Semiconductor Electronics

MATERIALS, DEVICES AND SIMPLE CIRCUITS

Semiconductors

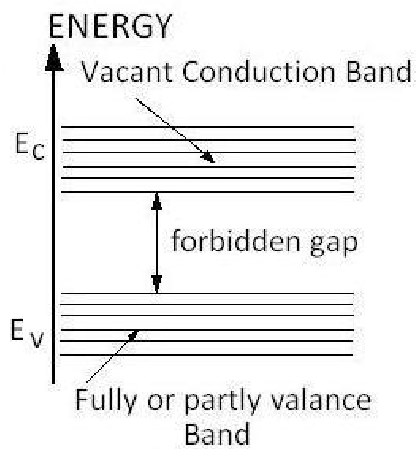
It has been observed that certain materials like germanium, silicon etc. have resistivity between good conductors like copper and insulators like glass. These materials are known as semiconductors. A material which has resistivity between conductors and insulators is known as semiconductor. The resistivity of a semiconductor lie approximately between 10^{-2} and 10^{4} Ω m at room temperature. The resistance of a semiconductor decreases with increase in temperature over a particular temperature range. This behavior is contrary to that of a metallic conductor for which the resistance increases with increase in temperature.

The elements that are classified as semiconductors are Si, Ge, In, etc. Germanium and silicon are most widely used as semiconductors

Energy band in solids

In the case of a single isolated atom, there are various discrete energy levels. In solids, the atoms are arranged in a systematic space lattice and each atom is influenced by neighboring atoms. The closeness of atoms results in the intermixing of electrons of neighboring atoms. Inside the crystal each electron has a unique position and no two electrons see exactly the same pattern of surrounding charges. Because of this, each electron will have a different energy level.

These different energy levels with continuous energy variation form what are called energy bands. The energy band which includes the energy levels of the valence electrons is called the valence band. The energy band above the valence band is called the conduction band. Energy difference between energy of conduction band and valance band is called band gap energy or forbidden energy gap With no external energy, all the valence electrons will reside in the valence band. If the lowest level in the conduction band happens to be lower than the highest level of the



valence band, the electrons from the valence band can move into the conduction band

Let us consider what happens in the case of Si or Ge crystal containing N atoms. For Si, the outermost orbit is the third orbit (n = 3), while for Ge it is the fourth orbit (n = 4).

The number of electrons in the outermost orbit is 4 (3s and 3p electrons for Si).

Hence, the total number of outer electrons in the crystal is 4N.

The maximum possible number of electrons in the outer orbit is 8 (2s + 6p electrons).

So, for the 4N valence electrons there are 8N available energy states. These 8N discrete energy levels can either form a continuous band or they may be grouped in different bands depending upon the distance between the atoms in the crystal

Conductors:

Normally the conduction band is empty. But when it overlaps on the valence band electrons can move freely into it. This is the case with metallic conductors.

Insulators:

In an insulator, the forbidden energy gap is very large. In general, the forbidden energy gap is more than 3eV and almost no electrons are available for conduction. Therefore, a very large amount of energy must be supplied to a valence electron to enable it to move to the conduction band. In the case of materials like glass, the valence band is completely filled at 0 K. The energy gap between valence band and conduction band is of the order of 10 eV. Even in the presence of high electric field, the electrons cannot move from valence band to conduction band. If the electron is supplied with high energy, it can jump across the forbidden gap. When the temperature is increased, some electrons will move to the conduction band. This is the reason, why certain materials, which are insulators at room temperature become conductors at high temperature. The resistivity of insulator approximately lies between 10^{11} and 10^{16} Ω m

Semiconductors:

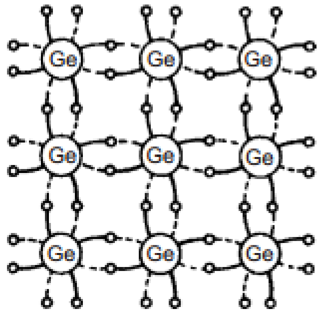
In semiconductors the forbidden gap is very small. Germanium and silicon are the best examples of semiconductors. The forbidden gap energy is of the order of 0.7eV for Ge and 1.1eV for Si. There are no electrons in the conduction band. The valence band is completely filled at 0 K. With a small amount of energy that is supplied, the electrons can easily jump from the valence band to the conduction band. For example, if the temperature is raised, the forbidden gap is decreased and some electrons are liberated into the conduction band. The conductivity of a semiconductor is of the order of 10^2 mho m^{-1}

INTRINSIC SEMICONDUCTOR

A semiconductor which is pure and contains no impurity is known as an intrinsic semiconductor. In an intrinsic semiconductor, the number of free electrons and holes are equal. Common examples of intrinsic semiconductors are pure germanium and silicon represent charge carriers at absolute zero temperature and at room temperature respectively.

(a) at absolute zero (b) at room temperature

The electrons in an intrinsic semiconductor, which move in to the conduction band at high temperatures are called as intrinsic carriers. In the valence band, a vacancy is created at the place where the electron was present, before it had moved in to the conduction band. This vacancy is called hole.
helps in understanding the creation of a hole. Consider the case of pure germanium crystal. It has four electrons in its outer or valence orbit. These electrons are known as valence electrons. When two atoms of gemanium are brought close to each other, a covalent bond is formed between the atoms. If some additional energy is received, one of the electrons contributing to a covalent bond breaks and it is free to move in the crystal lattice



While coming out of the bond, a hole is said to be created at its place, which is usually represented by a open circle. The hole behaves as an *apparent free particle* with effective positive charge. An electron from the neighboring atom can break the covalent bond and can occupy this hole, creating a hole at another place. Since an electron has a unit negative charge, the hole is associated with a unit positive charge. The importance of hole is that, it may serve as a carrier of electricity in the same manner as the free electron, but in the opposite direction.
In intrinsic semiconductors, the number of free electrons, ne is equal to the number of holes, n_h. That is
$n_e = n_h = n_i$

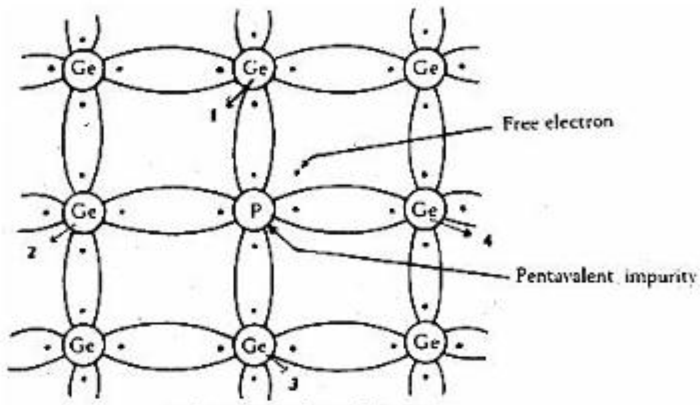
where ni is called intrinsic carrier concentration
Under the action of an electric field, these holes move towards negative potential giving the hole current, I_h. The total current, I is thus the sum of the electron current Ie and the hole current I_h: $I = I_e + I_h$
It may be noted that apart from the process of generation of conduction electrons and holes, a simultaneous process of recombination occurs in which the electrons recombine with the holes. At equilibrium, the rate of generation is equal to the rate of recombination of charge carriers. The recombination occurs due to an electron colliding with a hole.

EXTRINSIC SEMICONDUCTOR

Electrons and holes can be generated in a semiconductor crystal with heat energy or light energy. But in these cases, the conductivity remains very low. The efficient and convenient method of generating free electrons and holes is to add very small amount of selected impurity inside the crystal. The impurity to be added is of the order of 100 ppm (parts per million). The process of addition of a very small amount of impurity into an intrinsic semiconductor is called doping.
The impurity atoms are called dopants. The semiconductor containing impurity atoms is known as impure or doped or extrinsic semiconductor. There are three different methods of doping a semiconductor.
(i) The impurity atoms are added to the semiconductor in its molten state.
(ii) The pure semiconductor is bombarded by ions of impurity atoms.
(iii) When the semiconductor crystal containing the impurity atoms is heated, the impurity atoms diffuse into the hot crystal.
Usually, the doping material is either pentavalent atoms (bismuth, antimony, phosphorous, arsenic which have five valence electrons) or trivalent atoms (aluminium, gallium, indium, boron which have three valence electrons).
The pentavalent doping atom is known as donor atom, since it donates one electron to the conduction band of pure semiconductor.
The trivalent atom is called an acceptor atom, because it accepts one electron from the pure semiconductor atom.
Depending upon the type of impurity atoms added, an extrinsic semiconductor can be classified as N-type or P-type.

(A) N-TYPE SEMICONDUCTOR

When a small amount of pentavalent impurity such as arsenic is added to a pure germanium semiconductor crystal, the resulting crystal is called N-type semiconductor. Fig a shows the crystal structure obtained when pentavalent arsenic impurity is added with pure germanium crystal

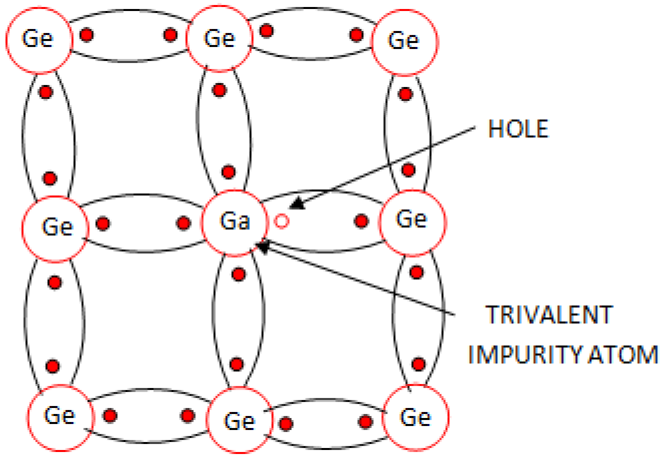


n–type semiconductor

The four valence electrons of arsenic atom form covalent bonds with electrons of neighboring four germanium atoms. The fifth electron of arsenic atom is loosely bound. This electron can move about almost as freely as an electron in a conductor and hence it will be the carrier of current. In the energy band picture, the energy state corresponding to the fifth valence electron is in the forbidden gap and lies slightly below the conduction band . This level is known as the donor level. When the fifth valence electron is transferred to the conduction band, the arsenic atom becomes positively charged immobile ion. Each impurity atom donates one free electron to the semiconductor. These impurity atoms are called donors. In N-type semiconductor material, the number of electrons increases, compared to the available number of charge carriers in the intrinsic semiconductor. This is because, the available larger number of electrons increases the rate of recombination of electrons with holes. Hence, in N-type semiconductor, free electrons are the majority charge carriers and holes are the minority charge carriers in an extrinsic therefore, known as n-type semiconductors. For n-type semiconductors, we have, $n_e >> n_h$

(b) P-type semiconductor

When a small amount of trivalent impurity (such as indium, boron or gallium) is added to a pure semiconductor crystal, the resulting semiconductor crystal is called P-type semiconductor. (Fig) shows the crystal structure obtained, when trivalent boron impurity is added with pure germanium crystal.



The three valence electrons of the boron atom form covalent bonds with valence electrons of three neighborhood germanium atoms. In the fourth covalent bond, only one valence electron is available from germanium atom and there is deficiency of one electron which is called as a hole. Hence for each boron atom added, one hole is created. Since the holes can accept electrons from neighborhood, the impurity is called acceptor. The hole, may be filled by the electron from a neighboring atom, creating a hole in that position from where the electron moves. This process continues and the hole moves about in a random manner due to thermal effects. Since the hole is associated with a positive charge moving from one position to another, this is called as P-type semiconductor. In the P-type semiconductor, the acceptor impurity produces an energy level just above the valence band. Since, the energy difference between acceptor energy level and the valence band is much smaller, electrons from the valence band can easily jump into the acceptor level by thermal agitation. In P-type semiconductors, holes are the majority charge carriers and free electrons are the minority charge carriers. Therefore, extrinsic semiconductors doped with trivalent impurity are called p-type semiconductors. For p-type semiconductors, the recombination process will reduce the number (n_i)of intrinsically generated electrons to n_e. We have, for p-type semiconductors $n_h >> n_e$ Note that the crystal maintains an overall charge neutrality as the charge of additional charge carriers is just equal and opposite to that of the ionised cores in the lattice. Conduction in p-type and n-type semiconductors The semiconductor's energy band structure is affected by doping. In the case of extrinsic semiconductors, additional energy states due to donor impurities and acceptor impurities also exist.

In the energy band diagram of n-type Si semiconductor, the donor energy level is slightly below the bottom of the conduction band and electrons from this level move into the conduction band with very small supply of energy. At room temperature, most of the donor atoms get ionised but very few (~10–12) atoms of Si get ionised. So the conduction band will have most electrons coming from the donor impurities, Similarly for p-type semiconductor, the acceptor energy level is slightly above the top of the valence band. With very small supply of energy an electron from the valence band can jump to the level and ionise the acceptor negatively. (Alternately, we can also say that with very small supply of energy the hole from level sinks down into the valence band. Electrons rise up and holes fall down when they gain external energy.) At room temperature, most of the acceptor atoms get ionised leaving holes in the valence band. Thus at room temperature the density of holes in the valence band is predominantly due to impurity in the extrinsic semiconductor. The electron and hole concentration in a semiconductor in thermal equilibrium is given by $n_e n_h = n_i^2$

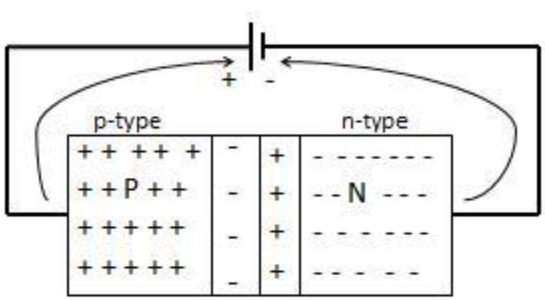
Solved Numerical

Q) Suppose a pure Si crystal has 5×10^{28} atoms m^{-3}. It is doped by 1 ppm concentration of pentavalent As. Calculate the number of electrons and holes. Given that $n_i = 1.5 \times 10^{16}$ m^{-3}.

Solution:
Note that thermally generated electrons ($n_i \sim 10^{16}$ m^{-3}) are negligibly small as compared to those produced by doping.
Therefore, $n_e \approx N_D$.
Since $n_e n_h = n_i^2$
The number of holes nh = $(2.25 \times 10^{32})/(5 \times 10^{22})$
$n_h \sim 4.5 \times 10^9$ m^{-3}

PN Junction diode

If one side of a single crystal of pure semiconductor (Ge or Si) is doped with acceptor impurity atoms and the other side is doped with donor impurity atoms, a PN junction is formed as shown in Fig



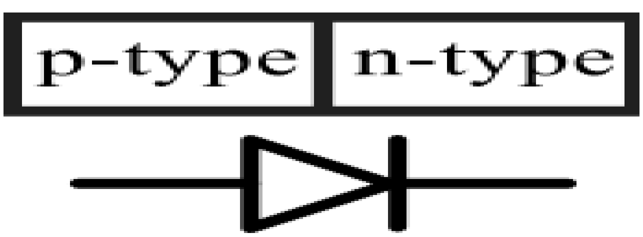
P region has a high concentration of holes and N region contains a large number of electrons. As soon as the junction is formed, free electrons and holes cross through the junction by the process of diffusion. During this process, the electrons crossing the junction from N-region into the P region, recombine with holes in the P-region very close to the junction.

Similarly holes crossing the junction from the P-region into the N-region, recombine with electrons in the N-region very close to the junction. Thus a region is formed, which does not have any mobile charges very close to the junction. This region is called depletion region. In this region, on the left side of the junction, the acceptor atoms become negative ions and on the right side of the junction, the donor atoms become positive ions

An electric field is set up, between the donor and acceptor ions in the depletion region. The potential at the N-side is higher than the potential at P-side. Therefore electrons in the N-side are prevented to go to the lower potential of P-side. Similarly, holes in the P-side find themselves at a lower potential and are prevented to cross to the N-side. Thus, there is a barrier at the junction which opposes the movement of the majority charge carriers. The difference of potential from one side of the barrier to the other side is called potential barrier. The potential barrier is approximately 0.7V for a silicon PN junction and 0.3V for a germanium PN junction. The distance from one side of the barrier to the other side is called the width of the barrier, which depends upon the nature of the material.

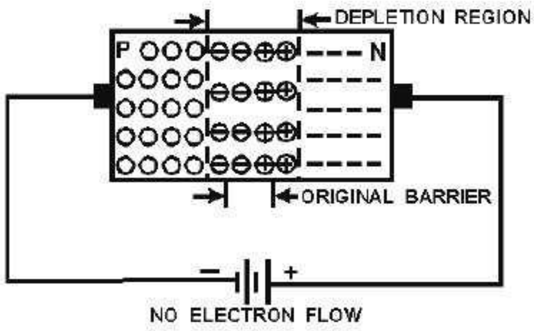
Symbol for a semiconductor diode

The diode symbol is shown in Fig The P-type and N-type regions are referred to as P–end and N–end respectively. The arrow on the diode points the direction of



Forward biased PN junction diode

When the positive terminal of the battery is connected to P-side and negative terminal to the N-side, so that the electric field across diode due to battery is in opposite direction to the electric field of barrier, then the PN junction diode is said to be forward biased.



When the PN junction is forward biased (Fig), the applied positive potential repels the holes in the P-region, and the applied negative potential repels the electrons in the Nregion, so the charges move towards the junction.
If the applied potential difference is more than the potential barrier, some holes and free electrons enter the depletion region.
Hence, the potential barrier as well as the width of the depletion region are reduced. The positive donor ions and negative acceptor ions within the depletion region regain electrons and holes respectively. As a result of this, the depletion region disappears and the potential barrier also disappears. Hence, under the action of the forward potential difference, the majority charge carriers flow across the junction in opposite direction and constitute current flow in the forward direction.

Forward bias characteristics

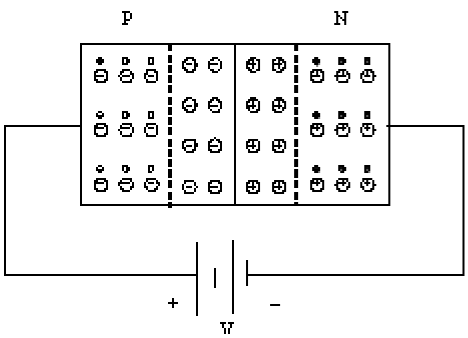
The circuit for the study of forward bias characteristics of PN junction diode is shown in Fig The voltage between P–end and N–end is increased from zero in suitable equal steps and the corresponding currents are noted down. (Fig) shows the forward bias characteristic curve of the diode. Voltage is the independent variable. Therefore, it is plotted along X– axis. Since, current is the dependent variable, it is plotted against Y–axis. From the characteristic curve, the following conclusions can be made.

(i) The forward characteristic is not a straight line. Hence the ratio V/I is not a constant (i.e) the diode does not obey Ohm's law. This implies that the semiconductor diode is a non-linear conductor of electricity.

(ii) It can be seen from the characteristic curve that initially, the current is very small. This is because, the diode will start conducting, only when the external voltage overcomes the barrier potential (0.7V for silicon diode). As the voltage is increased to 0.7 V, large number of free electrons and holes start crossing the junction. Above 0.7V, the current increases rapidly. The voltage at which the current starts to increase rapidly is known as cut-in voltage or knee voltage of the diode

Reverse biased PN junction diode

When the positive terminal of the battery is connected to the difference is in the same direction as that of barrier potential, the junction is said to be reverse biased.



When the PN junction is reverse electrons in the N region and holes in the Pregion are attracted away from the junction N-side and negative terminal to the P-side, so that the applied potential Because of this, the number of negative ions in the P-region and positive ions in the N-region increases. Hence the depletion region becomes wider and the potential barrier is increased.

Since the depletion region does not contain majority charge carriers, it acts like an insulator. Therefore, no current should flow in the external circuit. But, in practice, a very small current of the order of few microamperes flows in the reverse direction. This is due to the minority carriers flowing in the opposite direction. This reverse current is small, because the number of minority carriers in both regions is very small. Since the major source of minority carriers is, thermally broken covalent bonds, the reverse current mainly depends on the junction temperature.

Reverse bias characteristics The circuit for reverse bias characteristics of PN junction diode.

The voltage is increased from zero in suitable steps. For each voltage, the corresponding current readings are noted down. reverse bias characteristic curve of the diode. From the characteristic curve, it can be concluded that, as voltage is increased from zero, reverse current (in the order of microamperes) increases and reaches the maximum value at a small value of the reverse voltage. When the voltage is further increased, the current is almost independent of the reverse voltage upto a certain critical value. This reverse current is known as the reverse saturation current or leakage current. This current is due to the minority charge carriers, which depends on junction temperature.

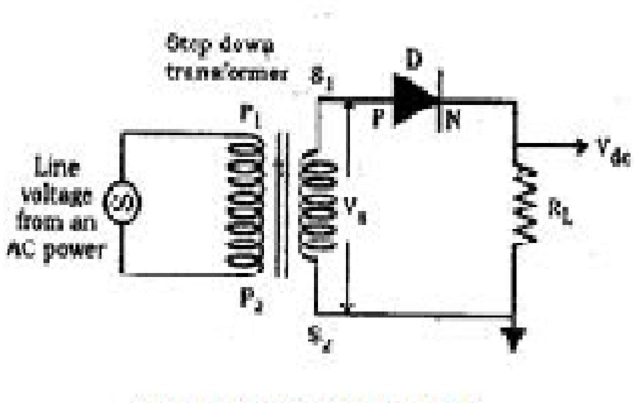
Avalanche breakdown :

When both sides of the PN junction are lightly doped and the depletion layer becomes large, avalanche breakdown takes place. In this case, the electric field across the depletion layer is not so strong. The minority carriers accelerated by the field, collide with the semiconductor atoms in the crystal.

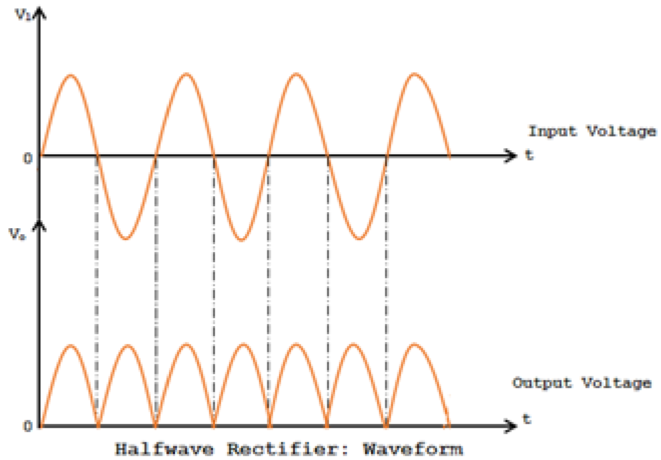
Because of this collision with valence electrons, covalent bonds are broken and electron hole pairs are generated. These charge carriers, so produced acquire energy from the applied potential and in turn produce more and more carriers. This cumulative process is called avalanche multiplication and the breakdown is called avalanche breakdown.

Half wave rectifier

A circuit which rectifies half of the a.c wave is called half wave rectifier. Fig shows the circuit for half wave rectification. The a.c. voltage (Vs) to be rectified is obtained across the secondary ends S_1 S_2 of the transformer.

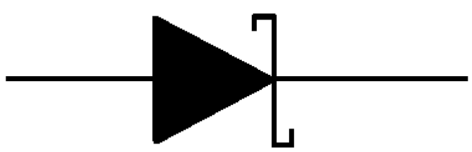


The P-end of the diode D is connected to S_1 of the secondary coil of the transformer. The N-end of the diode is connected to the other end S_2 of the secondary coil of the transformer, through a load resistance RL. The rectified output voltage Vdc appears across the load resistance RL. During the positive half cycle of the input a.c. voltage Vs, S_1 will be positive and the diode is forward biased and hence it conducts. Therefore, current flows through the circuit and there is a voltage drop across RL. This gives the output voltage as shown in Fig. During the negative half cycle of the input a.c. voltage (Vs), S_1 will be negative and the diode D is reverse biased. Hence the diode does not conduct. No current flows through the circuit and the voltage drop across RL will be zero. Hence no output voltage is obtained. Thus corresponding to an alternating input signal, unidirectional pulsating output is obtained. The ratio of d.c. power output to the a.c. power input is known as rectifier efficiency. The efficiency of half wave rectifier is approximately 40.6%.

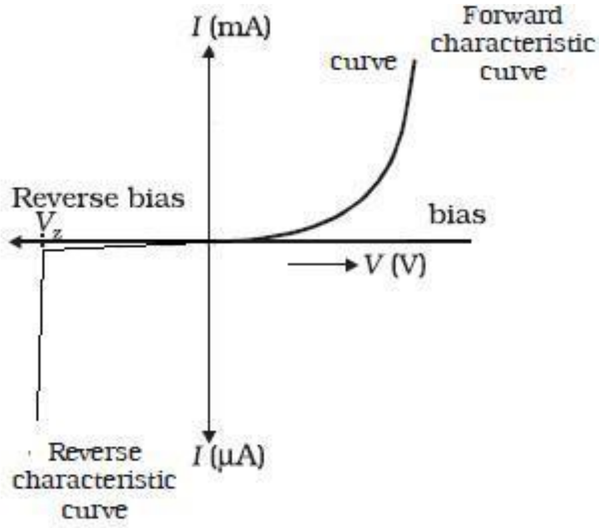


Zener diode

It is a special purpose semiconductor diode, named after its inventor C. Zener. It is designed to operate under reverse bias in the breakdown region and used as a voltage regulator. The symbol for Zener diode is shown in (a).



Zener diode is fabricated by heavily doping both p-, and n- sides of the junction. Due to this, depletion region formed is very thin ($<10^{-6}$ m) and the electric field of the junction is extremely high (~5×106 V/m) even for a small reverse bias voltage of about 5V. The I-V characteristics of a Zener diode is shown in Fig



It is seen that when the applied reverse bias voltage(V) reaches the breakdown voltage (V_z) of the Zener diode, there is a large change in the current. Note that after the breakdown voltage V_z, a large change in the current can be produced by almost insignificant change in the reverse bias voltage. In other words, Zener voltage remains constant, even though current through the Zener diode varies over a wide range. This property of the Zener diode is used for regulating supply voltages so that they are constant.

Zener breakdown :

We know that reverse current is due to the flow of electrons (minority carriers) from p → n and holes from n → p.

When both sides of the PN junction are heavily doped, consequently the depletion layer is narrow. Zener breakdown takes place in such a thin narrow junction As the reverse bias voltage is increased, the electric field at the junction becomes significant. When the reverse bias voltage V = V_z, then the electric field strength is high enough to pull valence electrons from the host atoms on the p-side which are accelerated These electrons account for high current observed at the breakdown. The emission of electrons from the host atoms due to the high electric field is known as internal field emission or field ionisation. The electric field required for field ionisation is of the order of 10^6 V/m.

Light emitting diode

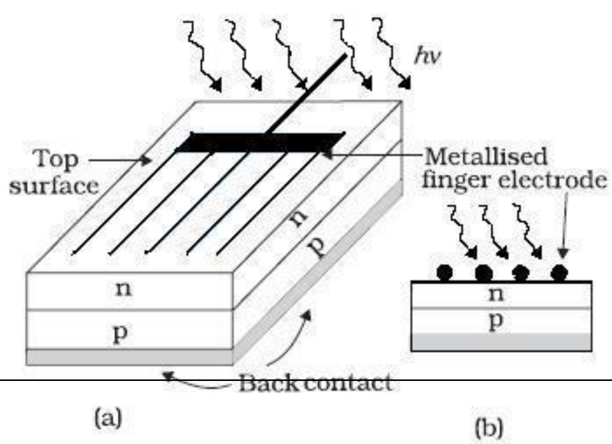
It is a heavily doped p-n junction which under forward bias emits spontaneous radiation. The diode is encapsulated with a transparent cover so that emitted light can come out. When the diode is forward biased, electrons are sent from n → p and holes are sent from p → n.

At the junction boundary the concentration of minority carriers increases compared to the equilibrium concentration (i.e., when there is no bias). Thus at the junction boundary on either side of the junction, excess minority carriers are there which recombine with majority carriers near the junction. On recombination, the energy is released in the form of photons. Photons with energy equal to or slightly less than the band gap are emitted. When the forward current of the diode is small, the intensity of light emitted is small. As the forward current increases, intensity of light increases and reaches a maximum. Further increase in the forward current results in decrease of light intensity .
LEDs are biased such that the light emitting efficiency is maximum .
The V-I characteristics of a LED is similar to that of a Si junction diode. But the threshold voltages are much higher and slightly different for each colour.
The reverse breakdown voltages of LEDs are very low, typically around 5V. So care should be taken that high reverse voltages do not appear across them .
LEDs that can emit red, yellow, orange, green and blue light are commercially available. The semiconductor used for fabrication of visible LEDs must at least have a band gap of 1.8 eV (spectral range of visible light is from about 0.4 µm to 0.7 µm, i.e., from about 3 eV to 1.8 eV).
The compound semiconductor Gallium Arsenide – Phosphide ($GaAs_{1-x}P_x$) is used for making LEDs of different colours.
$GaAs_{0.6}P_{0.4}$ (E_g ~ 1.9 eV) is used for red LED.
GaAs (E_g ~ 1.4 eV) is used for making infrared LED.
These LEDs find extensive use in remote controls, burglar alarm systems, optical communication, etc.
Extensive research is being done for developing white LEDs which can replace incandescent lamps.
LEDs have the following advantages over conventional incandescent low power lamps:

(i) Low operational voltage and less power.
(ii) Fast action and no warm-up time required.
(iii) The bandwidth of emitted light is 100 Å to 500 Å or in other words it is nearly (but not exactly) monochromatic.
(iv) Long life and ruggedness.
(v) Fast on-off switching capability.

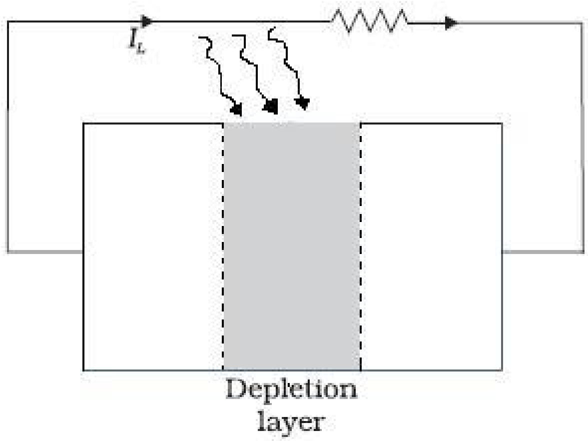
Solar cell

A solar cell is basically a p-n junction which generates emf when solar radiation falls on the p-n junction. It works on the same principle (photovoltaic effect) as the photodiode, except that no external bias is applied and the junction



area is kept much larger for solar radiation to be incident because we are interested in more power. A simple p-n junction solar cell is shown in figure A p-Si wafer of about 300 µm is taken over which a thin layer (~0.3 µm) of n-Si is grown on

one-side by diffusion process. The other side of p-Si is coated with a metal (back contact). On the top of n-Si layer, metal finger electrode (or metallic grid) is deposited. This acts as a front contact. The metallic grid occupies only a very small fraction of the cell area (<15%)so that light can be incident on the cell from the top.The generation of emf by a solar cell, when light falls on, it is due to the following three basic processes: generation, separation and collection—

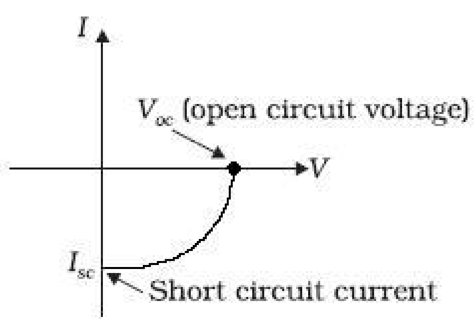


(i) generation of e-h pairs due to light (with hν > Eg) close to the junction;

(ii) separation of electrons and holes due to electric field of the depletion region. Electrons are swept to n-side and holes to p-side;

(iii) the electrons reaching the n-side are collected by the front contact and holes reaching p-side are collected by the back contact. Thus p-side becomes positive and n-side becomes negative giving rise to photovoltage.

When an external load is connected as shown in the Fig. a photocurrent IL flows through the load. A typical I-V characteristics of a solar cell is shown in the Fig.



(b)

Note that the I – V characteristics of solar cell is drawn in the fourth quadrant of the coordinate axes. This is because a solar cell does not draw current but supplies the same to the load. Semiconductors with band gap close to 1.5 eV are ideal materials for solar cell fabrication. Solar cells are made with semiconductors like Si (Eg = 1.1 eV), GaAs (Eg = 1.43 eV), CdTe (Eg = 1.45 eV), CuInSe2 (Eg = 1.04eV), etc. The important criteria for the selection of a material for solar cell fabrication are

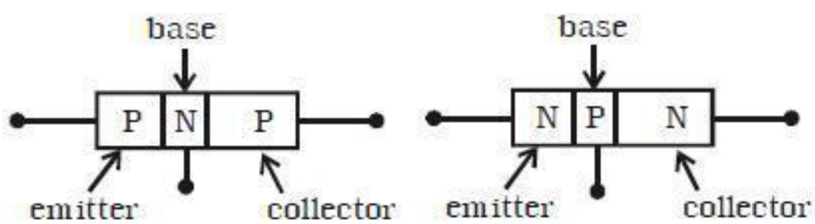
(i) Band gap (~1.0 to 1.8 eV),
(ii) High optical absorption (~104 cm–1), electrical conductivity,
(iii) Availability of the raw material, and
(iv) Cost. Note that sunlight is not always required for a solar cell. Any light with photon energies greater than the band gap will do. Solar cells are used to power electronic devices in satellites and space vehicles and also as power supply to some calculators.

(v) **Junction transistor**

A junction transistor is a solid state device. It consists of silicon or germanium crystal containing two PN junctions. The two PN junctions are formed between the three layers. These are called base, emitter and collector.

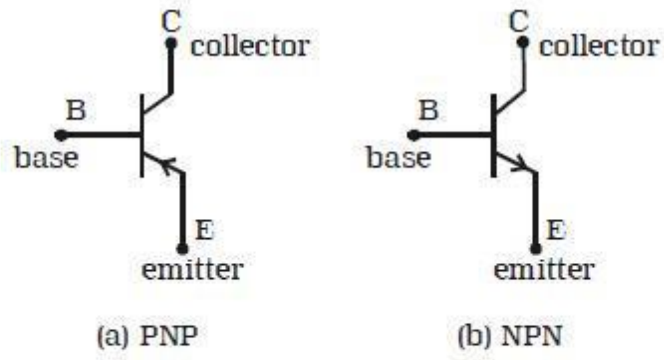
(i) Base (B) layer : It is a very thin layer, the thickness is about 25 microns. It is the central region of the transistor.

(ii) Emitter (E) and Collector (C) layers : The two layers on the opposite sides of B layer are emitter and collector layers. They are of the same type of the semiconductor. An ohmic contact is made to each of these layers. The junction between emitter and base is called emitter junction. The junction between collector and base is called collector junction. In a transistor, the emitter region is heavily doped, since emitter has to supply majority carriers. The base is lightly doped. The collector region is lightly doped. Since it has to accept majority charge carriers, it is physically larger in size. Hence, emitter and collector cannot be interchanged. The construction of PNP and NPN transistors are shown in Fig a and Fig b respectively.



(a) *Construction of PNP transistor* (b) *Construction of NPN transistor*

For a transistor to work, the biasing to be given are as follows :
(i) The emitter-base junction is forward biased, so that majority charge carriers are repelled from the emitter and the junction offers very low resistance to the current.
(ii) The collector-base junction is reverse biased, so that it attracts majority charge carriers and this junction offers a high resistance to the current. Transistor circuit symbols
The circuit symbols for a PNP and NPN transistors are shown in Fig



The arrow on the emitter lead pointing towards the base represents a PNP transistor When the emitter-base junction of a PNP transistor is forward biased, the direction of the conventional current flow is from emitter to base.

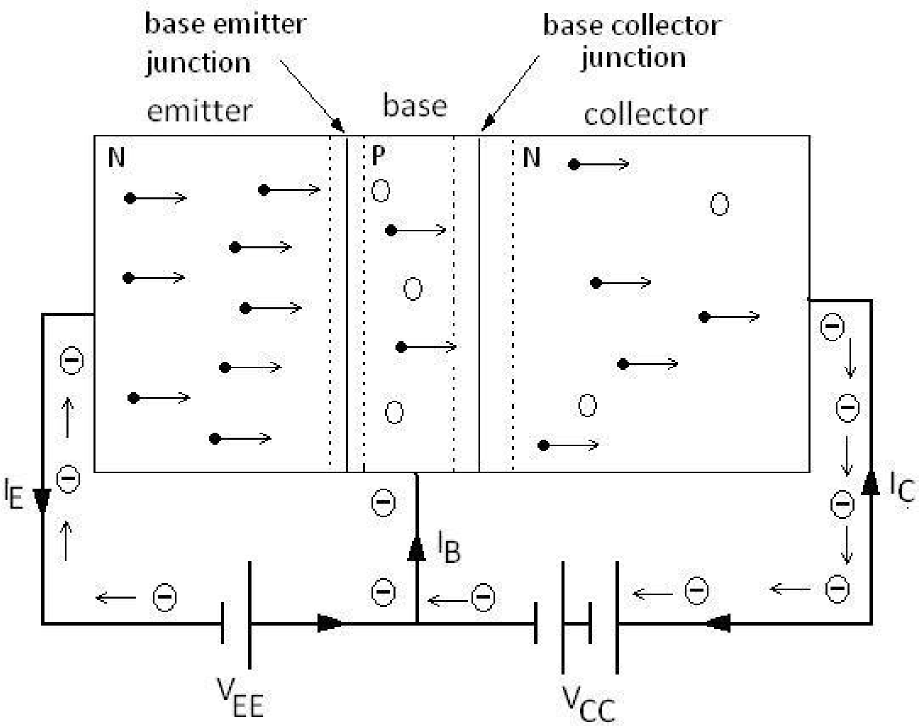
NPN transistor is represented by arrow on the emitter lead pointing away from the base. When the emitter base junction of a NPN transistor is forward biased, the direction of the conventional current is from base to emitter.

Working of a NPN transistor

A NPN transistor is like two PN junction diodes, which are placed back-to-back. At each junction, there is a depletion region which gives rise to a potential barrier. The external biasing of the junction is provided by the batteries V_{EE} and V_{CC} as shown in Fig. The emitter base junction is forward biased and the collector base junction is reverse biased.

Since the emitter-base junction is forward biased, a large number of electrons cross the junction and enters the base constitutes current I_E. The base width is small and has fewer concentration of impurity as a result only 5% of electrons entering base recombine with holes.

The rest of the electrons enters collector region due to influence of the battery V_{CC}. For every electron entering the collector one electron flows in the external circuit and constitutes the collector current I_C. Similarly for every electron combing with the hole in the base section, there is one electron which gets attracted by VEE and flows as base current IB in external circuit Applying Kirchoff's current law to the circuit, the emitter current is the sum of collector current and base current. $I_E = I_C + I_B$



This equation is the fundamental relation between the currents in a transistor circuit. This equation is true regardless of transistor type or transistor configuration. The action of PNP transistor is similar to that of NPN transistor.

Solved Numerical

Q) In NPN transistor about 1010 electrons enter the emitter in 1μs when it is connected to a battery. About 2% electrons recombine with the holes in the base. Calculate the values of I_E , I_B, I_C, α_{dc}, and β_{dc} (e = 1.6 ×10^{-19} C)

Solution:
As per the definition of current

$$I_E = Q/t = ne/t$$

$$I_E = 10^{10} \times 1.6 \times 10^{-19}/10^{-6} = 1600\mu A$$

2% of the total current entering the base from the emitter recombine with the holes which constitutes the base current IB. The rest of the 98% electrons reaches the collector and constitutes the collector current

I_B = 0.021, I_E = 0.02 × 1600 = 32 μA
I_C = 0.98, I_E = 0.98 × 1600 =1568 μA

$$\alpha dc = I_C/I_E = 1568 \times 10^{-6}/1600 \times 10^{-6} = 0.98$$

$$\beta dc = I_C/I_B = 1568 \times 10^{-6}/32 \times 10^{-6} = 49$$

Current amplification factors α and β and the relation between them

The current amplification factor or current gain of a transistor is the ratio of output current to the input current. If the transistor is connected in common base mode, the current gain α

$$\alpha = I_C/I_E$$

and if the transistor is connected in common emitter mode, the current gain β

$$\beta = I_C/I_B$$

Since 95% of the injected electrons reach the collector, the collector current is almost equal to the emitter current. Almost all transistors have α , in the range 0.95 to 0.99 We know that

$$\alpha = I_C/I_E$$

And $I_E = I_C + {}_{IB}$ Thus

$$\alpha = I_C/I_C + I_B$$

$$1/\alpha = I_C + I_B/I_C = 1 + I_B/I_C$$

$$1/\alpha - 1 = I_B/I_C$$

$$1/\alpha - 1 = 1/\beta$$

$$\beta = \alpha/1 - \alpha$$

Usually β lies between 50 and 300. Some transistors have β as high as 1000. Similarly we can prove

$$\alpha = \beta/1 + \beta$$

Cut off region :

There is very small collector current in the transistor, even when the base current is zero ($I_B = 0$). In the output characteristics, the region below the curve for $I_B = 0$ is called cut off region. Below the cut off region, the transistor does not function. In this region both baseemitter region and base-collector region are reverse biased.

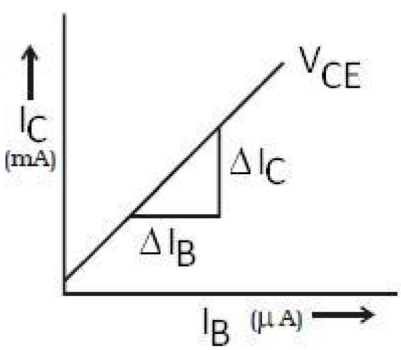
Active region : The central region of the curves is called active region. In the active region, the curves are uniform. In this region, E-B junction is forward biased and C-B junction is reverse biased. The output impedance r_o is defined as the ratio of variation in the collector emitter voltage to the corresponding variation in the collector current at a constant base current in the active region of the transistor characteristic curves. output impedance, r_o

$$\boldsymbol{r_o} = (\Delta \boldsymbol{V_{CE}}/\Delta \boldsymbol{I_C})\boldsymbol{I_B}$$

The output impedance of a transistor in CE mode is low. Its value can be found out from the input characteristic curve. Normally its value is found between 50kΩ to 100kΩ

Transfer characteristics

The transfer characteristic curve is drawn between I_C and I_B, when V_{CE} is kept constant at a particular value. The base current I_B is increased in suitable steps and the collector current I_C is noted down for each value of IB. The transfer characteristic curve is shown in Fig.



The current gain is defined as the ratio of a small change in the collector current to the corresponding change in the base current at a constant VCE. current gain, β

$$\beta = (\Delta I_C/\Delta I_B)V_{CE}$$

The common emitter configuration has high input impedance, low output impedance and higher current gain when compared with common base configuration. Taking the ratio of β and ri for ac circuit

$$\beta/r_i = \Delta I_C/\Delta I_B \,/\, \Delta V_{BE}/\Delta I_B \,/ = \Delta I_C/\Delta V_{BE}$$

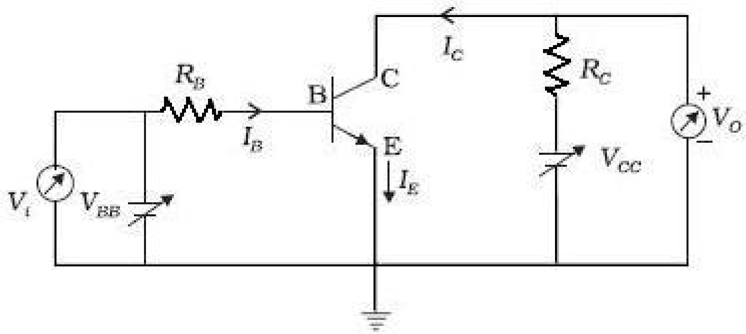
Ratio of the change in the current in the output circuit (ΔIC) to the change in the input voltage (ΔVBE) is known as the trans-conductance gm its unit is mho

$$g_m = \Delta I_C/\Delta V_{BE} = \beta/r_i$$

Transistor as switch

In an ideal ON/OFF switch, when it is OFF the current is not flowing in the circuit because switch offers infinite resistance.

When switch is in On condition, maximum current flows because its resistance is zero. We can prepare such an electronic switch by using the resistor. The operating point switch from cutoff to saturation along the load line.



We shall try to understand the operation of the transistor as a switch by analyzing the behavior of the base-biased transistor in CE configuration as shown in Fig.

Applying Kirchhoff's voltage rule to the input and output sides of this circuit, we get $V_{BB} = I_B R_B + V_{BE}$ and $V_{CE} = V_{CC} – I_C R_C$. We shall treat VBB as the dc input voltage V_i and V_{CE} as the dc output voltage V_O. So, we have

$V_i = I_B R_B + V_{BE}$ ----eq(1)and

$V_o = V_{CC} – I_C R_C$. ----eq(2)

Let us see how Vo changes as Vi increases from zero onwards.

(i) When input voltage Vi is zero or less than 0.6V for Si transistor (transistor cut in voltage), the base current IB will be zero. Hence the collector current will also zero IC = 0 From eq(2) VO = VCC In this situation resistance of output circuit is very large. Hence the current is not flowing through it. This is the 'OFF' or "cut off" condition of the transistor.

(ii) When the input voltage will be Vi = VCC , the base current is maximum, hence the collector current is maximum. The voltage drop (ICRL) across the load resistance RL will be approximately VCC . According to eq(2) VO = 0 In this condition resistance of the output circuit of the transistor is very small to the effect that maximum current is flowing through it. This is called the "ON" condition or saturation condition of the transistor.

Alternatively, we can say that a low input to the transistor gives a high output and a high input gives a low output. The switching circuits are designed in such a way that the transistor does not remain in active state. This circuit is used to make 'NOT' gate in the digital electronics

Transistor oscillators

An oscillator may be defined as an electronic circuit which converts energy from a d.c. source into a periodically varying output.

Oscillators are classified according to the output voltage, into two types viz. sinusoidal and non-sinusoidal oscillators.

If the output voltage is a sine wave function of time, the oscillator is said to be sinusoidal oscillator. If the oscillator generates non-sinusoidal waveform, such as square, rectangular waves, then it is called as non-sinusoidal oscillator (multivibrator).

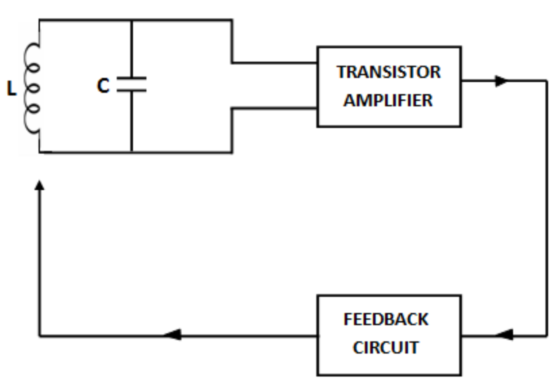
The oscillators can be classified according to the range of frequency as audio-frequency (AF) and radio-frequency (RF) oscillators.

Sinusoidal oscillators may be any one of the following three types:

(i) LC oscillators
(ii) RC oscillators
(iii) Crystal oscillators

Essentials of LC oscillator:

Fig shows the block diagram of an oscillator.

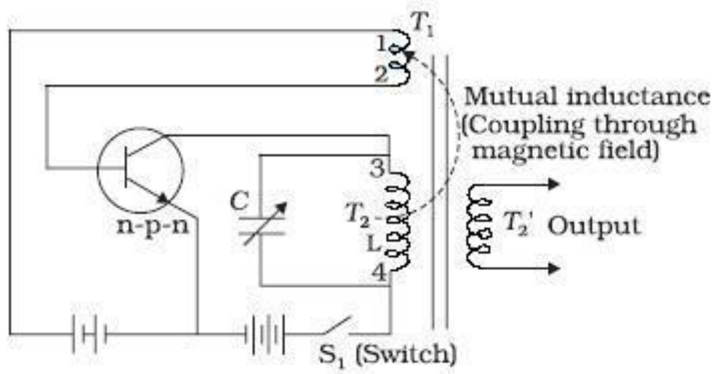


It's essential components are (i) tank circuit, ii) amplifier and (iii) feedback circuit.

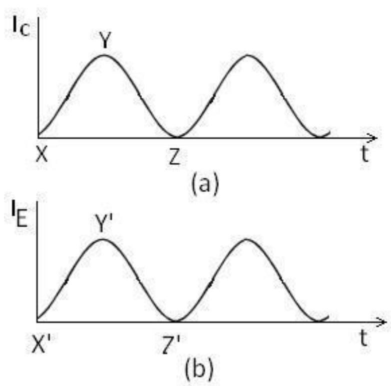
(i) Tank circuit : It consists of inductance coil (L) connected in parallel with capacitor (C). The frequency of oscillations in the circuit depends upon the values of inductance coil and capacitance of the capacitor.

(ii) Amplifier : The transistor amplifier receives d.c. power from the battery and changes it into a.c. power for supplying to the tank circuit.

(iii) Feedback circuit : It provides positive feedback (i.e.) this circuit transfers a part of output energy to LC circuit in proper phase, to maintain the oscillations



Suppose switch S_1 is put on to apply proper bias for the first time. Obviously, a surge of collector current flows in the transistor. This current flows through the coil T_2 where terminals are numbered 3 and 4 [Fig.]. This current does not reach full amplitude instantaneously but increases from X to Y, as shown in [Fig.] The inductive coupling between coil T_2 and coil T_1 now causes a current to flow in the emitter circuit (note that this actually is the 'feedback' from input to output). As a result of this positive feedback, this current (in T_1; emitter current) also increases from X′ to Y′ [Fig.]



The current in T_2 (collector current) connected in the collector circuit acquires the value Y when the transistor becomes saturated. This means that maximum collector current is flowing and can increase no further. Since there is no further change in collector current, the magnetic field around T_2 ceases to grow. As soon as the field becomes static, there will be no further feedback from T_2 to T_1. Without continued feedback, the emitter current begins to fall. Consequently, collector current decreases from Y towards Z [Fig]. However, a decrease of collector current causes the magnetic field to decay around the coil T_2. Thus, T_1 is now seeing a decaying field in T_2 (opposite from what it saw when the field was growing at the initial start operation). This causes a further decrease in the emitter current till it reaches Z' when the transistor is *cut-off*.
This means that both IE and IC cease to flow. Therefore, the transistor has reverted back to its original state (when the power was first switched on).

The whole process now repeats itself. That is, the transistor is driven to saturation, then to cut-off, and then back to saturation. The time for change from saturation to cut-off and back is determined by the constants of the tank circuit or tuned circuit (inductance L of coil T_2 and C connected in parallel to it). The resonance frequency (ν) of this tuned circuit determines the frequency at which the oscillator will oscillate.

$$\nu = 1/2\pi\sqrt{LC}$$

In the circuit of the tank or tuned circuit is connected in the collector side. Hence, it is known as *tuned collector oscillator.*

If the tuned circuit is on the base side, it will be known as tuned base oscillator. There are many other types of tank circuits (say RC) or feedback circuits giving different types of giving different types of oscillators like Colpitt's oscillator, Hartley oscillator, RCoscillator

Solved Numerical

Q) In transistor oscillator circuit an output signal of 1MHz frequency is obtained. The value of capacitance C = 100pF. What should be the value of the capacitor is a signal of 2MHz frequency is to be obtained

Solution:

C_1 = 100pF = 10 ×10^{-13} F, f_1 = 1MHz = 10^6 Hz

f_2 = 2 MHz = 2×10^6 Hz, C_2 = ?

$$f_1 = 1/2\pi\sqrt{LC_1}$$

And

$$f_2 = 1/2\pi\sqrt{LC_2}$$

$$f_1/f_2 = \sqrt{C_2/C_1}$$

$$C_2 = ({}_1/'f_2)^2 \times C_1$$

$$C_2 = (1/2)^2 \times 100 \times 10^{-12} = 25pF$$

Digital electronics

Digital Electronics involves circuits and systems in which there are only two possible states which are represented by voltage levels. Other circuit conditions such as current levels open or closed switches can also represent the two states. *Analog signal*

Semiconductor Electronics

The signal current or voltage is in the form of continuous, time varying voltage or current (sinusoidal). Such signals are called continuous or analog signals. A typical analog signal

Digital signal and logic levels

A digital signal (pulse). It has two discrete levels, 'High' and 'Low'. In most cases, the more positive of the two levels is called HIGH and is also referred to as logic 1. The other level becomes low and also called logic 0. This method of using more positive voltage level as logic 1 is called a positive logic system. A voltage 5V refers to logic 1 and 0 V refers to logic 0. On the other hand, in a negative logic system, the more negative of the two discrete levels is taken as logic 1 and the other level as logic 0. Both positive and negative logic are used in digital systems. But, positive logic is more common of logic gates. Hence we consider only positive logic for studying the operation of logic gates.

Logic gates

Circuits which are used to process digital signals are called logic gates. They are binary in nature. Gate is a digital circuit with one or more inputs but with only one output. The output appears only for certain combination of input logic levels. Logic gates are the basic building blocks from which most of the digital systems are built up. The numbers 0 and 1 represent the two possible states of a logic circuit. The two states can also be referred to as 'ON and OFF' or 'HIGH and LOW' or 'TRUE and FALSE'. *Basic logic gates using discrete components*

The basic elements that make up a digital system are 'OR', 'AND' and 'NOT' gates. These three gates are called basic logic gates. All the possible inputs and outputs of a logic circuit are represented in a table called TRUTH TABLE. The function of the basic gates are explained below with circuits and truth tables.

Boolean algebra

Boolean algebra, named after a mathematician George Boole is the algebra of logic, which is applied to the operation of computer devices. The rules of this algebra is simple, speed and accurate. This algebra is helpful in simplifying the complicated logical expression. Laws and theorems of Boolean algebra

The fundamental laws of Boolean algebra are given below which are necessary for manipulating different Boolean expressions.

Basic laws :

Commutative laws : A + B = B + A ; AB = BA

Associative Laws: A + (B + C) = (A + B) + C ; A (BC) = (AB) C

Distributive law: A (B+C) = AB + AC

Special theorems:

A + AB = A

(A + B) (A + C) = A + BC

A (A + B) = A

A + A B = A + B

A (A + B) = AB

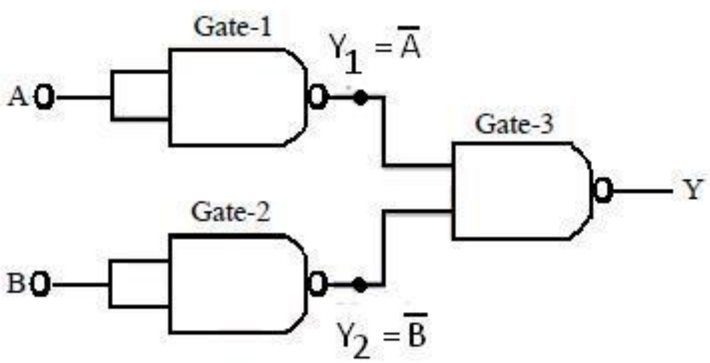
(A + B) (A + C) = AC + A B

AB + A C = (A + C) (A + B)

Theorems involving a single variable can be proved by considering every possible value of the variable.

Solved Numerical

Q) Show that the circuit drawn in figure comprising of three NAND gates behave as an OR gate



Solution:

Gate 1 and Gate 2 have identical inputs. Hence both behaves as NOT gate

Hence $y_1 = \bar{A}$ and $y_2 = \bar{B}$

Gate 3 is NAND hence output

$$y = \bar{y}_1\bar{y}_2$$

$$y = \bar{A}\,\bar{B}$$

By DeMorgan's theorem y = AB

Hence the above circuit behaves as OR gate and can be verified using truth table