Electronic Devices: PN Junction Theory

Intrinsic Vs. Extrinsic Semiconductor:

- Semiconductors fall into two broad categories- intrinsic and extrinsic.
- An intrinsic semiconductor is one, which is pure enough that impurities do not appreciably affect its electrical behavior. It is composed of only one kind of material. Silicon and germanium are two examples of this kind of semiconductor. They are also called "undoped" or "i-type" semiconductors.
- Equal numbers of electrons and holes are present in an intrinsic semiconductor.
- The concentration of carriers is strongly dependent on the temperature. At low temperatures, the valence band is completely full making the material an insulator. Increasing the temperature leads to an increase in the number of carriers and a corresponding increase in conductivity. This characteristic shown by intrinsic semiconductor is different from the behavior of most metals, which tend to become less conductive at higher temperatures.
- Both silicon and germanium are tetravalent, i.e. each has four electrons (valence electrons) in their outermost shell. Both elements crystallize with a diamond-like structure, i.e. in such a way that each atom in the crystal is inside a tetrahedron formed by the four atoms which are closest to it. Each atom shares its four valence electrons with its four immediate neighbors, so that each atom is involved in four covalent bonds.
- If we take a crystal of some pure material, such as silicon, we find that each silicon atom has four bonds (this is called the valency of the atom), and each bond links it to another silicon atom. This bonding forms a crystal lattice and occupies all of the electrons in the silicon atoms. Because there are no free electrons, this material cannot pass a current, and is an insulator.

Extrinsic Semiconductor:

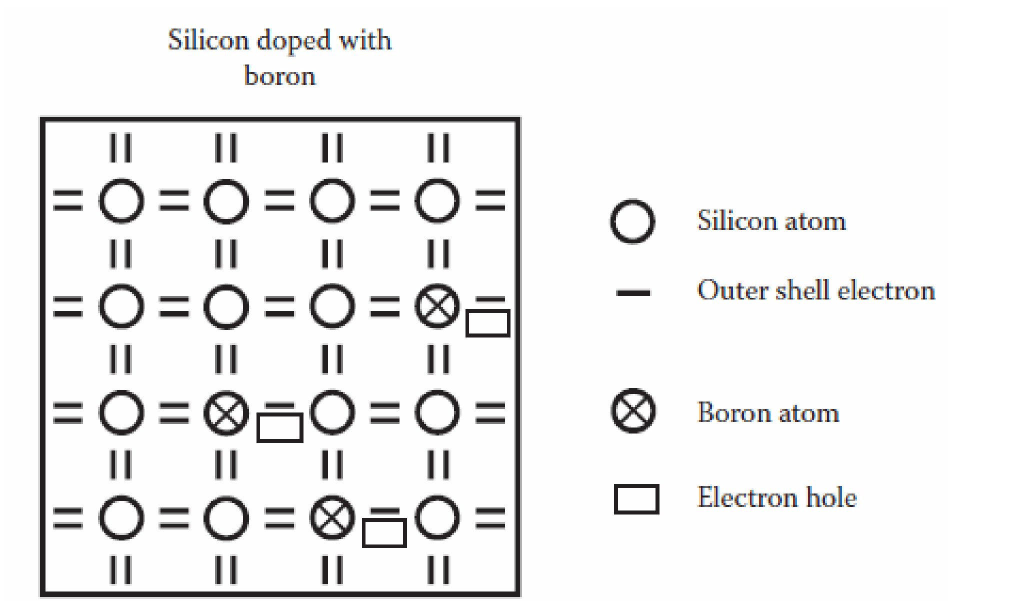
- Extrinsic semiconductors are made of intrinsic semiconductors that have had other substances added to them to alter their properties.
- The electronic properties and the conductivity of a semiconductor can be changed in a controlled manner by adding very small quantities of other elements, called "dopants", to the intrinsic material.
- If we carefully add an additional element to a pure material (such as silicon), the situation changes. If we add an atom with a valency of five instead of four, we find that wherever an atom of the introduced element is found, four of its bonds will attach to nearby silicon atoms, but the fifth bond will be unattached, leaving a free electron available, which will carry a current.

Doping and Dopant:

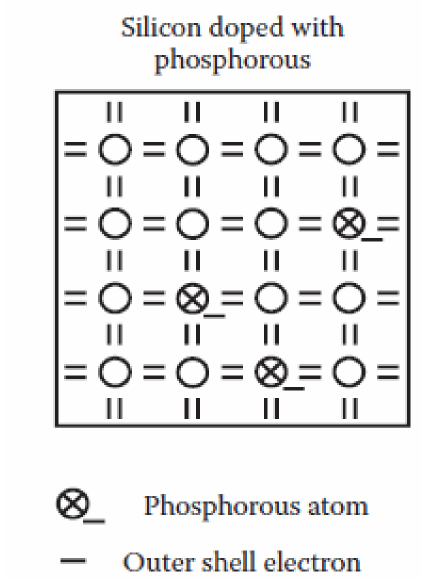
- The property of semiconductors that makes them most useful for constructing electronic devices is that their conductivity may easily be modified by introducing impurities into their crystal lattice. The process of adding controlled impurities to a semiconductor during the manufacturing process to increase its conductivity is known as doping. The impurities added are called dopants. Common dopants include arsenic, antimony, bismuth and phosphorous.
- The type and level of doping determines whether the semiconductor is N-type (current is conducted by excess free electrons) or P-type (current is conducted by electron vacancies).
- The amount of impurity, or dopant, added to an intrinsic (pure) semiconductor varies its level of conductivity. Doped semiconductors are often referred to as extrinsic.
- By adding impurity to pure semiconductors, the electrical conductivity may be varied not only by the number of impurity atoms but also, by the type of impurity atom.

Acceptor and Donor:

- The materials chosen as suitable dopants depend on the atomic properties of both the dopant and the material to be doped.
- In general, dopants are classified as either electron acceptors or donors. A donor atom donates weakly bound valence electrons to the material, creating excess negative charge carriers. These weakly bound electrons can move about in the crystal lattice relatively freely and can facilitate conduction in the presence of an electric field.
- Conversely, an acceptor atom accepts an electron from the material to be doped and thereby produces a hole.
- Semiconductors doped with donor impurities are called n-type, while those doped with acceptor impurities are known as p-type. The n and p type designations indicate which charge carrier acts as the material's majority carrier. The opposite carrier is called the minority carrier.
- For example, the pure semiconductor silicon has four valence electrons. Boron has three valence electrons. It can be used as an acceptor when used to dope silicon. Therefore, a silicon crystal doped with boron creates a p-type semiconductor.

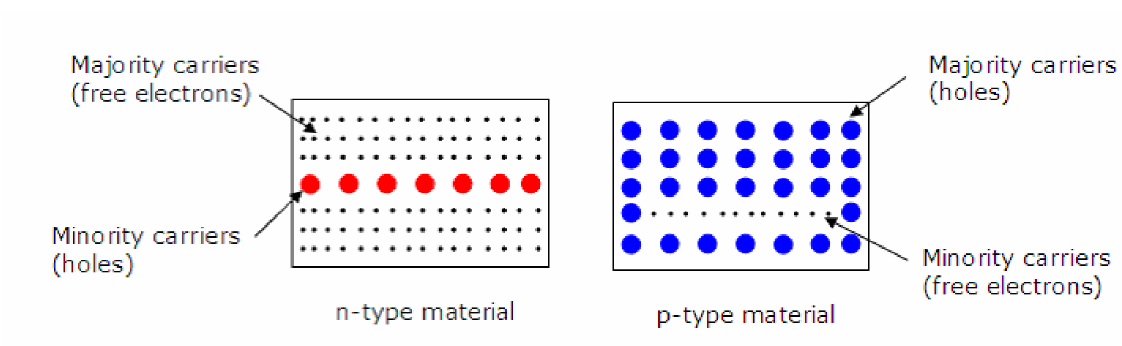


- On the other hand, phosphorous has five valence electrons. So it can be used as a donor when used to dope with silicon. Therefore, a silicon crystal doped with phosphorous creates an n-type semiconductor.



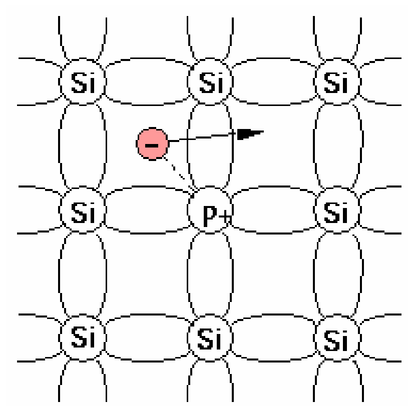
Majority and Minority Carriers:

- If we dope silicon with material having a valency of five, extra free electrons are created in the crystal. These electrons are called majority carriers- since the majority portion of current is by the flow of free electrons. Holes are the minority carriers in this case. Material doped in this way is referred to as n-type semiconductor.
- On the other hand, if we dope silicon with a material having a valency of three, holes are created in the crystal due to the absence of electrons. The holes carry a charge, equivalent to the charge on an electron, but of the opposite sign. Material doped in this way is referred to as p-type semiconductor. Holes outnumber the free electrons. Therefore, in a p-type material, holes are the majority carriers and free electrons are the minority carriers.



n-Type Semiconductor:

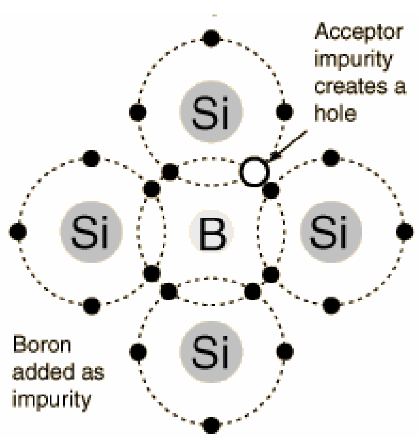
- When a small amount of pentavalent impurity is added to a pure semiconductor, it is known as n-type semiconductor.
- The addition of pentavalent impurity provides a large number of free electrons in the semiconductor crystal. Typical examples of pentavalent impurities are arsenic, antimony, phosphorous. Such impurities which produce n-type semiconductor are known as donor impurities, because they donate or provide free electrons to the semiconductor crystal.
- To explain the formation of n-type semiconductor, consider a pure silicon crystal.
  - We know that silicon atom has four valence electrons.
  - When a small amount of pentavalent impurity like phosphorous is added to silicon crystal, a large number of free electrons become available in the crystal.
  - The reason is simple. Phosphorous is pentavalent i.e. its atom has five valence electrons.
  - A phosphorous atom fits in the silicon crystal in such a way that its four valence electrons form covalent bonds with four silicon atoms.
  - The fifth valence electron of phosphorous atom finds no place in covalent bonds and is thus free as shown in the figure below.



  - Therefore, for each phosphorous atom added, one free electron will be available in the silicon crystal.
  - Though each phosphorous atom provides one free electron, yet an extremely small amount of phosphorous impurity provides enough atoms to supply millions of free electrons.

p-Type Semiconductor:

- When a small amount of trivalent impurity is added to a pure semiconductor, it is known as p-type semiconductor.
- The addition of trivalent impurity provides a large number of holes in the semiconductor crystal. Typical examples of trivalent impurities are boron, gallium, indium etc.
- Such impurities which produce p-type semiconductor are known as acceptor impurities, because the holes created can accept the electrons.
- To explain the formation of p-type semiconductor, consider a pure silicon crystal.
  - We know that silicon atom has four valence electrons.
  - When a small amount of trivalent impurity like boron is added to silicon crystal, there exist a large number of holes in the crystal.
  - The reason is simple. Boron is trivalent i.e. its atom has three valence electrons.
  - Each boron atom fits into the silicon crystal but now only three covalent bonds can be formed. It is because three valence electrons of boron atom can form only three single covalent bonds with three silicon atoms as shown in the figure below.



  - In the fourth covalent bond, only silicon atom contributes one valence electron while boron has no valence electron to contribute as all its three valence electrons are already engaged in the covalent bonds with neighboring silicon atoms. In other words,

****

fourth bond is incomplete; being short of one electron. This missing electron is called a hole.

- Therefore, for each boron atom added, one hole is created. A small amount of boron provides millions of holes.

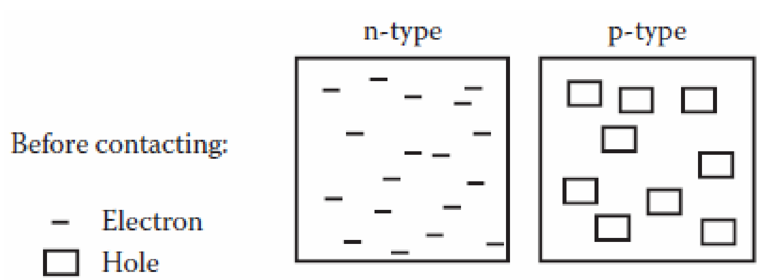
- When a doped semiconductor contains excess holes it is called "p-type", and when it contains excess free electrons it is known as "n-type", where p (positive for holes) or n (negative for electrons) is the sign of the charge of the majority mobile charge carriers.
- An *N-type* semiconductor carries current mainly in the form of negatively-charged electron, in a manner similar to the conduction of current in a wire.
- A P-type semiconductor carries current predominantly as electron deficiencies called holes. A hole is a place in a crystal where an electron normally would be, but is not. A hole has a positive electric charge, equal and opposite to the charge on an electron.
- In a semiconductor material, the flow of holes occurs in a direction opposite to the flow of electrons.
- The junctions which form where n-type and p-type semiconductors join together are called p-n junctions, which are the basis for all kinds of semiconductor devices.

P-N Junction

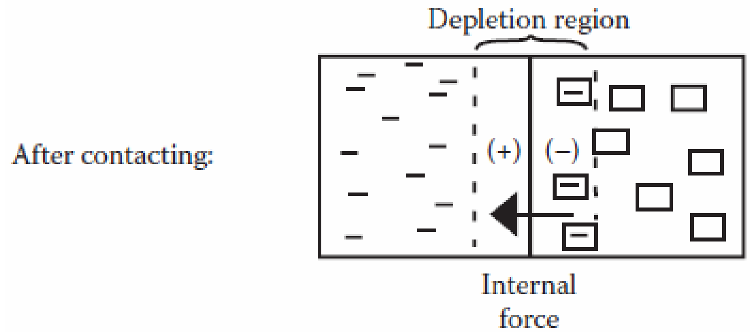
- On its own, a p-type or n-type semiconductor is not very useful. However very useful devices can be made when they are combined.
- When a p-type semiconductor is suitably joined to an n-type semiconductor, the contact surface is called pn junction.
- Most semiconductor devices contain one or more pn junctions.
- The pn junction is of great importance because it is in effect, the control element for semiconductor devices. Hence, a thorough knowledge of the formation and properties of pn junction is needed to properly understand the semiconductor devices.
- A diode is made of two joined semiconductor crystals—one p-type and the other n-type.
- The p-n junction can be formed by allowing a p-type material to diffuse into an n-type region at high temperatures.

Properties of P-N Junction:

- To explain the properties of a pn junction, let us consider two blocks of semiconductor materials: one p-type and another n-type as shown below.



- In the above figure, excess electrons of n-type are shown as minus symbols (–), and holes in p-type are shown by the box symbols.
- Holes and electrons are the majority charge carriers in p-type and n-type materials respectively.
- Now, suppose the two pieces of semiconductors are suitably joined to form pn junction which is shown below.



- In the above figure, p-type and n-type semiconductor materials are shown side by side.
- The region just around the contact is called a p-n junction.
- Keep in mind that n-type material has a high concentration of free electrons while p-type material has a high concentration of holes. Notice that at the junction, holes and electrons are close together, so some free electrons from the n-side are attracted across the junction to fill adjacent holes on the p-side. Similarly, some free holes from the p-side are attracted across the junction to fill adjacent electrons on the n-side. This process is called diffusion.
- The free electrons crossing the junction create negative ions on the p-side by giving some atoms one more electron than their total number of protons. The electrons also leave positive ions behind them on the n-side. Hence, a positive charge is established on the n-side and a negative charge is created on the p-side of the junction.
- Electrons are charged negatively, so if many electrons diffuse into the p-type semiconductor crystal, then the part of that crystal nearest the junction will become charged negatively, shown by the (−) symbol in the above figure. Then the part of the n-type crystal nearest the junction (initially neutral), from which the electrons have departed, will become charged positively, shown by the (+) symbol.
- When a sufficient number of electrons and holes cross the junction, further diffusion is prevented, because now positive charge on the n-side of the junction repels holes to cross from p-type to n-type. Similarly, now negative charge on the p-side of the junction repels free electrons to cross from n-type to p-type. Thus, a barrier is set up against further movement of charge carriers. This is called potential barrier or junction barrier V_0 which is of the order of 0.1 to 0.3 volt.

- At the junction of a p-type and an n-type semiconductor there is a thin region (about 10 μm in length) around the junction where the n-type piece is positively charged and the p-type piece is negatively charged. This region is called the depletion region or depletion layer which blocks current conduction from the n-type region to the p-type region, but allows current to conduct from the p-type region to the n-type region. In regions away from the depletion region, the crystals are neutral, as usual.

Applying Voltage Across the pn Junction:

The potential difference across a pn junction can be applied in two ways:

- Forward Biasing
- Reverse Biasing

- **Forward Biasing**
  - When an external voltage is applied to the pn junction making the p-side positive with respect to the n-side, the pn junction is said to be forward biased (F.B).
  - To apply forward bias, connect positive terminal of the battery to p-type and negative terminal to n-type as shown in the figure below:

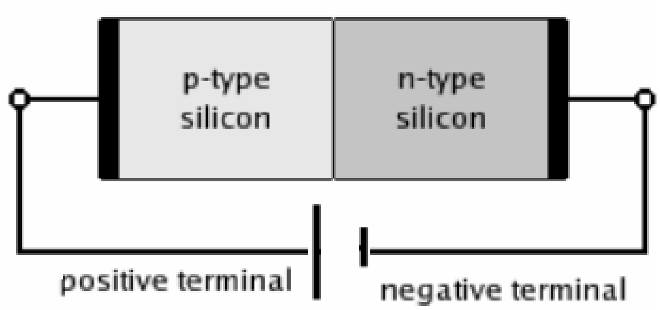


Figure: Forward biasing

  - The applied forward potential establishes an electric field which acts against the filed due to potential barrier. Therefore, the resultant field is weakened and the barrier height is reduced at the junction. (As potential barrier voltage is very small (0.1 v to 0.3 v), a small forward voltage is sufficient to completely eliminate the barrier).
  - Once the potential barrier is eliminated by the forward voltage, junction resistance becomes almost zero and a low resistance path

is established for the entire circuit. Therefore, current flows in the circuit. This is called forward current. The magnitude of this current depends upon the applied forward voltage.

- In this mode, potential barrier voltage is decreased or cancelled and depletion band narrows, thus permitting the movement of majority carriers across the junction and hence, current flows.
- A forward biased diode has a very low resistance.

Reverse Biasing

- When an external voltage is applied to the pn junction making the p-side negative with respect to the n-side, the pn junction is said to be reverse biased (R.B).
- To apply reverse bias, connect negative terminal of the battery to p-type and positive terminal to n-type as shown in the figure below:

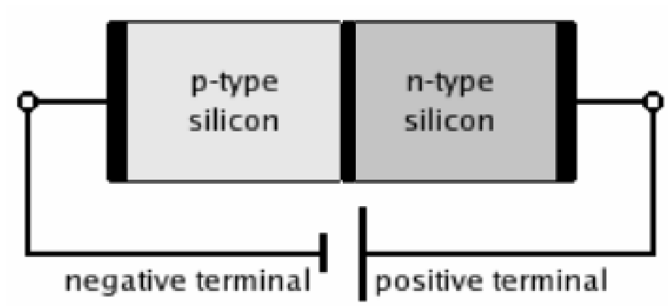


Figure: Reverse biasing

- It is clear that applied reverse voltage establishes an electric field which acts in the same direction as the field due to potential barrier. Therefore, the resultant field at the junction is strengthened and the barrier height is increased. The increased potential barrier prevents the flow of charge carriers across the junction. Thus, a high resistance path is established for the entire circuit and hence the current does not flow.
- In this mode, potential barrier voltage is increased and depletion band widens, thus preventing the movement of majority carriers across the junction and hence, no current flows.
- A reverse biased diode has a very high resistance.

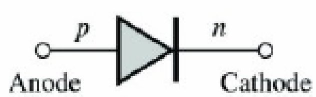
****

Semiconductor Diodes

- The first solid-state devices used in electronics were diodes.
- Diode is a semiconductor device made from joining one n-type semiconductor with one p-type semiconductor. This two-layer device consists of a single p-n junction, and is the simplest semiconductor device.
- A diode conducts electric current in only one direction, functioning as a one-way valve.
- Diodes typically are made from semiconductor materials such as silicon, germanium or selenium and are uses as voltage regulators, signal rectifiers, oscillators and signal modulators/demodulators.
- Understanding diode operation will prepare us to understand the operation of transistors—the most important component in modern computer circuits.
- A diode is made of two joined semiconductor crystals—one p-type and the other n-type, as in the figure below.

Symbol of a Diode:

Diode is a unijuction device (i.e. it is made from a single p-n junction). It has two distinct polarities namely anode which is negative and cathode which is positive.



Basic Operation of a diode

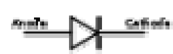
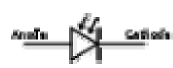
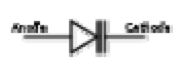
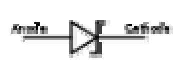
It allows current to flow only in one direction.

- Diode conducts when the polarities of diode is in reverse to the polarities of voltage source and forward biased voltage is greater than the break over voltage.
- Diode does not conduct when the polarities of diode is in the same as the polarities of voltage source or forward biased voltage is less than the break over voltage.

****

- When current is flowing through the device, it is said to be **forward biased** and has a fairly low resistance. If the voltage across the diode is reversed, the diode is said to be **reverse biased** and usually no current will flow.
- When reverse biased, free electrons in the n-type material, being negative, are attracted to the positive potential applied to the diode. At the other side, positively charged holes are attracted to the negative terminal. The carriers (holes and electrons) are pulled away from the junction area, leaving it depleted of any charge carriers. This forms a **depletion layer** and it is an insulator, so no current can flow through the device, except in Zener diodes.
- When the potential difference across the diode is reversed, the opposite happens - carriers are pushed away from the terminals towards the junction, until they reach the opposite type of material. Here, electrons are able to "fall into" holes, completing the circuit and allowing current to flow.

Various Types of Diode:

Types of Diode	**Symbol**
PN Diode	
Light Emitted Diode	
Schottky Diode	
Photo Diode	
Varactor Diode	
Zener Diode	

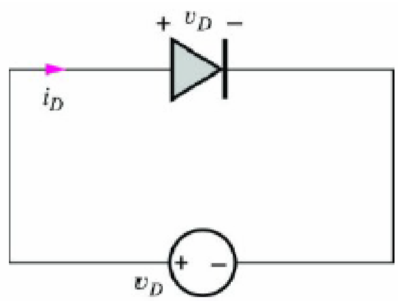
Diode Applications:

– Rectifiers

– Waveform clipping and clamping circuits

– DC-DC converters

Diode Terminal I-V Characteristics



$$i_D = I_S\left[\exp\left(\frac{v_D}{V_T}\right) - 1\right]$$

where I_S = reverse saturation current [A]

v_D = voltage applied to diode [V]

V_T = kT/q = thermal voltage [V] (25 mV at room temperature)

q = electronic charge [1.60×10^{-19} C]

k = Boltzmann's constant [1.38×10^{-23} J/K]

T = absolute temperature [K]

I_S is typically between 10^{-18} and 10^{-9} A. It depends on device parameters like junction area, doping levels, temperature, etc.

Reverse, Zero, and Forward Bias

Reverse bias, i.e., $v_D < 0$:

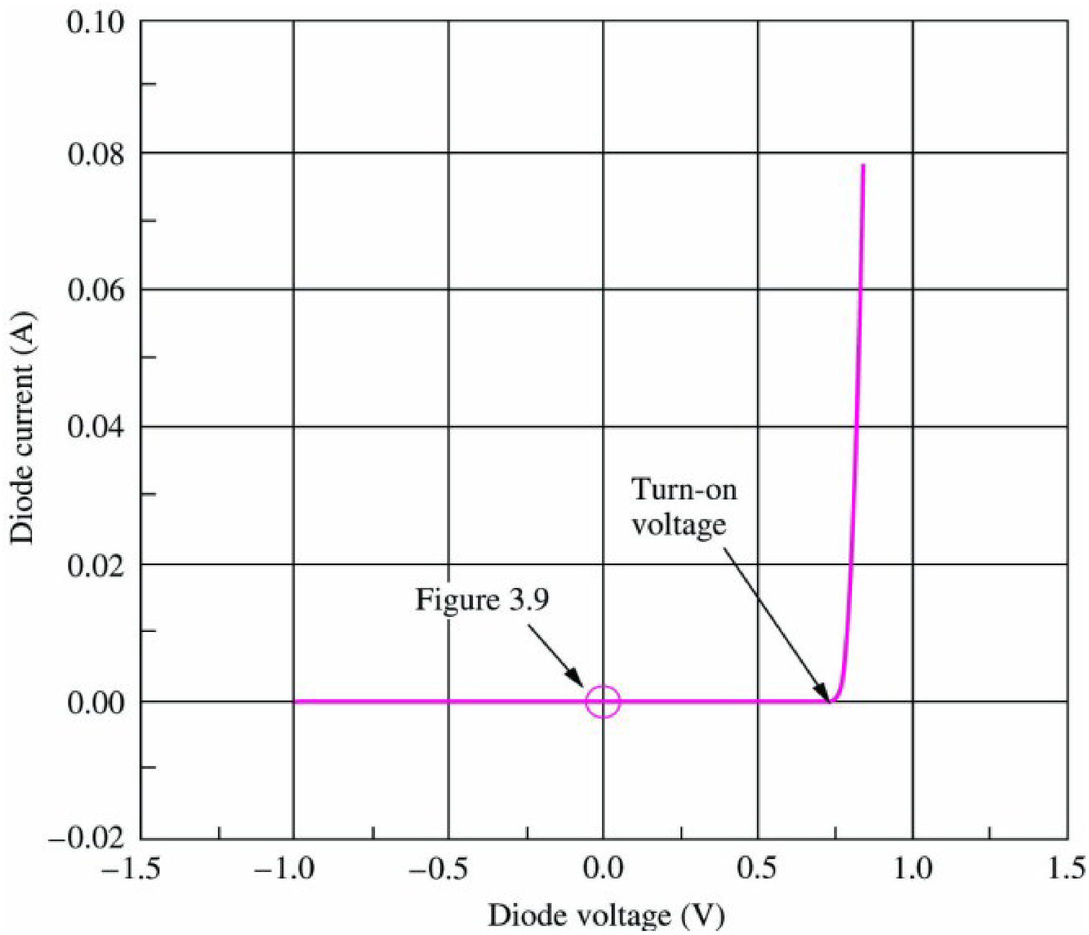
$$i_D = I_S\left[\exp\left(\frac{v_D}{V_T}\right) - 1\right] \approx I_S[0-1] \approx -I_S \qquad \text{for } v_D < -4V_T$$

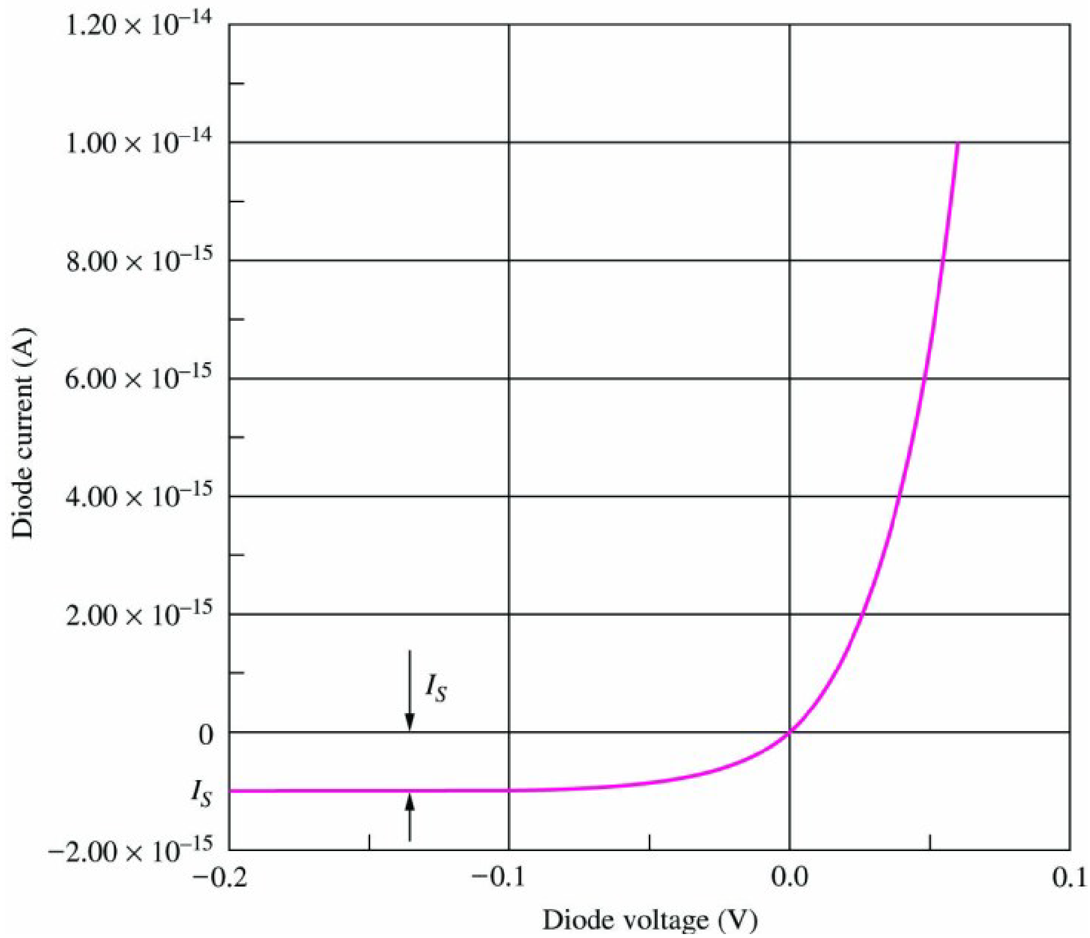
Zero bias, i.e., $v_D = 0$:

$$i_D = I_S\left[\exp\left(\frac{v_D}{V_T}\right)-1\right] = I_S[1-1] = 0$$

Forward bias, i.e., $v_D > 0$:

$$i_D = I_S\left[\exp\left(\frac{v_D}{V_T}\right)-1\right] \approx I_S\exp\left(\frac{v_D}{V_T}\right) \qquad \text{for } v_D > 4V_T$$




Diode current (A)
1.20 × 10^-14
1.00 × 10^-14
8.00 × 10^-15
6.00 × 10^-15
4.00 × 10^-15
2.00 × 10^-15
0
I_S
−2.00 × 10^-15
I_S
−0.2
−0.1
0.0
0.1
Diode voltage (V)

Transistor

- When a third doped element is added to a crystal diode in such a way that two pn junctions are formed, the resulting device is known as transistor.
- It is a kind of electronic device that is capable of achieving amplification of weak signal far far better than that of vacuum tube.
- Transistors are far smaller than vacuum tubes, have no filament and hence need no heating power and may be operated in any position.
- Transistors are mechanically strong, have practically unlimited life and can do some jobs better than vacuum tubes.
- Transistor was invented by J. Bardeen and W. H. Brattain of Bell Laboratory in 1948. It has now become the heart of most electronic applications.

Types of Transistors:

- A transistor consists of two pn junctions formed by sandwiching either p-type or n-type semiconductor between a pair of opposite types.
- Accordingly, there are two types of transistors:
    - (i) npn transistor
    - (ii) pnp transistor

- An npn transistor is composed of two n-type semiconductors separated by a thin section of p-type.
- A pnp transistor is formed by two p-type semiconductors separated by a thin section of n-type.

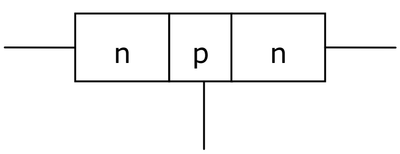


Figure: an npn transistor

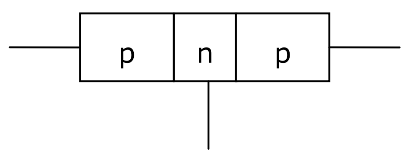


Figure: a pnp transistor

- As there are two pn junctions exist in a transistor, therefore a transistor may be regarded as a combination of two diodes connected back to back.

Origin of the name 'Transistor':

- When new devices are invented, scientists often try to devise a name that will appropriately describe the device.
- A transistor has two pn junctions: one is forward biased and another is reverse biased.
- The forward biased junction has a low resistance path whereas the reverse biased junction has a high resistance path.
- The weak signal is introduced in the low resistance circuit and the output is taken from the high resistance circuit.
- Therefore a transistor transfers a signal from a low resistance to high resistance.
- The prefix 'trans' means the signal transfer property of the device while 'istor' classifies it as a solid element in the same general family with resistor.

Naming the Transistor Terminals:

- A transistor has three sections, namely: emitter, base, and collector.
- The section on one side is the emitter and the section on the opposite side is the collector. The middle section (which is very thin as compared to other two sections) is called the base. This section forms two junctions between the emitter and collector.

(i) Emitter:

- It supplies charge carriers (electrons or holes).
- The emitter is heavily doped so that it can inject a large number of charge carriers (electrons or holes) into the base.
- The emitter is always forward biased with respect to base so that it can supply a large number of majority carriers.
- Figure A shows that the emitter of p-type (pnp) transistor is forward biased and supplies hole charges to its junction with the base. Similarly, figure B shows that the emitter of n-type

(npn) transistor is also forward biased and supplies free electrons to its junction with the base.

(ii) Collector:

- The collector is moderately doped and collects the charge.
- The collector is always reverse biased with respect to base.
- Its function is to remove charges from its junction with the base.
- Figure A shows that the collector of p-type (pnp) transistor is reverse biased and receives hole charges that flow in the output circuit. Similarly, figure B shows that the collector of n-type (npn) transistor is also reverse biased and receives electrons.

(iii) Base:

- The middle section of transistor which forms two pn junctions between emitter and collector is called the base.
- The base is lightly doped and very thin.
- It passes most of the emitter injected charge carriers to the collector.
- The base-emitter junction is forward biased, allowing low resistance for the emitter circuit.
- The base-collector junction is reverse biased and provides high resistance in the collector circuit.

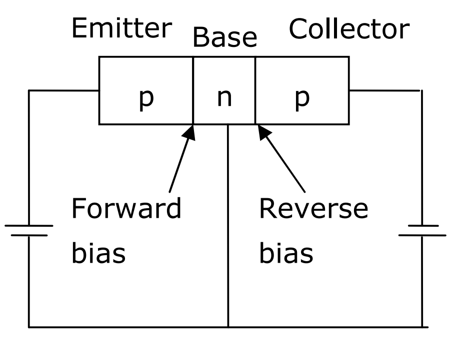


Figure: npn transistor

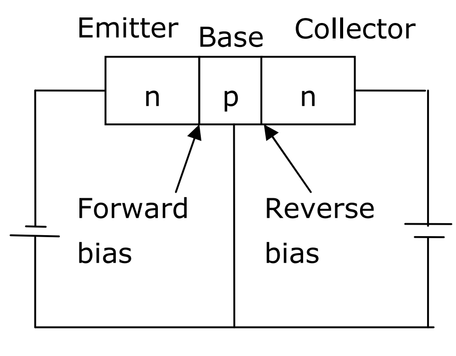


Figure: pnp transistor

Transistor Symbol:

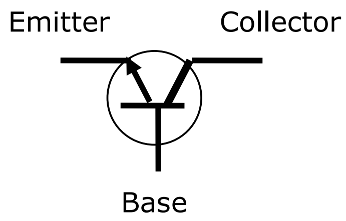


Figure: Symbol of npn transistor

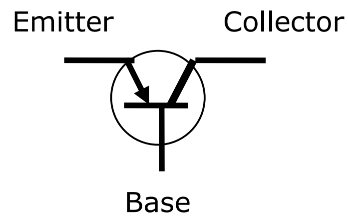


Figure: Symbol of pnp transistor