What is vacuum tube?

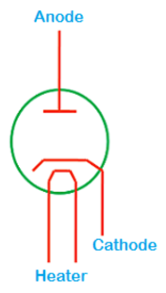
Generally, vacuum refers to a space where charged particles such as electrons, protons, neutrons and all other matter are absent. In other words, vacuum is nothing but the empty space.

Vacuum tube is an electronic device that controls the flow of electrons in a vacuum. It is also called as electron tube or valve. John Ambrose Fleming developed the first vacuum tube in 1904. Fleming's diode allows the flow of electric current in only one direction (from cathode to anode) and blocks the electric current in another direction (from anode to cathode). In 1906, American electrical engineer Lee De Forest invented Audion vacuum tube.

The invention of vacuum tubes has produced a new branch of engineering called electronics. In early days, vacuum tubes are used in television, radios, radar, electronic computers, and amplifiers. However, after the development of semiconductor devices, the usage of vacuum tubes in the electronic devices was reduced. Now-a-days, most of the electronic devices (computers, television, radar etc.) made from vacuum tubes are replaced by the semiconductor devices such as diodes, transistors, and integrated circuits.

Vacuum tubes are huge and occupy large amount of space. However, the construction and operation of vacuum tubes is easy to understand. Vacuum tubes are made from the materials such as glass and ceramics. Vacuum tubes are mostly depends on the thermionic emission process to emit the free electrons. In the thermionic process, heat is used to emit the free electrons. Vacuum tube that emits the free electrons by the application of heat is called thermionic valve or thermionic tube.

A vacuum tube consists of cathode (also called as filament), anode (also called as plate), and electrode (also called as grid). Cathode is an electron emitter that emits the free electrons whereas anode is an electron collector that collects the free electrons.

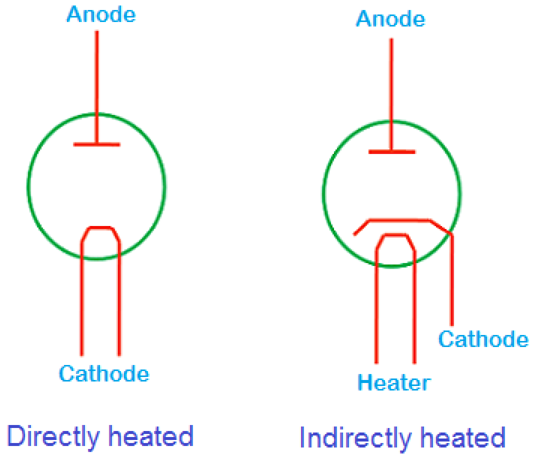


Grid or electrode controls the electric current or flow of electrons between anode and cathode. The free electrons that are emitted by the cathode are attracted towards the anode or plate. These free electrons carry the electric current while moving from cathode to anode.

Directly heated and indirectly heated cathode

In the thermionic tubes, the cathode is heated electrically to a desired temperature to emit the free electrons from the metal surface. This can be done in two ways: by directly heating the cathode or indirectly heating the cathode.

If the heat or heating electric current is passed directly to the cathode that emits the free electrons, the cathode is said to be a directly heated cathode or directly heated emitter. In the directly heated cathode, the cathode itself is the heating element or filament. Hence, the heat required to emit the free electrons from the metal surface is less compared to the indirectly heated cathode.



If the heat or heating electric current is passed indirectly to a cathode that emits the free electrons, the cathode is said to be an indirectly heated cathode or indirectly heated emitter.

In the indirectly heated cathode, there is no electrical connection between the cathode and the heater. Hence, the cathode itself is not a heating element. The heating electric current is passed through the heater or filament and the cathode is heated indirectly. Hence, the amount of heat required to emit the free electrons from the metal surface is more compared to the directly heated cathode.

Types of vacuum tubes

Vacuum tubes are generally classified into four types:

- **Vacuum diodes**
- **Vacuum triodes**
- **Vacuum tetrodes**
- **Vacuum pentodes**

Advantages and disadvantages of vacuum tubes

Advantages of vacuum tubes

- Vacuum tubes are replaced easily.
- Vacuum tubes can works at high temperature without any damage.
- Vacuum tubes produce superior sound quality.

Disadvantages of vacuum tubes

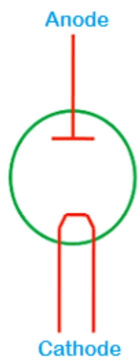
- Vacuum tubes are huge compared to the semiconductor devices such as diodes, transistors, and integrated circuits.
- Vacuum tubes generate more heat.
- High voltages are required to operate the vacuum tubes.
- Vacuum tubes consume more power.
- High cost.
- Failure rate is high.
- Vacuum tubes occupy more space than the transistors.

Vacuum diode

In 1904, Sir John Ambrose Fleming invented the first vacuum tube called vacuum diode. It is also called Fleming valve or thermionic tube. Vacuum diode is an electronic device that allows the electric current in one direction (cathode to anode) and blocks the electric current in another direction (anode to cathode).

Two electrodes of vacuum diode

Vacuum diode is the simplest form of vacuum tube. It consists of two electrodes, a cathode, and an anode or plate. The cathode emits the free electrons. Hence, it is called as emitter. The anode collects the free electrons. Hence, it is called as collector.



The cathode and anode are enclosed in an empty glass envelope. The anode is a hollow cylinder made of molybdenum or nickel and cathode is a nickel cylinder coated with strontium and barium oxide. The anode surrounds the cathode. In between the cathode and anode an empty space is present, through which the free electrons or electric current flow.

What is electrode?

Electrode is a conductor through which free electrons or electric current leaves or enters. In vacuum diode, cathode is an electrode or conductor from which the free electrons are emitted into the vacuum. On the other hand, anode is an electrode that collects the free electrons emitted by the cathode. In other words, free electrons leave the cathode and enter into anode.

Electron emission depends on the amount of heat applied and the work function

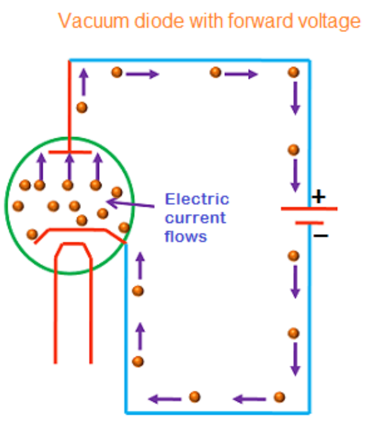
The number of free electrons emitted by the cathode is depends on two factors: amount of heat applied and work function.

If more amount of heat is applied, more number of free electrons is emitted. Similarly, if less amount of heat is applied, less number of free electrons is emitted.

The minimum amount of energy required to remove the free electrons from the metal is called work function. Metals with low work function will require less amount of heat energy to emit the free electrons. On the other hand, metals with high work function will require large amount of energy to emit the free electrons.

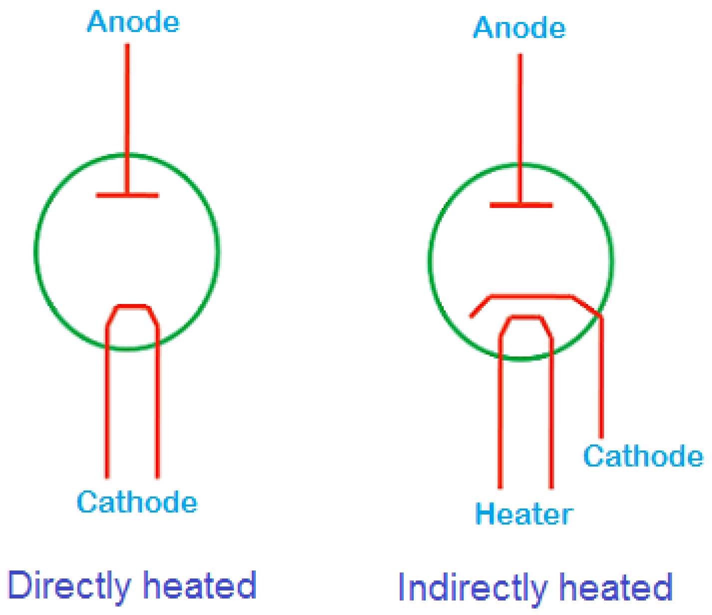
Hence, choosing a good material will increase the electron emission efficiency. Most commonly used thermionic emitters include oxide-coated cathode, tungsten, and thoriated tungsten.

Directly and indirectly heated cathode



When the cathode is indirectly or directly heated, free electrons are emitted from it.

In the directly heated cathode, the heat energy is supplied directly to the cathode. Hence, a small amount of heat energy is enough to emit the free electrons from the cathode. When the heat energy is directly supplied to the cathode, large number of free electrons gain sufficient energy and breaks the bonding with the cathode. The free electrons that break the bonding with the cathode are emitted into the vacuum. These emitted free electrons are attracted towards the anode.



In the indirectly heated cathode, no electrical connection is present between the cathode and the heater. Hence, the cathode is not heated directly. The heat energy is supplied to the heater and the heater will transfer its heat energy to the cathode. When the heat energy applied to the cathode is increased to a desired level, the free electrons in the cathode gain sufficient energy and break the bonding with the cathode. The free electrons that break the bonding with the cathode are emitted into the vacuum. These emitted free electrons are attracted towards the anode.

Vacuum diode with forward voltage

When the heat is supplied to the heater, it gains heat energy. This heat energy is transferred to the cathode. When the free electrons in the cathode gains sufficient energy, they breaks the bonding with the cathode and jumps into vacuum. The free electrons in the vacuum need sufficient kinetic energy to reach the anode.

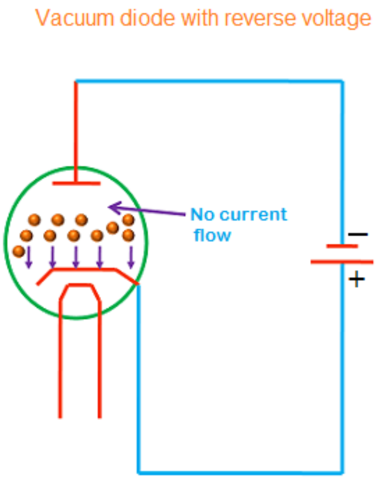
If voltage is applied to the vacuum diode, in such a way, that anode is connected to a positive terminal and cathode is connected to a negative terminal (anode is more positive with respect to the cathode), the free electrons in the vacuum gains enough kinetic energy to reach the anode.

We know that, if two opposite charged particles are placed close to each other they get attracted. In this case, anode is positively charged and free electrons emitted from the cathode are negatively charged. Hence, the free electrons that gain enough kinetic energy will move or attracted towards the anode. These free electrons carry the electric current while moving from cathode to anode.

If the positive voltage applied to plate or anode is increased, the number of free electrons attracted towards the anode is also increased. Thus, the electric current in the vacuum diode increases with increase in the anode or plate voltage.

Vacuum diode with reverse voltage

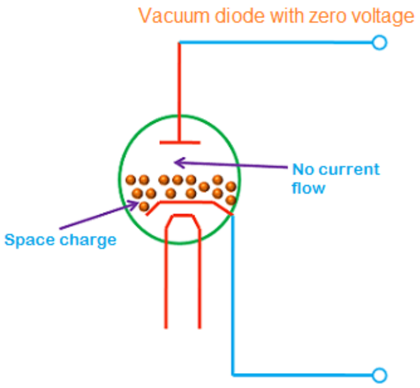
If voltage is applied to the vacuum diode, in such a way, that anode is connected to the negative terminal and cathode is connected to the positive terminal (anode is more negative with respect to cathode), the free electrons in the vacuum gains enough kinetic energy to reach the anode. However, anode repels the free electrons that try to move towards it.



We know that if two like charged particles are placed close to each other they get repelled. In this case, anode is negatively charged and the free electrons emitted from the cathode are also negatively charged. Hence, the anode repels the free electrons that are emitted by the cathode. Therefore, no electric current flows in the vacuum diode.

Vacuum diode with zero voltage

If no voltage is applied to the vacuum diode, anode or plate acts as neutral. It neither attracts nor repels the free electrons emitted from the cathode. Hence, the free electrons emitted from the cathode do not move or attracted towards the anode.



Therefore, no electric current occurs in the vacuum diode. However, the large number of free electrons emitted from the cathode is builds up at one place near the cathode and forms a cloud of free electrons. This cloud of free electrons near the cathode is called space charge.

Conclusion

Therefore, the vacuum diode allows the electric current from cathode to anode and do not allow the electric current from anode to cathode. This one-way direction of the electric current enables the vacuum diode to act like a switch. If the anode or plate is positive with respect to cathode, the vacuum diode acts like a closed switch. On the other hand, if the anode is negative with respect to the cathode, it acts like an open switch.

Vacuum triode

The basic vacuum tube (vacuum diode) is used to convert the alternating current into direct current. However, they cannot amplify the electric signal. In other words, they cannot amplify the voltage or power. To amplify the electrical signal, an extra electrode is required. When the extra electrode is placed between the cathode and anode, the resulting electronic device is called vacuum triode.

The name itself indicates that, it has three electrodes: cathode, anode, and control grid. American electrical engineer Lee De Forest invented the first electronic amplifying device (vacuum triode) in 1906 by adding an extra electrode (control grid) between the cathode and anode. Vacuum triode is a 3-electrode device that amplifies the electrical signal.

Electrodes of vacuum triode

Vacuum triode consists of three electrodes: anode, cathode and control grid. The anode, cathode and control grid are enclosed in an empty glass envelope. The cathode is surrounded by a control grid, which is in turn surrounded by anode. The construction of vacuum triode is similar to vacuum diode. However, vacuum triode contains an extra electrode (control grid).

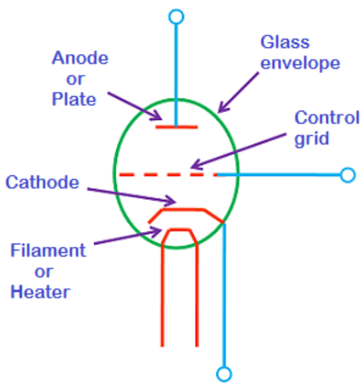


Fig: Symbol of vacuum triode

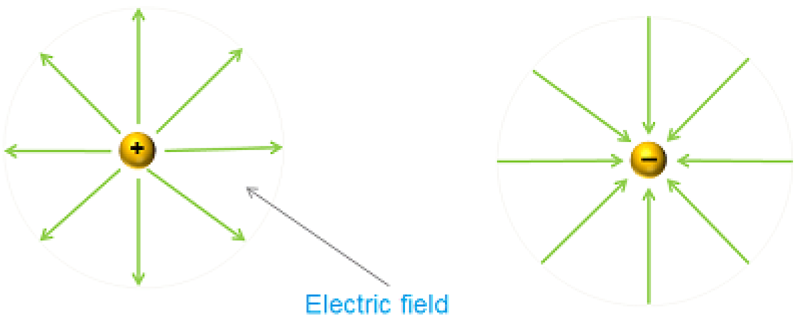
Cathode emits the free electrons when it is heated. Hence, cathode is also called as emitter. The process by which cathode emits the free electrons when it is heated is called thermionic emission. Anode collects the free electrons that are emitted by the cathode. Hence, anode or plate is also called as collector.

In between the anode and cathode, control grid is present. Control grid is placed more nearer to the cathode than anode to increase the electric current efficiently. Control grid will control the flow of electrons between the cathode and anode. Hence, control grid is also called as electron controller or electric current controller.

Control grid is made of network of wires that controls the electrons flow between the cathode and anode. The space between the network of wires in the grid is very large. Hence, the free electrons move easily from cathode to anode through the opening of the control grid. Free electrons that are moving from cathode to anode will carry the electric current.

Electric field

Electric field is the region around a charged particle within which other charged particle will experience a force. Positively charged particles have positive electric field around them whereas negatively charged particles have negative electric field around them.



If two opposite charged particles are placed close to each other, they get attracted. On the other hand, if two like or same charged particles are placed close to each other they get repelled.

In vacuum triode, if positive voltage is applied to the anode or plate, it becomes positively charged. Hence, anode produces positive electric field towards the free electrons. On the other hand, free electrons emitted from the cathode are negatively charged. Hence, free electrons produce negative electric field towards the anode.

The positive electric field of anode has more strength than the negative electric field of free electrons. Hence, free electrons are attracted towards the anode. However, the distance between the anode and cathode is high. Therefore, if small voltage is applied, small number of free electrons is attracted towards the anode.

On the other hand, the distance between the control grid and the cathode is less (control grid is much closer to the cathode than anode). Hence, a small positive voltage applied to the control grid is enough to attract the free electrons. The free electrons that are attracted towards the control grid will easily move towards the anode.

What is meant by electrode?

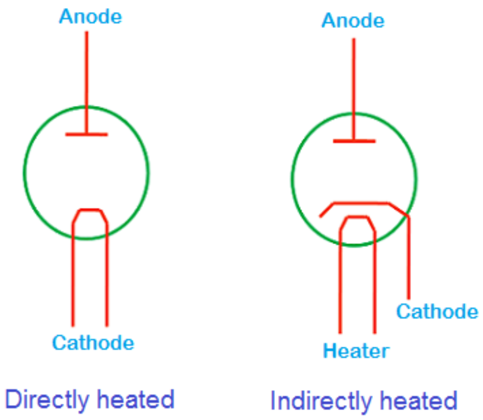
The conductor through which free electrons enter or leaves is called electrode. In vacuum triode, cathode is an electrode, which emits the free electrons. In other words, free electrons leave or go away from cathode and enter into vacuum. Anode is an electrode, which collects the free electrons emitted by the cathode. In other words, free electrons that are emitted by the cathode are entered into plate or anode. Control grid is also called as electrode because, it increases the flow of electrons between the cathode and anode.

Directly and indirectly heated cathode

In the vacuum triode, the cathode is heated to emit the free electrons. This can be done in two ways: by directly heating the cathode or indirectly heating the cathode.

If the heat is supplied directly to the cathode, the cathode is said to be directly heated. In this method, the cathode itself is a heater or heating element or filament. Hence, a small amount of heat energy will provides enough energyfor the free electrons to escape from the cathode.

The free electrons that are escaped from the cathode will enter into vacuum. These free electrons in the vacuum are attracted towards the anode. In the directly heated cathode, the amount of heat energy required to emit the free electrons is less compared to the indirectly heated cathode.



If the heat is supplied indirectly to the cathode, the cathode is said to be indirectly heated. In the indirectly heated cathode, there is no electrical connection between the heater and the cathode.

When the heat is supplied to the heater, it gains heat energy. The heat energy gained by the heater is supplied to the cathode. Thus, heat is indirectly supplied to the cathode. When the free electrons in the cathode gain enough energy in the form of heat, they break the bonding with the cathode and jumps into vacuum.

Electrons emitted from the cathode depends on the amount of heat applied and work function

The number of free electrons escaped from the cathode depends on the amount of heat applied to the cathode and the work function of the cathode

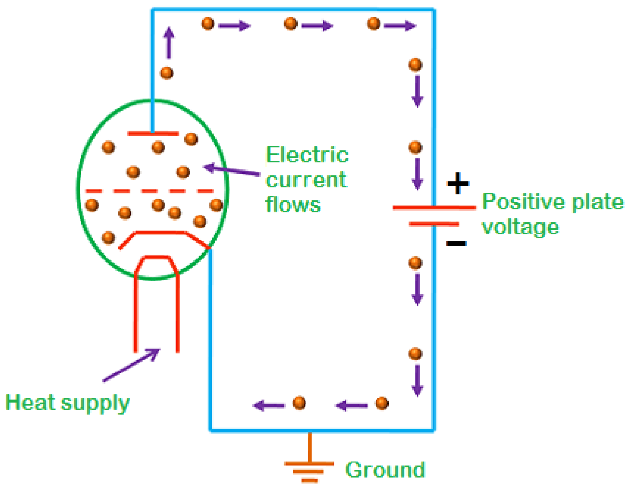
If large amount of heat energy is supplied to the cathode, large number of free electrons is emitted from the cathode. Similarly, if small amount of heat energy is supplied to the cathode, less number of free electrons is emitted from the cathode.

Work function is the minimum amount of heat energy required to remove the free electrons from the metal. Low work function metals require less amount of heat energy to emit the free electrons. On the other hand, high work function metals require large amount of heat energy to emit the free electrons.

<u>Vacuum triode with zero grid voltage</u>

If no voltage is applied to the control grid and positive voltage is applied to the plate, the vacuum triode behaves like normal vacuum diode, because control grid will not shows any effect on the free electrons emitted from the cathode.

Vacuum triode with zero grid voltage



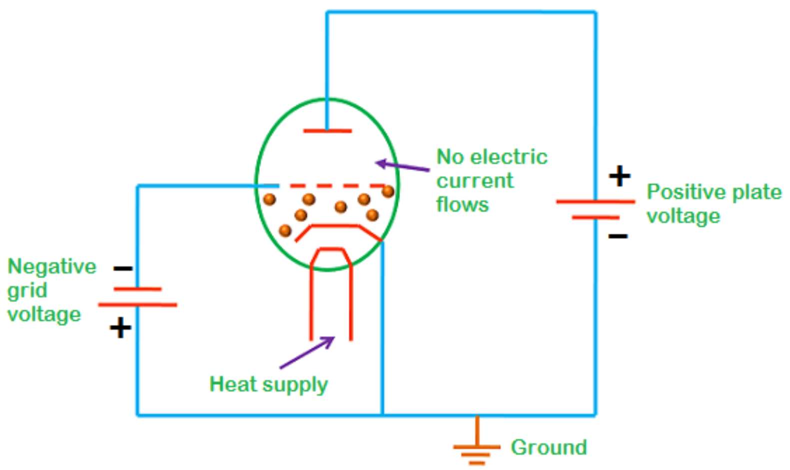
If voltage is applied to the control grid, it produces electric field. In this case, no voltage is applied to the control grid. Hence, control grid will not produce the electric field to attract or repel the free electrons. Therefore, the free electrons emitted from the cathode will easily moves towards the anode or plate from the openings of control grid.

Vacuum triode with negative grid voltage

If negative voltage is applied to the control grid without changing the positive plate voltage, no electric current flows in the vacuum triode, because the control grid opposes or repels the free electrons that try to move towards the anode.

Because of this supply of negative voltage, the control grid becomes negatively charged. Hence, it produces negative electric field. On the other hand, free electrons emitted from the cathode are also negative charged. Hence, free electrons also produce negative electric field.

Vacuum triode with negative grid voltage

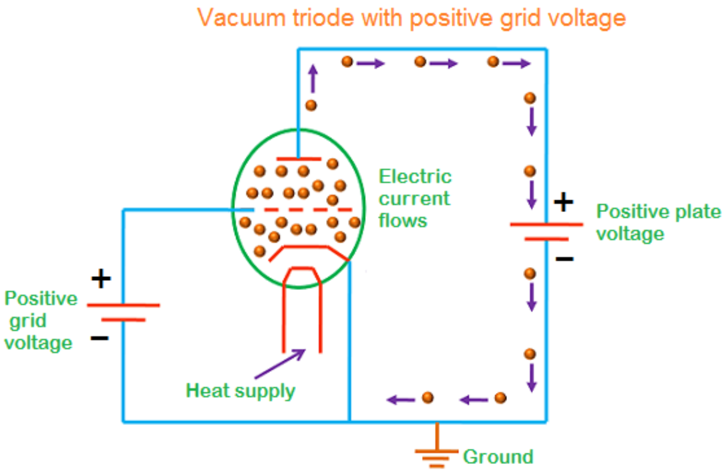


We know that, if two like or same charges are placed close to each other they get repelled. Hence, the control grid opposes or repels the free electrons emitted from the cathode. However, a small number of free electrons overcome the negative electric field of the grid and move towards the anode.

If the negative voltage applied to the control grid is increased, no electrons will move towards the anode. Hence, no electric current flows in the vacuum triode.

Vacuum triode with positive grid voltage

If positive voltage is applied to the control grid without changing the positive plate voltage, electric current flows in the vacuum triode, because the control grid attracts large number of free electrons. The free electrons that are attracted towards the control grid will move easily towards the anode.



If positive voltage is applied to the control grid, it becomes positively charged. Hence, it produces positive electric field towards the free electrons. On the other hand, free electrons emitted from the cathode are negatively charged. Hence, free electrons produce negative electric field towards the control grid.

We know that, if two opposite charged particles are placed close to each other they get attracted. Hence, the control grid attracts the free electrons. The free electrons that are attracted towards the control grid will easily move towards the anode. The free electrons carry the electric current while moving from cathode to anode.

If the positive voltage applied to the control grid is further increased then even more number of free electrons are attracted towards the control grid. Therefore, electric current increases with increase in the grid voltage.

Vacuum tetrode

We use triode to amplify the electrical signal, but at higher frequencies, it will acts as an oscillator rather than as an amplifier. Adding the extra grid (screen grid) between the control grid and the plate or anode reduces the unwanted capacitance between plate and the control grid.

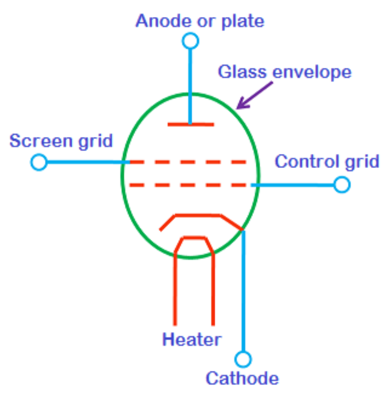
As the name suggests, vacuum tetrode consists of four electrodes: cathode, anode, control grid, and screen grid. The American physicist and electrical engineer Albert Wallace Hull invented the tetrode vacuum tube in 1926.

Electrodes of vacuum tetrode

Vacuum tetrode consists of four electrodes: cathode, anode, control grid, and screen grid. The cathode, anode, control grid, and screen grid are enclosed in an empty glass envelope. The cathode is surrounded by control grid. The control grid is surrounded by the screen grid. The screen grid is surrounded by the anode or plate.

The construction of vacuum tetrode is similar to vacuum triode. However, vacuum tetrode contains an extra electrode called screen grid.

Symbol of vacuum tetrode



Cathode

Cathode is an electrode that emits the free electrons when heated. It is also sometimes referred as emitter. Cathode has more number of negative charges (electrons) than positive charges (protons). Therefore, it is negatively charged.

Anode or plate

Anode is a positively charged electrode that collects the free electrons emitted from the cathode. It is also sometimes referred as collector. Anode has lesser number of negative charges (free electrons) than positive charges (protons). Therefore, it is positively charged.

Control grid

Control grid is placed between the cathode and plate. This grid is placed closer to the cathode than the plate to increase the electric current efficiently. Controls grid present between the cathode and plate controls the flow of electrons. Hence, control grid is also known as electron controller.

Screen grid

Screen grid is placed between the control grid and plate. This grid is placed closer to the control grid than plate to reduce the capacitance efficiently. Screen grid acts as an electrostatic shield to protect the control grid from the positive electric field or influence of the plate when its potential changes.

Main functions of screen grid

The main functions of screen grid are:

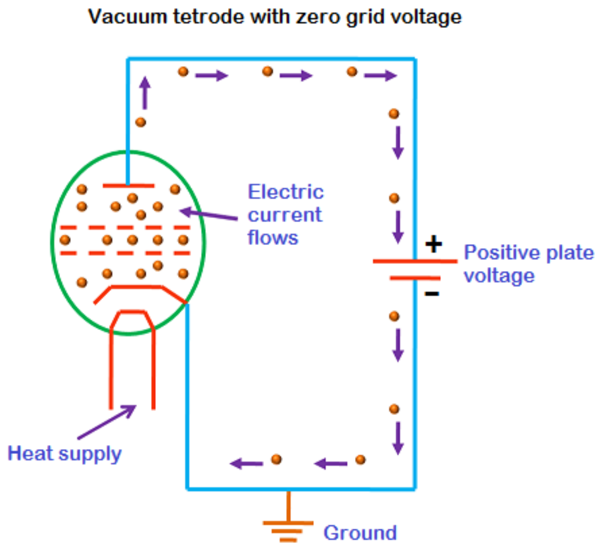
- To accelerate and attract free electrons to the plate or anode
- To reduce the capacitance between the plate and control grid
- To increase the control over electron flow
- To increase the vacuum tube efficiency
- Reduce distortion
- Increase gain

The control grid and screen grid are made of network of wires that controls the electron flow between cathode and anode. The space between the network of wires in the control grid and screen grid is very large. Hence, the free electrons move easily from cathode to anode through the opening of grids.

<u>Vacuum tetrode with zero grid voltage and positive plate voltage</u>

If no voltage is applied to the screen and control grids, and positive voltage is applied to the plate, the vacuum tetrode acts like a vacuum diode. Because control grid and screen grid does not show any effect on the free electrons emitted from the cathode.

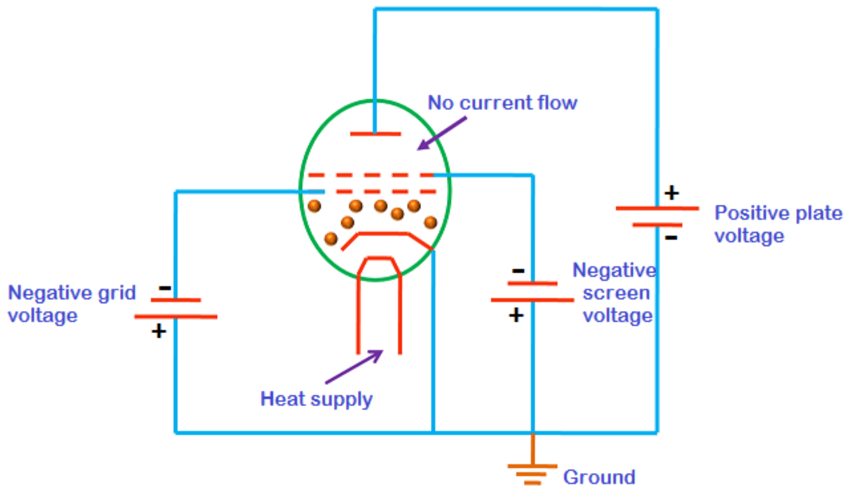
<u>Vacuum tetrode with negative grid</u> voltage and positive plate voltage



If negative voltage is applied to the control grid and screen grid without changing the plate voltage, no electric current flows in the vacuum tetrode. Because control grid and screen grid opposes or repels the free electrons that try to move towards the anode.

Because of the supply of negative voltage, the control grid and screen grid generates a strong negative electric field. On the other hand, free electrons emitted from the cathode are negatively charged. Hence, free electrons also generate negative electric field. However, the electric field generated by the control grid and screen grid is much stronger than the electric field of free electrons.

Vacuum tetrode with negative grid voltage



We know that, when two same charges are placed close to each other, they get repelled. Therefore, because of this strong negative electric field, the electrons emitted from the cathode are repelled by this strong negative electric field.

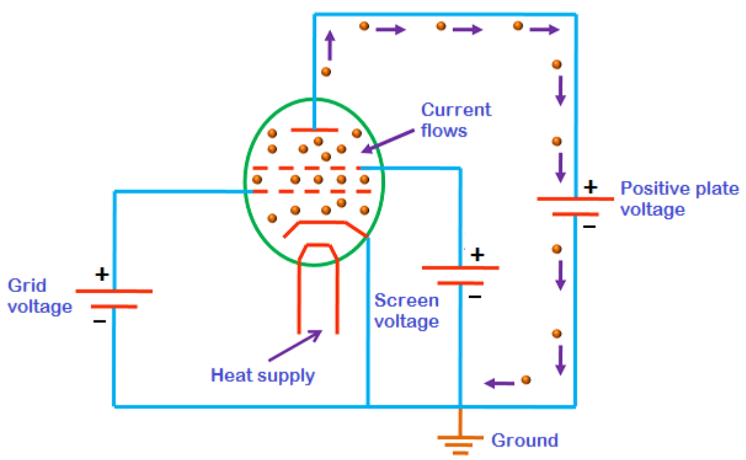
As a result, no electrons flow from cathode to anode. Therefore, no electric current flows through a vacuum tetrode, when negative voltage is applied to the grids.

Vacuum tetrode with positive grid voltage and positive plate voltage

If positive voltage is applied to the control grid and screen grid without changing the positive plate voltage, electric current flows in the vacuum tetrode.

Because of the supply of positively voltage, both control grid and screen grid will become positively charged and generates strong positive electric field.

Vacuum tetrode with positive grid voltage



The positive voltage applied to the screen grid is somewhat lower than the plate voltage. However, the electric field of the screen grid has greater influence on the free electrons. Because of the large space between the wires of the screen grid, most of the free electrons attracted to the screen grid will pass to the plate. Hence, screen grid acts as an electrostatic force pulling electrons from cathode to anode. Therefore, the screen grid acts as the primary anode and plate acts as the secondary anode.

However, some electrons attracted towards the screen grid were observed by the screen gird. As a result, an unwanted screen grid current occurs. However, this electric current is negligible compared to the electric current flow from cathode to anode.

As long as the plate voltage is higher than the screen grid voltage, plate current in the vacuum tetrode highly depends on the screen grid voltage and much less on the plate voltage.

The screen grid is much closer to the cathode than anode or plate. Therefore, the electric field produced by the screen grid has greater influence on the cathode space chare (free electrons).

If positive voltage applied to the screen grid is increased to a value higher than the plate voltage, the electrons transmitted to anode or plate will return towards the screen grid. This is because, if we increase the screen grid voltage to a value higher than the plate voltage, screen grid generates stronger electric field than the plate electric field. As a result, electrons from the plate are attracted towards the screen grid. As a result, a reverse current flows in the vacuum tetrode, which is not desirable.

Many large TV stations, radios, and industries use giant power tetrodes which works efficiently when used as RF power amplifiers.

Secondary electron emission in tetrode

The secondary electron emission is the emission of secondary electrons from the metal surface when primary electrons strike the metal surface at high speed.

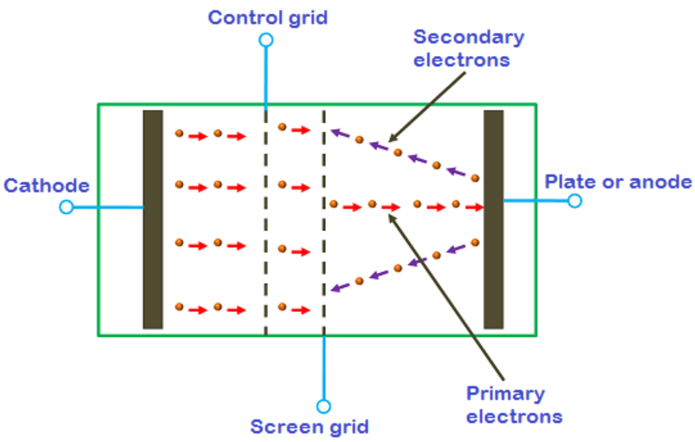
In tetrode, the secondary electron emission occurs mainly at anode or plate. When the electrons strike the anode or plate at high speed, some secondary electrons are emitted from the plate surface. What happens to these secondary electrons emitted from plate surface is depends on voltage applied across screen grid and plate.

Case 1: If voltage applied across screen grid is greater than plate voltage

In this case, screen grid has higher potential than plate. Hence, screen grid produces greater electric field than plate.

When primary electrons strike the plate at high speed, they transfer their kinetic energy to the electrons in the plate. As a result, the electrons in the plate gain enough energy and overcome the strong attractive force of the plate and escapes from the plate surface.

Secondary electrons when gird voltage > plate voltage



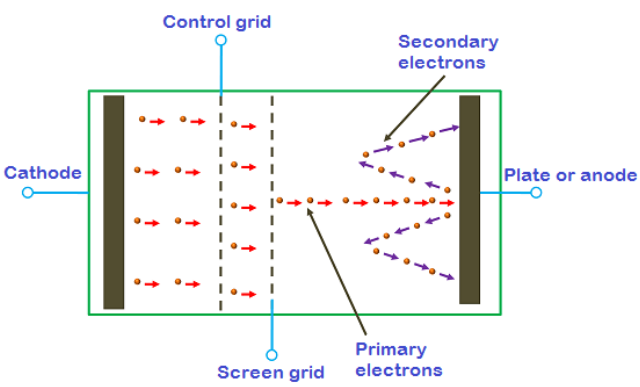
The electrons escaped from the plate surface experience a greater attractive force from the screen grid than the plate. As a result, the secondary electrons emitted from the plate surface are attracted to the screen grid. Therefore, electric current flows in reverse direction (from plate to screen grid), which is not desirable.

Case 2: If voltage applied across plate is greater than screen grid

In this case, plate or anode has higher potential than screen grid. Hence, plate produces greater electric field than screen grid.

When electrons strike plate at high speed, some secondary electrons are emitted from the plate surface. The electrons emitted from the plate surface experience greater attractive force from the plate than the screen grid. As a result, the secondary electrons emitted from the plate surface return to anode. Therefore, electric current flows in forward direction (from screen grid to anode).

Secondary electrons when grid voltage < plate voltage



We always need to keep the plate potential higher than the grids potential (screen grid and control grid), to efficiently increase the electric current.

Vacuum pentode

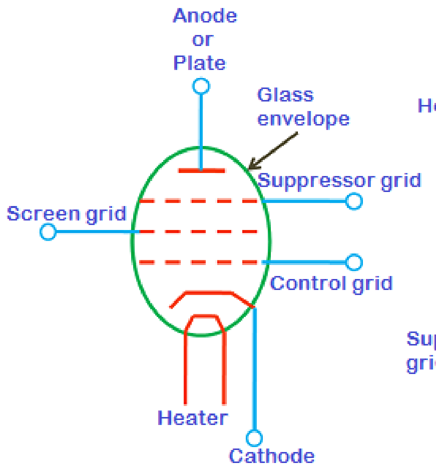
We use screen grid in tetrode to reduce capacitance between the control grid and plate (anode). However, tetrodes have one drawback. When the screen grid voltage is greater than the plate voltage, the secondary electrons emitted from the plate are attracted to the screen grid. Because of this, the electric current flows in reverse direction (from plate to screen grid) which is undesirable.

This drawback can be overcome by placing an extra grid called suppressor grid in between screen grid and the plate. The suppressor grid repels secondary electrons towards anode or plate.

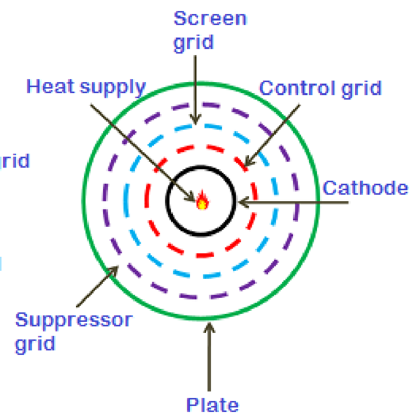
As the name suggests, vacuum pentode consists of five electrodes: cathode, control grid, screen grid, suppressor grid, and plate or anode. Bernard D. H. Tellegen invented the pentode in 1926.

Electrodes of vacuum pentode

The pentode is made of evacuated glass envelope containing 5 electrodes. The air inside the glass envelope is removed completely. The 5 electrodes of the pentode include cathode, control grid, screen grid, suppressor grid, and plate.



Symbol of pentode



Circular view of pentode

The cathode is surrounded by control grid. The control grid is surrounded by screen grid. The screen grid is surrounded by suppressor grid. The suppressor grid is surrounded by plate or anode.

The construction of vacuum pentode is similar to vacuum tetrode. However, vacuum pentode contains an extra grid (suppressor grid).

<u>Cathode</u>

Cathode is a negatively charged electrode, which emits free electrons when heat energy is supplied. It has more number of negative charges (electrons) than positive charges (protons). Therefore, it is negatively charged. Cathode is also sometimes referred as emitter.

Plate or anode

Anode is a positively charged electrode, which collects the free electrons emitted by the cathode. It has lesser number of negative charges (electrons) than positive charges (protons). Therefore, it is positively charged. Anode is also sometimes referred as collector.

Control grid

As the name suggests, it is used to control the flow of electrons. Control grid is placed between anode and cathode. This grid is placed closer to the cathode than anode to increase the flow of electric current efficiently. Control grid is also sometimes referred as electron controller.

Screen grid

Screen grid is mainly used to reduce the capacitance between control grid and anode, and to increase the velocity of free electrons. Screen grid is a positively charged electrode placed in between control grid and plate or anode. It is placed closer to the control grid than anode to reduce the capacitance efficiently.

Uses of screen grid

The screen grid is mainly used to:

- Reduce the capacitance between control grid and plate.
- Increase gain
- Increase control over electron flow
- Increase velocity of free electrons.
- Reduce distortion

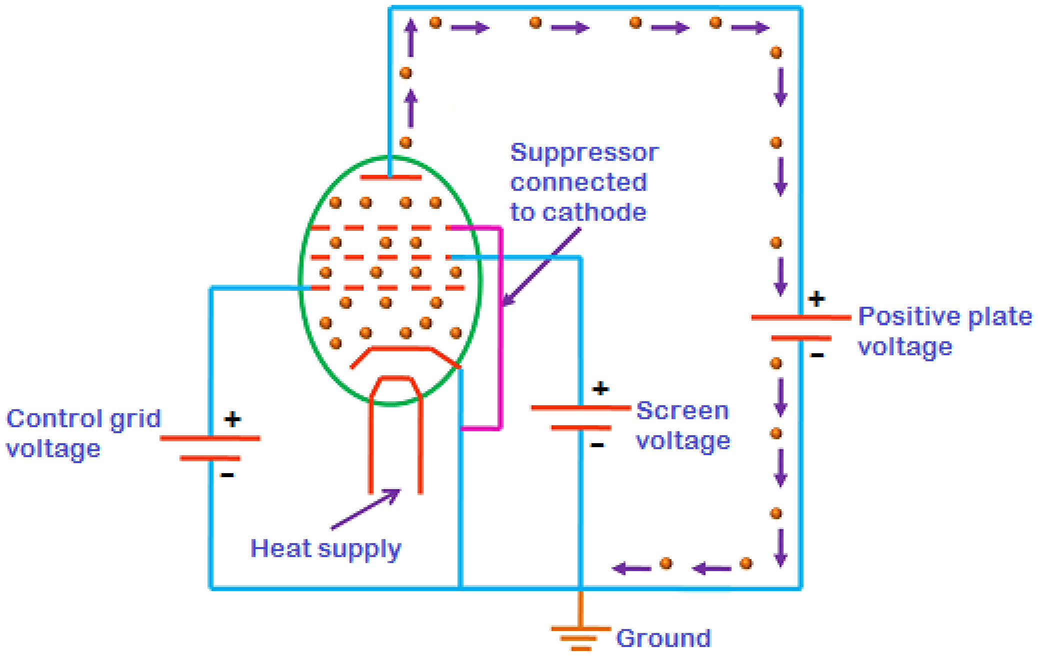
Suppressor grid

The suppressor grid is placed between the screen grid and plate. It is mainly used to repel the secondary electrons (emitted from anode) back to the anode. It is internally connected to the cathode. Therefore, a large number of free electrons are transmitted directly to the suppressor grid. Because of the gaining of excess electrons, suppressor grid acts as a negatively charged electrode and generates negative electric field. Therefore, it repels the secondary electrons (emitted from anode) back to anode.

Voltage applied across vacuum pentode

When positive voltage is applied to the control grid, screen grid and anode or plate, electric current starts flowing in the vacuum pentode. The suppressor grid is connected internally to the cathode.

The screen grid and control grid are made of network of wires that controls the electron flow between cathode and anode. The space between the network of wires in the screen grid and control grid is very large. Hence, the free electrons emitted from the cathode flows easily through the network of wires and attracted to the plate.



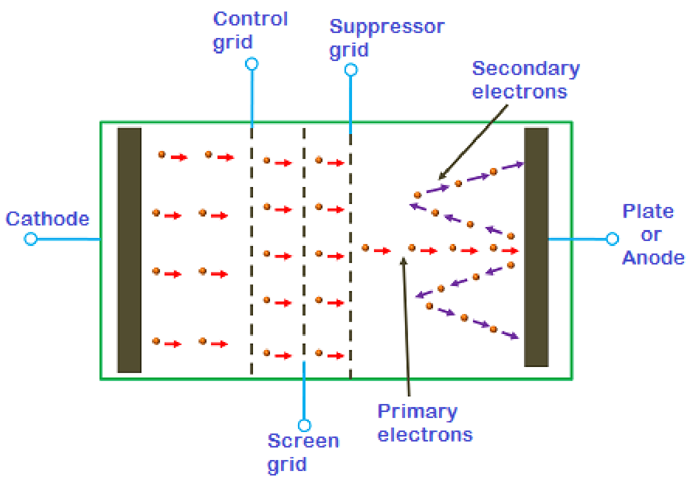
Because of the positive supply voltage to the control grid and screen grid, a large number of free electrons are pulled away from the grids towards the positive terminal. As a result, control grid and screen grid loses large number of electrons. Therefore, they become positively charged and generates strong positive electric field.

The electric field generated by the control grid and screen grid has greater influence on the free electrons than plate electric field. This is because, control grid and screen grid are closer to the cathode.

Because of the large space between the wires of screen grid, most of the free electrons attracted to the screen grid, will pass easily through the wires and reaches anode. However, some electrons attracted to the screen grid were observed by the wires of a screen grid. As a result, a loss of electric current occurs. However, this electric current is negligible compared to the electric current from cathode to anode.

The free electrons emitted from cathode travel at high speed. Therefore, when the electrons emitted from the cathode strike the plate surface at high speed, they transfer their kinetic energy to the valence electrons in the plate. As a result, the valence electrons in the plate gains sufficient energy and overcome the strong attractive force of the plate and escapes or emitted from the plate surface. These emitted electrons from the plate surface are called secondary electrons and the electrons, which hit the plate surface, are called primary electrons.

We know that suppressor grid is connected directly to the cathode. Therefore, the suppressor grid is also negatively charged. This negatively charged suppressor grid generates a negative electric field. The negative electric field generated by the suppressor grid is weaker than the anode electric field. However, this electric field is enough to control the secondary electrons emitted from anode or plate.



The secondary electrons emitted from the plate surface experience a repulsive force from the suppressor grid and attractive force from the plate. As a result, the secondary electrons return to the plate or anode.

Thus, the suppressor grid reduce the unwanted secondary electron current from plate to screen grid.

Applications of pentode

Pentodes are widely used in radios and televisions until 1960s. After 1960s they were replaced by transistors. However, they continued to be used in some applications such as electric guitar amplifiers, microphone preamplifiers, high-power radio transmitters, and professional audio applications.

Advantages of pentode

Pentodes are able to operate at high frequencies

Pentodes have high amplification factor than tetrodes

Metallic bond

What is metallic bond?

Metallic bonding is a type of chemical bonding formed in the metals. Metals are good conductors of heat and electricity.

In metals, each atom has 8 or 12 neighboring atoms surrounding them and the valence electrons in the metal atoms are less than four. Hence, it is not possible for the metal atoms to form 8 or 12 covalent bonds with the neighboring atoms because they do not have such large number of valence electrons to form covalent bonding with the 8 or 12 neighboring atoms. Thus, metal atoms cannot form a covalent bond with the neighboring atoms.

The atoms in the metals are stick together because of a special type of bonding known as metallic bonding. The concept of metallic bonding was first proposed in 1900 by Paul Drude, which is known as Drude model or electron sea model or free electron model. According to the free electron model, the metal atoms have low electro-negativity and low ionization energy.

Electro-negativity and ionization energy

Electro-negativity is the measure of attraction between the nucleus of an atom and the shared electrons. In metals, the attraction between the nucleus and the valence electrons is less.

Metal atoms contain only one, two, or three valence electrons. They do not have large number of valence electrons to form covalent bond with the neighboring atoms. Hence, the valence electron of one atom is not shared with another atom.

Ionization energy is the amount of energy required to remove an electron from the atom. In the metals, valence electrons or outermost orbit electrons are loosely held by the nucleus of an atom. Hence, a small amount of energy removes the valence electrons from the parent atom.

Metals at low temperature

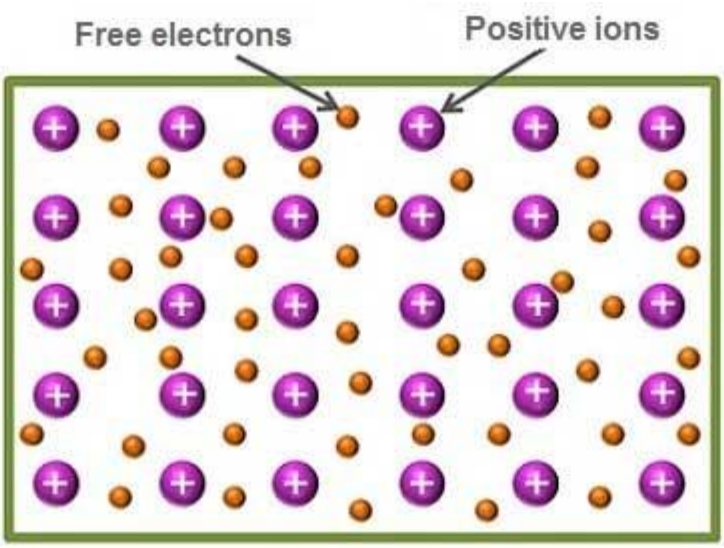
When a small amount of energy in the form of heat or light is applied to the metal atoms, the valence electrons gain enough energy to break the bonding with the metal atoms and then jumps into conduction band. The conduction band electrons move freely from one place to another place. Hence, they are called free electrons. Thus, a small amount of energy or heat generates free electrons in the metals.

Metals at room temperature

At room temperature, most of the valence electrons in the metals gain enough energy in the form of heat to break the bonding with the atom and then jumps into the conduction band. This will results in large number of electrons in conduction band. These electrons move freely from one place to another place within the conduction band. The large number of free electrons in the conduction band is called as sea of electrons or electron cloud.

Generally, atoms are electrically neutral (have equal number of electrons and protons). However, when the atom loses or gains a valence electron, it will become a charged atom or ion (has unequal number of electrons and protons). If the atom loses an electron, it will become a positively charged atom or ion (has lesser number of electrons than protons). Positive ions have more number of protons (positive charge carriers) than electrons (negative charge carriers). Hence, the positive charge carriers (protons) will creates more electric charge or electric field around the nucleus than negative charge carriers (electrons) do. The positive ions are also called as positively charged kernels.

At room temperature, large number of atoms loses most of their valence electrons, because of this, large number of positively charged atoms or positive ions are produced. The large number of positive ions creates a huge positive electric field around them. We know that electron itself is a negatively charged particle. In metals, large number of free electrons is present. Hence, the sea of free electrons or large number of free electrons in the metals will create a negative charge or negative electric field around them.



In metals, the negative charge or negative electric field around each free electron is not only belongs to a single positive ion, it belongs to all the positive ions. Hence, the electric field around each free electron is shared with all the positive ions. Similarly, the positive charge or positive electric field around a positive ion is belongs all the free electrons. Hence, the electric field around the each positive ion is shared with all the free electrons. Likewise, the sea of free electrons and the group of positive ions will share their electric field or charge.

We know that opposite charges attract each other. The sea of free electrons has negative charge or negative electric field around them where as the positive ions has positive charge or positive electric field around them. Hence, the electrostatic force of attraction between the sea of free electrons and the group of positive ions makes them stick together.

The electrostatic force of attraction between the positively charged atoms or positive ions and the sea of free electrons or conduction electrons, which holds them together, is called metallic bond.

Properties of metallic bonds

Good conductors of heat and electricity: The Sea of charge carriers in the metals easily carries the charge or heat energy from one place to another place. These charge carriers are responsible for electric current in the metals. Hence, metals are called as good conductors of heat and electricity.

High melting and boiling points: In metals, the electrostatic force of attraction between the sea of free electrons and the positive ions is very strong. Therefore, it is very difficult to break the bonding between them. Lot of energy is required to break the bonding between them. Thus, it is very difficult to melt or vaporize the metals.

Valence electrons: Metals have only one, two or three electrons in the outer most orbit of an atom. Hence, they cannot form covalent bonding with the neighboring atoms.

Smooth surface: The surface of the metal is very smooth. Hence, the light cannot penetrate through the metal. The light that falls on the metals is reflected.

Electron volt

The Joule is a unit of energy in the SI system. Energy is measured in joules. However, joule is very large unit while dealing with electrons and other tiny particles. Hence, electron volt is used while dealing with smaller particles like electrons.

Electron volt is defined as the amount of energy one electron gains by moving through a potential difference of one volt. Hence, one electron volt is equal to elementary charge (1.60217657 × 10-19 coulombs (C)) multiplied by one volt. It can be written as

$$ev = (1.60217657 \times 10\text{-}19\ C)\ (1V)$$

$$\text{Therefore, } 1ev = 1.60217657 \times 10\text{-}19\ J.$$

Energy band theory in solids

In a single isolated atom, the electrons in each orbit have definite energy associated with it. But in case of solids all the atoms are close to each other, so the energy levels of outermost orbit electrons are affected by the neighboring atoms.

When two single or isolated atoms are bring close to each other then the outermost orbit electrons of two atoms are interact or shared with each other. i.e, the electrons in the outermost orbit of one atom experience a attractive force from the nearest or neighboring atomic nucleus. Due to this the energies of the electrons will not be in same level, the energy levels of electrons are changed to a value which is higher or lower than that of the original energy level of the electron.

The electrons in same orbit exhibits different energy levels. The grouping of this different energy levels is called energy band.

However, the energy levels of inner orbit electrons are not much affected by the presence of neighboring atoms.

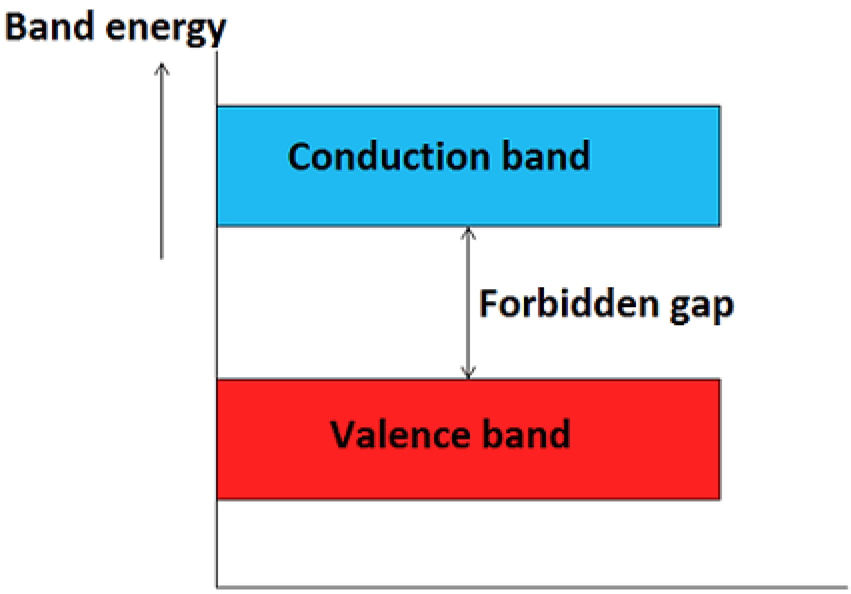
Important energy bands in solids

There are number of energy bands in solids but three of them are very important. These three energy bands are important to understand the behavior of solids. These energy bands are

Valence band

Conduction band

Forbidden band or forbidden gap



Valence band

The energy band which is formed by grouping the range of energy levels of the valence electrons or outermost orbit electrons is called as valence band.

Valence band is present below the conduction band as shown in figure. Electrons in the valence band have lower energy than the electrons in conduction band.

The electrons present in the valence band are loosely bound to the nucleus of atom.

Conduction band

The energy band which is formed by grouping the range of energy levels of the free electrons is called as conduction band.

Generally, the conduction band is empty but when external energy is applied the electrons in the valence band jumps in to the conduction band and becomes free electrons. Electrons in the conduction band have higher energy than the electrons in valence band.

The conduction band electrons are not bound to the nucleus of atom.

Forbidden gap

The energy gap which is present between the valence band and conduction band by separating these two energy bands is called as forbidden band or forbidden gap.

In solids, electrons cannot stay in forbidden gap because there is no allowed energy state in this region. Forbidden gap is the major factor for determining the electrical conductivity of a solid. The classification of materials as insulators, conductors and semiconductors is mainly depends on forbidden gap.

The energy associated with forbidden band is called energy gap and it is measured in unit electron volt (eV).

1 eV = 1.6 × 10-19 J

The applied external energy in the form of heat or light must be equal to to the forbidden gap in order to push an electron from valence band to the conduction band.

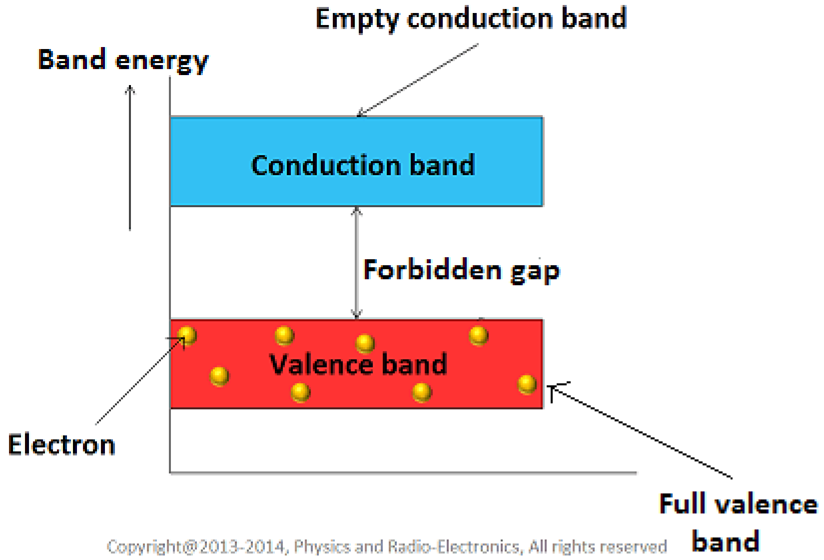
Classification of materials based on forbidden gap

Forbidden gap plays a major role for determining the electrical conductivity of material. Based on the forbidden gap materials are classified in to three types, they are

- Insulators
- Conductors
- semiconductors

Insulators

The materials which does not allow the flow of electric current through them are called as insulators. Insulators are also called as poor conductors of electricity.



Normally, in insulators the valence band is fully occupied with electrons due to sharing of outer most orbit electrons with the neighboring atoms. Where as conduction band is empty, I.e, no electrons are present in conduction band.

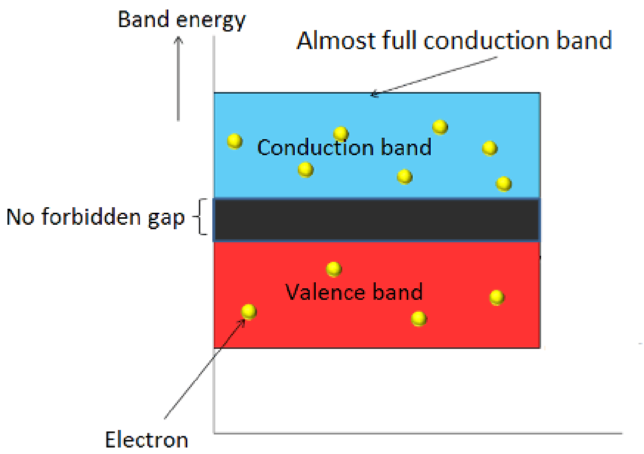
The forbidden gap between the valence band and conduction band is very large in insulators. The energy gap of insulator is approximately equal to 15 electron volts (eV).

The electrons in valence band cannot move because they are locked up between the atoms. In order move the valence band electrons in to conduction band large amount of external energy is applied which is equal to the forbidden gap. But in insulators, this is practically impossible to move the valence band electrons in to conduction band.

Rubber, wood, diamond, plastic are some examples of insulators. Insulators such as plastics are used for coating of electrical wires. These insulators prevent the flow of electricity to unwanted points and protect us from electric shocks.

Conductors

The materials which easily allow the flow of electric current through them are called as conductors. Metals such as copper, silver, iron, aluminum etc. are good conductors of electricity.



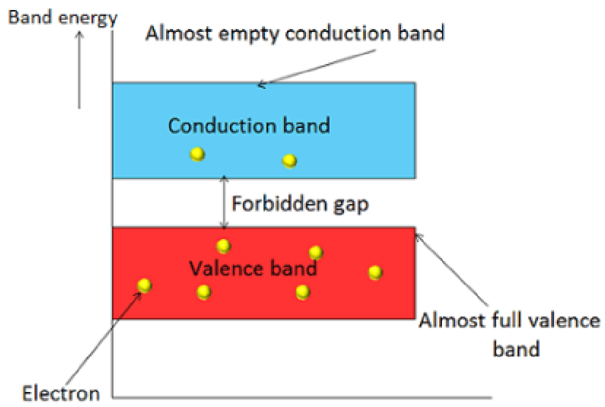
In a conductor, valence band and conduction band overlap each other as shown in figure. Therefore, there is no forbidden gap in a conductor.

A small amount of applied external energy provides enough energy for the valence band electrons to move in to conduction band. Therefore, more number of valence band electrons can easily moves in to the conduction band.

When valence band electrons moves to conduction band they becomes free electrons. The electrons present in the conduction band are not attached to the nucleus of a atom. In conductors, large number of electrons are present in conduction band at room temperature, I.e, conduction band is almost full with electrons. Where as valence band is partially occupied with electrons. The electrons present in the conduction band moves freely by carrying the electric current from one point to other.

Semiconductors

The material which has electrical conductivity between that of a conductor and an insulator is called as semiconductor. Silicon, germanium and graphite are some examples of semiconductors.



In semiconductors, the forbidden gap between valence band and conduction band is very small. It has a forbidden gap of about 1 electron volt (eV).

At low temperature, the valence band is completely occupied with electrons and conduction band is empty because the electrons in the valence band does not have enough energy to move in to conduction band. Therefore, semiconductor behaves as an insulator at low temperature.

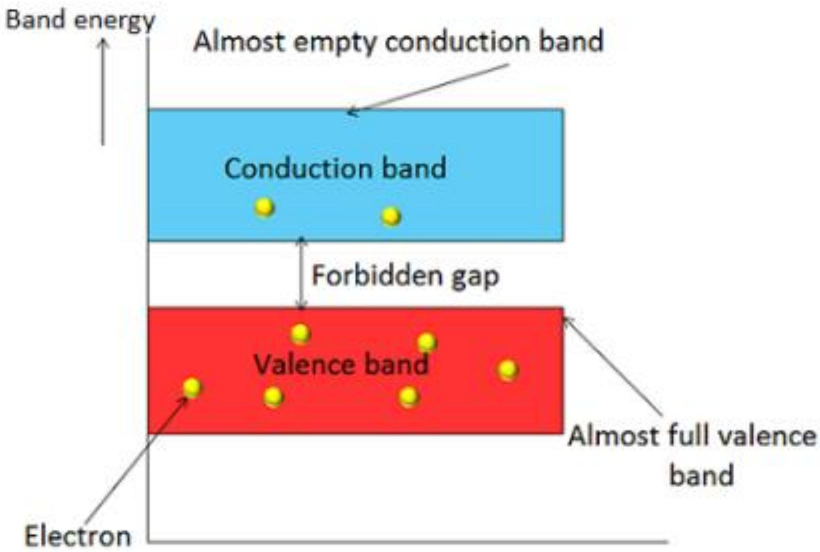
However, at room temperature some of the electrons in valence band gains enough energy in the form of heat and moves in to conduction band.

When the temperature is goes on increasing, the number of valence band electrons moving in to conduction band is also increases. This shows that electrical conductivity of the semiconductor increases with increase in temperature. I.e. a semiconductor has negative temperature co-efficient of resistance.

The resistance of semiconductor decreases with increase in temperature.

Introduction to semiconductor

The material which has electrical conductivity between that of a conductor and that of an insulator is called assemiconductor. Silicon, germanium and graphite are some examples of semiconductors. Semiconductors are the foundation of modern electronics, including transistors, Light-Emitting diodes, solar cells etc.



In semiconductors, the forbidden gap between valence band and conduction band is very small. It has a forbidden gap of about 1 electron volt (eV).

At low temperature, the valence band is completely occupied with electrons and conduction band is empty because the electrons in the valence band does not have enough energy to move in to conduction band. Therefore, semiconductor behaves as an insulator at low temperature.

However, at room temperature some of the electrons in valence band gains enough energy in the form of heat and moves in to conduction band. When the valence electrons moves in to conduction band they becomes free electrons. These electrons are not attached to the nucleus of a atom, So they moves freely.

The conduction band electrons are responsible for electrical conductivity. The measure of ability to conduct electric current is called as electrical conductivity.

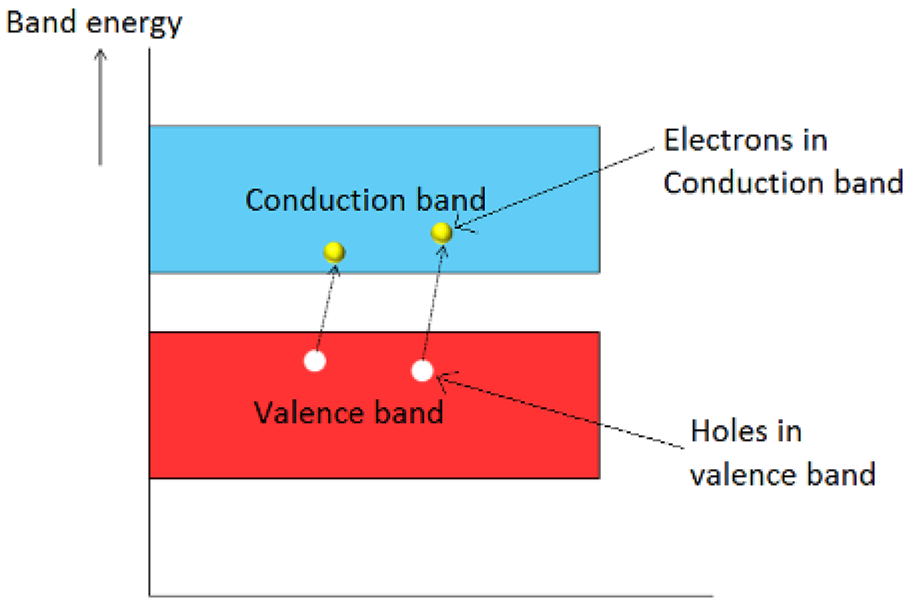
When the temperature is goes on increasing, the number of valence band electrons moving in to conduction band is also increases. This shows that electrical conductivity of the semiconductor increases with increase in temperature. i.e. a semiconductor has negative temperature co-efficient of resistance. The resistance of semiconductor decreases with increase in temperature.

In semiconductors, electric current is carried by two types of charge carriers they are electrons and holes.

Hole

The absence of electron in a particular place in an atom is called as hole.

Hole is a electric charge carrier which has positive charge. The electric charge of hole is equal to electric charge of electron but have opposite polarity.



When a small amount of external energy is applied, then the electrons in the valence band moves in to conduction band and leaves a vacancy in valence band. This vacancy is called as hole.

Intrinsic semiconductor

Introduction

Pure semiconductors are called intrinsic semiconductors. Silicon and germanium are the most common examples of intrinsic semiconductors. Both these semiconductors are most frequently used in the manufacturing of transistors, diodes and other electronic components.

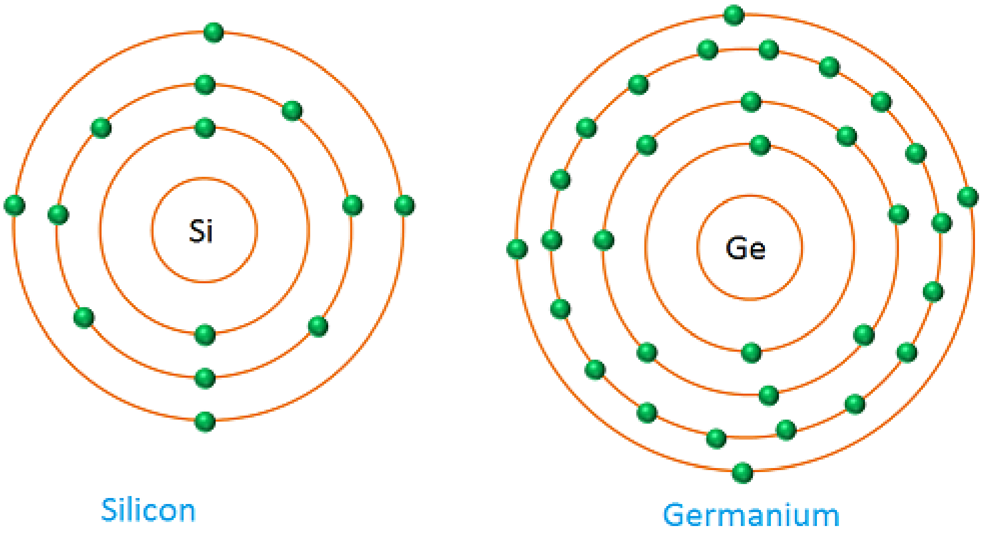
Intrinsic semiconductor is also called as undoped semiconductor or I-type semiconductor. In intrinsic semiconductor the number of electrons in the conduction band is equal to the number of holes in the valence band. Therefore the overall electric charge of a atom is neutral.

Atomic structure of silicon and germanium

The atomic structure of intrinsic semiconductor materials like silicon and germanium is as follows.

Atomic structure of silicon

Silicon is a substance consisting of atoms which all have the same number of protons. The atomic number of silicon is 14 i.e. 14 protons. The number of protons in the nucleus of an atom is called atomic number. Silicon atom has 14 electrons (two electrons in first orbit, eight electrons in second orbit and 4 electrons in the outermost orbit).



Atomic structure of germanium

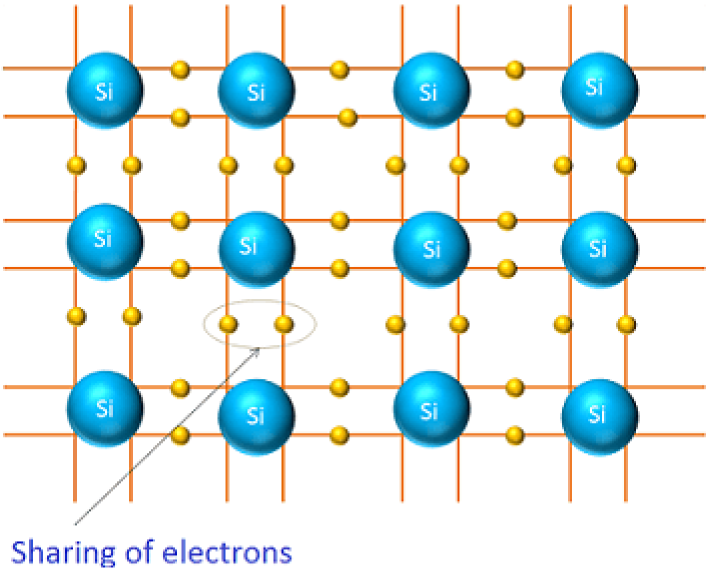
Germanium is a substance consisting of atoms which all have the same number of protons. The atomic number of germanium is 32 i.e. 32 protons. The number of protons in the nucleus of atom is called atomic number. Germanium has 32 electrons (2 electrons in first orbit, 8 electrons in second orbit, 18 electrons in third orbit and 4 electrons in the outermost orbit.

Intrinsic semiconductor

Covalent bonding in silicon and germanium

Covalent bonding in silicon

The outermost shell of atom is capable to hold up to eight electrons. The atom which has eight electrons in the outermost orbit is said to be completely filled and most stable. But the outermost orbit of silicon has only four electrons. Silicon atom needs four more electrons to become most stable. Silicon atom forms four covalent bonds with the four neighboring atoms. In covalent bonding each valence electron is shared by two atoms.

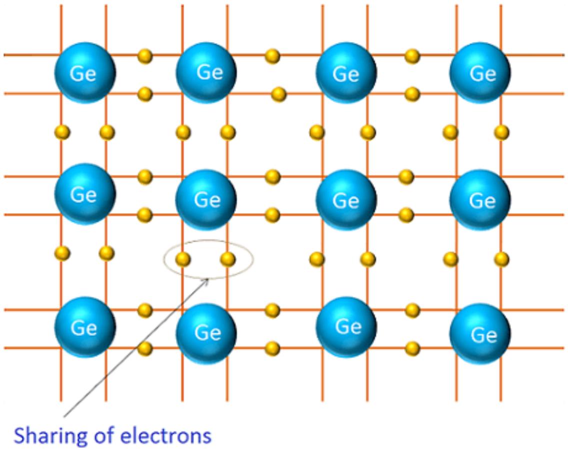


When silicon atoms comes close to each other, each valence electron of atom is shared with the neighboring atom and each valence electron of neighboring atom is shared with this atom. Likewise each atom will share four valence electrons with the four neighboring atoms and four neighboring atoms will share each valence electron with this atom. Therefore, total eight electrons are shared.

Covalent bonding in germanium

The outermost orbit of germanium has only four electrons. Germanium atom needs four more electrons to become most stable. Germanium atom forms four covalent bonds with the four neighboring atoms. In covalent bonding each valence electron is shared by two atoms.

When germanium atoms comes close to each other each valence electron of atom is shared with the neighboring atom and each valence electron of neighboring atom is shared with this atom. Likewise each atom will share four valence electrons with the four neighboring atoms and four neighboring atoms will share each valence electron with this atom. Therefore, total eight electrons are shared.



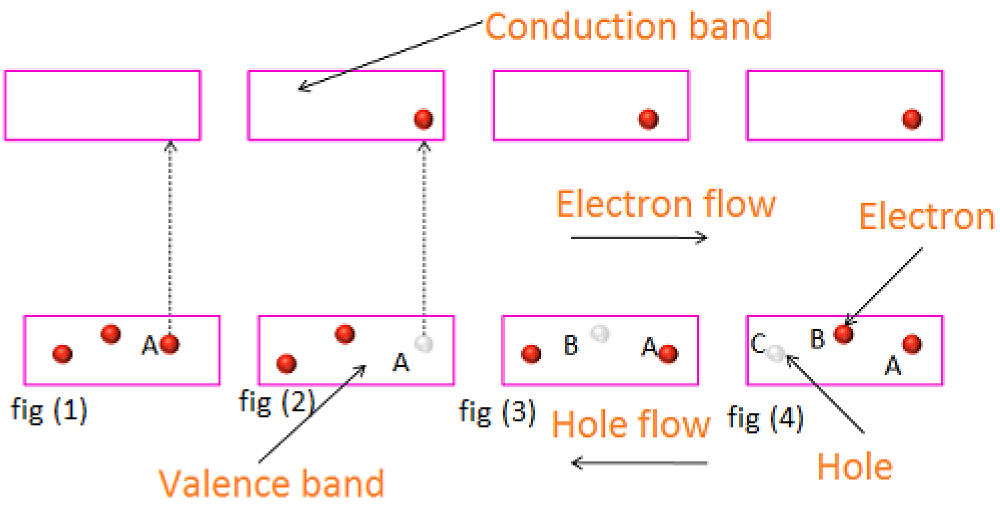
The outermost shell of silicon and germanium is completely filled and valence electrons are tightly bound to the nucleus of atom because of sharing electrons with neighboring atoms. In intrinsic semiconductors free electrons are not present at absolute zero temperature. Therefore intrinsic semiconductor behaves as perfect insulator.

Electron and hole current

In conductors current is caused by only motion of electrons but in semiconductors current is caused by both electrons in conduction band and holes in valence band. Current that is caused by electron motion is called electron current and current that is caused by hole motion is called hole current. Electron is a negative charge carrier whereas hole is a positive charge carrier.

At absolute zero temperature intrinsic semiconductor behaves as insulator. However, at room temperature the electrons present in the outermost orbit absorb thermal energy. When the outermost orbit electrons get enough energy then they will break bonding with the nucleus of atom and jumps in to conduction band. The electrons present in conduction band are not attached to the nucleus of an atom so they are free to move.

When the valence electron moves from valence band to the conduction band a vacancy is created in the valence band where electron left. Such vacancy is called hole.

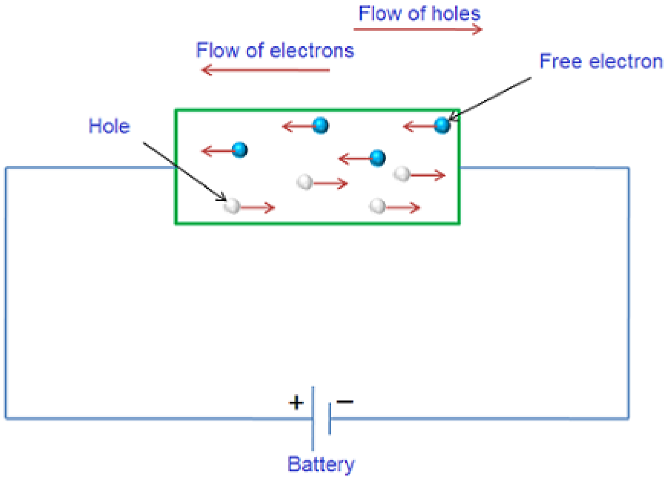


Let's take an example, as shown in fig there are three atoms atom A, atom B and atom C. At room temperature valence electron in an atom A gains enough energy and jumps in to conduction band as show in fig (1). When it jumps in to conduction band a hole (vacancy) is created in the valence band at atom A as shown in fig (2). Then the neighboring electron from atom B moves to atom A to fill the hole at atom A. This creates a hole at atom B as shown in fig (3). Similarly neighboring electron from atom C moves to atom B to fill the hole at atom B. This creates a hole at atom C as shown in fig (4). Likewise electrons moves from left side to right side and holes moves from right to left side.

Conduction in intrinsic semiconductor

The process of conduction in intrinsic semiconductor is shown in below fig. In the below fig, an intrinsic semiconductor is connected to a battery.

Here, positive terminal of battery is connected to one side and negative terminal of the battery is connected to other side. As we know like charges repel each other and opposite charges attract each other. In the similar way negative charge carriers (electrons) are attracted towards the positive terminal of battery and positive charge carriers (holes) attracted towards the negative terminal of battery.



Electrons will experience a attractive force from the positive terminal, so they move towards the positive terminal of the battery by carrying the electric current. Similarly holes will experience a attractive force from the negative terminal, so they moves towards the negative terminal of the battery by carrying the electric current. Thus, in a semiconductor electric current is carried by both electrons and holes.

In intrinsic semiconductor the number of free electrons in conduction band is equal to the number of holes in valence band. The current caused by electrons and holes is equal in magnitude.

The total current in intrinsic semiconductor is the sum of hole and electron current.

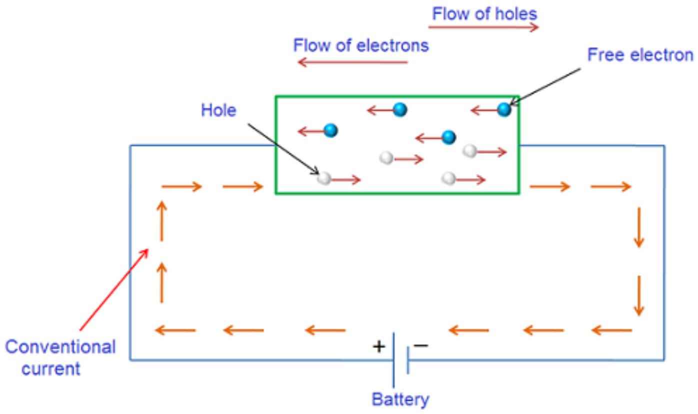
Total current = Electron current + Hole current

I = Ihole+ Ielectron

Conventional current

The electric current that flows from positive terminal of battery to the negative terminal of battery is called conventional current. The conventional current direction is in the same direction of flow of holes but opposite to the direction of flow of free electrons.

When Ben Franklin started experimenting with the electricity, he assumed that electric current (positive charge carriers) flow from positive to negative. But later world realized that it was wrong. Actually electric current flows from negative terminal to positive terminal of battery.



The flow of charge carriers is called current. Here charge carriers are protons or electrons. But specifically current is carried by electrons not protons because protons are strongly bound to the nucleus of an atom because of strong nuclear force. So electrons those are loosely bounded to the nucleus of an atom break bonding with the parent atom and become free. The electrons that are not bound to atoms flow freely and constitute current. Electrons flow from negative terminal to positive terminal of battery.

Majority of people in the world are following the conventional current direction. In some text books current direction is written in opposite to the conventional current direction i.e. considering electron flow as the direction of current. But majority of the textbooks uses conventional current direction. Most of the electronic devices are analyzed using this conventional current direction.

It is important to understand the difference between conventional current direction and electron flow current direction. We can follow any one. It does not matter whether we are using conventional current direction or electron flow direction. Both of them will give the same result. We can design a circuit by using any one of them.

Intrinsic carrier concentration

In intrinsic semiconductor, when the valence electrons broke the covalent bond and jumps into the conduction band, two types of charge carriers gets generated. They are free electrons and holes.

The number of electrons per unit volume in the conduction band or the number of holes per unit volume in the valence band is called intrinsic carrier concentration. The number of electrons per unit volume in the conduction band is called electron-carrier concentration and the number of holes per unit volume in the valence band is called as hole-carrier concentration.

In an intrinsic semiconductor, the number of electrons generated in the conduction band is equal to the number of holes generated in the valence band. Hence the electron-carrier concentration is equal to the hole-carrier concentration.

It can be written as,

$$n_i = n = p$$

Where, n = electron-carrier concentration

P = hole-carrier concentration

and ni = intrinsic carrier concentration

The hole concentration in the valence band is given as

$$p = N_V e^{\frac{-(E_F - E_V)}{K_B T}}$$

The electron concentration in the conduction band is given as

$$n = N_c e^{\frac{-(E_C - E_F)}{K_B T}}$$

Where KB is the Boltzmann constant

T is the absolute temperature of intrinsic semiconductor

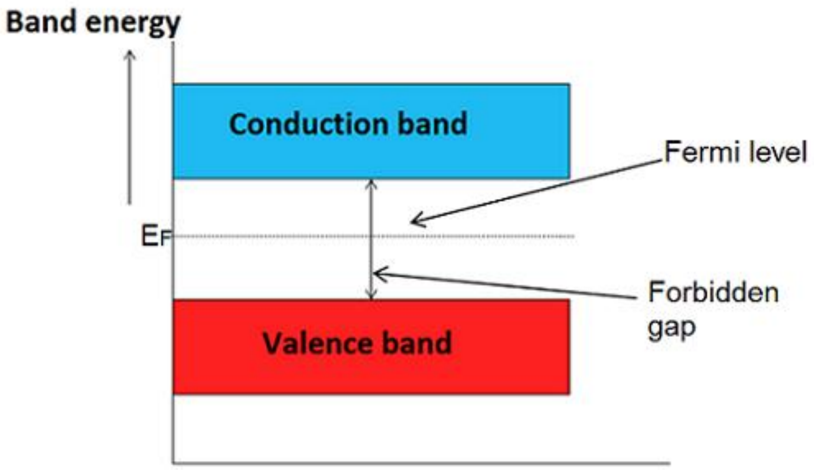
Nc is the effective density of states in conduction band.

Nv is the effective density of states in valence band.

Fermi level in intrinsic semiconductor

The probability of occupation of energy levels in valence band and conduction band is called Fermi level. At absolute zero temperature intrinsic semiconductor acts as perfect insulator. However as the temperature increases free electrons and holes gets generated.

In intrinsic or pure semiconductor, the number of holes in valence band is equal to the number of electrons in the conduction band. Hence, the probability of occupation of energy levels in conduction band and valence band are equal. Therefore, the Fermi level for the intrinsic semiconductor lies in the middle of forbidden band.



Fermi level in the middle of forbidden band indicates equal concentration of free electrons and holes.

The hole-concentration in the valence band is given as

$$p = N_V e^{\frac{-(E_F - E_V)}{K_B T}}$$

The electron-concentration in the conduction band is given as

$$n = N_c e^{\frac{-(E_C - E_F)}{K_B T}}$$

Where KB is the Boltzmann constant

T is the absolute temperature of the intrinsic semiconductor

Nc is the effective density of states in the conduction band.

Nv is the effective density of states in the valence band.

The number of electrons in the conduction band is depends on effective density of states in the conduction band and the distance of Fermi level from the conduction band.

The number of holes in the valence band is depends on effective density of states in the valence band and the distance of Fermi level from the valence band.

For an intrinsic semiconductor, the electron-carrier concentration is equal to the hole-carrier concentration.

It can be written as

$$p = n = n_i$$

Where P = hole-carrier concentration

n = electron-carrier concentration

and ni = intrinsic carrier concentration

The fermi level for intrinsic semiconductor is given as,

$$E_F = \frac{E_C + E_V}{2}$$

Where EF is the fermi level

EC is the conduction band

EV is the valence band

Therefore, the Fermi level in an intrinsic semiconductor lies in the middle of the forbidden gap.

Extrinsic semiconductor

Introduction

The semiconductor in which impurities are added is called extrinsic semiconductor. When the impurities are added to the intrinsic semiconductor, it becomes an extrinsic semiconductor. The process of adding impurities to the semiconductor is called doping. Doping increases the electrical conductivity of semiconductor.

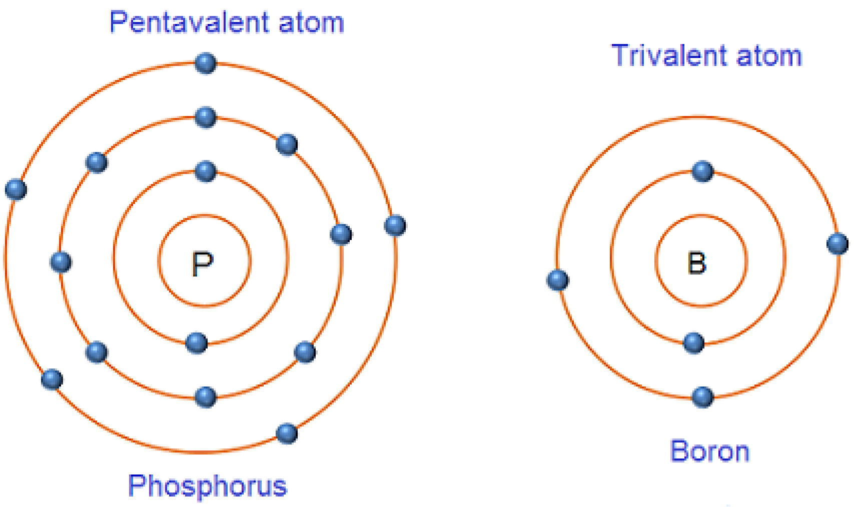
Extrinsic semiconductor has high electrical conductivity than intrinsic semiconductor. Hence the extrinsic semiconductors are used for the manufacturing of electronic devices such as diodes, transistors etc. The number of free electrons and holes in extrinsic semiconductor are not equal.

Types of impurities

Two types of impurities are added to the semiconductor. They are pentavalent and trivalent impurities.

Pentavalent impurities

Pentavalent impurity atoms have 5 valence electrons. The various examples of pentavalent impurity atoms include Phosphorus (P), Arsenic (As), Antimony (Sb), etc. The atomic structure of pentavalent atom (phosphorus) and trivalent atom (boron) is shown in below fig.



Phosphorus is a substance consisting of atoms which all have the same number of protons. The atomic number of phosphorus is 15 i.e. 15 protons. The number of protons in the nucleus of an atom is called atomic number. Phosphorus atom has 15 electrons (2 electrons in first orbit, 8 electrons in second orbit and 5 electrons in the outermost orbit).

Trivalent impurities

Trivalent impurity atoms have 3 valence electrons. The various examples of trivalent impurities include Boron (B), Gallium (G), Indium(In), Aluminium(Al).

Boron is a substance consisting of atoms which all have the same number of protons. The atomic number of boron is 5 i.e. 5 protons. Boron atom has 5 electrons (2 electrons in first orbit and 3 electrons in the outermost orbit).

Classification of extrinsic semiconductors based on impurities added

Based on the type of impurities added, extrinsic semiconductors are classified in to two types.

N-type semiconductor

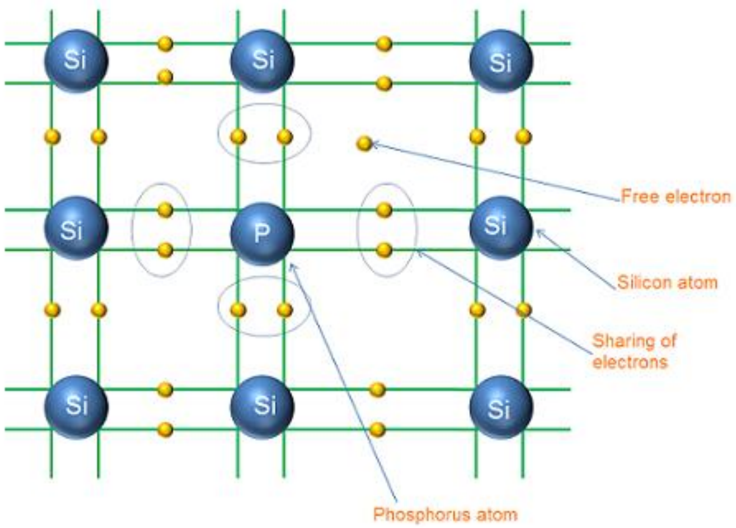
P-type semiconductor

N-type semiconductor

When pentavalent impurity is added to an intrinsic or pure semiconductor (silicon or germanium), then it is said to be an n-type semiconductor. Pentavalent impurities such as phosphorus, arsenic, antimony etc are called donor impurity.

Let us consider, pentavalent impurity phosphorus is added to silicon as shown in below figure. Phosphorus atom has 5 valence electrons and silicon has 4 valence electrons. Phosphorus atom has one excess valence electron than silicon. The four valence electrons of each phosphorus atom form 4 covalent bonds with the 4 neighboring silicon atoms. The fifth valence electron of the phosphorus atom cannot able to form the covalent bond with the silicon atom because silicon atom does not have the fifth valence electron to form the covalent bond.

Thus, fifth valence electron of phosphorus atom does not involve in the formation of covalent bonds. Hence, it is free to move and not attached to the parent atom.



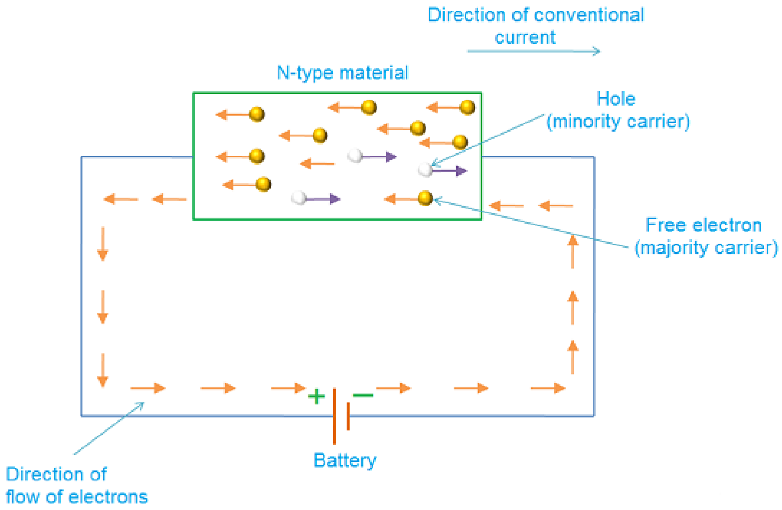
This shows that each phosphorus atom donates one free electron. Therefore, all the pentavalent impurities are called donors. The number of free electrons are depends on the amount of impurity (phosphorus) added to the silicon. A small addition of impurity (phosphorus) generates millions of free electrons.

Charge on n-type semicondctor

So many people think that n-type semiconductor has large number of free electrons. So, the total electric charge of n-type semiconductor is negative. But this assumption is wrong. Even though n-type semiconductor has large number of free electrons, but these free electrons is given by the pentavalent atoms that are electrically neutral. Therefore, the total electric charge of n-type semiconductor is also neutral.

Conduction in n-type semiconductor

Let us consider an n-type semiconductor as shown in below figure. When voltage is applied to n-type semiconductor; the free electrons moves towards positive terminal of applied voltage. Similarly holes moves towards negative terminal of applied voltage.

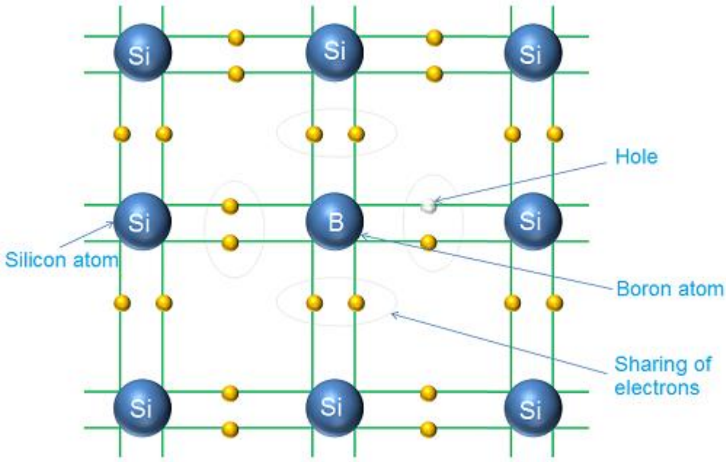


In n-type semiconductor, the population of free electrons is more whereas the population of holes is less. Hence in n-type semiconductor free electrons are called majority carriers and holes are called minority carriers. Therefore, in a n-type semiconductor conduction is mainly because of motion of free electrons.

P-type semiconductor

When the trivalent impurity is added to an intrinsic or pure semiconductor (silicon or germanium), then it is said to be an p-type semiconductor. Trivalent impurities such as Boron (B), Gallium (G), Indium(In), Aluminium(Al) etc are called acceptor impurity.

Let us consider, trivalent impurity boron is added to silicon as shown in below figure. Boron atomhas three valence electrons and silicon has four valence electrons. The three valence electrons of each boron atom form 3 covalent bonds with the 3 neighboring silicon atoms.



In the fourth covalent bond, only silicon atom contributes one valence electron, while the boron atom has no valence electron to contribute. Thus, the fourth covalent bond is incomplete with shortage of one electron. This missing electron is called hole.

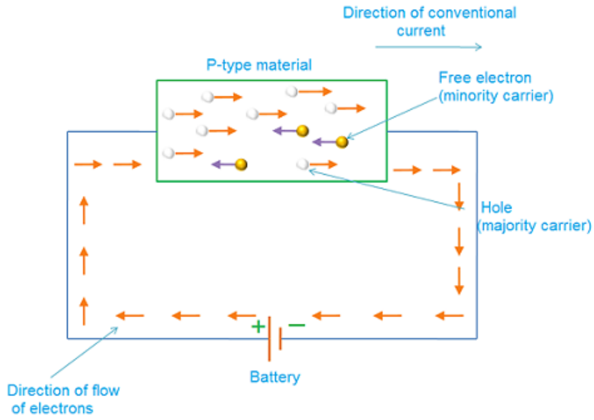
This shows each boron atom accept one electron to fill the hole. Therefore, all the trivalent impurities are called acceptors. A small addition of impurity (boron) provides millions of holes.

Charge on p-type semiconductor

So many people think that p-type semiconductor has large number of holes and current conduction is mainly due to these holes. So, the total electric charge of p-type semiconductor is positive. But this assumption is wrong. Even though p-type semiconductor has large number of holes, but these holes is provided by the trivalent atoms that are electrically neutral. Therefore, the total electric charge of p-type semiconductor is also neutral.

Conduction in p-type semiconductor

Let us consider a p-type semiconductor as shown in below figure. When voltage is applied to p-type semiconductor; the holes in valence band moves towards negative terminal of applied voltage. Similarly free electrons move towards positive terminal of applied voltage.



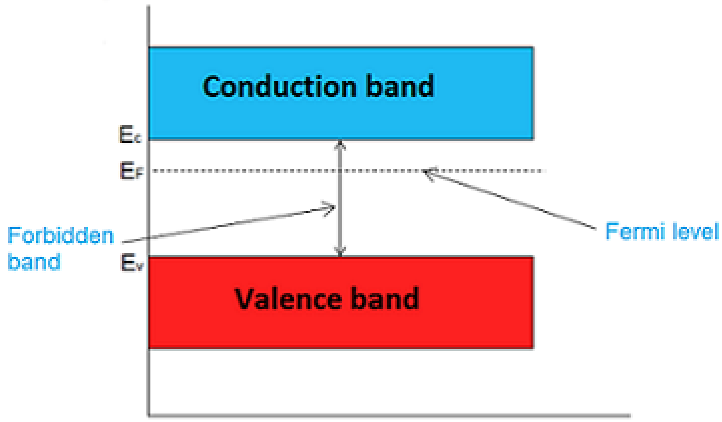
In p-type semiconductor, the population of holes in valence band is more, whereas the population of free electrons in conduction band is less. So, current conduction is mainly because of holes in valence band. Free electrons in conduction band constitute little current. Hence in p-type semiconductor, holes are called majority carriers and free electrons are called minority carriers.

Fermi level in extrinsic semiconductor

In extrinsic semiconductor, the number of electrons in the conduction band and the number of holes in the valence band are not equal. Hence, the probability of occupation of energy levels in conduction band and valence band are not equal. Therefore, the Fermi level for the extrinsic semiconductor lies close to the conduction or valence band.

Fermi level in n-type semiconductor

In n-type semiconductor pentavalent impurity is added. Each pentavalent impurity donates a free electron. The addition of pentavalent impurity creates large number of free electrons in the conduction band.



At room temperature, the number of electrons in the conduction band is greater than the number of holes in the valence band. Hence, the probability of occupation of energy levels by the electrons in the conduction band is greater than the probability of occupation of energy levels by the holes in the valence band. This probability of occupation of energy levels is represented in terms of Fermi level. Therefore, the Fermi level in the n-type semiconductor lies close to the conduction band.

The Fermi level for n-type semiconductor is given as

$$E_F = E_C - K_B T \log \frac{N_C}{N_D}$$

Where EF is the fermi level.

EC is the conduction band.

KB is the Boltzmann constant.

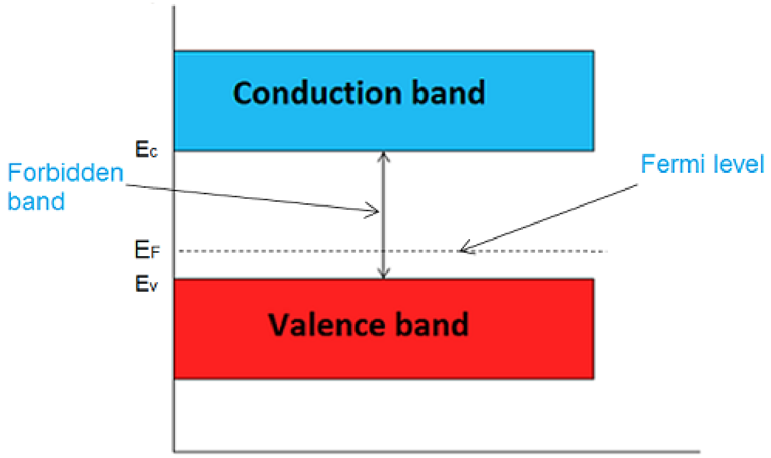
T is the absolute temperature.

NC is the effective density of states in the conduction band.

ND is the concentration of donar atoms.

Fermi level in p-type semiconductor

In p-type semiconductor trivalent impurity is added. Each trivalent impurity creates a hole in the valence band and ready to accept an electron. The addition of trivalent impurity creates large number of holes in the valence band.



At room temperature, the number of holes in the valence band is greater than the number of electrons in the conduction band. Hence, the probability of occupation of energy levels by the holes in the valence band is greater than the probability of occupation of energy levels by the electrons in the conduction band. This probability of occupation of energy levels is represented in terms of Fermi level. Therefore, the Fermi level in the p-type semiconductor lies close to the valence band.

The Fermi level for p-type semiconductor is given as

$$E_F = E_V + K_B T \log \frac{N_V}{N_A}$$

Where NV is the effective density of states in the valence band.

NA is the concentration of acceptor atoms.

Majority & minority carriers

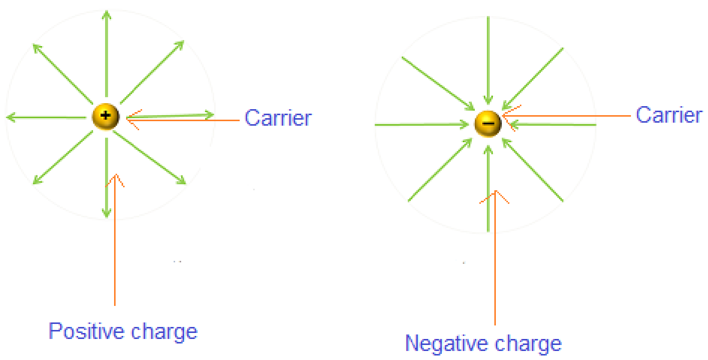
What is charge carrier?

Generally, carrier refers to any object that carry another object from one place to another place. For example, in countries such as India, Singapore and Brazil: Tiffin box or Tiffin carriers are widely used for carrying food from one place to another place. Here, the Tiffin box acts as a carrier that carries the food from one place to another place.

Let us take another example; People use vehicles such as buses, trains, airplanes, etc. to travel from one place to another place. Here, the vehicles act as carriers that carry people from one place to another place. In the similar way, particles such as free electrons and holes carry the charge or electric current from one place to another place.

Negative charge carriers

The negative charge carriers such as free electrons are the charge carriers that carry negative charge with them while moving from one place to another place. Free electrons are the electrons that are detached from the parent atom and moves freely from one place to another place.



Positive charge carriers

The positive charge carriers such as holes are the charge carriers that carry positive charge with them while moving from one place to another place. Holes are the vacancies in valence band that moves from one place to another place within the valence band.

Majority and minority charge carriers definition

The charge carriers that are present in large quantity are called majority charge carriers. The majority charge carriers carry most of the electric charge or electric current in the semiconductor. Hence, majority charge carriers are mainly responsible for electric current flow in the semiconductor.

The charge carriers that are present in small quantity are called minority charge carriers. The minority charge carriers carry very small amount of electric charge or electric current in the semiconductor.

Charge carriers in intrinsic semiconductor

The semiconductors that are in pure form are called intrinsic semiconductors. In intrinsic semiconductor the total number of negative charge carriers (free electrons) is equal to the total number of positive charge carriers (holes or vacancy).

Total negative charge carriers = Total positive charge carriers

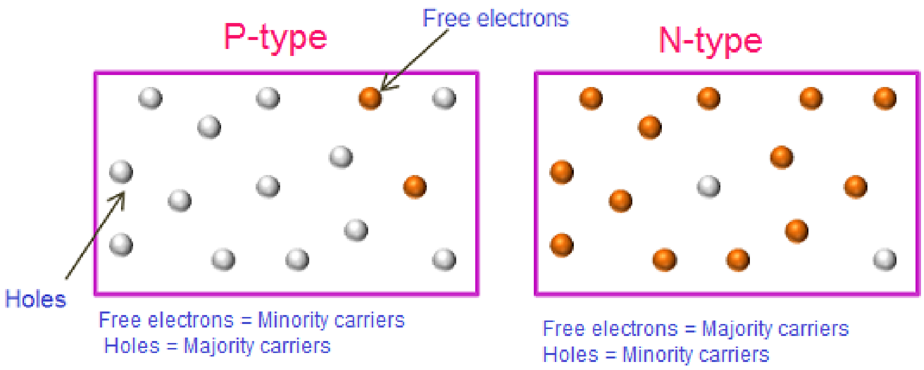
Majority and minority charge carriers in n-type semiconductor

When the pentavalent atoms such as Phosphorus or Arsenic are added to the intrinsic semiconductor, an n-type semiconductor is formed. In n-type semiconductor, large number of free electrons is present. Hence, free electrons are the majority charge carriers in the n-type semiconductor. The free electrons (majority charge carriers) carry most of the electric charge or electric current in the n-type semiconductor.

In n-type semiconductor, very small number of holes is present. Hence, holes are the minority charge carriers in the n-type semiconductor. The holes (minority charge carriers) carry only a small amount of electric charge or electric current in the n-type semiconductor.

The total number of negative charge carriers (free electrons) in n-type semiconductor is greater than the total number of positive charge carriers (holes) in the n-type semiconductor.

Total negative charge carriers > Total positive charge carriers



Majority and minority charge carriers in p-type semiconductor

When the trivalent atoms such as Boron or Gallium are added to the intrinsic semiconductor, a p-type semiconductor is formed. In p-type semiconductor, large number of holes is present. Hence, holes are the majority charge carriers in the p-type semiconductor. The holes (majority charge carriers) carry most of the electric charge or electric current in the p-type semiconductor.

In p-type semiconductor, very small number of free electrons is present. Hence, free electrons are the minority charge carriers in the p-type semiconductor. The free electrons (minority charge carriers) carry only a small amount of electric current in the p-type semiconductor.

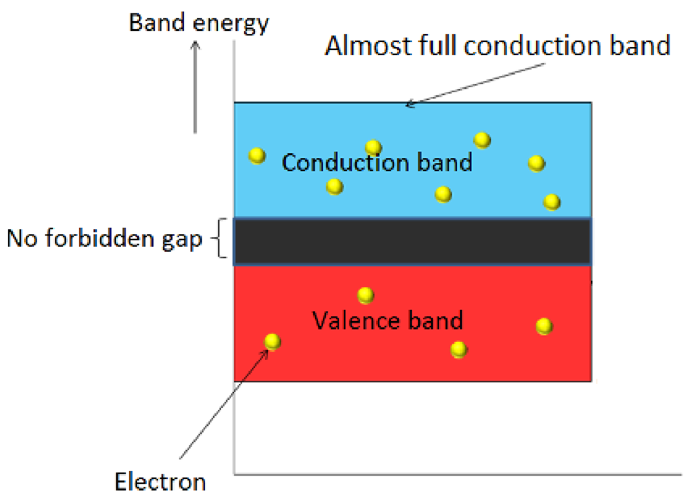
The total number of negative charge carriers (free electrons) in p-type semiconductor is less than the total number of positive charge carriers (holes) in the p-type semiconductor.

Total negative charge carriers < Total positive charge carriers

Heat & light effect on conductors

Heat effect on conductors

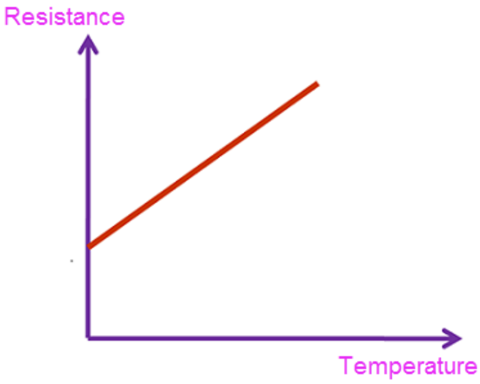
Conductors are the materials that easily allow the flow of electrons or electric current through them. In a conductor, no forbidden gap is present between the valence band and the conduction band. Hence, the valence electrons can easily jumps into the conduction band.



When a small amount of external energy in the form of heat is applied to the conductor, the valence band electrons or the valence electrons gain enough energy to break the bonding with the parent atom and they jumps into the conduction band. The electrons that are present in the conduction band are not attached to the parent atom. Hence, the electrons present in the conduction band moves freely from one place to another place within the conductor. The electrons present in the conduction band are called as free electrons. These free electrons carry the electric charge or electric current with them while moving from one place to another place within the conductor. In conductors, even at low temperature, a small number of free electrons are present, whereas at room temperature, a large number of free electrons are present in the conduction band.

When the heat energy is applied to the conductor, the atoms start vibrating at fixed positions and the free electrons moving in the conduction band will collides with these atoms. Due to this collision, the free electrons will lose their kinetic energy by transferring their kinetic energy to the atoms with which they collide.

After collision with the atoms, the free electrons again start moving in other direction and again the free electrons collides with another atom. This will result in loss of kinetic energy of the free electrons. Likewise, the free electrons will collide with the atoms repeatedly. This causes a reduction in the flow of free electrons that carry the electric current or electric charge from one place to another place.



The concept of collisions of free electrons with the atoms in the conductor can be easily understood with a simple example. Let us consider a Track running (hurdling) game of 100 meters in which 10 athletes are participating. In this Track running (hurdling) game, for each player 20 hurdles are placed in a series over the running track. The participants have to reach the destination by overcoming the hurdles.

When the athletes start running by increasing their speed, they gain the kinetic energy. When the 10 athletes collide with the first hurdle, they lose their kinetic energy and transfer the kinetic energy to the hurdle. After a collision with the first hurdle, the athletes again start running by increasing their speed (kinetic energy). If the athletes again collide at the second hurdle, they again lose their kinetic energy. Likewise, the athletes collide with the hurdles repeatedly. This results in reduction in the speed of athletes.

Here, the free electrons are act as 10 athletes that always try to move from one place to other place whereas the vibrating atoms at fixed positions are act as hurdles that oppose the flow of free electrons.

Reduction in current flow means increase in resistance. Hence, the resistance of a conductor increases with the increase in the temperature (heat). Therefore, conductors have a positive temperature co-efficient.

Light effect on conductors

The light effect on a conductor is exactly similar to the heat effect on a conductor. When a small amount of external energy in the form of light is applied to the conductor, the valence electrons gain enough energy from the light to break the bonding with the atom and they jumps into the conduction band. The electrons present in the conduction band are not attached to any atom. Hence, they move freely from one place to another place. These electrons are called as free electrons.

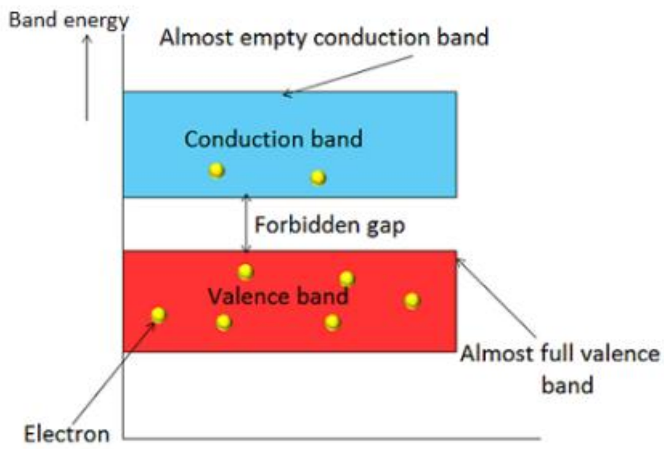
When the light energy is applied to the conductor, the atoms start vibrating and the free electrons moving in the conduction band will collides continuously with these atoms. Due to these collisions, the speed of the free electrons decreases and they do not move freely from one place to another place. This reduces the current flow in the conductor and increases the resistance (opposition to the electric current) in the conductor. Thus, the resistance (opposition to the electric current) of a conductor increases with the increase in the light energy.

Heat effect on semiconductors

A semiconductor is a material that has the electrical conductivity between that of a conductor and an insulator. In semiconductors, a small forbidden gap is present between the valence band and the conduction band. Forbidden gap is the energy gap present between the valence band and conduction band in which no electron energy levels are allowed. Hence, electrons do not stay or exist in the forbidden gap.

The valence electrons or the valence band electrons in the semiconductor need a small amount of external energy in the form of heat in order to jump from valence band to the conduction band. However, at absolute zero temperature, heat does not exist to transfer its energy to the valence electrons. Hence, the valence electrons do not get enough energy in the form of heat to break bonding with the parent atoms.

At absolute zero temperature, all the valence electrons are revolving around the nucleus of an atom. Hence, there are no free electrons present in the conduction band to carry electric current from one place to another place. Therefore, the semiconductor behaves as a perfect insulator at absolute zero temperature.

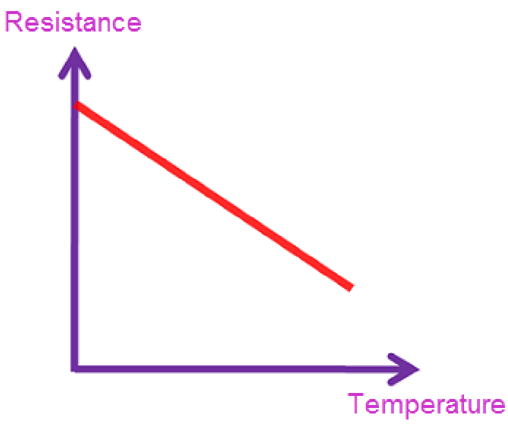


When the temperature is increased above the absolute zero temperature, some of the valence band electrons or valence electrons gain enough energy in the form of heat to break the bonding with the parent atom and they jumps into the conduction band. The electrons present in the conduction band are not attached to the parent atom. Hence, they move freely from one place to another place. These conduction band electrons are called as free electrons.

When the electron leaves the valence band and jumps into the conduction band, a vacancy is created at the electron position in the valence band. This vacancy is called as hole. Thus, both the free electrons in the conduction and holes in the valence band are generated at the same time. The free electrons carry the negative charge or electric current from one place to another place in the conduction band whereas the holes (vacancies) carry the positive charge or electric current from one place to another place in the valence band.

If the temperature or heat energy applied on the semiconductor is further increased then even more number of valence electrons gains enough energy to break the bonding with the parent atom and they jump into the conduction band. This results in increase in number of free electrons in the conduction band. If more number of electrons leaves the valence band and jumps into the conduction band then more number of holes (vacancies) are created in the valence band at the electrons position. Thus, a small increase in heat generates more number of charge carriers (electrons and holes).

In case of conductors, increase in the temperature increases the vibration of atoms. These vibrating atoms oppose the flow of electrons. This results in reduction of current flow in the conductor. The reduction in flow of electrons or current flow means increase in the resistance. Thus, the electric current in the conductor decreases with the increase in temperature.



Just like the conductors, the increase in the temperature increases the vibrations of atoms in the semiconductor. However, the vibrating atoms in the semiconductor oppose only few electrons and the remaining large number of electrons flows freely from one place to another place. This results, increase in current flow in the semiconductor. The increase in current flow means decrease in the resistance. Thus, the electric current in the semiconductor increases with the increase in temperature or heat (Semiconductor has negative temperature coefficient).

Light effect on semiconductors

The light effect on a semiconductor is exactly similar to the heat effect on a semiconductor. When a small amount of external energy in the form of light is applied to the semiconductor, the valence electrons gain enough energy to break the bonding with the parent atom and they jumps into the conduction band.

The electrons in the conduction band are not attached to the parent atom. Hence, they move freely from one place to the another place. The electrons that move freely from one place to another place in the conduction band are called as free electrons. When the electron leaves the valence band and jumps into the conduction band, a vacancy is created at the electron position. This vacancy is called as hole. Thus, in semiconductors both the free electrons and the holes are generated as pair at the same time.

If the light energy applied on the semiconductor is further increased then even more number of free electrons is generated. This increase in the number of free electrons increases the current flow and decreases the resistance in the semiconductor. Thus, the electric current in the semiconductor increases with the increase in the light energy.

Generation and recombination of carriers

Generation of carriers (free electrons and holes)

The process by which free electrons and holes are generated in pair is called generation of carriers.

When electrons in a valence band get enough energy, then they will absorb this energy and jumps into the conduction band. The electron which is jumped into a conduction band is called free electron and the place from where electron left is called hole. Likewise, two type of charge carriers (free electrons and holes) gets generated.

Recombination of carriers (free electrons and holes)

The process by which free electrons and the holes get eliminated is called recombination of carriers. When free electron in the conduction band falls in to a hole in the valence band, then the free electron and hole gets eliminated.

Law of mass action

The law of mass action states that the product of number of electrons in the conduction band and the number ofholes in the valence band is constant at a fixed temperature and is independent of amount of donor and acceptor impurity added.

Mathematically it is represented as

$$np = ni2 = constant$$

Where ni is the intrinsic carrier concentration

n is number of electrons in conduction band

p is number of holes in valence band

Law of mass action for extrinsic semiconductor

The law of mass action is applied for both intrinsic and extrinsic semiconductors. For extrinsic semiconductor the law of mass action states that the product of majority carriers and minority carriers is constant at fixed temperature and is independent of amount of donor and acceptor impurity added.

Law of mass action for n-type semiconductor

The law of mass action for n-type semiconductor is mathematically written as

$$nn\ pn = ni2 = constant$$

Where nn= number of electrons in n-type semiconductor

pn = number of holes in n-type semiconductor

The electrons are the majority carriers and holes are the minority carriers in n-type semiconductor.

In n-type semiconductor, as the number of electrons (majority) in the conduction band increases the number of holes (minority) in the valence band decreases.

Therefore, the product of electrons (majority) and holes (minority) remains constant at fixed temperature.

Law of mass action for p-type semiconductor

The law of mass action for p-type semiconductor is mathematically written as

$$pp\ np = ni2 = constant$$

Where pp = number of holes in p-type semiconductor

np = number of electrons in p-type semiconductor

The holes are the majority carriers and electrons are the minority carriers in p-type semiconductor.

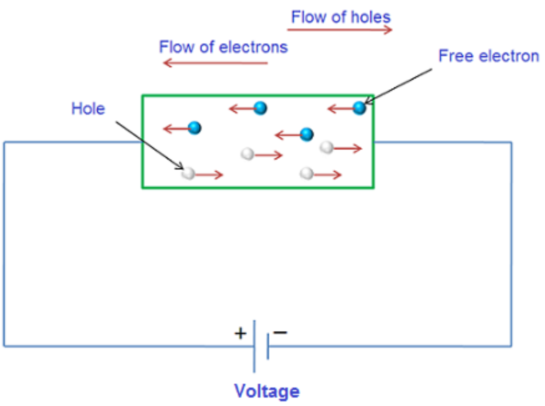
In p-type semiconductor, as the number of holes (majority) in the valence band increases the number of electrons in the conduction band (minority) decreases. Therefore, the product of holes (majority) and electrons (minority) remains constant at fixed temperature.

Drift current

The flow of charge carriers, which is due to the applied voltage or electric field is called drift current.

In a semiconductor, there are two types of charge carriers, they are electrons and holes. When the voltage is applied to a semiconductor, the free electrons move towards the positive terminal of a battery and holes move towards the negative terminal of a battery.

Electrons are the negatively charged particles and holes are the positively charged particles. As we already discussed that like charges repel each other and unlike charges attract each other. Hence, the electrons (negatively charged particle) are attracted towards the positive terminal of a battery and holes (positively charged particle) are attracted towards the negative terminal.



In a semiconductor, the electrons always try to move in a straight line towards the positive terminal of the battery. But, due to continuous collision with the atoms they change the direction of flow. Each time the electron strikes an atom it bounces back in a random direction. The applied voltage does not stop the collision and random motion of electrons, but it causes the electrons to drift towards the positive terminal.

The average velocity that an electron or hole achieved due to the applied voltage or electric field is called drift velocity.

The drift velocity of electrons is given by

$$V_n = \mu_n E$$

The drift velocity of holes is given by

$$V_p = \mu_p E$$

Where vn = drift velocity of electrons

vp = drift velocity of holes

μn = mobility of electrons

μp = mobility of holes

E = applied electric field

The drift current density due to free electrons is given by

$$J_n = e n \mu_n E$$

and the drift current density due to holes is given by

$$J_p = e p \mu_p E$$

Where Jn = drift current density due to electrons

Jp = drift current density due to holes

e = charge of an electron = 1.602 × 10-19 Coulombs (C).

n = number of electrons

p = number of holes

Then the total drift current density is

$$J = J_n + J_p$$

$$= e n \mu_n E + e p \mu_p E$$

$$J = e (n \mu_n + p \mu_p) E$$

Electron and hole mobility

Electron mobility

The ability of an electron to move through a metal or semiconductor, in the presence of applied electric field is called electron mobility.

It is mathematically written as

$$V_n = \mu_n E$$

$$\mu_n = \frac{V_n}{E}$$

Where vn = drift velocity of electrons

μn = mobility of electrons

E = applied electric field

Let us consider a semiconductor that consists of large number of free electrons. When there is no voltage or electric field applied across the semiconductor, the free electrons moves randomly.

However, when the voltage or electric field is applied across the semiconductor, each free electron starts to move more quickly in particular direction. Electrons move very fast in vacuum. However, in metals or semiconductors, free electrons do not move very fast instead they move with a finite average velocity, called drift velocity.

Drift velocity is directly proportional to electric field. Hence, when the electric field increases drift velocity also increases. However, mobility of electrons is independent of applied electric field i.e. change in electric field does not change the mobility of electrons.

The SI unit of electric field is V/m, and the SI unit of velocity is m/s. Thus, the SI unit of mobility is m2/ (V.s).

Hole mobility

The ability of an hole to move through a metal or semiconductor, in the presence of applied electric field is called hole mobility.

It is mathematically written as

$$V_p = \mu_p E$$

$$\mu_p = \frac{V_p}{E}$$

Where vp = drift velocity of holes

μp = mobility of holes

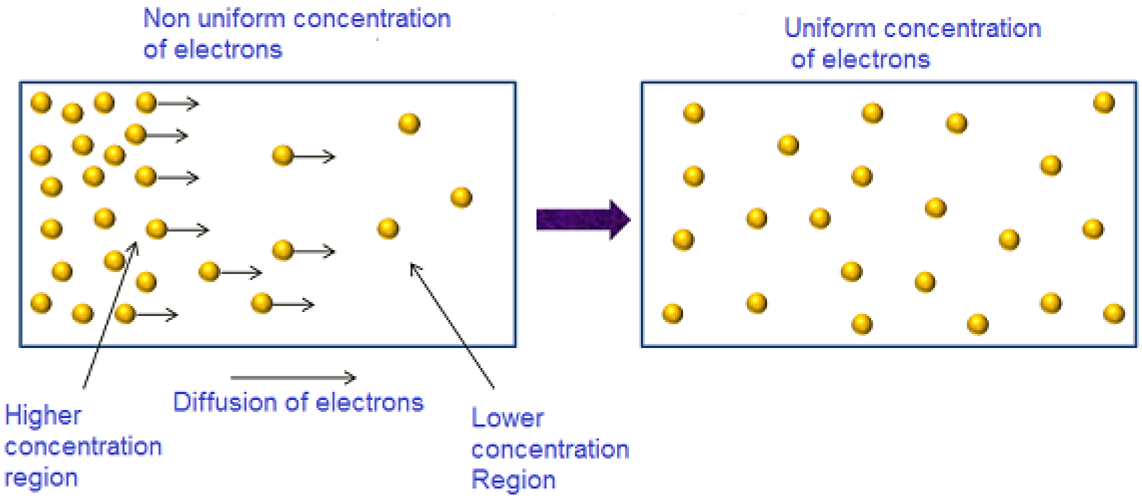
E = applied electric field

Diffusion current

The process by which, charge carriers (electrons or holes) in a semiconductor moves from a region of higher concentration to a region of lower concentration is called diffusion.

The region in which more number of electrons is present is called higher concentration region and the region in which less number of electrons is present is called lower concentration region. Current produced due to motion of charge carriers from a region of higher concentration to a region of lower concentration is called diffusion current. Diffusion process occurs in a semiconductor that is non-uniformly doped.

Consider an n-type semiconductor that is non-uniformly doped as shown in below figure. Due to the non-uniform doping, more number of electrons is present at left side whereas lesser number of electrons is present at right side of the semiconductor material. The number of electrons present at left side of semiconductor material is more. So, these electrons will experience a repulsive force from each other.



The electrons present at left side of the semiconductor material will moves to right side, to reach the uniform concentration of electrons. Thus, the semiconductor material achieves equal concentration of electrons. Electrons that moves from left side to right side will constitute current. This current is called diffusion current. In p-type semiconductor, the diffusion process occurs in the similar manner.

Both drift and diffusion current occurs in semiconductor devices. Diffusion current occurs without an external voltage or electric field applied. Diffusion current does not occur in a conductor. The direction of diffusion current is same or opposite to that of the drift current.

Concentration gradient

The diffusion current density is directly proportional to the concentration gradient. Concentration gradient is the difference in concentration of electrons or holes in a given area. If the concentration gradient is high, then the diffusion current density is also high. Similarly, if the concentration gradient is low, then the diffusion current density is also low.

The concentration gradient for n-type semiconductor is given by

$$J_n \propto \frac{dn}{dx}$$

The concentration gradient for p-type semiconductor is given by

$$J_p \propto \frac{dp}{dx}$$

Where Jn = diffusion current density due to electrons

Jp = diffusion current density due to holes

Diffusion current density

The diffusion current density due to electrons is given by

$$J_n = +e\, D_n \frac{dn}{dx}$$

Where Dn is the diffusion coefficent of electrons

The diffusion current density due to holes is given by

$$J_p = -e\, D_p \frac{dp}{dx}$$

Where Dp is the diffusion coefficent of holes

The total current density due to electrons is the sum of drift and diffusion currents.

Jn = Drift current + Diffusion current

$$J_n = en\mu_n E + e\, D_n \frac{dn}{dx}$$

(read topic : drift current)

The total current density due to holes is the sum of drift and diffusion currents.

Jp = Drift current + Diffusion current

$$J_p = ep\mu_p E - e\, D_p \frac{dp}{dx}$$

The total current density due to electrons and holes is given by

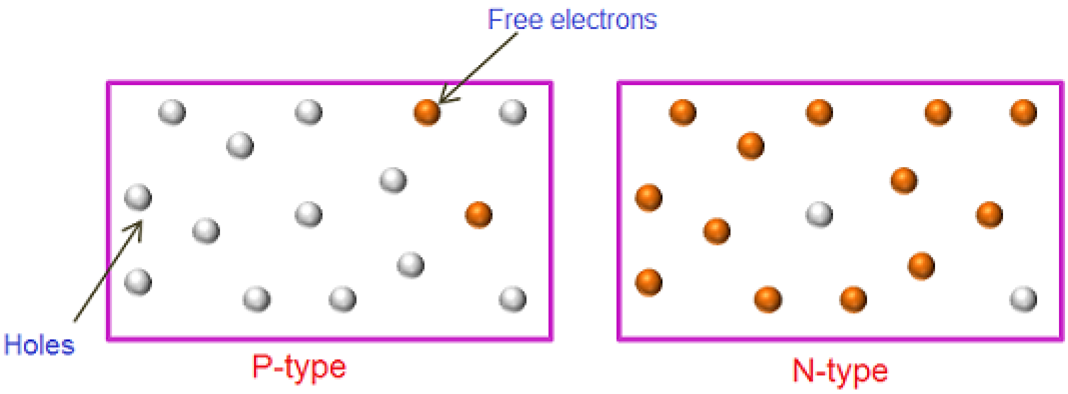
J = Jn + Jp

P-N Junction

We have already discussed about the p-type and n-type semiconductors in the previous tutorial. To read about p-type and n-type semiconductors click on this links (p-type semiconductor, n-type semiconductor).

The n-type semiconductor is formed by adding pentavalent impurities to the intrinsic semiconductor while p-type semiconductor is formed by adding trivalent impurities to the intrinsic semiconductor.

Also, in n-type semiconductors electrons are the majority carriers while holes are the minority carriers. On the other hand, in p-type semiconductors holes are the majority carriers while electrons are the minority carriers. The p-type and n-type semiconductors are not used separately for practical purpose because the overall charge of p-type and n-type semiconductors is electrically neutral. However, when p-type and n-type semiconductor materials are joined they behave differently.

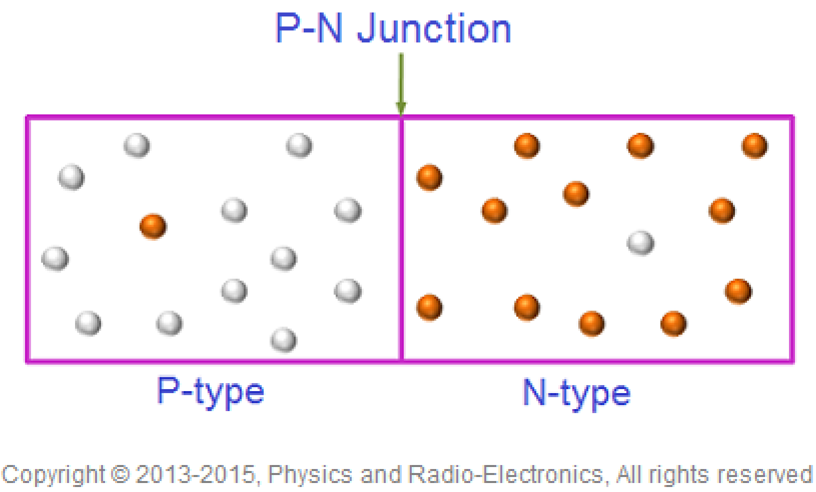


Remember that although n-type semiconductor has large number of free electrons, but these free electrons are provided by the pentavalent atoms that are electrically neutral. Thus, the n-type semiconductor is electrically neutral. Similarly, in p-type semiconductor holes are provided by the trivalent atoms that are electrically neutral. Therefore, overall charge of p-type semiconductor is also neutral.

Formation of p-n junction

Generally junction refers to a point where two or more things are joined. For example, when one or more railway tracks are joined a railway junction is formed. The region where the tracks meet or joined is called railway junction.

In the similar way, when an n-type semiconductor is joined with the p-type semiconductor, a p-n junction is formed. The region where the p-type and n-type semiconductors are joined is called p-n junction. It is also defined as the boundary between p-type and n-type semiconductor. This p-n junction forms a most popular semiconductor device known as diode.



P-n junction is also a fundamental building block of many other semiconductor electronic devices such as transistors, solar cells, light emitting diodes, and integrated circuits. The credit of discovery of the p-n junction goes to American physicist Russel Ohi of Bell Laboratories.

Semiconductor devices are the fundamental building blocks of all the electronic devices such as computers, control systems, ATM (Automated Teller Machine), mobile phones, amplifiers, etc. Diodes are the simplest form of all the semiconductor devices. The various applications of diodes include computers, power supplies, television, radios and so on.

Before the invention of semiconductor diode there was vacuum tubes which are large in size, takes more power, costly, and noisy. This problem was solved with invention of semiconductor diode. Semiconductor diodes are small in size, low cost and consume less power.

Zero bias P-N junction

In n-type semiconductor materials, pentavalent impurities are added while in p-type semiconductor materials, trivalent impurities are added. If the impurities added to p-type and n-type semiconductor materials are not uniform then at one region large number of charge carriers are present while at another region small number of charge carriers are present. Due to this non uniform doping charge carriers at high concentration region repel from each other and move towards the low concentration region to achieve uniform concentration all over the material.

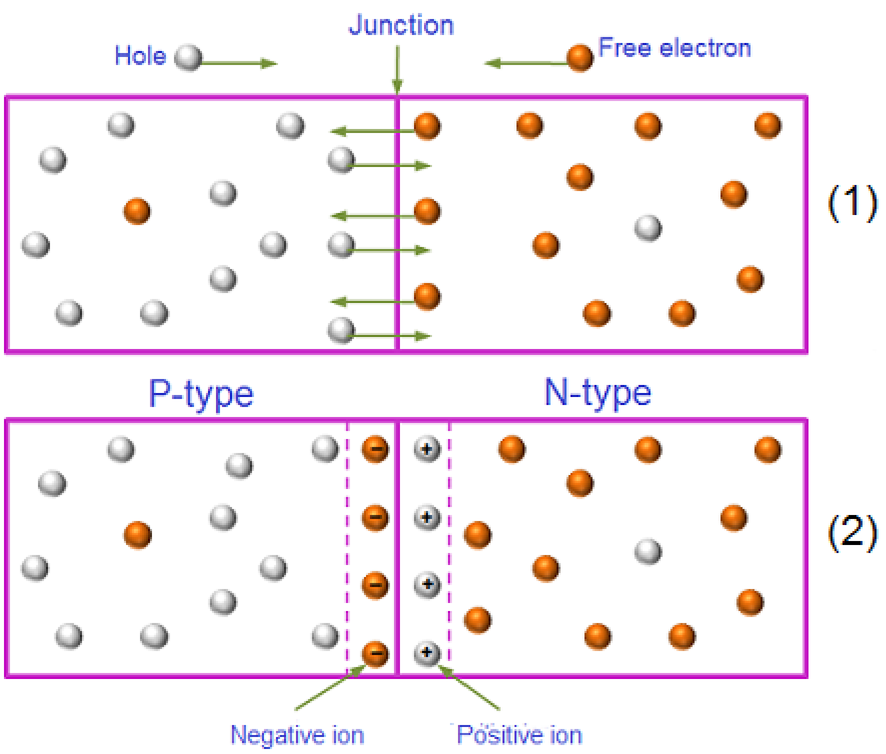
Charge carriers crossing the junction

The p-n Junction in which no external voltage is applied is called zero bias p-n junction. Zero bias p-n Junction is also called as unbiased p-n junction.

Let us consider a zero bias p-n junction as shown in below figure, at n-type semiconductor large number of free electrons is present while at p-type semiconductor small number of free electrons is present. Hence, the concentration of electrons at n-type semiconductor is high while the concentration of electrons at p-type semiconductor is low.

Due to this high concentration of electrons at n-side, they get repelled from each other. Hence they try to move towards the low concentration region. Also, since free electrons and holes at the junction are very close to each other. According to coulomb’s law there exist an electrostatic force of attraction between the opposite charges.

Hence, the free electrons from the n-side are attracted towards the holes at the p-side. Thus, the free electrons move from n-region (high concentration region) to p-region (low concentration region).



At p-type semiconductor large number of holes is present while at n-type semiconductor small number of holes is present. Hence, the concentration of holes at p-type semiconductor is high while the concentration of holes at n-type semiconductor is low. Also, since electrons and holes at the junction are very close to each other. According to coulomb's law there exist an electrostatic force of attraction between the opposite charges.

Hence, the holes from the p-side are attracted towards the free electrons at the n-side. Thus, the holes move from p-region (high concentration region) to n-region (low concentration region).

Formation of positive and negative ion

The free electrons crossing the junction provide the extra electrons to the p-side atoms by filling the holes in the p-side atoms. The atom which gains extra electron has more number of electrons than protons. We know that, when the atom gains an extra electron from the outside atom it will become a negative ion. Negative ions are also called as acceptors because they accept extra electrons from outside atoms.

Thus, each free electron that crosses the junction to fill the holes in the p-side creates negative ions on the p-side.

Similarly each free electron that left the n-side atom to fill the holes in the p-side atom creates a hole at n-side atom. The atom which loses an electron has more number of protons than electrons. We know that, when the atom loses an electron it will becomes a positive ion. Positive ions are also called as donors because they donate extra electrons to the outside atoms.

Thus, each free electron that left the n-side parent atom and crosses the junction to fill the holes in the p-side atom creates positive ion at n-side.

Barrier voltage

The n-type and p-type semiconductor materials are electrically neutral before the free electrons and holes had crossed the junction. However, when the free electrons and holes have crossed the junction, the n-type and p-type semiconductors become charged.

Charged atom

If the atom has unequal number of electrons and protons, the atom is said to be charged. This charged atom may be positive or negative. The atom which has more number of electrons than the protons is said to be negatively charged while the atom which has more number of protons than the electrons is said to be positively charged.

Positive and negative barrier voltage at the p-n junction

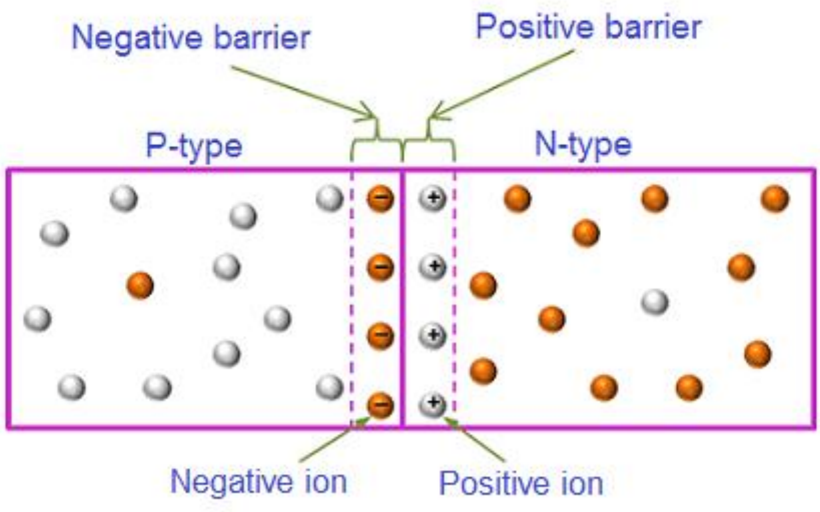
In positive ion, the total number of electrons orbiting the nucleus of an atom is less compared to the total number of protons present in the nucleus of an atom. Hence, it is positively charged. On the other hand, in negative ion, the total number of electrons orbiting the nucleus of an atom is more compared to the total number of protons present in the nucleus of an atom. Hence, it is negatively charged.

Thus, a net positive charge is built at the n-side of the p-n junction due to the positive ions at the n-side; similarly a net negative charge is built at the p-side of the p-n junction due to the negative ions at the p-side.

This net negative charge at the p-side of the p-n junction prevents the further flow of free electrons crossing from n-side to p-side because the negative charge present at the p-side of p-n junction repels the free electrons.

Similarly, the net positive charge at n-side of the p-n junction prevents the further flow of holes crossing from p-side to n-side. Hence, positive charge present at n-side and negative charge present at p-side of p-n junction acts as barrier between p-type and n-type semiconductor.

Thus, a barrier is build near the junction which prevents the further movement of electrons and holes.



The negative charge formed at the p-side of the p-n junction is called negative barrier voltage while the positive charge formed at the n-side of the p-n junction is called positive barrier voltage. The total charge formed at the p-n junction is called barrier voltage, barrier potential or junction barrier.

The size of the barrier voltage at the p-n junction is depends on, the amount of doping, junction temperature and type of material used. The barrier voltage for silicon diode is 0.7 volts and for germanium is 0.3 volts.

Barrier voltage allows minority carriers

The barrier voltage at the p-n junction opposes only the flow of majority charge carriers. Which means it prevents the flow of electrons from n-side and flow of holes from p-side.

However, it allows the flow of minority charge carriers. Any free electrons (minority carriers) produced by thermal energy on p-side of the semiconductor are attracted towards the positive barrier at the n-side of the p-n junction. Hence, free electrons (minority carriers) crosses from p-side to n-side.

Similarly, any holes (minority carriers) produced by thermal energy on n-side of the semiconductor are attracted towards the negative barrier at the p-side of the p-n junction. Hence, holes (minority carriers) cross from n-side to p-side.

Thus, the barrier voltage allows the flow of minority carriers (.I.e. free electrons at p-side and holes at n-side) to cross the junction.

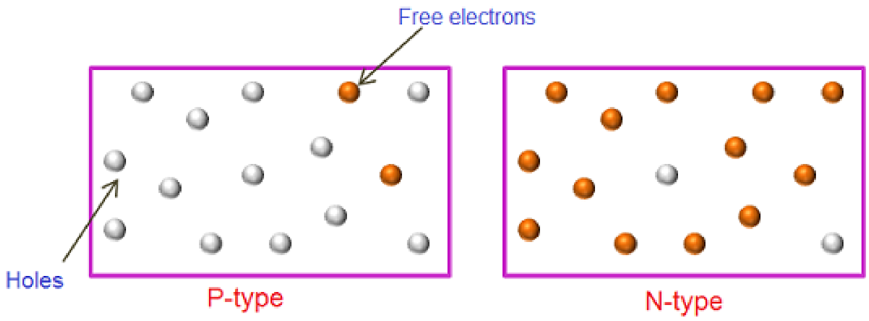
Depletion region

What is depletion region?

Generally, depletion refers to reduction or decrease in quantity of something. For example, oil depletion refers to decrease in oil production from a particular oil well, region, or geographic area over a given time. Similarly, insemiconductor physics, the depletion region refers to a region where flow of charge carriers are decreased over a given time and finally results in empty mobile charge carriers or full of immobile charge carriers.

P-type and n-type semiconductor

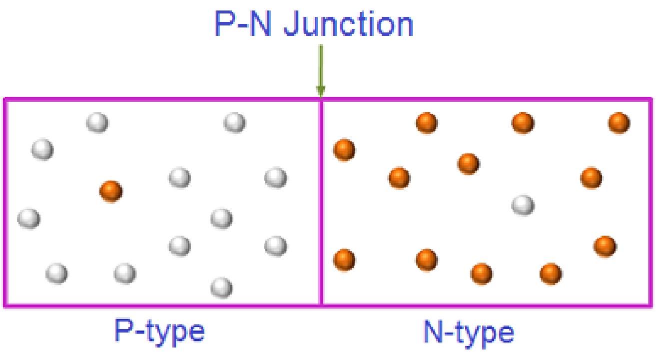
The p-type semiconductor is formed by adding trivalent impurities to the pure or intrinsic semiconductor while n-type semiconductor is formed by adding pentavalent impurities to the pure or intrinsic semiconductor.



In p-type semiconductors, holes are the majority charge carriers while free electrons are the minority charge carriers. On the other hand, in n-type semiconductors free electrons are the majority charge carriers while holes are the minority charge carriers.

Flow of free electrons and holes

If p-type semiconductor is joined with n-type semiconductor, a p-n junction is formed. The region in which the p-type and n-type semiconductors are joined is called p-n junction. This p-n junction separates n-type semiconductor from p-type semiconductor.

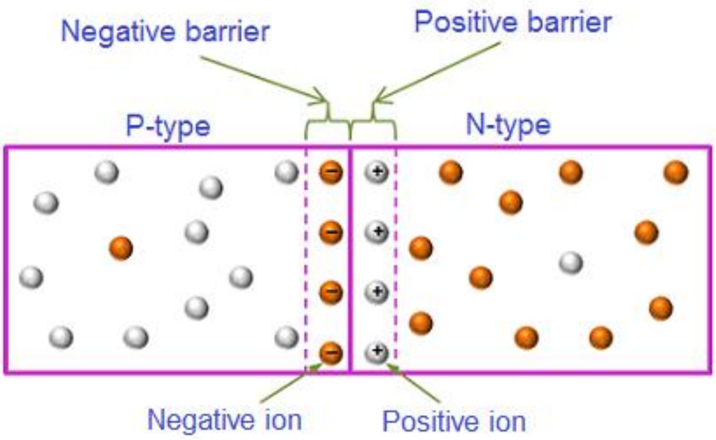


In n-type semiconductors, large number of free electrons is present due to this they get repelled from each other and try to move from a high concentration region (n-side) to a low concentration region (p-side). Moreover, near the junction free electrons and holes are close to each other. According to coulombs law there exist a force of attraction between opposite charges.

Hence, the free electrons from n-side attracted towards the holes at p-side. Thus, the free electrons move from n-side to p-side. Similarly, holes move from p-side to n-side.

Positive and negative charge at p-n junction

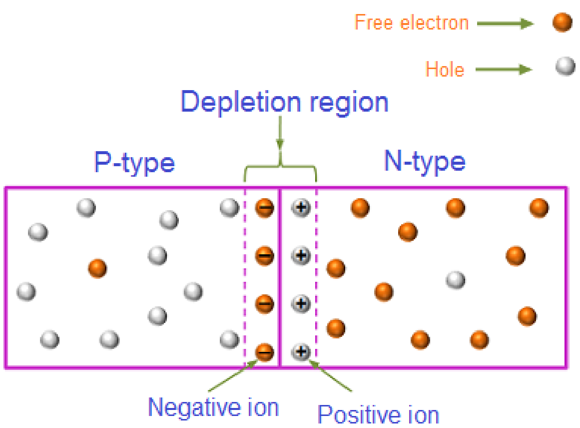
The free electrons that are crossing the junction from n-side provide extra electrons to the atoms on the p-side by filling holes in the p-side atoms. The atom that gains extra electron at p-side has more number of electrons than protons. We know that, when the atom gains an extra electron from the outside atom it will become a negative ion.



Thus, each free electron that is crossing the junction from n-side to fill the hole in p-side atom creates a negative ion at p-side. Similarly, each free electron that left the parent atom at n-side to fill the hole in p-side atom creates a positive ion at n-side.

Negative ion has more number of electrons than protons. Hence, it is negatively charged. Thus, a net negative charge is build at the p-side of p-n junction. Similarly, positive ion has more number of protons than electrons. Hence, it is positively charged. Thus, a net positive charge is build at n-side of the p-n junction.

The net negative charge at p-side of the p-n junction prevents further flow of free electrons crossing from n-side to p-side because the negative charge present at the p-side of the p-n junction repels the free electrons. Similarly, the net positive charge at n-side of the p-n junction prevents further flow of holes from p-side to n-side.



Thus, immobile positive charge at n-side and immobile negative charge at p-side near the junction acts like a barrier or wall and prevent the further flow of free electrons and holes. The region near the junction where flow of charges carriers are decreased over a given time and finally results in empty charge carriers or full of immobile charge carriers is called depletion region.

The depletion region is also called as depletion zone, depletion layer, space charge region, or space charge layer. The depletion region acts like a wall between p-type and n-type semiconductor and prevents further flow of free electrons and holes.

Width of depletion region

What is depletion region?

Depletion region is a region near the p-n junction where flow of charge carriers (free electrons and holes) is reduced over a given period and finally results in zero charge carriers. The width of depletion region is depends on the amount of impurities added to the semiconductor. Impurities are the atoms (pentavalent and trivalent atoms) added to the semiconductor to improve its conductivity.

If pentavalent atoms are added to the pure or intrinsic semiconductor, an n-type semiconductor is formed. On the other hand, if trivalent atoms are added to the pure semiconductor, a p-type semiconductor is formed.

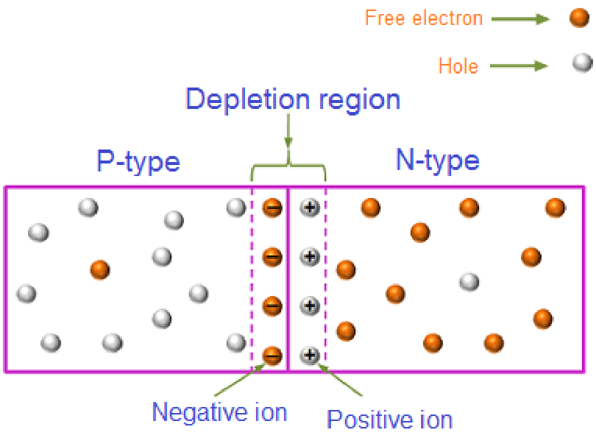
The number of charge carriers in the p-type and n-type semiconductors is depends on the amount of impurity atoms (pentavalent and trivalent atoms) added. If large number of impurity atoms is added to the p-type and n-type semiconductor, a large number of charge carriers (free electrons or holes) are generated. Similarly, if small number of impurity atoms is added to the p-type and n-type semiconductor, a small number of charge carriers are generated.

Depletion width: P-type and n-type semiconductors is heavily doped

The process of adding impurity atoms to the pure or intrinsic semiconductor is called doping. When a large number of pentavalent atoms are added to the intrinsic semiconductor, a large number of free electrons are generated. The semiconductor, which has large number of free electrons, is called n-type semiconductor.

When a large number of trivalent impurity atoms are added to the intrinsic semiconductor, a large number of holes are generated. The semiconductor, which has large number of holes, is called p-type semiconductor.

When a heavily doped n-type semiconductor is joined with the heavily doped p-type semiconductor, a p-n junction is formed. In n-type semiconductor, large number of free electrons is present; due to this, they get repelled from each other and try to move from higher concentration region (n-side) to a lower concentration region (p-side). Similarly, in p-type semiconductor, holes get repelled from each other and try to move from higher concentration (p-side) region to a lower concentration region (n-side). The free electrons that cross the junction provide extra electrons to the atoms at p-side by filling the holes. The atom, which gains extra electron at the p-side, becomes a negative ion. Similarly, the free electrons, which left the parent atoms to fill the holes at p-side, create a positive ion at n-side.



If the doping level is further increased, then even more number of free electrons and holes are generated. This will create a large electric field at n-side and p-side. This electric field dominates the opposing electric field from the ions (depletion region). Hence, the electric field pushes free electrons towards p-side and holes towards n-side.

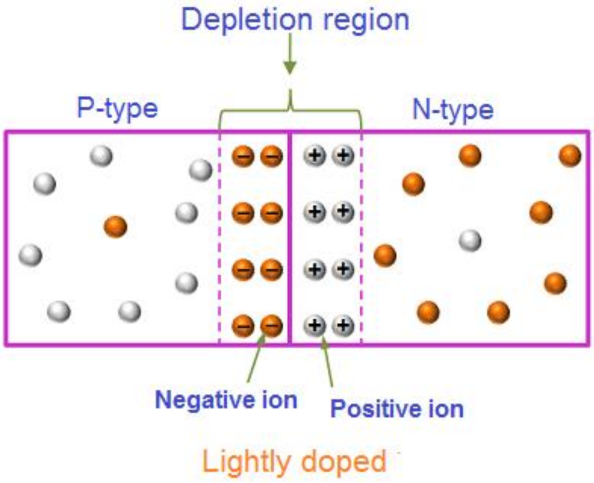
In heavily doped semiconductors, recombination rate is very fast because of large number of charge carriers. Hence, the free electrons (majority charge carriers) fill the holes in the positive ions at n-side before they cross the p-n junction. The positive ion (charged atom), which gains the electron, becomes a neutral atom. Similarly, the holes (majority charge carriers) occupy the electron place in the negative ion before they cross the p-n junction. The negative ion, which loses the free electron, becomes a neutral atom.

Hence, the positive ions at n-side and negative ions at p-side (which acts like a barrier) are decreased over a given period. This decreases the width of depletion region. Thus, the width of depletion region in the heavily doped semiconductor decreases over a given period.

Depletion width: P-type and n-type semiconductors is lightly doped

Just like the heavily doped n-type and p-type semiconductors, the free electrons and holes are generated in the lightly doped n-type and p-type semiconductors. However, the free electrons and holes generated in the lightly doped semiconductors are less when compared with the heavily doped semiconductors.

The free electrons in the n-type semiconductor experience a repulsive force from each other and try to move from a higher concentration region (n-side) to a lower concentration region (p-side). Similarly, the holes try to move from a higher concentration region (p-side) to a lower concentration region (n-side).



In the lightly doped semiconductors, the recombination rate is very slow. Hence, the free electrons from n-side cross the junction and fill the holes in the atoms at p-side before they recombine with the positive ions at n-side. The atom, which gains an extra electron at p-side, becomes a negative ion. Similarly, the holes from p-side cross the junction and occupy the electrons place at n-side before they recombine with the negative ions at p-side. The atom, which loses the valence electron at n-side, becomes a positive ion.

Hence, the positive ions at n-side and the negative ions at p-side are increased over a given period. This increases the width of depletion region. Thus, the width of depletion region increases in the lightly doped semiconductors over a given period.

P-N junction semiconductor diode

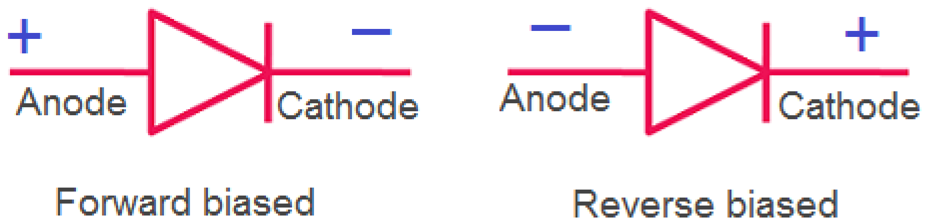
What is p-n junction semiconductor diode?

A p-n junction diode is two-terminal or two-electrode semiconductor device, which allows the electric current in only one direction while blocks the electric current in opposite or reverse direction. If the diode is forward biased, it allows the electric current flow. On the other hand, if the diode is reverse biased, it blocks the electric current flow.P-N junction semiconductor diode is also called as p-n junction semiconductor device.

In n-type semiconductors, free electrons are the majority charge carriers whereas in p-type semiconductors, holesare the majority charge carriers. When the n-type semiconductor is joined with the p-type semiconductor, a p-n junction is formed. The p-n junction, which is formed when the p-type and n-type semiconductors are joined, is called as p-n junction diode.

The p-n junction diode is made from the semiconductor materials such as silicon, germanium, and gallium arsenide. For designing the diodes, silicon is more preferred over germanium. The p-n junction diodes made from silicon semiconductors works at higher temperature when compared with the p-n junction diodes made from germanium semiconductors.

The basic symbol of p-n junction diode under forward bias and reverse bias is shown in the below figure



In the above figure, arrowhead of a diode indicates the conventional direction of electric current when the diode is forward biased (from positive terminal to the negative terminal). The holes which moves from positive terminal (anode) to the negative terminal (cathode) is the conventional direction of current.

The free electrons moving from negative terminal (cathode) to the positive terminal (anode) actually carry the electric current. However, due to the convention we have to assume that the current direction is from positive terminal to the negative terminal.

Biasing of p-n junction semiconductor diode

The process of applying the external voltage to a p-n junction semiconductor diode is called biasing. External voltage to the p-n junction diode is applied in any of the two methods: forward biasing or reverse biasing.

If the p-n junction diode is forward biased, it allows the electric current flow. Under forward biased condition, the p-type semiconductor is connected to the positive terminal of battery whereas; the n-type semiconductor is connected to the negative terminal of battery.

If the p-n junction diode is reverse biased, it blocks the electric current flow. Under reverse biased condition, the p-type semiconductor is connected to the negative terminal of battery whereas; the n-type semiconductor is connected to the positive terminal of battery.

Terminals of pn junction diode

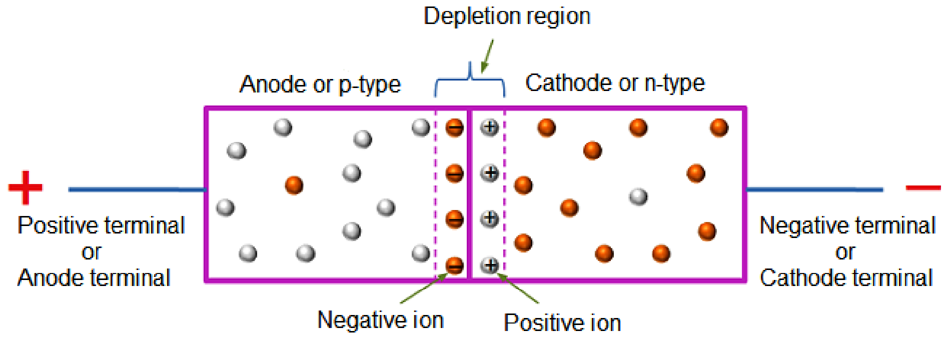
Generally, terminal refers to a point or place at which any object begins or ends. For example, bus terminal or terminus is a place at which all the buses begins or ends. Similarly, in a p-n junction diode, terminal refers a point at which charge carriers begins or ends.

P-n junction diode consists of two terminals: positive and negative. At positive terminal, all the free electrons will ends and all the holes will begins whereas at negative terminal all the free electrons will begins and all the holes will ends.

Terminals of diode under forward bias

In forward biased p-n junction diode (p-type connected to positive terminal and n-type connected to negative terminal), anode terminal is a positive terminal whereas cathode terminal is negative terminal.

Anode terminal is a positively charged electrode or conductor, which supplies holes to the p-n junction. In other words, anode or anode terminal or positive terminal is the source of positive charge carriers (holes), the positive charge carriers (holes) begins their journey at anode terminal and travel through the diode and ends at cathode terminal.



Forward biased

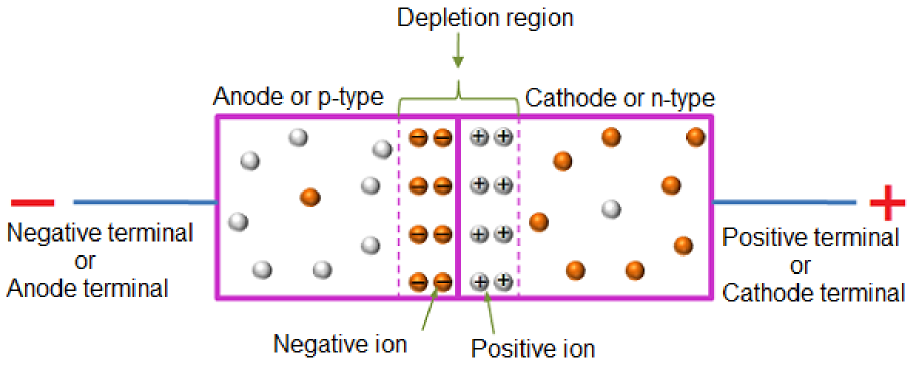
Cathode is the negatively charged electrode or conductor, which supplies free electrons to the p-n junction. In other words, cathode terminal or negative terminal is the source of free electrons, the negative charge carriers (free electrons) begins their journey at cathode terminal and travel through the diode and ends at anode terminal.

The free electrons are attracted towards the anode terminal or positive terminal whereas the holes are attracted towards the cathode terminal or negative terminal.

Terminals of diode under reverse bias

If the diode is reverse biased (p-type connected to negative terminal and n-type connected to positive terminal), the anode terminal becomes a negative terminal whereas the cathode terminal becomes a positive terminal.

Anode terminal or negative terminal supplies free electrons to the p-n junction. In other words, anode terminal is the source of free electrons, the free electrons begins their journey at negative or anode terminal and fills the large number of holes in the p-type semiconductor. The holes in the p-type semiconductor get attracted towards the negative terminal. The free electrons from the negative terminal cannot move towards the positive terminal because the wide depletion region at the p-n junction resists or opposes the flow of free electrons.



Reverse biased

Cathode terminal or positive terminal supplies holes to the p-n junction. In other words, cathode terminal is the source of holes, the holes begins their journey at positive or cathode terminal and occupies the electrons position in the n-type semiconductor. The free electrons in the n-type semiconductor gets attracted towards the positive terminal. The holes from the positive terminal cannot move towards the negative terminal because the wide depletion region at the p-n junction opposes the flow of holes.

Silicon and germanium semiconductor diodes

For designing the diodes, silicon is more preferred over germanium.

The p-n junction diodes made from silicon semiconductors works at high temperature than the germanium semiconductor diodes.

Forward bias voltage for silicon semiconductor diode is approximately 0.7 volts whereas for germanium semiconductor diode is approximately 0.3 volts.

Silicon semiconductor diodes do not allow the electric current flow, if the voltage applied on the silicon diode is less than 0.7 volts.

Silicon semiconductor diodes start allowing the current flow, if the voltage applied on the diode reaches 0.7 volts.

Germanium semiconductor diodes do not allow the electric current flow, if the voltage applied on the germanium diode is less than 0.3 volts.

Germanium semiconductor diodes start allowing the current flow, if the voltage applied on the germanium diode reaches 0.3 volts.

The cost of silicon semiconductors is low when compared with the germanium semiconductors.

Advantages of p-n junction diode

P-n junction diode is the simplest form of all the semiconductor devices. However, diodes plays a major role in many electronic devices.

A p-n junction diode can be used to convert the alternating current (AC) to the direct current (DC). These diodes are used in power supply devices.

If the diode is forward biased, it allows the current flow. On the other hand, if it is reverse biased, it blocks the current flow. In other words, the p-n junction diode becomes on when it is forward biased whereas the p-n junction diode becomes off when it is reversed biased (I.e. it acts as switch). Thus, the p-n junction diode is used as electronic switch in digital logic circuits.

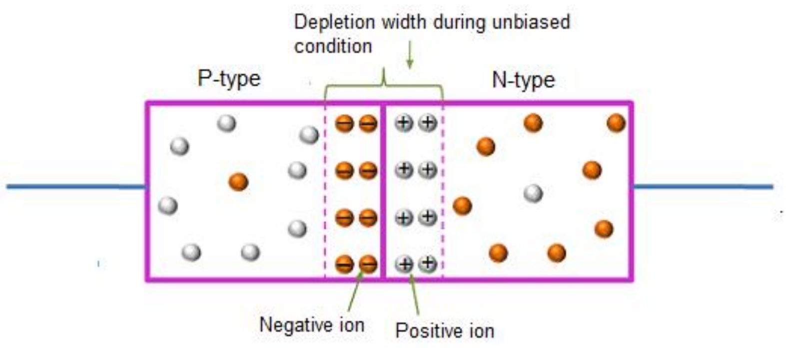
Forward biased p-n junction diode

The process by which, a p-n junction diode allows the electric current in the presence of applied voltage is called forward biased p-n junction diode.

In forward biased p-n junction diode, the positive terminal of the battery is connected to the p-type semiconductormaterial and the negative terminal of the battery is connected to the n-type semiconductor material.

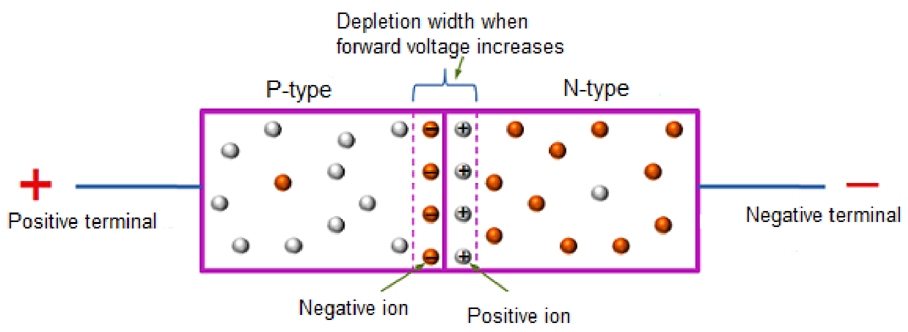
Unbiased diode and forward biased diode

Under no voltage or unbiased condition, the p-n junction diode does not allow the electric current. If the external forward voltage applied on the p-n junction diode is increased from zero to 0.1 volts, the depletion region slightly decreases. Hence, very small electric current flows in the p-n junction diode. However, this small electric current in the p-n junction diode is considered as negligible. Hence, they not used for any practical applications.



Unbiased

If the voltage applied on the p-n junction diode is further increased, then even more number of free electrons and holes are generated in the p-n junction diode. This large number of free electrons and holes further reduces the depletion region (positive and negative ions). Hence, the electric current in the p-n junction diode increases. Thus, the depletion region of a p-n junction diode decreases with increase in voltage. In other words, the electric current in the p-n junction diode increases with the increase in voltage.

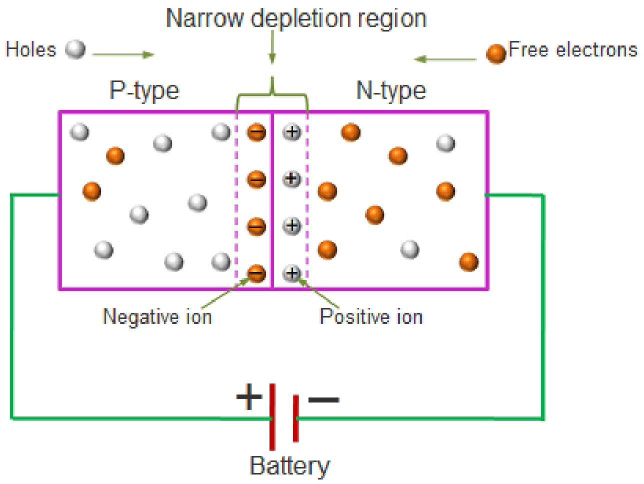


Forward biased

Electron and hole current

Electron current

If the p-n junction diode is forward biased with approximately 0.7 volts for silicon diode or 0.3 volts for germanium diode, the p-n junction diode starts allowing the electric current. Under this condition, the negative terminal of the battery supplies large number of free electrons to the n-type semiconductor and attracts or accepts large number of holes from the p-type semiconductor. In other words, the large number of free electrons begins their journey at the negative terminal whereas the large number of holes finishes their journey at the negative terminal.



Forward bias

The free electrons, which begin their journey from the negative terminal, produce a large negative electric field. The direction of this negative electric field is apposite to the direction of positive electric field of depletion region (positive ions) near the p-n junction.

Due to the large number of free electrons at n-type semiconductor, they get repelled from each other and try to move from higher concentration region (n-type semiconductor) to a lower concentration region (p-type semiconductor). However, before crossing the depletion region, free electrons finds the positive ions and fills the holes. The free electrons, which fills the holes in positive ions becomes valence electrons. Thus, the free electrons are disappeared.

The positive ions, which gain the electrons, become neutral atoms. Thus, the depletion region (positive electric field) at n-type semiconductor near the p-n junction decreases until it disappears.

The remaining free electrons will cross the depletion region and then enters into the p-semiconductor. The free electrons, which cross the depletion region finds the large number of holes or vacancies in the p-type semiconductor and fills them with electrons. The free electrons which occupy the holes or vacancies will becomes valence electrons and then these electrons get attracted towards the positive terminal of battery or terminates at the positive terminal of battery. Thus, the negative charge carriers (free electrons) that are crossing the depletion region carry the electric current from one point to another point in the p-n junction diode.

Hole current

The positive terminal of the battery supplies large number of holes to the p-type semiconductor and attracts or accepts large number of free electrons from the n-type semiconductor. In other words, the large number of holes begins their journey at the positive terminal whereas the large number of free electrons finishes their journey at the positive terminal.

The holes, which begin their journey from the positive terminal, produce a large positive electric field at p-type semiconductor. The direction this positive electric field is opposite to the direction of negative electric field of depletion region (negative ions) near the p-n junction.

Due to the large number of positive charge carriers (holes) at p-type semiconductor, they get repelled from each other and try to move from higher concentration region (p-type semiconductor) to a lower concentration region (n-type semiconductor). However, before crossing the depletion region, some of the holes finds the negative ions and replaces the electrons position with holes. Thus, the holes are disappeared.

The negative ions, which lose the electrons, become neutral atoms. Thus, the depletion region or negative ions (negative electric field) at p-type semiconductor near the p-n junction decreases until it disappears.

The remaining holes will cross the depletion region and attracted to the negative terminal of battery or terminate at the negative terminal of battery. Thus, the positive charge carriers (holes) that are crossing the depletion region carry the electric current from one point to another point in the p-n junction diode.

Conclusion

Remember, holes are nothing but vacancies created when the electrons left an atom. In p-type semiconductors, the valence electrons move from one atom to another atom whereas holes move in opposite direction. However, holes are the majority in p-type semiconductor. Hence, holes are considered as the charge carriers in the p-type semiconductor, which carry electric current from one point to another point.

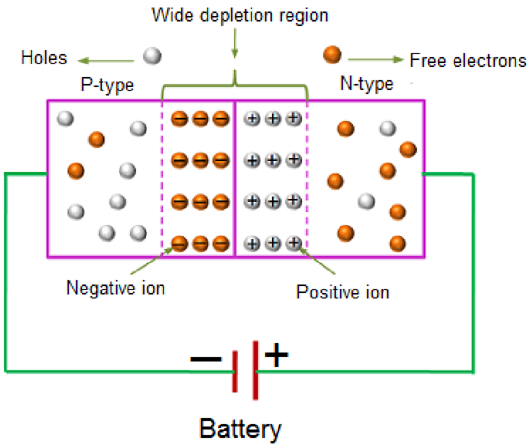
The actual direction of current is the direction of free electrons (from n-side to p-side). However, the conventional direction of electric current is the direction of holes (from p-side to n-side).

Reverse biased p-n junction diode

The process by which, a p-n junction diode blocks the electric current in the presence of applied voltage is called reverse biased p-n junction diode.

In reverse biased p-n junction diode, the positive terminal of the battery is connected to the n-type semiconductormaterial and the negative terminal of the battery is connected to the p-type semiconductor material.

When the external voltage is applied to the p-n junction diode in such a way that, negative terminal is connected to the p-type semiconductor and positive terminal is connected to the n-type semiconductor, holes from the p-side are attracted towards the negative terminal whereas free electrons from the n-side are attracted towards the positive terminal.



Reverse bias

In reverse biased p-n junction diode, the free electrons begin their journey at the negative terminal whereas holes begin their journey at the positive terminal. Free electrons, which begin their journey at the negative terminal, find large number of holes at the p-type semiconductor and fill them with electrons. The atom, which gains an extra electron, becomes a charged atom or negative ion or motionless charge. These negative ions at p-n junction (p-side) oppose the flow of free electrons from n-side.

On the other hand, holes or positive charges, which begin their journey at the positive terminal, find large of free electrons at the n-type semiconductor and replace the electrons position with holes. The atom, which loses an electron, becomes a charged atom or positive ion. These positive ions at p-n junction (n-side) oppose the flow of positive charge carriers (holes) from p-side.

If the reverse biased voltage applied on the p-n junction diode is further increased, then even more number of free electrons and holes are pulled away from the p-n junction. This increases the width of depletion region. Hence, the width of the depletion region increases with increase in voltage. The wide depletion region of the p-n junction diode completely blocks the majority charge carriers. Hence, majority charge carriers cannot carry the electric current.

However, p-n junction diode allows the minority charge carriers. The positive terminal of the battery pushes the holes (minority carriers) towards the p-type semiconductor. In the similar way, negative terminal of the battery pushes the free electrons (minority carriers) towards the n-type semiconductor.

The positive charge carriers (holes) which cross the p-n junction are attracted towards the negative terminal of the battery. On the other hand, the negative charge carriers (free electrons) which cross the p-n junction are attracted towards the positive terminal of the battery. Thus, the minority charge carriers carry the electric current in reverse biased p-n junction diode.

The electric current carried by the minority charge carriers is very small. Hence, minority carrier current is considered as negligible.

V-I characteristics of p-n junction diode

The V-I characteristics or voltage-current characteristics of the p-n junction diode is shown in the below figure. The horizontal line in the below figure represents the amount of voltage applied across the p-n junction diode whereas the vertical line represents the amount of current flows in the p-n junction diode.

Forward V-I characteristics of p-n junction diode

If the positive terminal of the battery is connected to the p-type semiconductor and the negative terminal of the battery is connected to the n-type semiconductor, the diode is said to be in forward bias. In forward biased p-n junction diode, VF represents the forward voltage whereas IF represents the forward current.

Forward V-I characteristics of silicon diode

If the external voltage applied on the silicon diode is less than 0.7 volts, the silicon diode allows only a small electric current. However, this small electric current is considered as negligible.

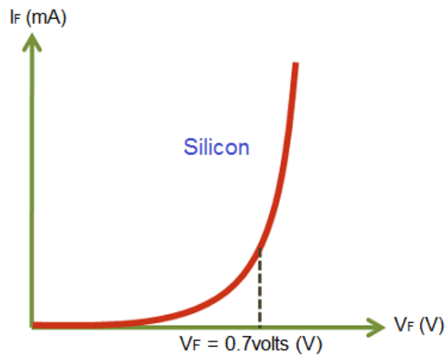


Fig: Forward characteristics of silicon diode

When the external voltage applied on the silicon diode reaches 0.7 volts, the p-n junction diode starts allowing large electric current through it. At this point, a small increase in voltage increases the electric current rapidly. The forward voltage at which the silicon diode starts allowing large electric current is called cut-in voltage. The cut-in voltage for silicon diode is approximately 0.7 volts.

Forward V-I characteristics of germanium diode

If the external voltage applied on the germanium diode is less than 0.3 volts, the germanium diode allows only a small electric current. However, this small electric current is considered as negligible.

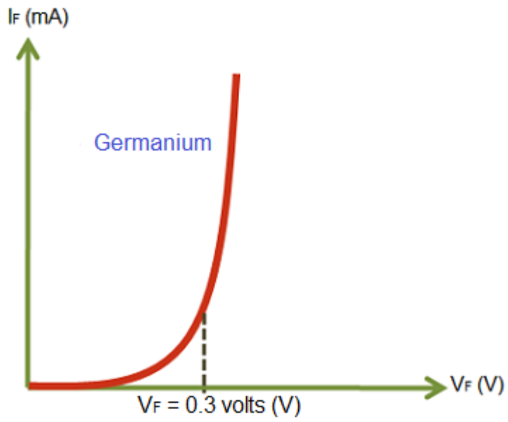


Fig: Forward characteristics of germanium diode

When the external voltage applied on the germanium diode reaches 0.3 volts, the germanium diode starts allowing large electric current through it. At this point, a small increase in voltage increases the electric current rapidly. The forward voltage at which the germanium diode starts allowing large electric current is called cut-in voltage. The cut-in voltage for germanium diode is approximately 0.3 volts.

Reverse V-I characteristics of p-n junction diode

If the negative terminal of the battery is connected to the p-type semiconductor and the positive terminal of the battery is connected to the n-type semiconductor, the diode is said to be in reverse bias. In reverse biased p-n junction diode, VR represents the reverse voltage whereas IR represents the reverse current.

If the external reverse voltage applied on the p-n junction diode is increased, the free electrons from the n-type semiconductor and the holes from the p-type semiconductor are moved away from the p-n junction. This increases the width of depletion region.

The wide depletion region of reverse biased p-n junction diode completely blocks the majority charge carrier current. However, it allows the minority charge carrier current. The free electrons (minority carriers) in the p-type semiconductor and the holes (minority carriers) in the n-type semiconductor carry the electric current. The electric current, which is carried by the minority charge carriers in the p-n junction diode, is called reverse current.

In n-type and p-type semiconductors, very small number of minority charge carriers is present. Hence, a small voltage applied on the diode pushes all the minority carriers towards the junction. Thus, further increase in the external voltage does not increase the electric current. This electric current is called reverse saturation current. In other words, the voltage or point at which the electric current reaches its maximum level and further increase in voltage does not increase the electric current is called reverse saturation current.

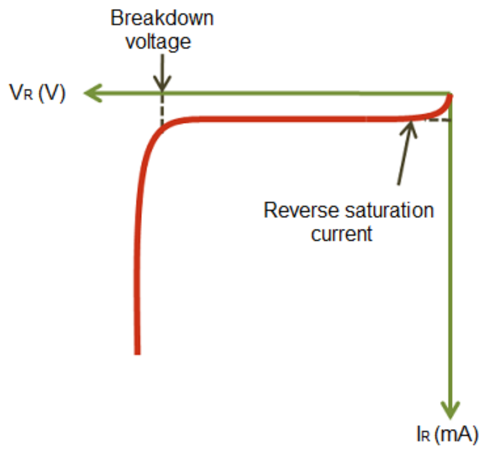


Fig: Reverse characteristics of diode

The reverse saturation current is depends on the temperature. If temperature increases the generation of minority charge carriers increases. Hence, the reverse current increases with the increase in temperature. However, the reverse saturation current is independent of the external reverse voltage. Hence, the reverse saturation current remains constant with the increase in voltage. However, if the voltage applied on the diode is increased continuously, the p-n junction diode reaches to a state where junction breakdown occurs and reverse current increases rapidly.

In germanium diodes, a small increase in temperature generates large number of minority charge carriers. The number of minority charge carriers generated in the germanium diodes is greater than the silicon diodes. Hence, the reverse saturation current in the germanium diodes is greater than the silicon diodes.

Depletion region breakdown

The process by which a depletion region at the p-n junction is destroyed and allows a large reverse current is called depletion region breakdown.

In reverse biased p-n junction diode, the negative terminal of the battery is connected to the p-type semiconductorwhereas the positive terminal of the battery is connected to the n-type semiconductor.

If the reverse voltage applied on the diode is increased, the width of depletion region increases. This depletion region blocks the majority carrier current. However, it allows the minority carrier current. This minority carrier current in the reverse biased p-n junction diode is called reverse current or reverse saturation current.

The width of the depletion region increases with the increase in voltage only up to a certain value or point. If the reverse voltage applied on the p-n junction diode is increased beyond that value or point, the junction breaks down and allows a large reverse current. At this point, a small increase in the voltage will rapidly increase the electric current.

The voltage or point at which junction breakdown occurs is called breakdown voltage. The breakdown voltage of a p-n junction diode is depends on the width of depletion region. The p-n junction diodes with wide depletion region have high breakdown voltage whereas the p-n junction diodes with narrow depletion region have low breakdown voltage.

The depletion region breakdown or junction breakdown occurs in two different methods. Those two different methods are zener breakdown and avalanche breakdown.

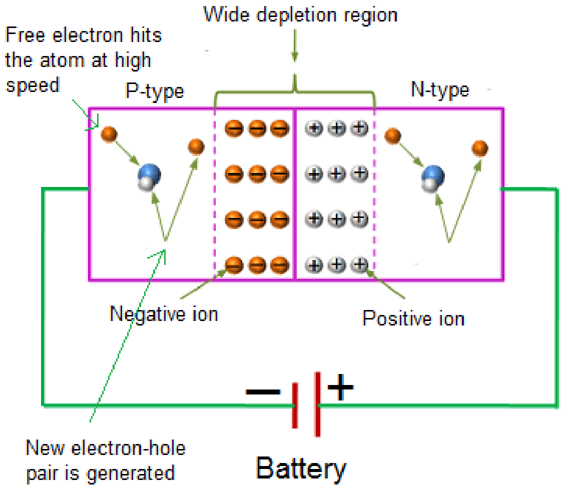
Avalanche breakdown

The avalanche breakdown occurs in lightly doped p-n junction diodes. Lightly doped p-n junction diodes have the wide depletion region. Hence, it is not possible for the external voltage or external electric field to destroy the depletion region directly.

However, it can destroy the depletion region or immobile charge carriers region with the help of minority carriers.

The generation of minority carriers in the reverse biased p-n junction diode does not depend on the applied reverse voltage. It depends only on the temperature. However, the applied reverse voltage on the p-n junction diode supplies energy to the minority carriers (free electrons at p-type and holes at n-type semiconductor). The free electrons, which gain energy from the external voltage are accelerated to greater velocities.

If the voltage applied to the p-n junction diode is further increased to a higher value, the free electrons gain large amount of energy and travel at very high speed.



When these high-speed free electrons collide with the atoms, ions, or valence electrons, they transfer their kinetic energy to the valence electrons. The valence electron, which gain enough energy from the high-speed free electron will breaks the bonding with the parent atom and becomes a free electron. The point at which the free electron left is called a hole. Thus, a free electron at the conduction band and a hole at the valence band is generated as pair.

The newly generated free electron gains energy from the external voltage source and travel at high speed. If this newly generated free electron collides with an atom or ion, another free electron and hole is generated. Likewise, a large number of minority carriers are generated with the few free electrons. This process is called carrier multiplication. In other words, carrier multiplication is the process by which collision of single free electron with the atom or ion leads to the generation of multiple free electrons and holes.

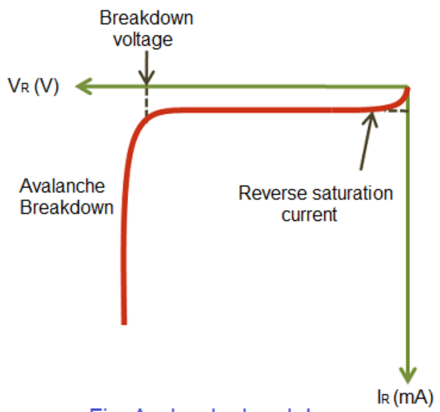


Fig: Avalanche breakdown

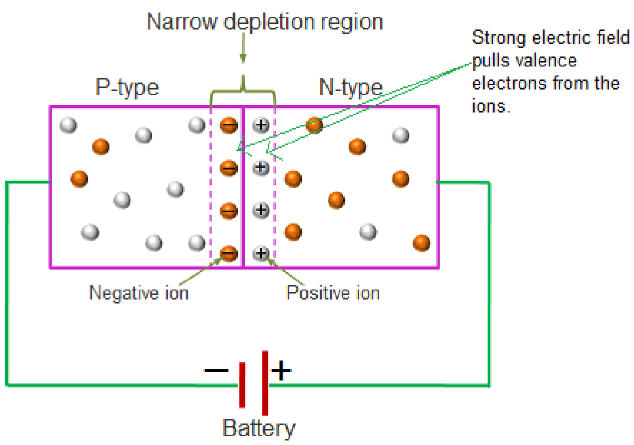
Thus, at high reverse voltage, large number of minority carriers is generated. This large number of minority carriers causes sudden increase in reverse current, which leads to junction breakdown. In other words, the sudden rise in reverse current destroys the motionless carriers region or depletion region.

The high-speed minority carriers, which causes the depletion breakdown or junction breakdown, is called avalanche breakdown.

Zener breakdown

The zener breakdown occurs in the heavily doped p-n junction diodes. Heavily doped p-n junction diodes have narrow depletion region.

If reverse voltage is applied on the narrow depletion p-n junction diode, the immobile ions in the depletion region gains energy from the external voltage. Hence, the electric field of the immobile ions increases. As a result, the overall electric field of the narrow depletion region increases.



If the voltage applied on the p-n junction diode is increased to a higher value, a very strong electric field is built in this narrow depletion region. This strong electric field of narrow depletion region applies force on the valence electrons and pulls them from the valence band.

Thus, free electrons and holes are generated as pair. Likewise, a large number of minority carriers are generated in the depletion region. At this point, a small increase in reverse voltage causes sudden rise in reverse current. This sudden rise of reverse current destroys the depletion region or p-n junction.

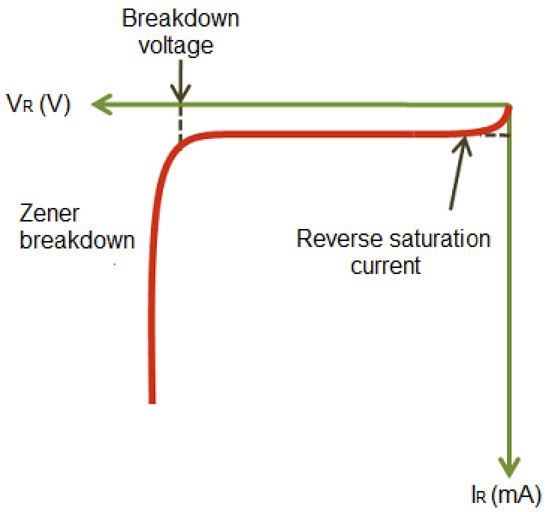


Fig: Zener breakdown

The strong electric field of the narrow depletion region, which causes the junction breakdown or depletion breakdown, is called avalanche breakdown.

Ideal diode

The ideal diode or perfect diode is a two terminal device, which completely allows the electric current without any loss under forward bias and completely blocks the electric current with infinite loss under reverse bias.

Ideal diodes actually do not exist. However, the V-I characteristics of ideal diodes is used to study the diode circuits. In other words, it is used to study the quality of a real diode by comparing it with the ideal diode.

It is actually not possible to design a real diode, which behaves exactly like an ideal diode. However, a well-designed diode behaves almost like a perfect diode or ideal diode.

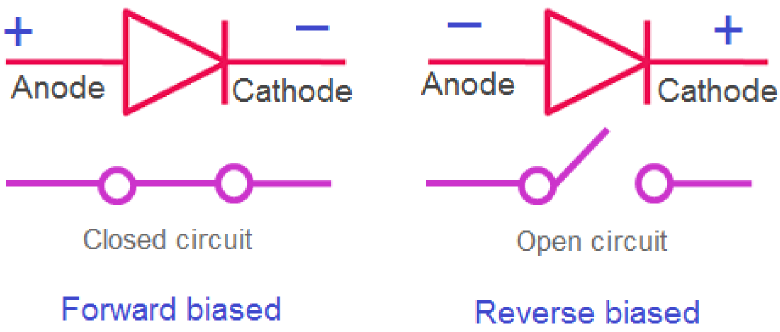
Ideal diode symbol

Ideal diode consists of two terminals: positive terminal and negative terminal. The positive end or positive terminal of the diode is called anode and the negative end or negative terminal of the diode is called cathode. The electric current always flow from anode or positive terminal to the cathode or negative terminal.

If the positive terminal of the battery is connected to the p-type semiconductor and the negative terminal of the battery is connected to the n-type semiconductor, the diode is said to be forward biased.

On the other hand, if the positive terminal of the battery is connected to the n-type semiconductor and the negative terminal of the battery is connected to the p-type semiconductor, the diode is said to be reverse biased.

The symbol of forward biased and reverse biased ideal diode is shown in the below figure.



Under forward biased condition, ideal diode acts as a perfect conductor with zero resistance whereas under reverse biased condition, it acts as a perfect insulator with infinite resistance. In other words, ideal diodes acts as closed circuit or short circuit under forward biased condition and acts as an open circuit or open switch under reverse biased condition.

Ideal diodes does not have depletion region or junction barrier, which resist the flow of electric current. Hence, ideal diode has no voltage drop or voltage loss.

Forward and reverse characteristics of ideal diode

The Forward and reverse characteristics of ideal diode under forward and reverse biased condition is shown in the below figure.

If the forward voltage or positive voltage (VF) (p terminal connected to p-side and n terminal connected to n-side) applied on the diode is equal to zero or greater than zero, the forward electric current (IF) in the ideal diode increases infinitely.

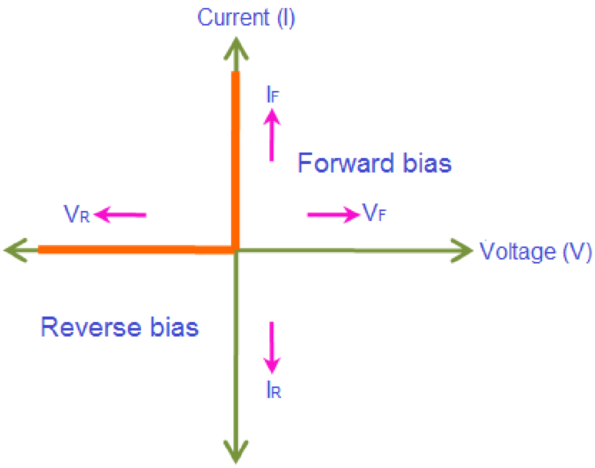


Fig: V-I characteristics of ideal diode

On the other hand, if the reverse voltage or negative voltage (VR) (p terminal connected to n-side and n-terminal connected to p-side) applied on the diode is less than zero, no forward electric current (IF) and reverse electric current (IR) flows in the ideal diode.

Real diode

The real diode or practical diode is a two terminal device, which allow most of the electric current under forward bias, and block most of the electric current under reverse bias.

Under forward bias, real diodes allow most of the electric current and block small electric current whereas under reverse bias, real diodes block most of the electric current and allow small electric current.

Real diode symbol

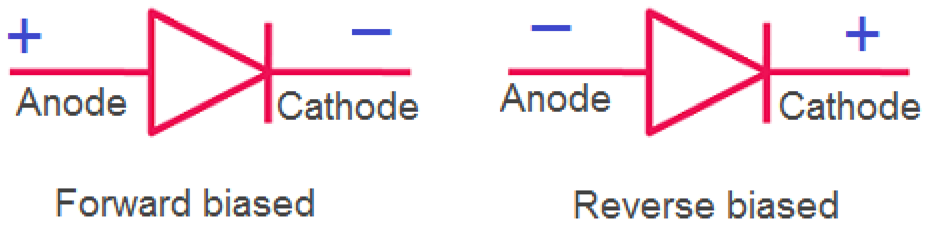
Real diode consists of two terminals. Those two terminals are the positive terminal and the negative terminal. The positive terminal of the real diode is called anode whereas the negative terminal of the diode is called cathode.

The conventional electric current always flows from positive terminal of the battery to the negative terminal of the battery. However, the actual electric current flows from negative terminal of the battery to the positive terminal of the battery.

If the p-type semiconductor is connected to the positive terminal of the battery and the n-type semiconductor is connected to the negative terminal of the battery, the diode is said to be forward biased.

On the other hand, if the p-type semiconductor is connected to the negative terminal of the battery and the n-type semiconductor is connected to the positive terminal of the battery, the diode is said to be reverse biased.

The circuit symbol of forward biased and reverse biased real diode is shown in the below figure.



Ideal diode starts allowing the electric current, once the forward voltage is applied. However, the real diode behaves differently.

The real diodes do not allow the electric current, if the forward voltage is less than the cut-in voltage. However, if the forward voltage applied on the real diode reaches the cut-in voltage, it starts allowing the electric current.

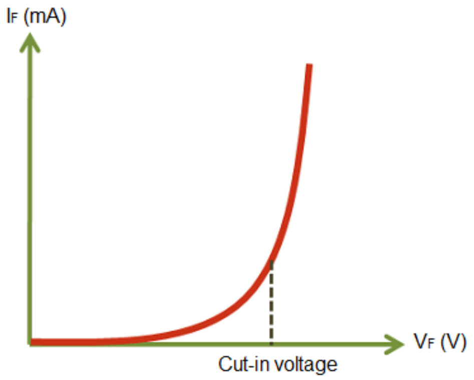
Forward biased real diode

Under forward bias, the positive terminal of the battery is connected to the p-type semiconductor and the negative terminal of the battery is connected to the n-type semiconductor.

In the n-type semiconductor, free electrons are the majority charge carriers and holes are the minority charge carriers whereas in p-type semiconductor, holes are the majority charge carriers and free electrons are the minority charges carriers.

In forward biased real diode, the majority charge carriers carry most of the electric current. The positive charge carriers from p-side to n-side carry the conventional electric current. However, the free electrons from n-side to p-side carry the actual electric current.

When the forward voltage is applied on the real diode, it does not allow the electric current up to a certain voltage because the depletion region present at the p-n junction block the electric current.



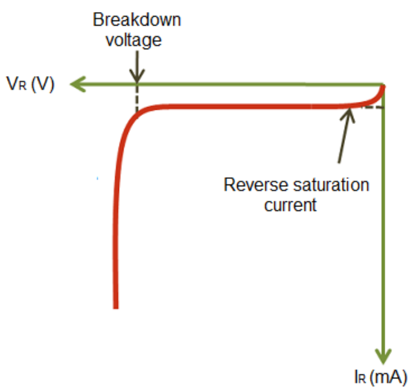
However, once it reaches that voltage, real diode starts allowing the electric current. In other words, if the voltage applied on the real diode reaches 0.7 for silicon and 0.3 for germanium, the electric current rises suddenly. The voltage at which real diodes starts allowing the electric current is called cut-in voltage.

This cut-in voltage is different for silicon and germanium diodes. For silicon diode, the knee voltage or cut-in voltage is approximately 0.7 volts whereas for germanium diode, the knee voltage or cut-in voltage is approximately 0.3 volts.

Reverse biased real diode

Under reverse bias, the negative terminal of the battery is connected to the p-type semiconductor and the positive terminal of the battery is connected to the n-type semiconductor.

When the reverse voltage is applied on the real diode, the free electrons at the n-side moves away from the p-n junction and attracted towards the positive terminal of the battery. Similarly, the holes at the p-side moves away from the p-n junction and attracted towards the negative terminal of the battery. As a result, the width of the depletion region increases.



This depletion region blocks the majority carrier current (free electrons at n-side and holes at p-side). However, it allows the minority charge carrier current (free electrons at p-side and holes at n-side). This minority charge carrier current in the reverse biased real diode is called reverse saturation current.

In the reverse biased real diode, the width of the depletion region increases with the increase in voltage up to a certain point. If it reaches that point, junction breaks down and reverse current increases rapidly. The voltage at which the junction breaks down is called the breakdown voltage.

Diode junction capacitance

In a p-n junction diode, two types of capacitance take place. They are,

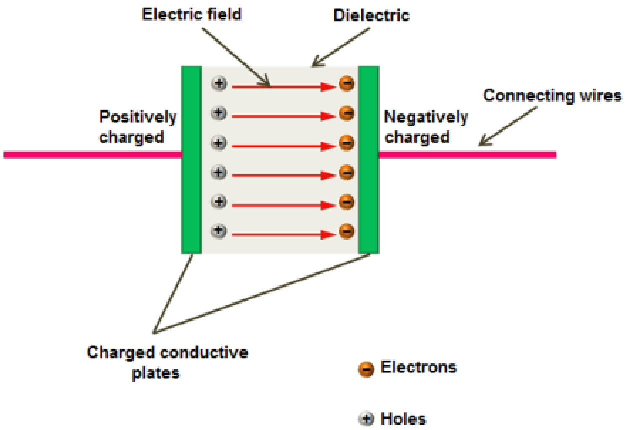
Transition capacitance (CT)

Diffusion capacitance (CD)

Transition capacitance (CT)

We know that capacitors store electric charge in the form of electric field. This charge storage is done by using two electrically conducting plates (placed close to each other) separated by an insulating material called dielectric.

The conducting plates or electrodes of the capacitor are good conductors of electricity. Therefore, they easily allow electric current through them. On the other hand, dielectric material or medium is poor conductor of electricity. Therefore, it does not allow electric current through it. However, it efficiently allows electric field.



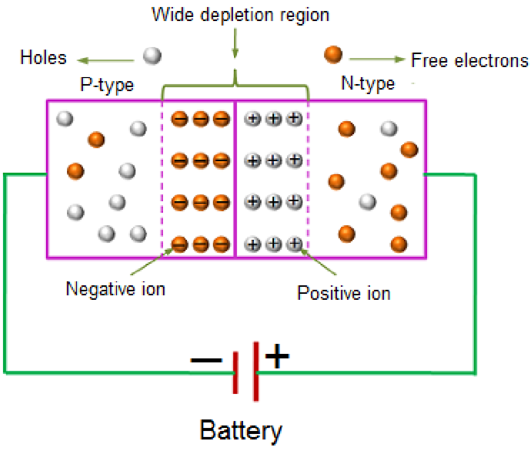
When voltage is applied to the capacitor, charge carriers starts flowing through the conducting wire. When these charge carriers reach the electrodes of the capacitor, they experience a strong opposition from the dielectric or insulating material. As a result, a large number of charge carriers are trapped at the electrodes of the capacitor. These charge carriers cannot move between the plates. However, they exerts electric field between the plates. The charge carriers which are trapped near the dielectric material will stores electric charge. The ability of the material to store electric charge is called capacitance.

In a basic capacitor, the capacitance is directly proportional to the size of electrodes or plates and inversely proportional to the distance between two plates.

Just like the capacitors, a reverse biased p-n junction diode also stores electric charge at the depletion region. The depletion region is made of immobile positive and negative ions.

In a reverse biased p-n junction diode, the p-type and n-type regions have low resistance. Hence, p-type and n-type regions act like the electrodes or conducting plates of the capacitor. The depletion region of the p-n junction diode has high resistance. Hence, the depletion region acts like the dielectric or insulating material. Thus, p-n junction diode can be considered as a parallel plate capacitor.

In depletion region, the electric charges (positive and negative ions) do not move from one place to another place. However, they exert electric field or electric force. Therefore, charge is stored at the depletion region in the form of electric field. The ability of a material to store electric charge is called capacitance. Thus, there exists a capacitance at the depletion region.



Reverse bias

The capacitance at the depletion region changes with the change in applied voltage. When reverse bias voltage applied to the p-n junction diode is increased, a large number of holes (majority carriers) from p-side and electrons (majority carriers) from n-side are moved away from the p-n junction. As a result, the width of depletion region increases whereas the size of p-type and n-type regions (plates) decreases.

We know that capacitance means the ability to store electric charge. The p-n junction diode with narrow depletion width and large p-type and n-type regions will store large amount of electric charge whereas the p-n junction diode with wide depletion width and small p-type and n-type regions will store only a small amount of electric charge. Therefore, the capacitance of the reverse bias p-n junction diode decreases when voltage increases.

In a forward biased diode, the transition capacitance exist. However, the transition capacitance is very small compared to the diffusion capacitance. Hence, transition capacitance is neglected in forward biased diode.

The amount of capacitance changed with increase in voltage is called transition capacitance. The transition capacitance is also known as depletion region capacitance, junction capacitance or barrier capacitance. Transition capacitance is denoted as CT.

The change of capacitance at the depletion region can be defined as the change in electric charge per change in voltage.

$$CT = dQ / dV$$

Where,

CT = Transition capacitance

dQ = Change in electric charge

dV = Change in voltage

The transition capacitance can be mathematically written as,

$$CT = \varepsilon A / W$$

Where,

ε = Permittivity of the semiconductor

A = Area of plates or p-type and n-type regions

W = Width of depletion region

Diffusion capacitance (CD)

Diffusion capacitance occurs in a forward biased p-n junction diode. Diffusion capacitance is also sometimes referred as storage capacitance. It is denoted as CD.

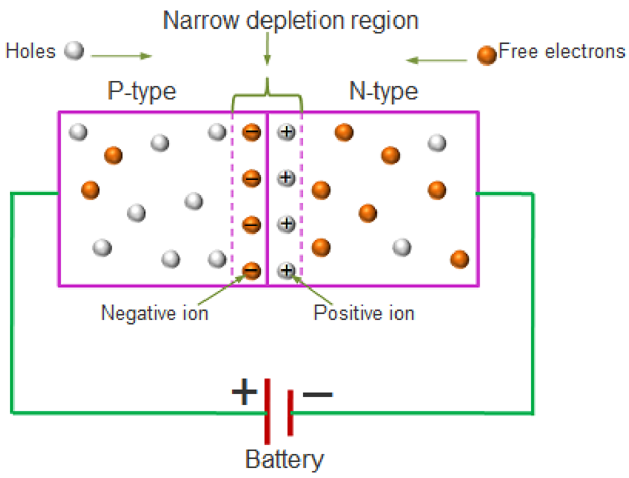
In a forward biased diode, diffusion capacitance is much larger than the transition capacitance. Hence, diffusion capacitance is considered in forward biased diode.

The diffusion capacitance occurs due to stored charge of minority electrons and minority holes near the depletion region.

When forward bias voltage is applied to the p-n junction diode, electrons (majority carriers) in the n-region will move into the p-region and recombines with the holes. In the similar way, holes in the p-region will move into the n-region and recombines with electrons. As a result, the width of depletion region decreases.

The electrons (majority carriers) which cross the depletion region and enter into the p-region will become minority carriers of the p-region similarly; the holes (majority carriers) which cross the depletion region and enter into the n-region will become minority carriers of the n-region.

A large number of charge carriers, which try to move into another region will be accumulated near the depletion region before they recombine with the majority carriers. As a result, a large amount of charge is stored at both sides of the depletion region.



Forward bias

The accumulation of holes in the n-region and electrons in the p-region is separated by a very thin depletion region or depletion layer. This depletion region acts like dielectric or insulator of the capacitor and charge stored at both sides of the depletion layer acts like conducting plates of the capacitor.

Diffusion capacitance is directly proportional to the electric current or applied voltage. If large electric current flows through the diode, a large amount of charge is accumulated near the depletion layer. As a result, large diffusion capacitance occurs.

In the similar way, if small electric current flows through the diode, only a small amount of charge is accumulated near the depletion layer. As a result, small diffusion capacitance occurs.

When the width of depletion region decreases, the diffusion capacitance increases. The diffusion capacitance value will be in the range of nano farads (nF) to micro farads (μF).

The formula for diffusion capacitance is

$$C_D = dQ / dV$$

Where,

CD = Diffusion capacitance

dQ = Change in number of minority carriers stored outside the depletion region

dV = Change in voltage applied across diode

Diode resistance

A p-n junction diode allows electric current in one direction and blocks electric current in another direction. It allows electric current when it is forward biased and blocks electric current when it is reverse biased. However, no diode allows electric current completely even in forward biased condition.

The depletion region present in a diode acts like barrier to electric current. Hence, it offers resistance to the electric current. Also, the atoms present in the diode provide some resistance to the electric current.

When charge carriers (free electrons and holes) flowing through the diode collides with atoms, they loseenergy in the form of heat. Thus, depletion region and atoms offer resistance to the electric current.

When forward biased voltage is applied to the p-n junction diode, the width of depletion region decreases. However, the depletion region cannot be completely vanished. There exists a thin depletion region or depletion layer in the forward biased diode. Therefore, a thin depletion region and atoms in the diode offer some resistance to electric current. This resistance is called forward resistance.

When the diode is reverse biased, the width of depletion region increases. As a result, a large number of charge carriers (free electrons and holes) flowing through the diode will be blocked by the depletion region.

In a reverse biased diode, only a small amount of electric current flows. The minority carriers present in the diode carry this electric current. Thus, reverse biased diode offer large resistance to the electric current. This resistance is called reverse resistance.

The two types of resistance takes place in the p-n junction diode are:

Forward resistance

Reverse resistance

Forward resistance

Forward resistance is a resistance offered by the p-n junction diode when it is forward biased.

In a forward biased p-n junction diode, two type of resistance takes place based on the voltage applied.

The two types of resistance takes place in forward biased diode are

Static resistance or DC resistance

Dynamic resistance or AC resistance

Static resistance or DC resistance

When forward biased voltage is applied to a diode that is connected to a DC circuit, a DC or direct current flows through the diode. Direct current or electric current is nothing but the flow of charge carriers (free electrons or holes) through a conductor. In DC circuit, the charge carriers flow steadily in single direction or forward direction.

The resistance offered by a p-n junction diode when it is connected to a DC circuit is called static resistance.

Static resistance is also defined as the ratio of DC voltage applied across diode to the DC current or direct current flowing through the diode.

The resistance offered by the p-n junction diode under forward biased condition is denoted as Rf.

$$R_f = \frac{\text{DC voltage}}{\text{DC current}}$$

Dynamic resistance or AC resistance

The dynamic resistance is the resistance offered by the p-n junction diode when AC voltage is applied.

When forward biased voltage is applied to a diode that is connected to AC circuit, an AC or alternating current flows though the diode.

In AC circuit, charge carriers or electric current does not flow in single direction. It flows in both forward and reverse direction.

Dynamic resistance is also defined as the ratio of change in voltage to the change in current. It is denoted as rf.

$$r_f = \frac{\text{Change in voltage}}{\text{Change in current}}$$

Reverse resistance

Reverse resistance is the resistance offered by the p-n junction diode when it is reverse biased.

When reverse biased voltage is applied to the p-n junction diode, the width of depletion region increases. This depletion region acts as barrier to the electric current. Hence, a large amount of electric current is blocked by the depletion region. Thus, reverse biased diode offer large resistance to the electric current.

The resistance offered by the reverse biased p-n junction diode is very large compared to the forward biased diode. The reverse resistance is in the range of mega ohms (MΩ).

Zener diode

A normal p-n junction diode allows electric current only in forward biased condition. When forward biased voltage is applied to the p-n junction diode, it allows large amount of electric current and blocks only a small amount of electric current. Hence, a forward biased p-n junction diode offer only a small resistance to the electric current.

When reverse biased voltage is applied to the p-n junction diode, it blocks large amount of electric current and allows only a small amount of electric current. Hence, a reverse biased p-n junction diode offer large resistance to the electric current.

If reverse biased voltage applied to the p-n junction diode is highly increased, a sudden rise in current occurs. At this point, a small increase in voltage will rapidly increases the electric current. This sudden rise in electric current causes a junction breakdown called zener or avalanche breakdown. The voltage at which zener breakdown occurs is called zener voltage and the sudden increase in current is called zener current.

A normal p-n junction diode does not operate in breakdown region because the excess current permanently damages the diode. Normal p-n junction diodes are not designed to operate in reverse breakdown region. Therefore, a normal p-n junction diode does not operate in reverse breakdown region.

What is zener diode?

A zener diode is a special type of device designed to operate in the zener breakdown region. Zener diodes acts like normal p-n junction diodes under forward biased condition. When forward biased voltage is applied to the zener diode it allows large amount of electric current and blocks only a small amount of electric current.

Zener diode is heavily doped than the normal p-n junction diode. Hence, it has very thin depletion region. Therefore, zener diodes allow more electric current than the normal p-n junction diodes.

Zener diode allows electric current in forward direction like a normal diode but also allows electric current in the reverse direction if the applied reverse voltage is greater than the zener voltage. Zener diode is always connected in reverse direction because it is specifically designed to work in reverse direction.

Zener diode definition

A zener diode is a p-n junction semiconductor device designed to operate in the reverse breakdown region. The breakdown voltage of a zener diode is carefully set by controlling the doping level during manufacture.

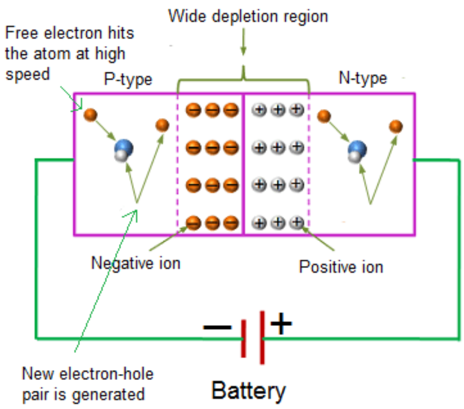
The name zener diode was named after the American physicist Clarance Melvin Zener who discovered the zener effect. Zener diodes are the basic building blocks of electronic circuits. They are widely used in all kinds of electronic equipments. Zener diodes are mainly used to protect electronic circuits from over voltage.

Breakdown in zener diode

There are two types of reverse breakdown regions in a zener diode: avalanche breakdown and zener breakdown.

Avalanche breakdown

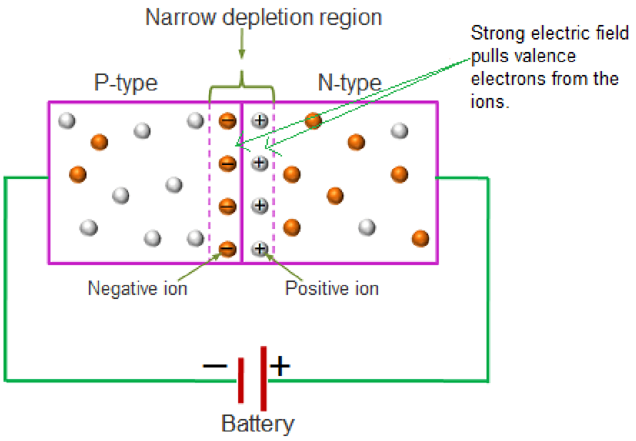
The avalanche breakdown occurs in both normal diodes and zener diodes at high reverse voltage. When high reverse voltage is applied to the p-n junction diode, the free electrons (minority carriers) gains large amount ofenergy and accelerated to greater velocities.



The free electrons moving at high speed will collides with the atoms and knock off more electrons. These electrons are again accelerated and collide with other atoms. Because of this continuous collision with the atoms, a large number of free electrons are generated. As a result, electric current in the diode increases rapidly. This sudden increase in electric current may permanently destroys the normal diode. However, avalanche diodes may not be destroyed because they are carefully designed to operate in avalanche breakdown region. Avalanche breakdown occurs in zener diodes with zener voltage (Vz) greater than 6V.

Zener breakdown

The zener breakdown occurs in heavily doped p-n junction diodes because of their narrow depletion region. When reverse biased voltage applied to the diode is increased, the narrow depletion region generates strongelectric field.



When reverse biased voltage applied to the diode reaches close to zener voltage, the electric field in the depletion region is strong enough to pull electrons from their valence band. The valence electrons which gains sufficient energy from the strong electric field of depletion region will breaks bonding with the parent atom. Thevalance electrons which break bonding with parent atom will become free electrons. This free electrons carry electric current from one place to another place. At zener breakdown region, a small increase in voltage will rapidly increases the electric current.

Zener breakdown occurs at low reverse voltage whereas avalanche breakdown occurs at high reverse voltage.

Zener breakdown occurs in zener diodes because they have very thin depletion region.

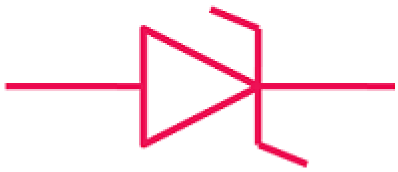
Breakdown region is the normal operating region for a zener diode.

Zener breakdown occurs in zener diodes with zener voltage (Vz) less than 6V.

Symbol of zener diode

The symbol of zener diode is shown in below figure. Zener diode consists of two terminals: cathode and anode.

Zener diode symbol



In zener diode, electric current flows from both anode to cathode and cathode to anode.

The symbol of zener diode is similar to the normal p-n junction diode, but with bend edges on the vertical bar.

VI characteristics of zener diode

The VI characteristics of a zener diode is shown in the below figure. When forward biased voltage is applied to the zener diode, it works like a normal diode. However, when reverse biased voltage is applied to the zener diode, it works in different manner.

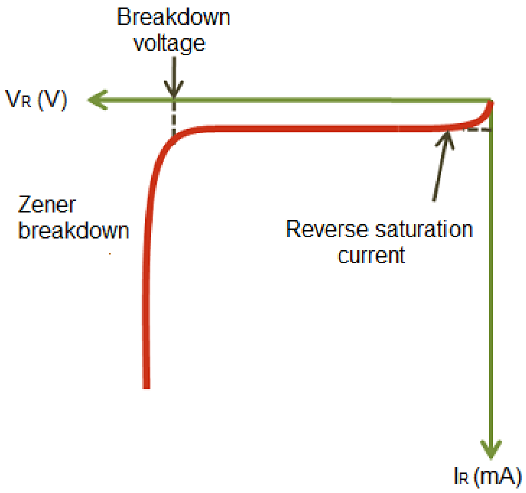


Fig: Zener breakdown

When reverse biased voltage is applied to a zener diode, it allows only a small amount of leakage current until the voltage is less than zener voltage. When reverse biased voltage applied to the zener diode reaches zener voltage, it starts allowing large amount of electric current. At this point, a small increase in reverse voltage will rapidly increases the electric current. Because of this sudden rise in electric current, breakdown occurs called zener breakdown. However, zener diode exhibits a controlled breakdown that does damage the device.

The zener breakdown voltage of the zener diode is depends on the amount of doping applied. If the diode is heavily doped, zener breakdown occurs at low reverse voltages. On the other hand, if the diode is lightly doped, the zener breakdown occurs at high reverse voltages. Zener diodes are available with zener voltages in the range of 1.8V to 400V.

Advantages of zener diode

Power dissipation capacity is very high

High accuracy

Small size

Low cost

Applications of zener diode

It is normally used as voltage reference

Zener diodes are used in voltage stabilizers or shunt regulators.

Zener diodes are used in switching operations

Zener diodes are used in clipping and clamping circuits.

Zener diodes are used in various protection circuits

Avalanche diode

What is avalanche diode?

An avalanche diode is a special type of semiconductor device designed to operate in reverse breakdown region. Avalanche diodes are used as relief valves (a type of valve used to control the pressure in a system) to protect electrical systems from excess voltages.

Construction of avalanche diode

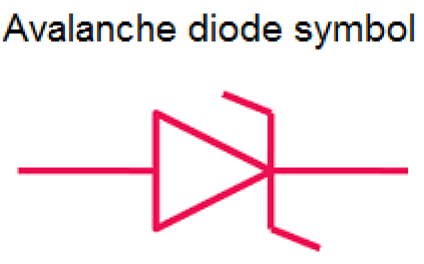
Avalanche diodes are generally made from silicon or other semiconductor materials. The construction of avalanche diode is similar to zener diode but the doping level in avalanche diode differs from zener diode.

Zener diodes are heavily doped. Therefore, the width of depletion region in zener diode is very thin. Because of this thin depletion layer or region, reverse breakdown occurs at lower voltages in zener diode.

On the other hand, avalanche diodes are lightly doped. Therefore, the width of depletion layer in avalanche diode is very wide compared to the zener diode. Because of this wide depletion region, reverse breakdown occurs at higher voltages in avalanche diode. The breakdown voltage of avalanche diode is carefully set by controlling the doping level during manufacture.

Symbol of avalanche diode

The symbol of avalanche and zener diode is same. Avalanche diode consists of two terminals: anode and cathode. The symbol of avalanche diode is shown in below figure.



The symbol of avalanche diode is similar to the normal diode but with the bend edges on the vertical bar.

How avalanche diode works?

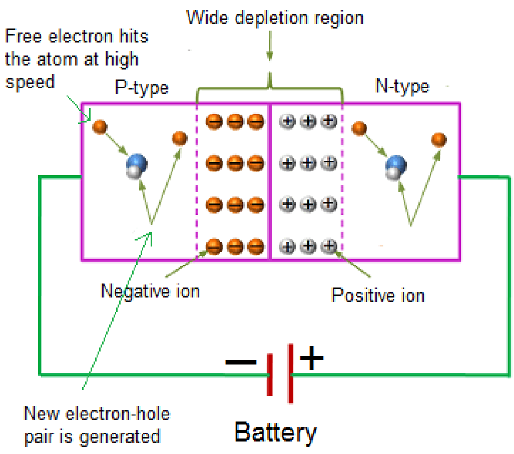
A normal p-n junction diode allows electric current only in forward direction whereas an avalanche diode allows electric current in both forward and reverse directions. However, avalanche diode is specifically designed to operate in reverse biased condition.

Avalanche diode allows electric current in reverse direction when reverse bias voltage exceeds the breakdown voltage. The point or voltage at which electric current increases suddenly is called breakdown voltage.

When the reverse bias voltage applied to the avalanche diode exceeds the breakdown voltage, a junction breakdown occurs. This junction breakdown is called avalanche breakdown.

When forward bias voltage is applied to the avalanche diode, it works like a normal p-n junction diode by allowing electric current through it.

When reverse bias voltage is applied to the avalanche diode, the free electrons (majority carriers) in the n-type semiconductor and the holes (majority carriers) in the p-type semiconductor are moved away from the junction. As a result, the width of depletion region increases. Therefore, the majority carriers will not carry electric current. However, the minority carriers (free electrons in p-type and holes in n-type) experience a repulsive force from external voltage.



As a result, the minority carriers flow from p-type to n-type and n-type to p-type by carrying the electric current. However, electric current carried by minority carriers is very small. This small electric current carried by minority carriers is called reverse leakage current.

If the reverse bias voltage applied to the avalanche diode is further increased, the minority carriers (free electrons or holes) will gain large amount of energy and accelerated to greater velocities.

The free electrons moving at high speed will collide with the atoms and transfer their energy to the valence electrons.

The valance electrons which gains enough energy from the high-speed electrons will be detached from the parent atom and become free electrons. These free electrons are again accelerated. When these free electrons again collide with other atoms, they knock off more electrons.

Because of this continuous collision with the atoms, a large number of minority carriers (free electrons or holes) are generated. These large numbers of free electrons carry excess current in the diode.

When the reverse voltage applied to the avalanche diode continuously increases, at some point the junction breakdown or avalanche breakdown occurs. At this point, a small increase in voltage will suddenly increases the electric current. This sudden increase of electric current may permanently destroys the normal p-n junction diode. However, avalanche diodes may not be destroyed because they are carefully designed to operate in avalanche breakdown region.

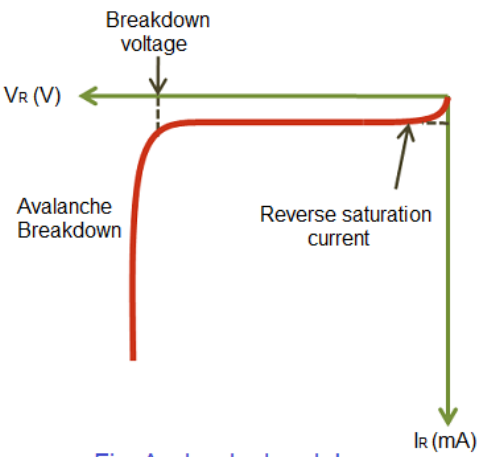


Fig: Avalanche breakdown

The breakdown voltage of the avalanche diode depends on the doping density. Increasing the doping density will decreases the breakdown voltage of the avalanche diode.

Applications of avalanche diodes

Avalanche diodes can be used as white noise generators.

Avalanche diodes are used in protecting circuits.

Photodiode

Introduction

A photodiode is a p-n junction or pin semiconductor device that consumes light energy to generate electric current. It is also sometimes referred as photo-detector, photo-sensor, or light detector.

Photodiodes are specially designed to operate in reverse bias condition. Reverse bias means that the p-side of the photodiode is connected to the negative terminal of the battery and n-side is connected to the positive terminal of the battery.

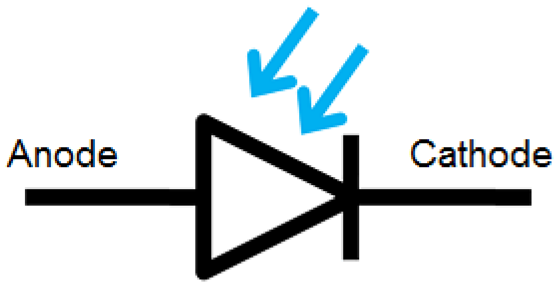
Photodiode is very sensitive to light so when light or photons falls on the photodiode it easily converts light into electric current. Solar cell is also known as large area photodiode because it converts solar energy or light energy into electric energy. However, solar cell works only at bright light.

The construction and working of photodiode is almost similar to the normal p-n junction diode. PIN (p-type, intrinsic and n-type) structure is mostly used for constructing the photodiode instead of p-n (p-type and n-type) junction structure because PIN structure provide fast response time. PIN photodiodes are mostly used in high-speed applications.

In a normal p-n junction diode, voltage is used as the energy source to generate electric current whereas in photodiodes, both voltage and light are used as energy source to generate electric current.

Photodiode symbol

The symbol of photodiode is similar to the normal p-n junction diode except that it contains arrows striking the diode. The arrows striking the diode represent light or photons.



Photodiode symbol

A photodiode has two terminals: a cathode and an anode.

Objectives and limitations of photodiode

Photodiode should be always operated in reverse bias condition.

Applied reverse bias voltage should be low.

Generate low noise

High gain

High response speed

High sensitivity to light

Low sensitivity to temperature

Low cost

Small size

Long lifetime

How photodiode works?

A normal p-n junction diode allows a small amount of electric current under reverse bias condition. To increase the electric current under reverse bias condition, we need to generate more minority carriers.

The external reverse voltage applied to the p-n junction diode will supply energy to the minority carriers but not increase the population of minority carriers.

However, a small number of minority carriers are generated due to external reverse bias voltage. The minority carriers generated at n-side or p-side will recombine in the same material before they cross the junction. As a result, no electric current flows due to these charge carriers. For example, the minority carriers generated in the p-type material experience a repulsive force from the external voltage and try to move towards n-side. However, before crossing the junction, the free electrons recombine with the holes within the same material. As a result, no electric current flows.

To overcome this problem, we need to apply external energy directly to the depletion region to generate more charge carriers.

A special type of diode called photodiode is designed to generate more number of charge carriers in depletion region. In photodiodes, we use light or photons as the external energy to generate charge carriers in depletion region.

Types of photodiodes

The working operation of all types of photodiodes is same. Different types of photodiodes are developed based on specific application. For example, PIN photodiodes are developed to increase the response speed. PIN photodiodes are used where high response speed is needed.

The different types of photodiodes are

PN junction photodiode

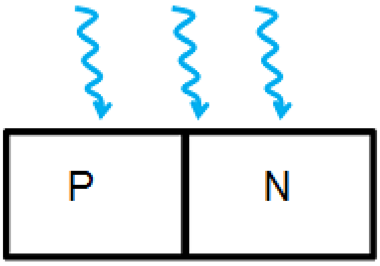
PIN photodiode

Avalanche photodiode

Among all the three photodiodes, PN junction and PIN photodiodes are most widely used.

PN junction photodiode

PN junction photodiodes are the first form of photodiodes. They are the most widely used photodiodes before the development of PIN photodiodes. PN junction photodiode is also simply referred as photodiode.Nowadays, PN junction photodiodes are not widely used.

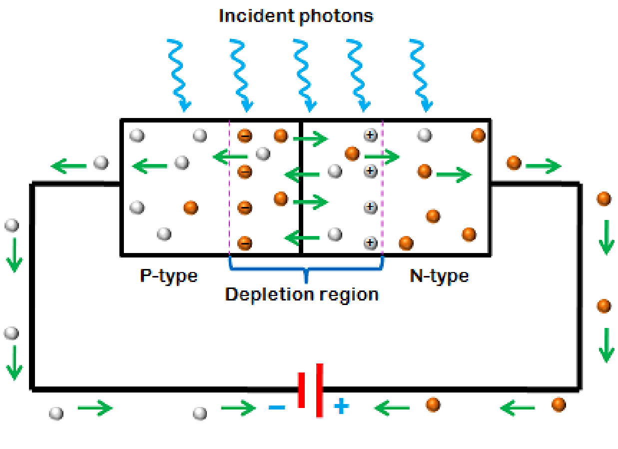


When external light energy is supplied to the p-n junction photodiode, the valence electrons in the depletion region gains energy.

If the light energy applied to the photodiode is greater the band-gap of semiconductor material, the valence electrons gain enough energy and break bonding with the parent atom. The valence electron which breaks bonding with the parent atom will become free electron. Free electrons moves freely from one place to another place by carrying the electric current.

When the valence electron leave the valence shell an empty space is created in the valence shell at which valence electron left. This empty space in the valence shell is called a hole. Thus, both free electrons and holes are generated as pairs. The mechanism of generating electron-hole pair by using light energy is known as the inner photoelectric effect.

The minority carriers in the depletion region experience force due to the depletion region electric field and the external electric field. For example, free electrons in the depletion region experience repulsive and attractive force from the negative and positive ions present at the edge of depletion region at p-side and n-side. As a result, free electrons move towards the n region. When the free electrons reaches n region, they are attracted towards the positive terminals of the battery. In the similar way, holes move in opposite direction.



PN Junction photodiode

The strong depletion region electric field and the external electric field increase the drift velocity of the free electrons. Because of this high drift velocity, the minority carriers (free electrons and holes) generated in the depletion region will cross the p-n junction before they recombine with atoms. As a result, the minority carrier current increases.

When no light is applied to the reverse bias photodiode, it carries a small reverse current due to external voltage. This small electric current under the absence of light is called dark current. It is denoted by I λ.

In a photodiode, reverse current is independent of reverse bias voltage. Reverse current is mostly depends on the light intensity.

In photodiodes, most of the electric current is carried by the charge carriers generated in the depletion region because the charge carriers in depletion region has high drift velocity and low recombination rate whereas the charge carriers in n-side or p-side has low drift velocity and high recombination rate. The electric current generated in the photodiode due to the application of light is called photocurrent.

The total current through the photodiode is the sum of the dark current and the photocurrent. The dark current must be reduced to increase the sensitivity of the device.

The electric current flowing through a photodiode is directly proportional to the incident number of photons.

PIN photodiode

PIN photodiodes are developed from the PN junction photodiodes. The operation of PIN photodiode is similar to the PN junction photodiode except that the PIN photodiode is manufactured differently to improve its performance.

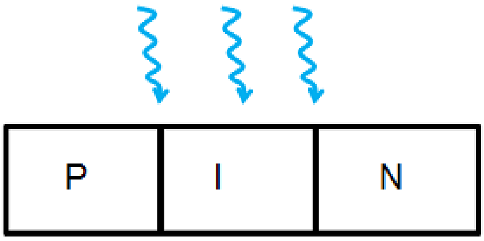
The PIN photodiode is developed to increase the minority carrier current and response speed.

PIN photodiodes generate more electric current than the PN junction photodiodes with the same amount of light energy.

Layers of PIN photodiode

A PN junction photodiode is made of two layers namely p-type and n-type semiconductor whereas PIN photodiode is made of three layers namely p-type, n-type and intrinsic semiconductor.

In PIN photodiode, an addition layer called intrinsic semiconductor is placed between the p-type and n-type semiconductor to increase the minority carrier current.



P-type semiconductor

If trivalent impurities are added to the intrinsic semiconductor, a p-type semiconductor is formed.

In p-type semiconductors, the number of free electrons in the conduction band is lesser than the number of holes in the valence band. Therefore, holes are the majority charge carriers and free electrons are the minority charge carriers. In p-type semiconductors, holes carry most of the electric current.

N-type semiconductor

If pentavalent impurities are added to the intrinsic semiconductor, an n-type semiconductor is formed.

In n-type semiconductors, the number of free electrons in the conduction band is greater than the number of holes in the valence band. Therefore, free electrons are the majority charge carriers and holes are the minority charge carriers. In n-type semiconductors, free electrons carry most of the electric current.

Intrinsic semiconductor

Intrinsic semiconductors are the pure form of semiconductors. In intrinsic semiconductor, the number of free electrons in the conduction band is equal to the number of holes in the valence band. Therefore, intrinsic semiconductor has no charge carriers to conduct electric current.

However, at room temperature a small number of charge carriers are generated. These small number of charge carriers will carry electric current.

PIN photodiode operation

A PIN photodiode is made of p region and n region separated by a highly resistive intrinsic layer. The intrinsic layer is placed between the p region and n region to increase the width of depletion region.

The p-type and n-type semiconductors are heavily doped. Therefore, the p region and n region of the PIN photodiode has large number of charge carriers to carry electric current. However, these charge carriers will not carry electric current under reverse bias condition.

On the other hand, intrinsic semiconductor is an undoped semiconductor material. Therefore, the intrinsic region does not have charge carriers to conduct electric current.

Under reverse bias condition, the majority charge carriers in n region and p region moves away from the junction. As a result, the width of depletion region becomes very wide. Therefore, majority carriers will not carry electric current under reverse bias condition.

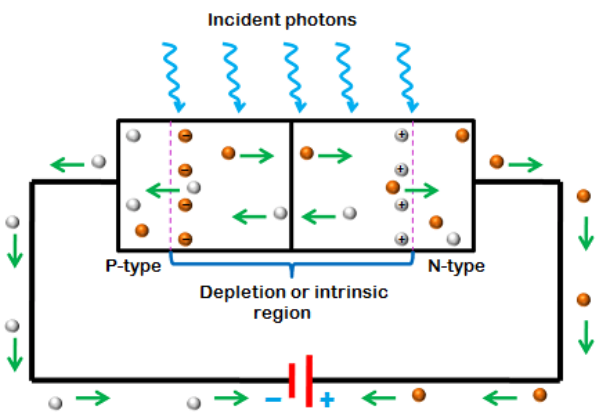
However, the minority carriers will carry electric current because they experience repulsive force from the external electric field.

In PIN photodiode, the charge carriers generated in the depletion region carry most of the electric current. The charge carriers generated in the p region or n region carry only a small electric current.

When light or photon energy is applied to the PIN diode, most part of the energy is observed by the intrinsic or depletion region because of the wide depletion width. As a result, a large number of electron-hole pairs are generated.

Free electrons generated in the intrinsic region move towards n-side whereas holes generated in the intrinsic region move towards p-side. The free electrons and holes moved from one region to another region carry electric current.

When free electrons and holes reach n region and p region, they are attracted to towards the positive and negative terminals of the battery.



PIN photodiode

The population of minority carriers in PIN photodiode is very large compared to the PN junction photodiode. Therefore, PIN photodiode carry large minority carrier current than PN junction photodiode.

When forward bias voltage is applied to the PIN photodiode, it behaves like a resistor.

We know that capacitance is directly proportional to the size of electrodes and inversely proportional to the distance between electrodes. In PIN photodiode, the p region and n region acts as electrodes and intrinsic region acts as dielectric.

The separation distance between p region and n region in PIN photodiode is very large because of the wide depletion width. Therefore, PIN photodiode has low capacitance compared to the PN junction photodiode.

In PIN photodiode, most of the electric current is carried by the charge carriers generated in the depletion region. The charge carriers generated in p region or n region carry only a small electric current. Therefore, increasing the width of depletion region increases the minority carrier electric current.

Advantages of PIN photodiode

Wide bandwidth

High quantum efficiency

High response speed

Avalanche photodiode

The operation of avalanche photodiode is similar to the PN junction and PIN photodiode except that a high reverse bias voltage is applied in case of avalanche photodiode to achieve avalanche multiplication.

Applying high reverse bias voltage to the avalanche photodiode will not directly increase the generation of charge carriers. However, it provides energy to the electron-hole pairs generated by the incident light.

When light energy is applied to the avalanche photodiode, electron-hole pairs are generated in the depletion. The generated electron-hole pairs experience a force due to the depletion region electric field and external electric field.

In avalanche photodiode, a very high reverse bias voltage supply large amount of energy to the minority carriers (electron-hole pairs). The minority carriers which gains large amount of energy are accelerated to greater velocities.

When the free electrons moving at high speed collides with the atom, they knock off more free electrons. The newly generated free electrons are again accelerated and collide with other atoms. Because of this continuous collision with atoms, a large number of minority carriers are generated. Thus, avalanche photodiodes generates more number of charge carriers than PN and PIN photodiodes.

Avalanche photodiodes are used in the applications where high gain is an important factor.

Advantages of avalanche photodiode

High sensitivity

Larger gain

Disadvantages of avalanche photodiode

Generates high level of noise than a PN photodiode

Photodiode operation modes

A photodiode can be operated in one of the two modes: photovoltaic mode or photoconductive mode.

Operation mode selection of the photodiode is depends upon the speed requirements of the application and the amount of dark current that is tolerable.

Photovoltaic mode

In the photovoltaic mode, the photodiode is unbiased. In other words, no external voltage is applied to the photodiode under photovoltaic mode.

In photovoltaic mode, dark current is very low. Photodiodes operated in photovoltaic mode have low response speed.

The photodiodes operated in photovoltaic mode are generally used for low speed applications or for detecting low light levels.

Photoconductive mode

In photoconductive mode, an external reverse bias voltage is applied to the photodiode.

Applying a reverse bias voltage increases the width of depletion region and reduces the junction capacitance which results in increased response speed. The reverse bias also increases the dark current.

Photodiodes operated in photoconductive mode has high noise current. This is due to the reverse saturation current flowing through the photodiode.

Dark current

Dark current is the leakage current that flows in the photodiode in the absence of light. The dark current in the photodiode increases when temperature increases. The material used to construct the photodiode also affects the dark current.

The different materials used to construct photodiodes are Silicon (Si), Germanium, (Ge), Gallium Phosphide (GaP), Indium Gallium Arsenide (InGaAs), Indium Arsenide Antimonide (InAsSb), Extended Range Indium Gallium Arsenide (InGaAs), Mercury Cadmium Telluride (MCT, HgCdTe).

Germanium, Indium Arsenide Antimonide, Indium Gallium Arsenide and Mercury Cadmium Telluride generates large dark current because they are very sensitive to temperature.

The response speed of Silicon, Gallium Phosphide, Indium Gallium Arsenide and Extended Range Indium Gallium Arsenide is very high.

Performance parameters of a photodiode

Responsivity

Responsivity is the ratio of generated photocurrent to the incident light power.

Quantum efficiency

Quantum efficiency is defined as the ratio of the number of electron-hole pairs (photoelectrons) generated to the incident photons.

Response time or transit time

The response time of a photodiode is defined as the time it takes for light generated charge carriers to cross p-n junction.

Photodiode applications

The various applications of photodiodes are

Compact disc players

Smoke detectors

Space applications

Photodiodes are used in medical applications such as computed tomography, instruments to analyze samples, and pulse oximeters.

Photodiodes are used for optical communications.

Photodiodes are used to measure extremely low light intensities.

Light Emitting Diode (LED)

What is light?

Before going into how LED works, let’s first take a brief look at light self. Since ancient times man has obtained light from various sources like sunrays, candles and lamps.

In 1879, Thomas Edison invented the incandescent light bulb. In the light bulb, an electric current is passed through a filament inside the bulb.

When sufficient current is passed through the filament, it gets heated up and emits light. The light emitted by the filament is the result of electrical energy converted into heat energy which in turn changes into light energy.

Unlike the light bulb in which electrical energy first converts into heat energy, the electrical energy can also be directly converted into light energy.

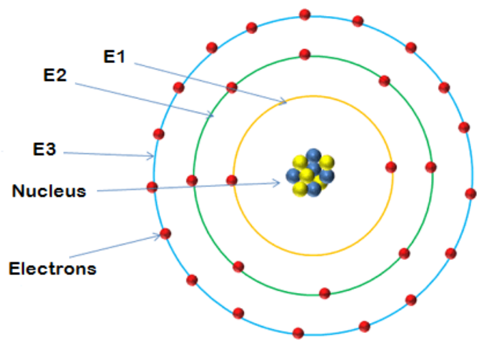
In Light Emitting Diodes (LEDs), electrical energy flowing through it is directly converted into light energy.

Light is a type of energy that can be released by an atom. Light is made up of many small particles called photons. Photons have energy and momentum but no mass.

Atoms are the basic building blocks of matter. Every object in the universe is made up of atoms. Atoms are made up of small particles such as electrons, protons and neutrons.

Electrons are negatively charged, protons are positively charged, and neutrons have no charge.

The attractive force between the protons and neutrons makes them stick together to form nucleus. Neutrons have no charge. Hence, the overall charge of the nucleus is positive.



The negatively charged electrons always revolve around the positively charged nucleus because of the electrostatic force of attraction between them. Electrons revolve around the nucleus in different orbits or shells. Each orbit has different energy level.

For example, the electrons orbiting very close to the nucleus have low energy whereas the electrons orbiting farther away from the nucleus have high energy.

The electrons in the lower energy level need some additional energy to jump into the higher energy level. This additional energy can be supplied by the outside source. When electrons orbiting the nucleus gains energy from outside source they jump into higher orbit or higher energy level.

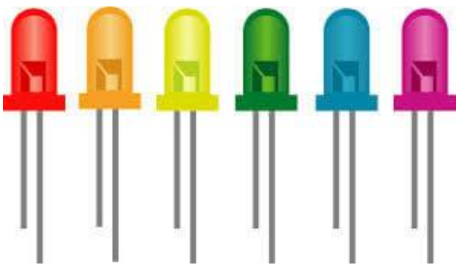
The electrons in the higher energy level will not stay for long period. After a short period, the electrons fall back to lower energy level. The electrons which jump from higher energy level to lower energy level will releases energy in the form of a photon or light. In some materials, this energy lose is released mostly in the form of heat. The electron which loses greater energy will release a greater energy photon.

What is Light Emitting Diode (LED)?

Light Emitting Diodes (LEDs) are the most widely used semiconductor diodes among all the different types of semiconductor diodes available today. Light emitting diodes emit either visible light or invisible infrared lightwhen forward biased. The LEDs which emit invisible infrared light are used for remote controls.

A light Emitting Diode (LED) is an optical semiconductor device that emits light when voltage is applied. In other words, LED is an optical semiconductor device that converts electrical energy into light energy.

When Light Emitting Diode (LED) is forward biased, free electrons in the conduction band recombines with theholes in the valence band and releases energy in the form of light.



The process of emitting light in response to the strong electric field or flow of electric current is called electroluminescence.

A normal p-n junction diode allows electric current only in one direction. It allows electric current when forward biased and does not allow electric current when reverse biased. Thus, normal p-n junction diode operates only in forward bias condition.

Like the normal p-n junction diodes, LEDs also operates only in forward bias condition. To create an LED, the n-type material should be connected to the negative terminal of the battery and p-type material should be connected to the positive terminal of the battery. In other words, the n-type material should be negatively charged and the p-type material should be positively charged.

The construction of LED is similar to the normal p-n junction diode except that gallium, phosphorus and arsenic materials are used for construction instead of silicon or germanium materials.

In normal p-n junction diodes, silicon is most widely used because it is less sensitive to the temperature. Also, it allows electric current efficiently without any damage. In some cases, germanium is used for constructing diodes.

However, silicon or germanium diodes do not emit energy in the form of light. Instead, they emit energy in the form of heat. Thus, silicon or germanium is not used for constructing LEDs.

Layers of LED

A Light Emitting Diode (LED) consists of three layers: p-type semiconductor, n-type semiconductor and depletion layer. The p-type semiconductor and the n-type semiconductor are separated by a depletion regionor depletion layer.

P-type semiconductor

When trivalent impurities are added to the intrinsic or pure semiconductor, a p-type semiconductor is formed.

In p-type semiconductor, holes are the majority charge carriers and free electrons are the minority charge carriers. Thus, holes carry most of the electric current in p-type semiconductor.

N-type semiconductor

When pentavalent impurities are added to the intrinsic semiconductor, an n-type semiconductor is formed.

In n-type semiconductor, free electrons are the majority charge carriers and holes are the minority charge carriers. Thus, free electrons carry most of the electric current in n-type semiconductor.

Depletion layer or region

Depletion region is a region present between the p-type and n-type semiconductor where no mobile charge carriers (free electrons and holes) are present. This region acts as barrier to the electric current. It opposes flow of electrons from n-type semiconductor and flow of holes from p-type semiconductor.

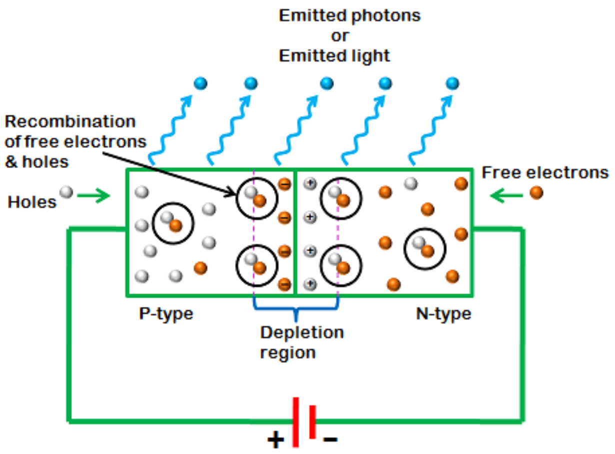
To overcome the barrier of depletion layer, we need to apply voltage which is greater than the barrier potential of depletion layer.

If the applied voltage is greater than the barrier potential of the depletion layer, the electric current starts flowing.

How Light Emitting Diode (LED) works?

Light Emitting Diode (LED) works only in forward bias condition. When Light Emitting Diode (LED) is forward biased, the free electrons from n-side and the holes from p-side are pushed towards the junction.

When free electrons reach the junction or depletion region, some of the free electrons recombine with the holes in the positive ions. We know that positive ions have less number of electrons than protons. Therefore, they are ready to accept electrons. Thus, free electrons recombine with holes in the depletion region. In the similar way, holes from p-side recombine with electrons in the depletion region.



Light Emitting Diode (LED)

Because of the recombination of free electrons and holes in the depletion region, the width of depletion regiondecreases. As a result, more charge carriers will cross the p-n junction.

Some of the charge carriers from p-side and n-side will cross the p-n junction before they recombine in the depletion region. For example, some free electrons from n-type semiconductor cross the p-n junction and recombines with holes in p-type semiconductor. In the similar way, holes from p-type semiconductor cross the p-n junction and recombines with free electrons in the n-type semiconductor.

Thus, recombination takes place in depletion region as well as in p-type and n-type semiconductor.

The free electrons in the conduction band releases energy in the form of light before they recombine with holes in the valence band.

In silicon and germanium diodes, most of the energy is released in the form of heat and emitted light is too small.

However, in materials like gallium arsenide and gallium phosphide the emitted photons have sufficient energy to produce intense visible light.

How LED emits light?

When external voltage is applied to the valence electrons, they gain sufficient energy and breaks the bonding with the parent atom. The valence electrons which breaks bonding with the parent atom are called free electrons.

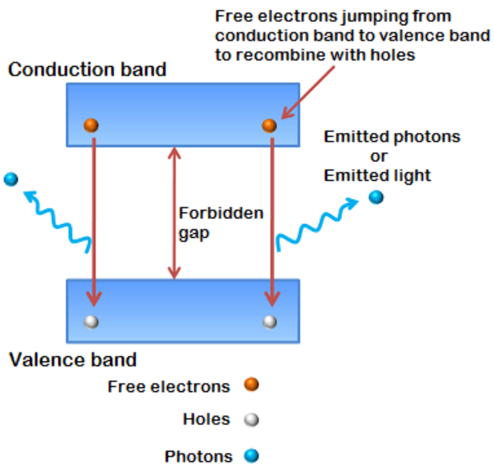
When the valence electron left the parent atom, they leave an empty space in the valence shell at which valence electron left. This empty space in the valence shell is called a hole.

The energy level of all the valence electrons is almost same. Grouping the range of energy levels of all the valence electrons is called valence band.

In the similar way, energy level of all the free electrons is almost same. Grouping the range of energy levels of all the free electrons is called conduction band.

The energy level of free electrons in the conduction band is high compared to the energy level of valence electrons or holes in the valence band. Therefore, free electrons in the conduction band need to lose energy in order to recombine with the holes in the valence band.

The free electrons in the conduction band do not stay for long period. After a short period, the free electrons lose energy in the form of light and recombine with the holes in the valence band. Each recombination of charge carrier will emit some light energy.



Process of light emission in LED

The energy lose of free electrons or the intensity of emitted light is depends on the forbidden gap or energy gap between conduction band and valence band.

The semiconductor device with large forbidden gap emits high intensity light whereas the semiconductor device with small forbidden gap emits low intensity light.

In other words, the brightness of the emitted light is depends on the material used for constructing LED and forward current flow through the LED.

In normal silicon diodes, the energy gap between conduction band and valence band is less. Hence, the electrons fall only a short distance. As a result, low energy photons are released. These low energy photons have low frequency which is invisible to human eye.

In LEDs, the energy gap between conduction band and valence band is very large so the free electrons in LEDs have greater energy than the free electrons in silicon diodes. Hence, the free electrons fall to a large distance. As a result, high energy photons are released. These high energy photons have high frequency which is visible to human eye.

The efficiency of generation of light in LED increases with increase in injected current and with a decrease in temperature.

In light emitting diodes, light is produced due to recombination process. Recombination of charge carriers takes place only under forward bias condition. Hence, LEDs operate only in forward bias condition.

When light emitting diode is reverse biased, the free electrons (majority carriers) from n-side and holes (majority carriers) from p-side moves away from the junction. As a result, the width of depletion region increases and no recombination of charge carriers occur. Thus, no light is produced.

If the reverse bias voltage applied to the LED is highly increased, the device may also be damaged.

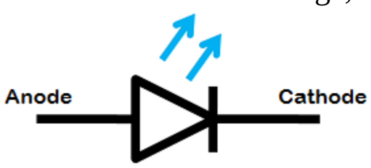
All diodes emit photons or light but not all diodes emit visible light. The material in an LED is selected in such a way that the wavelength of the released photons falls within the visible portion of the light spectrum.

Light emitting diodes can be switched ON and OFF at a very fast speed of 1 ns.

Light emitting diode (LED) symbol

The symbol of LED is similar to the normal p-n junction diode except that it contains arrows pointing away from the diode indicating that light is being emitted by the diode.

LEDs are available in different colors. The most common colors of LEDs are orange, yellow, green and red.

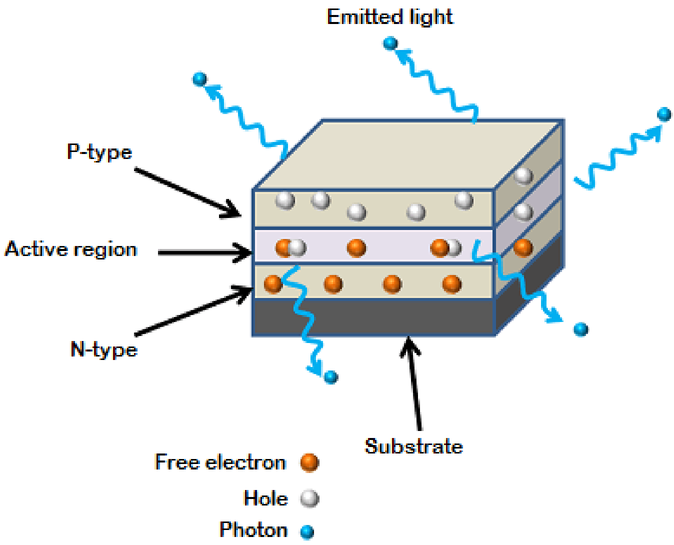


LED symbol

The schematic symbol of LED does not represent the color of light. The schematic symbol is same for all colors of LEDs. Hence, it is not possible to identify the color of LED by seeing its symbol.

LED construction

One of the methods used to construct LED is to deposit three semiconductor layers on the substrate. The three semiconductor layers deposited on the substrate are n-type semiconductor, p-type semiconductor and active region. Active region is present in between the n-type and p-type semiconductor layers.



Construction of LED

When LED is forward biased, free electrons from n-type semiconductor and holes from p-type semiconductor are pushed towards the active region.

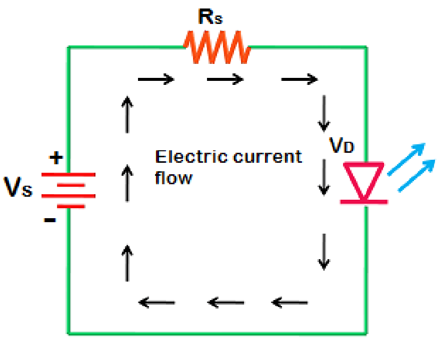
When free electrons from n-side and holes from p-side recombine with the opposite charge carriers (free electrons with holes or holes with free electrons) in active region, an invisible or visible light is emitted.

In LED, most of the charge carriers recombine at active region. Therefore, most of the light is emitted by the active region. The active region is also called as depletion region.

Biasing of LED

The safe forward voltage ratings of most LEDs is from 1V to 3 V and forward current ratings is from 200 mA to 100 mA.

If the voltage applied to LED is in between 1V to 3V, LED works perfectly because the current flow for the applied voltage is in the operating range. However, if the voltage applied to LED is increased to a value greater than 3 volts. The depletion region in the LED breaks down and the electric current suddenly rises. This sudden rise in current may destroy the device.



To avoid this we need to place a resistor (Rs) in series with the LED. The resistor (Rs) must be placed in between voltage source (Vs) and LED.

The resistor placed between LED and voltage source is called current limiting resistor. This resistor restricts extra current which may destroy the LED. Thus, current limiting resistor protects LED from damage.

The current flowing through the LED is mathematically written as

$$I_F = \frac{V_s - V_D}{R_s}$$

Where,

IF = Forward current

VS = Source voltage or supply voltage

VD = Voltage drop across LED

RS = Resistor or current limiting resistor

Voltage drop is the amount of voltage wasted to overcome the depletion region barrier (which leads to electric current flow).

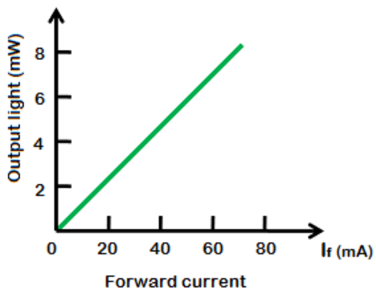
The voltage drop of LED is 2 to 3V whereas silicon or germanium diode is 0.3 or 0.7 V.

Therefore, to operate LED we need to apply greater voltage than silicon or germanium diodes.

Light emitting diodes consume more energy than silicon or germanium diodes to operate.

Output characteristics of LED

The amount of output light emitted by the LED is directly proportional to the amount of forward current flowing through the LED. More the forward current, the greater is the emitted output light. The graph of forward current vs output light is shown in the figure.



Visible LEDs and invisible LEDs

LEDs are mainly classified into two types: visible LEDs and invisible LEDs.

Visible LED is a type of LED that emits visible light. These LEDs are mainly used for display or illumination where LEDs are used individually without photosensors.

Invisible LED is a type of LED that emits invisible light (infrared light). These LEDs are mainly used with photosensors such as photodiodes.

What determines the color of an LED?

The material used for constructing LED determines its color. In other words, the wavelength or color of the emitted light depends on the forbidden gap or energy gap of the material.

Different materials emit different colors of light.

Gallium arsenide LEDs emit red and infrared light.

Gallium nitride LEDs emit bright blue light.

Yttrium aluminium garnet LEDs emit white light.

Gallium phosphide LEDs emit red, yellow and green light.

Aluminium gallium nitride LEDs emit ultraviolet light.

Aluminum gallium phosphide LEDs emit green light.

Advantages of LED

The brightness of light emitted by LED is depends on the current flowing through the LED. Hence, the brightness of LED can be easily controlled by varying the current. This makes possible to operate LED displays under different ambient lighting conditions.

Light emitting diodes consume low energy.

LEDs are very cheap and readily available.

LEDs are light in weight.

Smaller size.

LEDs have longer lifetime.

LEDs operates very fast. They can be turned on and off in very less time.

LEDs do not contain toxic material like mercury which is used in fluorescent lamps.

LEDs can emit different colors of light.

Disadvantages of LED

LEDs need more power to operate than normal p-n junction diodes.

Luminous efficiency of LEDs is low.

Applications of LED

The various applications of LEDs are as follows

Burglar alarms systems

Calculators

Picture phones

Traffic signals

Digital computers

Multimeters

Microprocessors

Digital watches

Automotive heat lamps

Camera flashes

Aviation lighting

Laser diode

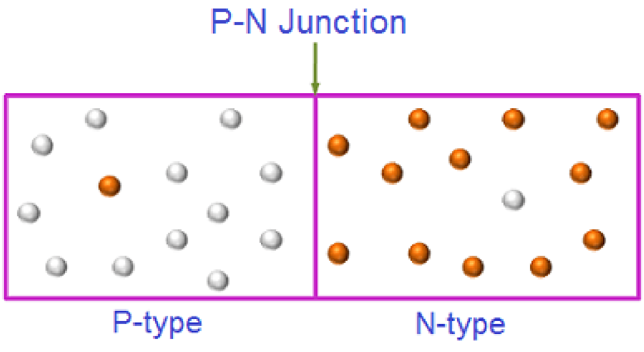
Laser diodes play an important role in our everyday lives. They are very cheap and small. Laser diodes are the smallest of all the known lasers. Their size is a fraction of a millimeter.

Laser diodes are also known as semiconductor lasers, junction lasers, junction diode lasers or injection lasers. Before going into laser diodes, let us first look at diode itself.

What is a p-n junction diode?

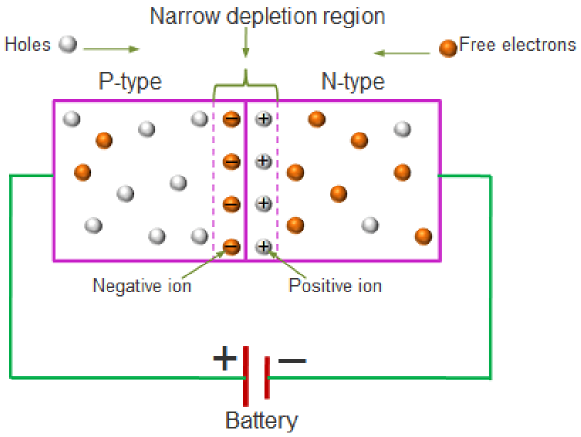
A p-n junction diode is a semiconductor device that allows the flow of current in only one direction.

The p-n junction diode is made of two types of semiconductor materials namely p-type and n-type semiconductor. The p-type semiconductor is joined with the n-type semiconductor to form a p-n junction. The device that results from the joining of a p-type and n-type semiconductor is called a p-n junction diode.



The p-n junction diode allows electric current in forward bias condition whereas it blocks electric current in reverse bias condition.

If the positive terminal of the battery is connected to the p-type semiconductor and the negative terminal of the battery is connected to the n-type semiconductor, the diode is said to be forward biased.



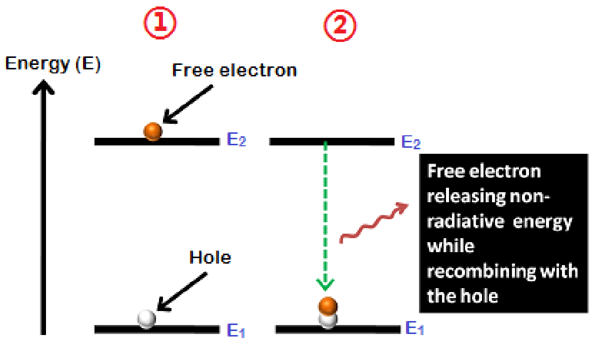
Forward bias

When a forward bias voltage is applied to the diode, free electrons start moving from the negative terminal of the battery to the positive terminal of the battery similarly holes start moving from the positive terminal of the battery to the negative terminal of the battery.

Because of these flow of charge carriers (free electrons and holes), electric current is generated in the p-n junction diode.

In ordinary p-n junction diodes, the electrons moving from n-type to p-type will recombines with the holes in the p-type semiconductor or junction. Similarly, the holes moving from p-type to n-type will recombines with the electrons in the n-type semiconductor or junction.

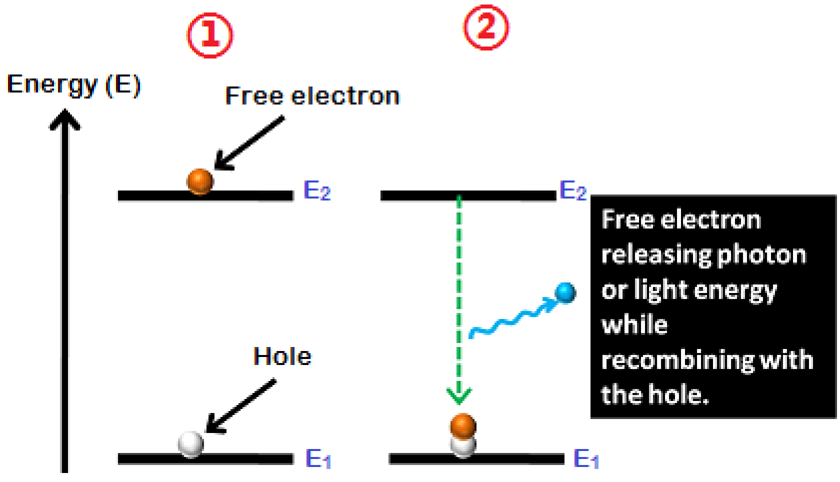
We know that the energy level of free electrons in the conduction band is high as compared to the holes in the valence band. Therefore, the free electrons will release their extra energy (non-radiative energy) while recombining with the holes.



Electron transition in ordinary diode

Physics and Radio-Electronics

In the light emitting diodes (LEDs) or laser diodes, the recombination takes place in a similar manner. However, the free electrons in LED's or laser diodes release energy in the form of light while recombining with the holes.



Electron transition in laser diode

What is a laser diode?

A laser diode is an optoelectronic device, which converts electrical energy into light energy to produce high-intensity coherent light. In a laser diode, the p-n junction of the semiconductor diode acts as the laser medium or active medium.

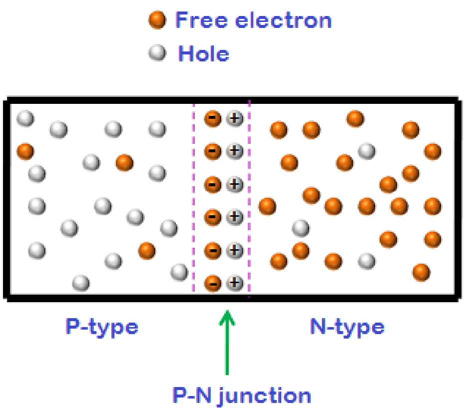
The working of the laser diode is almost similar to the light emitting diode (LED). The main difference between the LED and laser diode is that the LED emits incoherent light whereas the laser diode emits coherent light.

Laser diode construction

The laser diode is made of two doped gallium arsenide layers. One doped gallium arsenide layer will produce an n-type semiconductor whereas another doped gallium arsenide layer will produce a p-type semiconductor. In laser diodes, selenium, aluminum, and silicon are used as doping agents.

P-N junction

When a p-type layer is joined with the n-type layer, a p-n junction is formed. The point at which the p-type and n-type layers are joined is called p-n junction. The p-n junction separates the p-type and n-type semiconductors.



For the construction of laser diodes, gallium arsenide is chosen over silicon. In silicon diodes, the energy is released during recombination. However, this release of energy is not in the form of light.

In gallium arsenide diodes, the release of energy is in the form of light or photons. Therefore, gallium arsenide is used for the construction of laser diodes.

N-type semiconductor

Adding a small percentage of foreign atoms into the intrinsic semiconductor produces an n-type or p-type semiconductor.

If pentavalent impurities are added to the intrinsic or pure semiconductor, an n-type semiconductor is produced. In n-type semiconductors, free electrons are the majority charge carriers whereas holes are the minority charge carriers. Therefore, free electrons carry most of the electric current in n-type semiconductors.

P-type semiconductor

If trivalent impurities are added to the pure semiconductor, a p-type semiconductor is produced. In p-type semiconductors, holes are the majority charge carriers whereas free electrons are the minority charge carriers. Therefore, holes carry most of the electric current in p-type semiconductors.

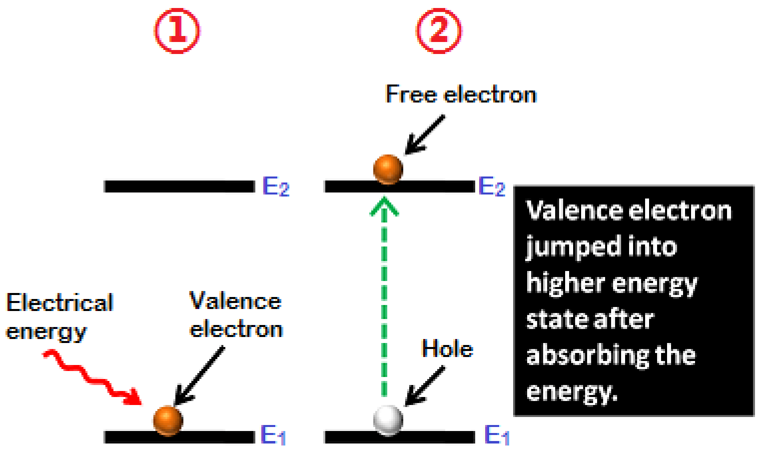
Main steps required for producing a coherent beam of light in laser diodes

The main steps required for producing a coherent beam of light in lasers diodes are: light absorption, spontaneous emission, and stimulated emission.

Absorption of energy

Absorption of energy is the process of absorbing energy from the external energy sources.

In laser diodes, electrical energy or DC voltage is used as the external energy source. When the DC voltage or electrical energy supplies enough energy to the valence electrons or valence band electrons, they break bonding with the parent atom and jumps into the higher energy level (conduction band). The electrons in the conduction band are known as free electrons.



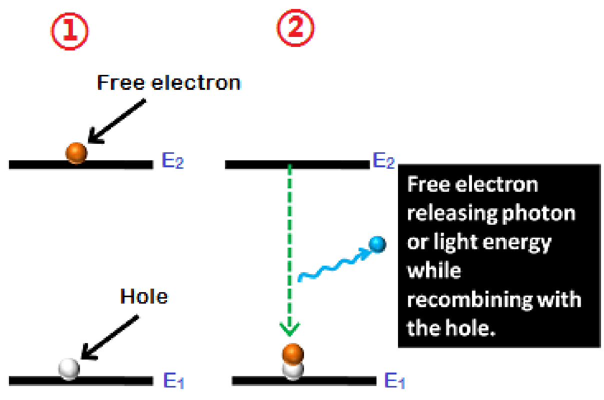
Absorption

When the valence electron leaves the valence shell, an empty space is created at the point from which electron left. This empty space in the valence shell is called a hole. Thus, both free electrons and holes are generated as a pair because of the absorption of energy from the external DC source.

Spontaneous emission

Spontaneous emission is the process of emitting light or photons naturally while electrons falling to the lower energy state.

In laser diodes, the valence band electrons or valence electrons are in the lower energy state. Therefore, the holes generated after the valence electrons left are also in the lower energy state.



Spontaneous emission

On the other hand, the conduction band electrons or free electrons are in the higher energy state. In simple words, free electrons have more energy than holes.

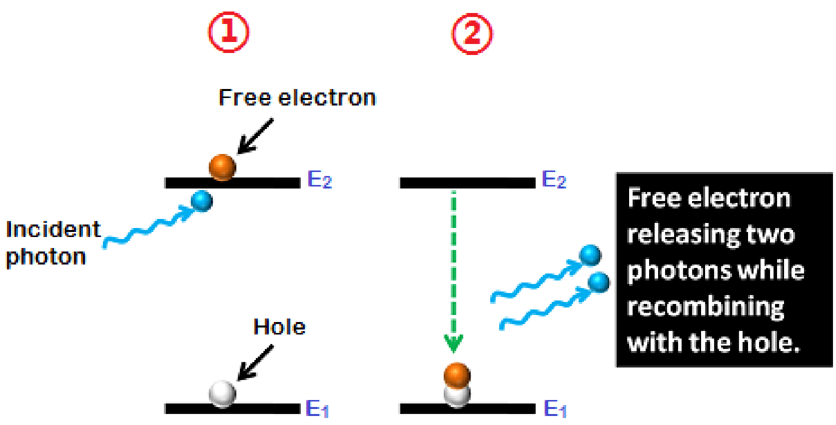
The free electrons in the conduction band need to lose their extra energy in order to recombine with the holes in the valence band.

The free electrons in the conduction band will not stay for long period. After a short period, the free electrons recombine with the lower energy holes by releasing energy in the form of photons.

Stimulated emission

Stimulated emission is the process by which excited electrons or free electrons are stimulated to fall into the lower energy state by releasing energy in the form of light. The stimulated emission is an artificial process.

In stimulated emission, the excited electrons or free electrons need not wait for the completion of their lifetime. Before the completion of their lifetime, the incident or external photons will force the free electrons to recombine with the holes. In stimulated emission, each incident photon will generate two photons.

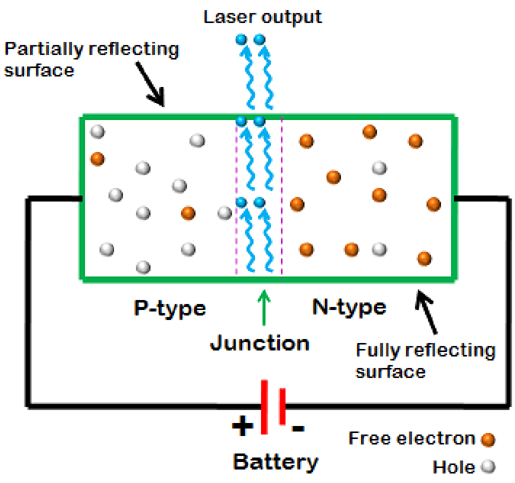


Stimulated emission

All the photons generated due to the stimulated emission will travel in the same direction. As a result, a narrow beam of high-intensity laser light is produced.

How laser diode works?

When DC voltage is applied across the laser diode, the free electrons move across the junction region from the n-type material to the p-type material. In this process, some electrons will directly interact with the valence electrons and excites them to the higher energy level whereas some other electrons will recombine with the holes in the p-type semiconductor and releases energy in the form of light. This process of emission is called spontaneous emission.



Laser diode

The photons generated due to spontaneous emission will travel through the junction region and stimulate the excited electrons (free electrons). As a result, more photons are released. This process of light or photons emission is called stimulated emission. The light generated due to stimulated emission will moves parallel to the junction.

The two ends of the laser diode structure are optically reflective. One end is fully reflective whereas another end is partially reflective. The fully reflective end will reflect the light completely whereas the partially reflective end will reflect most part of the light but allows a small amount of light.

The light generated in the p-n junction will bounce back and forth (hundreds of times) between the two reflective surfaces. As a result, an enormous optical gain is achieved.

The light generated due to the stimulated emission is escaped through the partially reflective end of the laser diode to produce a narrow beam laser light.

All the photons generated due to the stimulated emission will travel in the same direction. Therefore, this light will travel to long distances without spreading in the space.

Advantages of laser diodes

Simple construction

Lightweight

Very cheap

Small size

Highly reliable compared to other types of lasers.

Longer operating life

High efficiency

Mirrors are not required in the semiconductor lasers.

Low power consumption

Disadvantages of laser diodes

Not suitable for the applications where high powers are required.

Semiconductor lasers are highly dependent on temperature.

Applications of laser diodes

Laser diodes are used in laser pointers.

Laser diodes are used in fiber optic communications.

Laser diodes are used in barcode readers.

Laser diodes are used in laser printing.

Laser diodes are used in laser scanning.

Laser diodes are used in range finders.

Laser diodes are used in laser absorption spectrometry.

Tunnel diode

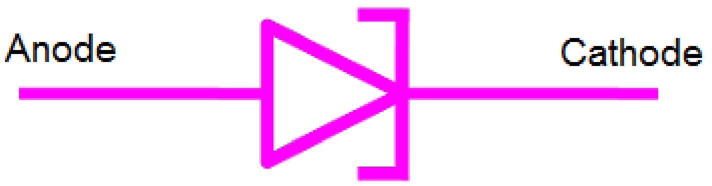
Tunnel diode definition

A Tunnel diode is a heavily doped p-n junction diode in which the electric current decreases as the voltageincreases.

In tunnel diode, electric current is caused by "Tunneling". The tunnel diode is used as a very fast switching device in computers. It is also used in high-frequency oscillators and amplifiers.

Symbol of tunnel diode

The circuit symbol of tunnel diode is shown in the below figure. In tunnel diode, the p-type semiconductor act as an anode and the n-type semiconductor act as a cathode.



Tunnel diode symbol

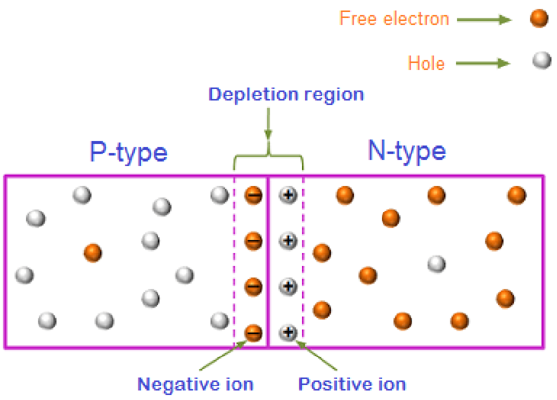
We know that a anode is a positively charged electrode which attracts electrons whereas cathode is a negatively charged electrode which emits electrons. In tunnel diode, n-type semiconductor emits or produces electrons so it is referred to as the cathode. On the other hand, p-type semiconductor attracts electrons emitted from the n-type semiconductor so p-type semiconductor is referred to as the anode.

What is a tunnel diode?

Tunnel diodes are one of the most significant solid-state electronic devices which have made their appearance in the last decade. Tunnel diode was invented in 1958 by Leo Esaki.

Leo Esaki observed that if a semiconductor diode is heavily doped with impurities, it will exhibit negative resistance. Negative resistance means the current across the tunnel diode decreases when the voltage increases. In 1973 Leo Esaki received the Nobel Prize in physics for discovering the electron tunneling effect used in these diodes.

A tunnel diode is also known as Esaki diode which is named after Leo Esaki for his work on the tunneling effect. The operation of tunnel diode depends on the quantum mechanics principle known as "Tunneling". In electronics, tunneling means a direct flow of electrons across the small depletion region from n-side conduction band into the p-side valence band.



Tunnel diode

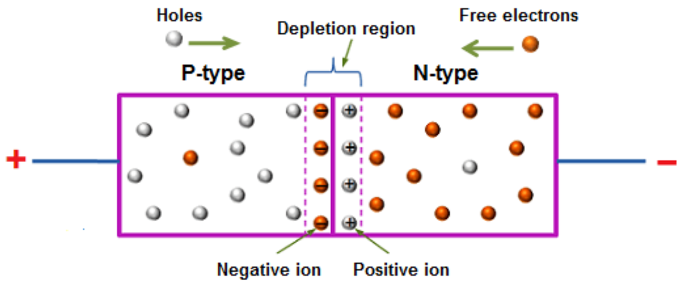
The germanium material is commonly used to make the tunnel diodes. They are also made from other types of materials such as gallium arsenide, gallium antimonide, and silicon.

Width of the depletion region in tunnel diode

The depletion region is a region in a p-n junction diode where mobile charge carriers (free electrons and holes) are absent. Depletion region acts like a barrier that opposes the flow of electrons from the n-type semiconductor and holes from the p-type semiconductor.

The width of a depletion region depends on the number of impurities added. Impurities are the atoms introduced into the p-type and n-type semiconductor to increase electrical conductivity.

If a small number of impurities are added to the p-n junction diode (p-type and n-type semiconductor), a wide depletion region is formed. On the other hand, if large number of impurities are added to the p-n junction diode, a narrow depletion region is formed.



Forward bias tunnel diode

In tunnel diode, the p-type and n-type semiconductor is heavily doped which means a large number of impurities are introduced into the p-type and n-type semiconductor. This heavy doping process produces an extremely narrow depletion region. The concentration of impurities in tunnel diode is 1000 times greater than the normal p-n junction diode.

In normal p-n junction diode, the depletion width is large as compared to the tunnel diode. This wide depletion layer or depletion region in normal diode opposes the flow of current. Hence, depletion layer acts as a barrier. To overcome this barrier, we need to apply sufficient voltage. When sufficient voltage is applied, electric current starts flowing through the normal p-n junction diode.

Unlike the normal p-n junction diode, the width of a depletion layer in tunnel diode is extremely narrow. So applying a small voltage is enough to produce electric current in tunnel diode.

Tunnel diodes are capable of remaining stable for a long duration of time than the ordinary p-n junction diodes. They are also capable of high-speed operations.

Concept of tunneling

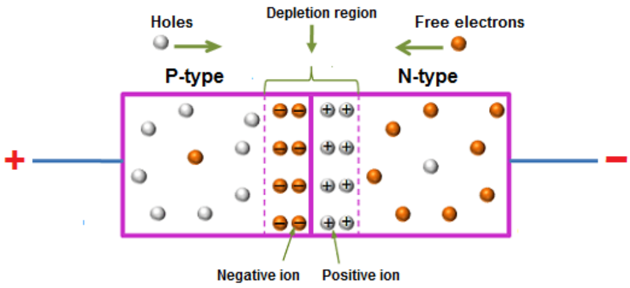
The depletion region or depletion layer in a p-n junction diode is made up of positive ions and negative ions. Because of these positive and negative ions, there exists a built-in-potential or electric field in the depletion region. This electric field in the depletion region exerts electric force in a direction opposite to that of the external electric field (voltage).

Another thing we need to remember is that the valence band and conduction band energy levels in the n-type semiconductor are slightly lower than the valence band

and conduction band energy levels in the p-type semiconductor. This difference in energy levels is due to the differences in the energy levels of the dopant atoms (donor or acceptor atoms) used to form the n-type and p-type semiconductor.

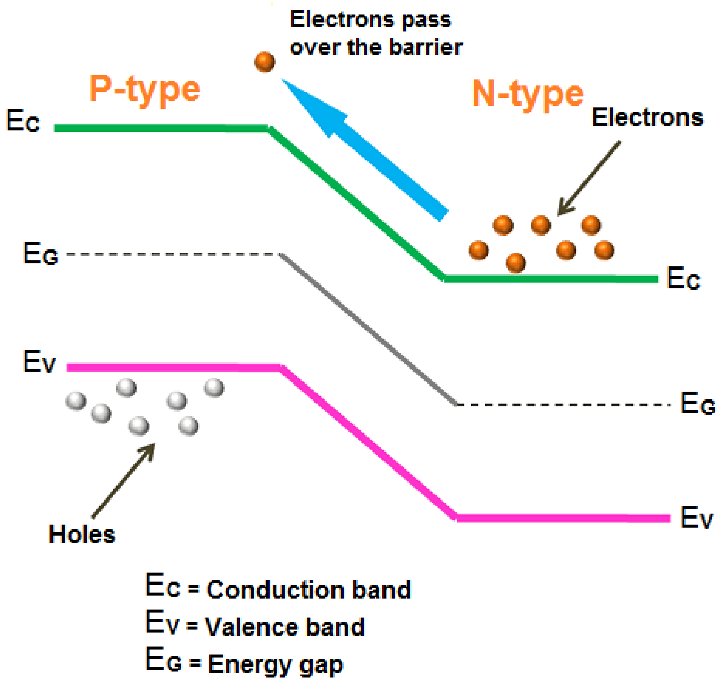
Electric current in ordinary p-n junction diode

When a forward bias voltage is applied to the ordinary p-n junction diode, the width of depletion region decreases and at the same time the barrier height also decreases. However, the electrons in the n-type semiconductor cannot penetrate through the depletion layer because the built-in voltage of depletion layer opposes the flow of electrons.

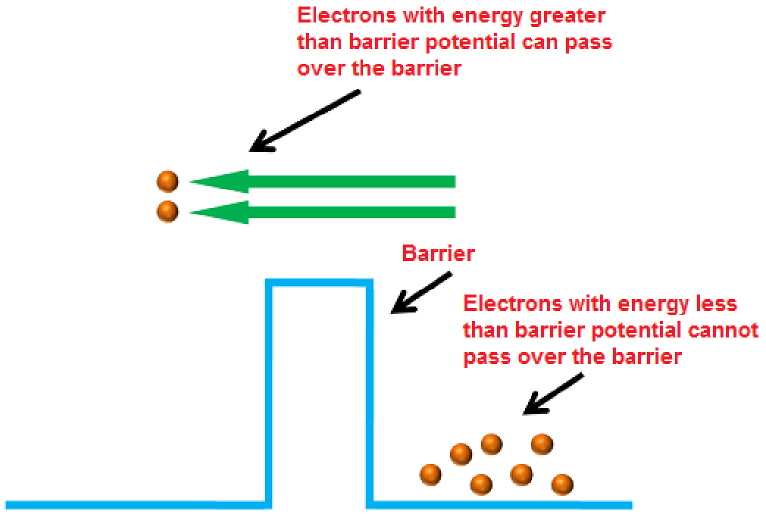


Forward bias p-n junction diode

If the applied voltage is greater than the built-in voltage of depletion layer, the electrons from n-side overcomes the opposing force from depletion layer and then enters into p-side. In simple words, the electrons can pass over the barrier (depletion layer) if the energy of the electrons is greater than the barrier height or barrier potential.



Ordinary P-N junction diode

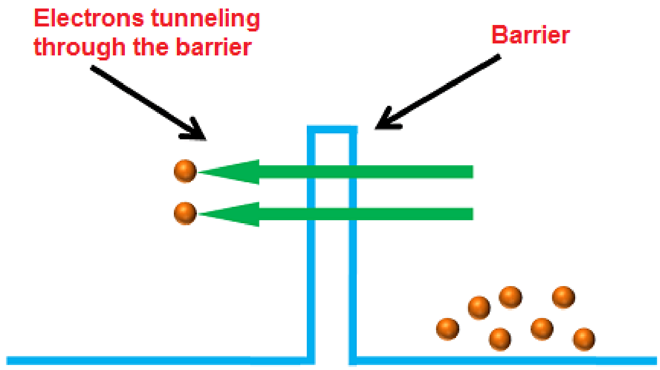


Ordinary p-n junction diode

Therefore, an ordinary p-n junction diode produces electric current only if the applied voltage is greater than the built-in voltage of the depletion region.

Electric current in tunnel diode

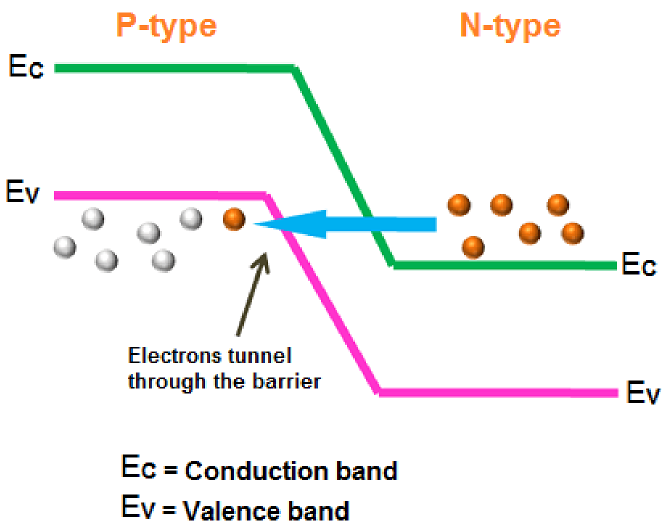
In tunnel diode, the valence band and conduction band energy levels in the n-type semiconductor are lower than the valence band and conduction band energy levels in the p-type semiconductor. Unlike the ordinary p-n junction diode, the difference in energy levels is very high in tunnel diode. Because of this high difference in energy levels, the conduction band of the n-type material overlaps with the valence band of the p-type material.



Tunnel diode

Quantum mechanics says that the electrons will directly penetrate through the depletion layer or barrier if the depletion width is very small.

The depletion layer of tunnel diode is very small. It is in nanometers. So the electrons can directly tunnel across the small depletion region from n-side conduction band into the p-side valence band.

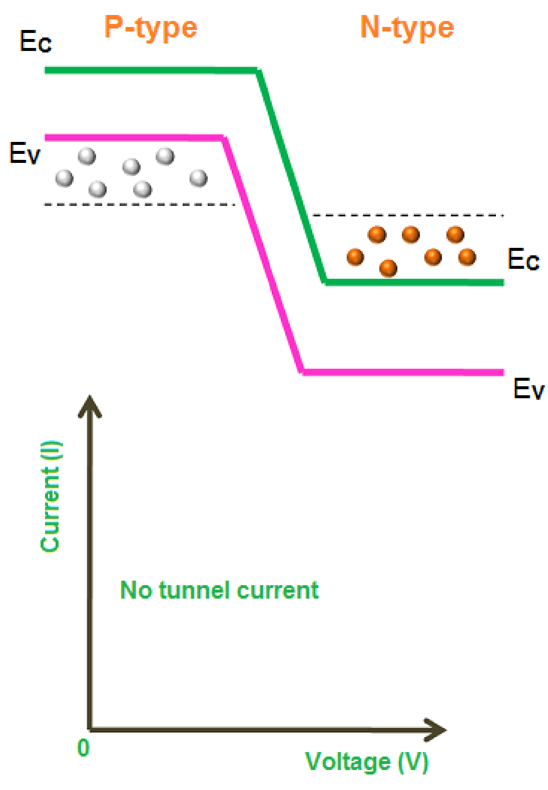


In ordinary diodes, current is produced when the applied voltage is greater than the built-in voltage of the depletion region. But in tunnel diodes, a small voltage which is less than the built-in voltage of depletion region is enough to produce electric current.

In tunnel diodes, the electrons need not overcome the opposing force from the depletion layer to produce electric current. The electrons can directly tunnel from the conduction band of n-region into the valence band of p-region. Thus, electric current is produced in tunnel diode.

How tunnel diode works?

Step 1: Unbiased tunnel diode



Unbiased tunnel diode

When no voltage is applied to the tunnel diode, it is said to be an unbiased tunnel diode. In tunnel diode, the conduction band of the n-type material overlaps with the valence band of the p-type material because of the heavy doping.

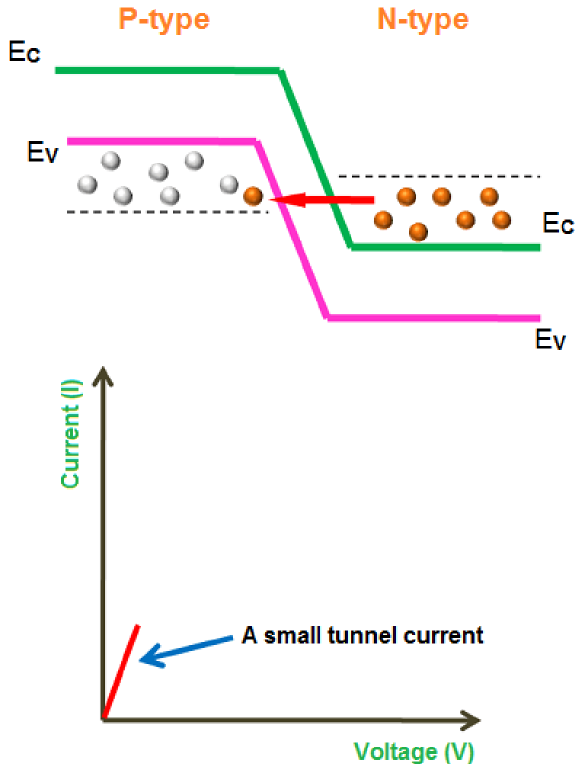
Because of this overlapping, the conduction band electrons at n-side and valence band holes at p-side are nearly at the same energy level. So when the temperature increases, some electrons tunnel from the conduction band of n-region to the valence band of p-region. In a similar way, holes tunnel from the valence band of p-region to the conduction band of n-region.

However, the net current flow will be zero because an equal number of charge carriers (free electrons and holes) flow in opposite directions.

Step 2: Small voltage applied to the tunnel diode

When a small voltage is applied to the tunnel diode which is less than the built-in voltage of the depletion layer, no forward current flows through the junction.

However, a small number of electrons in the conduction band of the n-region will tunnel to the empty states of the valence band in p-region. This will create a small forward bias tunnel current. Thus, tunnel current starts flowing with a small application of voltage.

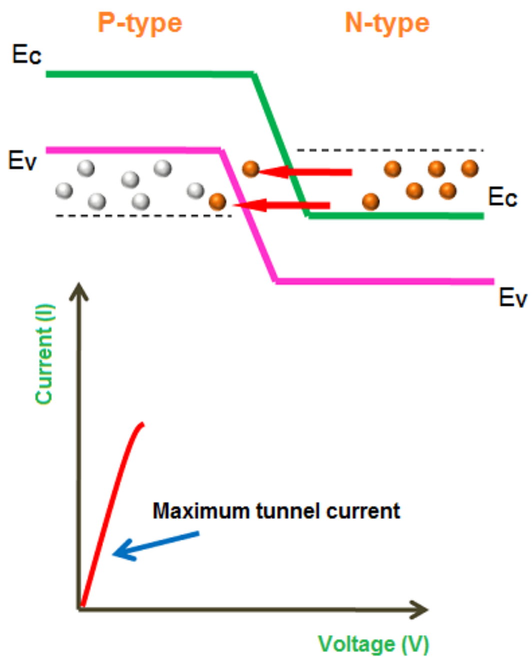


Small tunnel current

Step 3: Applied voltage is slightly increased

When the voltage applied to the tunnel diode is slightly increased, a large number of free electrons at n-side and holes at p-side are generated. Because of the increase in voltage, the overlapping of the conduction band and valence band is increased.

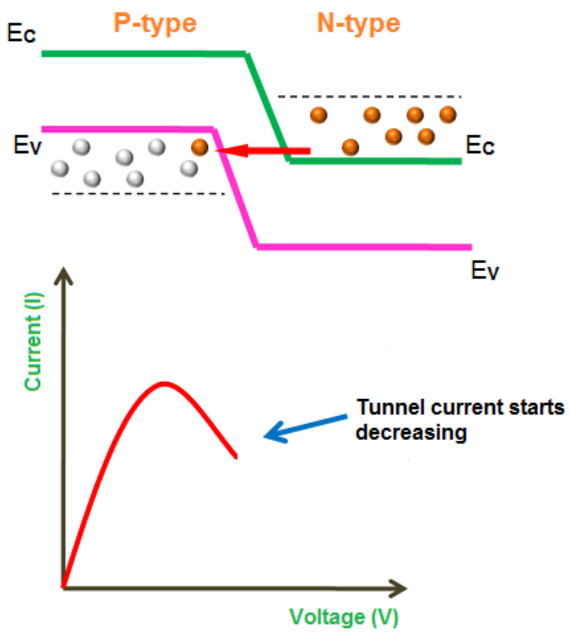
In simple words, the energy level of an n-side conduction band becomes exactly equal to the energy level of a p-side valence band. As a result, maximum tunnel current flows.



Maximum tunnel current

Step 4: Applied voltage is further increased

If the applied voltage is further increased, a slight misalign of the conduction band and valence band takes place.

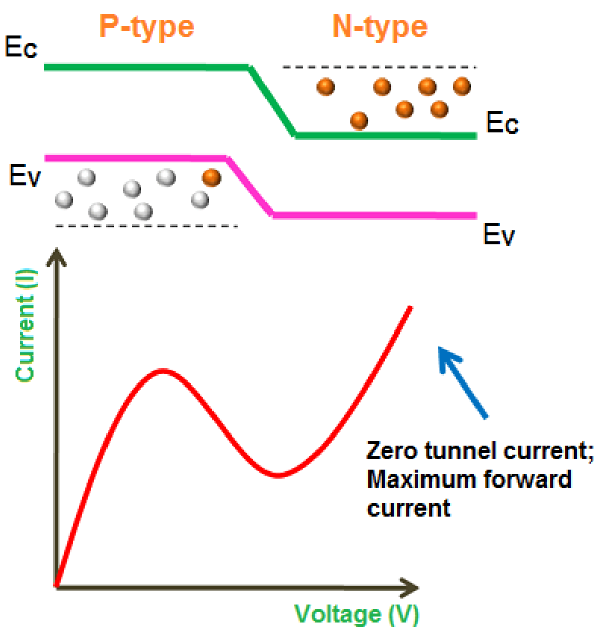


Tunnel current starts decreasing

Since the conduction band of the n-type material and the valence band of the p-type material sill overlap. The electrons tunnel from the conduction band of n-region to the valence band of p-region and cause a small current flow. Thus, the tunneling current starts decreasing.

Step 5: Applied voltage is largely increased

If the applied voltage is largely increased, the tunneling current drops to zero. At this point, the conduction band and valence band no longer overlap and the tunnel diode operates in the same manner as a normal p-n junction diode.



Zero tunnel current; maximum forward current

If this applied voltage is greater than the built-in potential of the depletion layer, the regular forward current starts flowing through the tunnel diode.

The portion of the curve in which current decreases as the voltage increases is the negative resistance region of the tunnel diode. The negative resistance region is the most important and most widely used characteristic of the tunnel diode.

A tunnel diode operating in the negative resistance region can be used as an amplifier or an oscillator.

Advantages of tunnel diodes

Long life

High-speed operation

Low noise

Low power consumption

Disadvantages of tunnel diodes

Tunnel diodes cannot be fabricated in large numbers

Being a two terminal device, the input and output are not isolated from one another.

Applications of tunnel diodes

Tunnel diodes are used as logic memory storage devices.

Tunnel diodes are used in relaxation oscillator circuits.

Tunnel diode is used as an ultra high-speed switch.

Tunnel diodes are used in FM receivers.

Schottky diode

Schottky diode definition

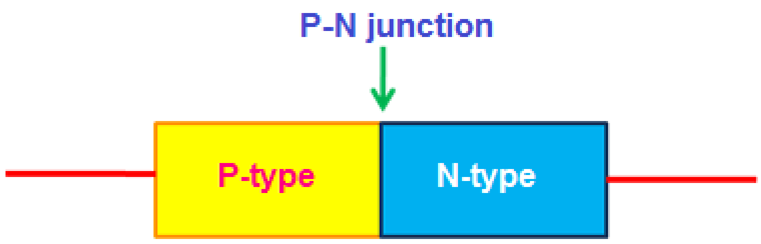
Schottky diode is a metal-semiconductor junction diode that has less forward voltage drop than the P-N junction diode and can be used in high-speed switching applications.

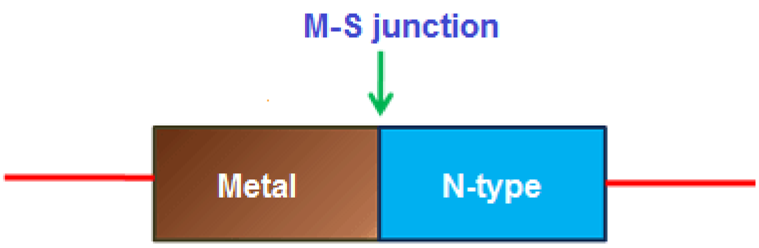
What is a schottky diode?

In a normal p-n junction diode, a p-type semiconductor and an n-type semiconductor are used to form the p-n junction. When a p-type semiconductor is joined with an n-type semiconductor, a junction is formed between the P-type and N-type semiconductor. This junction is known as P-N junction.

In schottky diode, metals such as aluminum or platinum replace the P-type semiconductor. The schottky diode is named after German physicist Walter H. Schottky.

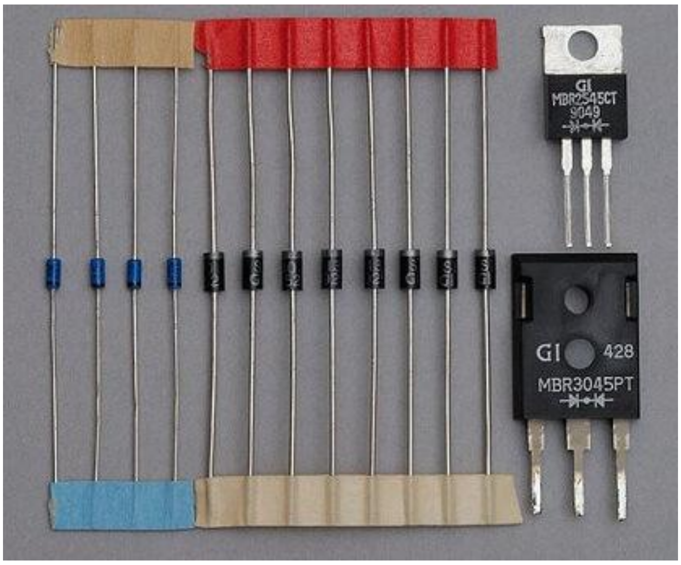
Schottky diode is also known as schottky barrier diode, surface barrier diode, majority carrier device, hot-electron diode, or hot carrier diode. Schottky diodes are widely used in radio frequency (RF) applications.





When aluminum or platinum metal is joined with N-type semiconductor, a junction is formed between the metal and N-type semiconductor. This junction is known as a metal-semiconductor junction or M-S junction. A metal-semiconductor junction formed between a metal and n-type semiconductor creates a barrier or depletion layer known as a schottky barrier.

Schottky diode can switch on and off much faster than the p-n junction diode. Also, the schottky diode produces less unwanted noise than p-n junction diode. These two characteristics of the schottky diode make it very useful in high-speed switching power circuits.



When sufficient voltage is applied to the schottky diode, current starts flowing in the forward direction. Because of this current flow, a small voltage loss occurs across the terminals of the schottky diode. This voltage loss is known as voltage drop.

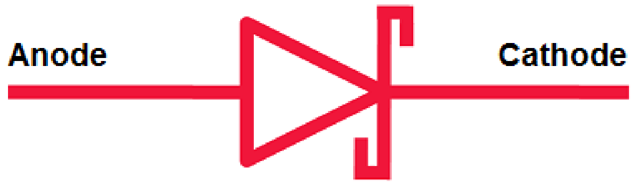
A silicon diode has a voltage drop of 0.6 to 0.7 volts, while a schottky diode has a voltage drop of 0.2 to 0.3 volts. Voltage loss or voltage drop is the amount of voltage wasted to turn on a diode.

In silicon diode, 0.6 to 0.7 volts is wasted to turn on the diode, whereas in schottky diode, 0.2 to 0.3 volts is wasted to turn on the diode. Therefore, the schottky diode consumes less voltage to turn on.

The voltage needed to turn on the schottky diode is same as that of a germanium diode. But germanium diodes are rarely used because the switching speed of germanium diodes is very low as compared to the schottky diodes.

Symbol of schottky diode

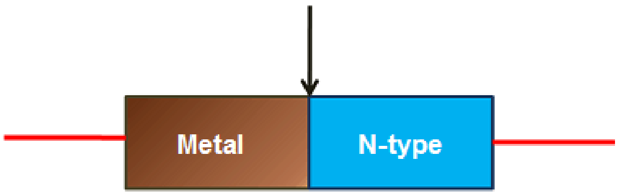
The symbol of schottky diode is shown in the below figure. In schottky diode, the metal acts as the anode and n-type semiconductor acts as the cathode.



Schottky diode symbol

Metal-semiconductor (M-S) junction

Metal-semiconductor (M-S) junction is a type of junction formed between a metal and an n-type semiconductor when the metal is joined with the n-type semiconductor. Metal-semiconductor junction is also sometimes referred to as M-S junction.

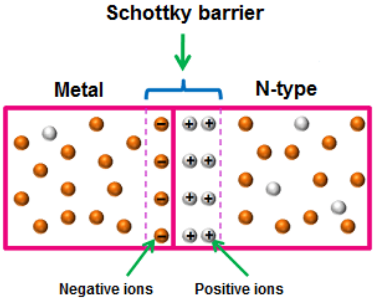


The metal-semiconductor junction can be either non-rectifying or rectifying. The non-rectifying metal-semiconductor junction is called ohmic contact. The rectifying metal-semiconductor junction is called non-ohmic contact.

What is a schottky barrier?

Schottky barrier is a depletion layer formed at the junction of a metal and n-type semiconductor. In simple words, schottky barrier is the potential energy barrier formed at the metal-semiconductor junction. The electrons have to overcome this potential energy barrier to flow across the diode.

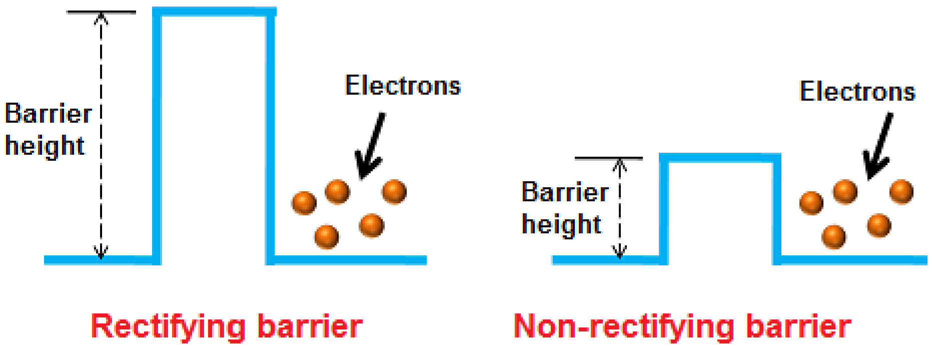
The rectifying metal-semiconductor junction forms a rectifying schottky barrier. This rectifying schottky barrier is used for making a device known as schottky diode. The non-rectifying metal-semiconductor junction forms a non-rectifying schottky barrier.



One of the most important characteristics of a schottky barrier is the schottky barrier height. The value of this barrier height depends on the combination of semiconductor and metal.

The schottky barrier height of ohmic contact (non-rectifying barrier) is very low whereas the schottky barrier height of non-ohmic contact (rectifying barrier) is high.

In non-rectifying schottky barrier, the barrier height is not high enough to form a depletion region. So depletion region is negligible or absent in the ohmic contact diode.



On the other hand, in rectifying schottky barrier, the barrier height is high enough to form a depletion region. So the depletion region is present in the non-ohmic contact diode.

The non-rectifying metal-semiconductor junction (ohmic contact) offers very low resistance to the electric current whereas the rectifying metal-semiconductor junction offers high resistance to the electric current as compared to the ohmic contact.

The rectifying schottky barrier is formed when a metal is in contact with the lightly doped semiconductor, whereas the non-rectifying barrier is formed when a metal is in contact with the heavily doped semiconductor.

The ohmic contact has a linear current-voltage (I-V) curve whereas the non-ohmic contact has a non-linear current-voltage (I-V) curve.

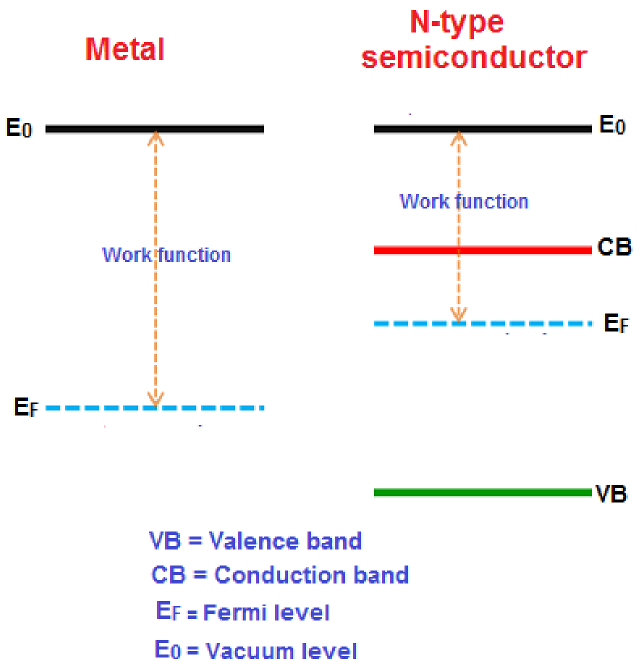
Energy band diagram of schottky diode

The energy band diagram of the N-type semiconductor and metal is shown in the below figure.

The vacuum level is defined as the energy level of electrons that are outside the material. The work function is defined as the energy required to move an electron from Fermi level (EF) to vacuum level (E0).

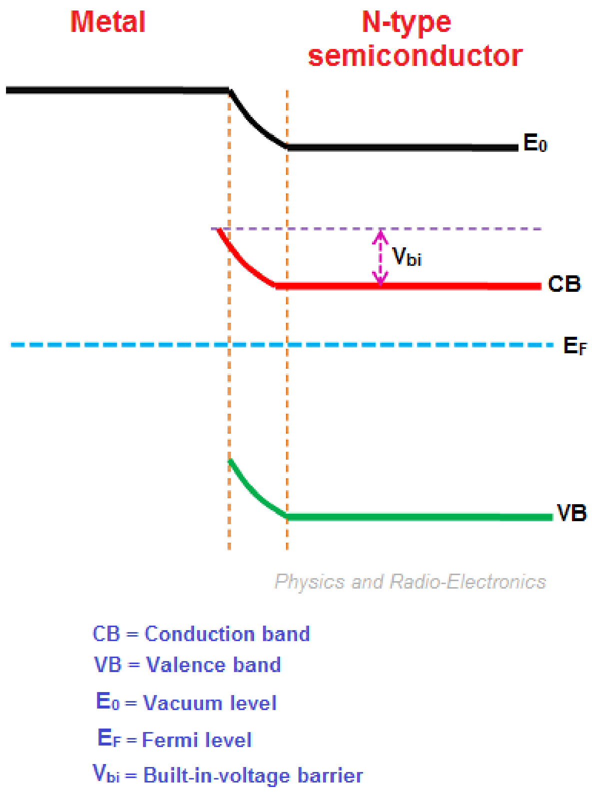
The work function is different for metal and semiconductor. The work function of a metal is greater than the work function of a semiconductor. Therefore, the electrons in the n-type semiconductor have high potential energy than the electrons in the metal.

The energy levels of the metal and semiconductor are different. The Fermi level at N-type semiconductor side lies above the metal side.



We know that electrons in the higher energy level have more potential energy than the electrons in the lower energy level. So the electrons in the N-type semiconductor have more potential energy than the electrons in the metal.

The energy band diagram of the metal and n-type semiconductor after contact is shown in the below figure.



When the metal is joined with the n-type semiconductor, a device is created known as schottky diode. The built-in-voltage (Vbi) for schottky diode is given by the difference between the work functions of a metal and n-type semiconductor.

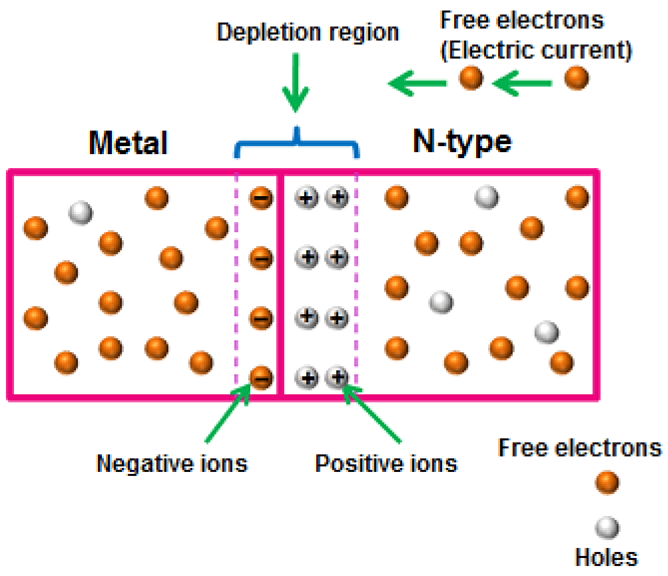
How schottky diode works?

Unbiased schottky diode

When the metal is joined with the n-type semiconductor, the conduction band electrons (free electrons) in the n-type semiconductor will move from n-type semiconductor to metal to establish an equilibrium state.

We know that when a neutral atom loses an electron it becomes a positive ion similarly when a neutral atom gains an extra electron it becomes a negative ion.

The conduction band electrons or free electrons that are crossing the junction will provide extra electrons to the atoms in the metal. As a result, the atoms at the metal junction gains extra electrons and the atoms at the n-side junction lose electrons.



Unbiased schottky diode

The atoms that lose electrons at the n-side junction will become positive ions whereas the atoms that gain extra electrons at the metal junction will become negative ions. Thus, positive ions are created the n-side junction and negative ions are created at the metal junction. These positive and negative ions are nothing but the depletion region.

Since the metal has a sea of free electrons, the width over which these electrons move into the metal is negligibly thin as compared to the width inside the n-type semiconductor. So the built-in-potential or built-in-voltage is primarily present inside the n-type semiconductor. The built-in-voltage is the barrier seen by the conduction band electrons of the n-type semiconductor when trying to move into the metal.

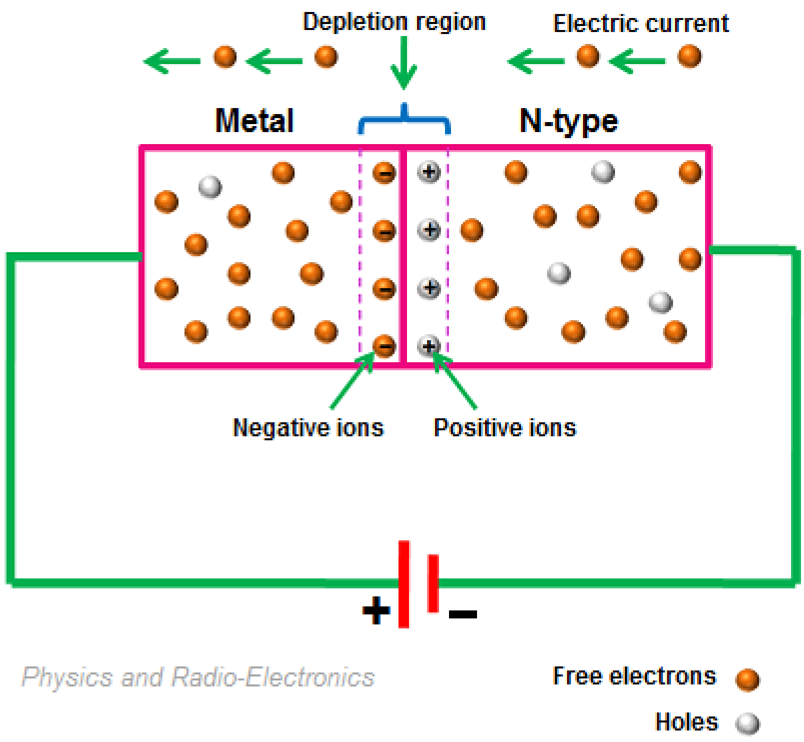
To overcome this barrier, the free electrons need energy greater than the built-in-voltage. In unbiased schottky diode, only a small number of electrons will flow from n-type semiconductor to metal. The built-in-voltage prevents further electron flow from the semiconductor conduction band into the metal.

The transfer of free electrons from the n-type semiconductor into metal results in energy band bending near the contact.

Forward biased schottky diode

If the positive terminal of the battery is connected to the metal and the negative terminal of the battery is connected to the n-type semiconductor, the schottky diode is said to be forward biased.

When a forward bias voltage is applied to the schottky diode, a large number of free electrons are generated in the n-type semiconductor and metal. However, the free electrons in n-type semiconductor and metal cannot cross the junction unless the applied voltage is greater than 0.2 volts.



Forward biased schottky diode

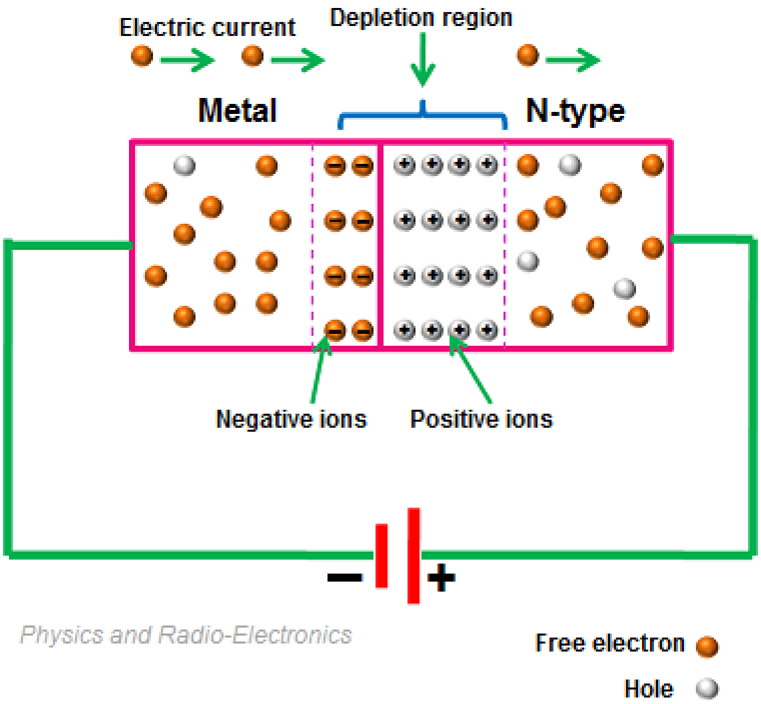
If the applied voltage is greater than 0.2 volts, the free electrons gain enough energy and overcomes the built-in-voltage of the depletion region. As a result, electric current starts flowing through the schottky diode.

If the applied voltage is continuously increased, the depletion region becomes very thin and finally disappears.

Reverse bias schottky diode

If the negative terminal of the battery is connected to the metal and the positive terminal of the battery is connected to the n-type semiconductor, the schottky diode is said to be reverse biased.

When a reverse bias voltage is applied to the schottky diode, the depletion width increases. As a result, the electric current stops flowing. However, a small leakage current flows due to the thermally excited electrons in the metal.



Reverse biased schottky diode

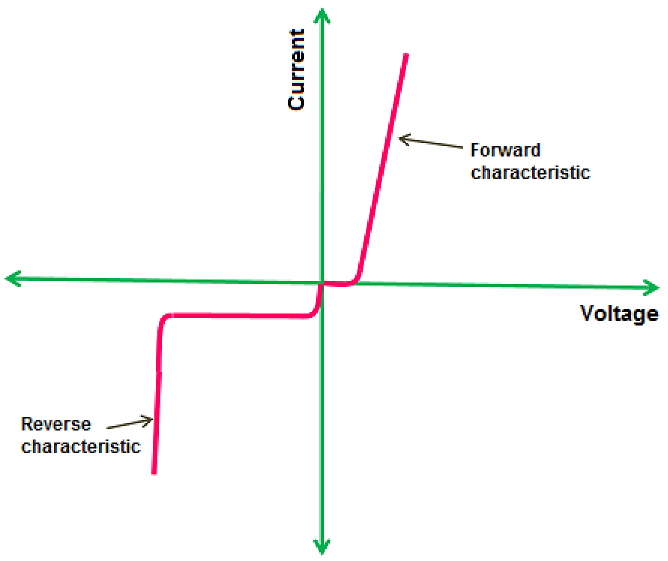
If the reverse bias voltage is continuously increased, the electric current gradually increases due to the weak barrier.

If the reverse bias voltage is largely increased, a sudden rise in electric current takes place. This sudden rise in electric current causes depletion region to break down which may permanently damage the device.

V-I characteristics of schottky diode

The V-I (Voltage-Current) characteristics of schottky diode is shown in the below figure. The vertical line in the below figure represents the current flow in the schottky diode and the horizontal line represents the voltage applied across the schottky diode.

The V-I characteristics of schottky diode is almost similar to the P-N junction diode. However, the forward voltage drop of schottky diode is very low as compared to the P-N junction diode.



V-I characteristics of schottky diode

The forward voltage drop of schottky diode is 0.2 to 0.3 volts whereas the forward voltage drop of silicon P-N junction diode is 0.6 to 0.7 volts.

If the forward bias voltage is greater than 0.2 or 0.3 volts, electric current starts flowing through the schottky diode.

In schottky diode, the reverse saturation current occurs at a very low voltage as compared to the silicon diode.

Difference between schottky diode and P-N junction diode

The main difference between schottky diode and p-n junction diode is as follows:

In schottky diode, the free electrons carry most of the electric current. Holes carry negligible electric current. So schottky diode is a unipolar device. In P-N junction diode, both free electrons and holes carry electric current. So P-N junction diode is a bipolar device.

The reverse breakdown voltage of a schottky diode is very small as compared to the p-n junction diode.

In schottky diode, the depletion region is absent or negligible, whereas in p-n junction diode the depletion region is present.

The turn-on voltage for a schottky diode is very low as compared to the p-n junction diode.

In schottky diode, electrons are the majority carriers in both metal and semiconductor. In P-N junction diode, electrons are the majority carriers in n-region and holes are the majority carriers in p-region.

Advantages of schottky diode

Low junction capacitance

We know that capacitance is the ability to store an electric charge. In a P-N junction diode, the depletion region consists of stored charges. So there exists a capacitance. This capacitance is present at the junction of the diode. So it is known as junction capacitance.

In schottky diode, stored charges or depletion region is negligible. So a schottky diode has a very low capacitance.

Fast reverse recovery time

The amount of time it takes for a diode to switch from ON state to OFF state is called reverse recovery time.

In order to switch from ON (conducting) state to OFF (non-conducting) state, the stored charges in the depletion region must be first discharged or removed before the diode switch to OFF (non-conducting) state.

The P-N junction diode do not immediately switch from ON state to OFF state because it takes some time to discharge or remove stored charges at the depletion region. However, in schottky diode, the depletion region is negligible. So the schottky diode will immediately switch from ON to OFF state.

High current density

We know that the depletion region is negligible in schottky diode. So applying is small voltage is enough to produce large current.

Low forward voltage drop or low turn on voltage

The turn on voltage for schottky diode is very small as compared to the P-N junction diode. The turn on voltage for schottky diode is 0.2 to 0.3 volts whereas for P-N junction diode is 0.6 to 0.7 volts. So applying a small voltage is enough to produce electric current in the schottky diode.

High efficiency

Schottky diodes operate at high frequencies.

Schottky diode produces less unwanted noise than P-N junction diode.

Disadvantages of schottky diode

Large reverse saturation current

Schottky diode produces large reverse saturation current than the p-n junction diode.

Applications of schottky diodes

Schottky diodes are used as general-purpose rectifiers.

Schottky diodes are used in radio frequency (RF) applications.

Schottky diodes are widely used in power supplies.

Schottky diodes are used to detect signals.

Schottky diodes are used in logic circuits.

Varactor diode

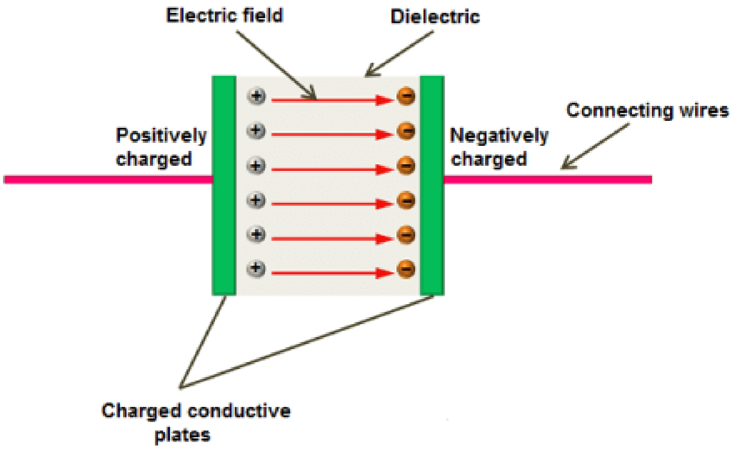
Varactor diode definition

Varactor diode is a p-n junction diode whose capacitance is varied by varying the reverse voltage. Before going to varactor diode, let's first take a look at the capacitor.

What is a capacitor?

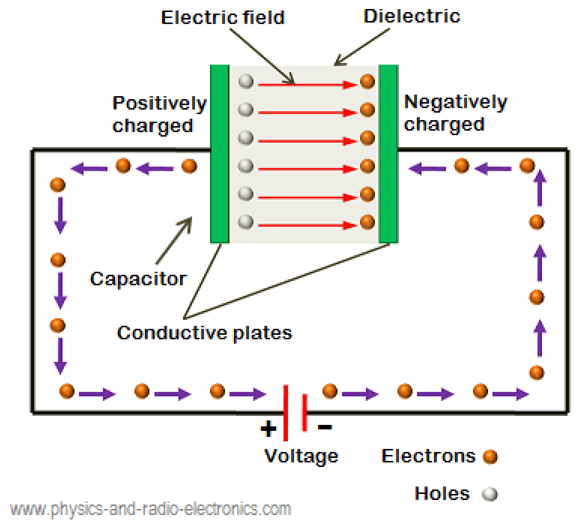
A capacitor is an electronic component that stores electrical energy or electric charge in the form of an electric field.

The basic capacitor is made up of two parallel conductive plates separated by a dielectric. The two conductive plates acts like electrodes and the dielectric acts like an insulator.



The conductive plates are good conductors of electricity so they easily allow electric current through them. On the other hand, a dielectric is poor conductor of electricity so it does not allow electric current through it but it allows electric field or electric force.

When voltage is applied to the capacitor in such a way that the negative terminal of the battery is connected to the right side electrode or plate and the positive terminal of the battery is connected to the left side electrode, the capacitor starts storing electric charge.



Because of this supply voltage, a large number of electrons start flowing from the negative terminal of the battery through a conductive wire. When these electrons enter into the right side plate, a large number ofatoms in the right side plate gains extra electrons. We know that any object that has a larger number of electrons (negative charge carriers) than protons (positive charge carriers) is said to be negatively charged. The right side plate has a larger number of electrons than protons. So the right side plate becomes negatively charged because of the gaining of extra electrons.

The free electrons in the right side plate or electrode will try move into the dielectric. However, dielectric blocks these electrons.

As a result, a large number of electrons are built up on the right side plate. Thus, the right side plate becomes a negatively charged electrode.

The dielectric blocks flow of charge carriers (free electrons) but allows electric force exerted by the negatively charged electrode.

On the other hand, the electrons on the left side plate experience a strong attractive force from the positive terminal of the battery. As a result, a large number of electrons leave the left side plate and flow towards the positive terminal of the battery. As a result, a positive charge is accumulated on the left side plate.

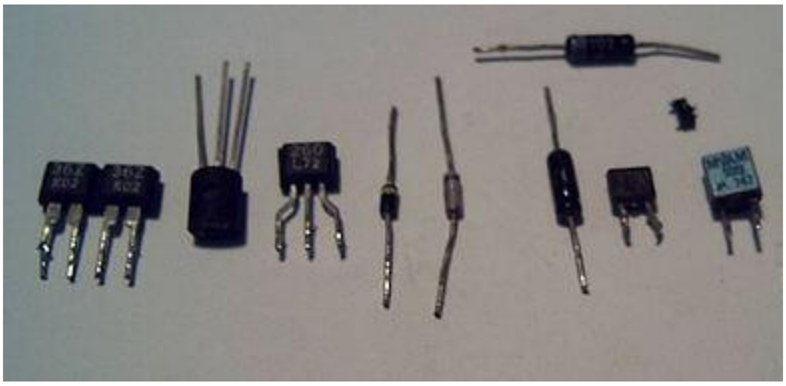
The positive and negative charges accumulated on both plates exert attractive force on each other. This attractive force between the plates is nothing but the electric field between the plates.

We know that the capacitance is the ability to store electric charge. So at both plates, the charge is stored. Thus, there exists a capacitance at both the plates.

What is varactor diode?

The term varactor is originated from a variable capacitor. Varactor diode operates only in reverse bias. The varactor diode acts like a variable capacitor under reverse bias.

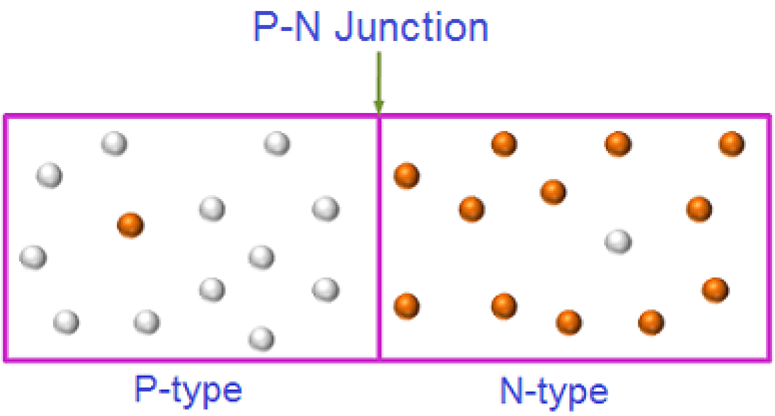
Varactor diode is also sometimes referred to as varicap diode, tuning diode, variable reactance diode, or variable capacitance diode.



The varactor diode is manufactured in such as way that it shows better transition capacitance property than the ordinary diodes.

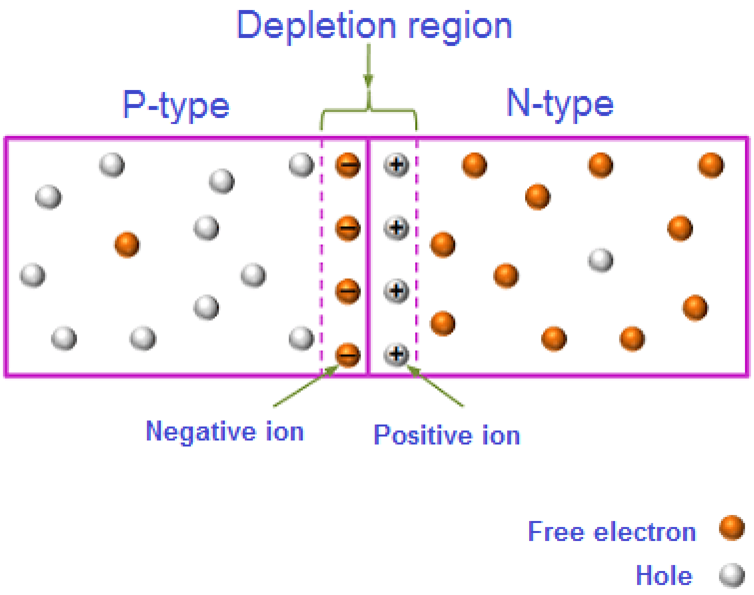
Varactor diode construction

The varactor diode is made up of the p-type and n-type semiconductor. In the n-type semiconductor, free electrons are the majority carriers and holes are the minority carriers. So the free electrons carry most of the electric current in n-type semiconductor. In the p-type semiconductor, holes are the majority carriers and free electrons are the minority carriers. So the holes carry most of the electric current in p-type semiconductor.



When a p-type semiconductor is in contact with the n-type semiconductor, a p-n junction is formed between them. This p-n junction separates the p-type and n-type semiconductor.

At the p-n junction, a depletion region is created. A depletion region is a region where mobile charge carriers (free electrons and holes) are absent.

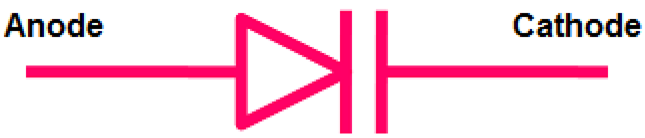


The depletion region is made up of positive and negative ions (charged atoms). These positive and negative ions does not move from one place to another place.

The depletion region blocks free electrons from n-side and holes from p-side. Thus, depletion region blocks electric current across the p-n junction.

Varactor diode symbol

The symbol of a varactor diode is shown in the below figure. The circuit symbol of the varactor diode is almost similar to the normal p-n junction diode.



Varactor diode symbol

Two parallel lines at the cathode side represents two conductive plates and the space between these two parallel lines represents dielectric.

Unbiased varactor diode

We know that in the n-type semiconductor, a large number of free electrons are present and in the p-type semiconductor, a large number of holes are present. The free electrons and holes always try to move from a higher concentration region to a lower concentration region.

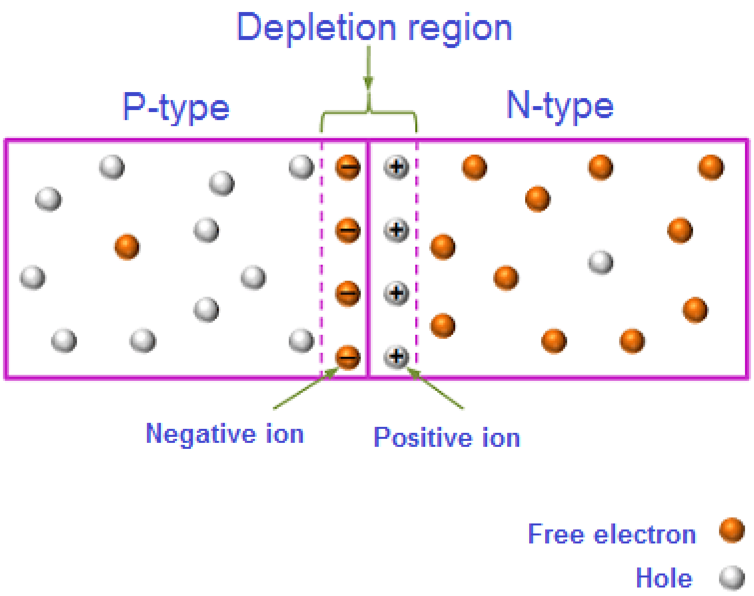
For free electrons, n-region is the higher concentration region and p-region is the lower concentration region. For holes, p-region is the higher concentration region and n-region is the lower concentration region.

Therefore, the free electrons always try to move from n-region to p-region similarly holes always try to move from p-region to n-region.

When no voltage is applied, a large number of free electrons in the n-region get repelled from each other and move towards the p-region.

When the free electrons reach p-n junction, they experience an attractive force from the holes in the p-region. As a result, the free electrons cross the p-n junction. In the similar way, holes also cross the p-n junction. Because of the flow of these charge carriers, a tiny current flows across diode for some period.

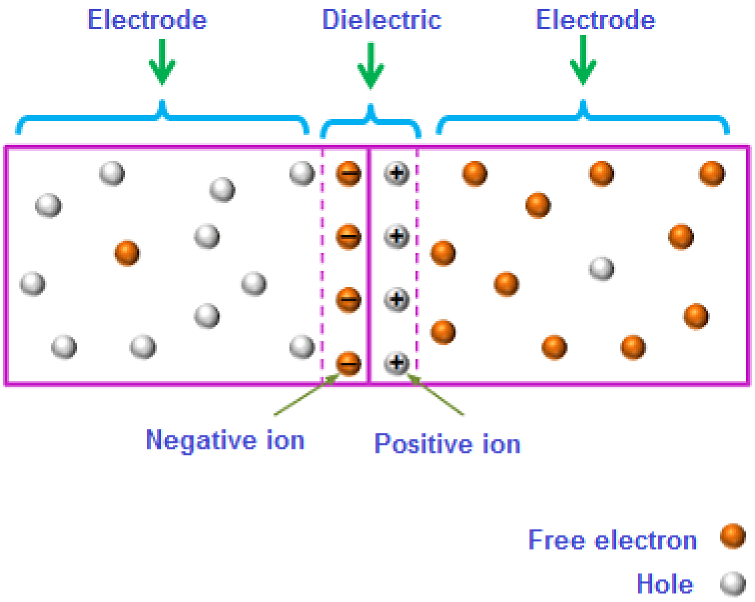
During this process, some neutral atoms near the junction at n-side loses electrons and become positively charged atoms (positive ions) similarly some neutral atoms near the junction at p-side gains extra electrons and become negatively charged atoms (negative ions). These positive and negative ions created at the p-n junction is nothing but depletion region. This depletion region prevents further current flow across the p-n junction.



The width of depletion region depends on the number of impurities added (amount of doping).

A heavily doped varactor diode has a thin depletion layer whereas a lightly doped varactor diode has a wide depletion layer.

We know that an insulator or a dielectric does not allow electric current through it. The depletion region also does not allow electric current through it. So the depletion region acts like a dielectric of a capacitor.



We know that electrodes or conductive plates easily allow electric current through them. The p-type and n-type semiconductor also easily allow electric current through them. So the p-type and n-type semiconductor acts like the electrodes or conductive plates of the capacitor. Thus, varactor diode behaves like a normal capacitor.

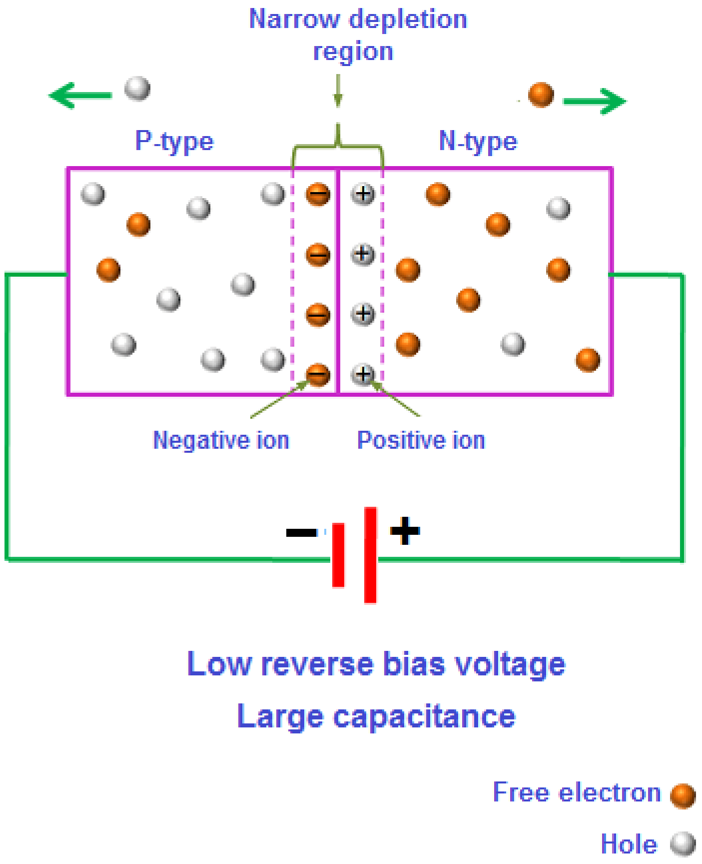
In an unbiased varactor diode, the depletion width is very small. So the capacitance (charge storage) is very large.

How varactor diode works?

The varactor diode should always be operated in reverse bias. Because in reverse bias, the electric current does not flow. When a forward bias voltage is applied, the electric current flows through the diode. As a result, the depletion region becomes negligible. We know that depletion region consists of stored charges. So stored charges becomes negligible which is undesirable.

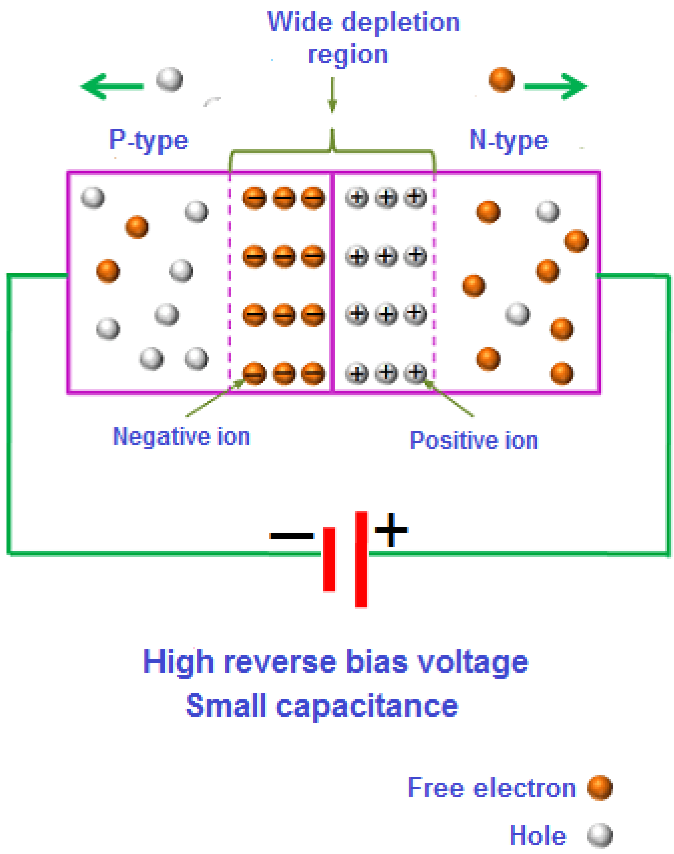
A varactor diode is designed to store electric charge not to conduct electric current. So varactor diode should always be operated in reverse bias.

When a reverse bias voltage is applied, the electrons from n-region and holes from p-region moves away from the junction. As a result, the width of depletion region increases and the capacitance decreases.



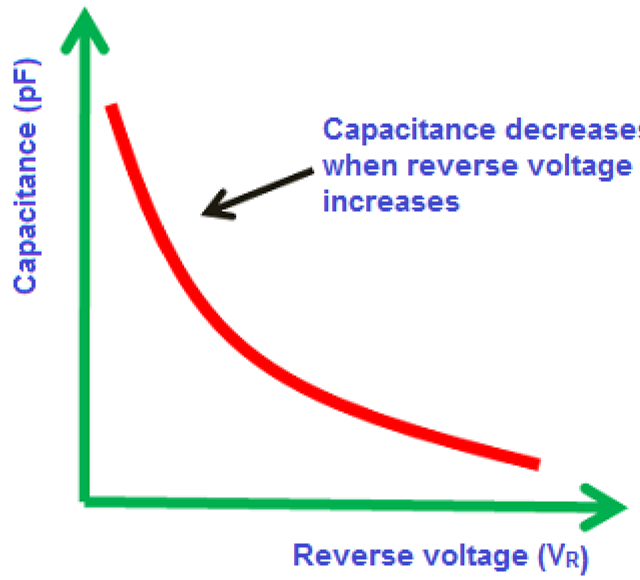
However, if the applied reverse bias voltage is very low the capacitance will be very large.

The capacitance is inversely proportional to the width of the depletion region and directly proportional to the surface area of the p-region and n-region. So the capacitance decreases as the as the width of depletion region increases.



If the reverse bias voltage is increased, the width of depletion region further increases and the capacitance further decreases.

On the other hand, if the reverse bias voltage is reduced, the width of depletion region decreases and the capacitance increases.



Thus, an increase in reverse bias voltage increases the width of the depletion region and decreases the capacitance of a varactor diode.

The decrease in capacitance means the decrease in storage charge. So the reverse bias voltage should be kept at a minimum to achieve large storage charge. Thus, capacitance or transition capacitance can be varied by varying the voltage.

In a fixed capacitor, the capacitance will not be varied whereas, in variable capacitor, the capacitance is varied.

In a varactor diode, the capacitance is varied when the voltage is varied. So the varactor diode is a variable capacitor. The capacitance of a varactor diode is measured in picofarads (pF).

Applications of varactor diode

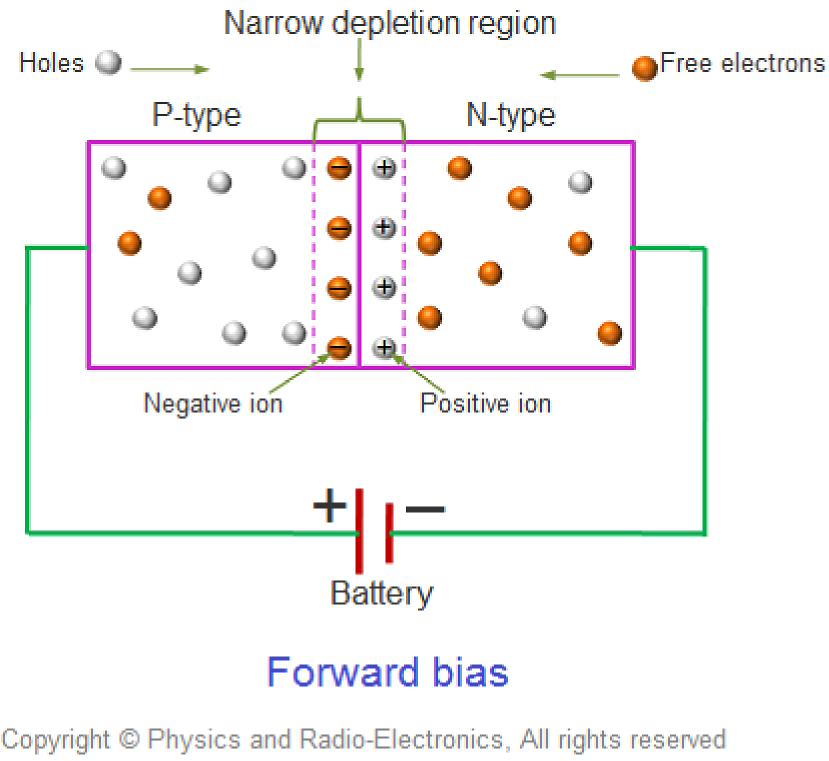
Varactor diode is used in frequency multipliers.

Varactor diode is used in parametric amplifiers.

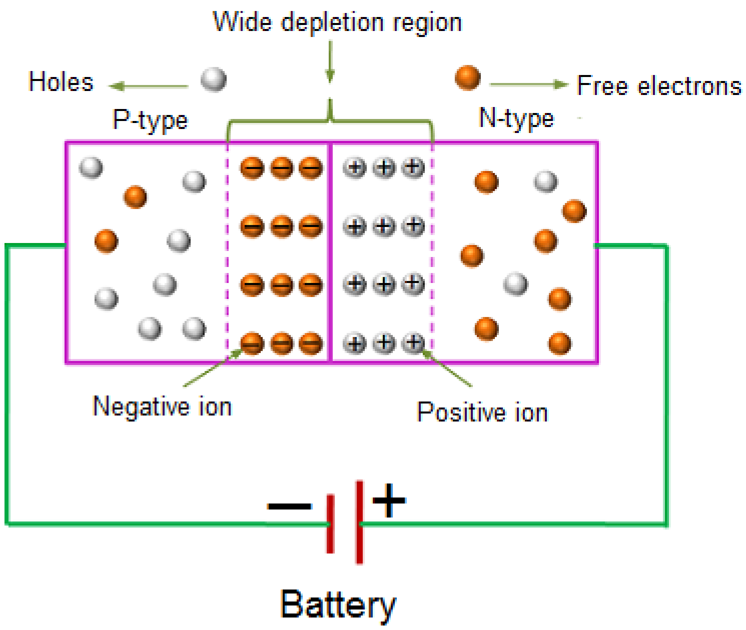
Varactor diode is used in voltage-controlled oscillators.

P-N Junction Diode Applications

A p-n junction diode is a two terminal device that allows electric current in one direction and blocks electric current in another direction.



In forward bias condition, the diode allows electric current whereas in reverse bias condition, the diode does not allow electric current.



Reverse bias

Due to this characteristic, the diode finds number of applications as given below:

Rectification

The conversion of alternating current into direct current is known as rectification. A p-n junction diode allows electric current when it is forward biased and blocks electric current when it is reverse biased. This action of p-n junction diode enables us to use it as a rectifier.

Diodes are used in clamping circuits for DC restoration.

Diodes are used in clipping circuits for wave shaping.

Diodes are used in voltage multipliers.

Diodes are used as switch in digital logic circuits used in computers.

Diodes are used in demodulation circuits.

Laser diodes are used in optical communications.

Light Emitting Diodes (LEDs) are used in digital displays.

Diodes are used in voltage regulators.