DEPARTMENT OF WEAPONS TRAINING
LOWRY AIR FORCE BASE
COLORADO

T R A N S I S T O R S

IN-SERVICE TRAINING COURSE

INTRODUCTION TO SEMICONDUCTOR DEVICES

CHAPTER 1

22 NOVEMBER 1960

COPY ____ OF 100 COPIES

FOR INSTRUCTIONAL PURPOSES ONLY

CHAPTER 1

INTRODUCTION TO SEMICONDUCTOR DEVICES

Author: A/1C Robert J. Widlar

Abstract -- The idea of conduction in solids by both holes and electrons is introduced. The existence of two distinct types of current carriers is demonstrated using the Hall effect. Without explaining the origin of these current carriers, an elementary description of junction diodes and two-junction transistors is given. Finally, it is shown how a transistor can be used in an amplifier circuit.

INTRODUCTION

It is difficult to find a piece of modern electronic equipment that does not incorporate some kind of semiconductor device, such as; a selenium rectifier, a crystal diode, or a transistor. This is true of almost all commercial, industrial, and military equipment because these semiconductor devices are generally smaller, more efficient and in some cases, more reliable than components used in the past.

The existence of semiconductors has been known for many years; but, until recently, they were considered a useless transition between conductors and insulators. They first enjoyed widespread use in the early days of radio when it was discovered that the contact between a fine wire and some substances - such as galina, iron pyrite, and silicon carbide - exhibited rectifying properties. The reason for this effect was not known. Therefore, vacuum tubes pushed semiconductors into the background, because conduction in a vacuum was well understood; and an intelligent approach could be used in the design and perfection of tubes.

The influx of radar and other complex electronic systems during World War II spurred the development of semiconductor diodes; but, because of the pressing need, an experimental approach was used. This effort resulted in practical silicon and selenium rectifiers; however, it did not provide a major breakthrough, as the phenomenon of conduction in solids was still not completely understood. The major advance came in 1948 with the development of the transistor, a semiconductor amplifying device. It appeared that the transistor could perform the same function as a vacuum tube, while using one hundredth the power and occupying far less space. This added impetus to semiconductor research, and workable theories on the operation of semiconductors were evolved.

The knowledge gained from the research on transistors was applied to other devices, and soon silicon and germanium diodes were made with ratings exceeding those of the best copper oxide and selenium rectifiers. Also, the silicon solar cell was developed which could convert the energy of the sun directly into electricity with reasonable efficiency. Diode amplifiers for example, the varactor and the tunnel diode were made which could perform at frequencies above the range of transistors. These and many other devices benefited from semiconductor research.

It can be seen that semiconductors have assumed an important position in the electronics industry. They are being used with increasing frequency because some semiconductor device might be able to replace a whole circuit of conventional components. It is therefore necessary for anyone connected with electronics to become familiar with semiconductor theory just as he was required to learn the theory of vacuum tubes.

One of the most important fundamentals in the theory of semiconductors is the mechanism of conduction in solids. But in order to understand this phenomenon, it is also necessary to learn the structure of solids. The chemistry and physics of a few important semiconductor materials will be discussed in chapter two.

Furthermore, the characteristics of these substances important to the understanding of practical semiconductor devices will be brought forth. In later chapters, the operation of transistors and other PN junction devices will be given a detailed explanation. Emphasis will be placed on the junction diode and the two-junction transistor because these will illustrate the more significant results that can be realized with semiconductors. Still later, the techniques used in the manufacture of these devices will be introduced; and most of the processes currently used in the production of diodes and transistors will be briefly described. The last chapter will be devoted primarily to the circuit applications of transistors although other devices of current interest will be mentioned.

Before going into a more exact explanation of semiconductor phenomenon, a beief preview of conduction in solids will be given and then related to the operation of the junction diode and the transistor. The purpose of this is to give the reader an opportunity to get an overall view of the material to be covered. Exact explanations and detailed proofs will not be used here.

CONDUCTION IN SEMICONDUCTORS

Semiconductors are a class of materials having conductivities somewhere between those of metalic conductors and insulators. The conductivity of these materials is several orders of magnitude greater than insulators but still very much less than metalic conductors. A simplified picture of conduction in metals attributes conduction to the existence of free electrons which can be moved under the influence of an electric field. It then seems reasonable to assume that free electrons are also responsible for conduction in semiconductors. This is the case for some semiconductor materials, but it has been shown that conduction can also take place by what appears to be a positive electron, or hole. The properties of semiconductors of concern at this point are that the current carriers can be either holes or electrons and that the concentration and type of current carrier can be controlled during production of the material. This much will be assumed here, but to help substantiate these statements a demonstration of the existence of holes follows:

It was found that the current carriers in semiconductors traveled at greater velocities than those in metals. It was thought, then, that these current carriers could be deflected appreciably from their normal path within the semiconductor by the application of a magnetic field. This was indeed the case as is shown in Figure 1.1. A current was passed between two metalic contacts on a block of semiconductor material. Two other electrodes were placed at right angles to the current flow as can be seen from the figure. If the current carriers were electrons, the results shown in Figure 1.1a. could be expected; with no magnetic field applied, the number of electrons reaching the two electrodes, at right angles to the current flow would be equal, and there would be no potential difference between these electrodes. When a magnetic field is applied into the page as shown, the electrons would be deflected upward; and more electrons would reach the upper contact than would reach the lower. The upper contact would then become more negative than the lower, and a potential difference could be measured between them. This was found to be the case for some semiconductor materials. However, with other materials, the opposite effect was observed: When the magnetic field was applied in the same direction, relative to current flow, the upper contact became positive with respect to the lower. This could not happen with electron current carriers. It was then assumed

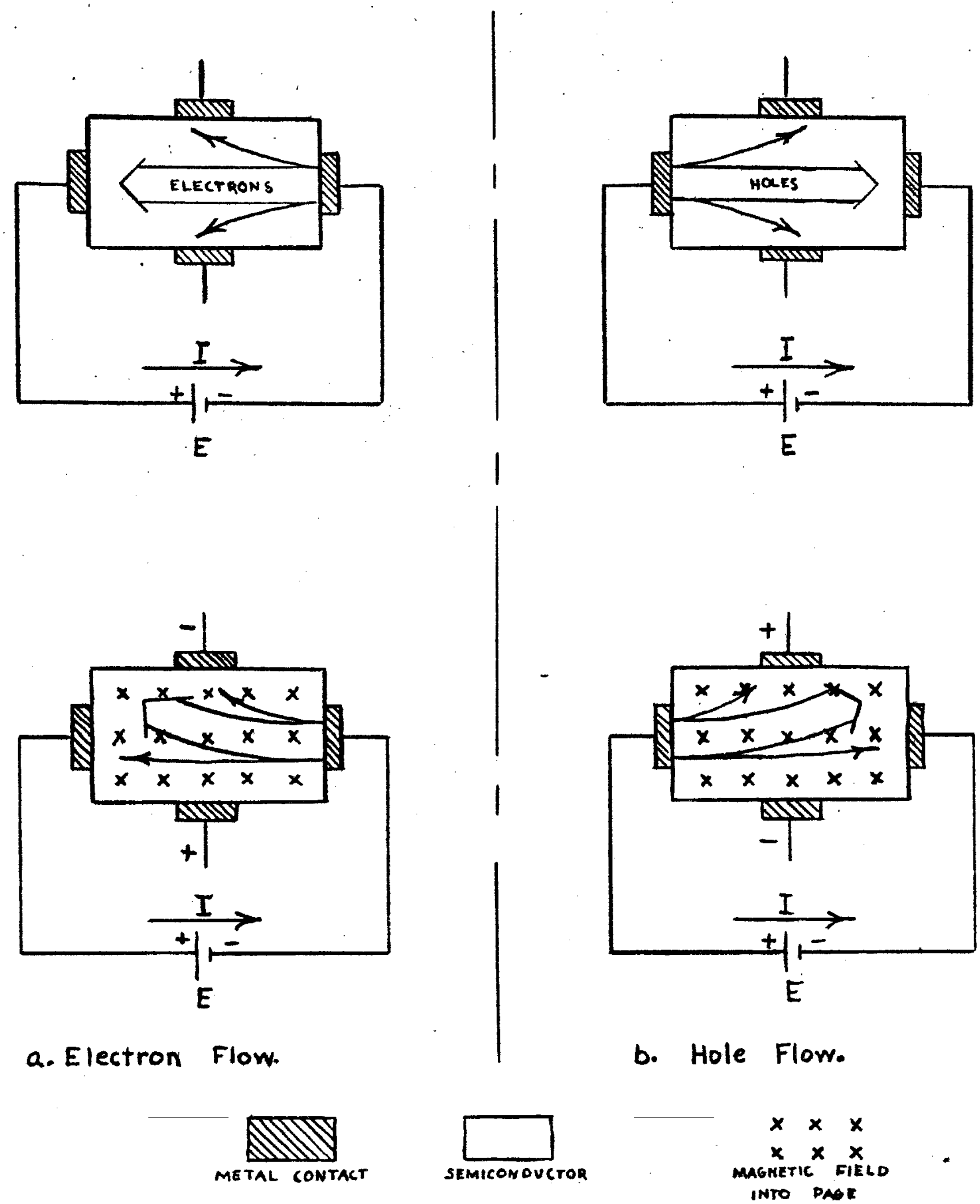


Figure I.I. Using the Hall Effect to Demonstrate Existence of Holes.

that there were two distinct types of current carriers in semiconductors, electrons and holes. This would explain the effect observed in Figure 1.1b.

This is known as the Hall effect, and in practice it is used to measure steady magnetic fields for the detection of submarines; for tape recorder heads; and in circulators, gyrators, and isolators where the direction of current flow is altered by a magnetic field.

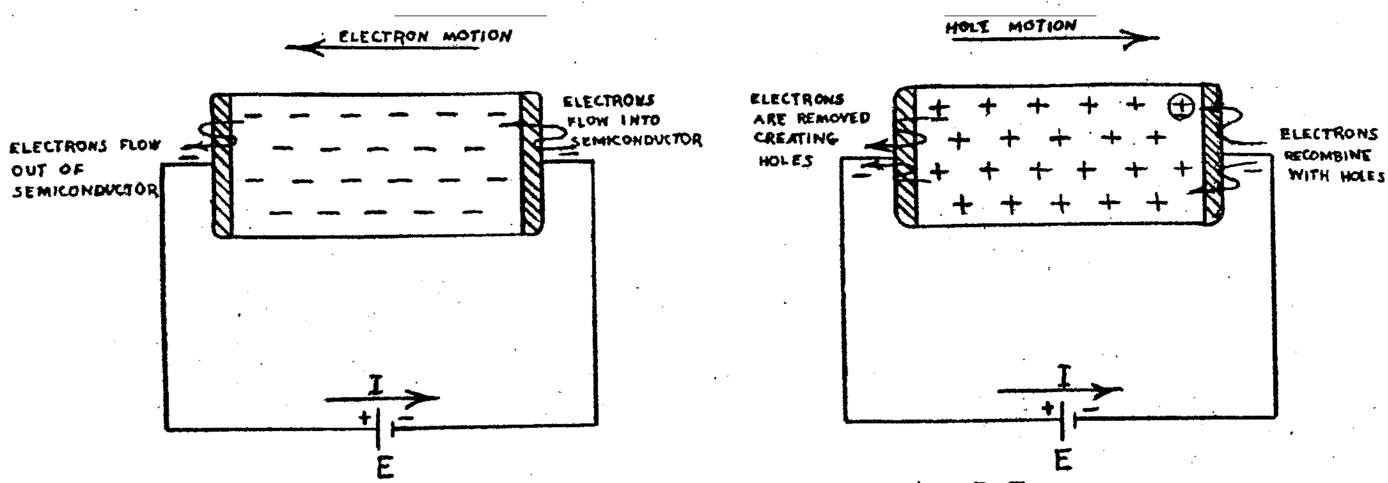
When the current carriers in a semiconductor are free electrons (Figure 1.2a) conduction is relatively easy to visualize. When a voltage is applied across the semiconductor, the electrons will move from negative to positive and into the external circuit. When an electron passes out of the semiconductor into the external circuit at the positive terminal, another electron immediately flows into the semiconductor at the negative terminal.

When holes carry the current in a semiconductor, there must be a transition at the circuit connections. This is because the current carriers in the external circuit are electrons while the current carriers in the semiconductor are holes. If an electric field is applied, the holes will flow from positive to negative until they reach the negative terminal as shown in Figure 1.2b. When a hole does reach the negative terminal, it captures a free electron from the external circuit and the hole disappears. At the same time the external circuit recovers an electron from the positive terminal, creating another hole. This new hole will travel to the negative terminal repeating the process. By this mechanism, current is carried by holes within the semiconductor, and by electrons in the metalic conductors of the external circuit.

Some helpful rules for predicting conduction in semiconductors will be stated here without proof. Whenever an electron is removed from a semiconductor under the influence of an applied electric field, it must soon be replaced by an electron from the potential source (external circuit) to keep the overall number of electrons constant. Similarly, whenever a hole is removed from a semiconductor by recombination with an external electron, the hole must be replaced by the removal of another electron to keep the overall number of holes constant. To appreciably change the number of holes or electrons would require high electric potentials which are never encountered in practice. Another detail concerns the production of holes by the removal of electrons: This can only occur to any appreciable extent at the terminals (contacts) of the semiconductor. Hole generation occurs at imperfections in the physical structure, and there are normally very few imperfections within the semiconductor. However, the surface conditions resulting from attaching a contact create many imperfections. Hole generation can therefore take place far more readily at the contacts than it can within the semiconductor.

JUNCTION DIODES

It can be seen then, that there are two distinct classes of semiconductors: N type in which the current carriers are free electrons, and P type in which the current carriers are equivalent to positive electrons (holes). This fact can be used to build many useful devices, one of which is the junction diode.



a. N Type.　　b. P Type.

Figure 1.2. Representations of P and N Type Semiconductors Indicating Mechanism of Conduction.

A junction diode is made by somehow joining together P and N type semiconductor materials. A diagram of a junction diode is shown in Figure 1.3. Contacts are made to both the P and N type materials so that the device can be connected to an external circuit. A device constructed in this manner will exhibit a low resistance to current flow in one direction and a high resistance in the other. A mechanical view of this action follows:

When P and N type semiconductors are joined, a junction is formed. The current carriers on one side of this junction are electrons, and those on the other are holes, as can be seen from Figure 1.3. When a voltage of the polarity shown in Figure 1.4 is applied across the diode, it will conduct. This is because the electrons and holes will be forced across the junction in opposite directions, establishing a current within the diode and also in the external circuit. More exactly, the applied voltage will move the electrons from negative to positive, across the junction, and into the P region. Once the electrons flow into the P region, two things can happen: The electrons can move through this region and into the external circuit via the positive diode contact; or they can recombine with the holes in the P region. If an electron does recombine with one of the holes, another electron will be drawn out of the P region, at the positive terminal, and flow into the external circuit; still another electron will flow into the diode at the negative terminal, to keep the overall number of holes and electrons constant in the P and N type regions. In either case, a current flow is established through the diode by the electrons.

Similarly, the holes will also contribute to current flow. Under the influence of the applied voltage, the holes will move across the junction into the N region. After crossing the junction, again two things can happen; the holes can move through the N region and recombine with electrons at the negative terminal of the diode, or they can recombine with electrons within the N region. In either case, when a hole recombines with an electron, another hole will be created at the positive terminal by the removal of an electron to the external circuit; and another electron will flow in at the negative terminal again to keep the overall number of holes and electrons constant. This action establishes a current through the diode by hole conduction.

It can be seen that a current will flow across the junction by both electron and hole conduction. Consequently, current will flow in the external circuit. The polarity of applied voltage that will cause this conduction is called forward or conducting bias.

If a potential of the polarity shown in Figure 1.5 is applied across the junction diode, no current will flow. The holes and electrons will be drawn away from the junction by the applied voltage, and no current carriers will flow across the junction. It can be seen from Figure 1.5 that the electrons will be influenced by the applied voltage to move away from the junction and toward the positive contact of the diode. However, a continuous current flow cannot be maintained by

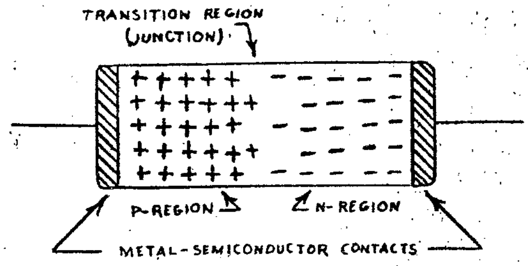


Figure 1.3. Diagram of PN Junction Diode.

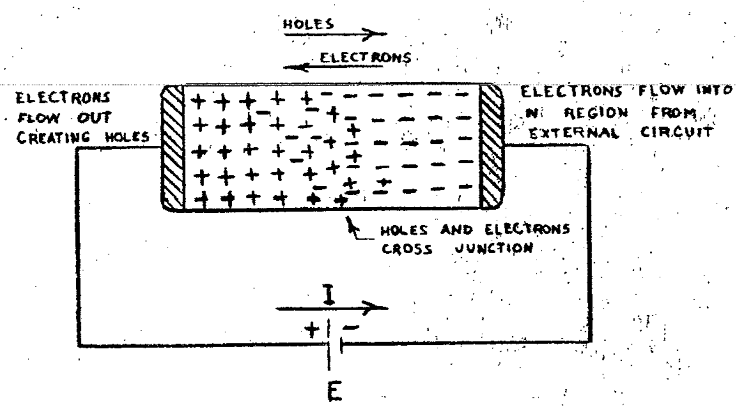


Figure 1.4. Action in a Junction Diode for Forward Bias.

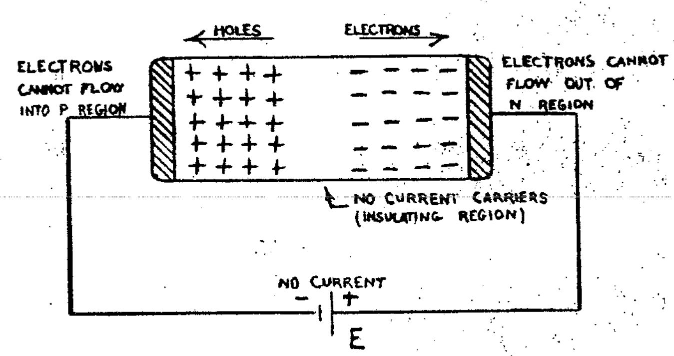


Figure 1.5. Action in a Junction Diode for Reverse Bias.

the electrons. If the electrons did flow out the positive terminal, they could not be replaced by electrons from the P region because there are no free electrons present.[1] Therefore, no appreciable amount of electrons can flow out the positive terminal without greatly reducing the number of electrons within the semiconductor.

A similar situation exists for the holes. The holes are moved away from the junction, and toward the negative terminal, under the influence of the applied potential. But again, the holes cannot recombine with electrons at the negative terminal to any appreciable extent. This would greatly reduce the number of holes because the holes cannot be replaced from the N region. Therefore, no current can be maintained through the diode by hole flow.

To summarize the description of a nonconducting P-N junction; When the voltage is first applied, a negligible current will flow. This makes the internal potential of the diode equal to the applied potential, by removing a minute quantity of holes and electrons. However, a continuous current cannot flow in the external circuit because there cannot be current flow within the diode: the holes and electrons cannot cross the junction. The holes and electrons will actually be pulled away from the junction forming an insulating region in which there are no current carriers. This condition is known as reverse bias.

The junction diode is used quite frequently in practice. It can replace vacuum tubes in a wide variety of circuits. This diode requires no filament power and physically, it is much smaller than equivalent vacuum diodes. The electrical characteristics of these diodes leaves little to be desired; the voltage drop across a typical junction diode is approximately 0.5 to 1.0 volt at rated current. This is much less than the 50 to 200 volt drop across high vacuum rectifiers and the 5 to 50 volt drop across gas diodes. This means that a small semiconductor diode functioning as a power rectifier can produce more D-C power at a higher efficiency (99 + percent) than can its larger vacuum counterparts. Semiconductor diodes are also used in low power and high frequency applications, and they are unchallenged in computer applications where thousands of diodes must be crowded into a small space.

Semiconductor diodes are made with a wide range of voltage and current ratings. Diodes are available with current capacities ranging from 10 ma to 500 amperes, and inverse voltage ratings of 10 to 1000 volts.

These diodes do have a small reverse current, but in most circuits it is of negligible magnitude; a fraction of a microampere for low current silicon diodes and a few milliamperes for high current germanium power rectifiers.

[1]In addition, the generation of appreciable numbers of holes and electrons cannot occur at the junction (or within the semiconductor) if it is properly made because there are few imperfections in the material.

TRANSISTORS

Although the development of the junction diode was an important contribution to the field of electronics, it represents only one component in a large family of solid-state devices. Another member of this family is the transistor, a semiconductor amplifying device. The operation of the transistor is closely related to the junction diode.

The first transistors made were point contact devices, but almost all of these are now obsolete. Hence, only junction transistors will be discussed here as they are by far the most popular type, both in theoretical considerations and in practical usage.

The physical construction of a PNP junction transistor is shown in Figure 1.6. A thin N type semiconductor wager is sandwiched between two larger P type slabs. Contacts are made to these three regions, and they form the basic elements of the transistor; emitter, base, and collector. The drawing is not made to scale; the base region is much thinner than is shown in the drawing. In a typical unit the cross section of the base region is 1/4 inch square while its thickness is only 0.001 inch. The emitter and collector are therefore separated by an extremely thin N type region. As will be seen, this is essential for efficient transistor operation.

As is indicated on the diagram, the current-carrier densities in the three regions are not alike. The emitter is made from a very high conductivity semiconductor. This means that there is a relatively large number of holes available for conduction. The base is a low conductivity semiconductor in that there are relatively few electrons available for conduction. The collector is a moderate conductivity P type material. The concentration of current-carriers in the collector is not too important for an elementary discussion. However, the high ratio of carriers between the emitter and base regions is essential as will be seen.

The collector-base and emitter-base junctions satisfy the requirements previously set down for a junction diode in that P and N type materials are joined together. The behavior of these junctions will indeed be similar to that of a diode, but they will be put to a somewhat different use.

In Figure 1.7 a voltage is applied to the transistor, negative on the collector and positive on the emitter. Under these conditions, the collector-base junction will behave like a reverse-biased diode: the holes and electrons will be pulled away from the collector junction, and no current will flow. This voltage is of the correct polarity to forward bias the emitter-base junction, but the emitter junction and collector junction are electrically in series. Hence, all the applied voltage appears across the collector junction and there is none left to forward bias the emitter junction. Under these circumstances, no current flows in the transistor; and it is said to be cut off.

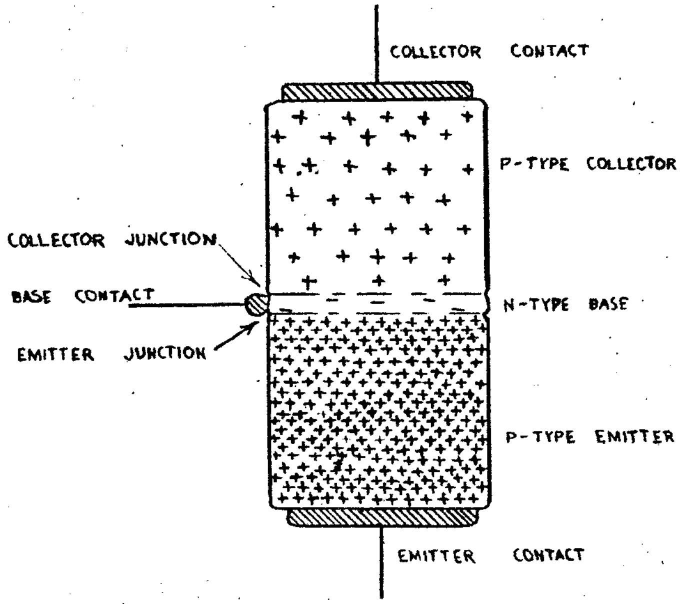


Figure 1.6. Diagram Showing Construction of PNP Transistor.

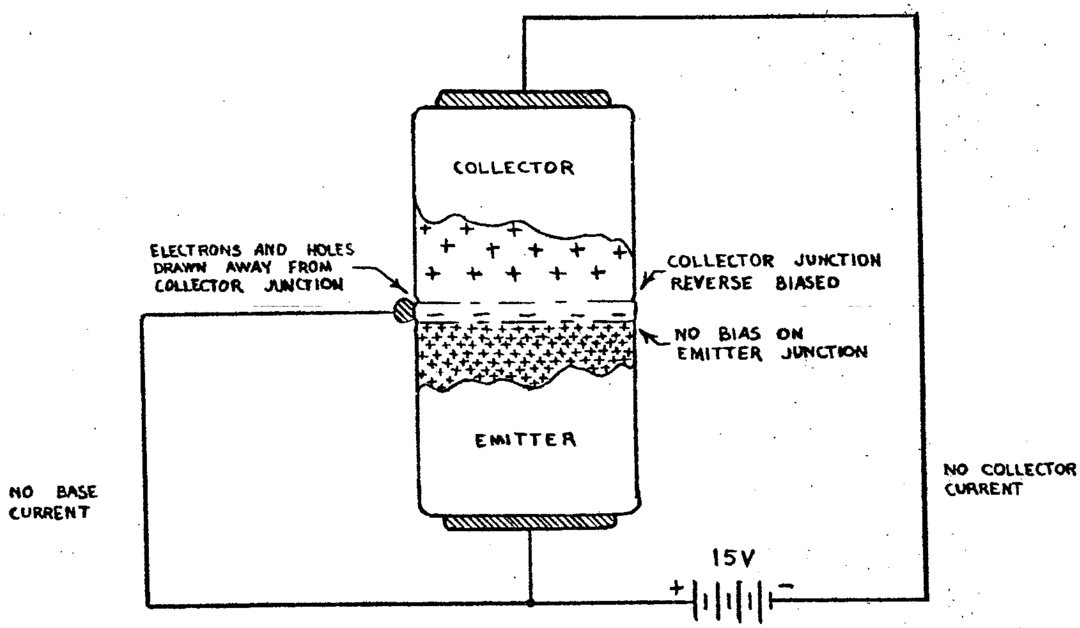


Figure 1.7. PNP Transistor in Nonconducting State.

If a forward bias is applied to the emitter junction, as in Figure 1.8, the amplifying action of a transistor becomes apparent. Feeding 0.2 volt at 1.0 ma into the base caused the collector current to increase from 0 to 50 ma. Furthermore, small changes in base current will result in correspondingly large changes in collector current. For example, if the base current is increased to 2.0 ma (approximately 0.4 volts base voltage) the collector current will increase to 100 ma. The base appears to exercise control over the collector current with a "current gain" of 50. The reasons for this must be investigated.

The amplifying action of a transistor takes place in a relatively small volume including the base and the two junctions. In Figure 1.9 this pertinent area has been redrawn in an expanded view to show this action more clearly.

When a forward bias is applied to the emitter junction, holes will flow across the junction from the emitter to the base; and electrons will flow from the base to emitter, as would be expected from the discussion of junction diodes. Here the similarity to diodes ends. After the holes flow across the emitter junction, there is a force acting on them that has not yet been brought forth. This force is the mutual repulsion that the holes have for each other because of their like charge.[1] When the holes are spilled across the emitter junction, they bunch up on the base side of the junction. Mutual repulsion will cause these holes to diffuse into the base. Because of the extreme thinness of the base, most of the holes will diffuse to the collector junction before they can recombine with electrons in the base or reach the base terminal. Once the holes reach the collector junction they are swept away into the collector by the negative collector voltage. When these holes flow into the collector, they will flow through the collector to the negative terminal. Thus, the forward bias on the emitter has resulted in a collector current.

It remains to be shown that the base current will be very much less than the collector current. Base current is caused by electrons flowing into the base from the external circuit for any of the following three reasons:

1. To replace electrons that have been forced across the emitter junction by the forward bias.

2. To replace electrons that have recombined with holes within the base region.

3. To recombine with holes reaching the base terminal.

This is where the ratio of conductivities between the emitter and base regions becomes important. It will be remembered that there is a far greater number of holes available for conduction in the emitter than there is electrons in the base. Therefore, when the emitter junction is forward biased, a greater

[1]This point will be discussed further in Chapter 2. The diffusion forces are of thermal rather than electrostatic origin.

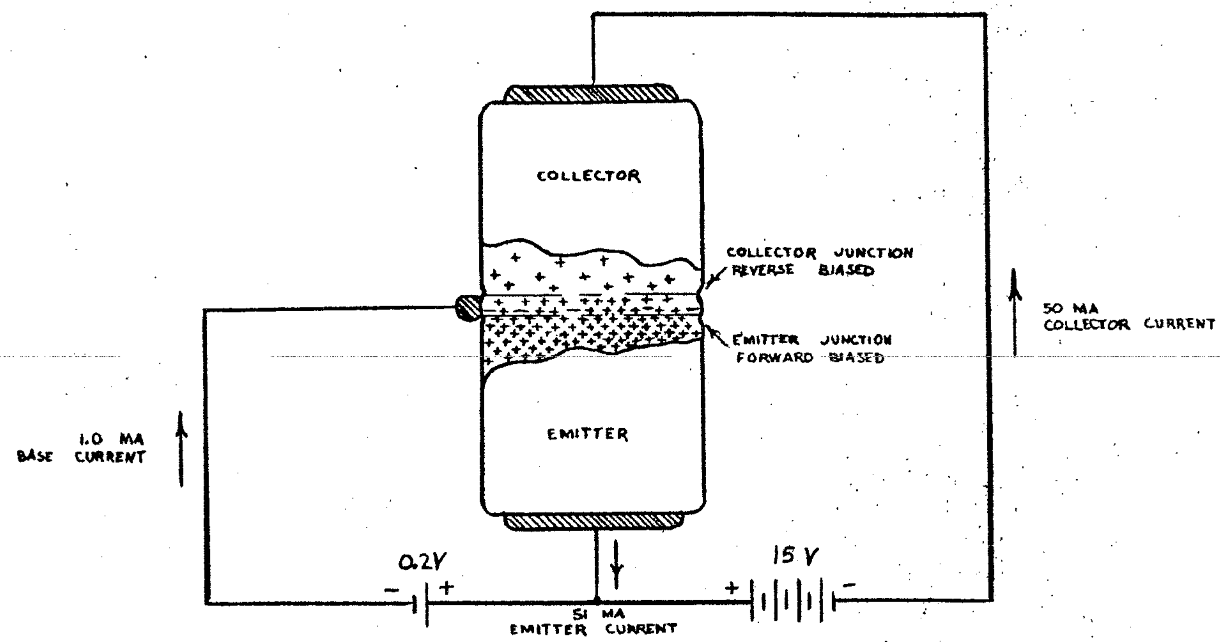
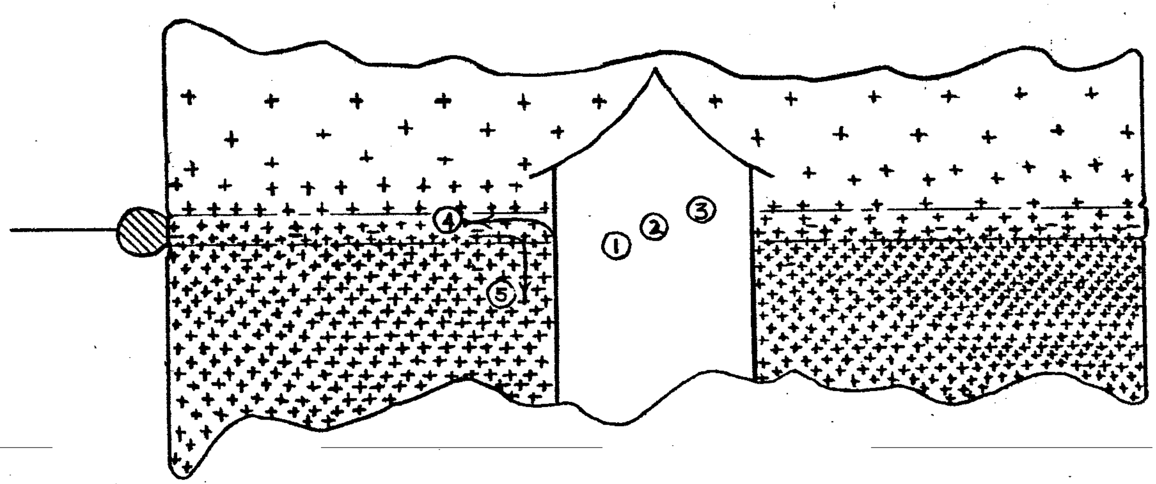


Figure 1.8. Conducting Transistor — Illustrating Control of Collector Current by Emitter-Base Junction.



1- Holes cross emitter junction because of foward bias.
2- Holes diffuse through base.
3- Holes are swept across collector junction by negative collector voltage.
4- Some holes do recombine with electrons or reach base terminal.
5- A far smaller number of electrons cross into emitter.

Figure 1.9. Expanded View of Base Region.

number of holes will flow across this junction into the base than will electrons into the emitter. Moreover, the holes in the base are not likely to recombine with electrons in appreciable numbers, both because of the extreme thinness of the base through which they must travel, and because of the small number of free electron carriers in the base. Finally, the number of holes, reaching the base and recombining with electrons, will be small because the holes will diffuse the short distance to the collector junction before traveling the comparatively long distance to the base contact. In this manner, small currents in the base can control large currents in the collector circuit. The ratio of the collector current to the emitter current is known as the beta (β) current gain of the transistor. In typical units it will range from 10 to 100.

The physics of the transistor will be set aside, for a moment, and its practical applications in electronic circuits will be observed. In particular, the use of the transistor as an amplifier will be discussed.

Figure 1.10 illustrates the PNP transistor connected as a medium-power amplifier. While a schematic representation of this transistor is used, the transistor elements and the directions of current flow are clearly labeled. The input and output waveforms are also shown.

As could be expected, -30 volts is applied on the collector to reverse bias the collector junction.[1] In addition, a small negative bias voltage is put on the base to forward bias the emitter junction. When the emitter junction is forward biased, a 50 ma collector current will pass through the 300 ohm load resistor dropping the collector voltage to -15 volts. The input signal is inserted in series with the bias source so that the forward bias can be increased and decreased alternately by the input signal. The changing forward bias will cause relatively large changes of collector current while there are only small changes in the base current. This varying collector current produces corresponding voltage drops across the series load resistor. The amplifying action of this circuit arises from the fact that small current and voltage changes in the base circuit can produce large variations in collector current. Further, because the collector is reverse biased, comparatively large voltages can be applied to the collector, so the current variations in the collector can cause large voltage changes. The tabulation of characteristics below give the type of performance that can be expected from a transistor amplifier. For this purpose the 300 ohm collector resistor is treated as the load:

1. Current gain --50.

2. Voltage gain --75.

3. Power input --0.025 mw.

4. Power output --94 mw.

5. Power gain --3760, or 36 db.

6. Maximum output --375 mw.

[1]Regardless of the configuration used, the collector junction of a transistor is always reverse biased.

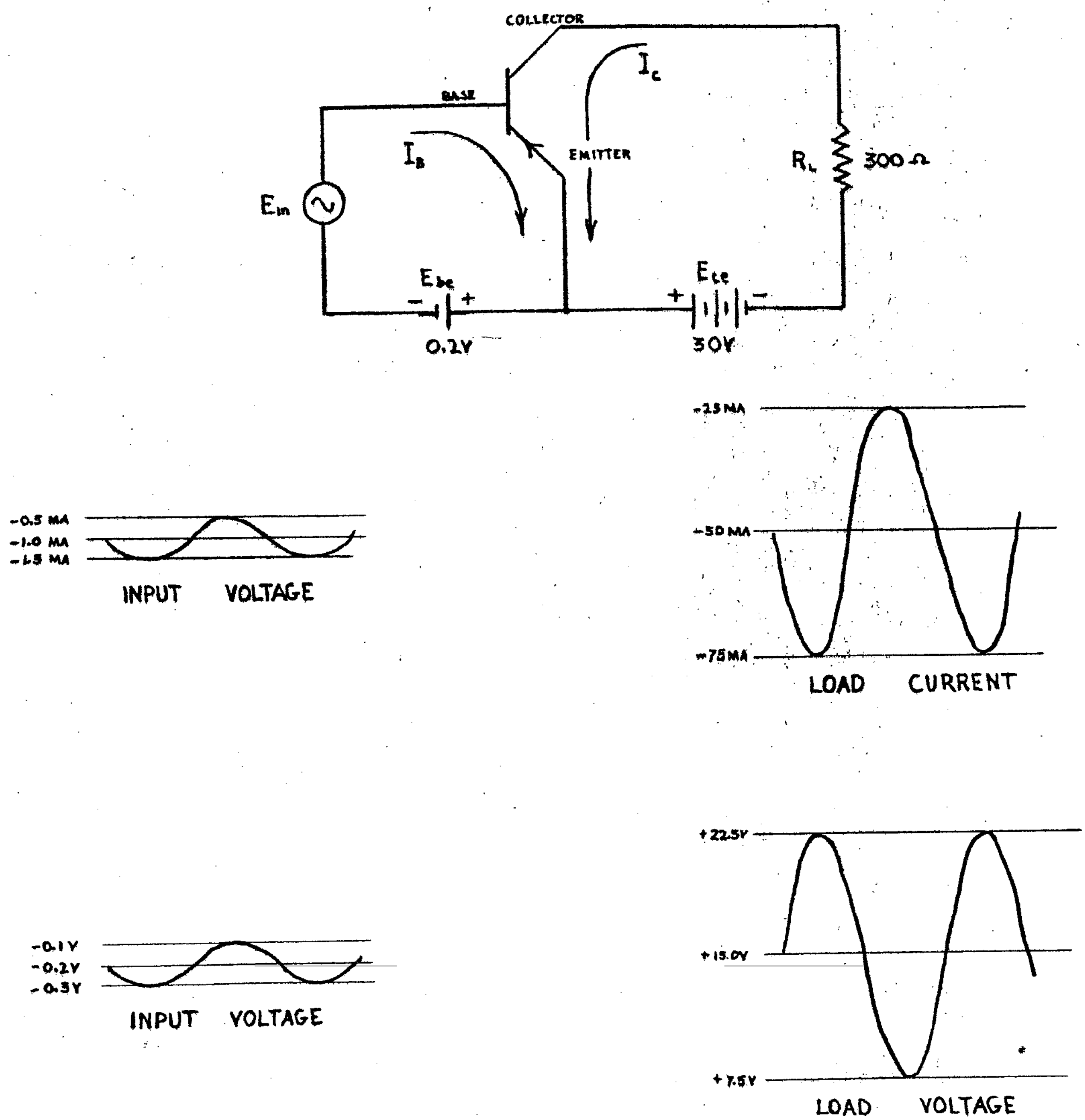


Figure 1.10. The Transistor as an Amplifier.

Although the listed power output of 375 mw may seem low, it must be remembered that this is about the same available from a large battery-powered portable radio.

This single circuit obviously does not begin to describe the possibilities of the transistor. Amplifiers and oscillators can be made that will operate with a d-c input power as low as a few microwatts or, using power transistors, with outputs in the order of several hundred watts. In addition, efficiencies approaching theoretical values can be attained with practical circuits in switching and audio applications. Furthermore, there are NPN transistors which operate in the same manner as PNP units except that the role of the holes and electrons are interchanged. The NPN transistors function with voltages of opposite polarity to the PNP types. This permits design of circuits that would not be possible with vacuum tubes, as this corresponds to having a tube that will operate with a negative plate voltage.

However, transistors do have many limitations that have not yet been mentioned, but these limitations can be overcome with proper circuit design. Therefore, any serious discussion of circuits using semiconductor diodes and amplifiers will be put off until a better understanding of the factors affecting semiconductor performance, mainly temperature and operating frequency, is gained.