2.2.2 Current Flow in the P-Type Material

Current flow through the P-type material is illustrated in *Figure 4-13*. Conduction in the P-material is by positive holes instead of negative electrons. A hole moves from the positive terminal of the P-material to the negative terminal. Electrons from the external circuit enter the negative terminal of the material and fill holes in the vicinity of this terminal. At the positive terminal, electrons are removed from the covalent bonds, thus creating new holes. This process continues as the steady stream of holes (hole current) moves toward the negative terminal.

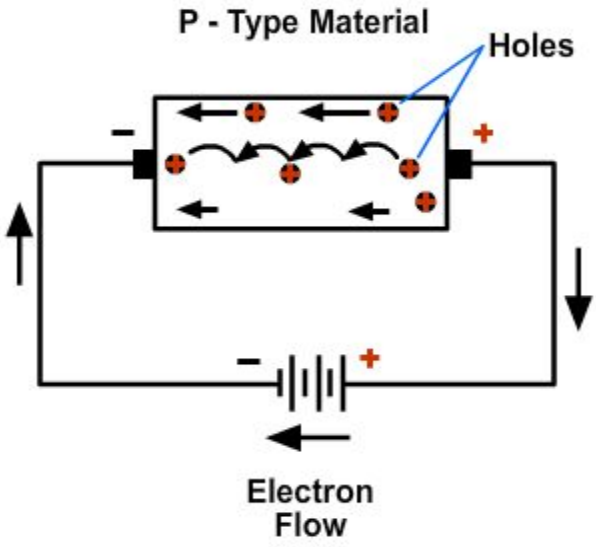


Figure 4-13 – Current flow in the P-type material.

Notice in both N-type and P-type materials, current flow in the external circuit consists of electrons moving out of the negative terminal of the battery and into the positive terminal of the battery. Hole flow, on the other hand, only exists within the material itself.

2.2.3 Junction Barrier

Although the N-type material has an excess of free electrons, it is still electrically neutral. This is because the donor atoms in the N-material were left with positive charges after free electrons became available by covalent bonding (the protons outnumbered the electrons); therefore, for every free electron in the N-material, there is a corresponding positively charge atom to balance it. The end result is that the N-material has an overall charge of zero.

By the same reasoning, the P-type material is also electrically neutral because the excess of holes in this material is exactly balanced by the number of electrons. Keep in mind that the holes and electrons are still free to move in the material because they are only loosely bound to their parent atoms.

It would seem that if we joined the N- and P-materials together by one of the processes mentioned earlier, all the holes and electrons would pair up. On the contrary, this does not happen. Instead the electrons in the N-material diffuse (move or spread out) across the junction into the P-material and fill some of the holes. At the same time, the holes in the P-material diffuse across the junction into the N-material and are filled by N-material electrons. This process, called junction recombination, reduces the number of free electrons and holes in the vicinity of the junction. Because there is a depletion, or lack, of free electrons and holes in this area, it is known as the depletion region.

The loss of an electron from the N-type material created a positive ion in the N-material, while the loss of a hole from the P-material created a negative ion in that material. These ions are fixed in place in the crystal lattice structure and cannot move. Thus, they make up a layer of fixed charges on the two sides of the junction, as shown in *Figure 4-14*. On the N side of the junction, there is a layer of positively charged ions; on the P side of the junction, there is a layer of negatively charged ions. An electrostatic field, represented by a small battery in the figure, is established across the junction between the oppositely charged ions. The diffusion of electrons and holes across the junction will

continue until the magnitude of the electrostatic field is increased to the point where the electrons and holes no longer have enough energy to overcome it, and are repelled by the negative and positive ions, respectively. At this point, equilibrium is established and, for all practical purposes, the movement of carriers across the junction ceases. For this reason, the electrostatic field created by the positive and negative ions in the depletion region is called a barrier.

The action just described occurs almost instantly when the junction is formed. Only the carriers in the immediate vicinity of the junction are affected. The carriers throughout the remainder of the N- and P- material are relatively undisturbed and remain in a balanced condition.

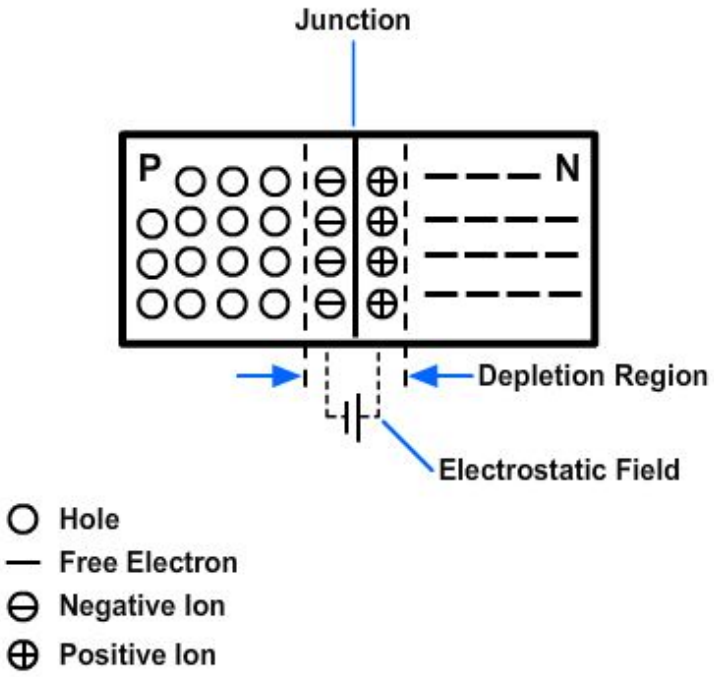


Figure 4-14 — PN junction barrier formation.

2.2.4 Forward Bias

An external voltage applied to a PN junction is called ***bias***. If, for example, a battery is used to supply bias to a PN junction and is connected so that its voltage opposes the junction field, it will reduce the junction barrier and therefore aid current flow through the junction. This type of bias is known as forward bias, and it causes the junction to offer only minimum resistance to the flow of current.

Forward bias is illustrated in *Figure 4-15*. Notice the positive terminal of the bias battery is connected to the P-type material and the negative terminal of the battery is connected to the N-type material. The positive potential repels holes toward the junction, where they neutralize some of the negative ions. At the same time, the negative potential repels electrons toward the junction where they neutralize some of the positive ions.

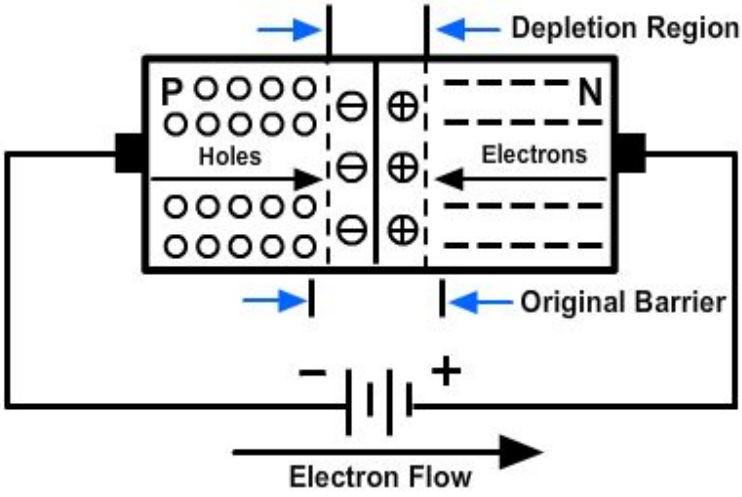


Figure 4-15 – Forward-biased PN junction.

Since ions on both sides of the barrier are being neutralized, the width of the barrier decreases; thus, the effect of the battery voltage in the forward-bias direction is to reduce the barrier potential across the junction and to allow majority carriers to cross the junction. Current flow in the forward-biased PN junction is relatively simple. An electron leaves the negative terminal of the battery and moves to the terminal of the N-type material. It enters the N-material, where it is the majority carrier and moves to the edge of the junction barrier. Because of forward bias, the barrier offers less opposition to the electron and it will pass through the depletion region into the P-type material. The electron loses energy in overcoming the opposition of the junction barrier, and upon entering the P-material, combines with a hole. The hole was produced when an electron was extracted from the P-material by the positive potential of the battery. The created hole moves through the P-material toward the junction where it combines with an electron.

It is important to remember that in the forward-biased condition, conduction is by majority current carriers (holes in the P-type material and electrons in the N-type material). Increasing the battery voltage will increase the number of majority carriers arriving at the junction and will therefore increase the current flow. If the battery voltage is increased to the point where the barrier is greatly reduced, a heavy current will flow and the junction may be damaged from the resulting heat.

2.2.5 Reverse Bias

If the battery mentioned earlier is connected across the junction so that its voltage aids the junction, it will increase the junction barrier and thereby offer a high resistance to the current flow through the junction. This type of bias is known as reverse bias.

To reverse bias a junction diode, the negative battery terminal is connected to the P-type material, and the positive battery terminal to the N-type material as shown in *Figure 4-16*. The negative potential attracts the holes away from the edge of the junction barrier on the P side, while the positive potential attracts the electrons away from the edge of the barrier on the N side. This action increases the barrier width because there are more negative ions on the P side of the junction, and more positive ions on the N side of the junction. Notice in the figure the width of the barrier has increased. This increase in the number of ions prevents current flow across the junction by majority carriers. However, the current flow across the barrier is not quite zero because of the minority carriers crossing the junction. As you recall, when the crystal is subjected to an external source of energy (light, heat, and so forth), electron-hole pairs are generated.

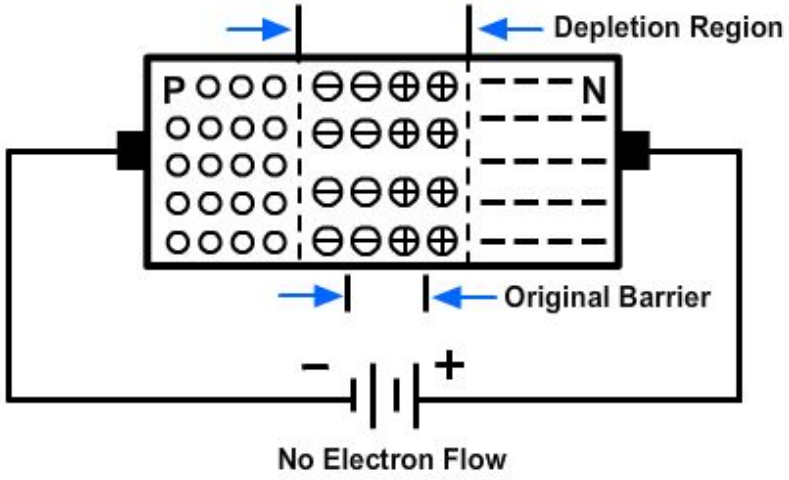


Figure 4-16 – Reverse-biased PN junction.

The electron-hole pairs produce minority current carriers. There are minority current carriers in both regions: holes in the N-material and electrons in the P-material. With reverse bias, the electrons in the P-type material are repelled toward the junction by the negative terminal of the battery. As the electron moves across the junction, it will neutralize a positive ion in the N-type material. Similarly, the holes in the N-type material will be repelled by the positive terminal of the battery toward the junction. As the hole crosses the junction, it will neutralize a negative ion in the P-type material. This movement of minority carriers is called minority current flow, because the holes and electrons involved come from the electron-hole pairs that are generated in the crystal lattice structure, and not from the addition of impurity atoms.

When a PN junction is reverse biased, there will be no current flow because of majority carriers but a very small amount of current because of minority carriers crossing the junction. However, at normal operating temperatures, this small current may be neglected.

The most important point to remember about the PN-junction diode is its ability to offer very little resistance to current flow in the forward-bias direction but maximum resistance to current flow when reverse biased. A good way of illustrating this point is by plotting a graph of the applied voltage versus the measured current. *Figure 4-17* shows a plot of this voltage-current relationship (characteristic curve) for a typical PN junction diode.

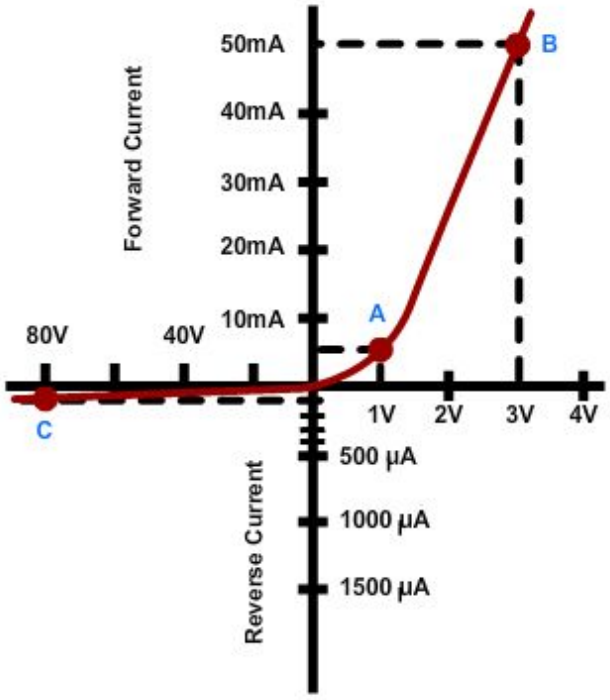


Figure 4-17 — PN junction diode characteristic curve.

To determine the resistance from the curve in *Figure 4-17*, you can use Ohm's law:

$$R = \frac{E}{I}$$

For example, at point A the forward-bias voltage is 1 volt, and the forward-bias current is 5 milliamperes. This represents 200 ohms of resistance (1 volt/5mA = 200 ohms). However, at point B, the voltage is 3 volts and the current is 50 milliamperes. This results in 60 ohms of resistance for the diode. Notice that when the forward-bias voltage was tripled (1 volt to 3 volts), the current increased 10 times (5mA to 50mA). At the same time, the forward-bias voltage increased, the resistance decreased from 200 ohms to 60 ohms. In other words, when forward bias increases, the junction barrier gets smaller and its resistance to current flow decreases.

On the other hand, the diode conducts very little when reverse biased. Notice at point C, the reverse bias voltage is 80 volts and the current is only 100 microamperes. This results in 800 k ohms of resistance, which is considerably larger than the resistance of the junction with forward bias. Because of these unusual features, the PN junction diode is often used to convert alternating current into direct current (rectification).

3.0.0 DIODE APPLICATION

Until now, only one application for the diode-rectification has been mentioned, but there are many more applications. Variations in doping agents, semiconductor materials, and manufacturing techniques have made it possible to produce diodes that can be used in many different applications. Examples of these type of diodes are signal diodes, rectifying diodes, Zener diodes (voltage protection diodes for power supplies), varactors (amplifying and switching diodes), and many more. Only applications for two of the most commonly used diodes, the signal diode and rectifier diode, will be presented in this section. The other diodes will be explained later in the chapter.

3.1.0 Half-Wave Rectifier

One of the most important uses of a diode is rectification. The normal PN junction diode is well suited for this purpose as it conducts very heavily when forward biased (low-resistance direction) and only slightly when reverse biased (high-resistance direction). If we place this diode in series with a source of ac power, the diode will be forward and reverse biased every cycle. Since in this situation current flows more easily in one direction than the other, rectification is accomplished. The simplest rectifier circuit is a half-wave rectifier, which consists of a diode, ac power source, and a load resistor (*Figure 4-18, View A* and *View B*).

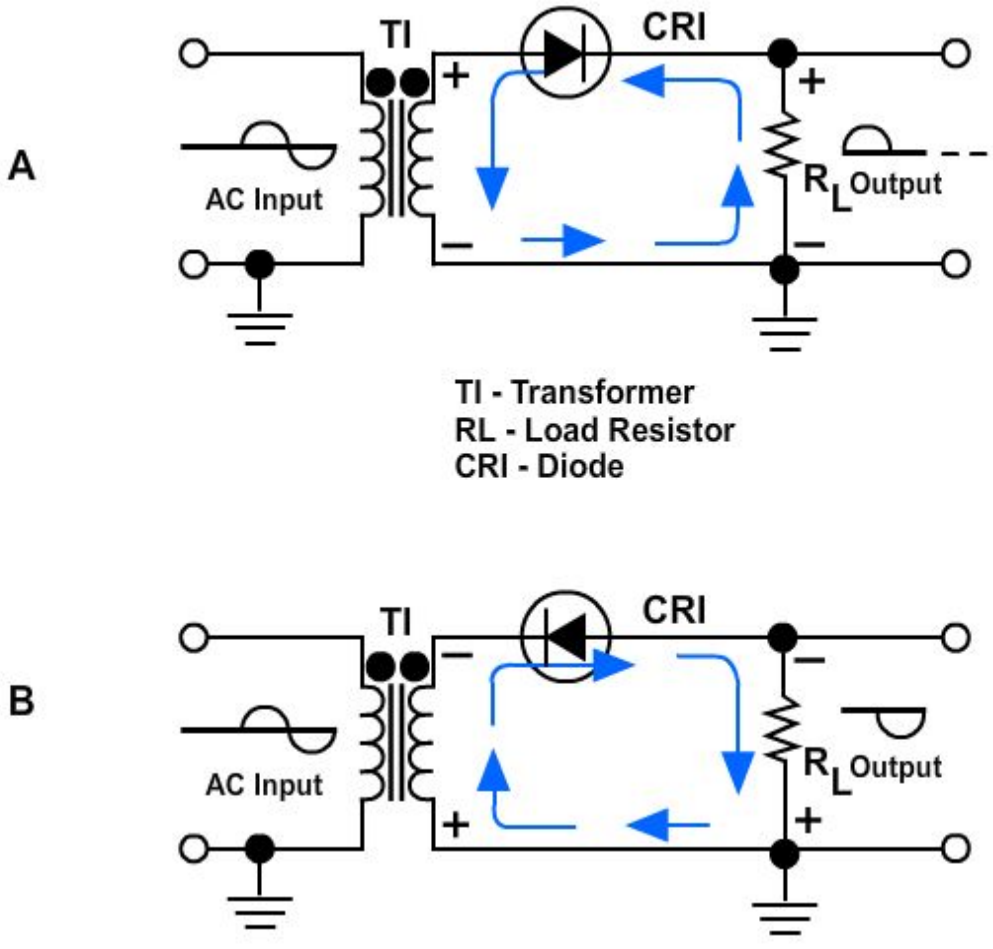


Figure 4-18 – Simple half-wave rectifier.

The transformer (T1) in the figure provides the ac input to the circuit; the diode (CR1) provides the rectification; and the load resistor (R_L) serves two purposes: (1) it limits the

amount of current flow in the circuit to a safe level and (2) it develops the output signal because of the current flow through it.

Before describing how this circuit operates, the definition of the word “load” as it applies to power supplies must be understood. Load is defined as any device that draws current. A device that draws little current is considered a light load, whereas a device that draws a large amount of current is a heavy load. Remember that when you speak of “load,” you are speaking about the device that draws current from the power source. This device may be a simple resistor or one or more complicated electronic circuits.

During the positive half-cycle of the input signal (solid line) in *Figure 4-18, View A*, the top of the transformer is positive with respect to ground. The dots on the transformer indicate points of the same polarity. With this condition, the diode is forward biased, the depletion region is narrow, the resistance of the diode is low, and current flows through the circuit in the direction of the solid lines. When this current flows through the load resistor, it develops a negative to positive voltage drop across it, which appears as a positive voltage at the output terminal.

When the ac input goes in a negative direction, the top of the transformer becomes negative and the diode becomes reversed biased (*Figure 4-18, View A*). With reverse bias applied to the diode, the depletion region increases, the resistance of the diode is high, and minimum current flows through the diode. For all practical purposes, there is no output developed across the load resistor during the negative alternation of the input signal, as indicated by the broken lines in the figure. Although only one cycle of input is shown, it should be realized that the action described above continually repeats itself as long as there is an input; therefore, because only the positive half-cycles appear at the output this circuit converted the ac input into a positive pulsating dc voltage. The frequency of the output voltage is equal to the frequency of the applied ac signal since there is one pulse out for each cycle of the ac input. For example, if the input frequency is 60 hertz (60 cycles per second), the output frequency is 60 pulses per second (pps).

However, if the diode is reversed as shown in *Figure 4-18, View B*, a negative output voltage is obtained. This is because the current is flowing from the top of R_L toward the bottom, making the output at the top of R_L negative with respect to the bottom or ground. Because current flows in this circuit only during half of the input cycle, it is called a half-wave rectifier.

The semiconductor diode shown in the figure can be replaced by a metallic rectifier and still achieve the same results. The metallic rectifier, sometimes referred to as a dry-disc rectifier, is a metal-to-semiconductor, large-area contact device. Its construction is distinctive: a semiconductor is sandwiched between two metal plates, or electrodes, as shown in *Figure 4-19*. Note in the figure that a barrier, with a resistance many times greater than that of the semiconductor material, is constructed on one of the metal electrodes. The contact having the barrier is a rectifying contact; the other contact is nonrectifying. Metallic rectifiers act the same as the diodes previously discussed in that they permit current to flow more readily in one direction than the other. However, the metallic rectifier is fairly large compared to the crystal diode, as can be seen in *Figure 4-20*. The reasons for this are: metallic rectifier units are stacked (to prevent inverse voltage breakdown), have large area plates (to handle high currents), and usually have cooling fins (to prevent overheating).

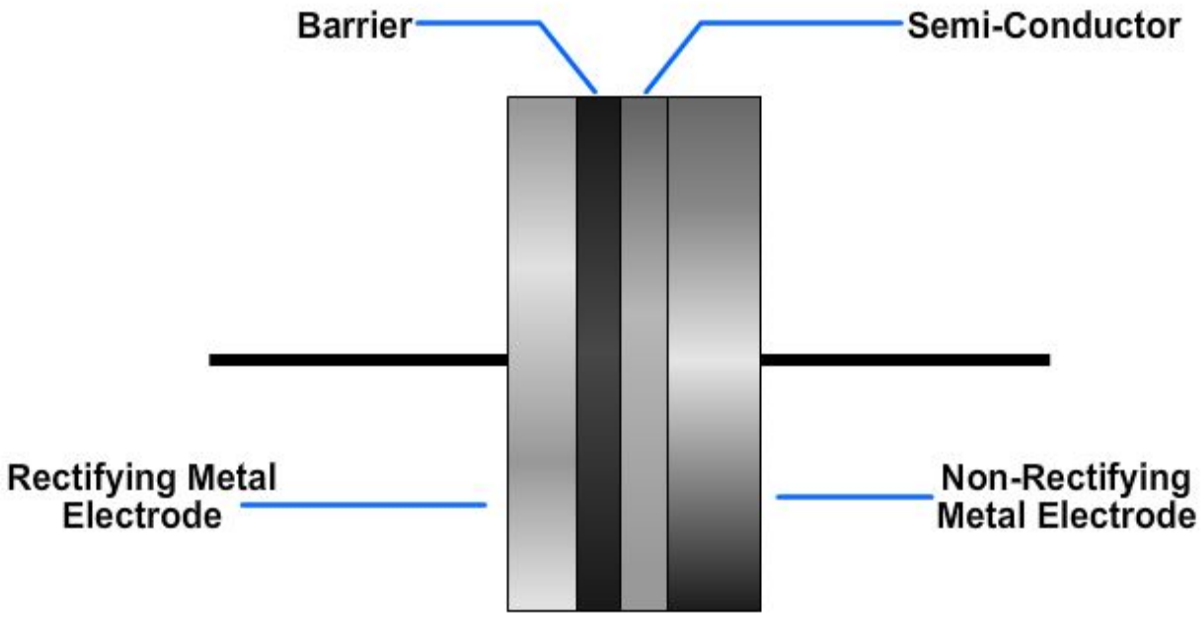


Figure 4-19 — Metallic rectifier.

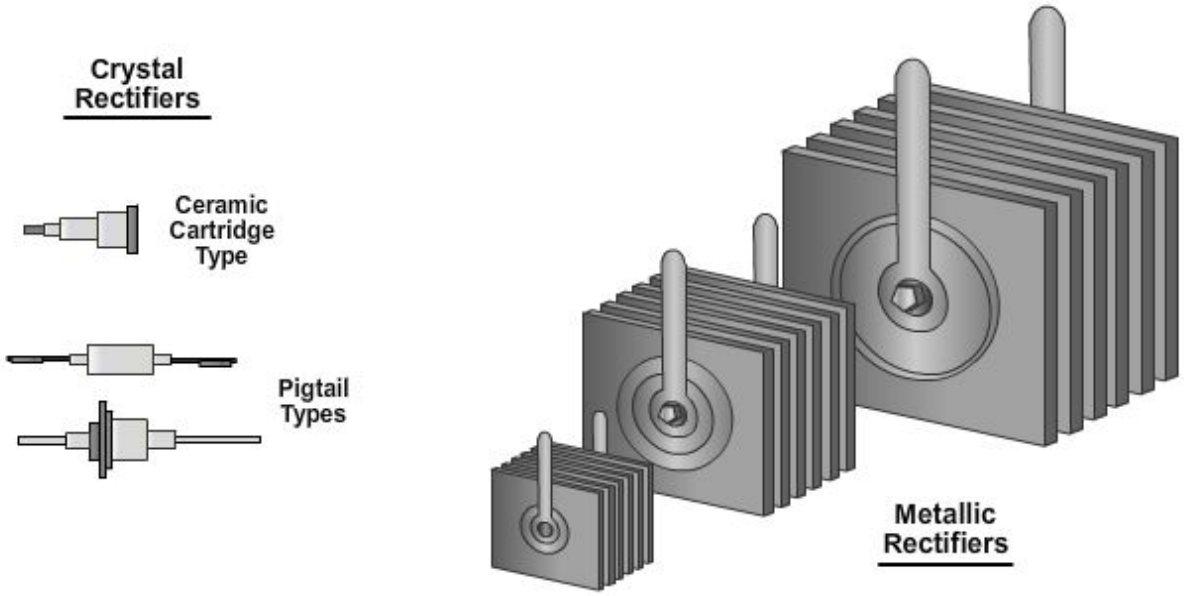


Figure 4-20 — Different types of crystal and metallic rectifiers.

There are many known metal-semiconductor combinations that can be used for contact rectification. Copper oxide and selenium devices are by far the most popular. Copper oxide and selenium are frequently used over other types of metallic rectifiers because they have a large forward current per unit contact area, low forward voltage drop, good stability, and a lower aging rate. In practical application, the selenium rectifier is used where a relatively large amount of power is required. On the other hand, copper-oxide rectifiers are generally used in small-current applications, such as ac meter movements or for delivering direct current to circuits requiring not more than 10 amperes.

Since metallic rectifiers are affected by temperature, atmospheric conditions, and aging (in the case of copper oxide and selenium), they are being replaced by the improved silicon crystal rectifier. The silicon rectifier replaces the bulky selenium rectifier as to current and voltage rating, and can operate at higher ambient (surrounding) temperatures.

3.2.0 Diode Switch

In addition to their use as simple rectifiers, diodes are also used in circuits that mix signals together (mixers), detect the presence of a signal (detector), and act as a switch to open or close a circuit. Diodes used in these applications are commonly referred to as signal diodes. The simplest application of a signal diode is the basic diode switch shown in *Figure 4-21*.

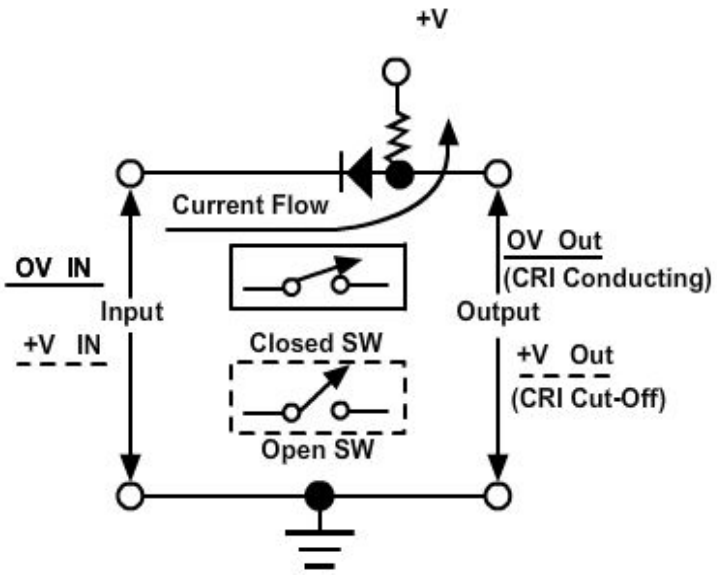


Figure 4-21 — Basic diode switch.

When the input to this circuit is a zero potential, the diode is forward biased because of the zero potential on the cathode and the positive voltage on the anode. In this condition, the diode conducts and acts as a straight piece of wire because of its very low forward resistance. In effect, the input is directly coupled to the output, resulting in zero volts across the output terminals; therefore, the diode acts as a closed switch when its anode is positive with respect to its cathode.

If you apply a positive input voltage (equal to or greater than the positive voltage supplied to the anode) to the diode's cathode, the diode will be reverse biased. In this situation, the diode is cut off and acts as an open switch between the input and output terminals. Consequently, with no current flow in the circuit, the positive voltage on the diode's anode will be felt at the output terminal; therefore, the diode acts as an open switch when it is reverse biased.

3.3.0 Diode Characteristics

Semiconductor diodes have properties that enable them to perform many different electronic functions. To do their jobs, engineers and technicians must be supplied with data on these different types of diodes. The information presented for this purpose is called diode characteristics. These characteristics are supplied by manufacturers either in their manuals or on specification sheets (data sheets). Because of the scores of manufacturers and numerous diode types, it is not practical to put before you a specification sheet and call it typical. Aside from the difference between manufacturers, a single manufacturer may even supply specification sheets that differ both in format

and content. Despite these differences, certain performance and design information is normally required. This information will be discussed in the next few paragraphs.

A standard specification sheet usually has a brief description of the diode. Included in this description is the type of diode, the major area of application, and any special features. Of particular interest is the specific application for which the diode is suited. The manufacturer also provides a drawing of the diode that gives dimension, weight, and, if appropriate, any identification marks. In addition to the above data, the following information is also provided: a static operating table (giving spot values of parameters under fixed conditions); sometimes a characteristic curve similar to the one in *Figure 4-17* (showing how parameters vary over the full operating range); and diode ratings (which are the limiting values of operating conditions outside which could cause diode damage).

Manufacturers specify these various diode operating parameters and characteristics with letter symbols in accordance with fixed definitions. The following is a list, by letter symbol, of the major electrical characteristics for the rectifier and signal diodes:

Rectifier Diodes

- **DC Blocking Voltage [V_R]** – the maximum reverse dc voltage that will not cause breakdown
- **Average Forward Voltage Drop [$V_{F(AV)}$]** – the average forward voltage drop across the rectifier given at a specified forward current and temperature
- **Average Rectifier Forward Current [$I_{F(AV)}$]** – the average rectified forward current at a specified temperature, usually at 60 H_Z with a resistive load
- **Average Reverse Current [$I_{R(AV)}$]** – the average reverse current at a specified temperature, usually at 60 H_Z
- **Peak Surge Current [I_{SURGE}]** – the peak current specified for a given number of cycles or portion of a cycle

Signal Diodes

- **Peak Reverse Voltage [PRV] –** the maximum reverse voltage that can be applied before reaching the breakdown point (PRV also applies to the rectifier diode.)
- **Reverse Current [I_R]** – the small value of direct current that flows when a semiconductor diode has reverse bias
- **Maximum Forward Voltage Drop at Indicated Forward Current [$V_F@I_F$]** – the maximum forward voltage drop across the diode at the indicated forward current.
- **Reverse Recovery Time [t_{rr}]** – the maximum time taken for the forward-bias diode to recover its reverse bias

The ratings of a diode, as stated earlier, are the limiting values of operating conditions, which, if exceeded, could cause damage to a diode by either voltage breakdown or overheating. The PN junction diodes are generally rated for: (1) maximum average forward current, (2) peak recurrent forward current, (3) maximum surge current, and (4) peak reverse voltage.

Maximum average forward current is usually given at a special temperature, usually 25° C (77° F) and refers to the maximum amount of average current that can be permitted to flow in the forward direction. If this rating is exceeded, structure breakdown can occur.

Peak recurrent forward current is the maximum peak current that can be permitted to flow in the forward direction in the form of recurring pulses.

Maximum surge current is the maximum current permitted to flow in the forward direction in the form of nonrecurring pulses. Current should not equal this value for more than a few milliseconds.

Peak reverse voltage (PRV) is one of the most important ratings. PRV indicates the maximum reverse-bias voltage that may be applied to a diode without causing junction breakdown.

All of the above ratings are subject to change with temperature variations. If, for example, the operating temperature is above that stated for the ratings, the ratings must be decreased.

3.4.0 Diode Identification

There are many types of diodes, varying in size from the size of a pinhead (used in subminiature circuitry) to large, 250-ampere diodes (used in high-power circuits). Because there are so many different types of diodes, some system of identification is needed to distinguish one diode from another. This is accomplished with the semiconductor identification system (*Table 4-1*).

Table 4-1 – Semiconductor Identification Codes.

XNYYY

XN	YYY
COMPONENT	IDENTIFICATION NUMBER

X- NUMBER OF SEMICONDUCTOR JUNCTIONS

N - A SEMICONDUCTOR

YYY - IDENTIFICATION NUMBER (ORDER OR REGISTRATION NUMBER)

ALSO INCLUDES SUFFIC LETTER (IF APPLICABLE) TO INDICATE:

1. MATCHING DEVICES
2. REVERSE POLARITY
3. MODIFICATION

EXAMPLE - 1N345A (AN IMPROVED VERSION OF THE SEMICONDUCTOR DIODE TYPE 345)

This system is not only used for diodes, but also transistors and many other special semiconductor devices as well. As illustrated in *Table 4-1*, the system uses numbers and letters to identify different types of semiconductor devices. The first number in the system indicates the number of junctions in the semiconductor device and is a number,

one less than the number of active elements. Thus, 1 designates a diode; 2 designates a transistor (which may be considered as made up of two diodes); and 3 designates a tetrode (a four-element transistor). The letter N following the first number indicates a semiconductor. The 2- or 3-digit number following the letter N is a serialized identification number. If needed, this number may contain a suffix letter after the last digit. For example, the suffix letter M may be used to describe matching pairs of separate semiconductor devices, or the letter R may be used to indicate reverse polarity. Other letters are used to indicate modified versions of the device which can be substituted for the basic numbered unit. For example, a semiconductor diode designated as type 1N345A signifies a two-element diode (1) of semiconductor material (N) that is an improved version (A) of type 345.

When working with these different types of diodes, it is also necessary to distinguish one end of the diode from the other (anode from cathode). For this reason, manufacturers generally code the cathode end of the diode with a k, +, cath, a color dot or band, or by an unusual shape (raised edge or taper) as shown in *Figure 4-22*. In some cases, standard color code bands are placed on the cathode end of the diode. This serves two purposes: (1) it identifies the cathode end of the diode and (2) it also serves to identify the diode by number.

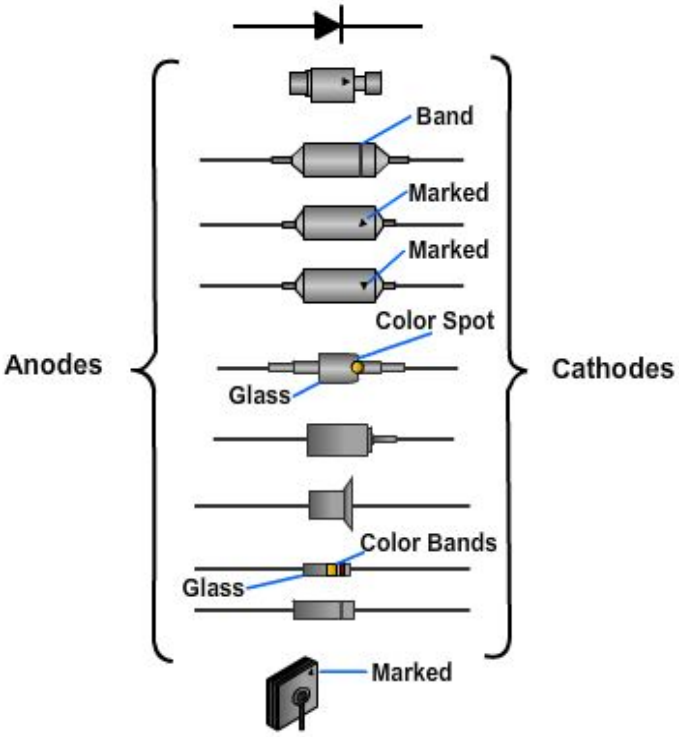
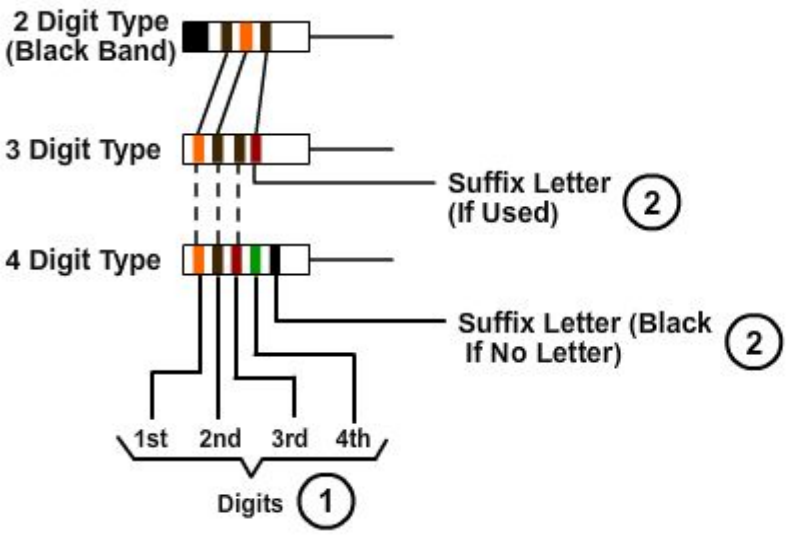


Figure 4-22 — Semiconductor diode markings.

The standard diode color code system is shown in *Figure 4-23*. Take, for example, a diode with brown, orange, and white bands at one terminal and figure out its identification number. With brown being a 1, orange a 3, and white a 9, the device is identified as a type 139 semiconductor diode, or specifically 1N139.

Keep in mind, whether the diode is a small crystal type or a large power rectifier type, both are still represented schematically, as explained earlier, by the schematic symbol shown in *Figure 4-9*.



Color	1 Digit	2 Diode Suffix Letter
Black	0	-
Brown	1	A
Red	2	B
Orange	3	C
Yellow	4	D
Green	5	E
Blue	6	F
Violet	7	G
Gray	8	H
White	9	J
Silver	-	-
Gold	-	-
None	-	-

Figure 4-23 — Semiconductor diode color code system.

4.0.0 FILTER CIRCUITS

While the output of a rectifier is a pulsating dc, most electronic circuits require a substantially pure dc for proper operation. This type of output is provided by single or multi-section filter circuits placed between the output of the rectifier and the load. There are four basic types of filter circuits:

- Simple capacitor filter
- LC choke-input filter
- LC capacitor-input filter (pi-type)
- RC capacitor-input filter (pi-type)

Filtering is accomplished by the use of capacitors, inductors, and/or resistors in various combinations. Inductors are used as series impedances to oppose the flow of alternating (pulsating dc) current. Capacitors are used as shunt elements to bypass the alternating components of the signal around the load (to ground). Resistors are used in place of inductors in low-current applications.

At this time, a brief review of the properties of a capacitor is necessary. First, a capacitor opposes any change in voltage. The opposition to a change in current is called capacitive reactance (X_C) and is measured in ohms. The capacitive reactance is determined by the frequency (f) of the applied voltage and the capacitance (c) of the capacitor.

$$X_C = \frac{1}{2\pi fC} \text{ or } \frac{.159}{fC}$$

From the formula, you can see that if frequency or capacitance is increased, the X_C decreases. Since filter capacitors are placed in parallel with the load, a low X_C will provide better filtering than a high X_C. For this to be accomplished, a better shunting effect of the ac around the load is provided, as shown in *Figure 4-24*.

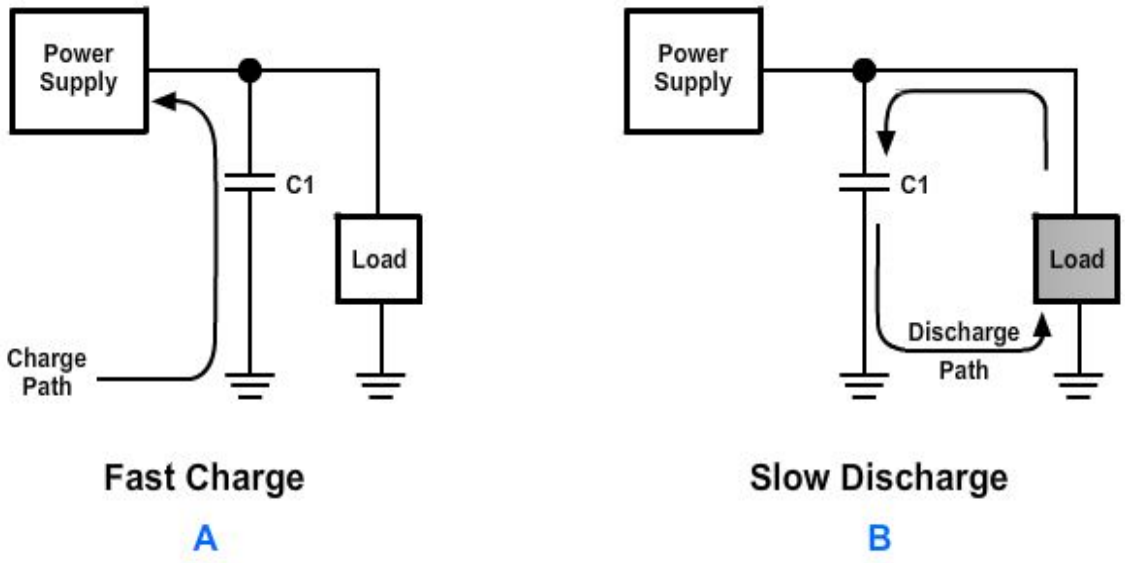


Figure 4-24 — Capacitor filter.

To obtain a steady dc output, the capacitor must charge almost instantaneously to the value of applied voltage. Once charged, the capacitor must retain the charge as long as possible. The capacitor must have a short charge time constant (*Figure 4-24, View A*). This can be accomplished by keeping the internal resistance of the power supply as small as possible (fast charge time) and the resistance of the load as large as possible for a slow discharge time, as illustrated in *Figure 4-24, View B*.

From your earlier studies in basic electricity, you may remember that one time constant is defined as the time it takes a capacitor to charge to 63.2 percent of the applied voltage or to discharge to 36.8 percent of its total charge. This action can be expressed by the following equation:

$T = RC$

Where: *R* represents the resistance of the charge or discharge path

And: *C* represents the capacitance of the capacitor

You should also recall that a capacitor is considered fully charged after five *RC* time constants (*Figure 4-25*). You can see that a steady dc output voltage is obtained when the capacitor charges rapidly and discharges as slowly as possible.

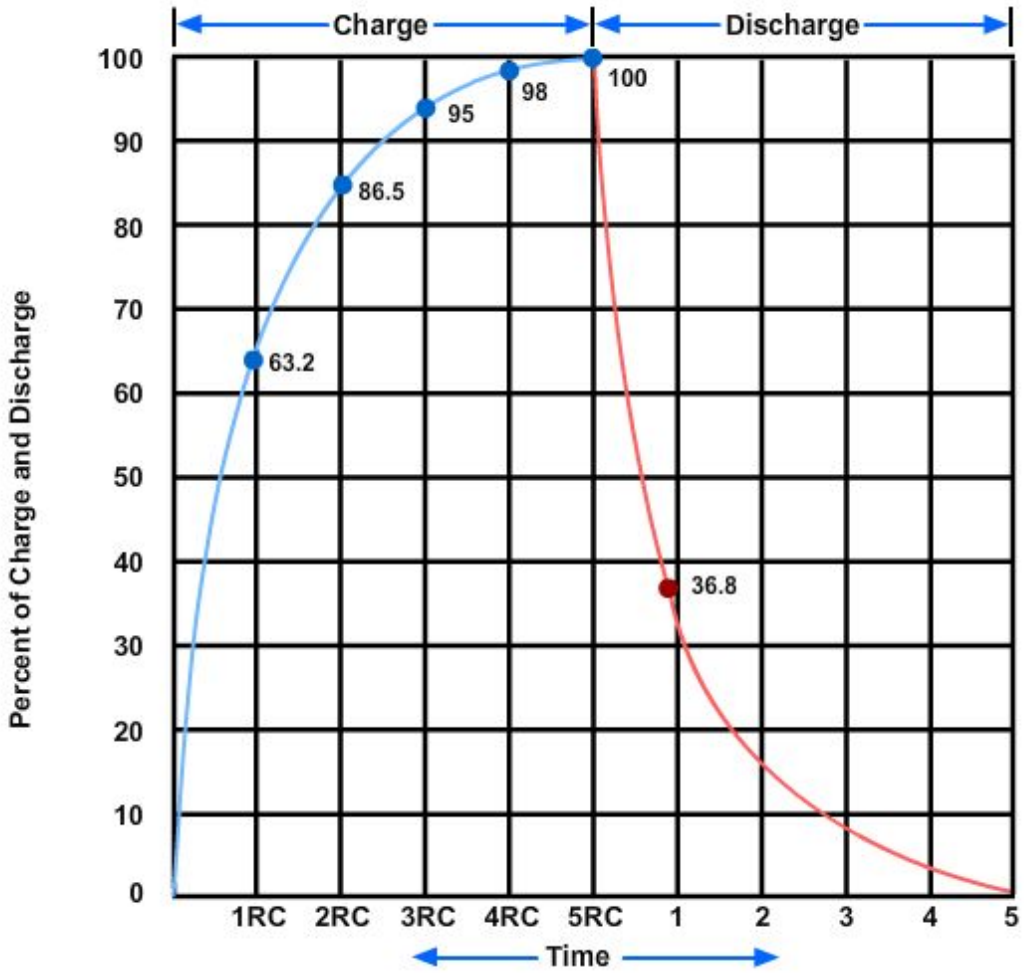


Figure 4-25 – *RC* time constant.

In filter circuits, the capacitor is the common element to both the charge and the discharge paths; therefore, to obtain the longest possible discharge time, you want the capacitor to be as large as possible. Another way to look at it is: The capacitor acts as a short circuit around the load (as far as the ac component is concerned, and since

$$X_C = \frac{1}{2\pi fC}$$

the larger the value of the capacitor (C), the smaller the opposition (X_C) or reactance to ac.

Now let us look at inductors and their application in filter circuits. Remember, an inductor opposes any change in current. In case you have forgotten, a change in current through an inductor produces a changing electromagnetic field. The changing field, in turn, cuts the windings of the wire in the inductor and thereby produces a counter electromotive force (CEMF). It is the CEMF that opposes the change in circuit current. Opposition to a change in current at a given frequency is called inductive reactance (X_L) and is measured in ohms. The inductive reactance (X_L) of an inductor is determined by the applied frequency and the inductance of the inductor. Mathematically, that is:

$$X_L = 2\pi fL$$

If frequency or inductance is increased, the X_L increases. Since inductors are placed in series with the load (*Figure 4-26*), the larger the X_L, the larger the ac voltage developed across the load.

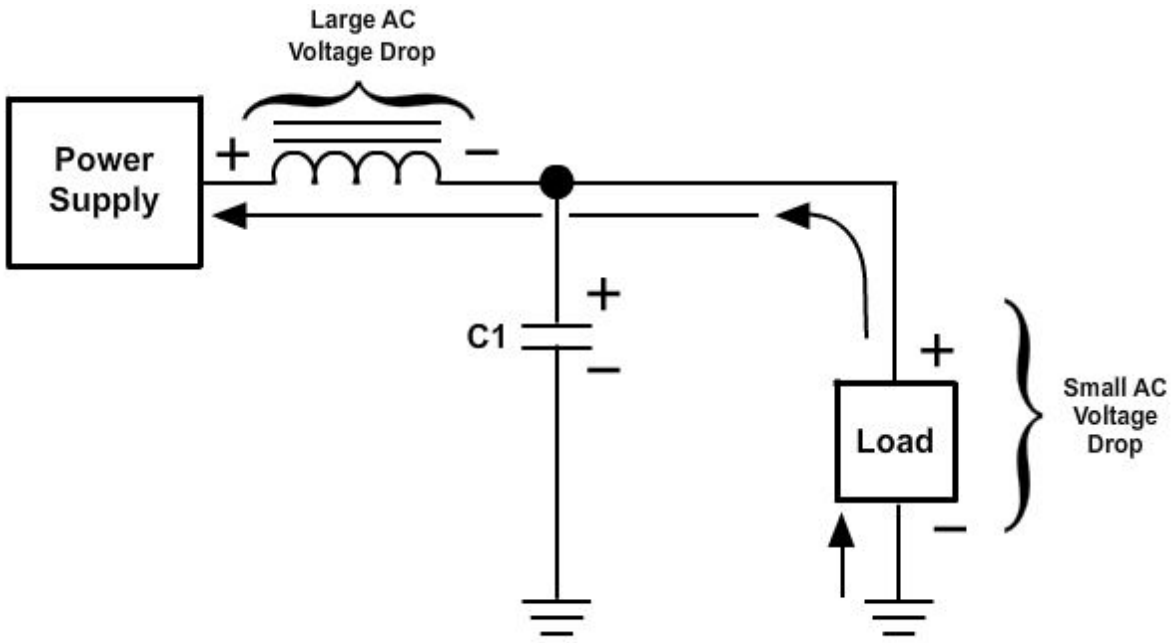


Figure 4-26 — Voltage drops in an inductive filter.

Now refer to *Figure 4-27*. When the current starts to flow through the coil, an expanding magnetic field builds up around the inductor. The magnetic field around the coil develops the CEMF that opposes the change in current. When the rectifier current decreases, the magnetic field collapses and again cuts the turns (windings) of wire, thus inducing current into the coil (*Figure 4-28*). This additional current merges with the rectifier current and attempts to keep it at its original level.

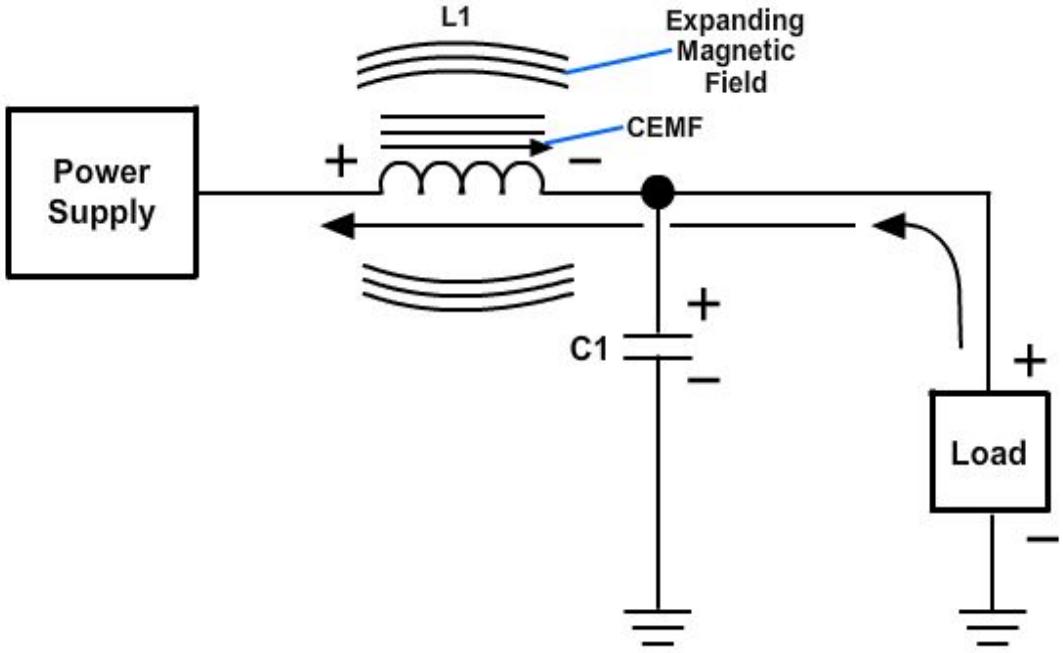


Figure 4-27 — Inductive filter (expanding field).

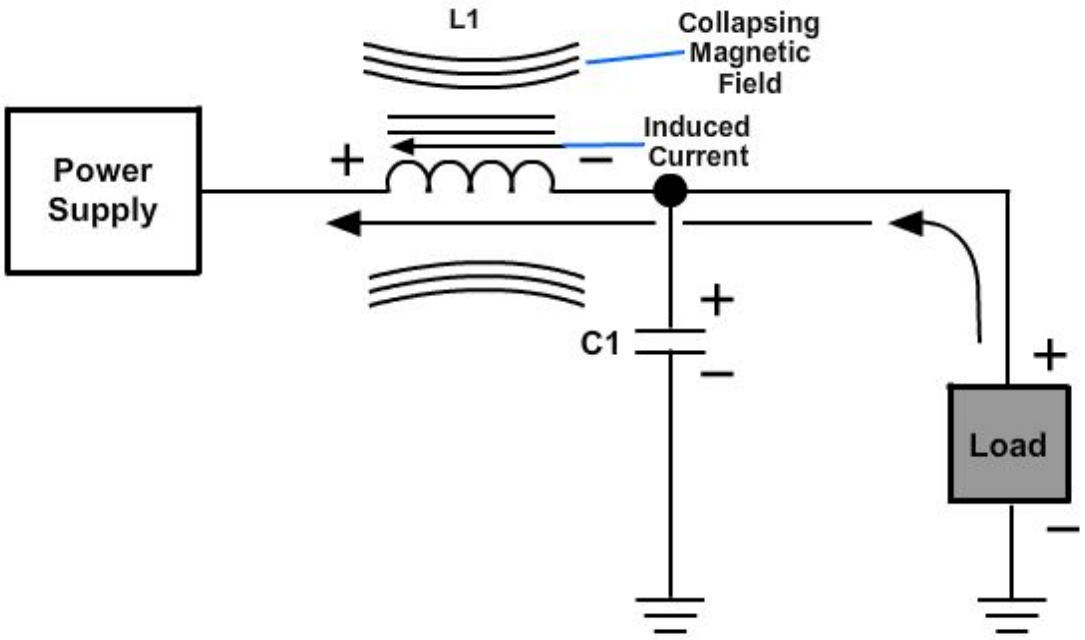


Figure 4-28 — Inductive filter (collapsing field).

Now that you have read how the components in a filter circuit react to current flow from the rectifier, the different types of filter circuits in use today will be discussed.

5.0.0 TYPES OF POWER SUPPLY FILTERS

5.1.0 Capacitor Filter

The simple capacitor filter is the most basic type of power supply filter. The application of the simple capacitor filter is very limited. It is sometimes used on extremely high-voltage, low-current power supplies for cathode-ray and similar electron tubes, which require very little load current from the supply. The capacitor filter is also used where the power-supply ripple frequency is not critical; this frequency can be relatively high. The capacitor (C1) is a simple filter connected across the output of the rectifier in parallel with the load (*Figure 4-29*).

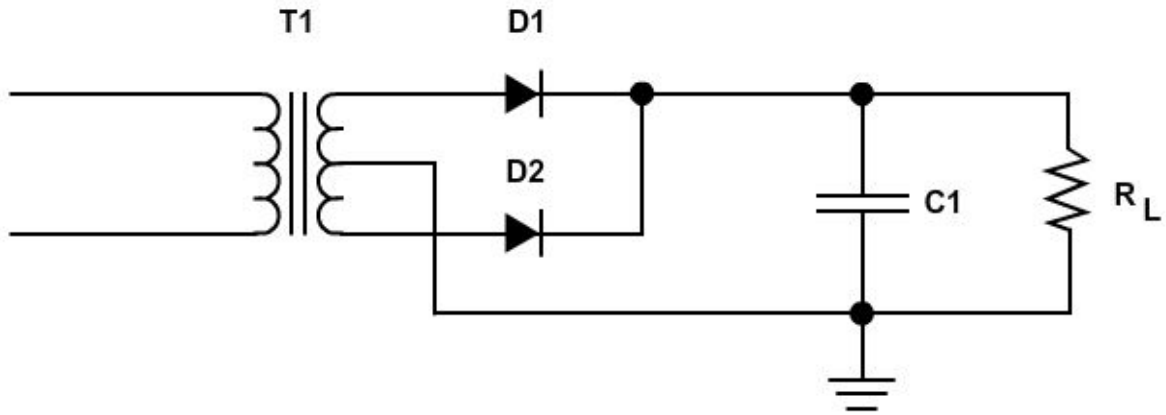


Figure 4-29 – Full-wave rectifier with a capacitor filter.

When this filter is used, the RC charge time of the filter capacitor (C1) must be short and the RC discharge time must be long to eliminate ripple action. In other words, the

capacitor must charge up fast, preferably with no discharge at all. Better filtering also results when the input frequency is high; therefore, the full-wave rectifier output is easier to filter than that of the half-wave rectifier because of its higher frequency.

For you to have a better understanding of the effect that filtering has on E_{avg}, a comparison of a rectifier circuit with a filter and one without a filter is illustrated in *Figure 4-30, Views A* and *B*. The output waveforms in *Figure 4-30* represent the unfiltered and filtered outputs of the half-wave rectifier circuit. Current pulses flow through the load resistance (R_L) each time a diode conducts. The dashed line indicates the average value of output voltage. For the half-wave rectifier, E_{avg} is less than half (or approximately 0.318) of the peak output voltage. This value is still much less than that of the applied voltage. With no capacitor connected across the output of the rectifier circuit, the waveform in *View A* has a large pulsating component (ripple) compared with the average or dc component. When a capacitor is connected across the output, *View B*, the average value of output voltage, (E_{avg}) is increased due to the filtering action of capacitor C1.

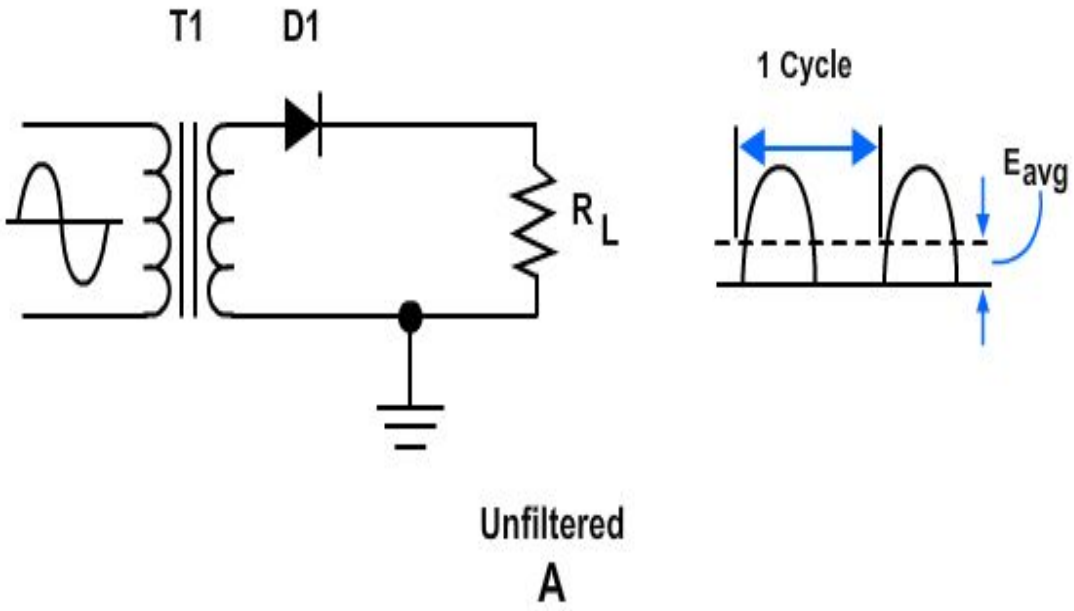


Figure 4-30 — Half-wave rectifier with and without filtering.

The value of the capacitor is fairly large (several microfarads), thus it presents a relatively low reactance to the pulsating current and it stores a substantial charge.

The rate of charge for the capacitor is limited only by the resistance of the conducting diode, which is relatively low; therefore, the RC charge time of the circuit is relatively short. As a result, when the pulsating voltage is first applied to the circuit, the capacitor charges rapidly and almost reaches the peak value of the rectified voltage within the first few cycles. The capacitor attempts to charge to the peak value of the rectified voltage anytime a diode is conducting, and tends to retain its charge when the rectifier output falls to zero. The capacitor cannot discharge immediately. The capacitor slowly discharges through the load resistance (R_L) during the time the rectifier is nonconducting.

The rate of discharge of the capacitor is determined by the value of capacitance and the value of the load resistance. If the capacitance and load-resistance values are large, the RC discharge time for the circuit is relatively long.

A comparison of the waveforms shown in *Figure 4-30, View A* and *View B*, illustrates that the addition of C1 to the circuit results in an increase in the average of the output voltage (E_{avg}) and a reduction in the amplitude of the ripple component (E_r), which is normally present across the load resistance.

Now, consider a complete cycle of operation using a half-wave rectifier, a capacitive filter (C1), and a load resistor (R_L). As shown in *Figure 4-31, View A*, the capacitive filter (C1) is assumed to be large enough to ensure a small reactance to the pulsating rectified current. The resistance of R_L is assumed to be much greater than the reactance of C1 at the input frequency. When the circuit is energized, the diode conducts on the positive half cycle and current flows through the circuit, allowing C1 to charge. C1 will charge to approximately the peak value of the input voltage. The charge is less than the peak value because of the voltage drop across the diode (D1). In *Figure 4-31, View A*, the charge on C1 is indicated by the heavy solid line on the waveform. The diode cannot conduct on the negative half cycle because the anode of D1 is negative with respect to the cathode (*Figure 4-31, View B*). During this interval, C1 discharges through the load resistor (R_L). The discharge of C1 produces the downward slope, as indicated by the solid line on the waveform in *Figure 4-31, View B*. In contrast to the abrupt fall of the applied ac voltage from peak value to zero, the voltage across C1, and thus across R_L, during the discharge period gradually decreases until the time of the next half cycle of rectifier operation. Keep in mind that for good filtering, the filter capacitor should charge up as fast as possible and discharge as little as possible.

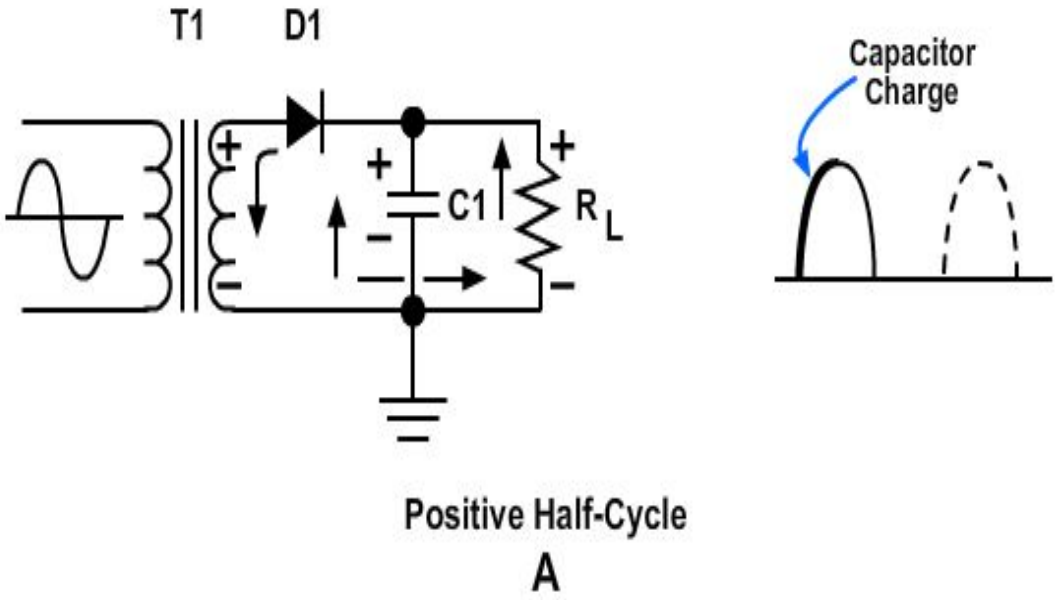


Figure 4-31 — Capacitor filter circuit (positive and negative half cycles).

Since practical values of C1 and R_L ensure a more or less gradual decrease of the discharge voltage, a substantial charge remains on the capacitor at the time of the next half cycle of operation. As a result, no current can flow through the diode until the rising ac input voltage at the anode of the diode exceeds the voltage on the charge remaining on C1. The charge on C1 is the cathode potential of the diode. When the potential on the anode exceeds the potential on the cathode (the charge on C1), the diode again conducts and C1 begins to charge to approximately the peak value of the applied voltage.

After the capacitor has charged to its peak value, the diode will cut off and the capacitor will start to discharge. Since the fall of the ac input voltage on the anode is considerably more rapid than the decrease on the capacitor voltage, the cathode quickly becomes more positive than the anode, and the diode ceases to conduct.

Operation of the simple capacitor filter using a full-wave rectifier is basically the same as that discussed for the half-wave rectifier. Referring to *Figure 4-32*, you should notice that because one of the diodes is always conducting on either alternation, the filter capacitor charges and discharges during each half cycle. Note that each diode conducts only for that portion of time when the peak secondary voltage is greater than the charge across the capacitor.

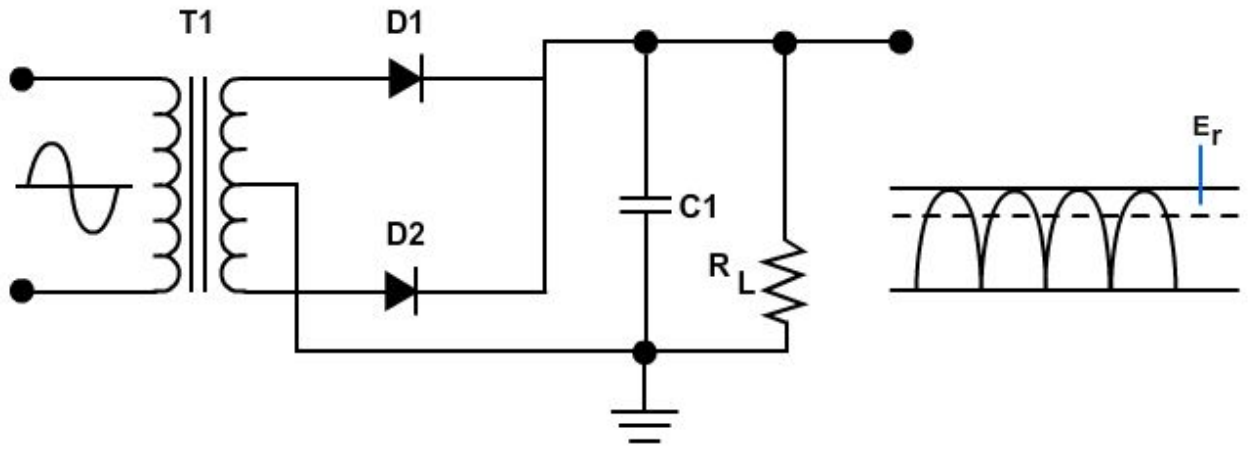


Figure 4-32 – Full-wave rectifier (with capacitor filter).

Another thing to keep in mind is that the ripple component (E) of the output voltage is an ac voltage and the average output voltage (E_{avg}) is the dc component of the output. Since the filter capacitor offers a relatively low impedance to ac, the majority of the ac component flows through the filter capacitor. The ac component is therefore bypassed (shunted) around the load resistance, and the entire dc component (or E_{avg}) flows through the load resistance. This statement can be clarified by using the formula for X_C in a half-wave and full-wave rectifier. First, you must establish some values for the circuit.

HALFWAVE RECTIFIER

Frequency at Rectifier Output: 60 H_Z

Value of Filter Capacitor: 30µF

Load Resistance: 10kΩ

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{.159}{fC}$$

$$X_C = \frac{.159}{60 \times .000030}$$

$$X_C = \frac{.159}{.0018}$$

$$X_C = 88.3\Omega$$

Frequency at Rectifier Output: 120 H_Z

Value of Filter Capacitor: 30µF

Load Resistance: 10kΩ

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{.159}{fC}$$

$$X_C = \frac{.159}{120 \times .000030}$$

$$X_C = \frac{.159}{.0036}$$

$$X_C = 44.16\Omega$$

As you can see from the calculations, by doubling the frequency of the rectifier, you reduce the impedance of the capacitor by one-half. This allows the ac component to pass through the capacitor more easily. As a result, a full-wave rectifier output is much easier to filter than that of a half-wave rectifier. Remember, the smaller the X_C of the filter capacitor with respect to the load resistance, the better the filtering action. Since $X_C = \frac{1}{2\pi fC}$, the largest possible capacitor will provide the best filtering. Remember also that the load resistance is an important consideration. If load resistance is made small, the load current increases, and the average value of output voltage (E_{avg}) decreases. The RC discharge time constant is a direct function of the value of the load resistance; therefore, the rate of capacitor voltage discharge is a direct function of the current through the load. The greater the load current, the more rapid the discharge of the capacitor, and the lower the average value of output voltage. For this reason, the simple capacitive filter is seldom used with rectifier circuits that must supply a relatively large load current. Using the simple capacitive filter in conjunction with a full-wave or bridge rectifier provides improved filtering because the increased ripple frequency decreases the capacitive reactance of the filter capacitor.

5.2.0 Choke-Input Filter

The LC choke-input filter is used primarily in power supplies where voltage regulation is important and where the output current is relatively high and subject to varying load conditions. This filter is used in high power applications, such as those found in radars and communication transmitters.

Notice in *Figure 4-33* that this filter consists of an input inductor (L1), or filter choke, and an output filter capacitor (C1). Inductor L1 is placed at the input to the filter and is in series with the output of the rectifier circuit. Since the action of an inductor is to oppose any change in current flow, the inductor tends to keep a constant current flowing to the load throughout the complete cycle of the applied voltage. As a result, the output

voltage never reaches the peak value of the applied voltage. Instead, the output voltage approximates the average value of the rectified input to the filter (*Figure 4-33*). The reactance of the inductor (X_L) reduces the amplitude of ripple voltage without reducing the dc output voltage by an appreciable amount. The dc resistance of the inductor is just a few ohms.

The shunt capacitor (C1) charges and discharges at the ripple frequency rate, but the amplitude of the ripple voltage (E_r) is relatively small because the inductor (L1) tends to keep a constant current flowing from the rectifier circuit to the load. In addition, the reactance of the shunt capacitor (X_C) presents a low impedance to the ripple component existing at the output of the filter, and thus shunts the ripple component around the load. The capacitor attempts to hold the output voltage relatively constant at the average value of the voltage.

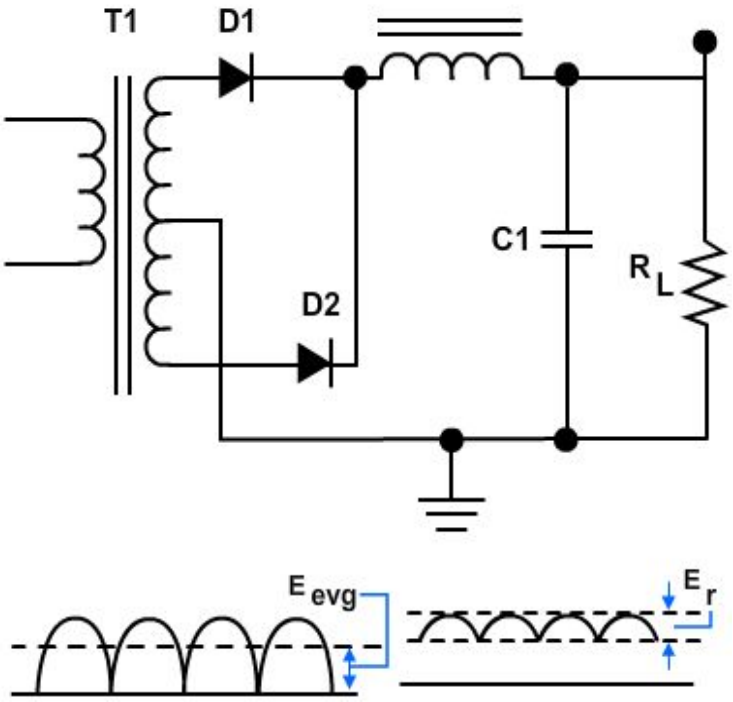


Figure 4-33 — LC choke-input filter.

The value of the filter capacitor (C1) must be relatively large to present a low opposition (X_C) to the pulsating current and to store a substantial charge. The rate of the charge for the capacitor is limited by the low impedance of the ac source (the transformer), by the small resistance of the diode, and by the counter ***electromotive force*** (***CEMF)*** developed by the coil; therefore, the RC charge time constant is short compared to its discharge time. This comparison in RC charge and discharge paths is illustrated in *Figure 4-34, View A* and *View B*. Consequently, when the pulsating voltage is first applied to the LC choke-input filter, the inductor (L1) produces a CEMF, which opposes the constantly increasing input voltage. The net result is to effectively prevent the rapid charging of the filter capacitor (C1). Thus, instead of reaching the peak value of the input voltage, C1 only charges to the average value of the input voltage. After the input voltage reaches its peak and decreases sufficiently, the capacitor (C1) attempts to discharge through the load resistance (R_L). C1 will only partially discharge, as indicated in *Figure 4-34, View B* because of its relatively long discharge time constant. The larger the value of the filter capacitor, the better the filtering action. However, because of physical size, there is a practical limitation to the maximum value of the capacitor.

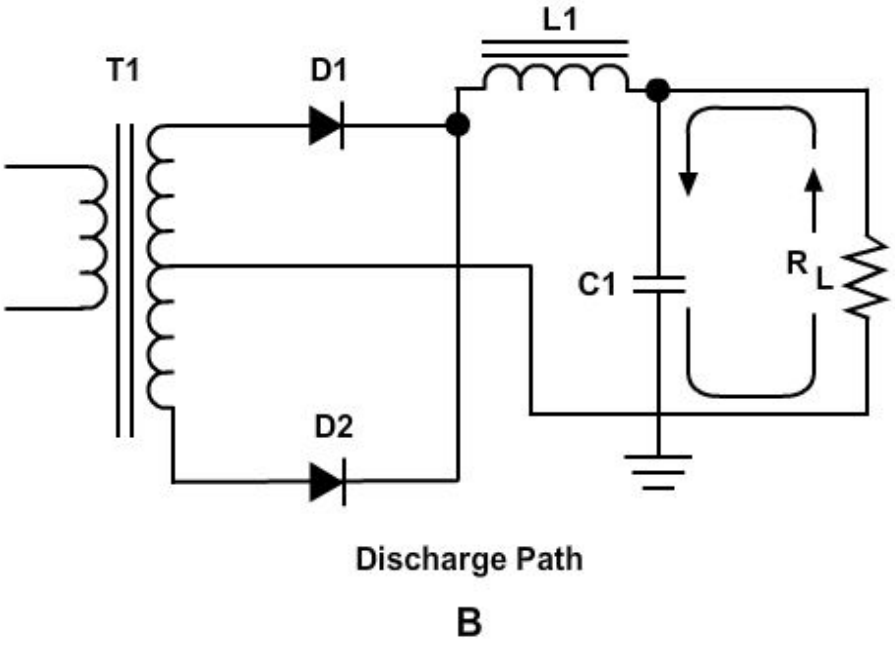


Figure 4-34 — LC choke-input filter (charge and discharge paths).

The inductor, also referred to as the filter choke or coil, serves to maintain the current flow to the filter output (R_L) at a nearly constant level during the charge and discharge periods of the filter capacitor. The inductor (L1) and the capacitor (C1) form a voltage divider for the ac component (ripple) of the applied input voltage. This is shown in *Figure 4-35, Views A* and *B*. As far as the ripple component is concerned, the inductor offers a high impedance (Z) and the capacitor offers a low impedance (*Figure 4-35, View B*). As a result, the ripple component (E_r) appearing across the load resistance is greatly attenuated (reduced). The inductance of the filter choke opposes changes in the value of the current flowing through it; therefore, the average value of the voltage produced across the capacitor contains a much smaller value of ripple component (E_r) than the value of ripple produced across the choke.

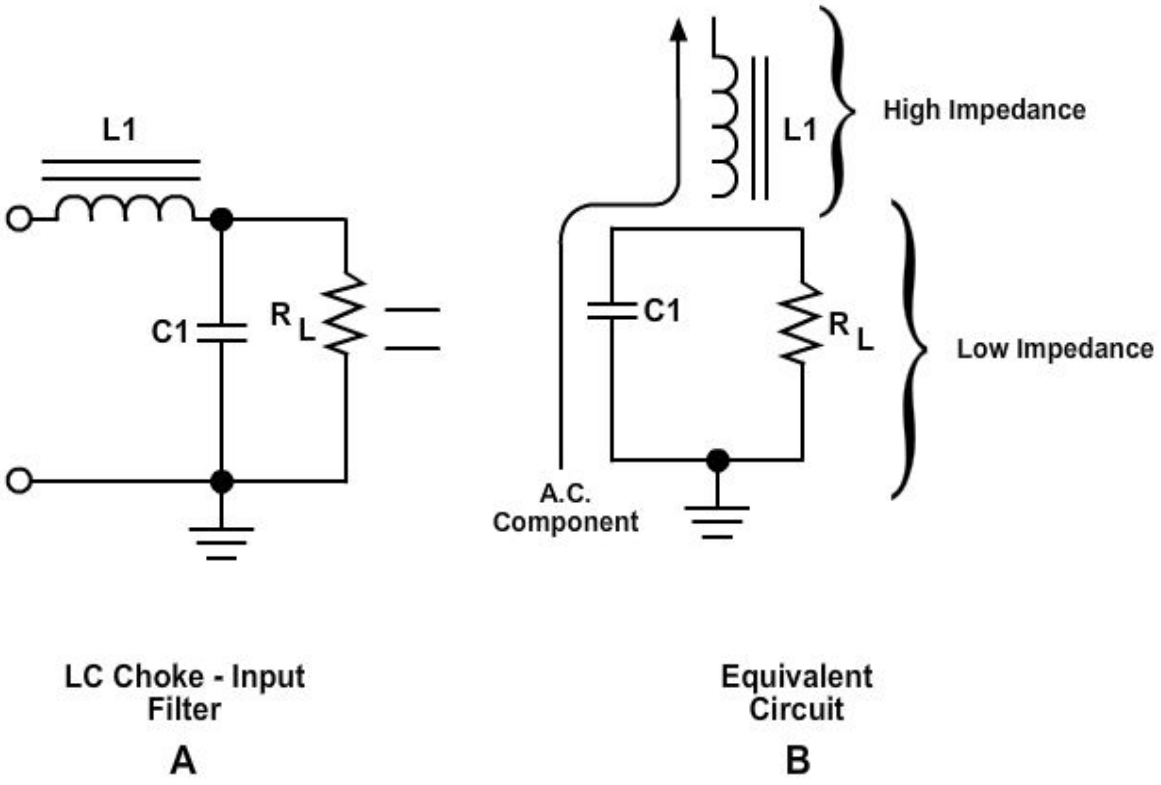


Figure 4-35 – LC choke-input filter.

Figure 4-36 illustrates a complete operation for a full-wave rectifier circuit used to supply the input voltage to the filter. The rectifier voltage is developed across the capacitor (C1). The ripple voltage at the output of the filter is the alternating component of the input voltage reduced in amplitude by the filter section. Each time the anode of a diode goes positive with respect to the cathode, the diode conducts and C1 charges. Conduction occurs twice during each cycle for a full-wave rectifier. For a 60-hertz supply, this produces a 120-hertz ripple voltage. Although the diodes alternate (one conducts while the other is nonconducting), the filter input voltage is not steady. As the anode voltage of the conducting diode increases (on the positive half of the cycle), capacitor C1 charges – the charge being limited by the impedance of the secondary transformer winding, the diode's forward (cathode-to-anode) resistance, and the counter electromotive force developed by the choke. During the nonconducting interval, when the anode voltage drops below the capacitor charge voltage, C1 discharges through the load resistor (R_L). The components in the discharge path have a long time constant; thus, C1 discharges more slowly than it charges.

The choke (L1) is usually a large value, from 1 to 20 henries, and offers a large inductive reactance to the 120-hertz ripple component produced by the rectifier; therefore, the effect that L1 has on the charging of the capacitor (C1) must be considered. Since L1 is connected in series with the parallel branch consisting of C1 and R_L, a division of the ripple (ac) voltage and the output (dc) voltage occurs. The

greater the impedance of the choke, the less the ripple voltage that appears across C1 and the output. The dc output voltage is fixed mainly by the dc resistance of the choke.

Now that you have read how the LC choke-input filter functions, it will be discussed with actual component values applied. For simplicity, the input frequency at the primary of the transformer will be 117 volts 60 hertz. Both half-wave and full-wave rectifier circuits will be used to provide the input to the filter.

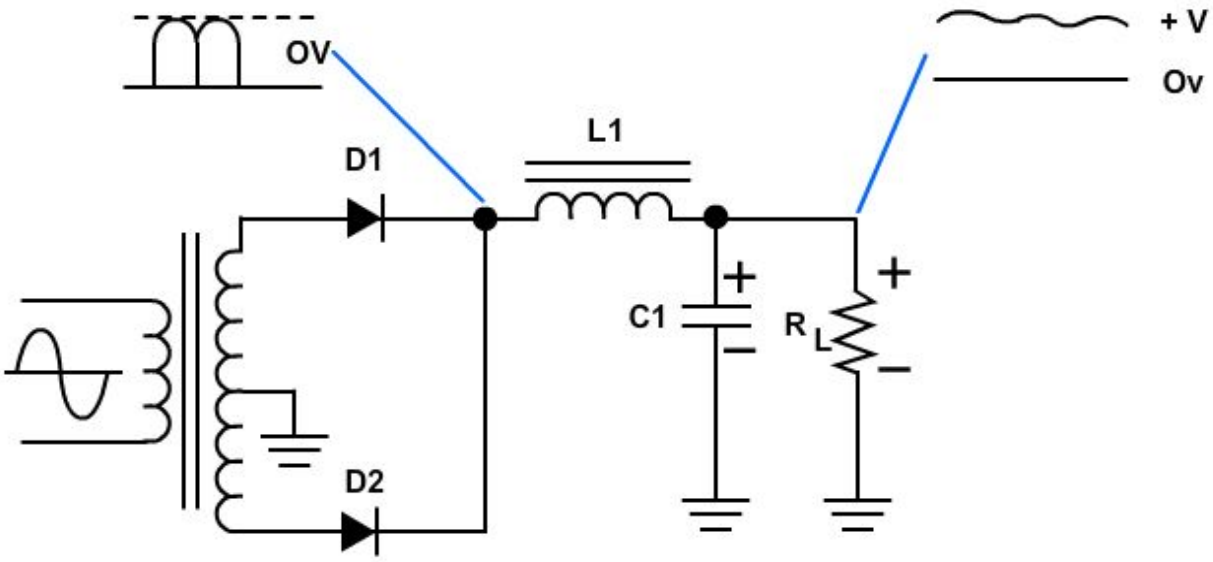


Figure 4-36 — Filtering action of the LC choke-input filter.

Starting with the half-wave configuration shown in *Figure 4-37*, the basic parameters are, with 117 volts ac rms applied to the T1 primary, 165 volts ac peak is available at the secondary [(117 V) x (1.414) = 165 V]. You should recall that the ripple frequency of this half-wave rectifier is 60 hertz; therefore, the capacitive reactance of C1 is:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(3.14)(60)(10)(10^{-6})}$$

$$X_C = \frac{(1)(10^6)}{3768}$$

$$X_C = 265\Omega$$

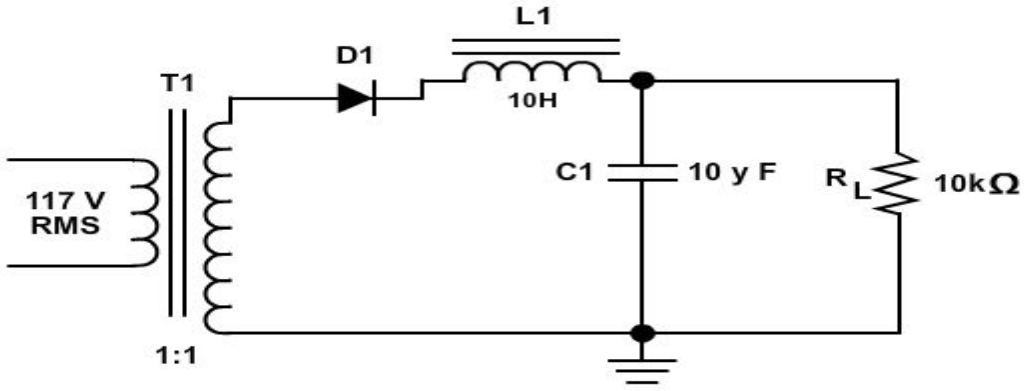


Figure 4-37 – Half-wave rectifier with an LC choke-input filter.

This means that the capacitor (C1) offers 265 ohms of opposition to the ripple current. Note, however, that the capacitor offers an infinite impedance to direct current. The inductive reactance of L1 is:

$$X_L = 2\pi fL$$

$$X_L = (2)(3.14)(60)(10)$$

$$X_L = 3.8\ kilohms$$

The above calculation shows that L1 offers a relatively high opposition (3.8 kilohms) to the ripple in comparison to the opposition offered by C1 (265 ohms). Thus, more ripple voltage will be dropped across L1 than across C1. In addition, the impedance of C1 (265 ohms) is relatively low with respect to the resistance of the load (10 kilohms); therefore, more ripple current flows through C1 than the load. In other words, C1 shunts most of the ac component around the load.

Now you can go a step further and redraw the filter circuit so that you can see the voltage divider action. Refer to *Figure 4-38, View A*. Remember, the 165 volts peak 60 hertz provided by the rectifier consists of both an ac and a dc component. This first discussion will be about the ac component. From *Figure 4-38,* you can see that the capacitor (C1) offers the least opposition (265 ohms) to the ac component; therefore, the greater amount of ac will flow through C1. The red line in *Figure 4-38, View B* indicates the ac current flow through the capacitor. Thus the capacitor bypasses, or shunts, most of the ac around the load.

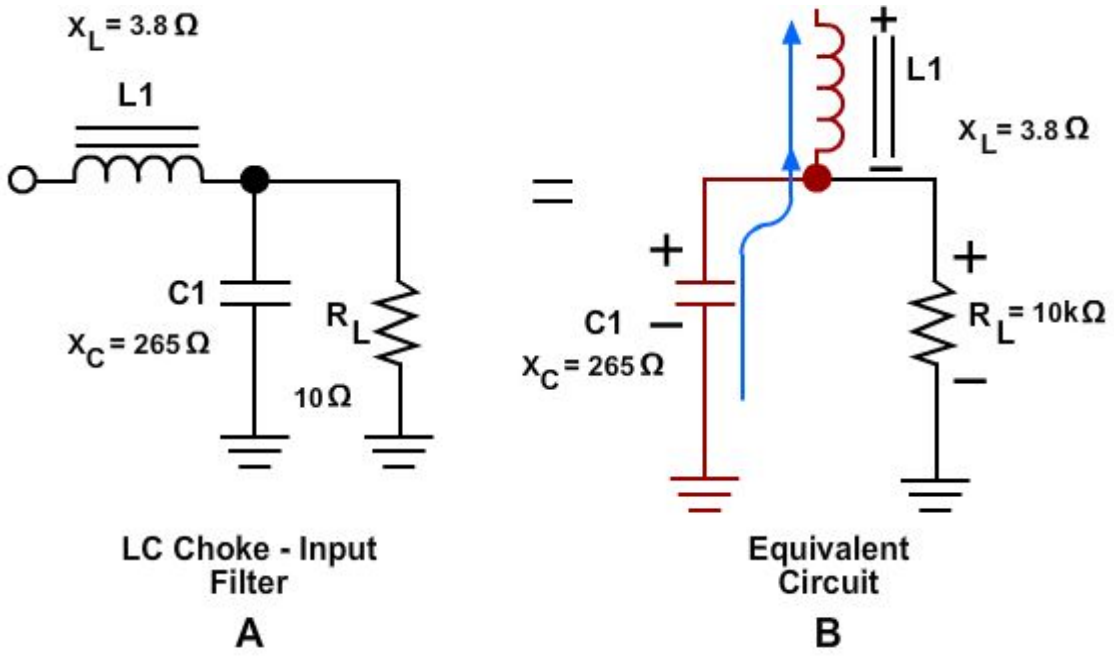


Figure 4-38 — AC component in an LC choke-input filter.

By combining the X_C of C1 and the resistance of R_L into an equivalent circuit as shown in *Figure 4-38, View B*, you will have an equivalent impedance of 265 ohms.

As a formula: $$RT = \frac{(R1)(R2)}{R1 + R2}$$

You now have a voltage divider as illustrated in *Figure 4-39*. You should see that because of the impedance ratios, a large amount of ripple voltage is dropped across L1, and a substantially smaller amount is dropped across C1 and R_L. You can further increase the ripple voltage across L1 by increasing the inductance ($X_L = 2\pi fL$).

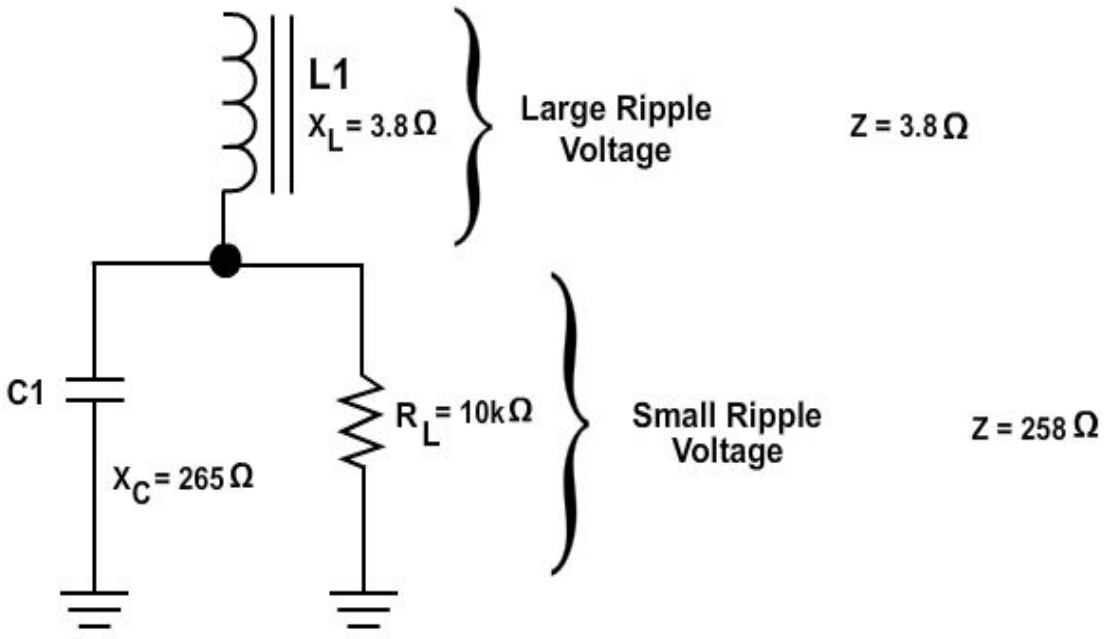


Figure 4-39 — Equivalent circuit of an LC choke-input filter.

Now let us discuss the dc component of the applied voltage. Remember, a capacitor offers an infinite (∞) impedance to the flow of direct current. The dc component,

therefore, must flow through R_L and L1. As far as the dc is concerned, the capacitor does not exist. The coil and the load are therefore in series with each other. The dc resistance of a filter choke is very low (50 ohms average). Consequently, most of the dc component is developed across the load and a very small amount of the dc voltage is dropped across the coil, as shown in *Figure 4-40*.

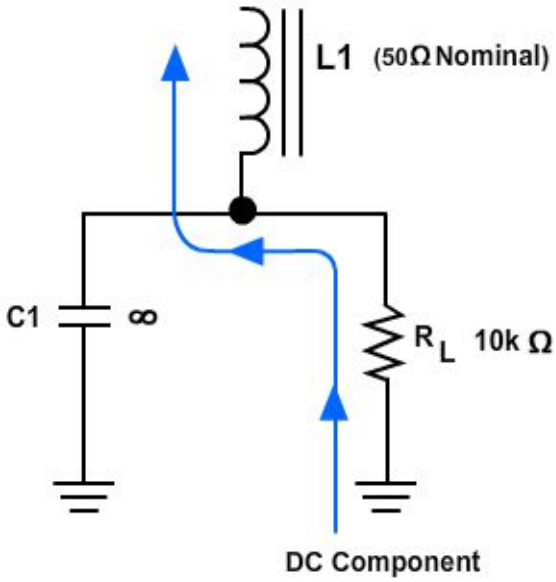


Figure 4-40 – DC component in an LC choke-input filter.

As you may have noticed, both the ac and the dc components flow through L1. Because it is frequency sensitive, the coil provides a large resistance to ac and a small resistance to dc. In other words, the coil opposes any change in current. This property makes the coil a highly desirable filter component. Note that the filtering action of the LC choke-input filter is improved when the filter is used in conjunction with a full-wave rectifier, as shown in *Figure 4-41*. This is due to the decrease in the X_C of the filter capacitor and the increase in the X_L of the choke. Remember, ripple frequency of a full-wave rectifier is twice that of a half-wave rectifier. For 60-hertz input, the ripple will be 120 hertz. The X_C of C1 and the X_L of L1 are calculated as follows:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(3.14)(120)(10)(10^{-6})}$$

$$X_C = \frac{(1)(10^6)}{7536}$$

$$X_C = 132.5\Omega$$

$$X_L = 2\pi fL$$

$$X_L = (2)(3.14)(120)(10)$$

$$X_L = 7.5\ kilohms$$

When the X_C of a filter capacitor is decreased, it provides less opposition to the flow of ac. The greater the ac flow through the capacitor, the lower the flow through the load. Conversely, the larger the X_L of the choke, the greater the amount of ac ripple developed across the choke; consequently, less ripple is developed across the load and better filtering is obtained.

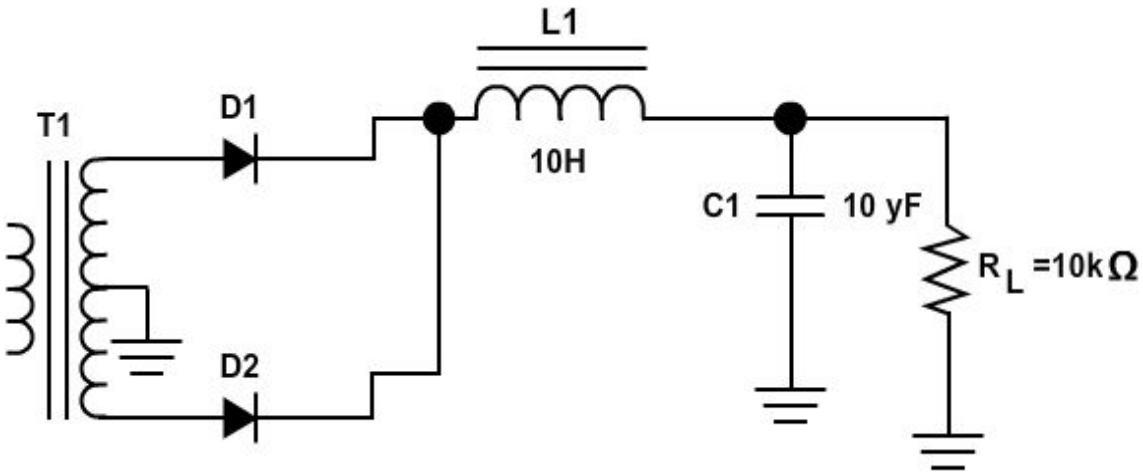


Figure 4-41 – Full-wave rectifier with an LC choke-input filter.

5.2.1 Failure Analysis of an LC Choke-Input Filter

The filter capacitors are subject to open circuits, short circuits, and excessive leakage; the series inductor is subject to open windings and, occasionally, shorted turns or a short circuit to the core.

The filter capacitor in the LC choke-input filter circuit is not subject to extreme voltage surges because of the protection offered by the inductor. However, the capacitor can become open, leaky, or shorted.

Shorted turns in the choke may reduce the value of inductance below the critical value. This will result in excessive peak-rectifier current, accompanied by an abnormally high output voltage, excessive ripple amplitude, and poor voltage regulation.

A choke winding that is open or shorted to the core will result in a no-output condition. A choke winding that is shorted to the core may cause overheating of the rectifier element(s) and blown fuses.

With the supply voltage removed from the input to the filter circuit, one terminal of the capacitor can be disconnected from the circuit. The capacitor should be checked with a capacitance analyzer to determine its capacitance and leakage resistance. When the capacitor is electrolytic, you must use the correct polarity at all times. A decrease in capacitance or losses within the capacitor can decrease the efficiency of the filter and can produce excessive ripple amplitude.

5.3.0 Resistor-Capacitor (RC) Filters

The RC capacitor-input filter is limited to applications in which the load current is small. This type of filter is used in power supplies where the load current is constant and voltage regulation is not necessary. For example, RC filters are used in high-voltage power supplies for cathode-ray tubes and in decoupling networks for multistage amplifiers.

Figure 4-42 shows an RC capacitor-input filter and associated waveforms. Both half-wave and full-wave rectifiers are used to provide the inputs. The waveform shown in *Figure 4-42, View A* represents the unfiltered output from a typical rectifier circuit. Note that the dashed lines in *Figure 4-42, View A* indicate the average value of output voltage (E_{avg}) for the half-wave rectifier. The average output voltage (E_{avg}) is less than half (approximately 0.318) the amplitude of the voltage peaks. The average value of output voltage (E_{avg}) for the full-wave rectifier is greater than half (approximately 0.637), but is

still much less than the peak amplitude of the rectifier-output waveform. With no filter circuit connected across the output of the rectifier circuit (unfiltered), the waveform has a large value of pulsating component (ripple) as compared to the average (or dc) component.

The RC filter in *Figure 4-42* consists of an input filter capacitor (C1), a series resistor (R1), and an output filter capacitor (C2). This filter is sometimes referred to as an RC pi-section filter because its schematic symbol resembles the Greek letter π.

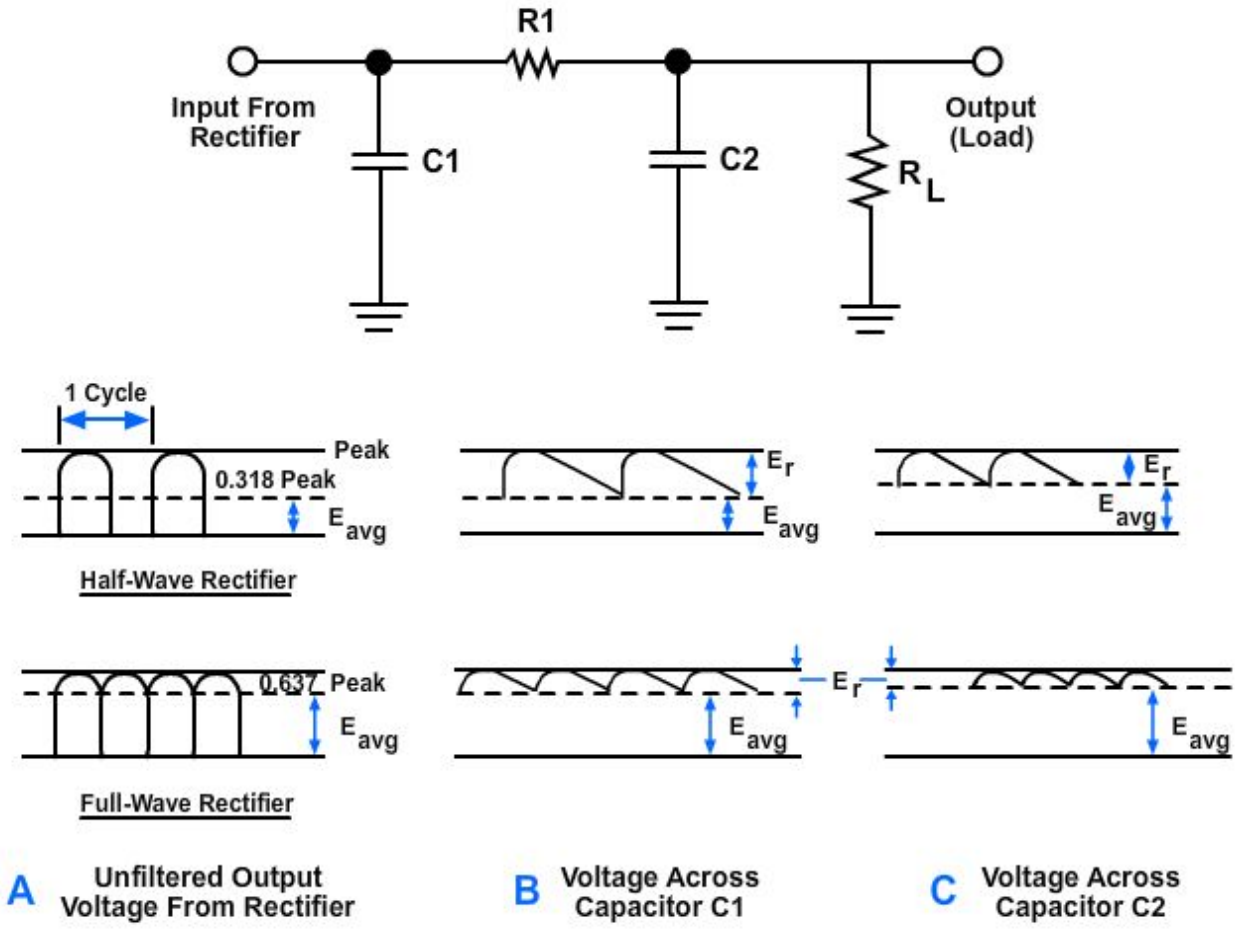


Figure 4-42 – RC filter and waveforms.

The single capacitor filter is suitable for many noncritical, low-current applications. However, when the load resistance is very low or when the percent of ripple must be held to an absolute minimum, the capacitor value required must be extremely large. While electrolytic capacitors are available in sizes up to 10,000 microfarads or greater, the large sizes are quite expensive. A more practical approach is to use a more sophisticated filter that can do the same job but that has lower capacitor values, such as the RC filter.

Figure 4-42, Views A, B, and *C* show the output waveforms of a half-wave and a full-wave rectifier. Each waveform is shown with an RC filter connected across the output. The following explanation of how a filter works will show you that an RC filter of this type does a much better job than the single capacitor filter.

C1 performs exactly the same function as it did in the single capacitor filter. It is used to reduce the percentage of ripple to a relatively low value. Thus, the voltage across C1 might consist of an average dc value of +100 volts with a ripple voltage of 10 volts peak-

to-peak. This voltage is passed on to the R1-C2 network, which reduces the ripple even further.

C2 offers an infinite impedance (resistance) to the dc component of the output voltage. Thus, the dc voltage is passed to the load, but reduced in value by the amount of the voltage drop across R1. However, R1 is generally small compared to the load resistance; therefore, the drop in the dc voltage by R1 is not a drawback.

Component values are designed so that the resistance of R1 is much greater than the reactance (X_C) of C2 at the ripple frequency. C2 offers a very low impedance to the ac ripple frequency. Thus, the ac ripple senses a voltage divider consisting of R1 and C2 between the output of the rectifier and ground; therefore, most of the ripple voltage is dropped across R1. Only a trace of the ripple voltage can be seen across C2 and the load. In extreme cases where the ripple must be held to an absolute minimum, a second stage of RC filtering can be added. In practice, the second stage is rarely required. The RC filter is extremely popular because smaller capacitors can be used with good results.

The RC filter has some disadvantages. First, the voltage drop across R1 takes voltage away from the load. Second, power is wasted in R1 and is dissipated in the form of unwanted heat. Finally, if the load resistance changes, the voltage across the load will change. Even so, the advantages of the RC filter overshadow these disadvantages in many cases.

5.3.1 Failure Analysis of the Resistor-Capacitor (RC) Filter

The shunt capacitors (C1 and C2) are subject to an open circuit, a short circuit, or excessive leakage. The series filter resistor (R1) is subject to changes in value and occasionally opens. Any of these troubles can be easily detected.

The input capacitor (C1) has the greatest pulsating voltage applied to it and is the most susceptible to voltage surges. As a result, the input capacitor is frequently subject to voltage breakdown and shorting. The remaining shunt capacitor (C2) in the filter circuit is not subject to voltage surges because of the protection offered by the series filter resistor (R1). However, a shunt capacitor can become open, leaky, or shorted.

A shorted capacitor or an open filter resistor results in a no-output indication. An open filter resistor results in an abnormally high dc voltage at the input to the filter and no voltage at the output of the filter. Leaky capacitors or filter resistors that have lost their effectiveness, or filter resistors that have decreased in value, result in an excessive ripple amplitude in the output of the supply.

5.4.0 LC Capacitor-Input Filter

The LC capacitor-input filter is one of the most commonly used filters. This type of filter is used primarily in radio receivers, small audio amplifier power supplies, and in any type of power supply where the output current is low and the load current is relatively constant.

Figure 4-43 shows an LC capacitor-input filter and associated waveforms. Both half-wave and full-wave rectifier circuits are used to provide the input. The waveforms shown in *Figure 4-43, View A* represent the unfiltered output from a typical rectifier circuit. Note that the average value of output voltage (E_{avg}), indicated by the dashed lines, for the half-wave rectifier is less than half the amplitude of the voltage peaks. The average value of output voltage (E_{avg}) for the full-wave rectifier is greater than half, but is still much less than the peak amplitude of the rectifier-output waveform. With no filter

connected across the output of the rectifier circuit (which results in unfiltered output voltage), the waveform has a large value of pulsating component (ripple) as compared to the average (or dc) component.

C1 reduces the ripple to a relatively low level (*Figure 4-43, View B*). L1 and C2 form the LC filter, which reduces the ripple even further. L1 is a large value iron-core induct (choke). L1 has a high value of inductance and therefore, a high value of X_L which offers a high reactance to the ripple frequency. At the same time, C2 offers a very low reactance to ac ripple. L1 and C2 are used for an ac voltage divider and, because the reactance of L1 much higher than that of C2, most of the ripple voltage is dropped across L1. Only a slight trace of ripple appears across C2 and the load (*Figure 4-43, View C*).

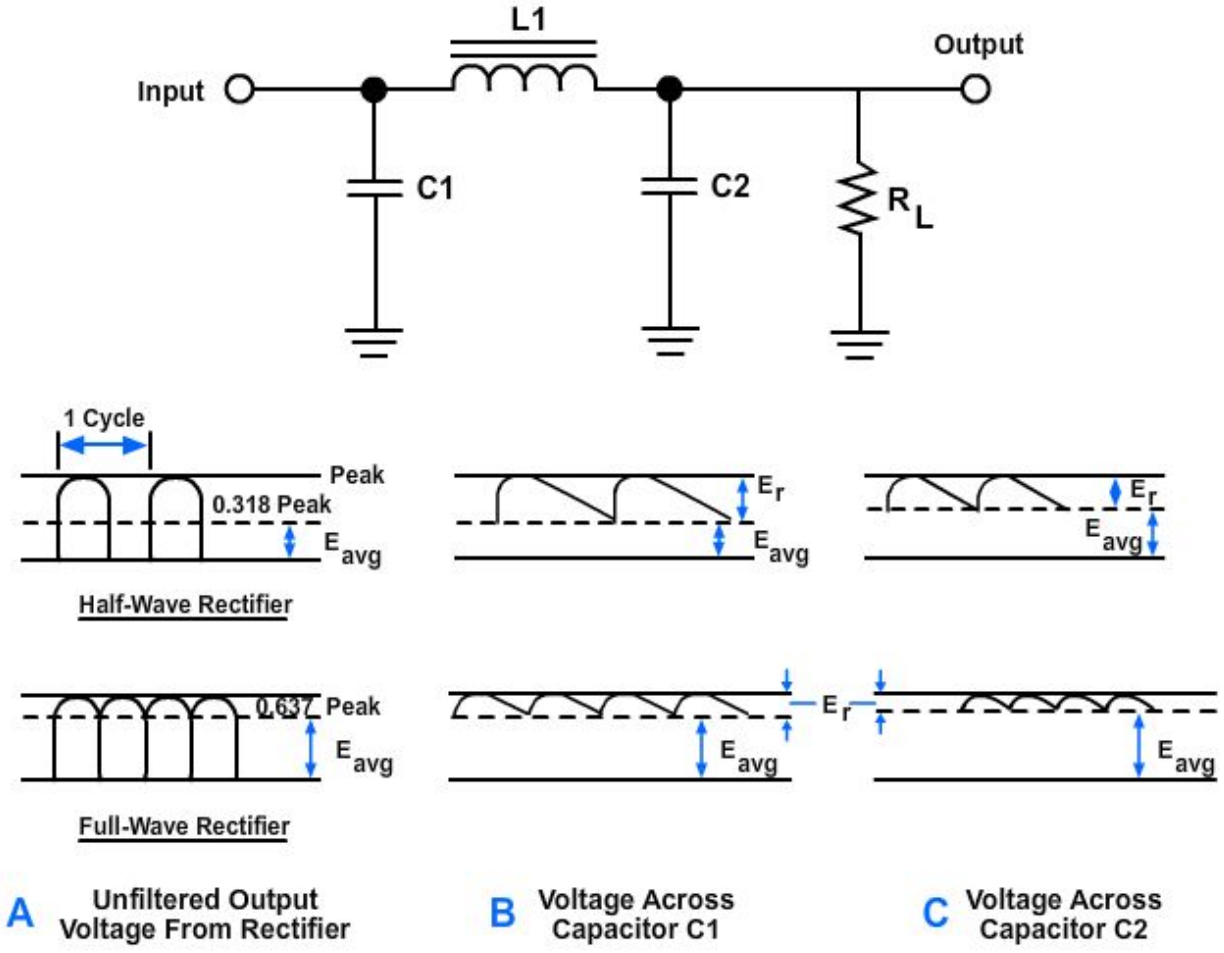


Figure 4-43 – LC filter and waveforms.

While the L1-C2 network greatly reduces ac ripple, it has little effect on dc. You should recall that an inductor offers no reactance to dc. The only opposition to current flow is the resistance of the wire in the choke. Generally, this resistance is very low and the dc voltage drop across the coil is minimal. Thus, the LC filter overcomes the disadvantages of the RC filter.

Aside from the voltage divider effect, the inductor improves filtering in another way. You should recall that an inductor resists changes in the magnitude of the current flowing through it. Consequently, when the inductor is placed in series with the load, the inductor maintains steady current. In turn, this helps the voltage across the load remain constant when size of components is a factor.

The LC filter provides good filtering action over a wide range of currents. The capacitor filters best when the load is drawing little current. Thus, the capacitor discharges very slowly and the output voltage remains almost constant. On the other hand, the inductor filters best when the current is highest. The complementary nature of these two components ensures that good filtering will occur over a wide range of currents.

The LC filter has two disadvantages. First, it is more expensive than the RC filter because an iron-core choke costs more than a resistor. The second disadvantage is size. The iron-core choke is bulky and heavy, a fact which may render the LC filter unsuitable for many applications.

5.4.1 Failure Analysis of the LC Capacitor-Input Filter

Shunt capacitors are subject to open circuits, short circuits, and excessive leakage; series inductors are subject to open windings and occasionally shorted turns or a short circuit to the core.

The input capacitor (C1) has the greatest pulsating voltage applied to it, is the most susceptible to voltage surges, and has a generally higher average voltage applied. As a result, the input capacitor is frequently subject to voltage breakdown and shorting. The output capacitor (C2) is not as susceptible to voltage surges because of the series protection offered by the series inductor (L1), but the capacitor can become open, leaky, or shorted.

A shorted capacitor, an open filter choke, or a choke winding that is shorted to the core results in a no-output indication. A shorted capacitor, depending on the magnitude of the short, may cause a shorted rectifier, transformer, or filter choke, and may result in a blown fuse in the primary of the transformer. An open filter choke results in an abnormally high dc voltage at the input to the filter and no voltage at the output of the filter. A leaky or open capacitor in the filter circuit results in a low dc output voltage. This condition is generally accompanied by an excessive ripple amplitude. Shorted turns in the winding of a filter choke reduce the effective inductance of the choke and decrease its filtering efficiency. As a result, the ripple amplitude increases.

6.0.0 INTRODUCTION to TRANSISTORS

The discovery of the first transistor in 1948 by a team of physicists at the Bell Telephone Laboratories sparked an interest in solid-state research that spread rapidly. The transistor, which began as a simple laboratory oddity, was rapidly developed into a semiconductor device of major importance. The transistor demonstrated for the first time in history that amplification in solids was possible. Before the transistor, amplification was achieved only with electron tubes. Transistors now perform numerous electronic tasks with new and improved transistor designs being continually put on the market. In many cases, transistors are more desirable than tubes because they are small, rugged, require no filament power, and operate at low voltages with comparatively high efficiency. The development of a family of transistors has even made possible the miniaturization of electronic circuits. *Figure 4-44* shows a sample of the many different types of transistors you may encounter when working with electronic equipment.

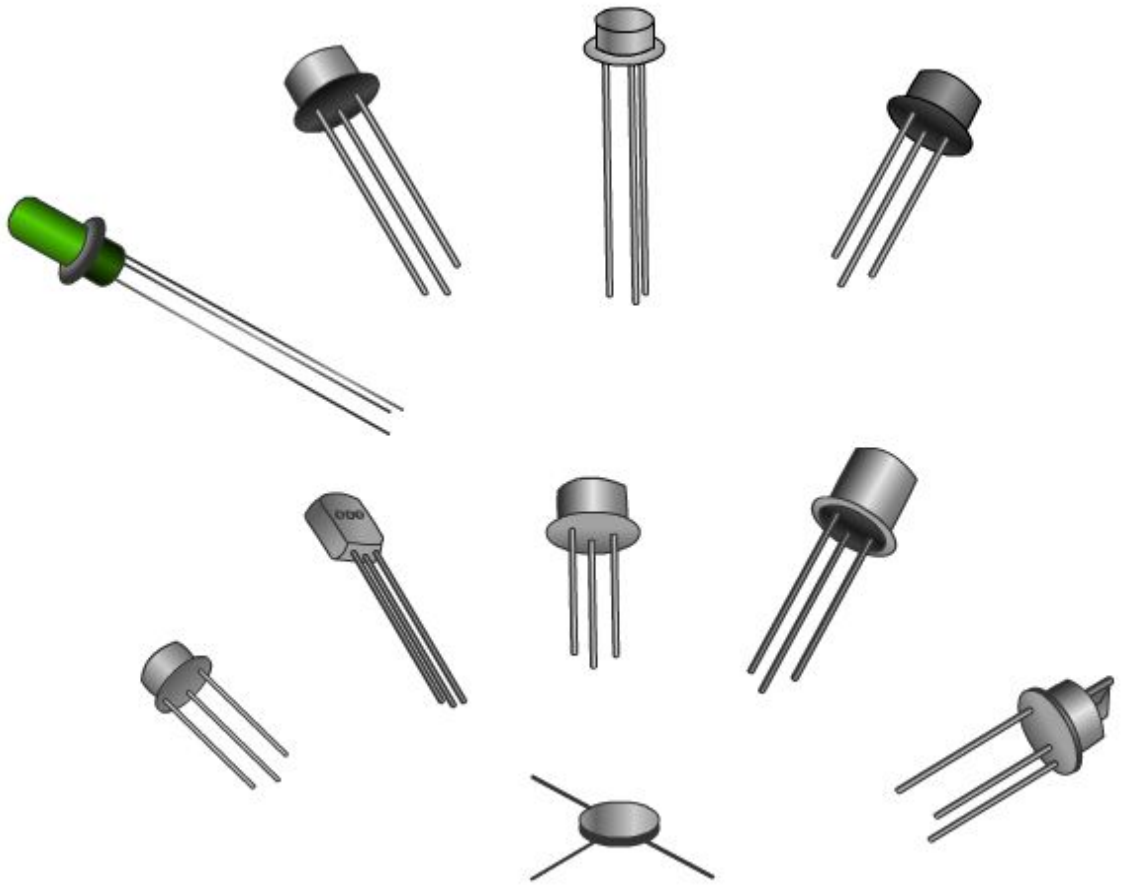


Figure 4-44 – An assortment of different types of transistors.

Transistors have infiltrated virtually every area of science and industry, from the family car to satellites. Even the military depends heavily on transistors. The ever-increasing uses for transistors have created an urgent need for sound and basic information regarding their operation.

From your study of the PN-junction diode, you now have the basic knowledge to grasp the principles of transistor operation. In this section, you will first become acquainted with the basic types of transistors, their construction, and their theory of operation.

The first solid-state device discussed was the two-element semiconductor diode. The next device on our list is even more unique. It not only has one more element than the diode, but it can amplify as well. Semiconductor devices that have three or more elements are called transistors. The term transistor was derived from the words transfer and resistor. This term was adopted because it best describes the operation of the transistor – the transfer of an input signal current from a low-resistance circuit to a high-resistance circuit. Basically, the transistor is a solid-state device that amplifies by controlling the flow of current carriers through its semiconductor materials.

There are many different types of transistors, but their basic theory of operation is all the same. As a matter of fact, the theory we will be using to explain the operation of a transistor is the same theory used earlier with the PN-junction diode except that now, two such junctions are required to form the three elements of a transistor. The three elements of the two-junction transistor are: (1) the emitter, which gives off, or emits,

current carriers (electrons or holes); (2) the base, which controls the flow of current carriers; and (3) the collector, which collects the current carriers.

6.1.0 Transistor Theory

You should recall from an earlier discussion that a forward-biased PN junction is comparable to a low-resistance circuit element because it passes a high current for a given voltage. In turn, a reverse-biased PN junction is comparable to a high-resistance circuit element. By using Ohm's law formula for power ($P = I^2R$) and assuming current is held constant, you can conclude that the power developed across a high resistance is greater than that developed across a low resistance. Thus, if a crystal were to contain two PN junctions (one forward biased and the other reverse biased), a low-power signal could be injected into the forward-biased junction and produce a high-power signal at the reverse-biased junction. In this manner, a power gain would be obtained across the crystal. This concept is merely an extension of the material covered in a previous section of this chapter and is the basic theory behind how the transistor amplifies.

6.1.1 NPN Transistor Operation

Just as in the case of the PN junction diode, the N-material comprising the two end sections of the NPN transistor contains a number of free electrons, while the center P section contains an excess number of holes. The action at each junction between these sections is the same as that previously described for the diode; that is, depletion regions develop and the junction barrier appears. To use the transistor as an amplifier, each of these junctions must be modified by some external bias voltage. For the transistor to function in this capacity, the first PN junction (emitter-base junction) is biased in the forward, or low-resistance, direction. At the same time, the second PN junction (base-collector junction) is biased in the reverse, or high-resistance, direction. A simple way to remember how to properly bias a transistor is to observe the NPN or PNP elements that make up the transistor. The letters of these elements indicate what polarity voltage to use for correct bias. For instance, notice the NPN transistor in *Figure 4-45.*

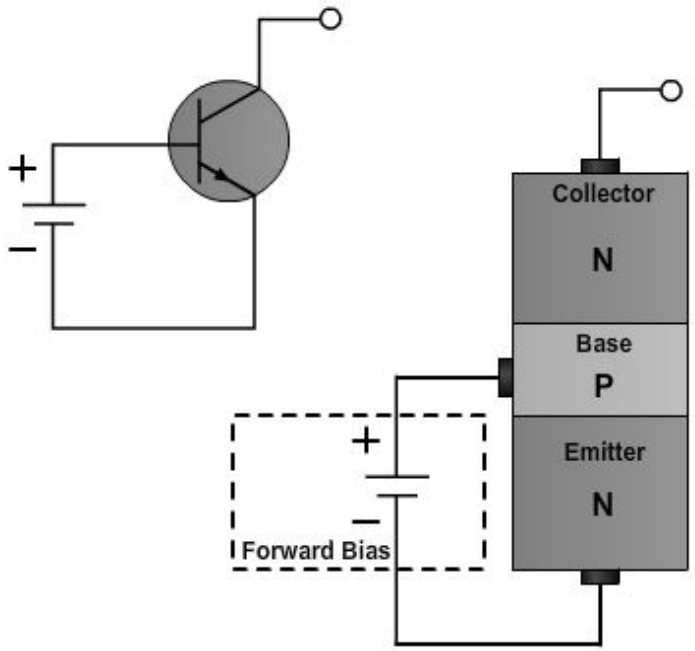


Figure 4-45 — NPN transistor.

The emitter, which is the first letter in the NPN sequence, is connected to the negative side of the battery, while the base, which is the second letter NPN, is connected to the positive side.

However since the second PN junction is required to be reverse biased for proper transistor operation, the collector must be connected to an opposite polarity voltage (positive) than that indicated by its letter designation (NPN). The voltage on the collector must also be more positive than the base (*Figure 4-46*).

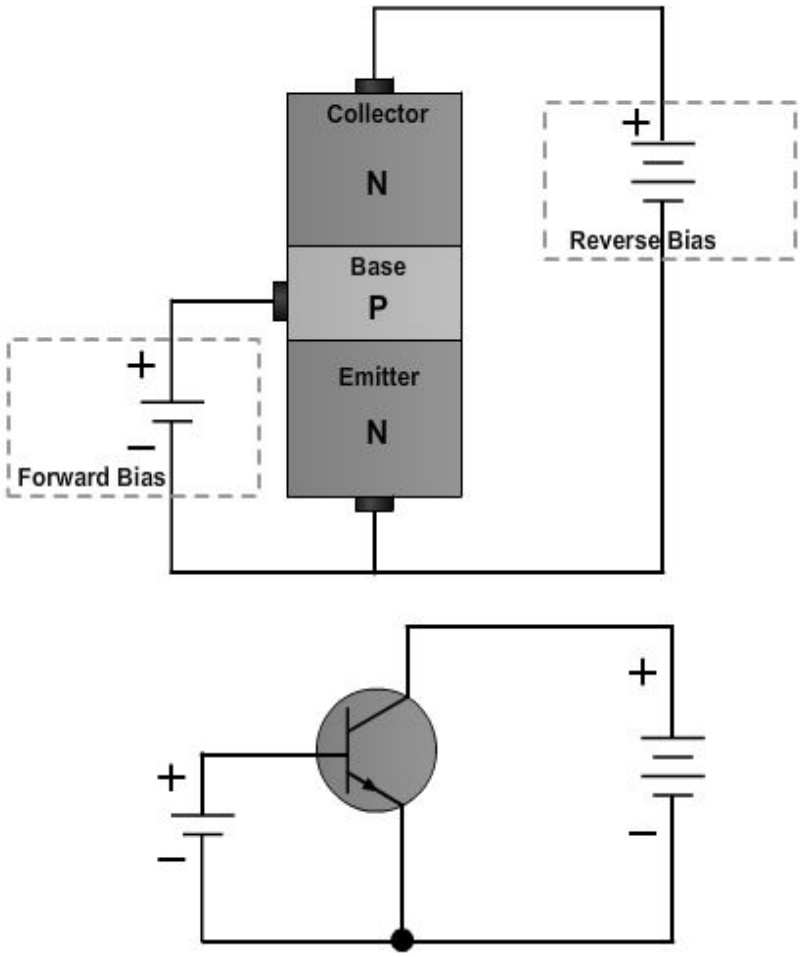


Figure 4-46 – Properly biased NPN transistor.

You now have a properly biased NPN transistor. In summary, the base of the NPN transistor must be positive with respect to the emitter, and the collector must be more positive than the base.

6.1.1.1 NPN Forward-Biased Junction

An important point to bring out at this time, which was not necessarily mentioned during the explanation of the diode, is the fact that the N-material on one side of the forward-biased junction is more heavily doped than the P-material. This results in more current being carried across the junction by the majority carrier electrons from the N-material than the majority carrier holes from the P-material. Therefore, conduction through the forward-biased junction is mainly by majority carrier electrons from the N-material (emitter) (*Figure 4-47*).

With the emitter-to-base junction in the figure biased in the forward direction, electrons leave the negative terminal of the battery and enter the N-material (emitter). Since electrons are majority current carriers in the N-material, they pass easily through the emitter, cross over the junction, and combine with holes in the P-material (base). For each electron that fills a hole in the P-material, another electron will leave the P-material (creating a new hole) and enter the positive terminal of the battery.

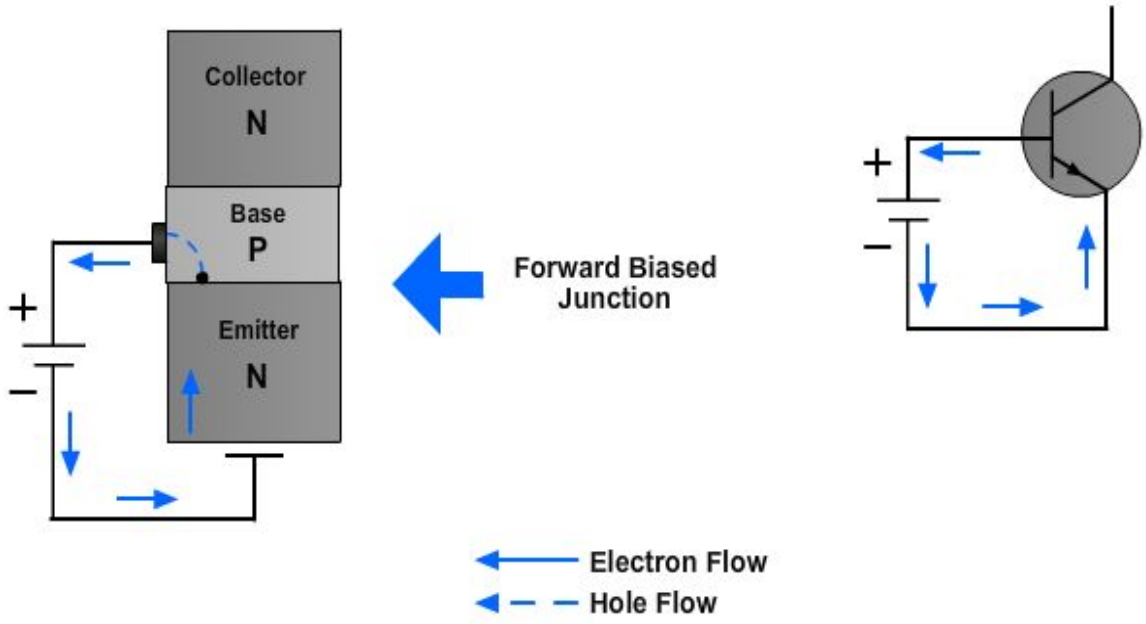


Figure 4-47 — Forward-biased junction in an NPN transistor.

6.1.1.2 NPN Reverse-Biased Junction

The second PN junction (base-to-collector), or reverse-biased junction as it is called, blocks the majority of current carriers from crossing the junction (*Figure 4-48*). However, there is a very small current, mentioned earlier, that does pass through this junction. This current is called minority current, or reverse current. As you recall, this current was produced by the electron-hole pairs. The minority carriers for the reverse-biased PN junction are the electrons in the P-material and the holes in the N-material. These minority carriers actually conduct the current for the reverse-biased junction when electrons from the P-material enter the N-material, and the holes from the N-material enter the P-material. However, the minority current electrons (as you will see later) play the most important part in the operation of the NPN transistor.

At this point, you may wonder why the second PN junction (base-to-collector) is not forward biased like the first PN junction (emitter-to-base). If both junctions were forward biased, the electrons would have a tendency to flow from each end section of the N P N transistor (emitter and collector) to the center P section (base). In essence, we would have two junction diodes possessing a common base, thus eliminating any amplification and defeating the purpose of the transistor. A word of caution is in order at this time. If you should mistakenly bias the second PN junction in the forward direction, the excessive current could develop enough heat to destroy the junctions, making the transistor useless; therefore, be sure your bias voltage polarities are correct before making any electrical connections.

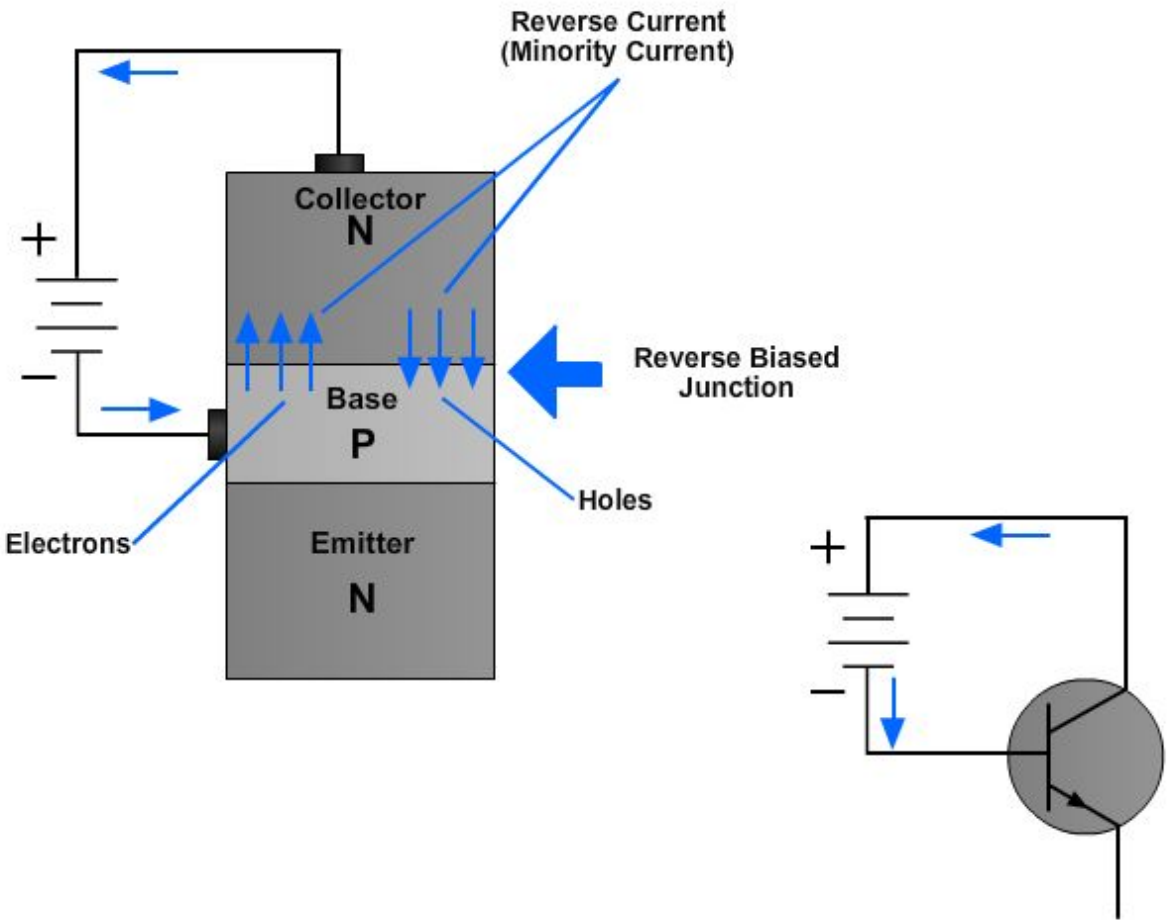


Figure 4-48 – Reverse-biased junction in an NPN transistor.

6.1.1.3 NPN Junction Interaction

You are now ready to see what happens when the two junctions of the NPN transistor are placed in operation at the same time. For a better understanding of just how the two junctions work together, refer to *Figure 4-49* throughout the discussion.

The bias batteries in *Figure 4-49* have been labeled V_{CC} for the collector voltage supply and V_{BB} for the base voltage supply. Also notice the base supply battery is quite small, as indicated by the number of cells in the battery, usually 1 volt or less. However, the collector supply is generally much higher than the base supply, normally around 6 volts. As you will see later, this difference in supply voltages is necessary to have current flow from the emitter to the collector.

As stated earlier, the current flow in the external circuit is always due to the movement of free electrons; therefore, electrons flow from the negative terminals of the supply batteries to the N-type emitter. This combined movement of electrons is known as emitter current (I_E). Since electrons are the majority carriers in the N-material, they will move through the N-material emitter to the emitter-base junction. With this junction forward biased, electrons continue on into the base region. Once the electrons are in the base, which is a P-type material, they become minority carriers. Some of the electrons that move into the base recombine with available holes. For each electron that recombines, another electron moves out through the base lead as base current (I_B) (creating a new hole for eventual combination) and returns to the base supply battery (V_{BB}). The electrons that recombine are lost as far as the collector is concerned;

therefore, to make the transistor more efficient, the base region is made very thin and lightly doped. This reduces the opportunity for an electron to recombine with a hole and be lost. Thus, most of the electrons that move into the base region come under the influence of the large collector reverse bias. This bias acts as forward bias for the minority carriers (electrons) in the base and, as such, accelerates them through the base-collector junction and on into the collector region. Since the collector is made of an N-type material, the electrons that reach the collector again become majority current carriers. Once in the collector, the electrons move easily through the N-material and return to the positive terminal of the collector supply battery (V_{CC}) as collector current (I_C).

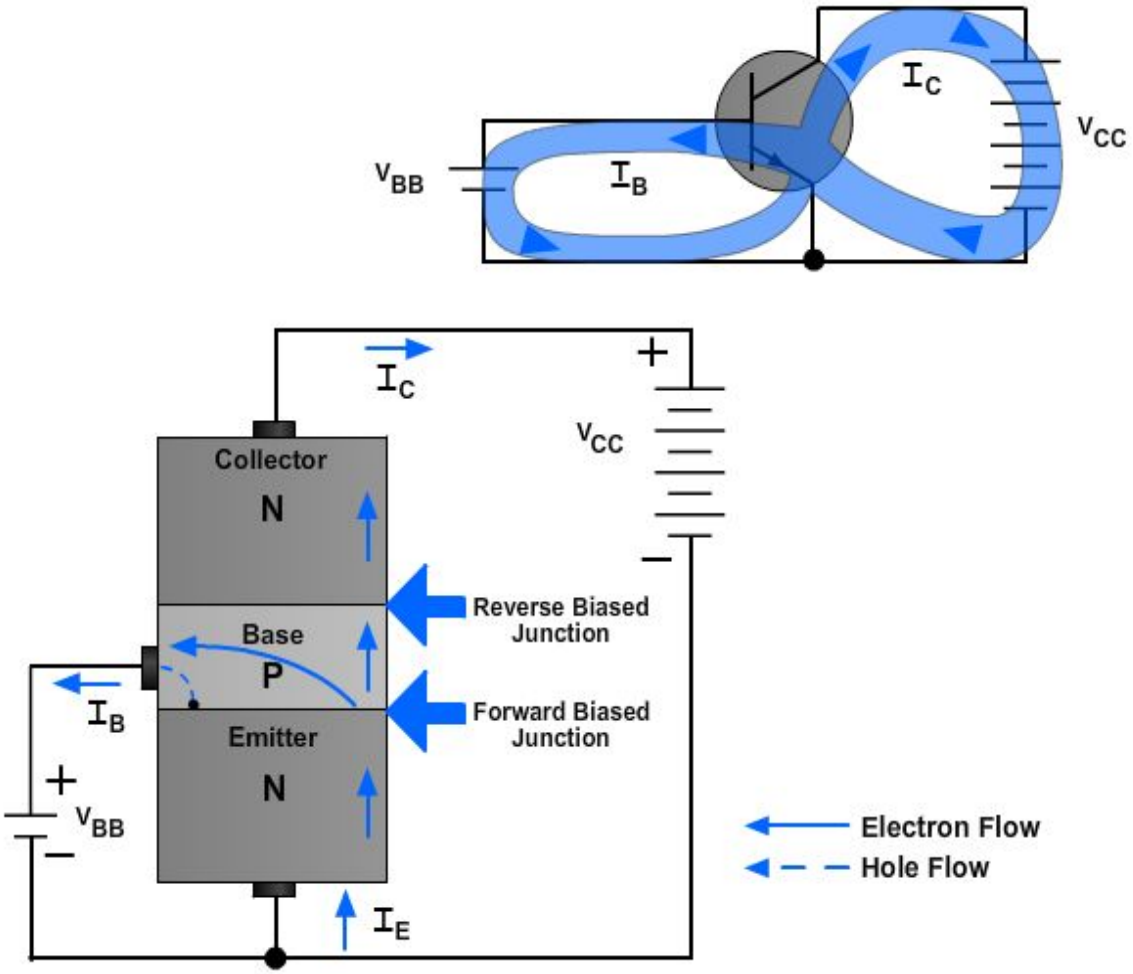


Figure 4-49 – NPN transistor operation.

To further improve on the efficiency of the transistor, the collector is made physically larger than the base for two reasons: (1) to increase the chance of collecting carriers that diffuse to the side as well as directly across the base region and (2) to enable the collector to handle more heat without damage.

In summary, total current flow in the NPN transistor is through the emitter lead; therefore, in terms of percentage, I_E is 100 percent. On the other hand, since the base is very thin and lightly doped, a smaller percentage of the total current (emitter current) will flow in the base circuit than in the collector circuit. Usually no more than 2 to 5 percent of the total current is base current (I_B) while the remaining 95 to 98 percent is collector current (I_C). A very basic relationship exists between these two currents:

$$I_E = I_B + I_C$$

In simple terms, this means that the emitter current is separated into base and collector current. Since the amount of current leaving the emitter is solely a function of the emitter-base bias, and because the collector receives most of this current, a small change in emitter-base bias will have a far greater effect on the magnitude of collector current than it will have on base current. In conclusion, the relatively small emitter-base bias controls the relatively large emitter-to-collector current.

6.1.2 PNP Transistor Operation

The PNP transistor works essentially the same as the NPN transistor. However, since the emitter, base, and collector in the PNP transistor are made of materials that are different from those used in the NPN transistor, different current carriers flow in the PNP unit. The majority current carriers in the PNP transistor are holes. This is in contrast to the NPN transistor, where the majority current carriers are electrons. To support this different type of current (hole flow), the bias batteries are reversed for the PNP transistor. A typical bias setup for the PNP transistor is shown in *Figure 4-50*. Notice that the procedure used earlier to properly bias the NPN transistor also applies here to the PNP transistor. The first letter (P) in the PNP sequence indicates the polarity of the voltage required for the emitter (positive), and the second letter (N) indicates the polarity of the base voltage (negative). Because the base-collector junction is always reverse biased, the opposite polarity voltage (negative) must be used for the collector. Thus, the base of the PNP transistor must be negative with respect to the emitter, and the collector must be more negative than the base. Remember, that just as in the case of

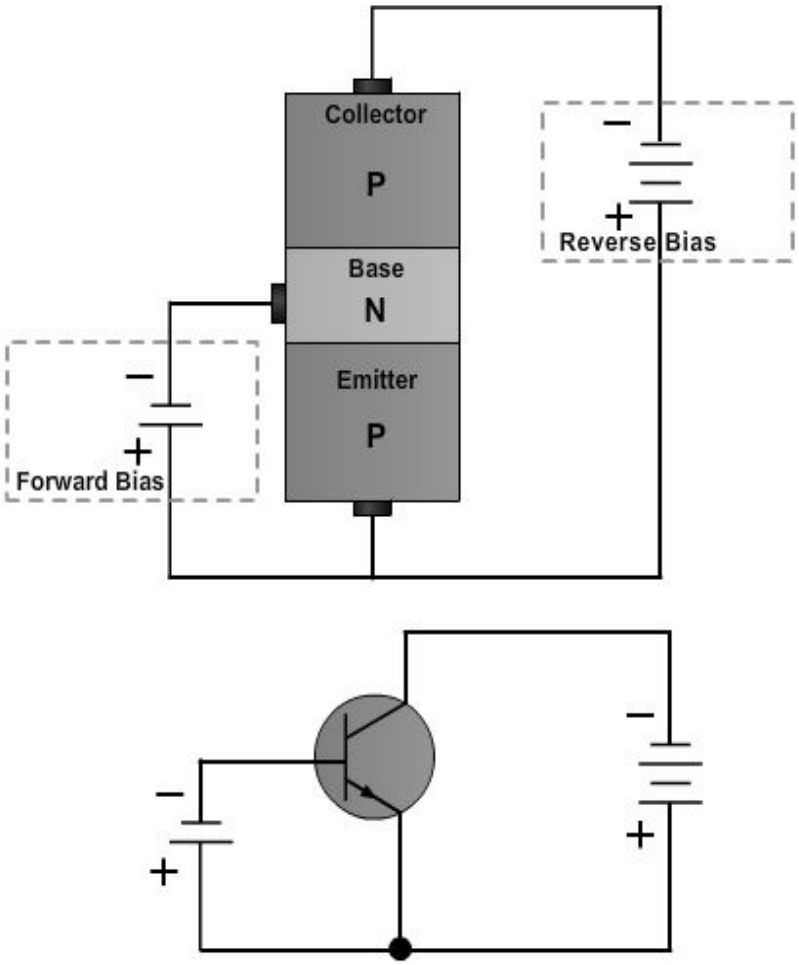


Figure 4-50 – Properly biased PNP transistor.

the NPN transistor, this difference in supply voltage is necessary to have current flow (hole flow in the case of the PNP transistor) from the emitter to the collector. Although hole flow is the predominant type of current flow in the PNP transistor, hole flow only takes place within the transistor itself, while electrons flow in the external circuit. However, it is the internal hole flow that leads to electron flow in the external wires connected to the transistor.

6.1.2.1 PNP Forward-Biased Junction

Now consider what happens when the emitter-base junction in *Figure 4-51* is forward biased. With the bias setup shown, the positive terminal of the battery repels the emitter holes toward the base, while the negative terminal drives the base electrons toward the emitter. When an emitter hole and a base electron meet, they combine. For each electron that combines with a hole, another electron leaves the negative terminal of the battery, and enters the base. At the same time, an electron leaves the emitter, creating a new hole, and enters the positive terminal of the battery. This movement of electrons into the base and out of the emitter constitutes base current flow (I_B), and the path these electrons take is referred to as the emitter-base circuit.

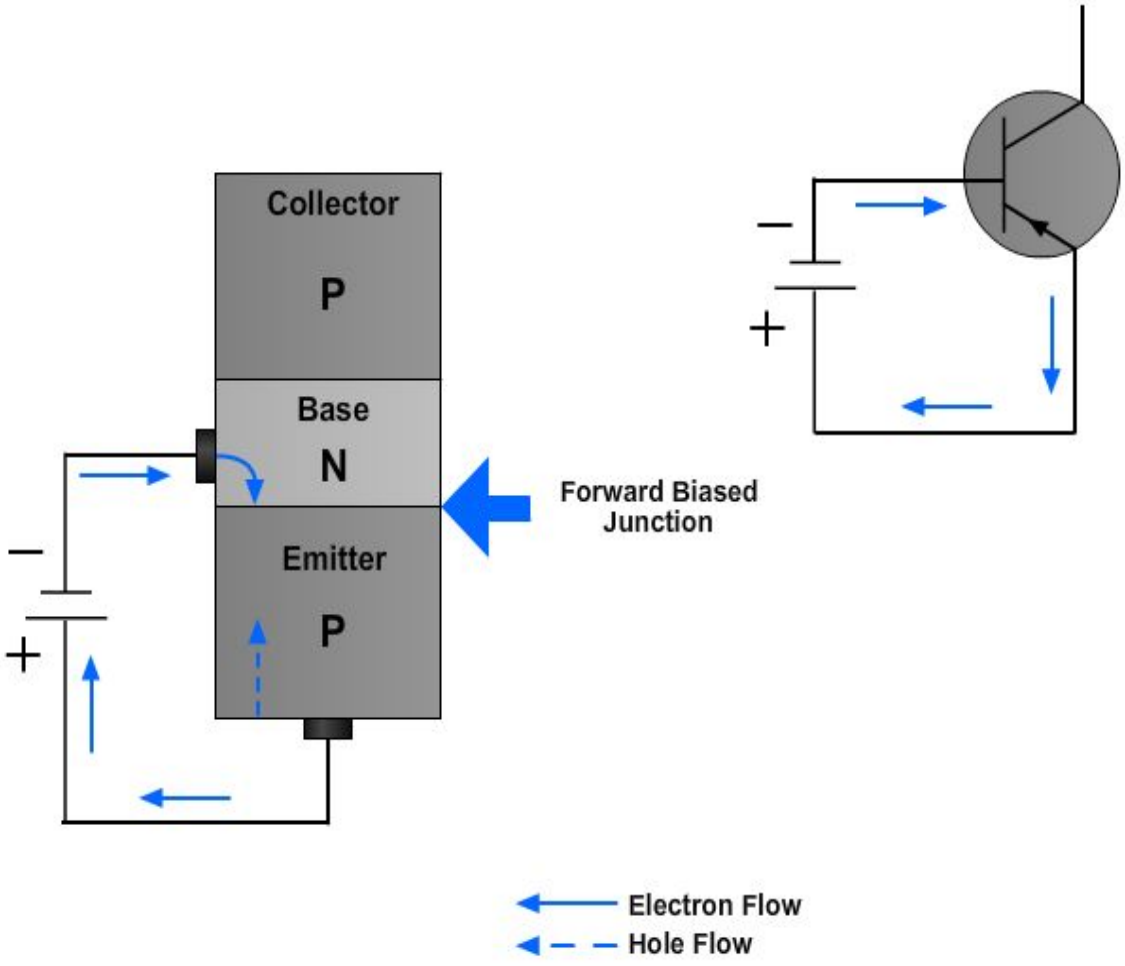


Figure 4-51 – Forward-biased junction in a PNP transistor.

6.1.2.2 Reverse-Biased Junction

In the reverse-biased junction, the negative voltage on the collector and the positive voltage on the base block the majority current carriers from crossing the junction (*Figure 4-52)*. However, this same negative collector voltage acts as forward bias for the minority current holes in the base, which cross the junction and enter the collector. The minority current electrons in the collector also sense forward bias (the positive base voltage) and move into the base. The holes in the collector are filled by electrons that flow from the negative terminal of the battery. At the same time that the electrons leave the negative terminal of the battery, other electrons in the base break their covalent bonds and enter the positive terminal of the battery. Although there is only minority current flow in the reverse-biased junction, it is still very small because of the limited number of minority current carriers.

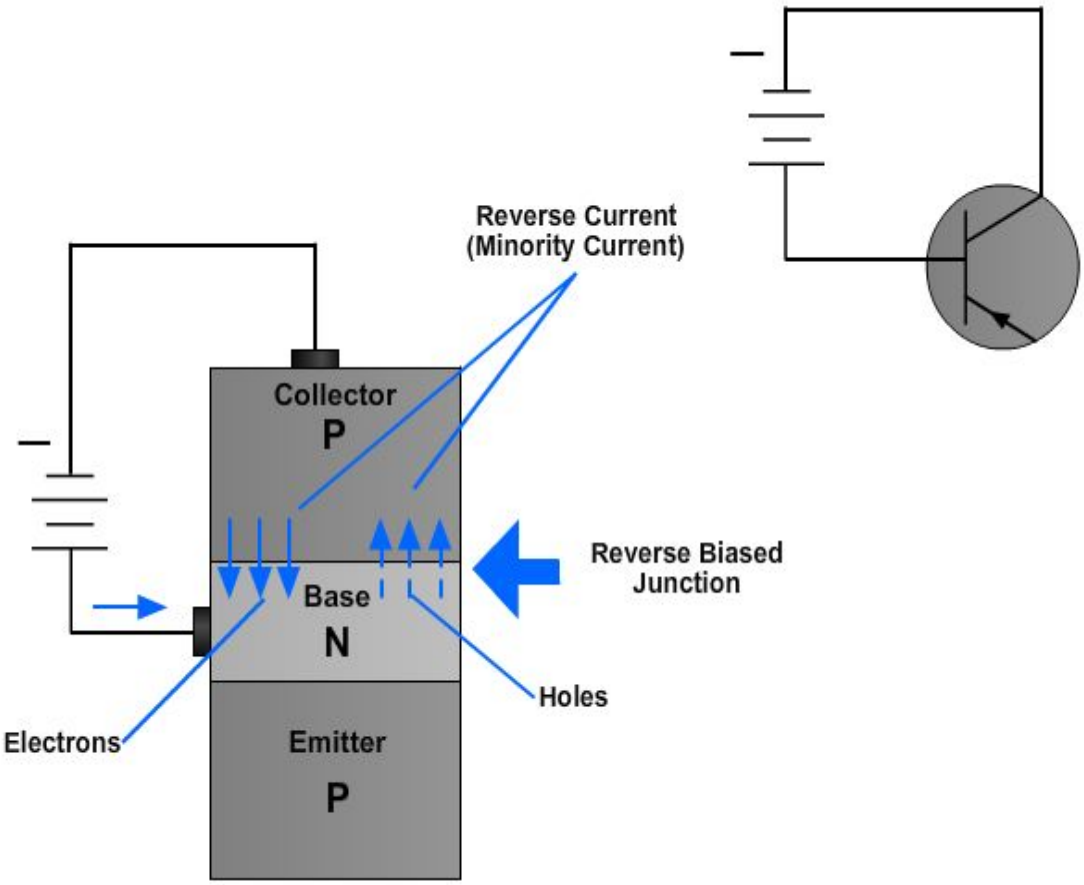


Figure 4-52 – Reverse-biased junction in a PNP transistor.

6.1.2.3 PNP Junction Interaction

The interaction between the forward- and reverse-biased junctions in a PNP transistor is very similar to that in an NPN transistor, except that in the PNP transistor, the majority current carriers are holes. In the PNP transistor shown in *Figure 4-53*, the positive voltage on the emitter repels the holes toward the base. Once in the base, the holes combine with base electrons. But again, remember that the base region is made very thin to prevent the recombination of holes with electrons; therefore, well over 90 percent of the holes that enter the base become attracted to the large negative collector voltage

and pass right through the base. However, for each electron and hole that combines in the base region, another electron leaves the negative terminal of the base battery (V_{BB}) and enters the base as base current (I_B). At the same time that an electron leaves the negative terminal of the battery, another electron leaves the emitter as IE (creating a new hole) and enters the positive terminal of V_{BB}. Meanwhile, in the collector circuit, electrons from the collector battery (V_{CC}) enter the collector as *Ic* and combine with the excess holes from the base. For each hole that is neutralized in the collector by an electron, another electron leaves the emitter and starts its way back to the positive terminal of V_{CC}.

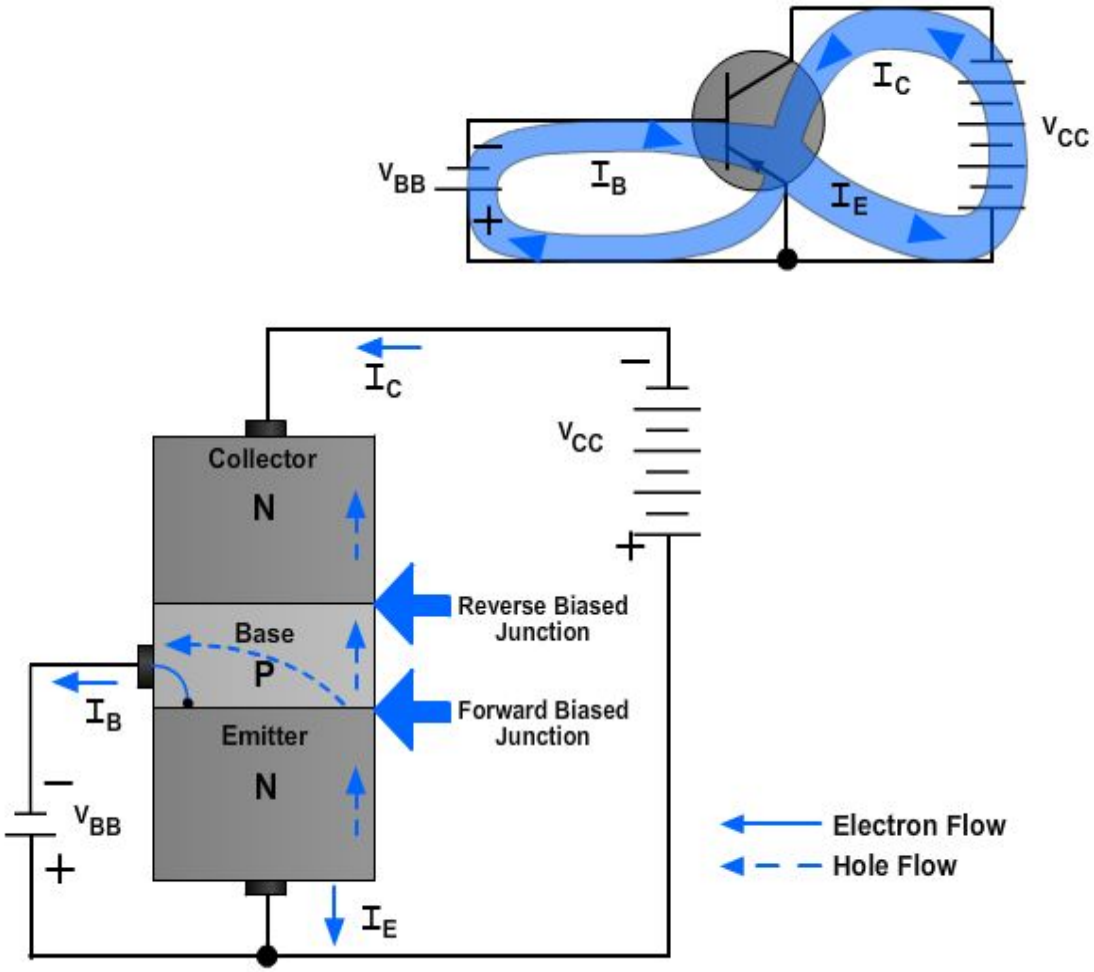


Figure 4-53 – PNP transistor operation.

Although current flow in the external circuit of the PNP transistor is opposite in direction to that of the NPN transistor, the majority carriers always flow from the emitter to the collector. This flow of majority carriers also results in the formation of two individual current loops within each transistor. One loop is the base-current path, and the other loop is the collector-current path. The combination of the current in both of these loops ($I_B + I_C$) results in total transistor current (I_E). The most important thing to remember about the two different types of transistors is that the emitter-base voltage of the PNP transistor has the same controlling effect on collector current as that of the NPN transistor. In simple terms, increasing the forward-bias voltage of a transistor reduces the emitter-base junction barrier. This action allows more carriers to reach the collector, causing an increase in current flow from the emitter to the collector and through the external circuit. Conversely, a decrease in the forward-bias voltage reduces collector current.

6.2.0 Transistor Classification

Transistors are classified as either NPN or PNP according to the arrangement of their N- and P-materials. Their basic construction and chemical treatment is implied by their names, NPN or PNP; that is, an NPN transistor is formed by introducing a thin region of P-type material between two regions of N-type material. On the other hand, a PNP transistor is formed by introducing a thin region of N-type material between two regions of P-type material. Transistors constructed in this manner have two PN junctions, as shown in *Figure 4-54.* One PN junction is between the emitter and the base; the other PN junction is between the collector and the base. The two junctions share one section of semiconductor material so that the transistor actually consists of three elements.

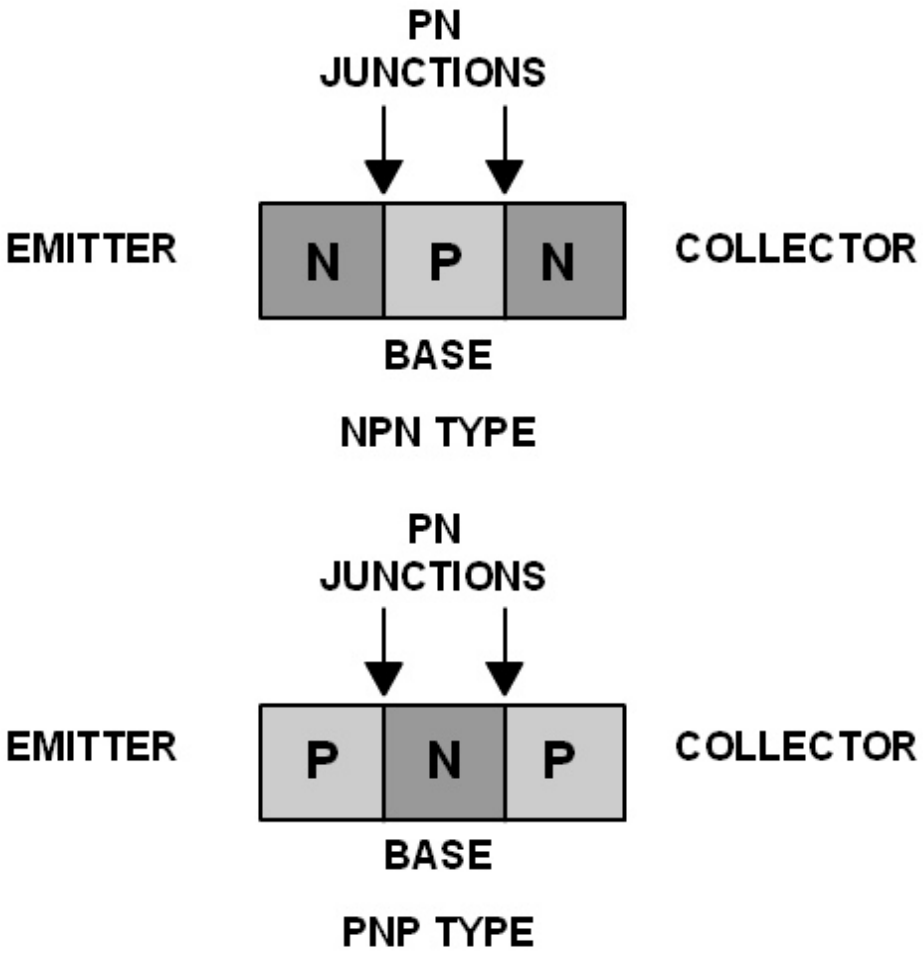


Figure 4-54 – Transistor block diagrams.

Since the majority and minority current carriers are different for N- and P-materials, it stands to reason that the internal operation of the NPN and PNP transistors will also be different. The theory of operation of the NPN and PNP transistors will be discussed separately in the next few paragraphs. Any additional information about the PN junction will be given as the theory of transistor operation is developed.

6.2.1 Amplifier Classes of Operation

In the previous discussions, we assumed that for every portion of the input signal there was an output from the amplifier. This is not always the case with amplifiers. It may be desirable to have the transistor conducting for only a portion of the input signal. The portion of the input for which there is an output determines the class of operation of the amplifier. There are four classes of amplifier operations. They are class A, class AB, class B, and class C.

6.2.1.1 Class A Amplifier Operation

Class A amplifiers are biased so that variations in input signal polarities occur within the limits of ***cutoff*** and ***saturation***. In a PNP transistor, for example, if the base becomes positive with respect to the emitter, holes will be repelled at the PN junction and no current can flow in the collector circuit. This condition is known as cutoff. Saturation occurs when the base becomes so negative with respect to the emitter that changes in the signal are not reflected in collector-current flow.

Biasing an amplifier in this manner places the dc operating point between cutoff and saturation and allows collector current to flow during the complete cycle (360 degrees) of the input signal, thus providing an output that is a replica of the input. *Figure 4-55* is an example of a class A amplifier. Although the output from this amplifier is 180 degrees out of phase with the input, the output current still flows for the complete duration of the input.

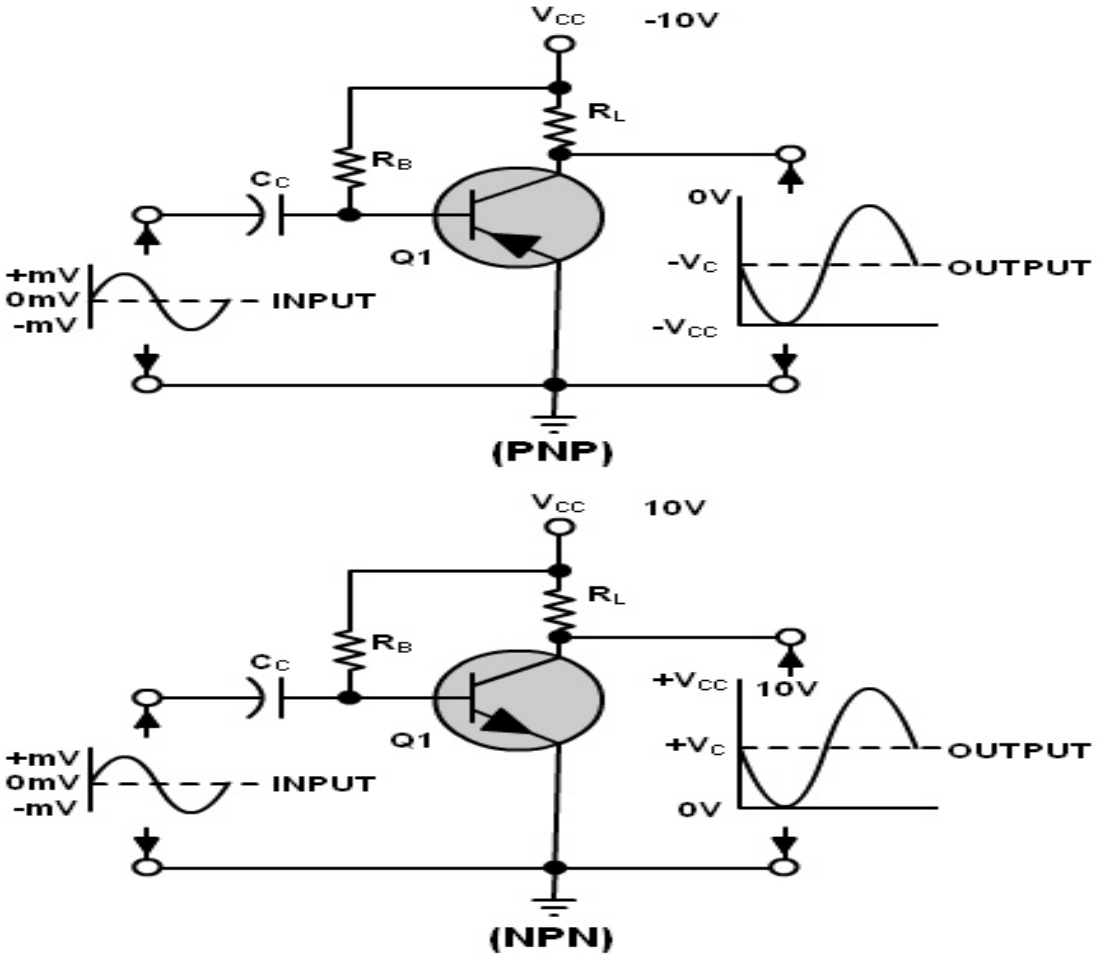


Figure 4-55 – Basic transistor amplifier.

The class A operated amplifier is used as an audio-frequency and radio-frequency amplifier in radio, radar, and sound systems, just to mention a few examples.

For a comparison of output signals for the different amplifier classes of operation, refer to *Figure 4-56* during the following discussion.

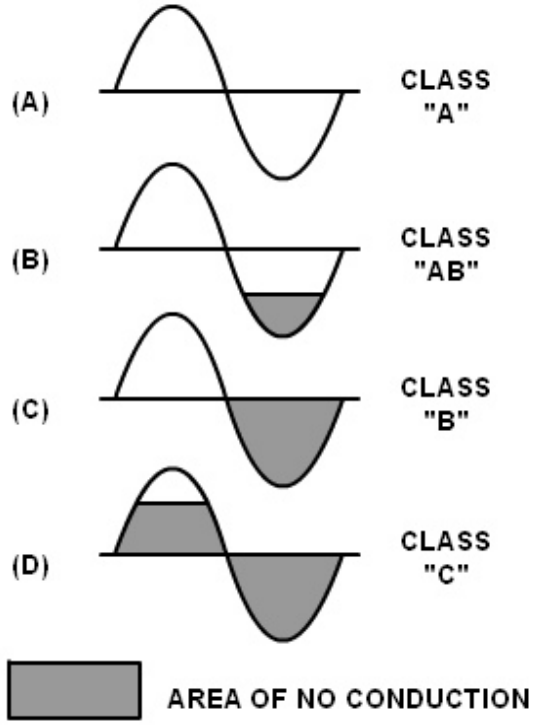


Figure 4-56 — Comparison of output signals for the different amplifier classes of operation.

6.2.1.2 Class AB Amplifier Operation

Amplifiers designed for class AB operation are biased so that collector current is zero (cutoff) for a portion of one alternation of the input signal. This is accomplished by making the forward-bias voltage less than the peak value of the input signal. By doing this, the base-emitter junction will be reverse biased during one alternation for the amount of time that the input signal voltage opposes and exceeds the value of forward-bias voltage; therefore, collector current will flow for more than 180 degrees but less than 360 degrees of the input signal (*Figure 4-56, View B*). As compared to the class A amplifier, the dc operating point for the AB amplifier is closer to cutoff.

The class AB operated amplifier is commonly used as a push-pull amplifier to overcome a side effect of class B operation called crossover distortion.

6.2.1.3 Class B Amplifier Operation

Amplifiers biased so that collector current is cut off during one-half of the input signal are classified class B. The dc operating point for this class of amplifier is set up so that base current is zero with no input signal. When a signal is applied, one half cycle will forward bias the base-emitter junction and I_C will flow. The other half cycle will reverse bias the base-emitter junction and I_C will be cut off. Thus, for class B operation, collector current will flow for approximately 180 degrees (half) of the input signal (*Figure 4-56, View C*).

The class B operated amplifier is used extensively for audio amplifiers that require high-power outputs. It is also used as the driver-amplifier and power-amplifier stages of transmitters.

6.2.1.4 Class C Amplifier Operation

In class C operation, collector current flows for less than one half cycle of the input signal (*Figure 4-56, View D*). The class C operation is achieved by reverse biasing the emitter-base junction, which sets the dc operating point below cutoff and allows only the portion of the input signal that overcomes the reverse bias to cause collector current

flow. The class C operated amplifier is used as a radio-frequency amplifier in transmitters.

From the previous discussion, you can conclude that two primary items determine the class of operation of an amplifier: (1) the amount of bias and (2) the amplitude of the input signal. With a given input signal and bias level, you can change the operation of an amplifier from class A to class B just by removing forward bias. Also, a class A amplifier can be changed to class AB by increasing the input signal amplitude. However, if an input signal amplitude is increased to the point that the transistor goes into saturation and cutoff, it is then called an overdriven amplifier.

You should be familiar with two terms used in conjunction with amplifiers – ***fidelity*** and ***efficiency***. Fidelity is the faithful reproduction of a signal. In other words, if the output of an amplifier is just the same as the input except in amplitude, the amplifier has a high degree of fidelity. The opposite of fidelity is a term we mentioned earlier – distortion. A circuit that has high fidelity has low distortion. In conclusion, a class A amplifier has a high degree of fidelity. A class AB amplifier has less fidelity, and class B and class C amplifiers have low or poor fidelity.

The efficiency of an amplifier refers to the ratio of output-signal power compared to the total input power. An amplifier has two input power sources: one from the signal, and one from the power supply. Because every device takes power to operate, an amplifier that operates for 360 degrees of the input signal uses more power than if operated for 180 degrees of the input signal. By using more power, an amplifier has less power available for the output signal; thus, the efficiency of the amplifier is low. This is the case with the class A amplifier. It operates for 360 degrees of the input signal and requires a relatively large input from the power supply. Even with no input signal, the class A amplifier still uses power from the power supply; therefore, the output from the class A amplifier is relatively small compared to the total input power. This results in low efficiency, which is acceptable in class A amplifiers because they are used where efficiency is not as important as fidelity.

Class AB amplifiers are biased so that collector current is cut off for a portion of one alternation of the input, which results in less total input power than the class A amplifier. This leads to better efficiency.

Class B amplifiers are biased with little or no collector current at the dc operating point. With no input signal, there is little wasted power; therefore, the efficiency of class B amplifiers is higher still.

The efficiency of class C is the highest of the four classes of amplifier operations.

6.2.2 Transistor Configurations

A transistor may be connected in any one of three basic configurations (*Figure 4-57*): (1) common emitter (CE), (2) common base (CB), and (3) common collector (CC). The term common is used to denote the element that is common to both input and output circuits. Because the common element is often grounded, these configurations are frequently referred to as grounded emitter, grounded base, and grounded collector.

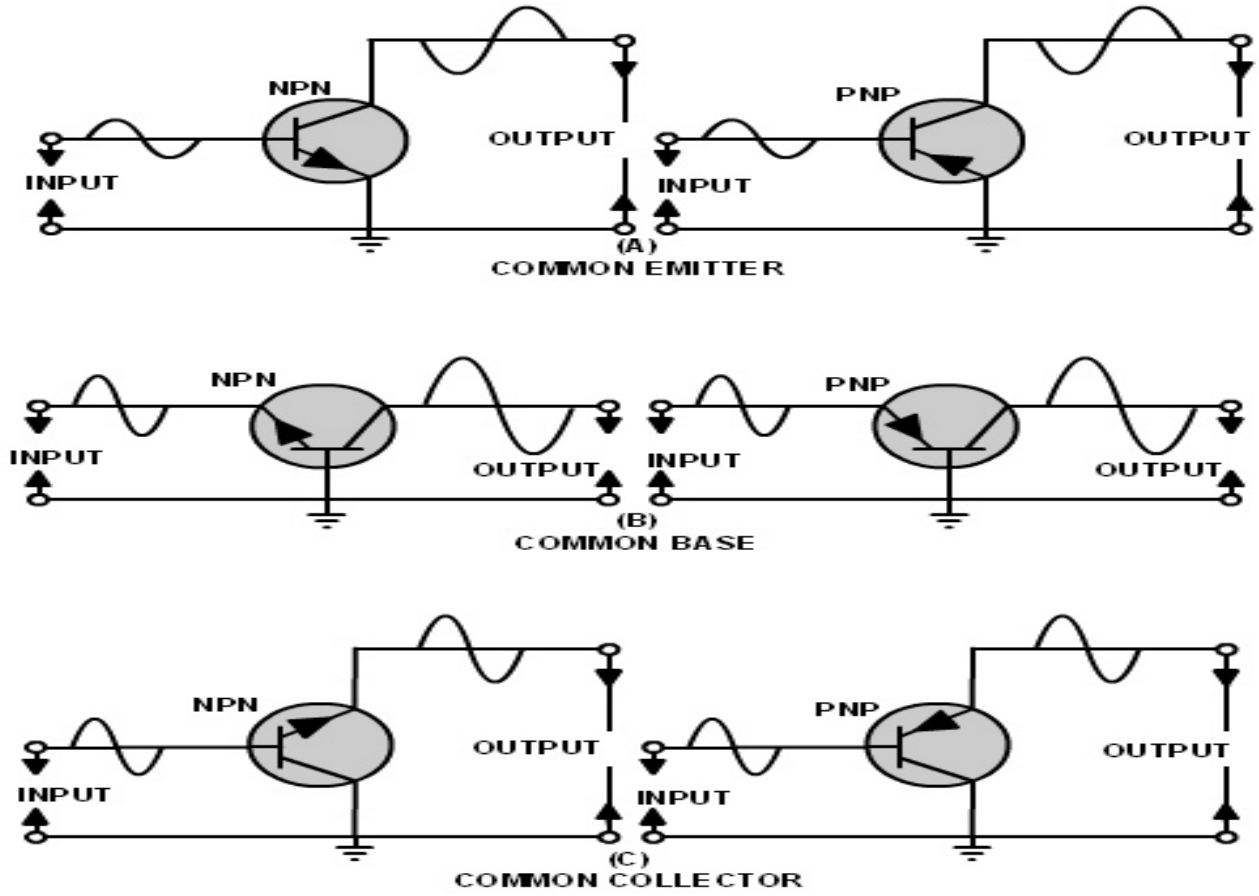


Figure 4-57 – Transistor configurations.

Each configuration, as you will see later, has particular characteristics that make it suitable for specific applications. An easy way to identify a specific transistor configuration is to follow three simple steps:

- Identify the element (emitter, base, or collector) to which the input signal is applied.
- Identify the element (emitter, base, or collector) from which the output signal is taken.
- The remaining element is the common element, and gives the configuration its name.

By applying these three simple steps to the circuit in *Figure 4-55*, you can conclude that this circuit is more than just a basic transistor amplifier; it is a common-emitter amplifier.

6.2.2.1 Common Emitter

The common-emitter configuration (CE) shown in *Figure 4-57, View A* is the arrangement most frequently used in practical amplifier circuits because it provides good voltage, current, and power gain. The common emitter also has a somewhat low input resistance (500 ohms-1500 ohms) because the input is applied to the forward-biased junction, and has a moderately high output resistance (30 kilohms-50 kilohms or more), because the output is taken off the reverse-biased junction. Because the input signal is applied to the base-emitter circuit and the output is taken from the collector-emitter circuit, the emitter is the element common to both input and output.

Because you have already covered what you now know to be a common-emitter amplifier (*Figure 4-55*), take a few minutes and review its operation using the PNP common-emitter configuration shown in *Figure 4-57, View A*.

When a transistor is connected in a common-emitter configuration, the input signal is injected between the base and emitter, which is a low-resistance, low-current circuit. As the input signal swings positive, it also causes the base to swing positive with respect to the emitter. This action decreases forward bias, which reduces collector current (I_C) and increases collector voltage (making V_C more negative). During the negative alternation of the input signal, the base is driven more negative with respect to the emitter. This increases forward bias and allows more current carriers to be released from the emitter, which results in an increase in collector current and a decrease in collector voltage (making V_C less negative or swing in a positive direction). The collector current that flows through the high-resistance, reverse-biased junction also flows through a high-resistance load, resulting in a high level of amplification.

Because the input signal to the common emitter goes positive when the output goes negative, the two signals (input and output) are 180 degrees out of phase. The common-emitter circuit is the only configuration that provides a phase reversal.

The common-emitter is the most popular of the three transistor configurations because it has the best combination of current and voltage gain. The term gain is used to describe the amplification capabilities of the amplifier. It is basically a ratio of output versus input. Each transistor configuration gives a different value of gain, even though the same transistor is used. The transistor configuration used is a matter of design consideration. However, as a technician, you will become interested in this output versus input ratio (gain) to determine whether or not the transistor is working properly in the circuit.

The current gain in the common-emitter circuit is called ***beta*** (*β*). Beta is the relationship of collector current (output current) to base current (input current). To calculate beta, use the following formula:

$$\beta = \frac{\Delta I_C}{\Delta I_B}$$ (*Δ* is the Greek letter delta; it is used to indicate a small change)

For example, if the input current (I_B) in a common emitter changes from 75 uA to 100 uA and the output current (I_C) changes from 1.5 mA to 2.6 mA, the current gain (*β*) will be 44:

$$\beta = \frac{\Delta I_C}{\Delta I_B} = \frac{11 \times 10^{-3}}{25 \times 10^{-6}} = 44$$

This simply means that a change in base current produces a change in collector current which is 44 times as large.

You may also see the term h_{fc} used in place of *β*. The terms h_{fc} and β are equivalent and may be used interchangeably. This is because h_{fc} means:

h = hybrid (meaning mixture)

f = forward current transfer ratio

e = common emitter configuration

The resistance gain of the common emitter can be found in a method similar to the one used for finding beta:

$$R = \frac{R_{out}}{R_{in}}$$

Once the resistance gain is known, the voltage gain is easy to calculate since it is equal to the current gain (β) multiplied by the resistance gain ($E = \beta R$). Also, the power gain is equal to the voltage gain multiplied by the current gain β ($P = \beta E$).

6.2.2.2 Common Base

The common-base configuration (CB) shown in *Figure 4-57, View B* is mainly used for impedance matching, because it has a low input resistance (30 ohms-160 ohms) and a high output resistance (250 kilohms-550 kilohms). However, two factors limit its usefulness in some circuit applications: (1) its low input resistance and (2) its current gain of less than 1. Because the CB configuration will give voltage amplification, there are some additional applications, which require both a low-input resistance and voltage amplification that could use a circuit configuration of this type, for example, some microphone amplifiers.

In the common-base configuration, the input signal is applied to the emitter, the output is taken from the collector, and the base is the element common to both input and output. Since the input is applied to the emitter, it causes the emitter-base junction to react in the same manner as it did in the common-emitter circuit. For example, an input that aids the bias will increase transistor current, and one that opposes the bias will decrease transistor current.

Unlike the common-emitter circuit, the input and output signals in the common-base circuit are in phase. To illustrate this point, assume the input to the PNP version of the common-base circuit in *Figure 4-57, View B* is positive. The signal adds to the forward bias, because it is applied to the emitter, causing the collector current to increase. This increase in I_C results in a greater voltage drop across the load resistor R_L, thus lowering the collector voltage V_C. The collector voltage, in becoming less negative, is swinging in a positive direction, and is therefore in phase with the incoming positive signal.

The current gain in the common-base circuit is calculated in a method similar to that of the common emitter except that the input current is I_E not I_B and the term alpha (α) is used in place of beta for gain. Alpha is the relationship of collector current (output current) to emitter current (input current). Alpha is calculated using the formula:

$$\alpha = \frac{\Delta I_C}{\Delta I_E}$$

For example, if the input current (I_E) in a common base changes from 1 mA to 3 mA and the output current (I_C) changes from 1 mA to 2.8 mA, the current gain (α) will be 0.90 or:

$$\alpha = \frac{\Delta I_C}{\Delta I_E} = \frac{18 \times 10^{-3}}{2 \times 10^{-3}} = 0.90$$ This is a current gain of less than 1.

Since part of the emitter current flows into the base and does not appear as collector current, collector current will always be less than the emitter current that causes it. (Remember that $I_E = I_B + I_C$) Therefore, alpha is always less than one for a common-base configuration.

Another term for α is h_f. These terms (and h_f) are equivalent and may be used interchangeably. The meaning for the term h_f is derived in the same manner as the term h_{fe} mentioned earlier, except that the last letter *e* has been replaced with *b* to stand for common-base configuration.

Many transistor manuals and data sheets only list transistor current gain characteristics in terms of β or h_{fe}. To find alpha (α) when given beta (β), use the following formula to convert β to α for use with the common-base configuration:

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

To calculate the other gains (voltage and power) in the common-base configuration when the current gain (α) is known follow the procedures described earlier under the common-emitter section.

6.2.2.3 Common Collector

The common-collector configuration (CC) shown in *Figure 4-57, View C* is used mostly for impedance matching. It is also used as a current driver because of its substantial current gain. It is particularly useful in switching circuitry because it has the ability to pass signals in either direction (bilateral operation).

In the common-collector circuit, the input signal is applied to the base, the output is taken from the emitter, and the collector is the element common to both input and output. The common collector is equivalent to our old friend, the electron-tube cathode follower. Both have high input and low output resistance. The input resistance for the common collector ranges from 2 kilohms to 500 kilohms, and the output resistance varies from 50 ohms to 1500 ohms. The current gain is higher than that in the common emitter, but it has a lower power gain than either the common base or common emitter. Similar to the common base, the output signal from the common collector is in phase with the input signal. The common collector is also referred to as an emitter-follower because the output developed on the emitter follows the input signal applied to the base.

Transistor action in the common collector is similar to the operation explained for the common base, except that the current gain is not based on the emitter-to-collector current ratio, alpha (α). Instead, it is based on the emitter-to-base current ratio called ***gamma*** (γ), because the output is taken off the emitter. Since a small change in base current controls a large change in emitter current, it is still possible to obtain high current gain in the common collector. However, since the emitter current gain is offset by the low output resistance, the voltage gain is always less than 1 (unity), exactly as in the electron-tube cathode follower. The common-collector current gain, gamma (γ), is defined as

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

and is related to collector-to-base current gain, beta (β), of the common-emitter circuit by the formula:

$$\gamma = \beta + 1$$

Since a given transistor may be connected in any of three basic configurations, there is a definite relationship, as pointed out earlier, between alpha (α), beta (β), and gamma (γ). These relationships are listed again for your convenience:

$$\alpha = \frac{\beta}{\beta + 1} \quad \beta = \frac{\alpha}{1 - \alpha} \quad \gamma = \beta + 1$$

Take, for example, a transistor that is listed on a manufacturer's data sheet as having an alpha of 0.90. We wish to use it in a common emitter configuration. This means we must find beta. The calculations are:

$$\beta = \frac{\alpha}{1 - \alpha} = \frac{0.90}{1 - 0.90} = \frac{0.90}{0.1} = 9$$

A change in base current in this transistor will thus produce a change in collector current that will be 9 times as large.

If you wish to use this same transistor in a common collector, you can find gamma (γ) by:

$$\gamma = \beta + 1 = 9 + 1 = 10$$

To summarize the properties of the three transistor configurations, a comparison chart is provided in *Table 4-2* for your convenience.

Table 4-2 – Transistor Configuration Comparison Chart

Amplifier Type	Common Base	Common Emitter	Common Collector
Input/Output Phase Relationship	0°	180°	0°
Voltage Gain	High	Medium	Low
Current Gain	Low (α)	Medium (β)	High (γ)
Power Gain	Low	High	Medium
Input Resistance	Low	Medium	High
Output Resistance	High	Medium	Low

Now that you have analyzed the basic transistor amplifier in terms of bias, class of operation, and circuit configuration, you will next apply what has been covered to *Figure 4-55*.

Figure 4-55 is not just a basic transistor amplifier, but a class A amplifier configured as a common emitter using fixed bias. From this, you should be able to conclude the following:

- Because of its fixed bias, the amplifier is thermally unstable.
- Because of its class operation, the amplifier has low efficiency but good fidelity.
- Because it is configured as a common emitter, the amplifier has good voltage, current, and power gain.

In conclusion, the type of bias, class of operation, and circuit configuration are all clues to the function and possible application of the amplifier.

6.3.0 Transistor Application

Transistors are frequently used as amplifiers. Some transistor circuits are current amplifiers with a small load resistance; other circuits are designed for voltage amplification and have a high load resistance; others amplify power.

As discussed earlier in the last section, transistors are broken down into three classes and their applications are as follows:

- Class A - The class A operated amplifier is used as an audio-frequency and radio-frequency amplifier in radio, radar, and sound systems, just to mention a few examples.
- Class AB - The class AB operated amplifier is commonly used as a push-pull amplifier to overcome a side effect of class B operation called crossover distortion.
- Class B - The class B operated amplifier is used extensively for audio amplifiers that require high-power outputs. It is also used as the driver-amplifier and power-amplifier stages of transmitters.
- Class C - The class C operated amplifier is used as a radio-frequency amplifier in transmitters.

6.4.0 Transistor Identification

Transistors are available in a large variety of shapes and sizes, each with its own unique characteristics. The characteristics for each of these transistors are usually presented on specification sheets or they may be included in transistor manuals. Although many properties of a transistor could be specified on these sheets, manufacturers list only some of them. The specifications listed vary with different manufacturers, the type of transistor, and the application of the transistor. The specifications usually cover:

1. A general description of the transistor that includes the following information:
    a. The kind of transistor, to include: the material used, such as germanium or silicon; the type of transistor (NPN or PNP); and the construction of the transistor (whether alloy-junction, grown, diffused junction, and so forth);
    b. Some of the common applications for the transistor, such as audio amplifier, oscillator, rf amplifier, and so forth; and
    c. General sales features, such as size and packaging mechanical data.
2. The Absolute Maximum Ratings of the transistor, which are the direct voltage and current values that, if exceeded in operation, may result in transistor failure. Maximum ratings usually include collector-to-base voltage, emitter-to-base voltage, collector current, emitter current, and collector power dissipation.
3. The typical operating values of the transistor. These values are presented only as a guide. The values vary widely, are dependent upon operating voltages, and also upon which element is common in the circuit. The values listed may include collector-emitter voltage, collector current, input resistance, load resistance, current-transfer ratio (another name for alpha or beta), and collector cutoff current, which is leakage current from collector to base when no emitter current is applied. Transistor characteristic curves may also be included in this section. A transistor characteristic curve is a graph that plots the relationship between

currents and voltages in a circuit. More than one curve on a graph is called a "family of curves."

4. Additional information for engineering-design purposes.

Transistors can be identified by a Joint Army-Navy (JAN) designation printed directly on the case of the transistor. The marking scheme explained earlier for diodes is also used for transistor identification. The first number indicates the number of junctions. The letter (N) following the first number tells you that the component is a semiconductor. The 2-digit or 3-digit number following the N is the manufacturer's identification number. If the last number is followed by a letter, it indicates a later, improved version of the device. For example, a semiconductor designated as type 2N130A signifies a three-element transistor of semiconductor material that is an improved version of type 130:

2	N	130	A
Number of Junctions (Transistor)	Semi-Conductor	Identification Number	First Modification

You may also find other markings on transistors that do not relate to the JAN marking system. These markings are manufacturers' identifications and may not conform to a standardized system. If in doubt, always replace a transistor with one having identical markings. To ensure that an identical replacement or a correct substitute is used, consult an equipment or transistor manual for specifications on the transistor.

6.4.1 Lead Identification

Transistor lead identification plays an important part in transistor maintenance; because, before a transistor can be tested or replaced, its leads or terminals must be identified. Because there is no standard method of identifying transistor leads, it is quite possible to mistake one lead for another; therefore, when you are replacing a transistor, you should pay close attention to how the transistor is mounted, particularly to those transistors that are soldered in, so that you do not make a mistake when you are installing the new transistor. When you are testing or replacing a transistor, if you have any doubts about which lead is which, consult the equipment manual or a transistor manual that shows the specifications for the transistor being used.

There are, however, some typical lead identification schemes that will be very helpful in transistor troubleshooting. These schemes are shown in *Figure 4-58*. In the case of the oval-shaped transistor shown in *Figure 4-58, View A*, the collector lead is identified by a wide space between it and the base lead. The lead farthest from the collector in line is the emitter lead. When the leads are evenly spaced and in line, as shown in *Figure 4-58, View B*, a colored dot, usually red, indicates the collector. If the transistor is round, as in *Figure 4-58, View C*, a red line indicates the collector, and the remitter lead is the shortest lead. In *Figure 4-58, View D,* the leads are in a triangular arrangement that is offset from the center of the transistor. The lead opposite the blank quadrant in this scheme is the base lead. When viewed from the bottom, the collector is the first lead clockwise from the base. The leads in *Figure 4-58, View E* are arranged in the same manner as those in *Figure 4-58, View D* except that a tap is used to identify the leads. When viewed from the bottom in a clockwise direction, the first lead following the tab is the emitter, followed by the base and collector.

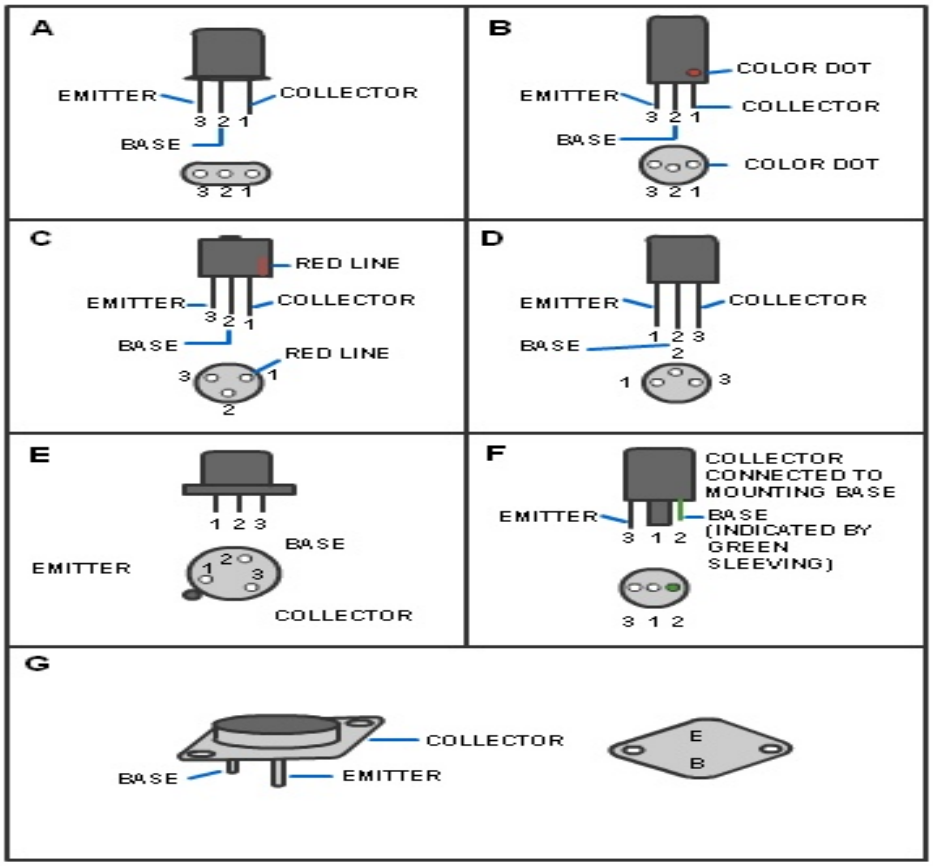


Figure 4-58 – Transistor lead identification.

In a conventional power transistor, as shown in *Figure 4-58, Views F* and *G*, the collector lead is usually connected to the mounting base. For further identification, the base lead in *Figure 4-58, View F* is covered with green sleeving. While the leads in *Figure 4-58, View G* are identified by viewing the transistor from the bottom in a clockwise direction (with mounting holes occupying 3 o'clock and 9 o'clock positions), the emitter lead will be either at the 5 o'clock or 11 o'clock position. The other lead is the base lead.

6.5.0 Transistor Schematic Symbol

The two basic types of transistors along with their circuit symbols are shown in *Figure 4-59*. It should be noted that the two symbols are different. The horizontal line represents the base, the angular line with the arrow on it represents the emitter, and the other angular line represents the collector. The direction of the arrow on the emitter distinguishes the NPN from the PNP transistor. If the arrow points in (Points iN), the transistor is a PNP. On the other hand if the arrow points out, the transistor is an NPN (Not Pointing iN).

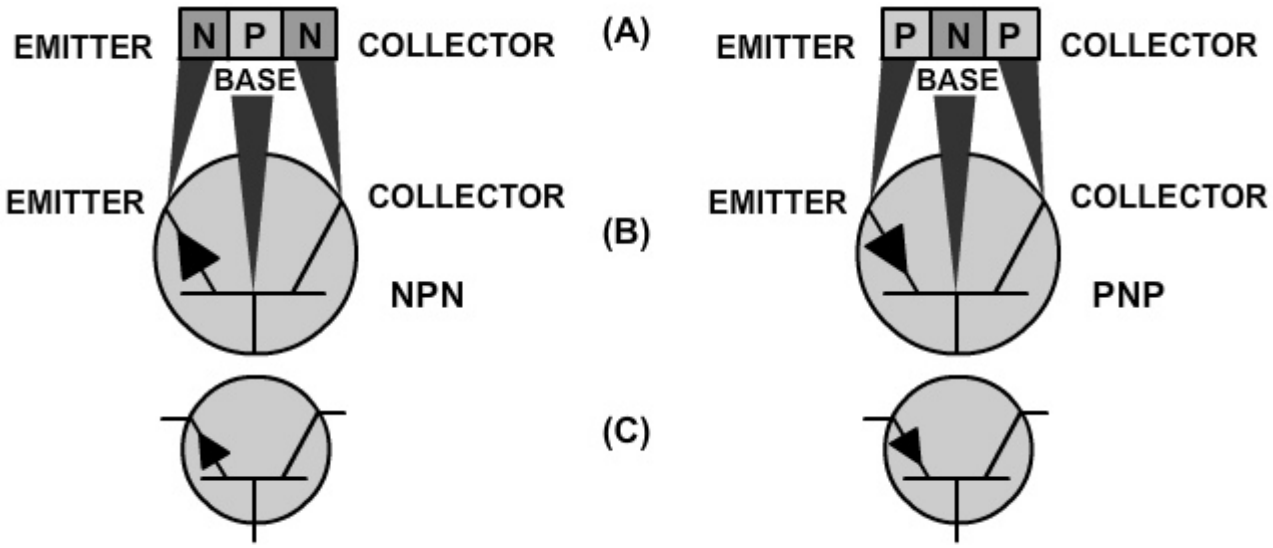


Figure 4-59 – Transistor representations.

Another point you should keep in mind is that the arrow always points in the direction of hole flow, or from the P to N sections, no matter whether the P section is the emitter or base. On the other hand, electron flow is always toward or against the arrow, just like in the junction diode.

6.6.0 Transistor Testing

There are several different ways of testing transistors. They can be tested while in the circuit, by the substitution method mentioned, or with a transistor tester or ohmmeter.

Transistor testers are nothing more than the solid-state equivalent of electron-tube testers (although they do not operate on the same principle). With most transistor testers, it is possible to test the transistor in or out of the circuit.

There are four basic tests required for transistors in practical troubleshooting: gain; leakage; breakdown; and switching time. For maintenance and repair, however, a check of two or three parameters is usually sufficient to determine whether a transistor needs to be replaced.

Since it is impractical to cover all the different types of transistor testers and since each tester comes with its own operator's manual, only the most frequently used testing device, the ohmmeter, will be covered here.

6.6.1 Testing Transistors with an Ohmmeter

Two tests that can be done with an ohmmeter are gain and junction resistance. Tests of a transistor's junction resistance will reveal leakage, shorts, and opens.

6.6.1.1 Transistor Gain Test

A basic transistor gain test can be made using an ohmmeter and a simple test circuit. The test circuit can be made with just a couple of resistors and a switch, as shown in *Figure 4-60*. The principle behind the test lies in the fact that little or no current will flow in a transistor between emitter and collector until the emitter-base junction is forward biased. The only precaution you should observe is with the ohmmeter. Any internal battery may be used in the meter provided that it does not exceed the maximum collector-emitter breakdown voltage.

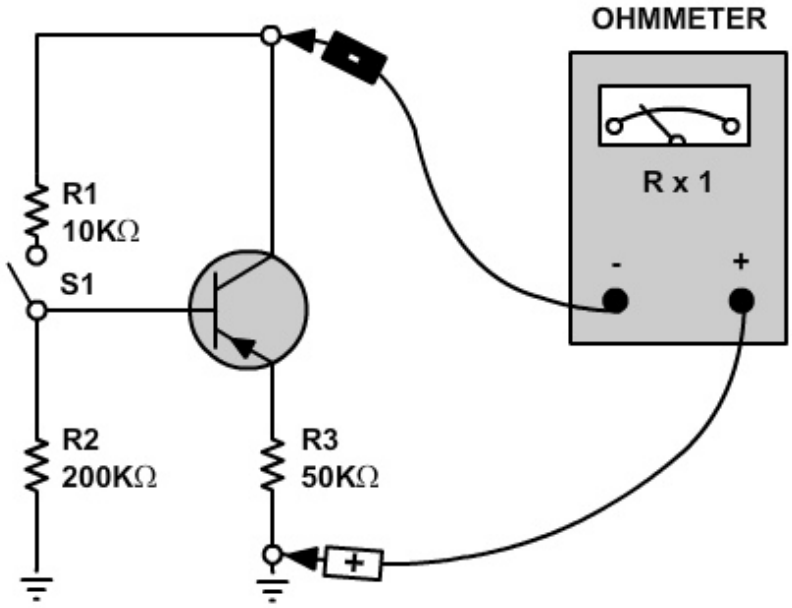


Figure 4-60 — Testing a transistor's gain with an ohmmeter.

With the switch in *Figure 4-60* in the open position as shown, no voltage is applied to the PNP transistor's base, and the emitter-base junction is not forward biased; therefore, the ohmmeter should read a high resistance, as indicated on the meter. When the switch is closed, the emitter-base circuit is forward biased by the voltage across R1 and R2. Current now flows in the emitter-collector circuit, which causes a lower resistance reading on the ohmmeter. A 10-to-1 resistance ratio in this test between meter readings indicates a normal gain for an audio-frequency transistor.

To test an NPN transistor using this circuit, simply reverse the ohmmeter leads and carry out the procedure described earlier.

6.6.1.2 Transistor Junction Resistance Test

An ohmmeter can be used to test a transistor for leakage (an undesirable flow of current) by measuring the base-emitter, base-collector, and collector-emitter forward and reverse resistances.

For simplicity, consider the transistor under test in each view of *Figure 4-61* (*View A, View B, and View C*) as two diodes connected back to back. Each diode will have a low

forward resistance and a high reverse resistance. By measuring these resistances with an ohmmeter as shown in *Figure 4-61*, you can determine if the transistor is leaking current through its junctions. When making these measurements, avoid using the R1 scale on the meter or a meter with a high internal battery voltage. Either of these conditions can damage a low-power transistor.

Now consider the possible transistor problems that could exist if the indicated readings in *Figure 4-61* are not obtained. A list of these problems is provided in *Table 4-3*.

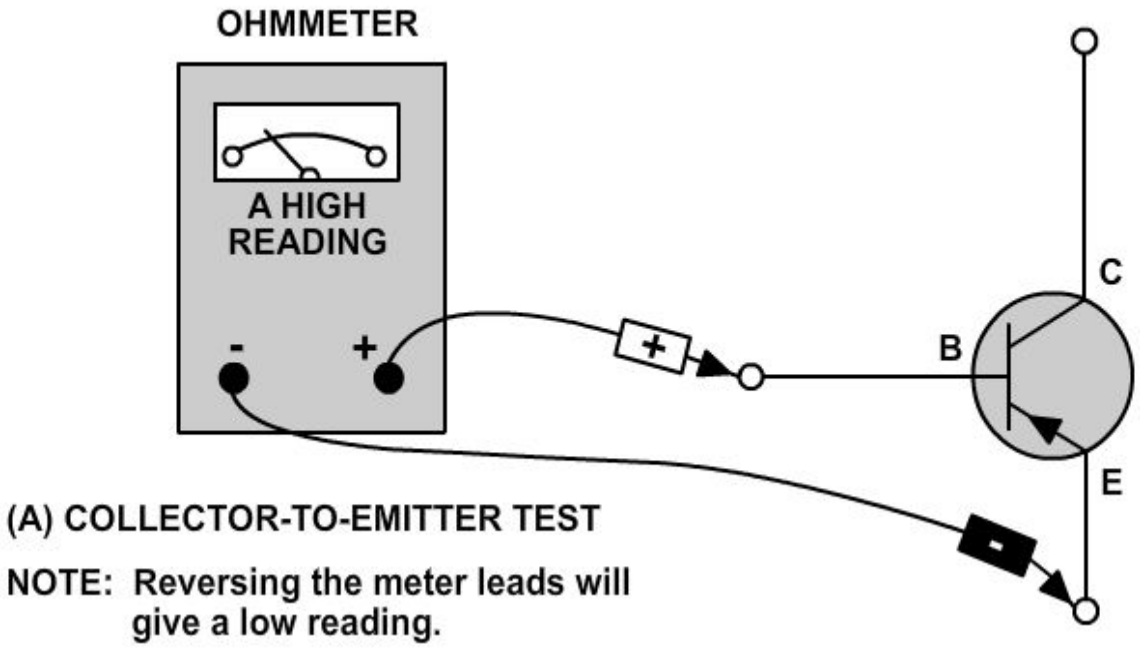


Figure 4-61 — Testing a transistor's leakage with an ohmmeter.

Table 4-3 – Possible Transistor Problems from Ohmmeter Readings.

Resistance Readings		Problems
Forward	**Reverse**	**The transistor is:**
Low (Not Shorted)	Low (Not Shorted)	Leaking
Low (Shorted)	Low (Shorted)	Shorted
High	High	Open*
*Except collector-to-emitter test.		

By now, you should recognize that the transistor used in *Figure 4-61, View A, B,* and *C* is a PNP transistor. If you wish to test an NPN transistor for leakage, the procedure is identical to that used for testing the PNP except the readings obtained are reversed.

When testing transistors (PNP or NPN), you should remember that the actual resistance values depend on the ohmmeter scale and the battery voltage. Typical forward and reverse resistances are insignificant. The best indicator for showing whether a transistor is good or bad is the ratio of forward-to-reverse resistance. If the transistor you are testing shows a ratio of at least 30 to 1, it is probably good. Many transistors show ratios of 100 to 1 or greater.

7.0.0 ZENER DIODE

When a PN-junction diode is reverse-biased, the majority carriers (holes in the P-material and electrons in the N-material) move away from the junction. The barrier or depletion region becomes wider and majority carrier current flow becomes very difficult across the high resistance of the wide depletion region (*Figure 4-62, Views A, B,* and *C*). The presence of minority carriers causes a small leakage current that remains nearly constant for all reverse voltages up to a certain value. Once this value has been exceeded, there is a sudden increase in the reverse current. The voltage at which the sudden increase in current occurs is called the ***breakdown*** voltage. At breakdown, the reverse current increases very rapidly with a slight increase in the reverse voltage. Any diode can be reverse biased to the point of breakdown, but not every diode can safely dissipate the power associated with breakdown. A Zener diode is a PN junction designed to operate in the reverse-bias breakdown region.

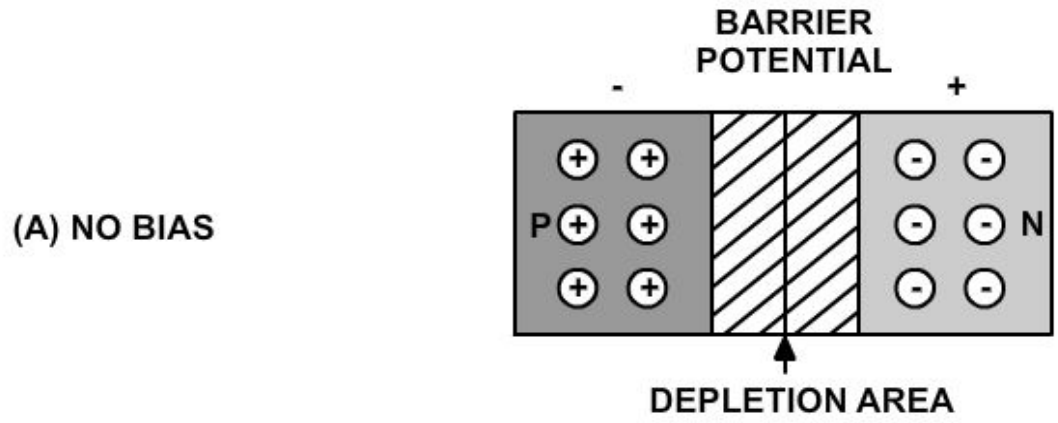


Figure 4-62 – Effects of bias on the depletion region of a PN junction.

There are two distinct theories used to explain the behavior of PN junctions during breakdown: one is the ***zener effect*** and the other is the ***avalanche effect***.

The zener effect was first proposed by Dr. Carl Zener in 1934. According to Dr. Zener's theory, electrical breakdown in solid dielectrics occurs by a process called ***quantum-mechanical tunneling***. The zener effect accounts for the breakdown below 5 volts, whereas above 5 volts, the breakdown is caused by the avalanche effect. Although the avalanche effect is now accepted as an explanation of diode breakdown, the term zener diode is used to cover both types.

The true zener effect in semiconductors can be described in terms of energy bands; however, only the two upper energy bands are of interest. The two upper bands, illustrated in *Figure 4-63, View A*, are called the conduction band and the valence band.

The conduction band is a band in which the energy level of the electrons is high enough that the electrons will move easily under the influence of an external field. Since current flow is the movement of electrons, the readily mobile electrons in the conduction band are capable of maintaining a current flow when an external field in the form of a voltage is applied; therefore, solid materials that have many electrons in the conduction band are called conductors.

The valence band is a band in which the energy level is the same as the valence electrons of the atoms. Since the electrons in these levels are attached to the atoms, the electrons are not free to move around as are the conduction band electrons. With the proper amount of energy added, however, the electrons in the valence band may be elevated to the conduction band energy level. To do this, the electrons must cross a gap that exists between the valence band energy level and the conduction band energy level. This gap is known as the forbidden energy band or forbidden gap. The energy difference across this gap determines whether a solid material will act as a conductor, a semiconductor, or an insulator.

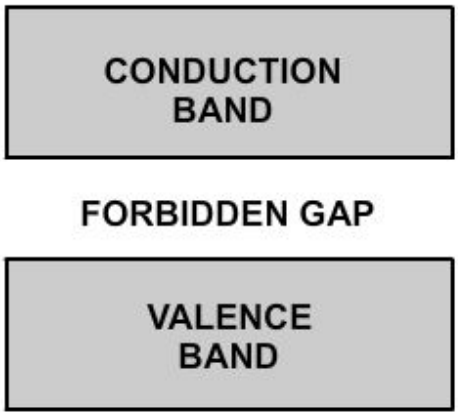


Figure 4-63 — Energy diagram for zener diode.

A conductor is a material in which the forbidden gap is so narrow that it can be considered nonexistent. A semiconductor is a solid that contains a forbidden gap (*Figure 4-63, View A*). Normally, a semiconductor has no electrons at the conduction band energy level. The energy provided by room temperature heat, however, is enough energy to overcome the binding force of a few valence electrons and to elevate them to the conduction band energy level. The addition of impurities to the semiconductor material increases both the number of free electrons in the conduction band and the number of electrons in the valence band that can be elevated to the conduction band. Insulators are materials in which the forbidden gap is so large that practically no electrons can be given enough energy to cross the gap; therefore, unless extremely large amounts of heat energy are available, these materials will not conduct electricity.

Figure 4-63, View B is an energy diagram of a reverse-biased Zener diode. The energy bands of the P- and N-materials are naturally at different levels, but reverse bias causes the valence band of the P-material to overlap the energy level of the conduction band in the N material. Under this condition, the valence electrons of the P-material can cross the extremely thin junction region at the overlap point without acquiring any additional energy. This action is called tunneling. When the breakdown point of the PN junction is reached, large numbers of minority carriers tunnel across the junction to form the current that occurs at breakdown. The tunneling phenomenon only takes place in heavily doped diodes, such as Zener diodes.

The second theory of reverse breakdown effect in diodes is known as avalanche breakdown and occurs at reverse voltages beyond 5 volts. This type of breakdown diode has a depletion region that is deliberately made narrower than the depletion region in the normal PN-junction diode, but thicker than that in the zener-effect diode. The thicker depletion region is achieved by decreasing the doping level from the level

used in zener-effect diodes. The breakdown is at a higher voltage because of the higher resistivity of the material. Controlling the doping level of the material during the manufacturing process can produce breakdown voltages ranging between about 2 and 200 volts.

The mechanism of avalanche breakdown is different from that of the zener effect. In the depletion region of a PN junction, thermal energy is responsible for the formation of electron-hole pairs. The leakage current is caused by the movement of minority electrons, which is accelerated in the electric field across the barrier region. As the reverse voltage across the depletion region is increased, the reverse voltage eventually reaches a critical value. Once the critical or breakdown voltage has been reached, sufficient energy is gained by the thermally released minority electrons to enable the electrons to rupture covalent bonds as they collide with lattice atoms. The released electrons are also accelerated by the electric field, resulting in the release of further electrons, and so on, in a chain or avalanche effect. This process is illustrated in *Figure 4-64*.

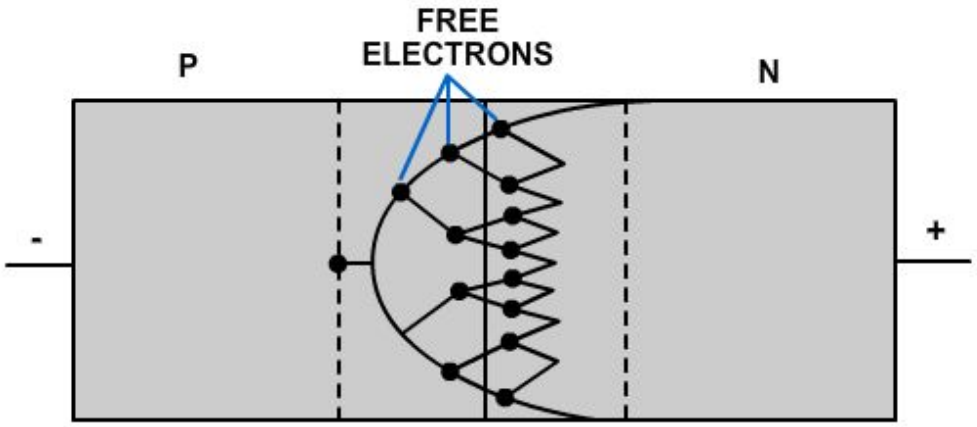


Figure 4-64 – Avalanche multiplication.

For reverse voltage slightly higher than breakdown, the avalanche effect releases an almost unlimited number of carriers so that the diode essentially becomes a short circuit. The current flow in this region is limited only by an external series current-limiting resistor. Operating a diode in the breakdown region does not damage it, as long as the maximum power dissipation rating of the diode is not exceeded. Removing the reverse voltage permits all carriers to return to their normal energy values and velocities.

Some of the symbols used to represent zener diodes are illustrated in *Figure 4-65*, *Views A, B, C, D,* and *E*. Note that the polarity markings that indicate electron flow is with the arrow symbol instead of against it as in a normal PN-junction diode. This is because breakdown diodes are operated in

Figure 4-65 – Schematic symbols for zener diodes.

the reverse-bias mode, which means the current flow is by minority current carriers.

Zener diodes of various sorts are used for many purposes, but their most widespread use is as voltage regulators. Once the breakdown voltage of a Zener diode is reached, the voltage across the diode remains almost constant regardless of the supply voltage; therefore, they hold the voltage across the load at a constant level. This characteristic makes Zener diodes ideal voltage regulators, and they are found in almost all solid-state circuits in this capacity.

8.0.0 SILICON CONTROLLED RECTIFIER (SCR)

The ***silicon controlled rectifier***, usually referred to as an SCR, is one of the family of semiconductors that includes transistors and diodes. A drawing of an SCR and its schematic representation is shown in *Figure 4-66, Views A* and *B*. Not all SCRs use the casing shown, but this is typical of most of the high-power units.

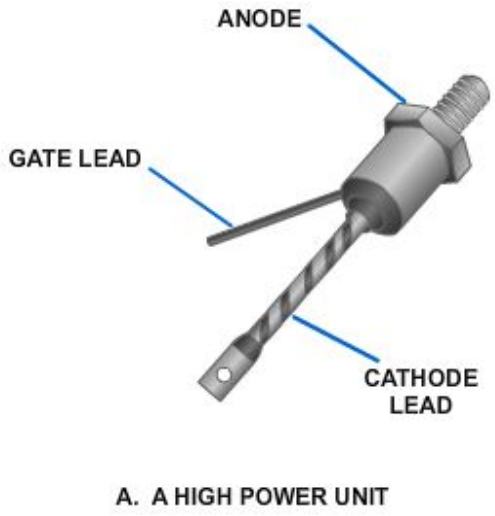


Figure 4-66 – Silicon controlled rectifier.

Although it is not the same as either a diode or a transistor, the SCR combines features of both. Circuits using transistors or rectifier diodes may be greatly improved in some instances through the use of SCRs.

The basic purpose of the SCR is to function as a switch that can turn on or off small or large amounts of power. It performs this function with no moving parts that wear out and no points that require replacing. There can be a tremendous power gain in the SCR. In some units a very small triggering current is able to switch several hundred amperes without exceeding its rated abilities. The SCR can often replace much slower and larger mechanical switches. It even has many advantages over its more complex and larger electron tube equivalent, the thyratron.

The SCR is an extremely fast switch. It is difficult to cycle a mechanical switch several hundred times a minute, yet some SCRs can be switched 25,000 times a second. It takes just microseconds (millionths of a second) to turn on or off these units. Varying the time that a switch is on as compared to the time that it is off regulates the amount of power flowing through the switch. Since most devices can operate on pulses of power (alternating current is a special form of alternating positive and negative pulse), the SCR can be used readily in control applications. Motor-speed controllers, inverters, remote switching units, controlled rectifiers, circuit overload protectors, latching relays, and computer logic circuits all use the SCR.

The SCR is made up of four layers of semiconductor material arranged PNPN. The construction is shown in *Figure 4-67, View A*. In function, the SCR has much in common with a diode, but the theory of operation of the SCR is best explained in terms of transistors.

Consider the SCR as a transistor pair, one PNP and the other NPN, connected as shown in *Figure 4-67, Views B* and *C*. The anode is attached to the upper P-layer; the cathode, C, is part of the lower N-layer; and the gate terminal, G, goes to the P-layer of the NPN triode.

In operation, the collector of Q2 drives the base of Q1, while the collector of Q1 feeds back to the base of Q2. Beta 1 ($\beta 1$) is the current gain of Q1, and Beta 2 ($\beta 2$) is the current gain of Q2. The gain of this positive feedback loop is their product, 1 times 2. When the product is less than one, the circuit is stable; if the product is greater than unity, the circuit is regenerative. A small negative current applied to terminal G will bias the NPN transistor into cutoff, and the loop gain is less than unity. Under these conditions, the only current that can exist between output terminals A and C is very high.

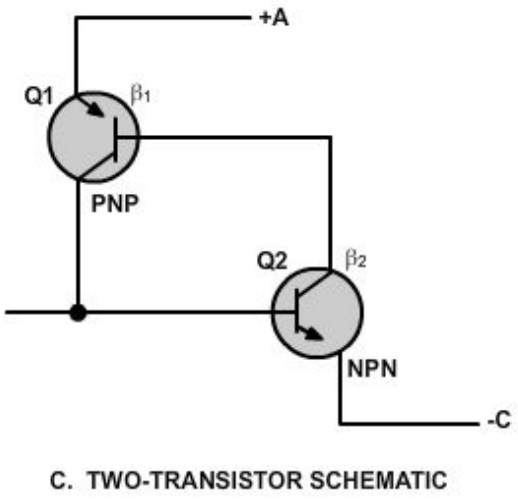


Figure 4-67 – SCR structure.

When a positive current is applied to terminal G, transistor Q2 is biased into conduction, causing its collector current to rise. Since the current gain of Q2 increases with increased collector current, a point (called the breakover point) is reached where the loop gain equals unity and the circuit becomes regenerative. At this point, collector current of the two transistors rapidly increases to a value limited only by the external circuit. Both transistors are driven into saturation, and the impedance between A and C is very low. The positive current applied to terminal G, which served to trigger the self-regenerative action, is no longer required since the collector of PNP transistor Q1 now supplies more than enough current to drive Q2. The circuit will remain on until it is turned off by a reduction in the collector current to a value below that necessary to

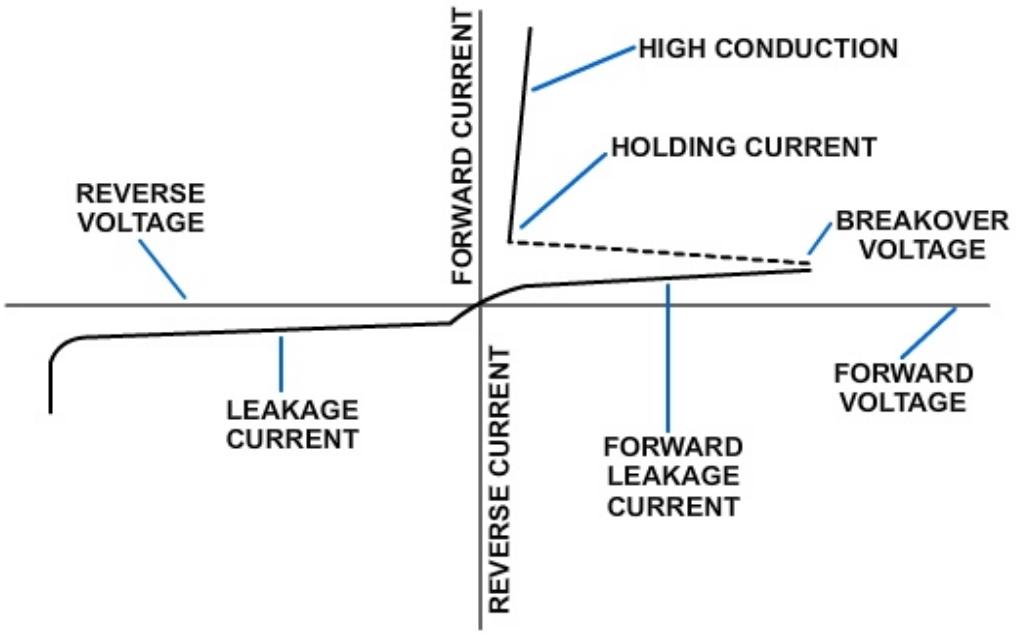


Figure 4-68 — Characteristic curve for an SCR.

maintain conduction.

The characteristic curve for the SCR is shown in *Figure 4-68*. With no gate current, the leakage current remains very small as the forward voltage from cathode to anode is increased until the breakdown point is reached. Here the center junction breaks down, the SCR begins to conduct heavily, and the drop across the SCR becomes very low.

The effect of a gate signal on the firing of an SCR is shown in *Figure 4-69*. Breakdown of the center junction can be achieved at speeds approaching a microsecond by applying an appropriate signal to the gate lead while holding the anode voltage constant. After breakdown, the voltage across the device is so low that the current through it from cathode to anode is essentially determined by the load it is feeding.

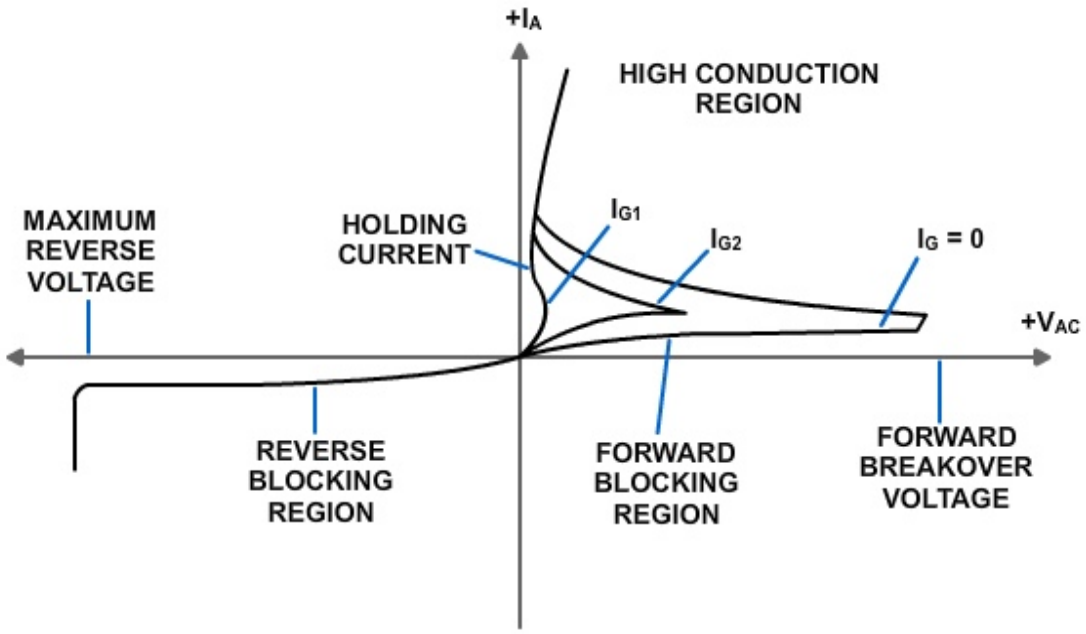


Figure 4-69 — SCR characteristic curve with various gate signals.

The important thing to remember is that a small current from gate to cathode can fire or trigger the SCR, changing it from practically an open circuit to a short circuit. The only way to change it back again (to commutate it) is to reduce the load current to a value less than the minimum forward-bias current. Gate current is required only until the anode current has completely built up to a point sufficient to sustain conduction (about 5 microseconds in resistive-load circuits). After conduction from cathode to anode begins, removing the gate current has no effect.

The basic operation of the SCR can be compared to that of the thyratron. The thyratron is an electron tube, normally gas filled, that uses a filament or a heater. The SCR and the thyratron function in a very similar manner. *Figure 4-70* shows the schematic of each with the corresponding elements labeled. In both types of devices, control by the input signal is lost after they are triggered. The control grid (thyratron) and the gate (SCR) have no further effect on the magnitude of the load current after conduction begins. The load current can be interrupted by one or more of three methods: (1) the load circuit must be opened by a switch, (2) the plate (anode) voltage must be reduced below the ionizing potential of the gas (thyratron), or (3) the forward-bias current must be reduced below a minimum value required to sustain conduction (SCR). The input resistance of the SCR is relatively low (approximately 100 ohms) and requires a current for triggering; the input resistance of the thyratron is exceptionally high and requires a voltage input to the grid for triggering action.

The applications of the SCR as a rectifier are many. In fact, its many applications as a rectifier give this semiconductor device its name. When alternating current is applied to a rectifier, only the positive or negative halves of the sine wave flow through. All of each positive or negative half cycle appears in the output. When an SCR is used, however, the controlled rectifier may be turned on at any time during the half cycle, thus controlling the amount of dc power available from zero to maximum, as shown in *Figure 4-71*. Since the output is actually dc pulses, suitable filtering can be added if continuous direct current is needed. Thus, any dc-operated device can have controlled amounts of power applied to it. Notice that the SCR must be turned on at the desired time for each cycle.

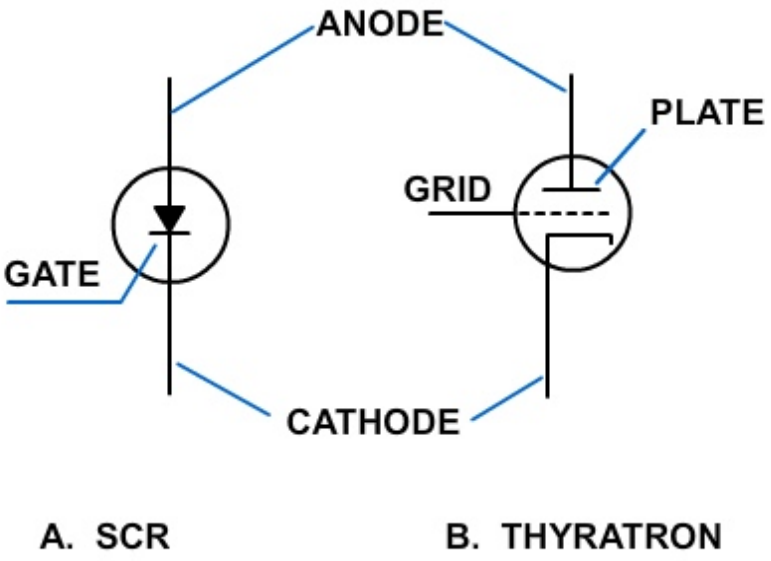


Figure 4-70 — Comparison of an SCR and a thyratron.

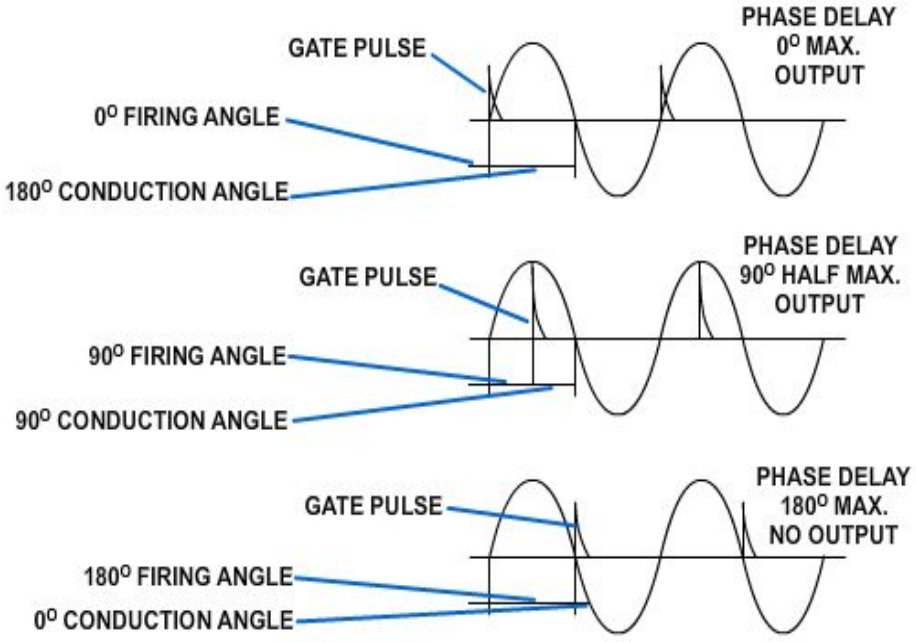


Figure 4-71 — SCR gate control signals.

When an ac power source is used, the SCR is turned off automatically because current and voltage drop to zero every half cycle. By using one SCR on positive alternations and one on negative, full-wave rectification can be accomplished, and control is obtained over the entire sine wave. The SCR serves in this application just as its name implies—as a controlled rectifier of ac voltage.

9.0.0 TRIAC

The ***triac*** is a three-terminal device similar in construction and operation to the SCR. The triac controls and conducts current flow during both alternations of an ac cycle instead of only one. The schematic symbols for the SCR and the triac are compared in *Figure 4-72*. Both the SCR and the triac have a gate lead. However, in the triac, the lead on the same side as the gate is main terminal 1, and the lead opposite the gate is main terminal 2. This method of lead labeling is necessary because the triac is essentially two SCRs back to back, with a common gate and common terminals. Each terminal is, in effect, the anode of one SCR and the cathode of another and either terminal can receive an input. In fact, the functions of a triac can be duplicated by connecting two actual SCRs, as shown in *Figure 4-73*. The result is a three-terminal device identical to the triac. The common anode-cathode connections form main terminals 1 and 2, and the common gate forms terminal 3.

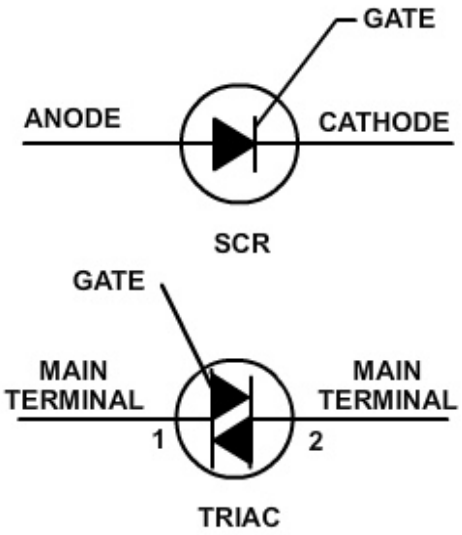


Figure 4-72 – Comparison of SCR and triac symbols.

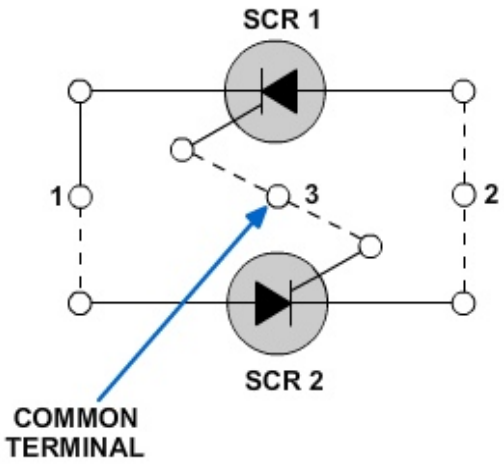


Figure 4-73 – Back-to-back SCR equivalent circuit.

The difference in current control between the SCR and the triac can be seen by comparing their operation in the basic circuit shown in *Figure 4-74*.