8. What is frequency modulation? Give the mathematical representation of FM.

Q.9. What is the modulation index of an frequency modulated (FM) carrier having carrier swing of 200 KHz and a modulating signal of 10 KHz.

TRANSDUCER

The words 'sensor' and 'transducer' are both widely used in the description of measurement systems. The former is popular in the USA whereas the latter has been used in Europe. A **'sensor'** is a device that detects a change in physical stimulus and turns it into a signal which can be measured or recorded, a corresponding definition of **'transducer'** is a device that transfers power from one system to another in the same or in the different form. The block diagram of a transducer is shown in Fig. 8.2.

Fig. 6.9. Block diagram of transducer.

The transducer receives energy from the process variable under measurement and transmits a proportional electrical energy to the output. This output energy can be further processed through liner operations like amplification, attenuation, differentiation, integration, addition, subtraction etc. so that it can be acceptable by the output unit for record or display. The processing of signal is called signal conductioning.

Qualities of a Good Transducer

- Ruggedness
- Repeatability
- Good dynamic response
- No hysteresis
- Linearity
- High output signal quality
- High reliability and stability

TYPES OF TRANSDUCER

There are various types of transducers depending upon the change in property or the energy they bring about to measure specified physical quantities. The input given to a transducer can be in the form of the displacement, strain, velocity, temperature, flow etc. and the output obtained fro, them can be in the form of voltage, current, change in resistance, inductance and capacitance etc. The output can be measured easily and it is calibrated against the input, thus enabling the measurement of the value of the input. The widely used transducer in measuring instruments are of following types.

Thermoelectric Transducers

- Thermocouple
- Thermistor
- Resistance Temperature Detector (RTD)
- Peltier Cooler

Electromechanical Transducers

- Strain Gauge
- Potentiometer
- Accelerometer
- Load Cell
- Galvanometer

Mechanical Transducer

- Bimetallic Strib
- Bourdon Tube
- Bellows
- Spring

Photoelectric Transducer

- Laser diodes
- Photo diode, photo resistor, photo transistor
- Light Emitting Diodes (LEDs)

Electro Acoustic Transducer

- Microphone
- Piezo electric Crystal
- Loudspeaker, earphone
- Hydrophone

Electrochemical Transducer

- pH Probes
- Electro-Galvanic Fuel Cell
- Hydrogen Sensors

Electromagnetic Transducers

- Antenna
- Hall-Effect Sensor
- Cathode Ray Tube (CRT)
- Magnetic Disk Reader/Writer

THERMOCOUPLE

Thermocouple is a transducer for the measurement of high temperature. It is a thermoelectric device, which works on the principle of **seeback effect** by converting thermal energy into electrical energy.

It consists of two wires of dissimilar metals joined together to form two junctions in a close circuit formation. If the two junctions are at different temperature, an electric current will flow in the circuit. This phenomenon is known as seebeck effect and was first observed by Thomas Johnn seebeck in 1981. He arranged 35 metals in order of their thermoelectric properties. The current flows through the hot junction from the former to the latter metal of the following series :

Bi-Ni-Co-Pd-Pt-U-Cu-Mn-Ti-Hg-Pb-

Sn-Cr-Mo-Ph-Ir-Au-Ag-Zn-W-Cd-Fe-As-Sb-Te

If the two metals used are copper (Cu) and Iron (Fe), then the current flows from copper to Iron at the hot junction (J_h) and from Iron to copper at the cold (reference) junction (J_c) which can be detected by galvanometer. This is explained in Fig 6.10.

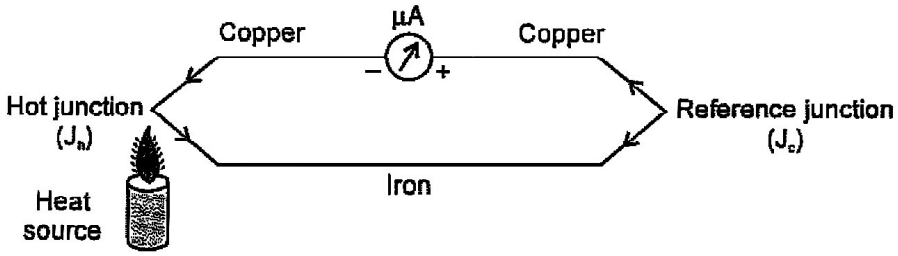


Fig. 6.10. Basic Thermocouple Phenomenon.

The amount of current (emf) produced depends on the difference in the temperature between the two junctions and on the characteristics of the two metals. The instrument that records the variations in current flow is calibrated in terms of temperature and is known as thermocouple pyrometer. When the junction, J_h is heated it produces voltage greater than the voltage across the cold junction (J_c). The difference between two voltage is measured and the voltmeter reading is converted into its corresponding temperatures. The conversion table is provided by thermo-couple manufacturer.

The emf produced in thermocouple is given by

$$\text{emf} = \int_{T_1}^{T_2} (S_1 - S_2)dt$$

Where T_1 and T_2 are the temperature of reference and measuring end respectively and S_1 and S_2 are the Seebeck coefficients of two thermoelements. The temperature-emf characteristics of some of the metals is shown in Fig. 6.11 where J_c is kept at 0°C and J_h is at variable temperature.

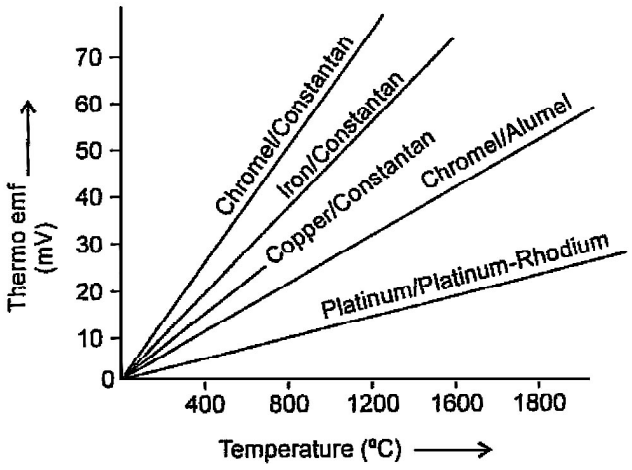


Fig. 6.11. Temperature-emf characteristics of some Thermocouples.

Advantages

i. Inexpensive
ii. Rugged construction
iii. High temperature range (200 to 2600°C)
iv. Fast response to change in temperature
v. Good accuracy

Disadvantages

i. Vulnerable to corrosion
ii. Calibration is difficult and tedious
iii. Complexity and prone to error
iv. Produces low volt output (mV)
v. None linear device
vi. Requires two temperature measurements

RESISTANCE TEMPERATURE DETECTOR (RTD)

Resistance temperature detector (or Resistance Thermometer) is basically a temperature sensitive resistor. It is a positive temperature coefficient device which means the resistance increases with temperature. The commonly used metals elements for RTD are platinum, nickel, copper and tungsten. The platinum has a linear temperature-resistance characteristics and is reproductable over wide range of temperatures. It is available in pure form and is relatively unaffected by environmental conditions, hence is used for precision thermometry. Nickel has a high temperature coefficient. The common forms of RTD elements are shown in Fig. 6.12.

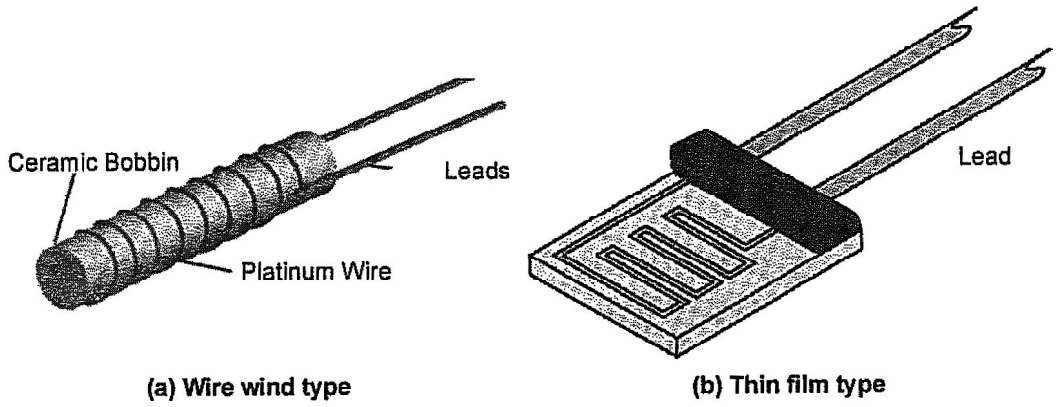


Fig. 6.12. Common Forms of RTD Elements.

The materials used for RTD have temperature coefficients of resistivity much higher than the coefficient of thermal expansion. The temperature coefficient of resistance, α, is given by :

$$\alpha = \frac{1}{\Delta T} \cdot \frac{\Delta \rho}{\rho_0} = \frac{1}{\Delta T} \cdot \frac{\Delta R}{R_0}$$

where ΔT = change in temperature (°C)

$\Delta \rho / \rho_0$ = fractional change in resistivity

$\Delta R / R_0$ = fractional change in resistance

ρ_0, R_0 = resistivity and resistance respectively (°C)

The resistance R_T at any other temperature is given as :

$$R_T = R_0(1 + \alpha \cdot \Delta T)$$

or $$R_T = R_0(1 + \alpha T)$$

But in case of non-linearly, the equation is modified as :

$$R_T = R_0(1 + \alpha_1 T + \alpha_2 T^2 + ... + \alpha_n T^n)$$

Where $\alpha_1, \alpha_2, ... \alpha_n$ are constants applicable for each metal.

Advantages

i. High Accuracy
ii. Linearity over wide range of temperature
iii. High temperature operation
iv. Fast response
v. Better stability at high temperature
vi. Easily replaceable
vii. More suitable for remote indicators
viii. Immune to electrical noise

Disadvantages

i. High cost
ii. Requires no point sensing
iii. Possibility of self heating
iv. Requires external DC power supply
v. Low sensitivity
vi. Affected by shock and vibration

Applications

i. Heating ovens/food processing
ii. Compression mouldings
iii. Plastics processing
iv. Air, gas and liquid temperature measurement
v. Textile production

STRAIN GAUGE

A strain gauge is a transducer which is used to measure the strain of an object. When an electrical conductor is stretched, it will become narrower and longer, both increases its electrical resistance end-to-end. Conversely, when a conductor is compressed (without buckling), it will broaden and shorten both decreases its electrical resistance end to end. If these stresses are kept within the elastic limit of the metal conductor (So that it does not deform permanently), the strip can be used as a measuring element for physical force, the amount of applied force inferred from measuring its resistance.

The most commonly used types of strain gauge are wire-type strain gauge, foil strain gauge and semiconductor strain gauge.

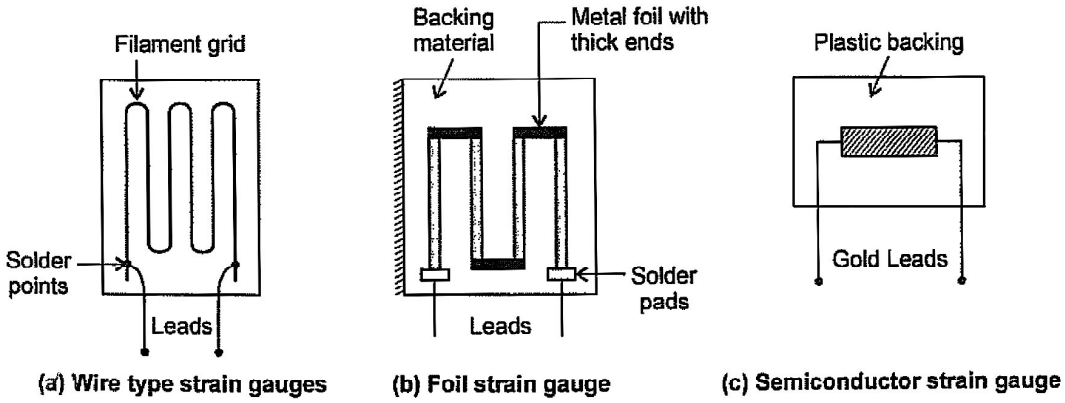


Fig. 6.13. Some Common Type of Strain Gauges.

Derivation for Gauge Factor

The fundamental formula for the resistance of a wire (conductor) of length, l, diameter d, cross section area A and resistivity ρ, can be expressed as :

$$R = \frac{\rho l}{A} = \frac{\rho l}{\pi\left(\frac{d^2}{4}\right)} = \frac{\rho l}{\frac{\pi}{4}\cdot d^2} \quad \text{.... (i)}$$

Let a tensile stress, S be applied to the wire, thereby a positive strain is produced which causes the length to increase by Δl and area to decrease by ΔA as shown in Fig. 6.14. This will also reduce the diameter by and thus resistance changes by ΔR.

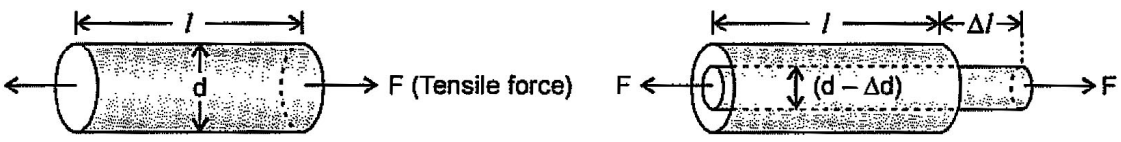


Fig. 6.14. Change in dimension of strain gauge wire when subjected to a tensile force.

The strain changes two important parameters of the wire i.e., ΔR and Δl. The measurement of sensitivity of a material to strain is known as gauge factor. It is defined as the ratio of change in resistance $\Delta R/R$ to change in length $\Delta l/l$.

Hence, Gauge factor $$(G) = \frac{\Delta R/R}{\Delta l/l} \quad \text{.... (ii)}$$

Since strain is defined as change in length divided by original length i.e $\sigma = \Delta l/l$. Hence equation (ii) can written as :

$$G = \frac{\Delta R/R}{\sigma} \quad \text{...(iii)}$$

Due to strain the length of wire increases by Δl and simultaneously the diameter decreases by Δd. Hence the new value of resistance is given by [Using equation (i)]

$$R_s = \frac{\rho(l+\Delta l)}{\frac{\pi}{4}\cdot(d-\Delta d)^2} = \frac{\rho\cdot(l+\Delta l)}{\frac{\pi}{4}\cdot(d^2-2d\cdot\Delta d+2d^2)}$$

Neglecting Δd^2 as Δd is very small. We get

$$R_s = \frac{\rho\cdot(l+\Delta l)}{\frac{\pi}{4}(d^2-2d\cdot\Delta d)} = \frac{\rho\cdot l\cdot\left(1+\frac{\Delta l}{l}\right)}{\frac{\pi}{4}\cdot d^2\left(1-\frac{2\cdot\Delta d}{d}\right)}$$

We know that Poisson's ratio, μ is defined as the ratio of strain in lateral direction to strain in axial direction. Thus

$$\mu = \frac{\Delta d/d}{\Delta l/l}$$

or $$\frac{\Delta d}{\Delta d} = \mu\cdot\frac{\Delta l}{l}$$

Substituting the value of $\frac{\Delta d}{d}$ in equation (iv), we get

$$R_s = \frac{\rho\cdot l\left(1+\frac{\Delta l}{l}\right)}{\frac{\mu}{4}\cdot d^2\left(1-2\cdot\mu\cdot\frac{\Delta l}{l}\right)}\times = \frac{\left(1+2\mu\frac{\Delta l}{l}\right)}{\left(1+2\mu\frac{\Delta l}{l}\right)}$$

$$= \frac{\rho l}{\left(\frac{\pi}{4}\right)d^2}\cdot\left[\frac{1+2\mu\cdot\frac{\Delta l}{l}+\frac{\Delta l}{e}+2\mu\cdot\frac{\Delta l^2}{l^2}}{1-4\mu^2\cdot\frac{\Delta l^2}{l^2}}\right]$$

Neglecting higher power of Δl, as Δl in very small.

$$R_s = \frac{\rho\cdot l}{\left(\frac{\pi}{4}\right)d^2}\cdot\left[1+2\mu\cdot\frac{\Delta l}{l}+\frac{\Delta l}{l}\right]$$

$$= \frac{\rho\cdot l}{\left(\frac{\pi}{4}\right)d^2}\cdot\left[1+(2\mu+1)\frac{\Delta l}{l}\right]$$

$$= \frac{\rho\cdot l}{\left(\frac{\pi}{4}\right)\cdot d^2}+\frac{\rho\cdot l}{\left(\frac{\pi}{4}\right)\cdot d^2}\cdot\left(\frac{\Delta l}{l}\right)\cdot(1+2\mu)$$

or $$R_s = R+\Delta R$$ from Equation (i)

where $$R = \frac{\rho\cdot l}{\left(\frac{\pi}{4}\right)d^2}$$ and

$$\Delta R = \frac{\rho \cdot l}{\left(\frac{\pi}{4}\right)d^2} \cdot \left(\frac{\Delta l}{l}\right) \cdot (1 + 2\mu)$$

$\therefore$ Gauge factor $$G = \frac{\Delta R/R}{\Delta l/l} = \frac{\left(\frac{\Delta l}{l}\right)(1+2\mu)}{\frac{\Delta l}{l}}$$

or $$G = 1 + 2\mu$$

Advantages

i. No moving Part
ii. Small in six
iii. Inexpensive

Disadvantages

i. They are non-linear
ii. Need to be calibrated
iii. Usually they are placed where it is really hard to install

LOAD CELL

Load cells are elastic devices that can be used for the measurement of force through indirect method i.e. through use of secondary transducer. Load cells utilize an elastic member as a primary transducer and strain gauge as secondary transducer. When the combination of the strain gauge-elastic member is used for weighing, it is called a **"load cell"**. It converts force into a measurable electrical output. Although there are many varieties of load calls, strain gauge based load cells are the most commonly used type.

Construction and Working

The strain gauge load cells shown in Fig. 8.15 utilizes four identical strain gauges attached to a steel cylinder. 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 gauge R_{g1} and R_{g4}. All the four gauges are connected into the bridge circuitry in such a manner as to make use of poisson's ratio.

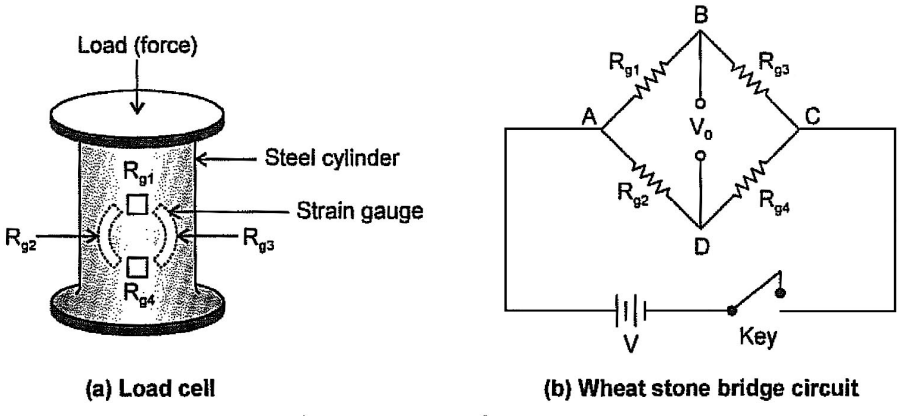


Fig. 6.15. Load cell Strain Gauge.

When there is no load on the cell, all the four gauges have the same resistance i.e. $R_{g1} = R_{g2} = R_{g3} = R_{g4}$. Hence the terminals B and D are at same potential, the bridge is balanced and the output voltage, $V_0 = 0$.

i.e. $$V_{AB} = V_{AD} = \frac{V}{2}$$

also $$V_{AB} - V_{AB} = V_0 = 0$$

Now when the load is applied the vertical gauges (R_{g1}, R_{g4}) undergo compression i.e. negative strain and thus decrease in resistance. Simultaneously the circumferential gauges (R_{g2}, Rg3) undergo tension i.e. positive strain which leads to increase in resistance. The two strain are not equal and are related to a factor μ, the Poisson's ratio. In this condition various resistance of the gauges are :

$$R_{g1} = R_{g4} = R - \Delta R \qquad \text{(Compression)}$$
$$R_{g2} = R_{g3} = R + \Delta R \qquad \text{(Tension)}$$

Potential at terminal B, $$V_{AB} = \frac{R_{g1}}{R_{g1} + R_{g3}} \cdot V = \frac{R - \Delta R}{(R - \Delta R) + (R + \mu \cdot \Delta R)} V$$

$$= \frac{R - \Delta R}{2R - \Delta R(1 - \mu)} \cdot V$$

Potential at terminal D, $$V_{AD} = \frac{R_{g2}}{R_{g2} + R_{g4}} \cdot V = \frac{R + \mu \cdot \Delta R}{(R + \Delta R) + (R + \mu \cdot \Delta R)} \cdot V$$

$$= \frac{R + \mu \cdot \Delta R}{2R - \Delta R(1 - \mu)} \cdot V$$

The changed output voltage

$$V_0 + \Delta V_0 = \frac{R - \Delta R}{2R - \Delta R(1 - \mu)} \cdot V - \frac{R + \mu \cdot \Delta R}{2R - \Delta R(1 - \mu)} \cdot V$$

$$= \frac{\Delta R(1 + \mu)}{2R} = 2(1 + \mu)\left(\frac{\Delta R}{R} \cdot \frac{V}{4}\right)$$

As the output voltage V_0 = 0 under unload conditions, hence change in output voltage due to applied load will be

$$\Delta V_0 = 2(1 + \mu)\left(\frac{\Delta R}{R} \cdot \frac{V}{4}\right)$$

This voltage is a measure of 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.

Advantages
i. Inexpensive
ii. Hermetically sealed & maintenance free
iii. Fast response to the load
iv. Small and compact size

Disadvantages
i. Needs protection against angular and non-axial loads.
ii. Overloading should be avoided beyond rated capacity.

Uses
i. Road vehicle or Aircraft weighing devices.
ii. Crane load monitoring
iii. Bridging lifting
iv. Torsion Test
v. Process Control

BIMETALLIC STRIP

A bimetallic strip is used to confert a temperature change into mechanical displacement. The strip consists of two strips of different metals of differing co-efficients of thermal expansion, bonded or welded together to form a single piece. At the bonding temperature the strip is flat and straight but when it is heated to a high temperature it starts bending or curling towards the side of metal with lower coefficients and when it is cooled below its normal temperature. The layer with higher coefficients α, is called active side and the other with lower coefficient is called the passive side. Usually the metals used are steel and copper or in some cases brass instead of copper. In some cases it is used as coil for compactness. Fig. 6.16 shows bimetallic elements.

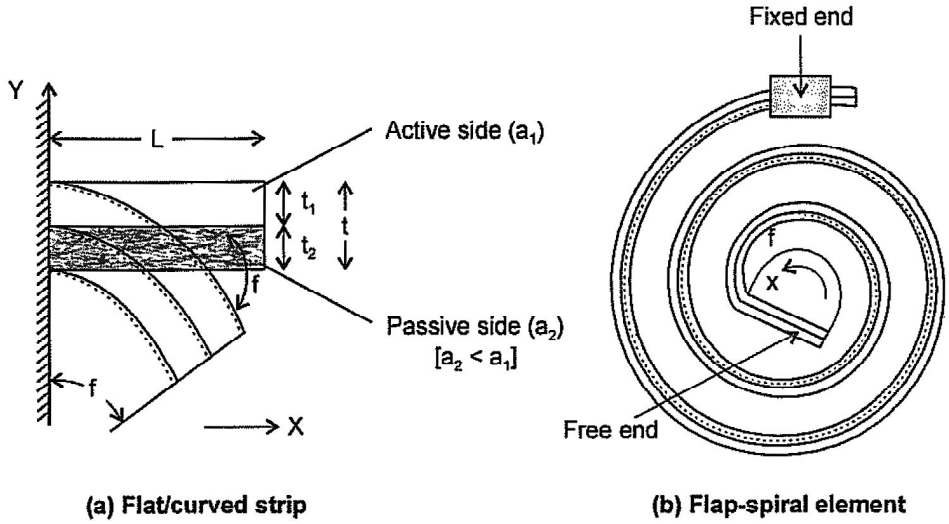


Fig. 6.16. Bimetallic Strip.

The strip is fixed at one end and the other end kept open in the environment whose temperature is to be measured. A rise in temperature by T°C causes the element to bend into an arc of a circle. The angle ϕ between the original plate axis and the tangent drawn from the tip to the deflection line is given by :

$$\phi = \frac{6\, E_1 E_2 t_1 t_2 (t_1 + t_2)(\alpha_1 + \alpha_2)}{4\, E_1 E_2 t_1 t_2 (t_1 + t_2)^2 + (E_1 t_1^2 - E_2 t_2^2)} \cdot L\, \Delta T$$

or $$\phi = k_T \cdot L \cdot \Delta T \quad \text{radians}$$

where E_1, E_2 = young's moduli of the materials of layer 1 and 2 respectively.

t_1, t_2 = Thickness of layer 1 and 2

α_1, α_2 = Coefficients of thermal expansion of the materials of layer 1 and 2 respectively.

L = Length of original element

ΔT = Rise in temperature

k_T = Bimetallic element sensitivity in radians per m/°C

The sensitivity will be maximum when the term $(E_1 t_1^2 - E_2 t_2^2)$ becomes zero, thus

$$\frac{t_1}{t_2} = \sqrt{\frac{E_2}{E_1}}$$

Under this condition K_T is given by :

$$K_T = \frac{3}{2}\frac{(\alpha_1 - \alpha_2)}{t}$$

The deflection (d) of the tip of the element from its original position is :

$$d = K_T \cdot \frac{L^2}{2} \cdot \Delta T$$

For circular strip $$\Delta\phi_T = K_T \cdot L \cdot \Delta T$$

Advantages

i. No moving part to avoid wear outs.
ii. Saving of expansive metals and alloys.
iii. Portability
iv. Independence from power supply.

Disadvantages

i. Railway track and roadway construction.
ii. Pipes carrying hot water.
iii. Inaccurate

Applications

i. Thermostats
ii. Mechanical clocks
iii. Heat engines
iv. Circuit breakers (restores when cooled down)

INTRODUCTION OF INTEGRATED CIRCUITS (ICS)

Until 1960 the technology of discrete components were used to design electronic circuits. It uses diodes, transistors in conjunction with other circuit elements like resistors, inductors, capacitors etc. which are joined with wires or plated conductors on printed circuit boards (PCBs). These discrete circuits have two main disadvantages. Firstly in a large circuit (like TV circuit, computer circuit) where there are hundred of components, the discrete assembly would occupy a large space. Secondly there will be hundred of soldered points which will cause a poor reliability. To avoid these problems researchers started a drive to miniaturize the circuits. One type of such circuit is the integrated circuits, generally abbreviated as IC.

The first commercially available IC came in 1961 from Fairchild Semiconductor Corporation. **An IC has various components such as resistor, capacitors, diodes, transistors etc. fabricated on the same small semiconductor chip.** ICs are now a days used in electronic industry, instrumentation, control system, computer industry, automobile industry and many other applications. This has now fulfilled the demand of industries for electronic equipment of smaller size, lighter weight, low power requirement and high reliability as compared with discrete components. The IC possesses the following advantages and disadvantages over discrete circuits :

Advantages

i. **Low cost** due to simultaneous production of a large number of components.
ii. **Extremely small size** due to fabrication of various circuit elements on the same chip of semiconductor material. Hence also light weight.
iii. **High Reliability** as components are fabricated sequentially without using soldering.
iv. **Low Power Consumption** due to smaller size components (capacitance, resistance etc.)
v. **High Temperature** handling capacity at extreme values.
vi. **Easy Replacement** as it is more economical to replace them than to repair them.
vii. **Increased Operating Speed** due to absence of parasitic capacitance effect.
viii. **Close matching of components and temperature coefficients** because of bull production in batches.
ix. **Improved functional performance** as more complex circuits can be fabricated for achieving better characteristics.
x. **Suitable for small signal operation** because of no change of stray electrical pickup as various components of an IC are located very close to each other on a silicon water.

Disadvantages

i. If a single component goes bad in an IC, it can bot be removed instead the whole IC is to be replaced.
ii. Limited power dissipation (< 10 watts).
iii. Inductors and transformers can not be integrated into it, instead they are connected exterier to the chip.
iv. High value of capacitance (> 30 pF) not convenient or economical to integrate, hence it is to be connected as discrete component exterior to IC Chip.
v. High grade p-n-p assembly is not easily possible.
vi. Low temperature coefficient is difficult to be achieved.
vii. Difficult to achieve low noise and high voltage operations.
viii. Large value of saturation resistance of transistors.
ix. Voltage dependence of resistors and capacitors.
x. Parameter modification not possible.
xi. Quite delicate in handling as these cannot withstand rough handling or excessive heat.

CLASSIFICATION OF IC'S

The integrated circuits may be classified according to a number of ways as shown in Figure 6.17.

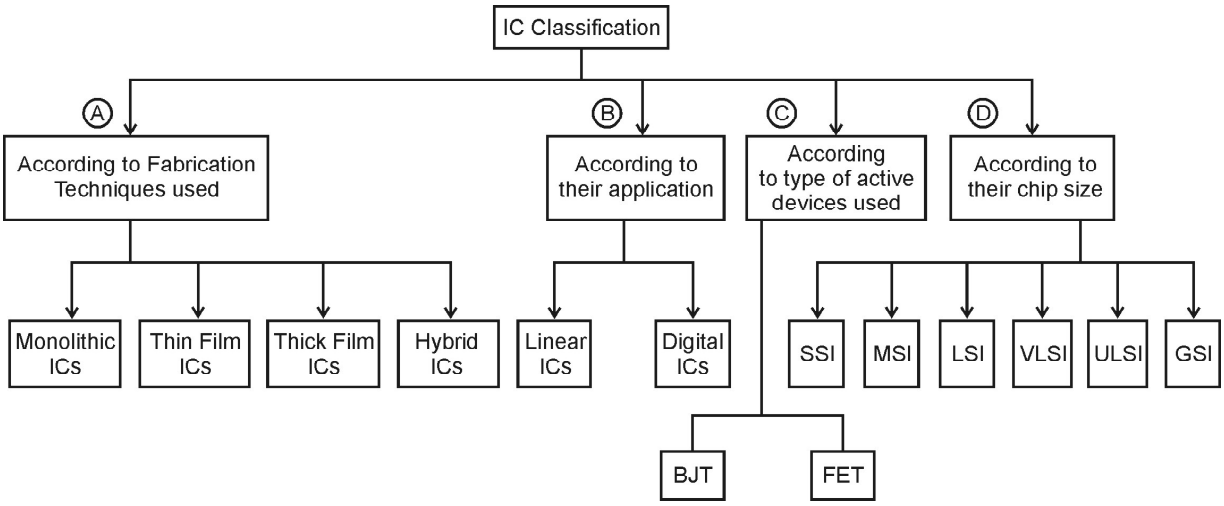


Fig. 6.17. Block Diagram of IC Classification.

A. On the basis of fabrication techniques used, the ICs can be divided into following four classes.

Monolithic ICs

The word 'monolithic' is derived from the greek 'monos' means single and 'lithos' means stone. Thus it is made into a single stone or single crystal. In monolithic ICs all circuit components both active components (transistor etc.) and passive components (diodes, resistor, capacitors etc.) and their interconnections are formed into a single chip of silicon. This technology is useful for manufacturing identical ICs in large quantities and therefore provides lowest per unit cost and greater reliability. These type of ICs are most commonly used. Figure 6.18 shows the monolithic IC in plastic and can type package.

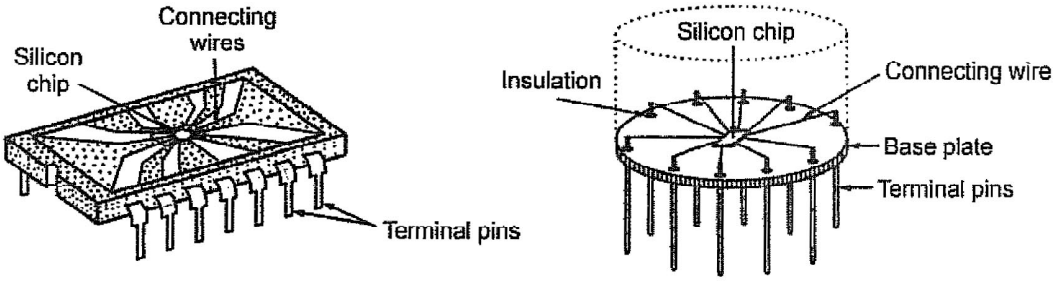


Fig. 6.18. Monolithic IC Packages.

Commercially available ICs of this types are used as amplifiers, voltage regulators, crowbars, AM receivers, TV circuits and computer circuits. However these ICs have following limitations.

- Low power rating (< 1 watt)
- Poorer isolation between components.
- Fabrication of inductors not possible.
- Passive components of only low values are possible.
- Lack of flexity in circuit design i.e. modification is not possible.

Thin and Thick Film ICs

These devices are larger than monolithic ICs but smaller than discrete circuits. These ICs can be used for comparatively higher power requirement. With a thin-or thick-film IC, the passive components like resistors and capacitors are integrated, but the transistors and diodes are connected as discrete components to form complete circuit. Therefore, commercially available thin-and thick-film circuits are combination of integrated and discrete components. The main difference between these is not their relative thickness but the method of deposition of film. Both have similar look, characteristic and features. Figure 6.19 shows the enlarged view of thick film IC.

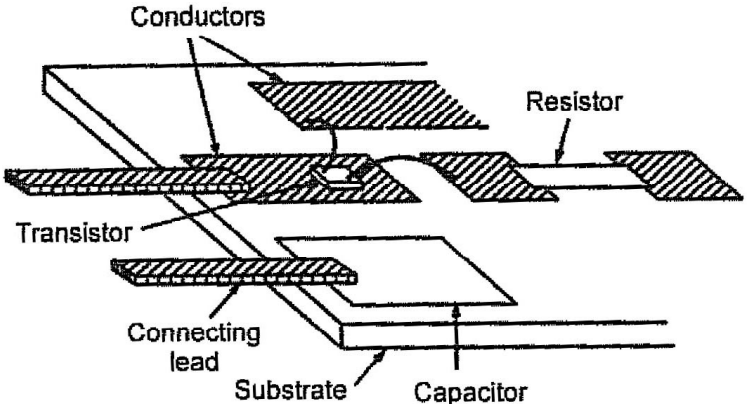


Fig. 6.19. Enlarged view of thick film IC.

Hybrid ICs

In Hybrid ICs the circuit is fabricated by interconnecting a number of individual chips. The active elements are diffused transistors or diodes. The passive components may be group of diffused resistors or capacitors on a single chip, or they may be thin-film components. Wiring provides interconnection between chips. These ICs are used for high power audio amplifiers applications from 5W to 50 Watts. The structure of a hybrid (multi chip ICs) is shown in Figure 6.20.

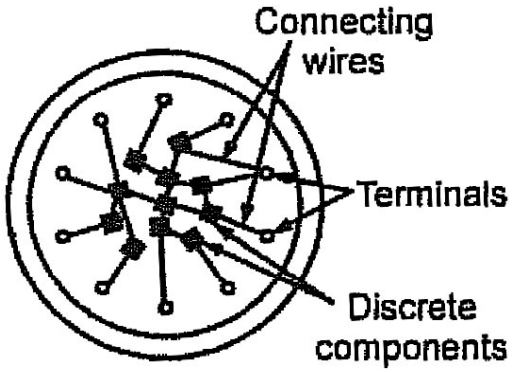


Fig. 6.20. Structure of Hybrid IC.

Like thin-and thick-film ICs, hybrid ICs usually have better performance than monolithic ICs. However the process is too expensive for mass production, multi chip techniques are quite economical for small quantity production and are more offen used as prototype for monolithic ICs.

b. On the basis of application ICs are of two types i.e. linear ICs and digital ICs. When the input/output relationship of a circuit is linear, linear ICs are used for example operational amplifiers (OP-Amps) which was originally designed for performing mathematical operations like addition, subtraction, multiplication, differentiation, integration, inversion etc.

When the circuit is either in on-state or off-state and not in between the two, the circuit is called the digital circuit. ICs used in such circuits are called digital ICs. They find their wide applications in computers and logic circuits.

c. Based upon the active devices used the ICs can be further classified as bipolar ICs using bipolar active devices (BJT) and unipolar ICs using unipolar active devices like (FET).

d. The ICs can also be classified on the basis of their chip size i.e. the number of components contains in it.

- Small Scale Integration (SSI) - 1 to 20 components/chip.
- Medium Scale Integration (MSI) - 20 to 100 components/chip.
- Large Scale Integration (LSI) - 100 to 1000 components/chip.
- Very Large Scale Integration (VLSI) - 1000 to 10,000 components/chip.
- Super Large Scale Integration (SLSI) - 10,000 to 1,00,000 components/chip.

SOLVED EXAMPLES

Example 8.1. A resistive wire strain gauge uses a soft iron wire of small diameter. The gauge factor is +4.2. Neglecting the piezoresistive effects, calculates the Poisson's ratio.

Solution : Given $G = + 4.2$

We know that $G = 1 + 2\mu + \frac{\Delta s / s}{\Delta l / l}$

But as piezoresistive effect is zero, hence

$$G = 1 + 2\mu$$

or $4.2 = 1 + 2\mu$

or $\mu = \frac{4.2 - 1}{2} = \frac{3.2}{2}$

$\therefore$ $\mu = 1.6$ **Ans.**

Example 8.2. A platinum thermometer has a resistance of 100 W at 25°C.

i. Find its resistance at 65°C, if the platinum has a resistance temp. coefficient of 0.003921°C.

ii. If the thermometer has a resistance of 150 W, calculate the temperature.

Solution : Using linear approximation, the resistances at any temperature θ°C.is given by

i. $R_\theta = R_0 (1 + \alpha_0 \Delta\theta)$

At 65°C $R_{65} = 100 [1 + 0.003921 (65-25)]$

or $R_{65} = 115.68\ \Omega$ **Ans.**

ii. Suppose θ is the unknown temperature then

$$150 = 100 [1 + 0.003921 (\theta - 25)]$$

$\theta = 152.55°C$ **Ans.**

■■■