DEPARTMENT OF WEAPONS TRAINING
LOWRY AIR FORCE BASE
COLORADO

T R A N S I S T O R S

IN SERVICE TRAINING COURSE

CHAPTER 3

THE PN JUNCTION

4 JANUARY 1961

FOR INSTRUCTIONAL PURPOSES ONLY

CHAPTER 3

Author: A1C Robert J Widlar

THE PN JUNCTION

Abstract - The characteristics of the PN junction diode and several other practical PN junction devices are investigated. The formation of an electric field in an insulating region between the P and N type materials is shown to explain the rectifying properties of such a junction. The effects of temperature on diode characteristics are discussed as are reverse breakdown mechanisms. The dependence of diode capacitance on reverse bias voltage is shown. The operation of various photoconductive and photovotaic cells is described, and the tunnel diode is given an elementary explanation. Both rectifying and ohmic metal-to-semiconductor contacts are mentioned.

GENERAL

The concepts of conduction in solids developed in the last chapter are a powerful tool for the understanding of semiconductor devices. In this chapter, they will be applied to the PN junction which has already been mentioned. Devices employing a single PN junction are capable of performing many useful functions that were not brought out in chapter 1. These functions include: voltage regulation (zener diode), frequency control (voltage variable capacitor), detection of radiant energy (photodiode), direct conversion of solar energy into electricity (solar cell), and the amplification of microwave signals (tunnel diode). An explanation of these devices will more completely describe the behaviour of a PN junction.

Since transistors and many other semiconductor devices employ multiple PN junctions, a study of the PN junction alone can explain certain phenomena which may be equally applicable to these more complex devices but more difficult to isolate in them. This chapter will build a foundation for the coming discussion of transistors.

THE PN JUNCTION

If a slab of P-type semiconductor is brought into intimate contact with a slab of N-type material, a PN junction will be formed. (This is not a practical method of making such a junction because of the discontinuous crystal structure that will exist at the interface of the two materials; but, for the present it will be assumed that in bringing the two slabs together, they are fused into a single crystal.) When the two materials are brought into electrical contact, holes will diffuse from the P-type material into the N-type material; and electrons will diffuse from the N-type into the P-type material. This is to be expected because these carriers, driven by thermal forces, will try to equalize their distribution throughout the crystal.

This diffusion will soon be brought to a halt by the unequal charge distribution set up by the displaced carriers. That is, if a free electron diffuses across the junction into the P-type region, it will leave behind an unneutralized donor ion which has a net positive charge. Moreover, after it diffuses into the P region, it will eventually recombine with a hole and create an unneutralized acceptor ion having a net negative charge. Similarly, if a hole diffuses across the junction, it will leave behind a negative acceptor ion. Crossing the junction, it will recombine with an electron and produce a positive donor ion. These unneutralized impurity ions will produce an electric field across the junction. As can be seen from the illustration in figure 3.1, this field will oppose diffusion of current carriers across the junction. This electric field is called a <u>barrier</u>.

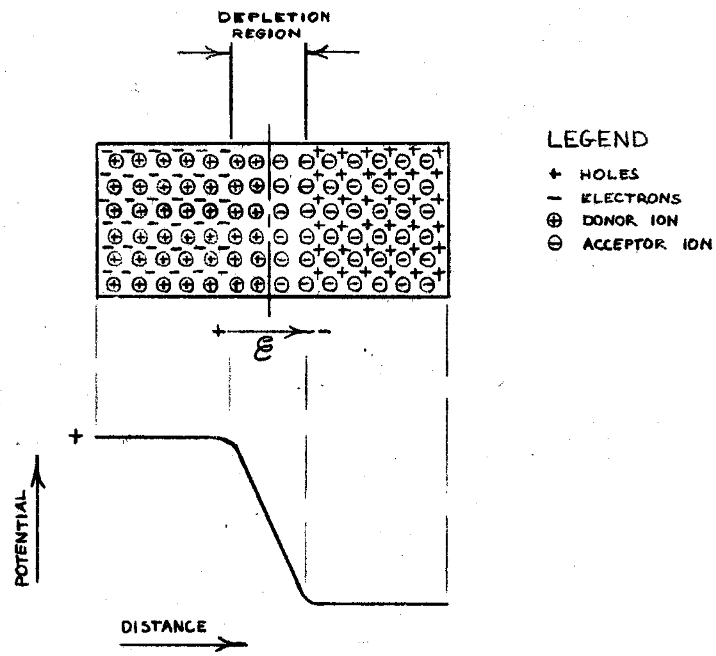

Figure 3.1. Formation of Barrier Potential in a PN Junction from the Diffusion and Recombination of Current Carriers which Exposes Unneutralized Impurity Ions in the Vicinity of the Junction.

The potential difference created across the junction by the unneutralized impurity ions is frequently refered to as the barrier potential. Although small, this potential (in the order of 0.3 volts for germanium and 0.5 volts for silicon) will set up a relatively strong electric field because it is confined to a narrow region near the junction. This region, called the depletion region will normally contain no current carriers. Any current carriers present within the depletion region will be swept away immediately by the electric field.

The barrier potential cannot be measured by ordinary means. A contact potential will be generated at any metalic contact on the crystal. If two probes are placed on the semiconductor, one on each side of the junction, the sum of the barrier and contact potentials will be zero so there will be no potential difference between the probes. A further explanation of this phenomenon involves a treatment of metal-semiconductor contacts which will be covered at the end of this chapter.

Forward Characteristics. If a small external potential is applied to a PN junction, positive to the P-type region and negative to the N-type region, it will produce an electric field in the depletion region opposing the barrier field. This will reduce the barrier field and permit diffusion of high energy current carriers across the junction. For applied potentials less than the barrier potential, the electric field within the crystal will be confined almost entirely to the depletion region because the number of current carriers present there is very much less than in the rest of the crystal, making it a high resistivity region. Under these conditions, current is established through the crystal primarily by thermal diffusion which is opposed by any existing barrier field. Reductions in the barrier potential permit lower energy carriers to diffuse across the junction in addition to the high energy carriers, thus increasing the current.

As current through the crystal increases, the concentration of current carriers in the depletion region also increases. The resistivity of this region will then approach that of the rest of the crystal, and a nearly uniform electric field will be established across the length of the crystal by the applied voltage. When the barrier is reduced to zero, even the lowest energy carriers will be able to cross the junction so further increases in current will come from an increased accelerating field. The potential variation with distance through the crystal is plotted in figure 3.2 for various values of applied voltage to illustrate these points.

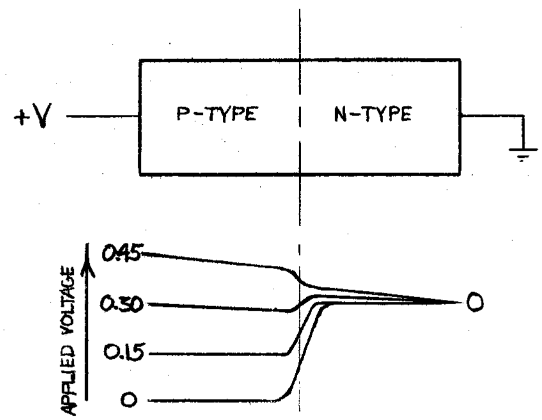

Figure 3.2. Variation of Potential with Distance across a PN Junction for Various Values of Forward Bias.

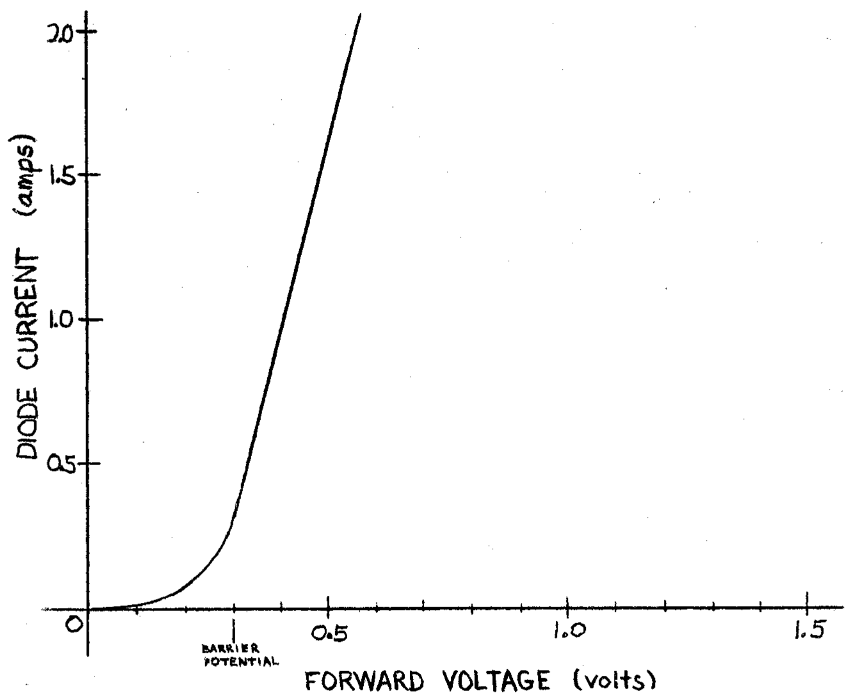

Figure 3.3. Forward Volt-Ampere Characteristics of a Junction Diode.

With this information, it is possible to explain the forward characteristics of a junction diode which are given in figure 3.3. For an applied voltage less than the barrier potential, the current does not increase linearly with voltage. If, for example, the voltage is doubled, the current will be more than doubled. This happens because reducing the barrier by one half will only allow a relatively small number of high energy carriers to diffuse across the junction, but doubling the voltage and eliminating the barrier will permit all the carriers to cross. After the barrier is eliminated, the current carriers are accelerated through the crystal by the relatively strong drift forces of the electric field in addition to the weaker diffusion forces. The velocity of the current carriers, and therefore the current, will depend on the applied voltage. In this region the current increases linearly with voltage.

Reverse Characteristics. If a reverse bias is applied to a PN junction, the barrier potential will be increased by an amount equal to the applied voltage. The entire reverse voltage will be dropped across the depletion region because the absence of current carriers produces a very high resistivity, or insulating, region. There is then no electric field acting on the current carriers in the P and N regions, and the barrier was already high enough to stop the diffusion of majority carriers. Therefore, there will be no current established through the diode.

This is an ideal condition, but the presence of minority carriers prevents its realization. There are holes present in the N-type region and free electrons present in the P-type region from the thermal generation of hole-electron pairs. Any of these minority carriers reaching the junction will be swept across by the barrier field. With no voltage applied to the diode, this reverse current is balanced by the diffusion of high energy carriers across the junction. Increasing the barrier height will reduce the number of high-energy carriers diffusing across the junction in the forward direction while the number of carriers crossing the junction in the reverse direction remains unchanged. Hence, a reverse current will be established.

When a reverse voltage of about one volt is reached, diffusion of even the highest energy carriers is stopped. From this point on, the reverse current is independent of voltage. The magnitude of the current will be determined by the diffusion rate of minority carriers to the junction. Once the carriers reach the junction, they will be swept across regardless of the junction potential.

The reverse characteristics of a junction diode are plotted in figure 3.4. The reverse current is found to increase until the diffusion of high energy carriers across the junction is stopped. Then it reaches a steady value determined only by the diffusion of minority carriers to the junction. This constant value of current is called the reverse saturation current.

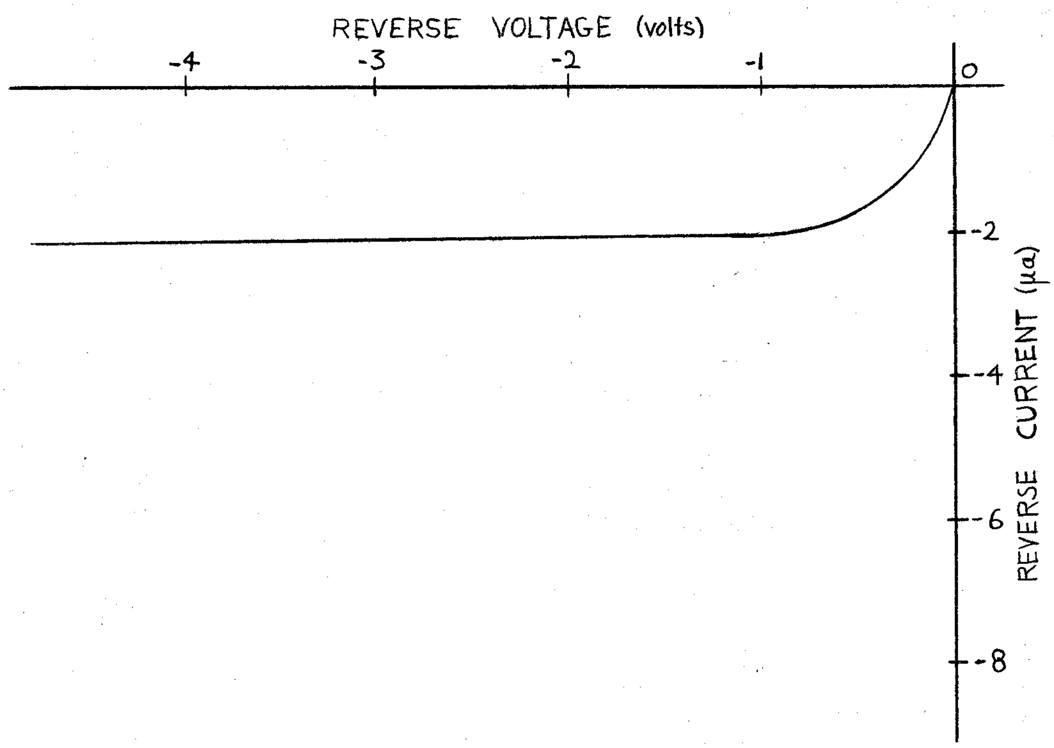

Figure 3.4. Typical Low Voltage Reverse Characteristics of a Junction Diode.

In practical diodes the reverse current will be found to increase slightly with voltage. This increase is caused by leakage on the surface of the semiconductor which contains water and other contaminants that conduct electricity. This leakage path will appear as a high resistance connected in parallel with the diode.

TEMPERATURE EFFECTS

Since the minority carriers in the P and N type regions are created by the thermal generation of hole-electron pairs, the reverse saturation current of a junction diode is dependent on temperature. As temperature is increased, the number of thermal generations increases quite rapidly, increasing the minority carrier concentrations in both the P and N type regions. Consequently, a greater number of carriers diffuse to the junction, thus increasing the reverse current.

The reverse saturation current of a germanium junction diode is plotted as a function of temperature in figure 3.5. At temperatures below 20°C, this current is indeed small; but it increases so rapidly above about 90°C that the diode becomes useless.

Because of the stronger covalent bonding of silicon, higher temperatures must be reached before an appreciable number of hole-electron pairs will be generated. This increases the maximum operating temperature of silicon devices. The plot of reverse saturation current *versus* temperature for a silicon diode has much the same shape as the curve in figure 3.5, except that the maximum operating temperature falls at approximately 200°C.

The forward characteristics of a junction diode are not too greatly affected by temperature. At higher temperatures, there is, in general, a small increase in the low voltage conductance and a small decrease in the conductance at higher voltages. This is shown in figure 3.6.

Increasing the temperature will increase the thermal energy of the majority carriers. Hence, when the barrier potential is reduced a given amount by an applied voltage, more carriers will be able to diffuse across at higher temperatures. After the barrier is completely eliminated, conduction through the diode is dependent on the resistivity of the semiconductor material in the body of the diode, so the conductance will decrease at higher temperatures due to the lowered carrier mobility.

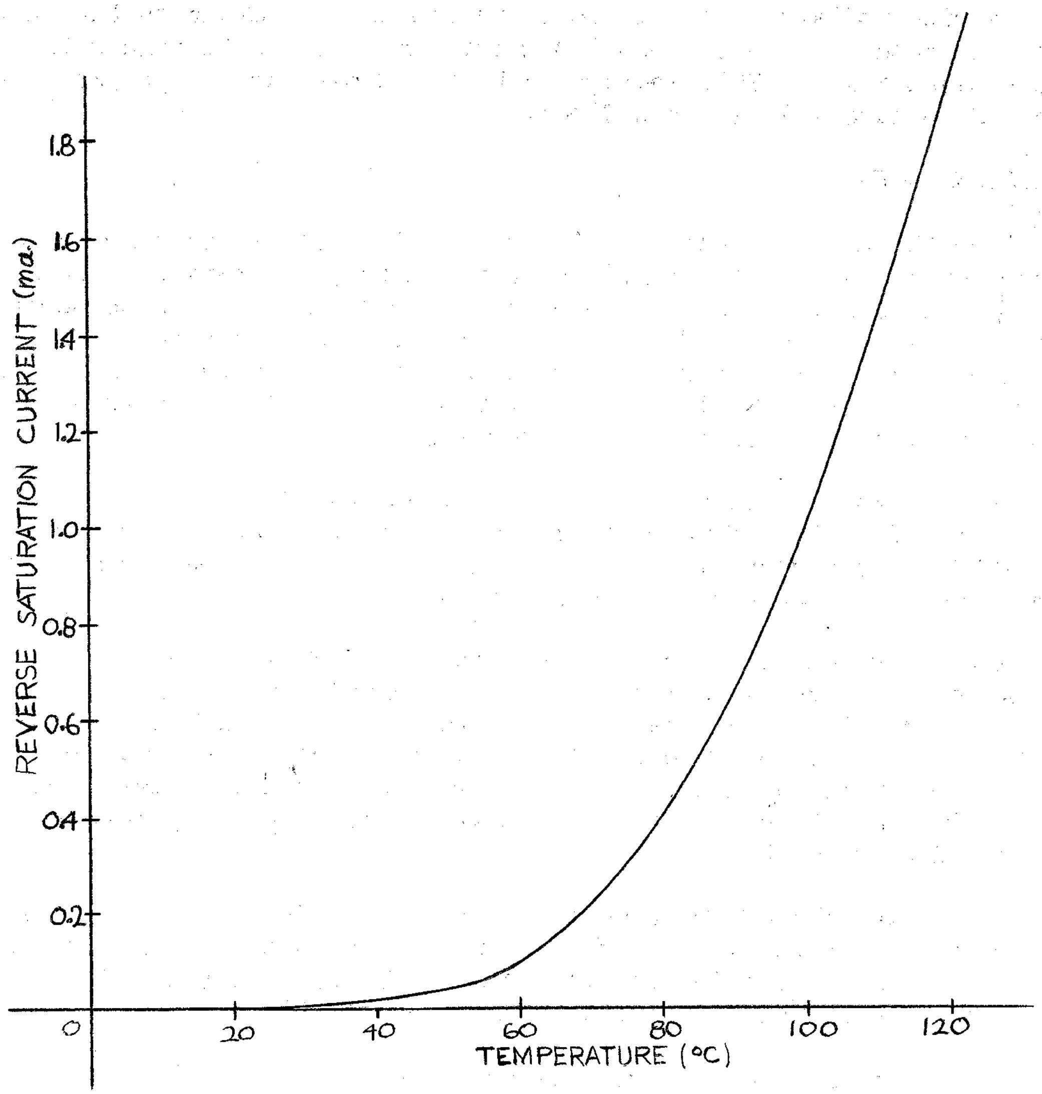

Figure 3.5. Plot of Reverse Saturation Current Versus Temperature for a Germanium Junction Diode.

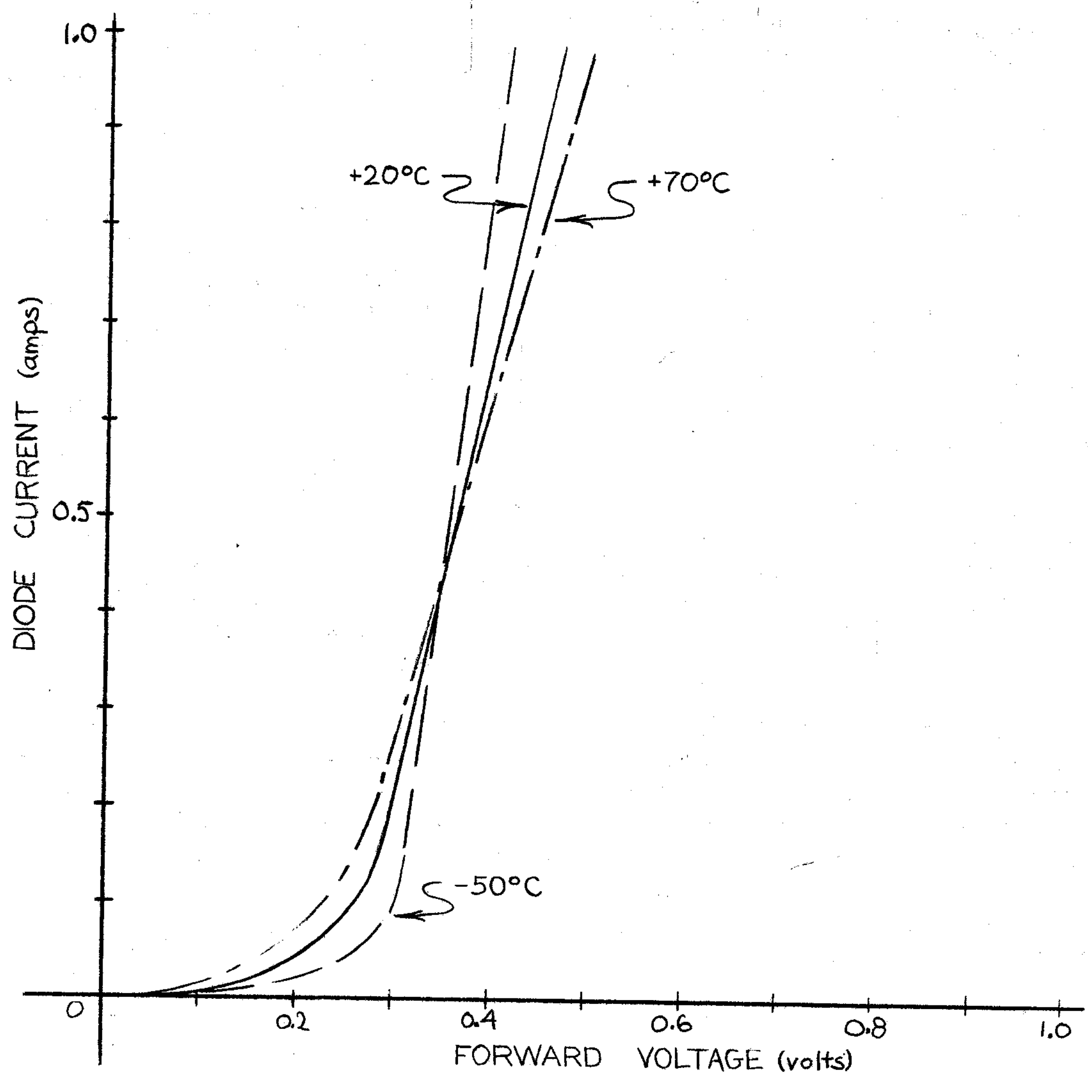

Figure 3.6. Change in the Forward Characteristics of a Germanium Junction Diode with Temperature.

REVERSE BREAKDOWN PHENOMENA

As the reverse voltage on a diode is increased, a point will eventually be reached where the diode begins to conduct heavily. This phenomena is known as reverse breakdown. There are three possible reverse - breakdown mechanisms: Thermal, avalanche, and zener breakdown. Whichever mechanism occurs first will temporarily destroy the reverse-blocking characteristic of the diode.

Thermal breakdown generally occurs in point contact diodes. In these diodes, the junction is formed near a point contact on the surface of the crystal. There are a large number of defects present near the surface, so point contact diodes are usually characterized by a high reverse current. At higher reverse voltages, this current gives rise to an appreciable power loss which is confined to a small volume near the point contact. This power dissipation will cause excessive heating of the junction, increasing the number of thermally generated current carriers and, therefore, the reverse current. This action is cumulative. After a certain voltage is reached (peak inverse voltage), the reverse current and the power dissipation will increase so rapidly that the reverse voltage will fall off with increasing current, producing the characteristics shown in figure 3.7. In the case of thermal breakdown, the diode will not be damaged if the power dissipation is kept low enough so that the junction does not melt.

Junction diodes usually undergo avalanche breakdown. As the reverse voltage is increased, the minority carriers that are swept across the junction by the barrier field are accelerated enough to excite electrons from the covalent bonds when they collide with the atoms of the crystal. These electrons will ionize other electrons giving rise to a cumulative action. This is similar in many ways to the breakdown of a gas diode. Avalanche breakdown is characterized by a sharp increase in reverse current at a nearly constant voltage as is shown in figure 3.8.

Zener breakdown, also common to junction diodes, occurs when the electric field across the junction becomes strong enough to rupture the covalent bonds. Even though the reverse voltage might be relatively small, the entire voltage is dropped across the narrow depletion region. This could create a intense field that is strong enough to ionize electrons directly from the covalent bonds. Zener breakdown is also characterized by the constant voltage characteristic shown in figure 3.8.

It is difficult to tell whether a particular diode undergoes avalanche or zener breakdown because of the great similurity of these two mechanisms. However, it is generally felt that the breakdown of junction diodes is caused by avalanche or in some cases a combination of avalanche and zener breakdown.

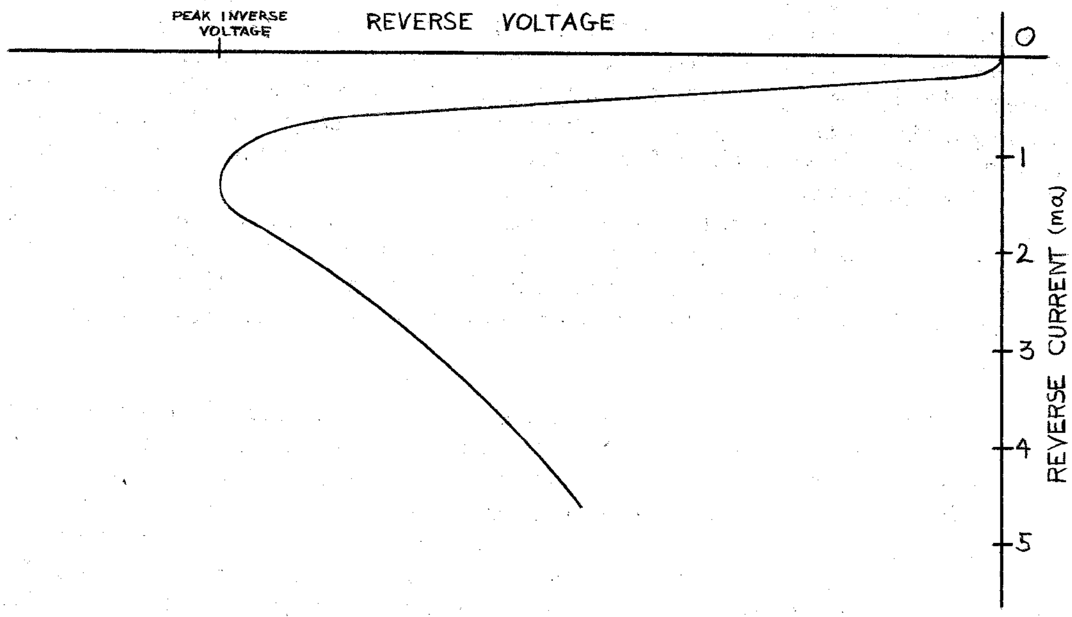

Figure 3.7. Thermal Breakdown Characteristics of a Point Contact Diode.

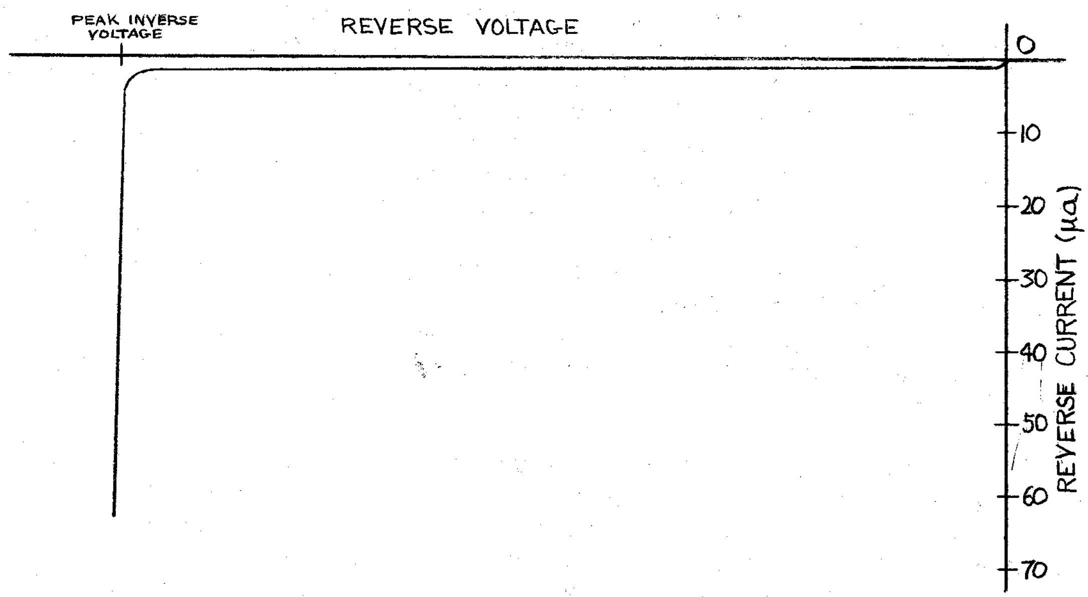

Figure 3.8. Avalanche or Zener Breakdown Characteristics of a Junction Diode.

<u>Zener Diodes</u>. A junction can be used as a voltage regulator by taking advantage of its constant voltage reverse breakdown. With avalanche or zener breakdown, the diode will not be damaged unless the power dissipated generates enough heat to physically alter the junction. In regulator circuits, it is frequently necessary to dissipate considerable power in the diode; therefore, provisions must be made to remove heat from the junction. This can be accomplished by the normal processes of radiation from an encapsulated diode; or, more efficiently, by soldering one end of the diode directly to a copper base and attaching this to a heat sink. Smaller units using the former method will dissipate about 150 mw while larger units mounted on a heat sink can handle up to about 50 watts.

Although their circuit applications are similar, zener diodes provide a greater flexability than gas tube regulators since zener diodes are available with reverse breakdown voltages anywhere between 0.5 and 400 volts. The breakdown voltage is determined by the doping of the P and N regions of the diode during manufacture. The degree of doping will determine the width of the depletion region and, therefore, the intensity of the electric field across the junction for a given reverse voltage (the entire reverse voltage is dropped across the depletion region). If the depletion region is made thinner, the field intensity will be greater; and the breakdown voltage will be lower.

If a diode is made from heavily doped P-type material and lightly doped N-type material, the depletion region will extend primarily into the N region. This happens because an equal number of donor and acceptor ions will be exposed in the formation of the barrier: When a hole diffuses across the junction, it will leave behind an unneutralized acceptor ion; and when it recombines with an electron in the N region, it will create an unneutralized donor ion, etc.. Therefore, the depletion region must extend farther into the lightly doped N region than into the heavily doped P region to unneutralize the same number of impurity atoms on both sides of the junction. In this case, then, the width of the depletion region and, consequently, the reverse breakdown voltage can be controlled during manufacture by the degree of doping in the N-type region. (This is one possible method. It is also possible to alter doping in the P-region or in both regions and achieve similar results.)

The name, zener diode, is somewhat misleading since either avalanche or zener breakdown could take place in these devices. It is generally felt that zener breakdown occurs in highly doped diodes with narrow junctions and a high electric field intensity for a given reverse voltage. Avalanche breakdown occurs in diodes with wider junctions where the accelerating field (potential rise over the mean free path of the carriers) is greater.

JUNCTION CAPACITANCE

A reverse biased PN junction will behave like a small capacitance. The depletion region, being devoid of current carriers, acts as an insulator between the conductive P and N regions, thus forming a capacitor. Electrically, this capacitance will appear to shunt the rectifying junction so it can limit the highest operating frequency of the diode. At high frequencies a voltage applied across the diode will be able to pass through the low reactance of the junction capacitance even though the diode is reverse biased.

The magnitude of the junction capacitance will depend upon the dielectric constant of the material from which the diode is made, the width of the depletion region, and the area of the function. The dielectric constant will be fixed by the choice of material (usually germanium or silicon); and the doping, which determines the width of the depletion region is adjusted to give the diode a low forward resistance while maintaining an acceptably high reverse breakdown voltage. Hence, reducing junction capacitance for high frequency operation is usually accomplished by reducing the junction area which also limits the maximum forward current.

In most low frequency applications, the junction capacitance (in the order of several micromicroforads) can be neglected because of its high reactance. Therefore, the junction area can be made quite large to give an increased current capacity.

VARIABLE CAPACTIANCE DIODES

The junction capacitance of a diode will be a function of the reverse voltage. The depletion region will become wider for increased reverse voltage because more of the immobile impurity atoms must be exposed to support the increased potential across the junction (the holes and electrons are pulled away from the junction by the reverse bias). This increases the width of the insulating region between the conductive P and N regions and, therefore, reduces the capacitance.

This effect is optimized in the variable capacitance diode. An abrupt transition between the P and N type materials is used to produce a maximum variation of capacitance with voltage. (A gradual transition would create a region near the junction which contained practically no impurities. This fixed insulating region would reduce the junction capacitance and also the change in junction capacitance with voltage). It is also necessary to reduce the series resistance of the diode body to give a high Q capacitance.

The characteristics of a typical variable capacitance diode are given in figure 3.9. As expected, a decreasing capacitance is shown for increasing reverse voltage.

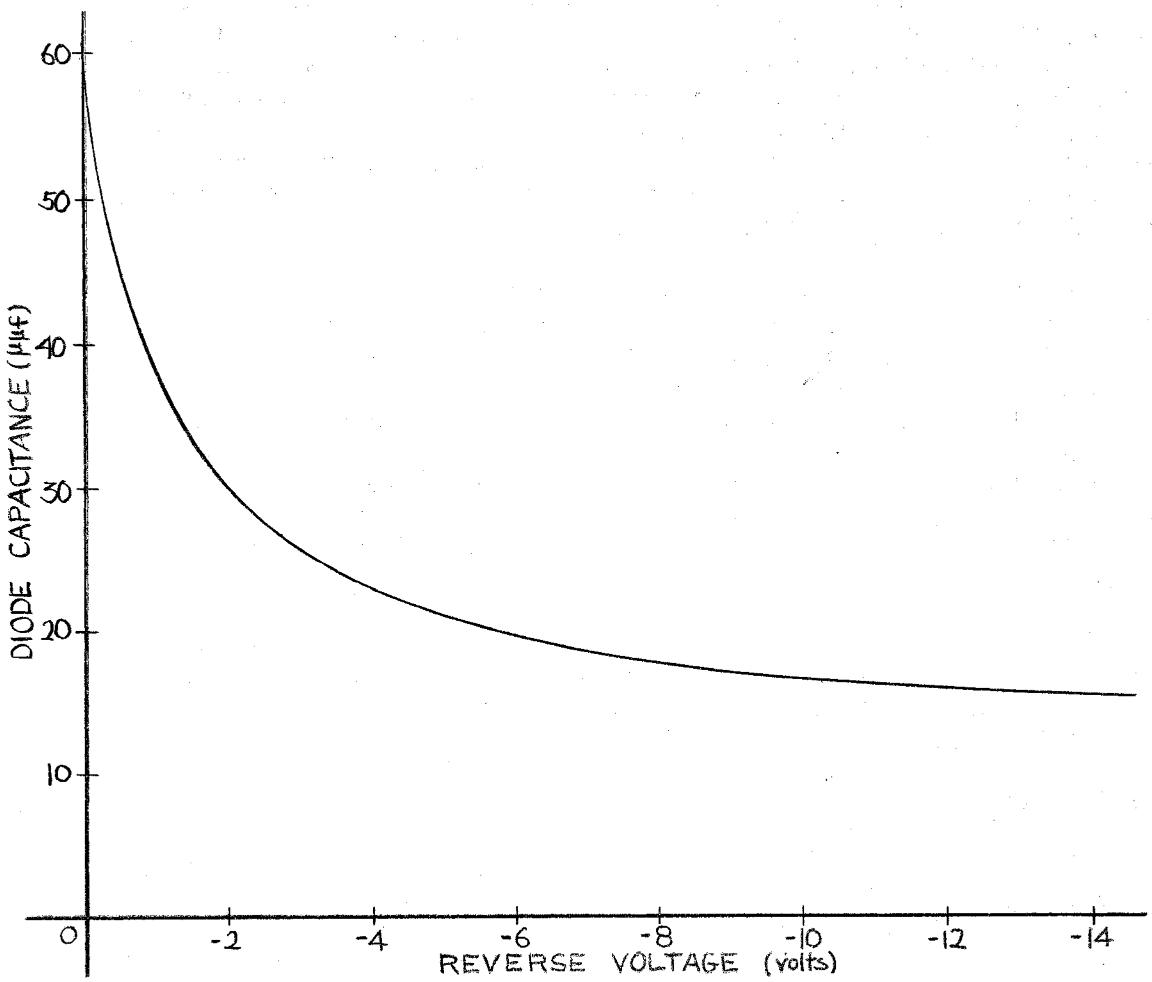

Figure 3.9. Variation of Diode Capacitance with Reverse Voltage for a Variable Capacitance Diode.

PHOTOELECTRIC EFFECTS

Light and other forms of radiant energy can ionize electrons from the covalent bonds of a semiconductor crystal. This will happen only if the photons making up the radiation have sufficient energy. Since the energy of the photons is directly porportional to the frequenty of the radiations, there will be a certain threshold frequency, for any particular material below which ionizations cannot be produced regardless of the radiation intensity. Radiant energy below the threshold frequency will pass through the semiconductor crystal without; but for frequencies very much above the threshold frequency, practically all the radiations are absorbed in producing ionizations. At higher frequencies, the radiations will not penetrate too far into the crystal so the ionizations will take place very close to the surface where recombination is likely to occur. Hence, for a given semiconductor material, there is a range of frequencies that will produce useful ionizations within the body of the semiconductor.

The threshold frequency for germanium and silicon is near the high end of the infrared spectrum, and the photoelectric response extends through visable light into the ultraviolet region. Other semiconductor materials are available (e.g. lead sulfide, lead selonide, indium arsenide, and indium antimonide) which have threshold frequencies farther down into infrared with useful response extending up to visable red.

Photoconductive Cells. Photoconductivity is the decrease in the resistance of a material caused by the increased number of current carriers made available by the absorption of radiant energy. High resistivity semiconductor materials will exhibit photoconductive properties. If a voltage is applied across a pure semiconductor in the absence of light, a small current, called the dark current, will be established due to the presence of uncontrolled amounts of impurities. However, when radiant energy in the proper frequency range is absorbed, electrons will be ionized from the covalent bonds causing an increase in current. This increase will be porportional to the light intensity as this will determine the number of current carriers liberated.

Because it is difficult to produce high resistivity semiconductor materials in practice, the dark current of this type of photoconductive device is quite high so a relatively high light flux must fall on the semiconductor to produce a noticable change in current. A PN junction can be used as a photocell in much the same way; only the dark current will be much less because of the high reverse resistance of such a junction.

If light falls on a back biased PN junction, current carriers generated in the depletion region will be swept away by the reverse voltage field, producing a reverse current. As before, the current will be porportional to the amount of light falling on the junction. Light falling on the bulk of the semiconductor material will be relatively ineffective in producing current because the electric field is confined entirely to the junction region.

Photovoltaic Cells. A PN junction can also be used as a self-generating photocell, or photovoltaic cell. A photovoltaic cell does not require an external voltage source: It will generate its own voltage. Its sensitivity is not as high as that of a photoconductive cell, but the self-generating feature is frequently more desirable than high sensitivity. Furthermore, since photovoltaic cells can convert sunlight directly into electrical energy, they can be used as a power source for portable equipment.

Even though a potential exists across the depletion region of a PN junction diode, it does not appear at the external terminals of the device. As already mentioned, the barrier potential is canceled by the combined contact potentials of the metal-semiconductor contacts to the diode. However, if any one of these potentials is increased or decreased, a voltage will appear at the diode terminals.

When a luminous flux falls on an unbiased PN junction, hole-electron pairs will be generated and swept out of the depletion region by the barrier field. This situation is shown in figure 3.10. The displacement of these charges will reduce the barrier potential making it less than the combined contact potentials; therefore, a voltage will appear at the external terminals of the diode. Furthermore, if a load resistance is placed across the diode, a current will be established, being supplied by the continuous generation of current carriers within the barrier field.

The short circuit current of a photovoltaic cell will depend on the number of current carriers generated in the barrier field so it will be directly porportional to the illumination intensity. The open circuit output voltage will be almost constant over a wide range of illumination intensities.

An output voltage is produced when the barrier potential is lowered by the displacement of current carriers which were generated in the barrier field. But lowering the barrier will also permit the forward diffusion of majority carriers across the junction which produces an opposing displacement of charged carriers. Hence, for a given light flux, the barrier potential will adjust itself until the photoelectric current produced by the generation of hole-electron pairs in the barrier field is balanced by the forward diffusion current across the lowered barrier. Since the number of current carriers with enough thermal energy to surmount the barrier increases quite rapidly as the barrier is lowered, a relatively large increase in illumination will necessitate only a small adjustment in barrier potential for a balanced condition to be reached. Therefore, small changes in illumination intensity will not produce significant changes in the no-load output voltage.

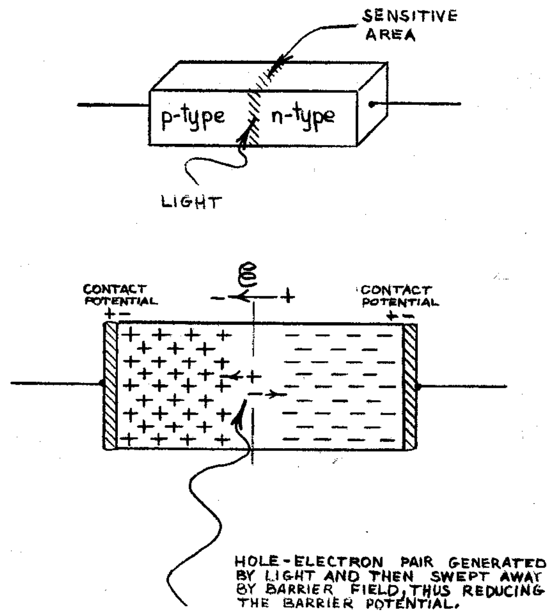

Figure 3.10. Operation of a Photovoltaic Cell.

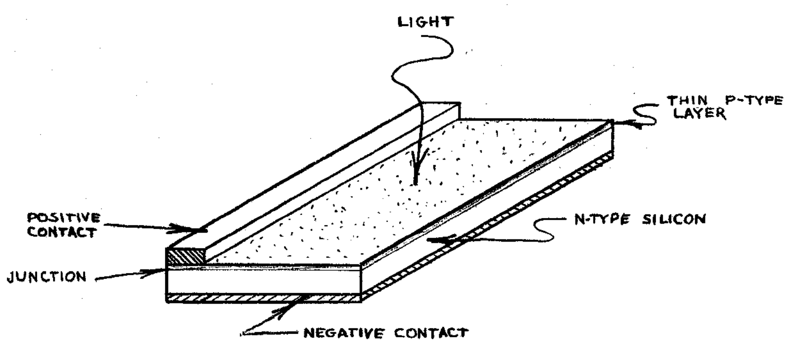

Figure 3.11. Construction of a Solar Cell.

The construction of a high efficiency photovoltaic cell is shown in figure 3.11. P-type impurities are diffused into a N-type crystal to produce a large area junction near the surface of the crystal. The photoelectric generations are confined to the junction region by using a material that absorbs most of the light near the surface, so the P-type layer must be very thin for high efficiency. Because the P region is so thin, the area of the positive contact must be as large as possible, without abscuring too much junction area, to reduce the internal resistance of the cell. A gradual transition between the P and N type materials is used to increase the width of the depletion region and, therefore, the active volume in which there is an electric field.

At the present time, silicon is used as the semiconductor material in solar cells because its photoelectric response is in the correct frequency range for operation on direct sunlight. Furthermore, larger barrier potentials are created in silicon diodes, as compared to germanium, so higher output voltages are possible (the selection of materials is a more rigid physical problem than might be thought because there is a direct relation between the strength of the covalent bonds, the photoelectric response, and the barrier potential). Silicon solar cells have been made that will produce an open circuit output voltage of 0.5 volt and a short circuit of 10 ma per square centimeter of active junction area when exposed to bright sunlight.

TUNNEL DIODES

The tunnel diode is a two terminal device that can be used as an amplifier, an oscillator or a switch. The tunnel diode will perform these functions by virtue of the fact that it exhibits negative resistance over certain ranges of operation. The electrical characteristics of a tunnel diode are shown in figure 3.12. Over a range of forward bias, the diode current decreases with increasing voltage. This is opposite to the behaviour of a normal (positive) resistance: Hence, the name, negative resistance.

A negative resistance will produce a power gain in a circuit, rather than a power loss. This is not in opposition to the conservation of energy; the negative resistance must be supplied power from another source, so it merely converts electrical energy of one form (usually d.c.) into electrical energy of another form. Normally, a negative resistance amplifies by canceling the loss of a positive resistance. However, there is a limit on the maximum gain that can be realized because if the total circuit resistance becomes negative, the circuit will become unstable and oscillate. More will be said about the applications of negative resistance devices in chapter 6.

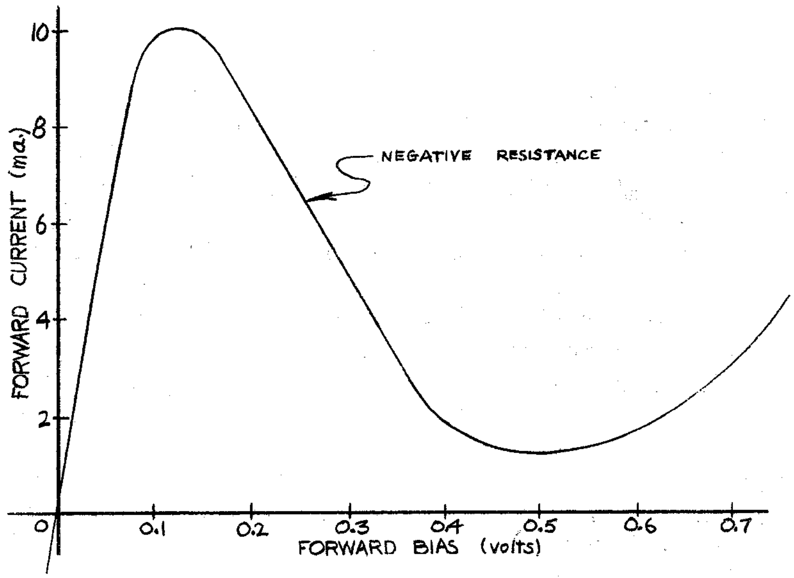

Figure 3.12. Electrical Characteristics of a Typical Gallium Arsenide Tunnel Diode.

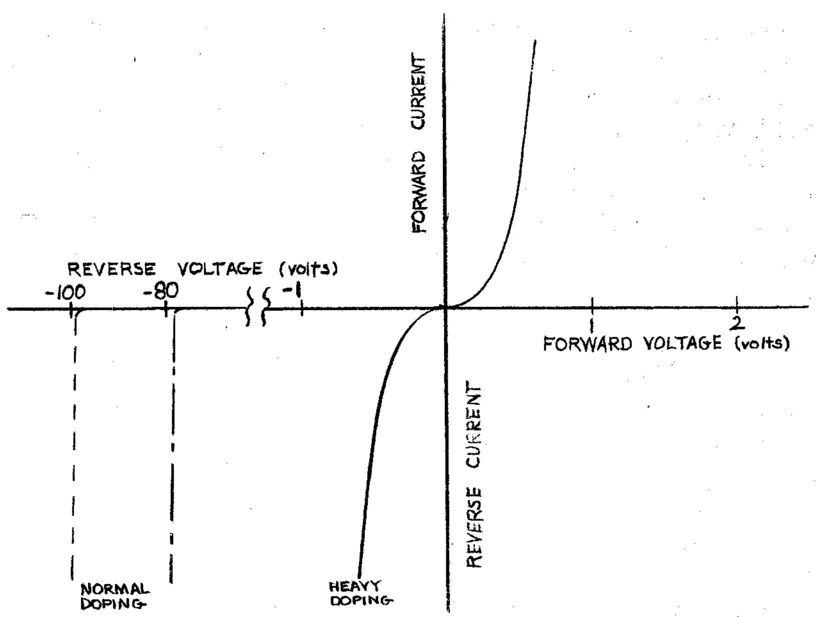

Figure 3.13. Effect of Doping on the Reverse Breakdown Voltage of a PN Junction.

The construction of a tunnel diode is similar to that of conventional diodes, except that the impurity concentration in the P and N regions is more than 1000 times greater. One consequence of such heavy doping has already been suggested in the treatment of zener diodes; that is, the depletion region will be very thin which causes an increased barrier field intensity for a given reverse bias and a lower breakdown voltage. The effect of doping on the reverse breakdown voltage is illustrated in figure 3.13. In the tunnel diode, the doping is carried to an extreme. In fact, the depletion region is so narrow that the normal barrier potential (no bias applied) is enough to induce zener breakdown. The tunnel diode, therefore, conducts heavily even for small reverse voltages.

The ultra thin depletion region (in the order of a few molecules thick) also permits a phenomenon known as quantum-mechanical tunneling, wherein the free electrons in the N type region can cross the junction to fill vacancies in the covalent bonds of the P-type material without being affected by the barrier field. This will only occur over a certain range of forward bias as is illustrated in figure 3.14. At higher forward biases, the conduction within a tunnel diode will take place by the diffusion of current carriers through a reduced barrier field as in a conventional diode.

METAL TO SEMICONDUCTOR CONTACTS

Metal to semiconductor contacts are an important part of every semiconductor device since they are required to make external circuit connections. There are, basically, two types of metal to semiconductor contacts: Rectifying and ohmic. A rectifying contact behaves much like a P-N junction in that it will only pass a current in one direction while an ohmic contact is insensitive to the direction of current and is used primarily to provide a low resistance contact to a semiconductor. Rectifying contacts will be discussed first since ohmic contacts are obtained by degenerating the performance of a rectifying contact.

<u>Rectifying Contacts</u>. When a metal is brought into contact with a P type semiconductor, free electrons from the metal will diffuse into the semiconductor to fill vacancies in the covalent bonds. This process will set up a charge unbalance, the semiconductor becoming negative with respect to the metal, and will continue until the potential difference established is sufficient to stop the diffusion of free electrons. If the semiconductor does not have too many imperfections near the surface where this contact is made, a barrier field will be set up in the semiconductor forming a depletion, or insulating, region as was the case with a PN junction. This depletion region will extend only into the semiconductor because of the vastly greater number of current carriers in the metal. If a voltage is applied to this contact in such a direction as to increase this barrier, there will be no current; but if the applied voltage reduces the barrier, diffusion will resume and a current will be established.

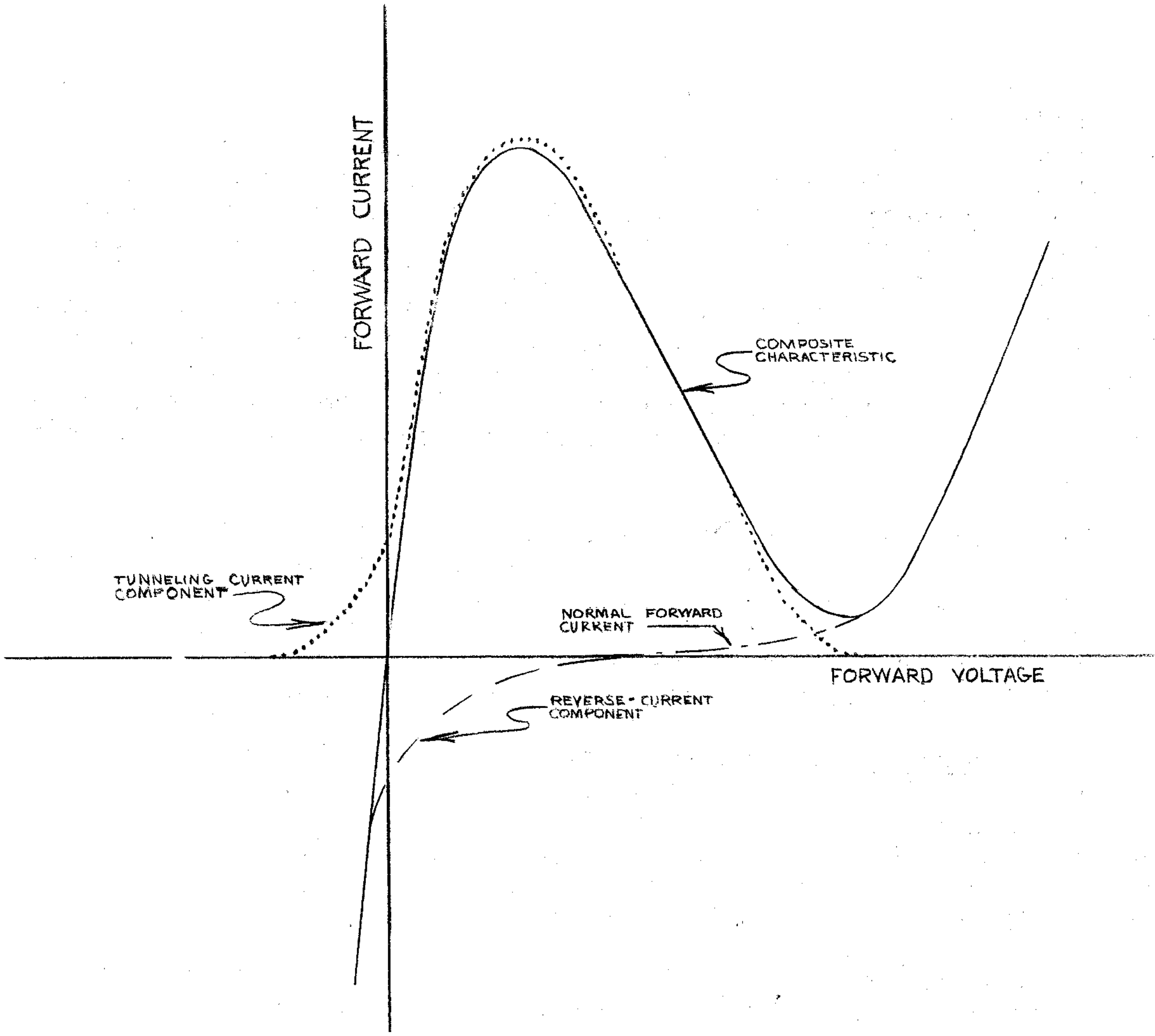

Figure 3.14. Resolving Tunnel Diode Characteristics into Individual Components for Purposes of Analysis.

Similar results can be realized by bringing a N-type semiconductor into contact with a metal. The free electrons in the semiconductor have a higher energy than those in the metal. When these materials are joined, electrons from the semiconductor will diffuse into the metal, becoming high energy free electrons in this material; however, the free electrons in the metal cannot diffuse into the semiconductor because their energy is too low. Therefore, electrons will diffuse into the metal from the semiconductor until a potential difference is set up that will halt the diffusion. Again, if the semiconductor near the contact surface is in good shape, a depletion region will be formed and the contact will be rectifying. A reverse bias will increase the barrier height so no current will flow; but a forward bias will lower the barrier, permitting diffusion.

There are other techniques available for forming rectifying contacts. If, for example, a metal containing a P-type impurity (eg. aluminum wire) is properly fused to a N-type crystal, a thin P-type region will be created around the contact; and a conventional PN junction will be formed. The same thing can be done in making a metalic contact on a P-type semiconductor by including N-type impurities in the metal.

Ohmic Contacts. Perhaps the simplest way to produce an ohmic contact on a semiconductor is to create defects in the crystal near the contact surface. This will destroy the rectifying properties of the contact by converting the depletion region into a region of high resistivity. However, because of this high resistivity, the contact area must be large if the contact resistance is to be low.

A practical method for producing small area, low resistance contacts is to dope the semiconductor very heavily near the contact surface. If the impurity concentration is made high enough, the semiconductor properties will be destroyed and an ohmic contact formed. This type of contact can be easily made by fusing a metal that is alloyed with suitable impurities to the crystal, thereby forming an ohmic P^{++} P or N^{++} N junction. Sucessful contacts can also be made by soldering on the contact using doped solders.

No matter how a contact is made between two dissimilar regions, there will be an initial diffusion of current carriers across the interface between the two materials, just as was the case with the rectifying contacts. This will set up a contact potential between the two materials, but the region in which a field is established is somehow distorted so that the contact does not exhibit rectifying properties.