26. SEMICONDUCTOR AND COMMUNICATION SYSTEM

SEMICONDUCTOR

1. INTRODUCTION

Certain substances like germanium, silicon etc. are neither good conductors like copper nor insulators like glass. In other words, the resistivity of these materials lies in between conductors and insulators. Such substances are classified as semiconductors. These substances have crystalline structure and are formed by covalent bonds. An important property of a semiconductor is that by adding a controlled amount of suitable impurity to it, its conductivity can be changed appreciably. This useful property is responsible for the widespread use of semiconductors in electronic devices. In this chapter, we shall discuss the electrical properties of semiconductors.

2. CLASSIFICATION OF SOLIDS

On the basis of electrical conductivity $\left(\sigma\right)$ or resistivity $\left(\rho = 1/\sigma\right)$, the solids can be classified into the following three classes:

(a) **Metal conductors.** These are those solids which possess high conductivity or low resistivity. This is due to the fact that metals have a large number of free electrons. The conductivity of metal conductors lies between 10^2 and 10^8 Sm^{-1} while their resistivity is in between 10^{-2} and 10^{-8} Ω m. Examples of metal conductors are: Al, Cu, Ag, Au, etc.

(b) **Insulators.** These are those solids which possess very low conductivity or very high resistivity. This is due to the fact that insulators have practically no free electrons. The conductivity of insulators lie between 10^{-11} and 10^{-19} Sm^{-1} while their resistivity is between 10^{11} and 10^{19} Ω m. Examples of insulators are: glass, rubber, plastic etc.

(c) **Semiconductors.** These are those solids which possess conductivity and resistivity in between metallic conductors and insulators. This is due to the fact that semiconductors have very few free electrons at room temperature and can be regarded as insulators for all practical purposes. The conductivity of semiconductors lies between 10^{-5} and 10^{0} Sm^{-1} while their resistivity in between 10^5 and 0.5 Ω m. Examples of semiconductors are: germanium, silicon, carbon, etc.

3. ENERGY BANDS IN SOLIDS

(a) Overlapping energy levels are termed as energy bands

(b) The energy band formed by the overlapping of valence electrons is known as valence energy band.

(c) The energy band formed by the overlapping of conduction electrons is known as conduction energy band.

(d) Electrical conduction in solid can take place only when electron remains present in its conduction energy band.

(e) The minimum energy required for exciting an electron from valence energy band to conduction energy band is known as forbidden energy gap $\left(\Delta E_g\right)$; $\Delta E_g = CEB_{min} - VEB_{max}$

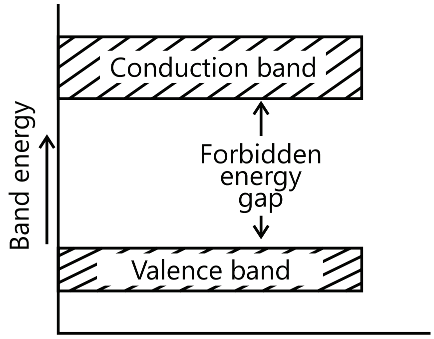


Figure 26.1

We have seen that energy levels of an isolated atom are converted into corresponding energy bands when the atom is in solid form. There is no appreciable modification in the energy levels of electrons in the inner orbits but there is a considerable modification of energy levels of valence electrons. It is because valence electrons are shared by more than one atom in the crystal. Therefore, valence electrons can be considered to be in either of the two energy bands viz lower valence band or upper or conduction band as shown in Fig. 26.1. No electron can have energy in the forbidden energy gap between two bands. Normally, the electrons reside in the valence band where they are held rather tightly to the individual atoms.

(a) **Valence band.** The range of energies (i.e. band) possessed by valence electrons is known as valence band.

The electrons in the outermost orbit of an atom are known as valence electrons. In a normal atom, valence band has the electrons of highest energy. This band may be completely or partially filled. For instance, in case of inert gases, the valence band is full whereas for other materials, it is only partially filled. The partially filled band can accommodate more electrons.

(b) **Conduction band.** In certain materials (e.g. metals), the valence electrons are loosely attached to the nucleus. Even at ordinary temperature, some of the valence electrons may get detached to become free electrons. In fact, it is these free electrons which are responsible for the conduction of current in a conductor. For this reason, they are called conduction electrons.

The range of energies (i.e. band) possessed by conduction band electrons is known as conduction band.

All electrons in the conduction band are free electrons. If a substance has empty conduction band, it means current conduction is not possible in that substance. Generally, insulators have empty conduction band. On the other hand, it is partially filled for conductors.

(c) **Forbidden energy gap.** The separation between conduction band and valence band on the energy level diagram is known as forbidden energy gap (E_g).

No electron of a solid can stay in a forbidden energy gap as there is no allowed energy state in this region. The width of the forbidden energy gap is a measure of the bondage of valence electrons to the atom. The greater the energy gap, more tightly the valence electrons are bound to the nucleus. In order to push an electron from valence band to the conduction band (i.e.to make the valence electron free), external energy equal to the forbidden energy gap must be supplied.

4. SEMICONDUCTORS

These are solids in which the forbidden energy gap between the valence band the conduction band is small, of the order of 1eV. At 0 kelvin, the valence band is completely filled and the conduction band is completely empty. At 0K, it behaves like an insulator (electron cannot absorb in infinitesimal energy because there is a forbidden gap just at the top of the valence band). At a finite temperature, (room temperature), some electrons gain energy due to thermal motion and jump from the top of the valence band to the conduction band. These electrons contribute to the conduction of electricity in a semiconductor.

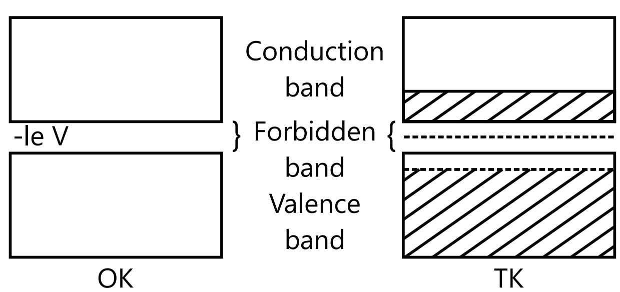


Figure 26.2

The forbidden gap in a semiconductor is small ~ 1eV. At finite temperature, some valance electrons go to conduction band. Then the formlessly is in middle of the gap

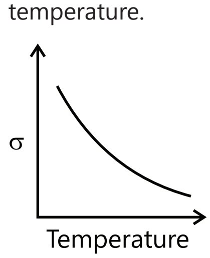
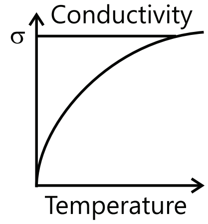
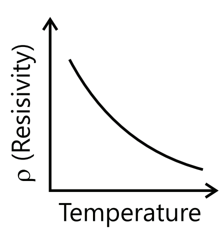
The energy gap is some semiconductors is as follows:

$\Delta E_g(\text{Silicon}) = 1.12\,\text{eV}$; $\Delta E_g(\text{germanium}) = 0.7\,\text{eV}$

$\Delta E_g(\text{Indium antimonide}) = 0.17\,\text{eV}$; $\Delta E_g(\text{Gallium arsenide}) = 1.43\,\text{eV}$; $\Delta E_g(\text{Tellurium}) = 0.33\,\text{eV}$

The energy gap decreases slightly with increases in temperature.

Table 26.1: Comparison between conductors, insulators and semiconductors

	Conductors	**Insulators**	**Semiconductors**
1	Valence band is partially filled or valence band conduction band are overlapped.	Completely empty conduction band separated from completely filled valence	At 0K, the conductive band is empty while valence band is full separated by small energy gap
2.	There is no forbidden energy gap	The forbidden gap is large E_g (diamond)~ $6\,\text{eV}$ E_g (diamond) ~ $6\,\text{eV}$	Separated by small energy gap E_g, Si=1.12 eV
3.	At room temperature, all electrons remains in the partially filled valence band or over lapped band	At room temperature, electrons do not get sufficient thermal energy to cross over and the forbidden energy band remains empty	At room temperature, many electrons have sufficient energy to go to the conduction band.
4.	Conducts electric current. Very small resistivity $\rho(\text{ohm meter})$ $\rho(\text{Cu}) = 1.7\times10^{-8}\,\Omega\text{m}$ $\rho(\text{Ag}) = 1.6\times10^{-2}\,\Omega\text{m}$ The conductivity is high $\sigma \approx 10^{7}$ to 10^{-9} ohm / m (or Siemen/m)	Does not conduct electric current (negligible conduction) very large resistively (ohm meter) $\rho(\text{glass}) \sim 10^{11} - 10^{12}\,\Omega\text{m}$ $\rho(\text{diamond}) \sim 10^{14}\,\Omega\text{m}$ very low conductivity $\sigma \sim 10^{-10}$ to $10^{-14}\,\Omega\text{m}$ very low conductivity $\sigma \sim 10^{-10}$ to 10^{-15} ohm / m (or Siemen/m)	May conduct electric current but conduction is small. Medium resistivity and medium conductivity $\rho(\text{Si}) = 2100\,\Omega\text{m}$ $\rho(\text{Ge}) = 0.47\,\Omega\text{m}$ $\sigma(\text{Ge}) \sim 2.13$ $\sigma(\text{Si}) \sim 4.7\times10^{-4}$ (ohm/m).
5.	Only electrons are the current carrier Number of free electrons (in Cu) $\sim 10^{28}$ per m^3	No current carrier (the electric conduction is almost zero for all practical purposes, see σ mentioned before)	Both electrons and holes contribute to current conduction. Number of free electrons (at room temperature) is in Ge $\sim 10^{19}$ per m^3 in Si $\sim 10^{16}$ per m^3
6.	Conductivity decreases with temperature. 	Conductivity negligibly small however increases slightly at very high temperatures. 	Conductivity increases with temperature (the resistivity / resistance decreases with temperature). The temperature coefficient of resistance of a semiconductor is negative 

4.1 Classification of Semiconductors

The semiconductors are of two types

A semiconductor in a pure form is called intrinsic semiconductor. The impurity must be less than 0.01 ppm (parts per million) i.e., less than 1 in 10^8 parts of semiconductor. At low temperature, the electrons are present in valence bonds of the semiconductor. As the temperature is increased a few electrons are raised to conduction band.

In intrinsic semiconductors: $n_e^{(0)} = n_n^{(0)} = n_i = AT^{3/2}\, e^{-\Delta Eg/2KT}$

Where $n_e^{(0)}$ is electron density in conduction band, $n_e^{(0)}$ is hole density in valance band and n_i is the density of intrinsic charge carriers.

At absolute zero temperature, all the valence electrons are tightly bound and no free electron is available for electrical condition. In the band picture, at absolute zero temperature, the conduction band is completely empty while, the valence band is fully filled. The semiconductor behaves as a perfect insulator at absolute zero temperature. At room temperature (~300K), some of the electrons may gain sufficient thermal energy and moved away from the influence of the nucleus, i.e. the covalent bond may be broken. The electron, so obtained is free to move in the crystal and conduct electricity (see Fig 26.4). The vacancy created in the covalent bond is called a hole.

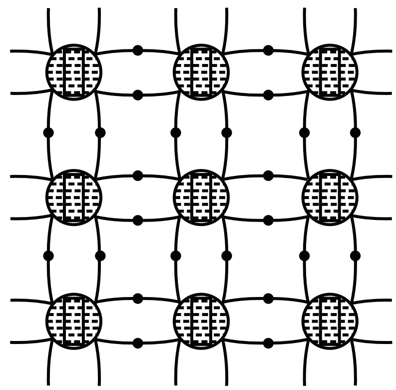


Figure 26.3

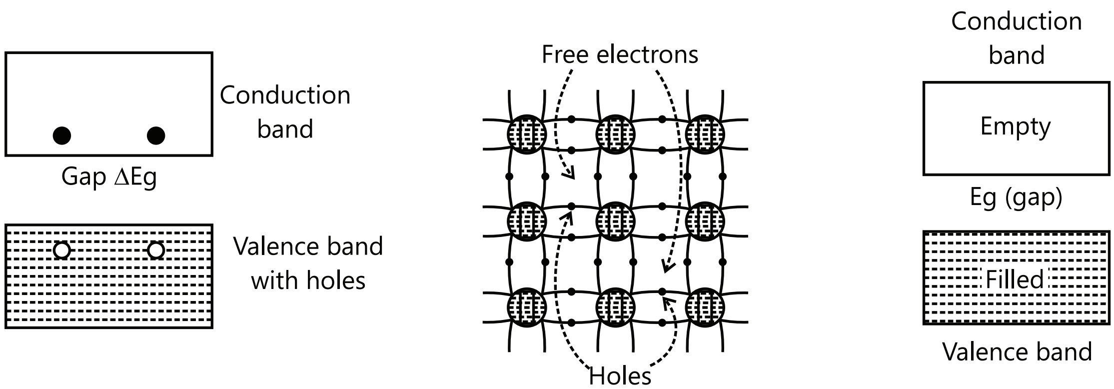


Figure 26.4

When a covalent band is broken, the electron hole pair is created. Thus, in an intrinsic semiconductor, Number of holes=number of free electrons $n_h = n_e$

4.2 Intrinsic Semiconductors

In intrinsic semiconductor, the number of free electrons and holes are equal. Both electrons and holes contribute in current conduction. For the purpose of flow of current, a hole, behaves like a positively charged particle having some effective mass. Therefore, while the electron moves from the negative electrode of the battery to the positive electrode through the semiconductor, the hole moves on opposite side.

The hole exists only inside a semiconductor. There are no holes in a metal. There, electric conduction through holes takes place inside the semiconductor only. Outside, in the metal wires, the electric current flow is due to electrons only. (In cell current flow is due to the motion of positive and negative ions).

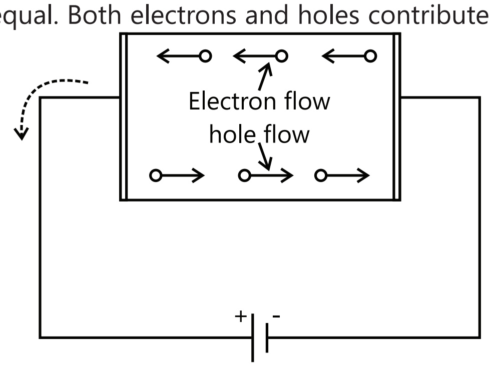


Figure 26.5

In an intrinsic semiconductor the current flow occurs due to the motion of both, the electrons and the holes. Let e=magnitude of charge on the electrons, n_h= number density of holes, μ_e =mobility of electrons an μ_h =mobility of holes, then the conductivity of intrinsic semiconductor is $\sigma = e\left(n_e\mu_e + n_h + \mu_h\right)$ Consider a block of semiconductor of length area of cross section A and having density of electron and holes as n_e as and n_h respectively when a potential difference say V is applied across it, current I flows through it as shown in Fig. 26.6 The current I is made of electron current I^e and hole current I_h.

Thus, $I = I_e + I_h$ ... (i)

If v_e is drift velocity of electrons, Then $I_e = en_e Av_e$ ...(ii)

Similarly, the hole current is given by $I_h = en_e Av_h$ ... (iii)

Using equation (ii) and (iii), the equation (i) becomes

$I = eA\,(n_e v_e + n_h v_h)$ ... (iv)

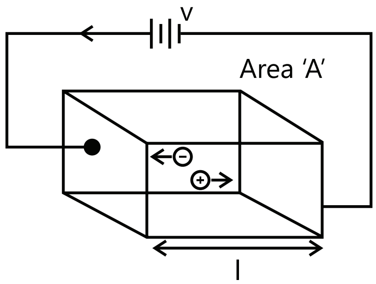


Figure 26.6

If R is the resistance offered by the semiconductor to the flow of current, then

$I = \frac{V}{R}$ or $\frac{V}{R} = eA\;(n_e v_e + n_h v_h)$ ... (v)

The electric field set up across the semiconductor is given by $E = \frac{V}{l}$ or $V = E$

Therefore, equation (v) becomes $\frac{El}{R} = eA\;(n_e v_e + n_h v_h)$ or $E\;\frac{E}{R\frac{A}{l}} = e\,(n_e v_e + n_h v_h)$

But $R\frac{A}{l} = \rho =$ resistivity of the material of the semiconductor. Therefore $\frac{E}{\rho} = e\;(n_e v_e + n_h v_h)$ ... (vi)

Mobility of electrons or holes is defined as the drift velocity acquired per unit electric field.

Therefore mobility of electrons and holes is given by $\mu_e = \frac{v_e}{E}$ and $\mu_h = \frac{v_h}{E}$

From equation (iv), we have

$$\frac{1}{\rho} = e\left\{n_e.\frac{ve}{E} + n_h.\frac{v_h}{E}\right\} \text{ or } \sigma = e\;(n_e\mu_e + n_h\mu_h) \quad \text{... (vii)}$$

Where $\sigma = e\frac{1}{\rho}$ is called conductivity of the material of the semiconductor and μ_e, μ_h are electron and hole mobilities respectively.

MASTERJEE CONCEPTS

- In pure semiconductors, at any temperature t, the carrier concentration n_e=n_h=n and the conductivity is determined by the value of E_g (width of the forbidden band)(see relations given above).
- In metal, however, the value of n is almost the same at different temperatures. The resistance arises due to interaction of free (conduction) electrons with the lattice vibrations.
- At absolute zero temperature, n=0, $\sigma = 0$ i.e., the pure semiconductor behaves like a perfect insulator. However, as temperature increases both n and σ increases. In germanium at T ≈ 300K, $n_e = n_h = 2.5 \times 10^{19}$ per m^3. The higher is the temperature, higher is the conductivity and lower is the resistivity.
- The temperature coefficient of the resistance of a semiconductor is negative.
- Pure semiconductors are of little use (may be used as heat or light sensitive resistance).

Vaibhav Krishnan (JEE 2009, AIR 22)

Illustration 1: Which one is more sensitive to heat, germanium or silicon? **(JEE MAIN)**

Sol: The band gap between conduction band and valance band in germanium (0.68 eV) is less than silicon (1.1 eV). Thus the electron in the valance band in germanium require less thermal energy (Order of KT) to transit from valance band to conduction band compared to silicon.

Germanium is more sensitive to heat. Electrons from the valence band of germanium require less energy to move from valence band to conduction band.

4.3 Extrinsic Semiconductors

A semiconductor in which impurities have been added in a controlled manner is called extrinsic semiconductor. The process of deliberately adding impurities to a semiconductor is called doping. The impurity atoms are either from V group such as arsenic (as), antimony (Sb), phosphorus (P) etc. or from III group (such as Aluminum (AI), gallium (Ga), indium (In) etc. There are two types of extrinsic semiconductors, (i) N-type (extrinsic) semiconductor and (ii) P-type (extrinsic) semiconductor. The conductivity of extrinsic semiconductor is controlled by the amount of doping, I part of a donor impurity per 10^9 parts of germanium increases its conductivity by a factor of nearly 10^3. The compounds of trivalent and pentavalent elements also behave like semiconductors, (indium antimonite), in P, GaP.

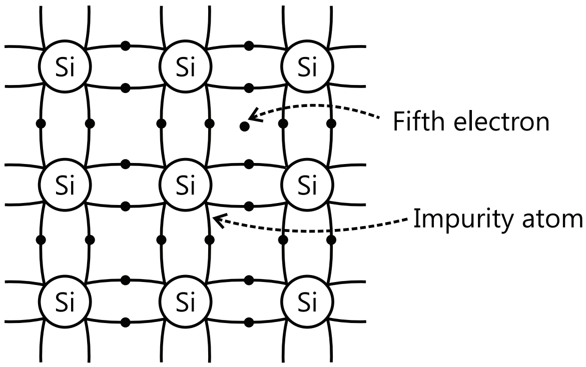


Figure 26.7

4.3.1 N-Type Semiconductor

N-Type (n-type) semiconductor is obtained by adding a small amount of pentavalent (V group) impurity to a sample of intrinsic semiconductor. The pentavalent impurities are P (phosphorus Z=15), As (Z=3), Sb (z=51), Bi(Z=83).

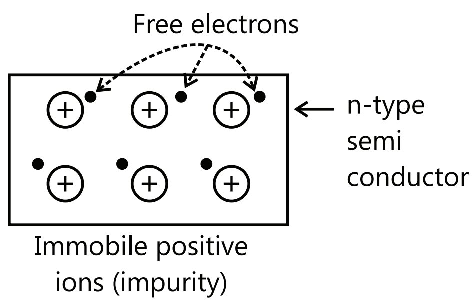


Figure 26.8

In the energy band picture we say that impurity atoms produce donor energy levels just below the conduction band. The electrons from these levels jump to the conduction band easily by gaining thermal energies (at room temperature). They may also break some covalent bonds producing electron hole pair, but their number is small. So in this type of extrinsic semiconductor, there are a large number of free electrons (donated by impurity atoms) and a negligible number of holes from broken covalent bond.

The impurity atom on donating electrons becomes positive ions. However the overall charge on the semiconductor is zero. The negative charge of the immobile positive charge of the immobile positive ions. The majority charge carriers are electrons (negative charge). Therefore, this type of extrinsic semiconductor is called n-type. The Fermi level does not lie in the middle of band gap, but it shifts towards the conduction band. The few holes formed by covalent bond breaking are called minority charge carries. The conductivity of the n-type semiconductors is controlled by the amount of impurity atoms added in it.

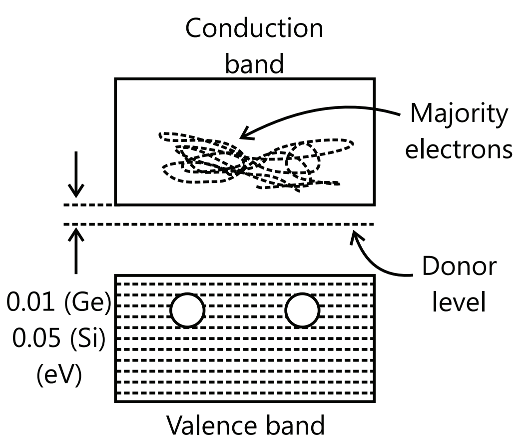


Figure 26.9

4.3.2 P-Type Semiconductor

P-Type (type) semiconductor is obtained by adding a small amount of trivalent (III group) impurity to intrinsic semiconductor. The impurities may be Boron (Z=5), AI(Z=13), Ga (Z=31), In (Z=49), TI (Z=81). For each acceptor iron there exist a hole in this type of semiconductor, there are a large number of holes present. The majority charge carriers are holes. Therefore it is called a P-type semiconductor.

In the picture, we say that accepter energy levels lie just above the valence band. These levels accept electrons from the valence band and create holes. The breaking of covalent bonds may create electron-hole pairs but their number is very little. The majority carriers are holes. The minority carriers are electrons.

The conduction takes place mainly through the motion of holes $n_h >> n_e$; $\sigma \approx e\mu_h n_h$

The overall charge on p-type semiconductor is zero. It is represented as shown in Fig. 26.11. The positive charge of free holes is balanced by the negative charge of immobile impurity ions.

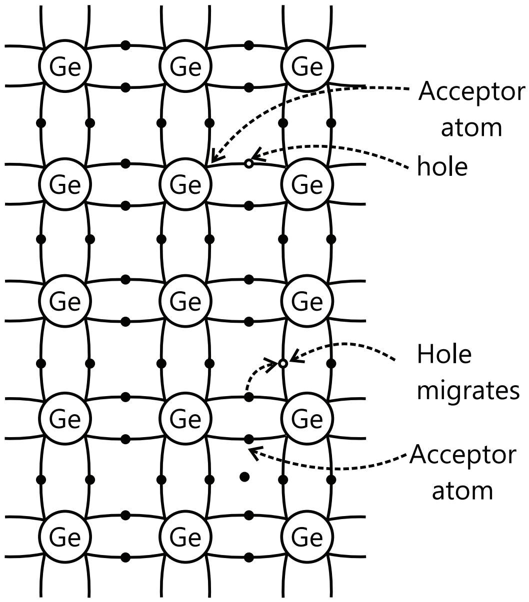


Figure 26.10

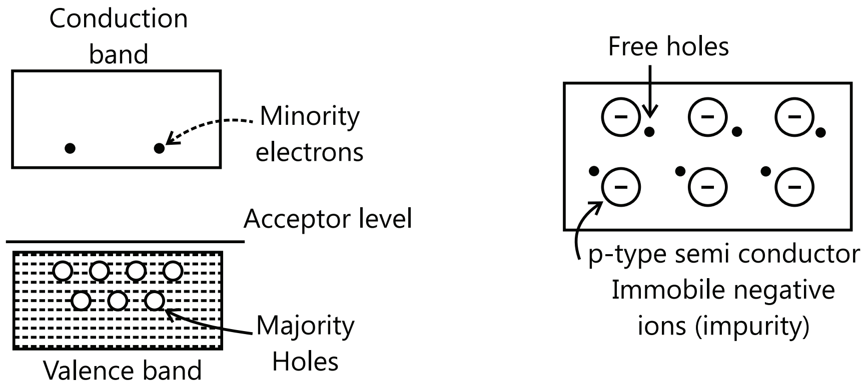


Figure 26.11

MASTERJEE CONCEPTS

When temperature is increased, covalent bonds break. This increases minority charge carriers. At very high temperature, it may happen when number of electron-hole pair obtained from bond breaking, far exceeds the charge carriers from impurities. Then the semiconductor behaves like intrinsic semiconductor. The critical temperature at which this happens is 85^0C germanium and 200^0C for silicon.

Chinmay S Purandare (JEE 2012, AIR 698)

Illustration 2: Calculate the conductivity and the resistivity of silicon crystal at 300K. It is given that $\mu_e = 1350\ cm^2/V\ s$, μ_h = 480 cm^2/ V s, and at 300K, the election- hole pair concentration is 1.072×10^{10} per cm^3. **(JEE MAIN)**

Sol: The conductivity of silicon is given as $\sigma = e \times (n_e\mu_e + n\mu_h)$ where n_e and n_h are electron and hole densities respectively. Here $n_e = n_h = n_i$ is the hole-electron pair concentration. The resistivity of the silicon is $\rho = \frac{1}{\sigma}$.

The conductivity for intrinsic semiconductor is $\sigma = e \times (n_e\mu_e + n\mu_h)$

Given, $n = 1.072\times10^{10}$ per cm^3 $= 1.072\times10^{16}$ per cm^3

Also $n_e = n_h$ for intrinsic semiconductor, further,

$\mu_e = 1350 cm^2 / V s$, = $0.1350\ m^2$/ volt ; $\mu_h = 0.048\ m^2/ V s$,

Therefore, $\sigma = 1.6\times10^{-19}\times1.072\times10^{16}\times\left(0.135+0.048\right) = 3.14\times10^{-4}\ \Omega$ m

$= 3.14\times10^{-4}$ Siemen per meter

The resistivity $r = \frac{1}{\sigma} = 10^{-4}/3.14 =$ 3185 Ω m.

Illustration 3: The concentration of acceptor atoms in a p-type germanium crystal is 4×10^{15} per cm^3. Find the conductivity of the crystal at 300K. The μ_h for germanium at 300K is 1900 cm^2 / V s. It is assumed that all the acceptor atoms are ionized at this temperature. **(JEE MAIN)**

Sol: In p-type germanium hole density is greater than electron density. The conductivity of the p-type germanium is given by $\sigma = n_h\ e\mu_h$

For extrinsic semiconductor (p-type) $\sigma = n_h\ e\mu_h$

Given $\sigma = n_h = 4\times10^{15}$ per $cm^3 = 4\times10^{21}$ per m^3

$\mu_h = 1900 cm^2$ /volt –sec

Thus $\sigma = 4\times10^{21}\times1.6\times10^{-19}\times0.190 = 1.216\times10^{2}$ ohm / m $= 1.21.6$ siemen / m

Table 26.2: Comparison of intrinsic and extrinsic semiconductors

S. No.	Intrinsic semiconductor	Extrinsic semiconductor
1.	It is a pure semiconductor with no impurity for this reason, it is called intrinsic (i.e. pure) semiconductor.	It is an impure semiconductor i.e. a controlled pentavalent or trivalent impurity added to a pure (intrinsic) semiconductor.
2.	The number of free electrons in the conduction band is equal to the number of holes in the valence band.	In an n-type semiconductor, free electrons exceed the hole. A p-type semiconductor, it is the reverse.
3.	Its electrical conductivity is low.	Its electrical conductivity is high.
4.	Its electrical conductivity depends on temperature alone.	Its electrical conductivity depends on temperature and the amount of doping.
5.	It is of no practical use.	It is used in electronic devices.

Table 26.3: Comparison of n-type and p-type semiconductors

S. No.	n-type semiconductor	p-type semiconductors
1	It is produced by adding controlled amount of pentavalent impurity to pure (intrinsic) semiconductor.	It is produced by adding controlled amount of trivalent impurity to pure (intrinsic) semiconductor.
2	The number of free electron far exceeds the number of holes.	The number of free holes far exceeds the number of electrons.
3	The current conduction is predominant by free electrons.	The current conduction is predominant by holes.
4	The donor energy level is just below the bottom of the conduction band.	The accepter energy level is just above the top of valence band.

Illustration 4: An intrinsic germanium has a resistivity of 0.47 Ω m at room temperature. Find the intrinsic carrier concentration if the mobility of electrons and holes are 0.39 m^2/ V s, and 0.19 m^2/volt-sec respectively. **(JEE MAIN)**

Sol: The resistivity of the intrinsic semiconductor is $\rho = \frac{1}{\sigma}$. Here σ is conductivity of the germanium and is given by $\sigma = n_i\left(\mu_e + \mu_h\right)$ where n_i is the concentration of electron-hole pair

Let n_i be the intrinsic carrier concentration.

Electrical conductivity, $\sigma = \frac{1}{\rho} = \frac{1}{0.47}$ Now $\sigma = n_i\left(\mu_e + \mu_h\right)$ Or $n_i = \frac{\sigma}{e\left(\mu_e + \mu_h\right)}$

Here σ =1/0.47 S/m, μ_e = 0.39 m^2/volt-sec, μ_h =0.19 m^2/volt-sec, $e = 1.6 \times 10^{-19}$ C

$$n_i = \frac{1}{0.47 \times 1.6 \times 10^{-19}\left(0.39 + 0.19\right)} = 2.3 \times 10^{-19}\ /\ m^3$$

Illustration 5: The resistivity of n-type germanium is 0.01 Ω m at room temperature. Find the donor concentration if the mobility of electrons 0.39m^2 /volt-sec. **(JEE MAIN)**

Sol: For n-type germanium the donor concentration is given by $n_d = \frac{1}{e\,\rho\,\mu_e}$

Let n_d be the donor concentration. $n_d = \frac{1}{e\,\rho\,\mu_e}$

Conductivity, $\sigma = \frac{1}{\rho} = \frac{1}{0.01} = 100\,S\,/\,m$ Now $\sigma = e\,N_d\,\mu_e$ ....n-type semiconductor

Or $n_d = \frac{\sigma}{e\,\mu_e} = \frac{100}{1.6 \times 10^{-19} \times 0.39} = 1.6 \times 10^{-21}\ /\ m^3$

Illustration 6: Mobilities of electrons and holes in a sample of intrinsic germanium at room temperature are 0.36m^2 / V s and 0.17 m^2 / V s respectively. If the electron and hole densities are each equal to 2.5×10^{19} /m^3, calculate the conductivity. **(JEE ADVANCED)**

Sol: The conductivity of intrinsic semiconductor is given by $\sigma = e\left(n_e\mu_e + n_h\mu_h\right)$ where n_e and n_h are the electrons and hole densities. As here both are equal then $n_i = n_e = n_h$ and conductivity is given by $\sigma = n_i\ e\left(\mu_e + \mu_h\right)$.

The conductivity of an intrinsic semiconductor is given by; $\sigma = n_i\ e\left(\mu_e + \mu_h\right)$

Here $n_i = 2.5 \times 10^{-19}\ /\ m^3; e = 1.6 \times 10^{-19} C; \mu_e = 0.36\ m^2\ /$ volt-sec; $\mu_h = 0.17 m^2\ /$ volt-sec

$$\sigma = 2.5 \times 10^{19} \times 1.6 \times 10^{-19}\left(0.36 + 0.17\right) = 2.12\ S/\ m$$

Illustration 7: A semiconductor is known to have an electron concentration of 8×10^{13} per cm^3 and a hole concentration of 5×10^{12} per cm^3.

(i) Is the semiconductor n-type or p-type?

(ii) What is the resistivity of the sample if the electron mobility is 23000 cm^2/ V s and hole mobility is 100 cm^2/ V s? **(JEE ADVANCED)**

Sol: For the semiconductor sample, if the hole density is less than electron density the semiconductor is N type in nature. The resistivity of the sample is given as $\rho = \frac{1}{\sigma}$ where σ is the conductivity of the sample.

(i) Since electron density $n_e\left(= 8 \times 10^{13}\ \text{per cm}^3\right)$ is greater than the hole density $n_h\left(= 5 \times 10^{12}\ \text{per cm}^3\right)$, the semiconductor is n-type.

(ii) The conductivity of the sample is given by; $\sigma = e\left(n_e\mu_e + n_h\mu_h\right)$

Here $n_e = 8\times10^{13}$ per cm^3; $n_h = 5\times10^{12}$ per cm^3

$\mu_e = 23000\ cm^2/Vs; \mu_h = 100 cm^2/Vs$

$\therefore \sigma = 1.6\times10^{-19}\times\left(8\times10^{13}\times23000+5\times10^{12}\times100\right)S\,cm^{-1} = 1.6\times184.05\times10^{-3}S\,cm^{-1}$

$\therefore$ Resistivity of the sample is given by; $\rho = \frac{1}{\sigma} = \frac{1}{1.6\times184.05\times10^{-3}} = 3.396\,\Omega cm$

Illustration 8: Determine the number density of donor atoms which have to be added to an intrinsic germanium semiconductor to produce an n-type semiconductor of conductivity $5\,\Omega^{-1}cm^{-1}$, given that mobility of conduction electrons in n-type Ge is 3900 cm^2/ Vs. Neglect the contribution of holes to conductivity. **(JEE ADVANCED)**

Sol: In n type semiconductor the number density of electrons is much greater than number density of holes. Thus we can neglect the number density of holes. Thus to produce the n type semiconductor, the donor of number density to added is found by $\sigma = en_e\mu_e$.

The conductivity of a semiconductor is given by; $\sigma = e\left(n_e\mu_e + n_h\mu_h\right)$

Neglect the contribution of holes to conductivity, we have, $\sigma = en_e\mu_e = eN_d\mu_e$ $\quad\left[\because N_d = n_e\right]$

$\therefore$ Number density of donor atoms or electron density is $N_d = \frac{\sigma}{e\,\mu_e}$

Here $\sigma = 5\Omega^{-1}\ cm^{-1}$; $\mu_e = 3900\,cm^2/Vs$; $e = 1.6\times10^{-19}C$

$$\therefore N_d = \frac{5}{3900\times1.6\times10^{-19}} = 8.01\times10^{15}\ cm^{-3}$$

Illustration 9: Suppose a pure Si crystal has 5×10^{28} atoms m^{-3}. It is doped by 1 ppm concentration of penta-valent impurity of Arsenic (As). Calculate the number of electrons and holes. Given that $n_i = 1.5\times10^{16}\ m^{-3}$.

(JEE ADVANCED)

Sol: The arsenic contains the 1 free electron in its conduction band and acts as donor impurity. The number of electron given by As to the semiconductor is given by

$$N = \frac{\text{Number of Si atom in parent crystal}\times10^{-6}}{|\text{y ppm}|}.$$

The number of holes in the semiconductor sample is given by $n_h = \frac{n_i^2}{n_e}$ here n_i is the concentration of holes and electron pair.

1 ppm =1 part per million $= \frac{1}{10^6}$

$\therefore$ Number of penta-valent atoms doped in Si crystal $= \frac{5\times10^{28}}{10^6} = 5\times10^{22}m^{-3}$

As one penta-valent impurity atom donates 1 free election to the crystal,

$\therefore$ Number of free elections in the crystal is $n_e = 5\times10^{22}\,m^{-3}$

$\therefore$ Number of holes, $n_h = \frac{n_i^2}{n_e} = \frac{\left(1.5\times10^{16}\right)^2}{5\times10^{22}} = 4.5\times10^9\ m^{-3}$

Illustration 10: The mean free path of conduction electrons in copper is about 4×10^{-8} m. Find the electric field which can give, on an average, 2 eV energy to a conduction electron in a block of copper. **(JEE ADVANCED)**

Sol: The band gap between conduction and valance band is given as 2 eV. The work done to lift electron from conduction band to the valance band is given by $W = F \times d$ where F=qE, the electric force applied on the electron.

Mean free path, d= 4×10^{-8} m; energy of electron =2eV

If E is the required electric field, then force on the electron is F=eE

The work done by the electric field on electron before it collides with copper atom is F_d. This work done is equal to the energy to the energy transferred to electron.

$\therefore F_d$=2 eV or eE×d =2eV $\therefore E = \frac{2V}{d} = \frac{2V}{4 \times 10^{-8}} = 5 \times 10^{7}\ V/m$

4.4 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.

(a) At absolute zero. At absolute zero temperature, all the electrons are tightly held by the semiconductor atoms. The inner orbit electrons are bound whereas the valence electrons are engaged in covalent bonding. At this temperature, the covalent bonds are very strong and there are no free electrons. Therefore, the semiconductor crystal behaves as a perfect insulator.

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.

(b) 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. 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 pure semiconductor, is too small to be of any practical value.

As the temperature of a semiconductor increases, 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 in the valence band. As we shall see in the next article, holes also contribute to current. In fact, hole current is the most significant concept in semiconductors.

Effect of temperature on conductivity of semiconductors

The electrical conductivity σ of a semiconductor is given by; $\sigma = e(n_e\mu_e + n_h\mu_h)$

As the temperature increases, the values of μ_e (electron mobility) and μ_h (hole mobility) decrease due to increase in their collision frequency. But with the increase in temperature, there is a large increase in charge carrier concentrations (μ_e and μ_h) due to the increase in breaking of covalent bonds. In fact, the increase in carrier concentration is so large that the decrease in the values of μ_e and μ_h has no significant effect. Hence, the conductivity of a semiconductor increases with the increase in temperature and vice-versa.

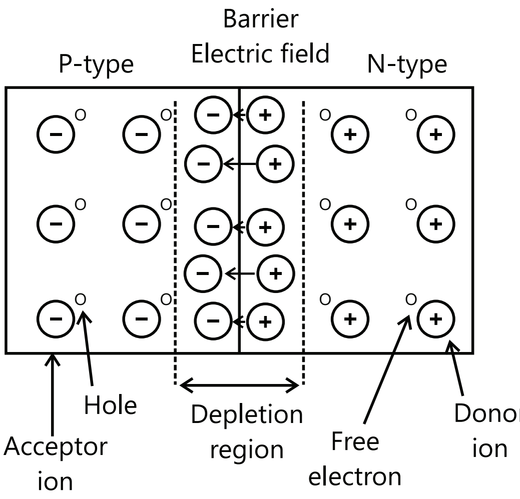


Figure 26.12

5. P-N JUNCTION

When a piece of p-type material and piece of n-type material are joined in such a manner that crystal structure remain continuous at the boundary, then a pn junction is formed. It is also called a pn junction (PN junction) diode.

AP-N junction cannot be made by simply pushing the two pieces together as it would not lead to a single crystal structure. There are special fabrication techniques to form a PN junction.

Immediately after a PN junction is formed, the following processes are initiated:

(a) The negative ions on P-sides and positive ions on N-sides are immobile. The majority holes from P region diffuse into N region, and the majority electrons from N region diffuse into P region,

(b) Due to the above, the electrons and holes at the junction region recombine and disappear (i.e. covalent bonds are completed).

(c) As a result, a layer of negative ions on P-side and a layer of positive ions on N-side is formed at the junction. In this region, due to recombination of electrons and holes, depletion of free charge carriers occurs. So this region is called depletion region. The charge density on the two sides of the junction (due to ion layers) is shown in Fig. 26.13 (A).

(d) The uncompensated ion layers in the depletion region generate an electric field in this region. The electric field points from N side to P side. This electric field prevents further diffusion of holes from P-sides. It also prevents further diffusion of electrons from the N side to P side. The electric field is called barrier electric field.

(e) The barrier electrical field gives rise to a difference of potential from one side to the other side. This is called barrier potential (or potential barrier). For silicon PN junction the barrier potential is about 0.7 V while for germanium PN junction, it is about 0.2V.

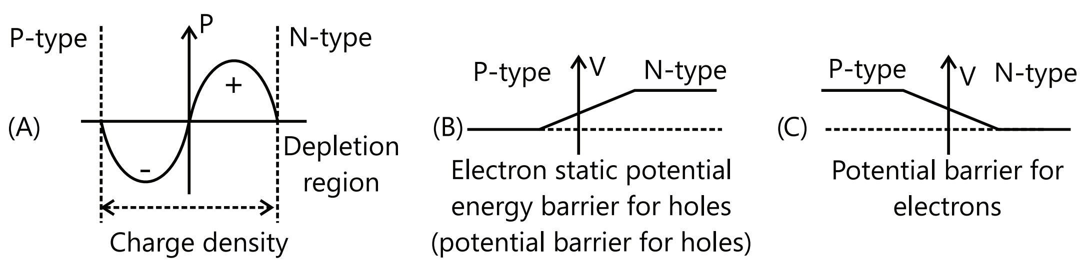


Figure 26.13

(f) For holes the potential on the N-sides is higher. Holes cannot cross the depletion region because of this barrier potential. Figure 26.13 (B). For electrons the potential barrier is shown in Figure 26.13 (C).

(g) On the average the potential barrier height in PN junction is ~0.5V and the width of the depletion region $\sim 1\mu m$ or 10^{-6} m. The barrier electric field is thus $E = \frac{V}{d} = \frac{0.5}{10^{-6}} = 5 \times 10^5$ Volt / m

5.1 P-N Junction with Forward Bias

(a) When the position terminal of a battery is connected to the P-side and the negative terminal to the N-side of a PN-junction, then it is said to be forward biased (See Fig 26.14).

(b) The holes are repelled from the positive terminal and completed to move towards the junction. The electrons are also repelled from the negative terminal and move towards the junction. This reduce the depletion region for a forward biased PN-junction

(c) The potential barrier is reduced. More charge carriers diffuse across the junction.

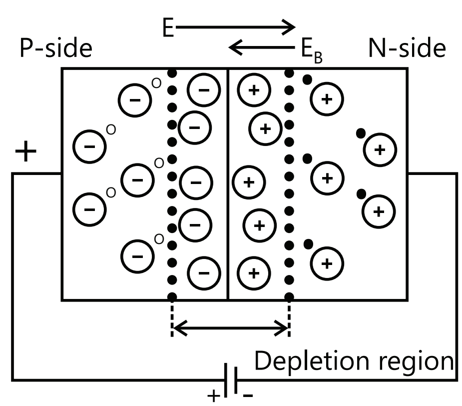


Figure 26.14

(d) In the P-type material, near the positive terminal, an electron breaks the covalent bond and goes to the battery. As a result a hole is created in P-sides. At the same time an electron enters the N sides from the negative terminal. The current in the P-region arises due to flow of the holes while the current in the N-region is due to electrons.

(e) The electric field at the barrier, due to the Battery is from P side to N side (forward bias). This is an opposition to the barrier electric field.

(f) If battery potential is increased, the potential barrier is further reduced. More majority carriers diffuse across the junction and the current increases.

5.2 P-N Junction with Reverse Bias

(a) When the positive terminal of a battery is connected to the N-side and negative terminal is connected to the P-side of the PN junction, then it is said to be reverse biased

(b) The holes in the P-region are attracted towards the negative terminal and the electrons in the N-region are attracted towards the positive terminal. Thus the majority carriers move away from the junction. The depletion region increases for a reverse biased PN-junction.

(c) The barrier potential increases, this makes it more difficult for the majority carriers to diffuse across the junction.

(d) A very little current called reverse saturation current flows due to minority carrier flow. It is of the order nano amperes (10^{-9} A) for silicon and micro amperes (10^{-6} A) for germanium PN-diodes.

(e) In reverse bias situation, the junction behaves like a high resistivity material sandwiched in between two regions.

(f) The effective capacitance of PN junction in the reverse bias condition is of the order of few pico farads.

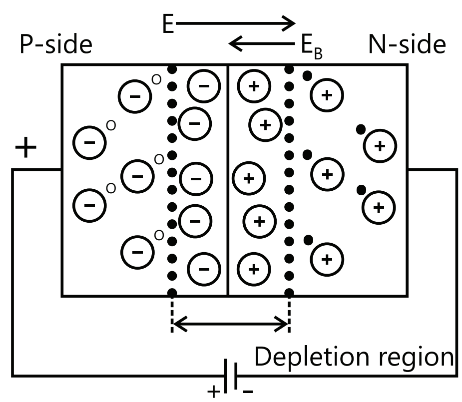


Figure 26.15

6. SEMICONDUCTOR DIODE

6.1 Forward and Reverse Bias Characteristics

Forward bias characteristics: The circuit diagram for studying the V-I characteristics of a PN junction diode in forward bias is shown in the Fig. 26.16.

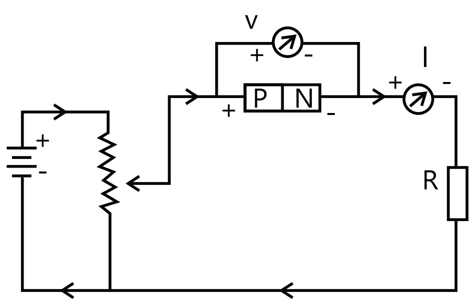


Figure 26.16

In forward bias the depletion region decreases, the barrier potential decrease, and the current flows due to diffusion of charge carriers across the junction. Majority holes from P side cross over to N side, and majority electrons from N sides cross over to P sides. The current voltage characteristic is shown in Fig. 26.17.

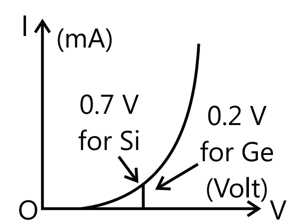


Figure 26.17

The diode current is negligibly small for first few tenths of a volt. The reason is that the diode does not conduct till the external voltage V, overcomes the barrier potential. The voltage at which the current starts to increase rapidly is called cut-in or knee voltage (V_0) of the diode. For a silicon diode V_0~0.7 volt for germanium V_0~0.2 volt junction diode in reverse bias is shown in Fig. 26.18.

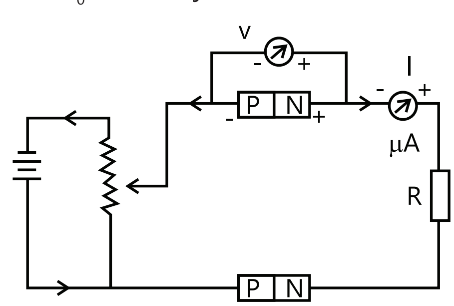


Figure 26.18

Reverse Bias characteristics: In reverse bias state, the depletion region increases and potential barrier also increases. The majority holes in P region and majority electrons in N region, now do not cross the junction. This does not give rise to any current.

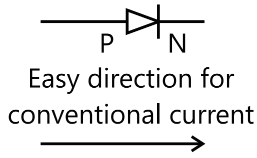


Figure 26.19

In reverse bias a very small current flows. This arises due to the flow of minority charge carriers across the junction. The reverse current is only few μ A for germanium diodes and only a few nA for silicon diodes. It remains small and almost constant for all reverse bias voltages less than the break down voltage V_z. At breakdown, the current increases rapidly for small increase in voltage. The full characteristics, forward and reverse bias are shown in the Fig. 26.19. The PN junction diode thus is a unidirectional device. Large current (mA) flows in one direction, but negligible current flows in the reverse direction.

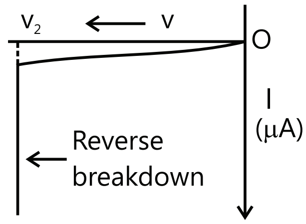


Figure 26.20

The symbol used for PN junction diode is shown in Fig. 26.20.

The equation for diode current is $\mu \quad I = I_o\left(e^{eV/kT} - 1\right)$

Where I_0 is called saturation current, V is positive for forward and negative reverse bias, k is Boltzmann constant, T is temperature and e= $1.6\times10^{-19}\,C$.

6.2 Reverse Breakdown

If the reverse bias voltage is made too high, the current through the PN junction increases rapidly at V_z (see Fig 26.21). The voltage at which this happens is called breakdown voltage or zener voltage.

There two mechanism which causes this breakdown. One is called zener breakdown and the other is called avalanche breakdown.

Zener breakdown: When reverse bias is increased the electric field at the junction also increases. At some stage the electric field becomes so high that it breaks the covalent bonds creating electron-hole pairs. Thus a large number of carriers are generated. This causes a large current to flow. This mechanism is known as zener breakdown.

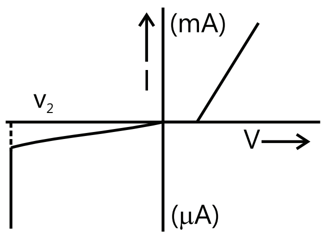


Figure 26.21

Avalanche breakdown: At high reverse voltage, due to high electric field, the minority charge carriers, while crossing the junction acquires very high velocities. These by collision breaks down the covalent bonds, generating more carriers. A chain reaction is established, giving rise to high current. This mechanism is called avalanche breakdown. In other words the covalent where the junction break down liberating a large number of electron hole pairs. Then the reverse current increases abruptly to high value. This is called avalanche break down and may damage the junction. This phenomenon is used to zener diode and used in voltage regulator.

Illustration 11: In a pure germanium sample, the electron and hole mobilities at room temperature are $3600 cm^2/V s$, $1700 cm^2/V s$ respectively. If the electron-hole concentration is $2.5\times10^{19}\ m^{-3}$, then the conductivity of Ge is **(JEE MAIN)**

Sol: As the electron-hole concentration is given the conductivity of the germanium is given by $\sigma = n_i e\left(\mu_e + \mu_n\right)$ where μ_e and μ_h are electron and hole mobilities respectively.

$\sigma = n_i e\left(\mu_e + \mu_n\right)$; $(\because \quad n_i = p_i)$

$\sigma = 2.5\times10^{19}\times1.6\times10^{-19}\left(.36 + .17\right) = 2.12\ \mho/m$

7. JUNCTION DIODE AS RECTIFIER

A device which converts alternating current (a.c.) into direction current (d.c.) is known as rectifier. The process of converting a.c. into d.c is known a rectification.

Principle: Junction diode conducts only when forward biased and not conduct when reverse biased. It acts as a valve. This fact makes the junction diode to work as a rectifier.

7.1 Junction Diode as a Half Wave Rectifier

The Rectifier which converts only one half of a.c. into d.c. is called halfwave rectifier. The circuit diagram of half wave rectifier is shown in Fig. 26.22.

The a.c. input signal to be rectified is fed to the primary (P) coil of the transformer. The secondary (S) coil is connected to the junction diode through a load resistance R_L. The output signal is obtained across the load resistance R_L.

Working: Since upper end of secondary coil is connected to p-region and lower end is connected to the n-region of the junction diode, so the junction diode is forward biased during the positive half of input a.c. The output voltage is obtained across the load resistance R_L. Upper end of the secondary coil becomes negative and lower and becomes positive. So the junction diode is reverse biased. Hence the junction does not conduct and we get no output across the load resistance during negative half of input a.c.

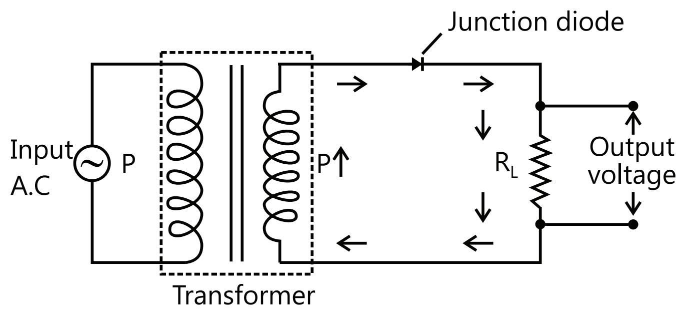


Figure 26.22

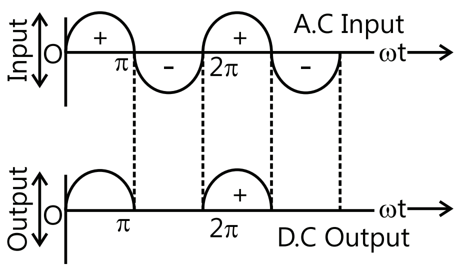


Figure 26.23

Disadvantages:

(a) Since the output signal is discontinuous, so the efficiency of half wave rectifier is small.

(b) The output is not pure d.c. but it is a fluctuating (or pulsating a.c.) which contains a.c. components or ripples also.

Expression for output d.c. voltage: Output d.c. voltage = Mean load current × load resistance i.e. $V_{d.c} = I_{d.c.} \times R_L$

But $I_{d.c.} = \frac{I_o}{\pi}$, where I_o is the maximum value of the secondary half wave current $\therefore V_{d.c.} = \frac{I_o}{\pi} \times R_L$

7.2 Junction Diode as a Full Wave Rectifier

Full wave rectifier converts both halves a.c. input signal to d.c. output.

The p-regions of both the diode the D_1 and D_2 are connected to the two ends of the secondary coil (s). The load resistance R_L across which output voltage is obtained is connected between common point of n region of diodes and central tapping of the secondary coil.

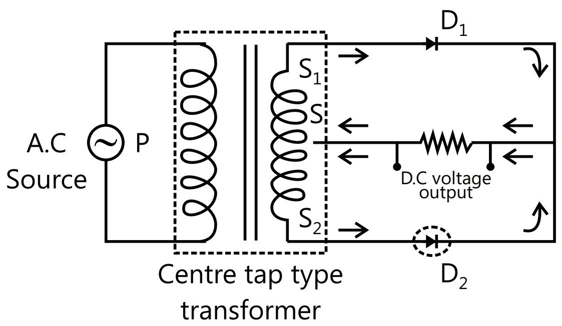


Figure 26.24

Working: The upper end of the secondary coil becomes positive while the lower end becomes negative. Thus, diode D_1 is forward biased and diode D_2 is reverse biased, so the current due to diode D_1 flows through the circuit in a direction shown by arrows (above R_L).The output voltage which varies in accordance with the input half cycle is obtained across the load resistance (R_L).

During negative half cycle of input a.c. signal. Diode D_1 is reverse biased and diode D_2 flows through the circuit in a direction shown by arrows (below R_L). The output voltage is obtained across the load resistance (R_L).

Since both the halves of input a.c. (wave) are rectified, so the junction diode is called a full wave rectifier.

Advantage: In full wave rectifier, output is continuous, so its efficiency is more than that of the half wave rectifier.

However, the output is again fluctuating (or pulsating d.c.) which can be smoothened by using a filter circuit.

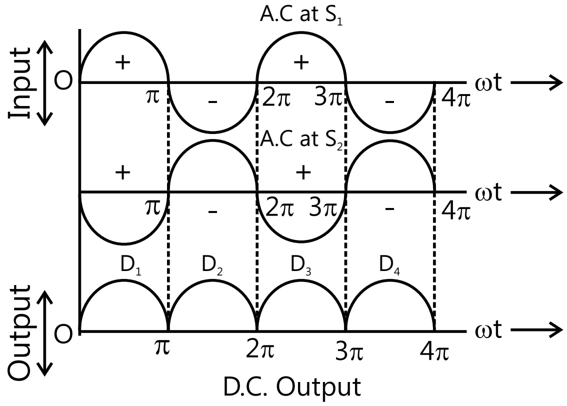


Figure 26.25

Expression for output d.c. voltage: Output d.c. voltage=Mean load current × load resistance i.e. $V_{d.c.} = I_{d.c.} \times R_L$

But $I_{d.c.} = \frac{2I_o}{\pi}$, where I_o is the maximum value of the secondary full wave current

$$\therefore V_{d.c.} = \frac{I_o}{\pi} \times R_L$$

Thus, output d.c. voltage in case of full wave rectifier is twice the output d.c. voltage in case of half wave rectifier.

MASTERJEE CONCEPTS

S. No.	Half wave rectifier	Full wave rectifier
1.	$I_{ac} = I_{dc} = \frac{I_o}{\pi}$	$I_{ac} = \frac{2I_o}{\pi}$
2.	$E_{ac} = E_{dc} = \frac{V_o}{\pi}$	$E_{ac} = \frac{2V_o}{\pi}$
3.	$r = 1.21 \ \therefore I_{ac} > I_{dc}$	r= 0.48, $\therefore I_{ac} < I_{dc}$
4.	$\eta = \frac{0.406}{r_p} ; 1 + \frac{r_p}{R_L}$	$\eta = \frac{0.812}{r_p} ; 1 + \frac{r_p}{R_L}$
5.	Form Factor =1.57	1.11
6.	Ripple frequency =ω	2ω
7.	Pulse frequency $= \frac{\text{input pulse frequency}}{2}$	Pulse frequency = input pulse frequency

Yashwanth Sandupatla (JEE 2012, AIR 821)

7.3 V-I Characteristics of Junction Diode

Volt-ampere or V-I characteristic of a junction diode (i.e. pn junction) is the curve between voltage across the diode and current through the diode. Since a junction diode may be forward biased or reverse biased, it has two type of V-I characteristics viz.

(a) Forward characteristics (b) Reverse characteristics.

(a) **Forward characteristics.** It is the graph between forward voltages (V_F) applied across the junction diode and the resulting forward current (I_F) through the diode. Figure 26.26(A) shows the circuit arrangement for determining the forward characteristics of a junction diode. Note that R is current limiting resistance and prevents the forward current from exceeding the permitted value.

The forward current is due to the majority carriers.

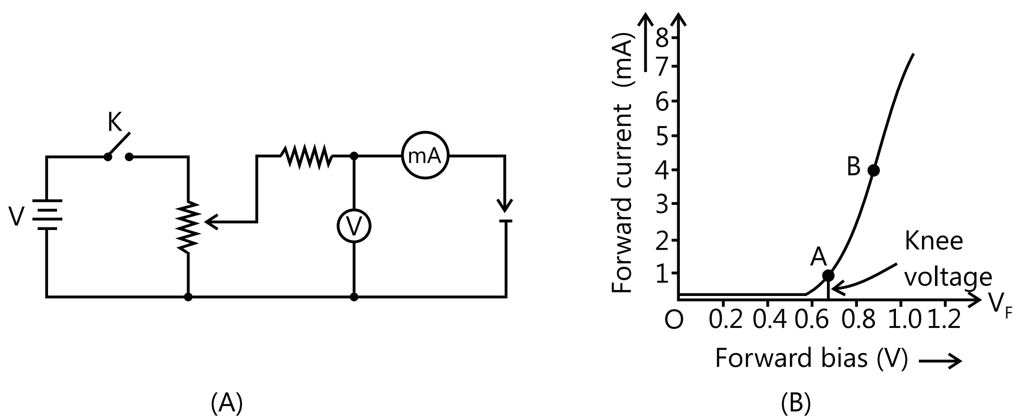


Figure 26.26

The forward voltage (V_F) across the junction diode is increased from zero in steps and the corresponding values of forward current (I_F) through the diode are noted. If we plot the graph between V_F and I_F we get the forward characteristic OAB of the junction diode as shown in Fig. 26.26 (B). The shape of this curve can be explained as under:

When the applied forward voltage is zero i.e. circuit is open at K, the barrier potential V_0 at the junction does not permit current flow. Therefore, the forward current I_F is zero as indicated by point 0 in Fig. 26.26 (b). As the forward voltage is increased from zero, the forward current increases very slowly (curve 0A) until the forward voltage across the diode reaches V_0 (=0.7V for silicon diode and 0.3V for germanium diode) at the knee of the curve. The forward voltage corresponding to knee of the curve is called knee voltage. Once the applied forward voltage exceeds the knee voltage, the forward current increases rapidly (curve AB).

The forward voltage at which the current through the diode starts to increase rapidly with increase in forward voltage is called knee voltage. For silicon diode, knee voltage=0.7V while for germanium diode, knee voltage= 0.3V.

Below the knee voltage, the curve is non-linear. But once the forward voltage exceeds the knee voltage, the diode behave like an ordinary conductor. Therefore, forward current rises sharply with increase in forward voltage (curve AB). The curve is now almost linear.

(b) **Reverse characteristics.** It is the graph between the reverse voltage (V_R) applied across the junction diode and the reverse current (I_R) through the diode. Figure 26.27(A) shows the circuit arrangement for determining the reverse characteristics of a junction diode. Note that the diode is reverse biased.

The reverse voltage (V_R) across the junction diode is increased from Fig. 26.27 (B) zero in steps and the corresponding value of reverse current (I_R) are noted. If we plot the graph between V_R and I_R, we get the reverse characteristics OCD of the junction diode as shown in Fig. 26.27(B). The shape of reverse characteristic of the diode can be explained as under:

Since the diode is reverse biased, its resistance is very high and practically no current flows through the circuit.

However, in practice, a very small reverse current (of the order of μA) flows with reverse bias as shown in Fig. 26.27(B). This is called reverse saturation current because its value practically remains constant until reverse breakdown voltage (V_{BR}) is reached. The reverse saturation current is due to minority carriers. It may be recalled that are a few free electrons in p-type material and a few holes in n-type material. These undesirable free electrons in p-type and hole in n-type are called minority carriers. To these minority carriers, the applied reverse bias appears as forward bias. Therefore, a small reverse current (I_R) flows in the circuit.

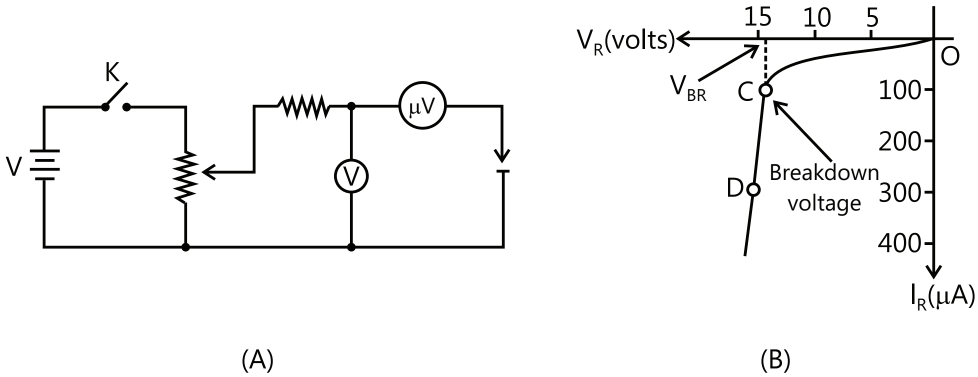


Figure 26.27

As shown in Fig. 26.27(b) when the reverse voltage becomes equal to reverse breakdown voltage V_{BR}, the reverse current increases very rapidly. Now reverse voltage remains approximately constant at V_{BR} but reverse current I_R increases very sharply resulting in overheating and possible damage. This large value of I_R is due to the fact that the kinetic energy of electrons (minority carriers) becomes high enough to knock out electrons from semiconductor atoms. Thus there is sudden decrease in resistance of the junction and abrupt rise of reverse current.

7.4 Dynamic or A.C. Forward Resistance of Junction Diode

It is the opposition offered by the junction diode to the changing forward current and may be defined as under:

The ratio of change in forward voltage across the diode to the resulting change in current through it is called a.c. forward resistance of the diode.

$$\text{a.c. forward resistance, } r_f = \frac{\text{change in forward voltage across the diode}}{\text{corresponding change in current through diode}}$$

The a.c. forward resistance is more significant as the diodes are generally used with alternating voltages. The a.c. forward resistance can be determined from the forward characteristic as shown in Fig. 26.28. If P is the operating point at any instant, then forward voltage is ob and forward current is OE. To find the a.c. forward resistance, vary the forward voltage on both sides of the operating point equally as shown in Fig. 26.28. where AB=BC. It is clear from this figure that:

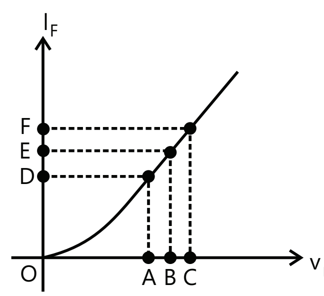


Figure 26.28

For forward voltage OA, circuit current is OD. For forward voltage OC, circuit current is of.

$$\therefore \text{a.c. forward resistance is; } r_f = \frac{\text{change in forward voltage}}{\text{change in forward current}} = \frac{OC - OA}{OF - OD} = \frac{AC}{DF}$$

It may be mentioned here that forward resistance of a crystal diode is very small. Ranging from 1 to 25 Ω. Note that above the knee point in the forward characteristic, the curve is linear. Therefore, above knee point, r, is independent of the forward applied voltage.

A.C. reverse resistance. The a.c. reverse resistance of a junction diode is very large and may be considered infinite for all practical purposes. For reason, a reverse diode practically conducts no current.

8. ZENER DIODE

A property doped P-N junction diode which works in the breakdown region without damaging itself is called a zener diode.

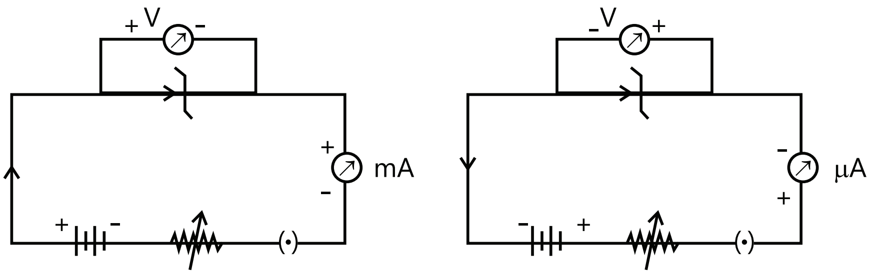
Zener diode is also known as breakdown diode. It is mainly as a voltage regular.

Symbolic representation of zener diode is made as 

The breakdown voltage zener voltage V_z depends on the concentration of doping. Both n and p regions of zener diode are heavily doped. The depletion layer is very thin. Since electric field,

E=-dV/dr, the electric field across the junction will be very high.

Volt-ampere characteristics of zener diode



(A) Forward biased (B) Reverse biased

Figure 26.29

Zener diode operates in the breakdown region (reverse bias).

Zener Diode as a voltage regulator: A zener diode can be used as voltage regular or stabilizer to provide a constant voltage from a source. The zener diode is connected across the fluctuating voltage source through a dropping resistor of resistance R_s. The constant voltage supply is obtained across the load R_L.

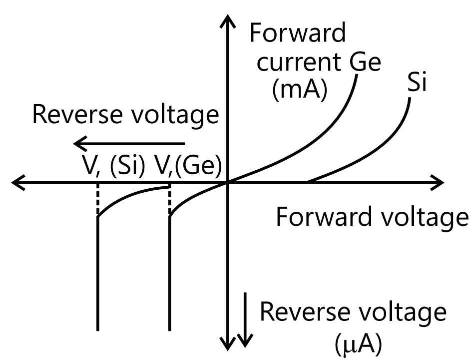


Figure 26.30

The 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_s absorbs the output voltage fluctuation so as to maintain constant voltage across the load R_L.

(a) 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. 26.32 (ii) It is clear that output voltage remains constant at V_z (=E_0). The excess voltage is dropped across the series resistance R_S. This will cause an load current remains constant. Hence, output voltage E_0 remains constant irrespective of the changes in the input voltage E_{in}.

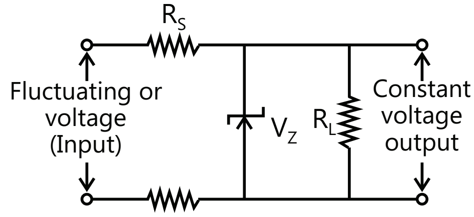


Figure 26.31

(b) Now suppose that input voltage is constant but the load resistance R_L decreases. This will cause an increase in load current. The extra current cannot come from the source because drop in R_S (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 I_z. Consequently, the output voltage stays at constant value.

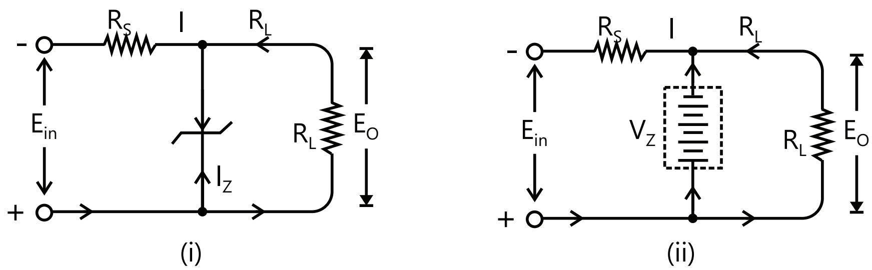


Figure 26.32

Voltage drop across $R_S = E_{in} - E_0$

Current through R_S, $I = I_Z + I_L$

Applying ohm's law, we have, $R_s = \frac{E_{in} - E_0}{I_Z + I_L}$

9. PHOTO DIODE

The junction diode which conducts when charge carriers are generated by the photons i.e., light incident on it is known as optoelectronic junction device.

A reverse biased special p-n junction diode having transparent window is known as photo diode and when it is illuminated with light, the reverse diode current varies linearly with the light flux.

Construction: A reverse biased p-n junction diode is enclosed in a clear plastic envelope. Light is allowed to fall on the surface of the plastic facing the diode. The output voltage is taken across the load resistance R_L.

Symbolic representation of a photo diode is shown in the Fig. 26.33.

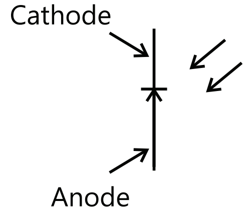


Figure 26.33

Principle. When a rectifier diode is reverse biased, it has a very small reverse leaked current. The same is true for a photo-diode. The reverse current is produced by thermally generated electron-hole pairs which are swept across the junction by the electric field create by the reverse voltage. In a rectifier diode, the reverse current increases with temperature due to an increase in the number of electron hole pairs. A photo-diode differs from a rectifier diode in that when its pn junction is exposed to light, the reverse current increases with the increase in light intensity and vice-versa. This is explained as follows. When light (photons) falls on the pn junction, the energy is imparted by the photons to the atoms in the junction. This will create more free electrons (and more holes). These additional free electrons will increase the reverse current. As the intensity of light incident on the pn junction increases, the reverse current also increases. In other words, as the incident light intensity increases, the resistance of the device (photo-diode) decreases. Figure 26.33 shows the schematic symbol of a photo-diode. The inward represent the incoming light.

Photo-diode operation. Figure 26.33 shows the basic photo-diode circuit. The circuit has reverse biased photos-diode, resistor R and d.c. supply. The operation of photo-diode is as under:

(a) When no light is incident on the pn junction of photo-diode, the reverse current I_r is extremely small. This is called dark current.

The resistance of photo-diode with no incident light is called dark resistance (R_R).

Dark resistance of photo-diode, $R_R = \frac{V_R}{\text{Dark current}}$

(b) When light is incident on the junction of the photo-diode, there is a transfer of energy from the incident light (photons) to the atoms in the junction. This will create more free electrons (and more holes). These additional free electrons will increase the reverse current.

(c) As the intensity of light increases, the reverse current I_R goes on increasing till it becomes maximum. This is called saturation current.

Reverse current versus illumination curve. Figure 26.34 shows the graph between reverse current (I_R) and illumination (E) of a photo-diode. The reverse current is shown on the vertical axis and is measured in μA. The illumination is indicated on the horizontal axis and is measured in m W/cm^2. Note that graph is a straight line passing through the origin.

$\therefore I_R = m\,E$ Where m= slope of the straight line

The quantity m is called is called the sensitivity of the photo-diode.

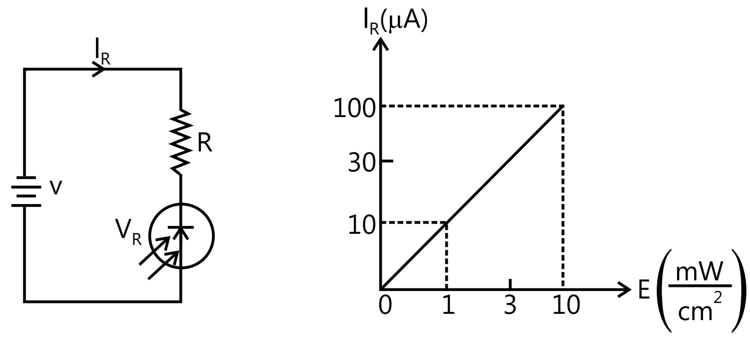


Figure 26.34

Volt- ampere characteristics of photo diode. When photo diode is reverse biased, then a constant current known as saturation current I_0 due to thermally generated minority carriers flows in the circuit. This current is also known as dark current.

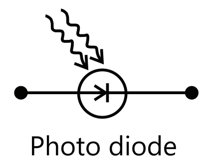


Figure 26.35

When light of energy (hv) more than the energy gap (E_g) of semiconductor falls on the photo-diode, additional electron-hole pairs are formed. The electron-hole pairs formed are proportional to the intensity of the incident light or the number of incident photons.

These electrons holes diffuse through the junction and hence current I_s also flows in addition to the dark current I_0. Thus the electric current I_0 is proportional to the intensity of incident light. Hence, the total reverse current is given by. $I = I_0 + I_s$

Total reverse current in a photo-diode increase with the increase in the intensity of the incident light.

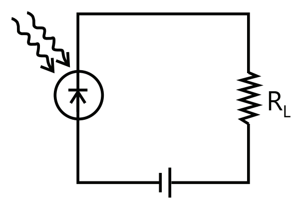


Figure 26.36

Uses of photo diodes:

(a) Photo diodes are used as photo detectors intensity of radiation.

(b) They are used as light operated switches.

(c) They are used in optical communication equipment's.

(d) They are used in fast reading of film sound tracks and tapes.

(e) They are used in logic circuits.

(f) They are used as optical demodulators.

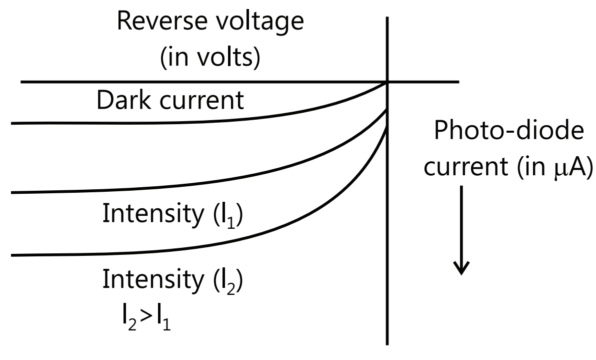


Figure 26.37

MASTERJEE CONCEPTS

Photo-diodes are operated in reverse photo-diode is used to detect photo radiation.

Light variations affect minority carrier based reverse current much more than the forward current

Anand KJEE 2011, AIR 47

10. LIGHT EMITTING DIODE (LED)

A special heavily doped P-N junction diode which emits spontaneous radiation when forward biased is known as light emitting diode (LED).

The symbolic representation of LED is snown in the Fig. 26.38:

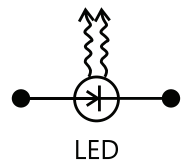


Figure 26.38

LEDs made of elemental semiconductor like germanium (Ge) and silicon (Si) emit energy in the form of infra-red (or heat) radiation.

LEDs made of Ga As of E_g~1.4 eV emits infrared radiations.

LED made of Ga As 0.6 P 0.4 eV E_g~1.9 eV emits red light.

Theory: When a p-n junction diode is forward biased (See Fig. 26.39), the electrons injected to p-side of the junction diode falls from the conduction band to the valence band and recombine with the holes in the valence band. [This is equivalent to the jumping of electrons from higher energy state (i.e., conduction band) to lower energy state (i.e., valence band)]. Hence, energy is known as electro-luminescence. The energy of the photon of visible light by hv =E_g, where E_g is the energy gap between conduction band and valence band v is the frequency of emitted visible radiation. The wavelength of the emitted light is given by $\frac{hc}{\lambda} = E_g$ or $\lambda \frac{hc}{E_g}$

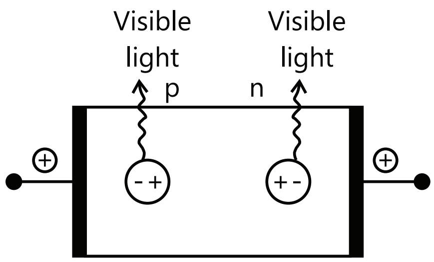


Figure 26.39

Advantage of LED

(a) Light emitting diodes are easily manufactured.

(b) LEDs have low cost.

(c) LED works at low voltage as compared to the incandescent bulb.

(d) LED has longer life than the incandescent bulb.

(e) They can be switched on and off very fast so they can be used as blinkers.

(f) Now warm up time is taken by them,

(g) They can emit monochromatic light as well as white light.

Uses of light emitting diodes

Light emitting diodes have the following uses:

(a) They are used as indicator lamps.

(b) They are used in digital displays in watches and calculators.

(c) Light emitting diodes which emit infra-red light are used in burglar alarm.

(d) They are used in remote control schemes.

(e) They are used as blinkers.

(f) They are used as decorating lights.

(g) They are used as light lamp and in torches.

11. SOLAR CELL OR PHOTO-VOLTAIC DEVICE

A special p-n junction diode which converts solar energy (sun light) into electrical is known as solar cell or photo-voltaic device. Junction surface of these diodes is kept large so that large radiations are caught.

A simple solar cell consists of a p-n junction of which n-region is very thin and p-region is thick.

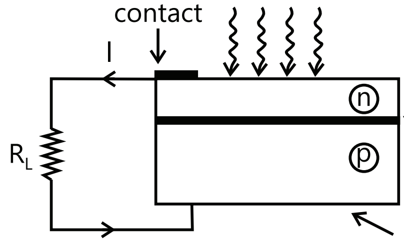


Figure 26.40

Action:

(a) When solar energy or light energy falls on the cells, electron-hole pairs are generated in both n-region and p-region of the junction in diode.

(b) The electrons from p-region diffuse through the junction to n-region and holes from n-region diffuse through the junction to the p-region due to electric field of depletion layer.

(c) If p-n junction diode is open circuited, then holes and electrons will collect or accumulate on the two sides of the junction. This gives rise to an open circuit voltage V_0.

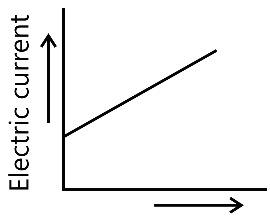


Figure 26.41

V-I characteristics: A typical V-I characteristic of a solar cell is shown in Fig.26.42. V_0 is open circuit voltage of the solar cell and I_s is the maximum current i.e., short circuit which can be drawn from the cell.

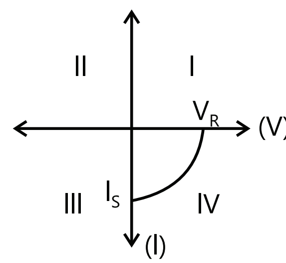


Figure 26.42

Uses of solar cell:

(a) Solar cells are used in street lights

(b) They are used in solar heaters.

(c) They are used in power supply of satellites and space vehicles.

(d) They are used in calculators.

Illustration 12: A diode used in the circuit shown in Fig. 26.43. has a constant voltage drop of 0.5 V at all currents and a maximum power rating of 100 mW. What should be the value of resistance R connected in series with diode for obtaining maximum current? **(JEE MAIN)**

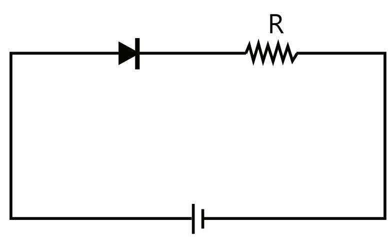


Figure 26.43

Sol: The power dissipated across diode is given by $P_{max} = V.I_{max}$ where I_{max} is the maximum current through the diode. The value of resistance across the diode is given by Ohm's law.

Applied voltage, E=1.5V

Voltage drop across diode V_d = 0.5V

Max. Power rating of diode, P_{max}=100 mW =0.1 W

The maximum current (I_{max}) that diode can carry safely is $I_{max} = \frac{P_{max}}{V_d} = \frac{0.1}{0.5} = 0.2A$

Voltage drop across resistance R=E-V_d=1.5-0.5=1.0V

$$\therefore\ R = \frac{\text{Voltage drop across R}}{I_{max}} = \frac{0.1}{0.2} = 5\Omega$$

Illustration 13: A battery of 2 V is connected across the points A and B as shown in Fig. 26.44. Find the current drawn from the battery if the positive terminal is connected to (i) The points A and (ii) The point B. Assume that the resistance of each diode is zero in forward bias and infinite in reverse bias. **(JEE ADVANCED)**

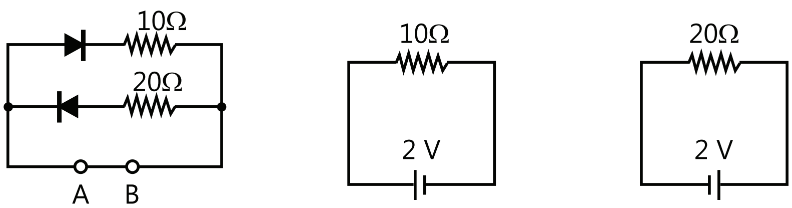


Figure 26.44

Sol: When the diode is connected in forward bias condition the resistance of diode is zero and hence the current through the diode is maximum. While the diode is in the reverse biased condition the resistance is infinite thus the circuit acts as open circuit. Thus the effective current in the circuit is obtained using Ohm's Law.

(a) When positive terminal of the battery is connected to point A, diode D_1 is forward biased and offers zero resistance while diode D_2 is reverse biased and offers infinite resistance therefore, diode D_1 may be replaced by a wire while diode D_2 is open- circuited.

The circuit then reduced to that shown in Fig. 26.44.

$\therefore$ Current drown from battery, $I = \frac{2V}{10\Omega} 0.2A$

(b) When the positive terminal of the battery is connected to point B, diode D_1 is reverse biased and offers infinite resistance while diode D_2 is forward biased and offers zero resistance. The circuit then reduces to that shown in Fig. 26.44.

$\therefore$ Current drown from battery, $I = \frac{2V}{20\Omega} 0.1A$

Illustration 14: In Fig, 26.45 what is the voltage needed at the source, to maintain 15 V across the load resistance R_L of $2k\Omega$, assuming that the series resistance R is $200\,\Omega$ and the zener requires a minimum current of 10 mA to work satisfactory? What is the zener rating required? **(JEE MAIN)**

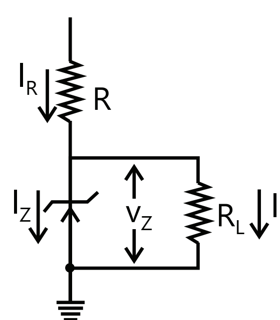


Figure 26.45

Sol: Current through the load resistor of resistance $2k\Omega$ is found using Ohm's law. As 10 mA current is still flowing through the zener when it is connected in reverse biased condition while the rest of current passes through load resistor, the rating of the zener diode should be higher than total current passing through the zener. The rating is given by $I_R + I_L$.

Voltage across R_L, is $V_L = 15$ V

Current through R_L, $I_L = \frac{V_L}{R_L} = \frac{15}{2\times10^{-3}} = 7.5\times10^{-3}$ A $= 7.5$mA

Zener current, $I_z = 10$ mA $\therefore$ Current through R, $I_R = I_z + I_L = 10 + 7.5$mA

Voltage drop across R, $V_R = I_R \times R = 17.5\times10^{-3}\times200 = 3.5$ V

Input voltage required, $V = V_R + V_L = 3.5 + 15 = 18.5$V

Therefore, zener diode should have a current rating of 17.5 mA and a breakdown voltage of 15V.

Illustration 15: 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.

(JEE ADVANCED)

Sol: The transformer is the step down transformer, thus the secondary voltage will be less than primary voltage applied. The DC output voltage obtained at the end of secondary coil is given by $V_{sm} = V_{pm} \times \frac{N_2}{N_1}$ where V_{pm} is peak voltage at the primary coil.

The inverted DC voltage obtained at the secondary coil is given by $V'_{dc} = \frac{V_{sm}}{\pi}$.

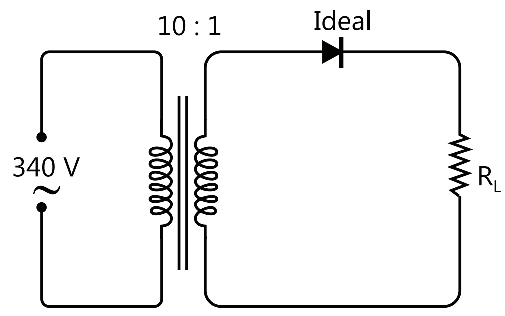


Figure 26.46

Primary to secondary turns is [see fig.26.46]

$\frac{N_1}{N_2} = 10$ R.M.S primary voltage =230 V

$\therefore$ Maximum primary voltage is $V_{pm} = (\sqrt{2}) \times$ r.m.s. primary voltage $= (\sqrt{2}) \times 230 = 325.3V$ Max. Secondary voltage is

$$V_{sm} = V_{pm} \times \frac{N_2}{N_1} = 325.3 \times \frac{1}{10} = 32.53V$$

(i) $V'_{dc} = \frac{V_{sm}}{\pi} = \frac{32.3}{\pi} = 10.36V$

(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

Illustration 16: A crystal diode having internal resistance r_i=20 Ω is used for half-wave rectification. If the applied voltage $V = 50 \sin\omega t$ and load resistance $R_L = 800$ Ω, find:

(i) I_m, I_{dc}, I_{rms} (ii) a.c. power input and d.c power output (iii) d.c output voltage (iv) Efficiency of rectification

(JEE MAIN)

Sol: The equation of AC voltage is given by $V = V_o \sin\omega t$ where V_o is the maximum voltage from the source. As a load resistance is applied in series to the supply, the current through the resistance and output voltage are obtained by Ohm's law. The power is given by $P = I_{eff}^2 \times R_{eff}$ where R_{eff} is the effective resistance in the circuit. The efficiency of the diode is given by $\varepsilon = \frac{P_{DC}}{P_{AC}}$ where P_{DC} and P_{AC} are DC and AC powers dissipated in the circuit.

Given that, V=50 sin ωt ; r_i = 20 Ω, and R_L = 800

$\therefore$ Maximum voltage, V_m=50V

(i) $I_m = \frac{V_m}{r_f + R_L} = \frac{50}{20 + 800} = 0.061A = 61$ mA; $I_{dc} = I_m / \pi = 61 / \pi = 19.4$ mA

$I_{rms} = I_m / 2 = 61 / 2 = 30.5$ mA

(ii) a.c. power input= $(I_{rms})^2 \times (r_f + R_L) = \left(\frac{30.5}{1000}\right)^2 \times (20 + 800) = 0.763$ watt

(iii) d.c power output $= I_{dc}^2 \times R_L = \left(\frac{19.4}{1000}\right)^2 \times 800 = 0.301$ watt

(iv) d.c. output voltage = $I_{dc} R_L = 19.4mA \times 800\Omega = 15.52$ volts

(v) Efficiency of rectification= $= \frac{0.301}{0.763} \times 100 = 39.5\%$

Illustration 17: A full wave rectifier uses two diodes, the internal resistance of each diode may be assumed constant at 20 Ω. The transformer r.m.s secondary voltage from center tap to each end of secondary is 50 V and load resistance is 980 Ω find.

(i) The mean load current (ii) The r.m.s. value of load current **(JEE MAIN)**

Sol: Mean load current is given by $I_{dc} = \frac{2I_m}{\pi}$ where I_m is the maximum current through load and rms current is given by $I_{rms} = \frac{I_m}{\sqrt{2}} = 0.707 I_m$.

$r_f = 20\Omega$, $R_L = 980\Omega$; Max. a.c. voltage, $V_m = 50 \times \sqrt{2} = 70.7V$

Max. Load current, $I_m = \frac{V_m}{r_f + R_L} = \frac{70.7V}{(20+980)\Omega} = 70.7mA$

(i) Mean load current, $I_{dc} = \frac{2I_m}{\pi} = \frac{2 \times 70.7}{\pi} = 45mA$

(ii) R.M.S value of load current is $I_{rms} = \frac{I_m}{\sqrt{2}} = \frac{70.7}{\sqrt{2}} = 50mA$

Illustration 18: In the center- tap circuit shown in Fig. 26.47, the diodes are assumed to be ideal i.e. having zero internal resistance. Find:

(i) d.c output voltage (ii) peak inverse voltage **(JEE ADVANCED)**

Sol: The average dc current in the circuit is $I_{dc} = \frac{2V_m}{\pi R_L}$ where V_m is the maximum voltage across secondary coil of center tap circuit. And thus the DC output voltage is obtained using Ohm's law. The peak inverse voltage is $V_o' = \sqrt{2} V_{rms}$

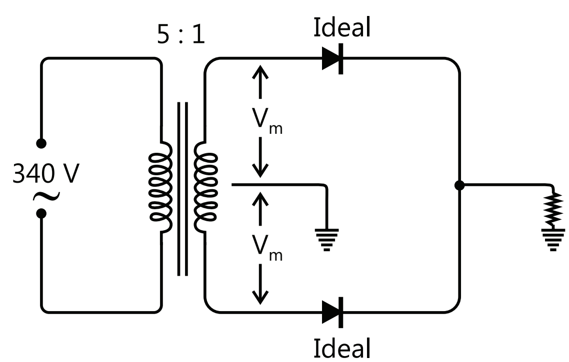


Figure 26.47

Primary to secondary turns, $N_1/N_2 = 5$

R.M.S. primary voltage = 230V

$\therefore$ R.M.S. secondary voltage $= 230 \times (1/5) = 46V$

Maximum voltage across half secondary winding is

$V_m = \frac{V_{rms}}{\sqrt{2}} = 32.5V$

(i) Average current, $I_{dc} = \frac{2V_m}{\pi R_L} = \frac{2 \times 32.5}{\pi \times 100} = 0.207\ A$

$\therefore$ d.c. output voltage, $V_{dc} = I_{dc} \times R_L = 0.207 \times 100 = 20.7V$

(ii) The peak inverse voltage (P_{IV}) is equal to the maximum secondary voltage i.e. P_{IV}=65V

Illustration 19: In a zener regulated power supply, a zener diode with V_z= 6V is used for regulation. The load current is to be 4.0 mA and unregulated input is 10.0 V what should be the value of series resistor R_s?

(JEE ADVANCED)

Sol: The current through the series resistance is $I = I_z + I_L$ where I_z is the zener current and I_L is the current through load resistance. The value of resistance is $R_S = \frac{V_{eff}}{I}$ where V_{eff} is the effective resistance across the resistance.

Here, I_L=4.0 mA; V_z=6V; E_{in} =10.0 V; R_s=?

For good regulation, the value of R_s should be such that current through the zener diode is much larger than the load current. Choosing zener current I_z to be 5 times the load current,

$I_z = 5I_L = 5 \times 4.0 = 20\text{mA}$; $\therefore$ Current through $R_{s,}$ is $I = I_z + I_L = 20 + 4 = 24$ mA

Voltage across $R_s = E_{in} - V_z = 10.0 - 6.0 = 4$; $\therefore R_s = \dfrac{\text{Voltage across } R_s}{I} = \dfrac{4V}{24mA} = 167\Omega$

Illustration 20: Figure 26.48 .shows the forward characteristic of a junction diode. Determine the d.c. and a.c. resistance of the diode when it operates at 0.3 V. **(JEE MAIN)**

Sol: The DC resistance of the diode is obtained by the slope of the curve at 0.3 V. The ac resistance is obtained by taking the ratio of the voltage difference at two point with their corresponding current.

Referring to Fig. 26.48 when V_F =0.3 V; I_F=4.3 mA=4.3×10^{-3} A.

$\therefore$ D.C resistance $r_{dc} = \dfrac{V_F}{I_F} = \dfrac{0.3VA}{4.3 \times 10^{-3}} = 69.77\Omega$

Referring to Fig. 26.48, $\Delta V_F = V_C - V_A$ and $\Delta I_F = I_C - I_A$

$\therefore$ A.C. resistance, $r_{ac} = \dfrac{\Delta V_F}{\Delta I_F} = \dfrac{V_C - V_A}{I_C - I_A} = \dfrac{0.35 - 0.25}{(6-3) \times 10^{-3}} = 33.33\Omega$

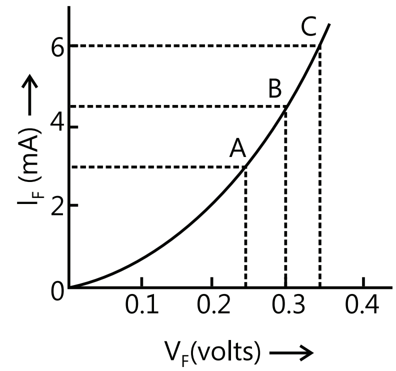


Figure 26.48

12. JUNCTION TRANSISTOR

A transistor is an electronic device formed by p and n-type of semiconductor which is used in placed of a triode valve. It was discovered in 1948 by American scientists Bardeen, Shockley and Brattain. Transistors are of two type: p-n-p transistor and n-p-n transistor.

12.1 P-N-P Transistor

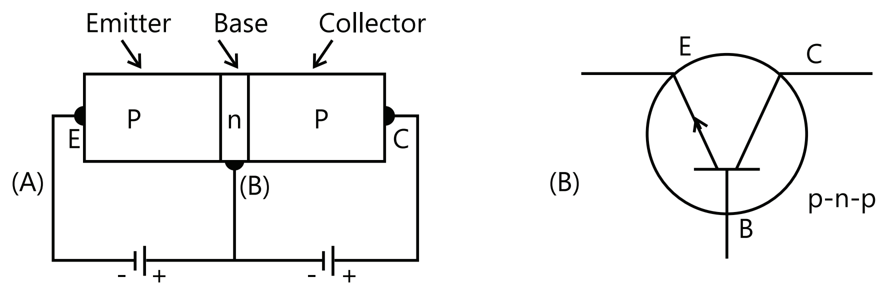


Figure 26.49

It consists of a very thin layer of n-type semiconductor sandwiched between two small p-type semi conductor (See Fig. 26.49). The central slice is called the 'base' while the left and right crystals are called the 'emitter' and the 'collector' respectively. The emitter is given a positive potential while the collector is given a negative potential with respect to the base. Thus, the emitter-base (p-n) junction on the left is under forward-bias (high resistance). While the base-collector (n-p) junction on the right is reverse-bias. The symbol for this transistor is shown in Fig. 26.49 (B) in which the direction of the arrow indicates the direction of current (direction of flow of holes).

Working: A 'common-base' circuit of p-n-p transistor is shown in Fig. 26.50. The emitter-base (p-n) junction on the left is given a small forward bias (fraction of a volt) while the base-collector (n-p) junction is given a large reverse-bias (a few volts).

Holes are the charge-carriers with in the p-n-p transistor, while electrons are the charge-carriers in the external circuit.

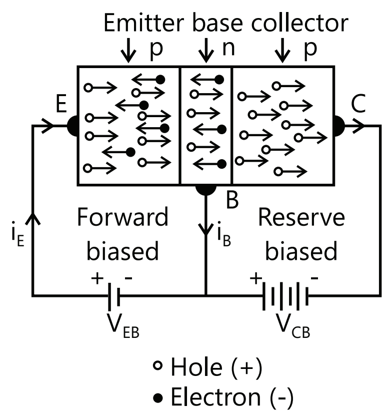


Figure 26.50

The small current which leaves the base terminal B is called the 'i_B' 'base-current' the larger current which leaves the collector terminal is called the 'collector-current' i_C. Both these currents combine to enter the emitter terminal E and constitute the emitter-current i_E. Clearly, $i_E=i_B+i_C$

The base being very thin, the number of hole-electron combinations in it is very small, and almost all the holes entering the base from the emitter reach the collector. Hence the collector-current i_C is only very slightly less than the emitter current i_E.

12.2 N-P-N Transistor

It consists of a very thin slice of p-type semiconductor (Fig. 26.51). In this transistor the emitter is given a negative potential while the collector is given a positive potential with respective to the base. Again, the emitter-base (n-p) junction on the left is under forward-bias, while the base collector (p-n) junction on the right is denser reverse-bias. *The symbol for the p-n-p transistor is shown in Fig. 26.51. (B) in which the direction of the arrow indicates the direction of current (opposite of the direction of flow of electrons).

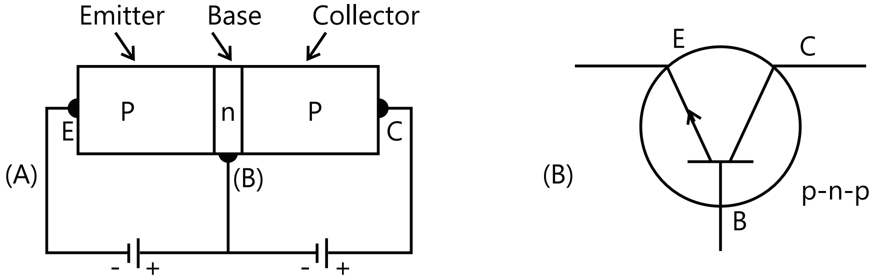


Figure 26.51

A transistor can be connector in a circuit in three different ways. They are:

(i) Common- base configuration, (ii) Common-emitter configuration and (iii) Common-collector configuration. The word 'common' is related with that electrode which is common in input and output circuit. This common electrode is generally grounded. Hence the above three configuration of connection are also called respectively as grounded- base configuration, grounded-emitter configuration, and groundcollector of connection are also called configuration. Each configuration has its own characteristics.

Working: A circuit known as' common-base' circuit of n-p-n transistor is shown in Fig. 26.52. The two n-regions contain the mobile electrons while the central thin p-region contains the mobile (positive) holes. The emitter-base by means of a battery V_{EB}, while the base-collector (p-n) junction on the right has been given a large reverse-bias by means of battery V_{CB}.

The electrons are charge-carriers with in the n-p-n transistor as well as in the external circuit (whereas holes are the charge- carriers with in p-n-p transistor).

The small current entering the base terminal B is the base current i_B, while the larger current entering that collector terminal C is the collector-current i_C. Both currents combine to leave the emitter terminal E and constitute the emitter current i_E. Thus, $i_E=i_B+i_C$

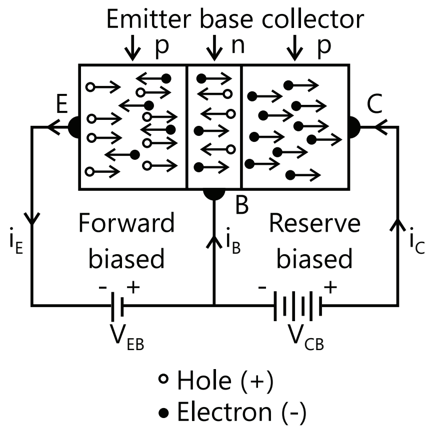


Figure 26.52

13. TRANSISTOR ACTION

There are four possible ways of biasing the two PN junctions (emitter junction and collector junction) of a transistor. There are tabulated below.

Emitter junction bias	Collector junction bias	Transistor operation
Forward	Reverse	Active

Emitter junction bias	Collector junction bias	Transistor operation
Forward	Forward	Saturation
Reverse	Reverse	Cut off
Reverse	Forward	Inverted

$i_E = i_B + i_C$; $i_{B<<}$ $i_{C,}$ and $i_{B<<}$ i_E

Thus i_C is always less than $i_{E,}$ but the difference is small.

Since the emitter junction is forward biased its resistance is small, while the collector junction is reverse biased, therefore its resistance is large. Thus, a transistor is a device which transfers i_E current from low resistance circuit to a high resistance circuit $\left(I_C < I_E\right)$. Thus it is,

Transfer+ resistor =transistor

(The name transistor originated from the action of the transistor).

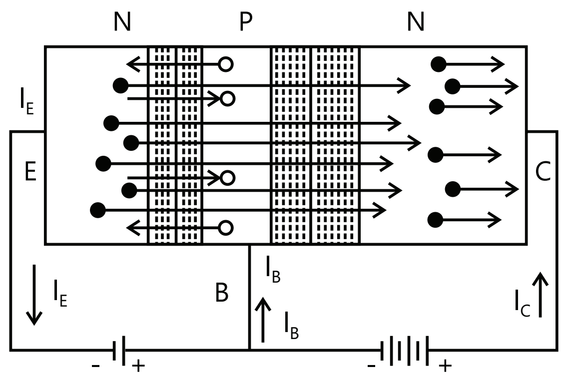


Figure 26.53

14. TRANSISTOR CHARACTERISTICS AND USE AS AN AMPLIFIER

A junction diode cannot amplify a signal. A transistor consisting of two p-n junctions, one is forward-biased and the other is reverse-biased can however, be used for amplifying a weak signal the forward-biased junction has a low-resistance path whereas the reverse-biased junction has a high-resistances path. The weak input signal is applied across the forward-biased (low resistance) junction and the output currents are signal is taken appears with a much higher voltage. The transistor thus acts as an amplifier.

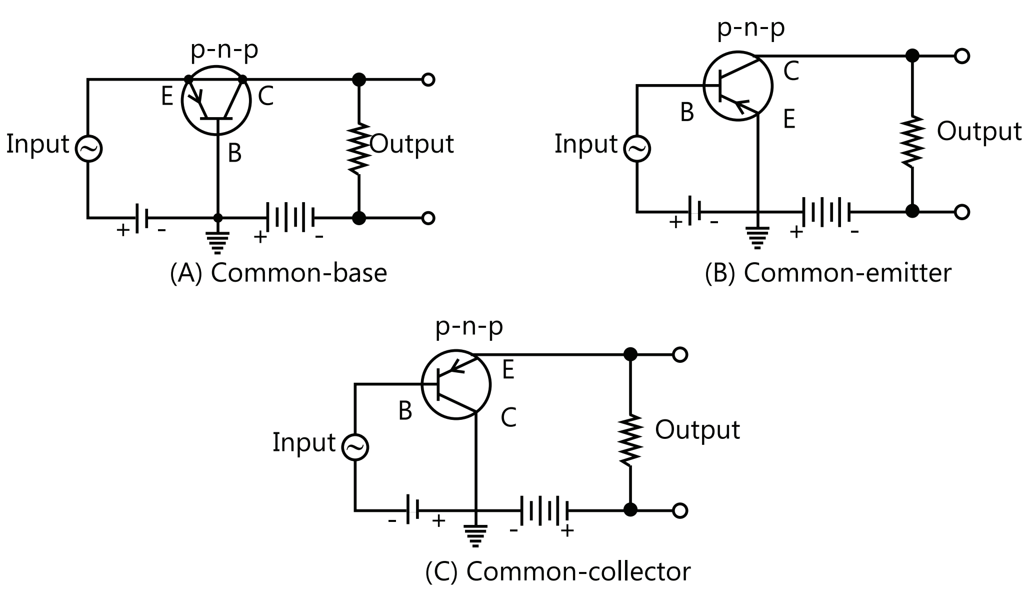


Figure 26.54

When a transistor is to be operated as amplifier, three different basic circuit connections are possible, as illustrated in Fig. 26.54. These are

(a) Common-base

(b) Common emitter and

(c) Common-collector circuits.

14.1 Transistor as Common-Base Amplifier (n-p-n)

Figure 26.55 shows the common-base amplifier circuit using an n-p-n transistor. The base is made common to the input and output circuits.

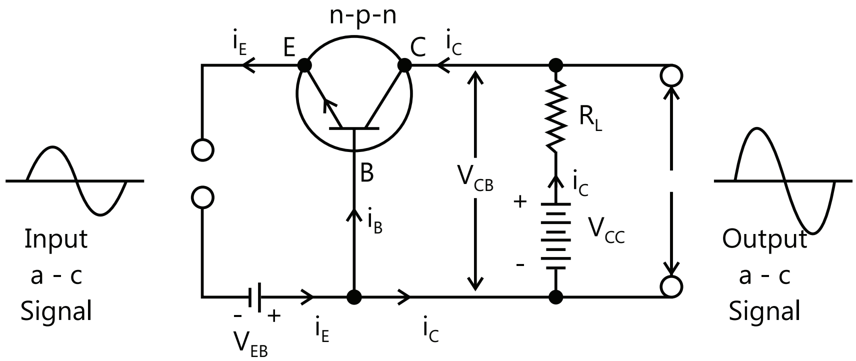


Figure 26.55

The emitter-base circuit is forward-biased by a low-voltage battery V_{EB} so that the resistance of the input circuit is small. The collector-base output circuit is reverse-biased by means of a high-voltage battery V_{CC} so the resistance at the output circuit is quite large. R_L is a load resistor connected in the collector-base circuit. The weak input ac voltage signal is applied across the emitter-base circuit and the amplified output signal is obtained across the collector-base circuit.

Let i_E, i_B and i_c be the emitter-current, base-current and collector-current irrespective when no ac voltage signal is applied to the input circuit. (The arrows represent the direction of hole current that is conventional current which is opposite to the direction of electron current). By Kirchhoff's first law, we have $i_E = i_B + i_C$ ... (i)

Due to the collector-current i_c, the voltage drop across R_L is $i_C R_L$. Therefore, the collector-to-base voltage (potential difference between collector and base) V_{CB} would be given by

$V_{CB} = V_{CC} - i_c R_L$ .... (ii)

When the input ac voltage signal is applied across the emitter-base circuit, it changes the emitter-base voltage and hence the emitter-current i_E which, in turn, changes the collector current i_C consequently the collector to base voltage V_{CB} varies in accordance with equation (ii). This variation in V_{CB}, when the input signal is applied, appears as an amplified output.

Phase relationship between input and output voltage signal in CB circuit:

The output voltage signal is in with the input voltage signal in the common-base amplifier.

14.1.1 Gains in Common-Base Amplifier

The various gains in a common-base amplifier are as follow:

(a) **ac current gain:** It is defined as the ratio of the change in the collector- current to the change in the emitter-current at a constant collector-to-voltage, and is denoted by α. Thus

$$\alpha_{(ac)} = \left(\frac{\Delta i_C}{\Delta i_E}\right)_{V_{CB}}$$

The value of α is slightly less than 1 (actually there is a little current loss).

(b) **ac Voltage Gain:** It is defined as the ratio of the changes in the output voltage to the change in the input voltage, and is denoted by A_V.

Suppose, on applying ac input voltage signal, the emitter current changes by Δi_C and correspondingly the collector-current changes by Δi_C. If R_{in} and R_{out} the resistances of the input and the output circuits respectively, then

$$A_v = \frac{\Delta i_C \times R_{out}}{\Delta i_E \times R_{in}} = \frac{\Delta i_C}{\Delta i_E} \times \frac{R_{out}}{R_{in}}$$

Now, $\Delta i_C / \Delta i_E$ is the ac current-gain and R_{out} / R_{in} is called the 'resistance gain'

$\because A_v = \alpha \times$ Resistance gain

Since the resistance gain is quite high, A_V is also quite high although α is slightly less than 1.

(i) ac Power Gain: It is defined as the ratio of the change in the output power to the change in the input power.

Since power=current$\times$ voltage, we have

ac power gain =ac current gain$\times$ac voltage-gain= α^2 $\times$resistance gain

14.2 Common Base Amplifier Using a p-n-p Transistor

Common-Emitter using an n-p-n transistor: Figure 26.56 show the common-emitter amplifier circuit using an n-p-n transistor. The emitter is made common to the input and output circuits.

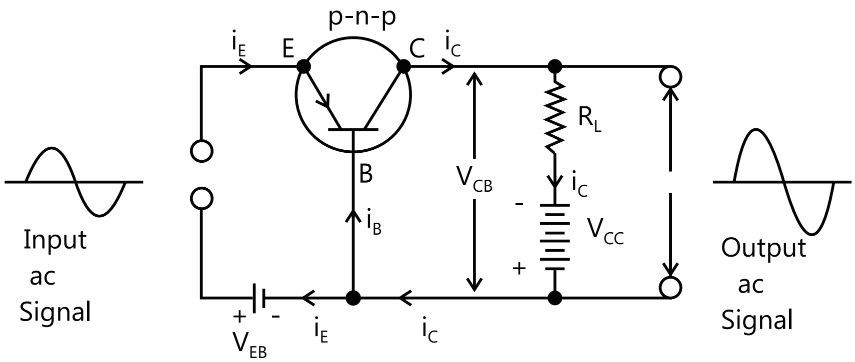


Figure 26.56

The input (base-emitter) circuit is forward-biased by a low-voltage battery V_{BE} so that the resistance of the input circuit is small. The output (collector-emitter) circuit is reverse-biased by means of a high voltage battery V_{CC} so that the resistance at the output circuit is high. R_L is a load resistance connected in the collected-emitter output circuit. The weak input ac signal is applied across the base-emitter circuit and the amplified output signal is across the collector-emitter circuit.

Let i_E, i_B and i_c be the emitter-current, base-current and collector-current irrespective when no ac voltage signal is applied to the input circuit. (The arrows represent the direction of hole current, that is conventional current which is opposite to the direction of electron current). By Kirchhoff's first law, we have $i_E=i_B+i_C$ ... (i)

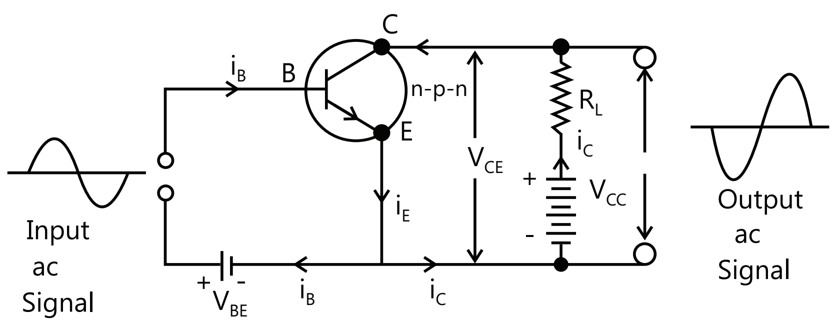


Figure 26.57

Due to the collector-current i_c, (which is only slightly smaller that i_E), the voltage drop across R_L is i_C R_L. Therefore, the collector-to-base voltage (potential difference between collector and base) V_{CB} would be given by $V_{CE} = V_{CC} - i_c$ R .... (ii)

When the input ac voltage signal is applied across the emitter-base circuit, it changes the emitter-base voltage and hence the emitter-current i_E which, in turn, changes the collector current i_C consequently the collector to base voltage V_{CE} varies in accordance with equation (ii). This variation in V_{CE}, when the input signal is applied, appears as an amplified output.

Phase Relationship between input and output voltage signals: In a common-emitter amplifier the input voltage signal and the output voltage signal obtained across the collector and the emitter are out of phase with each other.

The output voltage signal is 180° out of phase with the input voltage signal in the common-emitter amplifier.

14.3 Common-Emitter Amplifier Using a p-n-p Transistor

14.3.1 Gains in Common-Emitter Amplifier

The various gains in a common-emitter amplifier are as follows:

(a) dc current Gains: It is defined as the ratio of the collector current to the current, and is denoted by β (dc).

Thus $\beta(dc) = \frac{i_C}{i_B}$

In a typical transistor, a small base-current ($\approx 10\mu A$) produces large collector-current ($\approx 500\mu A$).

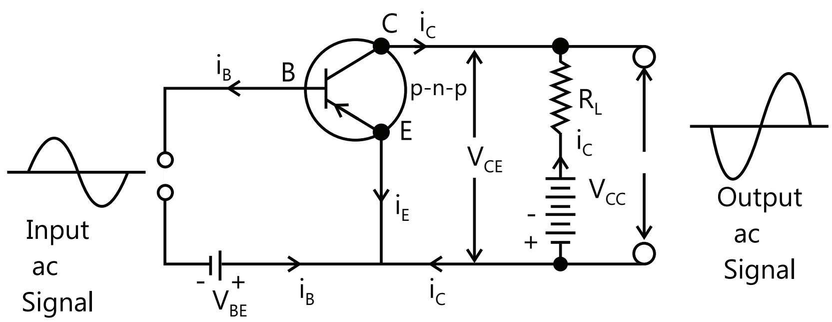


Figure 26.58

Thus $\beta(dc) = \frac{500}{10} = 50$

(b) ac current Gain: It is defined as the ratio of the collector-current to the change in the base-current at a constant collector to emitter voltage, and is denoted by $\beta(ac)$.

Thus $\beta(ac) = \left(\frac{\Delta i_C}{\Delta i_B}\right)_{V_{CE}}$

(c) Voltage gain: Suppose, on applying ac input voltage signal, the base-current changes by Δi_B and correspondingly the output collector-current change by Δi_c. If R_{in} and R_{out} be the resistance of the input and the output circuits respectively,

then $A_v = \frac{\Delta i_C \times R_{out}}{\Delta i_E \times R_{in}} = \frac{\Delta i_C}{\Delta i_E} \times \frac{R_{out}}{R_{in}}$ ... (i)

Now, $\Delta i_C / \Delta i_E$ is the ac current-gain (ac) and R_{out} / R_{in} is the 'resistance gain'

$\therefore Av = \beta(ac) \times \text{resistance gain}$ ... (ii)

Since $\beta(ac) >> \alpha$ (ac), the ac voltage gain in common-emitter amplifier is larger compared to the common-base amplifier., although the resistance gain is smaller.

From equation (i) and (ii), it follows that $A_v = g_m \times R_{out}$

(d) ac power gain: It is defined as the ratio of the change in the output power to the change in the input power.

Since power=current×voltage, we have

ac power gain=ac current gain × ac voltage gain= $\beta(ac) \times A_v$

$= \beta(ac) \times \{\beta(ac) \times \text{resistance gain}\} = \beta^2(ac) \times \text{resistance gain}$

Since $\beta(ac) >> \alpha(ac)$, the ac power gain in common-emitter amplifier is extremely large compared to that in common-base amplifier.

(e) Trans conductance: (g_m) is defined as the ratio of small change in the output current (i.e., collector current) to the corresponding small change in the voltage (V_B) at constant output voltage (V_c).

i.e $g_m = \left.\frac{\Delta I_0}{\Delta V_1}\right|_{V_0=\text{constant}} = \left.\frac{\Delta I_C}{\Delta V_B}\right|_{V_0=\text{constant}}$ or $g_m = \frac{\Delta I_C}{\Delta V_B} = \frac{\Delta I_C}{\Delta V_B} \times \frac{\Delta I_B}{\Delta V_B}$ (Multiplying and dividing by ΔV_B)

$= \left(\frac{\Delta I_C}{\Delta I_B}\right) \times \left(\frac{\Delta I_B}{\Delta V_B}\right)$ But $\frac{\Delta I_c}{\Delta V_B} = \beta_{a.c.}$ and $\frac{\Delta V_B}{\Delta I_B}$ = input resistance (R_i) $\therefore \quad g_m = \beta_{a.c.} / R_i$

Since voltage gain, $A_V = \beta_{a.c.} \times \frac{R_0}{R_i}$ $\therefore$ Using eqn. (i), we have $A_V = g_m \times R_0$ For $R_0 = R_L$ $\therefore A_V = g_m R_L$

14.3.2 Relation Between α and β

CB current gain (α): CB current gain (α) is the ratio of output current to the input current in common base configuration of a transistor. $\alpha_{dc} = \frac{I_C}{I_E}$; $\alpha_{ac} = \frac{\Delta Ic}{\Delta I_E}$

CE current gain (β): CE current gain (β) is the ratio of the output current to the input current in emitter configuration of the transistor. $\beta_{dc} = \frac{I_C}{I_B}$; $\beta_{ac} = \frac{\Delta Ic}{\Delta I_B}$

The CB current gain α and CE current gain β are related by the following relations.

$\frac{1}{\alpha} = 1 + \frac{1}{\beta}$; $\alpha = \frac{\beta}{\beta+1}$; $\beta = \frac{\alpha}{1-\alpha}$

The above relations are applicable for both dc and ac current gains.

MASTERJEE CONCEPTS

The value or α is always less than 1.α~0.9 to 0.99 or more. The value of β is always much greater than 1. β ~ 95 to 99 or so.

Chinmay S Purandare (JEE 2012, AIR 698)

Illustration 21: The dc current gain of a resistance in CB configuration is 0.99. Find its dc current gain in CE configuration. **(JEE MAIN)**

Sol: As the gain in common base configuration α_{dc} is given, the gain in common emitter configuration is given by $\beta_{dc} = \frac{\alpha_{dc}}{1-\alpha_{dc}}$.

Given α_{dc} 0.99 therefore, $\beta_{dc} = \frac{\alpha_{dc}}{1-\alpha_{dc}} = \frac{0.99}{1-0.99} = \frac{0.99}{0.1}$ or $\beta_{dc} = 99$

Illustration 22: In certain resistance α_{dc} =0.98 and I_E =1 mA. Find the corresponding values of base current and collector current. **(JEE MAIN)**

Sol: The collector current is found by $\alpha_{dc} = \frac{I_C}{I_E}$. In the circuit the base current is $I_B = I_E - I_C$.

(i) $\alpha_{dc} = \frac{I_C}{I_E}$ or $0.98 = \frac{I_C}{I(mA)}$ Thus I_C =0.98mA

(ii) Using $I_B = I_E - I_C$; we get $I_B = (1-0.98)mA = 0.02mA$

Illustration 23: In a common base connection, current amplified factor is 0.9. If the emitter current is 1mA, determine the value of base current. **(JEE MAIN)**

Sol: The collector current is found by $\alpha_{dc} = \frac{I_C}{I_E}$. In the circuit the base current is $I_B = I_E - I_C$.

Here α = 0.9, I_E = 1 mA. Now $\alpha = \frac{I_C}{I_E}$ Or $I_C = \alpha\, I_E = 0.9 \times 1 = 0.9mA$. Also $I_E = I_B + I_C$

$\therefore$ Base current, $I_B = I_E - I_C = 1 - 0.9 = 0.1$ mA

Illustration 24: For a CE amplifier (see Fig. 26.59), the audio signal voltage across the collector resistance $R_C = 2.0\ k\Omega$ is 2.0 V. Suppose the current amplification factor of the transistor is 100. What should be the value R_B in series with V_{BB} supply of 2.0 V if d.c. base current has to be to 10 time the supply current? Also calculate the d.c. collector current. **(JEE ADVANCED)**

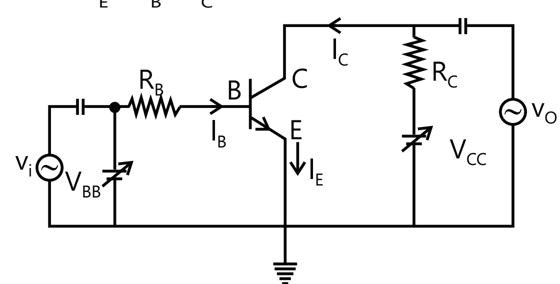


Figure 26.59

Sol: In the CE configuration the base current is obtained as $i_b = \frac{i_c}{\beta}$ where i_c is the AC collector current. The DC base and collector current are obtained as $i_B = 10 \times i_b$ and $I_C = \beta I_B$.

The output a.c. voltage is 2.0V.

$\therefore$ a.c. collector current $i_C = \frac{2.0V}{R_C} = \frac{2.0V}{2000\Omega} = 1$ mA

$\therefore$ a.c. base current $i_b = i_C / \beta = 1/100 = 0.01$ mA

$\therefore$ d.c. base current, $i_B = 10 \times i_b = 10 \times 0.01 = 0.1$ mA

Applying Kirchhoff's voltage law to the base circuit,

$V_{BB} = I_B R_B + V_{BE}$ or $R_B = \frac{V_{BB} - V_{BE}}{I_B} = \frac{(2.0-0.6)}{0.1\ mA} = 14k\Omega$ (Assume V_{BE} =0.6V)

d.c collector current $I_C = \beta I_B = 100 \times 0.1 = 10mA$

Illustration 25: In an NPN transistor, 10^{10} electrons enter the emitter in 10^{-6}s. If 2% electrons are lost in the base, calculate the current transfer ratio and current amplification factor. **(JEE MAIN)**

Sol: The amplification factor is given by $\beta = \frac{I_C}{I_B}$ and current transfer ratio is given by $\frac{I_C}{I_E}$.

Current = charge / time

$\therefore$ Emitter current, $I_E = \frac{N_e}{t} = \frac{10^{10} \times 1.6 \times 10^{-19}}{10^{-6}} = 1.6 \times 10^{-3} A = 1.6mA$

Base current, $I_B = 2\%$ of $I_E = \frac{2}{100} \times 1.6 = 0.032mA$

In a transistor, the currents relation is $I_E = I_B + I_C$ or $I_C = I_E - I_B = 1.6 - 0.032 = 1.568$ mA

$\therefore$ Current transfer ratio $= \frac{I_C}{I_E} = \frac{1.568}{1.6} = 0.98$

Current amplification factor, $\beta = \frac{I_C}{I_B} = \frac{1.568}{0.032} = 49$

15. TRANSISTOR OSCILLATOR

Oscillator is a device which delivers a.c. output waveform of desired frequency from d.c. power even without input signal excitation.

The electric oscillations are produced by L-C circuit (i.e. tank circuit containing inductor and capacitor in parallel). L-C circuit producing L-C oscillations consists of an inductor inductance L and a capacitor of variable capacitance C. Using positive feedback arrangement inductors L and L' are inductively coupled as both the coils are around same core acts as a positive feedback arrangement.

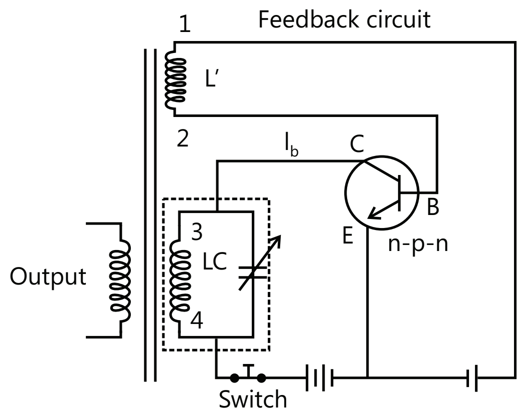


Figure 26.60

Working: When switch is closed, emitter-base junction is forward biased and the collector-emitter base junction reverse biased. The emitter current and hence collector current begins to flow. The inductor L' opposes the growth of collector current to its maximum value. Therefore, the current I grows slowly. As a result of this, the magnetic flux linked with L' changes. Since coil L' is inductively coupled with coil L, so magnetic flux linked with coil L also changes. Due to the change in magnetic flux linked with coil L, an induced e.m.f is set up across the coil L'. As a result of this positive feedback, collector current I_C is further increased. The process of increasing the collector current continued till the magnetic flux linked with the coil L' becomes maximum (i.e., constant). At this stage, collector current I_C becomes maximum (transistor becomes saturated) and the change in magnetic flux linked with the coil L' ceases. As a result of this, there is also no change in the magnetic flux linked with the coil L and hence the induced e.m.f. across the coil L becomes zero (transistor is in the cut off region). The capacitor C starts discharging through the coil L.

Now, the forward bias of emitter-base junction decreases. Hence the emitter current I_E and consequently collector current I_C begin to decrease. As the collector current I_C decreases, again magnetic flux linked with the coil L' decreases. Consequently, the magnetic flux linked with coil L' also decrease. Hence, induced e.m.f. is set up across the coil L' but now in opposite direction. The forward bias across emitter-base junction is further decreased and hence the emitter current I_E and collector current I_C is further decreased. The process continues till the collector becomes zero. At this stage, capacitor gets discharged through coil L but now in the opposite direction.

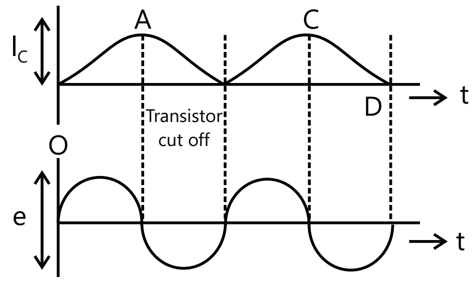


Figure 26.61

Now the emitter current and hence collector current increases in the opposite direction. This process repeats and the collector current oscillates between maximum and minimum values.

The frequency of oscillation is given by $v = \frac{1}{2\pi\sqrt{LC}}$

16. TRANSISTOR AS A SWITCH

We have already discussed that a crystal diode behaves like a switch. When the diode is forward biased, it conducts current easily and behaves like a closed switch. However, when diode is reverse biased, it practically conducts no current and behaves like an open switch. A transistor can also be used as a switch by making emitter-base junction either reverse biased or sufficiently forward biased. Figs. 26.62 (a) and (b) illustrate the operation of a transistor as a switch.

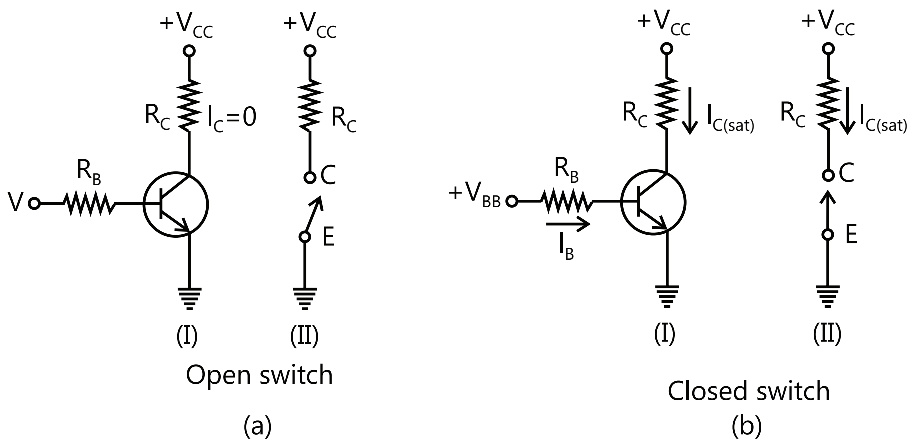


Figure 26.62

(a) Transistor as open switch. In Fig. 26.62 (a), the base-emitter junction is not forward biased. Therefore, base current I_B=0. As a result, collector current $I_C\left(=\beta I_B\right)$ is also zero. Under this condition, the transistor behaves as an open switch. In other words, there is an open between the collector and emitter as indicated in Fig. 26.62(a).

(b) Transistor as closed switch. In Fig. 26.62.(b), the base-emitter junction is sufficiently forward biased so that base current I_B is made large enough to cause maximum collector current to flow. This maximum value of collector current is called saturation current $I_{C(sad)}$. Under this collector and emitter as shown in Fig. 26.62.

Illustration 26: An LC oscillation has a tank circuit with L_1= 58.6 μH and C_1=300 PF. Calculate the frequency of oscillations. **(JEE MAIN)**

Sol: The frequency of oscillation is given by $f=\dfrac{1}{2\pi\sqrt{L_1C_1}}$.

$L_1 = 58.6H = 58.6\times10^{-6}\,H; C_1 = 300\times10^{-12}F$

The frequency of oscillations f is given by; $f=\dfrac{1}{2\pi\sqrt{L_1C_1}}=\dfrac{1}{2\pi\sqrt{58.6\times10^{-6}\times300\times10^{-12}}}$ Hz

$=1199\times10^3\,Hz = 1199$ kHz

Illustration 27: In Fig. 26.63, the V_{BB} supply can be varied from 0V to 5.0 V. The Si transistor has $\beta_{dc}=250$ and $R_B=100k\Omega$, $R_C=1k\Omega$, $V_{CC}=5.0V$. Assume that when the transistor is saturated, V_{CE}=0V and V_{BE}=0.8V. Calculate (a) the minimum base current for which the transistor will reach saturation. Hence (b) determine V_i when the transistor is switched on. (c) Find the range of V_i for which the transistor is switched off and switched on. **(JEE ADVANCED)**

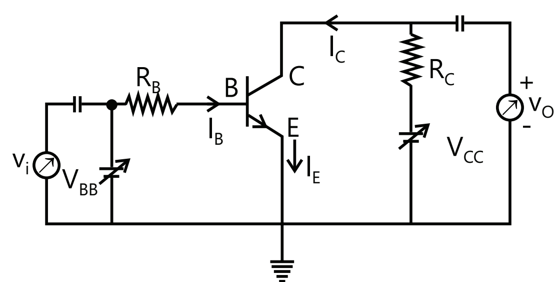


Figure 26.63

Sol: When the transistor reach saturation the collector current is maximum and given by $I_C = \frac{V_{CC} - V_{CE}}{R_C}$. The base current is $I_B = \frac{I_C}{\beta_{dc}}$ where β_{dc} is the DC gain of the circuit. The value of V_i is obtained by applying Kirchhoff's law to the circuit.

(a) When the transistor is saturated, the collector current is maximum at saturation, it is given that V_{CE}=0V; V_{BE}=0.8 V.

Now, $V_{CC} = V_{CE} + I_C R_C$ $\quad \therefore\ I_C = \frac{V_{CC} - V_{CE}}{R_C} = \frac{(5.0-0)V}{1k\Omega} = 5.0\ mA$

Corresponding to saturation current I_C=5.0 mA, the base current I_B is $I_B = \frac{I_C}{\beta_{dc}} = \frac{5.0 mA}{250} = 20\mu A$

(b) The input voltage at which the transistor goes into saturation is

$V_i = I_B R_B + V_{BE} = 20\mu A \times 100\ k\Omega + 0.8 = 2.8V$

(c) We know that for S_i transistor; the transistor will remain off if V_i is less than 0.6 V. Therefore, between 0Vand 0.6 V, the transistor will be in the switched off state. However, between 2.8V and 5.0 V, the transistor will be in switched on state.

17. LOGIC GATES

A logic gate is a digital circuit which is based on certain logical relationship between the input and the output voltage of the circuit.

The logic gates are built using the semiconductor diodes and transistors. Each logic gate is represented by its characteristic symbol. The operation of a logic gate is indicated in a table, known as truth table. This table contains all possible combination of inputs and the corresponding outputs. A logic gate is also represented by a Boolean algebraic expression. Boolean algebra is a method of writing equation showing how an output depends upon the combination of inputs. Boolean algebra was invented by George Boole.

Basic logic gates: (1) OR gate, (2) AND gate, and (3) NOT gate

The OR gate: The output and an OR gate attains the state 1 if one or more inputs attain the state

(a) Logic symbol of OR gate

The Boolean expression of OR gate is Y=A+B, read as Y equals A or B.

Truth table of a two input OR gate

Input	Output
A B	Y
0 0	0
0 1	1
1 0	1
1 1	1

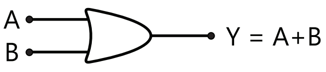


Figure 26.64

(b) The AND gate: The output of an AND gate attains the state 1 if and only if all the inputs are in state

Logic symbol of AND gate

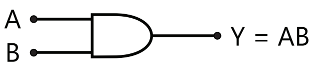


Figure 26.65

The Boolean expression of AND gate is Y=A.B it is read Y equal A and B

Truth table of a two input AND gate

Input	Output
A B	Y
0 0	0
0 1	0
1 0	0
1 1	1

(c) The NOT gate: The output of a NOT gate attains the state 1 if and only input does not attains the state 1.

Logic Symbol of NOT gate:

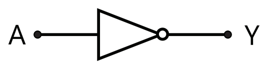


Figure 26.66

The Boolean expression is $Y=\bar{A}$, read as Y equal Not A.

Truth table of NOT gate

Input	Output
A	B
0	1
1	0

Combination of Gates: The three basic gates (OR, AND and NOT) when connected in various combinations give us logic gates such as NAND gates, which are the universal building blocks of digital circuits.

(a) The NAND gate: Logic symbol of NAND gate

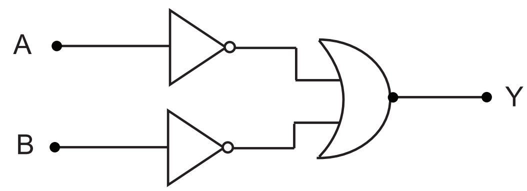


Figure 26.67

The Boolean expression on NAND gate is $Y=\overline{AB}$

Truth table of a NAND gate

Input	Output
A B	Y
0 0	1
0 1	1
1 0	1
1 1	0

(b) The NOR gate: Logic symbol of NOR gate

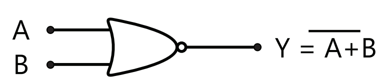


Figure 26.68

The Boolean expression of NOR gate is $Y=\overline{A+B}$

Truth table of a NOR gate

Input	Output
A B	Y

Input	Output
0 0	1
0 1	0
1 0	0
1 1	0

Universal gates: The NAND or NOR gate is the universal building block of all digital circuits. Repeated use of NAND gates (or NOR gates) gives other gates. Therefore, any digital system can be achieved entirely from NAND or NOR gates. We shall show how the repeated use of NAND (and NOR) gates will gives use different gates.

(a) The NOT gate from a NAND gates: When all the input of a NAND gate are connected together, as shown in the figure, we obtain a NOT gate

Truth table of signal-input gate

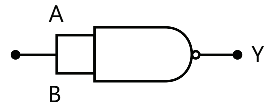


Figure 26.69

Input			Output
A	B=(A)		Y
0	0		1
1	1		0

(b) The AND gate from a NAND gates: If a NAND gate is followed by a NOT gate (i.e., a single input NAND gate), the resulting circuits is an AND gates as shown in figure and truth the table given show how an AND gates has been obtained from NAND gates.

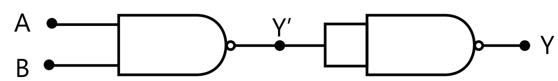


Figure 26.70

Truth table

A	B	Y′	Y
0	0	1	0
0	1	1	0
1	0	1	0
1	1	0	1

(c) The OR gate from NAND gates: If we invert the A and B and then apply them to the NAND gate, the resulting circuit is an OR gate.

Truth table

A	B	A	B	Y
0	0	1	1	0
0 1 1	1	1	0	1
	0	0	1	1
	1	0	0	1

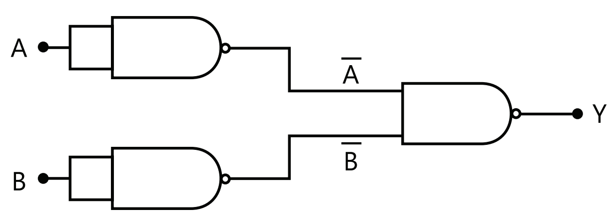


Figure 26.71

(d) The NOT gate from NOR gate: When all the inputs of a NOR gate are connected together as shown in the figure, we obtain a NOR gate.

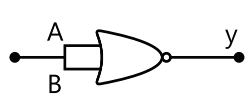


Figure 26.72

(e) The AND gate from NOR gates: If we invert A and B and then apply them to the NOR gate, the resulting circuit is an AND gate.

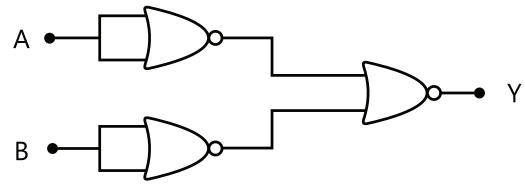


Figure 26.73

(f) The OR gate from NOR gate: If a NOR gate is followed by a single input NOR gate (NOT gate), the resulting circuit is an OR gate.

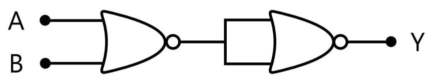


Figure 26.74

XOR AND XNOR gate: (i) The exclusive-OR gate (XOR gate)

The output of a two-input XOR gate attains the state 1 if one and input attains the state 1.

Logic symbol of XOR gate:

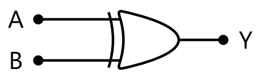


Figure 26.75

The Boolean expression of XOR gate is $Y = \overline{A+B} + \overline{AB}$ or $Y = A \oplus B$

Truth table of a XOR gate

Input	**Output**
A B	**Y**
0 0	0
0 1	1
1 0	1
1 1	0

Exclusive: NOR gate (XNOR gate). The output is in state 1 when it's both input are the same that is, both 0 or both 1.

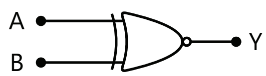


Figure 26.76

(a) Logic symbol of XNOR gate: The Boolean expression of XNOR

gate is Y=A.B+ $Y = \overline{A.B}$ or $Y = \overline{A \oplus B}$

Truth table of a XNOR gate

Input		**Output**
A	B	Y
0 0		1
0 1		0
1 0		0
1 1		1

Law of Boolean algebra: Basic OR, AND, and NOT operations are given below:

OR	AND	NOT
A+0=A	A.0=0	$A+\bar{A}=1$
A+1=1	A.1=A	$A.\ \bar{A}=0$
A+A=A	A. A=A	A. A=A

Boolean algebra obeys commutative, associative and distributive laws as given below:

(i) Commutative laws: AB=BA

(ii) Associative laws: A+ (B+C) = (A+B) + C; A. (B.C)=(A.B).C

(iii) Distributive laws: A (B+C) = AB+AC

Some other useful identities:

(i) A + AB = A;

(ii) A.(A + B) = A.

(iii) $A+\overline{A}B = A+B$

(iv) $A.(\overline{A}+B) = AB$

(v) A + BC = AB + AC

(vi) $(\overline{A} + B)(A + B) = B$

(b) De Morgan's Theorem: First theorem $\overline{A+B} = \overline{A}.\overline{B}$. Second theorem: = $\overline{A.B} = \overline{A}+\overline{B}$

MASTERJEE CONCEPTS

The NAND gate is a universal gate because its repeated use can produce other logic gates.

Ankit Rathore (JEE Advanced 2013, AIR 158)

Illustration 28: The output of an OR gate is connected to both the inputs of a NAND gate. Draw the logic circuit of this combination of gates and write its truth table. **(JEE MAIN)**

Sol: When output of OR gate is connected to the input of NAND gate the circuit behaves as NOR gate.

The logic circuit of the combination of the two gates is shown in Fig.26.77. It is clear that: $Y' = A+B$ and $Y = \overline{A+B}$

This means that NOR gate is formed. The truth table of the given logic circuit is given below:

A	B	Y'=A+B	$Y=\overline{A \bullet B}$
0	0	0	1
1	0	1	0
0	1	1	0
1	1	1	0

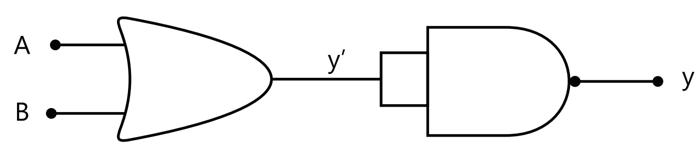


Figure 26.77

Illustration 29: The output of an OR gate is connected to both the inputs of a NOR gate. Draw the logic circuit of this combination of gates and write its truth table. **(JEE MAIN)**

Sol: When output of OR gate is connected to the input of NOR gate the circuit behaves as NAND gate.

The logic circuit of the combination of two gates is shown is Fig. 26.78. It is clear that:

$$Y' = A+B \text{ and } Y = \overline{(A+B)+(A+B)} = \overline{A+B}$$

The truth table of the given logic circuit is shown below:

A	B	Y′	Y
0	0	0	1
1	0	1	0
0	1	1	0
1	1	1	0

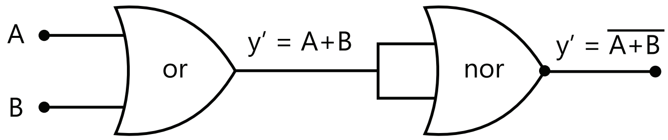


Figure 26.78

Illustration 30: Identify the logic gates marked X, Y in Fig. 26.79. Write down the output at y, when A=1, B=1 and A=0, B=0. **(JEE ADVANCED)**

A	B	Y′	Y
1	1	0	1
0	0	1	0

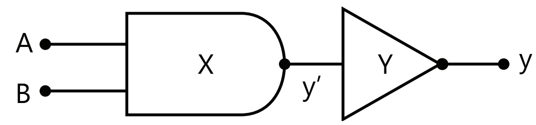


Figure 26.79

Sol: The gate X is AND gate while the gate Y is NOT gate.

The logic gate marked as X is NAND gate while the gate marked as Y is NOT gate. It is clear that:

$Y' = \overline{A.B}$ and $Y = \overline{y'}$

Therefore, the output at y will be as shown in the table.

COMMUNICATION SYSTEMS

1. INTRODUCTION

In the most fundamental sense communication involves the transmission of information from one point to another through a series of processes. In the early 1900s radio communications – the transmission and reception of voice and music through the air was probably the only application of electronics of any significance. The telephones at our hands, the radios and televisions in our living rooms, the computer terminals in our offices and homes are all capable of providing communications from every corner of the earth. In this chapter, we shall focus our attention on the principles of communication.

2. BASIC ELEMENTS OF A COMMUNICATION SYSTEM

Irrespective of the form of communication system being considered, there are three basic elements of every communication system viz, (i) Transmitter (ii) Communication channel (iii) Receiver

Figure shows the block diagram of the basic elements of a communication system.

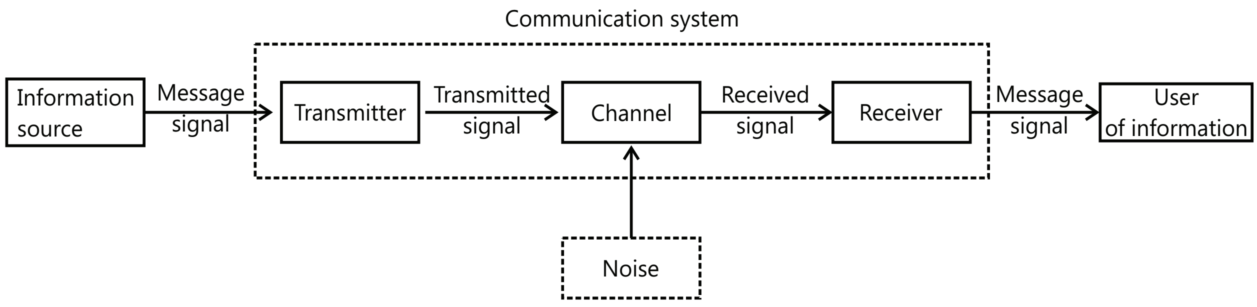


Figure 26.80

Generally, the transmitter is located at one place and the receiver at some other place. The communication channel is links the transmitter and the receiver.

(a) Transmitter. The function of the transmitter is to transform the message or information (e.g. music, speech, picture, written message etc.) into a suitable form and transmit it over the communication channel. Generally, the information is not electrical in nature. The transmitter first converts the message into equivalent electrical variations. It is then called signal. The signal modulates a high frequency wave called carrier wave and the resultant waves are called modulated waves. The actual method of modulation varies from one communication system to another. After modulation the modulated waves are transmitted over the communication channel.

(b) Communication channel. It is the medium by which the modulated waves are transmitted from the transmitter to the receiver. The communication channel or transmission medium may be free space, transmission lines or optical fibers. For example, in case of radio and TV transmission, the communication channel is the free space. However, in case of telegraphy and telephony, communication channel is transmission lines.

In the process of transmission, signals are contaminated by noise signals. This is called channel noise. Noise is unwanted energy usually of random character generated by numerous natural or man-made events e.g. lightning, turning on or off electrical equipment etc.

(c) Receiver. The function of the receiver is to receive the modulated waves transmitted by the transmitter and to do demodulation or detection or decoding. In this process, the original signal is separated from the carrier. This process is called demodulation or detection and is the reverse of the modulation process done in the transmitter. The recovered signal is then utilized as the situation demands.

Basic modes of communication: There are two basic modes of communication viz.

(i) Point-to-Point communication mode

(ii) Broadcast mode

(i) In Point-to-point communication mode the message is transmitted over a link between a single transmitter and single receiver. Conversation between two persons through a telephone is an example of point-to-point communication.

(ii) In broadcast mode (or point-to-many points communication) there is a single transmitter and a large number of receivers. Radio broadcasting and television telecast are the examples of this mode of communication.

3. COMMONLY USED TERMS IN ELECTRONIC COMMUNICATION SYSTEMS

In electronic communication system, the following terms are frequently used:

(a) Transducer. A device that converts one form of energy into another form of energy is called a transducer. For example, a microphone converts sound energy into electrical energy. Therefore, microphone is a transducer. Similarly, a loudspeaker is a transducer because it converts electrical energy into sound energy.

(b) Signal. The information converted into electrical form that is suitable for transmission is called a signal. For example, in a radio station, music, speech etc. are converted into electrical form by a microphone for transmission into space. This electrical form of sound (music, speech etc.) is the signal. The signal can be of two type's viz. (i) analog signal (ii) digital signal

(i) Analog signal. A continuously varying signal (voltage or current) is called an analog signal. For example, an alternating voltage varying sinusoidally is an analog signal (see figure). If such an analog signal is applied to the input of a transistor amplifier, the output voltage will also vary sinusoidally. This is the analog operation i.e., the output voltage can have an infinite number of values. Due to many-valued output, the analog operation is less reliable.

(ii) Digital signal. A signal (voltage or current) that can have two discrete values is called a digital signal. For example, a square wave is a digital signal (see figure). It is because this signal has only two values viz, +5V and 0V and no other value. These values are labeled as High and Low. The high voltage is +5V and the low voltage is 0V. If proper digital signal is applied to the input of a transistor the transistor can be driven between cut off and saturation. In other words, the transistor will have two-state operations i.e. output is either low or high. Since digital operation has only two states (i.e., ON or OFF), it is far more reliable than many-valued analog operation. It is because with two states operation all the signals are easily recognized as either low or high.

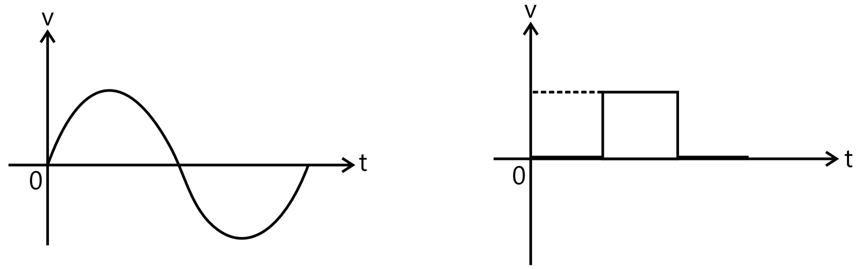


Figure 26.81

(c) Noise. The unwanted signal is called a noise. The noise is undesirable because it disturbs the transmission and processing of signals in a communication system. The source generating the noise may be located inside or outside the system. Efforts should be made to minimize the noise level in a communication system.

(d) Transmitter. An electronic system that broadcasts modulated electromagnetic signals toward one or more distant receivers is called a transmitter. In a transmitter the signal is processed to produce modulated waves. These modulated waves contain the signal and are sent to the receiver through the channel (e.g. space).

(e) Receiver. Any electronic system that enables the desired modulated wave to be separated from all other modulated waves coming into the antenna is called a receiver. In a receiver, the signal is extracted from the modulated wave.

(f) Attenuation. The loss of strength of the signal while propagating through the medium is known as attenuation. It occurs because the medium distorts, reflected and refracts the signal as it passes through it.

(g) Amplification. The process of raising the strength of a signal is called amplification and it is done by an electronic circuit called amplifier. Amplification is necessary to compensate for the attenuation of the signal in a communication system. The energy required for additional signal strength is obtained from a d.c. power source.

(h) Range. The range of a signal is the distance between the source and the destination up to which the signal can be received in sufficient strength.

(i) Bandwidth. The bandwidth of an electronic circuit is the range of frequencies over which it operates nicely. Fox example, suppose an amplifier has a bandwidth of 300 Hz to 3100Hz. It means that the amplifier will amplify the signals nicely (i.e. with least distortion) in this frequency range. For signals outside this range, the amplification will be drastically reduced.

(j) Modulation. The signals in communication system (e.g. music, speech etc.) are low frequency signal and cannot be transmitted to large distances. In order to transmit the signal to large distances, it is superimposed on a high frequency wave (called carrier wave). This process is called modulation. Modulation is done at the transmitter and is an important part of communication system.

(k) Demodulation. The process of extracting signal from the modulated wave is called demodulation and is carried out in the receiver. This is reverse process of modulation.

(l) Repeaters. Repeaters are signal boosters installed at suitable locations in between the transmitter and the receiver. Each repeater receives the transmitted signal, amplifies the signal and transmits the amplified signal to the next

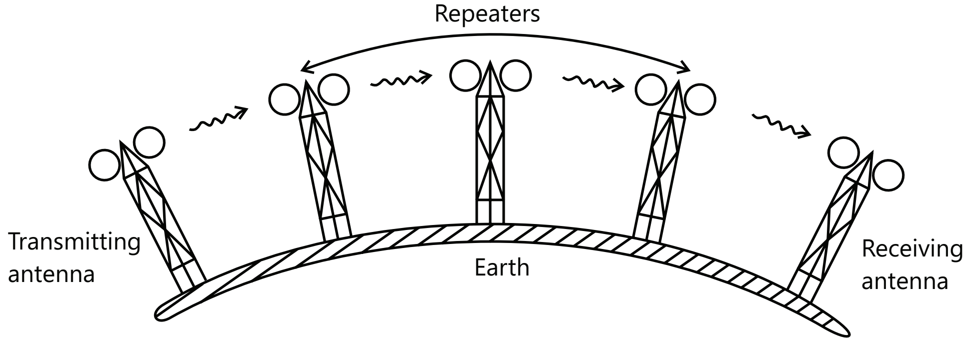


Figure 26.82

repeater (see figure). Obviously, a repeater is a combination of receiver, amplifier and transmitter. Repeaters are used to extend the range of a communication system. A communication satellite is essentially a repeater station in space.

4. BANDWIDTH OF SIGNALS

In electronic communication, message signals converted in the electrical form are transferred from one point to another point. These message signals are of two types viz.

1. Analog signals   2. Digital signals

(a) Bandwidth of analog signals. An analog signal is that in which the voltage or current varies continuously with time. In analog communication, the information or message to be transmitted is generally in continuous waveform. The range of frequencies which are necessary for satisfactory transmission of information or message contained in the analog signal is called Bandwidth of the analog signal.

Examples of analog signals are speech, music, sound produced by vibrating strings, picture (video) signals etc. These analog signals are converted into electrical form with suitable transducers and then transmitted to the required destination.

Different types of analog signals require different bandwidths.

(i) Bandwidth of speech signals. Speech signals contain frequencies between 300 Hz. And 3100 Hz. Therefore, speech signals require bandwidth=3100 Hz-300 Hz=2800 Hz.

(ii) Bandwidth of music signals. The audio range of frequencies produced by musical instruments is from 20 Hz to 20 kHz. Therefore, music signals require a bandwidth of about 20 kHz (20 kHz-20 Hz).

(iii) Bandwidth of video signals. For transmission of pictures, the video signals require a bandwidth of about 4.2 MHz. Since a TV signal contains both audio and video signals, it is usually allocated a bandwidth of 6 MHz for the transmission of TV signals.

(b) Bandwidth of digital signals. A digital signal is that in which voltage or current can have only two discrete values. Therefore, a digital signal is in the form of rectangular / square waves or pulses. Each pulse has two levels of voltage or current represented by 0 and 1. Examples of digital signals are: letters printed in a book, out-put of digital computer etc.

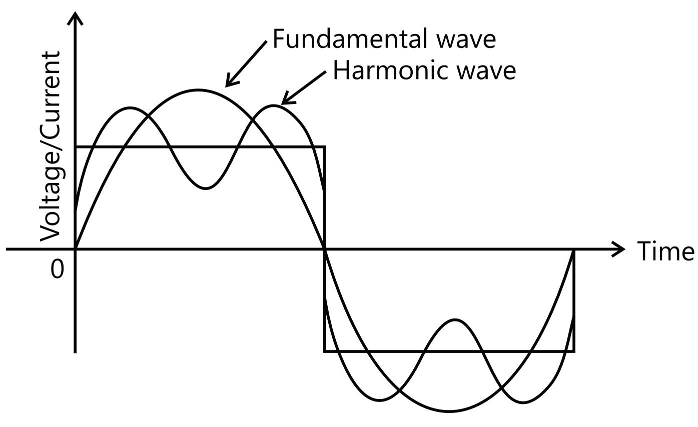


Figure 26.83

Theoretically infinite bandwidth is required for digital signals. This is illustrated in figure. We have seen that digital signals are in the form of rectangular waves. It can be shown that a rectangular waves. It can be shown that a rectangular waves can be considered as the superposition of a large number of sinusoidal waves of frequencies f_0, 2 f_0, 3 f_0, 4 f_0, ...n f_0 where n is an integer ranging from 1 to infinity. Here $f_0 = 1/t_0$ is called the fundamental frequency and 2 f_0, 3 f_0, 4 f_0,.... are called second harmonic, third harmonic, fourth harmonic, ... Therefore, the bandwidth of a digital signal is infinite. However, for all practical purpose, higher harmonics (e.g. 4 f_0, 5 f_0 .....) can be neglected because contribution of higher harmonics to shape of the wave is very small. Therefore, if the available bandwidth for a digital signal is large enough to accommodate a few harmonics the information contained in the digital signal is not lost and the rectangular digital signal is more or less recovered.

5. BANDWIDTH OF TRANSMISSION MEDIUM

Transmitter and receivers have some transmission medium between them so that message or information may be transferred from the transmitter to the receiver. Some commonly used transmission media are coaxial cables, optical fibers, free space etc. Different types of transmission media are suitable for different bandwidths.

(a) **Coaxial cables.** These are used for signals below 18 GHz. The commonly used coaxial cables have a bandwidth of 750 MHz

(b) **Optical fibers.** Optical fibers are suitable for microwaves and UV waves. The bandwidth of an optical fiber is about 10^{11} Hz.

(c) **Free space.** In radio communication, free space acts as the transmission medium. The frequency range of space communication is from 10^5 Hz to 10^9 Hz. This frequency range is further subdivided and allocated for different services.

Table 26.4: Some important wireless communication frequency bands

Service	Frequency bands	Comments
Standard AM broadcast	540-1600 kHz	
FM broadcast	88-108 MHz	
Television	54-72 MHz	VHF (very high frequencies)
	76-88 MHz	TV
	174-216 MHz	UHF (ultra high frequencies)
	420-890 MHz	TV
Cellular Mobile Radio	896-901 MHz	Mobile to base station
	840-935 MHZ	Base station to mobile
Satellite Communication	5.925-6.425 GHz	Uplink
	3.7-4.2 GHz	Downlink

6. PROPAGATION OF EM WAVES IN ATMOSPHERE

In radio communication, an antenna at the transmitter radiates the electromagnetic waves which travel through the space and reach the receiving antenna at the other end. Several factors influence the propagation of electromagnetic waves and the path they follow.

(a) As the electromagnetic waves travel away from the transmitter, their power goes on decreasing.

(b) The electromagnetic waves are little affected by the surrounding atmosphere, rain, snow etc and are able to penetrate non-metallic objects easily.

(c) The electromagnetic waves are stopped dead by metals or fine mesh screens.

The earth's atmosphere plays an important role in the propagation of electromagnetic waves. Therefore, it is desirable to discuss the composition of earth's atmosphere.

6.1 Earth's Atmosphere

The gaseous envelope around the earth is called atmosphere. It extends up to a height of about 400km above the surface of the earth. The earth's atmosphere is mainly composed of nitrogen (78%) and oxygen (21%). It also contains minute quantities of carbon dioxide, neon, water vapors, dust particles, etc. the density of air (ρ) goes on decreasing as we go up. The earth's atmosphere has been divided into the following regions (see figure).

(a) Troposphere: This region is up to a height of 12km from the earth's surface. It is in this region that clouds are formed. The temperature in troposphere decreases with height at the rate of about 6.5^0C per kilometer to a value of about -50^0C at its upper boundary. The density of air falls from ρ to $\rho/10$.

(b) Stratosphere: Above the troposphere lies the stratosphere. This region is from 12km to 50km from the earth's surface. An important part of stratosphere is the ozone layer which extends from 30km to 50km from the earth surface. The ozone layer contains ozone in abundance. It absorbs most of the ultraviolet radiation coming from the sun. The temperature falls from 280 K to 220 K and the density of air falls from $\rho/10$ to $\rho/1000$.

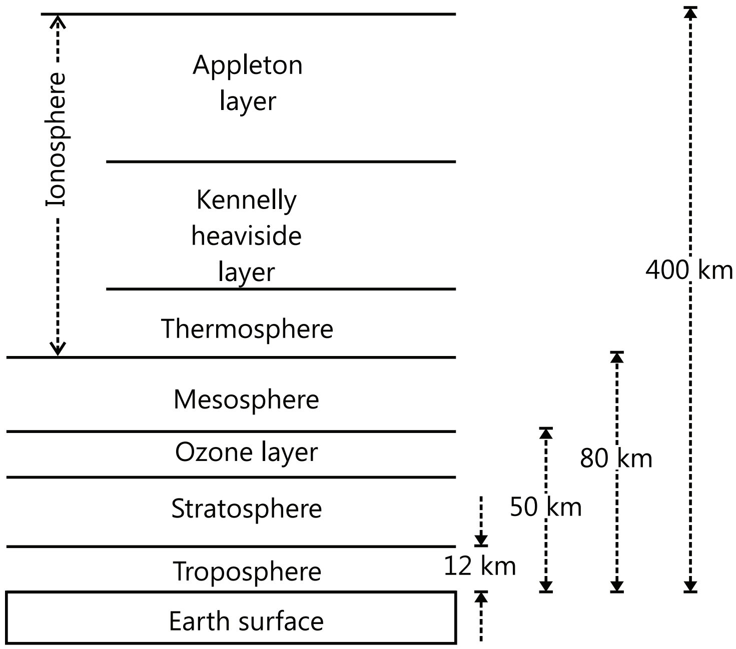


Figure 26.84

(c) Mesosphere: This region is from 50km to 80km from the earth's surface. Temperature falls from 220 K to 180 K. Density of air falls from $\rho/1000$ to $\rho/10^5$.

(d) Ionosphere: This region extends from 80 km to 400 km from the earth's surface. In this region, the constituent gases are ionized by ultraviolet radiation and X-rays from the sun. There are main layers viz, Heavy side layer and Appleton layer, in this region. The heights of these layers vary with the season and the day. Further, the layers are not fixed, but are irregular and of varying thickness.

Layers of Ionosphere: The ionosphere plays an important role in the propagation of radio waves in space communication. Because of the variation of the composition of air, there are several regions of varying ionization density within the ionosphere, dividing the ionosphere into several layers. The important layers of ionosphere are D, E, F_1 and F_2 layers as shown in figure. The ionization density (i.e. number of ions or electrons per unit volume) of each layer varies with time of the day, season, altitude etc.

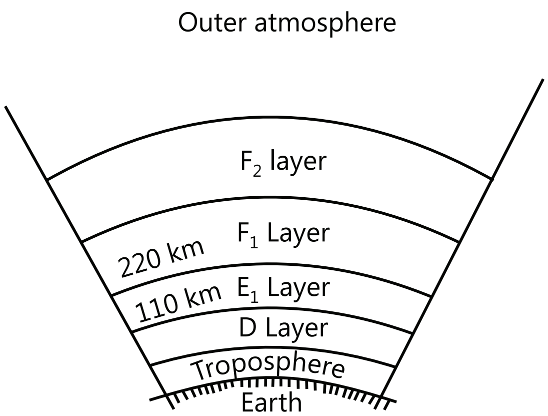


Figure 26.85

(a) The lowest layer, called D layer, exists only in the day time at an altitude of 50 to 90 km above the earth's surface. Ionization in this region is relatively weak and does not affect the direction of travel of radio waves.

(b) The next layer, the E layer, is in a region of about 90-140 km above the surface of earth. It has a maximum density at noon but is only weakly ionized at night.

(c) The last layer, the F layer, is quite variable. At night, it exists as a single layer in a region of about 140-400 km above the earth's surface. However in day time, it splits into two layers F_1 and F_2 as shown in figure.

When radio waves from a transmitter are directed towards the ionosphere, then radio waves in the frequency range 2 MHz to 40 MHz are reflected back to earth by the ionosphere. However, radio waves of frequencies greater than 40 MHz are not reflected back to earth by the ionosphere; they penetrate into ionosphere and escape. Thus ionosphere acts as a reflector for certain types of radio waves and helps in the long distance transmission of radio waves.

6.2 Classification of Radio Waves

In radio communication, free space acts as the transmission medium. The radio waves have a very wide frequency range from 500 kHz to 1000 MHz. This frequency range is divided into various categories as shown in the table below.

Frequency Band	Frequency Range	Wavelength Range	Typical Uses
Very low frequency (v.l.f.)	10-30 kHz	30,000-10,000 m	Long-distance point-to-point communication
Low frequency (l.f.)	30-300 kHz	10,000-1,000 m	Marine, navigational aids
Medium frequency (m.f.)	300-3,000 kHz	1,000-100 m	Broadcasting, marine
High frequency (h.f.)	3-30 MHz	100-10 m	Communication of all types
Very high frequency (v.h.f.)	30-300 MHZ	10-1 m	Television, f.m. broadcasting radar, air navigation, short-wave broadcasting
Ultra-high frequency (u.h.f.)	300-3,000 MHz	1m-10 cm	Radar, microwave relays, short-distance communication
Super-high frequency (s.h.f.)	3,000-30,000 MHz	10-1 cm	Radar, radio relay, navigation, experimental
Extremely high frequency (e.h.f.)	30,000-300,000 MHz	1-0.1 cm	Experimental

7. SPACE COMMUNICATION

The phenomenon of sending, receiving and processing information through space is called space communication.

In space communication, the signal is carried by high frequency electromagnetic wave (called carried wave) from the transmitter to the receiver in free space. Since no wires are used, it is also called wireless communication. The frequencies used in space communication lie in the range 10^4 Hz to 10^{11} Hz. Radio, television and satellite communication fall under this category.

Types of radio wave propagation. In space communication, the radio waves travel from the transmitting antenna to the receiving antenna in free space. Depending upon the frequency of radio waves, the distance between the transmitter and receiving antennas and the path (or paths) by which radio waves reach the receiving antenna, the radio wave propagation can be carried out in the following three ways.

(a) Ground or surface wave propagation

(b) Sky wave or ionosphere wave propagation

(c) Space wave or direct wave or tropospheric propagation.

7.1 Ground Wave Propagation

When the radio waves from the transmitting antenna propagate along the surface of earth so as to reach the receiving antenna, it is called ground wave propagation or surface wave propagation.

In ground wave propagation, the radio waves travel along the surface of earth as shown in figure. This mode of propagation is possible only when the transmitting and receiving antennas are close to the surface of the earth. As the ground wave glides over the surface of earth, it induces current in the earth. This gives rise to resistance losses and dielectric losses in the ground. The energy required to supply these losses must come from the ground wave. Therefore, the energy of a ground wave decreases as it passes over the earth surface. The ground losses increase rapidly with the

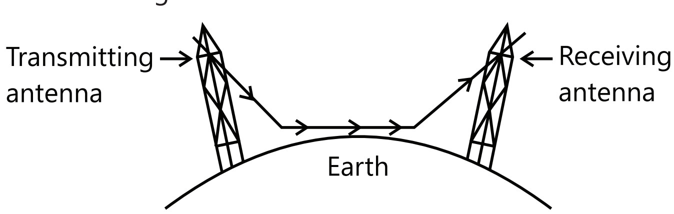


Figure 26.86

increase in frequency. Moreover, ground wave is also attenuated due to diffraction effect. As the wave propagates over the earth, it gradually tilts over. At some appreciable distance from the transmitting antenna, the wave lies down and dies.

We have seen that loss of power of a ground wave increases rapidly with the increase in frequency as well as with the increase in distance from the transmitting antenna. Therefore, ground wave propagation is limited to low frequency signals (500 kHz to 1500 kHz) and for short range communication. It cannot be used for high frequency and long range transmission.

(a) Below 500 kHz, reliable communication can be obtained over distance upto 1500 km by ground waves alone.

(b) Amplitude modulated radio waves in the medium frequency band are transmitted primarily via ground waves.

Advantages: Ground wave propagation has the following advantages:

(a) Ground wave propagation has excellent reliability.

(b) Reception is not affected by changing atmospheric conditions :

(c) Given enough transmitting power, ground wave communication can be maintained with any place in the world.

Disadvantages: Ground wave propagation has the following disadvantages:

(a) Ground wave propagation requires high transmitting power.

(b) Ground losses increase very rapidly with the increase in frequency of the signal.

(c) Ground losses very with surface material and composition.

7.2 Sky Wave Propagation

When the radio waves from the transmitting antenna reach the receiving antenna after reflection from the ionosphere, it is called Sky wave propagation or Ionospheric propagation.

We have already discussed that ionosphere has layers viz D, E, F_1 and F_2 layers of varying ionization density (i.e. number of ions or electrons per unit volume). These layers of ionosphere act as "radio mirror" for certain radio frequencies. It has been found that if radio waves in the frequency range 2 MHz to 30 MHz are directed towards the ionosphere, these waves are reflected back to earth by the various layers of the ionosphere as shown in figure. If the frequency of the radio waves is more than 30 MHz, these are not reflected back to earth from the ionosphere and penetrate into the ionosphere and escape. For this reason, we use sky wave escape. For this reason, we use sky wave propagation for radio waves lying in the frequency range 2 MHz to 30 MHz. The sky wave propagation is also known as ionosphere propagation because sky waves reach the receiver after reflection from the ionosphere.

The sky wave propagation is quite unreliable. It is because whether or not a radio wave is reflected back to earth by the ionosphere depends upon several factors including

(a) Frequency of radio wave

(b) Ionization density of ionosphere and

(c) The angel of incidence at which the radio wave enters the ionosphere.

Yet it is primary means of around the world short-wave communication.

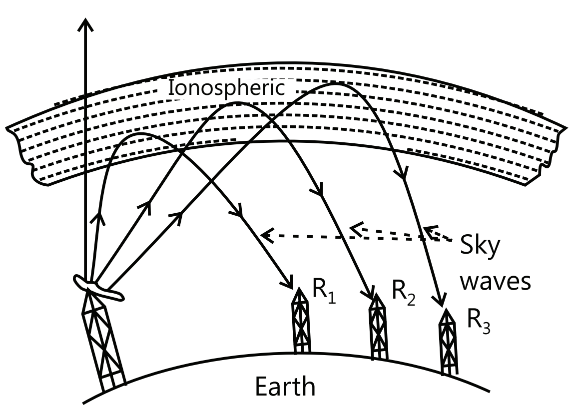


Figure 26.87

(a) **Critical frequency.** As the frequency of sky waves increases, the ionosphere becomes progressively less effective in reflecting the waves back to earth. At a certain maximum frequency of radio wave, the wave is not at all reflected back to earth. The highest frequency above which the ionosphere no longer returns the sky wave back to earth when transmitted in vertical direction is called critical frequency.

Since the critical frequency depends on the density of ionization, it will clearly vary with the time of day and season of the year. Furthermore, it is possible for a particular frequency to pierce the E layer but still be returned from the F layer because it has higher density of ionization. Of course, a still higher frequency will pierce both layers and be lost. The value of critical frequency (f_c) is given by ; $f_c = 9(N_{max})^{1/2}$ Where N_{max} = maximum electron density of ionosphere.

(b) Critical angle. As the vertical angle of the sky wave w.r.t. earth is increased, the ionosphere layers are no longer capable of reflecting sky waves back to earth. For a given frequency, the vertical angle above which the sky wave no longer returns to earth but travels outward into space is called critical angle.

Sky waves at or above critical angle may be refracted (bent) by ionosphere but they are not reflected back to earth. The critical angle primarily depends on density of ionization and on the frequency of the wave.

(c) Maximum usable frequency (MUF). It is defined as the highest frequency of the radio waves which when sent at a certain angle towards the given layer of ionosphere gets reflected from that layer and returns to earth. It is given by ;

Maximum usable frequency, $\text{MUF} = \frac{f_c}{\cos\theta} = f_c \sec\theta$

Where θ = Angle between normal and direction of incident waves.

Since the maximum usable frequency (MUF) depends on the density and height of ionosphere layers, it will vary from hour to hour, from day to day and from location to location.

(d) Skip distance. The distance between the transmitting aerial and the point where the sky wave is first received after returning to earth is called skip distance.

(e) Fading. When signals are received via sky waves, it often happens that the signal strength will increase and decrease periodically. In certain cases, the signal may be lost completely or may be drowned in the noise level. It is called fading. Fading is caused due to the following reasons:

(i) Multiple-path reception. For example, consider a receiving antenna that receives ground waves as well as sky waves. The signal strength will be the resultant of the two. This interference causes fading of the signal.

(ii) Ionospheric conditions. As the conditions in the ionosphere change, the phasing of sky wave will shift, and the resultant strength of the signal will also vary. This condition leads to fading of the signal.

(iii) Receiving antenna at the edge of skip distance. Fading also occurs when the receiving antenna is located at the edge of the skip distance. A slight change in the conditions of ionosphere may place the receiving antenna inside or beyond the skip distance. As a result, fading of signal may take place.

7.3 Space Wave Propagation (Line of Sight Propagation)

At frequencies above 30 MHz radio transmission cannot be carried out by ground waves or sky waves. It is because ground waves are quickly attenuated at such high frequencies and at frequencies above 30 MHz, the ionosphere is unable to reflect the sky waves back to earth. Therefore, we use direct waves (also called space waves) which travel directly in straight lines from the transmitting antenna to the receiving antenna as shown in figure. This mode of propagation is called space wave propagation or line of sight (LOS) propagation or tropospheric propagation.

When the radio waves from the transiting antenna travelling in straight line directly reach the receiving antenna, it is called space wave propagation.

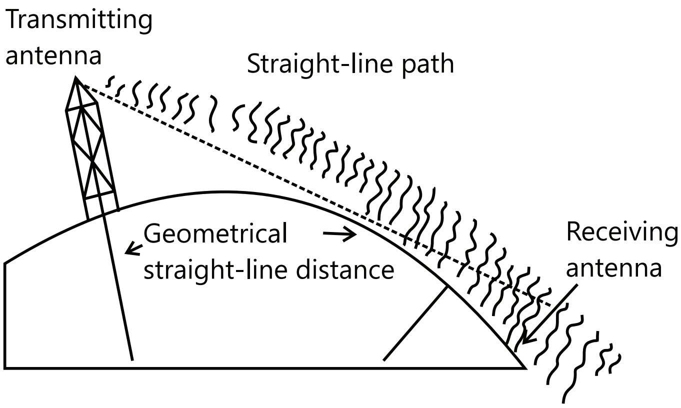


Figure 26.88

Wave propagation or line of Sight Propagation.

As the frequency increases, the radio waves tend to travel more and more in straight lines. Therefore, the receiving antenna must receive the signal directly from the transmitting antenna. The television frequencies lie in the range 100 MHz to 200 MHz. Therefore, reception of TV signals is possible only if the receiving antenna directly intercepts the signal. Similarly, radar, microwave relays and many other services solely depend on space wave propagation. The range of space wave propagation is limited by

(a) The line of sight distance i.e. the distance at which the transmitting and receiving antennas can see each other.

(b) The curvature of earth.

The space waves of direct waves travel essentially in a line of sight path. Due to curvature of earth, line of sight distance will depend upon the antenna height. Therefore, the maximum range of communication through space waves is determined by the height of the transmitting and receiving antennas. This is illustrated in figure.

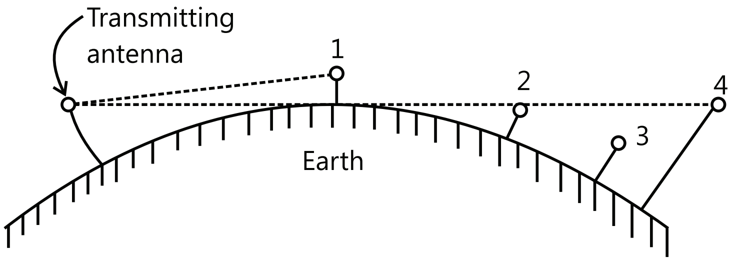


Figure 26.89

At receiving location 1, the height of the antenna is within the line of sight distance. For the same antenna height, receiving location 2 would be the maximum range of communication for direct waves. However, if we increase the height of the receiving antenna, the communication range can be increased to reach the location 4.

7.4 Range of TV Transmission

The TV signals are in the 100-200 MHz range. Therefore, transmission of such signals via ground waves or sky waves is not possible. In such situations, we use line of sight transmission i.e., TV signals are transmitted by directed waves.

Consider a TV transmitting antenna OP of height h located at point O on the surface of earth as shown in figure. When TV broadcast is made, the signal can reach up to tangent point A and B on the surface of earth. There will be no reception of the signal beyond points A and B. therefore, distance

On the earth surface is the range of TV transmission (d) and the height (h) of the transmitting antenna.

If C is the center of earth and R (=AC) is its radius, the from right angled triangle PAC, we have,

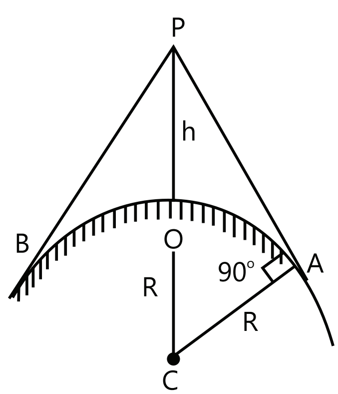


Figure 26.90

$(PC)^2 = (PA)^2 + (AC)^2$ or $(h+R)^2 = (PA)^2 + R^2$

Since the height (h) of the transmitting antenna is very small as compared to the radius (R) of the earth, $PA = PB \simeq d$.

$\therefore (h+R)^2 = d^2 + R^2$ or $d^2 = h^2 + 2Rh$

$\therefore d^2 = 2Rh$ or $d = \sqrt{2Rh}$

Note that the range of TV transmission depends upon the height of the transmitting antenna. The greater the height of the transmitting antenna, the larger is the range of TV transmission. For this reason, TV broadcasts are made from tall transmitting antennas.

Area covered by TV signal $\pi d^2 = \pi(2Rh)$ $[\because\ d = \sqrt{2Rh}]$

Population covered by TV signal = Population density x Area covered.

7.5 Maximum Communication Range for Space Waves

Figure shows space wave propagation between the transmitting antenna and the receiving antenna. Since space waves or direct waves follow the straight line path, they get blocked at some point C due to the curvature of earth. Thus earth presents a horizon to space wave propagation called radio horizon. The distance d_T is called radio horizon of the transmitting antenna while the distance d_R is called radio horizon for the receiving antenna. Therefore, maximum line of sight distance d_M between the transmitting and receiving antennas is

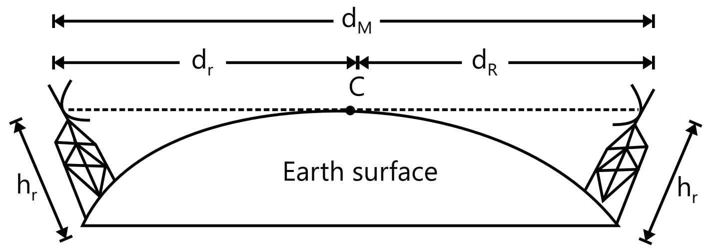


Figure 26.91

If h_T and h_R are the heights of transmitting and receiving antennas respectively and R is the radius of earth, then,

$d_r = \sqrt{2Rh_r}$ and $d_R = \sqrt{2Rh_R}$; $\therefore$ $d_M = \sqrt{2Rh_r} + \sqrt{2Rh_R}$

Clearly, if the signal is to be received beyond the horizon, then the height of the receiving antenna must be large enough to intercept the space or direct waves. However, space wave communication is greater than 100 km is hardly used in commercial communication. For an example, if h_T and h_R are 100m each, then using above formula, d_M comes out to be 71.4 km only. Sometimes, the range of TV transmission is increased by using repeater transmitting stations.

8. MODULATION

As discussed earlier, a high carrier wave is used to carry the audio signal. The question arises how the audio signal should be “added” to the wave. The solution lies in changing some characteristic of carrier wave in accordance with the signal. Under such conditions, the audio signal will be contained in the resultant wave. This process is called modulation and may be defined as under:

The process of changing some characteristic (e.g. amplitude, frequency or phase) of a carrier wave in accordance with the intensity of the signal is known as modulation. Modulation means “to change”. In modulation, some characteristic of carrier wave is called modulated wave or radio wave and contains the audio signal.

Need for modulation. Audio signals have a frequency range from 20 Hz to 20 kHz. These low frequency signals cannot be transmitted directly (i.e. without modulation) into space for the following reasons:

(a) Practical antenna length. For efficient transmission and reception, the transmitting and receiving antennas must have a length at least equal to $\lambda / 4$ where λ is the wavelength of the signal. For an audio signal of 15 kHz, the length of the antenna required will be:

Wavelength, $\lambda = \frac{c}{f} = \frac{3x10^8}{15x10^3} = 20000\text{m}$. $\therefore$ Length of antenna, $l = \frac{\lambda}{4} = \frac{20000}{4} = 5000\text{m}$

To set up a vertical antenna of this length is practically impossible. However, if a carrier wave of 1 MHz is used to carry the signal (i.e. modulation is done), the length of the antenna comes out of be 75 m only and this size can be easily constructed. Length of the antenna comes out to be 75m only and this size can be easily constructed.

(b) Effective power radiated by antenna. For a linear antenna of length l, it has been found that

Power radiated, $P \propto \left(\frac{l}{\lambda}\right)^2$. Therefore, for the same antenna length l, the power radiated will be large for signals of shorter wavelength or higher frequency. For good transmission, we need high powers. This requires the transmission to be carried out at high frequencies. For this reason, we use high frequency carrier wave to carry the audio signal.

(c) Mixing up of signals from different transmitters. All the audio signals from different transmitters have the same frequency range i.e. 20 Hz to 20 kHz. Therefore, if the audio signals from various transmitters are transmitted directly, they will get mixed up and there is no way to distinguish between them. This difficulty is solved by allotting different carrier frequencies to different transmitting stations. The above discussion shows the need for frequency transmission. For this purpose, the audio signal is superimposed on the high frequency waves. Hence modulation permits the transmission to occur at high frequency while it simultaneously allows the carrying of the audio signal.

8.1 Types of Modulation

As you will recall, modulation is the process of changing amplitude or frequency or phase of a carrier wave in accordance with the intensity of the signal. The carrier wave is a sinusoidal wave and can be represented as:

$e_c = E_C \cos(\omega_c t + \phi)$ where, e_c= Instantaneous voltage of carrier wave

E_C= Amplitude of carrier wave;

$\omega_c = 2\pi f_c$ Angular velocity at carrier frequency f_c

ϕ = Phase angle

Depending upon whether we change E_C (amplitude), (frequency) or ϕ (phase), modulation is of three types viz.

(a) Amplitude Modulation

(b) Frequency Modulation

(c) Phase Modulation

In India amplitude modulation is used in radio broadcasting. However, in television transmission, frequency modulation is used for sound signal and amplitude modulation for picture signal. We shall discuss these types of modulation in turn.

8.2 Amplitude Modulation

When the amplitude of high frequency carrier wave is changed in accordance with the intensity of the signal, it is called amplitude modulation.

In amplitude modulation, only the amplitude of the carrier wave is changed in accordance with the intensity of the signal. However, the frequency of the modulated wave remains the same i.e., carrier frequency. Figure shows the principle of amplitude modulation. Figure shows the audio electrical signal whereas Figure (ii) shows a carrier wave a constant amplitude. Figure (iii) shows the amplitude modulated (AM) wave.

Note that the amplitude of both positive and negative half-cycles of carrier wave is changed in accordance with the signal. For instance, when the signal is increasing in the positive sense, the amplitude of carrier wave also increases. On the other hand, during negative half-cycle of the signal, the amplitude of carrier wave decreases. Amplitude modulation is done by electronic circuit called modulator.

The following points are worth noting in amplitude modulation:

(i) The amplitude of the carrier wave changes according to the intensity of the signal.

(ii) The amplitude variation of the carrier wave is at the signal frequency f_s.

(iii) The frequency of the amplitude modulated wave remains the same i.e., carrier frequency f_c.

Modulation Factor. An important consideration in amplitude modulation is to describe the depth of modulation i.e., the extent to which the amplitude of carrier wave is changed by the signal. This is described by a factor called modulation factor which may be defined as under:

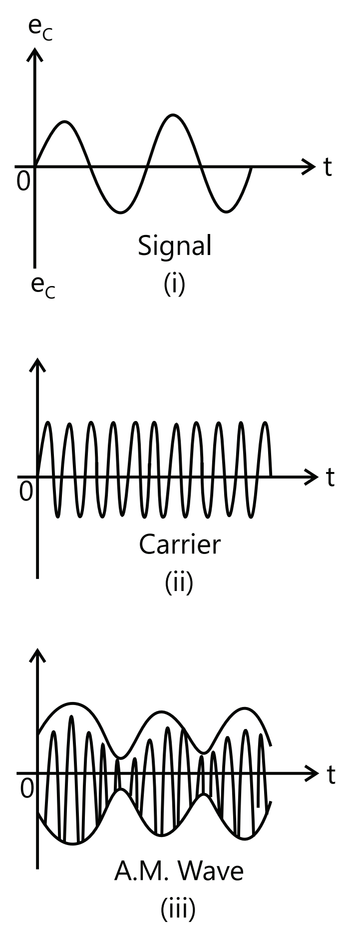


Figure 26.92

The ratio of change of amplitude of carrier wave to the amplitude of normal carrier wave is called the modulation factor m_a i.e.

$$\text{Modulation factor, } m_a = \frac{\text{Amplitude change of carrier wave}}{\text{Normal carrier wave (unmodulated)}} = \frac{E_S}{E_C}$$

Where E_S is the amplitude of the signal and E_c is the amplitude of normal carrier wave. Clearly, the modulation factor depends upon the amplitudes of the signal and carrier.

Obviously, $E_S = m_a E_C$

Figure shows the waveform of amplitude modulated wave. If the maximum and minimum voltage of AM wave are V_{max} and V_{min} respectively, then it is clear from figure that:

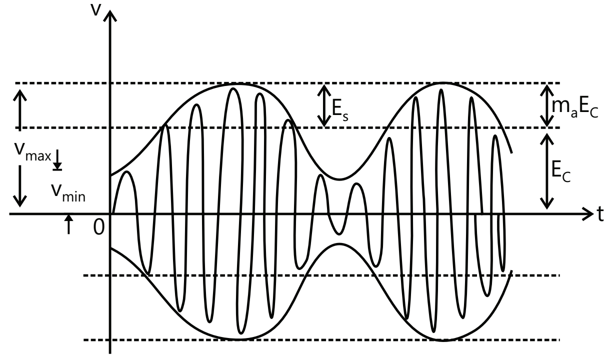


Figure 26.93

$$E_C = \frac{V_{max} + V_{min}}{2}\ ; \quad E_S = \frac{V_{max} - V_{min}}{2} \quad \therefore \quad \text{Modulation factor, } m_a = \frac{E_S}{E_C} = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

Modulation factor generally lies between 0 and 1.

8.2.1 Analysis of Amplitude Modulated Wave

A carrier wave may be represented by; $e_c = E_C \cos\omega_c t$

Where e= instantaneous voltage of carrier. E_C = Amplitude of carrier

$\omega_C = 2\pi f_c$ = Angular velocity at carrier frequency f_c.

In amplitude modulation, the amplitude E_C of the carrier wave is varied in accordance with the intensity of the signal as shown in figure. Suppose the modulation factor is m_a. It means that signal produces a maximum change of $m_a E_C$ in the carrier amplitude. Obviously, the amplitude of signal is $m_a E_C$. Therefore, the signal can be represented by:

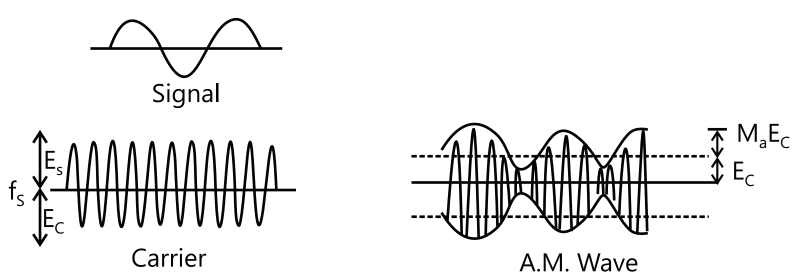


Figure 26.94

$e_s = m_a E_C \cos\omega_s t$

Where e_s = Instantaneous voltage of signal

$m_a E_C$ = Amplitude of signal

$\omega_s = 2 f_s$ = Angular velocity at frequency f_s.

The amplitude of the carrier wave varies at signal frequency f_s. Therefore, the amplitude of AM wave is given by

$$= E_C + m_a E_C \cos\omega_s t = E_C (I + m_a \cos\omega_s t)$$

The instantaneous voltage of AM wave is: e= Amplitude x $\cos\omega_c t$

$$= E_C (I + m_a \cos\omega_c t)\cos\omega_s t = E_C \cos\omega_c t + m_a E_C \cos\omega_s t \cos\omega_c t$$

$$= E_C \cos\omega_c t + \frac{m_a E_C}{2}(2\cos\omega_s t \cos\omega_c t) = E_C \cos\omega_c t + \frac{m_a E_C}{2}\left[\cos(\omega_c + \omega_s)t + \cos(\omega_c - \omega_s)t\right]^*$$

$$= E_C \cos\omega_c t + \frac{m_a E_C}{2}\cos(\omega_c + \omega_s)t + \frac{m_a E_C}{2}\cos(\omega_c - \omega_s)t$$

The following points may be noted from the above equation of amplitude modulated wave:

(a) The AM wave is equivalent to the summation of three sinusoidal waves; one having and the third having amplitude E_C and frequency $f_{c.,}$ the second having amplitude $m_a E_C / 2$ and frequency f_c+f_s) and the third having amplitude $m_a E_C / 2$ and frequency ($f_c - f_s$).

(b) The AM wave contains three frequencies viz. $f_{c.,}$ (f_c + f_s). The first frequency is the carrier frequency. Thus, the process of modulation does not change the original carrier frequency but produces two new frequencies (f_c + f_s) and (f_c – f_s) which are called sideband frequencies.

From trigonometry, we have the expansion formula: $2\cos A \cos B = \cos(A+B) + \cos(A-B)$

$$f_c = \frac{\omega_c}{2\pi}, \; f_c + f_s = \frac{\omega_c + \omega_s}{2\pi}, \qquad f_c - f_s = \frac{\omega_c - \omega_s}{2\pi}$$

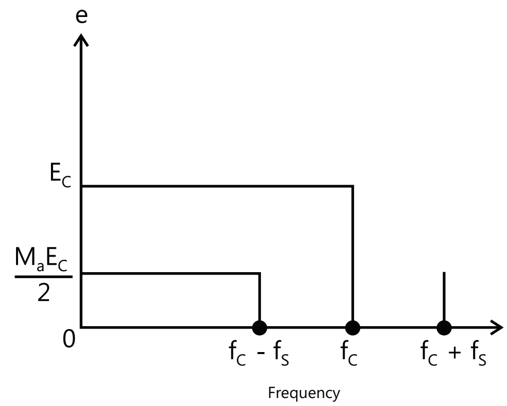


Figure 26.95

(c) The sum of carrier frequency and signal frequency i.e., $f_c - f_s$ i.e., the difference between carrier and signal frequencies.

8.2.2 Sideband Frequencies in AM Wave

In an amplitude modulated wave the sideband frequencies are of our interest. It is because the signal frequency f_s is contained in the sideband frequencies. Figure shows the frequency spectrum of an amplitude modulated wave. The frequency components in the AM wave are shown by vertical lines. The height of each vertical line is equal to the amplitude of the components present. It may be added here that in practical radio transmission, carrier frequency f_c is many times greater than signal frequency f_s. Hence, the sideband frequencies are generally close to the carrier frequency. It may be seen that a carrier modulated by a single frequency is equivalent to three simultaneous signals; the carrier itself and two other steady frequencies

i.e., $f_c + f_s$ and $f_c - f_s$.

Let us illustrate sideband frequencies with an example. Suppose the carrier frequency is 400 kHz and the signal frequency is 1 kHz. The AM wave will contain three frequencies viz, 400 kHz, 401 kHz and 399 kHz. It is clear that upper sideband frequency (401 kHz) and sideband frequency (399 kHz) are very close to the carrier frequency (400 kHz).

Bandwidth: In an AM wave, the bandwidth is from $\left(f_c - f_s\right)$ to $\left(f_c + f_s\right)$ i.e., $2f_s$. Thus in the above example, bandwidth is form 399 to 401 kHz which is twice the signal frequency. Therefore, we arrive at a very important conclusion that in amplitude modulation, bandwidth is twice the signal frequency. The tuned amplifier which is called upon to amplify the modulated wave must have the required bandwidth to include the sideband frequencies. If the turned amplifier has insufficient bandwidth, the upper sideband frequencies may not be reproduced by the radio receiver.

8.2.3 Power in AM Wave

The power dissipated in any circuit is a function of the square of voltage across the circuit and the effective resistance of the circuit. Equation of AM wave reveals that it has three components of amplitude $E_C, m_a E_C / 2$ and $m_a E_C / 2$. Clearly, power output must be distributed among these components.

Carrier power $$P_C = \frac{\left(E_C / \sqrt{2}\right)^2}{R} = \frac{E_C^2}{2R} \qquad \text{... (i)}$$

Total power of sidebands $$P_S = \frac{\left(m_a E_C / 2\sqrt{2}\right)^2}{R} + \frac{\left(m_a E_C / 2\sqrt{2}\right)^2}{R} = \frac{m_a^2 E_C^2}{8R} + \frac{m_a^2 E_C^2}{8R} = \frac{m_a^2 E_C^2}{4R} \qquad \text{... (ii)}$$

Total power of AM wave $P_T = P_C + P_S; = \frac{E_C^2}{2R} + \frac{m_a^2 E_C^2}{4R} = \frac{E_C^2}{2R}\left[1 + \frac{m_a^2}{2}\right]$ or $P_T = \frac{E_C^2}{2R}\left[\frac{2 + m_a^2}{2}\right]$ .... (iii)

Fraction of total power carried by sidebands is $\frac{P_S}{P_T} = \frac{\text{Exp. (ii)}}{\text{Exp. (iii)}} = \frac{m_a^2}{2 + m_a^2}$ ... (iv)

As the signal is contained in the sideband frequencies, therefore, useful power is in the sidebands. Inspection of exp. (iv) reveals that sideband power depends upon the modulation factor m_a. The greater the value of m_a, the greater is the useful power carried by the sidebands. This emphasizes the importance of modulation factor is

(a) When m_a =0, power carried by sidebands $= 0^2 / 2 + 0^2 = 0$

(b) When $m_a = 0.5$, power carried by sidebands $= \frac{(0.5)^2}{2 + (0.5)^2} = 11.1\%$ of total power of AM wave

(c) When m_a =1, power carried by sidebands $= \frac{(1)^2}{2 + (1)^2} = 33.3\%$ of total power of AM wave.

As an example, suppose the total power of an AM wave is 600 watts and modulation is 100%. Then sideband power is 600/3=200 watts and carrier power will be 600-200=400 watts.

The sideband power represents the signal content and the carrier power is that power which is required as the means of transmission.

Note. $P_C = \frac{E_C^2}{2R}$ $P_S = \frac{m_a^2 E_C^2}{4R}$ $\therefore \frac{P_S}{P_C} = \frac{1}{2} m_a^2$ or $P_S = \frac{1}{2} m_a^2 P_C$ .... (v)

Expression (v) gives the relation between total sideband power (P_S) and carrier power (P_C).

8.2.4 AM Modulator

A circuit which does amplitude modulation is called AM modulator. Figure shows the circuit of a simple AM modulator. It is essentially a CE amplifier having a voltage gain of A. the carrier signal is the input to the amplifier. The modulating signal is applied in the emitter resistance circuit.

Working: The carrier voltage e_c is applied at the amplifier and the modulating signal e_s is applied in the emitter resistance circuit. The amplifier circuit amplifies the carrier by a factor "A" so that the output is Ae_c. Since the modulating signal is a part of the biasing circuit, it produces low-frequency variations in the emitter circuit. This in turn causes *variations in "A". The result is that amplitude of the carrier varies in accordance with the strength of the signal. Consequently, amplitude modulated output is obtained across R_L. It may be noted that carrier should not influence the voltage gain A; only the modulating signal should be this. To achieve this objective, carrier should have a small magnitude and signal should have a large magnitude.

8.2.5 Limitations of Amplitude Modulation

Although theoretically highly effective, amplitude modulation suffers from the following drawbacks:

(a) **Noisy reception:** In an AM wave, the signal is in the amplitude variations of the carrier. Practically all the natural and man-made noises consist of electrical amplitude disturbances. As a radio receiver cannot distinguish between amplitude variations that represent noise and those that contain the desired signal, therefore, reception is generally noisy.

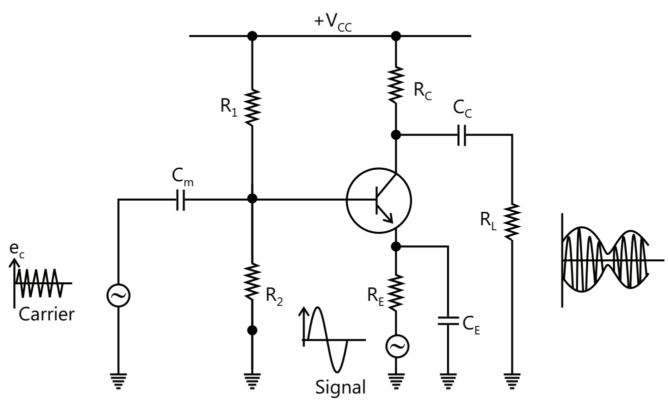


Figure 26.96

(b) Low efficiency: In amplitude modulation, useful power is in the sidebands as they contain the signal. As discussed before, an AM wave has low sideband power. For example, if modulation is 100%, the sideband power is only-third of the total power of AM wave. Hence the efficiency of this type of modulation is low.

(c) Small operating quality. Due to low efficiency of amplitude modulation, transmitters employing this method have a small operating range i.e., message cannot be transmitted over larger distances.

(d) Lack of audio quality: This is a distinct disadvantage of amplitude modulation. In order to attain high-fidelity reception, all audio frequencies up to 15 kHz must be reproduced. This necessitates bandwidth of only 10 kHz to minimize the interference from adjacent broadcasting stations. This means that the highest modulating frequency can be 5 kHz which is hardly sufficient to reproduce the music properly.

8.3 Demodulation

The process of recovering the audio signal from the modulated wave is known as demodulation or detection.

At the broadcasting station, modulation is done to transmit the audio signal over larger distance to a receiver. When the modulated wave is picked up by the radio receiver, it is necessary to recover the audio signal from it. This process is accomplished in the radio receiver and is called demodulation.

Necessity of demodulation. It was noted that amplitude modulated wave consists of carrier and sideband frequencies. The audio signal is contained in the sideband frequencies which are radio frequencies. If the modulated wave after amplification is directly fed to the speaker as shown in figure, no sound will be heard. It is because diaphragm of the speaker is not at all able to respond to such high frequencies. Before the diaphragm is able to move in one direction, the rapid reversal of current tends to move it in the opposite direction i.e. diaphragm will not move at all. Consequently, no sound will be heard.

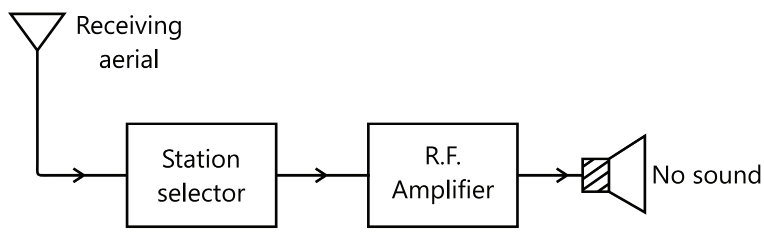


Figure 26.97

From the above discussion, it follows that audio signal must be separated from the carrier at a suitable stage in the receiver. The recovered audio is then amplified and fed to the speaker for conversion into sound.

8.3.1 Essentials in Demodulation

In order that a modulated wave is audible, it is necessary to change the nature of modulated wave. This is accomplished by a circuit called detector. A detector circuit performs the following two functions:

(a) It rectifies the modulated wave. i.e. negative half of the modulated wave is eliminated. As shown in figure (i) a modulated wave has positive and negative halves exactly equal. Therefore, average current is zero and speaker cannot respond. If the negative half of this modulated wave is eliminated as shown in figure, the average value of this wave will not be zero since the resultant pulses are now all in one direction. The average value is shown by the dotted line in figure (ii). Therefore, the diaphragm will have definite displacement corresponding to the average wave is similar to that of the modulation envelope. As the signal is of the same shape as the envelope, therefore, average wave shape is of the same form as the signal.

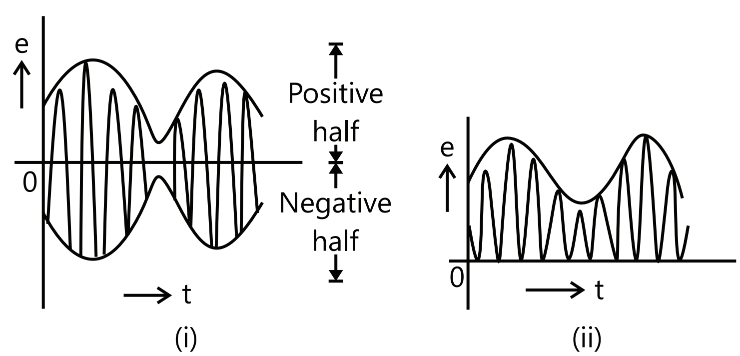


Figure 26.98

(b) It separates the audio signal from the carrier. The rectified modulated wave contains the audio signal and the carrier. It is desired to recover the audio signal. This is achieved by a filter circuit which removes the carrier frequency and allows the audio signal to reach the load i.e., speaker.

8.4 A. M. Diode Detector

Figure below shows a simple detector circuit employing junction diode and filter circuit. The modulated wave of desired frequency is selected by the parallel turned circuit L_1C_1 and is applied to the junction diode. During the positive half-cycles of modulated wave, the diode conducts while during negative half-cycles, it does not. The result of this rectifying action is that output of consists of positive half-cycles of modulated wave as shown.

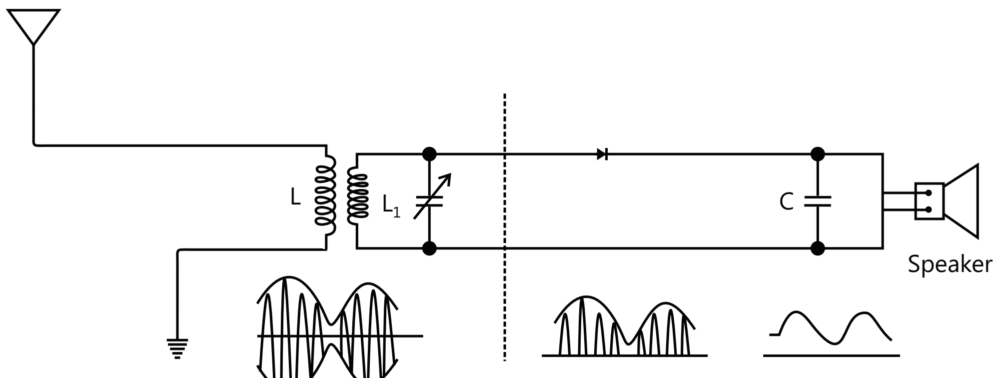


Figure 26.99

The rectified modulated wave contains radio frequency and the signal cannot be fed to the speaker for sound reproduction. If done so, no sound will be heard due to the inertia of speaker diaphragm. The r.f. (radio frequency) component is filtered by the capacitor C shunted across the speaker. The value of this capacitor is sufficiently large to present low reactance to the r.f. (radio frequency) component while presenting a relatively high reactance to the audio signal. The result is that the r.f. (radio frequency) component is by-passed by the capacitor C and the signal is passed on to the speaker for sound reproduction.

Illustration 1: Calculate the length of a half-wave dipole antenna at (i) 30 MHz (ii) 300 MHz and (iii) 3000 MHz. What inference do you draw from the results? **(JEE MAIN)**

Sol: The wavelength of half-wave dipole antenna at certain frequency is given by $\lambda = \frac{c}{f}$

Speed of radio waves, $c = 3x10^8\ ms^{-1}$

(i) $\lambda = \frac{c}{f} = \frac{3x10^{-8}}{30x10^6} = 10m$ $\therefore$ Length of half-wave dipole antenna is $l = \lambda/2 = 10/2 = 5m$

(ii) $\lambda = \frac{c}{f} = \frac{3x10^8}{300x10^6} = 1m$ $\therefore$ Length of half-wave dipole antenna is $l = \lambda/2 = 1/2 = 0.5m$

(iii) $\lambda = \frac{c}{f} = \frac{3x10^8}{3000x10^6} = 0.1m$ $\therefore$ $l = \lambda/2 = 0.1/2 = 0.05m$

It is clear that length of the dipole antenna decreases as the frequency of carrier wave increase.

Illustration 2: The maximum peak-to-peak voltage of an amplitude modulated wave is 16 mV and the minimum peak-to-peak voltage is 4 mV. Calculate the modulation factor. **(JEE MAIN)**

Sol: The modulation factor is $m_a = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$.

Figure shows the conditions of the problem.

Maximum voltage of AM wave, $V_{max} = \frac{16}{2} = 8mV$

Minimum voltage of AM wave, $V_{min} = \frac{4}{2} = 2mV$

$$\therefore m_a = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} = \frac{8-2}{8+2} = \frac{6}{10} = 0.6$$

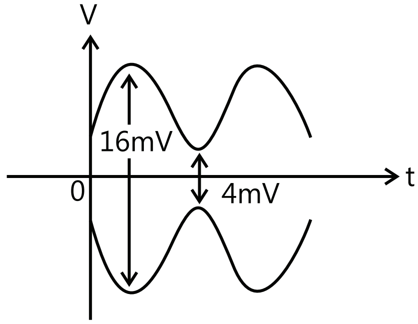


Figure 26.100

Illustration 3: An AM wave is represented by the expression:

$V = 5(1 + 0.6\cos 6280t)\ \sin 211x10^4\ t$ Volts **(JEE ADVANCED)**

(i) What are the minimum and maximum amplitudes of the AM wave?

(ii) What frequency components are contained in the modulated wave and what is the amplitude of each component?

Sol: Comparing the equation with standard wave equation $V = E_C(1 + m_a \cos\omega_s t)\ \sin\omega_c\ t$ we get the maximum carrier amplitude E_C, the modulation factor m_a and ω. Using this we get the minimum and maximum amplitude of the wave as $E_c - m_aE_c$ and $E_c + m_aE_c$ respectively. The frequency is given by $f = \frac{\omega}{2\pi}$.

The AM wave equation is given by $V = 5(1 + 0.6\cos 6280\ t)\ \sin 211x10^4\ t$ volts ... (i)

Compare it with standard AM wave $V = E_C(1 + m_a \cos\omega_s t)\ \sin\omega_c\ t$ ... (ii)

From eqs. (i) and (ii), we get, E_C = Carrier amplitude = 5V; m_a = Modulation factor = 0.6

f_s = Signal frequency = $\omega_s / 2\pi = 6280 / 2\pi = 1kHz$

f_c = Carrier frequency = $\omega_s / 2\pi = 211x10^4 / 2\pi = 336kHz$

(i) Minimum amplitude of AM wave = $E_c - m_aE_c = 5 - 0.6x5 = 2V$

Maximum amplitude of AM wave $= E_c + m_aE_c = 5 + 0.6x5 = \delta V$

(ii) The AM wave will contain three frequencies viz.

	$f_c - f_s$	f_c,	$f_c + f_s$
or	336-1	336,	336+1
or	335 kHz,	336 kHz,	337 kHz

The amplitudes of the three components of AM wave are:

	$\frac{m_a E_c}{2}$	E_c	$\frac{m_a E_c}{2}$
or	$\frac{0.6x5}{2}$	5,	$\frac{0.6x5}{2}$
or	1.5 V,	5 V,	1.5 V

Illustration 4 : An 50 kW carrier is to be modulated to a level of (i) 80% (ii) 10%. What is the total sideband power in each case? **(JEE MAIN)**

Sol: The power used in modulation m_a is given by $P_s = \frac{1}{2}m_a^2\ P_c$ where P_c is the power of carrier wave.

(i) $P_s = \frac{1}{2}m_a^2\ P_c = \frac{1}{2}(0.8)^2\ x50 = 16kW$

(ii) $P_s = \frac{1}{2}m_a^2\ P_c = \frac{1}{2}(0.1)^2\ x50 = 0.25kW$

Note the effect of modulation factor on the magnitude of sideband power. In the first case (m_a = 80%), we generated and transmitted 50 kW carrier in order to send 16 kW of intelligence. In the second case (m_a = 10%), the same carrier of operation decreases rapidly as modulation factor decreases. For this reason, in amplitude modulation, the value of m_a is kept as close to unity as possible.

Illustration 5: Consider an optical communication system operating at $\lambda = 800\,nm$. Suppose, only 1% of optical source frequency is the available channel bandwidth for optical communication. How many channels can be accommodated for transmitting (i) audio signals requiring a bandwidth of 8 kHz (ii) video TV signals requiring an approximate bandwidth of 4.5 MHz? **(JEE MAIN)**

Sol: The number of the channels for audio or video signal is obtained as

$N = \frac{\text{Total bandwidth}}{\text{Bandwidth of signal}}$. The total bandwidth is obtained as $0.01 \times f = \frac{0.01 \times c}{\lambda}$.

Optical wavelength, $\lambda = 800\,nm = 8x10^{-7}\,m$

Frequency of optical source is $f = \frac{c}{\lambda} = \frac{3x10^8}{8x10^{-7}} = 3.75\,x10^{14}\,Hz$

$\therefore$ Total bandwidth of the channels $= 1\%$ of $3.75 \times 10^{14} = 3.75 \times 10^{12}\,Hz$

(i) Number of channels for audio signals $= \frac{\text{Total bandwidth of channel}}{\text{Bandwidth of audio signal}} = \frac{3.75\,x10^{12}}{8x10^3} = 4.7 \text{ x } 10^8$

(ii) Number of channels for video TV signals $= \frac{3.75\,x10^{12}}{4.5\,x10^6} = 8.3 \text{ x } 10^5$

7.5. Frequency Modulation

When the frequency of carrier wave is changed in accordance with the intensity of the signal, it is called frequency modulation.

In frequency modulation, only the frequency of the carrier wave is changed in accordance with the signal. However, the amplitude of the modulated wave remains the same i.e. carrier wave amplitude. The frequency variations of carrier wave depend upon the instantaneous amplitude of the signal as shown in figure (iii). When the signal voltage is zero as a A, C, E and G, the carrier frequency is uncharged. When the signal approaches its positive peaks as a B and F, the carrier frequency is increased to maximum as shown by the closely spaced cycles. However, during the negative peaks of signal as at D, the carrier frequency is reduced to minimum as shown by the widely spaced cycles. The following points may be noted:

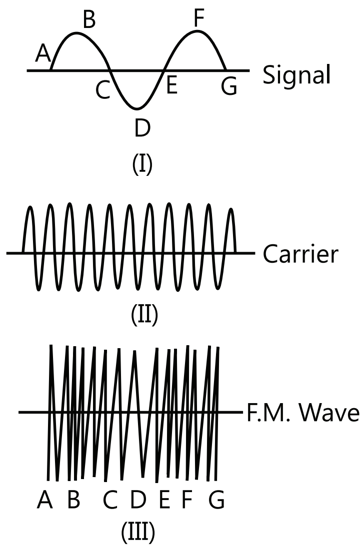


Figure 26.101

(a) All the signals having the same amplitude will change the carrier frequency by the same amount irrespective of their frequencies.

(b) All modulating signals of the same frequency, say 2 kHz, will change the carrier at the same rate of 2000 times per second irrespective of their individual amplitudes.

7.5.1 Advantages and Disadvantages of FM

Advantages

(a) It gives noiseless reception. As discussed before, noise is a form of amplitude variations and a FM receiver will reject such signals.

(b) The operating range is quite large.

(c) It gives high-fidelity reception,

(d) The efficiency of transmission is very high.

(e) Since FM has a large number of sidebands, it can be used for stereo sound transmission.

Disadvantages

(a) A much wider bandwidth is required by FM. The bandwidth required is 7 to 8 times as large as for AM.

(b) FM transmitting and receiving equipment's are complex, particularly for modulation and demodulation. Therefore, FM is more expensive than AM.

(c) FM reception is limited to line-of-sight.

PROBLEM-SOLVING TACTICS

1. For long distance transmission, we use electrical signals because they can be transmitted at very high speeds $\left(= 3\text{x}10^{8}\,\text{ms}^{-1}\right)$

2. The energy of a wave is directly proportional to its frequency. This permits modulated waves to carry the signals to long distances.

3. In amplitude modulation (AM), the amplitude of high frequency wave is changed in accordance with the intensity of the signal.

$$\text{Modulation factor, } m_a = \frac{\text{Amplitude change of carrier wave}}{\text{Normal carrier wave }\left(\text{unmodulated}\right)}$$

The value of m depends upon the amplitudes of carrier and signal.

4. In frequency modulation (FM), the frequency of high frequency wave (carrier) is changed in accordance with the intensity of the signal.

$$\text{Modulation Index, } m_f = \frac{\text{Maximum frequency deviation}}{\text{Modulating signal frequency}}$$

5. In AM, the power level of the carrier is not affected by the modulation index m.

6. In phase modulation, the phase angle of the high frequency wave (carrier) is changed in accordance with the strength of the modulating signal.

FORMULAE SHEET

1. In an n-type semiconductor, $n_e \cong N_d >> n_h$ where N_d is the number density of donor atoms.

In a p-type semiconductor, $n_h \cong N_a >> n_e$ where N_a is the number density of acceptor atoms.

In a doped semiconductor (n type or p-type). $n_e n_h >> n_i^2$

Where n_i is number density of intrinsic carriers?

2. In amplitude modulation (AM), the amplitude of high frequency wave is changed in accordance with the intensity of the signal.

$$\text{Modulation factor, } m_a = \frac{\text{Amplitude change of carrier wave}}{\text{Normal carrier wave (unmodulated)}}$$

3. In frequency modulation (FM), the frequency of high frequency wave (carrier) is changed in accordance with the intensity of the signal.

4. $$\text{Modulation Index, } m_f = \frac{\text{Maximum frequency deviation}}{\text{Modulating signal frequency}}$$

5. $$\sigma = e(n_e\mu_e + n_h\mu_h)$$

Where $\sigma = \frac{1}{\rho}$ is called conductivity of the material of semiconductor and μ_e, μ_h are electron and hole mobilities respectively.

6. The equation for diode current is $I = I_o\left(e^{cV/kT} - 1\right)$

Where I_o is called saturation current, V is positive for forward and negative for reverse bias, k is Boltzmann constant, T is temperature and $e = 1.6 \text{x} 10^{-19}\,C$.

7. Half wave Rectifier

Expression for output D.C. Voltage

Output d.c. voltage = Mean load current x load resistance i.e. $V_{d.c.} = I_{d.c.} R_L$. But

Where I_o is the maximum value of the secondary half wave current $\therefore\ V_{d.c.} = \frac{I_o}{\pi} \text{x } R_L$

8. Full-wave Rectifier

Expression for output D.C. Voltage

Output D.C. voltage = Mean load current x load resistance i.e. $V_{d.c.} = I_{d.c.} R_L$ but $I_{d.c.} = \frac{I_o}{\pi}$ where I_o is the maximum value of the secondary half wave current $\therefore\ V_{d.c.} = \frac{I_o}{\pi} \text{x } R_L$

Thus, output D.C. voltage in case of full wave rectifier is twice the output D.C. voltage in case of half wave rectifier.

9. $$\text{a.c. forward resistance, } r_f = \frac{\text{change in forward voltage across diode}}{\text{corresponding change in current through diode}}$$

10. Zener diode voltage regulation

Voltage drop across $R_S = E_{in} - E_o$; Current through $R_S.I = I_z + I_L$

Applying Ohm's law, we have $R_S = \frac{E_{in} - E_o}{I_z + I_L}$

Where R_S is the series resistance that absorbs voltage fluctuations, R_L is the load resistance across which output regulated voltage is desired, I_z is the zener current and I_L is the load current.

11. For a photodiode, $\therefore \; I_R = mE$ Where m = slope of the straight line

The quantity m is called the sensitivity of the photo-diode.

I_R is the reverse current and E is the illumination of the photo diode.

12. For a transistor, where $I_E = I_B + I_C$ is emitter current, I_B is base current and I_C is collector current.

13. Gains in Common-Base Amplifier

The various gains in a common-base amplifier are as follow:

(i) **ac Current Gain:** It is defined as the ratio of the change in the collector-current to the change in the emitter-current at a constant collector-to-base voltage, and is denoted by α.

Thus $\alpha_{(ac)} = \left(\frac{\Delta i_C}{\Delta i_B}\right)_{V_{CE}}$

The value of α is slightly less than 1 (actually, there is a little current loss).

(ii) **ac Voltage Gain:** It is defined as the ratio of the changes in the output voltage to the change in the input voltage, and is denoted by A.

Suppose on applying an ac input voltage signal, the emitter current changes by Δi and correspondingly the collector-current changes by Δi_C. If R_{in} and R_{out} be the resistances of the input and output circuits respectively, then

$$A_V = \frac{\Delta i_C \, x R_{in}}{\Delta i_E \, x R_{out}} = \frac{\Delta i_C}{\Delta i_E} x \frac{R_{in}}{R_{out}}$$

Now, $\Delta i_C / \Delta i_E$ is the ac current-gain and R_{in}/R_{out} is called the 'resistance gain'.

$\therefore$ Av = $\alpha \times$ Resistance gain

Since the resistance gain is quite high A_V is also high although α is slightly less than 1.

(iii) **ac Power Gain:** It is defined as the ratio of the change in the output power to the change in the input power.

Since power=current x voltage, we have ac power gain = ac current gain x ac voltage-gain = $\alpha^2 \times$ Resistance gain

14. Gain in Common emitter amplifier

(i) dc current Gains: It is defined as the ratio of the collector current to the base current, and is denoted by $\beta(dc) = \frac{i_C}{i_B}$

In a typical transistor, a small base-current $(\approx 10\mu A)$ produces a large collector-current $(\approx 500\mu A)$. Thus $\beta(dc) = \frac{500}{10} = 50$

(ii) ac Current Gain : It is defined as the ratio of the change in the collector-current to the change in the base-current at a constant collector to emitter voltage, and is denoted by

$\beta(ac)$. Thus $\beta(ac) = \left(\frac{\Delta i_C}{\Delta i_B}\right)_{V_{CE}}$

(iii) Voltage gain : Suppose, on applying an ac input voltage signal, the input base-current

Charges by Δi_B and correspondingly the output collector-current changes by Δi_C. If R_{in} and R_{out} be the resistance of the input and the output circuits respectively, then.

$$A_V = \frac{\Delta i_C \times R_{out}}{\Delta i_B \times R_{in}} = \frac{\Delta i_C}{\Delta i_B} \times \frac{R_{out}}{R_{in}} \quad \text{... (i)}$$

Now, $\Delta i_C / \Delta i_B$ is the ac current gain (ac) and R_{in}/R_{out} is the resistance gain

$$\therefore A_v = \beta(ac) \times \text{resistance gain} \quad \text{... (ii)}$$

Since $\beta(ac) >> \alpha(ac)$, the ac voltage gain in common-emitter amplifier is larger compared

To the common-base amplifier, although the resistance gain is smaller.

From equation (i) and (ii), it follows that $A_v = g_m x R_{out}$

(iv) ac Power gain : It is defined as the ratio of the change in the output power to the change in the input power.

Since power=current x voltage, we have ac power gain = ac current gain × ac voltage gain

$= \beta(ac) \times A_v = \beta(ac) \times \{\beta(ac) \times \text{resistance gain}\} = \beta^2(ac) \times \text{resistance gain}$

Since $\beta(ac) >> \alpha(ac)$, the ac power gain in common-emitter amplifier is extremely large

Compared to that in common-base amplifier.

15. The frequency of oscillations is given by $\boxed{v = \frac{1}{2\pi\sqrt{LC}}}$

16. Value of critical frequency in sky wave propagation is given by $f_C = 9(N_{max})^{1/2}$

Where N_{max} =Maximum electron density of ionosphere.

17. Maximum usable frequency, $MUF = \frac{f_C}{\cos\theta} = f_C \sec\theta$

Where θ = Angle between normal and direction of incident waves.

18. Modulation factor, $m_a = \frac{\text{Amplitude change of carrier wave}}{\text{Normal carrier wave (unmodulated)}} = \frac{E_S}{E_C} = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$

19. The instantaneous Voltage of AM wave is

$$= E_C \cos\omega_c t + \frac{m_a E_C}{2}\cos(\omega_c + \omega_s)t + \frac{m_a E_C}{2}\cos(\omega_c - \omega_s)t$$

20. In an AM wave, the bandwidth is form $(f_c - f_s)$ to $(f_c + f_s)$ i.e, $2f_s$.

21. Power In AM Wave

The power dissipated in any circuit is a function of the square of voltage across the circuit and the effective resistance of the circuit. Equation of AM wave reveals that it has three components of amplitude $E_C, m_a E_C/2$ and $m_a E_C/2$. Clearly, power output must be distributed among these components.

$$\text{Carrier power, } P_C = \frac{(E_C/\sqrt{2})}{R} = \frac{E^2c}{2R} \quad \text{... (i)}$$

$$\text{Total power of sidebands } P_C = \frac{(m_a E_C/2\sqrt{2})^2}{R} = \frac{(m_a E_C/2\sqrt{2})^2}{R} = \frac{m_a^2 E_C^2}{8R} + \frac{m_a^2 E_C^2}{8R} = \frac{m_a^2 E_C^2}{4R} \quad \text{.... (ii)}$$

$$\text{Total power of AM wave, } P_T = P_C + P_S = \frac{E_C^2}{2R} + \frac{m_a^2 E_C^2}{4R} = \frac{E_C^2}{2R}\left[1 + \frac{m_a^2}{2}\right]$$