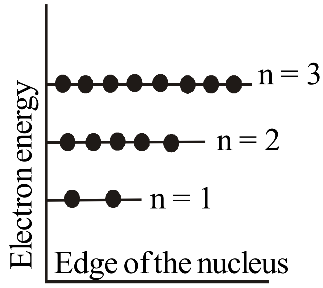
SOLIDS AND SEMICONDUCTING DEVICES

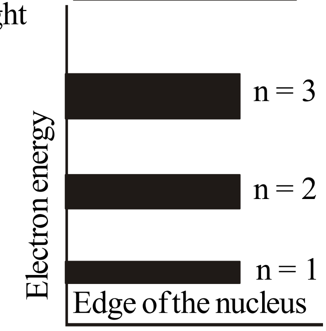


In an atom, electrons are distributed in various orbits. For an isolated atom, each orbit has discrete energy values. This is called *energy state or energy level*.

These energy levels are represented as follows:-

Energy Band

(The above mentioned case is for an isolated atom) When atoms are brought closer as in a solid, the electrons interact with each other, and also with the neighboring atomic cores. Hence the individual energy levels of electrons are disturbed. As a result, electrons (especially outermost electrons) can have a number of discrete but closely spaced energy states within a certain range rather than a single energy state. This range of energy states possessed by an electron is called *energy band*.



Valance Band

The range of energy states possessed by valance electrons is called valance band.

Conduction Band

The range of energy states possessed by free electrons which causes conduction is called conduction band.

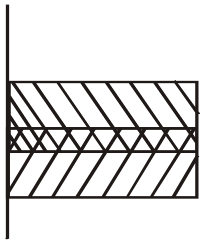
Forbidden band or Forbidden gap

The valance band and the conduction band are separated by energy gap called forbidden gap or forbidden band. The significance of energy gap is that electrons in a valance band can be excited into conduction band by supplying an amount of energy equal to band gap energy.

Depending on energy band structure solids are categorised into three: (i) conductors (ii) insulators (iii) semi conductors.

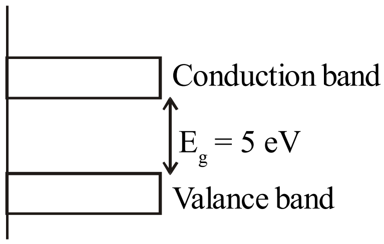
(i) **Conductors**

In conductors, there are a large number of free electrons (available for conduction). In conductors, valance band and conduction band overlap each other i.e. no energy is required to move electrons from valance band to conduction band.



(ii) **Insulators**

In an insulator, the forbidden band is very large (of the order of 5 eV or greater). Therefore electrons from valance band cannot reach the conduction band at room temperature. Hence the conduction band is empty and conduction is not possible through insulator.

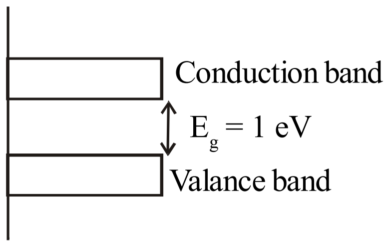


(iii) **Semiconductors**

(Semiconductor is a substance whose conductivity lies in between that of a good conductor and of an insulator)

In case of semiconductors, the energy gap is low i.e. about 1 eV. At room temperature, many electrons acquire sufficient energy to jump this narrow forbidden band so that there will be appreciable conductivity at room temperature.

Examples for semiconductors are Ge, Si etc.



Note: Usually if the outermost orbit contain less than four electrons it will act like a conductor and if the outer most orbit contain more than four electrons it will act like an insulator.

The outermost orbit of Si and Ge contains four electrons. It forms covalent bond by sharing one electron from each of the four neighboring atoms to make octet.

The atoms align themselves in a uniform 3 D pattern so that each atom is surrounded by four atoms. Such a pattern is called a crystal. At absolute zero (zero kelvin), all the valance electrons are tightly bonded. No free electrons are available for conduction. i.e. the semiconductor behaves as a perfect insulator at absolute zero.

At room temperature, some electrons acquire sufficient energy to break the covalent bond and arrives at the conduction band. So there is a deficiency of electrons in a covalent bond of the semiconductor. This is equivalent to a +ve charge and is called a hole. Both electrons in the conduction band and the holes, cause conduction in semiconductor.

Intrinsic semiconductor

A semiconductor in its pure form is called an intrinsic semiconductor. The number of electrons (n_e) and holes (n_h) will always be the the same in an intrinsic semiconductor.

Extrinsic semiconductor

Deliberately adding impurity atoms to a semiconductor to increase its conductivity is called doping. A doped semiconductor is called an extrinsic semiconductor. In an extrinsic semiconductor, $n_h \neq n_e$.

There are two types of extrinsic semiconductors:-

N - type semiconductor

By doping a pentavalent impurity atom like arsenic or antimony to an intrinsic semiconductor, the number of electrons are considerably increased (because each pentavalent atom makes covalent body with four neighboring intrinsic semiconductor atoms. So one electron is excess which cannot make covalent bond. It will reach the conduction band. A few mg of impurity atoms contributes millions of electrons to the conduction band.) Such *an extrinsic semiconductor in which electrons are majority carriers and holes are minority carriers is called N - type semiconductor.*

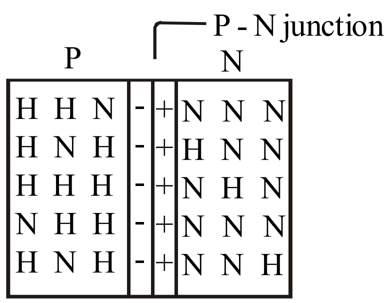
{Important Note: - An N - type semiconductor is electrically neutral because pentavalent impurity before doping and intrinsic semiconductor are electrically neutral} N - type never means that it is -vely charged.

P - type semiconductor

By doping with a trivalent impurity like Indium or Aluminium, the number of holes can be considerably increased. (because there is a deficiency of electron to form covalent bond which is equivalent to a hole. So each trivalent atom contribute a hole for conduction. A few mg of trivalent impurity contributes millions of holes. Apart from that there are few electrons in the conduction band already present in the intrinsic semiconductor) Such *an extrinsic semiconductor in which holes are majority carriers and electrons are minority carriers are called P - type semiconductors.*

A pentavalent impurity atom which supplies electrons is called a **donor type impurity** and the trivalent impurity atom which accepts electrons is called an **acceptor type impurity**.

P - N Junction



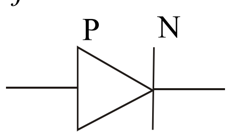
When a P - type semiconductor and an N - type semiconductor are joined together, a boundary junction is formed between them. This is called P - N junction.

When a P - N junction is formed electrons from N - region diffuses to P - region crossing the junction and neutralises the holes in the P - region. As a result, there will be a deficiency of electrons in N - region and deficiency of holes in P - region surrounding the junction. Due to this, N - type acquires a + ve potential and P - type acquires a - ve potential. Therefore a p.d is developed across the junction.

This p.d rated across the P - N junction due to the diffusion of electrons and holes which prevents further diffusion is called potential barrier.

The thin layer formed on both sides of the P - N junction which is devoid of majority carriers but immobile ions is called depletion layer.

A P - N junction can be represented as follows:



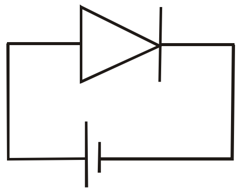
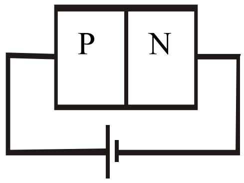
HSSLiVE.IN

Biasing of a P - N junction

Applying external pd across a P - N junction (i.e. connecting a battery to P - N junction) is called biasing of P - N junction. **A biased P - N junction is called a diode**. This can be done in two ways: -

(i) **Forward bias** - Here P - region is connected to +ve terminal of the source and N - region is connected to -ve terminal of the source.

(ii) **Reverse bias** - Here P - region is connected to -ve terminal of the battery and N - region is connected to +ve terminal of the battery.

1. Forward biasing

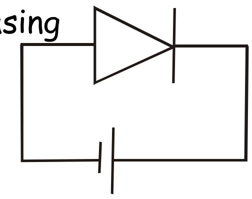
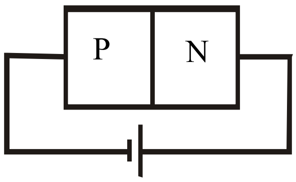
When a P - N junction is forward biased, the electrons in the N - region are repelled from the -ve terminal of the battery and reaches the P - region crossing the barrier. (Similarly for the holes)

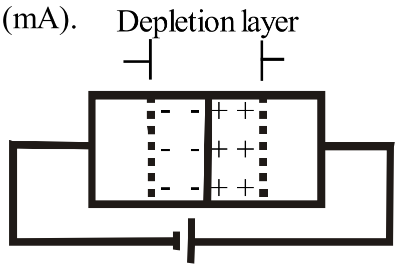
The electrons will recombine with the holes in the P - region. For each recombination of free electrons and holes that occur, an electron from -ve terminal of the battery enters the N - region. It then drifts towards the junction. Similarly in the P- type material, near the +ve terminal of the battery, an electron breaks a bond in the crystal and enters the +ve terminal of the battery. For each electron that breaks its bond, a hole is created. This hole drifts towards the junction. Note that there is continuous current in external circuit. (The current in N - region is due to movement of electrons whereas in P - region it is due to movement of holes).

If the battery voltage is increased, the barrier voltage is reduced because the biasing (pd) is opposite to potential barrier, thereby increasing the current through P - N junction.

During forward bias, (i) The barrier potential is reduced.
(ii) The width of the depletion layer is decreased.
(iii) The resistance offered by the junction called forward resistance will be reduced.
(iv) Electrons flow from N to P region through the junction and this gives rise to a forward current of the order of milli amperes (mA).

2. Reverse biasing



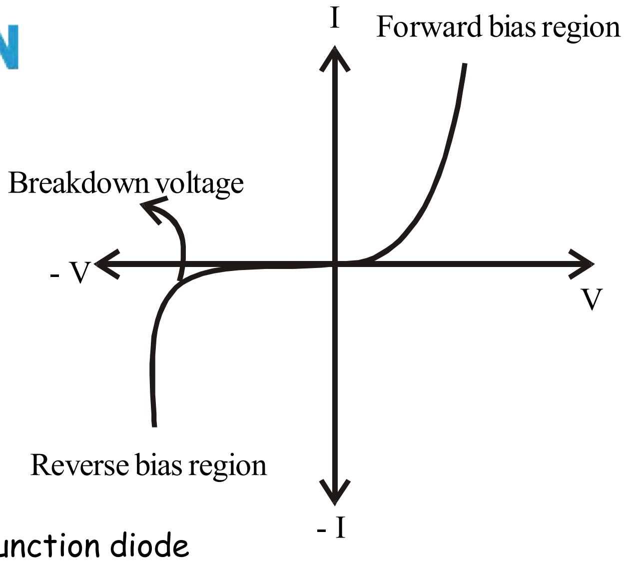
The holes in the P - region are attracted towards the -ve terminal of the battery. The electrons in the N - region are attracted towards the +ve terminal of the battery. Thus the majority carriers are drawn away from the junction. This action widens the depletion region and increases the value of barrier potential. However, the barrier potential is helpful for minority carriers in each region, and there will be a very small current due to minority carriers called reverse saturation current.

If the reverse bias voltage is increased, the KE of minority carriers increases and they knock out electrons from the covalent bond of the atoms. Thus large number of electrons and hole pairs are liberated and the reverse current increases abruptly to a high value. The breakdown of the junction occurs at this stage. *The reverse voltage at which P- N junction breaks down with sudden increase in reverse current is called breakdown voltage.*

During reverse bias, (i) The width of the depletion layer is increases.
(ii) The potential barrier is increased by the applied voltage.
(iii) The junction offers high resistance called reverse resistance.
(iv) For a particular reverse voltage (breakdown voltage), the current increases abruptly.

V - I characteristics of a diode.

A graph drawn connecting voltage and current is called V - I characteristics.

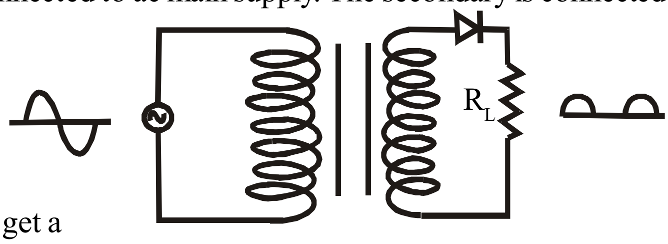


Applications of P - N junction diode

Diode can be used as a rectifier. *A device which converts ac into pulsating dc is called a rectifier.* During forward bias, the diode conducts and during reverse bias, the diode does not conduct. This is the basic principle behind a diode rectifier.

1. **Half wave rectifier**.

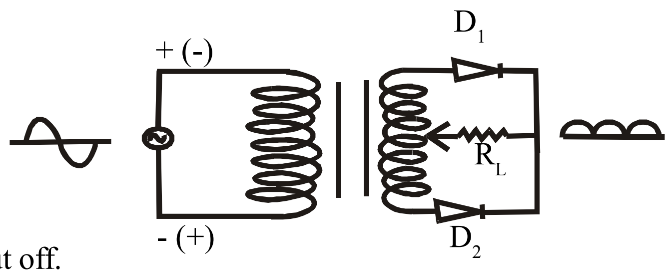
The primary of a step down transformer is connected to ac main supply. The secondary is connected in series with the diode and load resistance (R_L)



During +ve half of ac, diode is forward biased and hence conducts. The corresponding output voltage is obtained across the load resistance. During -ve half of ac diode becomes reverse biased and does not conduct. Hence we get a unidirectional intermittent current.

2. Full wave rectifier.

The primary of a step down transformer is connected to an ac supply. The secondary of the central tapped transformer is connected to two diodes D_1 and D_2. The output is taken across the load resistance R_L.

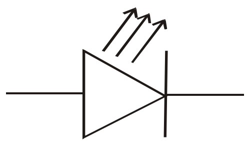


During +ve half of ac, D_1 is forward biased and D_2 is reverse biased. Hence D_1 conducts and D_2 cut off.

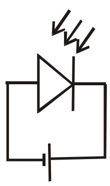
During -ve half, the polarity gets reversed. D_1 becomes reversed biased and D_2 becomes forward biased. Hence D_2 conducts D_1 cuts off. Hence for both half of ac, we get unidirectional current across load resistance R_L.

Special types of P - N junctions:

(i) **LED (Light Emitting Diode)**



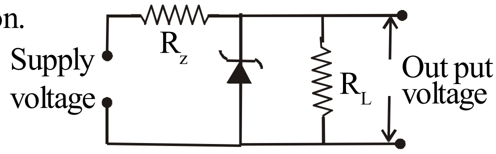
During forward bias, due to recombination of electrons and holes, energy is released. For Si and Ge diodes, these energy radiations belong to infrared region. If the P - N junction is made of Gallium Arsenide energy will be in infrared region. For Indium phosphate the energy radiations emitted will be in the visible region. Such a diode is called LED. LED's emit no light when reversed biased. Rather they will be destroyed when reversed biased.



(ii) **Photodiode**: A photodiode is a photodetector which is used to measure intensity of light.

A photodiode is a P - N junction made from photosensitive semiconducting material. The photodiode has to be reverse biased with voltage applied less than breakdown voltage. When light of suitable frequency fall on it, electrons and holes are produced and the reverse current increases with the intensity of the incident light.

(iii) **Zener diode**: It is a specially designed P - N junction which works in the reverse breakdown region. This is used for voltage regulation. For that the circuit is made as shown:



Working: When the supplied DC voltage increases beyond a d desired value, the voltage across the diode becomes breakdown voltage. Hence the current in the diode increases rapidly and causes a sufficient voltage drop across R to lower the output voltage back to normal value. (Due to parallel connection, the pd across Zener and load resistance R_L is same.)

(iv) **Solar cell**: It is a P - N junction that can convert light energy into electrical energy. When it is exposed to light, photons are absorbed and electron - hole pairs are generated in both P and N side of the junction. These electrons move towards N side and holes move towards the P - side. If no external voltage is connected, the electrons and holes are collected at the two sides of the junction which forms photo voltage.

Uses: Used in space satellites to provide power for several equipments.

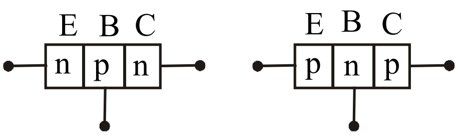
Junction Transistor

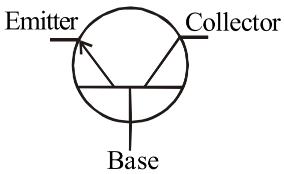
If two p-n junctions are formed on the same semiconductor crystal, the device is called a junction transistor.

Transistors are of two types – pnp and npn. In npn transistor, two n regions are separated by a p region. In pnp transistor, two p regions are separated by a n region.

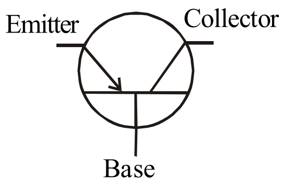
The transistor has three regions namely Emitter, Base and Collector.

1. **Emitter**; It is a region on one side of a transistor having large conductivity. *Emitter is always forward biased*. The emitter supplies a large number of majority carriers.
2. **Collector**: It is the region on the other side of the transistor. It collects the charge carriers emitted by the emitter. The collector always has low conductivity. *Collector is always reversed biased.*
3. **Base**: It is the middle region between the emitter and the collector. It is a thin region and is lightly doped.



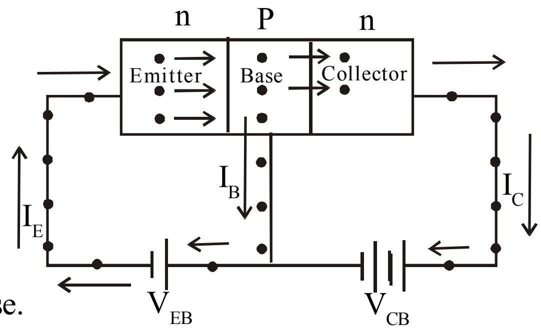


npn



pnp

Transistor action



To understand the transistor action, consider a npn transistor. Since the emitter base junction is forward biased, the electrons from the emitter diffuse into the base and holes from base diffuse into the emitter. Since the base is very thin, majority of the electrons from the emitter will reach the collector. Since the collector base junction is reverse biased, the collector is at a +ve potential wrt the base. As a result, the collector collects the electrons. This produces a current in the collector, called collector current (I_C). At the same time, the holes in the base region combines with the electrons from the emitter and a small current will flow in the emitter base circuit. This current is called base current (I_B).

Now applying Kirchoff's law, we get emitter current $I_E = I_C + I_B$. Thus emitter current is the sum of collector and base currents.

Same is the result in case of a pnp transistor.

Note: The collector current is of the order of milli amperes and base current of the order of microamperes.

$\therefore I_C >>> I_B$ and we can write, $I_E \approx I_C$.

The resistance of base - emitter junction (R_{BE}) is very low due to forward bias and that of collector - base junction (R_{CB}) is very high due to reverse bias. Therefore, collector voltage $I_C R_{CB}$ is large compared to that of emitter voltage $I_E R_{BE}$. (Let us consider collector voltage as output voltage and emitter voltage as input) The output voltage and hence output power is much greater than input power. This is the amplifying action of a transistor.

HSSLiVE.IN

There are two current gains in a transistor α and β. They are defined as: $\alpha = \frac{I_C}{I_E}$ and $\beta = \frac{I_C}{I_B}$

(If the currents are varying, $\alpha = \frac{\Delta I_C}{\Delta I_E}$ and $\beta = \frac{\Delta I_C}{\Delta I_B}$)

Relation between α and β.

$$I_E = I_C + I_B.$$

Divide by I_C.

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

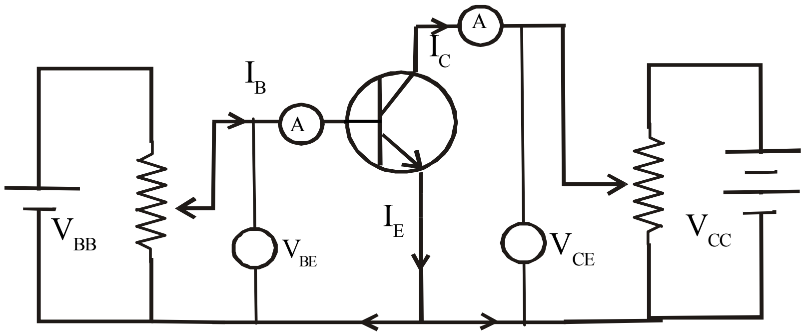
$$\therefore \alpha = \frac{\beta}{1+\beta} \quad \text{OR} \quad \beta = \frac{\alpha}{1-\alpha}$$

Transistor configurations

The method by which a transistor is connected in a circuit is called transistor configuration. There are three types of configurations – Common Base (CB) configuration, Common Emitter (CE) configuration and Common Collector (CC) configuration. In each of these modes, one terminal is common to the input and output parts. For eg; in the common base configuration, the input is supplied to emitter and base and the output is taken out from collector and base. Out of these, common emitter configuration has most wide ranging applications.

Common Emitter characteristics

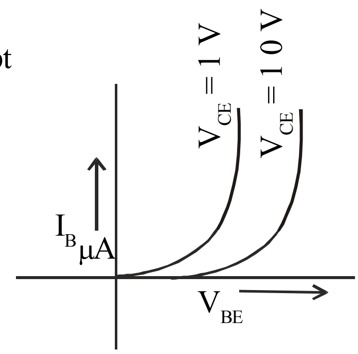
This is the widely used circuit. Here input is fed between B and E and output is taken across C and E. The circuit diagram is as shown.



1. **Input characteristics**

A graph drawn connecting the input voltage V_{BE} and input current I_B is called input characteristics.

To study the input characteristics, the collector emitter voltage (V_{CE}) is kept constant and the base current (I_B) is noted for different base emitter potentials (V_{BE}).
The variation of IB against V_{BE} is plotted in graph.



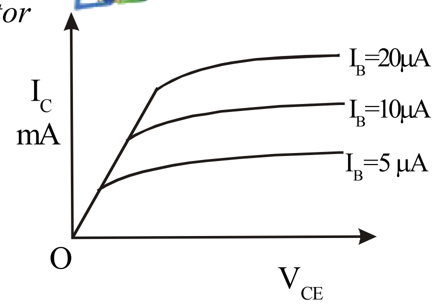
From graph, input resistance, $r_i = \left(\frac{\Delta V_{BE}}{\Delta I_C}\right)_{V_{CE}}$

The input resistance is low and is of the order of a few ohms.

HSSLiVE.IN

2. **Output characteristics**

The output characteristic is a graph connecting the collector current (I_C) and the collector – emitter voltage (V_{CE}) at constant base current (I_B).



It is obtained by measuring the collector current at different collector voltages by keeping the base current fixed.
The graph is as shown.
The ratio of small change in collector voltage to the correspondingchange in collector current at common base current is called output resistance r_o

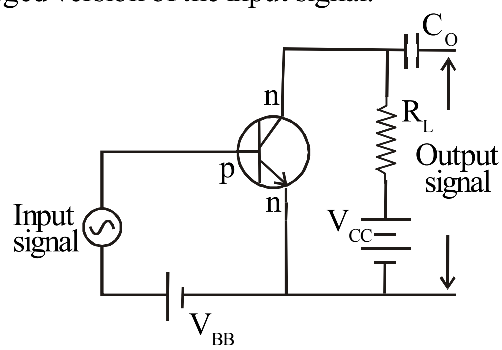
$$r_O = \left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B}$$

The output resistance has high value.

Transistor Amplifier

An amplifier is a device which is used for increasing the amplitude of variation of alternating voltage, current or power. The amplifier thus produces an enlarged version of the input signal.

Common Emitter Amplifier



The circuit diagram is as shown. Here emitter is common to both input and output. The input (Emitter - base) is forward biased with battery V_{BB} of voltage V_{EB} and output circuit (Collector - Emitter) is reverse biased with battery V_{CC} of voltage V_{CE}. Hence the resistance of input is low and that of output is high.

When no ac signal is applied to input, the emitter current,

$I_E = I_B + I_C$ ....(1)\

If V_C is the collector voltage, then $V_{CE} = V_C + I_C R_L$

i.e. $V_C = V_{CE} - I_C R_L$ ....(2)

Now signal is applied between base and emitter. So a small change in base voltage produces a change in base current I_B. This small variation in base current produces a large variation in collector current hence output voltage is increased.

Voltage gain (A_V)

It is the ratio of output voltage to the input voltage or it is the ratio of change in output voltage to a small change in input voltage.

Voltage gain $A_V = \frac{V_O}{V_i} = \frac{I_C R_L}{I_B R_i}$ where R_i is the input resistance. But current gain,

$\beta = \frac{I_C}{I_B}$ Therefore, $A_V = \beta \frac{R_L}{R_i}$

Current gain (β) Current gain is the ration of output current to the input current. i. e. $\beta = \frac{I_C}{I_B}$

Power gain (A_P)

Power gain is the ratio of output power to input power.

$$\text{Power gain,} \quad A_P = \frac{\text{output power}}{\text{input power}} = \frac{I_C V_O}{I_B V_i} = \beta A_V$$

But $A_v = \beta \frac{R_L}{r_i}$ ie, power gain, $A_P = \beta^2 \frac{R_L}{r_i}$

HSSLiVE.IN

Phase relationship between input and output

During the +ve half cycle of input ac signal, the EB circuit is more forward biased and emitter current and hence collector current increases. Then according to (2) collector voltage VC decreases. Thus the collector will become less +ve i.e, -ve wrt initial value. Thus during +ve half cycle of input ac signal voltage, the output signal voltage at collector varies through a - ve half cycle.

The -ve half cycle fo input ac signal opposes the forward biasing of EB circuit due to which the emitter current and hence collector current decreases. Hence collector voltage VC increases. Thus during -ve half cycle of input ac signal voltage, the output signal voltage varies through +ve half cycle.

Thus in CE amplifier, the input signal voltage and output voltage are opposite in phase i.e 180^0 out of phase

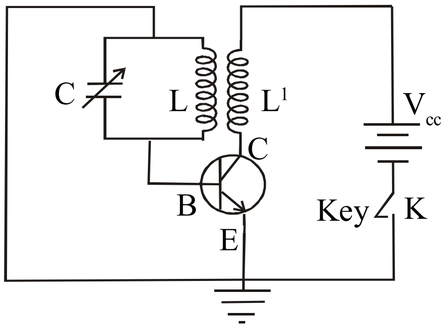
Transistor as an oscillator.

Oscillator is a device, which gives an output of any desired frequency without an external input signal.

The oscillator circuit has two principle sections:

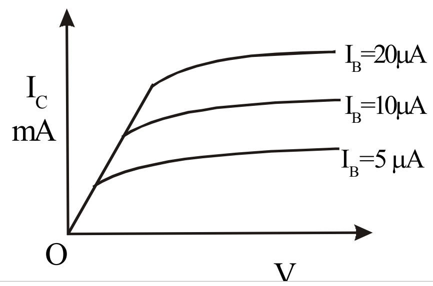
(i) Amplifying section:- This section is just a common - emitter amplifier with high voltage gain.

(ii) Feed back section :- It has a network of LC oscillatory circuit. Inductor - capacitor) to provide +ve feedback in correct phase.

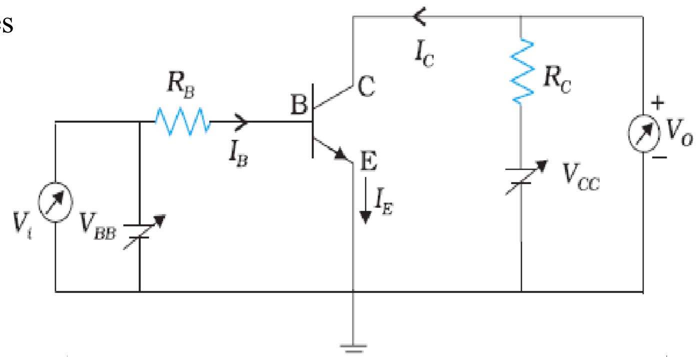


When the key K is switched on, LC oscillations are set up in the feed back section. Some part of this is fed back to the input circuit. Thus the transistor supply its own input and starts oscillating at a frequency which is determined by the value of inductor and capacitor in the feed back section. (The capacitor gets charged and discharged through the inductor and the frequency of oscillation is given by $f = \frac{1}{2\pi\sqrt{LC}}$)

Transistor as a switch.



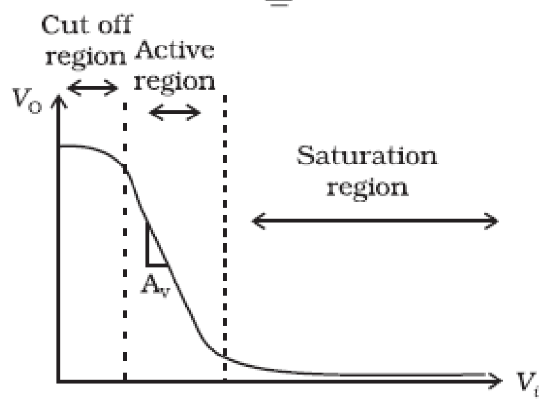
The output characteristics of a CE configuration is a graph connecting I_C and V_{CE} at constant I_B. We can explain the switching action of transistor using this curve. If the input base voltage is zero or negative, the transistor is said to be in OFF state. In this case, $I_B = 0$ and the output voltage is equal to collector supply voltage V_{CC}.



When the input voltage is increases to +ve values saturation value and transistor is said to be in ON condition. The output voltage in the on condition is V_{knee}.

The region between cutoff and saturation is called active region.

The transistor acts as a switch when it operates in cutoff or saturation state.



Applying Kirchoff's voltage law to the input side of this circuit, $V_{BB} = I_BR_B + V_{BE}$.

If $V_{BB} = V_i$ = dc input voltage,

$$V_i = I_BR_B + V_{BE}. ....(1)$$

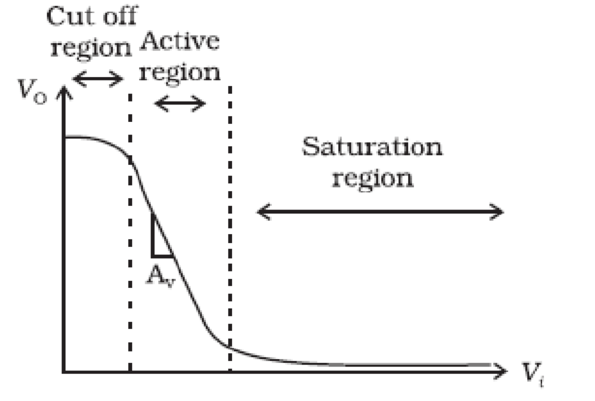
For output side of this circuit, on applying the same law,

$$V_{CE} = V_{CC} - I_CR_C$$

If $V_{CE} = V_O$ = dc output voltage.

HSSLiVE.IN

$V_O = V_{CC} - I_C R_C$ ....(2)



For a Silicon transistor, $V_{BE} = 0.7$ V Hence the transistor will be in cut off state as long as V_i is less than 0.7 V. The collector current I_C will be zero in this state. So $V_O = V_{CC}$.

When the input voltage is greater than 0.7 V, the transistor is in active state and collector current I_C starts to flow. Then the term $I_C R_C$ increases thereby reducing the value of V_O. The variation of I_C with V_i is linear. So as V_i increases, I_C increases linearly and V_O decreases linearly till its value becomes less than 1V. On further increasing V_i, the change of V_O with V_i becomes non - linear and the transistor inters into saturation state. The variation of V_O with V_i is shown in fig. This is called transfer characteristics.

When V_i is very low, the output V_O is high ($=V_{CC}$). When the transistor is not conducting, it is said to be switched off. If V_i is very high to drive the transistor into saturation state, V_O is low. The transistor driven into saturation state is said to be switched on. Thus a low input to the transistor gives a high output and a high input gives a low output. So a transistor can be used in switching circuits if it si allowed to operate either in cut off or in saturation region.

Logic Gates

A logic gate is an electronic circuit or device with two or more inputs but only a single output which occurs only for certain combinations of inputs.

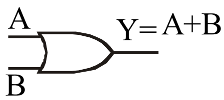
There are three basic gates: OR, AND and NOT.

1. Truth Table: It is a table which shows input and output possibilities of a given logic gate.
2. Boolean equation or Boolean expression: It is a simple way of representing the input combinations and output.

OR gate

OR gate has two or more inputs and one output. *It is based on the logic that the output is in the one state if one or more inputs are in the one state.* i.e. the output is high if one or more inputs are high.

(i) Logic symbol



Truth table

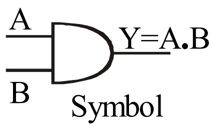
input		output
A	B	Y=A+B
0	0	0
0	1	1
1	0	1
1	1	1

(ii) Truth Table: Truth table of a two input OR gate is as shown: Here A and B are inputs and Y is the output.

(iii) Boolean Expression: $A + B = Y$ {Read as A or B = Y}

AND gate

An AND gate has two or more inputs and one output. *It is based on the logic that the output is in the one state (high) if and only if all the inputs are in the one state (high).*



Symbol

Truth table

input		output
A	B	Y=A.B
0	0	0
0	1	0
1	0	0
1	1	1

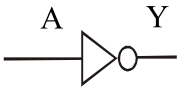
Boolean Expression : $A \bullet B = Y$ {Read as A and B = Y}

HSSLiVE.IN

NOT gate

A NOT gate is an electronic circuit in which out put is not present when input is present. A NOT gate converts 0 to 1 and 1 to 0.

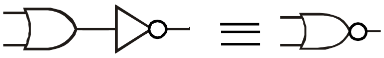
Boolean expression : $Y = \overline{A}$



Input A	output $Y=\overline{A}$
0	1
1	0

NOR gate

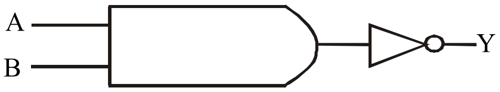
A **NOR** *gate is a* **NOT** *ed* **OR** *gate. It first performs an OR operation, the result of which is subjected to a NOT operation.*

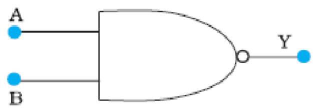


Input		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

NAND gate

A **NAND** *gate is* **NOT** *ed* **AND** *gate. It first performs AND operation, the result of which is subjected to a NOT op*

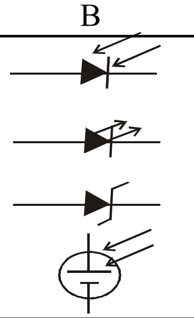




Input		Output
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

NAND gate is called a **universal gate** because it can be connected to other NAND gates to generate any logic function such as OR, AND and NOT. The *NOR gate is also a universal gate.*

$ Match the following:

A	B
zener diode	
solar cell	
laser diode	
daniel cell	
photo diode	