Semiconductor Diodes and Transistors

By: Sarah Marsh

November 28, 2012

The world has gone digital. In this digital world, silicon controlled rectifiers are everything. It would be hard to find a piece of modern technology that does not use a diode or transistor. Not only are these devices found in cutting-edge technology, they are responsible for the ability of these technologies to function at all. They are the building blocks of the technological revolution. Their function can be as simple as ensuring a calculator is not ruined by inserting the batteries backward, or it can be as complex as utilizing millions of embedded transistors to form the foundation of a computer. The power of diodes and transistors come from their ability to control electricity. This phenomenon makes them invaluable in circuit design and has formed the basis of integrated circuitry. Transistor packed chips called microprocessors exist in virtually every electronic device we interact with. Diodes are integral to radio broadcasting, LED lighting, radiation detection, laser technology, and so much more.

"Semiconductors fall between insulators and conductors in their ability to pass current" (Heller 1968). The basis for semiconductor diodes and transistors come from the intrinsic properties of the material they are made from. These materials are known as pure semiconductor materials (Hosch 2009). The most common type of pure semiconductor material is silicon, although many other materials can also be used.

Atoms are known to have electrons that encircle them. The electrons are in cloudlike orbits around the nucleus. Some electrons are in closer orbits than others. Each of the orbits is associated with some amount of the atom's total energy. The different orbits can now be referred to as energy levels, with the lowest energy level being closest to the nucleus. In its solid form, silicon is crystalline in structure. Atoms in crystals are bonded covalently. A covalent bond is one in which atoms share their outermost, or valence, electrons (Heller 1968). The two atoms thus share their highest energy level. The energy level of the electrons being shared is known as the valence level. When a crystal is formed, many atoms must be brought together to form it. The energy levels of the atoms become a "band" of energy levels which applies to all the atoms that make up the crystal. This band is referred to as the valence band. Above the valence band there is a "gap" where no energy levels can be occupied by an electron, this is known as the energy gap. Above that, the band is known as the "conduction band." In the conduction band, as the name suggests, electrons can be conducted, and a current can flow in the material.

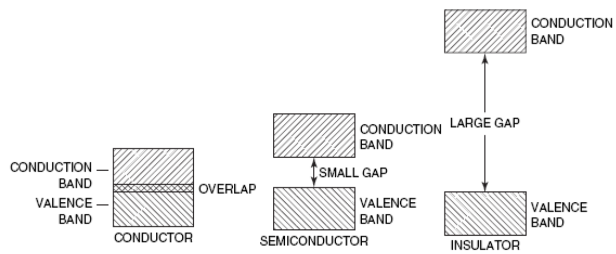

Figure 9-1. Valence and conduction bands in conductor, semiconductor, and insulator.

Figure 1 (Hecht 2008)

Figure 1 shows the energy gap, the valence band, and the conduction band. The number of electrons in the conduction band of a pure semiconductor material is given by the equation

$n =$ ——— (Hunter 1956)

where m_e is the mass of the electron, *h is* Planck's constant, *k* is Boltzmann's constant, *T* is the absolute temperature, is the energy level at the threshold of the conduction band, and is the Fermi level. "The Fermi level in a metal or semiconductor is located at that value of energy at which there is a 50 per cent probability of an energy state at that level being occupied by an electron" (Warschauer 1959). The Fermi level of a pure semiconductor material with no impurities is the energy level midway between the band edges of the energy gap. Analyzing the equation, it is seen that the number of electrons in the conduction band increases with greater temperatures and decreases exponentially with the energy difference between the Fermi level and the energy at the conduction “band edge.” An electron’s energy can be raised by absorbing light energy, another particle “bombarding” it, and thermal motion in the center of its atom. If there is a large gap between the valence and conduction band, it is very hard for an electron to gain enough energy through natural occurrences to move between them. This makes the material an insulator (Hunter 1956). This is true for silicon. Figure 1 demonstrates an energy gap’s relationship to semiconductors, conductors, and insulators. So, pure semiconductor materials must undergo a process known as “doping” to achieve the conductivity required for them to be classified as semiconductors (Heller 1968).

Doping is the process of adding small amounts of impurities to a pure semiconductor material to give it a desired conductivity. Doping yields two types of semiconductors, P-type and N-type. The junction between these two is the basis for modern transistors and diodes. An N-type semiconductor is formed when an impurity is added that leaves unbounded electrons present in the semiconductor material. This occurs because silicon has four valence electrons, and the impurity added has five valence electrons. Some common impurities added to create an N-type semiconductor are arsenic, antimony, and phosphorous. When the impurity is added, four of its valence electrons readily form covalent bonds with the four valences of silicon, but one electron has no available electron to form a bond with and is left unbounded (Heller 1968).

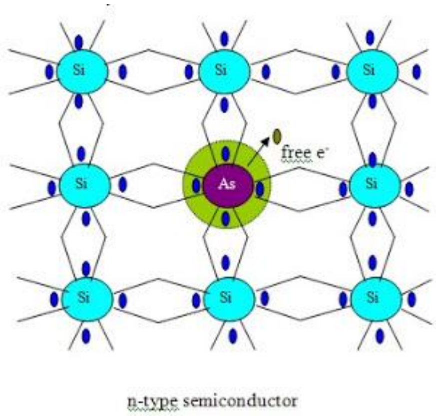

Figure 2 (Ashwin 2009)

Figure 2 shows arsenic added to silicon. P-type semiconductors are formed when an impurity is added that leaves the material electron deficient, and a "hole" occurs where a bond should be formed. Impurities added to create a P-type semiconductor include aluminum, gallium, and boron. The impurities that create P-type semiconductors have three valence electrons. When the impurity is added, three of its valence electrons bond covalently with three of silicon's four valence electrons. A "hole" occurs where the bond between silicon's fourth valence electron and another valence electron should have formed.

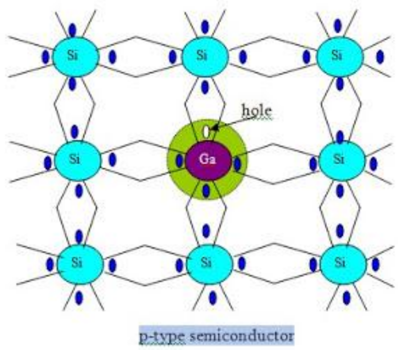

Figure 3(Ashwin 2009)

Figure 3 shows Gallium added to silicon. Current in P-type material is considered the flow of holes. Current does not necessarily "flow" in P-type material, it "hops." This is because a covalently bonded electron experiencing a force from the negative terminal of the voltage will "hop" into a nearby hole. There, it exerts a repelling force on a nearby electron which then breaks its bond and forms another hole. So, when an external voltage is applied, the holes move toward the negative side of the voltage source as the electrons move toward the positive side (Heller 1968). The equation for the number of holes found in the conduction band of a pure semiconductor material is identical to the equation for the number of electrons found, with the substitution of the energy at the top of the valence band for the energy at the threshold of the conduction band. The mass of a hole is equal to the mass of an electron. Thus, the equation for the number of holes is

$$n = \frac{\quad}{\quad} \qquad \text{(Hunter 1956)}$$

“where is the energy at the top of the valence band” (Hunter 1956). In a pure semiconductor material, the density of holes is equal to the density of electrons.

The majority current carrier is the carrier which has the higher density in a material, and the minority carrier is that with the lower density. Holes are the majority carriers in P-type, while electrons are the majority carriers in N-Type. Minority carriers exist in both materials (i.e. free electrons are still present in P-type, and holes are still present in N-type) (Warschauer 1959).

As was stated earlier, the basis for modern transistors and diodes is the junction between N-type and P-type semiconductors. This junction is known as the P-N junction. Because of thermal vibrations in the atoms on both sides of the junction, the holes on the P side and electrons on the N side tend to spread out. This process of spreading out is known as “diffusion” (Warschauer 1959). Holes from the P side diffuse to the N side, as electrons from the N side diffuse to the P side. When each respective carrier crosses the barrier, it encounters an excess of the other type of carriers, and electrons and holes recombine. Electrons leaving the N side create positive ions on the N side. Holes filled by the incoming electrons on the P side become negative ions. Because of this, an electric field is created. The potential difference impedes more holes and electrons from continuing to flow, so the potential difference is concentrated at the boundary between the P and N types. The field formed is known as a “potential barrier” because electrons and holes cannot flow (Heller 1968).

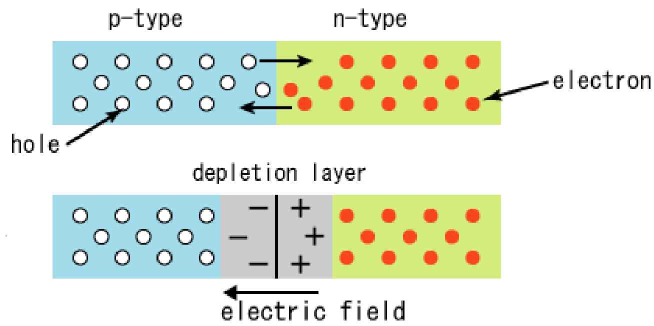

Figure 4 (Shahbuddin 2010)

Figure 4 demonstrates the formation of the potential barrier. One P-N junction is considered a diode. Two P-N junctions create a transistor (Heller 1968).

When an external voltage is applied to a diode by a battery, the polarity of the voltage source determines the behavior of current in the circuit. When the battery is connected in this orientation:

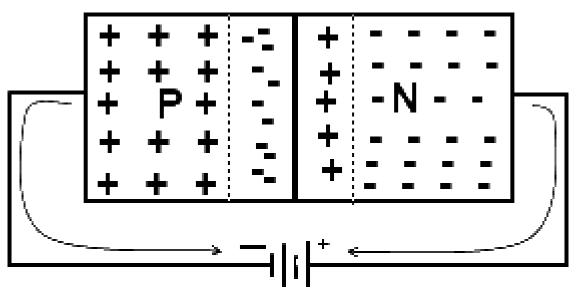

Figure 5 (“Formation of Depletion Layer in Diode”)

the junction is said to be reverse- biased. The electrons in N-type are attracted to the positive side of the terminal, and the holes in P-type move toward the negative side of the terminal. The majority carriers are pulled even farther away from the junction, and, as a result, the potential at the junction grows larger. The larger potential and farther distance makes it almost impossible for carriers cannot cross the junction. No current flows when a diode is connected in this orientation. When a battery is connected in this orientation:

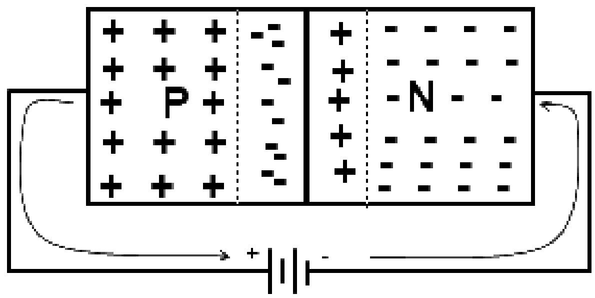

Figure 6 (“Formation of Depletion Layer in Diode”)

the junction is said to be forward-biased. When holes diffuse from P-type to N-type, they are filled by electrons from the negative terminal of the battery. Electrons diffuse from N-type to P-type where they are attracted by the positive terminal. The ratio of the holes and electrons diffusing is 1:1. So, a current flows through the diode. These two states give the diode the unique property of being able ensure current only flows one way through a circuit (Heller 1968).

When a P-N junction is formed, minority carriers still exist on both sides of the junction. Minority carriers can easily cross a reverse-biased junction because it is the equivalent of a forward-biased junction for them. Electrons in the P-Type move toward the positive potential in the N-type. This electron flow is known as “reverse current” (Heller 1968). Because the amount of minority carriers is limited, a maximum current is reached. When the maximum amount of minority carriers are flowing a state known as “saturation” is reached. The current flowing at this time is known as the “saturation reverse current” (Heller 1968). The current in a semiconductor circuit depends on this saturation current and is modeled by the equation:

I = — (Warschauer 1959)

where I_s is the saturation current, q is the charge on one electron, V is the voltage applied, K is Boltzmann's constant, and T is the absolute temperature. This model is known as the diffusion model and represents currents where there is not a large forward or reverse voltage (Warschauer 1959). In semiconductor circuits, a small voltage can produce a very large current because of the exponential.

A transistor is a device with two P-N junctions. In circuits, they serve two purposes: amplification and switching. "Put simply, amplification consists of magnifying a signal by transferring energy to it from an external source, whereas a transistor switch is a device for controlling a relatively large current between or voltage across two terminals by means of a small control current or voltage applied at a third terminal" (Muthuswamy 2007). A transistor can be a P section between two N sections, known as an NPN transistor. Or, it can be an N section between two P sections, known as a PNP transistor. The sections, in order, are named emitter, base, and collector.

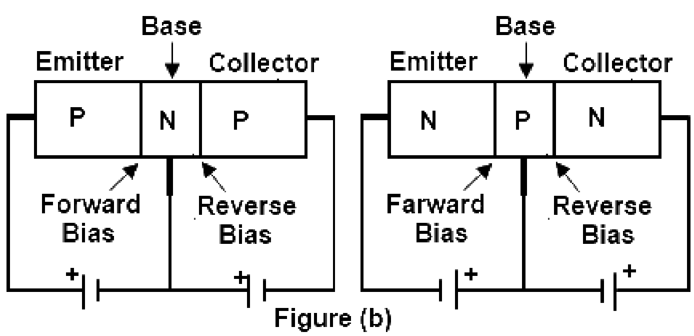

Figure 7 ("BJT Bipolar junction Transistor")

Figure 7 diagrams the two different transistors and their sections. The base region is always smaller than the two regions around it. In an NPN transistor, the current is dominated by electrons, while a PNP transistor current is dominated by holes. The two behave almost identically, but the sign of the current and voltage are reversed. Therefore, a PNP transistor is essentially a negative NPN transistor (Muthuswamy 2007). In an NPN transistor, the PN junction between the base and emitter is basically a forward-biased diode. Many of the electrons crossing the potential barrier at the base-emitter junction will not recombine with holes in the base region. This phenomenon occurs because the PN junction between the collector and base has a positive potential that attracts the electrons, and the base region is so small that it is quickly crossed (Heller 1968). A current arises opposite the flow of electrons and consequently flows from collector to emitter. Because there is also a current of holes from base to emitter, the current flowing from the emitter is larger than the current put out by the external batteries.

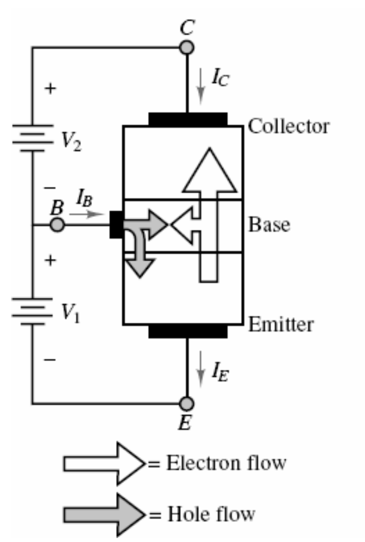

Figure 7(Muthuswamy 2007)

Kirchoff's law states that the incoming currents must equal the outgoing currents at any junction. By applying Kirchoff's law to Figure 7, it is clear that:

$$I_E = I_C + I_B$$ (Muthuswamy 2007)

Where I_E is the current in the emitter, I_C is the current in the collector, and I_B is the current in the base. In this kind of transistor, the "most important property…is that the small base current controls the much larger collector current" (Muthuswamy 2007).

$I_C =$ (Muthuswamy 2007)

Where is a “current amplification factor” that depends on the intrinsic physical properties of the materials that make up the transistor. It typically ranges from 20 to 200. The constant also determines the current needed to turn a transistor on and off (Muthuswamy 2007).

The unique and fascinating properties of diodes and transistors described in this paper have revolutionized technology. Their uses are being explored and expanded upon every single day. With companies beginning to produce computer chips with billions of transistors, technology is becoming faster and more capable than ever. An entire section of the U.S. is named after the developments made in the silicon controlled rectification industry: Silicon Valley. Some of the world’s most prevalent methods of travel, communication, and entertainment rely on devices no larger than a centimeter. Diodes and transistors are a beautiful representation of the power of physics and chemistry. They control our digital world and will for years to come.

Works Cited

1. Heller, Saul. "Semiconductor Fundamentals." In *Understanding Silicon Controlled Recitifiers*. 1968. Reprint, New York City: Hayden Book Company, Inc., 1969. 1-22.
2. Hosch, William . "semiconductor device." Britannica Online Encyclopedia. http://www.britannica.com/EBchecked/topic/533976/semiconductor-device/34322/The-p-n-junction#toc34323 (accessed November 8, 2012).
3. Hunter, Lloyd Philip. *Handbook of semiconductor electronics*. 1. ed. New York [usw.: McGraw-Hill, 1956.
4. Hecht, Jeff. "Semiconductor Basics." In *Understanding lasers: an entry-level guide*. 3rd ed. Hoboken, NJ: John Wiley & Sons ;, 2008. 236-260.
5. Ashwin, S. "N-Type semiconductor." Free Engineering Notes. http://freeengineeringnotes.blogspot.com/2009/01/n-type-semiconductor.html (accessed November 15, 2012).
6. Ashwin, S. "P-Type semiconductor." Free Engineering Notes. http://freeengineeringnotes.blogspot.com/2009/01/p-type-semiconductor.html (accessed November 15, 2012).
7. Warschauer, Douglas M.. *Semiconductors and transistors*. New York: McGraw-Hill, 1959.
8. Shahbuddin, Sarosh . "Optoelectronics." Sarosh Shahbuddin. http://saroshshahbuddin.wordpress.com/2010/08/18/optoelectronics/ (accessed November 15, 2012).

9. "Formation of Depletion Layer in Diode." Electronics engineering notes, Lectures, Projects. http://www.daenotes.com/electronics/devices-circuits/formation-depletion-layer-diode#axzz2DGlWFEUl (accessed November 25, 2012).

10. Muthuswamy, Bharathwaj. "Diodes and Transistors Handout." Berkeley.edu. inst.eecs.berkeley.edu/~ee100/su07/handouts/DiodeTransistorNotes.pdf (accessed November 14, 2012).

11. "BJT Bipolar junction Transistor." Electronics engineering notes, Lectures, Projects. http://www.daenotes.com/electronics/devices-circuits/bjt-bipolar-junction-transistor#axzz2DR5cCixj (accessed November 25, 2012).