ElectronicsTutorials

SEMICONDUCTOR DIODES

Semiconductor Diodes

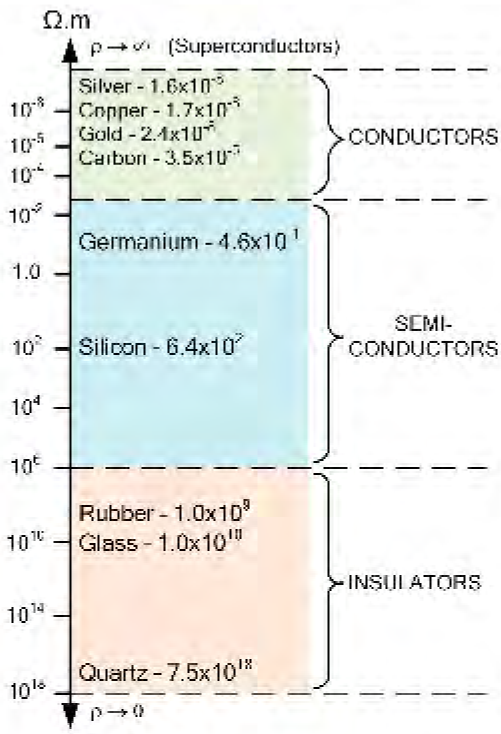
A **Semiconductor Diode** is a two lead solid-state device that allows current to pass through itself in one direction only, acting as a sort off one way valve to electron current flow. A diode is formed by joining together two pieces of semi-conductor material. One of these being a *n*-type material, and the other a *p*-type material.

The resulting semiconductor material has a positive "p-region" at one end and a negative "n-region" at the other end, producing a device which has resistivity value somewhere between that of a conductor and an insulator. But what is a "Semiconductor" material, and what is meant by the terms, n-type and p-type materials. So to understand the operation of a semiconductor diode, we first need to look at what makes something either a **Conductor** or an **Insulator**.

Insulators and Conductors

The electrical **Resistance** of an electrical or electronic component or device is generally defined as being the ratio of the voltage difference across it to the current flowing through it, basic Ohm's Law principals. The problem with using resistance as a measurement is that it depends very much on the physical size of the material being measured as well as the material out of which it is made.

The electrical resistivity of a particular material on the other hand is a measure of how strongly the material opposes the flow of electric current through it and is measured in Ohm-metres, (Ω.m). Thus if the resistivity of various materials is compared, they can be classified into three main groups, Conductors, Insulators and Semiconductors as shown.



Materials which permit the flow of electrons through themselves are called **Conductors**, such as gold, silver, copper, etc. Conductors have low resistivity.

Materials which block the flow of electrons are called **Insulators**, such as rubber, glass, quartz, etc. Insulators have high resistivity.

Materials whose resistivity falls somewhere between those of a conductor and an insulator are commonly called **Semiconductors.**

The ability of semiconductors to conduct electricity can be greatly improved by replacing or adding certain donor or acceptor atoms to their crystalline structure and silicon is by far the most commonly used semiconductor material for making electronic devices.

Notice that there is a very small margin between the resistivity of the conductors such as silver and gold, compared to a much larger margin for the resistivity of the insulators between glass and quartz.

This difference in resistivity between the various materials is due in part to their ambient temperature as metals are much better absorbers and conductors of heat than are insulators.

Semiconductor Basics

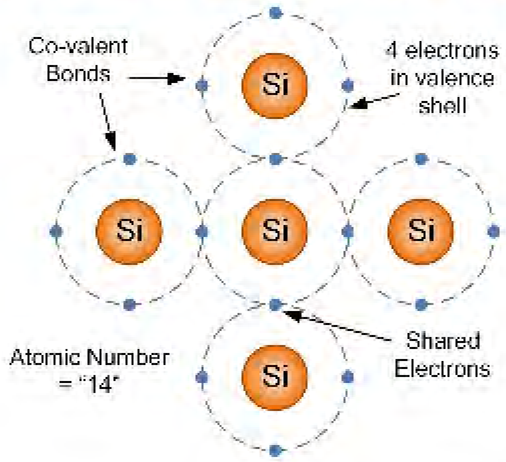
Semiconductors materials such as Silicon (Si), Germanium (Ge) and Gallium Arsenide (GaAs), have electrical properties somewhere between those of a "conductor" and an "insulator". Semiconductors consists of atoms that have no net electric charge (having the same number of electrons as protons), so they neither are good conductors of electricity nor are they good insulators (hence their name "semi"-conductors).

Semiconductor materials made from silicon are used to produce diodes, mosfet's bipolar transistor and all types of integrated circuits

These semiconductor materials have very few "fee electrons" in their valence shell because their atoms are closely grouped together in a tight crystalline pattern called a "crystal lattice". However, their ability to conduct electricity can be greatly improved by adding certain "impurities" to this crystalline structure thereby, producing more free electrons than holes or vice versa.

By controlling the amount of another material added to the semiconductor, it is possible to control its conductivity. These impurities are called donors or acceptors depending on whether they produce electrons or holes respectively. This process of adding impurity atoms to semiconductor atoms (the order of 1 impurity atom per 10 million (or more) atoms of a semiconductor) is called **Doping** and produces an impurity semiconductor.

A SILICON ATOM STRUCTURE



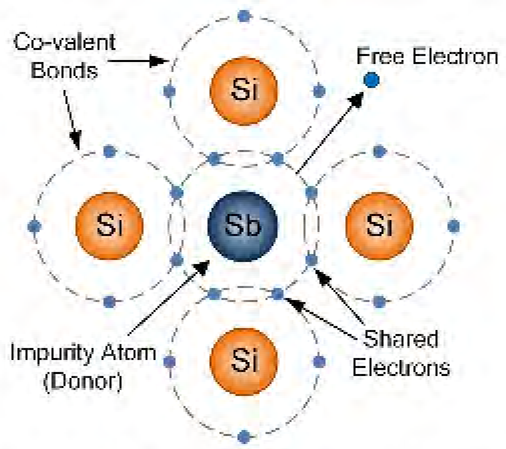
A single silicon atom consists of fourteen negatively charged electrons surrounding a nucleus of fourteen positively charged protons. Thus silicon has the atomic number 14.

As there are an equal number of positive and negative charges, the silicon atom has no net electric charge. Of the fourteen electrons, only the four outer electrons (shown as those in the outermost valence shell) are available for sharing.

The outer four electrons of a silicon atom bond to other silicon atoms with each bond consisting of two electrons, one electron from each silicon atom. This type of bonding, where the outermost electrons are shared equally between neighbouring atoms is called "covalent bonding".

So in order for a silicon crystal consisting of many such bonds to conduct electricity, we need to introduce an impurity atom that has five outer electrons in its outermost valence shell to share with its neighbouring atoms.

N-TYPE SEMICONDUCTOR STRUCTURE

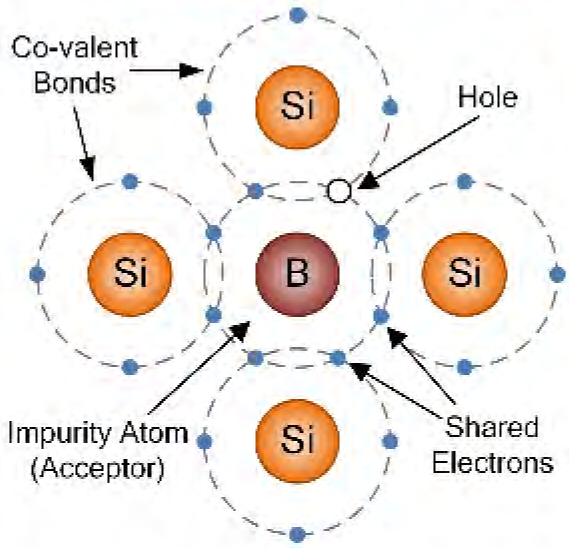


The most common type of "*pentavalent*" (5-electron) impurity used to dope silicon atoms is **Antimony** (symbol Sb) or **Phosphorus** (symbol P), because they have 51 electrons arranged in five shells around their nucleus and having five (5) electrons in its outermost shell for sharing.

After bonding the resulting semiconductor material has an excess of current-carrying free electrons, each with a negative charge, making it an "**n-type**" (*n* for negative) material. These free electrons are called "Majority Carriers" and resulting holes minority carriers.

Then n-type semiconductors have pentavalent impurity atoms (Donors) added to them and conduct by "electron" movement. In these types of materials the donors are positively charged so there are a large number of free electrons. Thus n-type semiconductor materials contain more electrons than holes.

P-TYPE SEMICONDUCTOR STRUCTURE



If we now go the other way, and introduce a "*trivalent*" (3-electron) impurity into the crystalline structure, such as **Boron** (symbol B) or **Indium** (symbol In), which have only three outermost electrons available for bonding, the fourth complete bond cannot be formed.

Therefore, a complete connection is not possible, giving the semiconductor material an abundance of positively charged carriers known as "holes" in the structure of the crystal where electrons are effectively missing.

The doping of Boron atoms causes conduction to consist mainly of positive charge carriers resulting in what is called a "**p-type**" material with the positive holes being called “Majority Carriers” and any free electrons called minority carriers.

Then p-type semiconductors are a material which have trivalent impurity atoms (Acceptors) added and conducts by the movement of "holes". In these types of materials the acceptors are negatively charged and there are a large number of holes for free electrons to fill. Thus p-type semiconductor materials contain more holes than electrons.

So by using different doping agents to a base semiconductor material of either Silicon (S) or Germanium (Ge), it is possible to produce different types of basic semiconductor materials, either n-type or p-type for use in electronic semiconductor components, microprocessor and solar cell applications. Thus we could think of semiconductors as being “part-time” conductors whose conductivity can be controlled by doping.

The PN-junction

The Positive-Negative-Junction, or PN-junction for short, is the basic structure of diodes, rectifiers, transistors, solar cells, light-emitting diodes and any other electronic device which is manufactured using semiconductor materials.

Donor atoms produce a n-type material with more electrons than holes. Acceptor atoms produce a p-type material with more holes than electrons.

A doped silicon crystal will have either free electrons or stationary holes depending upon whether it's an n-type or p-type material And we now know that an n-type material has a large number of free electrons (being negatively charged) and that the p-type material has a large number of free holes (being positively charged).

As both electrons and holes each have an opposite electrical charge, they will be attracted to each other when brought together. This attraction of opposite charges is called diffusion and is important in understanding the formation of pn- junctions.

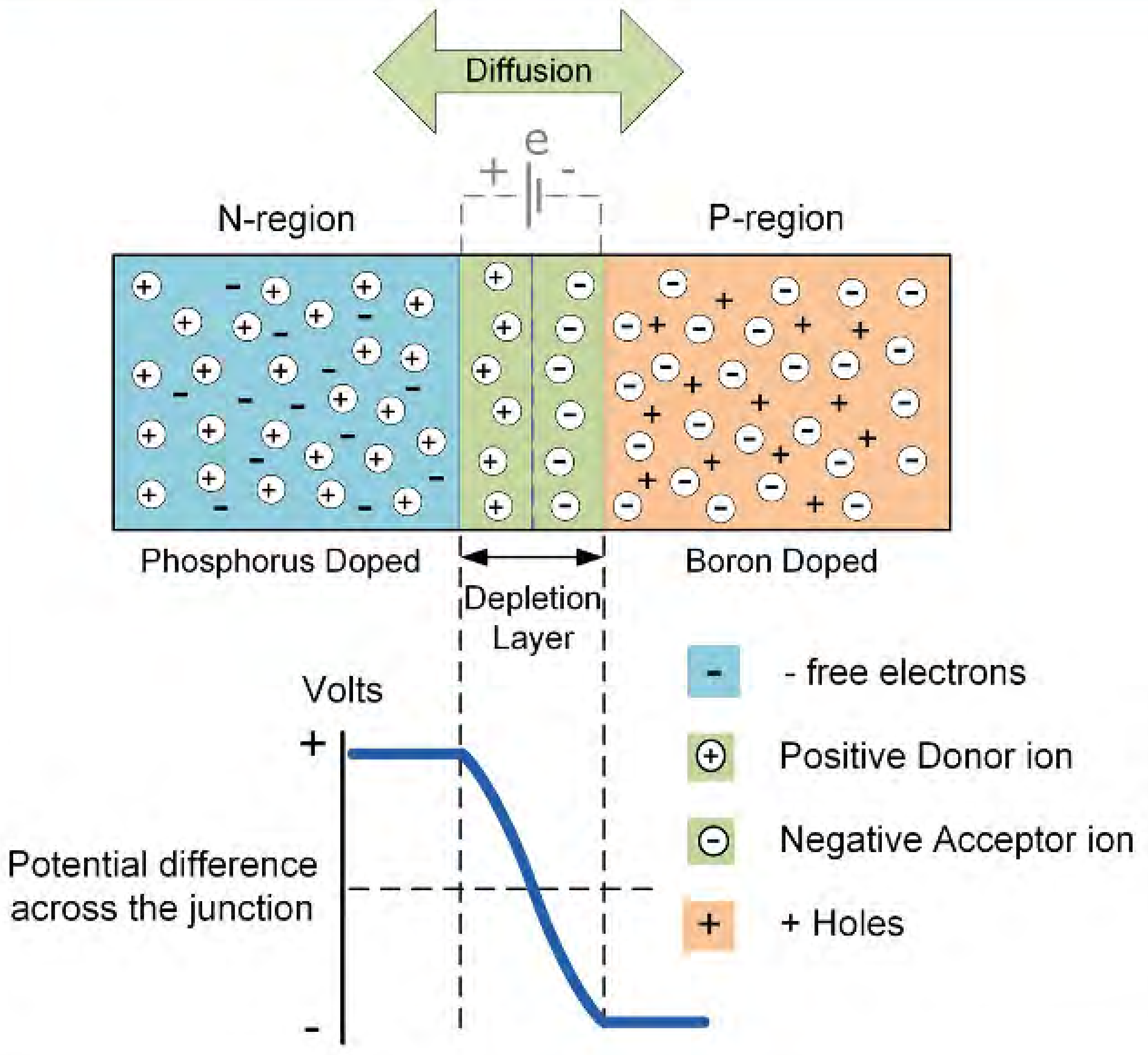
When the n-type and p-type materials are first joined together a very large density gradient exists between both sides of the newly created pn-junction. The result is that some of the free electrons from the negative side begin to migrate across this newly formed junction to fill up the holes in the p-type material producing negative ions.

However, because the electrons have moved across the junction from the negative side to the positive side, they leave behind positively charged donor ions on the negative side. This causes holes from the acceptor impurity to migrate across the junction in the opposite direction into the region where there are large numbers of free electrons.

As a result of this movement from one side to the other, the charge density of the p-type along the junction is filled with negatively charged acceptor ions and likewise, the charge density of the n-type along the junction becomes positive. This charge transfer of electrons and holes across the junction is known as **diffusion**.

This process continues back and forth until the number of electrons which have crossed the junction have a large enough electrical charge to repel or prevent any more charge carriers from crossing over the junction. Eventually a state of equilibrium (electrically neutral situation) will occur producing a "potential barrier" zone around the area of the junction as the donor atoms repel the holes and the acceptor atoms repel the electrons. Thus diffusion establishes a "built-in" electric field.

DEPLETION LAYER FORMATION



Since no free charge carriers can rest in a position where there is a potential barrier, the regions on either sides of the junction now become completely "depleted" of any more free carriers in comparison to the n- and p-type materials further away from the junction. As a result, the depletion region becomes highly resistive and behaves as if it were a perfect insulator. The area around the **PN-Junction** is now called the **Depletion Layer**.

The width of the p- and n-type areas inside the depletion layer depends on how heavily each side is doped. The total charge on each side of a pn-junction must be equal and opposite to maintain a neutral charge condition around the junction.

The electric field created by the diffusion process has produced a potential difference across the junction with an open-circuit (zero bias) condition. The voltage across the depletion layer for silicon is about 0.6 – 0.7 volts and for germanium is about 0.3 – 0.35 volts. This potential barrier will always exist even if the device is not connected to any external power source, as seen in diodes.

Diffusion is the movement of electrons and holes across the pn-junction establishing a built-in electric field around it.

The significance of this built-in potential across the junction is that it opposes both the flow of holes and electrons across it and is why it is called the potential barrier. In practice, a **pn-junction** is formed within a single crystal of semiconductor material rather than just simply joining or fusing together two separate pieces.

Biasing the PN-junction

Without any external voltage being applied to the newly created pn-junction, the junction just sits there in a state of equilibrium. However, if we were to make electrical connections at the ends of both the n-type and the p-type materials and connect them to a battery source, additional energy now exists to overcome the potential barrier.

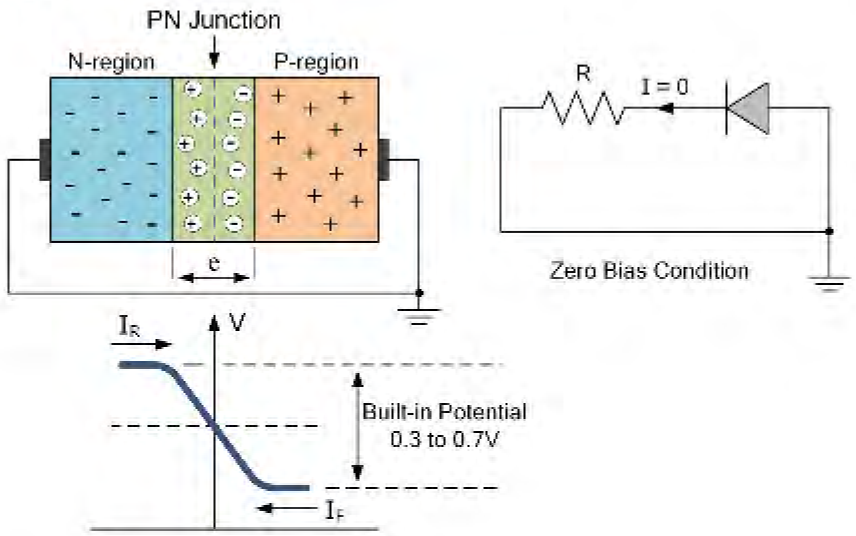
The effect of adding this additional energy source results in the free electrons being able to cross the depletion region from one side to the other.
The biasing of the pn-junction with regards to the potential barrier's width produces an asymmetrical conducting two terminal device, better known as the **PN-Junction Diode**.

There are two operating regions and three possible "biasing" conditions for the standard *Junction Diode* and these are:

- **Zero Bias** – No external voltage potential is applied to the pn-junction and is in a state of equilibrium.
- **Reverse Bias** – The voltage potential is connected negative, (-ve) to the p-type material and positive, (+ve) to the n-type material across the junction which has the effect of **Increasing** the pn-junction's width.
- **Forward Bias** – The voltage potential is connected positive, (+ve) to the p-type material and negative, (-ve) to the n-type material across the junction which has the effect of **Decreasing** the pn-junction's width.

ZERO BIASING THE PN-JUNCTION

When a diode is connected in a **Zero Bias** condition, no external voltage is applied to the pn-junction.

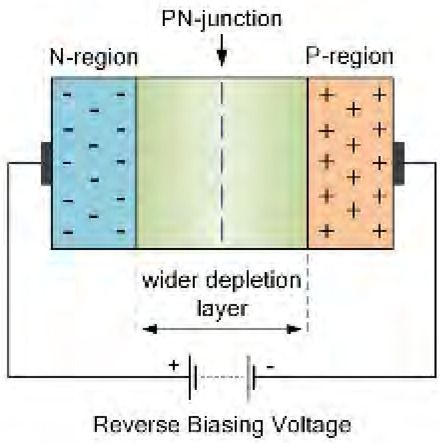


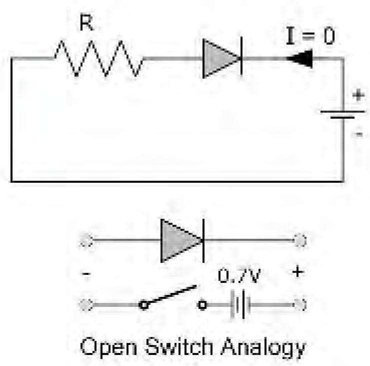
However if the diodes terminals are shorted together, a few holes in the p-type material with enough energy to overcome the potential barrier will move across the junction against this barrier potential. This is known as the **Forward Current**, I_F. Likewise, holes generated in the n-type material move across the junction in the opposite direction. This is known as the Reverse Current, I_R.

Although a voltage difference appears across the junction due to the diffusion process, no electric current flows since no circuit has been connected to the pn-junction.

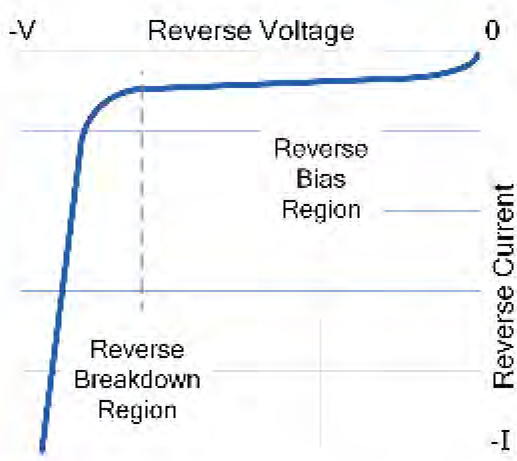
REVERSE BIASING THE PN-JUNCTION

When the pn-junction is connected in a **Reverse Bias** condition, a positive voltage is applied to the n-type material and a negative voltage is applied to the p-type material. The positive voltage applied to the n-type material attracts electrons towards the positive electrode and away from the junction, while the holes in the p-type end are also attracted away from the junction towards the negative electrode.





The net result is that the junctions built-in electric field and the reverse external voltage add together as they are both in the same voltage direction causing the depletion layer to grow even wider.



Due to a lack of electrons and holes in the depletion layer it presents a high impedance path (an open switch) preventing current from flowing through the junction. Thus for a reverse bias condition, a pn-junction acts as an open-switch.

However, a very small **leakage current** does flow through the junction which can be measured in micro-amperes, (μA).

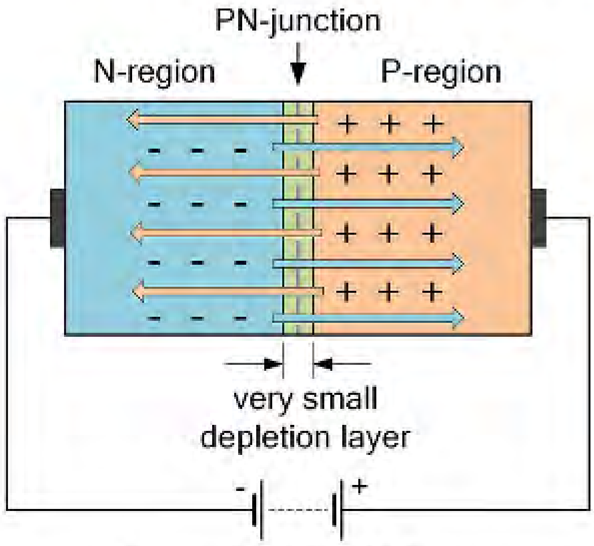
Increasing the external source voltage further makes the depletion layer become wider and more resistive until the point where the external voltage is sufficiently high enough to cause permanent failure of the junction as shown on the I-V curve.

Having said that, the failure of the pn-junction due to an increase in the external reverse voltage has practical applications in many voltage stabilising circuits. It is possible to design the reverse operating characteristics of the pn-junction in such a way that the breakdown voltage is at a specific desired value and where a series resistor is used to limit the reverse breakdown current to a preset maximum value thus producing a fixed voltage output across the diode junction. These types of diodes are commonly known as Zener Diodes and are discussed later.

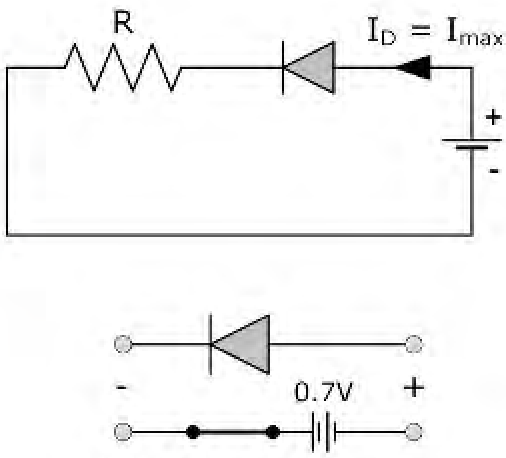
FORWARD BIASING THE PN-JUNCTION

When the pn-junction is connected in a **Forward Bias** condition, a negative voltage is applied to the n-type material and a positive voltage is applied to the p-type material. The net result is that the junctions built-in electric field and the reverse external voltage are in opposite directions causing the depletion layer to become thinner in magnitude than the magnitude of the original "built-in" zero biased condition.

If this external voltage becomes greater than the value of the potential barrier, approx. 0.6 to 0.7 volts for silicon semiconductors and 0.25 to 0.3 volts for germanium semiconductors, the potential barriers opposition will be overcome and electrical current will start to flow unimpeded through the junction.

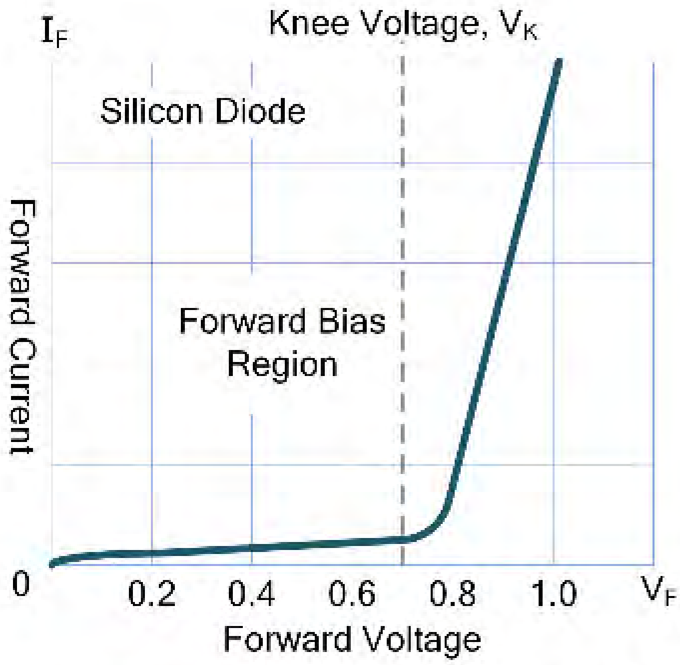


Forward Biasing Voltage



Closed Switch Analogy

This is because the negative side of the external voltage pushes or repels electrons towards the junction giving them the energy to cross over and combine with the holes being pushed in the opposite direction towards the junction by the positive side of the external voltage.



This condition represents the low impedance path (a closed switch) condition through the pn-junction allowing very large currents to flow with only a small increase in the external biasing voltage. Thus for a forward bias condition, a pn-junction acts as a closed-switch.

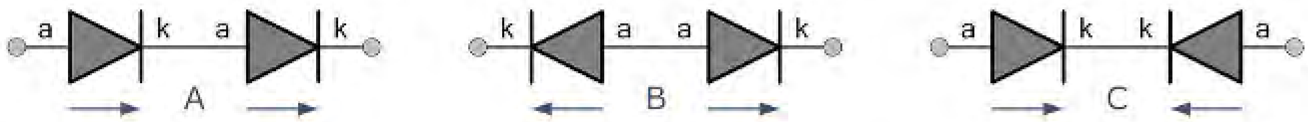
This results in a characteristics curve of virtually zero current flowing up to this voltage point, called the "knee" on the static curves. Above the 0.7 V point (for silicon) current flows through the diode with little increase in the external voltage as shown.

Since the diode can conduct "infinite" current above this knee point as it effectively becomes a closed switch, resistors are used in series with the diode to limit the current flow to a safe value and as long as the junction is conducting current, the forward voltage drop V_F across the diode is approximately equal to the knee voltage, V_K.

The Semiconductor Diode

As we have seen, the pn-junction is the basic building block of the semiconductor diode. Current rises rapidly through the junction when forward biased and practically zero current flows when reverse biased, up to the junctions reverse breakdown voltage. Then we can use this "open" or "closed" switching effect of the junction to good use in many different circuits.

Note that diodes being one way switching devices cannot be connected together in series randomly as only circuit combination "A" will conduct current.

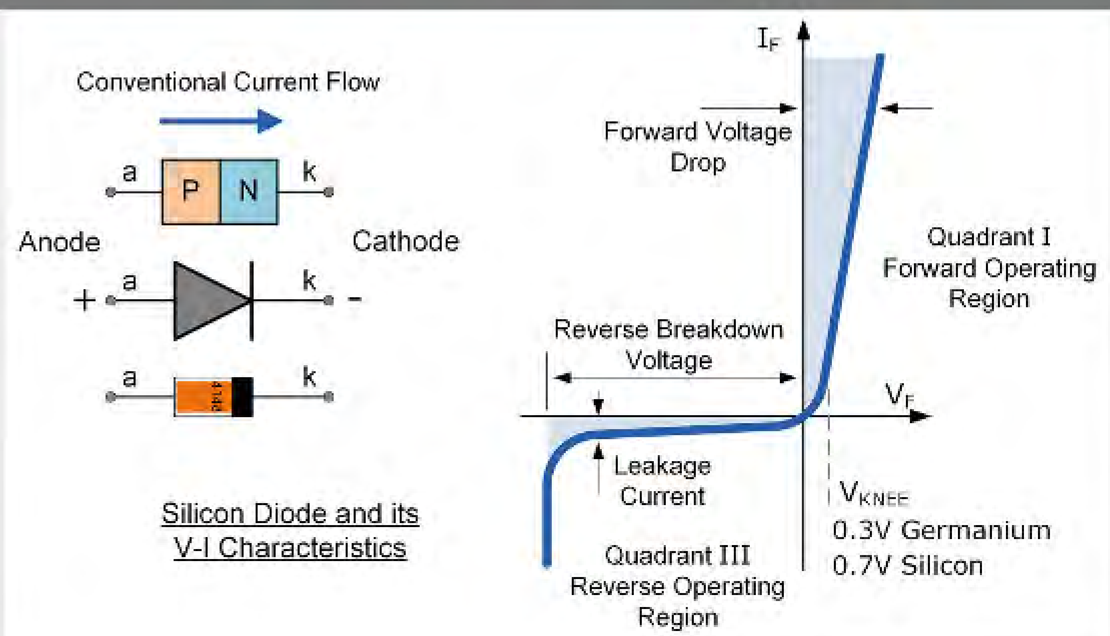


There are basically two types of pn-junction semiconductor diode, the small *Signal Diode* and the *Power Diode*. The signal diode (for example the 1N4148) is a small semiconductor device generally used in electronic circuits, where small currents or high frequencies are involved such as in clamping, wave-shaping or digital logic circuits.

Power diodes (for example the 1N4007) on the other hand have much larger semiconductor pn-junctions compared to the smaller signal diode, and are therefore capable of passing much large forward currents at high voltage values for use in rectification circuits.

The electronic symbol given for any type of diode, either signal or power diode, is that of an arrow with a bar or line at its end and this is illustrated below along with the Steady State I-V Characteristics Curve.

SEMICONDUCTOR DIODE I-V CHARACTERISTIC CURVE



The arrow always points in the direction of conventional current flow through the diode meaning that the diode will only conduct if a positive supply is connected to the *Anode*, (a) terminal and a negative supply is connected to the *Cathode* (k) terminal. The cathode (negative end) is often marked with a band for identification.

Semiconductor Diode Parameters

PN-junction diodes are manufactured in a wide variety of voltage and current ratings which must be taken into account when choosing a diode for a particular application or circuit. Typical diode parameters to considered are:

- Maximum Forward Current, (I_F)
- Forward Voltage Drop, (V_F)
- Peak Inverse Voltage, (PIV)
- Maximum Operating Temperature, (T)

When forward biased and conducting, a diodes pn-junction has a small forward dynamic resistance, so $I^2 * R$ power must be dissipated across it as it conducts causing heat to be generated directly at the junction. Therefore to prevent overheating of the pn-junction, a diode has a *Maximum Forward Current* rating.

The *Forward Voltage* Drop is the voltage drop across anode and cathode terminals of the diode at a defined current level when it is forward biased. It is given by the relationship of: P_D/I_F.

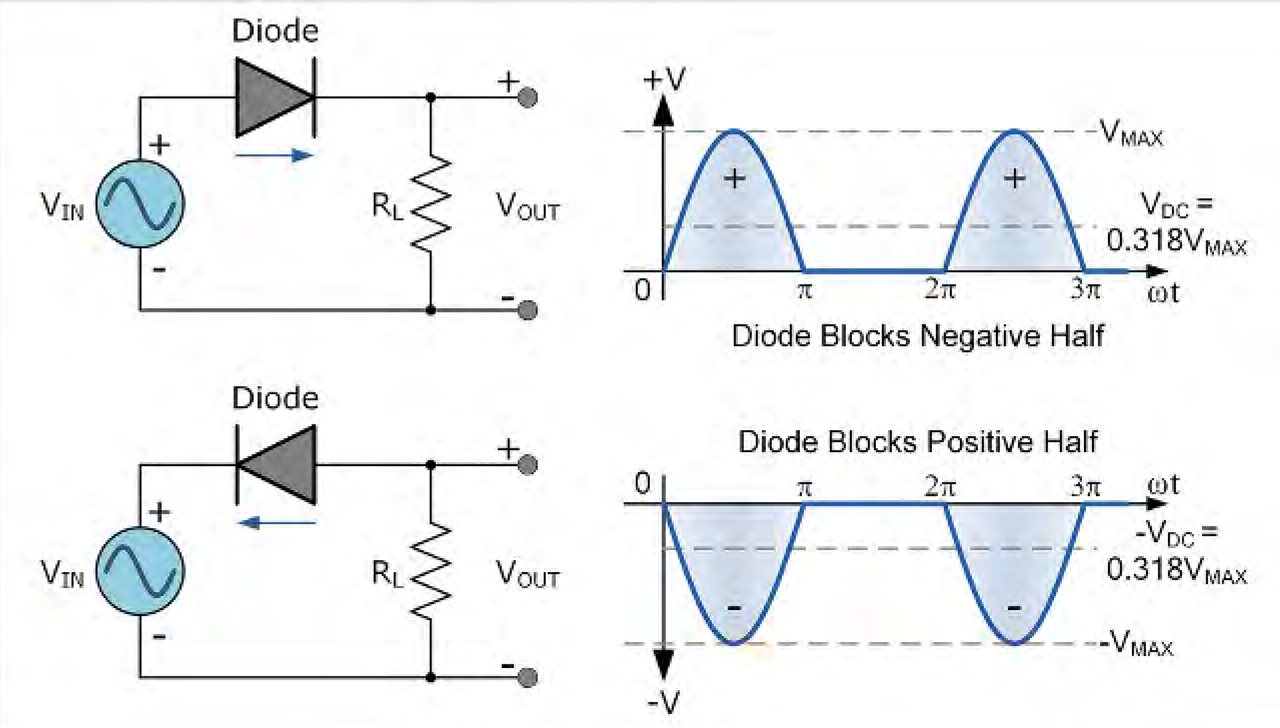
The *Peak Inverse Voltage* is the maximum voltage value allowed across the A and K terminals of a diode in the reverse direction without reverse breakdown occurring, Thus PIV represents the maximum reverse voltage that may be applied to a diode.

The *Maximum Operating Temperature* is the maximum ambient temperature that the diode can operate in before the pn-junction of the diode deteriorates. The maximum forward current is chosen so that the junction temperature does not exceed this.

Applications of the Diode

One particular application of a pn-junction diode is that of the half-wave rectifier where either the positive or negative half of the alternating input waveform is blocked depending on the orientation of the diode. In fact one of the fundamental applications of the diode is in ***Rectification****. Rectification is the process of changing an alternating AC input supply into a pulsed or continuous DC output supply that has only one single polarity.*

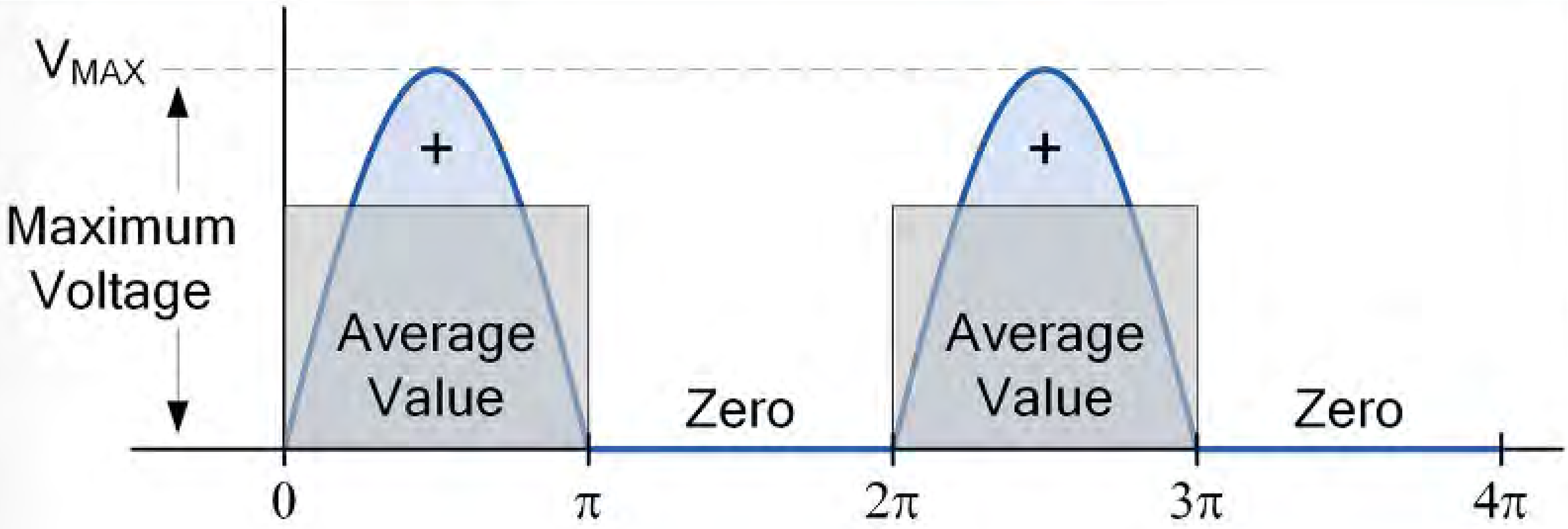
HALF-WAVE RECTIFIER CIRCUITS



During each half cycle of the AC sine wave in which the diode is forward biased, that is the Anode is positive with respect to the Cathode, results in current flowing through the diode. Thus $I_{OUT} = I_{IN}$. During the other half cycle of the AC input waveform, the diode is reverse biased as the Anode is negative with respect to the Cathode. Therefore, no current flows through the diode so $I_{OUT} = 0$.

The current on the DC side of the circuit flows in one direction only making the circuit **Unidirectional**. As the load resistor receives from the diode one half of the waveform, then zero volts, then one half of the waveform, then zero volts etc, producing an output which consists of half-cycles only. This pulsating output waveform not only varies ON and OFF every cycle, but is only present 50% of the time and with a purely resistive load, this high voltage and current ripple content is at its maximum.

HALF-WAVE RECTIFIER AVERAGE OUTPUT VALUE



This pulsating voltage means that the equivalent DC value dropped across the load resistor, RL is therefore only one half of the sinusoidal waveforms average value. So the equivalent DC voltage value taken over one-half of a waveform is defined as: 0.637 x maximum amplitude value. Thus the equivalent DC voltage is given as: $0.318*V_{MAX}$ for the maximum voltage amplitude or $0.45*V_{RMS}$ for the root mean squared value of each half-cycle.

$$V_{AVE} = 0.318 * V_{MAX} = 0.45 * V_{RMS}$$

and

$$I_{AVE} = 0.318 * I_{MAX} = 0.45 * I_{RMS}$$

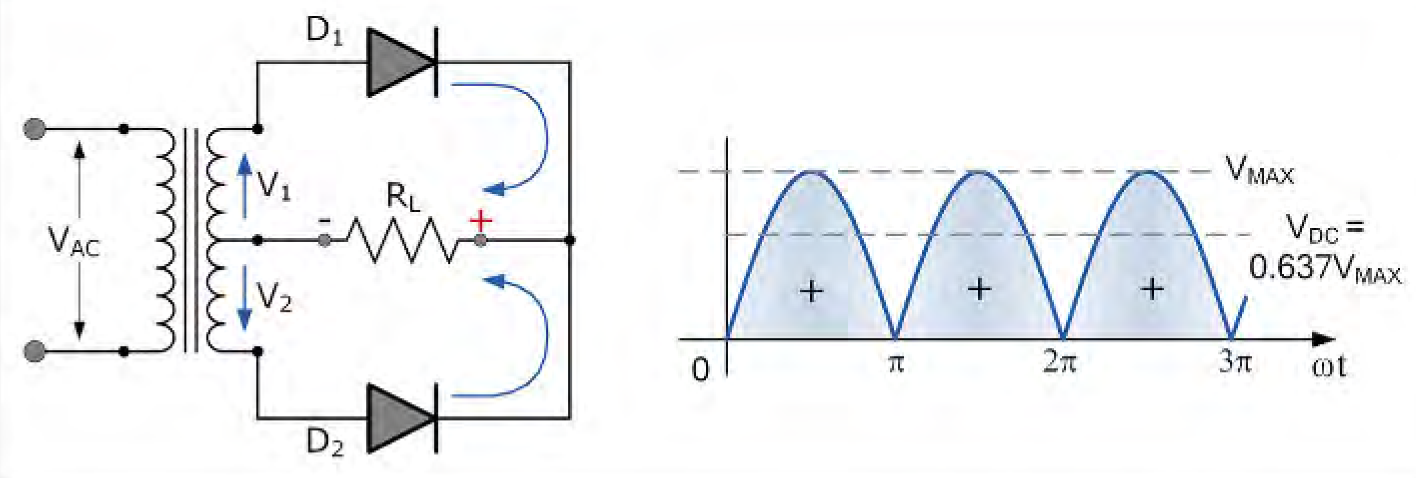
Then we can see that a half-wave rectifier circuit converts either the positive or negative halves of an AC waveform into a pulsed DC output that has an average value of $0.318*A_{MAX}$ or $0.45*A_{RMS}$.

FULL-WAVE RECTIFICATION

The full-wave rectifier utilises both halves of the input sinusoidal waveform to provide a unidirectional output that basically consists of two half-wave rectifiers connected together to feed the connected load.

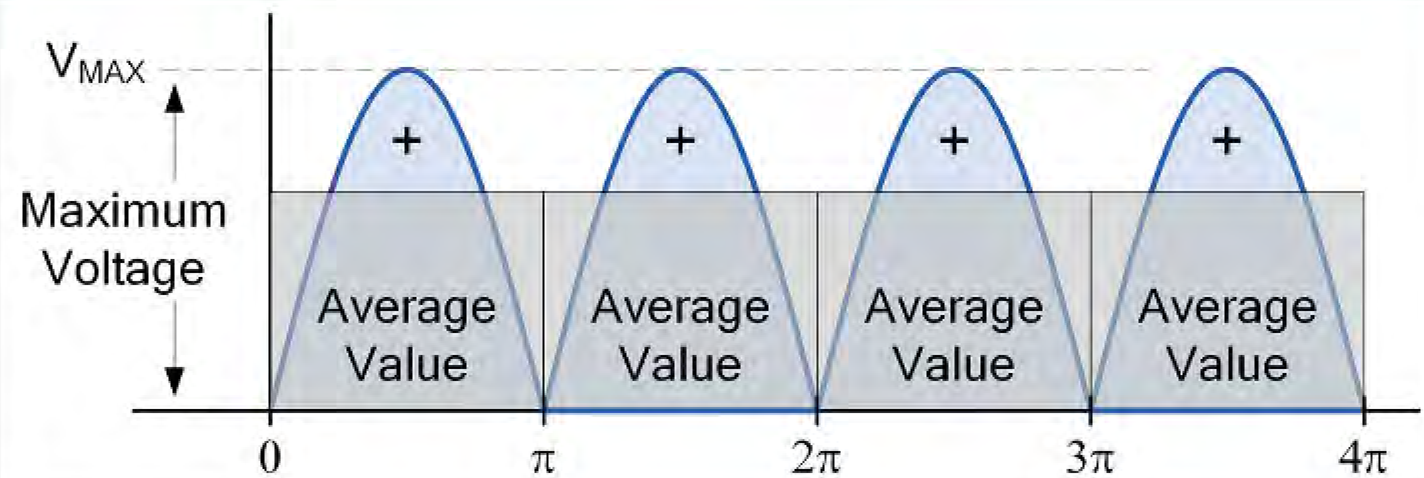
The full-wave rectifier does this by using two diodes to pass both the negative and positive half cycles of the input waveform, while inverting the negative half of the input waveform to create a pulsating DC output. Even though the voltage and current output from the rectifier is pulsating, it does not reverse direction using the full 100% of the input waveform and thus providing full-wave rectification.

FULL-WAVE RECTIFIER CIRCUIT



Although this pulsating output waveform uses 100% of the input waveform, its average DC voltage (or current) is not at the same value. We remember from above that the average or mean DC value taken over one-half of a sinusoid is defined as: 0.318 x maximum amplitude value. However unlike half-wave rectification above, full-wave rectifiers have two positive half-cycles per input waveform giving us a different average value as shown.

FULL-WAVE RECTIFIER AVERAGE OUTPUT VALUE



Here we can see that for a full-wave rectifier, for each positive peak there is an average value of $0.318*V_{MAX}$ and as there are two peaks per input waveform, this means there are two lots of average value summed together. Thus the DC output voltage of a full-wave rectifier is twice that of the previous half-wave rectifier giving an average value of $0.637*V_{MAX}$.

Thus the equivalent DC voltage is given as: $0.637 * V_{MAX}$ for the maximum voltage amplitude or $0.9 * V_{RMS}$ for the root mean squared value of each cycle.

$$V_{AVE} = 0.637 * V_{MAX} = 0.9 * V_{RMS}$$

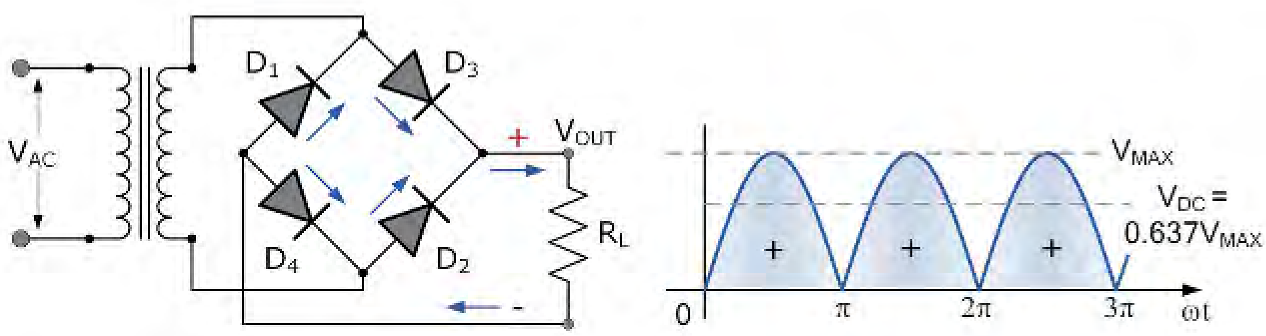
and

$$I_{AVE} = 0.637 * I_{MAX} = 0.9 * I_{RMS}$$

Then we can see that a full-wave rectifier circuit converts both the positive or negative halves of an AC waveform into a pulsed DC output that has an average value of $0.637 * A_{MAX}$ or $0.9 * A_{RMS}$. Note that a full-wave rectifier circuit reduces the output ripple by a factor of two. Also the output DC ripple is twice the input supply frequency.

THE BRIDGE RECTIFIER

The full-wave bridge rectifier circuit uses four diodes arranged in a bridge arrangement to pass both halves of the input waveform to the output producing a pulsating DC output. At any time two of the four diodes in the bridge are forward biased while the other two are reverse biased. Thus there are two diodes in the conduction path instead of the single one for the half-wave rectifier.



During the positive half cycle of V_{IN}, diodes D_3 and D_4 are forward biased while diodes D_1 and D_2 are reverse biased. Then for the positive half cycle of the input waveform, current flows along the path of: $D_3 - R_L - D_4$ and back to the supply.

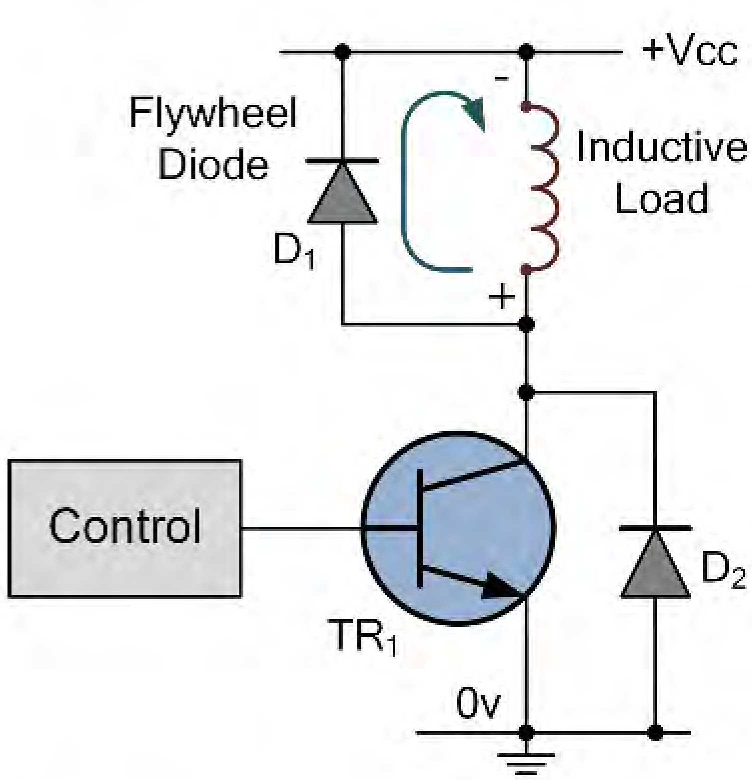
During the negative half cycle of V_{IN}, diodes D_2 and D_1 are forward biased while diodes D_3 and D_4 are reverse biased. Then for the negative half cycle of the input waveform, current flows along the path of: $D_2 - R_L - D_1$ and back to the supply.

Although we can use four individual power diodes to make a full-wave bridge rectifier, ready-made bridge rectifier components are available "off-the-shelf" in a range of different voltage and current sizes and which can be soldered directly onto PCB circuit boards or be connected by spade connectors to wires.

THE FLYWHEEL DIODE

When a transistor, MOSFET, switch or sensor is used to control an inductive load, for example a relay coil, solenoid, transformer, small DC motor, etc. the energy stored in the magnetic field of the load will subject switching device to a high reverse voltage when the load current is disconnected (turned-off).

This is because opening of the switching device does not immediately stop current flowing in the inductive load. Thus we need a way to safely dissipate the energy stored in the inductance of the load when the switching device is opened and we can do this by using a transient suppression diode.



The diode in the circuit is called a flywheel diode, flyback diode, freewheeling diode, or suppression diode. The diode can be connected in parallel across the inductive load as shown by D_1 or across the semiconductor switching device as shown by D_2, or both.

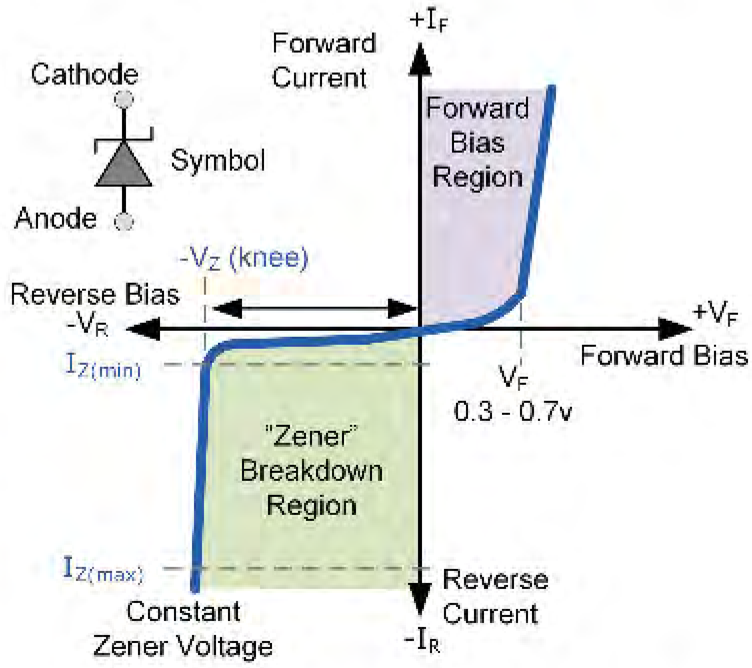
When the inductive load is switched-on and passing current, diode D_1 is reverse-biased and blocking. When the load is switched-off, the rapid collapse of the coils magnetic field causes a large back emf of possibly hundreds of volts to build up across the coil.

As this emf is the reversal of the DC supply, the diode now becomes forward biased and clamps the reverse voltage across the coil to about 0.7V dissipating the stored energy and protecting the switching transistor.

Flywheel diodes are only applicable when the supply is a polarised DC voltage. Transient suppression of an AC coil requires a different protection method, and for this an RC snubber network or MOV is used.

Zener Diodes

Previously we said that if a normal pn-junction diode is reverse-biased with a sufficiently large enough reverse voltage, it will suffer from premature breakdown and become damaged. The **Zener Diode** or "Breakdown Diode", as they are sometimes referred too, is a type of semiconductor diode which is designed to take advantage of the reverse voltage applied to it operating in its breakdown region.



A *zener diode* behaves just like a normal general-purpose diode when biased in the forward direction, which is Anode positive with respect to its Cathode, passing the rated forward current.

However, unlike a conventional diode that blocks any flow of current through itself when reverse biased, that is the Cathode becomes more positive than the Anode, as soon as the reverse voltage reaches a pre-determined value, the zener diode begins to conduct in the reverse direction.

A **Zener Diode** is used in its "reverse bias" or reverse breakdown mode for voltage regulation in that the diodes anode connects to the negative side of the supply. When sufficient reverse voltage is applied (cathode end biased positively), the zener is driven into its reverse breakdown avalanche mode of operation.

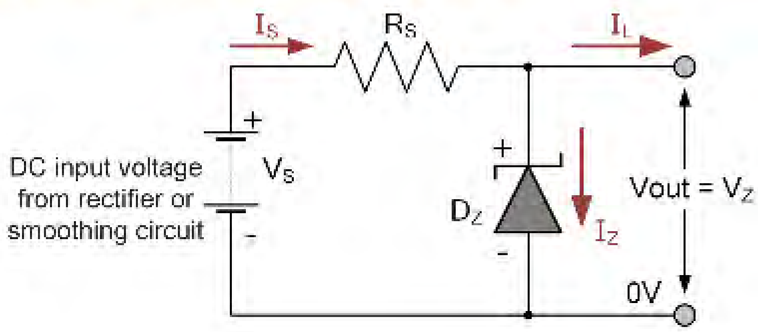
From the I-V characteristics curve we can see that the transition into avalanche breakdown, the "knee" point, a zener diode has an almost constant negative voltage, V_Z regardless of the value of the current flowing through the diode.

This zener value remains nearly constant even with large variations in forward current as long as the zener diodes current remains between its breakdown current, $I_{Z(MIN)}$ and its maximum handling current rating, $I_{Z(MAX)}$.

Thus a zener diode can conduct current in both directions with the forward current I_F being a function of forward voltage V_F, and the reverse zener current, I_Z being a function of the reverse breakdown voltage V_Z.

This ability to control itself can be used to great effect to regulate or stabilise a voltage source against supply or load variations. The fact that the voltage across the diode in the breakdown region is almost constant turns out to be an important characteristic of the zener diode as voltage regulation is the most common application of a zener diode.

ZENER DIODE REGULATION



The basic voltage regulator circuit shows resistor, R_S is connected in series with the zener diode to limit the current flowing through the diode. The supply voltage, V_S which is greater than V_Z is connected across the combination.

The stabilised output voltage V_{OUT} which is also the zener voltage V_Z is taken from across the zener diode.

The zener diode is connected with its cathode terminal connected to the positive rail of the DC supply so it is reverse biased and will be operating in its breakdown condition. Resistor R_S is selected so to limit the maximum current flowing in the circuit and power dissipation in a zener diode.

Zener Diodes can be used to produce a stabilised output

With no load connected to the circuit, the load current will therefore be zero, ($I_L = 0$), so all the circuit current passes through the zener diode which in turn dissipates its maximum power.

The load resistance RL is connected in parallel with the zener diode thus $V_{OUT} = V_Z$. The supply current is now split between the zener diode and the load with the voltage drop across the series resistor being: $V_S - V_Z$.

Under light load (high-impedance) conditions, nearly all the supply current flows through the zener diode and so the zener diode will attempt to limit the output voltage to the zener potential. Under heavier load (low-impedance) conditions, most of the supply current is taken by the load with less current flowing through the zener diode.

If the current drawn by the load current becomes too excessive, then there is no supply current available for the zener diode so it stops conducting and so output voltage regulation is lost. Then there is a minimum zener current, IZ for which the stabilization of the zener voltage, VZ is effective so the zener current must stay above this value operating under load within its breakdown region at all times. The upper limit of current is of course dependent upon the power rating of the zener device.

Zener Diode Clipping

As well as voltage regulation circuits, zener diodes can also be used to limit the maximum amplitude of a signal. Zener diode clipping and clamping circuits can used to shape or modify an input AC waveform (or any sinusoid) producing a differently shape output waveform depending upon the circuit arrangement.

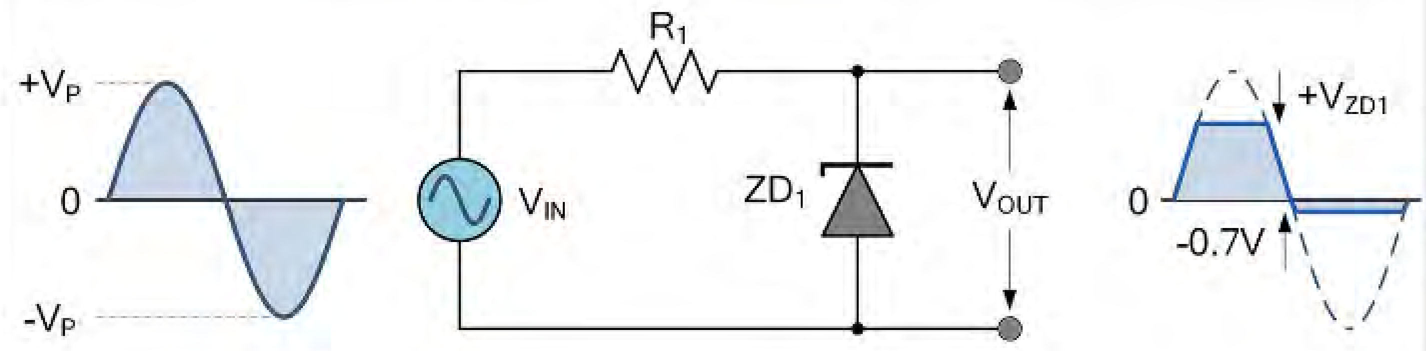
Diode clipper circuits are also called limiters because they limit or clip-off the positive (or negative) part of an input AC signal. As zener clipper circuits limit or cut-off part of the waveform across them, they are mainly used for circuit protection or for use in waveform shaping circuits.

For example, if we wanted to clip a sinusoidal waveform at +7.5 volts, we would use a 7.5V zener diode. If the signal waveform tries to exceed the 7.5V limit, the zener diode will "clip-off" the excess voltage from the input producing a waveform with a flat top keeping the output constant at the required +7.5 volts.

Note that in the forward bias condition a zener diode is still an ordinary semiconductor diode and when the AC waveform goes negative below -0.7V, (for silicon) the zener diode turns "ON" (conducts) like any normal silicon diode would and clips the output at -0.7V as shown. Thus a single zener can produce a half-wave clipping circuit.

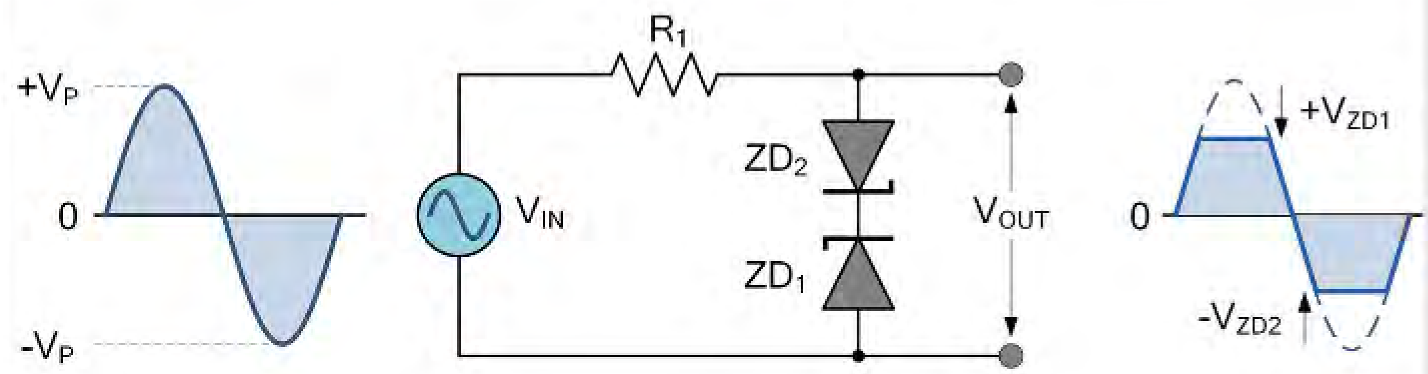
Zener Diodes can also be used for wave-shaping and the clipping or clamping of signal waveforms

HALF-WAVE ZENER DIODE CLIPPING



We can develop this idea further by using the zener diodes reverse-voltage characteristics of two zener's to clip both halves of a signal waveform using series connected back-to-back zener diodes as shown.

FULL-WAVE ZENER DIODE CLIPPING



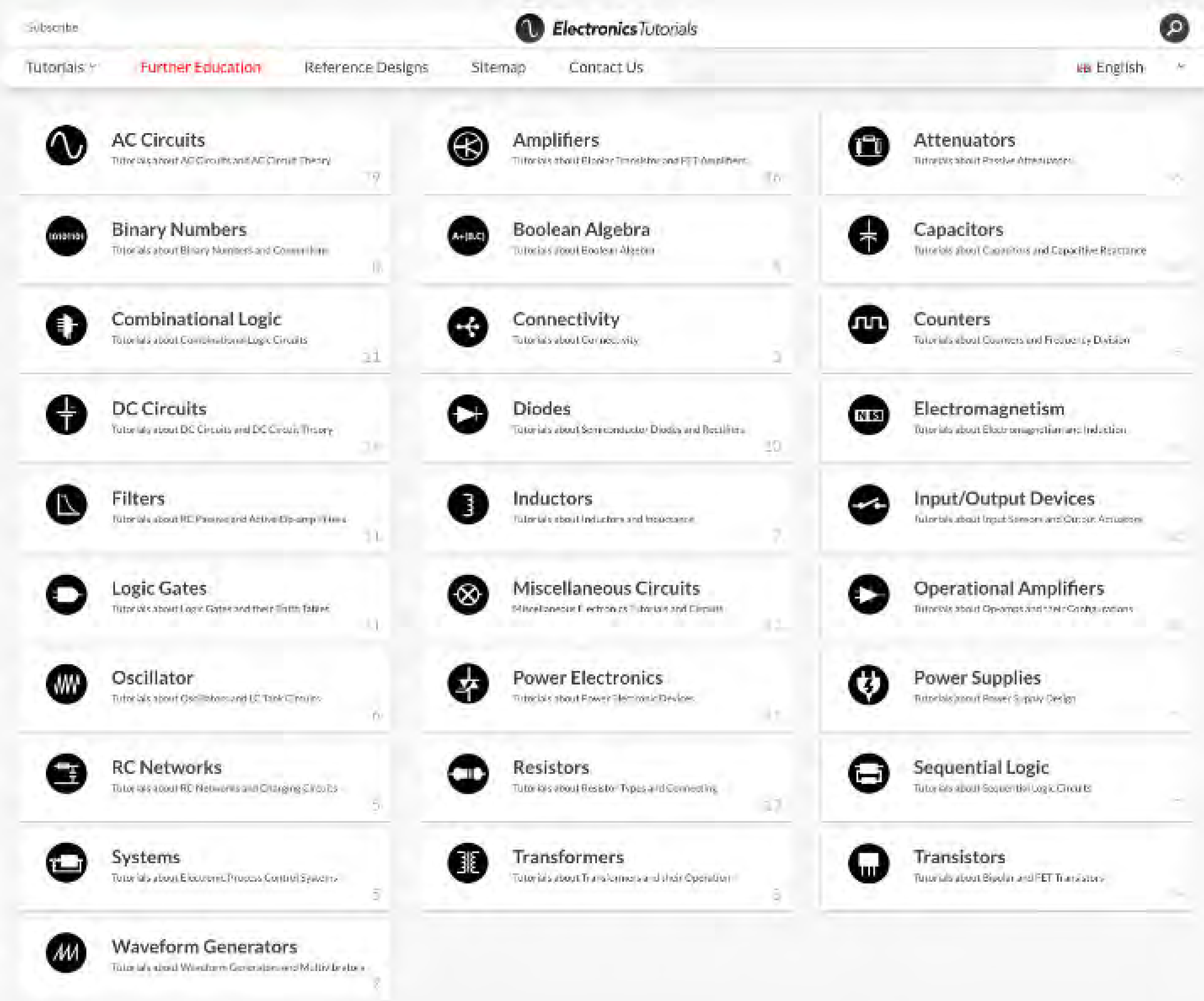
The output waveform from a full-wave zener diode clipping circuit will be clipped at the zener voltage plus the 0.7 volt (silicon) forward volt drop of the other diode. So for example, the positive half-cycle will be clipped at the sum of zener diode, ZD_1 plus the forward-biased 0.7 volts from ZD_2 and vice versa for the negative half-cycle.

Thus zener diode clipping circuits can be used to limit the maximum amplitude of a signal waveform and can be designed to clip only the positive part of the signal waveform, the negative part of the signal waveform, or both polarities of the signal waveform producing an almost square-wave waveform output signal from a sinusoid.

The positive and negative clipping levels can be adjusted independently by using two back-to-back zener diodes with different zener voltage clamping ratings. For example, the positive half-cycle could be at 4.5 volts and the negative half-cycle at 7.2 volts.

Zener diodes are manufactured with a wide range of voltages and can be used to give different voltage references on each half cycle, the same as above. Zener diodes are available with nominal zener breakdown voltages, V_Z ranging from 2.4 to over 100 volts, (for example the 1N4728A to 1N4764A range) with a typical tolerance of 1 or 5%. Note that once conducting in the reverse breakdown region, full current will flow through the zener diode so a suitable current limiting resistor, R_Z must be chosen.

ElectronicsTutorials



ASPENCORE

Aspencore Media GmbH
Rauwagnerstr. 5
85560 Ebersberg / Germany

Editorial Contact
Wayne Storr
wstorr@aspencore.com

For Reader Inquiries please contact:
Wayne Storr
wstorr@aspencore.com

Editorial Design
ARTPOOL Communication Services
www.artpool.de