IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 13, Issue 1 Ver. I (Jan. – Feb. 2018), PP 52-57
www.iosrjournals.org

Bipolar Junction Transistor as a Switch

Ali Habeb Aseeri[1], Fouzeyah Rajab Ali[2]
[1](*Switching Dep, High institute of telecommunication and navigation/PAAET, Kuwait,ah.aseeri@paaet.edu.kw*)
[2](*Switching Dep, High institute of telecommunication and navigation/PAAET, Kuwait,fozrmha@gmail.com*)

Abstract: *Understanding the application of a bipolar Junction transistor or BJT as a switch requiers understanding the general working principles behind a transistor and the specific working principles behind a BJT. A transistor is essentially a semiconductor device with physical properties that make it ideal for amplifying or switching electric current and other signal. At the heart of this device is a doped semiconductor with engineered properties to alter its conductivity for a particular use. A BJT is a type of transistor with two major semiconductor materials that constitute three major areas or regions, each doped according to requirements. This architectural characteristics of a BJT brings forth effective applications in implications or on-off switching operations. Nonetheless, understanding BJT as a switch requires understanding the working principles underneath the device, the functions of each of the three major regions within this transistor, and the role of electron movement or current flow in the switching mechanism*

Keywords: *BJT-collector-emitter-base-collector voltage*

Date of Submission: 22-12-2017 Date of acceptance: 12-01-2018

I. Introduction

A bipolar junction transistor or BJT is a fundamental electronic component because of its expansive and essential applications across numerous electronic devices such as radio transmitters (1), televisions, desktop computers, and portable computing devices such as smartphones and tablet computers, among others. This component is specifically found on the circuit boards or logic circuits of electronic devices. Prime modern examples of BJT applications are microprocessor and solid-state memory storage hardware technologies. Note that because it is a particular type of transistor, the primary applications of BJT center on either amplifying electronic signals or switching signals and electrical power(3). The special physical properties of a BJT that further translate into operational reliability and cost efficiency make it a suitable electronic component for switching operation or for opening and closing a circuit. This explanatory research paper examines and discusses the working principles behind the bipolar junction transistor and its specific application as a switch.

II. Bipolar Junction Transistor as a Switch

A Historical Overview of Transistors

The invention of transistors has revolutionized the modern world because it was primary instrumental in launching the electronics age and further supporting the digital information age. Note that it was physicist Julius Edgar Lilienfeld who first filed a patent containing the working principles for an earlier version of a field-effect transistor or FET in 1925 in Canada(5). He also filed a similar patent in 1925 and 1928 in the United States. This FET From Lilienfeld was a metal-oxide semiconductor that used copper sulfide as the semiconductor material and alumina as gate insulator (Colinge & Greer, 2016). German electrical engineer and inventor Oskar Heil also filed a patent for a metal-oxide semiconductor FET around the 1930s in Britain. His patent described the prospect of applying an electric field in a semiconducting material to control resistance (Arns, 1998). The separate patents from Lilienfeld and Heil were meant to introduce an invention that was intended as a solid-state replacement for the triode. Furthermore, the term "transistor" was not used in their patents and there were no operational device made from their descriptions of working principles (Lilienfeld, 1930; Lilienfeld, 1932; Heil, 1934)(6).

It was in 1947 when one of the first transistors was actually invented and produced at Bell Labs of telecommunication company AT&T by William Shockley and coworkers John Bardeen and Walter Brattain(7). Their device was based on experiments that resulted from observing the production of a signal with an output power greater than the input power upon applying two gold point contacts to a crystal of germanium. This device was eventually called a point-contact transistor (Colinge & Greer, 2016). Note that the term "transistor" was first coined at the Bell Labs, specifically by American engineer John R. Pierce who described the device introduced by Shockley et al. and other similar devices or concepts as transfer resistor. Shockley et al. expanded further the ideas behind the transistor through further research and by sharing the licenses to other researchers.

In 1956, Shockley, Bardeen, and Brattain received the Nobel Prize in Physics for their discovery of the transistor effect (Jones, 2014).

The transistor is a revolutionary invention nonetheless. To further understand its impact, it is worth mentioning that prior to its discovery and subsequent widespread applications, the triode was the widely used electronics-related device(4). To be specific, a triode is an electronic amplification device characterized by three electrodes housed inside a vacuum tube that was widely used in radio devices and early television appliances. The problem with the triode is that it is a bulky and fragile device that consumes a considerable amount of energy when under operation (Tyne, 1977). The introduction of the transistor marked a new era in electronics because it was instrumental in further driving the transition from triodes and other electron vacuum technologies to solid-state electronics. It eventually overtook the worldwide sales of vacuum tubes in 1959 due initially to the efforts of Japanese manufacturers that used the patents from Bell Labs to mass produce transistors and promote their applications in consumer electronics products and telecommunications equipment. As a specific example, the transistor radios made in Japan became popular across the world during the 1950s and 1960s. These products were hot commodities that time. The Japanese also introduced other all-solid-state consumer electronics products and telecommunications equipment that dominated their respective global market shares for years to come (Huurdeman, 2003). Throughout the years since it was first patented, invented, and public introduced, the transistor has become the fundamental component of modern electronic devices.

The Working Principles Behind BJT

Understanding the working principles behind a bipolar junction transistor requires understanding the essential or basic principles behind the transistor, in addition its properties. Thus, as a backgrounder, a transistor is a solid-state electronic component made primarily of semiconductor materials used generally to control comparatively large electrical current between two areas of materials made of semiconductor crystals using a very small current or voltage that s applied to a transitional intermediate area. This basic working principle behind a transistor allows it to act as an amplifier of electronic signals or as a switch for signals and electrical power (Mims, 1990). Take note that both operational purposes of a transistor, specifically as an amplifier and as a switch, were originally reserved only for electron vacuum tubes. A transistor operates better than vacuum technologies because it is relatively smaller and mechanically more robust, thus making it more durable because it produces little heat when compared against, say for example, a triode (Mims, 1990; Amos, 1981)(8).

Another critical fact to consider when understanding the working principle behind a transistor is that it is a semiconductor. By definition nonetheless, a semiconductor is a material that has an electrical conductivity value that is considered both a conductor that is similar to a copper and as an insulator that is similar to a glass (Amos, 1981; Grundmann, 2010). Shockley from Bell Labs concluded through a series of studies that the only semiconductor materials suitable for electronics applications were germanium and silicon (Lojek, 2007). Nonetheless, the first utilized semiconductor material for transistors was elemental germanium. This element is obtained as a byproduct of metal refining. The problem with germanium is that transistors built from this materials tend to have limited output power and are prone to current leakages at high ambient temperature, thus resulting to severe limitations in applications. To solve the problem, elemental silicon replaced germanium as a primary semiconductor material (Amos, 1981). Note that one of the first use of silicon as a material for transistor was around 1950s and 1960s when the first metal-oxide FET were developed and used (Colinge & Greer, 2016). Nevertheless, the aforementioned primary physical property makes a semiconductor an ideal material to use in electronic components, especially those components that generally function as mediums for conducting and controlling electric current and other related signals(9).

It is important to note that the usability of a semiconductor depends on the properties of the primary material employed. Regardless of the elemental material used, the conducting properties of a semiconductor can be altered through the process of doping. This process basically involves introducing impurities into the crystal structure of the elemental material. Without the impurities, the semiconductor material has concentrations of electrons and electron holes that are equivalent under thermal equilibrium. For example, a pure silicon, which is also referred as an intrinsic semiconductor, has a low electrical conductivity and thus, is essentially useless because it needs to have electrons within it to absorb some energy and become free electrons to conduct electricity. Through doping, this silicon can be inserted with phosphorus so that one electron is free to move in the entire system or with boron so that there will be a vacant position for an electron called electron hole. A doped semiconductor or extrinsic semiconductor with free moving electrons is called an n-type n-doped semiconductor or simply N. On the other hand, a doped semiconductor or extrinsic semiconductor with electron holes is called a p-type or p-doped semiconductor or simply P. An n-doped semiconductor is an electron charge carrier while a p-doped semiconductor is a hole charge carrier. The doping of a semiconductor is essentially tantamount to the engineering of a transistor. In other words, a doped semiconductor is a transistor in itself and the introduction of impurities allow it to conduct electricity.

Using BJT as a switch

Shockley was also credited as the first to make and patent the first bipolar junction transistor or BJT in 1948 while working at the Bell Labs (Colinge & Greer, 2016). This particular type of transistor features both electron charge carriers and hole charge carriers, and thus, n-doped semiconductor and p-doped semiconductor and was initially introduced as an alternative to field-effect transistors (Moskowitz, 2016)(11). Furthermore, the general structure of a BJT consists of regions or areas called the base, collector, and the emitter. The base is primarily responsible for activating the transistor while the collector serves as the positive lead and the emitter serves as the negative lead (Mims, 1990). Thus, within a BJT, electric current or related signal passes through these three regions. The constitution and resulting property of the semiconductor influences the behavior of this electric current or signal.

There are two specific types of BJT. These are NPN and PNP. In an NPN BJT, the entire material is engineered by having its both ends made of n-type semiconductors and the middle area made of an p-type semiconductor. On the other hand, a PNP BJT has both ends made of p-type semiconductors and the middle area made of n-type semiconductor (Mims, 1990). To illustrate the specific working principle, take note of an NPN BJT as an example.

With an NPN BJT and the application of current through an external source such as a battery, electrons travel from the base p-type semiconductor material to the emitter region made of n-type semiconductor material. This electrons travel further to the base made of p-type semiconductor and branches out to travel again to the emitter region or the collector region made of n-type semiconductor. Electrons that moved to the collector region will further travel back to the emitter region. The entire process illustrates a two-fold cycle in which the first cycle of electron movement involves the base-emitter loop and the second cycle involves the emitter-base-collector loop. In this process, any output device attached between the paths running from the base to the emitter will demonstrate faint reproduction of energy or signal. On the other hand, any output device attached between the paths running from the collector and emitter will demonstrate amplified reproduction of energy or signal. This is because more electrons are collected in the collector region than the base region.

Regardless if it is an NPN BJT or a PNP BKT, the base region is always doped and the geometric profile is always thin. This is illustrated in Figure 1 and Figure 2 that both provided a schematic structure and circuit symbol for NPN and PNP BJT. This means that it is made of a thin extrinsic semiconductor material. Nonetheless, this gives the base region a property that allows more electrons to enter before all electron holes are filled, thus providing more movement of free electrons to the adjacent region—either to the collector region or the emitter region—or to any available medium for conductivity. In consideration of this, the base region should always be sufficiently thin and carriers can dominantly reach the collector region by diffusion.

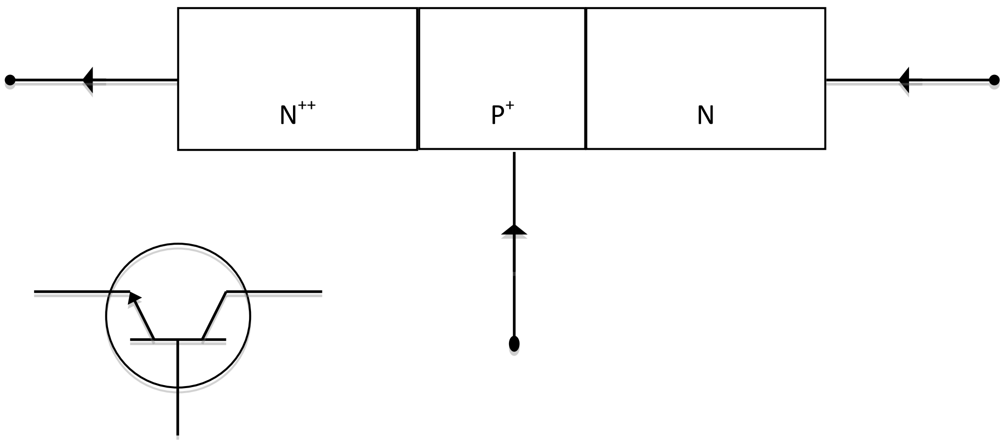

Figure 1. Schematic Structure and Circuit Symbol for an NPN Type Bipolar Junction Transistor

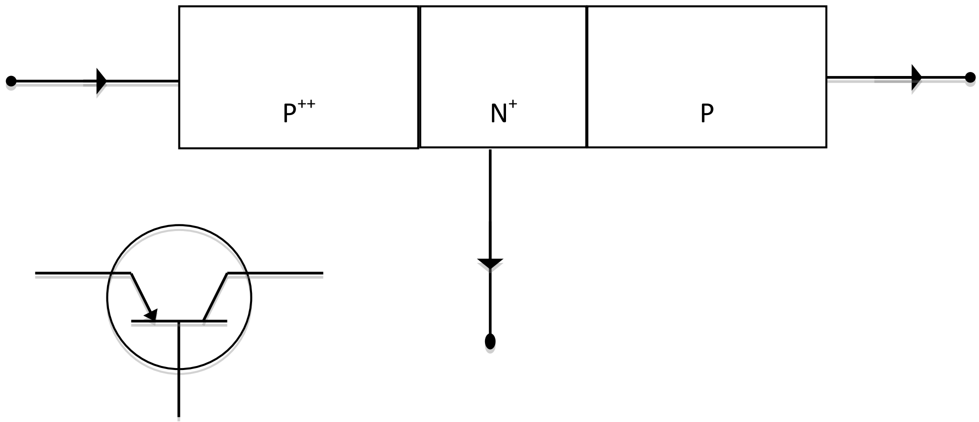

Figure 2. Schematic Structure and Circuit Symbol for a PNP Type Bipolar Junction Transistor

Note that transistors, including BJT are used either as amplifiers or as a switch. When used as a switch, the transistor has two states: either an on state or an off state. In an on state, the transistor behaves as a conductive material that has a significant collector current and a very low collector-emitter voltage. On the other hand, while an the off state, the transistor behaves as a non-conductive material with a negligible collector current and a collector-emitter voltage equal to the voltage of the supply (Amos, 1981). Under a level or point of saturation, the ratio of the current in the base region to the current of the emitter-collector regions is constant. However, when under saturation, the entire BJT achieves its highest possible point of conductivity, thus undergoing an on state. In addition, while under saturation, additional increases in the base current will fail to increase the emitter-collector current (Mims, 1990). Take note that a relatively small base current is enough to saturate the entire BJT to enter an on state and further into full conductivity, thus stimulating it to behave as a type of solid-state on-off switch. This entire mode of operation of BJT, specifically when used as a switch, has many applications, especially in digital electronics circuits.

III. Relevant Equations in BJT Applications

The measurement of the effectiveness of a BJT is solely dependent on the proportion of electrons that are able to move from the base region and reach further to the collector region. The level of doping of the different regions within a BJT, especially the heavy doping of the emitter region and the light doping of the base region result in more electrons moving from the emitter and into the base(4). When used as a switch, the operability of BJT is measured by the amount of current or voltage it needs to enter saturation mode or in other words, to under go an on state and become a full conductive material. The following examples illustrate different equations.

An Example of NPN BJT Switch

Consider the following example values: (1) Base Resistance as symbolized by Rb is equal to 50 k ohm; (2) Collector Resistance as symbolized by Rc is equal to 07 k ohm; (3) Collector Current Voltage as symbolized by Vcc is 5V; and (4) Beta Value is 125.

Taking into consideration the aforementioned values, below is the initial calculation:

$I_c = V_{cc}/R_c$ when $V_{CE} = 0$
$I_c = 5V/0.7k$ ohm
$I_c = 7.1$ mA

Additional calculation for the base current:

$I_b = I_c / \beta$
$I_b = 7.1$ mA$/125$
$I_b = 56.8$ µA

Based from the calculations above, the maximum value of the collector current in the NPN BJT circuit is 7.1 mA. Note that this is when Vce is equal to zero while the base current to which the collector current flows is at 56.8 µA. These results essentially mean that increasing the base current beyond 56.8 µA will result in the entire transistor to enter into a saturation mode and thus, into an on state with full conductivity properties.

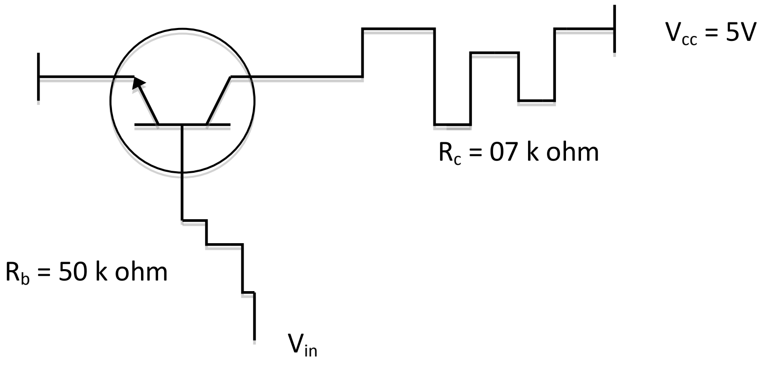

Figure 3. Schematic representation of an NPN BJT Used As A Switch

An Example of PNP BJT Switch

A PNP BJT switch works essentially the same as an NPN BKT switch. The only difference is that current flows from base region of the PNP BJT. In real world applications, this type of transistor switch is commonly used in negative ground configurations. It is important to note that the base region of a PNP BJT is always negatively biased when compared against the emitter region. When used as a switch, a more negative base voltage results in the flow of base current. A low voltage or a more negative voltage makes a PNP BJT a short circuit.

The following equation is used calculating the base and collector currents, and most especially the current requirements:

$$Ic = Ie - Ib$$
$$Ic = \beta . Ib$$
$$Ib = Ic / \beta$$

Consider the following example values to illustrate the use of PNP BKT as a switch: A particular load requires 100 mA current while the beta value of the transistor is 100.
Taking into consideration the aforementioned values, below is the initial calculation:

$$Ib = Ic / \beta$$
$$Ib = 100\text{ mA} / 100$$
$$Ib = 1\text{ mA}$$

The aforementioned result of the calculation revealed that the application of 1 mA current to the base region would result in the PNP BJT to enter saturation mode and in turn, entering into an on state with full conductivity properties.

IV. Conclusion

A bipolar junction transistor or BJT is a type of transistor that is also regarded as a fundamental electronic component because of its expansive and essential applications across numerous electronic devices. Aside from applications for amplifications, another notable application of BJT is in switching or more appropriately, controlling the flow of electric current or other related signals. Stimulating the switching properties of an BJT requires the application of suitable amount of current. This application of a current comes from the use of external source that would stimulate the movement of electron form the external and within the BJT component and across the entire circuitry. The movement of electrons represent current flow. With a suitable amount of current, the BJT enters a mode called saturation that would enable it to enter further the on state. Under this state or mode, the BJT becomes a fully conductive material. Removing and applying the current represents a transition between an on state and an off state. The entire process represents a solid-state on-off switching mechanism. Note that current application and reapplication can transpire in cycles over short periods, thus the on-off cycle can be fast. The entire process is applied in digital electronics

References

[1]. Amos, S. W. (1981). Principles of transistor circuits: Introduction to the design of amplifiers, receivers, and digital circuits. (6th ed.). London: Butterworths
[2]. Arns. R. G. (1998). The other transistor: Early history of the metal-oxide semiconductor field-effect transistor. Engineering Science and Education Journal. 7: 223-240
[3]. Colinge, J. P. & Greer, J. C. (2016). Nanowire transistors: Physics of devices and materials in one dimension. Cambridge: Cambridge University Press
[4]. Grundmann, M. (2010). The physics of semiconductors: An introduction including nanophysics applications. (2nd ed.). Berlin: Springer Science and Business Media
[5]. Heil, O. (1934). GB439457 (A): Improvements in or relating to electrical amplifiers and other control arrangements and devices. Germany: European Patent Office
[6]. Huurdeman, A. A. (2003). The worldwide history of telecommunications. New Jersey: John Wiley & Sons
[7]. Jones, C. (2014). The technical and social history of software engineering. New Jersey: Addison-Wesley
[8]. Lilienfeld, J. E. (1930). US 1745175 A: Method and apparatus for controlling electric currents. United States: U.S. Patent and Trademark Office
[9]. Lilienfeld, J. E. (1932). US 1877140 A: Amplifier for electric current. United States: U.S. Patent and Trademark Office
[10]. Lojek, B. (2007). History of semiconductor engineering. New York: Springer Science and Business Media
[11]. Mims, F. M. (1990). "Transistor." In Lexicon Universal Encyclopedia. New York: Lexicon Publishing Inc.
[12]. Moskowitz, S. L. (2016). Advanced materials innovation: Managing global technology in the 21st century. New Jersey: John Wiley & Sons
[13]. Tyne, G. F. J. (1977). Saga of the Vacuum Tube. (1st ed.). Indiana: Howard W. Sams & Co.

Ali Habeb Aseeri "Bipolar Junction Transistor as a Switch." IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) 13.1 (2018): 52-57.