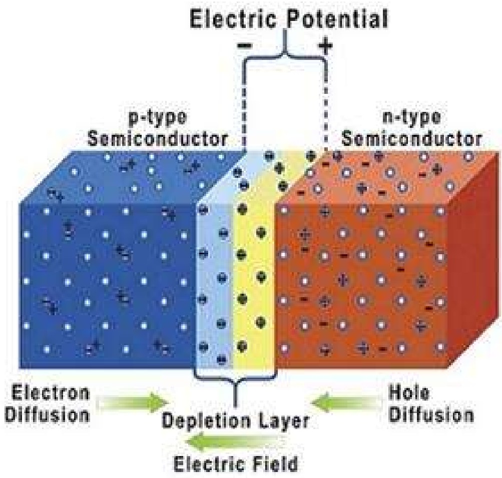
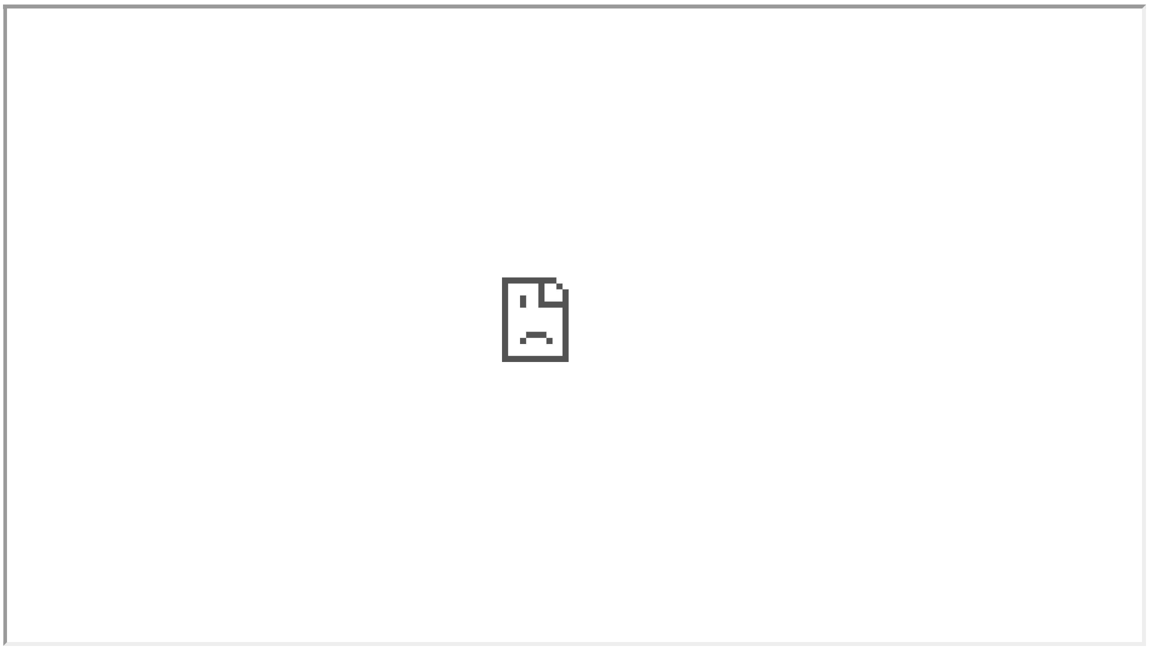
How is PN Junction Formed? Basics and Examples

Author: Apogeeweb **Date:** 11 May 2021 25



Introduction

A **p-n junction** is an interface or a boundary between n-type and p-type semiconductor materials, inside a semiconductor. One of the crucial keys to solid state electronics is the nature of the P-N junction. For example, a PN Junction Diode is one of the simplest semiconductor devices around, and which has the characteristic of passing current in only one direction only. And the p-side or the positive side of the semiconductor has an excess of holes and the n-side or the negative side has an excess of electrons. Why pn junction exists? and How does it work? What is p-n junction diode?



PN Junction Introduction

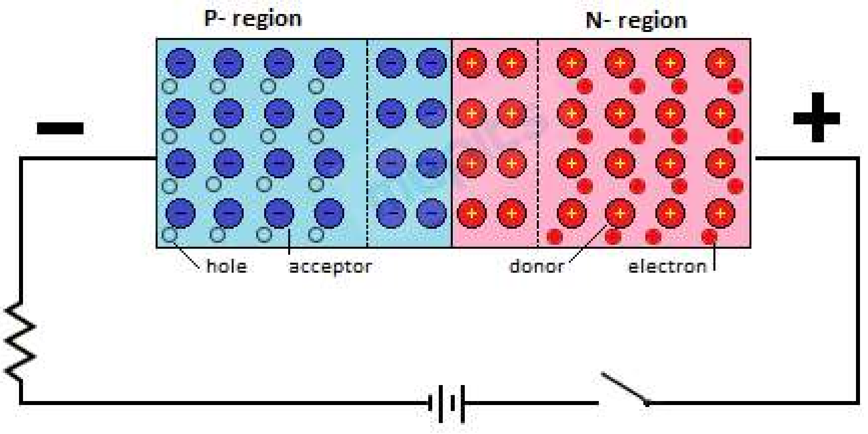
Catalog

I PN Junction Basic

1.1 PN Semiconductor

- N-type Semiconductor

In silicon crystal (or germanium crystal) doped with a small amount of impurity phosphorus element (or antimony element), since semiconductor atoms (such as silicon atoms) are replaced by impurity atoms, among the five outer electrons in the outer layer of phosphorus atoms four of them form covalent bonds with the surrounding atoms, and the extra electron is almost unbound and becomes a free electron more easily. Therefore, the N-type semiconductor has become a semiconductor with a higher concentration of electrons, and its conductivity is mainly due to the conduction of free electrons.



- P-type Semiconductor

In silicon crystal (or germanium crystal) doped with a small amount of impurity boron element (or indium element), since semiconductor atoms (such as silicon atoms) are replaced by impurity atoms, the three outer electrons in the outer layer of boron atoms and a semiconductor atom form a covalent bond, at this time, a "hole" is generated. This hole may attract bound electrons to "fill", making the boron atom a negatively charged ion. In this way, this type of semiconductor has a higher concentration of "holes" ("corresponding to" positive charges) and becomes a substance capable of conducting electricity.

1.2 PN Junction Review

P-N junction is formed by joining **n-type and p-type semiconductor** materials, which is a two terminal device that allows electric current in one direction and blocks electric current in another direction.

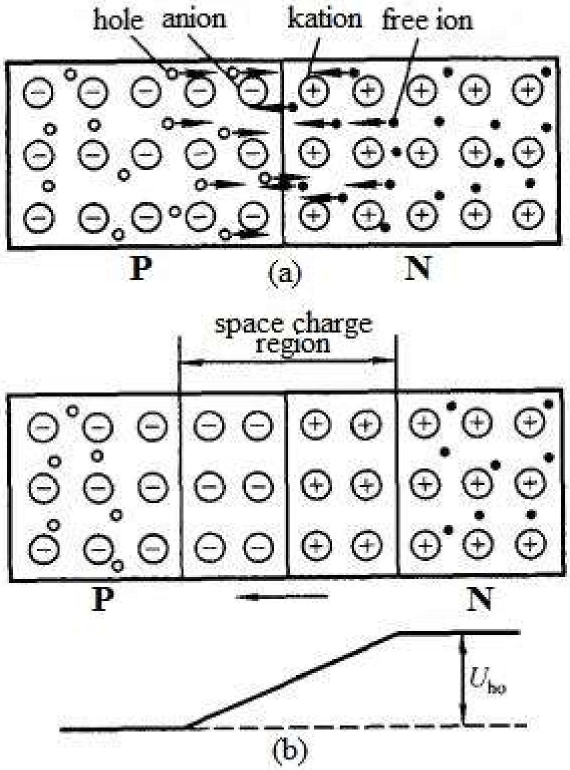
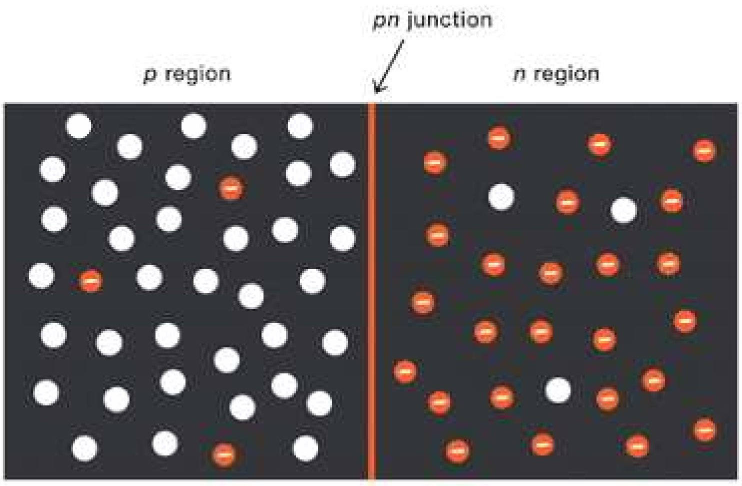


Figure 1. How is PN Junction Formed

On a silicon wafer, different doping processes are used to form an N-type semiconductor on one side and a P-type semiconductor on the other side. We call the area near the interface of the two semiconductors a PN junction.

After the P-type semiconductor and the N-type semiconductor are combined, since the free electrons in the N-type region are more and the holes are less, the concentration difference between electrons and holes appears at their junction. Due to the difference in the concentration of free electrons and holes, some electrons will diffuse from the N-type region to the P-type region, and some holes will diffuse from the P-type region to the N-type region. As a result of their diffusion, the P region loses holes, leaving negatively charged impurity ions, and the N region loses electrons, leaving positively charged impurity ions. The ions in the semiconductor cannot move arbitrarily in an open circuit, so they do not participate in conduction. These immovable charged particles form a space charge zone near the interface between the P and N zones. The thickness of the space charge zone is related to the concentration of dopants.



After the space charge region is formed, due to the interaction between the positive and negative charges, an internal electric field is formed in the space charge region, the direction of which is from the positively charged N region to the negatively charged P region. Obviously, the direction of this electric field is opposite to the direction of carrier diffusion, which used to prevent diffusion.
On the other hand, this electric field will cause the minority carrier holes in the N region to drift to the P region, and the minority carrier electrons in the P region to drift to the N region. The direction of the drift movement is just opposite to the diffusion movement. The holes drifting from the N region to the P region supplement the holes lost in the P region on the original interface, and the electrons drifting from the P region to the N region supplement the electrons lost in the N region on the original interface, which makes the electric charge is reduced and the internal electric field is weakened. Therefore, the result of drift motion is to narrow the space charge region and strengthen the diffusion motion.
Finally, the diffusion of multiple carriers and the drift of minority carriers reach a dynamic balance. On both sides of the junction surface of the P-type semiconductor and the N-type semiconductor, a thin ion layer is left. The charge area formed by this thin ion layer is called a PN junction. The direction of the internal electric field of the PN junction points from the N to the P. It is also called the depletion layer, because lack of electrons.

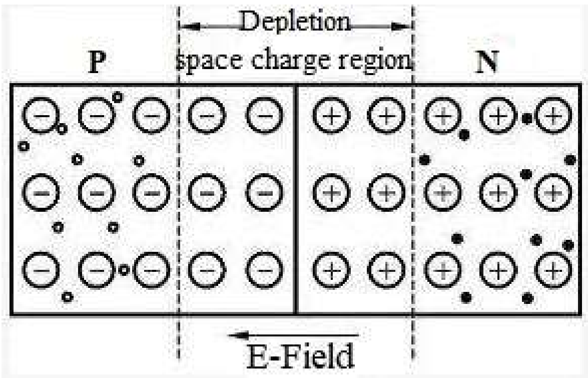


Figure 2. PN Junction Depletion Region

II PN Junction Characteristic

2.1 Unidirectional Conductivity

(1) The PN junction is turned on when the **forward voltage** is applied.
If the positive pole of the power supply is connected to the P area and the negative pole is connected to the N area, a part of the applied forward voltage is in the PN junction area, at this time, the PN junction is in a forward bias. The current flows from the P to the N, and the holes and electrons move to the interface, which narrows the space charge area. In addition, the current can pass smoothly. Its direction is opposite to the direction of the electric field in the PN junction, which weakens the internal electric field. As a result, the resistance of the internal electric field to the diffusion movement of the multitons weakens, and the diffusion current increases. The diffusion current is much larger than the drift current, and the influence of the drift current can be ignored, and the PN junction is in low resistance.
(2) PN junction is cut off when **reverse voltage** is applied.
If the positive pole of the power supply is connected to the N zone, the negative pole is connected to the P zone, and a part of the applied reverse voltage applies in the PN junction zone, and it is in reverse bias. Then the holes and electrons move away from the interface, which widens the space charge area, and the

current cannot flow. The direction is the same as the direction of the electric field in the PN junction, which strengthens the internal electric field. The resistance of the internal electric field to the multiton diffusion movement is enhanced, and the diffusion current is greatly reduced. At this time, the drift current formed by the minority carriers in the PN junction region under the action of the internal electric field is greater than the diffusion current. The diffusion current can be ignored, and the PN junction exhibits high resistance.

2.2 Reverse Breakdown

When a reverse voltage is applied to the PN junction, the space charge region becomes wider and the electric field in the region strengthens. When the reverse voltage increases to a certain level, the reverse current will suddenly increase. If the external circuit cannot limit the current, the current will be so large that it will burn the PN junction. At this time, it is called the breakdown voltage. There are two basic breakdown ways, namely tunnel breakdown (also called Zener breakdown) and avalanche breakdown. The former has a breakdown voltage of less than 6V and has a negative temperature coefficient, and the latter has a breakdown voltage of greater than 6V and a positive temperature coefficient.

2.3 Volt-Ampere Characteristic

The volt-ampere characteristics of the PN junction are shown in the Figure 3, which visually shows the unidirectional conductivity.

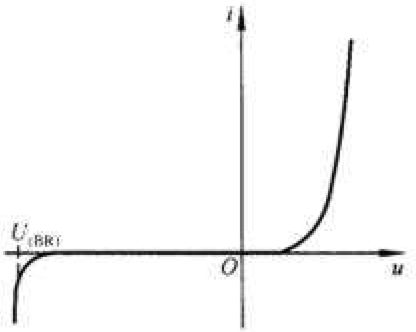


Figure 3. Volt-Ampere Curve of the PN Junction

The volt-ampere characteristic shows by $i_D = I_S\left(e^{\frac{V_D}{V_T}} - 1\right)$

Where i_D is the current passing through the PN junction, V_D is the applied voltage at both ends of the PN junction, and V_T is the voltage equivalent of temperature.

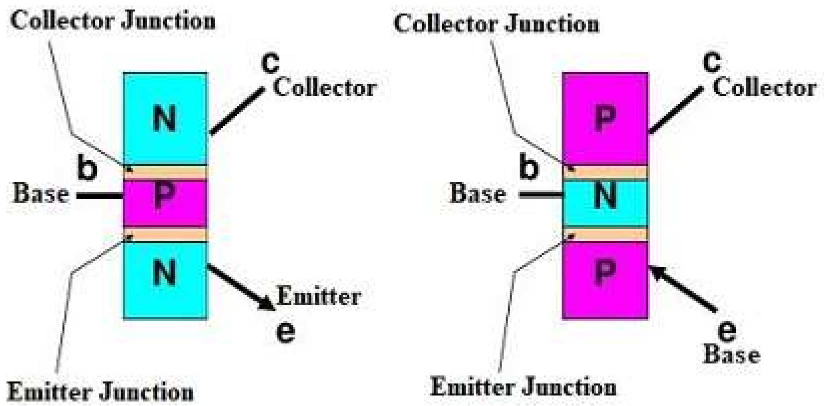
2.4 Capacitance Characteristic

When a reverse voltage is applied to the PN junction, the positive and negative charges in the space charge region constitute a capacitive device. Its capacitance changes with the applied voltage, mainly including barrier capacitance (CB) and diffusion capacitance (CD). Both of them are non-linear capacitors.

III Typical Example: Transistor PN Junction

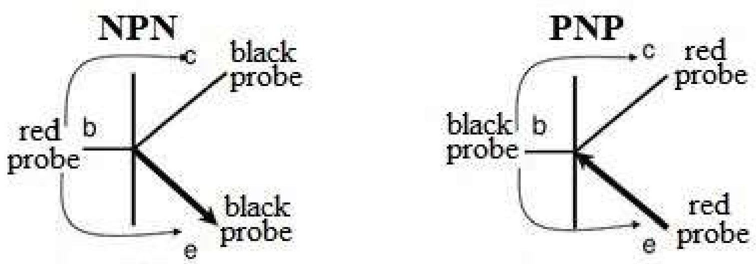
Transistors are one of the basic building blocks of modern electronics. In the diode tutorials we saw that simple diodes are made up from two pieces of semiconductor material to form a simple pn-junction. While the transistor is a three terminal solid state device which is formed by connecting two diodes back to back. Hence it has got two PN junctions.

- Transistor NPN-Type and PNP-Type Junctions



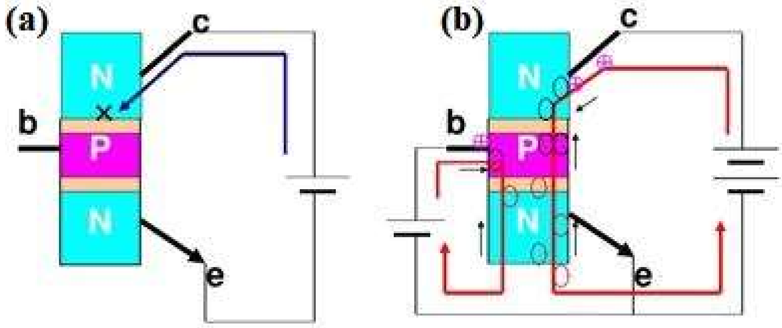
- Transistor Working State

The transistor works like an electronic switch. It can turn a current ON and OFF. The basic idea behind a transistor is that it lets you control the flow of current through one channel by varying the intensity of a much smaller current that's flowing through a second channel.



1) Cut-off State (**C**): The base current is zero.
2) Amplified State (**A**): The transmitter junction is forward biased (that is, the voltage direction is P->N), and the collector junction is reverse biased.
3) Saturation State (**S**): Both the emitter junction and the collector junction are forward biased.

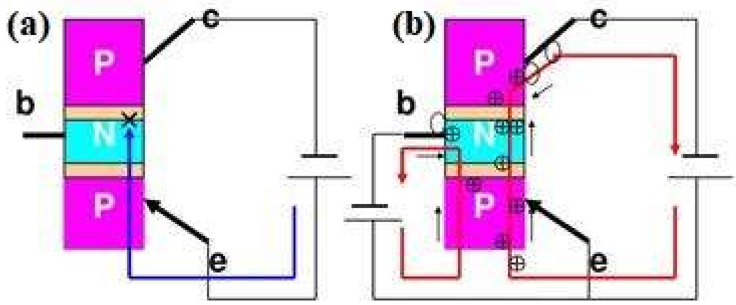
Working State	
NPN-Type Transistor	PNP-Type Transistor
Vb<Ve (C) Vc>Vb>Ve (A) Vb>Ve Vb>Vc (S)	Vb>Ve (C) Vc<Vb<Ve (A) Vb<Ve Vb<Vc (S)



In Figure (a), when there is no voltage input at b of the NPN transistor, no current flows between c and e, and the triode is in the cut-off state.

In Figure (b), when a positive voltage is input to b of the NPN transistor, the negative electrons in the N region of e are attracted by the positrons of P region in b. Due to the effect of the power plant, and they rush (diffuse) to the base region, however, only part of the negative electrons collide with the positive electrons (recombination) to generate the base current, and the other part of the negative electrons gather near the collector junction. The negative electrons gathered in the collector junction pass through (drift) the collector junction due to the action of the electric field. After reaching the collector area, it collides with the positrons gathered in c (N-type semiconductor terminal) to generate a collector current.

It can be seen from this that the greater the base current, the greater the collector current. That is, when a small current is input to the collector, a large current can be obtained by the collector, and now the transistor is in an amplified state.



It should be noted that when the base current reaches a certain level, the collector current no longer rises. At this time, the transistor loses its current amplification effect, and the voltage between the collector and the emitter is very small. The collector and emitter are equivalent to the on-state of the switch. At this moment, the transistor is in a saturated state.

The working principle of the PNP transistor is the same as that of the NPN transistor, except that the direction of the bias voltage and the current are opposite, and the roles of electrons and holes are reversed. The PNP transistor uses Veb to control the positrons incident on the collector area from the emitter area through the base area, while the NPN transistor uses Vbe to control the negative electrons that enter the collector area from the emitter area through the base area.

In addition, in a low-power design, the transistor control circuit will have a certain impact on the circuit. No matter it is NPN or PNP, there will be leakage current in the PN junction of the transistor. When the I/O controls the base voltage, in order to stabilize the base voltage, a pull-down resistor is generally added to the base of the NPN switch circuit. In the design of the PNP switch circuit, a pull-down resistor is added to the base. The pull-up and pull-down resistors are selected according to the control chip, transistor and circuit voltage.

Frequently Asked Questions about PN Junctions Formed

1. What is PN junction and how it is formed?

P-n junctions are formed by joining n-type and p-type semiconductor materials, as shown below. ... However, in a p-n junction, when the electrons and holes move to the other side of the junction, they leave behind exposed charges on dopant atom sites, which are fixed in the crystal lattice and are unable to move.

2. What is p-type and n-type?

In silicon doping, there are two types of impurities: n-type and p-type. In n-type doping, arsenic or phosphorus is added in small quantities to the silicon. ... In p-type doping, boron or gallium is used as the dopant. These elements each have three electrons in their outer orbitals.

3. What is a PN junction diode?

A PN Junction Diode is one of the simplest semiconductor devices around, and which has the characteristic of passing current in only one direction only. ... By applying a negative voltage (reverse bias) results in the free charges being pulled away from the junction resulting in the depletion layer width being increased.

4. What happens in a PN junction?

A forward-biased PN junction conducts a current once the barrier voltage is overcome. The external applied potential forces majority carriers toward the junction where recombination takes place, allowing current flow. A reverse-biased PN junction conducts almost no current.

5. What is a PN junction used for?

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.