Training

Automotive Electronics 101
by Bernie Thompson - Automotive Test Solutions

Electronics have changed the world and the automobile is not exempt from these changes. Vehicles have changed dramatically over the years with the electrical and electronic systems that have been added to the vehicle. The modern vehicle is a marvel of engineering that rivals that of a spaceship. Electronic control of the modern vehicle is a given and this is not going away anytime soon. Modern automotive electronics will provide an ever increasing complication when servicing vehicles in your service bays. It will be important to understand how electronics work, not only for the vehicles in your services bays today but for the vehicles that will be there tomorrow.

Electrical system control uses passive effects such as resistance, capacitance, and inductance to control current flow. Whereas electronic control uses active effects such as amplification and rectification to control current flow. Both electrical and electronic principals are in use with the application of electronics. Electronic control can incorporate analog electronics or digital electronics. Analog electronics use continuously varying voltage signals, whereas digital electronics use a system where there are only two voltage levels. Both analog and digital systems are incorporated within the modern vehicle.

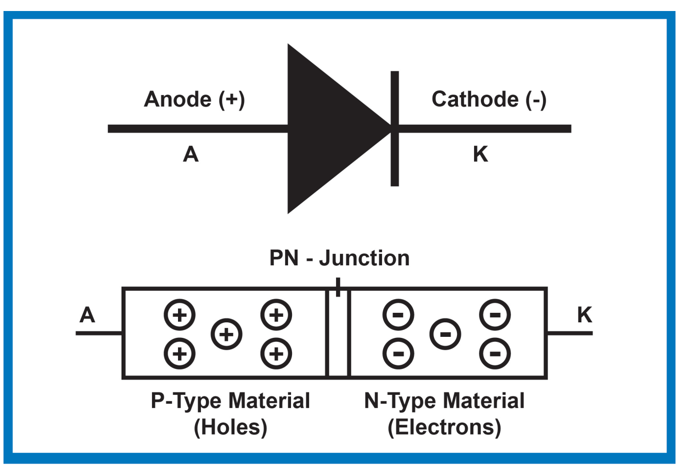

Figure 1

First let's take a look at a semi-conductor device referred to as a Diode, illustrated in Figure 1. A diode is a two terminal device that typically only allows the current to flow in one direction. The diode allows the current to flow in one direction by lowering the resistance (typically zero) when a current with a forward bias is applied. When the current has a reversed bias applied the resistance increases (typically infinite). One can think of this as a switch that conducts current when the contacts are closed (on-state) and shuts off the current flow when the contacts are open (off-state). However this on-off switching mechanism is not mechanical but is solid state. Solid state refers to a device with no moving parts. Therefore, solid state electronic devices are made up of solid components that do not move. These solid state devices are made up of solid materials produced in the form of a crystal lattice.

In order to understand this switching mechanism one will need to look into the subatomic world. In this world one can see Atoms are made up of particles containing Protons (+), Neutrons (0), and Electrons (-). The atom's nucleus is at the center of the atom and is made of Protons and Neutrons, while the Electrons are in orbit around the nucleus. One might think of this as a solar system where the sun is the mass that is at the center of the solar system, and the planets are in orbit around the sun. These Protons carry a positive charge (+) while the Neutrons carry a neutral charge (0) and the Electrons carry a negative charge (-). The number of protons and neutrons are equal to the number of electrons when the atom is in balance, or in equilibrium. When an electron is forced away from the atom, the atom becomes deficient of electrons and becomes positively charged (more protons than electrons). When an electron is forced to an atom, the atom then has an excess of electrons so it becomes negatively charged (more electrons than protons). In either of these states, positively charged or negatively charged, the atom is out of equilibrium. In this state the atom will want to get back to a state of equilibrium.

The force acting on the atom to create an imbalanced condition will need to continue to be applied in order to keep the atom out of balance. When the force is no longer applied on the atom, the atom will then move to a state of equilibrium once again. These forces that will create imbalance to the atom are light, magnetism, heat, and chemical reactions. These are very strong forces within the universe. The out of balance atoms will create an electrical charge (voltage) that is referred to as static charge. If the conditions are correct, and the static charge can create the movement of electrons, then dynamic charge will be present. Dynamic charge is the flow of electrons which is electricity. Electrons flow from a negative potential to a positive potential, whereas the charge flows from a positive potential to a negative potential.

Electricity can be moved through a conductor which has a small resistance. A conductor has less than three electrons in the outermost valance band. A valence electron is an electron that is contained within the outermost shell that is associated with an atom. With three or less valence electrons there is room for the electrons to freely move in and out of the valence band. Electricity can also be moved through a semiconductor, however semiconductors are poor conductors due to these atoms having four electrons within the outermost shell that is associated with an atom. With four electrons in the valance band there is little room for the electrons to move freely into and out of the valance band, thus adding resistance to the current flow. An insulator has more than five electrons within the outermost shell that is associated with an atom. With five or more electrons in the valence band there is no room for the electrons to freely move into and out of the valence band, thus they offer a great resistance to current flow.

A semiconductor's conductivity, having four electrons, falls in-between a conductor such as copper, and an insulator such as glass. The most common semiconductor used is Silicon (silica sand). A silicon atom has 14 electrons around the nucleus with 4 of these located in the valence band on the outermost orbital. When this is made into a single crystal, it can be used as a material for semiconductor products. Pure silicon is an insulator so in order to use silicon in electronic applications the material will be doped with an impurity atom. Doping is where a chemical element is added to the pure silicon base. In order to produce a P-Type material a trivalent atom, an atom with three valence electrons, is used, for example; Indium, Gallium, Aluminum, Boron ,etc.. In order to produce an N-Type material a pentavalent atom is used. This is where an atom having five valence electrons is used to add electrons to the outermost orbital or valence band, for example; Phosphorus, and Arsenic.

Now one can see by adding the impurities to the silicon that two structures are produced. One structure where the material is deficient of electrons, and the other structure where the material has excessive electrons. When the material is deficient in electrons (P-Type material) there are holes in the material. When the material is excessive in electrons (N-Type material) there are free electrons in the material, as illustrated in Figure 2. Either P-Type material or N-type material by themselves acts as a resistor, however when these materials are combined, as shown in Figure 1, their properties change dramatically.

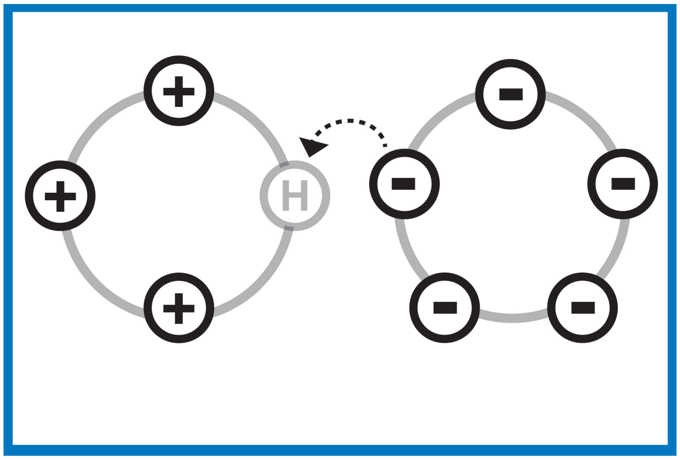

Figure 2

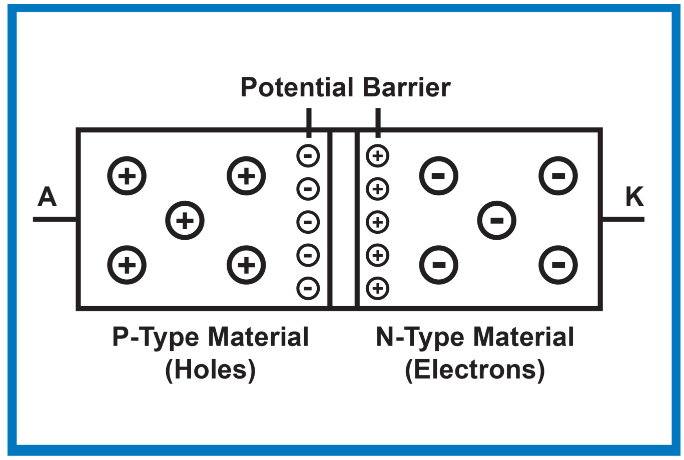

Figure 3

As seen in Figure 1 where the PN-Junction is formed between the P-type and N-type materials (the depletion region); the P-type having holes allows the N-type material with free electrons to have electrons naturally migrate across to fill some of the holes. This produces a small region where the P-type material becomes negatively charged and the N-type material becomes positively charged, as illustrated in Figure 3. This charge that is set up in the depletion region, referred to as the potential barrier, will prevent any natural electron migration from occurring. This is due to there being no free charge carriers available in this region. If a power source, such as a battery, is used it will change this depletion region by applying charge carriers back into the depletion region, as illustrated in Figure 4. If the power source is used to apply a forward bias to this depletion region, and it has enough voltage to overcome the potential barrier (.7 Volts), it will allow the N-type material to give electrons to the holes in the P-type material. In this forward basing position current will flow from the N-type region to the P-type region. This is referred to as drift current. In this condition the diode is in a conduction state, or the diode is now in an on-state. If the power supply is reversed it will move the electrons in the opposite direction. This reverse basing will allow the

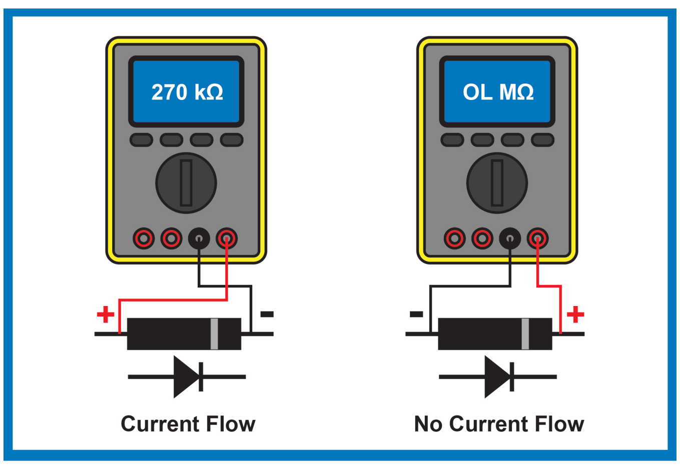

Figure 4

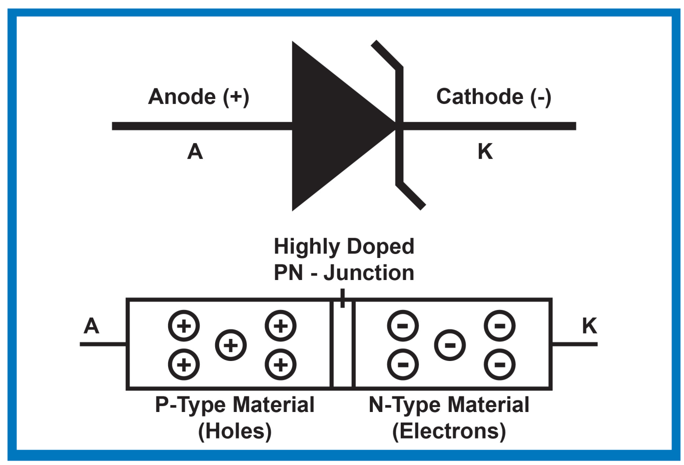

Figure 5

depletion region to widen, stopping the current flow. In this reverse biasing condition the electrons and holes are pulled away from the junction by the power source. In this condition the depletion region is further away from the conduction state and the diode is now in an off-state.

Another diode referred to a Zener diode can be used in certain application, illustrated in Figure 5. A Zener diode is a semiconductor that allows the current to flow in both directions. All diodes have what is referred to as avalanche. Avalanche is the breakdown voltage level where the diode will conduct current backwards. The Zener diode has a highly doped NP-Junction area where this avalanche voltage is set to a reliable reverse voltage level. Zener diodes exhibit a controlled breakdown that allows the current to keep the voltage across the Zener diode very close to the

Zener breakdown voltage. This voltage level is much lower than a conventional diodes avalanche voltage. The Zener voltage level is adjusted by the amount of doping in the PN-Junction area. This reverse conduction occurs due to electron quantum tunneling in the small space between the P and N regions. This is known as the Zener effect, and is referred to as the Zener set voltage.

Yet another diode referred to a Light-Emitting Diode (LED) is used in vehicle electronics, as illustrated in Figure 6. The diode produces light when the electrons flow through the device. This light is produced when an electron moves from the N-type material across the depletion region and fills the holes in the P-type material. When this occurs energy in the form of a photon is generated. A photon is a type of particle that has an electromagnetic field including electromagnetic radiation such as radio waves and light. A photon has no mass and moves at the speed of light. The color of the light (corresponding to the wavelengths of the photons) is determined by the energy required for electrons to cross the depletion region of the semiconductor. So, in order to produce different colors the depletion region is doped with impurity. It can also be produced by layering different impurities in the depletion region. In addition to doping the material, the voltage level used to forward bias the diode can produce different colors. LEDs have many advantages over traditional light means, mainly in their power consumption. LEDs are far more efficient, smaller in size, more robust, and have fast switching times.

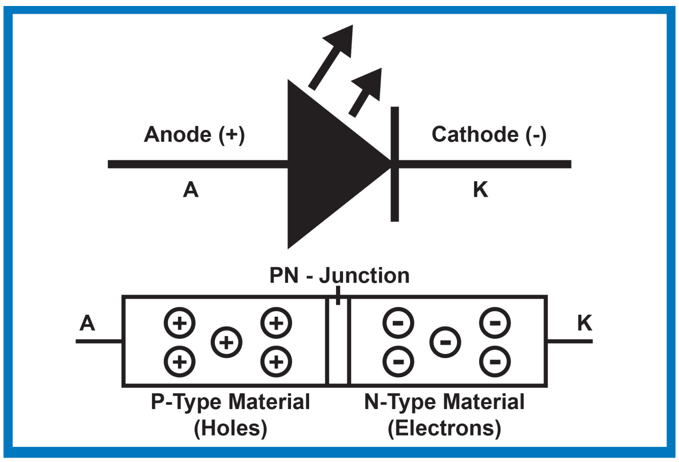

Figure 6

Diodes are used for many different functions on the vehicle. For example, the charging system uses diodes to rectify the induction in the alternators three stator windings. This is accomplished with a diode trio. The diode trio allows only the positive voltage to pass through the diode. This positive voltage is then used by the vehicle electrical system. Diodes are also used for spike suppression in windings. For example; purge control solenoids, vent control solenoids, VVT solenoids, fuel injector solenoids, etc.. Zener diodes are widely used for voltage regulation. This regulation can provide regulated voltage for a supply circuit (reference voltage) or regulate the voltage cut level on a fuel injector. Light-Emitting Diodes are used in the vehicle for; indication lamps, warning lamps, illumination, and lighting systems.

When testing a diode the dual resistance measurement of the diode in both the forward bias and reverse bias direction is referred to as the Front-To-Back Resistance Ratio of the Diode. A ten time's difference in diode resistance is a minimum testing difference for most diodes. This means the resistance should be ten times greater when a reverse bias is applied to the diode. In most cases this indicates the diode is good. It will be important to understand that a digital multimeter resistance test uses a lower testing battery voltage than an analog ohm meter. The digital meter may not have enough voltage to overcome the diode depletion region in the diode to turn it on. This means when using a digital multimeter it will need to have a special "Diode Test Function" in order to forward bias the diode correctly. If the digital multimeter does not incorporate a "Diode Test Function" it may not work properly. In this case use an analog multimeter for testing diodes. When testing a diode within a circuit, such as a spike suppression diode on a relay winding, always use an analog multimeter or a device that you know can forward bias the diode in this condition.

Another Semiconductor device that one will need to understand is the Transistor. The transistor is a semiconductor device that operates on the same principles that the diode operated with. The transistor is a solid-state device that usually has three terminals and is used as an amplifier and/or a switch. The transistor is the electronic component that has changed the world. This device was first conceived in in 1926, however it was not able to be produced at that time. Some twenty years later in 1946 the first working transistor was produced at Bell Labs. The most widely used transistor is the Metal–Oxide–Semiconductor Field-Effect Transistor (MOSFET). The MOSFET was the first truly compact transistor that could be miniaturized and mass-produced for a wide range of uses.

The first type of transistor we will examine is the Bipolar Junction Transistor (BJT). In order to operate, the BJT uses P-type material that has holes and N-type material that has excess electrons. The PN-junction diode that we have discussed above has two terminals that are connected to two elements; an anode and a cathode. The BJT transistor has three terminals that are connected to three elements; the Emitter, the Collector, and the Base. Each of these are connected to an external circuit through the transistor terminals. This type of transistor can be thought of as two diodes connected together in series, as illustrated in Figure 7.

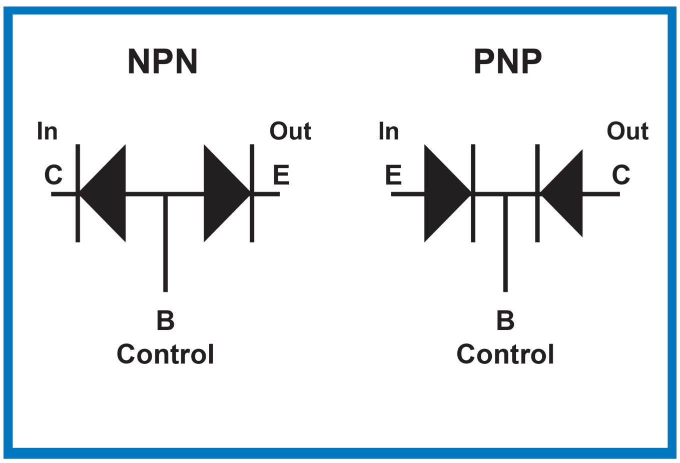

Figure 7

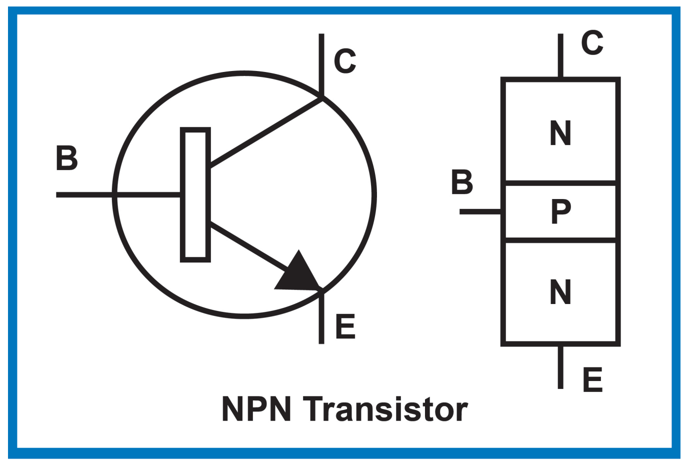

Figure 8

First we will analyze the BJT NPN Transistor, illustrated in Figure 8. The emitter and base are made up of N-type material. Whereas the base is made up of P-type material the BJT has two junction areas. These are formed between the N-type material and P-type material. Since there are two junction areas there are two depletion regions that set up potential barriers. These are formed at each of the edges of the PN-junction. In order to turn this device on, voltage will be applied to the P-type material. This will forward bias the transistor and allow the P-type material to give electrons to the N-type material. To control the NPN transistor the base voltage is increased or elevated. Just as the NP-junction diode needs .7 volts to overcome the potential barrier and turn on the diode, so will the BJT. However with the NP-junction diode this .7 volts will fully turn the device on. This is not the case with the NPN transistor. This base voltage sets the transistor's output or amplification. This .7 volt turn on point just starts the current flow through the device.

In this condition the transistor is in the on-state where electrons (current) are moving from the collector to the emitter. These devices are operated by current. A small current flowing through the base will enable a large current flow through the emitter to the collector. The greater the base current level the greater the transistor's output or amplification will be. The voltage applied on the base allows a small current to flow through the base (P-type more electrons) to the emitter (N-type less electrons, holes) and then to ground. This will forward bias the transistor's junction and turn the device on. The main current will flow through the collector to the emitter. The collector is connected to the more positive lead, whereas the emitter is connected to the more negative lead, as illustrated in

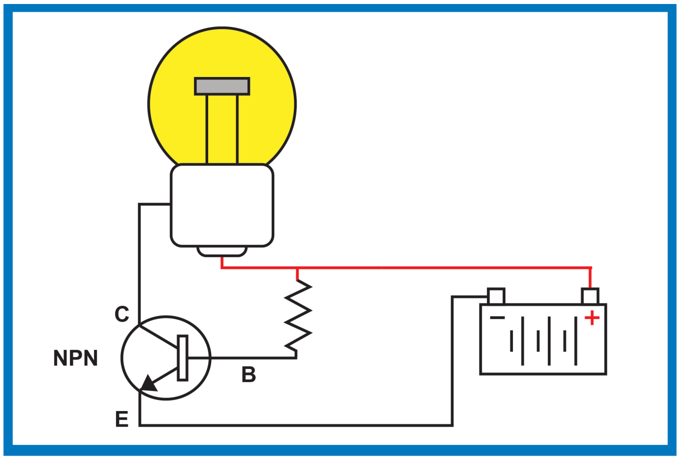

Figure 9

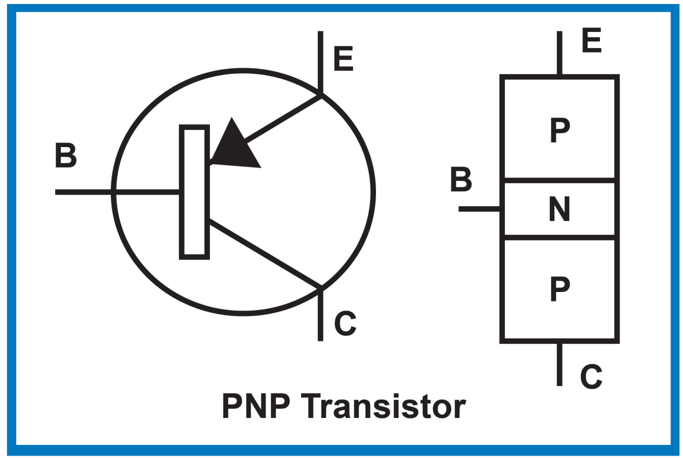

Figure 10

Figure 9. The base is the lead in the middle and is connected to the P-type material. The emitter will always be connected to the lead with the arrow. On the NPN transistor the arrow will always be pointing out (Not Pointing IN, NPN). These type of transistors usually control a ground source.

Second we will analyze the BJT PNP Transistor, illustrated in Figure 10. The emitter and base are made up of P-type material. Whereas the base is made up of N-type material. The BJT has two junction areas. These are formed between the N-type material and P-type material. Since there are two junction areas there are two depletion regions that set up potential barriers. These are formed at each of the edges of the PN-junction. In order to turn this device on, voltage will be decreased from the N-type material (base). When the voltage is decreased there will then be a potential difference between the base and emitter. This decrease of voltage to the N-type material will forward bias the transistor and allow the P-type material (P-type more electrons) to give electrons to the N-type material (N-type less electrons, holes). To control the PNP transistor the base voltage is decreased or dropped. Just as the NP-junction diode needs .7 volts to overcome the potential barrier and turn on the diode, so to will the BJT. However with the NP-junction diode this .7 volts will fully turn the device on. This is not the case with the PNP transistor. This base voltage sets the transistors output or amplification. This .7 volt drop turn on point just starts the current flow through the device.

In this condition the transistor is in the on-state where electrons (current) are moving from the emitter to the collector. The greater the base voltage level drops, the greater the current flow through the emitter to the base. These devices work on current flow. The greater the current flow through the base, the greater the current flow is from the emitter to the collector, or the greater the transistor's output or amplification will be. The voltage decrease on the base allows a small current to flow through the emitter to the base and then to ground. This will forward bias the transistor's junction and turn the device on. Whereas the main current will flow through the emitter to the collector. The emitter is connected to the more positive lead, whereas the collector is connected to the more negative lead, as illustrated in Figure 11. The base is the lead in the middle connected to the N-type material. The

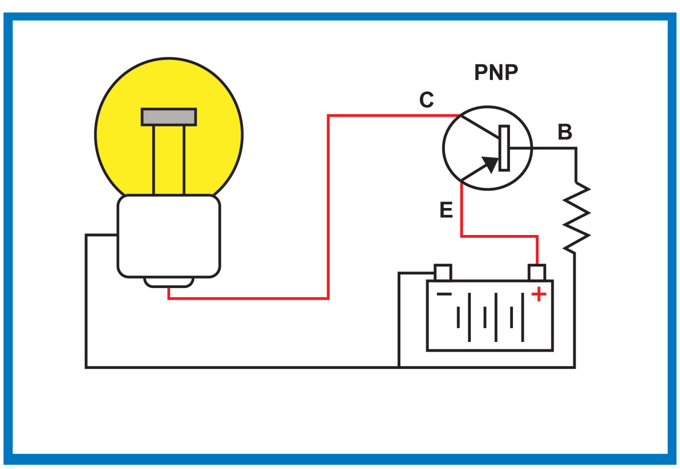

Figure 11

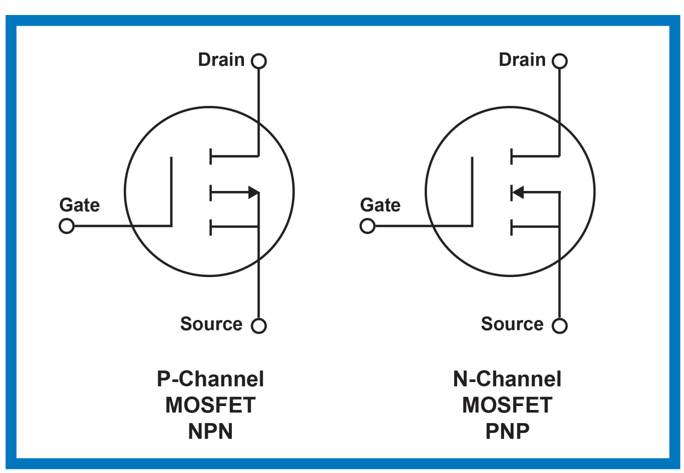

Figure 12

emitter will always be connected to the lead with the arrow. On the PNP transistor the arrow will always be pointing in (Pointing IN, PNP). These type of transistors usually control a power source.

Yet another type of semiconductor transistor is the Field-Effect Transistor (FET), as illustrated in Figure 12. There are many different types of FETs, this design is referred to as an E channel MOSFET. These type of transistors do not allow current to flow through the control device (gate/base) of the transistor, but instead use an electric field (voltage) to control the device. The FET uses an insulated-gate to control the current flow through the device. This gate is fabricated by a controlled oxidation of the semiconductor which is usually silicone. This controlled oxidation is produced by a process of thermal oxidation that can be performed on selected areas of the silicon wafer and blocked on others. This process is commonly referred to as the local oxidation of silicon. This silicon dioxide is used as an insulator.

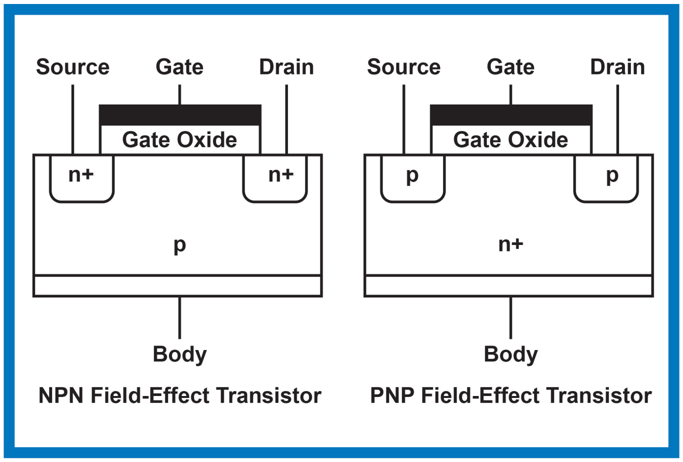

Figure 13

This gate is controlled, or turned on by voltage. These transistors use three terminals that connect to an external circuit. These terminals are the Source, Gate, and Drain.. These are much like the BJT emitter, base, and collector. The NPN FET has a source and drain which are produced with ion implementation of two N-type material strips that are embedded in the substrate body produced with P-type material, as shown in Figure 13. The insulated-gate is formed to bridge the two N-type strips, and is made of silicon dioxide. Above the silicon dioxide is a layer of aluminum (metal) that is bonded to the silicon dioxide. This insulator material keeps the metal from touching the N-type and P-type materials. The terminals are etched to the N-type strips and the metal in the silicon dioxide insulated-gate. If there is no positive potential (voltage) on the gate then there are no available carriers to move the current from the source to the drain, thus no current flow. In this condition the transistor is in an off-state. When positive voltage is applied to the metal and it becomes charged, it works like a capacitor and forms an equal but opposite charge in the P-type channel. This attracts negative charge carriers (electrons) from the N-type material. These charge carriers can now conduct the electrons from the source to the drain. In this condition the transistor is in an on-state. The greater the positive voltage applied on the insulated gate the wider the channel becomes. This allows more current to flow from the source to the drain.

The PNP FET has a source and drain which are produced with ion implementation of two P-type material strips that are embedded in the substrate body produced with N-type material, as shown in Figure 13. The insulated-gate is formed to bridge the two P-type strips and is made of silicon dioxide. Above the silicon dioxide insulated-gate a layer of aluminum (metal) is bonded to the silicon dioxide. The silicon dioxide is used to insulate the metal from the P-type and N-type materials. The terminals are etched to the P-type strips and the metal in the silicon dioxide insulated-gate. If there is not a positive potential (voltage) on the gate then there are no available carriers to move the current from the source to the drain, thus no current flow. In this condition the transistor is in an off-state. When a positive voltage is applied to the metal and it becomes charged, it works like a capacitor and forms an equal but opposite charge in the N-type channel. This attracts positive charge carriers (holes) from the P-type material. These charge carriers can now conduct the electrons from the source to the drain. In this condition the transistor is in an on-state. The greater the positive voltage applied on the insulated gate the wider the channel becomes. This allows more current to flow from the source to the drain.

Additionally, some FETs have a fourth terminal that is connected to the body. This fourth body terminal serves to bias the transistor into operation. The source terminal and body are usually connected together since the source is often connected to the highest or lowest voltage within the circuit, depending on the type of FET used (N-channel or P-channel). The voltage applied to the covered gate determines the electrical conductivity of the device. If the level of voltage on the gate is changed, the flow of current between the source and drain is changed. This ability to change conductivity with the amount of applied voltage allows FETs to be used as amplifiers and/or switches.

Transistors have a very wide use in the modern vehicle. These devices are used to turn on all types of devices such as; fuel injectors, ignition coils, VCT control solenoids, purge control solenoids, transmission solenoids, HVAC doors, lights, etc.. Additionally these transistors are used to setup the logic in all the vehicle modules. The module program works with the logic states within the hardware to allow the program to control the system. At the base of these logic states is the logic gate.

Transistors are used to produce Logic Gates within an integrated circuit. A logic gate is a device that works on a principal of having either an on-state or off-state. These devices are the basic building blocks of any digital system. Logic gates are made up of electronic circuits that have one or more inputs but only one output. The relationship between the inputs and the output is based on a certain logic. The logic for each type of logic gate is put into a truth table. A truth table is a good way to show the function of a logic gate. It shows the output states for every possible combination of input states. The symbols 0 (false) and 1 (true) are usually used in truth tables. There are several logic gate configurations such as; NOT gate, AND gate, OR gate, XOR gate or when these same gates are inverted; NAND gate, NOR Gate, XNOR Gate. There are several logic gate circuit diagrams that we will analyze.

NOTE: There are switches drawn in these circuits only to explain how these gates work. These logic gates do not have switches when they are in the integrated circuit. These switches would be solid state devices that are turned on and off in an actual integrated circuit.

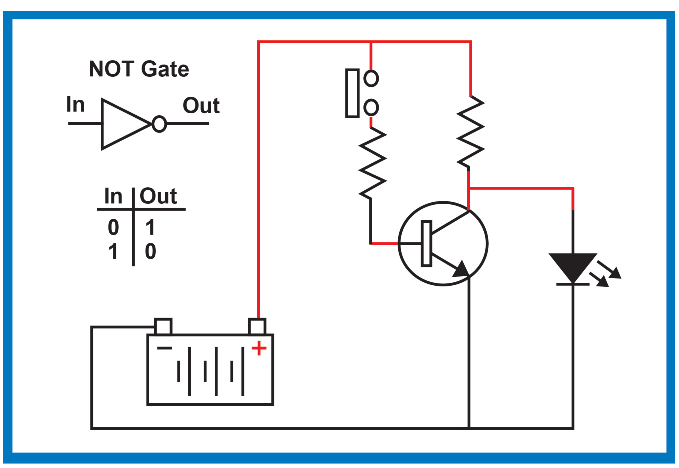

Figure 14

Let's start with looking at a NOT Gate, as illustrated in Figure 14. In this diagram there is a power source in the lower left corner. This power source will supply current to the circuit so it can operate. The current will be supplied to the LED on the right side. The LED will be "on" as the default condition. If the switch is closed (1) the transistor is turned on so the current path now moves through the transistor. This leg of the circuit that goes to the LED is now a ground circuit. This lowers the voltage to the LED so there is not a potential difference between the LED input and output. So, in this condition the LED is turned off (0). When one looks at the truth table it is clear that when the switch is off (0) the output is (1). When the switch is closed (1) the output is (0). The NOT gate is used as an inverter.

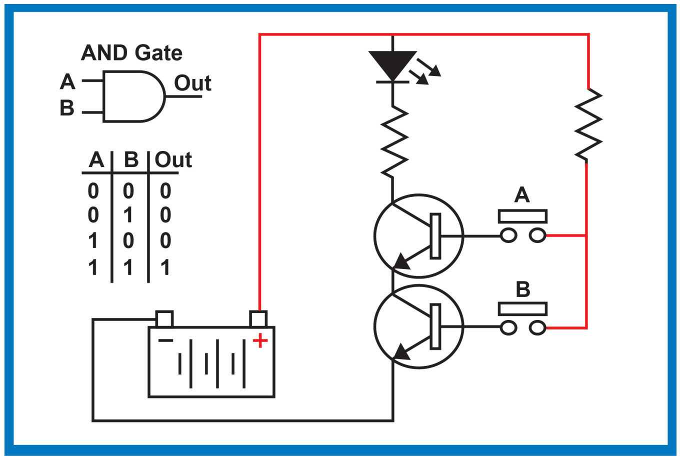

Figure 15

Now let's look at an AND Gate, as illustrated in Figure 15. In this diagram there is a power source in the lower left corner. This will supply current to the circuit so it can operate. The circuit supply's current to the LED only if both transistor A and transistor B are turned on. The LED default state is off (0). If switch A is closed (1) then voltage is supplied to the base of transistor A, turning it on (1). However this will only close (1) transistor A, transistor B is off (0), so there is not a complete circuit to ground. If switch B is closed (1) then voltage is supplied to the base of transistor B, turning it on (1). However this will only close (1) transistor B, transistor A is off (0), so there is not a complete circuit to ground. If switch A is closed (1), and switch B is closed (1) at the same time, then the ground circuit is complete and the LED is turned on (1). This shows the function of the AND gate where both gates must be closed A(1) B(1) to get an output that is on (1).

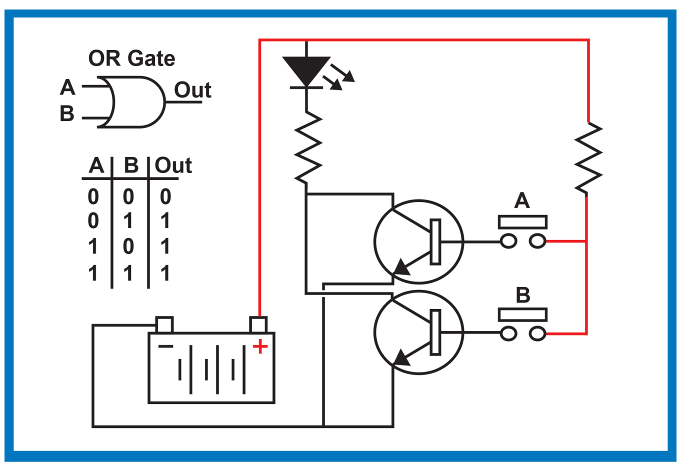

Figure 16

Now let's look at an OR Gate, as illustrated in Figure 16. In this diagram there is a power source in the lower left corner. This will supply current to the circuit so it can operate. The circuit supplies current to the LED only if either transistor A is on, or if transistor B is on, or if both transistor A and transistor B are on. The LED default state is off (0). If transistor A switch is open (0) and transistor B switch is open (0), the LED output state is off (0). If transistor switch A is closed then voltage is applied to transistor A's base turning transistor A on (1). When transistor A is turned on (1) the ground circuit is completed. The LED output is now turned on (1). If transistor switch B is closed then voltage is applied to transistor B's base, turning transistor B on (1). When transistor B is turned on (1) the ground circuit is completed. The LED output state is now turned on (1). If both transistor A and transistor B are both turned on A(1) B(1) the ground circuit is completed and the output state is on (1). This shows the function of the OR Gate, where either gate can be closed A(1), B(1), or A(1) and B(1) to get an output that is on (1).

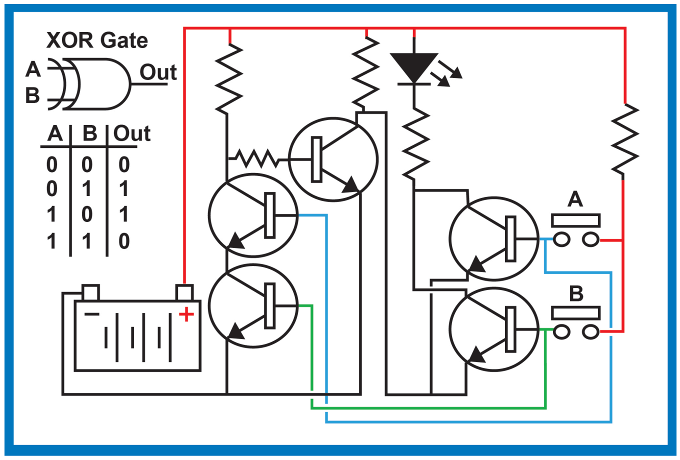

Figure 17

Now let's look at an XOR Gate, as illustrated in Figure 17. In this diagram there is a power source in the lower left corner. This will supply current to the circuit so it can operate. The circuit supplies current to the LED only if transistor A is on, or if transistor B is on. If transistor A and transistor B are on at the same time there is not a current path so the LED is turned off. The XOR gate LED default state is off (0). If transistor A switch is open (0), and transistor B switch is open (0), then the output state is off (0). If transistor A switch is closed it supplies current to both transistor A's bases turning them on. This will turn on both A transistors (1)(1) attached to switch A. This will complete the ground leg through transistor T1 that is defaulted to an on-state. This will turn on the LED, thus the output state is on (1). If transistor B switch is closed it supplies current to both transistor B's bases turning them on. This will turn on both B transistors (1)(1) attached to switch B. This will complete the ground leg through transistor T1 that is defaulted to an on-state. This will turn on the LED, thus the output state is on (1). If transistor A switch is closed, and transistor B switch is closed, then the current path is now through transistors A and B that are in series. This A B transistor circuit is now a ground circuit so the base voltage to transistor T1 is low, thus turning off transistor T1. In turn this opens the ground circuit for the LED turning it off. This output state is now in an off-state (0). This shows the function of the XOR Gate, where either gate can be closed A(1) or B(1), to get an output that is on (1). But If A and B are closed (1)(1) then the output is off (0).

The modern world is full of electronics and so is the modern vehicle. It will be important for the modern technician to have a clear understanding of electricity and electronics. This article is just a very small introduction into the world of electronics. In order for the modern technician to be successful in automotive repair, one will need to continue one's education in these fields.