SEMICONDUCTOR DIODES AND TRANSISTORS

MANUFACTURERS OF CATHODE-RAY OSCILLOSCOPES

VOLUME 6
REFERENCE FOR VOLS. 1-3

TEKTRONIX PUBLICATION 062-0422-00

SEMICONDUCTOR DIODES AND TRANSISTORS

VOLUME 6

REFERENCE FOR VOLUMES 1, 2 AND 3

TEKTRONIX, INC.
P. O. BOX 500
BEAVERTON, OREGON

062-0422-00

SIXTH PRINTING DECEMBER 1968

© 1965, 1966, 1967 TEKTRONIX, IN
BEAVERTON, OREGON
ALL RIGHTS RESERVED

INDEX

TITLE | PAGE

Prerequisites ---------- 1
Objectives ---------- 2

SECTION 1

Basic Atomic Theory Review ---------- 3
Fermi Level ---------- 16
Doping ---------- 18
N and P Type Semiconductors ---------- 20
Carrier Mobility ---------- 23
Basic Semiconductor Diodes Forming a Junction ---------- 25
Biasing and Diode Action ---------- 29
Diode EI Curves ---------- 33
Minority Carrier Lifetime ---------- 35
Carrier Distribution and Stored Charge ---------- 37
Diode Thermal Considerations Power Dissipation ---------- 38
Thermal to Electrical Analogy ---------- 40
Diode External Heat Sinks ---------- 44
Pulse Power and Thermal Time Constant ---------- 47
Calculating a Required Heat Sink ---------- 48
Temperature Coefficients ---------- 50
Pulse and Switching Diode Thermal Characteristics ---------- 51
Questions for Section 1 ---------- 57
Table 1 ---------- 66
Table 2 ---------- 69
Table 3 ---------- 70
Table 4 ---------- 72
Table 5 ---------- 74
Table 6 ---------- 76
Table 7 ---------- 80

SECTION 2

DIODE DEVICES

TITLE	PAGE
Doping by Diffusion	82
Diode Rectifiers	84
Chart 2A - Diode Rectifier Specifications	85
Diode Rectifier Measurements	88
Stacking Silicon Rectifiers	92
Typical Silicon Rectifier Circuit	95
Voltage Variable Capacitors	96
Zener Diodes	100
Temperature Compensation - Zener Diodes	108
Tunnel Diodes	113
Tunnel Diode Amplifiers	125
Measuring Tunnel Diodes	136
Tunnel Diode Switches	141
Tunnel Diode Stabilization	148
Backward Diode	150
Fast Signal and Switching Diodes	154
Diode Switching Time Measurements	155
Diode Switching Parameters	159
Measurement System Degradation	162
Snap-Off Diode	165
Four Layer Diode	169
Silicon Controlled Rectifier	173
Questions for Section 2	178
Table 8	189
Table 9	211
Table 10	213
Table 11	214
Table 12	215

SECTION 3

TRANSISTORS

TITLE	PAGE
Constant Current and Voltage Sources	216
Basic Transistors - Carrier Injection	219
Saturation	224
Biasing	226
Carrier Paths	231
Transistor Currents	234
Beta d-c	236
Beta Small Signal	239
Transistor Thermal Currents	241
Symbols	243
Transistor Configurations	245
Characteristic Curves	249
Load Lines	251
Saturation and Stored Charge	255
Review of Transistor Characteristics	261
Collector Breakdown (Avalanche)	266
Power Dissipation - Thermal To Electrical Analogy	274
Maximum Power Dissipation Curves	280
Pulse Power Limitations	285
The Transistor Switch	290
Switching Time Measurements	292
Speed-Up Capacitor	297
Current Mode Switching	300
Avalanche Mode Switching	303
High Frequency Transistors	306
$f\alpha_b$ or f_{hfb}	307
$f\alpha_e$ or f_{hfe}	308
f_T	309

TITLE	PAGE

C_{ob} and C_{oe} ---------- 311
Rate Grown and Alloy Transistors ---------- 312
Micro-Alloy and Micro-Alloy Diffused Transistors ---------- 314
Mesa and Epitaxial Mesa Transistors ---------- 316
Planar Transistors ---------- 318
Questions for Section 3 ---------- 321
Answers for Section 1 Questions ---------- 333
Answers for Section 2 Questions ---------- 334
Answers for Section 3 Questions ---------- 335
Figure 5 (fold-out) ---------- 336

SEMICONDUCTOR DIODES AND TRANSISTORS

VOLUME 6

REFERENCE FOR VOLUMES 1-3

This volume supplements and provides reference material for Volumes 1 through 3 of the Semiconductor Diodes and Transistors programed instruction series. It covers the same information presented in Volumes 1 through 3 in a conventional text form rather than the programed instruction approach. This volume also includes additional material and mathematical approaches that are not necessary to meet the objectives outlined in the individual volumes of the programed series.

Although this volume is intended to supplement the programed instruction series, it is complete in itself and can be used without the programed instruction series. Volumes 1 through 3 will not be listed as prerequisites. Since it is designed to supplement the programed instruction series, the section numbers in this volume relate to the volume with the same number; i.e., Section 1 relates to Volume 1, etc.

This volume does not have as an objective the teaching of the material contained herein. It is intended as a reference publication to supplement the programed instruction series. Therefore, the objectives will not indicate that the learner shall have reached a certain level of understanding or comprehension. Since programed instruction is designed to teach the information presented, using proven methods to enhance the learning process, predictable results can be listed in the objectives and tested for with a post-test. Using text type materials such as this volume, predictable results are not possible and therefore, the objectives will be listed in general terms.

PREREQUISITES:

This volume assumes the reader's comprehension of the following concepts. If he does not have this background, some outside study is indicated before starting this volume.

1. Basic electrical current, voltage, resistance, and power.

2. Ohms Law: $E = IR$ (voltage = current x resistance)
3. Power (watts) = I^2R (current^2 x resistance) = IE (current x voltage)
4. Kirchhoff's voltage and current laws
5. Thevinen's, Norton's, and Millman's theorems.
6. Basic alternating current theory including capacitance, inductance, and reactance
7. Algebra to the level of linear equations
8. Centigrade and Kelvin Temperature scales
9. Basic physics including the atomic structure of matter
10. Definitions of the words and prefixes:
    a. Diode
    b. Transistor
    c. Semiconductor
    d. Semiconductor Device
    e. Ions
    f. Molecules
    g. Proton
    h. Electron
    i. Nucleus
    j. Kinetic Energy
    k. Milli (m)
    l. Micro (μ)
    m. Nano (n)
    n. Pico (p)
    o. Kilo (K)
    p. Mega (M)

BROAD OBJECTIVES:

The aim of this volume is to provide supplementary and reference material for the programed instruction series on Semiconductor Diodes and Transistors. Each section within this volume relates to the programed instruction volume of the same number, and provides additional information that may be pursued by the learner having successfully completed that volume of the programed instruction series, or having gained this same background from some other source. It has the added objective of combining all of the information in the first three volumes of the programed instruction series in one volume and providing charts and graphs that may be conveniently found when a reference is required. It also provides a deeper coverage of the material from a mathematical standpoint that in the programed instruction series which allows the learner that wants to pursue the information further, a source to gain added information on the subject. Since it is hard to predict educational objectives with a text-type publication, specific objectives will be neglected, however, sample problems and test questions will be found at the end of each section, and the answers are listed at the back of the volume. This will allow the reader some measure of his gains using this volume.

SECTION ONE

BASIC SEMICONDUCTORS AND DIODES

BASIC ATOMIC THEORY REVIEW:

Since we have assumed in the prerequisites that the reader will have an understanding of basic physics, we will move right into the basic atomic theory as relates to semiconductors and the electrical and chemical properties of matter.

If we were able to isolate the most basic atom in space so that no external energy were affecting the atom, we would find an atom with one proton in the nucleus and one orbiting electron, as shown in Figure 1.

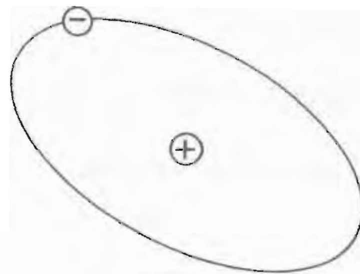


FIGURE 1

Figure 1 shows the diagram or model of an atom commonly referred to as the Bohr atom. This is really not a very complete model, since it is two dimensional, and merely shows the electron orbiting about the nucleus. It does not tell us much about the electrical characteristics of the atom, other than the atom is electrically neutral. The proton in the nucleus has one positive charge and the orbiting electron has one negative charge, and the atom is balanced or has no charge. Removing the orbiting electron in some fashion, perhaps by applying external energy, would leave one proton in the nucleus and no orbiting electron. Since the proton is assigned a positive charge and there is no electron to balance it, the atom would have a positive charge.

An ion is an atom with a different number of protons in the nucleus than it has orbiting electrons. If an extra electron were forced in some way into orbit about the nucleus of this atom, there would be more orbiting electrons than protons in the nucleus, and the atom would take on a negative charge. This is perhaps an

over-simplification, when we are dealing with an atom which we have assumed is isolated in space and it is not being affected by external energy. This simple model, however, allows the explanation of the electrical characteristics of the isolated atom. The model that we use to represent the atom, an electrical circuit, or an electronic device can be simple or complex, depending on the application. If the information or the analysis required is complex, it will probably require a complex model. If the information or the analysis required is simple, then a simple model will be sufficient. Within this publication, we will try and keep the model as simple as possible and still give the required information.

The simple Bohr model of the atom is not sufficient to explain the electrical characteristics of conductors, insulators, and semiconductors. Therefore, we must construct a model that gives us the required information. Figure 2 shows a different model of the simple atom in Figure 1.

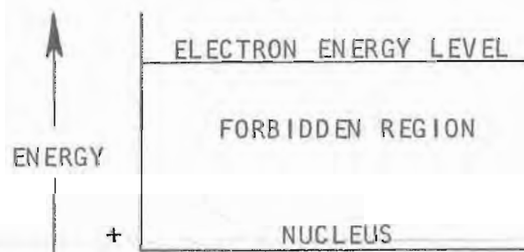
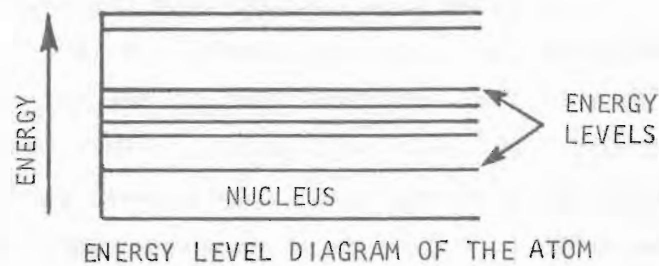


FIGURE 2

Figure 1 shows the Bohr model with the proton in the nucleus and an orbiting electron. Figure 2 shows another model of the atom that is a graph of the energy that the electron possesses with respect to the nucleus. Since there is an attraction between the proton in the nucleus and the orbiting electron, it is safe for us to assume that the electron must possess energy in order to maintain its orbit about the nucleus. The energy is kinetic energy, and the amount of energy that the electron possesses will determine its placement in orbit about the nucleus. With the atom isolated in space as we first assumed, it is safe to draw the diagram that we have in Figure 2. Note that the base line or the horizontal plot has no dimension in the diagram. We are dealing with one isolated atom and simply plotting the energy that the electron possesses with respect to the nucleus at any point in its orbit. The electron energy level shown is the level of the one electron orbiting the nucleus, and no other electron can exist in this level. The area between the proton or nucleus and the orbiting electron can be considered a forbidden region because the electron cannot exist in this area.

If a diagram is drawn of a more complex atom (complex meaning an atom with a larger number of protons in the nucleus and orbiting electrons), the diagram might resemble the one in Figure 3.



ENERGY LEVEL DIAGRAM OF THE ATOM

FIGURE 3

Here the energy levels of the electrons are once again drawn with respect to the nucleus, and the energy levels indicate the levels that have electrons orbiting about the nucleus. It is well to remember that this diagram is for an atom that is isolated in space with no external fields or other atoms affecting the energy levels of the electrons. It should also be remembered that this is nothing more than a model, and is just a method of expressing the energy that the electrons possess with respect to the nucleus of the atom.

When atoms bond together or interconnect in some fashion to form molecules in structures, there is interaction between the atoms and the orbits of the individual electrons are distorted. The electron in one portion of its orbit may exist in one energy level and be forced to move to a different energy level in another portion of its orbit. The electron in orbit (when affected by external fields) does not remain in one distinct energy level, but perhaps in several levels within its orbit. Constructing an energy diagram of atoms in a structure, the energy levels of the electrons cannot be shown as distinct energy levels with respect to the nucleus. Therefore, what is termed an energy band diagram is constructed, plotting the band of energy levels in which the electrons can exist with respect to the nucleus, as shown in Figure 4.

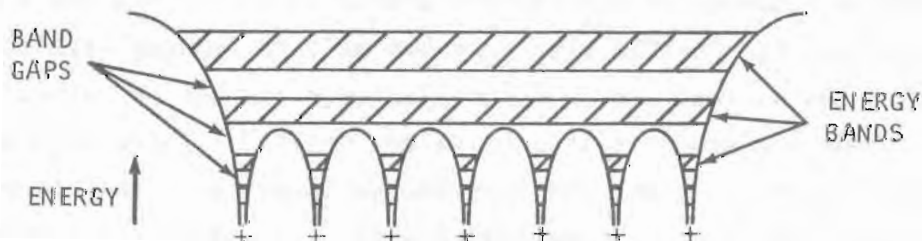


FIGURE 4

You will note that, in the energy band diagram, the horizontal has a dimension representing the inter-atomic spacing between the nuclei of the atoms. The energy bands nearest the nucleus are shown separate for each individual atom because there is no interconnection between them; however, there is a band that is shown common to all of the atoms. This is the outer-most band of the individual atoms that normally contains electrons. The bonding together or interconnection between atoms in molecules of structures takes place in this outer-most band. Since there is interaction between the atoms in the band, it is shown common to all of the atoms. The band in which the interconnection between atoms takes place is termed the valence band. The areas between the energy bands in the diagram in Figure 4 are referred to as band gaps or forbidden regions. No electrons can exist in the forbidden regions between the permissible energy bands. There are permissible energy levels above the valence band in the energy band diagram, in which electrons can exist. Electrons do not normally exist in these levels above the valence band, unless some source of external energy is applied to move them out of the valence band. Electrons existing above the valence band have been moved there by some external energy.

Adding energy to the electron moves it out of its normal orbit to some higher energy level. The electron will not move out of the valence band to exist in the band gap; therefore, in order to move an electron out of the valence band, sufficient energy must be applied to cause the electron to cross the band gap into a permissible energy level. The band of energy levels above the valence band that can contain electrons if external energy is applied is referred to as the conduction band. Electrons that are moved from the valence band to the conduction band are free of the individual atom and can be moved between the atoms if external energy is applied. When electrons are moved out of their normal orbits by application of external energy, the atom is termed "excited".

There is only a certain number of permissible energy levels within the energy bands. The innermost band that can contain electrons has only two energy levels; therefore, the band closest to the nucleus can only contain two electrons per atom. When it contains two electrons per atom, it is considered "filled". The second energy band in an energy band diagram has eight distinct energy levels and, therefore, can contain eight electrons. It is considered filled when it contains eight electrons per atom. The next band can contain eighteen electrons, and the next band thirty-two

electrons. With no external energy applied, atoms in structures will fill the bands from the nucleus out. It should be remembered that each unexcited atom in an energy band diagram has the same number of protons in the nucleus as electrons in the bands. The atom is considered neutral, or unexcited, when it contains the same number of orbiting electrons as protons in the nucleus.

To get a better feel for an energy band diagram, we will refer to Figure 5 (fold out at end of book), which is a periodic table of the elements. The maximum number of electrons per band is listed down the left hand column of Figure 5. These are the energy bands in our model. From the periodic table of the elements in Figure 5, the first shell can contain two electrons per atom; the second shell eight, and so forth. The second column on the periodic table lists the number of bands of the atoms to the right. The maximum number of electrons per atom in the first shell is two.

Hydrogen has only one shell and is the most simple atom listed. Hydrogen has the symbol 1-H-1. This is the atomic number of hydrogen. The first number, the one, indicates that hydrogen has one orbiting electron. The H indicates hydrogen, and the second 1 indicates that there is one particle in the nucleus. 1-H-1 is the symbol for hydrogen, and it indicates that there is one orbiting electron and one particle in the nucleus. Since hydrogen has only one orbiting electron, the number of electrons in the outer band of hydrogen is one.

The valence is listed across the top of the periodic table of the elements. We note that hydrogen has a valence of one, or one electron in the valence band. Let's pick another item at random in the valence one column. Copper has a symbol indicated by 29-CU-63. This indicates that copper has 29 orbiting electrons, CU stands for copper, and that there are 63 particles in the nucleus. There are particles in the nucleus of an atom, other than the protons. There are 29 protons in the nucleus to balance the 29 orbiting electrons in copper, and the other particles in the nucleus add to the atomic weight of copper. The shells, or energy bands, are filled from the nucleus out. The first shell can contain two electrons, the second eight, and the third eighteen, etc. We see in Figure 5 that copper has a total of four shells. The inner shells, or energy bands of copper are filled. Since copper has 29 orbiting electrons and only 28 are needed to fill the first three bands, there is one electron left in the outer most band of copper. The outer most shell, or band, is a valence band; therefore, copper will have a

valence of one (valence indicating the number of electrons existing in the outer most energy band of the unexcited atom, which is termed the valence band).

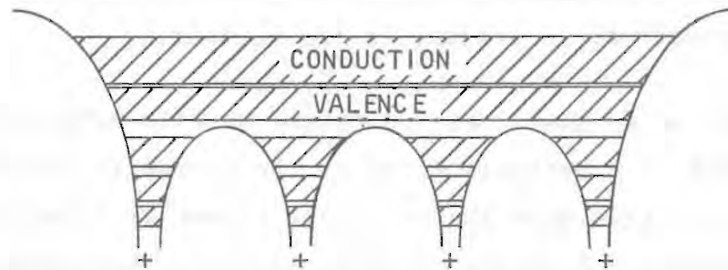


FIGURE 6

Figure 6 is an energy band diagram of copper. Note that the inner-most bands contain two, eight, and eighteen electrons respectively, and that the valence band contains one electron. The conduction band and the valence band are only separated by a very narrow band gap. We should already know that copper is a very good conductor, and that the band gap between the valence and the conduction band must be overcome to free electrons from the nucleus of the individual atoms. Copper has a very narrow band gap between the valence and the conduction band. (Actually at room temperature, 25°C, the bands overlap.)

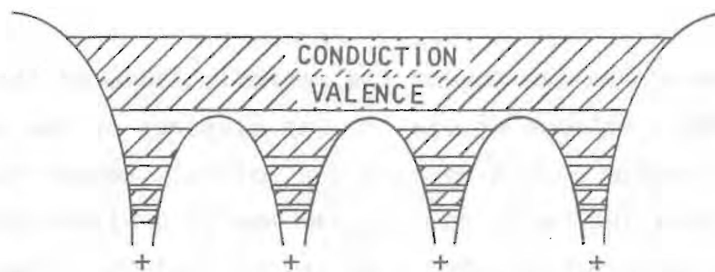


FIGURE 7

Figure 7 is an energy band diagram of silver. Looking at the periodic table of the elements in Figure 5, we find that silver has an atomic number of 47. This indicates that it has 47 orbiting electrons. Unlike copper, silver has five bands or shells. Silver, like copper, is in the valence one column, indicating that it has one valence electron; however, it has one more shell than copper. Notice in the energy band diagram of silver in Figure 7 that the valence band exists at a higher energy level with respect to the nucleus than copper. The valence electron of silver has more energy than the valence electron of copper.

We should already know that silver is a better conductor than copper, even though it has the same number of valence electrons. The greater number of bands gives an indication of why. Silver has one valence electron; however, this valence electron is existing in a higher energy level than the valence electron of copper, and the band gap between the valence band and conduction band in silver is very narrow and actually seems to overlap. It will take less applied energy to move the valence electrons of silver into the conduction band than it will to move the valence electrons of copper; therefore, silver is a better conductor than copper. Another way of putting it, there will be less opposition or resistance in silver to the movement of electrons than in copper. The resistance of a material can be related to the amount of energy it takes to move the electrons out of the valence band into the conduction band, and allow them to become carriers of current. From this, it is evident that conductors should have a small band gap between the valence and the conduction band.

To this point, we have been drawing an energy band diagram of a structure, and indicating all of the bands from the nucleus out. We will find, however, that the valence band and the conduction band are the two most important bands when dealing with the electrical characteristics of the atom. The inner bands are present, but generally disregarded when the energy band diagram is constructed. The primary part that the inner bands play in the electrical characteristics of a material is the number of bands that are present. The greater the total number of energy bands, the farther the valence electrons will exist from the nucleus energy-wise, and the more energy the valence electrons will possess. In any energy diagrams that follow in this publication, the inner-most bands will be disregarded, and only the valence band and conduction band will be shown. It should be remembered that the inner-most bands are present and affect the electrical characteristics by the number of bands that are present; however, as this is their major affect on the electrical characteristics, they will be disregarded.

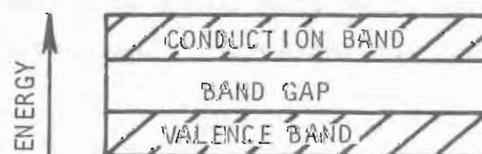


FIGURE 8

Figure 8 indicates the type of energy band diagram that we will use from this point on. The energy that the electrons in the bands possess is plotted vertically, and the inner-most bands are disregarded. Only the valence and conduction bands are shown. The width of the diagram, if the entire energy band diagram were shown, would indicate the inter-atomic spacing. For most purposes this is not necessary.

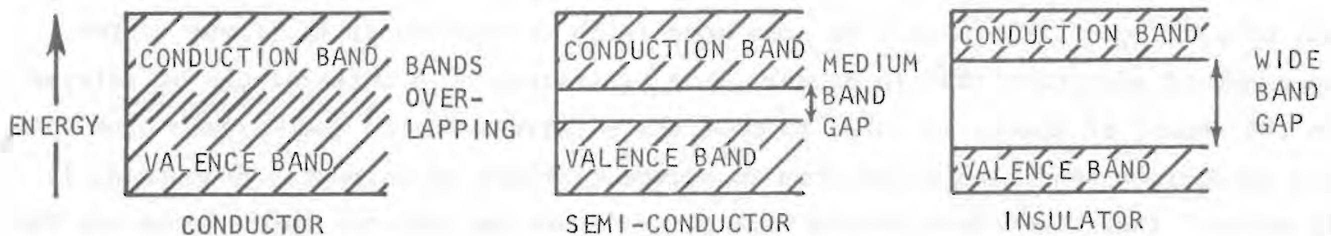


FIGURE 9

Figure 9 shows the energy band diagram for a conductor, an insulator, and a semiconductor. Note that the energy band diagram of the conductor has the conduction band and the valence band effectively overlapped at room temperature, and electrons are available in the conduction band with little effort. Room temperature for this publication will be assumed to be approximately 25°C, 77°F, or 300°K. Note that the energy band diagram of an insulator shows a very wide band gap between the valence and the conduction band, when compared to that of a conductor. This indicates that it will take a large amount of applied energy to move electroms from the valence band to the conduction band, and the material will have a high electrical resistance. Materials with a wide band gap between the valence and conduction band offer a high resistance and are considered electrical insulators. Those materials that fall between the insulator and conductor are termed semiconductors. Semiconductors have a fairly wide band gap when compared to the conductor, but a fairly narrow band gap when compared to the insulator. Semiconductors are not good conductors or good insulators.

Figure 10 lists some typical conductors, semiconductors, and insulating materials.

Silver and aluminum are listed as typical conductors, and we see that they have 10^{-5} and 10^{-6} ohms per centimeter^3, or a fairly low electrical resistance. They are considered conductors. Mica has 10^{12} to 10^{13} ohms per centimeter^3, and polyethelene has 10^{15} to 10^{16} ohms per centimeter^3, and they are considered insulators. Note that germanium and silicon have 50 to 60 and 50,000 to 60,000 ohms per centimeter^3 respectively, and these are considered semiconductors. (They fall somewhere between the lower resistance of the conductor and the very high resistance of the insulator.)

In looking at the periodic table of the elements in Figure 5, we find that germanium and silicon (the examples of semiconductors that we just used) lie in the column marked valence four. Germanium has an atomic number of 32 and has four shells, giving it four valence electrons. Silicon has an atomic number of fourteen; however, it only has three shells and the inner two shells are filled, leaving four electrons in the valence band. It should be noted at this time that the valence electrons in germanium are at higher energy levels than the valence electrons of silicon. Silicon has a higher electrical resistance than germanium, although they both are in the valence four column in Figure 5.

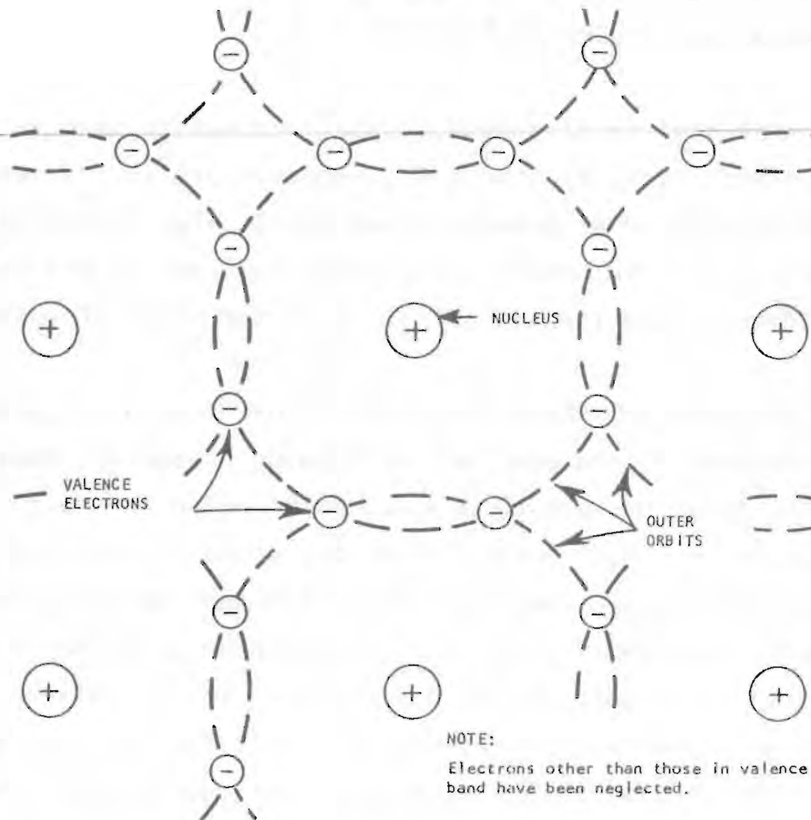
The first two conductors that we discussed, copper and silver, are in the valence one column. The semiconductors, silicon and germanium, are in the valence four column. Aluminum is considered a conductor, and yet it lies in the valence three column. This indicates that the number of valence electrons is not the only governing factor in determining the electrical characteristics of matter.

The type of bonding together or interconnection of the atoms also governs the electrical characteristics. There are several bonding processes; however, in semiconductor work the covalent bonding process is of prime interest. We all are familiar with H_2O, or water. H_2O is made up of two atoms of hydrogen and one atom of oxygen. Referring to Figure 5 once again, we find that hydrogen has a valence of one and oxygen has a valence of six. When the valence of hydrogen and oxygen is stated, it indicates that hydrogen has one electron in its valence band per atom and oxygen has eight electrons in its valence band per atom. Regardless of the number of electrons that can exist in a given shell or energy band, the outer-most band of the unexcited atom or the valence band is considered filled when it contains eight electrons per atom. If the valence band is the fourth shell, such as

ELEMENT (Symbol)	GROUP IN PERIODIC TABLE	NUMBER VALENCE ELECTRONS	APPLICATIONS IN SEMICONDUCTOR DEVICES
boron (B), aluminum (Al), gallium (Ga), indium (In)	III	3	Doping Elements to form P-Type semiconductors.
germanium (Ge), Silicon (Si)	IV	4	Basic semiconductor materials when in crystal form
phosphorus (P), arsenic (As), antimony (Sb)	V	5	Doping elements to form N-Type semiconductor

MATERIAL	RESISTANCE IN OHMS PER CENTIMETER CUBE (R/CM^3)	CATEGORY
silver, aluminum	10^{-6} 10^{-5}	Conductor
pure germanium pure silicon	50 - 60 50,000 - 60,000	Semiconductor
mica, polyethlene	10^{12} - 10^{13} 10^{15} - 10^{16}	Insulator

FIGURE 10



Each atom shares its valence electrons with four other atoms.

FIGURE 11

it is with germanium, the fourth shell can contain thirty-two electrons; however, when the fourth shell is the valence band, it is considered filled when it contains eight electrons. The bonding together of atoms forms octet structures. In an octet structure, the material has a valence of eight. When two atoms of hydrogen and one atom of oxygen bond together, the result is water. The bonding process in water is termed ionic bonding. Ionic bonding occurs when atoms with a small number of valence electrons give up their valence electrons to fill the valence band of atoms with a large number of valence electrons. Atoms with less than four electrons will tend to give up their electrons more readily than atoms with greater than four valence electrons. When oxygen and hydrogen are brought together, the hydrogen atoms tend to give up their valence electron to the oxygen and the bonding process occurs. Hydrogen takes on a positive charge. A positive charge results when the one electron leaves the hydrogen, leaving behind the proton in the nucleus. Since it has more orbiting electrons than protons in the nucleus when it accepts the electrons from hydrogen, the oxygen atom takes on a negative charge. Unlike charges attract and there is an attraction between the oxygen and the hydrogen. The hydrogen, having taken on a positive charge, is an ion. The oxygen, having taken on a negative charge, is an ion, and the process is referred to as ionic bonding. The ionization and de-ionization of the atoms goes on continuously, and the result is a fluid.

In semiconductor work, the concern is with the characteristics of solids. Taking germanium as an example, the bonding process is termed covalent bonding. In covalent bonding, the atoms of germanium arrange themselves equidistant from one another, and tend to share valence electrons rather than forming ions. Germanium has a valence of four (refer to Figure 5), and shares its four valence electrons with four adjacent atoms of germanium. Each atom of germanium appears to have a valence of eight and the atoms are tightly bonded together. No ions are formed and there is a lattice work that extends through the entire germanium structure.

Figure 11 shows a simplified diagram of the bonding process in germanium. Note that the path of the orbits of the valence electrons brings them into the valence band of the adjacent atoms and all of the atoms share valence electrons with their neighbors. Each atom appears to have a valence of eight. Since the valence electrons remain with the parent atom, no ions are formed and the atoms remain electrically neutral. An energy band diagram of germanium shows the valence band

filled as a result of the sharing of valence electrons and a distinct band gap between the valence band and the conduction band.

An element can be broken down into its smallest form and still maintain its characteristics. All of the listings in Figure 5 are elements. When broken down into their smallest form, they still exhibit the properties of the whole. When water is broken down into its smallest form, there are two atoms of hydrogen and one atom of oxygen, and no longer water as such. The smallest particle of water that still maintains the characteristics of water is a molecule (the molecule containing two atoms of hydrogen and one atom of oxygen). The molecule is the smallest particle of matter that still exhibits the original characteristics of the matter and, when pure elements are involved, the smallest particle is the atom of the element. If more than one element makes up the matter, the material is said to be a compound, if a chemical action is required to separate the elements. If no chemical action is required to separate the elements, the matter is simply a mixture. A compound is a bond of two or more dissimilar atoms or dissimilar elements. A bond of germanium atoms and arsenic atoms results in a compound that has a characteristic different than either of the two elements in their basic form.

Referring to Figure 5 again, arsenic is in the valence five column, and gallium is in the valence three column. Arsenic has five valence electrons and gallium has three valence electrons. When atoms of gallium and arsenic bond together, they form a structure that has an effective valence of eight, by sharing their valence electrons. The five valence electrons of arsenic bond with the three valence electrons of gallium and form a structure. The structure that they form is termed a compound and is called gallium arsenide. Gallium arsenide is a compound of gallium and arsenic.

Atoms with eight valence electrons do not lend themselves to bonding with other atoms in a covalent bond to form a structure. Atoms with a small number of valence electrons tend to give up their valence electrons very easily and make electrical conductors. Semiconductors generally fall in the valence four, five, and three columns with the possibility of using valence six and valence two materials (referring to Figure 5). Germanium and silicon structures are made up of elements, while such semiconductor materials as gallium arsenide are compounds.

We have seen that the number of valence electrons and the type of interconnection or bonding between the atoms affects the electrical characteristics of solids. Materials made up of elements with small number of valence electrons tend to be good electrical conductors. Materials in the three, four, and five valence columns in a covalent bond are termed semiconductors, and materials that are bonded together in a tight bond with a wide band gap between the valence and conduction band are termed insulators.

If external energy is applied to the atoms in a solid, electrons can be moved from the valence band to the conduction band and provide current carriers. The energy can be applied in the form of heat, light, or an electric field. Sufficient energy must be applied to the valence electrons to move them across the band gap before they can become current carriers. With no external energy applied including heat, light, or an electric field, a semiconductor structure has the valence band filled and no electrons existing in the conduction band. The material is essentially an electrical insulator. At room temperature, or approximately 25°C, sufficient energy in the form of heat is applied to some of the electrons in the semiconductor structure to move them to the conduction band. This will provide current carriers, and the material is termed a semiconductor. It should be noted that when heat energy moves electrons from the valence band to the conduction band, that the structure is not complete at the point that the electron left. Since the atoms all share valence electrons with adjacent atoms, an electron removed from the structure forms a hole in the semiconductor structure. When an electron is moved to the conduction band, a hole is left in the valence band, and it can be said that a <u>hole electron pair</u> has been formed by the application of external energy. With a pure semiconductor structure, moving an electron to the conduction band will leave an incomplete structure at that point.

It should be noted that the heat energy can provide current carriers and the resistance of material is reduced. Therefore, the electrical characteristics of the material are not only dependent on the number of valence electrons of the individual atoms and the type of bonding, but also dependent on the temperature. The greater the temperature, the more hole electron pairs are formed, and the more current carriers are provided. This lowers the electrical resistance of the semiconductor material. This is just the opposite of the temperature affects on conductors where an increase in temperature increases the resistance of the conductor.

Table 1 is included here for the people that would like to pursue the mathematics involved in determining the number of free holes and electrons in intrinsic semiconductors at a given temperature. Intrinsic semiconductor refers to semiconductors that have no impurities added. In other words, a piece of germanium which contains only germanium atoms, or a piece of silicon that contains only silicon atoms is termed intrinsic. Table 2 lists the characteristics of pure semiconductor structures. The important relationship that can be gained from table 1 is that the number of hole electron pairs is related exponentially to the temperature. When the number of free electrons and holes at room temperature is compared to the total number of atoms per centimeter^3, it is found that only one atom in a billion has lost its electron to the conduction band in germanium. Silicon has a wider band gap between the valence and conduction band and, at room temperature, only one atom in about 73 billion atoms of silicon has lost its electron to the conduction band. This indicates the reason for the large difference in the electrical resistance of silicon and germanium at room temperature. The density of holes and electrons increases by about 5% per degree centigrade in germanium, and about 7% per degree centigrade in silicon when temperature increases.

FERMI LEVEL:

An energy band diagram of intrinsic or pure silicon or germanium, assuming no external energy applied, has the valence band filled and no electrons existing in the conduction band. The highest level in the valence band is considered to be the 100% probability level for the existence of electrons, and the lowest level in the conduction band is considered to be the 0% probability level for the existence of electrons. Midway between these two points is the 50% electron probability level. Probability level indicates the probability of the existence of electrons at that point. Figure 12 shows the energy band diagram of a pure semiconductor with the probability levels marked in percentages. Figure 13 shows the Fermi level.

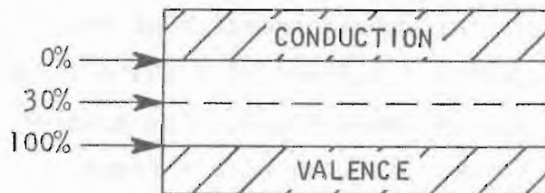


FIGURE 12

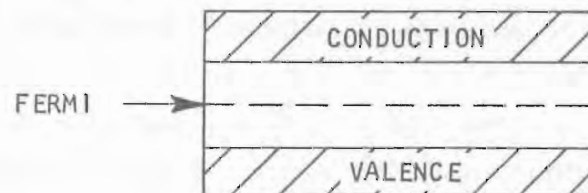


FIGURE 13

These probability levels are assuming no external energy to include heat is being applied. The highest level in the valence band is the 100% occupation level, or probability level. (We would expect to find electrons in this level.) Electrons cannot exist in the band gap, however, they can exist in the lowest level of the conduction band. With no external energy applied and a pure semiconductor, we would not expect to find electrons in the conduction band, therefore, it is assigned as the 0% probability level. The point midway between these points is, of course, the 50% probability level. The 50% probability level is termed the Fermi level. Figure 14 shows the typical energy band diagram of an intrinsic semiconductor with the Fermi level indicated.

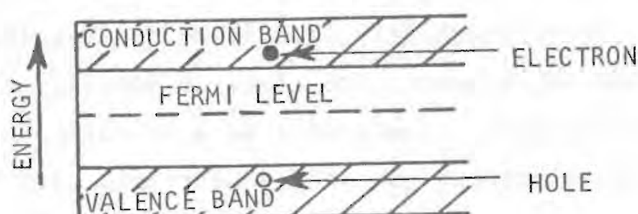


FIGURE 14

Notice that the fermi level is midway in the band gap between the valence band and the conduction band. The lowest level in the conduction band is the 0% probability level, and the highest level in the valence band is the 100% probability level. Table 3 is listed for those that want to pursue the mathematics involved in the Fermi distribution. For our purposes, assuming the Fermi level to be the 50% probability level of the existence of electrons is sufficient.

When hole electron pairs are formed in intrinsic semiconductors as a result of heat energy, the electron will exist in an energy level the same number of energy units above the Fermi level as the hole exists below the Fermi level in the valence band. The application of an external source of voltage to intrinsic semiconductor at room temperature results in a small current flow. The amount of current flowing will depend on the number of hole electron pairs that have been formed due to thermal energy and the magnitude of applied voltage. The electrons will move as current carriers in the conduction band of the material, and the holes will move as current carriers in the valence band of the material. The current in the external circuit is the sum of the current carried by the holes and the electrons. Table 4 shows the mathematics involved in determining the current in intrinsic semiconductors for those who want to pursue this farther.

It is important to note at this time that semiconductors can have two types of predominate current carriers. In intrinsic semiconductors, there are equal amounts of holes and electrons and both will serve as current carriers. There is no predominate current carriers and the current must be dealt with in terms of both holes and electrons as current carriers. It is also well to note that the electron moves in the conduction band, the hole moves in the valence band, and their velocities are different. Note in table 2 that the electron lattice mobility in germanium is about 3900 centimeters^2/volt second at room temperature, while the hole mobility is about 1900 centimeters^2/volt second at room temperature. This indicates that the electron mobility in germanium at room temperature is nearly twice that of the hole mobility. This can make a significant difference in later studies. Note also in table 2 that the electron mobility in silicon is less than half that of the electron mobility in germanium; however, the electron mobility in silicon is three times the hole mobility in silicon. There will be a pronounced difference in the mobility or the velocity of electrons as current carriers, with respect to holes as current carriers in semiconductors. We will pursue this further at a later point. For now, the fact that we have two current carriers in semiconductors and that they do have different velocities is sufficient. This is really not a new concept, since the familiar voltage regulators or gas-filled tube had two types of current carriers. When the gas in such a tube was ionized, the current was carried by both positive and negative ions, and the total current was the sum of the current carried by both the positive and negative ions.

DOPING:

In dealing with intrinsic semiconductors, the only current carriers available are those thermally generated by surrounding air temperature or carriers that are set free due to the application of an electric field. The intrinsic semiconductor material can be termed ohmic, or non-rectifying. Referring to Figure 15, the same amount of current flows, regardless of the polarity of the applied source.

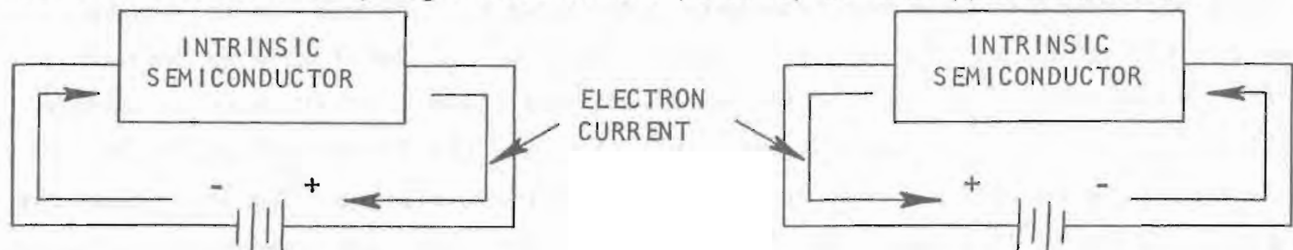


FIGURE 15

Note in Figure 15 that electrons leave the negative side of the source and enter the intrinsic material, travel through the intrinsic material and on to the positive side of the bias source. Electrons are repelled by the negative side of the bias source and move as current carriers in the conduction band of the intrinsic semiconductor material. At the same time, the holes in the valence band are traveling in the opposite direction in the valence band of the material, and the current in the external circuit is the sum of the current carried by the two types of carriers. If the polarity of the biasing source is reversed, current simply flows in the opposite direction. Connections that are referred to as ohmic are non-rectifying. The resistance of the semiconductor material may be changed by adding imperfections or impurities to the structure. Adding impurities to the intrinsic semiconductor that will increase the number of current carriers available at a given temperature, will reduce the electrical resistance of the semiconductor. Since the current can be carried by either holes or electrons, impurities may be added that will provide more holes <u>or</u> electrons as current carriers.

Referring to Figure 5, arsenic has an atomic number of 33 and falls in the valence 5 column. Arsenic has five valence electrons, one more than germanium or silicon. Germanium and silicon are covalent bond structures in that the atoms arrange themselves equidistant from each other and share valence electrons. Each germanium or silicon atom shares a valence electron with four adjacent atoms. If a few atoms of arsenic are added to the germanium or silicon structure, the atom of arsenic will provide more valence electrons than needed to complete the covalent bond. Four of the valence electrons of arsenic will bond with four adjacent germanium or silicon atoms, and the valence band will be filled. The fifth valence electron of the arsenic atom cannot exist in the valence band since it is filled and, therefore, is not needed to complete the structure. It would exist somewhere in the band gap, however, at room temperature, sufficient energy in the form of heat is applied to ionize the arsenic atom. At room temperature, the electron is available in the conduction band. When arsenic is added as an impurity to intrinsic germanium or silicon, electrons can be available in the conduction band without holes having to exist in the valence band. The electrons in the conduction band will be available as current carriers as the result of adding the impurities to the intrinsic germanium or silicon and not as a result of thermal energy forming hole electron pairs. At room temperature, of course, thermal energy will also have formed a few hole electron pairs, providing more electrons in the conduction band, and also a few

holes in the valence band. The number of electrons in the conduction band will be much greater than the number of holes in the valence band when the material is properly doped. Therefore, electrons will be the predominate or majority current carrier available in this type of semiconductor. When imperfections are added to intrinsic semiconductor with more valence electrons than needed to complete the covalent bond, electrons are provided as current carriers as a result of doping. The majority current carriers provided by the doping process are electrons in the conduction band and the material is referred to as N type semiconductor. Although at room temperature there will be both holes and electrons available as current carriers in N type semiconductor as a result of thermal energy, there will be many more electrons than holes. (The doping process has provided electrons in the conduction band that do not require a corresponding hole in the valence band.)

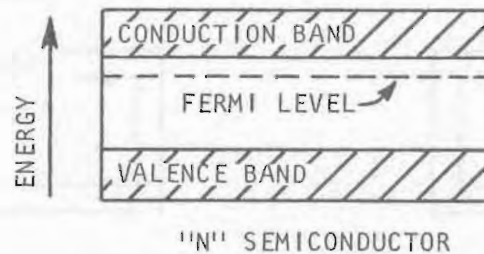
Figure 16 shows N type semiconductor with a biasing voltage supply added.

FIGURE 16

Note that the current in the semiconductor is carried by electrons moving in the conduction band, and holes moving in the valence band, but that the electrons in the conduction band are the predominate current carriers. When the battery polarity is reversed, the current direction simply reverses, and we can say that N type semiconductor is ohmic in that it has a resistance but does not rectify. N type semiconductors will have a lower resistance than the intrinsic semiconductor before doping.

The addition of impurities increases the number of current carriers available and reduces the resistance of the material. The greater the number of impurities added, the lower the resistance of the material after doping. Constructing an energy band diagram of N type semiconductor, the Fermi level is no longer placed in the center of the band gap between the valence and the conduction band. With intrinsic semiconductors, the top of the valence band is the 100% electron probability level, and the lowest energy level in the conduction band is the 0% electron

probability level, and the 50% level is midway in the band gap. The addition of impurities to form N type semiconductors provides electrons at a higher energy level than an intrinsic semiconductor. That is, it is possible now to have electrons existing in the conduction band without corresponding holes in the valence band. Therefore, the 0% and, as a result, the 50% electron probability level moves up in the diagram.



"N" SEMICONDUCTOR

FIGURE 17

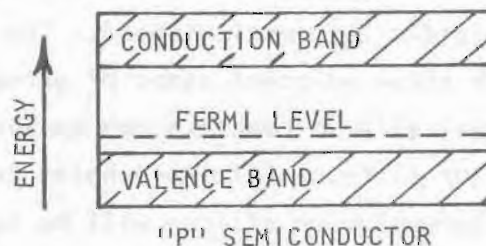
Note in Figure 17 that the Fermi level exists closer to the conduction band than the valence band in N type semiconductor. The greater the number of impurity atoms added in the doping process, the higher the Fermi level will exist in the energy band diagram. It is possible to add enough impurities to move the Fermi level up into the conduction band of the N type semiconductor.

Referring to Figure 5 again, Indium falls in the valence three column. Indium has three valence electrons, one less than germanium or silicon. If a small amount of Indium is added to intrinsic silicon or germanium, there will be an insufficient number of valence electrons to complete the covalent bond. The three valence electrons of Indium will bond with three adjacent atoms of germanium or silicon and there is no fourth electron available to complete the covalent bond. In other words, adding Indium to germanium or silicon will leave holes in the valence structure. The structure of the germanium or silicon will be incomplete at points, and there will be holes available as current carriers without electrons having to exist in the conduction band. Impurities with less valence electrons than needed to complete the covalent bond form P type semiconductors which have holes in the valence band as a result of the doping process. The holes provide carriers of current in P type semiconductors. At room temperature, of course, there will be hole electron pairs formed as a result of heat energy; however, the number of holes will be large compared to the number of electrons in the conduction band. The holes will be the predominate or the majority current carrier in P type semiconductors and the electrons existing in the conduction band as a result of thermal

energy will be the minority current carriers. Impurities added to form P type semiconductors are referred to as acceptor impurities, and they provide holes in the valence band of the structure. Impurities that are added to form N type semiconductors are referred to as donor impurities, and they provide electrons as current carriers in the conduction band of N type semiconductors.

FIGURE 18

Figure 18 shows a piece of P type semiconductor with a bias voltage applied. Current is carried by holes moving in the valence band of the P type semiconductor, which results in electron movement in the opposite direction in the external circuit. Current is also carried by electrons moving in the conduction band of P type semiconductors; however, since the number of holes is large compared to the number of electrons, the current is carried primarily by hole movement in the valence band. If the polarity of the biasing battery is reversed, hole movement changes direction in the P type semiconductor and current flows in the opposite direction. P type semiconductors are ohmic in that they are non-rectifying.



"P" SEMICONDUCTOR
FIGURE 19

Figure 19 shows the energy band diagram of P type semiconductors. Note that the Fermi level exists closer to the valence band than the conduction band in this type of semiconductor. The top level in the valence band is no longer the 100% probability level for the existence of electrons, because the doping impurities have provided holes in the valence band without corresponding electrons in the conduction band. As a result of doping, the 100% probability level and, therefore, the 50% probability level are moved down in the energy band diagram. The greater

the amount of acceptor doping impurities added, the lower the Fermi level will exist in the energy band diagram. It is possible to add a sufficient number of impurities to move the Fermi level down into the valence band of the material.

It should be remembered that both N and P type semiconductors are electrically neutral and there is a balance between the number of orbiting electrons and the protons in the nucleus of the atoms in the structure. There are electrons available as current carriers in N type germanium and holes available as current carriers in P type germanium, as a result of doping; however, there is no net charge, because a balance exists between the electrons and protons in the atoms in the structure. The unbalance exists in the covalent bond. The number of impurities added in the doping process might be only one part in several million parts of the intrinsic semiconductor. This will be sufficient to provide current carriers and reduce the resistance of the intrinsic material. When holes move in the valence band of the P material, as a result of the application of an external biasing source, a hole is formed at one side of the material by removing an electron from the valence band. There is a net drift of holes through the valence band of the material and the holes recombine with electrons at the opposite side of the P type material. Holes move internally in the opposite direction of the electron movement in the external circuit.

CARRIER MOBILITY:

When there are large numbers of electrons available in the conduction band such as in N type semiconductors, the movement of electrons in the valence band can be compared to the movement of electrons in a conductor. On the application of applied voltage, there is a net drift of electrons through the conduction band from the negative to the positive side of the source, resulting in a current in the external circuit. The mobility of the electrons is set by the type of semiconductor material used.

Table 2 lists the characteristics of three types of semiconductor materials most commonly used, and the lattice mobility of electrons is shown for all three. When the electrons are in the majority and are the majority current carriers, the mobility is governed by the lattice mobility of the intrinsic semiconductor that has been doped. With N type germanium, this is 3900 centimeters^2/volt second; with

silicon this is 1500 centimeters^2/volt second, as shown in table 2. When the electron is in the minority, such as it would be in P type semiconductor, there are few electrons existing in the conduction band, and there is not the net drift of electrons that occurs when the electrons are in the majority. The electrons are not caused to drift due to a potential gradient across the material and they are transported by a means termed "diffusion". The electrons, when in the minority, tend to take a random erratic path in the semiconductor, and travel at a different velocity than when in the majority. Electrons existing as minority carriers can be said to diffuse and, consulting table 2, the electron diffusion constant in germanium is 100 centimeters^2/second; in silicon it is 38 centimeters^2/second. The electron traveling as a minority carrier has a different velocity than when traveling as a majority carrier. The same is true of holes as majority and minority carriers. Referring to table 2 once again, the hole mobility in germanium is 1900 centimeters^2/volt second, while the hole diffusion constant for germanium is 49 centimeters^2/second. Holes have a different velocity when in the majority than when in the minority (serving as current carriers).

The important things to remember at this time are that doping can provide either electrons or holes as current carriers; that there will be both electrons and holes available as current carriers at room temperature, even though the doping was aimed at providing only electrons or only holes. The majority current carrier will have a different velocity than the minority current carrier. Holes moving as majority current carriers in P type semiconductors will have a different velocity than holes moving as minority current carriers in N type semiconductors. Electrons moving as majority current carriers in N type semiconductors will have a different velocity than electrons traveling as minority current carriers in P type semiconductors. The majority current carriers are said to have a certain mobility set by the type of basic semiconductor used. The transport of minority carriers is by diffusion, which is a random movement of minority carriers through the material. The diffusion constant of minority carriers is also set by the intrinsic semiconductor that has been doped. The mobilities and diffusion constants of the three commonly used semiconductors can be found in table 2.

Table 5 shows the math involved in determining the number of holes and electrons available in intrinsic material, the number of electrons provided as current carriers when donor impurities are added, the number of holes provided as

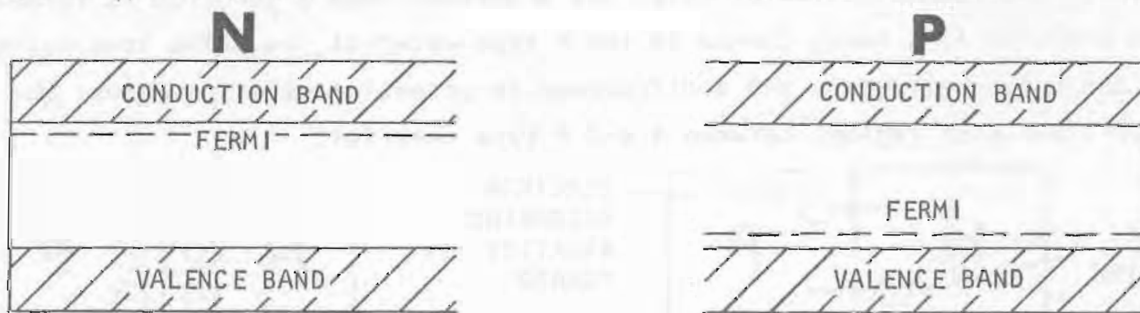
current carriers when acceptor impurities are added, the number of holes as minority carriers that are present at a given temperature in N type, and the number of electrons that are present at a given temperature as minority carriers in P type semiconductors.

Table 6 covers the math involved in the mobility of majority carriers and the diffusion of minority carriers in semiconductors. The parameters mobilance and diffusion are formulated, and the relationships of the majority carrier current and the minority carrier current to these parameters are formulated, for those that want to pursue this further. An important point is the relationship of the mobility of majority carriers with respect to the diffusion constant of minority carriers. The ratio of the mobility of holes (μ_p) to the diffusion constant of holes (D_p) is equal to the ratio of the mobility constant of electrons (μ_n) to the diffusion constant of electrons (D_n), and these ratios are equal to $\frac{q}{KT}$, where q is the charge on an electron, K is Boltzman's constant, and T is the temperature. This will be important in dealing with junctions of P and N type semiconductors a little bit later.

BASIC SEMICONDUCTOR DIODES:

FORMING A JUNCTION:

The construction of a semiconductor, PN junction or diode is accomplished by doping the opposite sides of a semiconductor with donor and acceptor impurities, and the transition point between P and N material becomes a junction. To analyze what happens when a junction is formed, we will hypothetically bring together two pieces of N and P type material.



BEFORE JUNCTION IS FORMED

FIGURE 20

Figure 20 shows the energy band diagram of both N and P type semiconductors, and in this diagram, it is assumed that impurities have been added to the same type of intrinsic semiconductor. The nucleus is assumed to be the 0 energy level, and the energy in the bands is plotted with respect to the nucleus. The Fermi level in the N type exists closer to the conduction band than the valence band. The Fermi level in the P type exists closer to the valence band than the conduction band and, if the two materials are brought together, there will be electrons existing in the conduction band of the N type and holes existing in the valence band of the P type, between which there is an attraction. The electron has a net negative charge and the hole has a net positive charge, and there is an attraction between the two. When the two pieces of N and P type material are joined, electrons will migrate into the conduction band of the P type from the N type; holes will migrate from the valence band of the P type into the valence band of the N type, and recombination will occur. By recombination, we speak of an electron filling a hole and neither being present as current carriers. If an electron leaves a conduction band to fill a hole in the valence band, it is no longer available as a current carrier in the conduction band. When the electron fills a hole in the valence band, the hole is no longer available as a current carrier. An area about the junction of the two materials will start to become depleted of current carriers, since the electrons filling the holes reduce the number of carriers available. The electron, on leaving the N material, leaves behind the nucleus of the impurity atom which has one more proton than it has orbiting electrons. The electron, on leaving the N material, leaves behind a positive ion in the N material - the positive ion being the impurity atom that has lost an electron. The electron, on entering the P material and recombining with a hole fills the hole in the valence band of the P material and satisfies the covalent bond; however, the impurity atom in the P type now has more orbiting electrons than it has protons in the nucleus, and becomes a negative ion. The recombination of holes and electrons when a junction is formed results in positive ions being formed in the N type material, negative ions being formed in the P type material, and a difference in potential existing across the junction or transition region, between N and P type material.

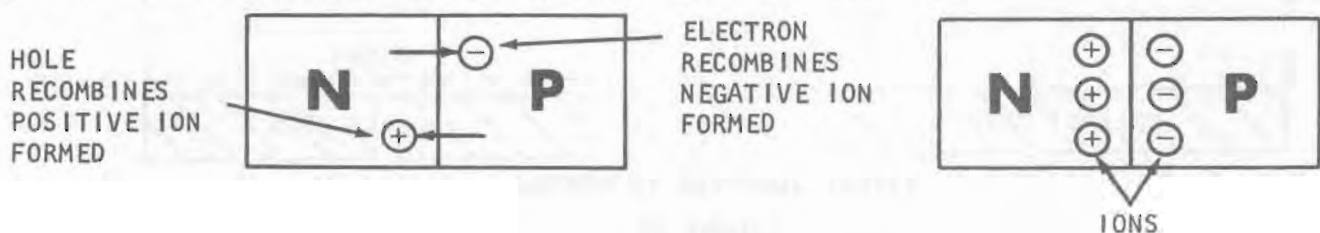
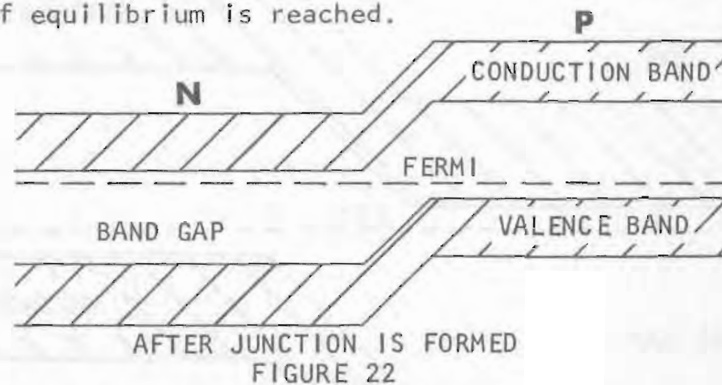


FIGURE 21

Figure 21 is a rough sketch of a PN junction with the depletion region and the positive and negative ions indicated. The recombination of holes and electrons occurs until a point of equilibrium is reached.



AFTER JUNCTION IS FORMED
FIGURE 22

Figure 22 shows the energy band diagram of a PN junction at equilibrium. Note that when equilibrium occurs, the Fermi levels of the two sides are lined up. The conduction band of the N side is opposite the band gap in the P side, the valence band of the P side is opposite the band gap in the N side, and there is a potential difference or built-in voltage existing across the junction or transition region. This built-in potential at room temperature is typically 300 millivolts for a germanium PN junction, and about 700 millivolts for a silicon PN junction. Table 7 covers the math involved in determining the built-in voltage or potential difference that exists across the junction. The math in table 7 is not necessary for this level of coverage; however, it is included for those that want to pursue this a bit further.

It should be noted at this time that the PN junction diode can be likened to a charged capacitor, since the N and P regions are essentially low resistance areas due to the heavy concentration of majority carriers, while the depletion region, having been depleted of majority carriers is essentially an area of effective insulation between the N and P regions. The N and P regions separated by the depletion region can be likened to a leaky capacitor, whose capacitance may be changed if the width of the depletion region is changed. (This is the same as changing the distance between the plates of a capacitor.) The voltage variable capacitor, which is a device that will be discussed later in this volume, takes advantage of the capacitance at the junction and uses it in circuitry. The junction at equilibrium has the Fermi levels lined up and the two sides are no longer electrically neutral. The N side has a net positive charge, and the P side has a net negative charge. Recall that the energy band diagrams are a plot of negative electron volts with respect to the nucleus and, prior to forming the junction, the

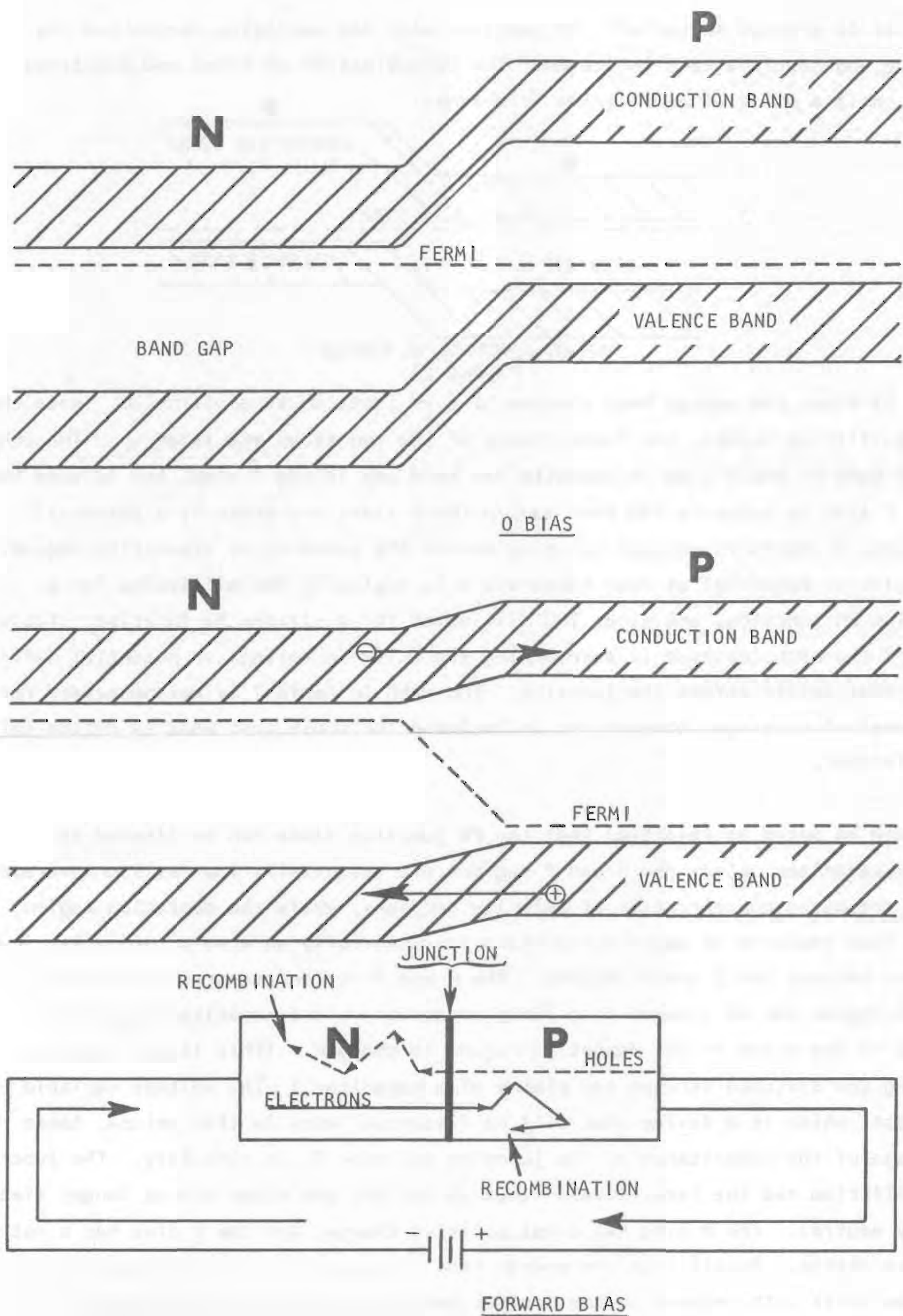


FIGURE 23

nucleus is assumed to have 0 electron volts. When the N side starts taking on a positive charge as the junction is formed, the nuclei are no longer at 0, but start taking on a positive charge. The entire N side of the diagram moves down. The P side, on taking on a negative charge, has its energy band diagram existing higher than before, and Figure 22 illustrates how far the diagrams move before equilibrium is reached. At equilibrium, the Fermi levels are lined up and the two sides of the diagram have shifted due to the built-in voltage, and a potential difference is existing across the junction. To have external current, some external energy must be applied to bring the junction out of equilibrium.

BIASING AND DIODE ACTION:

Consider the application of a battery with the polarity shown in Figure 23. Note that the negative side of the battery is reducing the positive charge of the N side, and the positive terminal of the battery is reducing the negative charge of the P side. The energy band diagram is distorted by the application of the external energy and electrons can move from the conduction band of the N side into the conduction band of the P side, and holes can move from the valence band of the P side into the valence band of the N side. This polarity of bias allows majority carriers (electrons in the conduction band of the N side, and holes in the valence band of the P side) to cross the junction, or transition region. Notice that once these carriers have crossed the junction they are no longer majority carriers. An electron on leaving the conduction band of the N side and entering the conduction band of the P side becomes a minority carrier. A hole on leaving the valence band of the P side and entering the valence band of the N side becomes a minority carrier

We previously have discussed the transport of minority carriers. The transport of minority carriers is by diffusion and they have a different velocity than when transported as majority carriers. The majority carriers that cross the junction with forward voltage applied become minority carriers and seek impurities or imperfections in the opposite side in order to accomplish recombination. That is, an electron crossing from the N material into the P material must find an impurity atom or some imperfection in the P material that will allow recombination. A hole, on migrating from the P material into the N material as a result of forward bias, must find an impurity or an imperfection in order to recombine with an electron. When an electron or hole that has crossed the junction finds an impurity atom or

some imperfection, it can accomplish recombination and is lost as a current carrier. The hole can be said to be replaced from the positive side of the bias source by the removal of an electron at the contact with the P side. The electron is replaced by the negative terminal of the battery by injection into the N side, and current can be considered continuous through the junction. It should be noted, however, that electrons, on crossing the junction, must move into the valence band to accomplish recombination. In order to do this, they must give up energy. Since electrons move from a higher energy level to a lower energy level, they must give up energy during recombination.

When recombination occurs, energy is released in the form of heat and light. The total external current that will flow is dependent on the number of minority carriers adjacent to the junction or transition region of the diode. The number of minority carriers is related exponentially to the applied voltage and the temperature.

Table 8 covers the math involved in determining the current as related to the number of minority carriers adjacent to the junction. For our purposes, it is sufficient that we understand that the application of a forward voltage (negative to N type, positive to P type) will result in majority carriers crossing the junction where they become minority carriers and recombine. One point that should be remembered when forward voltage is applied, is that after crossing the junction, the carriers must diffuse, since they are minority carriers. The total current in the external circuit is the sum of the electron recombination current and the hole recombination current when forward voltage is applied. It should also be remembered that the current will vary exponentially rather than directly as the applied voltage varies.

An increase in temperature increases the availability of minority carriers and increases the current for a given applied voltage. It can be seen in table 8 that the relationship of current to temperature is also an exponential function. A diode with forward voltage applied and conducting forward current is considered to be in its on state. A diode with reverse voltage applied has a small reverse current flowing at a given temperature and is considered to be in its off state.

Figure 24 shows the application of a reverse voltage to a PN junction. The negative side of the voltage source is increasing the negative ionic charge of the P material, and the positide side of the bias source is increasing the positive ionic charge of the N side. This widens the depletion region and increases the electric field or potential difference across the junction. Note in Figure 24 that the conduction band of the N material has been moved farther away from the conduction band of the P material. The valence band of the N material has been moved farther away from the valence band of the P material, reducing the probability of majority carriers crossing the junction. Minority carriers (electrons in the P conduction band and holes in the N valence band) that enter the transition region with reverse bias applied see a high electric field across the transition region that tends to accelerate them into the other side of the junction. Note in Figure 24 that an electron, on diffusing from the conduction band of P side into the transition region would be accelerated across the junction into the conduction band of the N side. Once into the N side, it simply becomes a majority carrier and a current results in the external circuit. A hole in the valence band of the N side, on entering the transition region, sees the high electric field due to the applied reverse voltage and is accelerated into the valence band of the P side. Once into the valence band of the P side, the hole becomes a majority carrier and current results in the external circuit. The current, with reverse voltage applied, is a result of minority carriers which are accelerated across the junction by the reverse voltage. Reverse voltage opposes the movement of majority carriers across the junction, but enhances the movement of minority carriers across the junction. Once across the junction, the minority carriers, of course, become majority carriers. The reverse current that flows when reverse voltage is applied is dependent on the number of minority carriers present adjacent to the transition region in the two sides of the diode. The minority carriers are present primarily as a result of thermal energy forming hole electron pairs. This current is normally referred to as saturation current and given the symbol I_S. A further increase in voltage after application of sufficient reverse voltage to cause all of the minority carriers to cross the junction results in no further increase in current. The total reverse current is the sum of the electrons in a conduction band of the P side which cross the junction and the holes in the valence band of the N side which cross the junction. Table 8 covers the math involved in determining the current with reverse voltage applied as well as with forward voltage applied.

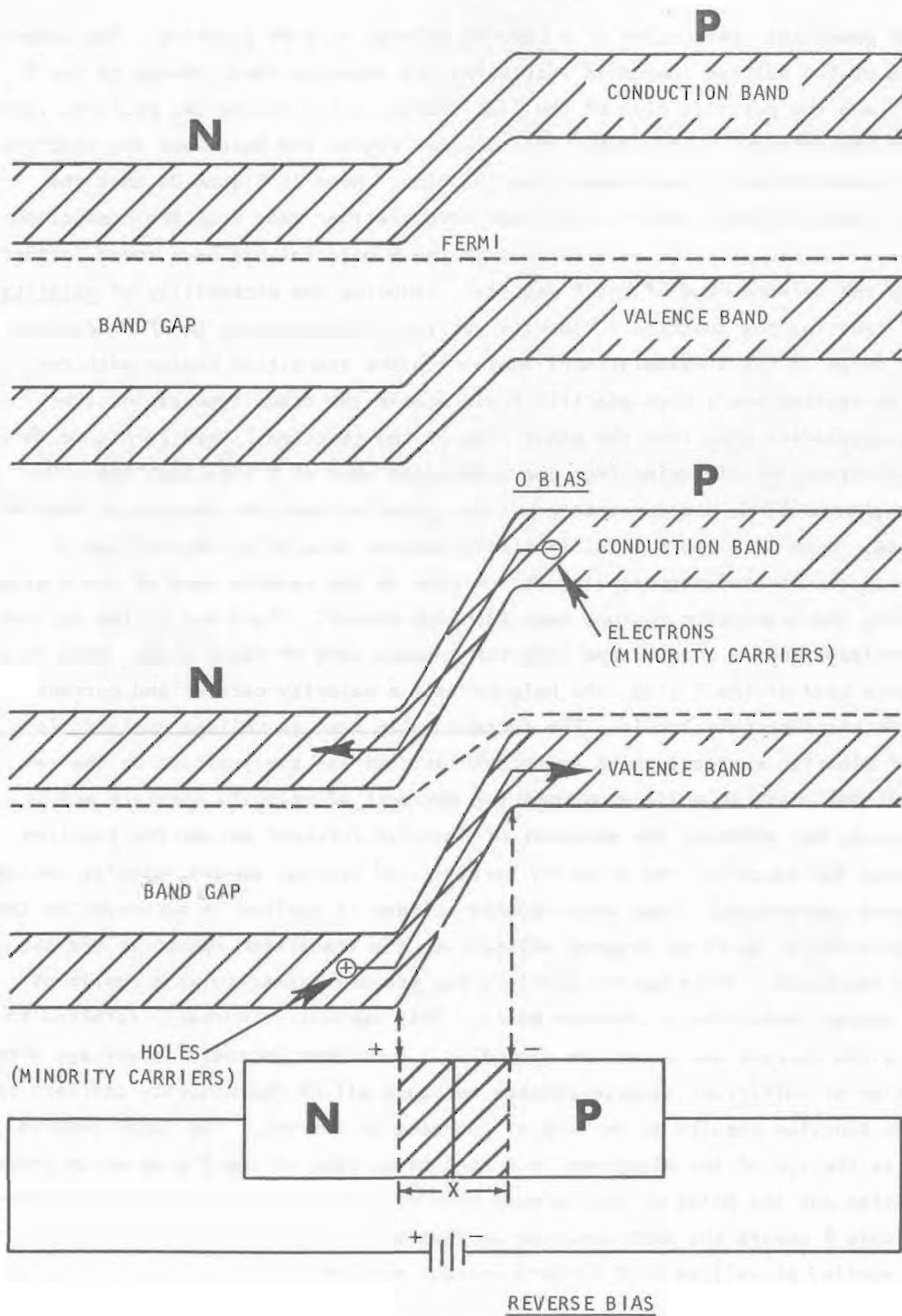


FIGURE 24

At this time realize that a current flows with reverse voltage applied and its magnitude depends on the number of minority carriers available in the two sides of the junction. These carriers primarily result from the formation of hole electron pairs by thermal energy. The number of minority carriers available relates exponentially to temperature therefore, the current with reverse voltage applied varies exponentially as the temperature varies. A plot of current versus voltage at a given temperature develops the curve of figure 25. This figure represents a typical germanium diode. Equilibrium occurs at the origin (X-Y axis intersection). Applying forward voltage results in an exponential curve from the origin extending to infinity. Reverse voltage causes a small quantity of saturation current flow (I_s).

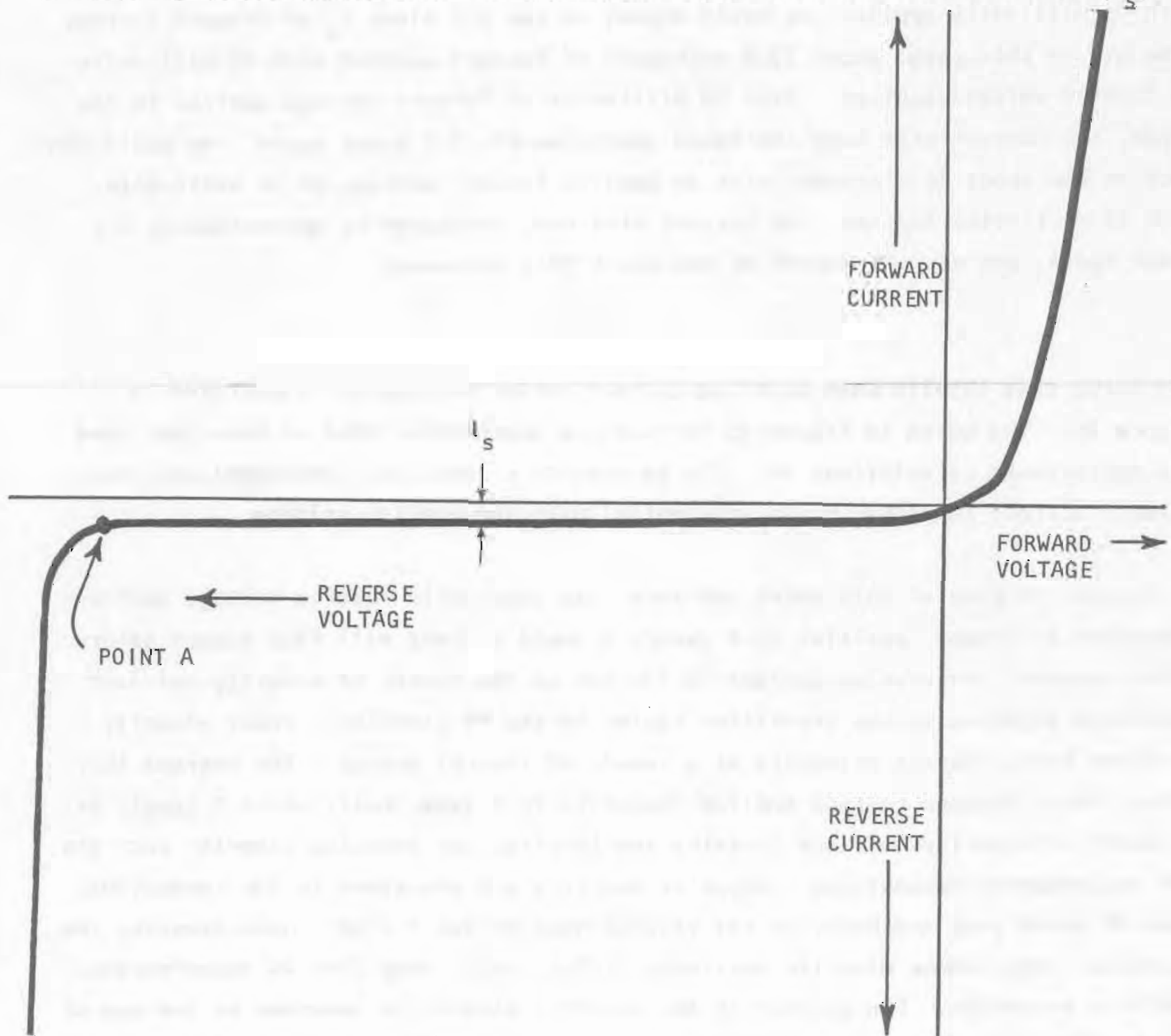


FIGURE 25

At the point marked A, the total minority carriers available in the adjacent junction sides govern current quantity. This is the maximum current at a given temperature with a reverse voltage applied. As the reverse voltage is decreased, the current will decrease to 0 and, as forward voltage is applied, forward current will start to increase on the same exponential curve. This current is formulated mathematically in table 8; however, expressed more simply, the current will change by a factor of approximately 2.7 for each 25 millivolts of change in the applied voltage. In other words, current will change 2.7 times for a change of 25 millivolts. As an example, assume that I_s is 5 microamps. The current that flows with forward voltage applied will change by 2.7 times for each change of 25 millivolts. With 25 millivolts applied, we would expect to see 2.7 times I_s of forward current flow or, in this case, about 13.5 microamps of forward current with 25 millivolts of forward voltage applied. With 50 millivolts of forward voltage applied to the diode, the current will have increased approximately 2.7 times again. We would expect to see about 36 microamps with an applied forward voltage of 50 millivolts. With 75 millivolts applied, the current will have increased by approximately 2.7 times again, and we will expect to see about 98.5 microamps.

The curve that results when plotting current versus voltage is illustrated in Figure 25. The curve in Figure 25 follows the exponential that we have just used the approximate calculations on. The saturation current is 5 microamps and the forward current increases on an exponential with the applied voltage.

It is well to stop at this point and take note that, with reverse voltage applied (netative to P type, positive to N type), a small current will flow termed saturation current. Saturation current is limited by the number of minority carriers available adjacent to the transition region of the PN junction - these minority carriers being present primarily as a result of thermal energy. The current that flows with a forward voltage applied (negative to N type, positive to P type), is a result of majority carriers crossing the junction and becoming minority carriers, and subsequently recombining. Majority carriers are electrons in the conduction band of the N side and holes in the valence band of the P side. Upon crossing the junction, they become minority carriers, diffuse until they find an imperfection, and then recombine. The current in the external circuit is governed by the sum of the electrons that cross the junction and recombine, and the holes that cross the

junction and recombine. The current in the external circuit with forward voltage applied is once again governed by the number of minority carriers adjacent to the junction. With forward voltage applied, however, the minority carriers are primarily majority carriers that have crossed the junction and are seeking imperfections to accomplish recombination. With reverse voltage applied, the current is limited because there is a limited number of minority carriers available; however, with forward voltage applied, the number of minority carriers is increased since majority carriers cross the junction and become minority carriers. The current in either case is related exponentially to the applied voltage and the temperature. At a given temperature, the current can be plotted with respect to the voltage, assuming that the current will increase approximately 2.7 times for each 25 millivoltsincrease in applied forward voltage. This is true for either forward or reverse voltage; however, with reverse voltage, the sign of the voltage is negative and the result is a negative current. We can say that a forward voltage will result in a fairly large current if the applied voltage is sufficient, while a reverse voltage will result in a fairly small current at a given temperature. The current with reverse voltage applied will be limited by the number of minority carriers available as a result of thermal energy and any leakage current that might be apparent in the diode.

MINORITY CARRIER LIFETIME:

The average time that minority carriers exist in a material before recombining is termed the minority carrier lifetime of the material. With small amounts of imperfections in a semiconductor, the minority carrier lifetime is fairly long. As the number of impurities, or imperfections, is increased, either by doping or some other means, the minority carrier lifetime is reduced. The minority carrier lifetime can be considered a figure of merit for a semiconductor. The longer the minority carrier lifetime (i.e.; the average time the minority carriers exist as such before recombining), the greater the purity of the semiconductor material. As the doping is increased, or some method used to provide recombination traps or imperfections in the structure that allow recombination, the shorter the minority carrier lifetime. Since the current in a semiconductor diode is dependent on the recombination of minority carriers, the minority carrier lifetime becomes important, especially where speed considerations are involved. Recall that when majority carriers are allowed to cross the junction, they become minority carriers and must

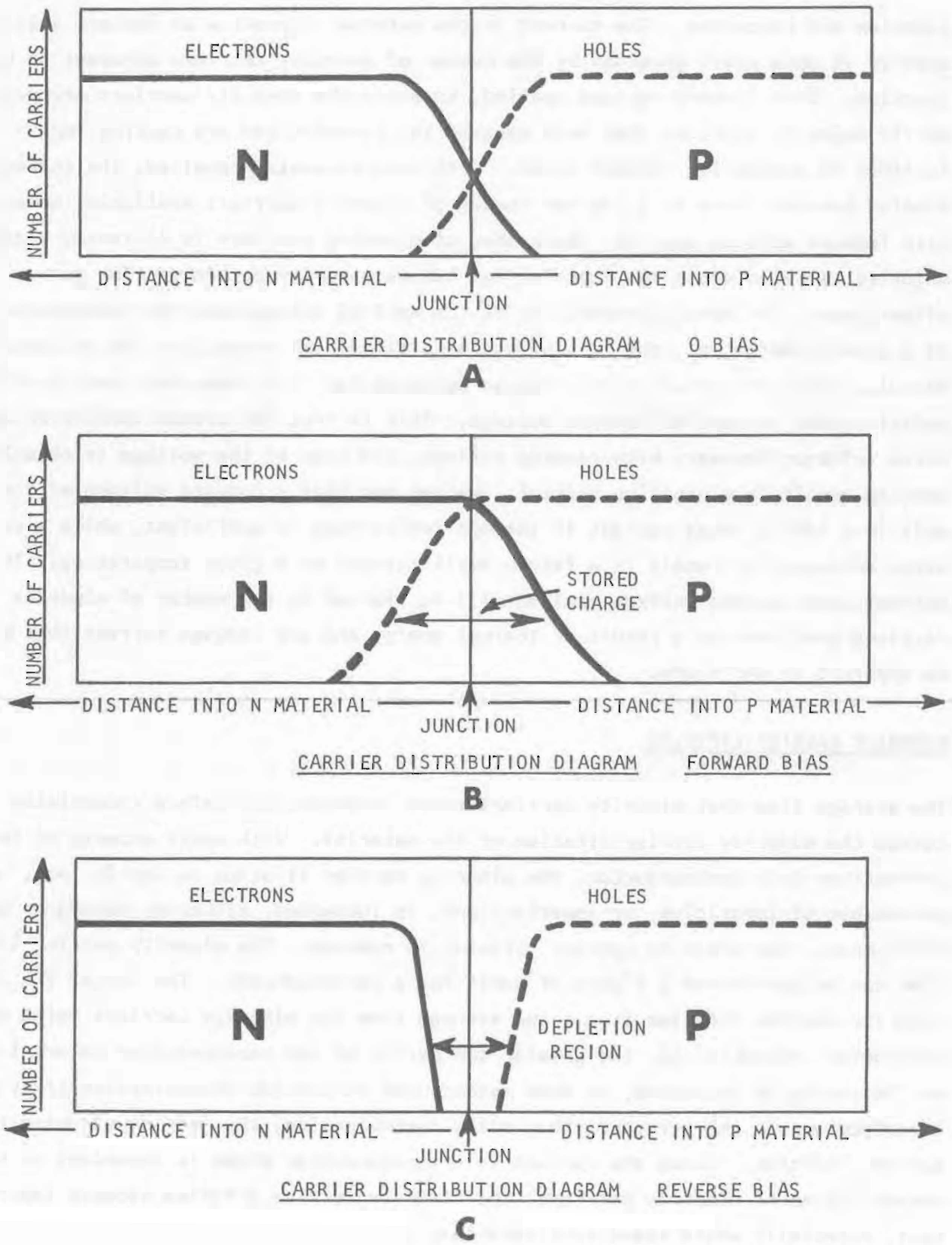


FIGURE 26

diffuse until they find imperfections and accomplish recombination. While they are diffusing and before recombination occurs, they represent a stored charge. In other words, these electrons and holes, having crossed the junction, are existing as excess carriers in the two sides and represent a charge. If the imperfections or impurities near the junction are few, the carriers might have to diffuse for a fairly long period of time before accomplishing recombination. The stored charge about the junction could be relatively large for a given forward current. If the imperfections in the two sides of the diode are many for the same forward current, the stored charge would be much smaller. It can be said then that the effective charge, stored about the junction of the forward conduction diode, is directly proportional to the minority carrier lifetime. The greater the minority carrier lifetime, the greater the stored charge for a given forward current. Of course, if the forward current is increased, more carriers cross the junction and more carriers will be diffusing, searching for imperfections to accomplish recombination and the stored charge increases. It is generally considered that stored charge has a linear relationship with forward current. In other words, if forward current were to double for a given diode, the charge stored would double.

Let's stop and review that. For a given diode, the charge stored about the forward conducting junction represented by the diffusing minority carriers is directly related to the forward current and the lifetime of the minority carriers. The greater the lifetime of the minority carriers, the greater the stored charge for a given forward current. For a given forward current, the longer the minority carrier lifetime and the greater the stored charge. Since the diffusing minority carriers must be removed before the diode can go into its normal non or low conducting state, the stored charge becomes important when a diode is used in fast switching applications, since switching from a forward conducting mode to a reverse bias mode results in a fairly large reverse current initially.

Figure 26 shows the carrier distribution for a forward, reverse, and zero bias state. Note that the carrier distribution diagram indicates that there are some minority carriers existing in the two sides with zero volts applied. These minority carriers are primarily the result of thermal energy forming hole electron pairs. As forward bias is applied, the number of minority carriers increases greatly in the two sides. This is a result of forward bias allowing the majority carriers to cross the junction and become minority carriers. Note that some of these minority

carriers must diffuse well into the opposite side before they accomplish recombination. The minority carriers represented by the cross hatched area in the forward bias curve in Figure 26 represent the stored charge about the forward conduction junction. τ_q is the diode parameter given that relates the minority carrier lifetime and forward current to stored charge. τ_q assumes a linear relationship between stored charge and forward current. τ_q is given in pico-coulombs per milliampere and when taken as a product with diode forward current (in milliamperes) gives the value of diode stored charge (in pico-coulombs).

DIODE THERMAL CONSIDERATIONS AS RELATED TO POWER DISSIPATION:

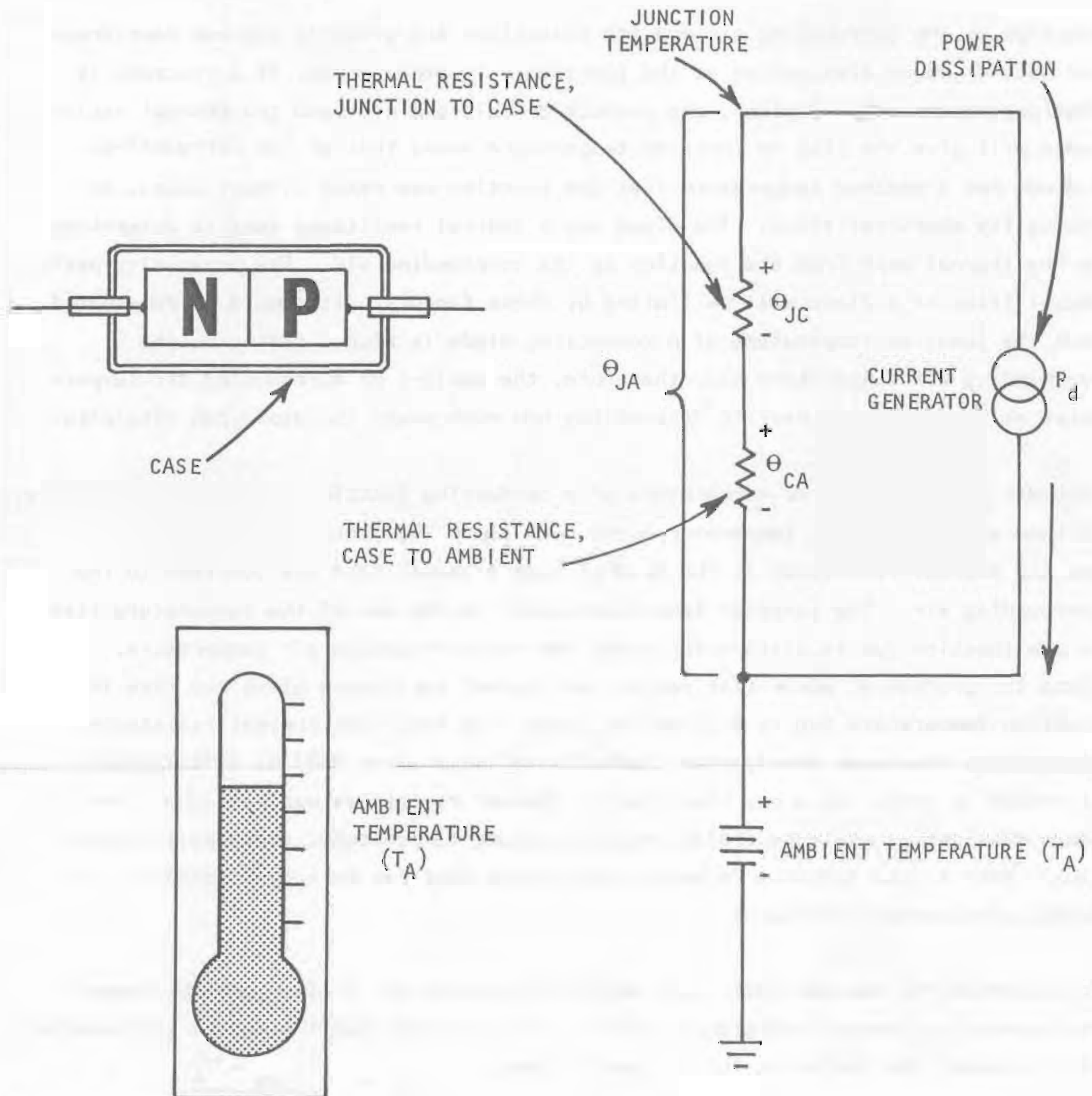
It has already been stated that the current, voltage, and resistance of a semiconductor PN junction is related exponentially to the temperature. A temperature rise at the junction will cause a change in the current, voltage, and resistance of the semiconductor junction. The change in temperature at the junction can be caused externally or internally of the diode. A change in the ambient or surrounding air temperature will change the temperature of the junction for a given power dissipation, and a change in power dissipation at the junction will change the junction temperature for a given ambient temperature. A conducting junction will generate heat since the carriers changing bands must give up their energy in the form of heat and light. Any heat generated at the junction will raise the junction temperature about that of the surrounding air. At a given temperature, a PN junction will no longer serve as a rectifying device. Since the current is related exponentially to the temperature, a temperature limit will be reached where a further change in temperature will cause the diode to be essentially a short circuit and damage to the circuit or the diode can result. With germanium PN junctions, this is typically a 100° centigrade, while with silicon PN junctions, this is typically 175° centigrade. Depending on the intrinsic material and the type of doping and the construction of the device, this may vary above and below the typical values given. The diode is normally enclosed in some type of encapsulation or case, and there is opposition to the transfer of heat to the surrounding air. The opposition offered in series with the path of heat transfer from the junction to the surrounding air is related analogically to electrical resistance and is termed thermal resistance. Thermal resistance is the opposition to the transfer of heat from the

junction to the surrounding air and its dimensions are given in degrees centigrade per watt of power dissipation at the junction. In other words, if a junction is dissipating one watt of power, the product of this one watt and the thermal resistance will give the rise in junction temperature above that of its surroundings. A diode has a maximum temperature that the junction can reach without damage or losing its characteristics. The diode has a thermal resistance that is determined by the thermal path from the junction to the surrounding air. The power dissipation capabilities of a diode will be limited by these factors. It should be remembered that the junction temperature of a conducting diode is always riding on the surrounding air temperature and, therefore, the ambient or surrounding air temperature will also have a part in determining how much power the diode can dissipate.

In order to determine the temperature of a conducting junction, it will be necessary to know the surrounding temperature, how much power the junction is dissipating, and the thermal resistance in the path of heat transfer from the junction to the surrounding air. The junction temperature will be the sum of the temperature rise at the junction due to dissipating power and the surrounding air temperature. Since the product of power dissipation and thermal resistance gives the rise in junction temperature due to dissipating power, the lower the thermal resistance the greater the power dissipation capabilities for a given ambient temperature. It should be noted, however, that even if thermal resistance was zero ohms, the power dissipation would be limited by the ambient temperature, so there is a practical limit to the increase in power dissipation that can be accomplished by reducing the thermal resistance.

To determine the maximum power that may be dissipated for a given ambient temperature when the thermal resistance of the diode from the junction to the surrounding air is known, the following formula can be used: $P_{max} = \frac{T_{Jmax} - T_A}{\theta_{JA}}$

P_{max} indicates the maximum allowable power dissipation, T_{Jmax} indicates the maximum junction temperature of the device under operating conditions, T_A indicates the ambient temperature, and θ_{JA} is the symbol given thermal resistance junction to ambient. When power dissipation is limited by the total thermal resistance of a diode, an external heat sink can be added to aid in transferring the heat to the surrounding air. When external heat sinking is added, the total thermal resistance



JUNCTION POWER DISSIPATION = CURRENT

OPPOSITION TO HEAT TRANSFER = RESISTANCE

AMBIENT TEMPERATURE = VOLTAGE

JUNCTION TEMPERATURE = VOLTAGE

ANALOGICAL ASSOCIATION OF ELECTRICAL AND THERMAL CHARACTERISTICS

FIGURE 27

junction to ambient is reduced. An example of heat sinking might be connecting the diode thermally to the chassis or to some type of radiating element. Low power diodes, switching diodes, etc., do not normally use an external heat sink, and the typical thermal resistance might be on the order of 250° centigrade per watt. Rectifier diodes and diodes designed to handle more power might be mounted in an encapsulation that has a heavy stud which can be protruded through the chassis and bolted down tight. This allows the chassis to aid in radiating the heat generated at the junction to the surrounding air. It is possible using external heat sinking to reduce the total thermal resistance down to a few degrees centigrade per watt.

Figure 27 shows the analogical association of electrical and thermal characteristics of a PN junction. A model is shown of the diode in its case in Figure 27, and no external heat sinking is being used. The predominate thermal resistances involved are the thermal resistance between the junction and the case and the thermal resistance between the case and the surrounding air; in other words, the opposition offered in the transfer of heat from the junction to the case or encapsulation and the opposition offered to heat transfer from the case to the surrounding air. The sum of the two predominate thermal resistances will give the total thermal resistance from junction to ambient. Note that thermal resistance is related analogically to electrical resistance, while ambient temperature is related analogically to electrical voltage, and power dissipation is related analogically to electrical current. Once these associations have been made, Ohm's Law may be used in solving for junction temperature, or maximum power dissipation, etc.

Since the voltage rises across the thermal resistances are aiding in polarity to the ambient temperature, the junction temperature measured from the point shown in Figure 27 will be some voltage greater than the ambient temperature. Since power dissipation is related to an electrical current, the product of this current and the thermal resistances gives the voltage drops across the thermal resistances in the analogy. Therefore, the product of power dissipation and thermal resistance will give the voltage (or temperature rise) at the junction. This, added to the ambient temperature, will give the junction temperature. Expressing this as a formula gives:

$$T_J = \Theta_{JC} P_d + \Theta_{CA} P_d + T_A$$

where T_J is the junction temperature, P_d is power dissipation, Θ_{JC} is the thermal resistance junction to case, Θ_{CA} is the thermal resistance case to ambient, and

T_A is the ambient temperature. It is well to remember that these formulas are possible as a result of relating the thermal characteristics to electrical characteristics and using Ohm's Law. The formula for the maximum power dissipation for a given ambient temperature and thermal resistance is simply the allowable change in junction temperature divided by the thermal resistance, or in formula form:

$$P_{max} = \frac{\Delta T_{J\ allowable}}{\Theta_{JC} + \Theta_{CA}} = \frac{T_{Jmax} - T_A}{\Theta_{JA}}$$

As an example, consider a PN junction diode that has a maximum operating junction temperature of 100° centigrade and the ambient temperature is 50° centigrade. In this case, the total allowable change in junction temperature due to dissipating power is the difference in these two temperatures, or 50° centigrade. If the thermal resistance is 250° centigrade per watt, the maximum power dissipation is 0.2 watts, or 200 milliwatts. Notice that we did nothing more than divide the change in ° centigrade per watt that would occur at the junction into the allowable change in junction temperature, and this gave us the maximum power dissipation of the junction.

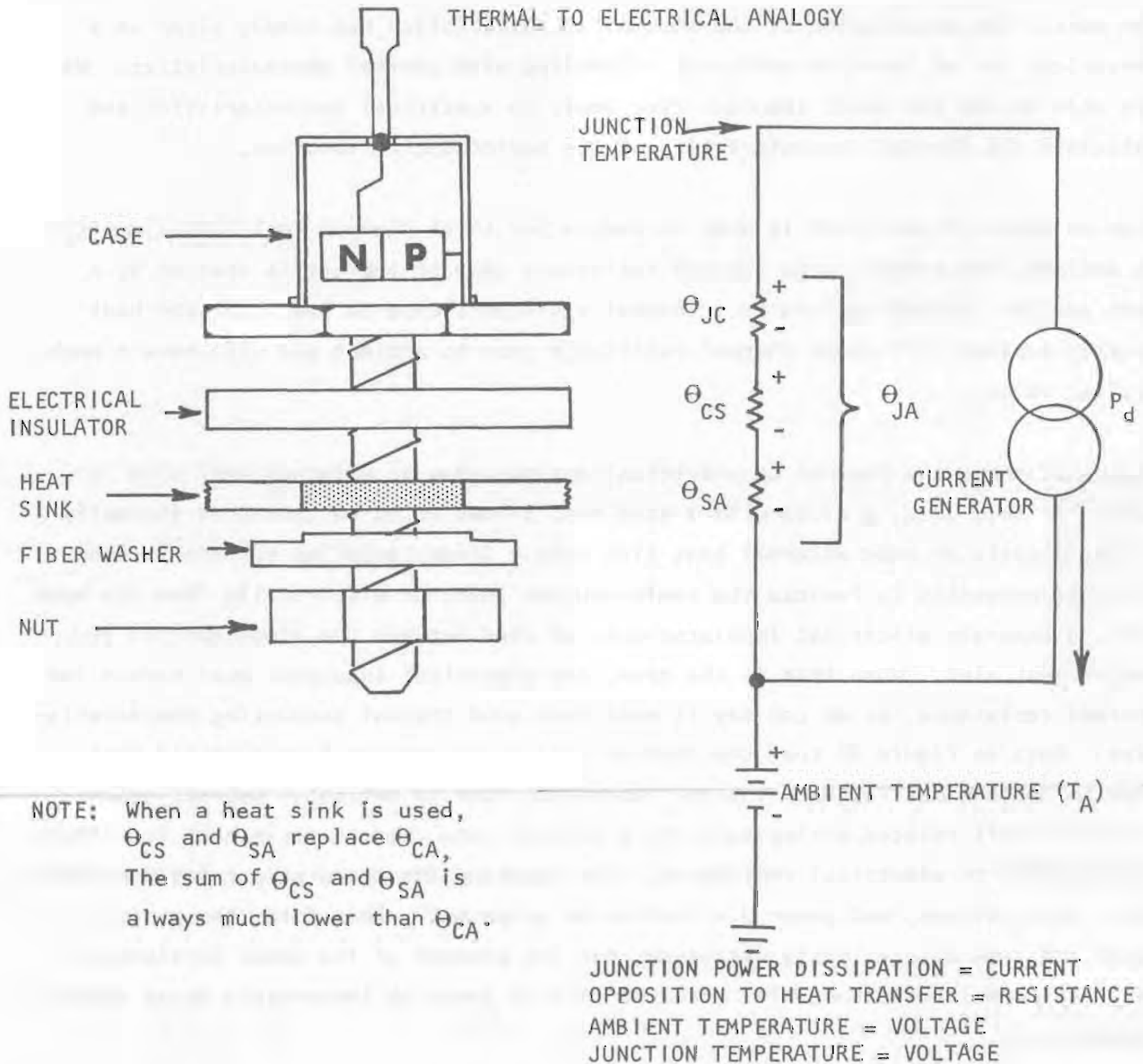
Let's take another example, only this time let's determine how many ° centigrade above the ambient temperature the junction is existing. A diode has a total thermal resistance of 25° centigrade per watt and is dissipating two watts of power. What is the junction temperature if the ambient temperature is 25° centigrade? In this case, the rise in junction temperature is simply the product of the thermal resistance and the power dissipation at the junction. Two watts of power times 25° centigrade per watt indicates that the junction will be 50° centigrade above the surrounding air. Since we have stated that the surrounding air is at 25° centigrade, we would expect to measure 75° centigrade at the junction. Looking at the same problem, if we had the ability to make temperature measurements at the junction and of the surrounding air, we can certainly determine the thermal resistance of the device. Considering the previous problem where the surrounding air temperature was 25° centigrade and the junction temperature was 75° centigrade, we can see that the product of junction current and voltage would give the amount of power we were dissipating; in this case, two watts. Dividing two watts into the rise in junction temperature above ambient would give us the number of degrees centigrade per watt, or the thermal resistance of the junction; in this case 25° centigrade

per watt. The association of the thermal characteristics has simply given us a convenient set of tools to work with in dealing with thermal characteristics. We are able to use the basic theorems that apply to electrical characteristics and calculate the thermal characteristics of the semiconductor junction.

When an external heat sink is used to reduce the total thermal resistance junction to ambient, the rather large thermal resistance case to ambient is shunted by a much smaller thermal resistance. Thermal resistance case to heat sink and heat sink to ambient will shunt thermal resistance case to ambient and will have a much smaller value.

Figure 28 shows the thermal to electrical analogy when an external heat sink is used. In this case, a diode with a stud that allows it to be connected thermally to the chassis or some external heat sink with a large radiating surface is used. If it is necessary to isolate the semiconductor junction electrically from the heat sink, a separate electrical insulator must be used between the diode and the external heat sink. When this is the case, the electrical insulator must have a low thermal resistance, or we can say it must have good thermal conducting characteristics. Note in Figure 28 that the thermal resistance case to heat sink and heat sink to ambient has replaced thermal resistance case to ambient. Ambient temperature is still related analogically to a voltage, and thermal resistance is related analogically to electrical resistance. The junction temperature is related analogically to a voltage, and power dissipation to a current. Therefore, the analogy remains the same as previously discussed, and the product of the power dissipation and the thermal resistance will give the rise in junction temperature above ambient temperature.

Note in Figure 28 that even if the thermal resistance of the heat sink and the connection of the diode to the heat sink were zero, the power dissipation is limited by the thermal resistance junction to case. Junction temperature will always be greater than ambient temperature by the product of power dissipation and the inherent thermal resistance. Thermal resistance junction to case is determined by the semiconductor device in its encapsulation. Thermal resistance heat sink to ambient is determined by the radiating surface of the heat sink, and thermal resistance case to heat sink is set by the thermal connection between the device and the heat sink. If no electrical insulator is needed, the case of the diode can be connected



Insulating Washer	Typical Thermal Resistance (Θ_{CS}) in °C/W	
	Dry	W/Silicon Lubricant
None	0.2	0.1
Teflon	1.45	0.8
Mica	0.8	0.4
Anodized Aluminum	0.4	0.35

directly to the heat sink. The chart at the bottom of Figure 28 indicates some typical thermal resistances case to heat sink for direct connection and for several types of insulating washers. When no insulating washer is needed and the case is connected directly to the heat sink, the typical thermal resistance is in the order of 0.2° centigrade per watt. Note in the chart at the bottom of Figure 28 that when silicon lubricant is added, the thermal resistance is reduced to 0.1° centigrade per watt. Silicon lubricant is made available by several manufacturers and, when applied to the insulating washer or between the semiconductor device and its heat sink, reduces the thermal resistance. Reducing the thermal resistance case to heat sink, of course, reduces total thermal resistance and allows a higher power dissipation for a given ambient temperature. The silicon lubricant has good heat conducting properties and fills in the gaps between the semiconductor device and the heat sink or insulating washer. Much the same effects can be obtained by using an abrasive and highly polishing the surfaces that are going to come together. The silicon lubricant minimizes the need for polishing the surface of the semiconductor case and the heat sink for good heat transfer.

When an insulating washer is required, silicon lubricant will still reduce the inherent thermal resistance case to heat sink. Figure 28 lists three typical insulating washers that are used with semiconductor devices. Note that the thermal resistances will be substantially reduced by the application of silicon lubricants.

It should be brought out at this time that it takes a period of time for the junction temperature to change when power dissipation changes. In other words, if we were to suddenly increase the power dissipation, it would take a period of time for the junction temperature to change to its new value. In order to gain a set of tools to deal with the period of time involved in the change in junction temperature, a time constant of thermal resistance and thermal capacitance is assumed. Figure 29 shows the thermal capacitance in shunt with the thermal resistance of the device. To simplify the thermal to electrical analogy, only one thermal capacitance is shown in shunt with all three of the thermal resistances in Figure 29. It should be remembered that there are individual time constants made up of thermal resistance junction to case and it's associated thermal capacitance, thermal resistance case to heat sink and it's associated thermal capacitance, and thermal resistance heat sink to ambient and its thermal capacitance. To keep the model as simple as we can and still have tools to work with, we have assumed one thermal time constant of

THERMAL TO ELECTRICAL ANALOGY

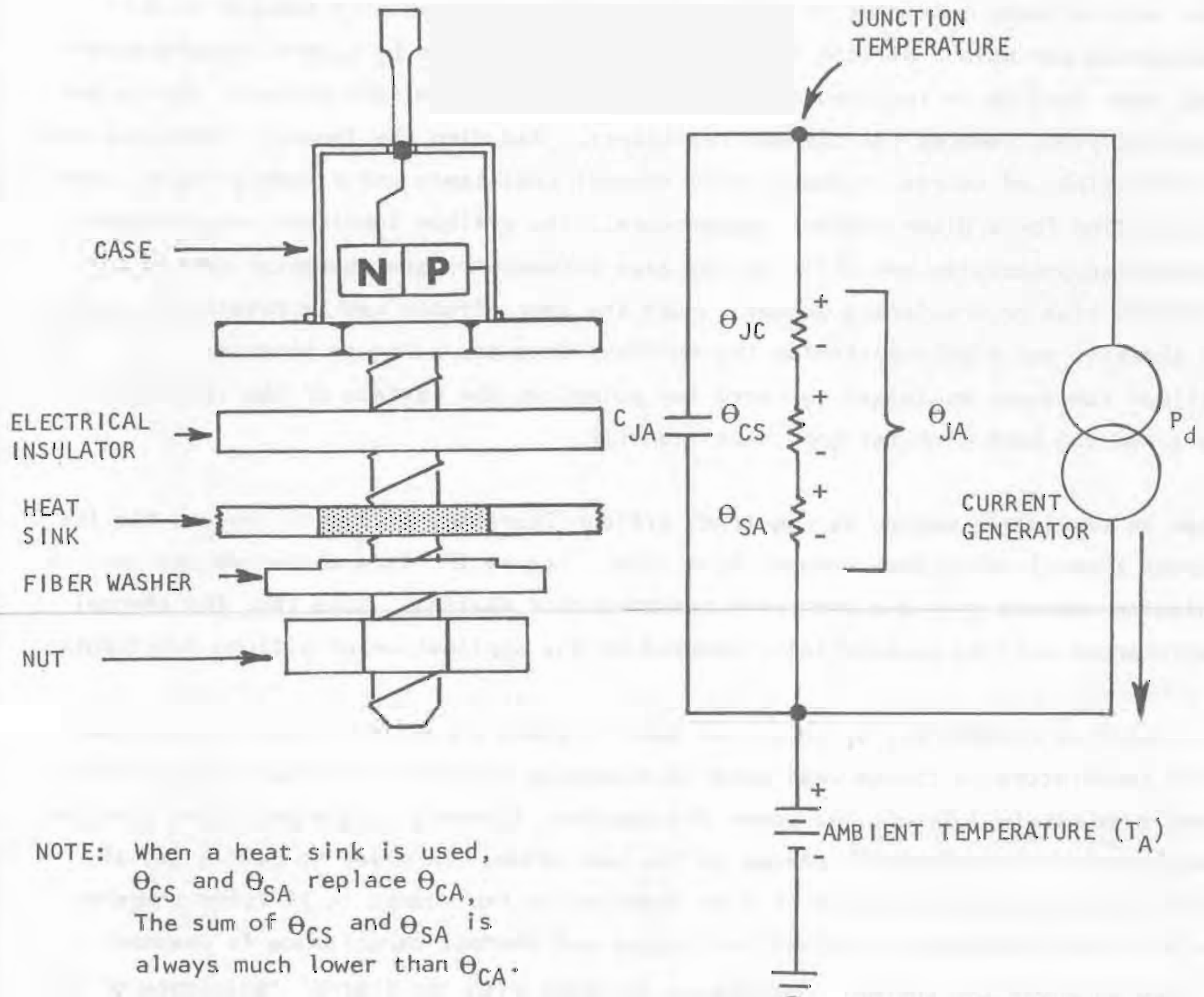


FIGURE 29

total thermal resistance in shunt with thermal capacity.

Thermal capacity is given in dimensions of watt-seconds per degree centigrade. The product of thermal resistance in degrees centigrade per watt and thermal capacity in watt-seconds per degrees centigrade will give a resultant time. The time, as a result of taking the product of thermal capacitance and thermal resistance will be one thermal time constant. Assuming that, at time zero, there is no power dissipation and suddenly the diode is caused to conduct resulting in a power dissipation, it will take approximately five thermal time constants for the change in junction temperature to occur. In other words, when we start dissipating power, it will take about five thermal time constants for the junction temperature to get to its final value above the surrounding air temperature. We can also say that, if the junction temperature is at some level above ambient temperature and the power dissipation is removed or reduced, it will take approximately five thermal time constants for the junction temperature to reach ambient once again.

Let's stop and take stock for a moment. First of all, the thermal characteristics are related analogically to electrical characteristics to give a set of tools to work with when dealing with thermal characteristics. Once the tools have been gained, the basic electrical theorems can be used to solve problems dealing with thermal characteristics. First of all, thermal resistance and power dissipation, being related analogically to current and resistance, allows junction temperature to be calculated as a voltage at some potential difference above ambient temperature. This also allows the calculation of the maximum power dissipation for a given thermal resistance and ambient temperature. Transposing the formulas involved allows the calculation of any one quantity when the other two are known. When external heat sinking is used, the total thermal resistance is reduced. Thermal resistance case to ambient is replaced by a lower thermal resistance made up of a thermal resistance case to heat sink and heat sink to ambient. Silicon lubricant added between the diode and the heat sink, or to the insulating washer when it is required, reduces the total thermal resistance. If it is necessary to deal with the time involved in the temperature change at the junction, another component is added and termed thermal capacitance. Thermal capacitance is given in watt-seconds per degree centigrade. The produce of thermal capacitance and the thermal resistance gives a time constant that will allow us to deal with the time involved in the change of junction temperature when power dissipation changes. It might be difficult at this

time to see the need for thermal capacitance; however, we will deal with thermal capacitance when we deal with pulse power effects.

CALCULATING A REQUIRED HEAT SINK:

Tables 10, 11, and 12 plot the thermal resistances in typical metals for a variety of areas and thicknesses. Using these charts, the size of the heat sink required to give us a specified thermal resistance when dealing with a semiconductor device can be determined. As an example, consider table 10. Suppose the required thermal resistance is 4° centigrade per watt or less. Looking at the horizontal plot in table 10, find 4° centigrade per watt. Next find where the line indicating 4° centigrade per watt intersects one of the curves. Notice that it will intersect the 1/4" thickness aluminum plate and the 1/8" thickness aluminum plate. Taking the point of intersection with the 1/4" aluminum plate and moving to the left, the graph indicates that it will take over 70 square inches of dissipating area in order to give a thermal resistance of 4° centigrade per watt. Dealing with 1/8" plate in table 10 and finding the point of intersection with the 4° centigrade per watt curve, it would require around 180 square inches of dissipating area in order to give a thermal resistance of 4° centigrade per watt. When thinking of heat sinking a diode or some other semiconductor device, one might first determine the type of insulating washer that is going to be used (if it is required), and the table at the bottom of Figure 28 gives the typical thermal resistance case to heat sink. The manufacturer normally lists the thermal resistance or thermal conductance (which is simply the reciprocal of thermal resistance) for the semiconductor device. Knowing the amount of power the device is going to dissipate, and what ambient temperature is expected, the maximum allowable thermal resistance of the heat sink could be calculated. Tables 10, 11, or 12 could be used to determine how much dissipating area would be required to hold the thermal specifications.

In order to calculate the required thermal resistance of the heat sink, we might take the formula:

$$P_{Dmax} = \frac{T_{Jmax} - T_A}{\Theta_{JA}}$$

and transpose it to solve for thermal resistance:

$$\Theta_{JA} = \frac{T_{Jmax} - T_A}{P_{Dmax}}$$

Total thermal resistance is equal to the sum of the individual thermal resistances

$$\theta_{JA} = \theta_{JC} + \theta_{CS} + \theta_{SA}$$

and, therefore, the thermal resistance of the heat sink is equal to:

$$\theta_{SA} = \theta_{JA} - \theta_{JC} - \theta_{CS}$$

In terms of temperature and power dissipation, the thermal resistance of the heat sink is equal to:

$$\theta_{SA} = \left(\frac{T_{Jmax} - T_A}{P_{Dmax}}\right) - \theta_{JC} - \theta_{CS}$$

In other words, if the allowable change in junction temperature is divided by the maximum power that it is expected to dissipate, total thermal resistance is obtained. Subtracting the known values of thermal resistance from the total allows the calculation of the maximum allowable thermal resistance of the heat sink. Tables such as shown in table 10, 11, and 12 might then be used to determine the total dissipating area required to give the desired thermal resistance.

Size limitations might require that the heat sink be of some configuration that will give the required radiating area and yet hold it to a small size. Some manufacturers have come forth with heat sinks with radiating fins. This allows the radiating area to be larger while keeping the size small. Some chassis on which the devices are mounted are vented or finned to increase the radiation surface.

It has not been stated as yet, however, the considerations that we are presently covering apply to all semiconductor junction devices and include all types of diode devices and transistors. Although we are more accutely aware of the thermal considerations and limitations with higher power devices, it is well to remember that these considerations apply to low power devices as well. Any semiconductor device will have its maximum power dissipation limited by the surrounding air temperature, the total thermal resistance and the maximum operation temperature of the junction. Changes in temperature at the junction as a result of power dissipation as well as changes in the surrounding air temperature will change the currents and voltages associated with the device. In diode devices, the predominant effects are changes in the leakage or saturation current of the diode with reverse voltage applied, changes in the voltage across the diode with forward voltage applied, and changes in the voltage across the diode when the diode is in a reverse breakdown condition.

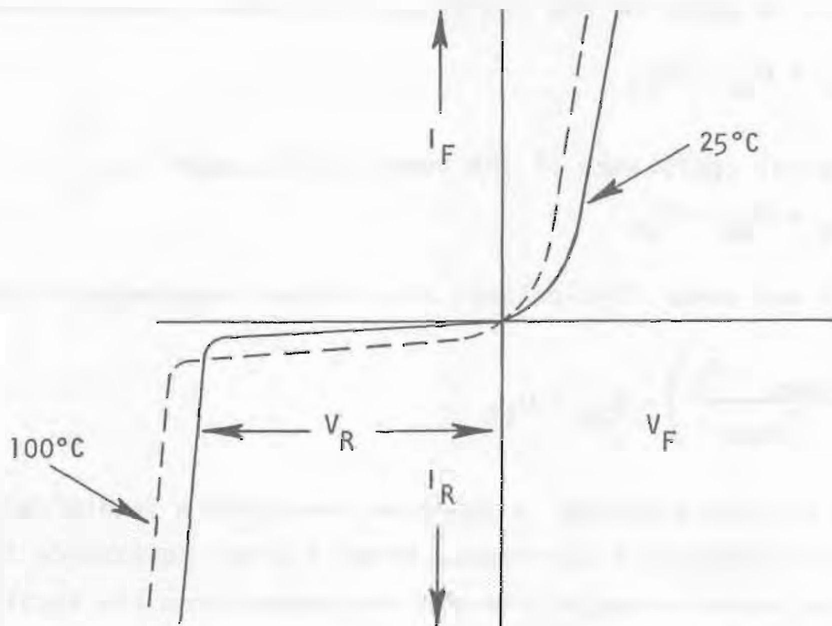


FIGURE 30

Figure 30 is a voltage versus current curve for a semiconductor diode with forward and reverse voltage applied. The solid curve indicates the characteristics at 25° centigrade, and the dotted curve indicates the characteristics at 100° centigrade. You will note that with forward voltage applied, the forward d-c resistance and the forward d-c voltage decreases with an increase in temperature. With reverse voltage applied, the d-c resistance of the diode decreases with an increase in temperature and the reverse or saturation current increases. The voltage across the device in its reverse breakdown condition, however, increases with an increase in temperature. This, of course, is assuming that the diode is in an avalanche breakdown condition and, therefore, we show the breakdown occuring above 6 volts. A temperature can be reached at which the device starts losing its characteristics and a decrease in the reverse d-c voltage across the diode can be observed. Within the normal operating temperatures of the device, we can say that the forward d-c resistance and voltage varies inversely as temperature varies; therefore, the forward voltage and resistance characteristics of a semiconductor diode have a negative temperature coefficient of voltage. The reverse d-c resistance of a semiconductor device not in a reverse breakdown mode can be said to vary inversely as the temperature varies. The reverse d-c resistance decreases as the temperature increases. This is a result of the formation of more hole electron pairs with an increase in temperature, increasing the number of current carriers available and decreasing the resistance of the device. When the semiconductor device is in an avalanche breakdown mode at

greater than 6 volts of reverse voltage, the d-c voltage across the device will increase with an increase in temperature. If the device is doped heavily enough to have it operating in the tunnel breakdown mode, an increase in temperature will cause the voltage across the reverse biased diode to decrease. Showing the breakdown voltage to be above 6 volts, we can then refer to the voltage of the device in breakdown as having a positive temperature coefficient of voltage. It should be kept in mind that the temperature change as indicated here can originate externally or internally of the diode. Internal temperature changes as a result of power dissipation and external temperature changes as a result of a change in the surrounding air temperature.

PULSE AND SWITCHING THERMAL CHARACTERISTICS:

To this point, the discussion about diode thermal considerations has related to steady state power dissipation. If the diode or transistor, as the case may be, is responding to pulses or operating in a switching mode or configuration, the steady state power dissipation may be small or even considered to be zero. Considerable power may be dissipated during the time of switching, however. The amount of time that the diodes spend dissipating a given amount of power starts to become important. The thermal capacitance previously discussed and the thermal time constant of the device come into play.

Consider a diode that is initially off, but has a pulse applied that will turn it on to a fairly high current for a short period of time and then allow the diode to return to an off state once again. The diode in the off state might be dissipating very little or no power at all, but when turned to its on state, can be dissipating a considerable amount of power. If the diode is allowed to remain in the on state for five thermal time constants, the temperature of the junction will reach a temperature given by the product of the peak power dissipation and the device's thermal resistance, plus the ambient temperature, or:

$$T_J = P_{Dpeak} \Theta_{JA} + T_A$$

where T_J is the junction temperature, P_{Dpeak} is the power dissipation in the on state, Θ_{JA} is the total thermal resistance of the device, and T_A is the ambient temperature. It is well to note, however, that if the diode does not stay in the on state for the full five thermal time constants, the temperature of the junction

will be somewhat less than this. If the junction is then allowed to cool for five thermal time constants, when the next pulse appears, the junction will once again require five thermal time constants to reach the temperature set by this value of peak power. If it is not allowed to cool for five thermal time constants, it will be at some temperature above the ambient temperature at the time the pulse is applied, and will start heating on an exponential time curve towards its maximum value again. From this, it can be seen that the duty cycle will also play a part in governing the temperature of the junction as related to peak power when dealing with pulses. It is possible, under certain conditions, to handle considerably more peak power than steady state power for a given diode and ambient temperature.

The ratio of the peak power dissipation capabilities of the junction with respect to its steady state power dissipation capabilities is termed the coefficient of power of the diode. Steady state or d-c coefficient of power is, of course, unity. This would be referred to as a 100% duty cycle. If the pulse width is very narrow with respect to the thermal time constant of the device, the coefficient of power is essentially the reciprocal of the duty cycle. In other words, a 10% duty cycle has a coefficient of power of 10 (assuming the pulse widths are very narrow with respect to the thermal time constant). With a coefficient of power of 10, we would expect to be able to dissipate 10 times the peak power as the diode could dissipate as steady state power. Carrying this farther, if the duty cycle were 1% (still assuming that the pulse width is very narrow with respect to the thermal time constant), we would expect the diode to be able to dissipate 100 times the peak power than it could dissipate as steady state power. If the pulse is a single transient (absolutely non-reoccuring), the coefficient of power is equal to the thermal time constant divided by the pulse width:

$$C_{PJA} = \frac{\tau_{JA}}{P_W}$$

where C_{PJA} is the coefficient of power, τ_{JA} is the thermal time constant, and P_W is pulse width.

Consider a diode with a thermal time constant of 50 milliseconds, the thermal time constant being determined by the product of thermal capacitance and thermal resistance. If a single non-reoccuring transient with a pulse width of 0.1 milliseconds is applied to the diode, the diode should be able to dissipate 500 times the power that the diode could dissipate if it were steady state power. If the non-reoccuring

transient has a pulse width of one millisecond, the diode should be able to dissipate fifty times the power it could if the power were steady state, and so forth. When the pulse <u>is</u> <u>re-occuring</u>, the duty cycle must be taken into account. The coefficient of power is no longer the ratio of the thermal time constant to the pulse width when the pulse re-occurs.

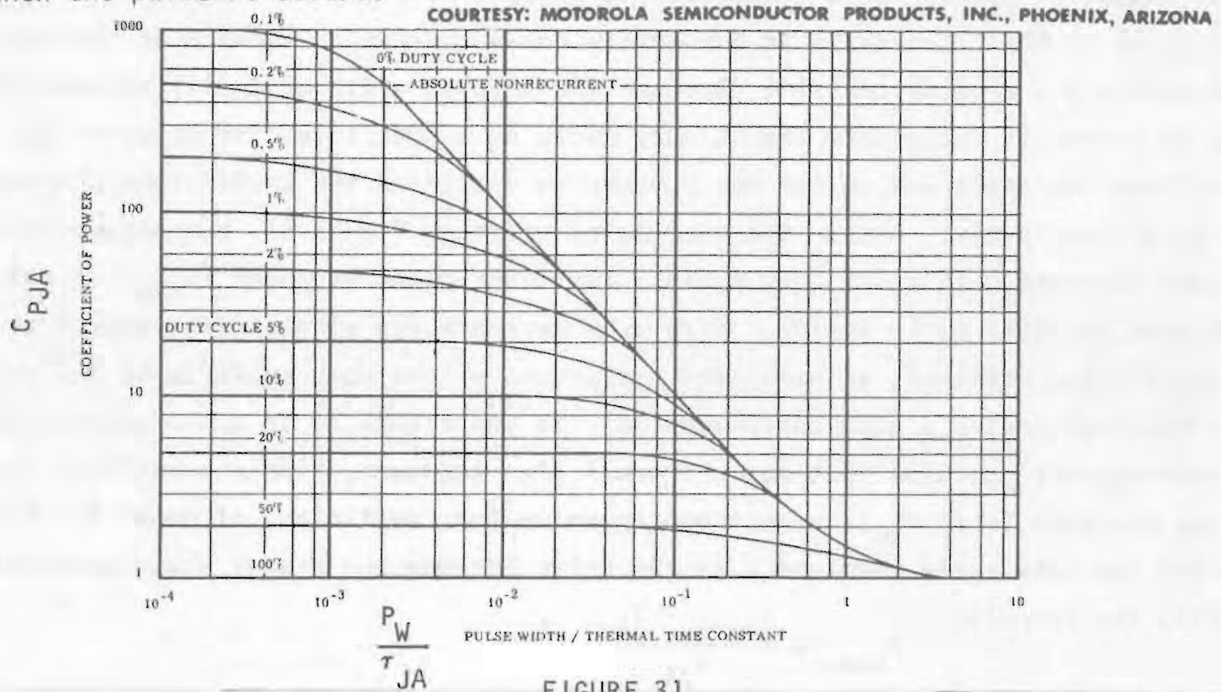


FIGURE 31

Figure 31 may be used to determine the coefficient of power when the duty cycle must be taken into account. The horizontal in Figure 31 is a plot of the ratio of pulse width to the thermal time constant, and the vertical lists the coefficient of power. The running parameter is duty cycle. Note that, with an absolute non-reoccuring pulse, the slope of the line is essentially the ratio of the thermal time constant to the pulse width. In other words, to find the coefficient of power when the duty cycle is zero, simply divide the thermal time constant by the pulse width. Note that when the pulse width is very narrow with respect to the thermal time constant, the coefficient of power is essentially the reciprocal of the duty cycle. A 10% duty cycle gives a coefficient of power of ten. We can say that, if the pulse width is very narrow with respect to the thermal time constant, the 10% duty cycle would allow ten times the peak power than the power that could be dissipated if it were steady state power.

Let's apply Figure 31. Assuming the diode has a thermal time constant of 100 milliseconds, the line marked 1 on the horizontal in Figure 31 indicates the point

at which pulse width and thermal time constant are equal $\left(\frac{\tau_{JA}}{P_W} = 1\right)$. The horizontal designations in Figure 31 are as follows. The line marked $1 = \tau_{JA}$, $10^{-1} = 0.1\ \tau_{JA}$, $10^{-2} = 0.01\ \tau_{JA}$, etc. In other words, if pulse width is $0.01\ \tau_{JA}$, the line marked 10^{-2} would be used. For a 2 millisecond pulse width, we could expect a coefficient of power of 50 if the pulse is absolutely non-reocurring. Looking at the horizontal in Figure 31, finding the line representing a pulse width of 2 milliseconds ($0.02\ \tau_{JA}$), we find that it intersects the 0% duty cycle at a coefficient of power of 50. This verifies the scale and we can not proceed to determine the coefficient of power for a given duty cycle. Assume the same pulse width in Figure 31, but assume that the pulse reoccurs with a 10% duty cycle. The coefficient of power (C_{PJA}) is reduced to 10 when the duty cycle is 10%. With very narrow pulse widths with respect to the thermal time constant, assuming the reciprocal of the duty cycle to be the coefficient of power will give a good approximation. To use Figure 31 in conjunction with any semiconductor junction that has a thermal time constant, simply substitute the junction's time constant in as τ_{JA}, and proceed to solve for coefficient of power for the pulse width and duty cycle involved. We can solve for maximum steady state power dissipation using the formula:

$$P_{Dmax} = \frac{T_{Jmax} - T_A}{\theta_{JA}}$$

We can find the coefficient of power using Figure 31 if there is no steady state power dissipation. Peak power is found by simply taking the product of maximum steady state power and the coefficient of power, or:

$$P_{Pmax} = P_{Dmaxss}\ C_{PJA}$$

Then substituting the formula for maximum steady state power, we get

$$P_{Pmax} = \left(\frac{T_{Jmax} - T_A}{\Theta_{JA}}\right) C_{PJA}$$

where P_{Pmax} is the maximum peak power, T_{Jmax} is the maximum operating junction temperature, T_A is the ambient temperature, Θ_{JA} is the total thermal resistance, and C_{PJA} is the coefficient of power.

If the device is quiescently operating at some steady state power dissipation before the pulse is applied, the steady state power dissipation must be taken into account. The formula for maximum peak power becomes:

$$P_{Pmax} = \left(\frac{T_{Jmax} - T_A - P_{ss}\ \Theta_{JA}}{\Theta_{JA}}\right) C_{PJA}$$

where P_{Pmax} is the maximum peak power, T_{Jmax} is the maximum junction temperature, T_A is the ambient temperature, P_{ss} is the steady state power dissipation, Θ_{JA} is the total thermal resistance, and C_{PJA} is the coefficient of power. Note in the formula that the product of P_{ss} and Θ_{JA} simply gives the rise in junction temperature as a result of steady state power dissipation. Subtracting this from T_{Jmax} takes care of the rise in junction temperature as a result of dissipating steady state power. The total allowable change in junction temperature is then divided by the total thermal resistance. The product of this quantity and the coefficient of power gives the maximum peak power when there is steady state power to take into account. Keep in mind that this is still simply the total allowable change in junction temperature divided by the total thermal resistance and modified by a coefficient of power when dealing with pulses. The formula below illustrates this:

$$P_{Pmax} = \left(\frac{\Delta T_{J\ allowable}}{\Theta_{JA}}\right) C_{PJA}$$

where $\Delta T_{J\ allowable} = T_{Jmax} - T_A - P_{ss}\ \Theta_{JA}$

It is evident that a significant increase in the power dissipation capabilities of a semiconductor device can be accomplished when dealing with pulses and switching if proper precautions are taken.

Let's stop and review for a moment. When dealing with steady state power dissipation, the allowable change in junction temperature divided by the thermal resistance determines the amount of power that may be dissipated. When dealing with pulses or switching applications, the thermal time constant of the semiconductor device must be taken into account. Since it will take five thermal time constants for the heating effect at the junction to take place, greater power can be dissipated on peaks than might be dissipated in a steady state condition. This assumes that the junction does not get a chance to heat to its final value. If the pulse is absolutely non-reoccuring, the coefficient of power is essentially the thermal time constant divided by the pulse width. When the pulses do re-occur, the junction does not get a chance to completely cool before another pulse is applied, and the duty cycle must be taken into account. To take into account the duty cycle, we plot pulse widths as a fraction of thermal time constant against coefficient of power and let duty cycle be the running parameter. We can then determine the coefficient of power for a switching application and determine the peak power that

the diode can dissipate for a given pulse width and duty cycle. If steady state power must be taken into account when dealing with pulses, we simply calculate the heating effect due to stady state power dissipation and subtract this from the allowable change in junction temperature. In the formula for peak power when steady state power is also being dissipated, the heating effect due to steady state power dissipation is simply subtracted from the total allowable change in junction temperature. We should also keep in mind that these same approaches can be applied to transistors and other semiconductors devices and are not just limited to diodes or diode devices.

QUESTIONS FOR SECTION 1

Study each question carefully, including any diagrams provided, and select the most correct answer.

1. Electrons exist in energy levels about the nucleus of an atom. An electron that gives up some of its energy will move to an orbit _________ to/from the nucleus.

   a. farther
   b. 90 degrees
   c. 45 degrees
   d. closer
   e. at right angles

2. Atoms in structures are bonded together in the _________ _________ band.

   a. conduction energy
   b. valence energy
   c. nucleus energy
   d. kinetic energy
   e. forbidden energy

3. Atoms in structures can be excited by heat energy. This moves electrons from the _________ band to the _________ band.

   a. valence, conduction
   b. valence, forbidden
   c. conduction, valence
   d. conduction, forbidden
   e. forbidden, conduction

4. The band gap between the valence and conduction band is _________ in conductors and _________ in insulators.

   a. wide, narrow
   b. filled, empty
   c. narrow, wide
   d. empty, narrow
   e. filled, narrow

5. Atoms bonded tightly in a pure covalent bond structure provide carriers of current at room temperature due to _________ energy forming _________ _________ _________.

   a. electrical
      high field emission
   b. kinetic
      accelerated proton action
   c. heat
      fermi tunneling electrons
   d. heat
      hole-electron pairs
   e. electrical
      high potential deceleration

6. Heat energy lowers the resistance of a pure covalent bond structure by providing __________ __________ __________ as current carriers.

   a. holes and electrons
   b. electrons, not holes
   c. holes, not electrons
   d. free excited protons
   e. free light photons

7. Electrons moving between energy bands in a structure must take on or give up energy. The electron gives up its energy in the form of __________.

   a. Kinetic energy
   b. heat
   c. light
   d. both b and c
   e. both a and b

8. Doping intrinsic semiconductor with donor impurities results in forming __________ type semiconductor.

   a. P
   b. N
   c. P and N
   d. junction
   e. intrinsic

9. Doping intrinsic semiconductor with acceptor impurities provides __________ as majority current carriers.

   a. electrons
   b. photons
   c. phonons
   d. holes
   e. protons

10. At room temperature (25°C), there will be __________ as current carriers in N type semiconductor.

    a. electrons
    b. holes
    c. phonons
    d. photons
    e. both a and b

11. Doping a pure semiconductor with donor impurities provides current carriers as a result of the doping process that will travel in the __________ energy band of the material.

    a. valence
    b. equilibrium
    c. fermi
    d. conduction
    e. lowest

12. An increase in temperature will provide more ________ as current carriers in P type semiconductors.

    a. electrons
    b. holes
    c. both a and b
    d. both a and b but more of b
    e. both a and b but more of a

13. The N and P properties of doped semiconductors are not lost due to forcing electrons in or out of the material since the dopent atoms are immobile and become ________.

    a. hole-electron pairs.
    b. ions
    c. thermally excited.
    d. charged
    e. either b or d

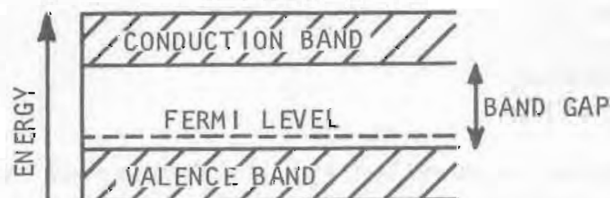
14. Forcing electrons into P material without a path for them to leave results in forming ________.

    a. positive ions
    b. negative ions
    c. neutral atoms
    d. no charge
    e. a state of equilibrium

15. The fermi level is the ________ electron probability level and exists ________ the valence and conduction bands in an intrinsic semiconductor.

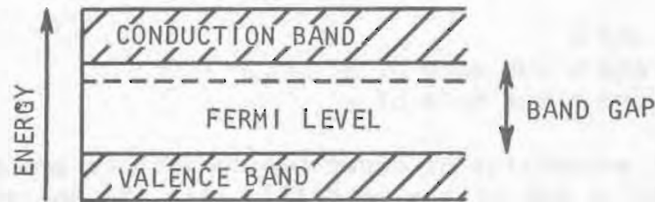
    a. 100%, near the bottom of the band gap between
    b. 0%, midway in the band gap between
    c. 50%, near the top of the band gap between
    d. 50%, midway in the bandgap between
    e. 100%, near the top of the band gap between

16. The energy band diagram shown indicates ________ type semiconductor.



    a. P
    b. intrinsic
    c. either a or b
    d. N
    e. either a or d

17. The energy band diagram shown is a semiconductor material which has had ________________ inpurities added.

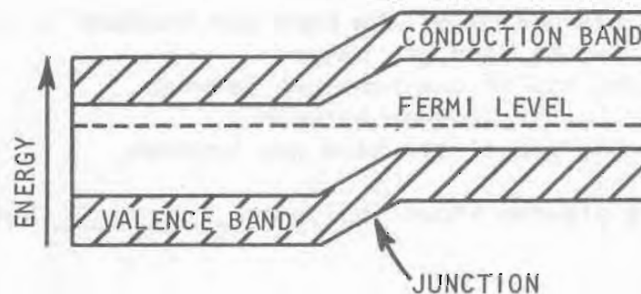


a. acceptor
b. donor
c. N type
d. P type
e. either a or d
f. either b or c

18. When a PN junction is formed, recombination of holes and electrons takes place until __________ is reached (with no external energy applied).

a. tunneling
b. avalanche
c. equilibrium
d. zener breakdown
e. carrier lifetime

19. The diagram shown indicates a PN junction with __________ bias applied.

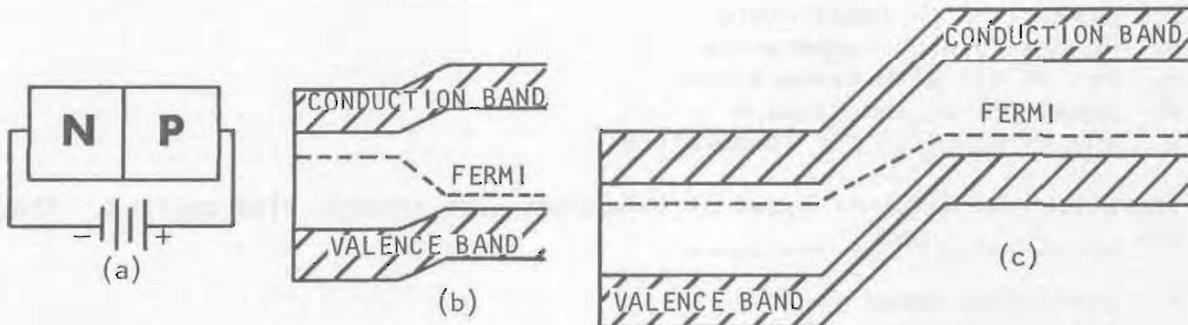


a. 0
b. forward
c. reverse
d. high reverse
e. either c or d

20. In the diagram in question 19, the P material is on the ________ and the N material is on the _________ of the junction shown.

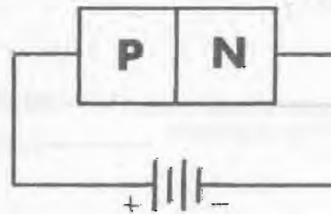
a. left, right
b. top, bottom
c. bottom, top
d. right, left

21. The battery in diagram (a) would result in ____________ carriers crossing the junction and the energy band diagram in diagram _________.



a. majority, (c)
b. minority, (b)
c. minority, (c)
d. majority, (b)

22. The external current due to the application of the battery of the polarity shown in the diagram (assuming insufficient voltage to cause breakdown) are the result of _____________ carriers crossing the junction.



a. majority
b. minority

23. The current that flows when reverse bias is applied is primarily carried by carriers which are present as a result of _____________ (after any stored charge has been removed).

a. the formation of hole-electron pairs
b. doping
c. heat energy
d. added impurities
e. either a or c
f. either b or d

24. Holes from the P side and electrons from the N side cross the junction when ____________ bias is applied.

a. forward
b. reverse
c. either a or b

25. The resistance of a reverse biased diode that is not in a breakdown condition, varies ______________.

a. directly with temperature
b. inversely with temperature
c. not at all with temperature
d. sometimes a, sometimes b
e. the surrounding air temperature

26. There are two distinct types of breakdown with reverse bias applied. They are __________ and __________.

a. avalanche, zener
b. zener, tunneling
c. zener, thermal
d. tunneling, avalanche
e. either a or d

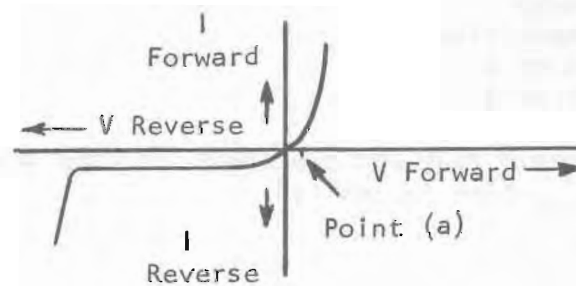
27. The voltage across the terminals of a diode in avalanche breakdown varies __________ as temperature varies and the voltage across the terminals of a diode in __________ breakdown varies __________ as temperature varies.

a. directly, zener, directly
b. inversely, tunneling, directly
c. directly, zener, inversely
d. inversely, zener, directly
e. either b or d

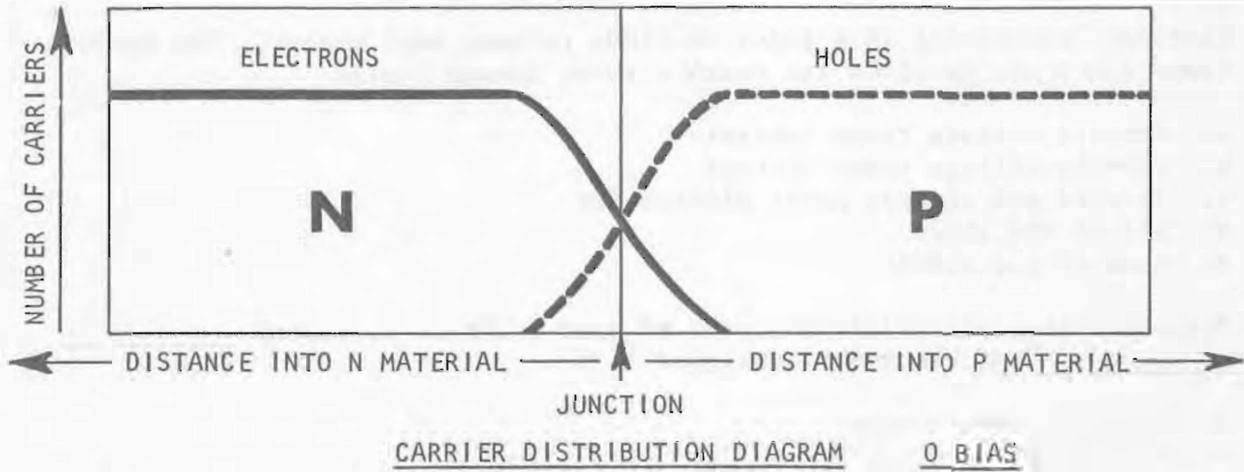
28. An increase in temperature __________ the voltage level at which tunneling occurs and an increase in temperature __________ the voltage at which avalanche occurs.

a. reduces, increases
b. increases, increases
c. reduces, reduces
d. increases, reduces

29. The curve shown is a voltage versus current curve for a semiconductor diode. Point (a) is about ________ volts for a silicon diode and about ________ volts for a germanium diode.

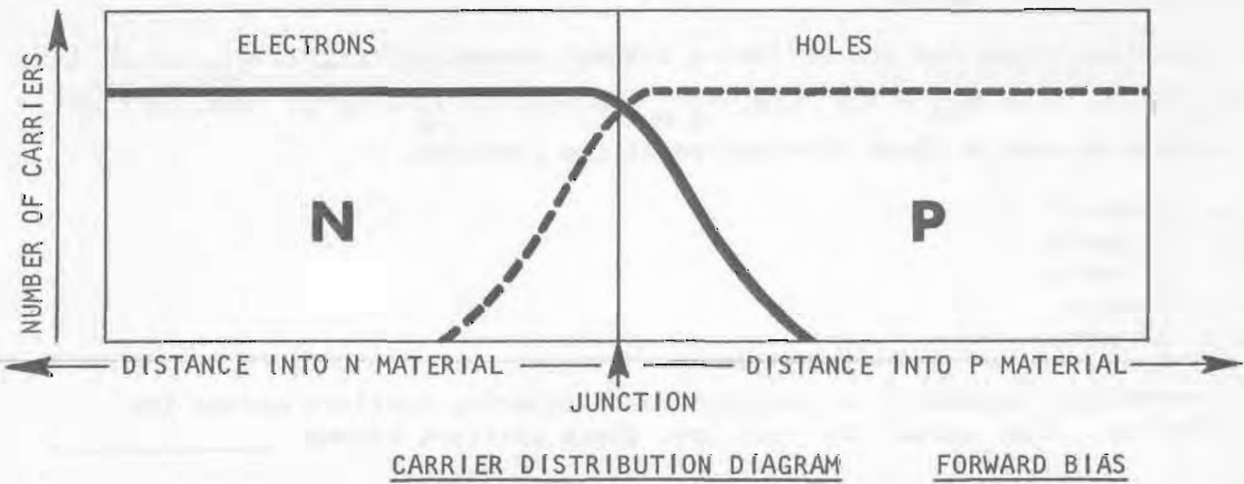


a. 7, 3
b. 0.03, 0.07
c. 3, 7
d. 0.7 0.3
e. 0.3, 0.7



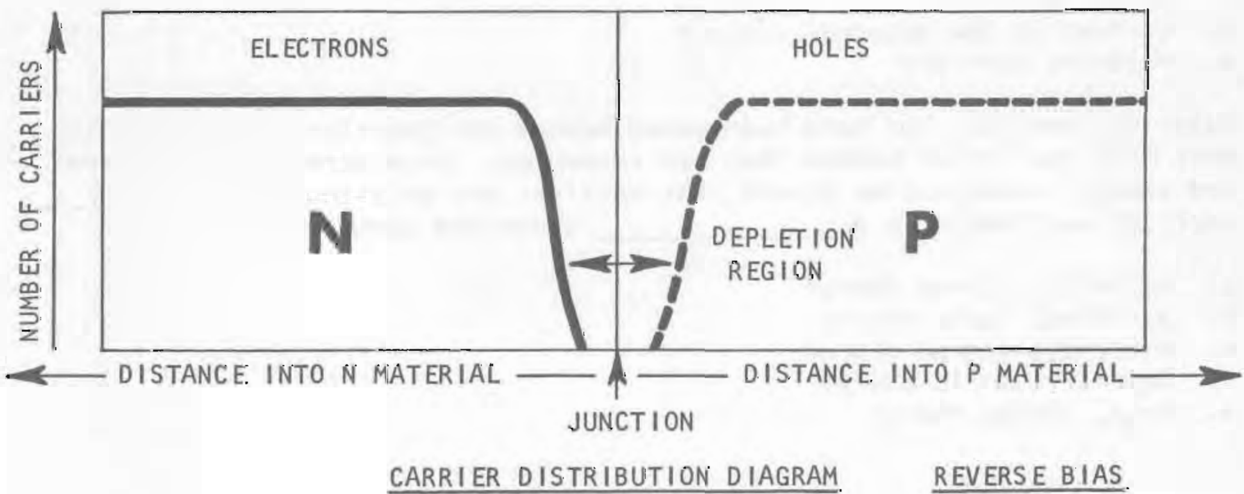
CARRIER DISTRIBUTION DIAGRAM   0 BIAS

FIGURE 31A



CARRIER DISTRIBUTION DIAGRAM   FORWARD BIAS

FIGURE 31B



CARRIER DISTRIBUTION DIAGRAM   REVERSE BIAS

FIGURE 31C
TEST

30. Carriers recombining in a junction diode release heat energy. The maximum temperature the junction can reach without damage limits ________.

a. forward voltage times current
b. reverse voltage times current
c. forward and reverse power dissipation
d. all of the above
e. none of the above

31. The opposition offered in the path of heat transfer is termed ________ ________ and the symbol assigned is Θ.

a. radiation resistance
b. heat sink
c. radiation conductance
d. thermal resistance
e. thermal radiation

32. A junction diode has the following thermal characteristics: $\Theta_{JC} = 0.8° C/W$, $\Theta_{CS} = 0.6° C/W$, $\Theta_{SA} = 1.6° C/W$, $T_{J\ max} = 150° C$, $T_A = 60°C$. What is the maximum allowable power dissipation at the junction.

a. 3 watts
b. 10 watts
c. 90 watts
d. 9 watts
e. 30 watts

33. Forward bias applied to a junction moves majority carriers across the junction. Once across the junction, these carriers become ________.

a. recombined
b. minority carriers
c. holes
d. current in the external circuit
e. majority carriers

34. Majority carriers that have been moved beyond the junction by forward bias must find impurities before they can recombine. Once across the junction and before recombination occurs, the carriers are existing as ________ carriers and result in a ________ about the junction.

a. majority, stored charge
b. electron, ionic charge
c. minority, stored charge
d. majority, ionic charge
e. hole, stored charge

35. The amount of stored charge about the junction is determined by the amount of forward current and the ________.

    a. minority carrier lifetime
    b. majority carrier lifetime
    c. junction width time constant
    d. thermal time constant
    e. minority carrier time constant.

36. In figure 31A, the lines indicating holes in the P side and electrons in the N side extend into the opposite sides of the material indicating the existence of ________ in the two sides of the junction.

    a. current flow
    b. minority carriers
    c. majority carriers
    d. a thermal resistance
    e. a thermal time constant

37. The solid and dashed lines in figure 31B show that more holes and electrons have been moved across the junction with forward bias applied. These carriers are existing as ________ carriers, and represent a ________ about the junction.

    a. majority, ionic charge
    b. majority, stored charge
    c. minority, thermal charge
    d. minority, stored charge
    e. majority, depletion region

38. In figure 31C, the area about the junction is depleted of carriers as a result of applied reverse bias. To change from the condition in figure 31B to the condition in 31C, the ________ must be removed.

    a. donor impurities
    b. bias source
    c. stored charge
    d. impurities near the junction
    e. ions in the two sides

39. Impurities may be added near the junction to reduce the ________ and the ________ for a given forward current.

    a. minority carrier lifetime, stored charge
    b. majority carrier transit time, junction voltage
    c. junction voltage, power dissipation
    d. junction voltage, junction temperature

40. τ_q is given in ________ per ________ and can be used to calculate ________.

    a. pico-farads, volt, junction capacity
    b. pico-coulombs, milli-ampere, stored charge
    c. nano-seconds, centimeter, risetime
    d. pico-seconds, μinch, recovery time

TABLE 1

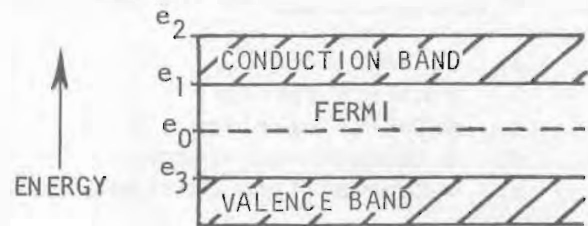
HOLE ELECTRON PAIRS IN INTRINSIC SEMICONDUCTORS

Considering an idealized intrinsic semiconductor with no added impurities,

$$n_i = p_i$$

where n_i = density of free electrons in intrinsic material and p_i = density of free holes in intrinsic material

$$n_i = \int_{e_1}^{e_2} \frac{[Ds] \quad de}{\epsilon^{\frac{(e-e_o)}{KT}} + 1}$$



Where:

e_1 = energy at the bottom of the conduction band

e = energy of an electron in the material

e_o = fermi energy level

e_2 = energy at the top of the conduction band

[Ds] = electron density states in conduction band

K = Boltzmann's constant = 1.38×10^{-23} joules/°K

T = temperature

ϵ = 2.718

Evaluating the integral with assumptions:

$$n_i = \left(\frac{2\pi m_n KT}{h^2}\right)^{3/2} \epsilon^{-\frac{(e_i - e_o)}{KT}}$$

where: h = Planck's constant = 6.6254×10^{-27} erg/sec

m_n = mass of the electron

Since the $n_i = p_i$ the number of holes is given by:

$$p_i = \left(\frac{2\pi m_p KT}{h^2}\right)^{3/2} \epsilon^{-\frac{(e_3 - e_o)}{KT}}$$

where: e_3 = the energy at the top of the valence band

m_p = mass of the holes

h = Planke's constant

TABLE 1 (continued)

Since in pure semiconductor, the Fermi level (e_o) is midway between the top of the valence band and the bottom of the conduction band (e_3 and e_1), the product of $n_i\ p_i$ is related exponentially to the band gap width between the valence and conduction band as shown by:

$$n_i\ p_i = \left(\frac{2\pi KT}{h^2}\right)^3 \ (m_n\ m_p)^{3/2} \ \epsilon^{\frac{-(e_1 - e_3)}{KT}}$$

Lumping the constants for a given material

$$n_i\ p_i = n_i^{\ 2} = C^2\ T^3\ \epsilon^{\frac{-(e_1 - e_3)}{KT}}$$

$$n_i = p_i = CT^{\ 3/2}\ \epsilon^{\frac{-(e_1 - e_3)}{2\ KT}}$$

Where: C = the constant of proportionality for the given material.

At room temperature (25°C):

$p_i = n_i = 2.5 \times 10^{13}\ cm^3$ (for germanium)

$p_i = n_i = 6.8 \times 10^{10}\ cm^3$ (for silicon)

$n_i\ p_i = n_i^{\ 2}$

$n_i^{\ 2} = 6.25 \times 10^{26}\ cm^6$ (for germanium)

$n_i^{\ 2} = 4.62 \times 10^{21}\ cm^6$ (for silicon)

Typical constants of proportionality (C in the formulas) are:

C (silicon) $\approx 4.74 \times 10^{22}$

C (germanium) $\approx 1.76 \times 10^{22}$

The number of hole-electron pairs at room temperature is typically:

Germanium = $2.5 \times 10^{13}\ cm^3$

Silicon = $6.8 \times 10^{10}\ cm^3$

TABLE 1 (continued)

The number of atoms in a cubic centimeter is:

Germanium = 4.42×10^{22} cm^3

Silicon = 4.99×10^{22} cm^3

From this it can be seen that germanium at room temperature has less than one electron in the conduction band per billion germanium atoms and that there is a factor of 10^3 difference in the free electrons in intrinsic germanium as opposed to silicon at room temperature.

TABLE 2

CHARACTERISTIC	GERMANIUM (Ge)	SILICON (Si)	GALLIUM ARSENIDE(GaAs)
Melting Point in °C	937	1415	1240
Density in grams/CM^3 at 25°C	5.32	2.33	5.31
Atomic Weight	72.6	28.09	144.6
Dielectric Constant (Air = 1)	16	11.8	11.1
Intrinsic Resistivity (25°C) in Ohms/CM^3	46	2.3×10^5	3.7×10^8
Band Gap Energy (e_1-e_3) in Electron Volts	0.67	1.106	1.4
Lattice Constant in Angstrom Units	5.657	5.431	5.654
Number of Atoms/CM^3	4.42×10^{22}	4.99×10^{22}	4.43×10^{22}
Electron Lattice (μn) Mobility CM^2/ volt-sec (25°C)	3900	1350	10,000
Hole Mobility (μp) CM^2/volt-sec (25°C)	1900	480	450
Electron Diffusion (Dn) Constant, CM^2/sec (25°C)	100	38	310
Hole Diffusion (Dp) Constant, CM^2/sec (25°C)	49	13	11.5
Thermal Conductivity in Watt Units	.63	.84	.52
Thermal Coefficient of Expansion/°C	6.1×10^{-6}	4.2×10^{-6}	5.9×10^{-6}

TABLE 3

The Fermi level or Fermi energy is defined here as the 1/2 or 50% carrier occupation probability level. Thus defined we can write an equation for the probability that a given energy level is occupied by a carrier as:

$$f_{(e)} = \frac{1}{\frac{(e - e_o)}{KT} + 1}$$

where $f_{(e)}$ = electron probability function

e = energy of the electron in the material

e_o = fermi energy level

K = Boltzman's Constant = 1.38×10^{-23} joules/°K

T = Temperature - °K

The placement of the Fermi level in an energy band diagram is found by taking the product of the probability function and the density of state function. If the density of electrons in the conduction band is equal to the density of holes in the valence band, the Fermi level exists midway in the band gap between the top of the valence band and the bottom of the conduction band. This is the case with intrinsic semiconductors.

$$\text{(intrinsic) } e_o = \frac{(e_1 - e_3)}{2} + e_3$$

where e_1 = energy at the bottom of the conduction band

e_3 = energy at the top of the valence band

e_o = Fermi level

If the number of electrons in the conduction band is greater than the number of holes in the valence band, the Fermi level is somewhat above the center of the band gap $(e_1 - e_3)$. This is the case in N type semiconductors.

If the number of holes in the valence band is greater than the number of electrons in the conduction band, the Fermi level exists somewhat below the center of the band gap $(e_1 - e_3)$. This is the case in P type semiconductors.

Taking the product of the density function and the Fermi probability function, the

TABLE 3 (Continued)

number of electrons in the conduction band can be calculated by:

$$n_i = \int_{e_1}^{e_2} \frac{[D_s] \; d_e}{\epsilon^{\frac{(e - e_o)}{KT}} + 1}$$

where n_i = number of electrons in the conduction band

e_1 = energy at bottom of conduction band

e_2 = energy at top of conduction band

e_3 = energy at top of valence band

e = energy of electron in the material

$[D_s]$ = density of states in the conduction band

K = Boltzman's Constant

T = Temperature

The integral of this is evaluated in Table 1.

TABLE 4

The current in semiconductors can be expressed as:

Current = number of charge carriers x charge per carrier x average velocity per carrier

For electrons as carriers:

$$I = n_i \, q \, \upsilon_n$$

and for the holes as carriers:

$$I = p_i \, q \, \upsilon_p$$

where I = current

n_i = number of electrons

q = electronic charge = 1.6×10^{-19} coulomb

p_i = number of holes

υ_n = average velocity of the electron in the material

υ_p = average velocity of the hole in the material

and where both holes and electrons are present:

$$I = n_i \, q \, \upsilon_n + p_i \, q \, \upsilon_p$$

The average velocities of electrons and holes can be expressed as the average velocity per unit of electric field.

$$\upsilon_n = E_{\mu n}$$

$$\upsilon_p = E_{\mu p}$$

where υ_n = average velocity of electron

υ_p = average velocity of hole

E = applied electric field

μn = average velocity of electron per unit electric field

μp = average velocity of hole per unit electric field

and the equation for current can be:

$$I = (n_i \, q \, \mu_n + p_i \, q \, \mu_p)E$$

TABLE 4 (Continued)

Conductivity is defined as:

$$g = \frac{I}{E}$$

therefore

$$g = \frac{(n_i\, q\, \mu_n + p_i\, q\, \mu_p)\, E}{E} = n_i\, q\, \mu_n + p_i\, q\, \mu_p$$

μ_n and μ_p are referred to as the mobility constants and have dimensions of:

Silicon	μ_n = 1,350 CM^2/volt-sec
	μ_p = 480 CM^2/volt-sec
Germanium	μ_n = 3,900 CM^2/volt-sec
	μ_p = 1,900 CM^2/volt-sec
Gallium Arsenide	μ_n = 10,000 CM^2/volt-sec
	μ_p = 450 CM^2/volt-sec

TABLE 5

ELECTRONS AND HOLES IN N AND P TYPE SEMICONDUCTORS

From Table 1, the formula for the density of electrons and holes in intrinsic semiconductor was reduced to:

$$p_i = n_i = CT^{3/2} \epsilon^{\frac{-(e_1 - e_3)}{2KT}}$$

and the product of the two:

$$n_i p_i = n_i^2 = C^2 T^3 \epsilon^{\frac{-(e_1 - e_3)}{KT}}$$

The product of n_i and p_i is dependent on the width of the band gap and the temperature, but not the placement of the Fermi level.

The recombination rate is given as αnp

where α = proportionality factor for the particular material
n = free electrons (n_i in intrinsic)
p = free holes (p_i in intrinsic)

There is the thermal generation of hole-electron pairs and the recombination of hole-electron pairs. Stating the density of holes or electrons is the same as stating the recombination rate. At any given temperature, there is a density of holes and electrons at which the recombination rate (αnp) equals the thermal-generation rate (GT). This density is the thermal equilibrium density, and for thermal equilibrium in intrinsic semiconductors:

$$GT = \alpha n_i p_i = \alpha n_i^2$$

$$GT = \alpha n_i^2$$

$$n_i^2 = \frac{GT}{\alpha}$$

$$n_i = \sqrt{\frac{GT}{\alpha}}$$

$$n_i = CT^{3/2} \epsilon^{\frac{-(q E_g)}{2KT}}$$

where q = charge on the electron = 1.6×10^{-19} coulomb
E_g = band-gap voltage $(e_1 - e_3)$

TABLE 5 (Continued)

In doped semiconductor, the recombination rate becomes: (N material as an example)

$$\alpha np = \alpha N_n p$$

where N_n = number of donor impurity atoms.

The density of donor atoms is such that the density of free electrons is approximately determined by the density of donor atoms and at thermal equilibrium.

$$GT = \alpha n_i^2 = \alpha N_n p$$

The recombination process dominates the generation process and the density of free holes is reduced while the density of free electrons is increased so that:

$$pn = n_i^2 = n_n p_n = C^2 T^3 \epsilon \frac{-(q E_g)}{KT}$$

where n_n = free electrons in N type semiconductor
p_n = free holes in N type semiconductor

The number of free electrons in N type material is essentially equal to the number of donor atoms added in the doping process, while the number of free holes is reduced over intrinsic at a given temperature:

$$n_n = N_n \qquad p_n = \frac{n_i^2}{N_n}$$

The same can be said for P type semiconductor, except that holes are in the majority and the number of free holes is essentially equal to the number of acceptor impurities added or:

$$p_p = N_p \qquad n_p = \frac{n_i^2}{N_p}$$

where p_p = free holes in P semiconductor
n_p = free electrons in P semiconductor
N_p = number of acceptor impurity atoms added in the doping process

TABLE 6

For a given piece of semiconductor, parameters may be derived based on the constants of the type of material used in the given piece of semiconductor.

From Table 4: (holes) $I = p\, q\, \mu_p\, E$

where p = free holes

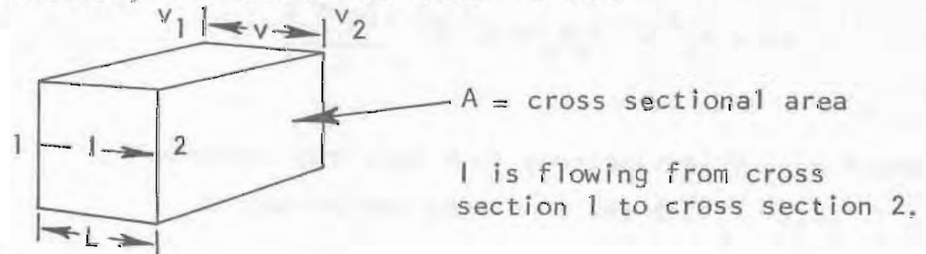
q = electronic charge = 1.6×10^{-19} coulombs

μ_p = mobility constant, holes

E = applied electronic field

I = current density

For a given piece of semiconductor, this can be rewritten to read:



$$I_{p(1\to2)} = \frac{p_1 + p_2}{2}\,\frac{v_1 - v_2}{L}\,A\,q\,\mu_p$$

where I_p = hole current from point 1 to point 2

p_1 = hole density at 1

p_2 = hole density at 2

$v_1 - v_2$ = electric field, 1 to 2

A = cross sectional area at right angles to direction of current

q = electronic charge = 1.6×10^{-19} coulombs

μ_p = mobility constant, holes

L = length, 1 to 2

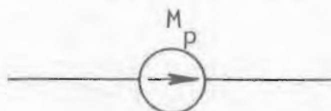
This can be rewritten:

$$I_{p(1\to2)} = \frac{p_1 + p_2}{2}\,(v_1 - v_2)\,M_p$$

which makes $M_p = \dfrac{A\,q\,\mu_p}{L}$

TABLE 6 (Continued)

and M_p is the parameter "hole mobilance" and is given the symbol



Electron mobilance can be formulated in the same manner and found to be:

$$M_n = \frac{A\ q\ \mu_n}{L}$$

and $$I_{n\ (1 \rightarrow 2)} = \frac{n_1 + n_2}{2}\ (v_1 - v_2)\ M_n$$

where: n_1 = the density of electrons at point 1

n_2 = the density of electrons at point 2

When carriers are in the minority, their transport is not mobilance but the carriers move by diffusion. The formula for current can be written:

$$I_{p_d} = \frac{p_1 - p_2}{L}\ A\ q\ D_p$$

where: I_{p_d} = hole diffusion current

p_1 = density of holes at point 1

p_2 = density of holes at point 2

This can be written:

$$I_{p_d} = (p_1 - p_2)\ H_p$$

where: H_p = the parameter hole diffusance

and it is evident that $H_p = \frac{A\ q\ D_p}{L}$, and given the symbol

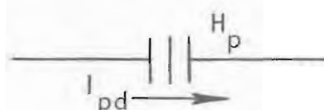


TABLE 6 (Continued)

The parameter, electron diffusance can be formulated in the same way and found to be:

$$H_n = \frac{A\ q\ D_n}{L}$$

and the electron diffusion current can be found by:

$$I_{n_d} = (n_1 - n_2)\ H_n$$

where: I_{n_d} = electron diffusion current

n_1 = density of electrons at point 1

n_2 = density of electrons at point 2

These parameters will be used in later tables. The density of carriers at 2 (or any point if the length is substituted for L) can be found by the formula:

$$n_2 = n_1\ \epsilon^{\alpha(L)}$$

where: n_2 = carrier density at 2

n_1 = carrier density at 1

L = length

$$\alpha = \frac{-\mu E\tau + \left[(\mu E\tau)^2 + 4D\tau\right]^{1/2}}{2D\tau}$$

where: E = electric field

μ = ambipolar mobility constant

D = ambipolar diffusion constant

τ = carrier lifetime

solving for μ and D:

$$\mu = \frac{p - n}{\frac{n}{\mu_p} + \frac{p}{\mu_n}}$$

$$D = \frac{p + n}{\frac{n}{D_p} + \frac{p}{D_n}}$$

TABLE 6 (Continued)

and by Einstein's relation:

$$D_n = \frac{\mu_n KT}{q}$$

and therefore:

$$\frac{D_n}{\mu_n} = \frac{KT}{q} \qquad \frac{\mu_n}{D_n} = \frac{q}{KT}$$

$$\frac{D_p}{\mu_p} = \frac{KT}{q} \qquad \frac{\mu_p}{D_p} = \frac{q}{KT}$$

at 300°K (room temperature):

$$\frac{KT}{q} = 26\text{mv}$$

$$\frac{q}{KT} = 4 \times 10^{-2}$$

TABLE 7

POTENTIAL BARRIER IN PN JUNCTIONS

The potential on the P side of a PN junction is given by the formula:

$$\Delta_{pv} = -\frac{KT}{q} \ln\left(\frac{N_p}{P_i}\right)$$

where: Δ_{pv} = potential on the P side

N_p = number of acceptor impurity atoms added in the doping process

P_i = intrinsic carrier density

(Note: We have previously established that the number of free carriers is $\approx$ to the number of added impurity atoms.)

The potential on the N side is given as:

$$\Delta_{nv} = \frac{KT}{q} \ln\left(\frac{N_n}{n_i}\right)$$

where: Δ_{nv} = potential on the N side

N_n = number of donor impurity atoms

n_i = intrinsic carrier density

The potential difference or built-in voltage (Δ_v) is:

$$\Delta_v = \Delta_{nv} - \Delta_{pv}$$

and reduces to:

$$\Delta_v = \frac{KT}{q} \ln\left(\frac{N_n N_p}{n_i^2}\right)$$

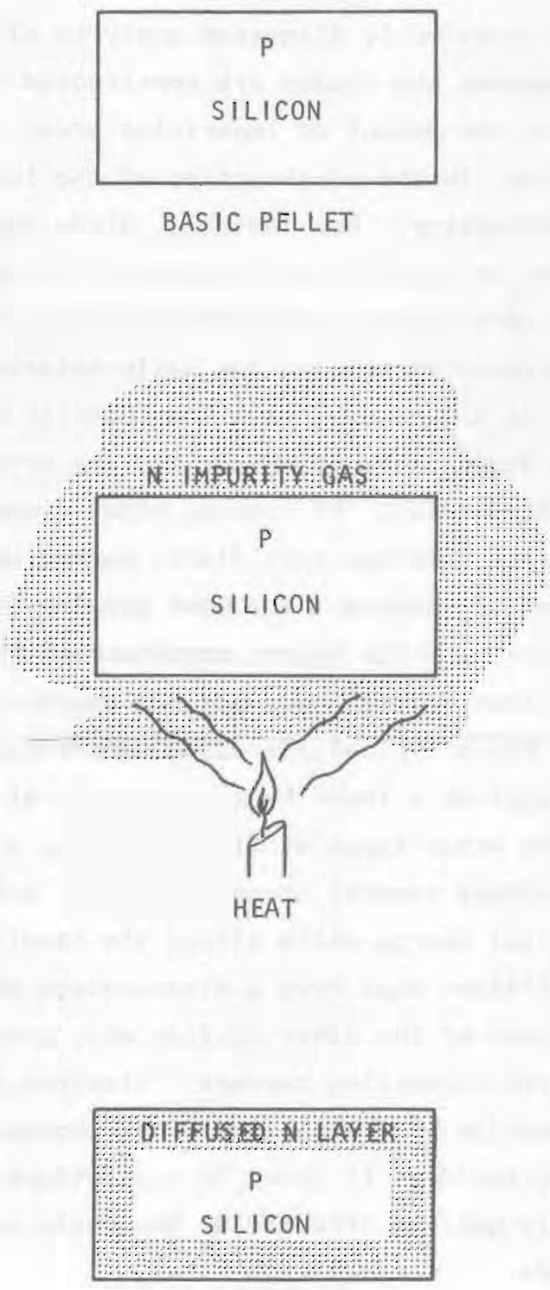


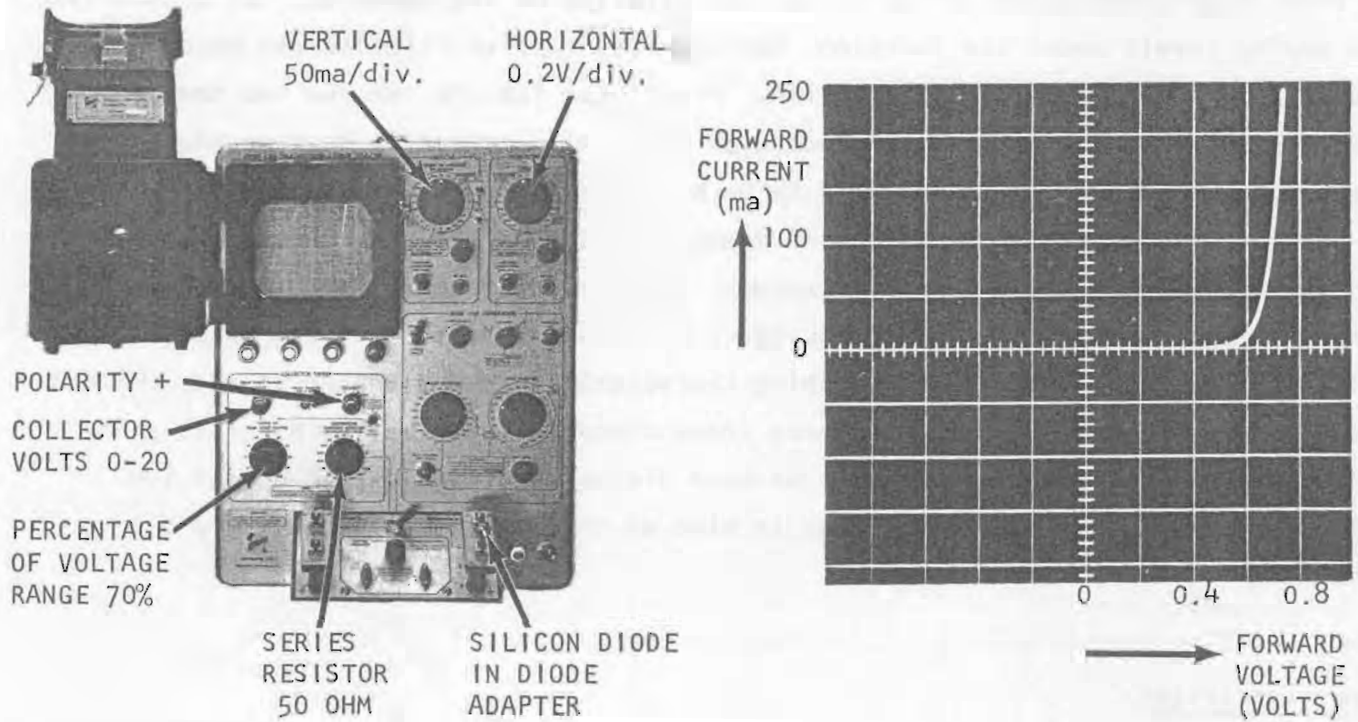
FIGURE 1-2

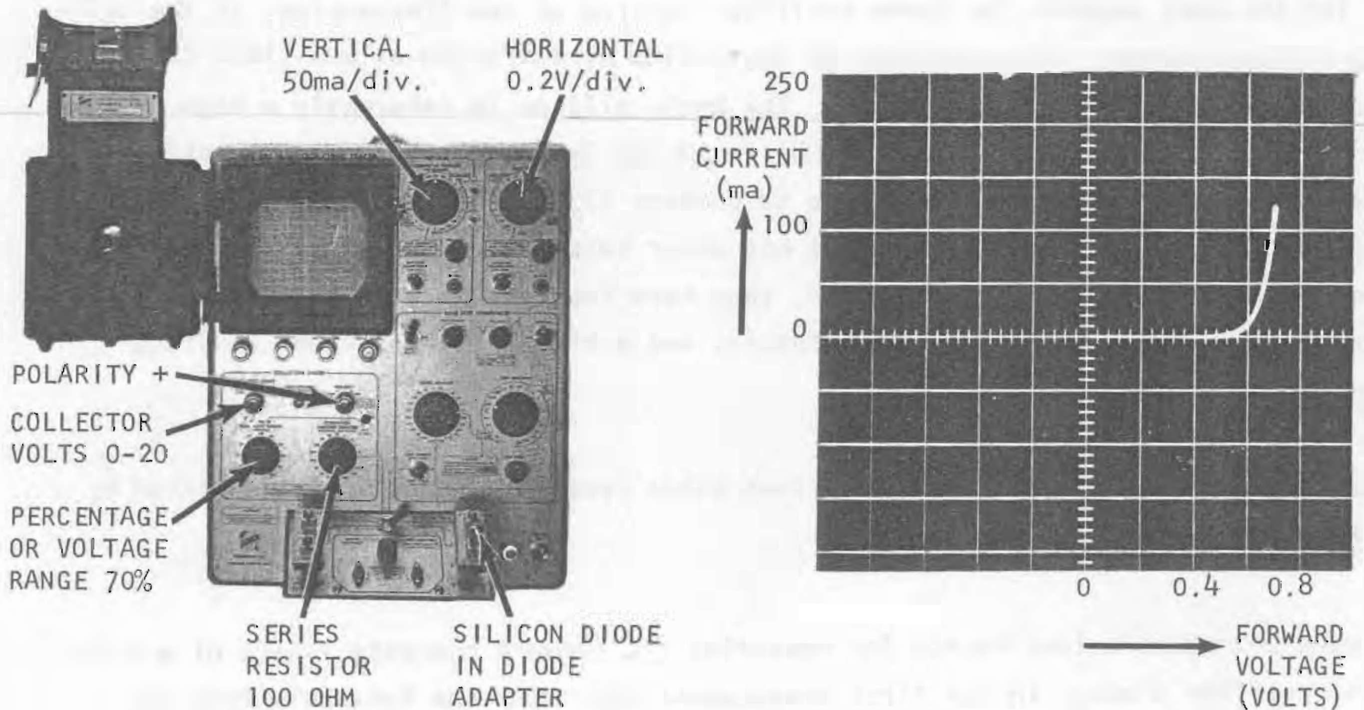
SECTION TWO

DIODE DEVICES

Although the basic concepts previously discussed apply to all diodes and diode devices, for specific applications the diodes are constructed in different fashions. The differences might lie in the amount of impurities added in the doping process, in the geometry of the device, in the construction of the junction itself, or the basic material used in construction. For instance, diode devices used for high power rectifier applications are now primarily made of silicon as a basic material with the dopent impurities added by the diffusion process. The diffusion process of adding impurities is the process of heating the basic material while surrounding it with the doping impurities in a gaseous form. The impurity atoms enter the basic material when it reaches a temperature at which they can enter the structure. This process is illustrated in Figure 1-2. Of course, other diode devices are still in use as high power rectifiers. Selenium rectifiers, germanium diodes, and some copper oxide devices are in use; however, the most popular of course is the silicon diode. The silicon diode can tolerate higher temperatures than the other devices. Semiconductor diodes other than silicon have maximum operating temperatures of about 100°C, while silicon has a typical operating temperature of 150°C. Silicon has the other added advantages of a lower reverse current at a given temperature and applied voltage than the other types of diode devices, a much higher reverse breakdown voltage (up to perhaps several thousand volts), and the ability to handle a higher density of electrical charge which allows the handling of a greater magnitude of forward current. Silicon does have a disadvantage of having a higher conducting voltage drop than some of the other devices and, therefore, a little higher power dissipation for a given conducting current. Electron mobility and hole mobility are faster in germanium than in silicon. Of course, the fastest mobility of all to date is gallium arsenide. If speed is a consideration, it might warrant use of germanium or possibly gallium arsenide as the basic material used to construct a semiconductor diode.

If the device is to be designed for fast switching applications, speed, a small stored charge, and small junction capacity will be three of the predominate considerations. The amount of stored charge is governed by the number of minority carriers diffusing and seeking to recombine when the diode is forward biased. This





TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER MOD 122C WITH DIODE ADAPTER AND TYPE C-12 OSCILLOSCOPE CAMERA

FIGURE 2-2

in turn is governed by the minority carrier lifetime of the material. By controlling the doping levels about the junction, the minority carrier lifetime can be changed and the stored charge can be controlled. From these factors, we can see that the approach to construction of a semiconductor diode will primarily be determined by the application to which it is going to be put. If the diode is to be designed for 60 cycle rectification, a high reverse breakdown and a low reverse leakage current might be two of the primary considerations, along with the ability to handle large amounts of power. If the diode is going to be used for fast switching applications, a low stored charge and a fast switching characteristic would be the primary aim and the device would be designed to enhance these characteristics. The point is that all of the basic considerations that we have discussed will apply to all of the diode devices and this should be kept in mind as the individual devices are discussed.

DIODE RECTIFIERS:

By far the most popular for diode rectifier service at low frequencies, is the silicon diffused diode. The addition of impurities by diffusion allows close control of the diode's characteristics. The basic silicon is inherently a high resistance semiconductor material and the diffusion of impurities to form a junction results in a junction that will operate to perhaps 175°C. Silicon diodes can be designed so that reverse breakdown does not occur below several hundred, or even a thousand volts. Properly constructed, they have low reverse leakage currents or saturation currents at a given temperature, and a high forward current handling capability.

Chart 2A lists some of the more important diode rectifier specifications listed by the manufacturer.

Figure 2-2 shows a test set-up for measuring the forward characteristics of a silicon rectifier diode. In the first measurement made with the Tektronix Type 575 Transistor Curve Tracer, the series resistance is set at 50Ω with a total applied voltage of approximately 70% of 20V, or about 14V. A little less than 1V is across the diode rectifier, and the remaining voltage is dropped across the series resistance. In this case, the current is limited to slightly less than 250ma.

DIODE RECTIFIER SPECIFICATIONS

Peak Inverse Voltage (PIV):

Maximum reverse bias that can be placed across a diode without the diode breaking down. For a sine wave input, using a resistive or inductive load, the PIV should be greater than the peak input voltage. Using a capacitive load, the PIV should be greater than twice the peak input voltage.

Maximum Sine Wave Input Voltage (rms):

This is valid for resistive or inductive loads, but not for capacitive loads for the reasons listed above. The actual reverse voltage applied across the diode can be greater than the rms value listed if capacitive loads are used. Therefore, this value should be below the maximum rating.

Average Half-Wave Rectified Forward Current:

(Resistive Load or D.C. Output Current) This is the maximum D.C. current that can be passed through the diode without exceeding the maximum allowable temperature of the junction. Ambient temperature must be specified, and this current must be derated at higher temperatures.

Peak One Cycle Surge Current:

(60cps) This is the value of peak recurrent forward current that, if exceeded, will damage the diode. Temperature must be specified and the average value of this current must be kept below the maximum DC output current to prevent excessive heating of the diode.

Thermal Resistance:

Used to determine the rise in junction temperature when power dissipation is known. The product of thermal resistance and the power dissipated will give the rise in junction temperature. This will be with respect to case or ambient temperature depending on which thermal resistance is used. For different diodes, the thermal resistance is given with respect to case or to surrounding air.

Thermal Capacitance:

The energy capacity of the junction, given in w sec/°C.

Thermal Time Constant:

Thermal resistance times thermal capacitance.

Maximum Forward Voltage Drop:

The maximum voltage drop across the diode with a specified DC current and temperature.

Maximum Rated Reverse Current at Maximum Rated Reverse DC Voltage:

The reverse leakage current that will flow when the rectifier is back biased to PIV. The temperature must be specified as this current is temperature sensitive.

Operating and Storage Temperatures:

Maximum operating and storage temperatures that will not damage the device.

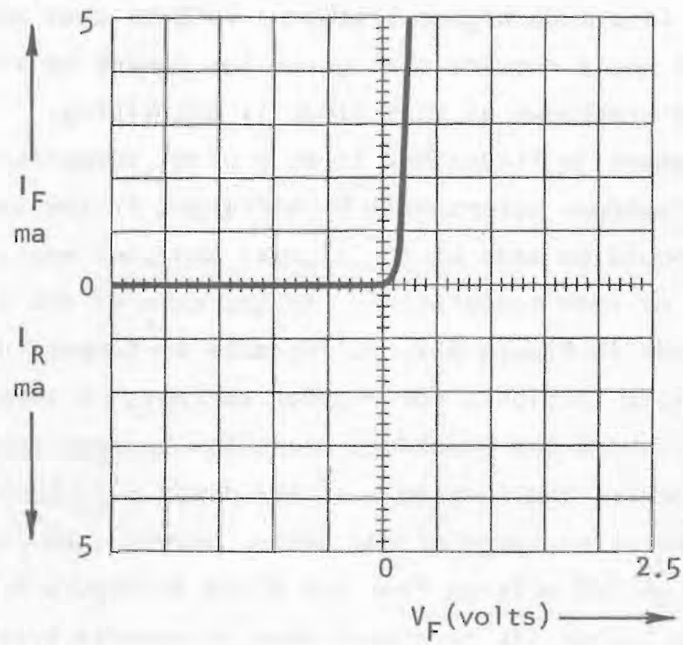
CHART 2A

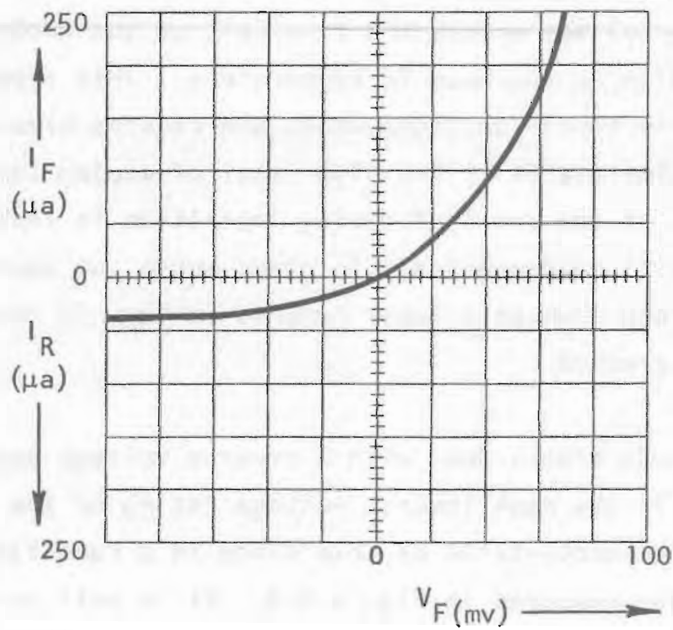
In the second measurement in Figure 2-2, the resistance has been increased to 100Ω and the total current for the same applied voltage has been halved. Since the current in the silicon rectifier diode is related exponentially to the applied voltage, the current that flows with greater than about 1V of forward voltage applied to the diode will be primarily limited by the resistance in series. Assume an applied forward voltage of 50V and the desire to have the current limited to 100ma. It becomes a simple problem to subtract slightly less than 1V from the total applied forward voltage and determine what resistance would be required to limit the current to 100ma. In this case, assuming approximately 0.7V across the diode, about 49.3V must be dropped across the series resistance. Dividing 49.3V by 100ma indicates a required resistance of 4.93KΩ. The important point here is that once a sufficient forward voltage has been applied to cause a significant forward current in the diode, the current will be primarily impedance limited, external of the diode. It must be insured that the total series resistance will limit the forward current to a safe or desired value.

Figure 3-2 shows the area at low currents with a forward voltage applied to a silicon diode. It is well to note that the current is not zero at any point with a forward voltage applied. Like the basic diode that was discussed earlier, the current varies as the exponent containing voltage varies. It can be seen that the current is related exponentially to the voltage when we expand the curve and look at the area near zero volts.

Since we are typically dealing with a fairly large forward current in silicon rectifier applications, the small current that flows below approximately 0.6 or 0.7V is not considered significant. A silicon diode in such an application can be considered to be in its off state until about 0.7V is applied and then considered to be in its on state (referring to an applied forward voltage).

Consider Figure 4-2. In this measurement set-up, the reverse characteristics of th diode are being checked. The Tektronix Type 575 Transistor-Curve Tracer, Mod 122C, is a conventional Type 575 with an additional voltage range for checking the reverse breakdown characteristics of the diodes and transistors. With the Mod 122C, a reverse voltage up to 1.5KV is possible under a limited current situation. In this measurement approach, however, the point at which reverse breakdown occurs is the desired measurement and the fact that the current must be limited is of





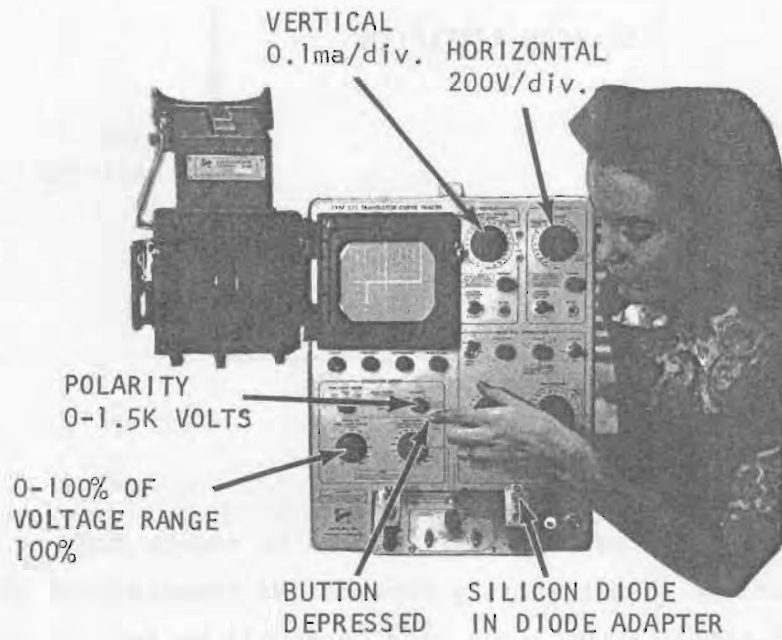
TEST SET-UP TEKTRONIX TYPE 575
TRANSISTOR-CURVE TRACER

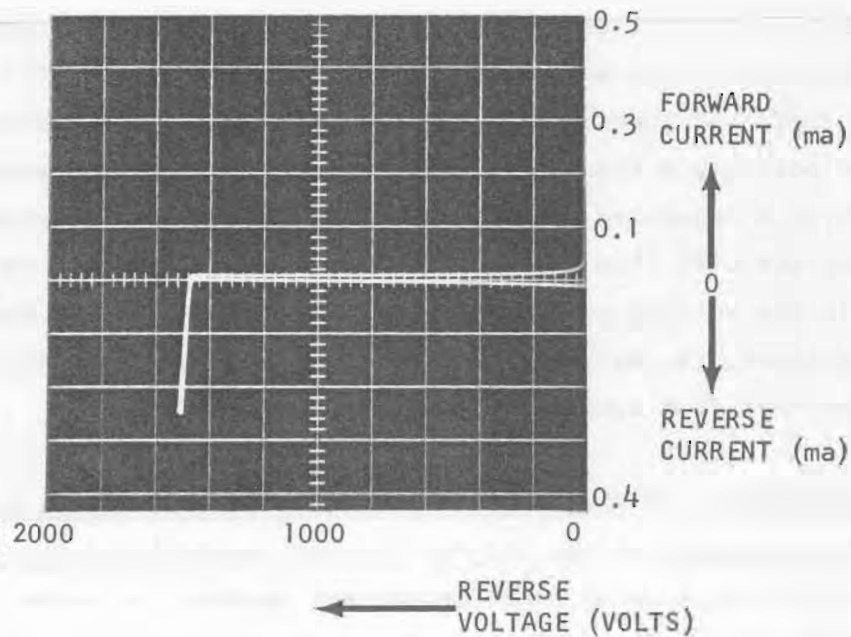
FIGURE 3-2

little consequence. The diode being measured in Figure 4-2 breaks down at about 1480V. This, of course, is a much higher breakdown voltage than comparable germanium rectifier diodes. It would require that germanium diodes be stacked in series to accomplish such a high breakdown as this diode is exhibiting. It should be remembered that the measurement in Figure 4-2 is at a given temperature, and the voltage point at which breakdown occurs will be different if the temperature changes. Ideally, a measurement should be made at the highest expected ambient temperature, along with a measurement at room temperature. In the case of the diode in avalanche breakdown such as the diode in Figure 4-2, an increase in temperature will increase the voltage across the diode terminals for a given current. A temperature point will be reached, however, where the breakdown mechanism becomes more complex and a decrease in the voltage across the terminals of the diode will be encountered. Over the normal temperature operating range of the diode, however, we can expect a positive temperature coefficient of voltage from the diode in Figure 4-2. This means we can expect the voltage across its terminals when in reverse breakdown to increase with an increase in temperature.

There is another breakdown mechanism that will be discussed later that has a negative temperature coefficient of voltage. By negative temperature coefficient of voltage, we mean that the voltage across the terminals of the diode in reverse breakdown will decrease with an increase in temperature. This type of breakdown only occurs in very heavily doped junctions where the reverse breakdown voltage has been reduced below approximately 6V by the high level of doping impurities. This brings up the point that, as the level of doping impurities is increased, the voltage breakdown point will be decreased. In other words, we would expect the diode in Figure 4-2 to break down at a lower reverse voltage if the amount of doping impurities were increased.

The point at which the diode breaks down with a reverse voltage applied (such as measured in Figure 4-2), is the peak inverse voltage rating of the diode rectifier. The total voltage excursion encountered by this diode in a rectifier application should be below the voltage measured in Figure 4-2. It is well to remember that when reactive filters are used in a power supply application, the peak voltage can be significantly larger than the peak input or applied voltage. For this reason, it is wise to insert a safety factor when calculating peak inverse voltage ratings.





TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR CURVE TRACER
MOD 122C WITH DIODE ADAPTER AND TYPE C-12 OSCILLOSCOPE CAMERA

FIGURE 4-2

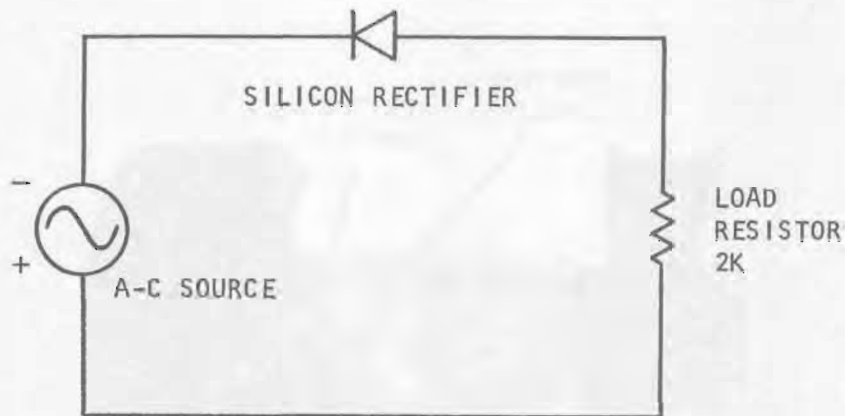


FIGURE 5-2

If the silicon rectifier is placed in series with an AC source such as shown in Figure 5-2, the a-c source will alternately forward and reverse bias the silicon rectifier. On one half cycle, the silicon rectifier will be forward biased, and will have approximately 0.7V across it with the remaining voltage dropped across the series resistance in the circuit. If this were a power supply application, the resistor might be the load that the rectifier is supplying. On the alternate half cycle, the silicon rectifier will be reverse biased and the only current is the diode saturation current governed by the temperature and the applied voltage. Placed in series with an a-c source, the silicon rectifier will deliver pulsating d-c to the load. Of course, filter circuits, etc., can be added to remove the pulsations and possibly a regulating circuit added to further remove the pulsations and to deliver a regulated d-c to the load. Considering electron current, a large electron current will flow into the arrowhead of the silicon rectifier symbol, resulting in the voltage polarity across the load as shown. When dealing with conventional current, it can be assumed that a large current will flow with the arrow in a silicon rectifier symbol.

In this publication, if the type of current is not indicated, we will assume that we are talking about electron current and not conventional current. Therefore, we will expect to find a large electron current against the arrow in the rectifier symbol. This will hold true in all of our discussions.

Refer to the simple circuit in Figure 5-2 once again and assume a peak input voltage of 50V. When the input is at its peak value and forward biasing the silicon recti-

fier, we would expect to find about 0.7V across the silicon rectifier and the remaining voltage, or 49.3V, across the load. With 49.3V across the load resistor of 2KΩ, we would expect about 24.8ma of peak current to be flowing. Note that we merely subtracted the voltage drop across the silicon rectifier from the total peak voltage applied to find the peak voltage applied across the load. Dividing the peak voltage across the load by the load resistance gave us the peak current involved. If reactive filters are involved, this becomes more complex. We might make the same calculations for average and effective value of a-c input if this were the required information. Note that on the alternate half cycle of the a-c input, we would want to insure that the peak inverse voltage specification of the silicon rectifier was greater than the expected peak input voltage of the input generator.

If the maximum reverse current of the diode at 25° is given at 10μa, we could also determine how much voltage would be across the load in a reverse bias condition of the silicon rectifier. In this case, with a reverse current of 10μa and a load resistance of 2KΩ, the product gives 20mv of reverse voltage across the load on the alternate half cycle. Suppose the 10μa were measured at a peak inverse voltage of 500V for this particular rectifier. To determine the reverse resistance of the silicon rectifier, we might divide the reverse current of 10μa into the peak inverse voltage of 500V and find that the reverse resistance of the silicon rectifier is approximately 50 megΩ at 25°C. The reverse resistance of a silicon rectifier can be found simply by dividing the reverse voltage by the reverse current. In this case, 500V divided by 10μa gave a reverse resistance of 50 megΩ. If the reverse or saturation current is stated at 150°C as 500μa, the reverse resistance will have decreased to approximately 1 megΩ at 150°C.

We might also find the forward resistance of the silicon rectifier by dividing the maximum forward voltage drop by the current stated for that value. For the diode that was measured in Figure 2-1, there was approximately 0.75V across the diode when there was 250ma of forward current. By dividing 0.75V by 250ma, we find that the forward resistance of the diode is approximately 3Ω at 250ma of forward current. Of course, this would have to be stated at the temperature that the measurement was taken. It is well to note that the forward resistance of 3Ω with respect to a reverse resistance of 50 megΩ indicates a pretty fair rectifying device and this is true of silicon rectifiers in general. Even a forward resistance of 3Ω or less with respect to 1 megΩ of reverse resistance would have to be considered a good rectifying

characteristic, and we found that with the particular rectifier measured, that the reverse resistance was only down to 1 megΩ at 150°C.

Let's stop for a moment and review. The silicon rectifier has advantages over its other semiconductor counterparts in that it has a low reverse current, a high reverse breakdown voltage, and the ability to handle large forward currents as a result of being able to handle a high concentration of electric charge. It's one disadvantage for low frequency operation might be the somewhat larger voltage drop across its terminals than its other semiconductor counterparts, and it might have an added disadvantage at the higher frequencies in that it is an inherently slower device than perhaps germanium or other semiconductor diodes. For low frequency rectifier service in power applications, we find the silicon rectifier in very extensive use. If we desired to determine the maximum power that the device could handle, we could use the thermal characteristics as outlined in section one. A large number of silicon diodes are stud mounted so that they might be placed on external heat sinks to help dissipate the heat generated in high power applications.

STACKING SILICON RECTIFIERS IN SERIES:

A higher peak inverse voltage rating may be obtained by stacking the rectifiers in series. In applications such as this, the forward voltage across the series stack is increased; however, much larger peak inverse voltage ratings may be obtained.

Figure 6-2 illustrates a simple circuit. The silicon rectifiers shown in Figure 6-2 have individual peak inverse voltage ratings of 1000V. Stacking the three in series, we would expect to find a total peak inverse voltage rating of 3000V.

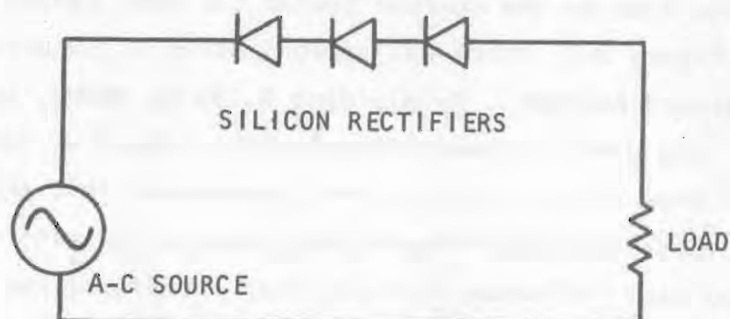


FIGURE 6-2

Of course, when dealing with the forward bias condition, we have the sum of the three voltage drops across the silicon rectifiers subtracting from the total applied voltage. This appears to be a simple solution to the reverse breakdown problem at first; however, it is a bit more complicated than this.

When stacking the silicon rectifiers in series to increase the reverse breakdown rating, it should be insured that the silicon rectifiers have near equal reverse resistances. If they do not have near equal reverse resistances, the applied reverse voltage will not divide equally among the silicon rectifiers. One rectifier will have a larger voltage applied across it than the others if its reverse resistance is higher than the others. A second consideration is the stored charge or storage time of the diodes when stacked in series. If one diode has a greater stored charge than the others, when the applied voltage changes from forward to reverse, one diode will continue to conduct for a longer period of time than the other diodes. The greater portion of the reverse voltage can be distributed across the diodes that become non-conducting and cause them to break down. In other words, the diodes should have near equal stored charges and reverse resistances if they are to be stacked in series to increase the reverse breakdown rating that one device would offer. Some manufacturers offer silicon rectifier stacks made of diodes cut from identical crystals in order to insure similar reverse resistances and stored charges.

If it is not possible to have the diodes that are to be stacked in series cut from the same crystal, the differences in reverse resistances can be minimized by adding shunting resistors such as shown in Figure 7-2. The shunting resistors placed across the individual diodes reduce the effects of the differences in reverse resistances of the diodes.

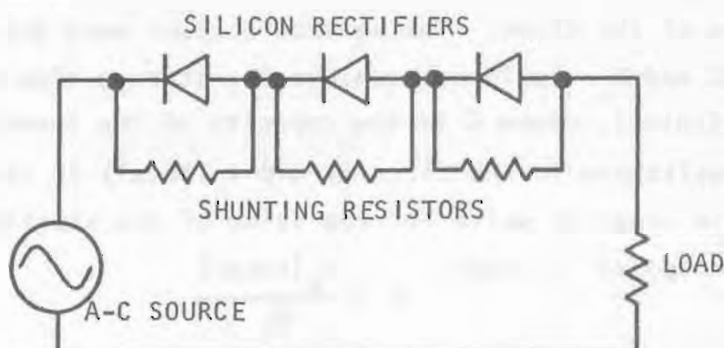


FIGURE 7-2

It is evident that care should be taken that the shunting resistors have near equal values. The shunting resistors should be high enough in value to limit the reverse current to a low value, but small enough in resistance to aid in distributing the peak inverse voltage equally among the diodes. We previously have calculated the reverse resistance of a diode. Typical values of shunting resistors are one half the reverse resistance of the individual diodes. This will be small enough to aid in the distribution of the peak inverse voltage, while large enough to limit reverse current to a reasonable value.

Since the stored charge and, as a result, storage time is a limiting factor involved, shunting capacitors can be added to minimize the effects of different stored charges in the individual diodes. Figure 8-2 illustrates this.

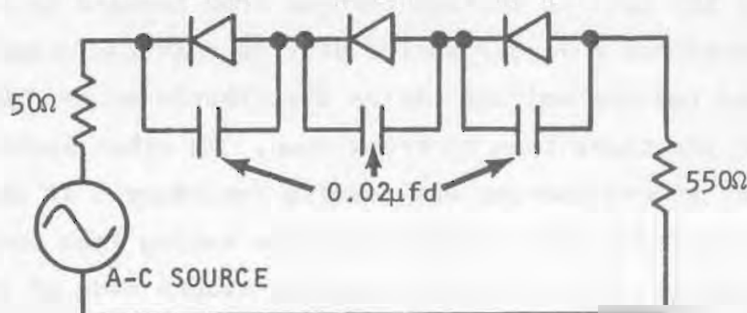
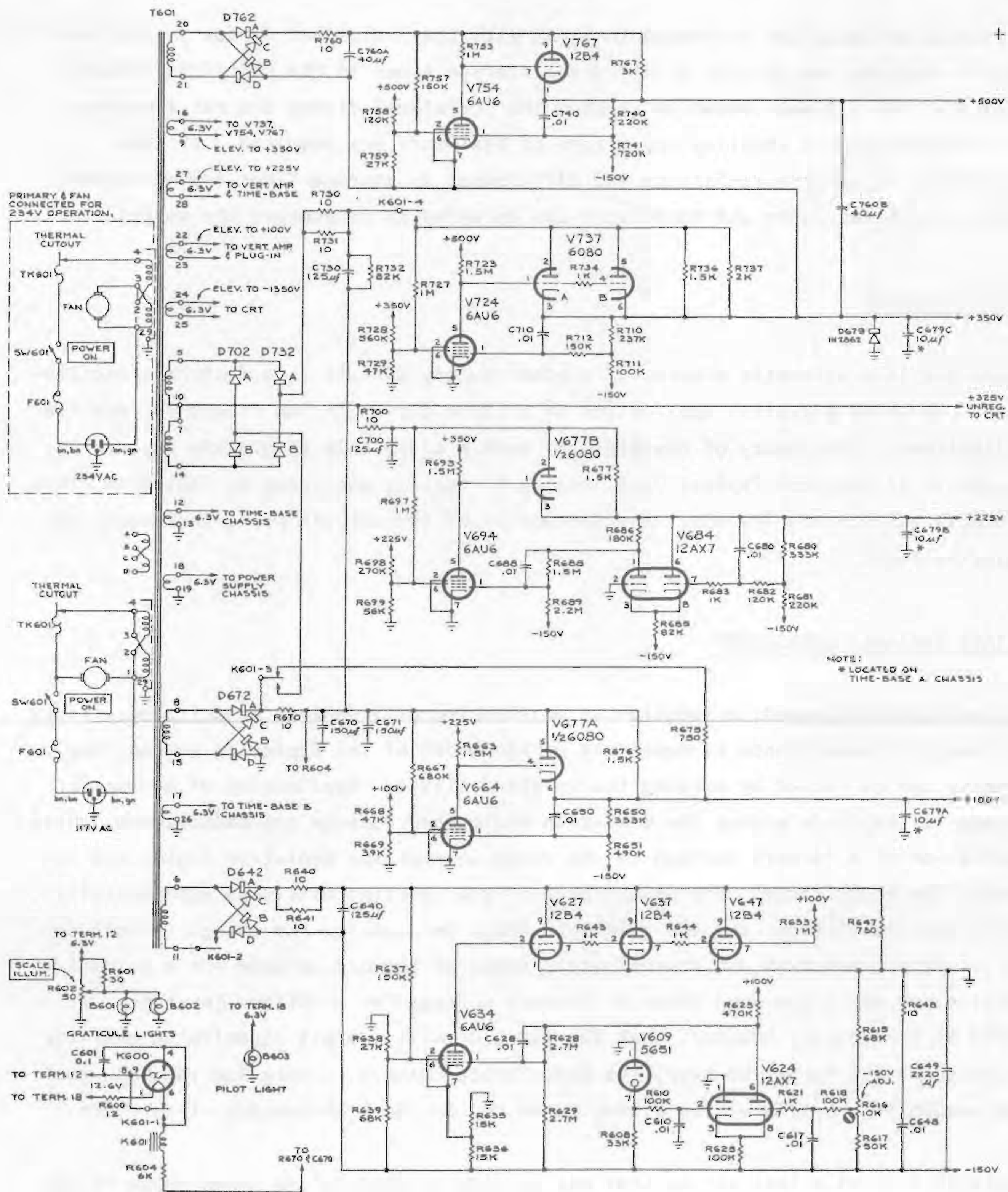


FIGURE 8-2

The values of shunting capacitors should be selected to have near equal capacitance for proper compensation. In order to compensate the difference in storage times of the individual diodes, the time constant of the shunting capacitor and the load resistance should be equal to the storage time of the diode and the load resistance or, in other words, $C R_L = t_s$, where C is the capacity of the shunting capacitor, R_L is the resistance of the load or the total series resistance in the circuit, and t_s is the storage time of the diode. Taking into account more than one diode then, the time constant of C and R_L should be equal to the storage time of all of the diodes, or $C R_L = t_s$ (total), where C is the capacity of the shunting capacitor R_L is the total series resistance in the circuit, and t_s(total) is the total storage time of the diodes. In order to solve for the value of the shunting capacitor, the formula can be transposed to read:

$$C = \frac{t_s(\text{total})}{R_L} .$$

The proper value of a shunting capacitor can be found by this formula and, when



TYPE 545A OSCILLOSCOPE

MRH 11-27-61

POWER SUPPLY

CIRCUIT NUMBERS 600 THRU 769

FIGURE 9-2

this value of capacitor is placed in shunt with the individual diodes in the stack, it will minimize the effects of different storage times in the individual diodes. Of course, the optimum situation is when the individual diodes are cut from the same crystals and no shunting capacitors or resistors are required. If both differences in reverse resistance and differences in storage times are a problem, both shunting resistors and capacitors can be added to compensate the series stack.

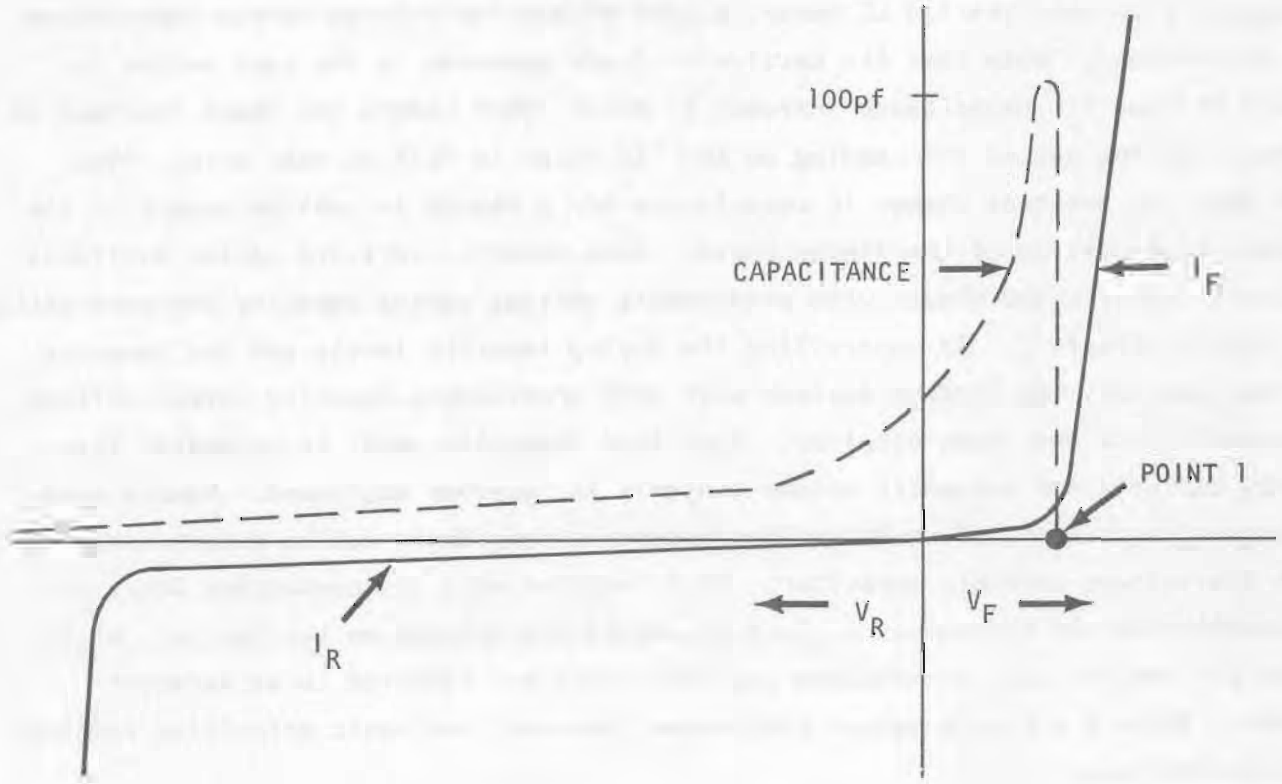
TYPICAL CIRCUIT:

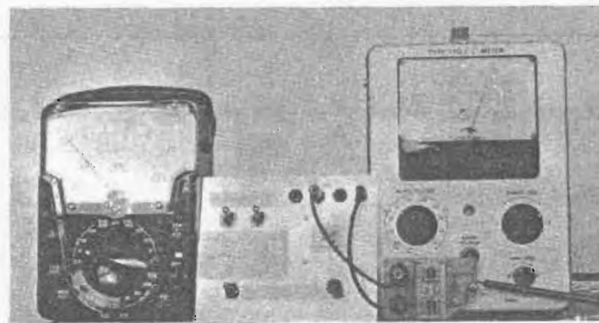
Figure 9-2 is a schematic diagram of a power supply circuit in a Tektronix oscilloscope. It shows a typical application of silicon diodes in low frequency rectifier applications. The theory of operation of such a circuit is adequately covered in Chapter 12 of the book Typical Oscilloscope Circuitry, published by Tektronix, Inc., Tektronix publication 070-253. Any discussion of the circuit will, therefore, be neglected here.

VOLTAGE VARIABLE CAPACITORS:

As previously discussed, a junction at equilibrium will exhibit capacitance. Since the amount of capacitance is dependent on the width of the depletion region, the capacity can be varied by varying the applied voltage. Application of a reverse voltage to the diode widens the depletion region and reduces the capacitance, while application of a forward voltage to the diode narrows the depletion region and increases the capacitance. The capacitance of the junction will vary exponentially as the applied voltage, and the greatest change in capacity for change in voltage will occur between zero and approximately 300mv of forward voltage for a germanium junction and about zero and 700mv of forward voltage for a silicon junction. It should be remembered, however, that the junction will exhibit capacitance over its entire operating range; however, the capacitance falls to a very low value in the high conducting portion of the diodes curve due to the high density of carriers.

Figure 10-2 shows a test set-up that may be used to measure the capacitance of the PN junction. The Tektronix Type 130 LC Meter has a guard voltage that allows the stray capacitance of the circuit to be compensated and allows only the capacitance of the diode to be measured. By varying the applied voltage, and noting the changes





TEST SET-UP

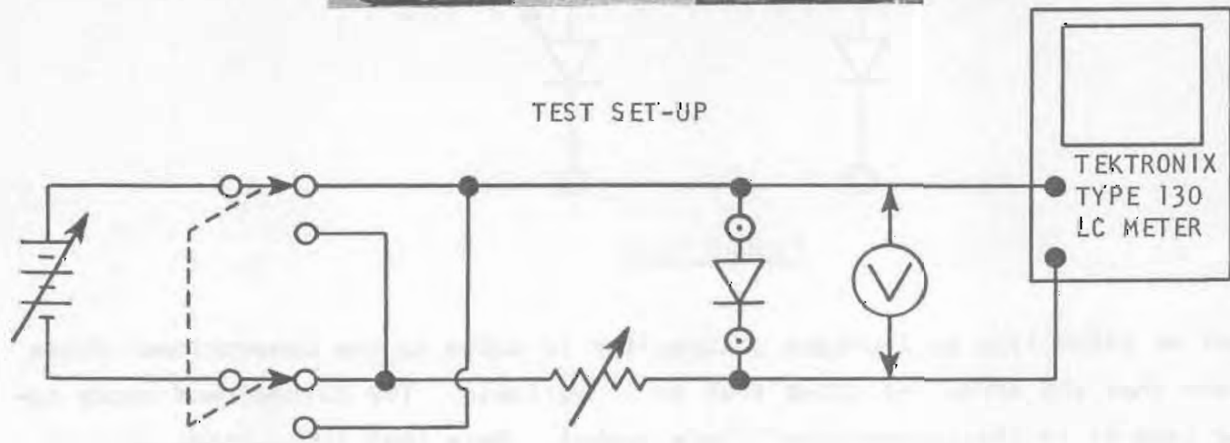


FIGURE 10-2

in capacity on the Type 130 LC Meter, a plot of applied voltage versus capacitance can be obtained. Note that the particular diode measured in the test set-up in Figure 9-2 had its capacitance increase to about 100pf before the sharp increase in forward current caused the reading on the 130 Meter to fall to near zero. Also note that the greatest change in capacitance for a change in voltage occurs in the forward bias portion of the diodes curve. Some manufacturers are making available specially constructed diodes with predictable voltage versus capacity characteristics for use in circuitry. By controlling the doping impurity levels and the geometry of the junction, low leakage devices with good predictable capacity versus voltage characteristics have been obtained. They lend themselves well to automatic frequency controls and automatic volume controls in receiver equipment. Remote capacitance adjustments that would perhaps require a long shaft can be accomplished with the voltage variable capacitor. This requires only the connecting leads to a potentiometer in a convenient place to adjust the voltage on the device. Higher power devices for use in microwave and radar work are referred to as Varactor diodes. There are also numerous tradenames; however, the basic principles involved are covered here.

SYMBOL:

Figure 11-2 shows the conventional diodes symbol along side of that of a voltage variable capacitor.

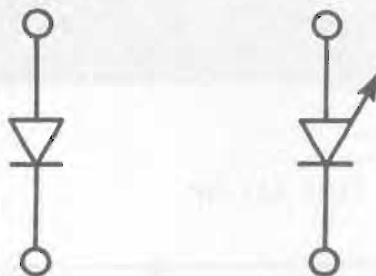


FIGURE 11-2

Note that an added line to indicate a capacitor is added to the conventional diode symbol and that the arrow indicates that it is variable. The cathode and anode remain the same as in the conventional diode symbol. Note that the voltage variable capacitor will typically be operated in what would normally be considered the off

mode of the diode.

In Figure 12-2, the applied voltage is forward biasing the voltage variable capacitor and increasing its capacitance. In Figure 13-2, the polarity of bias applied is reverse biasing the voltage variable capacitor and decreasing its capacitance.

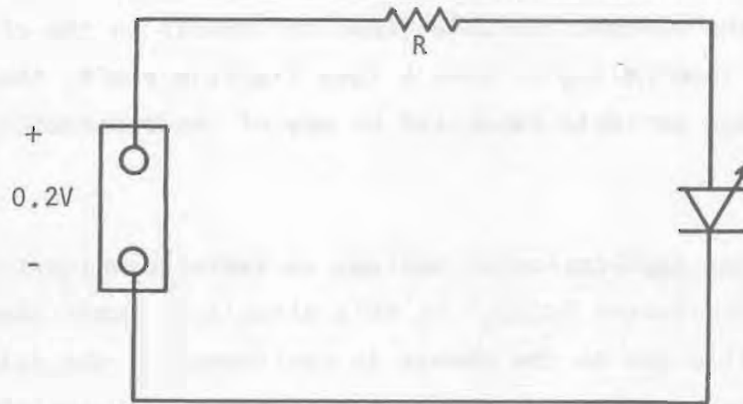


FIGURE 12-2

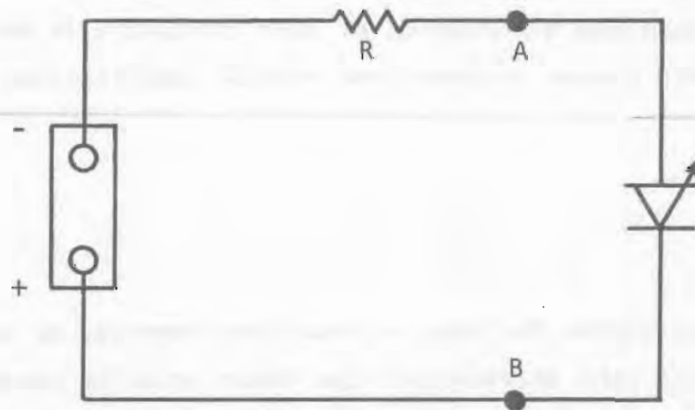


FIGURE 13-2

If the d-c source shown in Figures 12-3 and 13-2 was replaced with an a-c generator, we would find the a-c generator alternately forward and reverse biasing the voltage variable capacitor and varying its capacitance. If the voltage variable capacitor were made part of a tuned circuit, in an FM transmitter we would find the resonant frequency of the tuned circuit in the FM transmitter changing as the applied a-c generator moved through its cycle. This would frequency modulate the transmitter. Therefore, the voltage variable capacitor might lend itself well to a frequency

modulator circuit.

Figure 14-2A is a schematic diagram of a portion of the Tektronix Type 526 Television Vector Scope. In Figure 14-2A, the voltage variable capacitor is being used as an external adjustment that may be controlled from the front panel. In the schematic diagram shown, the adjustment can be made from the front panel by way of a potentiometer, while the voltage variable capacitor exists in the circuitry inside the instrument. Rather than having to have a long flexible shaft, the connections can be made to the voltage variable capacitor by way of interconnecting wires, and the shaft is not needed.

Figure 14-2B shows another application of voltage variable capacitors in the Tektronix Type 526 Television Vector Scope. In this circuitry, small changes in phase of the precision delay line due to the change in resistance of the delay line used, are compensated by feeding the d-c voltage developed across the resistance of the delay line to the voltage variable capacitors and having them exhibit either a capacitance or an inductance by changing the resonant frequency of the tuned network they are a part of. A complete explanation of this circuitry is not important as these circuit diagrams are simply to show some typical applications of voltage variable capacitors.

ZENER DIODES:

When dealing with silicon diodes for use in rectifier service, or for use as a voltage variable capacitor, the entrance of the diode into the reverse breakdown condition was undesirable. If a diode is constructed to purposely operate in the reverse or avalanche breakdown condition, the voltage across the diode terminals will remain fairly constant over a wide range of currents. This is assuming the diode remains in the reverse breakdown condition. We might draw an analogy of a diode in avalanche breakdown to the familiar voltage regulator tube. As the voltage regulator tube has the voltage across it raised to a level at which the gas ionizes, further increase in voltage will be dropped across the series resistance in the circuit. This occurs since more ionization will take place decreasing the tube's resistance and increasing the current. The voltage across the terminals of a voltage regulator tube stays relatively constant once into the ionized condition. The same is true of the PN junction when in avalanche breakdown. The voltage across its

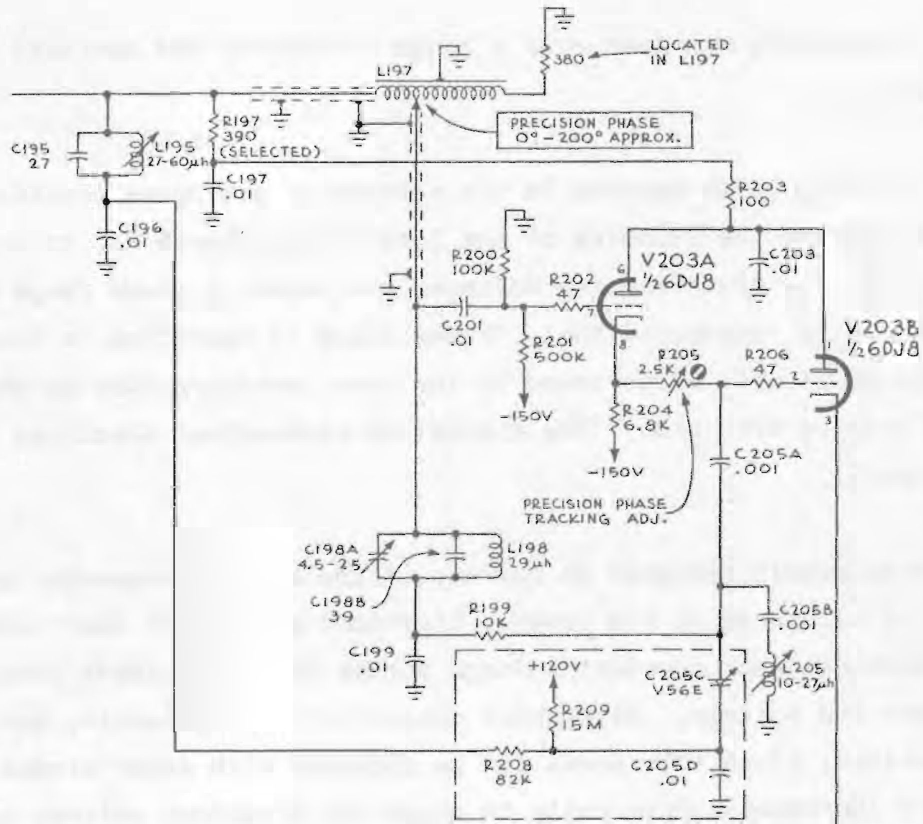


FIGURE 14-2A

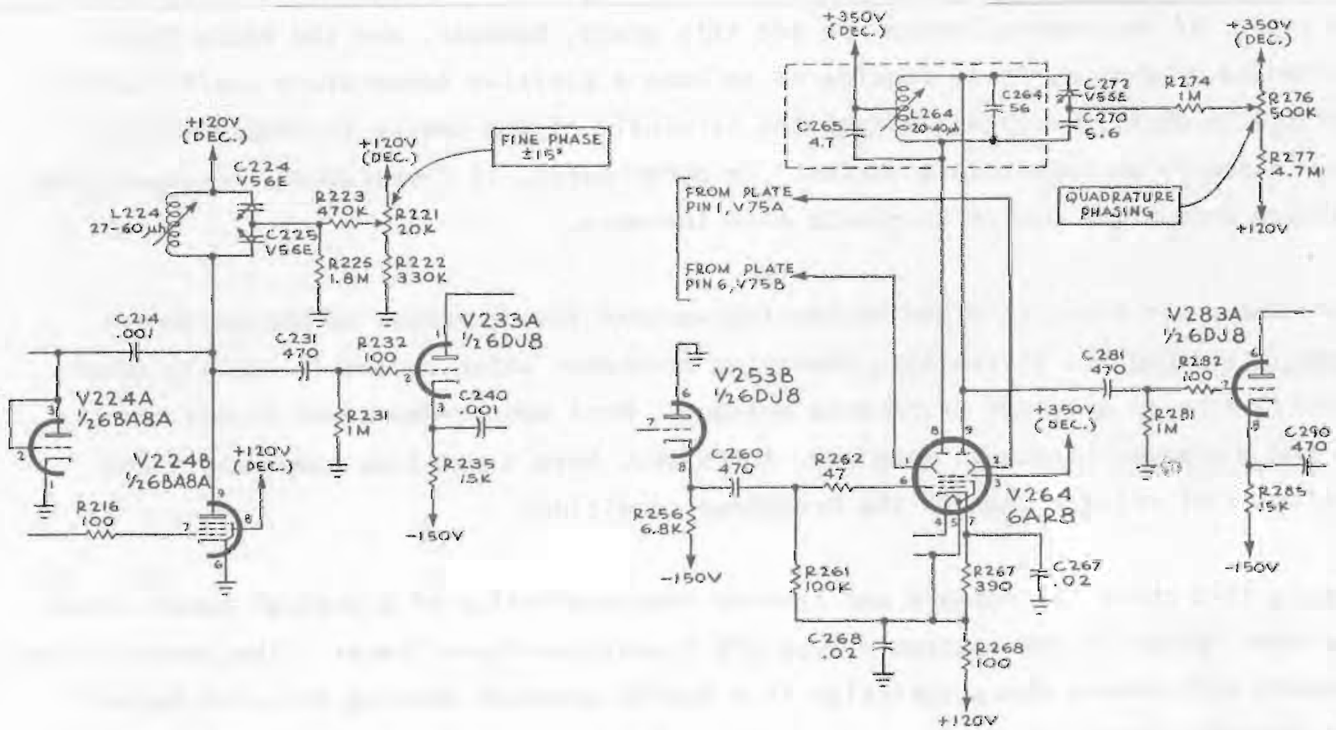


FIGURE 14-2B

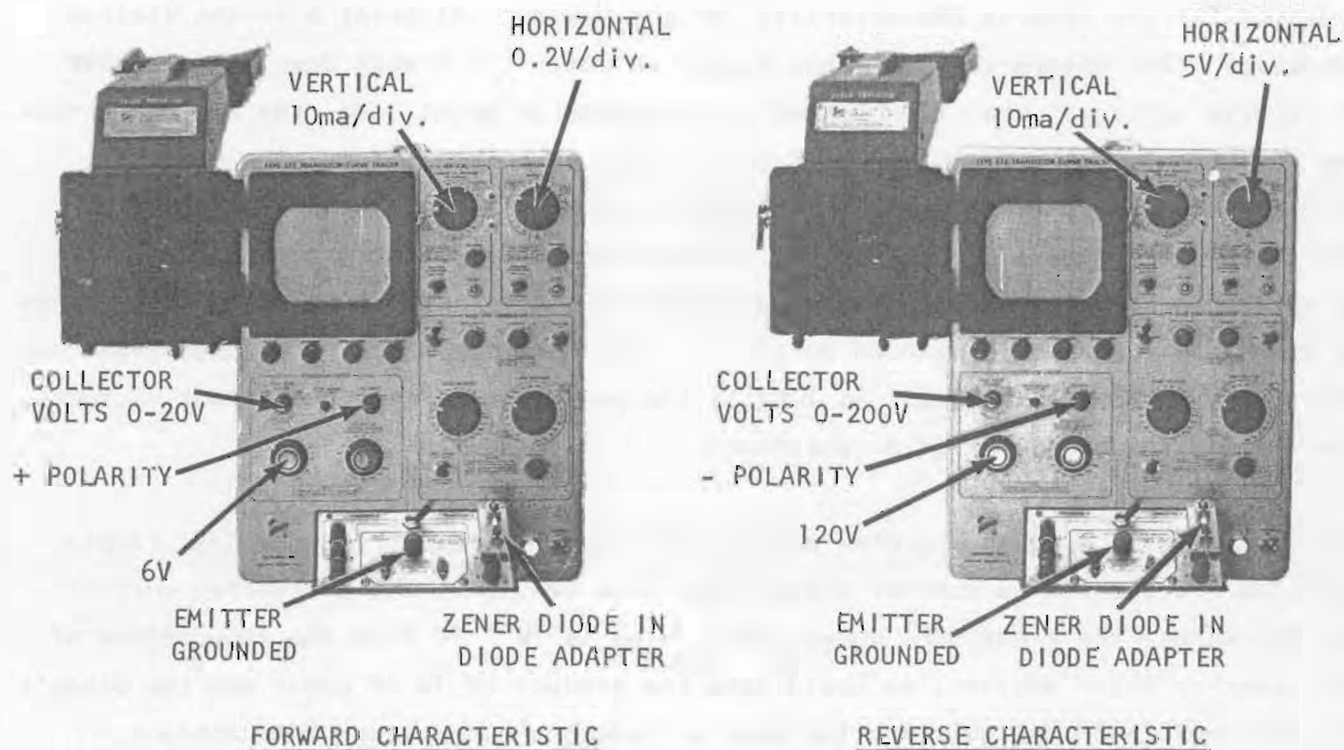
terminals stays relatively constant over a range of voltage and currents once into the breakdown mode.

When the diode is designed to operate in the reverse or avalanche breakdown mode, the amount of doping and the geometry of the junction can be varied to cause the diode to break down at a given reverse voltage, and handle a given range of reverse currents. It should be remembered that, if the diode is operating in reverse breakdown, the maximum power will be governed by the same considerations as when it is operated in the forward direction. The thermal considerations discussed in section one will still apply.

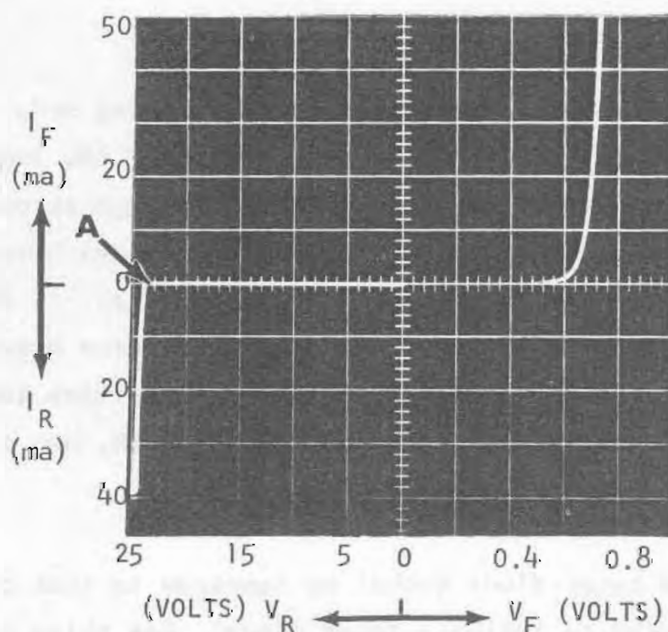
Zener diodes are purposely designed to operate in the reverse breakdown portion of their characteristics. When in the reverse breakdown portion of their characteristics, they maintain a near constant voltage across their terminals with large changes in current and voltage. By careful control of doping levels, and the geometry of the junction, close tolerances can be obtained with zener diodes. If the doping levels are increased sufficiently to cause the breakdown voltage to be below approximately 6V, another breakdown mechanism called tunneling is enhanced and the reverse breakdown voltage of the device has a negative temperature coefficient of voltage. If the doping levels are not this great, however, and the diode is in avalanche breakdown, it is considered to have a positive temperature coefficient of voltage in that the voltage across the terminals of the device in breakdown will vary directly as temperature varies. In other words, if temperature increases, the voltage across the device terminals will increase.

The name zener diode is often misleading because the breakdown mechanism termed zener breakdown is, in reality, tunneling breakdown which occurs in heavily doped junctions below about 6V of reverse voltage. Most modern-day zener diodes operate in the avalanche breakdown mode and, therefore, have a positive temperature coefficient of voltage when in the breakdown condition.

Figure 15-2 shows the forward and reverse characteristics of a typical zener diode. The test set-up is the Tektronix Type 575 Transistor-Curve Tracer. The photo of the forward and reverse characteristics is a double exposure showing both the forward and reverse characteristics of the device. Note that the horizontal scale is 0.2V per major division for the forward characteristic of the device, and 5V per major



TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER WITH DIODE ADAPTER AND TYPE C-12 OSCILLOSCOPE CAMERA



I_F = FORWARD CURRENT

I_R = REVERSE CURRENT

V_F = FORWARD VOLTAGE

V_R = REVERSE VOLTAGE

COMPOSITE: DOUBLE EXPOSURE SHOWING BOTH FORWARD AND REVERSE EI CHARACTERISTICS OF A ZENER DIODE

FIGURE 15-2

division for the reverse characteristic of the device. At point A in the diagram, the zener diode enters the avalanche breakdown mode. It breaks down at about 24V of reverse voltage. When the current is increased to about 40ma, the voltage across the device is approximately 25V. If the current is maintained at this point at this temperature, the zener diode will tend to have 25V across its terminals. If the temperature were to increase, the voltage across the diodes terminals at 40ma of reverse current would also increase. Note in Figure 15-2 that, although designed to work in the reverse breakdown portion of its characteristics, the zener diode can serve as a rectifying device. As long as the power considerations are not exceeded, the zener diode can serve as a rectifier.

To calculate the power being dissipated by the zener diode in Figure 15-2, simply take the product of the current and voltage involved. At 40ma of reverse current and 25V across the diode, the power dissipation is 1W. To find the temperature of the junction above ambient, we could take the product of 1W of power and the diode's thermal resistance and gain the increase in junction temperature over ambient. This is included as a reminder that the same basic considerations apply to zener diodes as to other types of diodes.

The product of power dissipation and the thermal resistance of the diode when added to ambient temperature must not exceed the maximum operating temperature of the zener diode.

The zener diode in Figure 15-2, when in the forward conducting mode, will have about 0.75V across its terminals when forward current is 50ma. If the temperature increases for this same current, the diode will have the voltage across its terminals reduced. Therefore, zener diodes like other PN junction diodes have a negative temperature coefficient of voltage in the forward biased mode. It is well to note at this time that when a zener diode is operated in the reverse breakdown mode and is in avalanche breakdown (above V_R equals 6V), it has a positive temperature coefficient of voltage, but when in the forward conducting mode, has a negative temperature coefficient of voltage.

Figure 16-2 illustrates the zener diode symbol as compared to that of a conventional diode. Note that a Z is added to indicate zener diode. One thing that is well to note is that a zener diode is normally operated in a reverse bias mode and depends

on minority carriers in the breakdown condition for current.

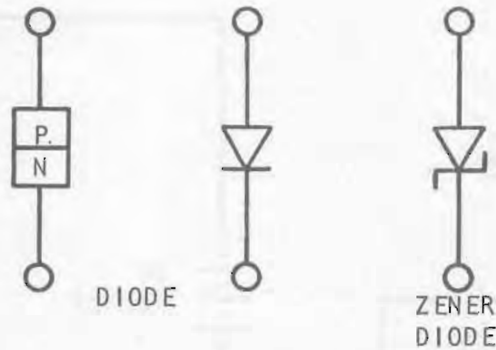


FIGURE 16-2

Therefore, electron current will flow with the arrow in the zener diode symbol or, if conventional current is used, it will flow against the arrow in the zener diode symbol. This is, of course, assuming that reverse voltage is being applied to the zener diode and that it is being operated in the reverse breakdown portion of its characteristics.

Figure 17-2 shows a zener diode in a typical application. In this case, the zener diode is being used as a voltage regulator and the applied voltage will be greater than the zener breakdown voltage.

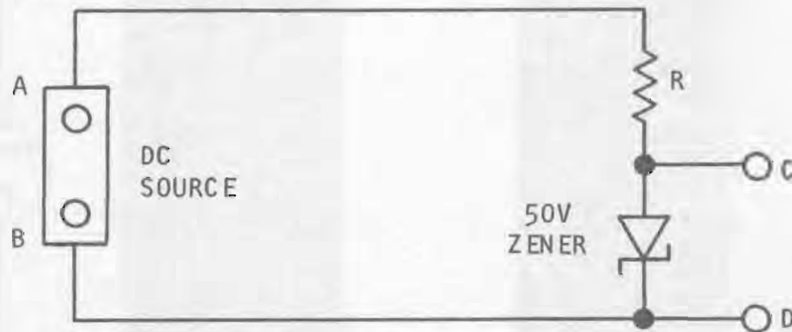
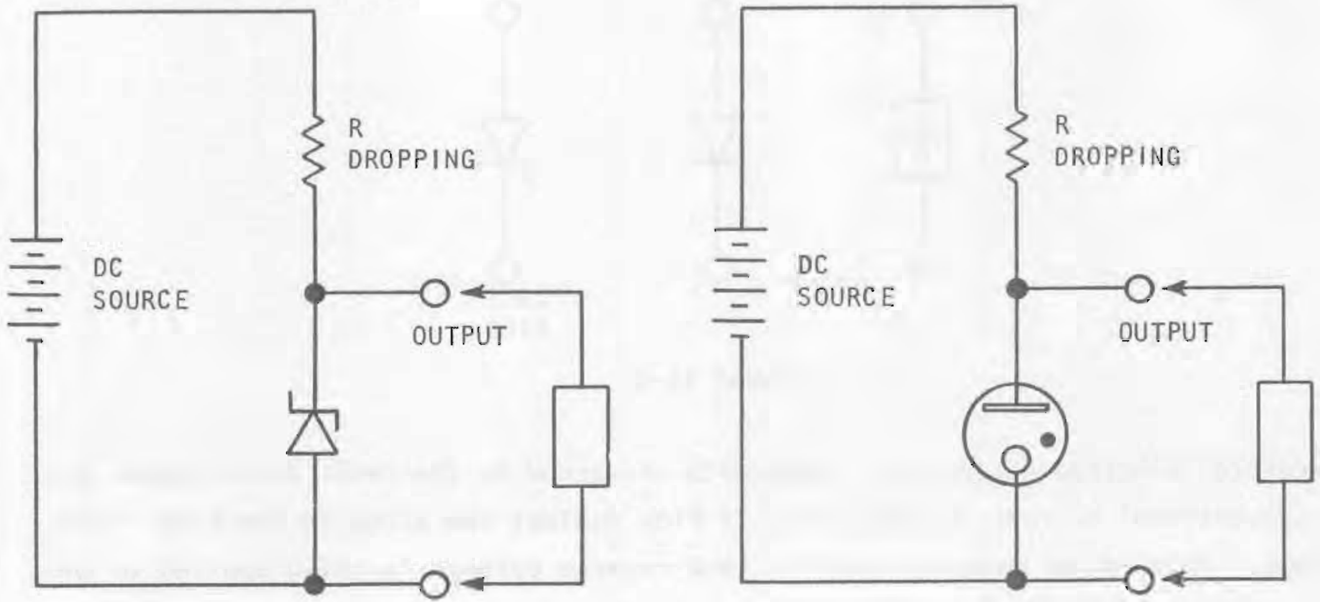
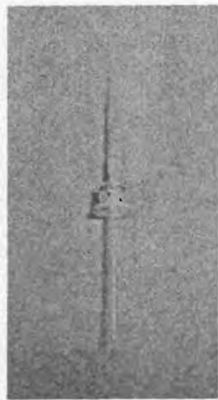


FIGURE 17-2

For normal operation, point A in the diagram will be negative and greater than 50V. The series resistance (R) will drop the added voltage and the output at points C and D will be a 50V regulated with C negative with respect to D. Note that electron current flows with the arrow in the zener diode symbol (conventional current flows against the arrow in the zener diode symbol). Any changes in the d-c source cause a change in the voltage drop across the resistor (R) and the voltage across the zener diode is maintained relatively constant. Any changes in the load current





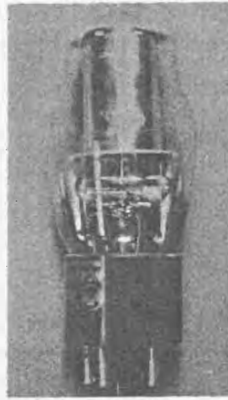


FIGURE 18-2

would cause the zener diode to change its current, dropping the same voltage across resistor (R) and maintaining the 50V output.

Figure 18-2 shows the zener diode in a typical regulating configuration and also the conventional VR tube or gas filled regulator tube in its typical configuration. Although the zener diode operation is similar to the VR tube, it has some distinct advantages. First of all, the minimum voltage of the VR tube is limited to around 70V which is the ionization potential of the gas used in VR tubes. The zener diode can operate over a range of voltages from a few volts to several hundred volts. The current through the VR tube is limited between the levels set by the internal geometry of the tube and the minimum current needed to maintain the gas in an ionized state. The current through the zener diode can be varied from a very small value to a maximum value set by the maximum power dissipation of the zener diode at a given temperature. There will be a limitation set by the geometry of the zener diode as well. The noise generated in the zener diode in the VR tube are about the same in magnitude since the collision of accelerated carriers in the zener diode in the avalanche mode and the ionization of the gas in the VR tube generates about the same level of noise. The minimum ionization potential of the VR tube causes it to break into relaxation oscillation if a capacitor is placed in shunt with it. This is not so of the zener diode since it can operate over a wide range of currents and voltages, therefore, a shunt capacitor or RC time constant can be used with the zener diode to filter out the noise. Such a filter used with the VR tube would cause it to break into relaxation oscillations.

Figure 19-2 illustrates a noise suppression circuit.

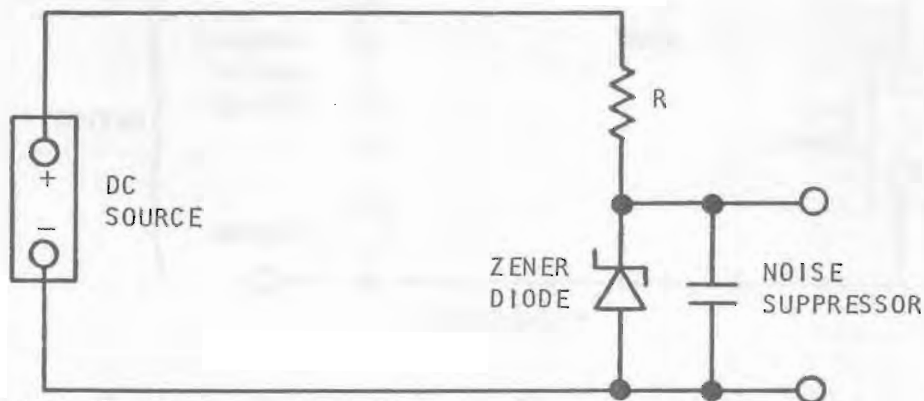


FIGURE 19-2

A shunt capacitor of from 0.01 to 0.1μf will tend to compensate the noise generated in a zener diode. The same capacitor placed in shunt with the conventional VR tube would cause it to break into relaxation oscillations. Smaller size is a big factor in the advantages of the zener diode over that of the conventional VR tube. The fact that the zener diode can be stud mounted and bolted to the chassis for external heat sinking results in a much smaller package when compared to a VR tube with similar power handling capabilities.

Figure 20-2 shows a typical application of a zener diode in a voltage regulator configuration. In this case, the voltage is being regulated by two zener diodes, the first of which is taking care of line voltage fluctuations and the second being used to compensate for load current variations.

If the changes in voltage across the zener diode terminals must be compensated for changes in temperature, the zener diode can be placed in series with a forward biased diode. Temperature compensation is accomplished when the output is taken across both devices. Figure 21-2 illustrates this. The forward biased diode in the diagram has a negative temperature coefficient of voltage while the zener diode in avalanche has a positive temperature coefficient of voltage and they tend to temperature compensate each other. If one forward bias diode does not have sufficient change in voltage with temperature, more than one forward bias diode can be added in series with the zener diode. It should be remembered that when this is done, the total output voltage will be the sum of the zener voltage and the voltage drops across the two forward biased diodes. Figure 22-2 illustrates this.

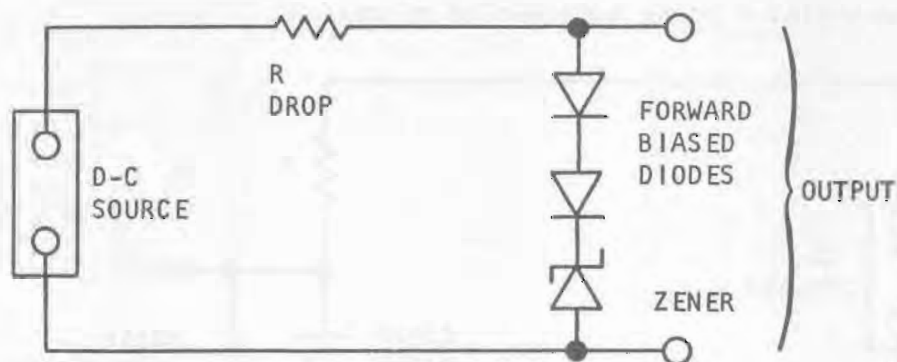
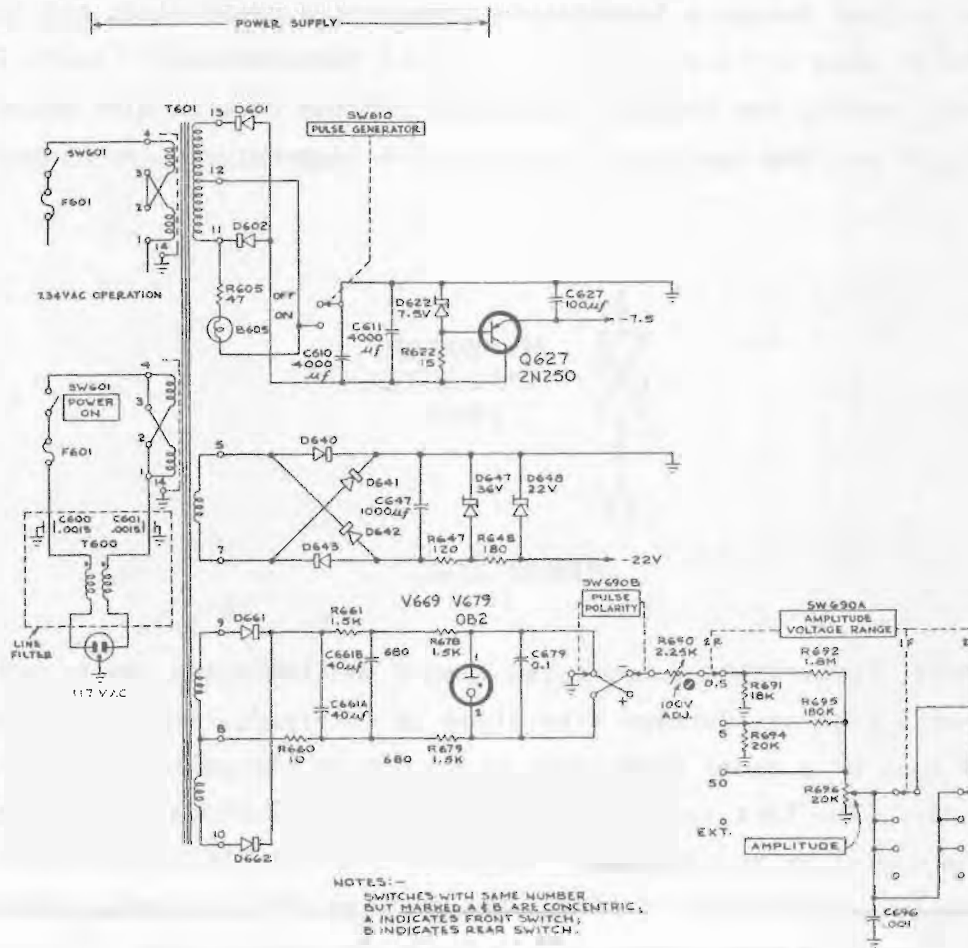


FIGURE 22-2

Some manufacturers are producing zener diodes in which the zener diode is constructed with a comparable zener in series opposition in the same encapsulation.



TYPE 110 FIGURE 20-2

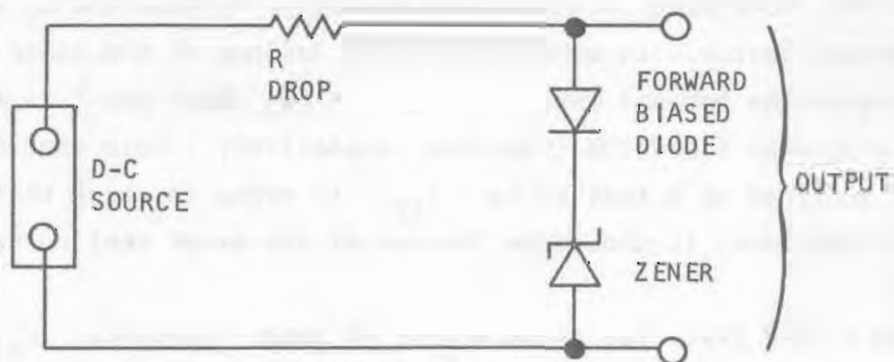


FIGURE 21-2

The zener diode is then termed a temperature compensated zener diode and the tolerances can be listed very critically over a range of temperatures. Figure 23-2 illustrates this, showing the forward bias zener and the reverse bias zener in series in a circuit and the equivalent symbol for a temperature compensated zener.

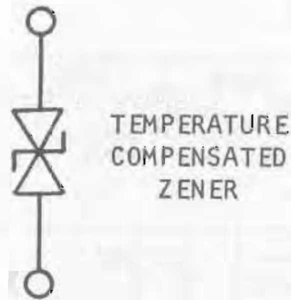


FIGURE 23-2

There are, however, temperature compensated zeners available and the symbol does not indicate that a separate forward bias diode is involved. Since the functional symbol is still that of a zener diode and the output is the voltage across the terminals of the device, the fact that a separate forward bias diode is involved is neglected in the symbology and the symbol shown is that of a conventional zener diode. The temperature coefficient of the zener diode will indicate, however, a very close tolerance in zener voltage over a range of temperatures.

Figure 24-2 is a specification sheet for Motorola zener diodes and has been reprinted with the consent of Motorola Semiconductor Products, Inc., Phoenix, Arizona. Note that the Motorola type number has the first number indicating the wattage rating of the zener diode followed by the letter indicating Motorola and the nominal zener voltage. The first zener diode indicated in Figure 24-2 is a 10M6.8Z. This indicates a 10W, 6.8V zener diode. When no number follows the Z, it is assumed to be a 20% tolerance device. In other words, the voltage of the zener diode could be 20% above or below the nominal zener voltage listed. When the Z is followed by a 5 or a 10, it indicates 5 and 10% tolerance respectively. Note that the nominal zener voltage is measured at a test current I_{ZT}. In order to check this diode against its specifications, it should be checked at the zener test current.

The chart in Figure 24-2 lists two measurements of zener impedance. Z_{ZT} is the dynamic impedance measured at the zener test current I_{ZT}. Note in the case of the

COURTESY: MOTOROLA SEMICONDUCTOR PRODUCTS, INC., PHOENIX, ARIZONA

ELECTRICAL CHARACTERISTICS (@ 30°C Case unless otherwise specified)

JEDEC TYPE NUMBER	MOTOROLA TYPE NO.	Nominal Zener Voltage @ I_{ZT} (V_Z) Volts	Test Current (I_{ZT}) mA	Max Zener Impedance Z_{ZT} @ I_{ZT} ohms	Z_{ZK} @ I_{ZK} ohms	I_{ZK} mA	Max DC Zener Current 55°C Base (I_{ZM}) mA	Max Forward Voltage (V_F) Volts @I_F = 2 Amps	Typical Zener Voltage Temp. Coeff. %/°C
1N2970	10M6.8Z	6.8	370	1.2	500	1.0	1,320	1.5	.040
1N2971	10M7.5Z	7.5	335	1.3	250	1.0	1,180	1.5	.045
1N2972	10M8.2Z	8.2	305	1.5	250	1.0	1,040	1.5	.048
1N2973	10M9.1Z	9.1	275	2.0	250	1.0	960	1.5	.051
1N2974	10M10Z	10	250	3	250	1.0	860	1.5	.055
1N2975	10M11Z	11	230	3	250	1.0	780	1.5	.060
1N2976	10M12Z	12	210	3	250	1.0	720	1.5	.065
1N2977	10M13Z	13	190	3	250	1.0	660	1.5	.065
1N2979	10M15Z	15	170	3	250	1.0	560	1.5	.070
1N2980	10M16Z	16	155	4	250	1.0	530	1.5	.070
1N2982	10M18Z	18	140	4	250	1.0	460	1.5	.075
1N2984	10M20Z	20	125	4	250	1.0	420	1.5	.075
1N2985	10M22Z	22	115	5	250	1.0	380	1.5	.080
1N2986	10M24Z	24	105	5	250	1.0	350	1.5	.080
1N2988	10M27Z	27	95	7	250	1.0	300	1.5	.085
1N2989	10M30Z	30	85	8	300	1.0	280	1.5	.085
1N2990	10M33Z	33	75	9	300	1.0	260	1.5	.085
1N2991	10M36Z	36	70	10	300	1.0	230	1.5	.085
1N2992	10M39Z	39	65	11	300	1.0	210	1.5	.085
1N2993	10M43Z	43	60	12	400	1.0	195	1.5	.090
1N2995	10M47Z	47	55	14	400	1.0	175	1.5	.090
1N2997	10M51Z	51	50	15	500	1.0	160	1.5	.090
1N2999	10M56Z	56	45	16	500	1.0	150	1.5	.090
1N3000	10M62Z	62	40	17	600	1.0	130	1.5	.090
1N3001	10M68Z	68	37	18	600	1.0	120	1.5	.090
1N3002	10M75Z	75	33	22	600	1.0	110	1.5	.090
1N3003	10M82Z	82	30	25	700	1.0	100	1.5	.090
1N3004	10M91Z	91	28	35	800	1.0	85	1.5	.090
1N3005	10M100Z	100	25	40	900	1.0	80	1.5	.090
1N3007	10M110Z	110	23	55	1,100	1.0	72	1.5	.095
1N3008	10M120Z	120	20	75	1,200	1.0	67	1.5	.095
1N3009	10M130Z	130	19	100	1,300	1.0	62	1.5	.095
1N3011	10M150Z	150	17	175	1,500	1.0	54	1.5	.095
1N3012	10M160Z	160	16	200	1,600	1.0	50	1.5	.095
1N3014	10M180Z	180	14	260	1,850	1.0	45	1.5	.095
1N3015	10M200Z	200	12	300	2,000	1.0	40	1.5	.100

JEDEC TYPE NO.	MOTOROLA TYPE NO.	Nominal Zener Voltage @ I_{ZT} (V_Z) Volts	Test Current (I_{ZT}) mA	Max Zener Impedance Z_{ZT} @ I_{ZT} Ohms	Z_{ZK} @ I_{ZK} Ohms	I_{ZK} mA	Max DC Zener Current 55°C Base (I_{ZM}) mA	Max Forward Voltage V_F @ 10 amps	Typical Zener Voltage Temp. Coeff. %/°C
1N2804	50M6.8Z	6.8	1850	0.2	70	5	6,600	1.5	.040
1N2805	50M7.5Z	7.5	1700	0.3	70	5	5,900	1.5	.045
1N2806	50M8.2Z	8.2	1500	0.4	70	5	5,200	1.5	.048
1N2807	50M9.1Z	9.1	1370	0.5	70	5	4,800	1.5	.051
1N2808	50M10Z	10	1200	0.6	80	5	4,300	1.5	.055
1N2809	50M11Z	11	1100	0.8	80	5	3,900	1.5	.060
1N2810	50M12Z	12	1000	1.0	80	5	3,600	1.5	.065
1N2811	50M13Z	13	960	1.1	80	5	3,300	1.5	.065
1N2813	50M15Z	15	830	1.4	80	5	2,800	1.5	.070
1N2814	50M16Z	16	780	1.6	80	5	2,650	1.5	.070
1N2816	50M18Z	18	700	2.0	80	5	2,300	1.5	.075
1N2818	50M20Z	20	630	2.4	80	5	2,100	1.5	.075
1N2819	50M22Z	22	570	2.5	80	5	1,900	1.5	.080
1N2820	50M24Z	24	520	2.6	80	5	1,750	1.5	.080
1N2822	50M27Z	27	460	2.8	90	5	1,500	1.5	.085
1N2823	50M30Z	30	420	3.0	90	5	1,400	1.5	.085
1N2824	50M33Z	33	380	3.2	90	5	1,300	1.5	.085
1N2825	50M36Z	36	350	3.5	90	5	1,150	1.5	.085
1N2826	50M39Z	39	320	4.0	90	5	1,050	1.5	.090
1N2827	50M43Z	43	290	4.5	90	5	975	1.5	.090
1N2829	50M47Z	47	270	5.0	100	5	880	1.5	.090
1N2831	50M51Z	51	245	5.2	100	5	810	1.5	.090
1N2833	50M62Z	62	200	7	120	5	660	1.5	.090
1N2834	50M68Z	68	180	8	140	5	600	1.5	.090
1N2835	50M75Z	75	170	9	150	5	540	1.5	.090
1N2836	50M82Z	82	150	11	160	5	490	1.5	.090
1N2837	50M91Z	91	140	15	180	5	420	1.5	.090
1N2838	50M100Z	100	120	20	200	5	400	1.5	.090
1N2840	50M110Z	110	110	30	220	5	365	1.5	.095
1N2841	50M120Z	120	100	40	240	5	335	1.5	.095
1N2842	50M130Z	130	95	50	275	5	310	1.5	.095
1N2843	50M150Z	150	85	75	400	5	270	1.5	.095
1N2844	50M160Z	160	80	80	450	5	250	1.5	.095
1N2845	50M180Z	180	68	90	525	5	220	1.5	.095
1N2846	50M200Z	200	65	100	600	5	200	1.5	.100

FIGURE 24-2

10M6.8Z that the dynamic impedance is 1.2Ω at the nominal zener voltage and test current. Z_{ZK} is the dynamic impedance measured at a specified reverse current, in this case 1ma. The measured dynamic impedance at 1ma is 500Ω. The relationship of Z_{ZT} to Z_{ZK} gives a measure of the sharpness of the zener characteristic at the entry into a breakdown condition. It also indicates the dynamic impedance that would be encountered at these two operating points if the zener diode were being used in an a-c circuit configuration. The maximum d-c current rating of a zener diode indicates the maximum steady state current that can be passed by the zener diode. Maximum forward voltage is the maximum voltage that would be found across the zener diodes terminals in a forward bias mode at the specified value of forward current. The typical zener voltage temperature coefficient in percent per degree centigrade is the percent of a nominal zener voltage per degree centigrade that the voltage will change with a change in temperature. All of the temperature coefficients listed in Figure 24-2 are positive temperature coefficients of voltage.

Figure 25-2 shows the characteristics for zener diodes above and below 6V.

COURTESY: MOTOROLA SEMICONDUCTOR PRODUCTS, INC., PHOENIX, ARIZONA

MOTOROLA 400 MILLIWATT ZENER DIODES (3.3 — 12 VOLTS) ELECTRICAL CHARACTERISTICS (@ 25°C Ambient unless otherwise specified)

Type Number	Nominal Zener Voltage V_Z — Volts @ $I_{ZT} = 20mA$	Maximum Zener Impedance Z_{ZT} — Ohms @ $I_{ZT} = 20mA$	Maximum Reverse Leakage Current $T_A = 25°C$ I_R — μA @ $V_R = 1V$	Maximum Reverse Leakage Current $T_A = 150°C$ I_R — μA @ $V_R = 1V$	Maximum Zener Current I_{ZM} — mA	Maximum Forward Voltage V_F — Volts $I_F = 200mA$	Zener Voltage Coefficient Temperature %/°C
1N746	3.3	28	10	30	110	1.5	—.060
1N747	3.6	24	10	30	100	1.5	—.055
1N748	3.9	23	10	30	95	1.5	—.050
1N749	4.3	22	2	30	85	1.5	—.035
1N750	4.7	19	2	30	75	1.5	—.020
1N751	5.1	17	1	20	70	1.5	—.010
1N752	5.6	11	1	20	65	1.5	+.005
1N753	6.2	7	0.1	20	60	1.5	+.020
1N754	6.8	5	0.1	20	55	1.5	+.035
1N755	7.5	6	0.1	20	50	1.5	+.045
1N756	8.2	8	0.1	20	45	1.5	+.050
1N757	9.1	10	0.1	20	40	1.5	+.055
1N758	10.0	17	0.1	20	35	1.5	+.060
1N759	12.0	30	0.1	20	30	1.5	+.060

TOLERANCES

The Motorola Type Numbers shown have a standard tolerance on nominal zener voltage of ± 10%. To specify ± 5% tolerance on nominal voltage, add suffix letter "A" to the Motorola Type Number. Example: Type 1N746A is a 3.3 volt nominal ± 5% tolerance zener diode.

FIGURE 25-2

The zener diodes listed have positive temperature coefficients until the voltage is reduced below 6V and the diode enters a different breakdown mode. The 1N752 in Figure 25-2 has a positive temperature coefficient while the 1N751 has a negative temperature coefficient of voltage. The breakdown mode has changed from an avalanche breakdown mode to a tunnel breakdown mode and the temperature coefficient of voltage has changed from positive to negative. As the doping levels are increased and the nominal zener voltage is decreased, the negative temperature coefficient of voltage becomes more and more pronounced. The 1N746 which has a nominal zener voltage of 3.3V has a -0.06% °/C negative temperature coefficient of voltage. From the characteristics listed in Figure 25-2, a 1N752 has the lowest temperature coefficient of voltage of any of the devices listed. Selecting this diode as a voltage reference would give a very stable voltage with respect to temperature. The critical point of transition between avalanche and tunnel breakdown is being approached and the 1N752 has a very small temperature coefficient of voltage.

TUNNEL DIODES:

As previously discussed, increasing the amount of doping impurities added to the intrinsic semiconductor when forming a junction reduces the voltage at which the diode enters a reverse breakdown condition. This is illustrated in Figure 26-2.

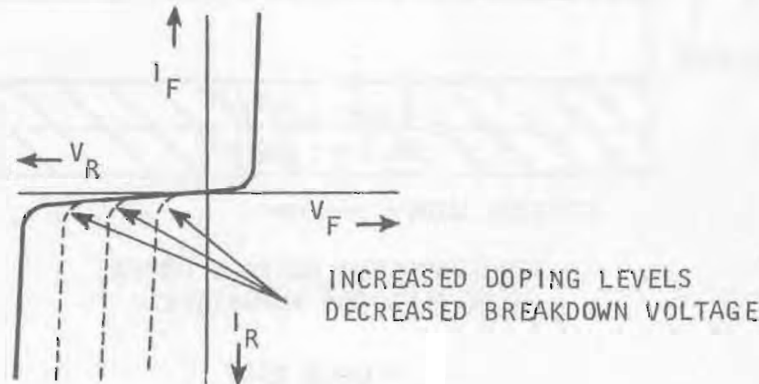
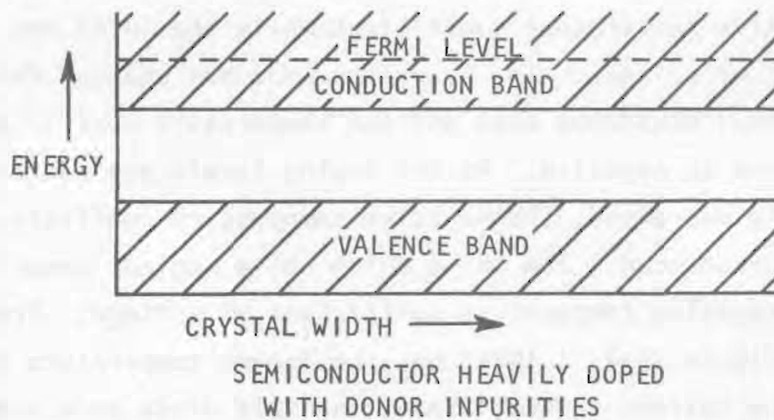


FIGURE 26-2

Increasing the number of donor impurities added to the N side of the junction moves the fermi level farther from the center of the band gap towards the conduction band. In construction of tunnel diodes, the fermi level is moved into the conduction band of the N side by very heavy doping. In other words, sufficient impurities are added to the N side of the junction to cause the fermi level to exist

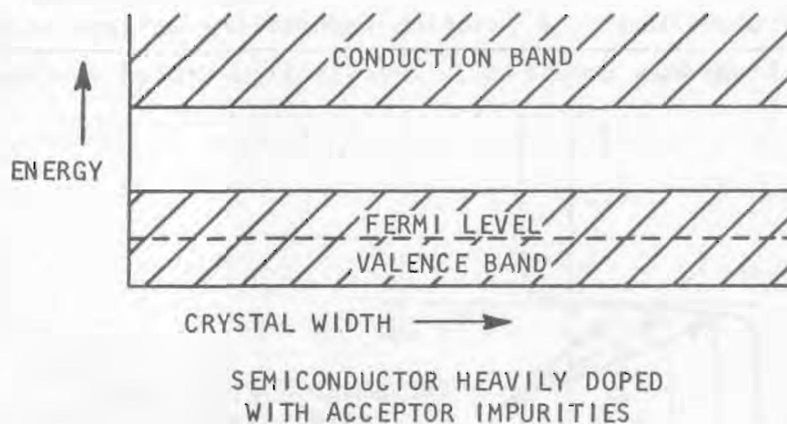
in the conduction band of the N side. This is illustrated in Figure 27-2.



SEMICONDUCTOR HEAVILY DOPED
WITH DONOR INPURITIES

FIGURE 27-2

At the same time, the P side of a tunnel diode has sufficient impurities added to move the fermi level down into the valence band of the P side. This is illustrated in Figure 28-2.



SEMICONDUCTOR HEAVILY DOPED
WITH ACCEPTOR IMPURITIES

FIGURE 28-2

The amount of impurities added in the doping process determines the reverse voltage at which the diode enters breakdown. A point can be reached where a further increase in the doping level actually causes the diode to be in what is termed a reverse breakdown condition with forward voltage applied. This is illustrated in Figure 29-2.

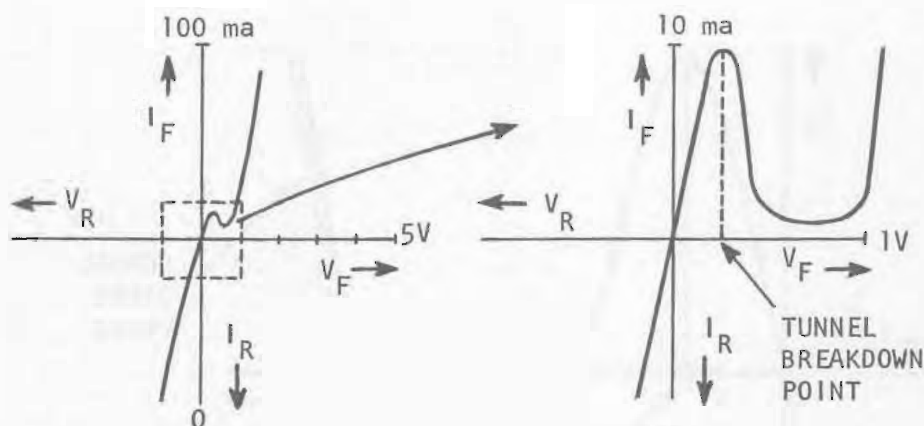


FIGURE 29-2

Sufficient impurities have been added to the diode whose characteristics are shown in Figure 29-2 to have the diode in a breakdown condition with forward voltage applied. The small section below 1V that is in the dotted box in the left diagram in Figure 29-2 is expanded and shown in the diagram at the right. Notice that, with 0V applied, there is no net external current. With the application of reverse voltage, current increases in the negative direction immediately and with the application of forward voltage, the current increases in the forward direction immediately. A forward voltage point is reached at which the current no longer continues to increase but starts to decrease with a further increase in forward voltage. The dynamic resistance in this region of the curve represents a negative value. A positive change or an increase in forward voltage results in a decrease or a negative current excursion, and this represents a negative resistance. A further increase in the forward voltage causes the diode to enter the normal forward conduction characteristic and the normal diode curve results. The increased doping levels result in an area of negative resistance or conductance between the zero voltage point and the normal forward conducting portion of the diodes characteristic. It has been found that controlling the geometry of the junction can control the amount of current flowing at the peak just to the left of the negative resistance (or conductance) region in Figure 29-2. In Figure 29-2, the current at the peak is approximately 10ma. Changing the geometry of the junction can change the peak current at this point and, of course, change the amount of negative resistance shown by the diode.

Figure 30-2 shows the tunnel diode EI curve related to a normal diode EI curve.

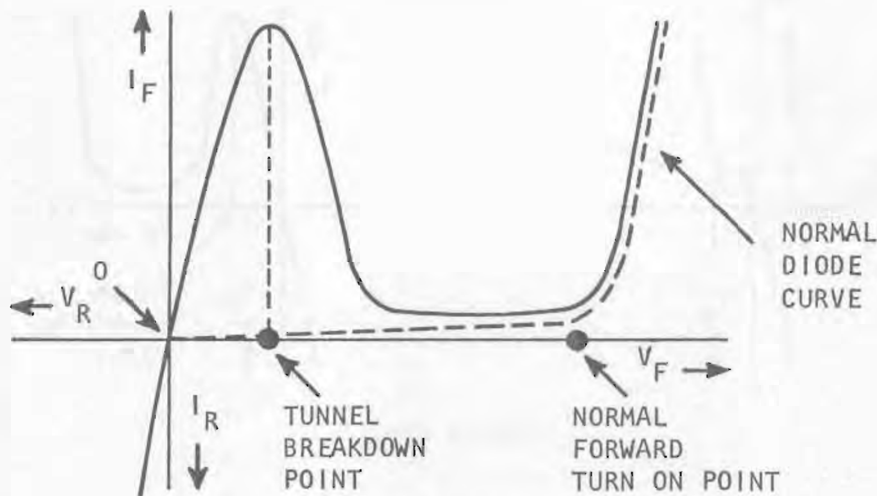


FIGURE 30-2

The tunnel diode curve illustrates the results of adding a large amount of impurities when forming the junction. We can say that the tunnel diode is a heavily doped junction that exhibits a negative resistance characteristic over a portion of its operating range. You will recall that we stated that adding a large amount of impurities can actually move the fermi level of the N side of the junction into the conduction band of the N side and move the fermi level into the valence band of the P side. The result, when the junction is formed, is a very narrow depletion region or transition region between the N and the P material, and an overlap of the valence and conduction bands between the N and the P side separated by the band gap. This is illustrated in Figure 31-2. The width of the junction or transition region between the N and the P material of a tunnel diode can be on the order of 100 Å units or about 0.4 μinches. The electrons in the conduction band of the N side are opposite the electrons in the valence band of the P side with zero volts applied (junction at equilibrium), and there is no external current. With the transition region made this narrow as a result of a large amount of doping impurities, a process referred to as quantum mechanical tunneling is enhanced.

For this discussion, the process known as quantum mechanical tunneling will be simplified to the process of an electron on one side of the junction or barrier crossing the junction without having sufficient energy to do so in the normal fashion. It is possible in very narrow junctions when the bands are overlapped, for electrons

N P N P N P

CONDUCTION BAND
FERMI
EMPTY STATES
VALENCE BAND
ELECTRONS
BAND GAP
BAND GAP

A B C

N P N P N P

BAND GAP
EMPTY STATES
BAND GAP
BAND GAP

D E F

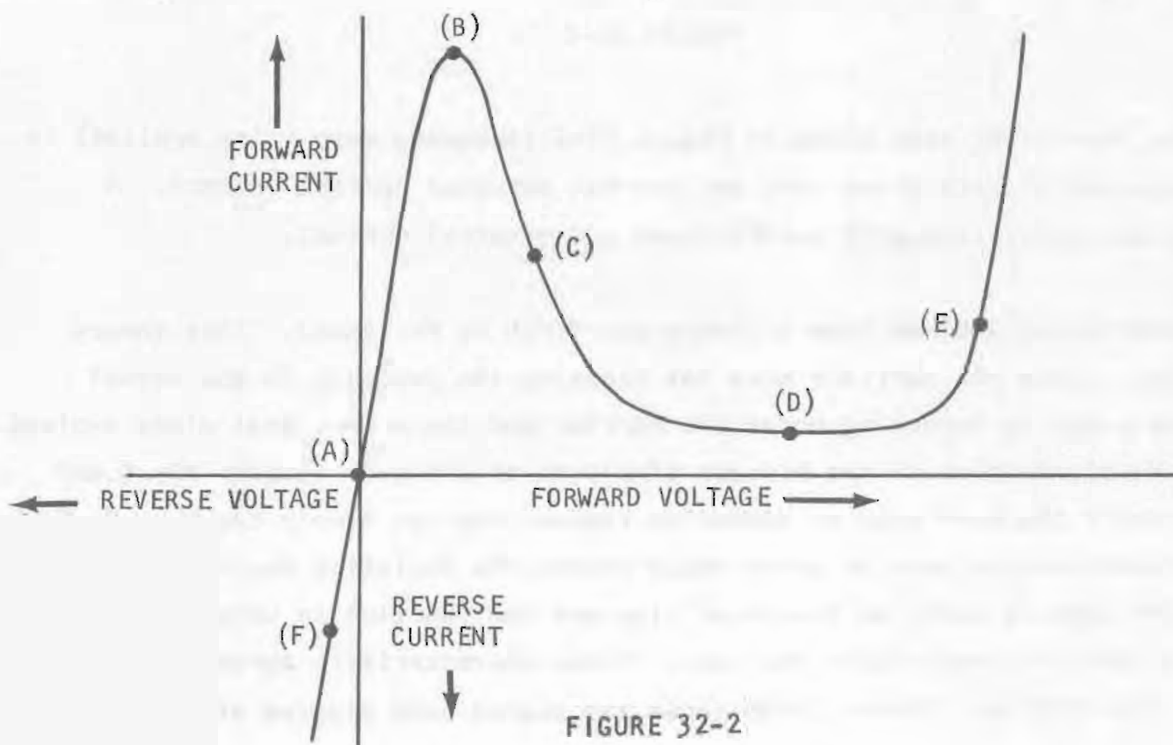


FIGURE 32-2

to move with great speed from the valence band of the P side to the conduction band of the N side and vice versa, instead of moving from conduction band to conduction band.

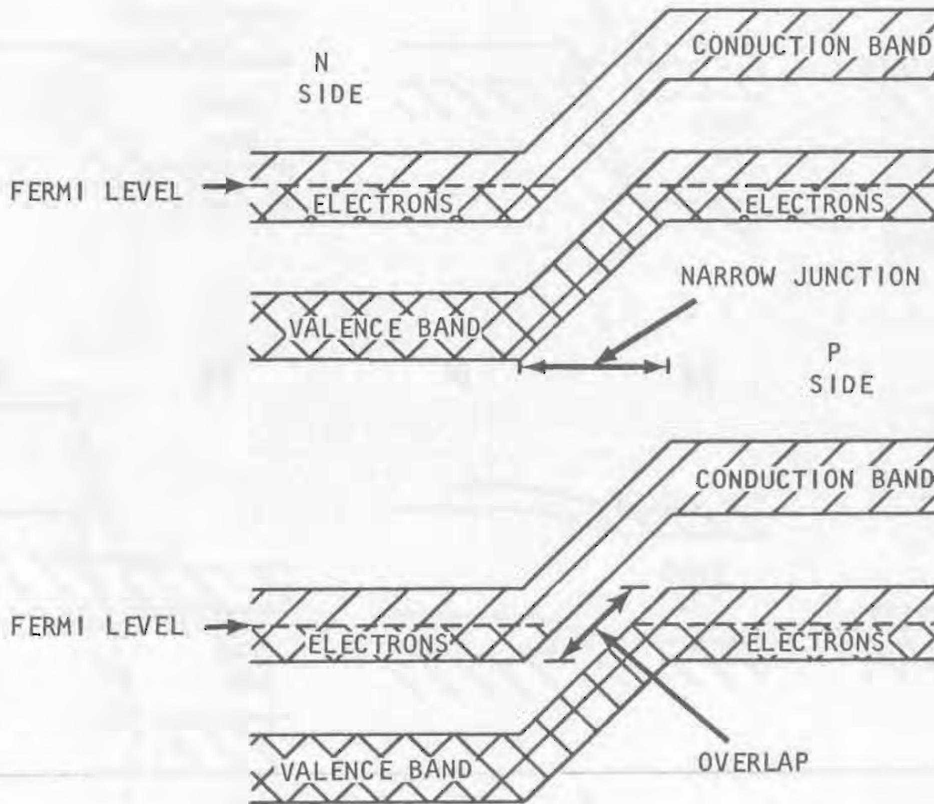


FIGURE 31-2

Any tunneling that might take place in Figure 31-2 (assuming zero volts applied) is by the same amount in both directions and the net external current is zero. A tunnel'diode at equilibrium will exhibit zero net external current.

The name tunnel diode evolved from a theory put forth by Mr. Esaki. This theory indicated that, since the carriers were not crossing the junction in the normal fashion, they might be tunneling under the barrier and the name tunnel diode evolved from this. We might think of the process simply as an arc over between the N and the P side across the band gap, or depletion region. We can simply say that it is possible for carriers to move at great speed across the depletion region if there is a place for them to exist on the other side and the junction is very narrow. With this in mind, we can explain the tunnel diode characteristic curve using the diagrams in Figure 32-2. Figure 32-2A shows the energy band diagram of the tunnel

diode at equilibrium. Note that the fermi level exists in the conduction band of the N side and in the valence band of the P side and when the fermi levels are aligned, there is an overlap of the valence and conduction bands across the junction separated by a narrow transition region. It is possible for carriers to cross the narrow barrier without appearing to have sufficient energy to do so.

Figure 32-2B shows the energy band diagram with a small amount of forward voltage applied. This point on the tunnel diode EI curve is indicated by (B) in the diagram at the bottom of Figure 32-2. The small amount of forward bias has lifted the carriers on the N side to a little higher energy point while lowering the carriers on the P side to a lower energy point, and it is now possible for an increase in tunneling to occur from the N to the P side. The tunnel diode curve at the bottom of Figure 32-2 indicates that with forward bias applied, the current increases in a positive direction.

A further increase in forward voltage starts lifting some of the carriers in the N side opposite the band gap as shown in Figure 32-2C. Since the number of carriers that can tunnel is reduced, the current starts to decrease or move in a negative direction.

A further increase in forward voltage moves most of the carriers opposite the band gap as shown in Figure 32-2D, and the current is reduced to its minimum value.

Continuing to increase the applied forward voltage results in the carriers moving from the conduction band of the N side into the conduction band of the P side and the normal diode curve results as shown in Figure 32-2E.

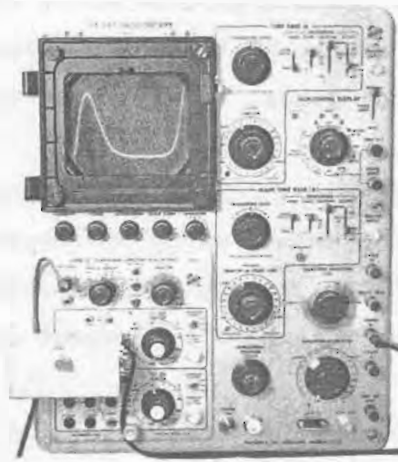
Figure 32-2F shows the energy band diagram with reverse voltage applied indicated by point F on the diode EI curve. The tunneling from the P side to the N side is enhanced when reverse voltage is applied and the current increases in the negative direction.

Solid understanding of quantum mechanics and quantum mechanical tunneling is not necessary to understand the action of a tunnel diode as used in circuitry. Therefore, further discussion will be neglected. It should be noted, however, that the speed of the electrons transition from the N side to the P side and vice versa will

be much faster than the time required for an electron to move from the conduction band of the N side into the conduction band of the P side, diffuse until finding an imperfection, and then recombining. The speed of transport of the carriers across the junction will be set by the speed of tunneling and the speed of the majority carrier transport in the particular type of semiconductor used in construction of a tunnel diode. Gallium arsenide tunnel diodes have the fastest electron mobility of any of the three major types of semiconductors used. In order of their electron mobility, we find gallium arsenide, then germanium, and finally silicon. For speed consideration, we would choose gallium arsenide over the other two types of semiconductors for construction of tunnel diodes. However, other considerations might cause us to choose a different type of semiconductor.

Figure 33-2 shows the voltage versus current curve for an actual tunnel diode. The magnitude of current at the first positive peak in Figure 33-2 is termed peak current. This is marked with an A in Figure 33-2. The current at this point is determined by the geometry of the junction. When tunnel diodes are manufactured, they are etched to size to set the peak current at this point. The amount of current flowing at peak current will determine the total negative resistance of the tunnel diode and also determine the total current change involved in the negative resistance region of the tunnel diodes curve. The current magnitude indicated by the letter B in Figure 33-2 is termed valley current. This is the minimum value of current in the valley between the first and second positive current slopes on the tunnel diodes characteristic curve. This is the point where the bands uncross when forward voltage is applied to the tunnel diode. The voltage at the point of peak current is referred to as peak voltage and is indicated by the C in Figure 33-2. The voltage at the point of valley current is termed valley voltage and is indicated by the D in Figure 33-2. The voltage at the point on the second positive current slope of the tunnel diode curve that is equal in magnitude to peak current is normally referred to as voltage at the forward point, or V_{FP}. This is the voltage at which the current at the second positive slope is equal to peak current and it becomes important when dealing with the tunnel diode as a switch.

Figure 34-2 shows the tunnel diode curve once again, with the points indicated by their symbols (I_p, I_v, V_p, V_v, and V_{FP}). The values of peak voltage, valley voltage, and forward point voltage are determined by the type of semiconductor used in construction of the tunnel diode. Peak current is determined by the geometry of the



TEST SET-UP: TEKTRONIX TYPE 547 OSCILLOSCOPE, TYPE "O" OPERATIONAL AMPLIFIER PLUG-IN WITH TUNNEL DIODE DRIVER ADAPTER*. PHOTO BELOW TAKEN WITH TEKTRONIX TYPE C-12 OSCILLOSCOPE CAMERA.

* NON-PRODUCTION ITEM

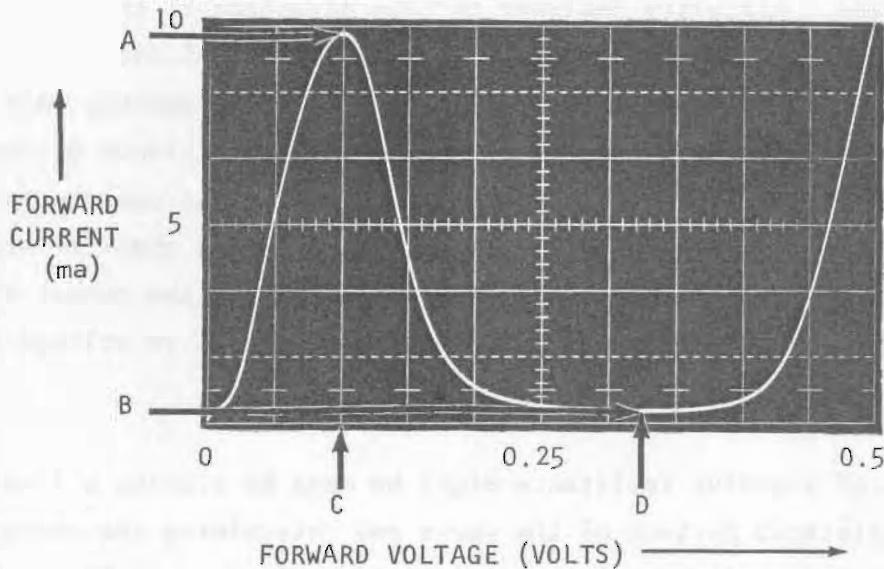


FIGURE 33-2

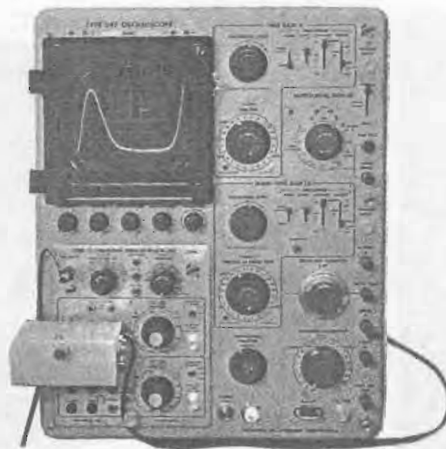
junction. Valley current seems to be a figure of merit of a tunnel diode. The lower the magnitude of valley current, the lower the amount of noise generated in the tunnel diode, and the greater the peak to valley current ratio.

Typical values of voltages for germanium tunnel diodes are peak voltage ≈55mv, valley voltage ≈350mv, and V_{FP} ≈500mv. Typical voltages for a gallium arsenide tunnel diode are peak voltage ≈150mv, valley voltage ≈500mv, and V_{FP} ≈1100mv. These voltages are typical at room temperature or about 25°C (300° Kelvin).

From the curves that have been shown, it can be seen that the tunnel diode will have areas of both positive and negative conductance when operated in the forward bias condition. The area from 0V to peak voltage will exhibit a positive resistance or conductance, the area between peak voltage and valley voltage will exhibit a negative conductance or resistance, and the area beyond valley voltage will once again exhibit a positive conductance or resistance.

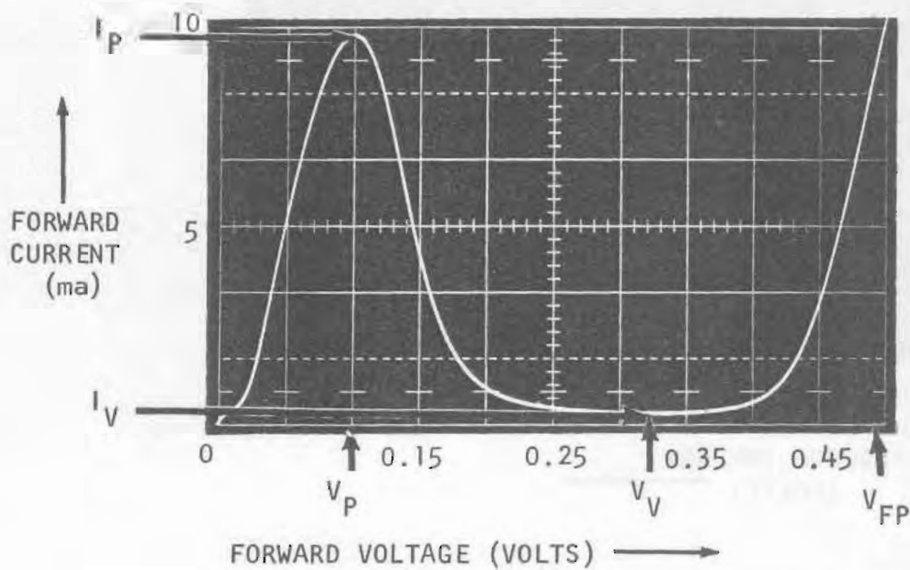
Figure 35-2 shows a test set-up for the measurement of tunnel diode conductance versus applied forward voltage and a double exposure shows this plot of conductance referenced to the tunnel diode voltage versus current curve. Notice that starting at zero, the tunnel diode conductance becomes negative at peak voltage and becomes positive at valley voltage. Circuitry designed to take advantage of the negative conductance characteristic of the tunnel diode must have the tunnel diode operating in the area between peak voltage and valley voltage. We can also express this in terms of current. In order to take advantage of the negative resistance or conductance characteristics of the tunnel diode, the tunnel diode must operate in the area between peak current and valley current. Note in Figure 35-2 that the area of the greatest magnitude of negative conductance is that portion of the tunnel diode's EI curve where the rate of change of negative current versus positive voltage is fastest.

An accurate measurement of negative resistance might be made by placing a line tangent to the negative resistance portion of the curve and determining the change in current for a given change in voltage. A test set-up such as shown in Figure 35-2 is not normally available. If the values of peak current, peak voltage, valley current, and valley voltage are known, these values will not give a measurement of negative conductance. Normally, a factor of two is added when peak and valley



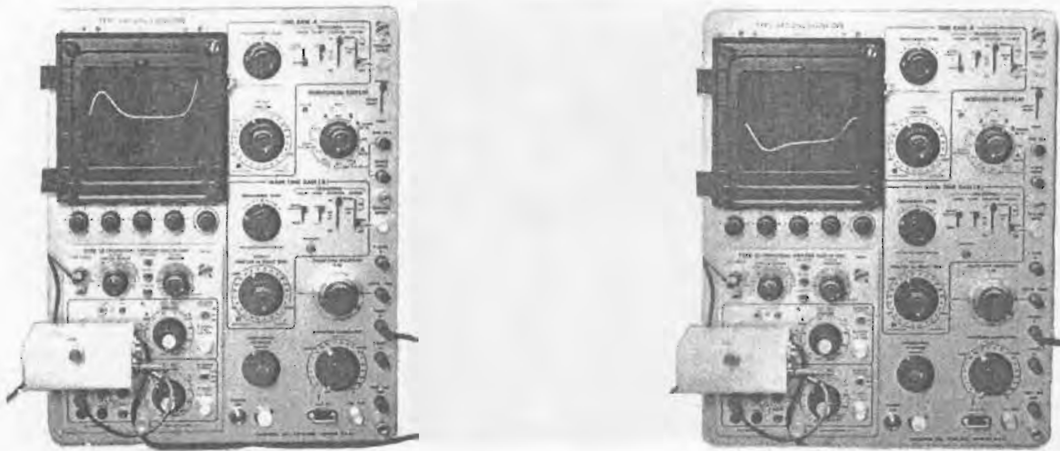
TEST SET-UP: TEKTRONIX TYPE 547 OSCILLOSCOPE, TYPE "O" OPERATIONAL AMPLIFIER PLUG-IN WITH TUNNEL DIODE DRIVER ADAPTER*. PHOTO BELOW TAKEN WITH TEKTRONIX TYPE C-12 OSCILLOSCOPE CAMERA.

* NON-PRODUCTION ITEM



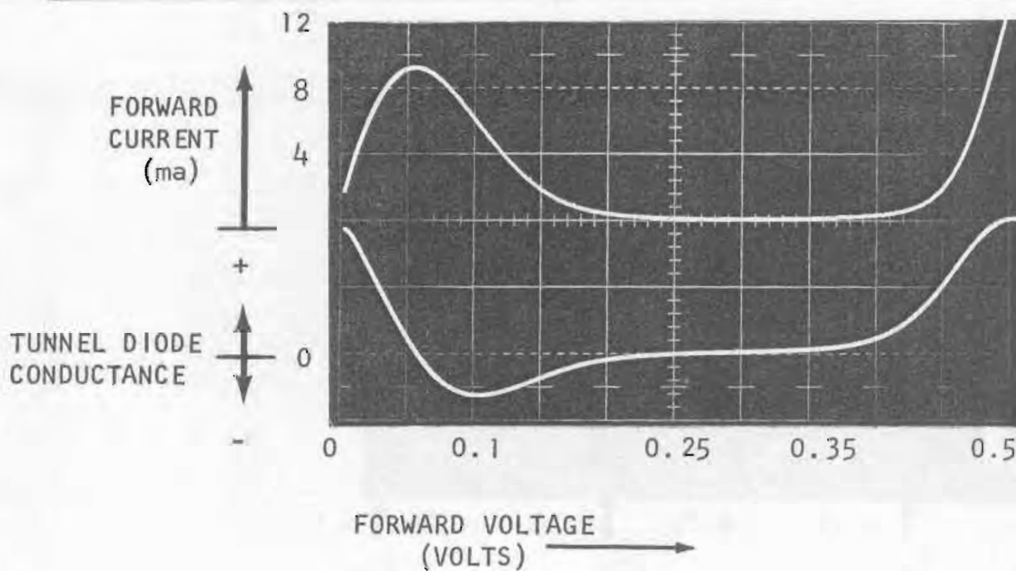
TUNNEL DIODE EI CURVE

FIGURE 34-2



TEST SET-UP: TEKTRONIX TYPE 547 OSCILLOSCOPE, TYPE "O" OPERATIONAL AMPLIFIER PLUG-IN WITH TUNNEL DIODE DRIVER ADAPTER*. PHOTO BELOW TAKEN WITH TEKTRONIX TYPE C-12 OSCILLOSCOPE CAMERA.

* NON-PRODUCTION ITEM



DOUBLE EXPOSURE SHOWING TUNNEL DIODE EI CURVE (TOP) AND CONDUCTANCE CURVE (BOTTOM)

FIGURE 35-2

voltage and currents are used to approximate the negative conductance of a tunnel diode. The formula:

$$-gd \approx \frac{2(i_p - i_v)}{v_p - v_v}$$

This formula gives a close approximation of the negative conductance offered by the tunnel diode. The 2 added in the formula is used to modify the answer since using peak and valley voltage directly will give only about half the negative conductance that the tunnel diode would offer in circuitry.

TUNNEL DIODE AMPLIFIERS:

To illustrate how the tunnel diode might serve as an amplifier, we might start by defining positive and negative resistance. First of all, by definition, a positive resistance dissipates power. Therefore, by definition, a negative resistance will generate power. Operated in the negative resistance region of its characteristic, the tunnel diode will show a negative resistance to a source. If the tunnel diode is properly biased and in suitable circuitry, to take advantage of its negative resistance characteristic, we would expect it to generate power, or more simply, serve as a power amplifier. Let's carry this a bit farther and look at the circuit configuration indicated by Figure 36-2.

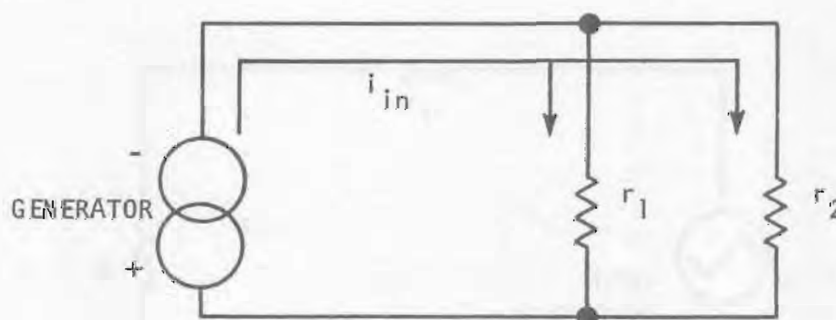


FIGURE 36-2

Note in Figure 36-2 that other than signal considerations have been neglected and the generator is supplying two positive resistances, R1 and R2. If we assume that the generator is a constant current generator (referring to magnitude), the current i_{in} will divide between R1 and R2, and the amount of current that will flow in each resistance is dependent on the amount of resistance of the individual branches. The current for either R1 or R2 will be less than the input current i_{in}, and no current amplification will occur.

Now refer to the diagram in Figure 37-2. The same generator is supplying i_{in} to a negative and a positive resistance in parallel.

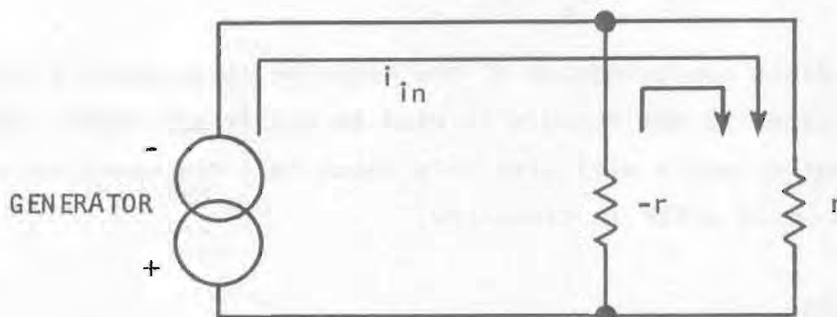


FIGURE 37-2

The negative resistance can supply current to the positive resistance and the positive resistance can have a current that is actually larger than i_{in}. In such a configuration, current gain can be accomplished. The voltage across the parallel branch of the negative resistance and positive resistance is the same; therefore, any power gain will be governed by the amount of current gain obtained.

Now refer to Figure 38-2. An input voltage generator is supplying a negative resistance in series with a positive resistance.

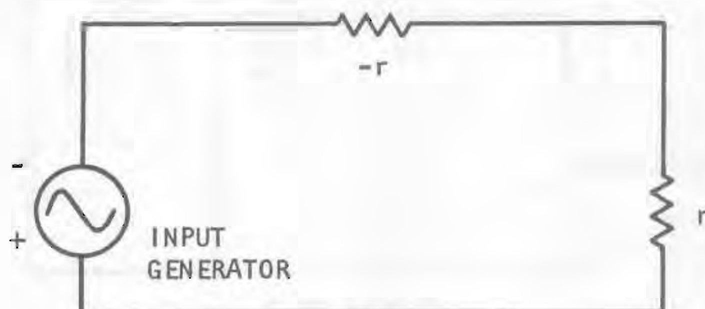


FIGURE 38-2

If we assume that the input generator is a perfect generator, it has zero internal impedance and the voltage across the individual resistances will add to equal the input generator voltage. The input generator will see the algebraic sum of the negative and positive resistances as a total resistance, and the amount of current in the circuit will be governed by this. Assume that the input generator has a

value of 2mv, the negative resistance is -50Ω, and the positive resistance is +48Ω. This is illustrated in Figure 39-2.

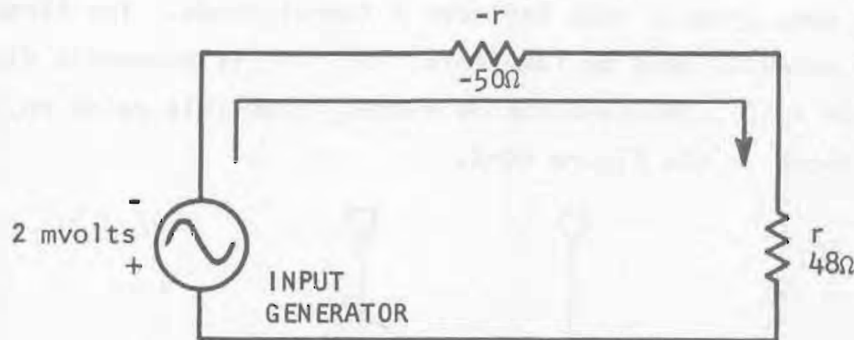


FIGURE 39-2

The input generator sees a negative resistance of -50Ω and a positive resistance of +48Ω, or an equivalent series resistance of -2Ω. By Ohm's law, the current is equal to 2mv divided by negative 2Ω, or a -1ma of current. The product of a negative 1ma of current and the -50Ω will give a positive 50mv across the negative resistance. The product of a negative 1ma and the positive 48Ω will give a negative 48mv across the positive resistance. The algebraic sum of the two voltages is, of course, the original input voltage of 2mv. However, if the output is taken across the negative resistance, we would find an output of +50mv. If the output is taken across the positive resistance, the output is a -48mv. The gain in the latter case is -48mv divided by 2mv $\left(\frac{e_{out}}{e_{in}}\right)$, or a voltage gain of 24.

Since the current through the generator is the same as the current through the negative and positive resistance, the current gain is unity, and the power gain is proportional to the magnitude of voltage gain.

From these examples, it can be seen that a negative resistance placed in series with a positive resistance and a generator can offer voltage and power gain. A negative resistance placed in shunt with an input generator and a positive resistance can offer current and power gain. Since we have already determined that a tunnel diode can exhibit a negative resistance characteristic if operated in the proper portion of its EI curve, a tunnel diode might be used as a negative resistance in these examples. We can accomplish amplification with a tunnel diode. Amplification is normally limited to those devices with three or more active terminals, while the tunnel diode allows amplification with a two-terminal device.

TUNNEL DIODE SYMBOLS:

Figure 40-2 shows some symbols that indicate a tunnel diode. The first symbol listed is the one normally used by Tektronix, Inc. in its schematic diagrams. The other symbols shown also indicate a tunnel diode. From this point on, we will use the first symbol shown in the Figure 40-2.

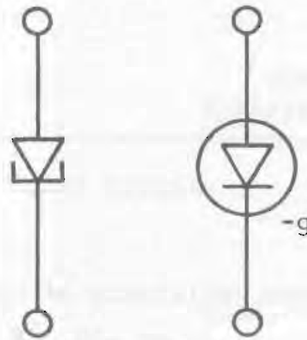


FIGURE 40-2

THE TUNNEL DIODE AS A CURRENT AMPLIFIER:

When a negative resistance is placed in parallel with a positive resistance and driven by a generator, current gain can be accomplished. Figure 41-2 shows the voltage versus current curve for the tunnel diode with a dashed line, the voltage versus current curve for a parallel resistance with the dotted line, and the resultant voltage versus current curve of the two in parallel by the solid line. It is a requirement for linear amplification that the resultant curve shown in Figure 41-2 be positive at all points. This means that the positive conductance, of necessity, must be predominate when a resistor is placed in parallel with a negative resistance or a tunnel diode in a linear amplifier configuration. If the negative conductance is predominate, resultant negative conductance is exhibited by the circuit and oscillations or switching results.

Since conductance is the reciprocal of resistance, it is nececessary that the positive resistance in a linear tunnel diode amplifier configuration be less than the negative resistance of the tunnel diode. If this is so, the positive conductance will be predominate and the resultant curve will be positive at all points. If the tunnel diode is properly biased in its negative conductance region and placed in

FORWARD CURRENT (ma)

20

10

5

A

B

$\Delta I \approx$

C

D

FORWARD VOLTAGE

0.1 0.2 0.3 0.4 0.5 (VOLTS)

$\Delta V \approx 0.1$ VOLTS

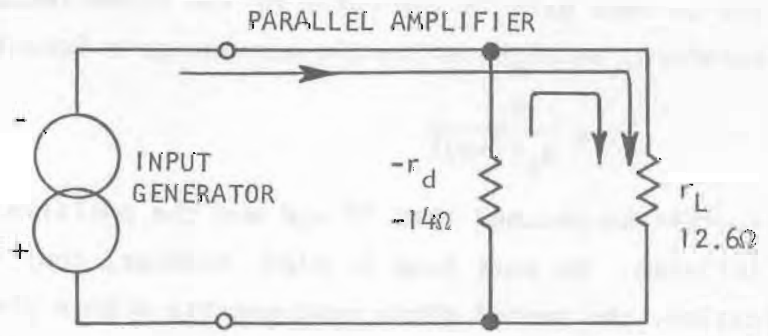


FIGURE 41-2

a suitable parallel configuration with a positive resistance that is less than the diodes negative resistance, linear amplification can be accomplished. A configuration such as shown in Figure 42-2 might be used to bias the tunnel diode quiescently.

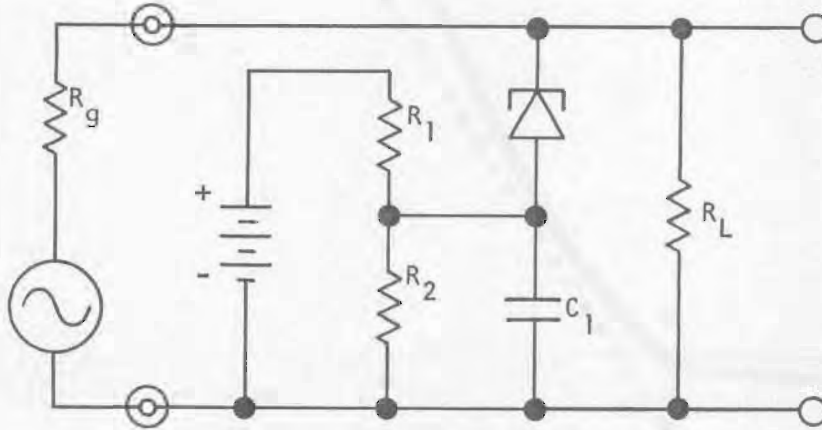
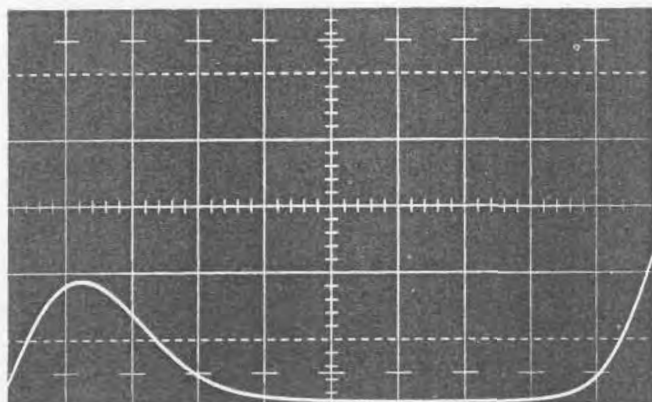


FIGURE 42-2

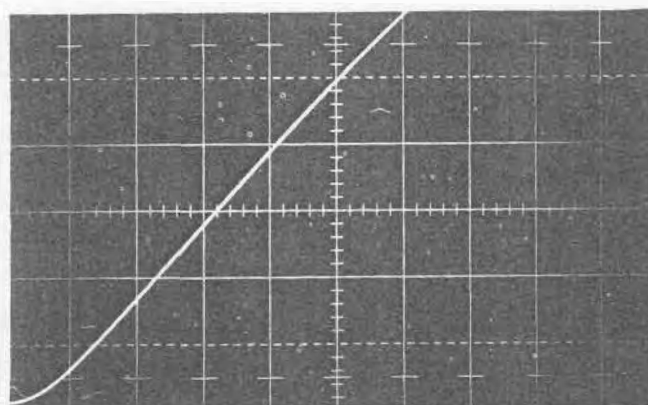
The total conductance seen by the generator in Figure 41-2 will be the algebraic sum of the negative and positive conductances. In the case of the circuit in Figure 41-2, the negative conductance of the tunnel diode is -71,400μmhos (the reciprocal of -14Ω), and the positive conductance of the resistor R_L is 79,400μmhos (the reciprocal of 12.6Ω). The algebraic sum of these two conductances is 8,000μmhos. The input generator then sees 8,000μmhos as an equivalent conductance. This is a resistance of 125Ω. Since the predominate conductance is positive, the resistance is a positive 125Ω. We can formulate an input resistance equation for a parallel amplifier indicating that the input resistance is equal to the reciprocal of the algebraic sum of the conductances, or $r_{input} = \frac{1}{g_L+(-gd)}$. The current gain in this configuration is the ratio of the current flowing in R_L to the input current. We can also say that the current gain is the ratio of the conductance of R_L to the input conductance; therefore, we might write the current gain formula as:

$$A_i = \frac{g_L}{g_L+(-gd)}$$

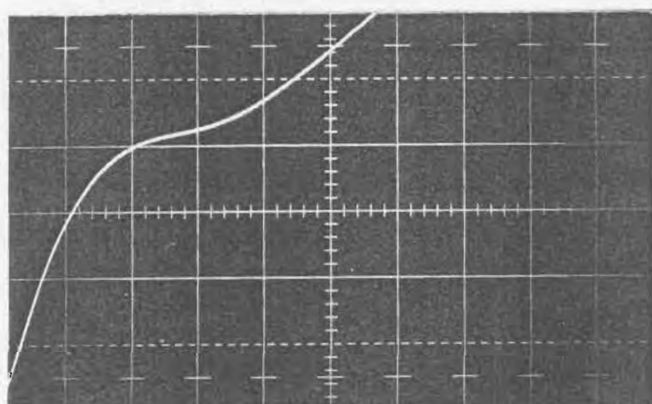
At first glance, it might be assumed that if -gd and the positive g_L were equal, the gain would be infinite. We must keep in mind, however, that in order to accomplish amplification, the tunnel diode must operate within the negative resistance or conductance region of its characteristic curve. The total current swing



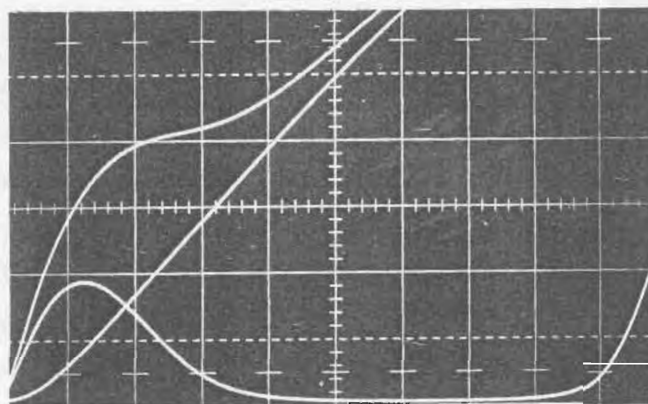
TUNNEL DIODE EI CURVE



RESISTOR EI CURVE



RESULTANT OF RESISTOR AND
TUNNEL DIODE IN PARALLEL



TRIPLE EXPOSURE COMPARING
THE THREE CURVES

TEST SET-UP: TEKTRONIX TYPE 547 OSCILLOSCOPE, TYPE "O" OPERATIONAL AMPLIFIER PLUG-IN WITH TUNNEL DIODE DRIVER ADAPTER* AND TYPE C-12 OSCILLOSCOPE CAMERA

* NON-PRODUCTION ITEM

NOTE: The non-linearity of the resistor EI characteristic at the start (left hand) side is due to using the oscilloscope sweep waveform to drive the adapter. The sweep starts below ground and a diode in the adapter disconnects during this period. The non-linearity at the end (right hand) of the resistor EI characteristic is due to limited current in the adapter power supply. The area of interest is linear, however.

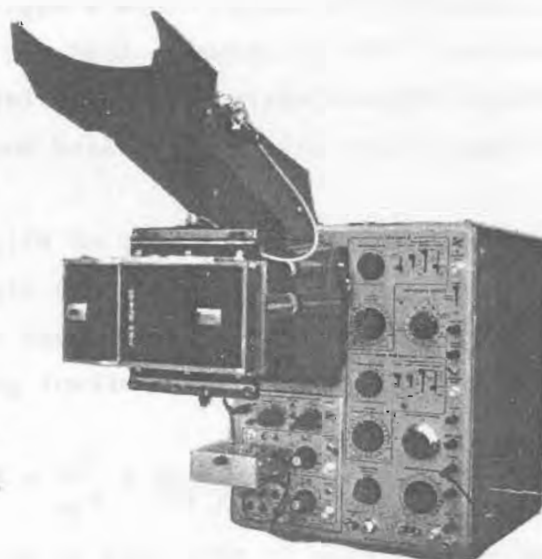


FIGURE 43-2

is, therefore, limited between peak current and valley current. It is also a requirement that the load resistance be less than the absolute value of the negative resistance of a tunnel diode. There are other requirements that will be discussed later.

Figure 43-2 illustrates the combination of a parallel positive and negative resistance. The first diagram in Figure 43-2 shows the tunnel diode voltage versus current curve. The apparent non-linearity at the start of the resistor EI curve (lower left hand corner) is due to using the oscilloscope sweep waveform to drive the tunnel diode. There is a diode in the tunnel diode adapter that disconnects the sweep during the first portion since the sweep waveform starts below ground. At the right hand side, or the end of the resistor voltage versus current curve, there tends to be a bit of non-linearity and this is due to the limited current in the power supply in the driver adapter that is being used in the test set-up. The area of interest is in the negative resistance region of the tunnel diode curve, and the resistor curve is linear in this area. We will neglect the non-linear portions of the curve. The third diagram in Figure 43-2 shows the resultant of the combination of the positive and negative conductances and in this case, the positive resistance is less than the negative resistance of the tunnel diode and the resultant curve is positive at all points. The triple exposure compares the three curves. A tunnel diode in this configuration properly biased in the negative conductance region with a parallel positive resistance having a lower value than the absolute value of the tunnel diode's negative resistance can be used as an amplifier configuration. This is assuming that too much reactance is not added and results in the tunnel diode breaking into oscillations. The stability requirements of the tunnel diode circuit will be discussed shortly.

Let's review just a moment. First of all, the input <u>conductance</u> to a parallel tunnel diode amplifier is equal to the algebraic sum of the negative and positive <u>conductances</u>. The current gain is equal to the load conductance divided by the input conductance; therefore, the current gain is equal to:

$$\frac{g_L}{g_L+(-gd)} = \frac{g_L}{g_{in}} = A_i = \text{current gain}$$

Since the voltage gain in this case is unity(if no series resistance is in the circuit to drop a voltage), the power gain is proportional to the current gain.

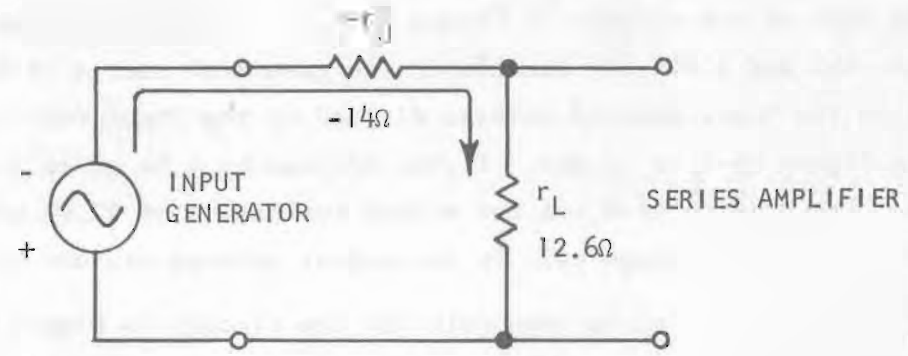


FIGURE 44-2

We can assume that the tunnel diode will be limited primarily to high frequency operation, the reason being that the layout for low frequency operation adds sufficient inductance and capacitance to cause the tunnel diode to break into oscillations.

THE TUNNEL DIODE AS A VOLTAGE AMPLIFIER:

When a negative resistance is placed in series with a positive resistance, voltage gain is made possible. We have already discussed that the negative resistance of the tunnel diode can be used in circuitry and accomplish voltage amplification and current amplification.

Figure 44-2 shows a negative resistance in series with a positive resistance driven by a voltage generator. The voltage versus current curve for the tunnel diode is shown with a dashed line, the voltage versus current curve for the resistor is shown with a dotted line, and the resultant curve is shown by a solid line. The generator sees an input resistance that is the algebraic sum of the series resistances, or $r_{input} = r_L + (-r_d)$. The current is equal to the applied input generator voltage divided by the input resistance. The voltage across R_L is the product of the circuit current and the value of the load resistance R_L. Since the input voltage is the product of the circuit current and the input resistance and the voltage across R_L is the product of R_L and the circuit current, the voltage gain can be expressed as the ratio of the load resistance to the input resistance. In formula form:

$$A_v = \frac{r_L}{r_{in}} = \frac{r_L}{r_L+(-r_d)}$$

In the case of the circuit in Figure 44-2, the input resistance is the algebraic sum of the -14Ω and a +12.6Ω; therefore, the generator sees a -1.4Ω. The current is equal to the input applied voltage divided by the input resistance. The input voltage in Figure 44-2 is 11.9mv. 11.9mv divided by -1.4Ω gives a circuit current of -8.5ma. The product of -8.4ma and a load resistance of 12.6Ω gives an output voltage of -107.1mv. The voltage gain is the output voltage divided by the input voltage, or $\frac{-107.1mv}{11.9mv}$, or 9. The voltage gain for the circuit in Figure 44-2 is 9. Expressing the voltage gain in terms of resistance, we might say that 12.6Ω (the resistance of the load resistor), divided by 1.4Ω (the input resistance), is equal to 9, and we

have validated our gain formula. Notice that the total conductance offered by the amplifier configuration must still be positive; therefore, the positive resistance must be less than the absolute value of the diode's negative resistance. The other stability criteria is also the same as for the parallel amplifier configuration. Since the current gain is unity, the power gain will be proportional to the voltage gain and the circuit will be limited to high frequency operation. From the voltage gain formula, it would at first seem that the gain would be infinity if $-r_d$ were equal to r_L. However, we must remember that the total voltage swing must be kept within the negative resistance portion of the diode's characteristic curve. The total voltage swing is limited between peak voltage and valley voltage. The maximum available power from the tunnel diode can be approximated from the following formulas:

$$P_{max(\text{voltage amplifier})} \approx \frac{|-g|(V_v - V_p)^2}{8}$$

and

$$P_{max(\text{current amplifier})} \approx \frac{(I_p - I_v)^2}{8\,|-g|}$$

These formulas take into account that we must operate within the negative resistance region of the tunnel diode's characteristic in order to serve as an amplifier. I_p is the peak point current, I_v is the valley current, V_v is the valley voltage, V_p is the peak voltage, and $|-g|$ is the absolute value of the negative conductance of the tunnel diode. We are using valley voltage and peak voltage, valley current and peak current, and the 8 in the formula gives an approximate answer that is close to the maximum power that the tunnel diode can supply a load. Recall that we modified the formula for negative conductance with a factor of 2. This same factor must be applied when dealing with peak and valley voltage and current points to determine the power that can be supplied to a load. Dealing with power, the factor becomes squared or 4 and then even optimum power transfer provides only a 50% efficient circuit and the factor becomes 8.

HIGH FREQUENCY LIMITATIONS:

There are two main high frequency limitations of the tunnel diode. One is the frequency at which the tunnel diode no longer exhibits a negative resistance characteristic and this is termed the resistive cut-off frequency. The other is the self-resonant frequency of the tunnel diode determined by the reactive components

associated with the diode and its encapsulation and any connecting leads. The resistive cut-off frequency is given the symbol f_{rco}, and is given by the formula:

$$f_{rco} = \frac{|-g|}{2\pi C}\sqrt{\frac{1}{|-g|\,R_s} - 1}$$

where C is the junction capacity, $|-g|$ is the absolute value of the diode's negative conductance, and R_s is the series resistance of the diode's leads.

The reactive cut-off frequency f_{xco} is given by the formula:

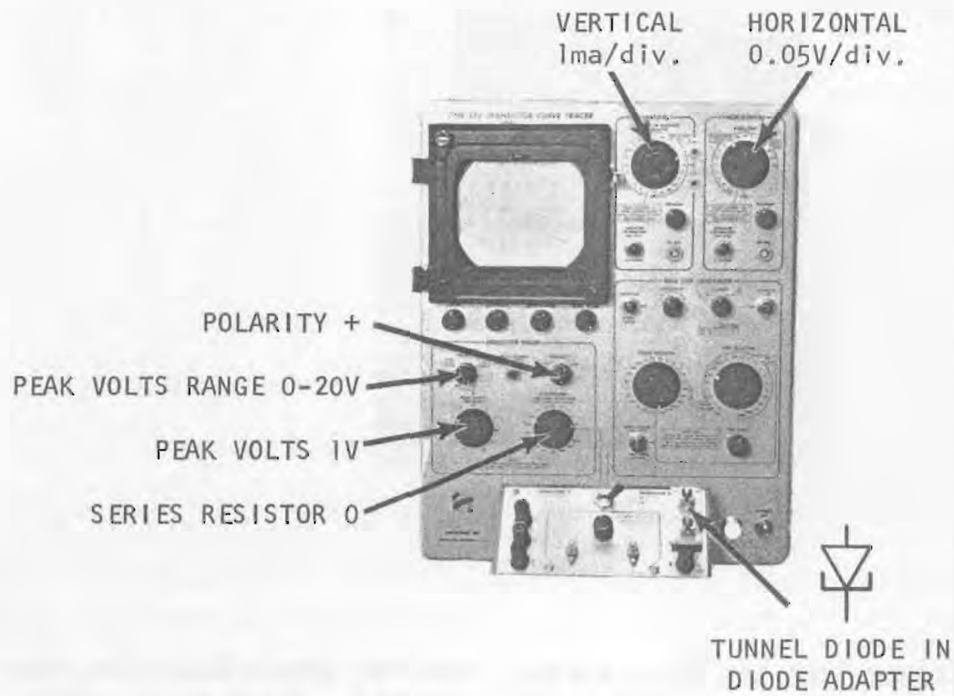
$$f_{xco} = \frac{1}{2\pi}\sqrt{\frac{1}{L_s C} - \left(\frac{|-g|}{C}\right)^2}$$

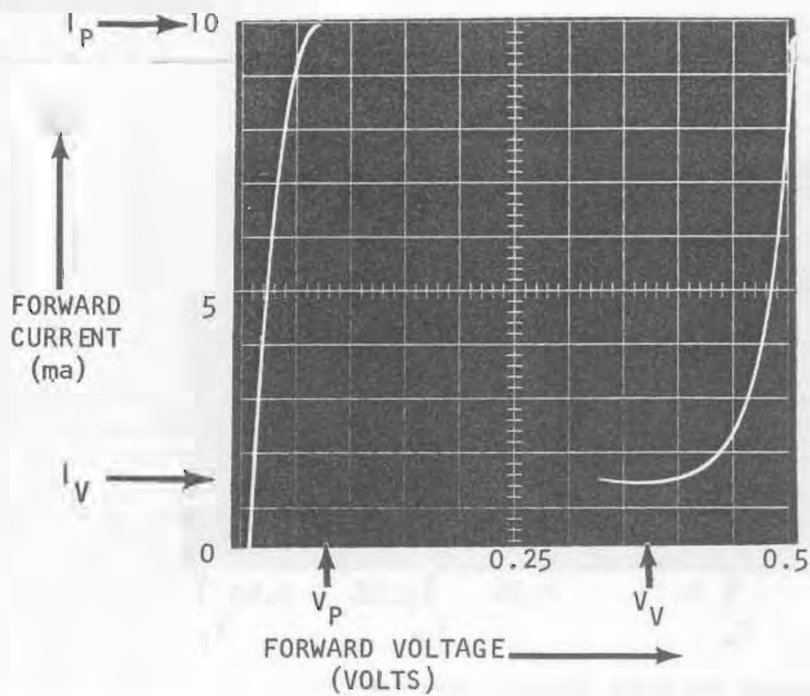
where C is the total capacity, L_s is the lead inductance, and $|-g|$ is the absolute value of the diode's negative conductance.

MEASURING TUNNEL DIODES WITH THE TEKTRONIX TYPE 575 TRANSISTOR CURVE TRACER:

Figure 45-2 shows a typical tunnel diode curve that might be observed on the Tektronix Type 575 Transistor-Curve Tracer. Note that the negative resistance region of the curve is not visible. Figure 46-2 shows the same tunnel diode when placed in a strip-line environment and observed with the Tektronix Type O Operational Amplifier plug-in and a tunnel diode driver adapter in a Type 547 oscilloscope. Note the difference in the curve in Figures 45-2 and 46-2. The main reason for the difference is the resistance associated with the circuitry in the 575 Transistor-Curve Tracer. The series resistance in the Tektronix Transistor Curve Tracer causes the tunnel diode to act as a switch in most measurement situations. In other words, the total value of positive resistance is greater than the absolute value of the tunnel diode's negative resistance, and the tunnel diode cannot enter its negative resistance region. Interpretation of the display can be misleading. For instance, calculation of the tunnel diode's negative resistance can be off by a factor of 2 when using the display in Figure 45-2. As a result of this, the formula

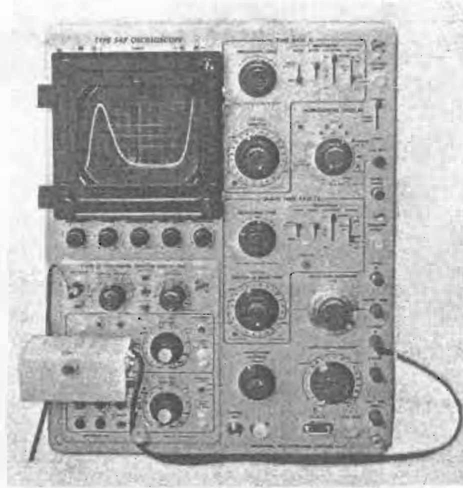
$$-r_d \approx \frac{V_p - V_v}{2(I_p - I_v)}$$





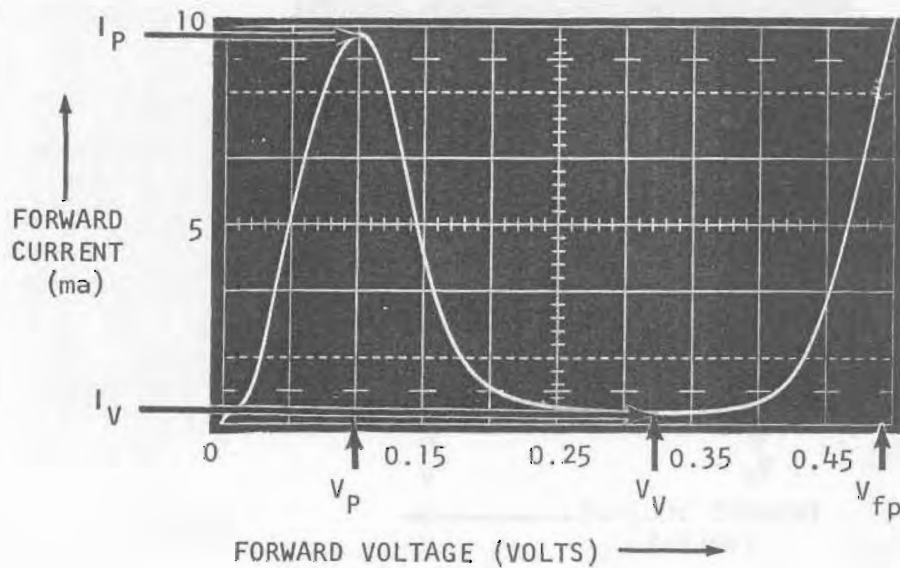
TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER WITH DIODE ADAPTER. PHOTO ABOVE TAKEN WITH TEKTRONIX TYPE C-12 OSCILLOSCOPE CAMERA.

FIGURE 45-2



TEST SET-UP: TEKTRONIX TYPE 547 OSCILLOSCOPE, TYPE "O" OPERATIONAL AMPLIFIER PLUG-IN WITH TUNNEL DIODE DRIVER ADAPTER*. PHOTO BELOW TAKEN WITH TEKTRONIX TYPE C-12 OSCILLOSCOPE CAMERA.

* NON-PRODUCTION ITEM



TUNNEL DIODE EI CURVE

FIGURE 46-2

can be used to determine the approximate negative resistance of a tunnel diode when using the Type 575 Transistor-Curve Tracer. For the tunnel diode in Figure 45-2, peak current is about 10ma, valley current is about 1.5ma, peak voltage is about 0.075V, and valley voltage is about 0.375V. Using the formula for negative resistance, the results are:

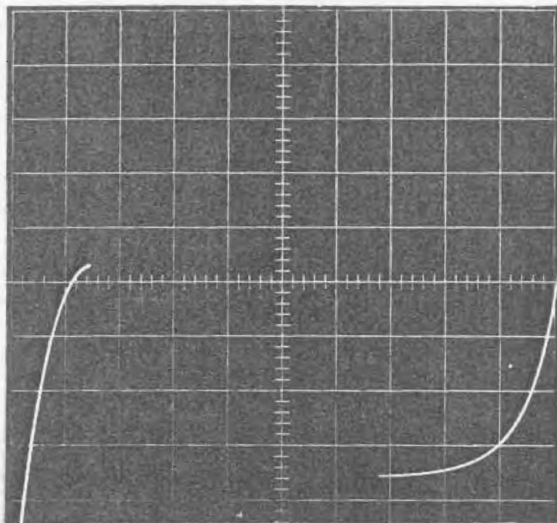
$$-r_d \approx \frac{V_p - V_v}{2(I_p - I_v)} \approx \frac{0.075V - 0.375V}{2(10ma - 1.5ma)} \approx -17.6\Omega$$

and solving for negative conductance, we simply take the reciprocal of the formula and the result is:

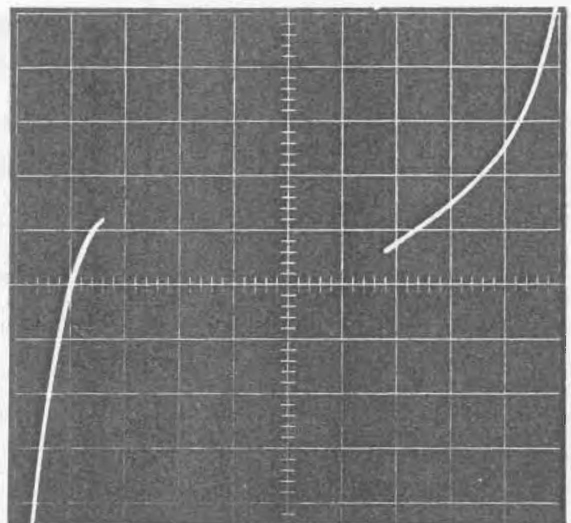
$$-g_d \approx \frac{2(I_p - I_v)}{V_p - V_v} \approx \frac{2(10ma - 1.5ma)}{0.075V - 0.375V} \approx -56{,}666\mu mhos$$

This approximates the negative resistance of a tunnel diode in Figure 45-2 as a negative 17.6Ω and the negative conductance at approximately -56,666μmhos.

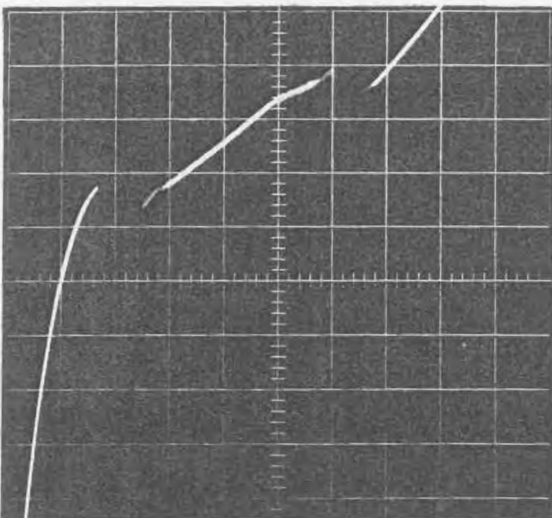
Figure 47-2 shows a method by which the negative resistance can be measured with the Type 575 Transistor Curve Tracer. With this method, a variable resistor is placed in parallel with the tunnel diode, and its resistance adjusted. The resistor is started at a high value so that it has very little effect on the tunnel diode curve, and then its value is slowly reduced. Note from the measurements in Figure 47-2 that, as the resistance is reduced, the positive conductance of the variable resistor becomes more and more pronounced in the equivalent shown on the cathode ray tube. When R_L is very much greater than the negative resistance of the tunnel diode, the tunnel diode operates in a switching mode. Notice that the tunnel diode will continue to operate as a switch and will not enter the negative resistance region until the variable resistor is adjusted so that it is just slightly less than the absolute value of the tunnel diode's negative resistance. When this occurs, the tunnel diode will stop switching to its second positive slope and will show a positive conductance curve at all points. There might be a little bit of oscillation associated in the display on the CRT face; however, the point at which the tunnel diode stops switching to its second positive slope will be very pronounced. At the point that the tunnel diode stops switching to its second positive slope, the tunnel diode's negative resistance is slightly greater than the positive resistance of the variable resistor. The 575 is turned off and the variable resistor removed and measured. The positive resistance of the variable resistor will be slightly less



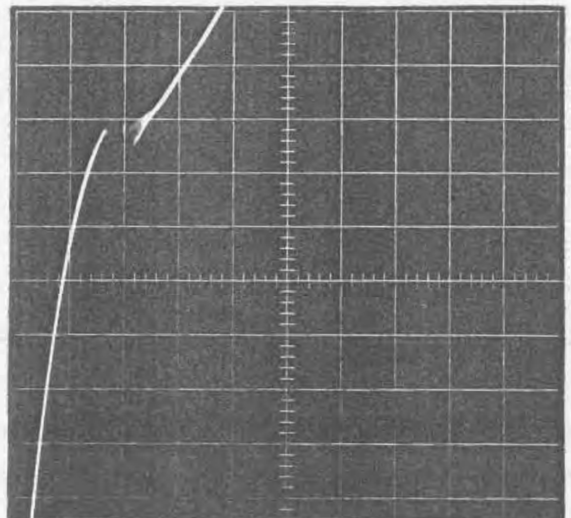
1. r_L VERY MUCH GREATER THAN $-r_d$



2. r_L MUCH GREATER THAN $-r_d$

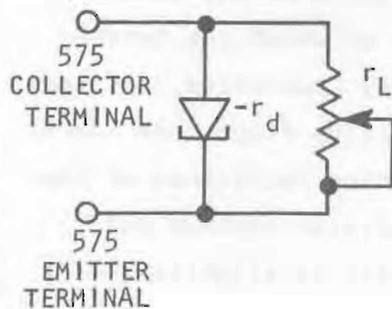


3. r_L GREATER THAN $-r_d$



4. r_L SLIGHTLY LESS THAN $-r_d$

TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER WITH TYPE C-12 CAMERA AND CIRCUIT SHOWN BELOW.



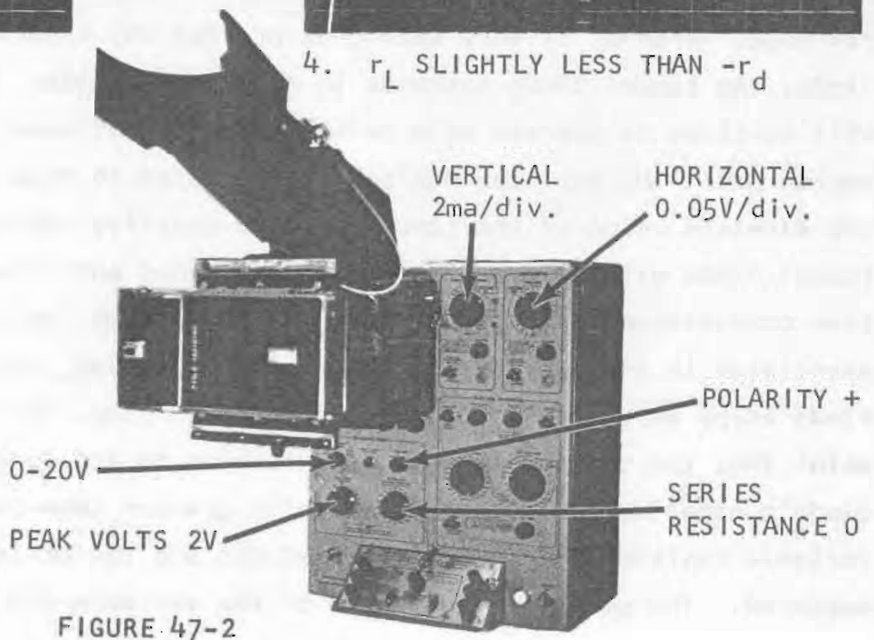


FIGURE 47-2

than the negative resistance of the tunnel diode. In the measurement set up in Figure 47-2, the negative resistance of the tunnel diode was measured as -14Ω. You will recall that we approximated the negative resistance of this tunnel diode at -17.6Ω.

If the 575 is to be used to check tunnel diodes for possible failures or for characteristics that will cause it to malfunction in a circuit, a known good tunnel diode can be observed and the curve recorded. When a tunnel diode is removed from a circuit and checked, deviations from the known good tunnel diode curve can be used as an indication of a fault. A significant change in the peak to valley ratio of the unknown tunnel diode when compared to the known tunnel diode might be an indication of a loss of the tunnel diode's characteristics. Use care because oscillations occuring can be the result of the measurement set-up. However, if the known good tunnel diode does not offer oscillations in the measurement set-up, it is well to expect the unknown tunnel diode to perform satisfactorily in the same measurement set-up.

For more precision measurements of the tunnel diode, a strip line or transmission line environment must be observed. For many fast tunnel diodes, the reactance associated with the measurement set-up will cause it to break into oscillations. By making the tunnel diode a part of a transmission line where the reactances make up part of the transmission system, the tunnel diode can be observed without oscillations. It might even be necessary with very fast tunnel diodes to make them a part of a transmission line in order to measure them and use a little different approach than with devices that do not exhibit a negative resistance characteristic.

THE TUNNEL DIODE AS A SWITCH:

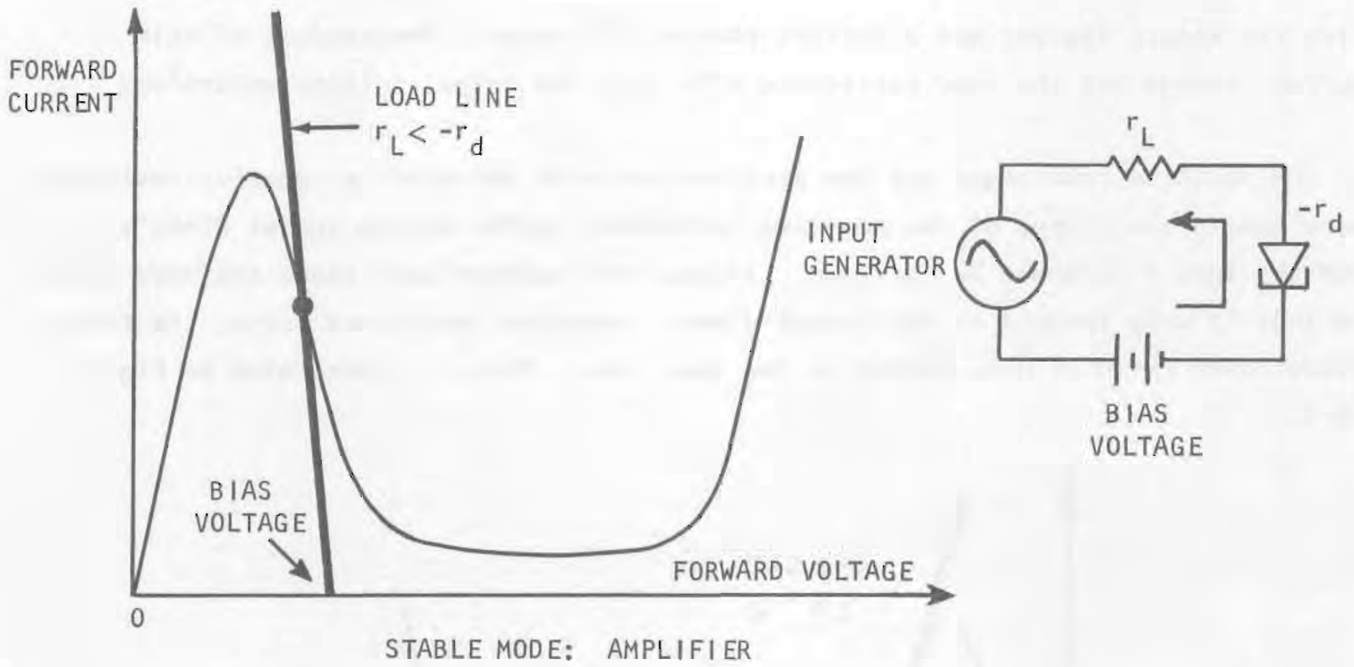
The 575 Transistor-Curve Tracer does not allow the tunnel diode to enter the negative resistance region of its voltage versus current curve. (The 575 resistance is greater than the tunnel diode's negative resistance.) The criteria for the tunnel diode to serve as a switch is that the positive resistance in the circuit be greater than the tunnel diode's negative resistance. Of course, proper biasing and circuitry

are also involved.

Figure 48-2 shows the voltage versus current curve for a tunnel diode with a load line for amplifier action and a load line for switching action. Note that the load line with the positive resistance less than the absolute value of the tunnel diode's negative resistance and the magnitude of the bias voltage shown in Figure 48-2A has the load line intersecting the tunnel diode curve at only one point. In this case, it intersects in the negative resistance region. Of course, this is necessary for amplifier action. You will note that the slope of the load line is steeper than the slope of the tunnel diode's negative resistance curve. If the load line were not steeper, it would be impossible for it to intersect the negative resistance region at only one point and not also intersect the positive resistance regions.

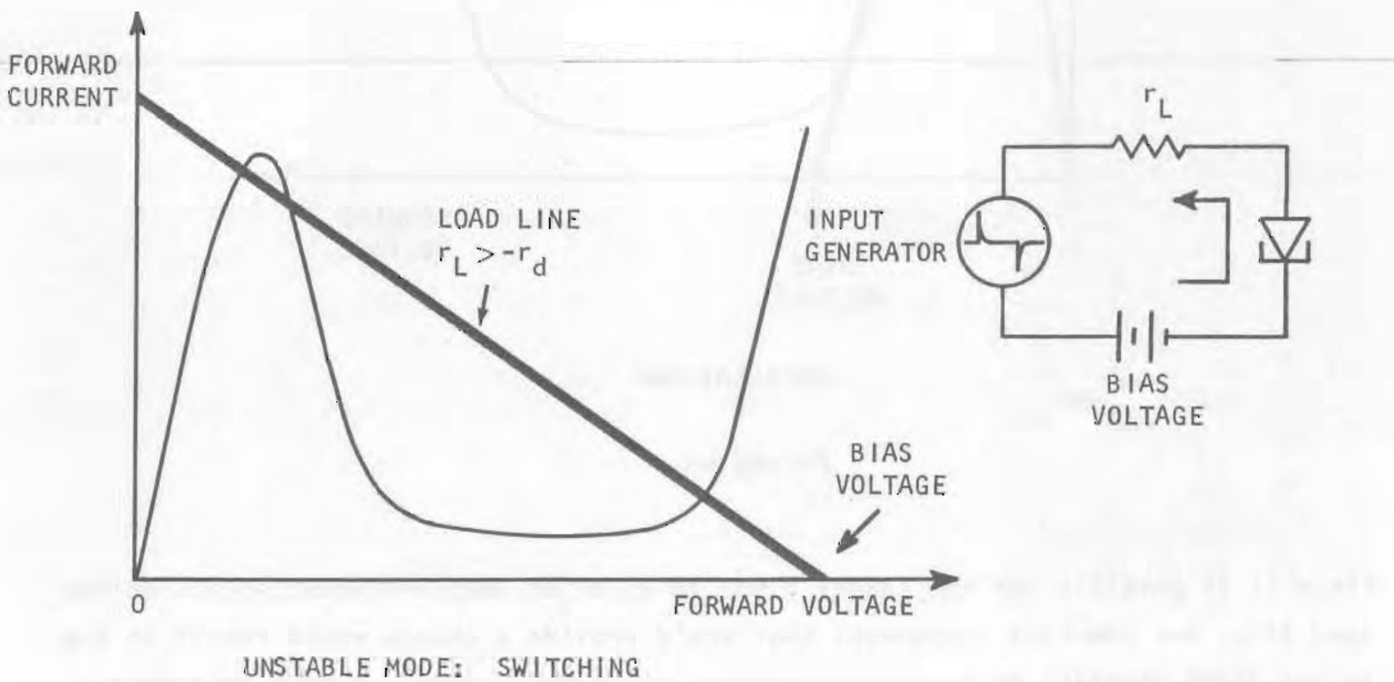
Figure 48-2B has the load line intersecting the tunnel diode curve at more than one point, and with the bias voltage shown, the tunnel diode could exist at any one of three points on the load line. The tunnel diode will have to exist somewhere on the load line. In Figure 48-2B, it could exist on the first positive slope where the load line intersects, in the negative resistance region where the load line intersects, or on the second positive slope where the load line intersects. With the bias voltage shown, however, the tunnel diode will exist on the first positive slope at its intersection with the load line if no external energy is applied to cause it to change resting points. If the input generator increases the bias voltage so that the load line no longer intersected the first positive slope, the tunnel diode will attempt to change to a point on the load line where it can exist. When the load line is moved as a result of an increase in the bias voltage by the input generator, it moves to a point where the tunnel diode cannot exist on the first positive slope. The load line no longer intersects the negative resistance area and it is impossible for the tunnel diode to enter the negative resistance region. The tunnel diode will simply switch to the second positive slope where it can rest once again on the load line. This is the reason for a linear amplifier design criteria to be that the positive resistance must be less than the absolute value of the tunnel diode's negative resistance. If not, it is impossible for the tunnel diode to serve as a linear amplifier.

When serving as an amplifier in a series configuration such as shown in Figure 48-2A, the input voltage will vary the bias voltage, the load line will move in the nega-



STABLE MODE: AMPLIFIER

FIGURE 48-2A

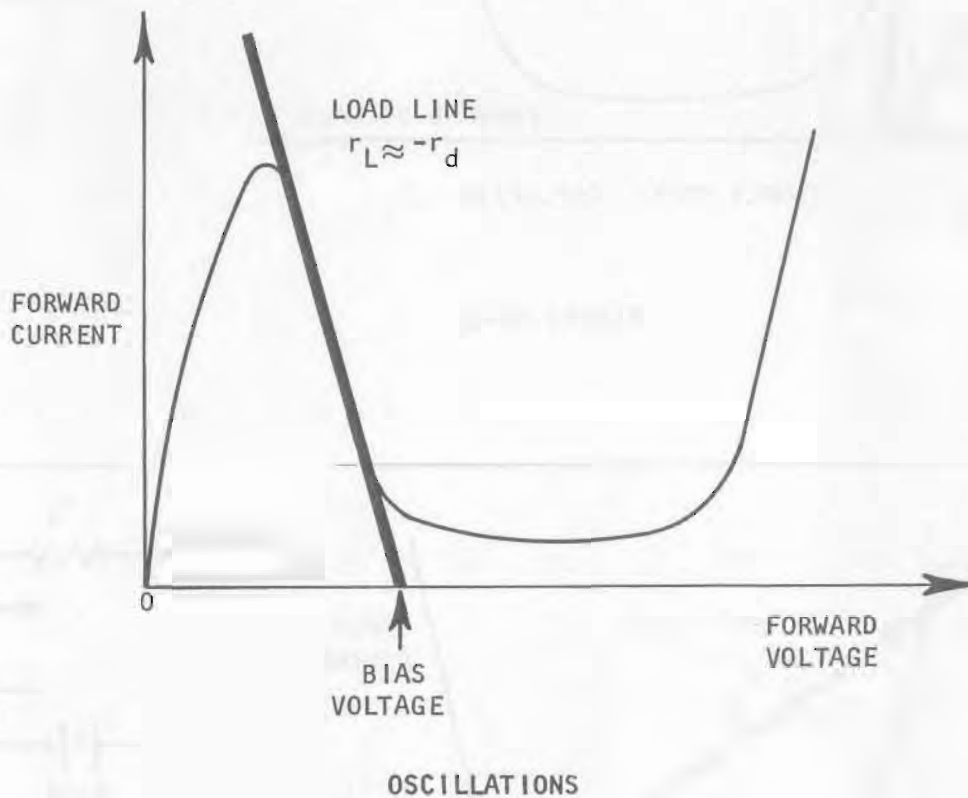


UNSTABLE MODE: SWITCHING

FIGURE 48-2B

tive resistance region, and a current change will occur. The product of this current change and the load resistance will give the output voltage excursion.

If the positive resistance and the absolute value of the diode's negative resistance were equal, the slopes of the negative resistance region of the tunnel diode's curve and the load line would be the same. If the bias voltage positioned the load line so that it were tangent to the tunnel diode's negative resistance curve, the tunnel diode could exist at many points on the load line. This is illustrated in Figure 49-2.



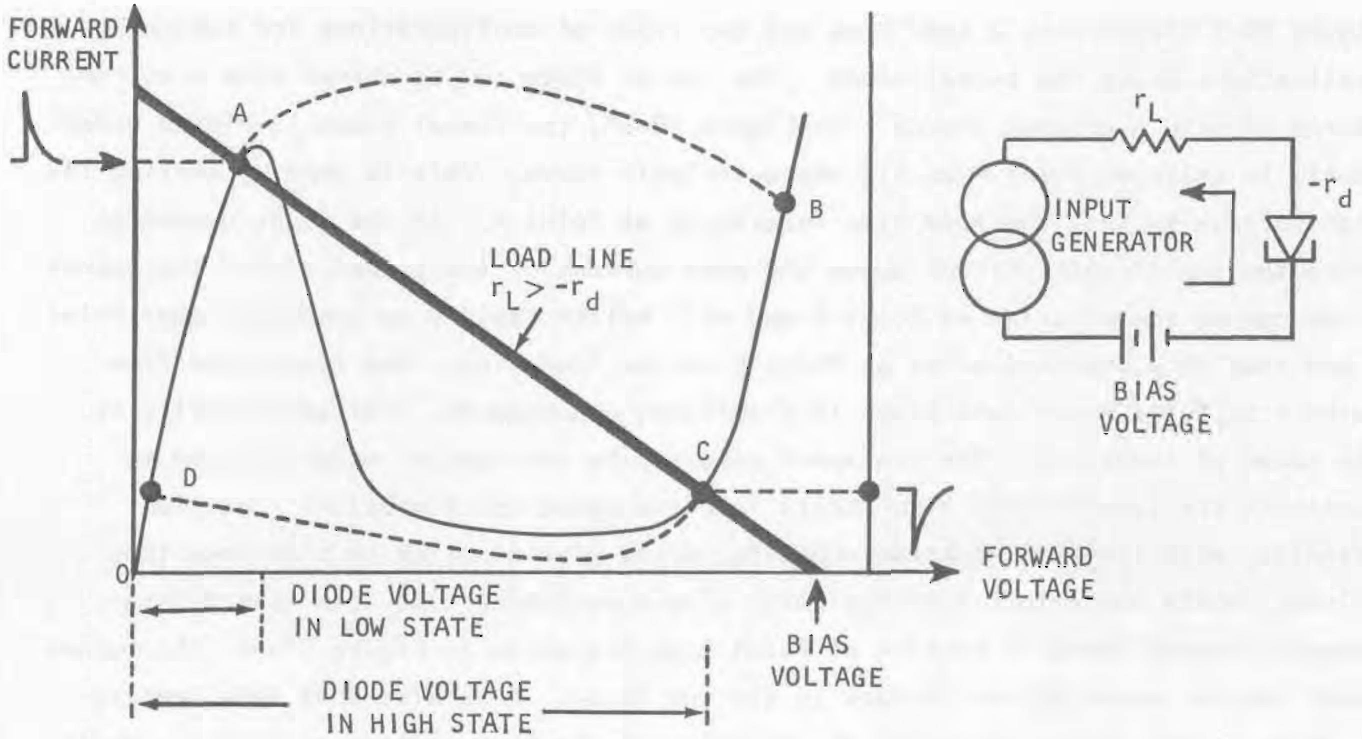
OSCILLATIONS

FIGURE 49-2

Since it is possible for the tunnel diode to exist at many different points on the load line, any reactive components that would provide a change would result in the tunnel diode existing at a different point on its load line. In this type of a configuration with the negative resistance and the positive resistance being near equal and tangent, the tunnel diode will serve as an oscillator. This is assuming that the other criterion are met and this will be discussed shortly.

Figure 50-2 illustrates a load line and two types of configurations for switching applications using the tunnel diode. The tunnel diode can be driven with a current source or with a voltage source. In Figure 50-2A, the tunnel diode is biased quiescently to exist at Point A on its characteristic curve. This is done by setting the bias voltage so that the load line intersects at Point A. If the input generator increases the circuit current above the peak current of the tunnel diode, the tunnel diode can no longer exist at Point A and will switch rapidly to somewhere near Point B and then move down and exist at Point C on the load line. The transition from Point A to Point B can take place in fractional nanoseconds, limited primarily by the speed of tunneling. The transport time set by the type of material used to construct the tunnel diode also enters into the speed considerations. Gallium arsenide, with its fast electron mobility, might give a switching time less than 0.1nsec, while germanium might typically give a switching time less than 0.3nsec. Once the tunnel diode is resting at Point C on its curve in Figure 50-2A, the tunnel diode has no reason to switch back to its low state. Note also that when resting at Point C, the tunnel diode has the majority of the bias voltage across it. While in its low state, it had only a small portion of the bias voltage across it. Reviewing Figure 50-2A, the tunnel diode can be biased quiescently to rest somewhere on its first positive slope.(in this diagram, at Point A). The input generator increasing the current above the tunnel diode's peak current forces the tunnel diode to switch to its second positive slope and then rest on its load line. This is indicated by Points B and C. Once switched to its second positive slope or its high state, it must be triggered once again in order to return to its low state. If the input generator reduces the circuit current below valley current, the tunnel diode will make an excursion from Point C to Point D and then rest once again at Point A with the circuit configuration shown. The switching excursion between A and B and between C and D can take place in fractional nanoseconds. The circuit is bi-stable in that it takes a trigger to switch it from Point A to Point B, and it also takes a trigger to switch it back into its low state.

In Figure 50-2B, we have a similar action in that the tunnel diode is being used as a bi-stable switch, but in this case, the applied voltage is being varied to cause the tunnel diode to switch. If the same considerations are established with the tunnel diode resting at Point A, an increase in the bias voltage by the input generator will move the load line so that it no longer intersects the first positive slope. The tunnel diode will switch to Point B and rest at Point C on the load line. The



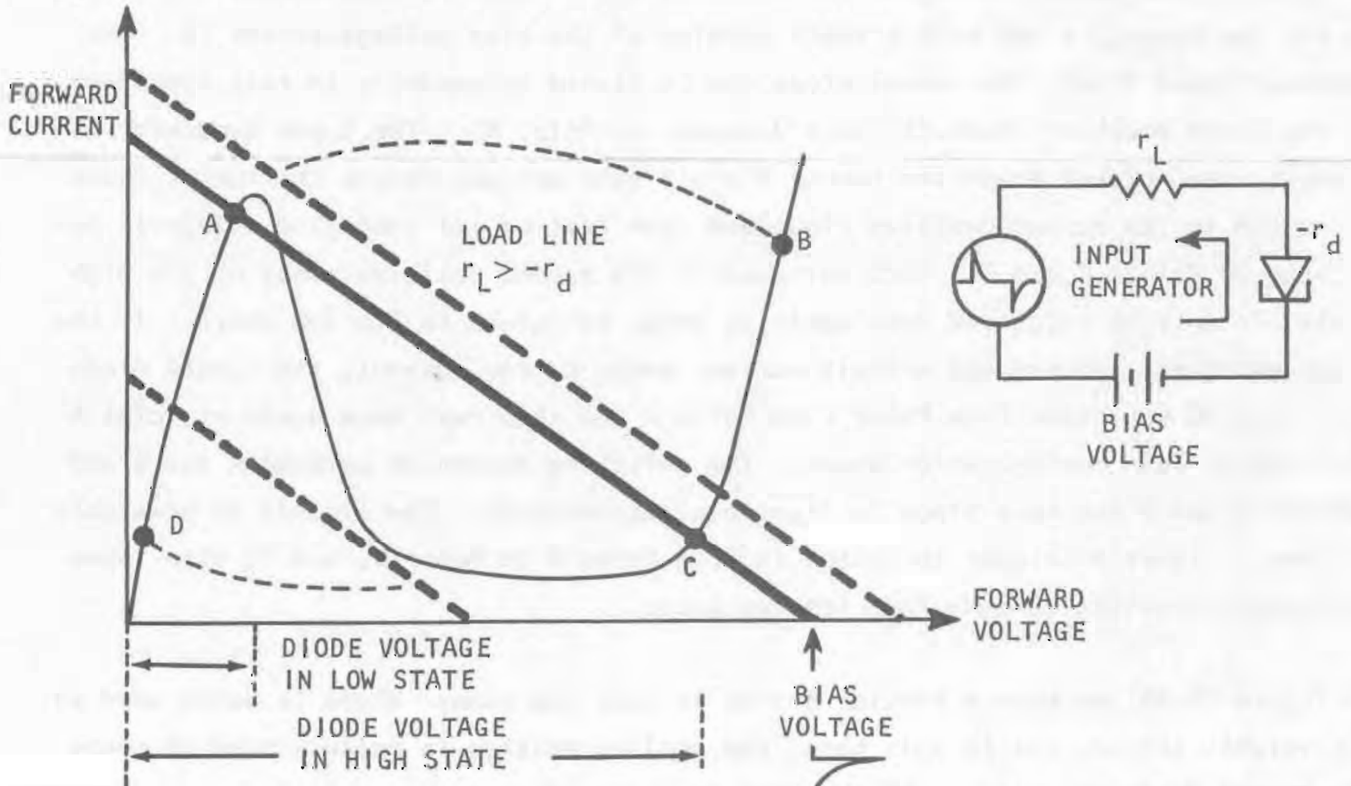


FIGURE 50-2

tunnel diode switching from its low state to its high state can take place in a fractional nanosecond. To have the tunnel diode switch back to its low state, the input generator must reduce the bias voltage below valley voltage so that the load line no longer intersects the second positive slope. When this is accomplished, the tunnel diode will switch back to near Point D and then once again rest at Point A.

Reviewing Figure 50-2B, the tunnel diode is a bi-stable, voltage driven switch. The switching levels are peak voltage and valley voltage, in that the input generator must move the load line beyond peak voltage in order to have the diode switch to its high state, and the generator must move the load line below valley voltage to have the tunnel diode switch back to its low state. The transitions can take place in fractional nanoseconds and the circuit is bi-stable. Point A and Point B on the tunnel diode characteristic are pretty well set (referring to voltage) by the type of semiconductor used in the manufacture of the diode, and Points A and B (referring to current) are set by the junction geometry which establishes the peak current of the tunnel diode. A germanium tunnel diode has Point A occuring at 55mv typically, and Point B at 500mv typically. The total voltage excursion with a germanium tunnel diode when switching from its low state to its high state is somewhat less than 450mv. Gallium arsenide tunnel diodes have a Point A voltage of 100 to 150mv. Point B in Figure 50-2 occurs at about 1100mv. This gives a possible total excursion of approximately one volt for a gallium arsenide tunnel diode.

The signal generated when the tunnel diode switches from its low state to its high state can be used to turn on a non-conducting transistor and the small but very fast transition can then be found at the output circuit of the transistor in an amplified form. Some speed is sacrificed from input to output; however, with present day transistors, a sufficiently fast triggering signal can be obtained in this fashion. The output is sufficiently large to serve as a sweep trigger in oscilloscopes.

Since sweep triggers for oscilloscopes are covered adequately in Chapter 10 and 11 of "Typical Oscilloscope Circuitry" published by Tektronix, Inc., they will not be discussed here. In any count-down circuit, however, such as an oscilloscope sweep trigger, the device used to generate the trigger must have a risetime significantly faster than the risetime of the input signal at the highest expected input frequency. This fast rise characteristic may only occur over a portion of the device's operating characteristics as long as this is the portion used in generating the trigger. The

tunnel diode transition from its low state to its high state can take place in fractional nanoseconds. If this portion of the tunnel diode's switching characteristic is used to generate a triggering signal, the tunnel diode will allow interrogation of very fast risetime, very high frequency pulses, or signals very close together without the inherent jitter or loss of triggering encountered when using a trigger source with a much slower risetime. Figure 51-2 shows a tunnel diode being used in such a configuration.

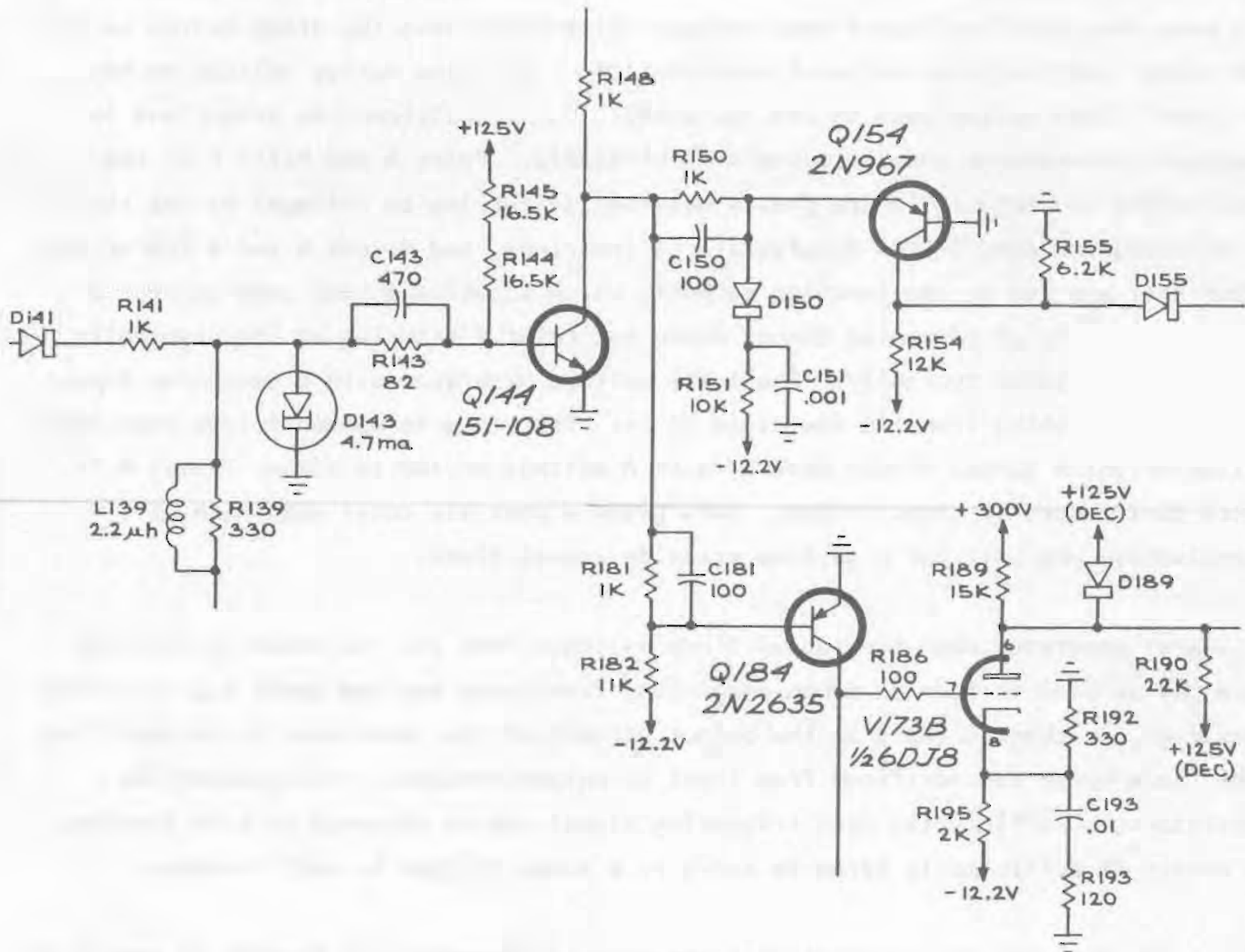


FIGURE 51-2

STABILIZING THE TUNNEL DIODE:

Figure 52-2 lists the stability requirements of the tunnel diode for its three modes

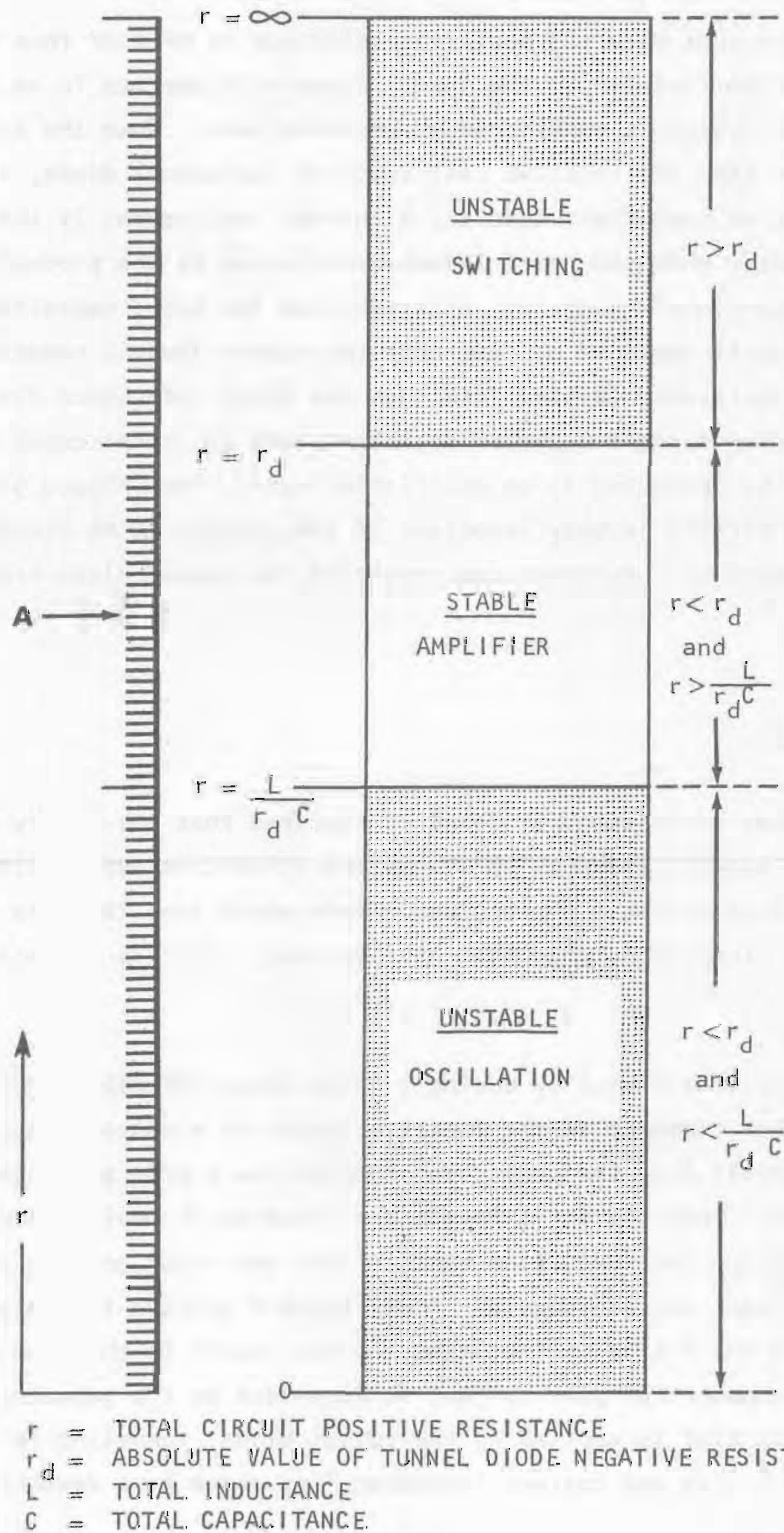


FIGURE 52-2

of operation. Note that when the positive resistance is greater than the absolute value of the negative resistance, the tunnel diode will operate in no mode except the switching mode, which is considered an unstable mode. When the total positive resistance is less than the negative resistance of the tunnel diode, it is possible for it to serve as an amplifier; however, a further requirement is that the positive resistance be greater than the total inductance divided by the product of the absolute value of the diode's negative resistance and the total capacitance. When the positive resistance is equal to or less than the tunnel diode's negative resistance, and the positive resistance is also less than the total inductance divided by the product of the tunnel diode's negative resistance and the total capacitance, the tunnel diode will be operating in an oscillation mode. From Figure 52-2, the total inductance in the circuit is very important if the circuit is to serve as an amplifier. A small amount of inductance can result in the tunnel diode breaking into oscillations.

THE BACKWARD DIODE:

There are times when a conventional diode is required that has a very low conducting voltage drop. An example might be in a coupling circuit in tunnel diode configurations where the drop across a conventional diode would tend to cause a tunnel diode to switch states. There are many other applications, but this is an important example.

The tunnel diode is constructed by adding a large amount of impurities to a semiconductor diode when compared to the impurity levels of a conventional diode. This moves the fermi levels into the conduction band of the N side and into the valence band of the P side. When the fermi levels are lined up at equilibrium, there is an overlap of the valence and conduction bands of the two sides and tunneling is enhanced across the very narrow junction. When forward voltage is applied, tunneling from the N side to the P side is increased and the result is the peak current of the tunnel diode. The amount of peak current is dependent on the geometry of the junction. When reverse bias is applied to the tunnel diode, tunneling is increased from the P side to the N side and current increases just above zero reverse volts.

A diode very similar in construction to the tunnel diode with heavy doping levels is

TUNNEL DIODE

FORWARD CURRENT

BACKWARD DIODE

(B)

(A)

REVERSE VOLTAGE

FORWARD VOLTAGE

CONVENTIONAL DIODE

(C)

REVERSE CURRENT

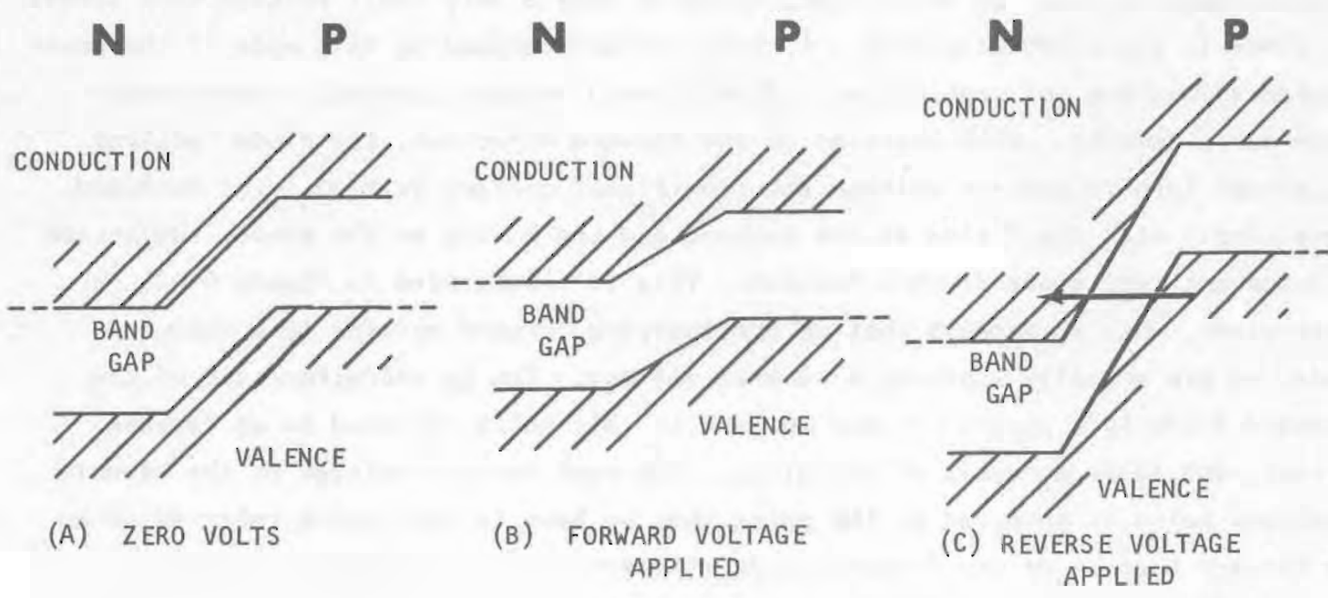
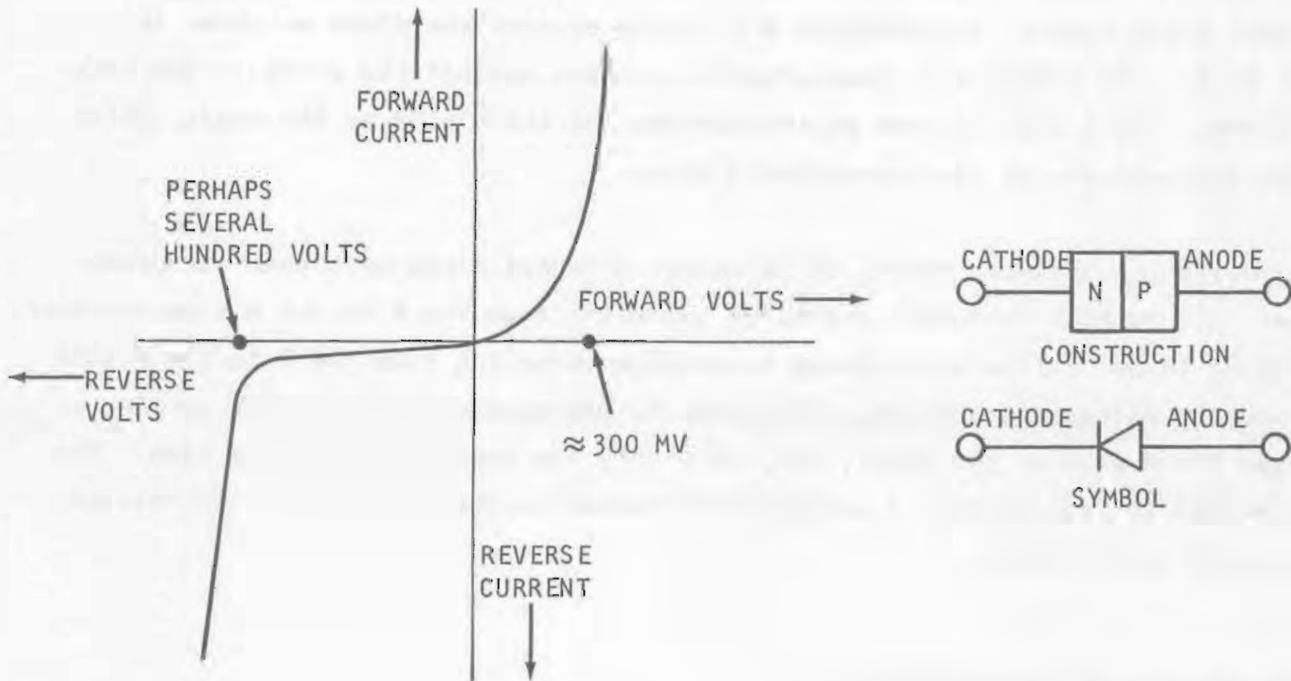


FIGURE 53-2

termed the backward diode. The backward diode, however, is doped with just enough impurities to have the fermi level in the N side exist at the bottom of the conduction band and the fermi level in the P side exist at the top of the valence band. This is illustrated in Figure 53-2A. With the fermi levels placed as they are in the figure, the diode at equilibrium has no band overlap. When forward voltage is applied, tunneling from the N to the P side is <u>not</u> enhanced. In Figure 53-2B, forward voltage raises the conduction band of the N side up opposite of the band gap and the peak current that results is very low with respect to the tunnel diode. In Figure 53-2C, applied reverse voltage does enhance tunneling from the P side into the N side. This is still referred to as the tunnel breakdown mode and it does show a reverse voltage versus current curve similar to avalanche breakdown. Any diode that is doped sufficiently heavy to reduce the breakdown voltage below approximately 6V of reverse voltage will break down primarily as a result of tunneling occuring. The big difference between this type breakdown and avalanche breakdown as far as temperature considerations are concerned is that an increase in temperature results in tunnel breakdown occuring at a lower reverse voltage. Therefore, a diode in tunnel breakdown will exhibit a negative temperature coefficient of voltage, while a diode in breakdown as a result of avalanche occuring has a positive temperature coefficient of voltage.

Note in the voltage versus current curve in Figure 53-2 that with reverse voltage applied, reverse current flows at just above 0V. If the diode is operated with reverse voltage applied, to turn it on, there is only a very small voltage drop across the diode in the conducting mode. A diode can be operated in this mode if the power considerations are not prohibitive. With forward voltage applied, a near normal diode curve results. When operated in the forward direction, the diode requires the normal forward turn-on voltage for significant current to flow. The backward diode simply uses the P side as the cathode and the N side as the anode, and serves as a conventional diode in this fashion. This is illustrated in Figure 54-2. In other words, when it appears that we are applying forward voltage to a backward diode, we are actually applying a reverse voltage. The <u>on</u> characteristic of the backward diode is a result of what we have to this point referred to as reverse current, and it is a result of tunneling. The peak inverse voltage or the reverse breakdown point is governed by the point that we have to this point referred to as the forward turn-on or the forward conducting voltage of the diode. Therefore, the backward diode will have a very low reverse breakdown characteristic since it is

COMPARING A CONVENTIONAL AND A BACKWARD GERMANIUM DIODE



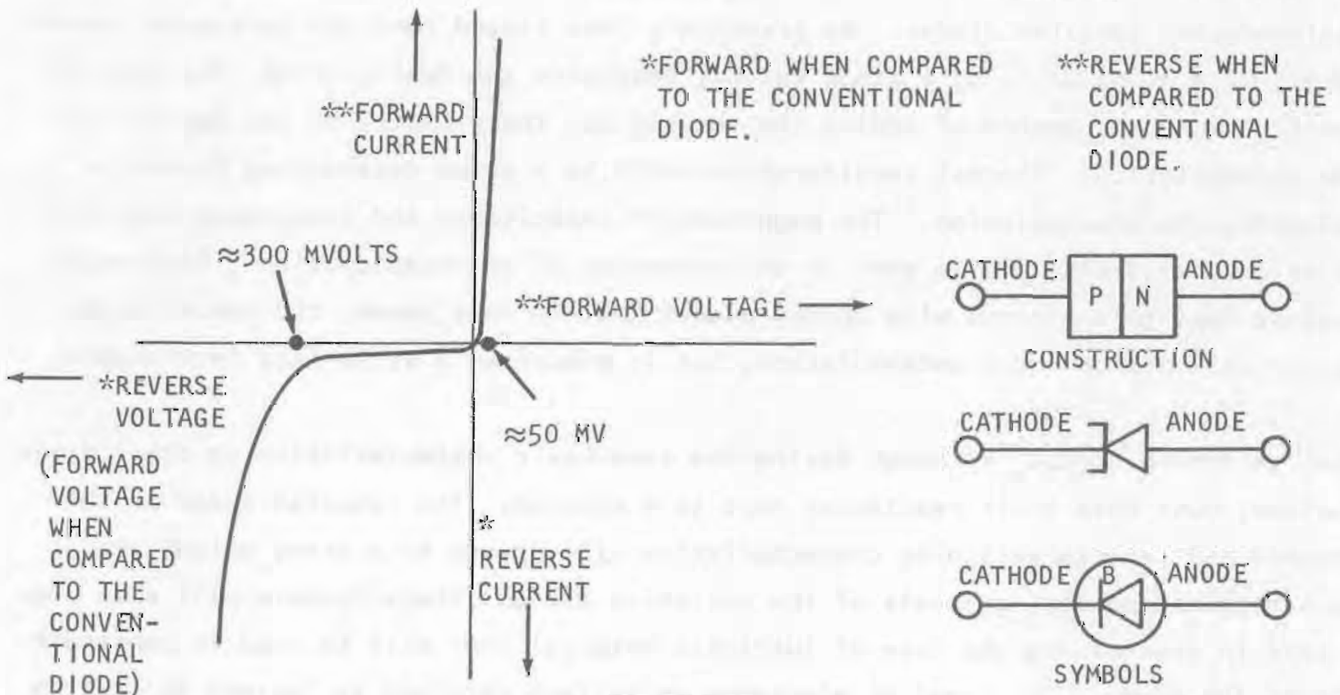


FIGURE 54-2

set by the forward turn-on voltage of the conventional diode. It will have a near zero turn-on voltage in the opposite bias direction, however, and a very low voltage drop when conducting. The symbol for the backward diode is shown simply as a conventional diode symbol, or sometimes a B is placed near the diode as shown in Figure 54-2. The symbol will have electron current against the arrow in the backward diode. The P side is used as the cathode and the N side as the anode, which is just the opposite of the conventional diode.

To summarize the backward diode, it is simply a tunnel diode with near zero peak current. It is heavily doped to enhance tunneling from the P to the N side; however, the doping is not sufficiently heavy to enhance tunneling from the N to the P side when forward voltage is applied. The diode is operated with the P side as the cathode, the N side as the anode, and has a very low conducting voltage drop. The peak inverse voltage rating is essentially limited to the forward turn-on voltage of a conventional diode.

SIGNAL AND FAST SWITCHING DIODES:

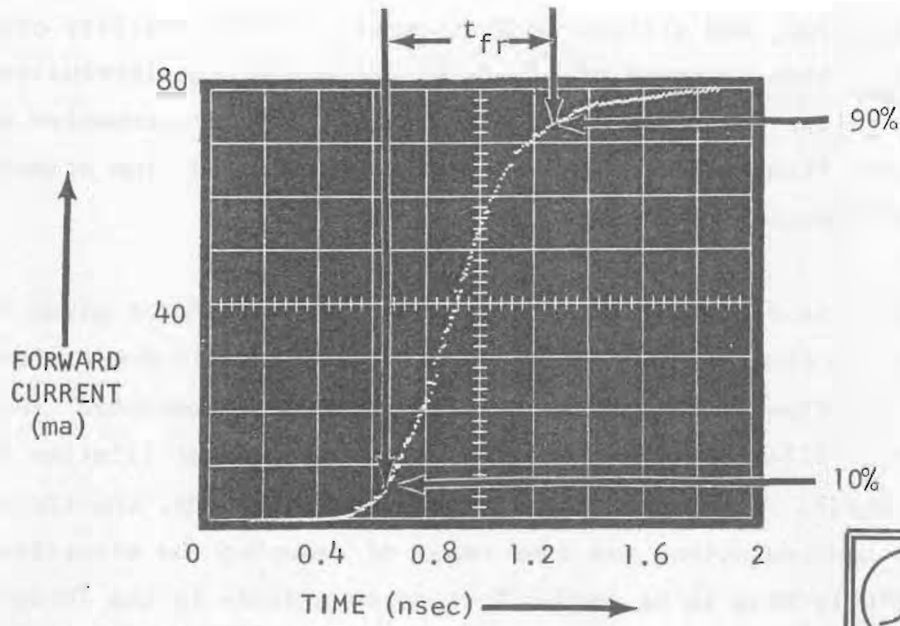
The basic characteristics that have been discussed in general for diodes and most of the characteristics that have been discussed for specific diodes apply to all semiconductor junction diodes. We previously have stated that the particular application of a diode will, to a great extent, determine the doping level, the type of construction, the method of adding the impurities, the geometry of the device, and the encapsulation. Thermal considerations will be a prime determining factor in selecting the encapsulation. The magnitude of capacitance and inductance that can be allowed will also play a part in the selection of an encapsulation. Reactances must be kept to a minimum with tunnel diodes and, in some cases, the tunnel diode is not even placed in an encapsulation, but is placed in a strip-line environment.

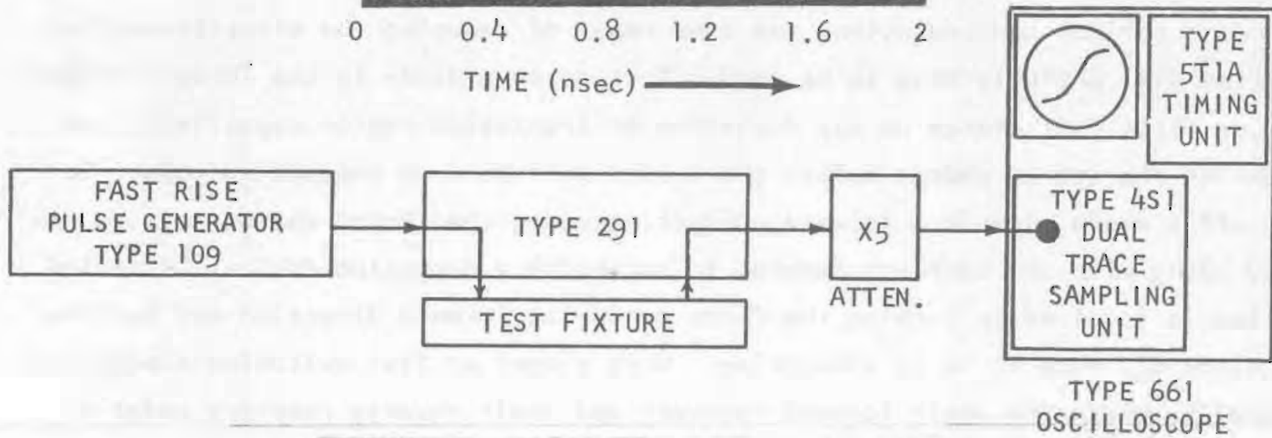
Fast switching diodes, although having the same basic characteristics as other diode devices, must have their reactances kept to a minimum. The required speed of the forward and reverse switching characteristics will govern to a great extent the construction and doping levels of the switching diode. These factors will also play a part in determining the type of intrinsic material that will be used in construction of the diode. The speed of electrons as current carriers is fastest in gallium

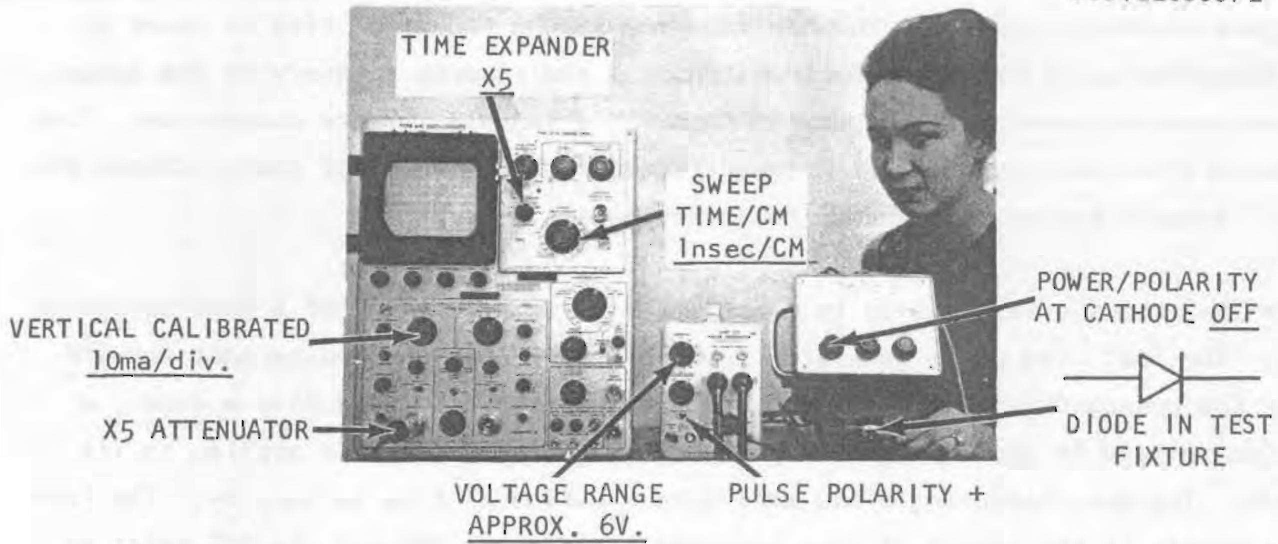
arsenide, then germanium, and silicon in that order. If the mobility of majority carriers or the diffusion constant of minority carriers of the intrinsic material is not a limiting factor for speed considerations, silicon or germanium can be used. For sub-nanosecond switching, it might be necessary to use gallium arsenide or some other type of semiconductor material.

As discussed with basic diodes, the amount of stored charge for a given forward current and hence the storage time of the junction is directly proportional to the minority carrier lifetime. The less impurities in the semiconductor, the longer the minority carrier lifetime (limited to the minority carrier lifetime of the intrinsic material used). With fast signal and switching diodes, the storage time becomes a serious consideration, and some means of reducing the minority carrier lifetime will probably have to be used. To turn on a diode in the forward direction, we first must charge up any depletion or transition region capacitance and establish the stored charge before the diodes will be in a conducting mode. To turn off a diode once in a forward conducting mode, the stored charge must be removed along with the carriers removed to establish a depletion region. A period of time is involved in turning the diode on in the forward direction and turning the diode off once it is in conduction. Most signal or fast switching diodes are checked by measuring their forward recovery and their reverse recovery under a given set of conditions. The forward recovery is the amount of time it takes to turn the diode on in the forward direction, and the reverse recovery is the amount of time involved turning the diode off once it is in full forward conduction. Fast diodes will be constructed for low capacitance and a low value of stored charge per unit of forward current to enhance fast switching times.

Figure 55-2 shows a test set-up to measure the forward recovery of a fast switching diode. The fast rise pulse generator in Figure 55-2 applies a pulse that forward biases the non-conducting diode in the test fixture. To forward bias a diode, a positive voltage is applied to its anode or a negative voltage is applied to its cathode. The non-conducting diode will take a period of time to turn on. The forward recovery is the amount of time measured between the 10% and the 90% point on the rise of the diode forward current waveform. Note that the measurement in Figure 55-2 shows that the diode has a forward recovery time of approximately 650picosec, or 0.65nsec. We could say that the diode measured in Figure 55-2 will turn on in about 650picosec. Actually, the diode will turn on a bit faster than this because







TEST SET-UP: TEKTRONIX TYPE 661 OSCILLOSCOPE WITH TYPE 5T1A TIMING UNIT AND TYPE 4S1 DUAL TRACE SAMPLING VERTICAL, TYPE 109 PULSE GENERATOR, AND TYPE 291 DIODE SWITCHING TIME TESTER. TOP PHOTO TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA.

FIGURE 55-2

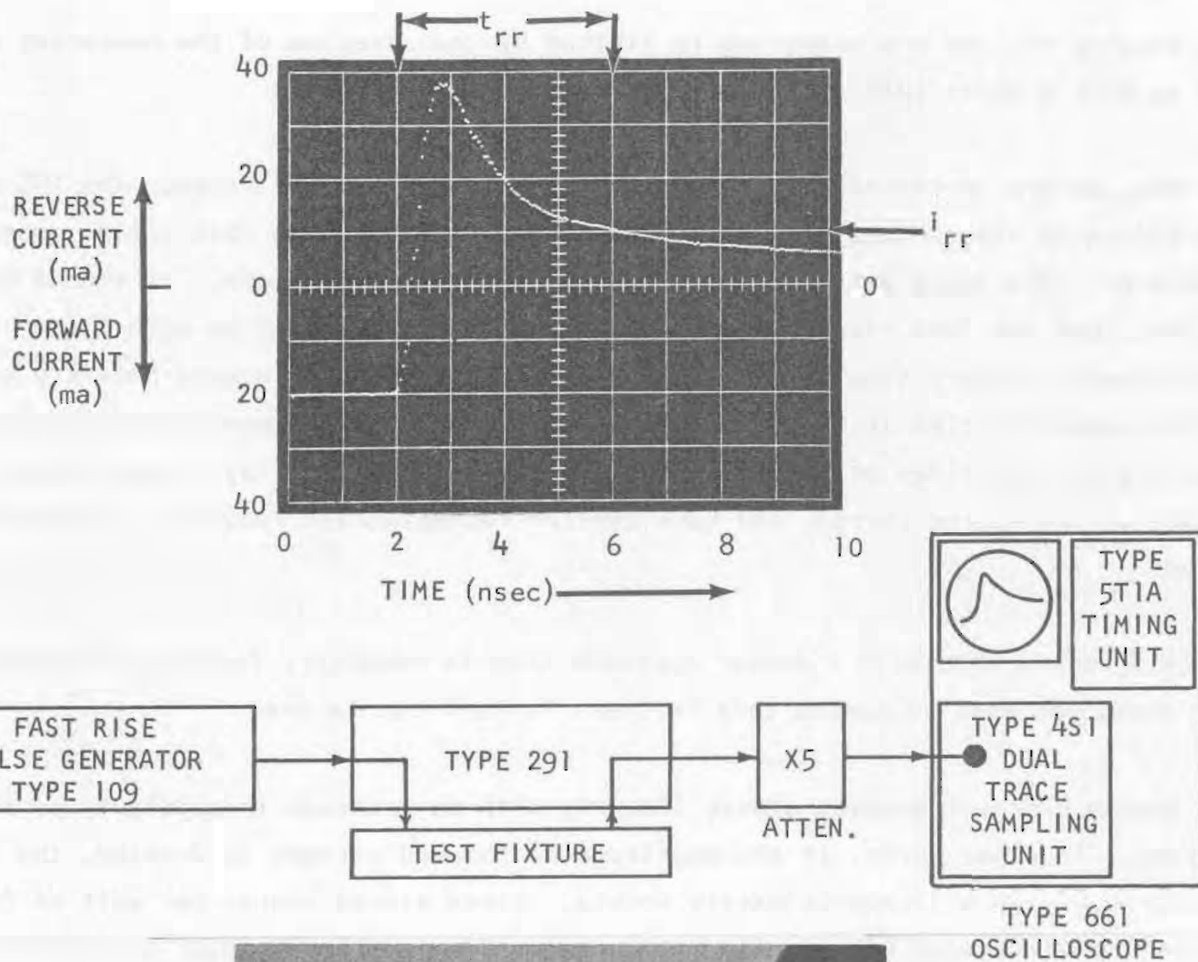
the display that we are measuring is limited by the risetime of the measuring system, and we will discuss this a little bit later.

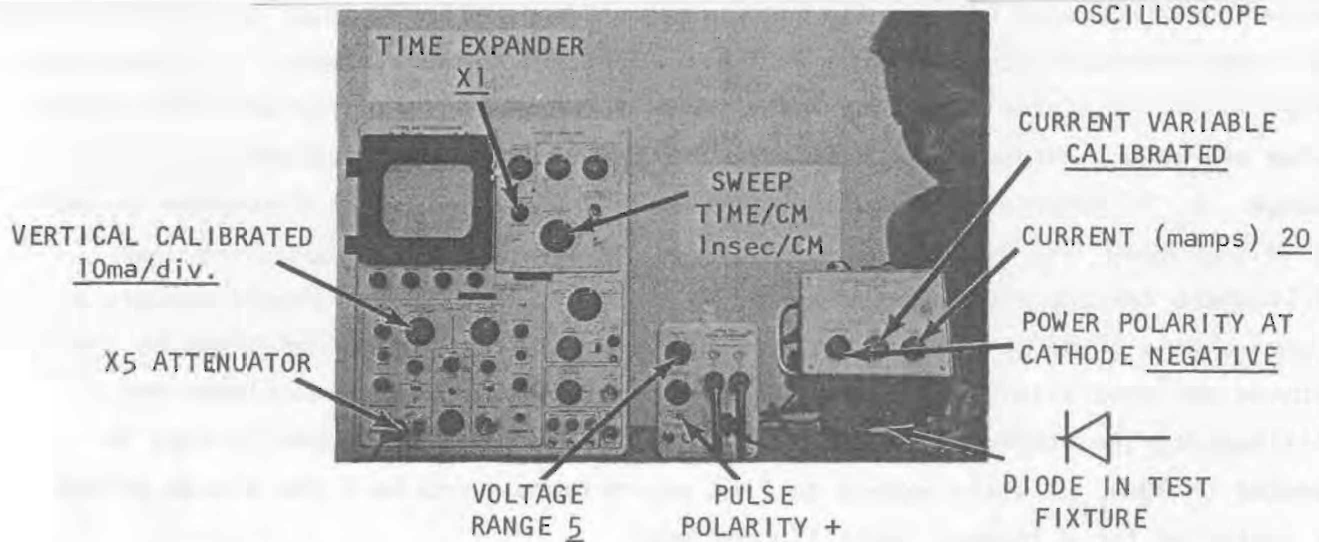
For now, we are concerned with measuring the forward recovery between the 10% and 90% points on the current rise of a diode that is driven by a fast rising pulse generator. The pulse must forward bias the non-conducting diode. It should be evident that the fast rise pulse generator should have a risetime much faster than the forward recovery time of the diode being measured. The forward recovery time is the amount of time it takes for the diode to increase the number of minority carriers on both sides of the junction, charge the depletion layer capacitance, establish the stored charge, and have carrier recombination reach an equilibrium value.

Table 8 covers this with a deeper approach than is necessary for this discussion. For those who want to pursue this farther, Table 8 can be used.

The stored charge increases almost linearly with an increase in magnitude of forward current. In other words, if the magnitude of forward current is doubled, the amount of stored charge will approximately double. Since stored charge per unit of forward current is a function of the minority carrier lifetime, a parameter can be formulated that will represent stored charge as a function of forward current. τ_q is the parameter and it is stated in picocoulombs per milliampere. τ_q indicates that a given value of forward current will result in a given number of picocoulombs of stored charge. τ_q is governed by minority carrier lifetime and this is discussed in Table 8, if you would like to pursue it further. Given a τ_q of 10 picocoulombs per milliampere for the diode in the test set-up in Figure 55-2, we should measure a stored charge of 800 picocoulombs. The forward current is 80ma, as shown by the cathode ray tube display. Taking the product of 80ma and 10 picocoulombs per milliampere, the stored charge is 800 picocoulombs. If the forward current is reduced to 40ma, we would expect to find approximately one-half the stored charge as indicated for a forward conduction of 80ma.

Figure 56-2 shows a test set-up to measure the reverse recovery of a diode. The diode is first established in a forward conducting mode and then a fast rising pulse generator is applied. The fast rising pulse attempts to reverse bias the diode and turn it off. A reverse voltage is a positive voltage applied to the





TEST SET-UP: TEKTRONIX TYPE 661 OSCILLOSCOPE WITH TYPE 5T1A TIMING UNIT AND TYPE 4S1 DUAL TRACE SAMPLING VERTICAL, TYPE 109 PULSE GENERATOR, AND TYPE 291 DIODE SWITCHING TIME TESTER. TOP PHOTO TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA.

FIGURE 56-2

diode's cathode, or a negative voltage applied to the diode's anode.

The diode in Figure 56-2 is initially conducting 20ma of forward current. To calculate the approximate stored charge of the diode, we could take the product of 20ma and the diode's τ_q. In the reverse recovery measurement in Figure 56-2, we can measure the reverse recovery time of the diode. The reverse recovery time is measured between the time of application of a reverse switching voltage to the forward conducting diode, and the time the diode reaches a specified value of reverse recovery current. The measurement in Figure 56-2 indicates a reverse recovery current of 10ma. This is indicated by the symbol i_{rr}.

The conditions under which reverse recovery time t_{rr} is measured must be stated when stating t_{rr}. In this case, the reverse recovery would be stated as recovering to a reverse current of 10ma. The output impedance of the Type 109 Pulse Generator is 50Ω, the impedance of the Type 291 is 50Ω, and the input impedance of the Type 4S1 Dual Trace Sampling Unit is 50Ω. The diode sees 50Ω looking in either direction or a total loop impedance of 100Ω. We might also state that the reverse recovery was measured in a 100Ω loop. Table 3-2 is added to allow the student that is interested to pursue this farther than is required for this discussion.

The following is a list of some symbols and characteristics used with fast switching diodes:

FORWARD TURN-ON CURRENT (I_{fr}):

Maximum value to which the forward current rises after the diode is switched with a forward switching voltage during a forward recovery time measurement.

FORWARD RECOVERY TIME (t_{fr}):

The time is takes the diode's forward current to rise from 10% to 90% of the forward turn-on current (I_{fr}), after the diode is switched with a forward switching voltage in a forward recovery time measurement.

FORWARD CURRENT (I_f):

The amount of forward current passing through the diode before a reverse switching voltage is applied when making a reverse recovery time measurement.

REVERSE RECOVERY CURRENT (i_{rr}):

Specified amount of reverse current to which a diode recovers when measuring reverse recovery time (i_{rr}).

REVERSE RECOVERY TIME (t_{rr}):

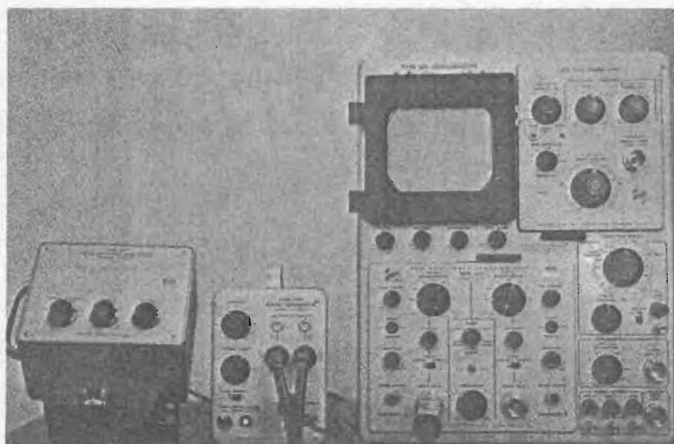
Time interval measured between the time a reverse switching voltage is applied to a forward conducting diode and the diode current reverses (passes through the 0 current level), to the time the diode reverse current reaches the reverse recovery current (i_{rr}) level.

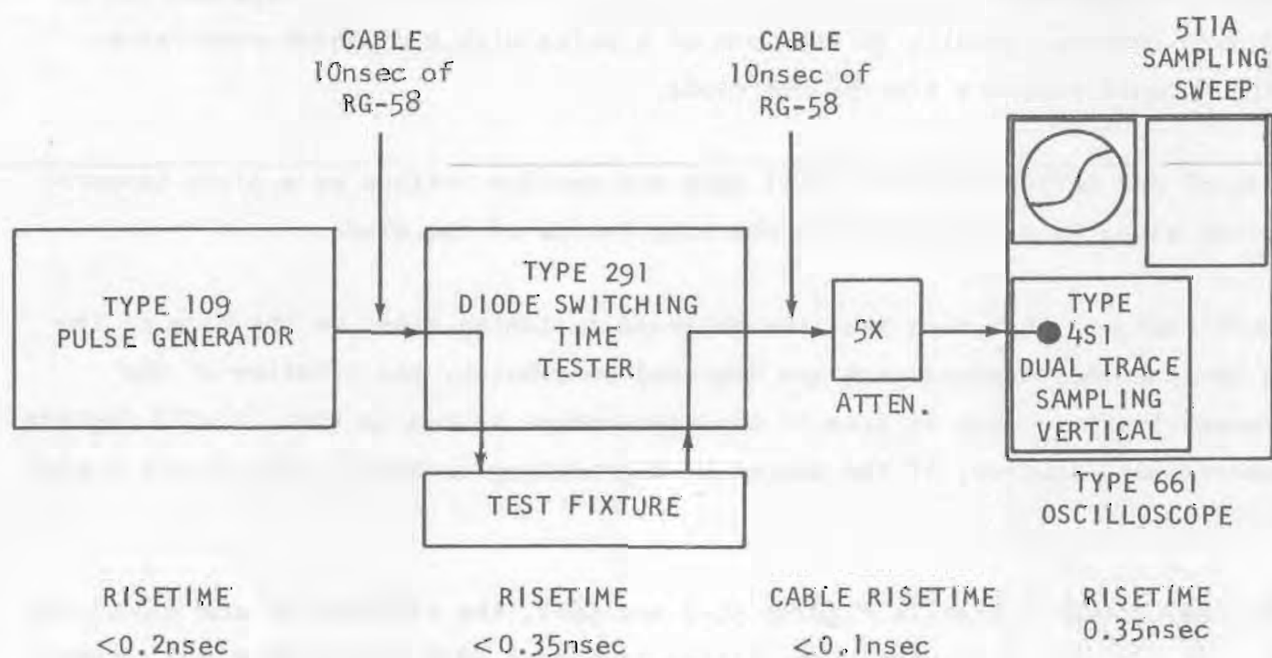
STORED CHARGE (Q_f):

Amount of charge stored by the diode that is recovered during a reverse recovery measurement, generally stated in picocoulombs.

τ_q:

The amount of charge stored by the diode that is recovered during a reverse recovery measurement for each unit of forward current in milliamperes ($\frac{Qf}{If}$). A convenient unit (usually picocoulombs per milliampere) for comparing the amount of diode stored charge charged for each unit of forward current. τ_q assumes a linear relationship between stored charge and diode forward current.





NOTE: Coaxial cable risetime is generally given as 0 to 50% risetime. To convert this to 10% to 90% risetime, use the factor 30 (i.e. 30 x 0-50% risetime ≈10 - 90% risetime).

FIGURE 57-2

FINAL RECOVERY VOLTAGE (V_r):

Final recovery voltage across a diode when measuring reverse recovery characteristics, when the diode is said to recover to a given voltage level in 100Ω loop.

FORWARD SWITCHING VOLTAGE:

A negative voltage applied to the diode's cathode, or a positive voltage applied to the diode's anode, usually in the form of a pulse with a risetime much faster than the forward recovery time of the diode.

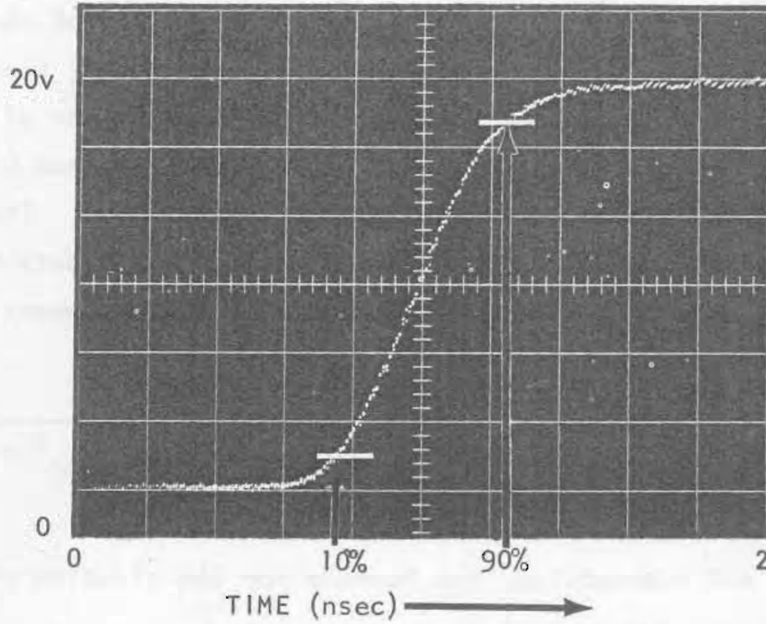
REVERSE SWITCHING VOLTAGE:

A negative voltage applied to the diode's anode, or a positive voltage applied to the diode's cathode, usually in the form of a pulse with a risetime much faster than the forward recovery time of the diode.

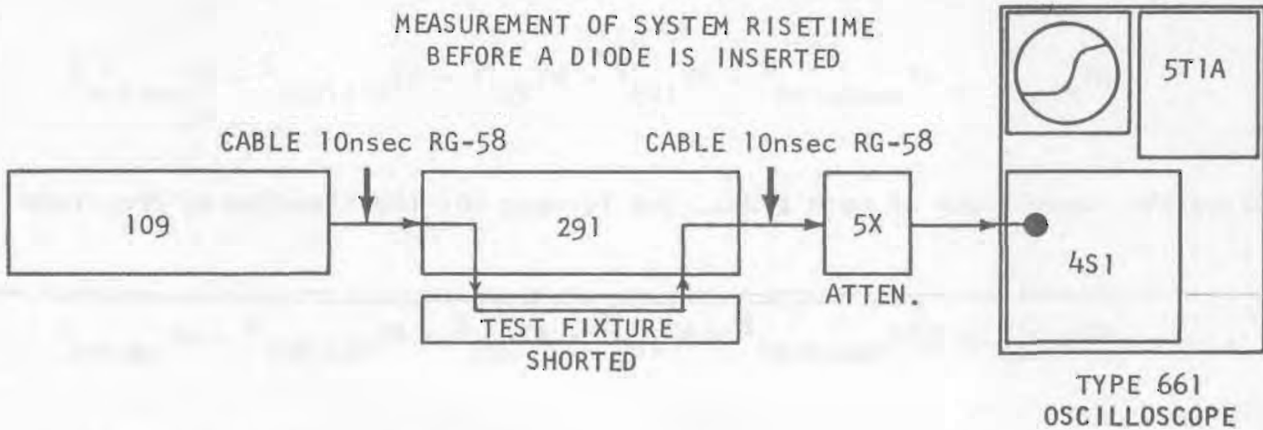
Fast signal and switching diodes will have the maximum ratings at a given temperature given along with the switching characteristics of the diode.

We should state at this time that the observed switching times on the face of the CRT in the previous measurements are degraded somewhat by the risetime of the measurement system. This is true of any measurement system in that it will degrade the measurement; however, if the amount of degradation is known, this is not a problem.

For the measurement system in Figures 55-2 and 56-2, the risetime of the individual parts of the measurements system are listed in Figure 57-2. Note that the risetime of the sampling oscilloscope (Type 661 with a 4S1 Dual Trace Sampling Vertical Plug-In) is 0.35nsec or less. The cable risetime is normally stated as a 0 to 50% risetime. In other words, coaxial cable risetimes are normally given for the rise of the cable from 0 to 50% rather than from 10% to 90%. In order to take the cable risetime into account, the risetime of the cable must be converted to the 10% to 90% risetime figure. A quick approximation for converting cable 0 to 50% risetime to



MEASUREMENT OF SYSTEM RISETIME
BEFORE A DIODE IS INSERTED



(These settings give a time/CM of 0.2 nanosec/CM.)

FIGURE 58-2

a risetime that corresponds to 10% to 90% is to take the product of the risetime given for 0 to 50% and the factor 30. In this case, we state that the risetime of the cable is somewhat less than 0.1nsec after the conversion factor of 30 has been applied. The risetime of the 291 Switching Time Tester is less than 0.35nsec, and the risetime of the Type 109 Pulse Generator is less than 0.2nsec. The risetime measured on the cathode ray tube is approximately equal to the square root of the sum of the squares of all of the risetimes involved in the measurement system, and the risetime of the diode. In formula form, this would appear as:

$$RT_{measured} \approx \sqrt{RT_{diode}^2 + RT_{109}^2 + RT_{291}^2 + RT_{4S1/661}^2 + RT_{cables}^2}$$

By squaring both sides and transposing, the formula for the risetime of the diode reads:

$$RT_{diode}^2 \approx RT_{measured}^2 - RT_{109}^2 - RT_{291}^2 - RT_{4S1/661}^2 - RT_{cables}^2$$

By taking the square root of both sides, the formula for the risetime of the diode reads:

$$RT_{diode} \approx \sqrt{RT_{measured}^2 - RT_{109}^2 - RT_{291}^2 - RT_{4S1/661}^2 - RT_{cables}^2}$$

All of the individual risetimes must be squared and then the square root is taken of the sum of all the squares. To simplify the calculations, the risetime of the system without the diode can be measured and used for solving the diode's risetime. To do this, the test fixture is shorted out and the risetime of the system is measured on the face of the CRT. This is illustrated in Figure 58-2. With the test fixture shorted, the risetime measures approximately 0.5nsec. Nothing more has been done than simply shorting out the test fixture and applying a fast rising signal from the Type 109 directly through the system to the oscilloscope. A measurement is made between the 10% and 90% points to find the measurement system risetime.

The formula to solve for diode risetime is now simplified to:

$$RT_{diode} \approx \sqrt{RT_{measured}^2 - RT_{system}^2}$$

The diode is inserted back in the test fixture and its risetime measured. The measured risetime is inserted in the formula to solve for the actual diode risetime. The risetime of the diode is the square root of the risetime measured squared minus the risetime of the system squared. Suppose the measured risetime on the CRT is 0.8nsec. This risetime was measured on the system in Figure 48-2. The risetime of the system is 0.5nsec and inserting it into the formula:

$$RT_{diode} \approx \sqrt{RT_{measured}^2 - RT_{system}^2}$$

$$\approx \sqrt{0.8nsec^2 - 0.5nsec^2}$$

$$\approx \sqrt{(0.64 - 0.25)\ nsec}$$

$$\approx \sqrt{0.39\ nsec}$$

$$\approx 0.625\ nsec$$

A measured risetime of 0.8nsec actually is a risetime of approximately 0.625nsec. Since we know the degradation offered by the system, calculating the diode's risetime is no problem.

SNAP-OFF DIODE:

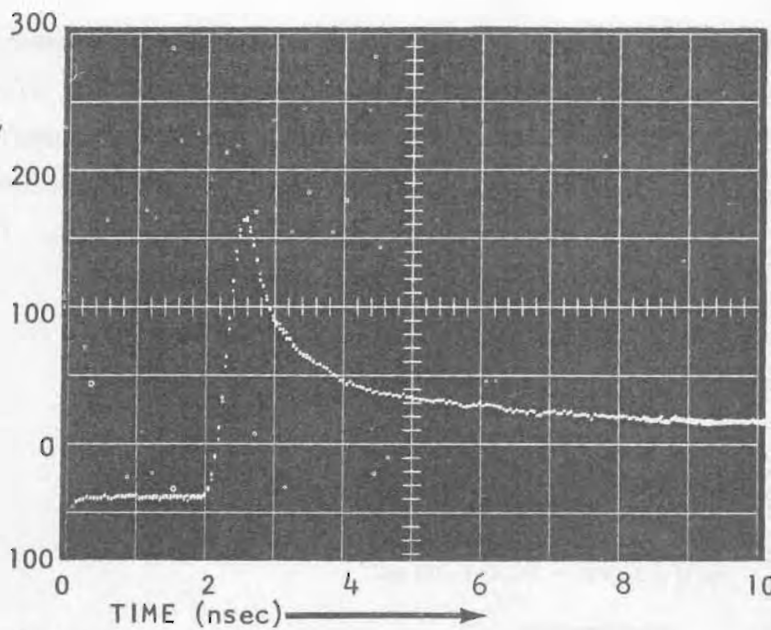
A diode device that finds use in pulse generation and shaping is the snap-off diode. The term "snap-off" refers to the speed with which the diode reverse current cuts off when the forward conducting diode is switched with a reverse switching voltage. The conventional fast switching diode is designed to have a low τ_q for a low stored charge. This involves adding more impurities to reduce the minority carrier lifetime. The snap-off diode is designed to have a fairly large τ_q with respect to other switching diodes, but care is taken to see that the reverse current, as a result of stored charge, will decrease abruptly or "snap-off".

Figure 59-2 shows a reverse recovery measurement of a conventional switching diode and a snap-off diode. The snap-off diode has a much greater stored charge per unit of forward current than the conventional switching diode, but the reverse current fall time is much faster in the snap-off diode.

CONVENTIONAL SWITCHING DIODE REVERSE RECOVERY

REVERSE CURRENT (ma)

FORWARD CURRENT (ma)

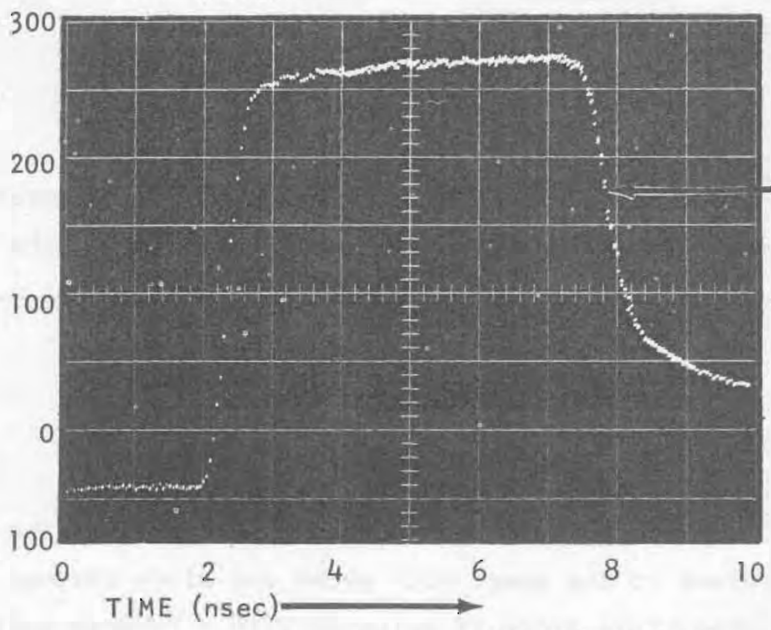


TEST SET-UP:
TEKTRONIX TYPE 661 SAMPLING OSCILLOSCOPE WITH TYPE 4S1 DUAL TRACE SAMPLING VERTICAL UNIT AND 5T1A TIMING UNIT, TYPE 109 PULSE GENERATOR, AND TYPE 291 DIODE SWITCHING TIME TESTER POWER SUPPLY WITH DIODE TEST FIXTURE. TOP TWO PHOTOS TAKEN WITH TEKTRONIX TYPE C-12 OSCILLOSCOPE CAMERA.

SNAP OFF DIODE REVERSE RECOVERY

REVERSE CURRENT (ma)

FORWARD CURRENT (ma)



SNAP OFF REGION

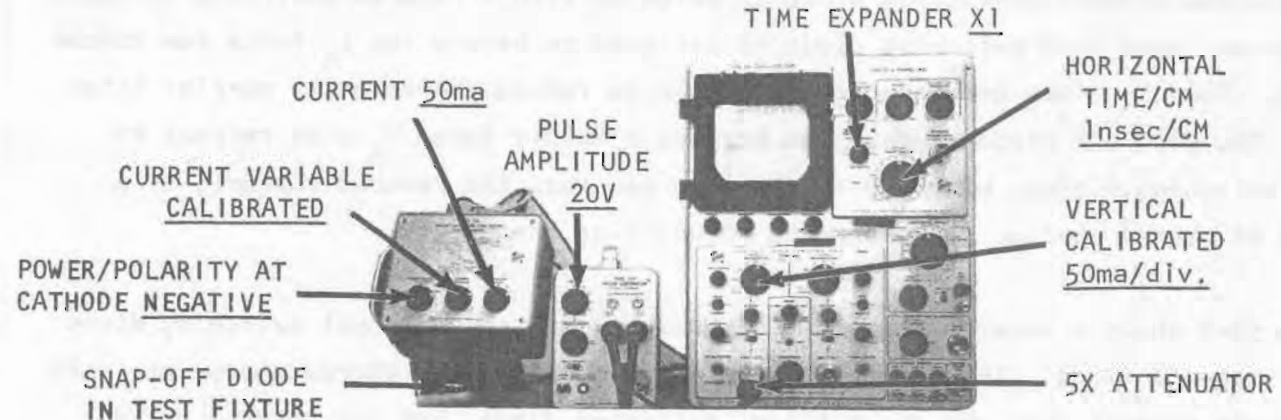


FIGURE 59-2

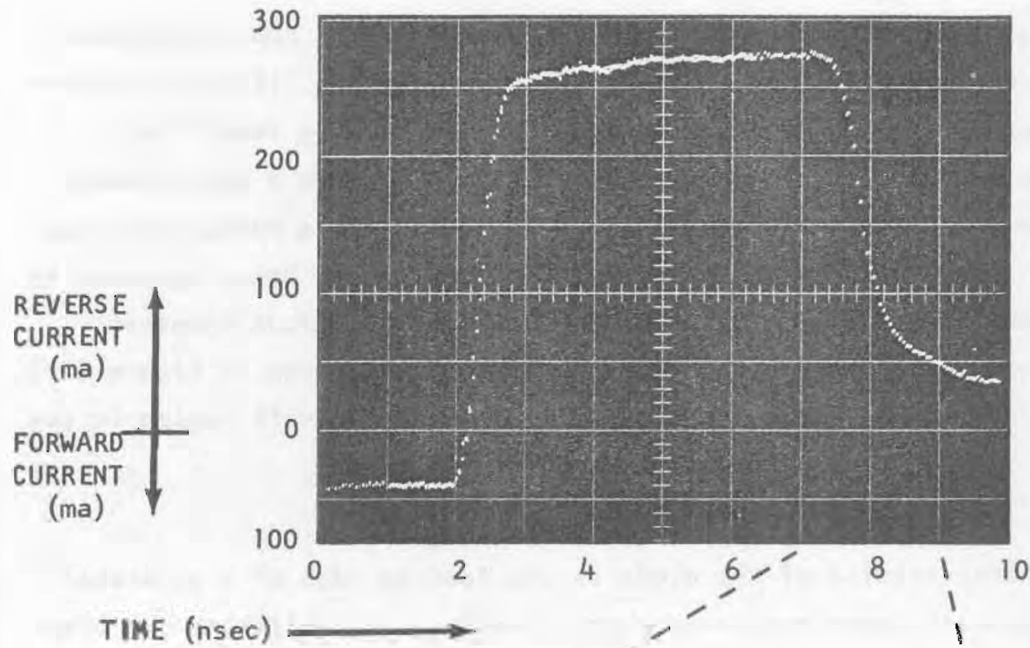
The snap-off diode has controlled doping levels and, as a result, controlled minority carrier lifetime throughout the device. The minority carrier lifetime is maintained long (few imperfections) near the junction; however, doping impurities (gold is typical) are diffused into the ends of the device to give a sharp change from a long to a short minority carrier lifetime. The result is a rather high reverse current during the first portion of reverse recovery time. Once the minority carriers near the junction are depleted, however, the reverse current snaps-off abruptly. The snap-off can occur in fractional nanoseconds as shown in Figure 60-2. The lower photo in Figure 60-2 is simply an expansion of the snap-off region in the upper photo.

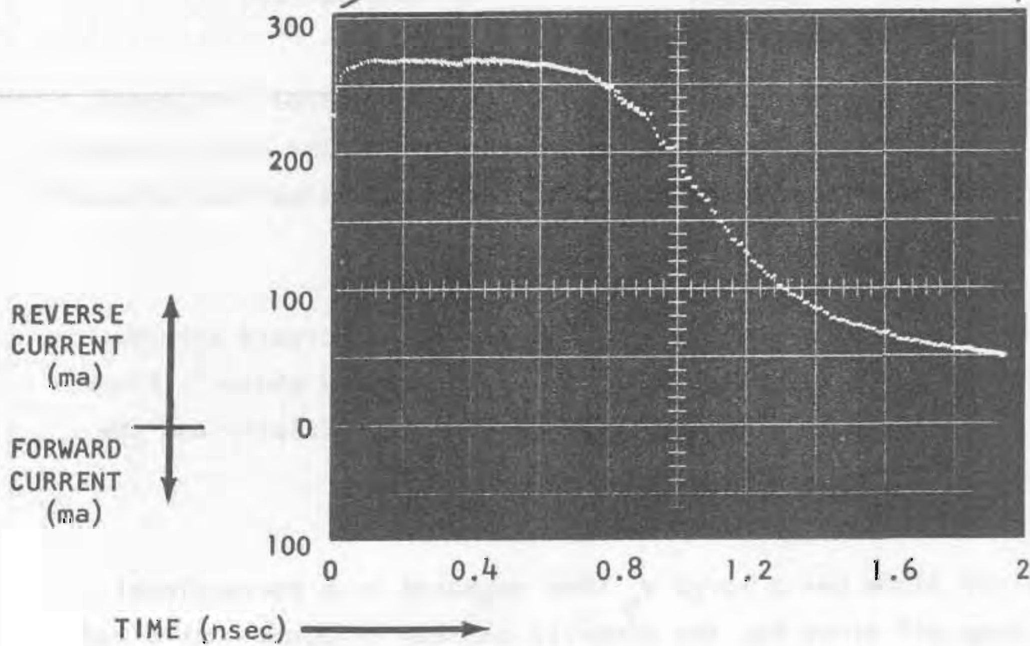
Using the snap-off characteristic of the diode as the leading edge of a generated pulse can give a pulse with fractional nanosecond risetime. Establishing the snap-off diode in forward conduction in circuitry and then switching it with a reverse switching voltage can result in the generation of a very fast rising pulse provided proper circuitry is used. The diode will continue to conduct for the duration of the stored charge and then snap-off abruptly in fractional nanoseconds.

Controlling the initial forward current of the snap-off diode controls the total stored charge and the storage time. This also allows control of the delay between the time of the application of the reverse switching voltage and time that snap-off occurs.

Figure 61-2 top photo has the snap-off diode conducting 40ma of forward current initially and the storage time is approximately 6nsec. The bottom photo in Figure 61-2 has the snap-off diode conducting 50ma of forward current initially and the storage time has increased to over 7nsec.

To summarize, the snap-off diode has a large τ_q when compared to a conventional switching diode. The snap-off diode has the minority carrier lifetime controlled by controlled doping, which results in the reverse current snapping-off abruptly in fractional nanoseconds. The snap-off characteristic is used in fast risetime pulse generation and shaping.





TEST SET-UP: SAME AS FIGURE 59-2 EXCEPT LOWER PHOTO HAS TIMF EXPANDER SET AT X5.

FIGURE 60-2

FOUR LAYER DEVICES:

The four layer or Shockley diode is a four layer triple junction device used in switching applications. Figure 62-2 shows a rough cross section and the symbols for this device.

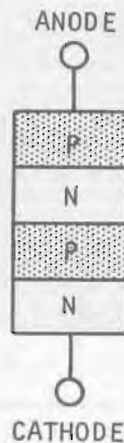


FIGURE 62-2

N material serves as a cathode while the second P material serves as the anode. Figure 63-2 shows the four layer diode with a forward voltage applied.

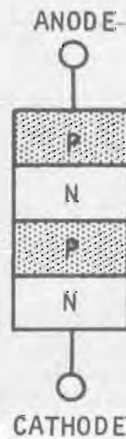
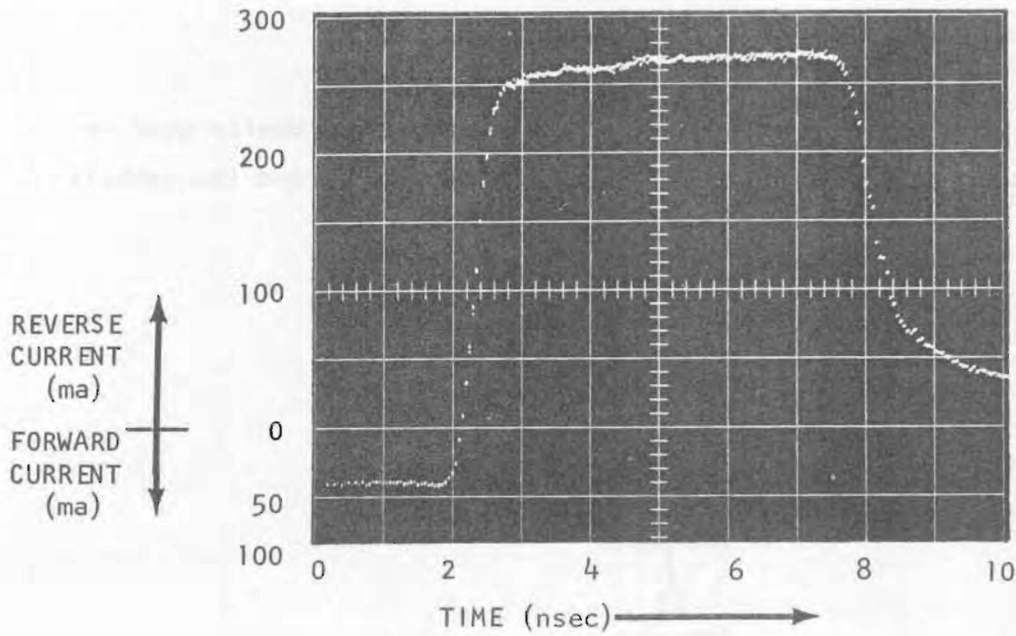
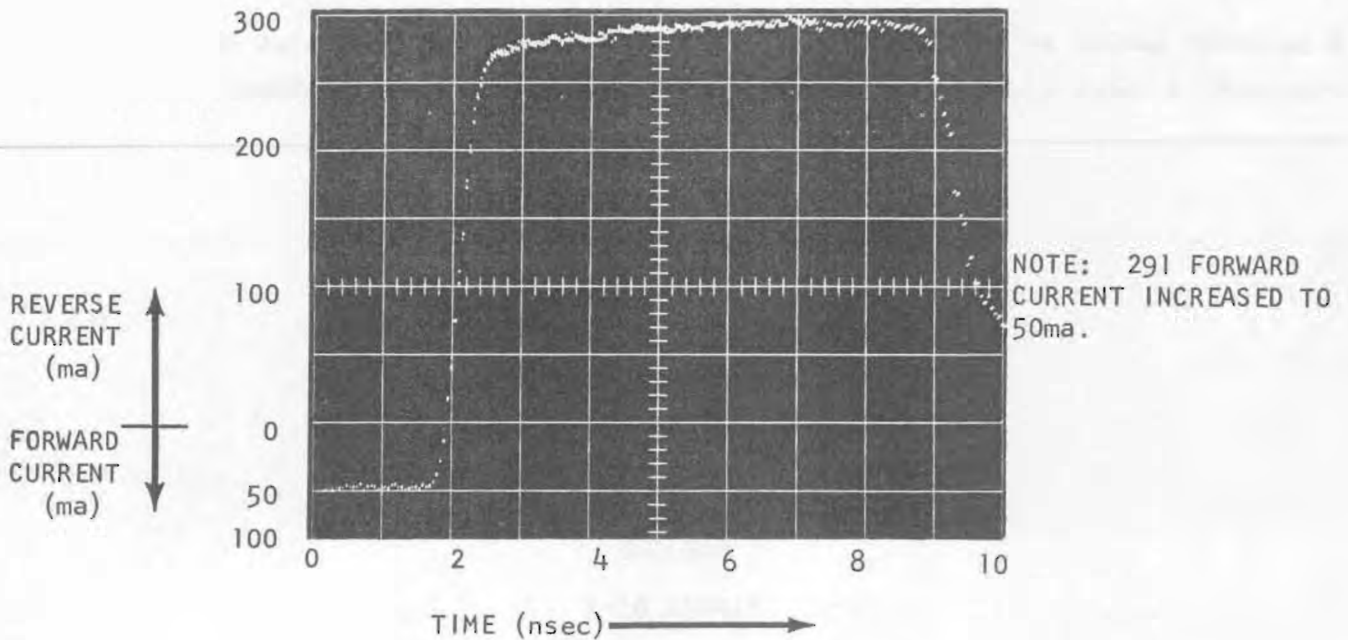


FIGURE 63-2

A forward voltage applied to the diode will attempt to forward bias the junction nearest the cathode and the junction nearest the anode. However, the center junction is reverse biased. The only current that flows with this polarity of bias applied is the diode's saturation current of the center junction. If the forward voltage is large enough to cause the center junction to go into an avalanche condition, a



SNAP-OFF DIODE REVERSE RECOVERY
WITH 40ma OF FORWARD CURRENT



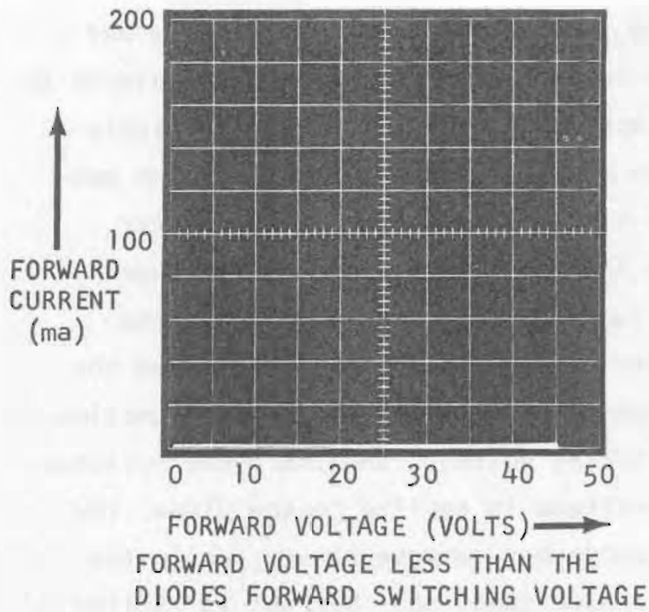
SNAP-OFF DIODE REVERSE RECOVERY
WITH 50ma OF FORWARD CURRENT

TEST SET-UP: SAME AS FIGURE 40A WITH THE EXCEPTION OF THE TYPE 291 CURRENT INCREASED TO 50ma FOR THE LOWER PHOTO

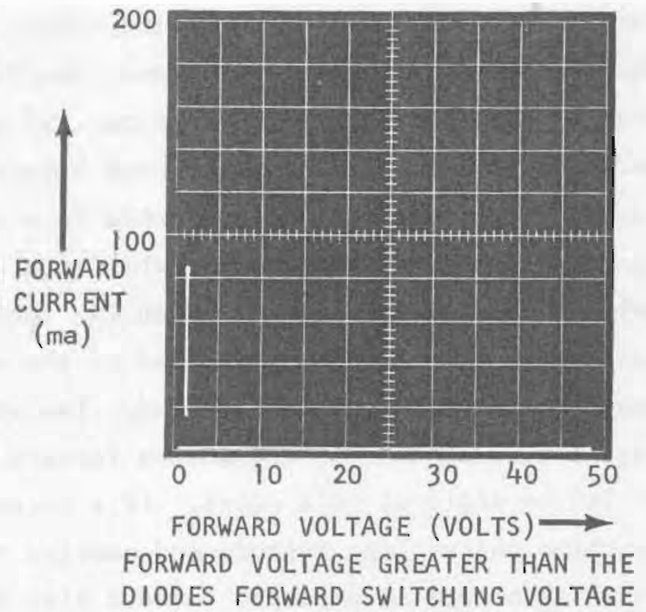
FIGURE 61-2

large current can flow through the diode. The junctions nearest the cathode and the anode are simply forward biased junctions and when the center junction starts to avalanche, a large current can flow. If the applied forward voltage is variable and adjusted from zero to a voltage beyond the point that the center junction avalanches, the four layer diode exhibits a very high resistance until the center junction breaks down and then exhibits a very low resistance. Once the forward voltage point is reached at which the center junction begins to avalanche, the current in the circuit is limited by the series resistance in the circuit and the amount of forward voltage applied. The voltage point at which the center junction begins to avalanche is termed the forward switching voltage, and the diode switches to its on state at this point. If a reverse voltage is applied to the diode, the junction nearest the cathode and nearest the anode are reverse biased, while the reverse voltage attempts to forward bias the center junction. Current is limited by the diode saturation current of the two outer junctions until sufficient reverse voltage is applied to cause both of these junctions to enter breakdown. The four layer diode with reverse voltage applied exhibits a reverse characteristic similar to a conventional diode; however, the breakdown voltage is approximately twice that of a conventional diode made of the same material. Since the four layer diode is not dependent on diffusion current in a narrow base (as in a transistor), a large amount of current may be handled for a given device size when compared to a transistor.

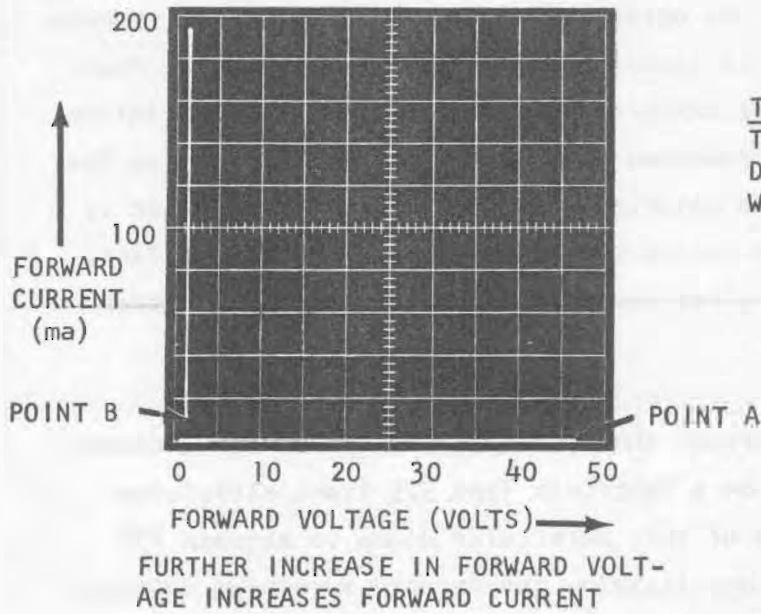
Figure 64-2 shows the voltage versus current characteristic of a four layer diode with forward voltage applied (measured on a Tektronix Type 575 Transistor-Curve Tracer). The forward switching voltage of this particular diode is between 45V and 50V. When the applied forward voltage is below the forward switching voltage, there is a very small current and a relatively large voltage across the diode. There is approximately 45V across the diode in the first CRT display in Figure 64-2. When the forward switching voltage of the diode is exceeded, there is a fast transition from the diode's high resistance state to its low resistance state. Once into the low resistance state, the current is primarily limited by the series resistance in the circuit and the applied voltage. When the resistance of the diode changes rapidly from a high resistance to a low resistance, the voltage across the diode changes from a large voltage to a small voltage. The time of this transition takes place in a microsecond or less typically, and the amount of switching power can be very large. Maximum power dissipation is governed by the same factors previously



FORWARD VOLTAGE LESS THAN THE DIODES FORWARD SWITCHING VOLTAGE



FORWARD VOLTAGE GREATER THAN THE DIODES FORWARD SWITCHING VOLTAGE



FURTHER INCREASE IN FORWARD VOLTAGE INCREASES FORWARD CURRENT

TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER MOD 122C WITH DIODE ADAPTER. PHOTOS ABOVE TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA.

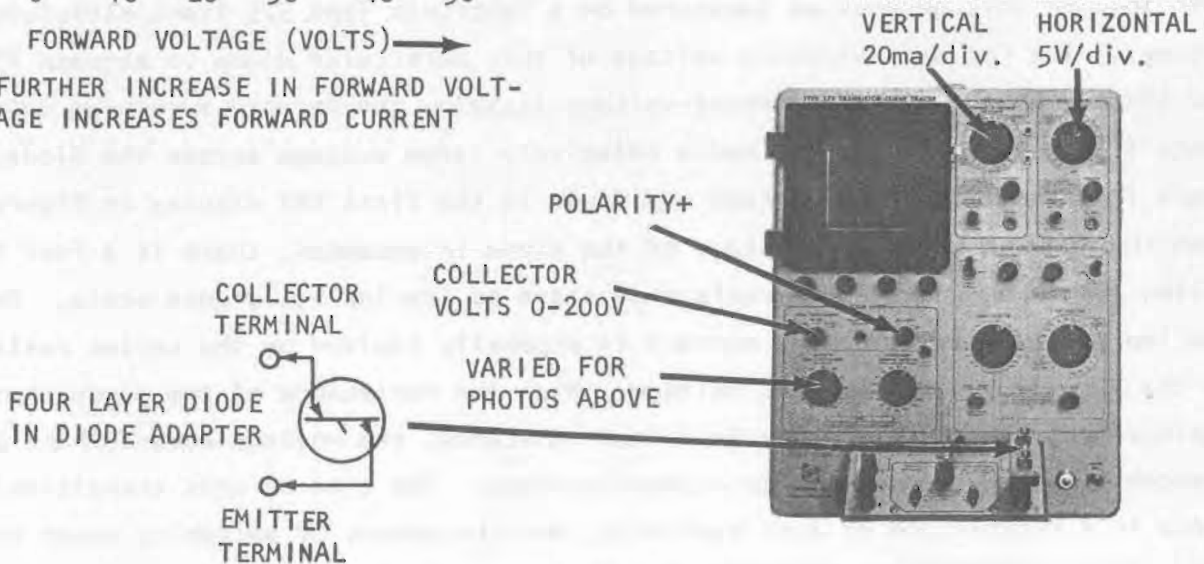


FIGURE 64-2

discussed for conventional diodes.

To return a conducting four layer diode to its off state, the current must be reduced below the point indicated by the arrow marked Point B in Figure 64-2. This current point is termed the holding current. Once the diode has been switched to its conducting state, and the current is above the holding current level, the diode will remain in its conducting state. To switch the diode to its non-conducting or off state, the current must be reduced below the holding current level. This might be done by reducing the applied voltage below a level at which the series resistance will limit the current below holding current level, or the circuit can be triggered with a current source.

The four layer or Shockley diode can be switched by increasing the forward voltage above the forward switching level. This results in the diode switching to its on state. The current must be reduced below holding current to switch the diode back to its off state.

Figure 65-2 shows the symbols for the four layer or Shockley diode. The symbol indicated with the A is the one used by Tektronix, Inc. Electron current flows against the arrow shown in the four layer diode symbol, while conventional current flows with the arrows in the symbols.

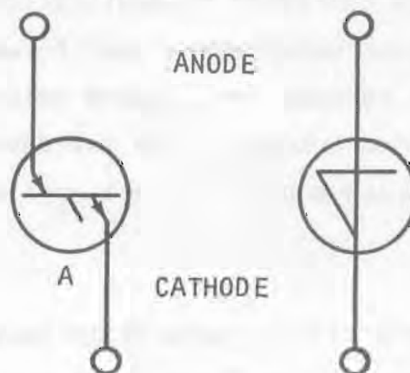


FIGURE 65-2

THE SILICON CONTROLLED RECTIFIER:

Another four layer triple junction device similar to the four layer diode, with the addition of a gate contact, is the silicon controlled rectifier Figure 66-2 shows

the symbol and cross section of the silicon controlled rectifier.

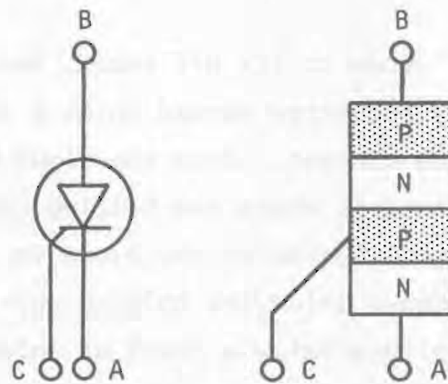
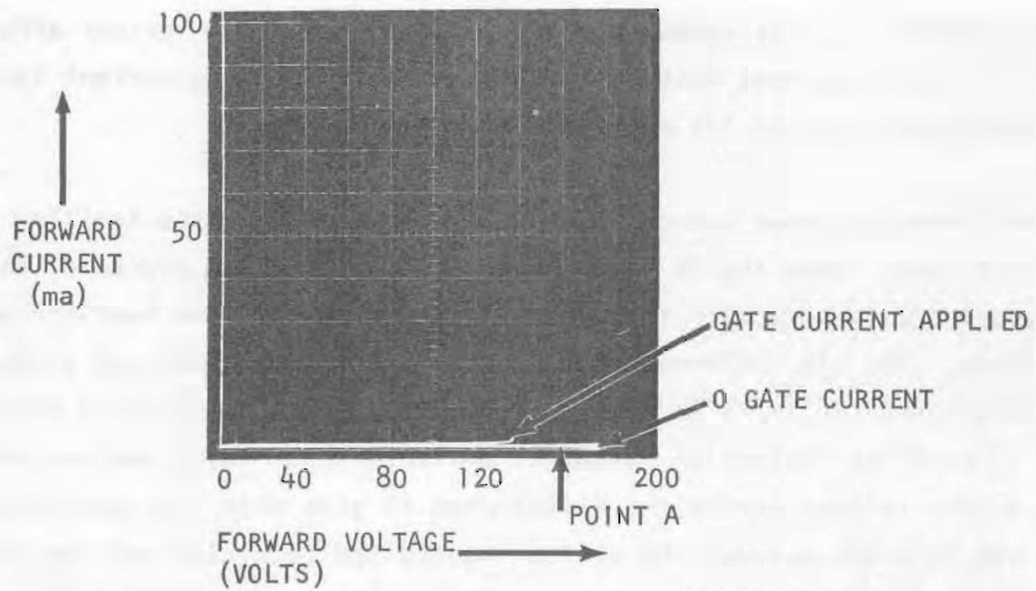


FIGURE 66-2

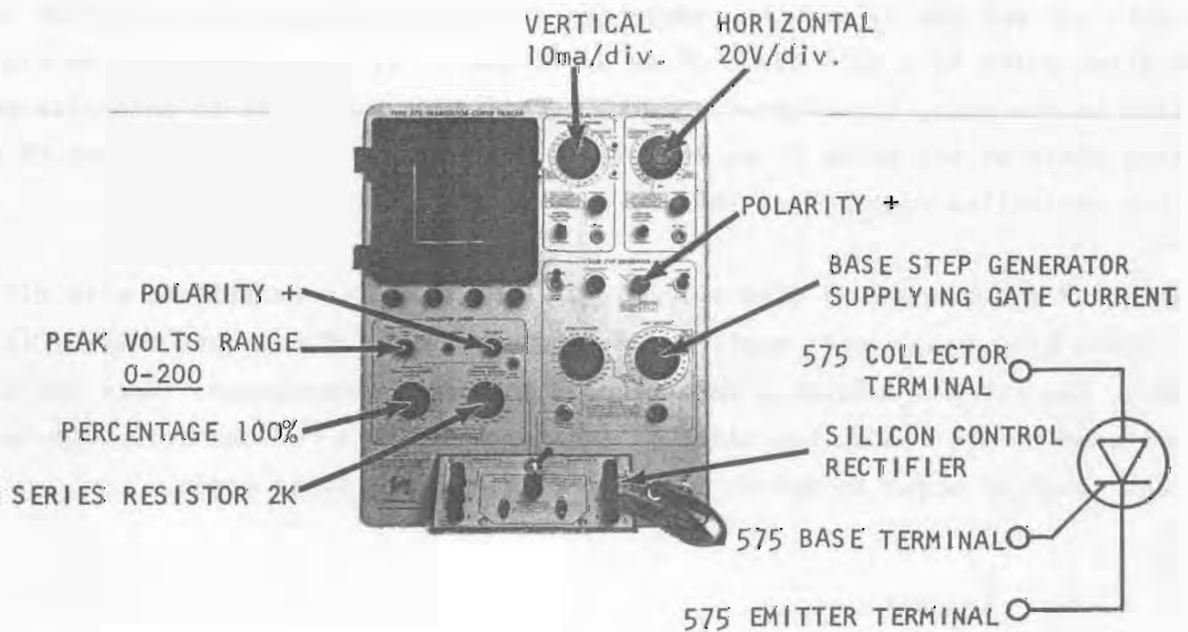
The added gate connection allows control of the current in the junction nearest the cathode. This controls the number of minority carriers in the P material of the junction nearest the cathode, and controls the point at which the center junction will go into an avalanche breakdown condition. The point at which the center junction enters avalanche breakdown determines the forward switching voltage of the diode.

The four layer diode has one forward switching voltage which must be exceeded in order to switch the diode to its on state. The silicon controlled rectifier can have its forward switching voltage decreased by application of a gate current. The diode can be quiescently biased just below its normal forward switching voltage. The application of a gate current reduces the forward switching voltage to a level below the bias voltage and the diode automatically switches. Since the bias voltage is above the new switching level with a gate current applied, the diode switches on. Figure 67-2 illustrates this.

Figure 67-2 is a measurement of the silicon controlled rectifier's characteristics on the Tektronix Type 575 Transistor-Curve Tracer. Biasing the diode at Point A in Figure 67-2 with no gate current applied, the diode remains in its non-conducting state. Applying a gate current of the magnitude required to reduce the switching voltage below Point A results in the diode automatically switching to its conducting state. The reason it switches is that the applied bias voltage is above the new switching voltage required to cause the diode to go into its conducting state.



SILICON CONTROL RECTIFIER EI CHARACTERISTIC



TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER MOD 122C
TOP PHOTO TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA

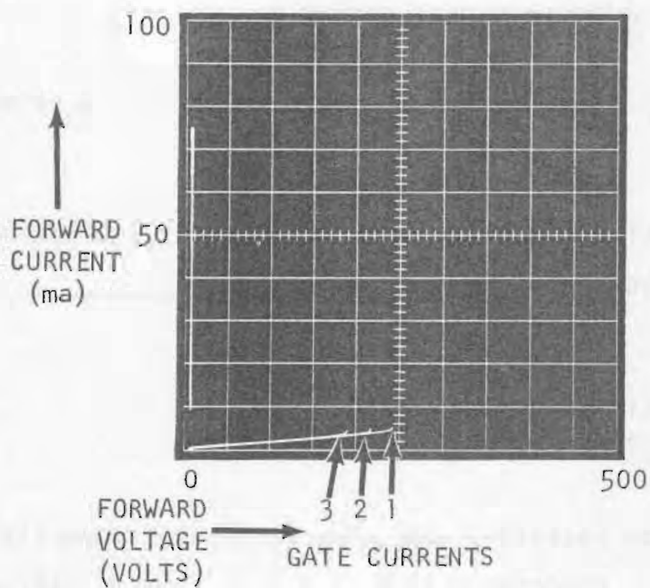
FIGURE 67-2

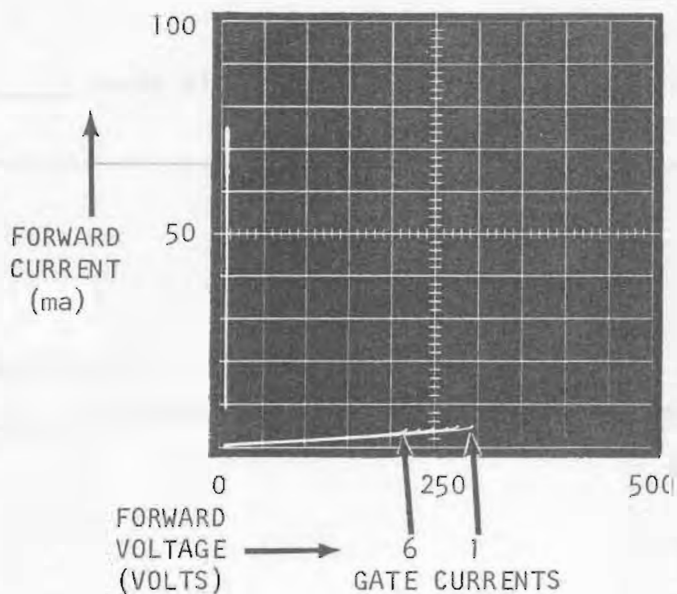
Once the diode is in its conducting state, the gate has no further affect on the diode. The diode current must be reduced below the holding current level in order to switch the diode back to its non-conducting state.

An analog might be drawn between the four layer diode and the familiar voltage regulator (VR) tube. Once the VR tube ionization potential is overcome, the resistance of the device will vary with the applied voltage as will the resistance of the four layer diode. The big difference between the four layer diode and a gas filled regulator tube characteristics is that the gas filled tube maintains a voltage drop across it near its ionization potential while the four layer device has a large change in the voltage across the device when it goes into its conducting state. An analog can be drawn between the silicon controlled rectifier and the thyratron tube. Silicon controlled rectifiers are replacing thyratron tubes in many applications. The advantages of the SCR are that it is a solid state device, smaller size, larger power handling capabilities, and better stability.

By properly selecting the bias voltage applied to the silicon controlled rectifier, the gate current can be made to reduce the switching voltage below the bias voltage at a given point in a half cycle of an a-c signal. By phase shifting the signal applied to the gate, the silicon control rectifier can be made to enter its conducting state at any point in an a-c signal. There are many applications of a silicon controlled rectifier. This is simply an example.

Figure 68-2 shows some CRT displays of silicon controlled rectifiers with different magnitudes of gate currents applied. As the magnitude of the gate current is increased, the forward switching voltage of the diode is decreased. Once the diode is switched to its conducting state, the current must be reduced below the holding current level in order to switch to its non-conducting state again.





NOTE: GATE CURRENTS INCREASE WITH 1 THE LOWEST MAGNITUDE.

TEST SET-UP: TEKTRONIX TYPE 575 TRANSISTOR CURVE TRACER, MOD 122C.
PHOTOS TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA.

FIGURE 68-2

QUESTIONS FOR SECTION 2

Read each question carefully studying any diagrams provided and select the most correct answer.

1. Surrounding the material to be doped with the dopent in a gaseous form, and then subjecting it to heat is termed doping by ____________.

   a. thermodynamics
   b. diffusion
   c. epitaxial deposition
   d. planar masking

2. Close control of silicon rectifier characteristics is accomplished by doping with the ____________ process in high ____________ silicon.

   a. epitaxial layers, resistance
   b. rate growing, resistance
   c. diffusion, resistance
   d. rate growing, reactance

3. Maximum operating temperature of silicon rectifiers is about ____________ degrees centigrade.

   a. 175
   b. 100
   c. 325
   d. 75

4. When sufficient voltage is applied to turn on a diode in the forward direction, or cause breakdown in the reverse direction, the current is ____________ limited.

   a. resistance
   b. depletion layer
   c. stored charge
   d. majority carrier

5. A forward biased silicon diode exhibits a ____________ temperature coefficient of voltage at low currents, which means the diode voltage will ________ with an increase in temperature.

   a. positive, increase
   b. negative, increase
   c. positive, decrease
   d. negative, decrease

6. A silicon diode in avalanche breakdown has a ______________ temperature coefficient of voltage.

   a. negative
   b. positive
   c. 0%
   d. -1.2V/degree centigrade

7. The peak value of a-c voltage that can be applied to a silicon rectifier is limited by the diodes' ______________ rating.

   a. peak forward voltage drop
   b. maximum forward current
   c. maximum reverse current
   d. peak inverse voltage

8. The forward voltage drop of a silicon diode will vary with ______________ ______________ and ______________ ______________.

   a. forward current, series resistance
   b. junction capacity, ambient temperature
   c. forward current, ambient temperature
   d. ambient temperature, stored charge

9. The diode's forward current and voltage can be used to calculate diode's forward ______________ and ______________.

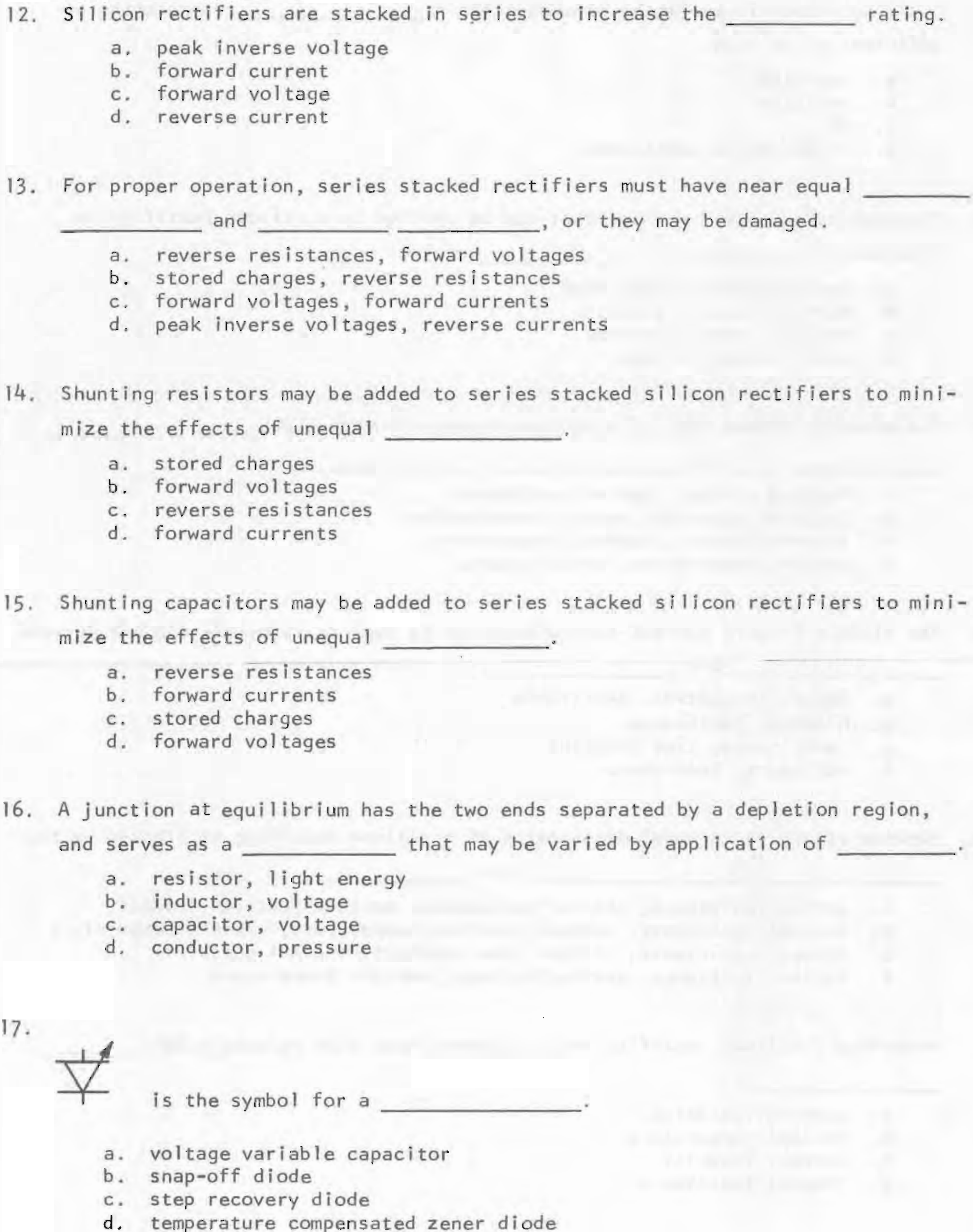
   a. power dissipation, resistance
   b. biasing, resistance
   c. capacitance, time constant
   d. reactance, resistance

10. Maximum steady state power dissipation of a silicon rectifier is limited by the ______________, ______________, and ______________.

    a. series resistance, thermal resistance, maximum junction capacity
    b. thermal resistance, maximum junction temperature, ambient temperature
    c. thermal resistance, thermal time constant, thermal capacity
    d. series resistance, applied voltage, ambient temperature

11. Attaching a silicon rectifier to an external heat sink reduces total ______________ ______________.

    a. power dissipation
    b. ambient temperature
    c. thermal capacity
    d. thermal resistance



12. Silicon rectifiers are stacked in series to increase the ________ rating.

    a. peak inverse voltage
    b. forward current
    c. forward voltage
    d. reverse current

13. For proper operation, series stacked rectifiers must have near equal ________ ________ and ________ ________, or they may be damaged.

    a. reverse resistances, forward voltages
    b. stored charges, reverse resistances
    c. forward voltages, forward currents
    d. peak inverse voltages, reverse currents

14. Shunting resistors may be added to series stacked silicon rectifiers to minimize the effects of unequal ____________.

    a. stored charges
    b. forward voltages
    c. reverse resistances
    d. forward currents

15. Shunting capacitors may be added to series stacked silicon rectifiers to minimize the effects of unequal ____________.

    a. reverse resistances
    b. forward currents
    c. stored charges
    d. forward voltages

16. A junction at equilibrium has the two ends separated by a depletion region, and serves as a __________ that may be varied by application of ________.

    a. resistor, light energy
    b. inductor, voltage
    c. capacitor, voltage
    d. conductor, pressure

17. is the symbol for a ____________.

    a. voltage variable capacitor
    b. snap-off diode
    c. step recovery diode
    d. temperature compensated zener diode

18. The name zener diode is misleading because ________________________.

    a. it was developed by Mr. Esaki
    b. many zeners operate in avalanche breakdown
    c. it is not really a diode at all
    d. most zeners operate in tunnel breakdown

19. The zener diode normally operates in the ______________ region of its EI characteristic.

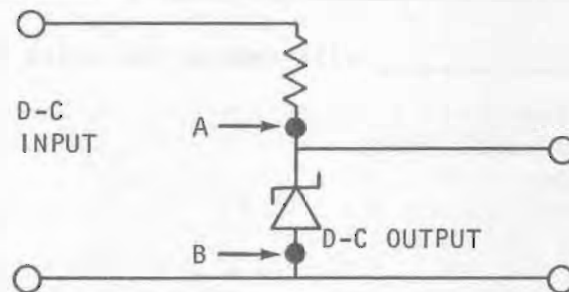
    a. reverse breakdown
    b. low reverse current
    c. high forward current
    d. forward saturation

20. The point of entry into the normal zener diode operating region of the EI curve is termed the _________________.

    a. zener zone
    b. push through point
    c. diffusion point
    d. zener knee

21. When operated as a voltage reference, the zener diode circuit shown will have electron movement from ____________ to ____________ through the diode.

    a. B, A
    b. A, B



22. Power dissipation by the zener diode is limited by ____________ considerations as/than other diodes.

    a. different
    b. the same

23. The number 10M47Z5 indicates a ____________ watt zener diode.

    a. 47
    b. 5
    c. 10
    d. 1/4

24. The number 50M47Z indicates that the zener diode has a ________% nominal zener voltage tolerance.

    a. 50
    b. 47
    c. 5
    d. 20
    e. 10

25. The voltage regulator (VR) tube is limited to a minimum voltage of about ____________ volts, while the zener is available over the entire voltage range up to several hundred volts.

    a. 12.3
    b. 70
    c. 105
    d. 28

26. The noise level generated in a zener diode is ______________ as/than the noise level generated in a VR tube.

    a. much less
    b. about the same
    c. much greater

27. A shunting ____________ will reduce the noise level in the zener diode.

    a. capacitor
    b. conventional diode
    c. resistor
    d. inductor

28. A zener diode operating in ___________ breakdown has a positive temperature coefficient of voltage.

    a. zener
    b. tunnel
    c. avalanche
    d. punch through

29. A forward biased diode at low currents has a ____________ temperature coefficient of voltage.

    a. positive
    b. negative
    c. zero

30. A zener diode operating in avalanche breakdown can be temperature compensated with a ______________ placed in ______________ with the zener.

    a. capacitor, shunt
    b. forward biased diode, series
    c. inductor, series
    d. forward biased diode, parallel

31. Zener diodes connected series opposing will tend to temperature compensate each other when the nominal zener voltage is above approximately _______ volts.

    a. 6
    b. 2
    c. 1.5
    d. 3

32. Tunnel breakdown is enhanced when the doping levels reduce the reverse breakdown voltage below approximately ________ volts.

    a. 15
    b. 6
    c. 3
    d. 1.5
    e. zero

33. Tunnel diodes have tunnel breakdown occuring with ____________ voltage applied.

    a. forward
    b. reverse
    c. both of the above

34. The tunnel diode has a region of ______________ between the tunnel breakdown point and the normal forward turn on point of a conventional diode.

    a. zero conductance
    b. zero resistance
    c. negative resistance
    d. positive conductance

35. Tunnel diodes have the conduction band of the N side and the valence band of the P side ______________ at zero bias.

    a. spread apart by 6 electron volts
    b. separated by a wide junction
    c. overlapped
    d. about 10 μinches apart

36. Reverse voltage applied to a tunnel diode results in increased tunneling from the ____________ to the ____________ side.

    a. N, P
    b. P, N

37. The magnitude of tunnel diode peak current (I_p) is determined by junction ____________.

    a. geometry
    b. doping levels
    c. forward voltage
    d. reverse voltage

38. Valley current is measured at a point between ____________ and ____________.

    a. tunnel breakdown, avalanche breakdown
    b. zero bias, peak current
    c. peak current, normal forward turn on
    d. normal forward turn on, avalanche breakdown

39. A negative resistance can be defined as a resistance that will ____________ power.

    a. generate
    b. dissipate
    c. store

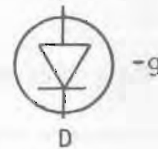
40. The tunnel diode will serve as an amplifier when placed in proper circuitry and operated ____________.

    a. on the first and second positive slopes
    b. on the first positive slope only
    c. in the negative resistance region and the first positive slope
    d. in the negative resistance region only.

41. Current amplification can be accomplished by operating the tunnel diode in an amplifier configuration in ____________ with its load resistor.

    a. series
    b. parallel

42.



A B C D

Symbol ____________ is <u>not</u> a symbol for a tunnel diode.

    a. A
    b. B
    c. C
    d. D

43. To serve as an amplifier, the negative resistance of the tunnel diode must be ________ to/as/than the total circuit positive resistance.

    a. greater
    b. less
    c. as large
    d. equal

44. The total resultant conductance of a tunnel diode amplifier circuit must be ________.

    a. inductive
    b. capacitive
    c. negative
    d. positive

45. The tunnel diode will switch between positive slopes when the circuit positive resistance is ________ the tunnel diode negative resistance.

    a. less than
    b. equal to
    c. greater than

46. The positive resistance of the Type 575 Transistor Curve Tracer generally makes it impossible to observe the negative resistance region of the tunnel diode EI curve, since it forces the tunnel diode to operate in the ________ mode.

    a. switching
    b. amplifier
    c. stable
    d. carrier insertion

47. A tunnel diode switch can be current or voltage driven and will switch between ________ slopes in fractional ________ seconds.

    a. negative, pico
    b. positive, milli
    c. positive, atto
    d. positive, nano

48. To switch a tunnel diode from its high state to its low state requires that the current be ________.

    a. increased above peak current
    b. decreased below valley current
    c. cut-off completely
    d. increased to diode saturation

49. The majority of the supply voltage in a tunnel diode switching configuration will be across the diode when it is in its ____________ state.

    a. high
    b. low

50. A diode with doping levels and construction similar to a tunnel diode, but with near zero peak current, is termed a ____________ diode.

    a. voltage variable capacitor
    b. snap-off
    c. half tunnel
    d. backward

51. A diode that offers a very low conducting voltage drop when used as a conventional rectifying diode is the ____________ diode.

    a. snap-off
    b. snap-on
    c. step recovery
    d. half tunnel
    e. backward

52. Forward recovery in a fast switching diode occurs during the time the current changes from ____________ to ____________ when switched with a switching voltage.

    a. zero, a forward equilibrium value
    b. a forward equilibrium value, zero
    c. a forward equilibrium value, a designated recovery level
    d. on, off

53. Forward recovery is measured between the ________% and ________% current points when the diode is switched with a ____________ switching voltage.

    a. 90, 10, reverse
    b. 50, 50, forward
    c. 10, 90, forward
    d. +50, -50, reverse

54. Reverse recovery is the time interval between the application of a ____________ switching voltage to a conducting diode, and the point at which reverse current reaches ____________.

    a. reverse, designated recovery level
    b. forward, designated recovery level
    c. reverse, zero
    d. forward, zero

55. Forward switching voltage is a ____________ voltage applied to a diodes cathode, or a ____________ voltage applied to a diodes anode.

    a. positive, negative
    b. negative, positive

56. The carriers recovered during reverse recovery time make up the diodes ________________.

    a. depletion region
    b. surface leakage
    c. contact resistance
    d. stored charge

57. Q_s is the symbol for diode stored charge, but a convenient unit of measure usually given in pico-coulombs per milliampere, is given the symbol ________.

    a. ϕ_q
    b. τ_q
    c. Q_q
    d. Φ_q

58. Stored charge varies directly as __________ and ________________.

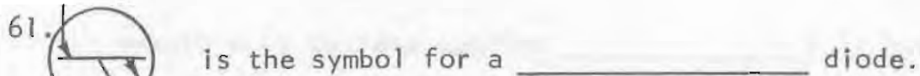
    a. minority carrier lifetime, forward current
    b. minority carrier lifetime, reverse switching voltage
    c. forward current, reverse switching voltage
    d. forward switching voltage, minority carrier lifetime

59. The snap-off diode is designed for a ____________ stored charge per unit of forward current, and a fast ______________ time of the reverse recovery waveform.

    a. low, fall
    b. high, rise
    c. high, fall
    d. low, rise

60. The snap-off diode continues to conduct for the duration of the ____________ when switched with a reverse switching voltage.

    a. reverse switching voltage
    b. stored charge
    c. carrier cancellation time
    d. week

61.  is the symbol for a ________________ diode.

a. shockley
b. snap-off
c. back to back zener
d. four layer
e. either a or d

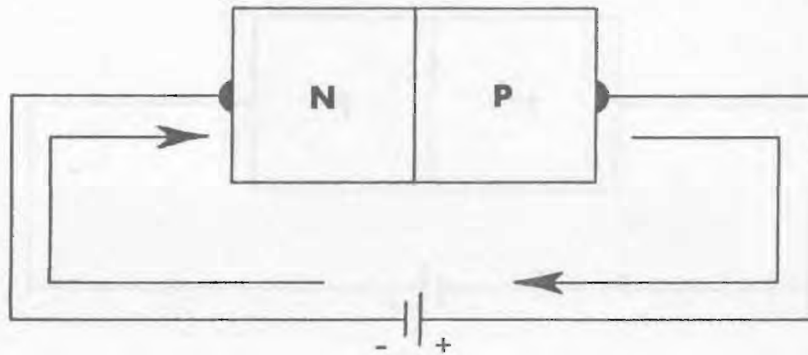
62.

is the symbol for a ____________, and A indicates the ______________.

a. snap-off, injector
b. four layer, control lead
c. shockley, shield
d. silicon control rectifier, gate
e. shockley, gate

TABLE 8

A PN diode with forward voltage applied, conducts electron current in the direction shown by the arrows in the circuit, external of the diode. If you prefer the conventional current approach, change the direction of the arrows.



The process within the diode is not this simply shown. Plotting the minority carrier densities with respect to the forward conducting junction (minority carriers being electrons in P type (n_p) and holes (p_n) in N type) indicates that several transports are involved.

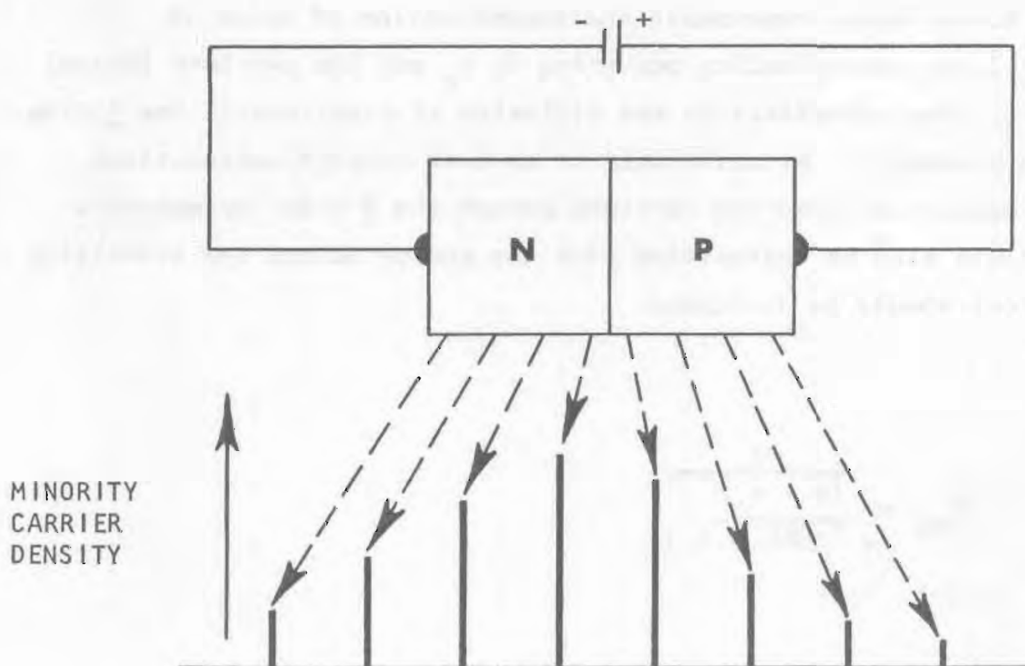
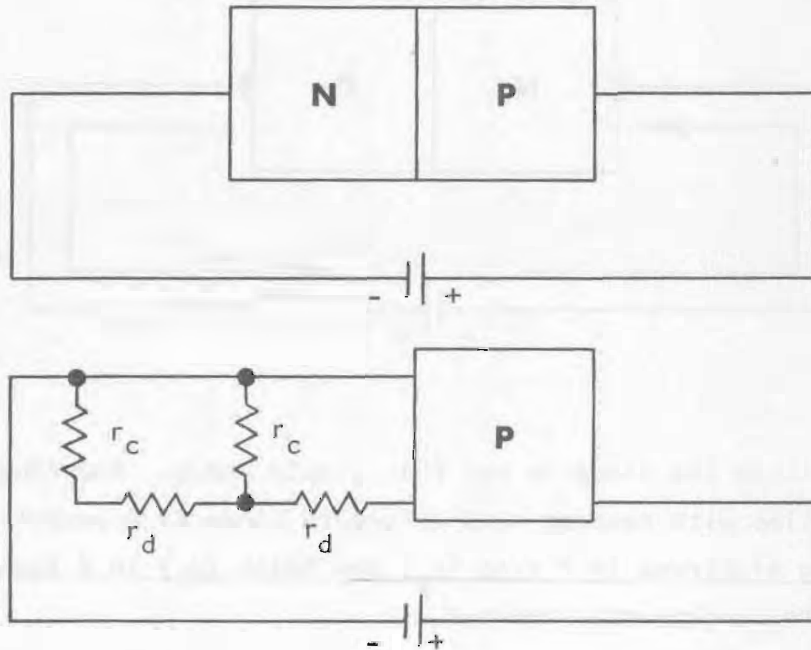


TABLE 8 (Continued)

It could be assumed that holes and electrons recombine right at the transition region of the diode and the electron current in the external circuit results, but in most cases this model is not a complete enough. One must consider the decrease in minority carriers with respect to distance (time) from the junction. This indicates a more complex model, perhaps such as shown below.



The model, as shown above, represents the recombination of holes in the N side (p_n), the recombination occurring in r_c and the carriers (holes) diffusing in r_d. The recombination and diffusion of electrons in the P side could be shown similarly. Actually this is an over simplification since the different density of minority carriers across the N side represents a charge that should also be represented plus the charge across the transition region (junction) should be indicated

From Table 3:

$$\mathcal{F}_{(e)} = \frac{1}{\epsilon^{\frac{(e - e_o)}{KT}} + 1}$$

TABLE 8 (Continued)

If $e - e_o \gg$ (several times) KT the occupation probability function reduces to:

$$\mathcal{f}_{(e)} = \epsilon^{\frac{-(e - e_o)}{KT}}$$

(Maxwell-Boltzman statistics which may be used when the probability is small)

For a PN junction, the number of electrons at the top of the barrier can be found by the formula:

$$n_B = n_n \, \epsilon^{\frac{-q\Delta_V}{KT}}$$

where n_B = density of electrons at the top of the barrier Δ_V with no external voltage applied

n_n = density of electrons in N side of junction

Δ_V = barrier or built-in voltage (see Table 7, Page 80)

With a forward voltage applied across the junction, the number of electrons at the top of the barrier is given by:

$$n_{Bv} = n_n \, \epsilon^{\frac{-q(\Delta_V - V)}{KT}}$$

where n_{Bv} = number of electrons at the top of the barrier with forward voltage applied

Δ_V = barrier voltage

V = applied voltage

The number of electrons in excess (injected) in the P region with forward voltage applied is given by:

$$n_e = n_{Bv} - n_p = n_{ne} \, \epsilon^{\frac{-q(\Delta_V - V)}{KT}} - n_p$$

where n_e = excess electrons in the P region

n_{Bv} = number of electrons at the top of the barrier

n_p = electrons normally present in the P region due to thermal generation

When the junction is at equilibrium with no external voltage applied, the number of electrons at the top of the barrier (n_B) must equal the normal number of

TABLE 8 (Continued)

electrons in the P region (n_p).

Therefore:

$$n_p = n_B = n_n \epsilon \frac{-q\Delta_v}{KT}$$

and with applied voltage, substituting in the foregoing formula:

$$n_{Bv} - n_p = n_p (\epsilon \frac{qv}{KT} - 1)$$

The number of electrons diffusing (injected) into the P region can be found by the formula:

$$I_{Dn} = \frac{q D_n n_p (\epsilon \frac{qv}{KT} - 1)}{L_n}$$

where I_{Dn} = electron diffusion current into the P region

L_n = diffusion length for electrons in the P region

$$L_n = \sqrt{D_n \tau_n}$$

where D_n = electron diffusion constant

τ_n = lifetime of electrons in the P region

The number of holes diffusing (injected) into the N region can be found by the formula:

$$I_{Dp} = \frac{q D_p p_n (\epsilon \frac{qv}{KT} - 1)}{L_p}$$

where I_{Dp} = holes diffusing into the N region

D_p = hole diffusion constant

L_p = hole diffusion length

$$L_p = \sqrt{D_p \tau_p}$$

where D_p = hole diffusion constant

τ_p = lifetime of the holes in N region

TABLE 8 (Continued)

The total current in the junction is the sum of the two currents, or:

$$I_{Dj} = I_{Dp} + I_{Dn} = \left(\frac{q\,D_p\,p_n}{L_p} + \frac{q\,D_n\,n_p}{L_n}\right)\left(\epsilon^{\frac{qv}{KT}} - 1\right)$$

When the applied voltage is negative (reverse voltage), the expression $(\epsilon^{\frac{qv}{KT}}$ approaches -1 and the reverse current or saturation current is essentially:

$$I_s = \left(\frac{q\,D_p\,p_n}{L_p} + \frac{q\,D_n\,n_p}{L_n}\right) \quad (-1)$$

Factoring out q and substituting for L_p and L_n, the formula becomes:

$$I_s = \left(p_n\sqrt{\frac{D_p}{\tau_p}} + n_p\sqrt{\frac{D_n}{\tau_n}}\right) q \ (-1)$$

and the diode current can be found by the formula:

$$I = I_s \ \left(\epsilon^{\frac{qv}{KT}} - 1\right)$$

The current in the external circuit with forward voltage applied is dependent on the number of minority carriers adjacent to the junction which is related exponentially to the applied voltage and temperature. The voltage on the N side (N_E) is proportional to the number of holes in the N material adjacent to the junction (p_n) as shown by:

$$N_E = p_n \ \left(\epsilon^{\frac{qv}{kT}} - 1\right)$$

and the voltage on the P side (P_E) is dependent on the number of electrons (n_p) adjacent to the junction shown by:

$$P_E = n_p \ \left(\epsilon^{\frac{qv}{kT}} - 1\right)$$

TABLE 8 (Continued)

K = boltzmanns constant, T = temperature, V = applied voltage, and q = charge on the electron.)

The same expressions hold with reverse voltage applied; however the sign of the voltage changes and the current changes direction. The minority carriers adjacent to the junction with reverse voltage applied, are primarily available due to the formation of hole-electron pairs by thermal energy, while the minority carriers adjacent to the junction with forward voltage applied are primarily injected from the opposite side of the junction.

If, however, the diode is switched rapidly from a forward voltage to a reverse voltage, there is a large number of minority carriers adjacent to the junction at the time of switching. These are the minority carriers that have not had the opportunity to recombine during the forward conduction process. This is sometimes referred to as diode stored charge.

We can expect a fairly large current after a forward conducting diode is switched with a reverse voltage, until the stored charge is depleted. The current will then be limited by the number of minority carriers adjacent to the junction as a result of thermal energy (disregarding any surface leakage, etc.) and is termed I_s, diode saturation current.

The hole current (I_p) varies as N_E and the electron current (I_n) varies as P_E, and the sum of the two ($I_n + I_p$ = I external) is the current in the external circuit, thus $\frac{N_E}{P_n} = \frac{P_E}{n_p} = (\epsilon^{\frac{qv}{kT}} - 1)$, so I external varies as the

$(\epsilon^{\frac{qv}{kT}} - 1)$ or; I external $= I_s\ (\epsilon^{\frac{qv}{kT}} - 1)$.

From this, we can see that we have to deal with the transport of electrons and the transport of holes and the recombination of both, plus the charge represented by the difference in minority carrier density for both holes and electrons in the two sides of the juncti[illegible]

TABLE 8 (Continued)

Lumped Semiconductor Models:

Hedgedus (4) outlines a more rigorous measurement approach; however, the less demanding models of John Linvill and J. F. Gibbons (1, 2) and the continuance and expansion of these by P. E. Gray, D. DeWitt, A. R. Boothroyd, along with Gibbons (3) offer a refreshing approach to semiconductors.

In developing a model for a PN junction diode, the following key approximations are made:

a. Holes move in N material primarily by diffusion and electrons move in P material primarily by diffusion.

b. The minority carrier densities adjacent to the junction are exponentially related to the applied voltage and hole and electron currents are continuous through the junction.

$$I \text{ external} = I_s \left(\frac{qv}{kT} - 1\right)$$

c. Essentially all the applied voltage appears across the junction. The voltage drops across the N and P regions are negligible because of the high density of majority carriers. Space charge neutrality applied.

$$P_E = n_p \left(\epsilon \frac{qv}{KT} - 1\right) \qquad N_E = p_n \left(\epsilon \frac{qv}{KT} - 1\right)$$

where: P_E = voltage on the P side

N_E = voltage on the N side

n_p = electron density in the P side adjacent to the junction

p_n = hole density in the N side adjacent to the junction

q = charge on the electron

v = voltage applied

K = Boltzmanns Constant

T = temperature

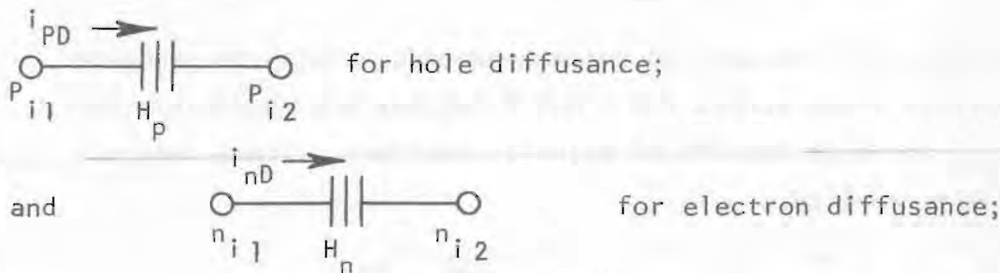
$P_E + N_E$ = voltage across the transition region or junction

TABLE 8 (Continued)

d. The N and P regions are connected to ohmic contacts of negligible resistance.

Since the relationship of current and voltage in a PN diode is exponential, it is desired to deal with current in terms of charge and time rather than voltage. The time constants involved are determined by the physical and electrical characteristics of the semiconductor material used. Among these are the lifetime of the minority carriers, the mobility constant of majority carriers, the diffusion constant of the minority carriers, etc. Rather than using conventional electrical symbols when dealing with models of semiconductors, the parameters are represented by their own symbols.

John Linvill and J. F. Gibbons (1, 2, 3) use the following symbols to indicate the diffusion transport mechanism and the recombination mechanism rather than the resistors, r_c and r_d, that were used in the simple model previously discussed. The parameters are: "<u>diffusance</u>" indicated by the symbols (formulated in Table 6):



Where: n_i = electron density

p_i = hole density

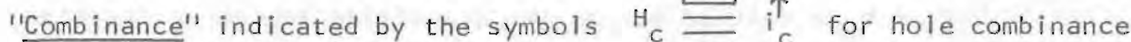
i_{nD} = electron diffusion current

i_{pD} = hole diffusion current

H_p = the parameter, hole diffusance

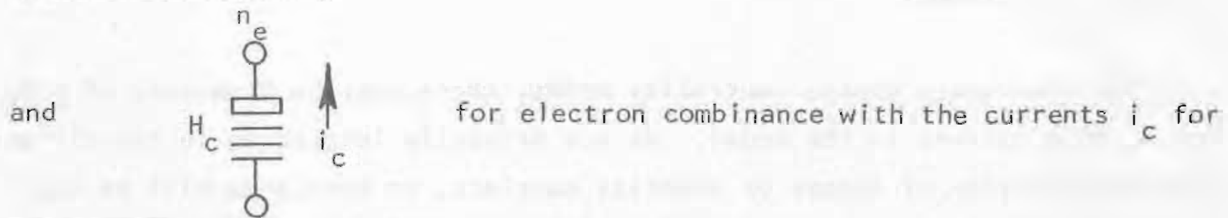
H_n = the parameter, electron diffusance

and the current $i_{PD} = (p_{i1} - P_{i2})\ H_p$; and the current $i_{nD} = (n_{i1} - n_{i2})\ H_n$.

"<u>Combinance</u>" indicated by the symbols 

for hole combinance

TABLE 8 (Continued)



and [symbol] for electron combinance with the currents i_c for hole combinance = $p_e H_C$: i_c for electron combinance = $n_e H_C$.

Where: n_e = excess electron density

p_e = excess hole density

i_c = combinance current

H_c = the parameter, hole combinance

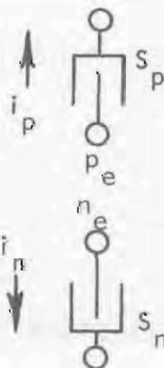
H_c = the parameter, electron combinance

$H_c = \frac{A L q}{\tau}$

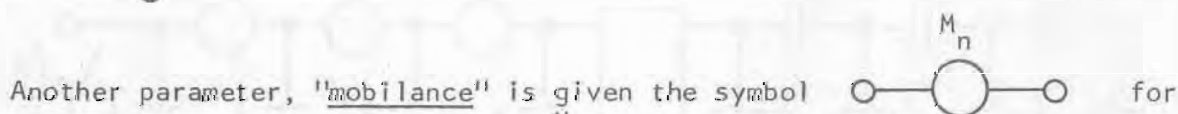
τ = carrier lifetime (refer to diagram in Table 6)

NOTE: n_e and p_e are used rather than n_i and p_i to denote those minority carriers in excess (injected) of the minority carriers present as a result of the formation of hole-electron pairs by thermal energy.

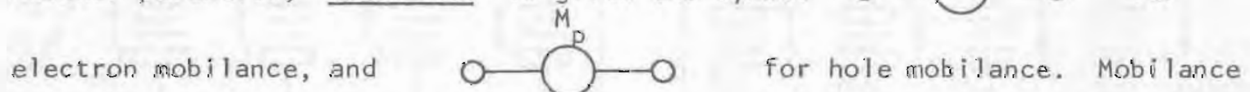
In addition, "<u>storance</u>" is the parameter representing the storage (lumped difference in density) of excess and minority carriers and given the symbols



for <u>hole storance</u>. The current $i_p = S_p \frac{dp}{dt}$. The symbol is for <u>electron storance</u>. The current $i_n = S_n \frac{dn}{dt}$.

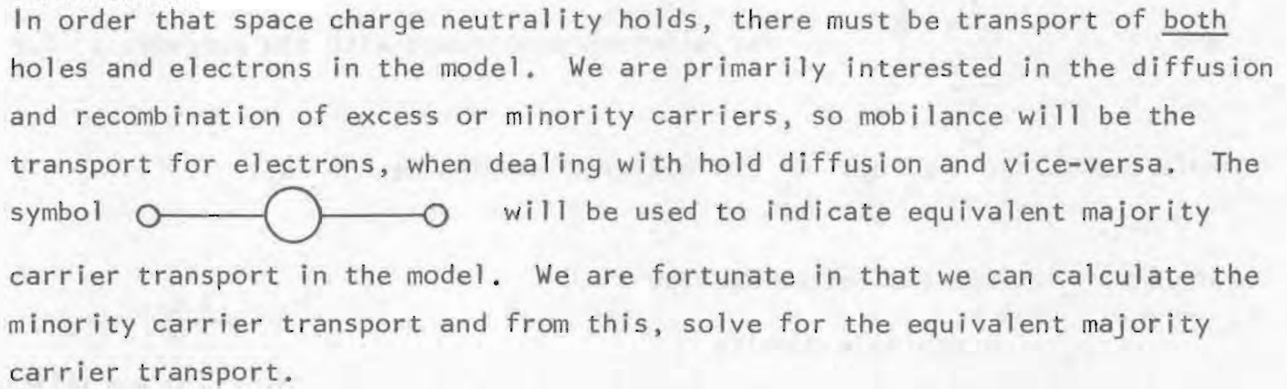


Another parameter, "<u>mobilance</u>" is given the symbol [symbol] for



electron mobilance, and [symbol] for hole mobilance. Mobilance is the parameter dealing with the transport of majority carriers and will be governed by the majority carriers mobility constant of the particular material used. It is well to remember that the important approximations stated, "<u>electrons move by diffusion in P material and holes move by diffusion in N</u>

TABLE 8 (continued)

In order that space charge neutrality holds, there must be transport of <u>both</u> holes and electrons in the model. We are primarily interested in the diffusion and recombination of excess or minority carriers, so mobilance will be the transport for electrons, when dealing with hold diffusion and vice-versa. The symbol  will be used to indicate equivalent majority carrier transport in the model. We are fortunate in that we can calculate the minority carrier transport and from this, solve for the equivalent majority carrier transport.

(The ratio of the mobility constant (μ_n or μ_p) to the diffusion constant (D_n or D_p) is equal to:

$$\frac{\mu_p}{D_p} = \frac{\mu_n}{D_n} = \frac{q}{KT}$$

where q is the charge on the electron, K is boltzmanns constant, and T is temperature.)

Neglecting the storance, an equivalent circuit for a diode can be shown using diffusance and combinance parameters and separating the electron and hole transports in the equivalent.

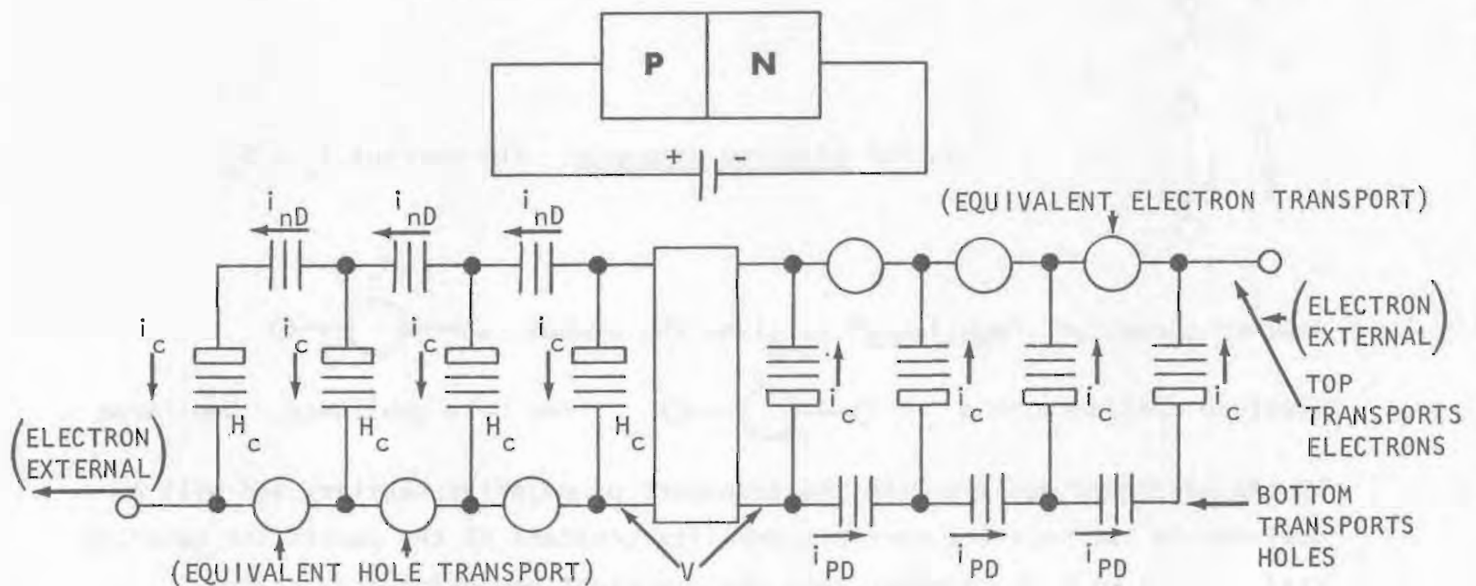
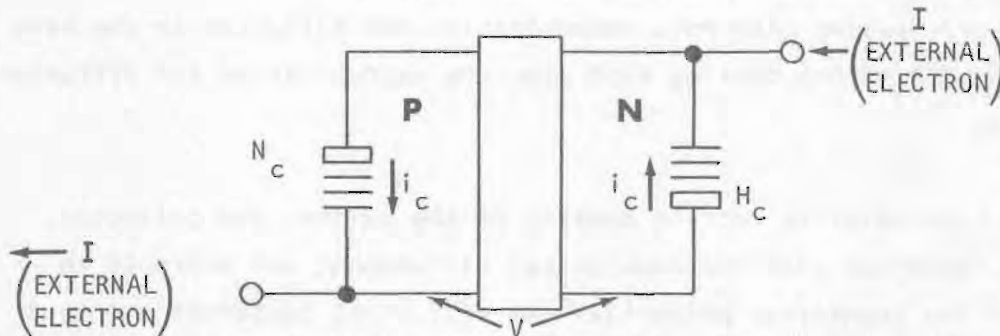
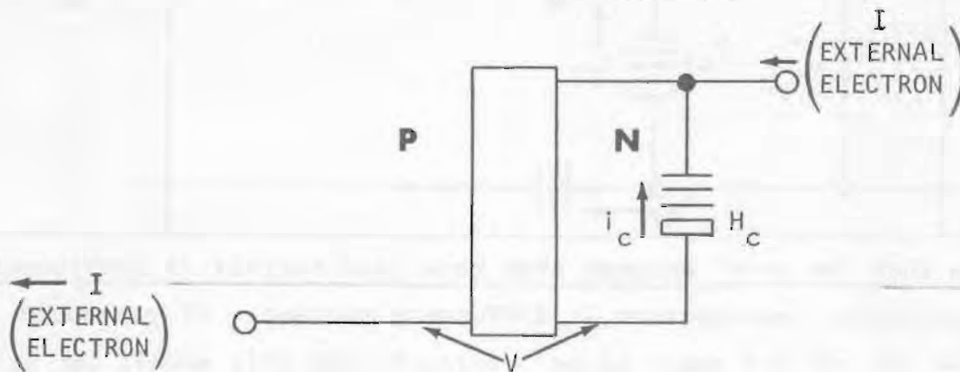


TABLE 8 (Continued)

Typically, the combinances are "lumped" and represented by one symbol with the diffusances disregarded. (When combinance current is the prime interest)



Since from the external diode terminals it is difficult to determine whether it is recombination of holes or electrons that is predominate in the diode, a further lumping is sometimes used in which it is assumed that recombination of one or the other is predominant and the diagram reduces to: (one or the other will be the predominant factor).



To take into account the storage of carriers in the device, the parameter storance is added.

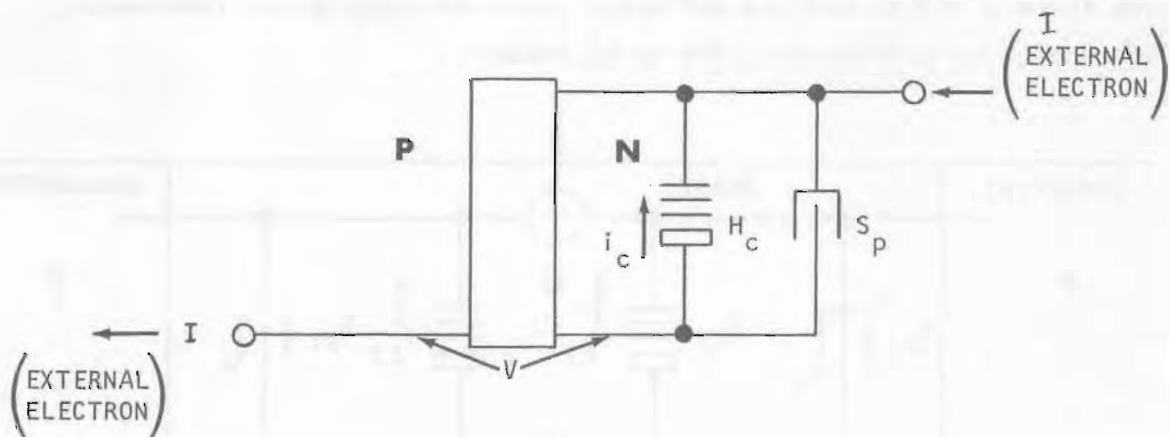
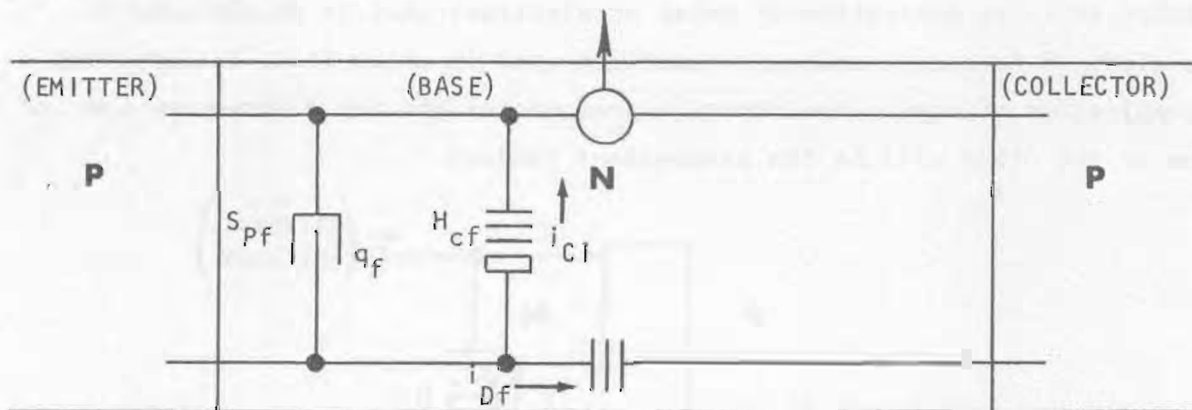


TABLE 8 (Continued)

When dealing with a transistor, the type of transistor tells us which type of minority carriers will be predominant in the transport in the base. In the PNP we are dealing with hole recombination and diffusion in the base and with the NPN we are dealing with electron recombination and diffusion in the base.

Due to the high majority carrier density in the emitter and collector, we will be concerned with the combinance, diffusance, and storance in the base of the transistor primarily. We will first construct a lumped model using the PNP transistor unsaturated, as an example:



You will note that the model assumes that base lead current is combinance current and collector lead current is diffusance current. If you would like to pursue the d-c and small signal approach with this model, the references listed (1, 2, 3) are the source. You will find more parameters such as "driftance" which will be neglected in this writing.

When a transistor becomes saturated, the collector and emitter junctions are forward biased and the collector junction contributes combinance, storance, and diffusance. The model becomes:

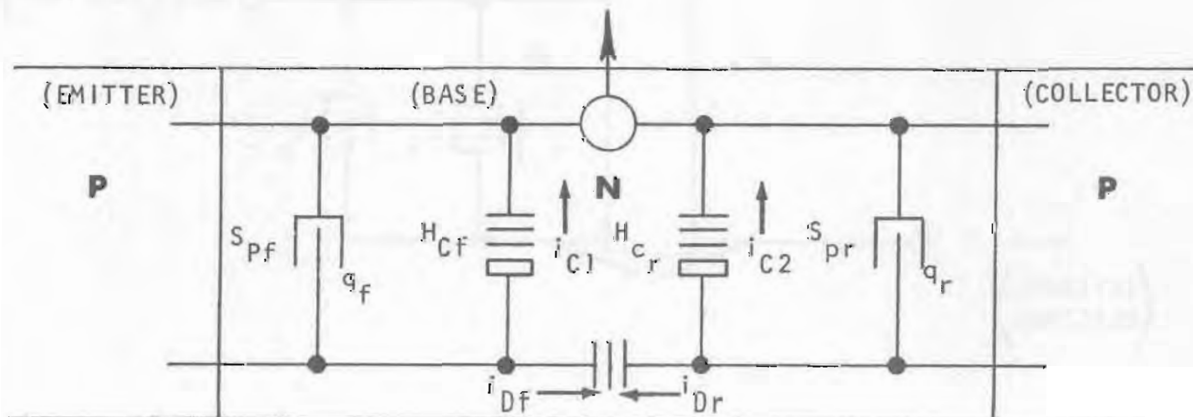
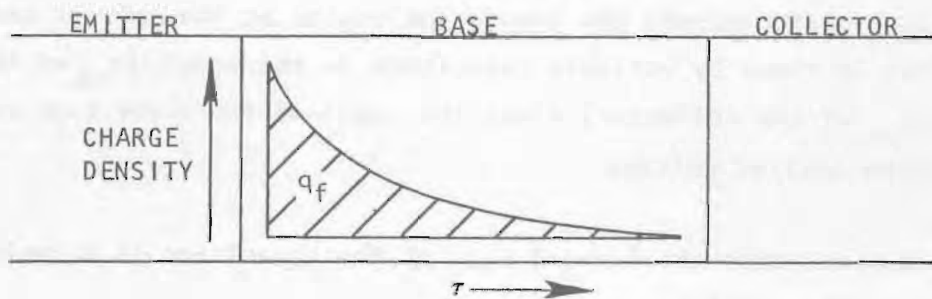


TABLE 8 (Continued)

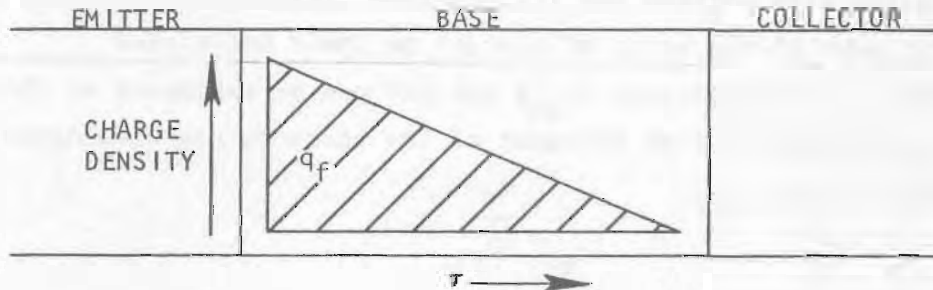
(You will note that in saturation, the total diffusance current becomes the algebraic sum of i_{Df} and i_{Dr}).

In the model, the subscript f deals with the forward biased emitter junction and the subscript r deals with collector junction effects when it becomes forward biased (saturation).

The distribution of excess carriers in the base with the emitter junction forward biased is approximated by the diagram below:



In lumping the storance, the parameter q_f approximates a distribution shown by the diagram below which throws in some error by assuming a linear distribution:



The collector becoming forward biased causes a distribution approximated by the diagram below. (Once again, assuming a linear distribution).

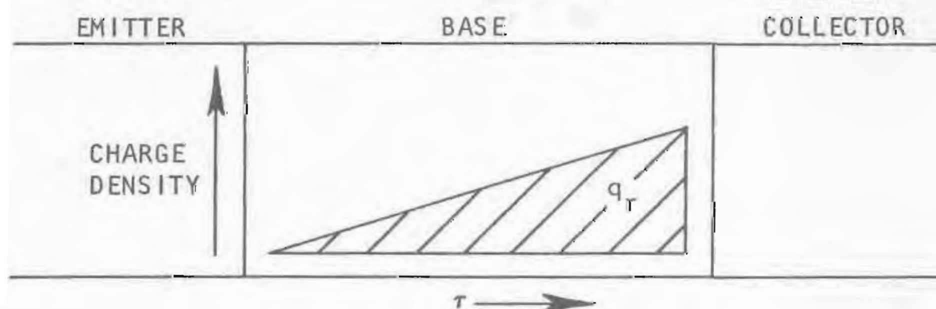
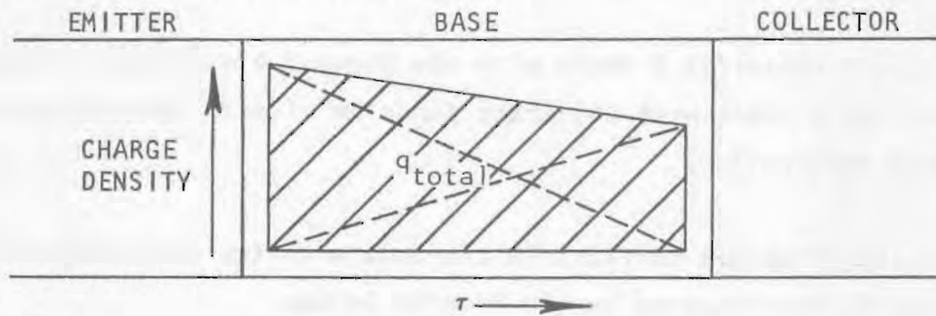


TABLE 8 (Continued)

and the resultant approximate total distribution with the transistor in saturation is shown in the diagram below:



There is also a charge across the transition region at the emitter and collector that is shown by variable capacitors in the model (q_{ve} at the emitter and q_{vc} at the collector) since the width of the transition region varies with the applied voltage

It becomes apparent that the forward beta of the transistor is a ratio of the currents i_{Df} and i_{cl}.

$$\beta_f = \frac{i_{Df}}{i_{cl}}$$

To avoid dealing with voltage and its problems, the currents can be expressed in terms of the ratio of charges to their associated time constants. The combinance (i_{cl}) current can be expressed as the ratio of the charge q_f and the time constant of the storance and combinance which has the symbol τ_{β_f}.

$$i_{c_l} = \frac{q_f}{\tau_{\beta_f}}$$

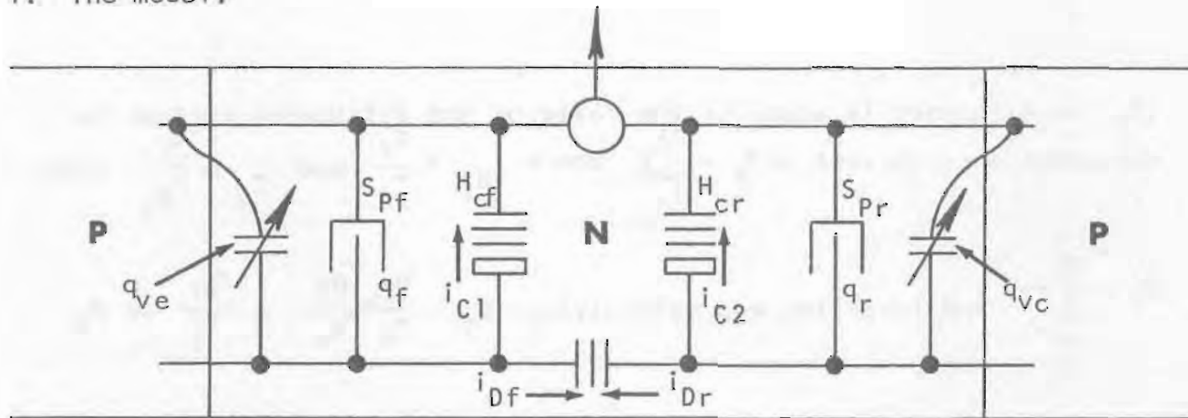
The diffusion current (i_{Df}) can be expressed as the ratio of the charge q_f to the time constant of the diffusance and the storance termed τ_f.

$$i_{Df} = \frac{q_f}{\tau_f}$$

TABLE 8 (Continued)

To summarize the model:

1. The model:



2. The charge in S_{pf} due to the difference in excess carrier density in the forward conducting base is designated q_f.

3. The time constant of the diffusance and forward storance is designated τ_f.

4. The forward control time constant (the time constant of the forward combinance and storance) is designated τ_{β_f}.

5. The collector (reverse) control time constant (the time constant of the reverse storance and combinance) is designated τ_{β_r}.

6. The charge S_{pr} in the base due to the collector being forward biased is designated q_r.

7. The time constant of the reverse storance and the diffusance is designated τ_r.

8. The forward diffusance current is designated i_{D_f} and is equal to the charge q_f divided by the time constant τ_f.

$$i_{D_f} = \frac{q_f}{\tau_f}$$

9. The forward combinance current is designated i_{C1} and is equal to the forward conducting charge in the base, q_f divided by the time constant τ_f.

$$i_{C1} = \frac{q_f}{\tau_f}$$

TABLE 8 (Continued)

10. (β_f) Beta forward is equal to the ratio of the diffusance current to the combinance current or $\beta_f = \frac{i_{D_f}}{i_{C_1}}$ where $i_{D_f} = \frac{q_f}{\tau_f}$ and $i_{C_1} = \frac{q_f}{\tau_{\beta_f}}$ then

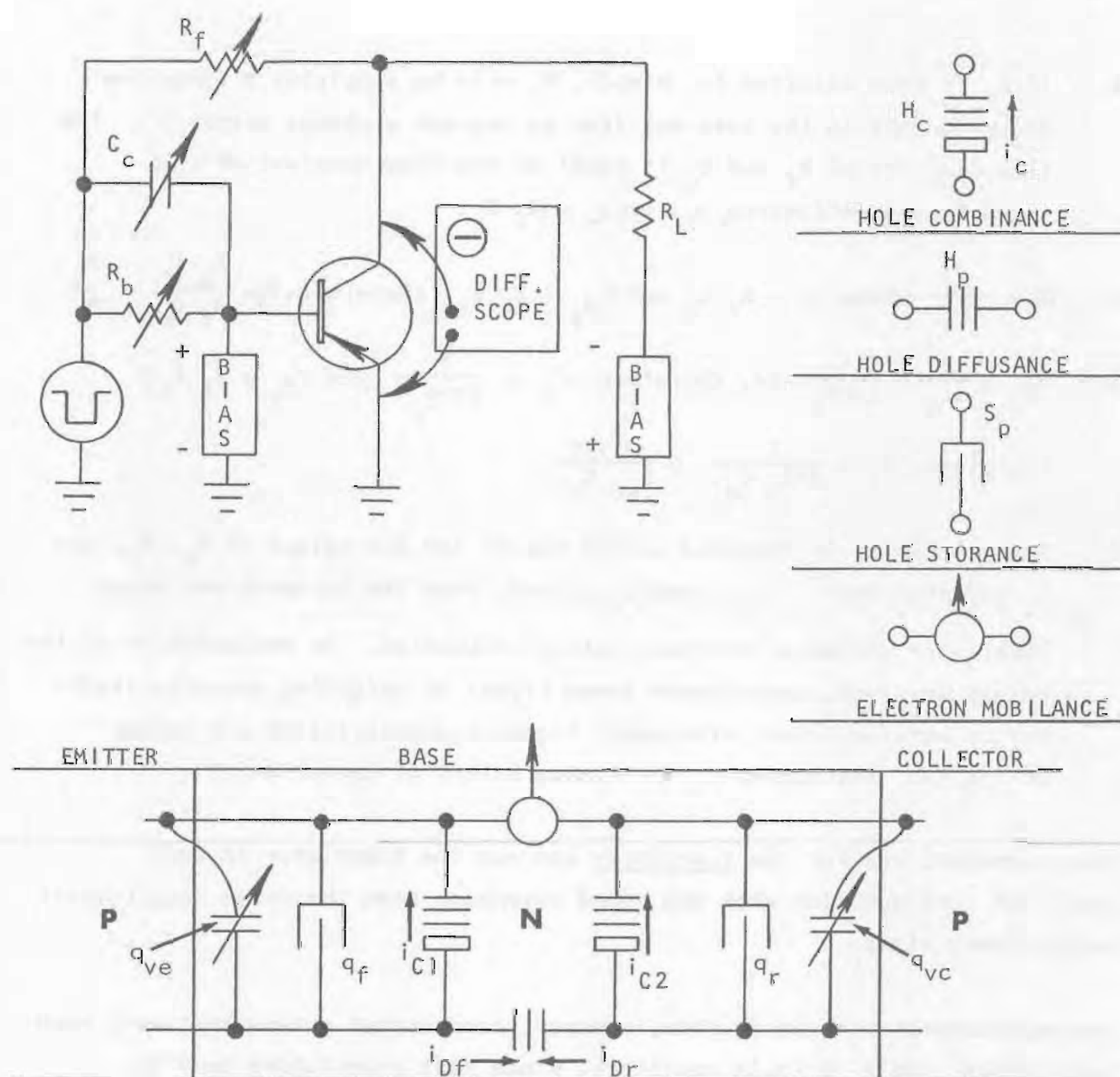
$$\beta_f = \frac{\frac{q_f}{\tau_f}}{\frac{q_f}{\tau_{\beta_f}}}$$ and inverting and multiplying; $\beta_f = \frac{q_f \, \tau_{\beta_f}}{\tau_f \, q_f} = \frac{\tau_{\beta_f}}{\tau_f}$ or β_f

is equal to the control (combinance) time constant divided by the (diffusance) transport time constant.

11. When the collector becomes forward biased, whether by applied bias or by being driven into saturation, there is storance, diffusance, and combinance as a result, and the currents are formulated in the same manner as the forward condition, with the subscript r indicating those values associated with the collector junction. i_{D_r} opposes i_{D_f} which explains the minus sign in the formula for i_{D_r}.

12. τ_{β_f}, the control charge time constant is essentially the reciprocal of ω_t or: $\tau_{\beta_f} = \frac{1}{\omega_t} = \frac{1}{2\pi f_t}$.

TABLE 8 (Continued)



Measurements:

The following measurement approach was used by a graduate student under Linvill at Stanford.

A. The transistor in the schematic diagram shown above is biased quiescently non-conducting. A fast rising pulse generator attempts to turn on the transistor and the values of R_b and C_c adjusted for proper speed-up action. When this is accomplished, the time constant $R_b\ C_c$ is equal to the time constant of the storance and combinance, or $\tau_{\beta_f} = R_b\ C_c$.

TABLE 8 (Continued)

B. If R_f is then adjusted for a null, R_f will be supplying β times the drive current in the base resistor to prevent a change across R_L. The time constant of R_f and C_c is equal to the time constant of the storance and diffusance, τ_f, or: $\tau_f = R_f\ C_c$.

C. $\beta_f = \frac{\tau_{\beta_f}}{\tau_f}$ where $\tau_f = R_f\ C_c$ and $\tau_{\beta_f} = R_b\ C_c$, therefore, $\beta_f = \frac{R_b\ C_c}{R_f\ C_c} = \frac{R_b}{R_f}$.

D. $\tau_{\beta_f} = \frac{1}{\omega_t} = \frac{1}{2\pi f_t}$ and, therefore, $f_t = \frac{1}{2\pi\tau_{\beta_f}}$ and $\tau_{\beta_f} = R_b\ C_c$,

therefore, $f_t = \frac{1}{2\pi R_b\ C_c} = \frac{0.159}{R_b\ C_c}$.

E. The transistor is reversed in its socket and the values of R_b, R_f, and C_c adjusted and τ_r, τ_{β_r}, and β_r gained, then the currents and other reverse or collector characteristics calculated. By manipulation of the values involved, approximate large signal or switching characteristics may be obtained along with small signal characteristics and values of internal reactances at difference points of operation.

The parameters are for the <u>transistor</u> and not the transistor in some specified configuration with the added advantage that the scope requirements are not very rigid.

The measurements are easily made, however, compensated attenuators work much better than simple variable resistors, since fast transistors tend to oscillate in the basic circuit shown.

Formulas for the measurement configuration on page 14:

<u>Forward</u>:

$$i_{D_f} = \frac{q_f}{\tau_f}$$

$$i_{C_1} = \frac{q_f}{\tau_{\beta_f}}$$

$$\beta_f = \frac{\tau_{\beta_f}}{\tau_f}$$

TABLE 8 (Continued)

$$\tau_f = R_f\ C_c$$

$$\tau_{\beta_f} = R_b\ C_c$$

$$\tau_{\beta_f} = \frac{1}{\omega_t} = \frac{1}{2\pi f_t}$$

$$\beta_f = \frac{\tau_{\beta_f}}{\tau_f} = \frac{R_b\ C_c}{R_f\ C_c} = \frac{R_b}{R_f}$$

$$\beta_f = \frac{R_b}{R_f}$$

$$\tau_{\beta_f} = \frac{1}{\omega_t} = \frac{1}{2\ \pi f_t} = R_b\ C_c$$

$$R_b\ C_c = \frac{1}{2\ \pi f_t}$$

$$f_t = \frac{1}{2\ \pi R_b\ C_c} = \frac{0.159}{R_b\ C_c}$$

Reverse: *(Note, transistor reversed in socket)

$$i_{D_r} = \frac{-q_r}{\tau_r}$$

$$i_{C_2} = \frac{q_r}{\tau_{\beta_r}}$$

$$\beta_r = \frac{\tau_{\beta_r}}{\tau_r}$$

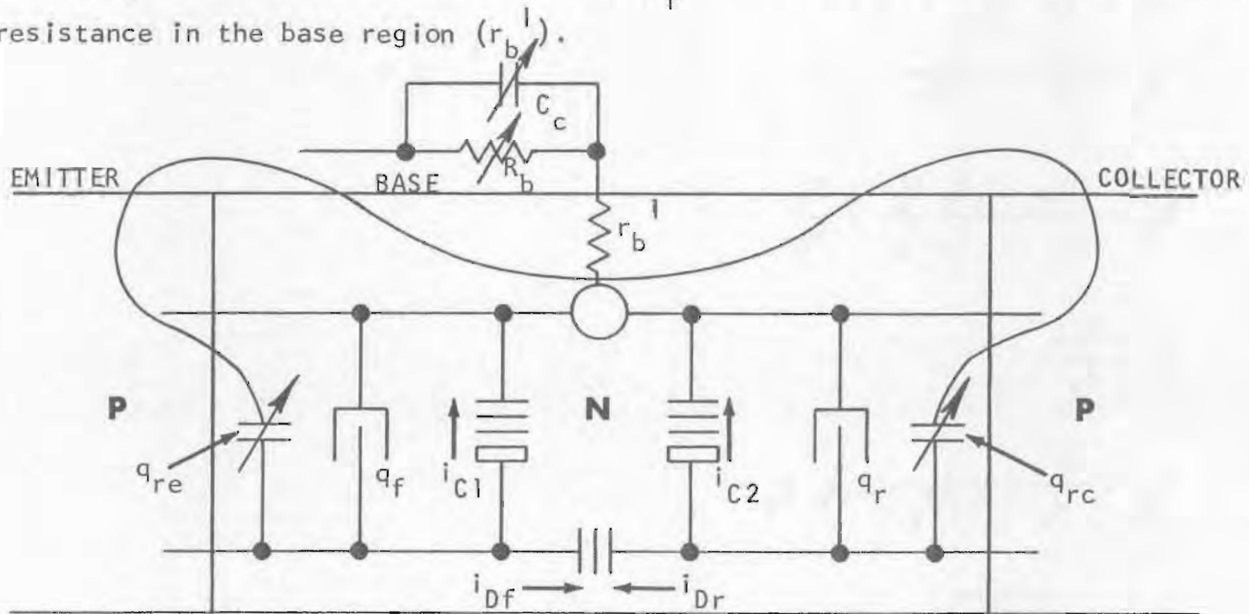
$$\tau_r = {*}R_f\ C_c$$

$$\tau_{\beta_r} = {*}R_b\ C_c$$

TABLE 8 (Continued)

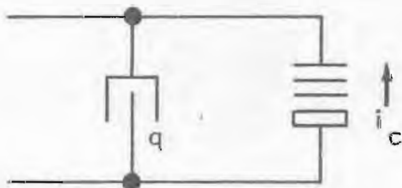
Limitations Of This Simple Measurement System:

The R_b C_c time constant used to measure τ_{β_f} cannot be exact, due to the finite resistance in the base region (r_b').



This model suggests why the aim in high frequency transistors is for a low r_b' and a low emitter and collector capacitance.

For fast switching operation, it is desired to have a low storance and combinance time constant.

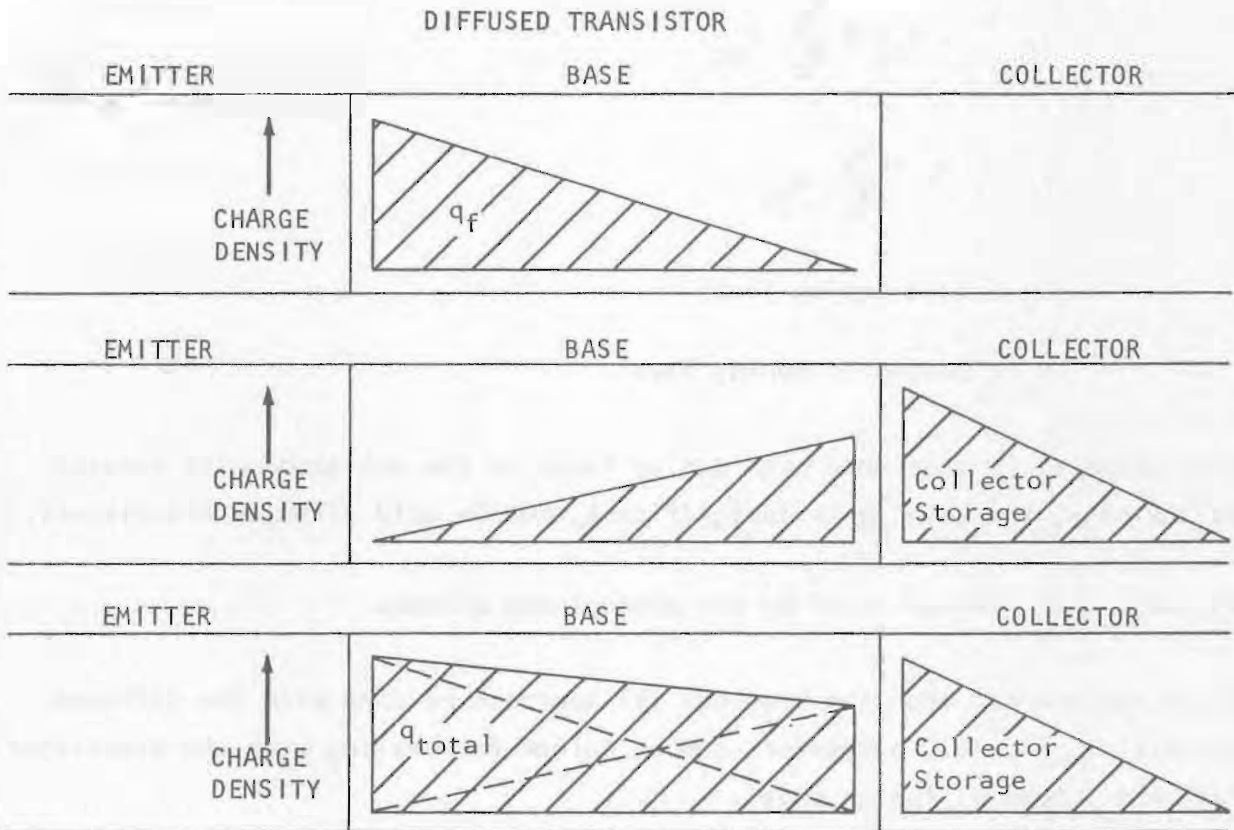


$$\tau = \frac{q}{i_c} \qquad \tau_{\downarrow} = \frac{q}{i_{c\uparrow}}$$

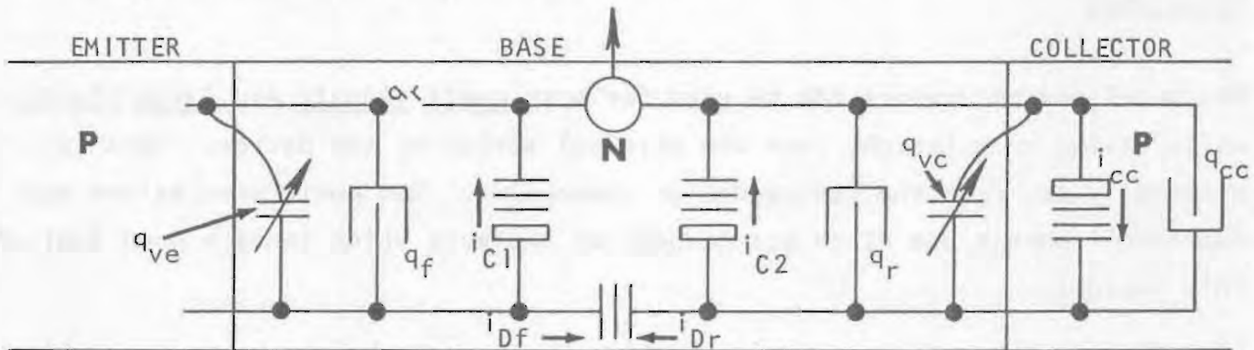
For the decrease in the time constant, gold can be diffused (diffused to indicate the doping process by diffusion and not carrier diffusion) into the base to reduce the minority carrier lifetime and increase the combinance current.

Dealing with a majority of transistors such as the typical alloy type transistor, the model previously discussed and the method of measurements will give close approximations. With the gold diffused transistor, a plot of the excess carrier distribution shows that when the transistor is in saturation, there is storage in the collector that cannot be neglected.

TABLE 8 (Continued)



This suggests a more complete model such as:



and this model can then be lumped to:

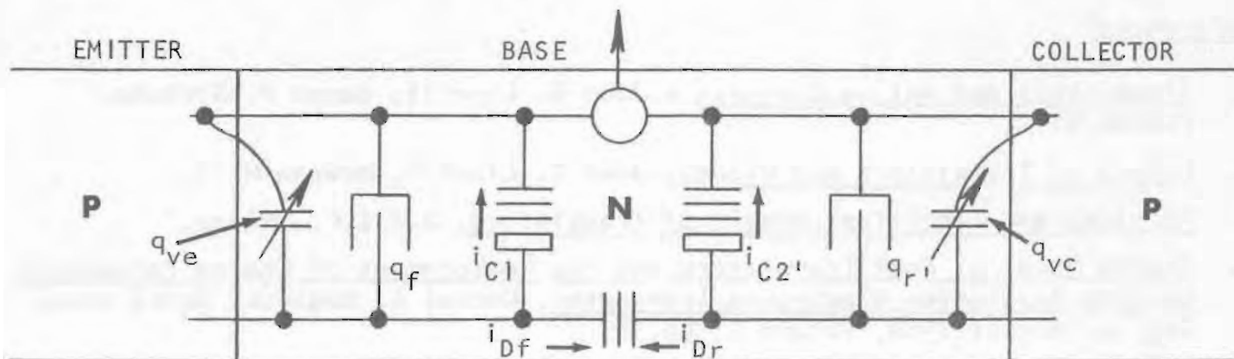


TABLE 8 (Continued)

Where:

$$i_{C_2}' = i_{C_2} + \frac{N_b}{N_c} i_{cc}$$

$$q_r' = q_r + \frac{N_b}{N_c} q_{cc}$$

N_b = base doping level

N_c = collector doping level

With transistors that have high doping level in the collector with respect to the base, the term $\frac{N_b}{N_c}$ is insignificant, but in gold diffused transistors, the collector storage could be the predominate effect.

It is recommended that the Hedgedus (4) approach be used with the diffused transistor, however, parameters can be gained for dealing with the transistor with the lumped diffusion model.

Advantages:

The model and parameters can be used for both small signals and large signals while giving much insight into the physical action of the device. This is especially so, from the device design standpoint. Two port formulations and Ebers-Moll models use black box methods of analysis which lacks a good deal of this insight.

References:

1. Transistors and Active Circuits - John G. Linvill, James F. Gibbons, McGraw Hill.
2. Models of Transistors and Diodes, John G. Linvill, McGraw Hill.
3. Physical and Electrical Models of Transistors, S.E.E.C., Wiley.
4. Charge Model of Fast Transistors and the Measurement of Charge Parameters By High Resolution Electronic Integrator, Cornel L. Hegedus, Solid State Design, August 1964, Volume 5, No. 8.
5. Transistors and Active Circuits, Linvill and Gibbons, McGraw Hill.

TABLE 9

EMITTER INJECTION

Since of necessity the base of a transistor has a width less than the diffusion length of the injected carriers, the formulations for current in a junction in Table 8 are not entirely valid when applied to transistors.

Consider an NPN transistor with forward bias applied to the emitter-base junction and reverse bias applied to the collector-base junction. The hole current from Table 8:

$$I_p = \frac{q\,A\,D_p\,p_e}{L_{pe}}\left(\epsilon^{\frac{q\,V_e}{KT}} - 1\right)$$

where q = charge on the electron

A = area of the junction

D_p = hole diffusion constant

p_e = hole density in emitter adjacent to the junction

L_{pe} = diffusion length of holes in the emitter

V_e = applied emitter-base forward voltage

The electron current, however, is effected by the collector bias and the formula is modified to take into account the width of the base:

$$I_n = -\frac{q\,A\,D_n\,n_b}{W}\left(\epsilon^{\frac{q\,V_e}{KT}} - 1\right)$$

where D_n = electron diffusion constant

n_b = density of electrons in the base adjacent to the junction

W = width of the active region of the base

and the emitter current is:

$$I_e = -\,q\,A\left(\frac{D_n\,n_b}{W} + \frac{D_p\,p_e}{L_{pe}}\right)\left(\epsilon^{\frac{q\,V_e}{KT}} - 1\right) - \frac{q\,A\,D_p\,p_e}{L_{pe}}$$

and for a PNP transistor:

TABLE 9 (Continued)

$$I_e = - q A \left(\frac{D_p p_b}{W} + \frac{D_n n_e}{L_{ne}} \right) \left(\epsilon^{\frac{q V_e}{KT}} - 1 \right) - \frac{q D_n n_e}{L_{ne}}$$

where p_b = holes in the base adjacent to the junction

n_e = electrons in the emitter adjacent to the junction

L_{ne} = diffusion length for electrons in the emitter

THERMAL RESISTANCE CHART, ALUMINUM PLATE

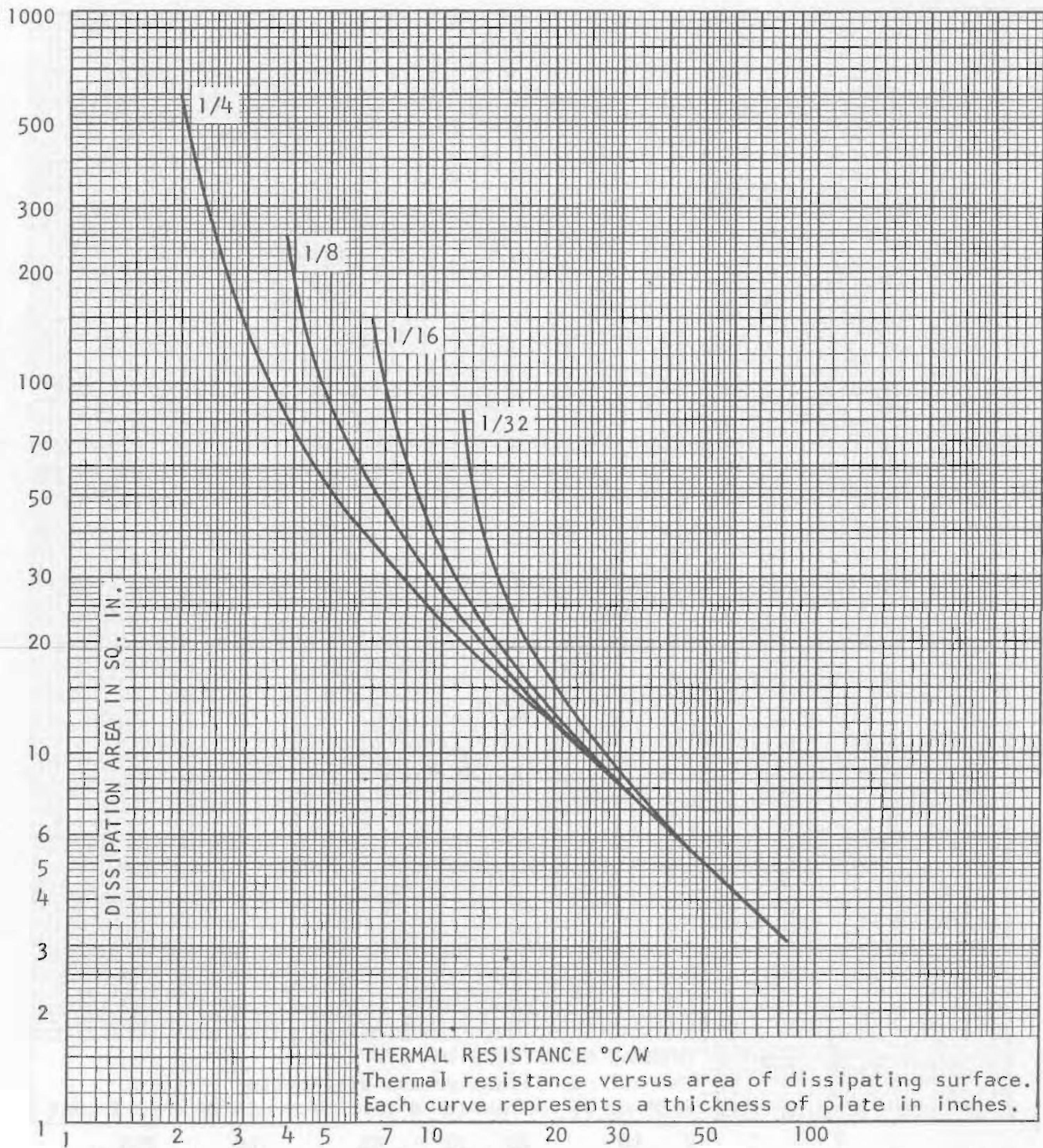


TABLE 10

THERMAL RESISTANCE CHART, COPPER PLATE

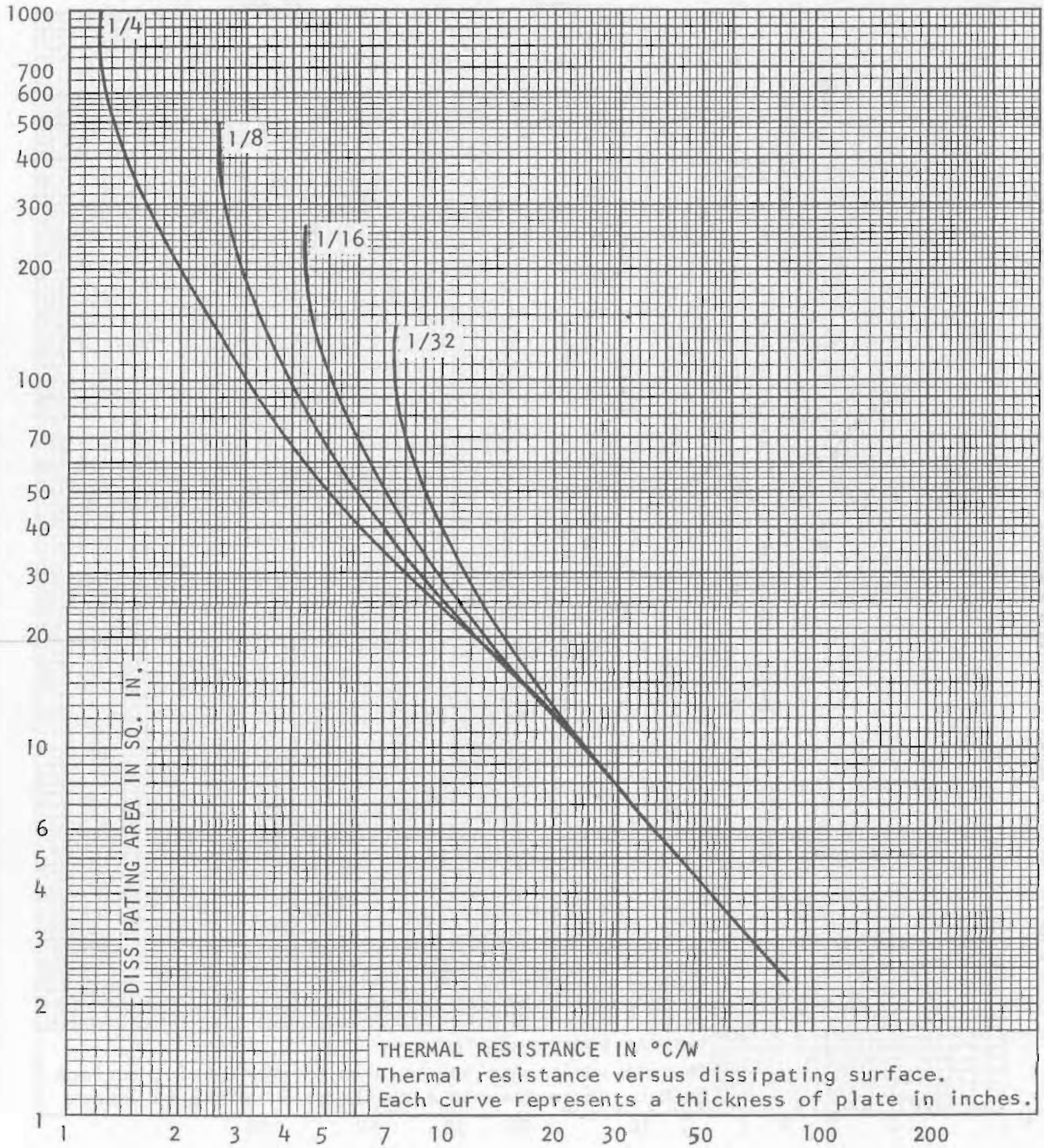


TABLE 11

THERMAL RESISTANCE CHART, BRASS PLATE

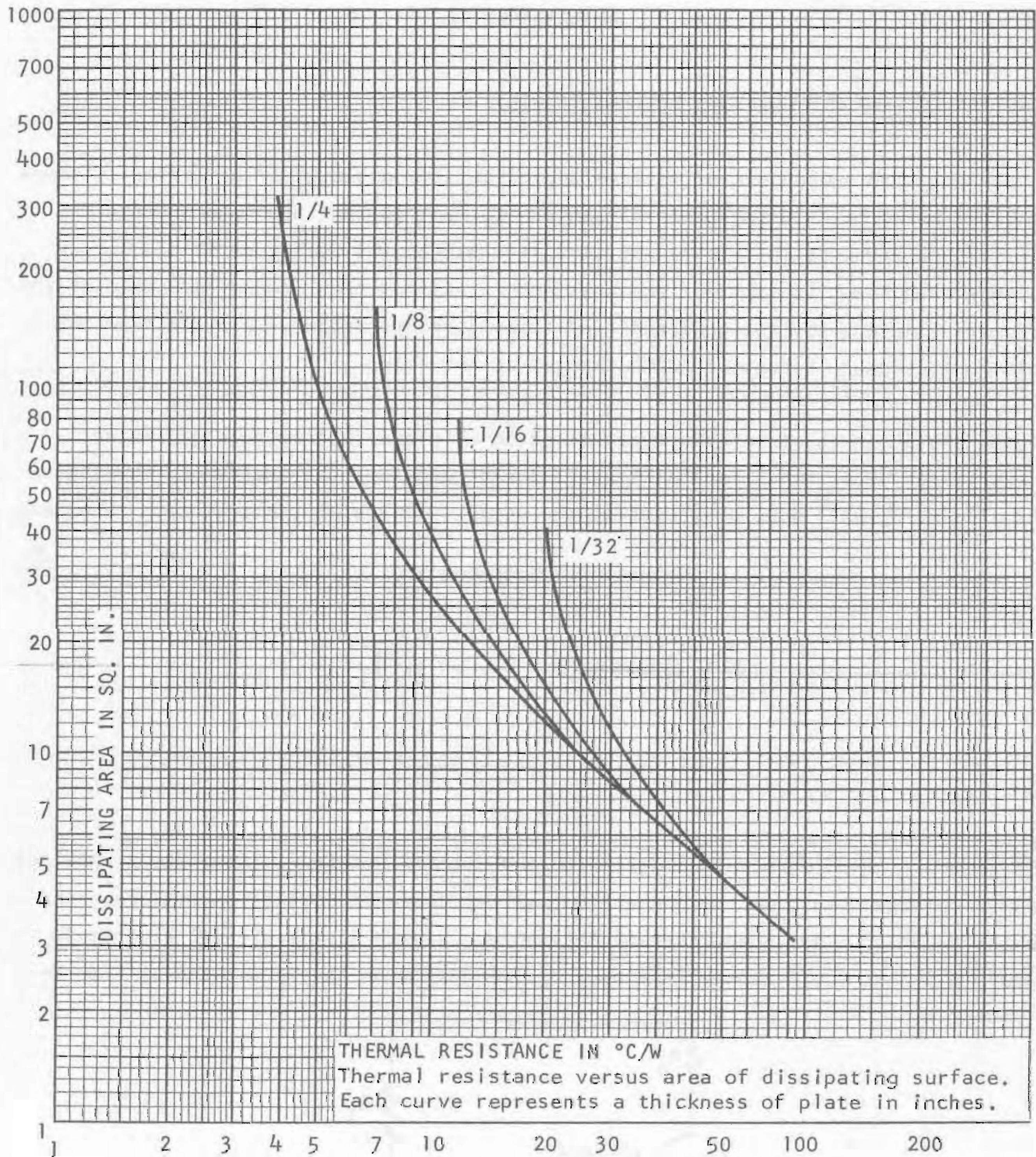


TABLE 12

SECTION THREE

TRANSISTORS

CONSTANT CURRENT AND CONSTANT VOLTAGE SOURCES:

In dealing with semiconductor devices and circuitry, a source or generator is sometimes referred to as a constant voltage or constant current source or generator. They are referred to in this manner to indicate that the magnitude of the output voltage or current is constant and does not relate to the frequency or shape of the output. Constant voltage generator indicates that the output amplitude from the voltage generator is constant with changes in the load.

Figure 1-3 shows the schematic representation of a constant voltage generator. Assuming the generator shown in Figure 1-3 has zero internal impedance, changes in the size of R_L will have no affect on the output voltage at the generator terminals.

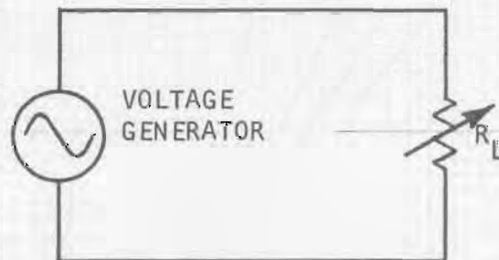


FIGURE 1-3

The generator in Figure 1-3 connected in any circuit configuration gives a constant output voltage magnitude. There are no perfect constant voltage generators. Any source of voltage has some internal impedance. Notice in the circuit diagram in Figure 2-3 that the internal impedance of the generator is represented by the resistance R_g.

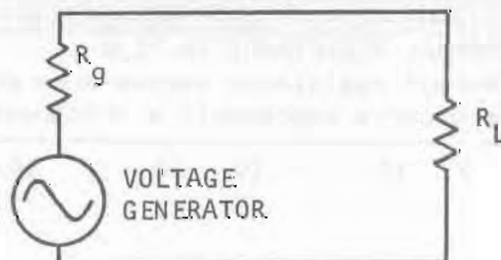


FIGURE 2-3

This resistance is internal of the voltage generator; however, in a schematic diagram, it can be represented as external of the voltage generator.

The output of the voltage generator is divided across R_g and R_L in Figure 2-3. If the voltage across R_L is equal to the voltage across R_g, then R_L and R_g are equal. Optimum power transfer occurs when this is the case. When the impedance of the load being driven is equal to the internal impedance of the driving source, optimum transfer of power occurs. There are times, however, when optimum transfer of power is not the desired result and the output current or voltage will be the prime consideration. If the ohmic value of R_g in Figure 2-3 is very small with respect to the value of R_L, the generator can be said to be a constant voltage generator (referring to magnitude) and changes in R_L or the current in the load circuit will only result in very small changes in the voltage delivered to R_L.

A generator can also be said to be a constant current generator. The diagram in Figure 3-3 shows a current generator supplying the load R_L.

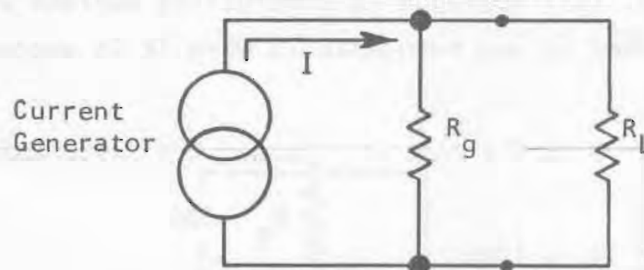


FIGURE 3-3

If the internal resistance of the generator (R_g) in Figure 3-3 is assumed to be infinity, changes in the load will have no effect on the current in the circuit. The current generator can be said to be a constant current generator (referring to magnitude).

There are no perfect constant current generators. Any current generator has a finite internal impedance. If the generator has an internal impedance that is very large with respect to the circuit, etc., that it is supplying, it can be said to be essentially a constant current source. R_g in Figure 3-3 indicates the imperfection in the current generator.

Consider the diagram in Figure 4-3. The resistor R_g is very large with respect to R_L. R_L will have only a small effect in determining the current in the circuit.

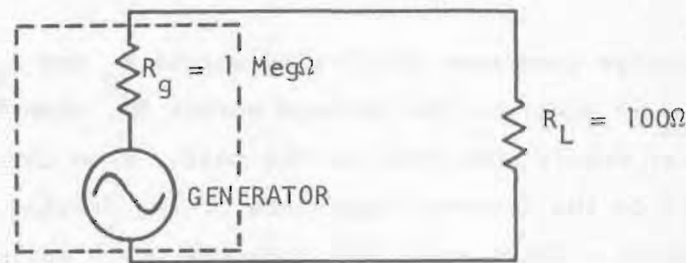


FIGURE 4-3

It can be said that using a higher than required voltage source with a high value of internal resistance is <u>synthesizing</u> a constant current source. The control of current is taken away from the circuit that the generator is supplying.

Figure 5-3 shows a generator with a variable internal resistance, R_g. Changes in the value of R_g will force changes in the current in R_L. Later in this publication and the volumes that follow, this approach to controlling current will be used. It is important that you remember it and recognize it when it is encountered again.

FIGURE 5-3

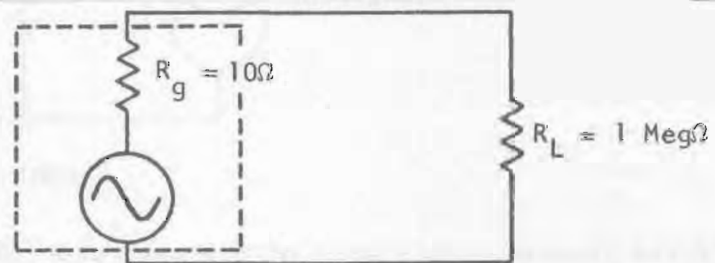


FIGURE 6-3

A semiconductor diode is a non-linear device in that the current does not change linearly with changes in applied voltage. It is typical to find a semiconductor device biased or driven by constant current sources. This is done to avoid the non-linear results when biasing or driving the device with voltage.

Consider the circuit in Figure 6-3. Is this a constant current or voltage generator? If you answered constant voltage, you are correct. The resistance of the generator is very low with respect to the load resistance.

BASIC TRANSISTORS:

As discussed in the section on diodes, the number of carriers that recombine is dependent on the number of imperfections or recombination traps that these carriers can find. A forward biased PN junction has minority carriers injected into the two sides. N side injects electrons into the P side, and the P side injects holes into the N side. If the minority carrier lifetime is short (many imperfections), the carriers recombine rapidly and the current in the external circuit is primarily a result of recombination. The carriers must diffuse before they can recombine and a period of time is involved in the recombination process. The average time that the carriers take in the recombination process is termed the minority carrier lifetime of the material. The minority carrier lifetime for intrinsic or pure semiconductor structure is long because of the small number of impurities. As doping impurities are added to a semiconductor material, the minority carrier lifetime is reduced. Figure 7-3 shows a PN junction which has the N side doped relatively heavy, and the P side doped lightly. The statement doped lightly indicates that relatively few impurities or imperfections have been added in the doping process.

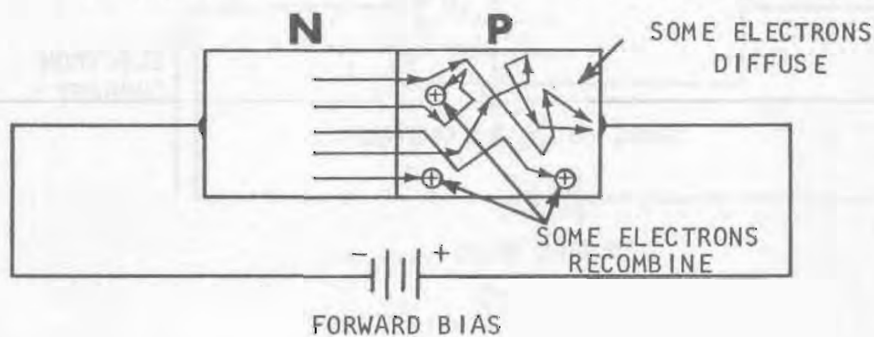


FIGURE 7-3

When forward bias is applied to the PN junction, electrons are injected from the N side into the P side and holes are injected from the P side into the N side. The number of electrons injected into the P side is large with respect to the number of holes injected into the N side with the doping levels in Figure 7-3. The number of electrons injected into the P side is much greater than the imperfections that these electrons must find to accomplish recombination. Many of the electrons diffuse through the P side of the material without recombining when the P side is the lightly doped side. Current in the external circuit in such a configuration can be made up of current as a result of recombination and current carrier by the diffusion

of electrons through the P material. The longer the lifetime of the P material, the more electrons will diffuse across the P material and the less will recombine. We might also say that the narrower the P material, the more electrons will diffuse across rather than recombining. If the width of the P side is narrow, with respect to the minority carrier lifetime, most of the current in the external circuit can be a result of diffusion of electrons in the P side of the junction.

A similar result can be obtained by doping the P side relatively heavy and using light doping in the N side. When this is done, the number of holes injected from the P side into the N side is large with respect to the number of electrons injected from the N side into the P side. Since the minority carrier lifetime is relatively long in the N side, many of the holes diffuse through the N side without accomplishing recombination. The current in the external circuit is a result of both recombination and diffusion of carriers. Figure 8-3 illustrates this.

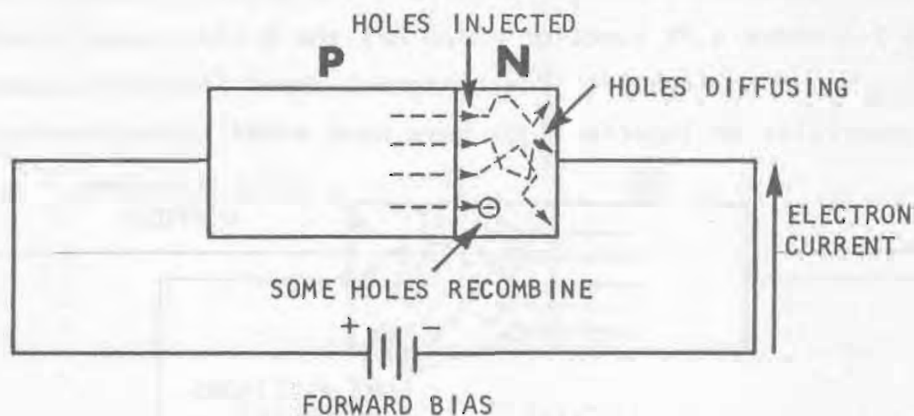


FIGURE 8-3

Light doping or adding few imperfections to one side of a PN junction results in some current by diffusion. Narrowing the lightly doped side increases the amount of current carried by diffusion. External circuit current still governs the number of minority carriers adjacent to the junction. But transport time differs because transport by diffusion is much slower than the transport of majority carriers.

The N side of Figure 8-3 has been doped relatively light and made narrow. Diffusion through the N material causes a large portion of the external circuit current. The characteristics discussed in section one apply to this junction. Constants listed in section one govern the speed of carrier movement: the specific material

diffusion constants for electrons and holes as minority carriers; and the specific material mobility constants for electrons and holes as majority carriers. Doping adds a very small number of impurity atoms compared to the atom total therefor, intrinsic material constants before doping remain valid for semiconductor diode carriers.

To summarize a bit, there are two current carrier transports involved in a semiconductor PN junction. They are mobility and diffusion. When majority carrier transport is dealt with, the mobility constant is used and when minority carrier transport is dealt with, the diffusion constant is used. The transports can become more involved if there are internal fields built into the PN junction device and this will be discussed later. At this time, we will be primarily concerned with the mobility of majority carriers and the diffusion of minority carriers. Doping one side of a PN junction lightly (giving it a long minority carrier lifetime), doping the other side relatively heavy, and making the lightly doped side narrow can result in a majority of the external current being a result of diffusing minority carriers. These carriers have been injected from the more heavily doped side.

To differentiate between minority carriers present as a result of thermal energy and those carriers that have been injected due to the application of external energy such as applied bias, the carriers that have been injected are sometimes referred to as excess minority carriers or simply excess carriers. Therefore, when we speak of excess carriers, we are speaking of carriers that have been injected by the application of some external bias, and are in excess of the number required to balance the protons in the nucleus of the particular material that they have been injected into. Table 9 is additional information and some math involved that is not necessary for the coverage in this volume. It is presented here for those that wish to pursue this further.

A simple PN junction transistor can be constructed by doping two junctions into a

single piece of germanium, silicon, or some other type of semiconductor material. In the doping process, the two ends are doped relatively heavy, while the center is doped lightly. In addition, the ends of the semiconductor are doped oppositely of the center. In other words, if the two ends are doped with N type impurities, the center would be doped with P type impurities. If the two ends are doped with P type impurities, the center will be doped with N type impurities. This is illustrated in Figure 9-3.

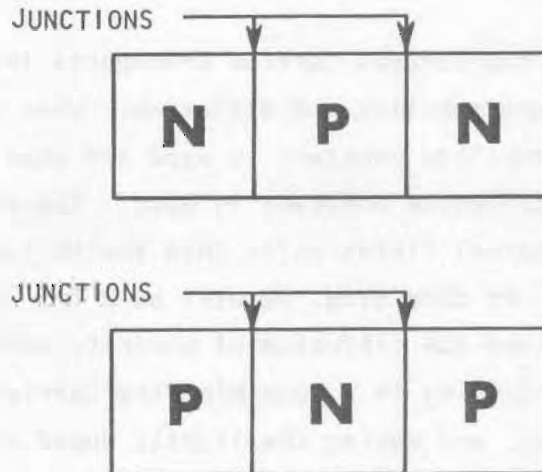


FIGURE 9-3

A transistor formed in this fashion will take on either an NPN or a PNP configuration, and there will be two distinct junctions in the material. The center portion is doped lightly and this indicates that the current in the center portion can be a result of both recombination and diffusion. With light doping in the center, the minority carrier lifetime is long and enhances the possibility of current by diffusion. There will be some imperfections in the center portion and some recombination will take place. There will be two current carrier transports involved in the center portion of the transistor. Recombination can take place in the center portion when excess carriers are injected into the center portion, and diffusing minority carriers represent a stored charge.

Section one went to great lengths discussing the action at the junction when it is formed. When two junctions are formed in a single crystal, identical action takes place at both junctions, as when the simple junction was formed in section one. Recombination will take place at the junctions until a state of equilibrium exists,

and the fermi levels are aligned. Note in Figure 10-3 that the fermi levels of the NPN transistor are aligned and that the doping is relatively light in the center portion.

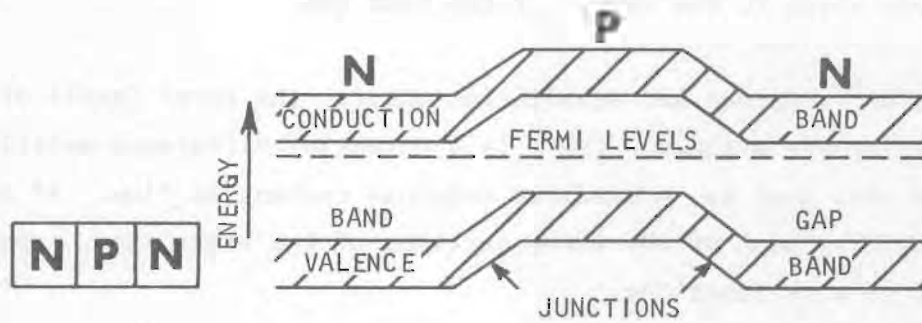


FIGURE 10-3

This is indicated by the fact that the fermi level has not been moved too far from the center of the band gap. The fermi level in the N side of the junction, however, has been moved very close to the conduction band of the N side as a result of heavier doping. When the two junctions are formed, a state of equilibrium occurs, the fermi levels line up, and there is no net current in the transistor.

If ohmic contacts (non-rectifying) are made to the three parts of the transistor, both junctions can be biased in the same manner that the basic diode is biased. It should be made clear at this point that for transistor action, one cannot simply put two diodes back to back and expect to accomplish transistor action. This will be better illustrated later; however, it is well to note that the center portion is doped lightly with respect to the ends, and that the center portion will have to be a continuous piece of material. Figure 11-3 shows the construction and energy band diagram of a PNP type transistor.

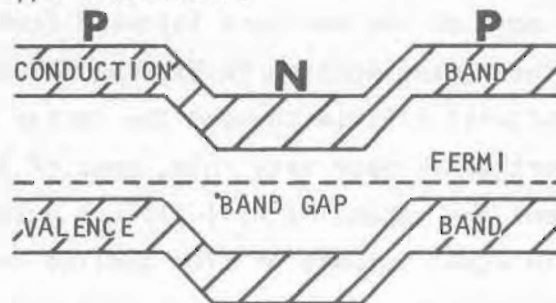


FIGURE 11-3

In this case, the center portion or lightly doped section of a transistor is of N material and the ends are of P type material. The ends are heavily doped, which is

indicated by the nearness of the fermi level to the valence bands. The center portion is relatively lightly doped, indicated by the fact that the fermi level is still relatively close to the center of the band gap.

Once the junction is formed and equilibrium occurs, the fermi levels of the three pieces of material are aligned. There is a potential difference existing across both junctions that must be reduced for external current to flow. If separate contacts are made to each of the three sections of the transistor, separate biasing can be applied to both junctions.

Figure 12-3 shows both junctions of a transistor forward biased. With both junctions of the transistor forward biased, holes and electrons cross both junctions to attempt recombination.

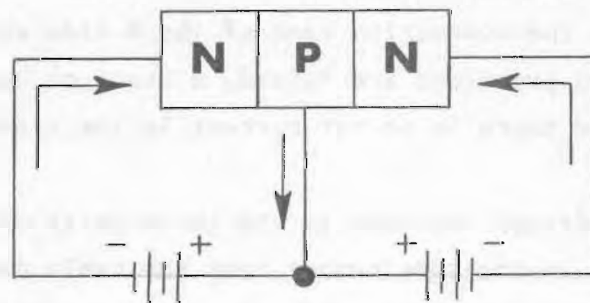


FIGURE 12-3

Since the center portion is doped so much lighter than the ends, there are many more electrons injected from the ends into the center portion than there are holes injected from the center portion into the ends. Since there are so few imperfections in the center portion, most of the carriers injected from the ends will diffuse. The carriers injected into the center portion in Figure 12-3 will either recombine if they find imperfections or will diffuse through the center portion to the opposite side. If the center portion is made very thin, most of the injected carriers that do not find imperfections and recombine will diffuse across the center portion into the opposite side. With equal amounts of bias applied to both junctions of the transistor in Figure 12-3 and assuming that both ends are doped with equal amounts of impurities, the total current through the center portion by diffusion is zero. The algebraic sum of the two diffusion currents is equal to zero. Note, however, that those carriers that do recombine in the center portion result in a current through the common lead out of the center portion. The total current in

the external circuit is the recombination current only, plus any leakage current on the surface. The condition that we have just discussed with both junctions forward biased is termed transistor saturation and will be discussed in detail later. It should be noted, however, that when a transistor is in saturation (both junctions forward biased), that the total current is the algebraic sum of the two diffusion currents. Also, the total number of minority carriers in the base is increased greatly when both junctions are forward biased, since both ends of the transistor inject carriers into the center portion. Since the center portion is very lightly doped, the minority carrier lifetime is long, and the carriers must diffuse or find imperfections and recombine.

With the simple transistor that we have discussed to this point, either end of the transistor can be used to inject carriers into the center portion. The center portion is made very narrow and lightly doped to enhance current by diffusion. The type of transistor construction depends to a great extent on the application in which the transistor will be used, the required frequency response, and the desired power dissipation capabilities of the transistor. One end of the transistor might be made large to dissipate more heat. The center portion might be doped a bit heavier to improve high frequency characteristics. The base may be made thinner to improve high frequency characteristics. The construction of transistors for different applications will be discussed at a later point; however, the basic transistor that we are now using can be used to discuss the characteristics of all transistors. The different applications can then be discussed showing the differences that exist between the basic transistor and those used in a specific application.

One end of the transistor will be termed the emitter and will serve as a source of current carriers injected into the center portion. The opposite end of the transistor will be termed the collector and will rely on the carriers injected into the center portion for current flow. The center portion is termed the base of the transistor and the name evolves from the early transistors being of a point contact construction. Metal wires termed "cat whiskers" were applied to a piece of semiconductor. The piece of semiconductor itself was large and made up what is now the base of the junction type transistor. The cat whiskers that were applied served as the emitter and collector, hence the center portion generally made up the largest bulk of the transistor. The piece of semiconductor served as a mounting and was

referred to as the base. This term has been carried over into present day transistors, and the center portion is still referred to as the base. As we have discussed, either end of the basic transistor (assuming that they are identical) could be used as the emitter. The type of transistor that is used as an example here is referred to as a rate grown transistor and is discussed later in this section. For power handling considerations, the removal of heat is a prime consideration. A collector might be made of a larger piece of semiconductor material for a greater heat radiating surface, or the collector might be attached thermally to the case or encapsulation of the transistor. The emitter might be made of a smaller piece of material. Special construction might be used to make the base very thin; however, the general considerations that we will now discuss apply to all types of transistors and only their construction will be different.

BIASING:

For amplifier operation, the transistor is normally quiescently biased with the emitter to base junction forward biased, and the collector to base junction reverse biased. This is illustrated in Figure 13-3.

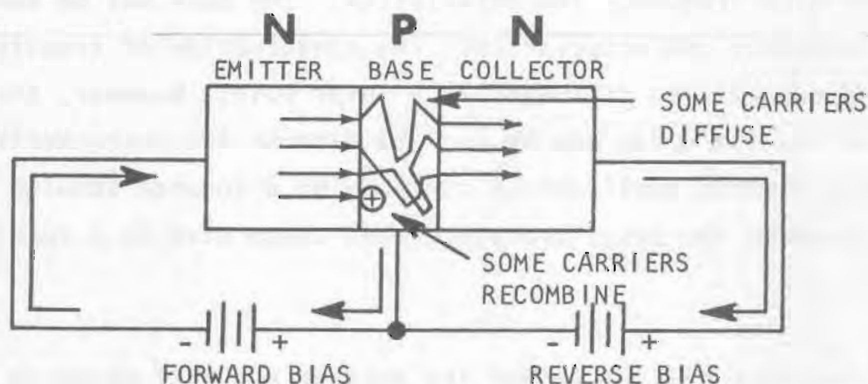


FIGURE 13-3

The emitter is doped much heavier than the base, and the base is doped lightly to enhance diffusion in the base. When the emitter to base junction is forward biased, a few holes are injected from the base into the emitter; however, a greater number of electrons are injected from the emitter into the base. Majority carriers from the emitter and the base cross the junction and attempt recombination. Since the transistor in Figure 13-3 is an NPN configuration, the majority carriers injected from the emitter are electrons. Many more electrons will be ejected from the emitter into the base than holes from the base into the emitter in Figure 13-3.

The injected electrons will diffuse in the base seeking impurities. Those electrons that find impurities or imperfections (recombination traps) will recombine and make up a current in the external base lead. Those electrons that do not find imperfections will diffuse in the base and, as a result of the base being very thin, most of these electrons will reach the collector. With the collector reverse biased, the depletion region about the collector junction is widened and the only current that will flow in the reverse biased collector junction is any surface leakage current and current as a result of minority carriers in the collector and the base. With no carriers injected from the emitter, only saturation current flows in the collector-base diode. With the emitter injecting carriers into the base, there are a large number of minority carriers in the base as a result of the forward bias applied to the emitter-base junction. The collector sees this as a continuing source of minority carriers and a fairly large current can flow in the collector junction, even though it is reverse biased. It is well to examine this at this time and remember that without the emitter injecting carriers into the base, the only current that will flow is any surface leakage current across the collector base junction and the current carried by thermally agitated minority carriers present in the collector and the base. Since thermal (heat) energy forms hole-electron pairs, there will be some holes in the N type collector and some electrons in the P type base at room temperature without external electrical energy applied. Without carriers injected from the emitter, the only current that will flow is the result of these hole-electron pairs (disregarding any surface leakage current). This current is referred to as <u>current</u>, <u>collector</u> to <u>base</u>, with the <u>emitter</u> <u>open</u> <u>circuited</u> and given the symbol I_{CBO}. I_{CBO} is the saturation current of the collector-base diode.

The collector to base current with the emitter open circuited (or not injecting carriers into the base) is limited by the number of minority carriers available as a result of thermal energy (plus any surface leakage) as long as sufficient reverse voltage is <u>not</u> applied to cause the collector-base junction to enter the avalanche breakdown mode. If sufficient voltage is applied to cause a collector junction to enter the avalanche breakdown mode, the current is primarily limited by the resistance in the circuit. Although a few transistors are designed to operate in the collector breakdown mode (avalanche), most transistors avoid this region of the transistor's characteristics.

With the collector junction reverse biased, and the emitter-base junction forward biased and injecting carriers into the base, the collector current is the sum of thermally generated minority carriers and the current injected from the emitter that diffuses across the base to the collector. Figure 14-3 shows the transistor with forward bias applied to the emitter-base junction, reverse bias applied to the collector junction, and the energy band diagram that results.

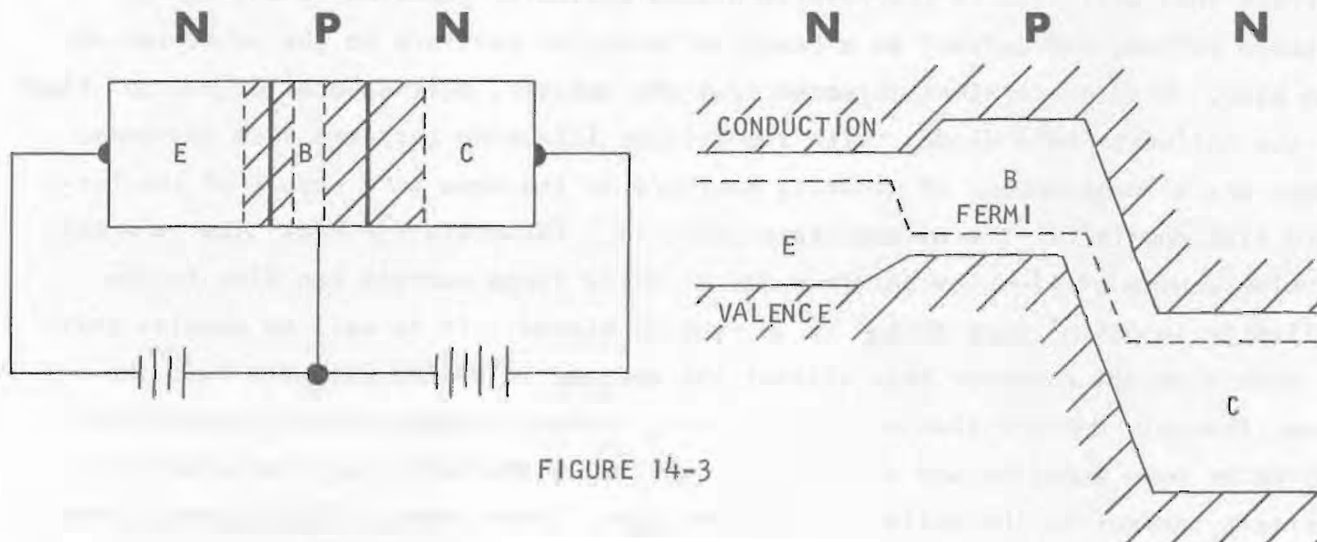


FIGURE 14-3

The application of forward bias to the emitter-base junction misaligns the fermi levels and allows the movement of electrons in the conduction band of the emitter into the conduction band of the P type base. These electrons cross the junction into the P type material. Some holes cross from the valence band of the P type base into the valence band of the emitter at the same time. The number of electrons injected from the conduction band of the emitter is much greater than the holes injected from the base into the emitter in Figure 14-3. The electrons injected into the base conduction band start diffusing until they find imperfections and recombine. Those electrons that do not find imperfections can diffuse across the base thickness and enter the high electric field established by the reverse bias on the collector junction. Minority carriers that enter the high electric field established by the collector reverse bias are accelerated across the junction. Electrons that do recombine in the base result in a small current in the base lead. If any surface leakage current and any current that is the result of minority carriers present as a result of thermal energy is disregarded, the collector current is made up of the diffusion current in the base.

The collector current can be controlled by varying the number of carriers injected

from the emitter into the base. Collector-base diode saturation current is established by the number of minority carriers present at a given temperature. A change in the collector current can be accomplished by varying the number of carriers injected from the emitter into the base. This can be summarized by saying that the collector current is dependent on the total number of minority carriers in the base of the transistor. By varying the number of minority carriers in the base, the collector current can be varied. The number of minority carriers in the base can be varied by changing either the emitter or the base lead current. The injected carriers can also be varied by changing the voltage on the base to emitter junction.

The number of minority carriers and, as a result, the current in a semiconductor junction is related exponentially to the applied voltage. Therefore, if an attempt is made to vary the number of injected carriers from the emitter into the base by applying a changing voltage to the base-emitter junction, the current change will vary exponentially as the applied voltage varies. If a linear change in current is applied to the emitter or base leads, a linear change in current can result in the collector lead.

For normal amplifier operation, the emitter-base junction is forward biased. Forward bias indicates that the junction is biased for forward conduction. The current in the collector circuit is controlled by the current in the emitter and the base circuits. Therefore, we might say that the transistor is essentially a current operated device. There is a linear relationship between the current in the collector, base, and the emitter. The same is <u>not</u> true of the current voltage relationship. An exponential relationship exists between the current in the collector circuit and the voltage on the emitter-base junction. The current in the collector circuit is relatively independent of the voltage applied on the collector to base junction as long as this junction is <u>not</u> forward biased. When both junctions are forward biased, the transistor is in saturation. As long as the collector junction is not forward biased, the transistor is not in a saturated mode and the collector voltage has only a small affect on the collector current.

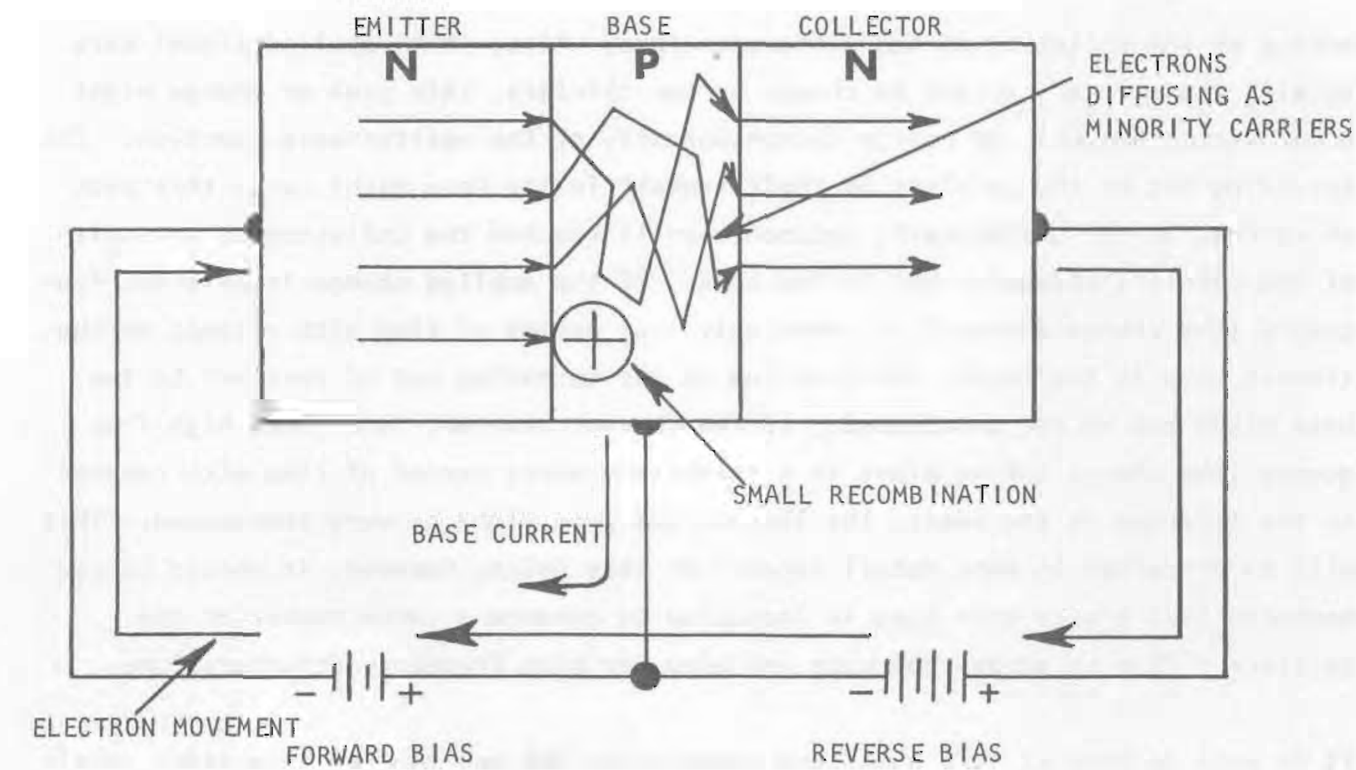
An analogy might be drawn between the transistor's output characteristics and that of the pentode vacuum tube. The pentode vacuum tube plate voltage has only a small affect on plate current as long as the plate voltage is higher than the

screen voltage. Transistor characteristics are similar in that, as long as the collector junction is not forward biased, the collector voltage has only a small affect on the collector current.

Figure 15-3 illustrates the transport of carriers in the base of the transistor. The diagram indicates only a small recombination of electrons in the base, and the other carriers diffusing across the base thickness to the collector. Note in the diagram that the movement of the injected carriers across the base is shown very erratic. We have already discussed in sections one and two the erratic movement of diffusing carriers. The diffusion constants and mobility constants of majority and minority carriers in intrinsic semiconductors are valid for use with transistors.

It should be remembered, however, that there will be a period of time involved in the crossing from the emitter to the collector. This period of time is set by the diffusion constant for the intrinsic material used in construction of the transistor. The number of carriers that do migrate or diffuse across the base will be dependent on the number of carriers injected into the base, the minority carrier lifetime of the base material and the width of the active area of the base. If the width is narrow with respect to the lifetime of the minority carriers, most of the diffusing carriers will reach the collector. If the lifetime is short with respect to the width of the base, a good number of the carriers can accomplish recombination and few carriers will migrate to the collector. This will govern the percentage of the emitter current that flows in the collector, and also govern to a great extent the frequency characteristics of the transistor.

Although the mobilities of majority carriers and the diffusion constant of minority carriers enter into the frequency considerations, the width of the base with respect to the minority carrier lifetime has a great deal to do with the high frequency characteristics of a transistor. For instance, if the base is fairly thick with respect to the minority carrier lifetime, ten carriers injected together into the base might arrive at the collector at ten different times. This is due to the erratic path taken by the carriers in their diffusion in the base. This is generally referred to as the spreading out of the carriers in the base, and this spreading out tends to limit the high frequency characteristics of the transistor. Remember that ten carriers injected into the base at exactly the same time might



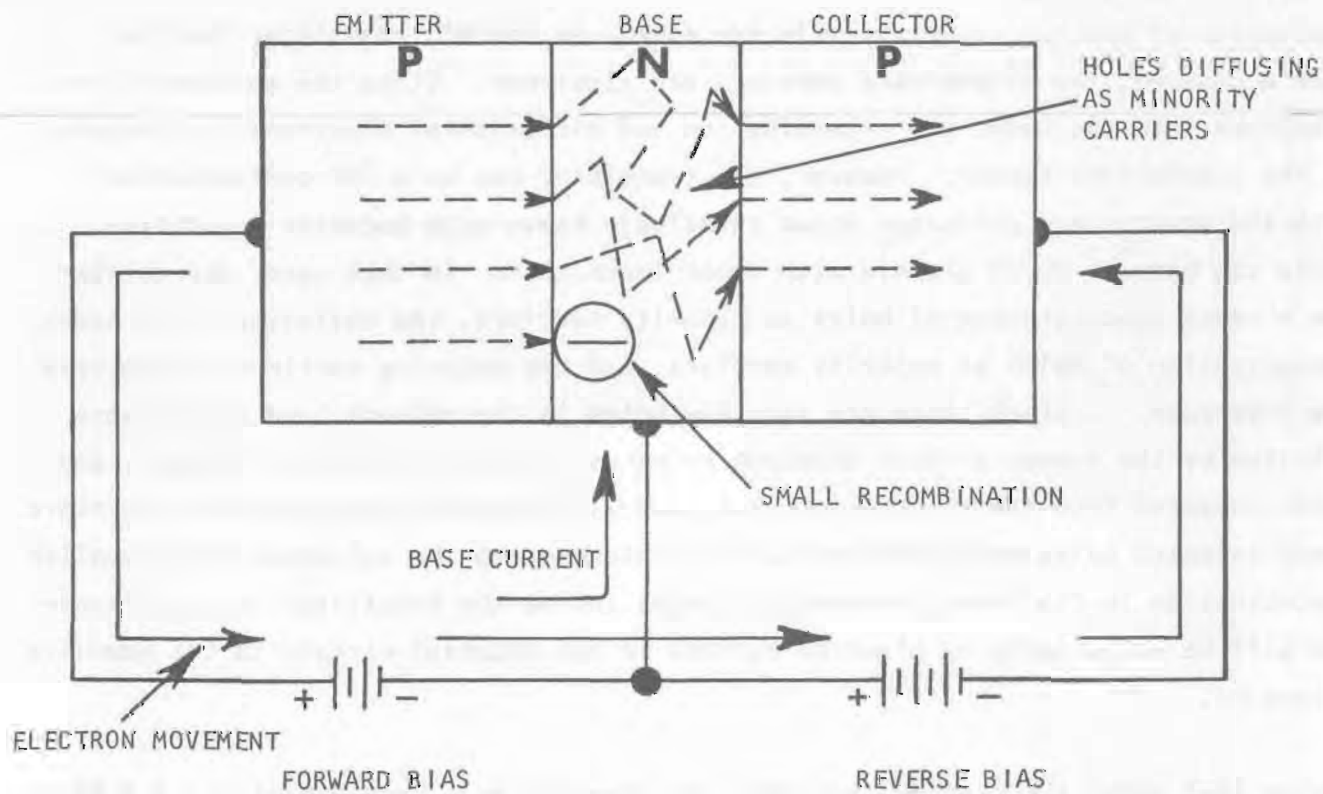
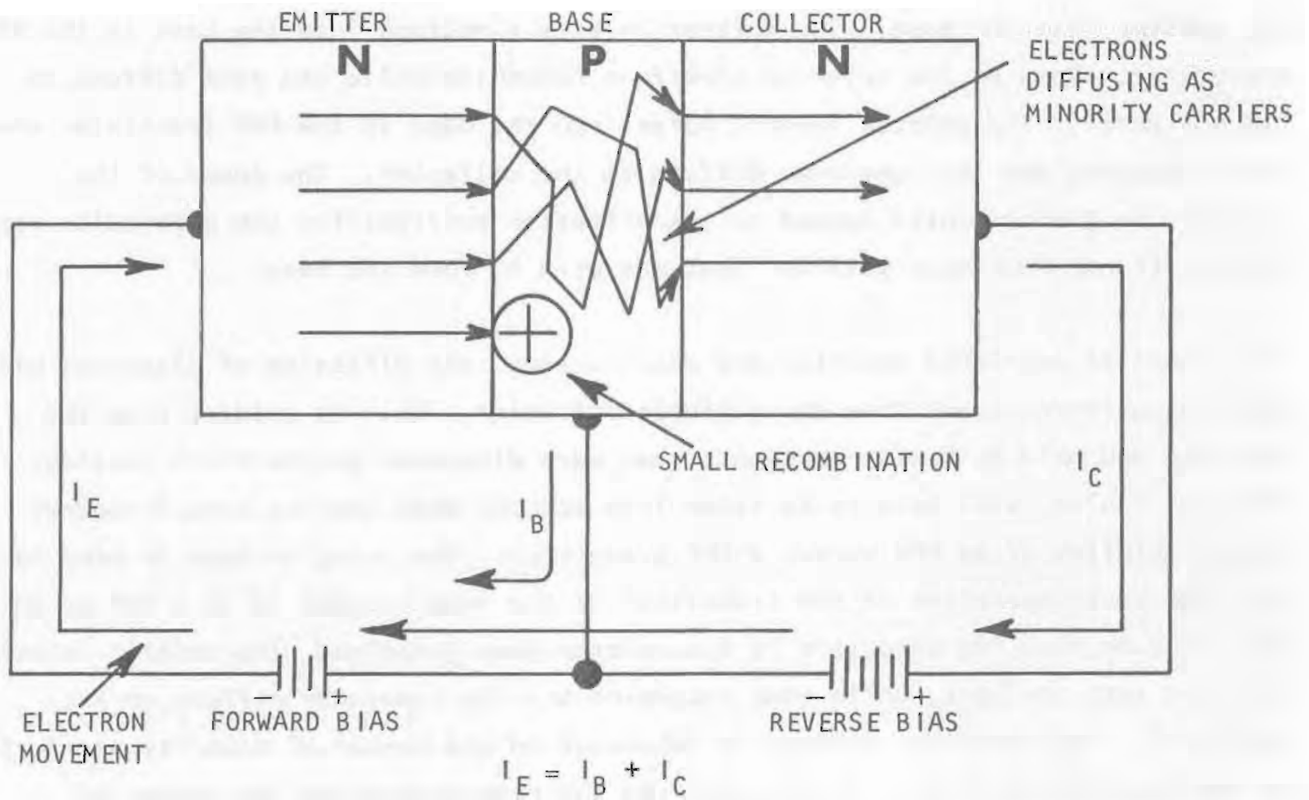


FIGURE 15-3

arrive at the collector at ten different times. Also, if an applied signal were causing the emitter current to change by ten carriers, this peak or change might occur instantaneously or nearly instantaneously at the emitter-base junction. The spreading out of the carriers in their transit in the base might cause this peak of current to be substantially reduced when it reached the collector as a result of the carriers spreading out in the base. If the applied change is of a low frequency (the change occurs in a relatively long period of time with respect to the transit time in the base), the loss due to the spreading out of carriers in the base might not be too pronounced. If the change, however, is of some high frequency (the change taking place in a relatively short period of time with respect to the lifetime in the base), the loss in the base might be very pronounced. This will be discussed in more detail later. At this point, however, it should be remembered that a very thin base is indicated to enhance a large number of the carriers diffusing across the base and also for high frequency considerations.

It is well to stop at this point and compare the NPN and the PNP transistor, their energy band diagrams, and the predominate carriers in the two. Generally speaking, the action of the two transistors is the same. In the NPN transistor that has been discussed, the predominate carriers are electrons. Since the emitter injects electrons into the base, the recombination and diffusion of electrons in the base is the predominate factor. However, the transistor can be a PNP configuration with the emitter and collector doped relatively heavy with acceptor impurities while the base is doped lightly with donor impurities. In this case, the emitter has a heavy concentration of holes as majority carriers, the collector has a heavy concentration of holes as majority carriers, and the majority carriers in the base are electrons. Since there are very few holes in the valence band of the base (limited by the number of hole electron pairs as a result of thermal energy), any holes injected from the emitter into the base will be existing as minority carriers. These injected holes must diffuse the base thickness to the collector or accomplish recombination in the base. Movement of holes inside the transistor in one direction will be accompanied by electron current in the external circuit in the opposite direction.

Figure 16-3 shows the carriers moving in the transistor. Recombination and diffusion is indicated in the base. Note in Figure 16-3 that when the emitter-base junction of either the NPN or the PNP is forward biased, carriers are injected from



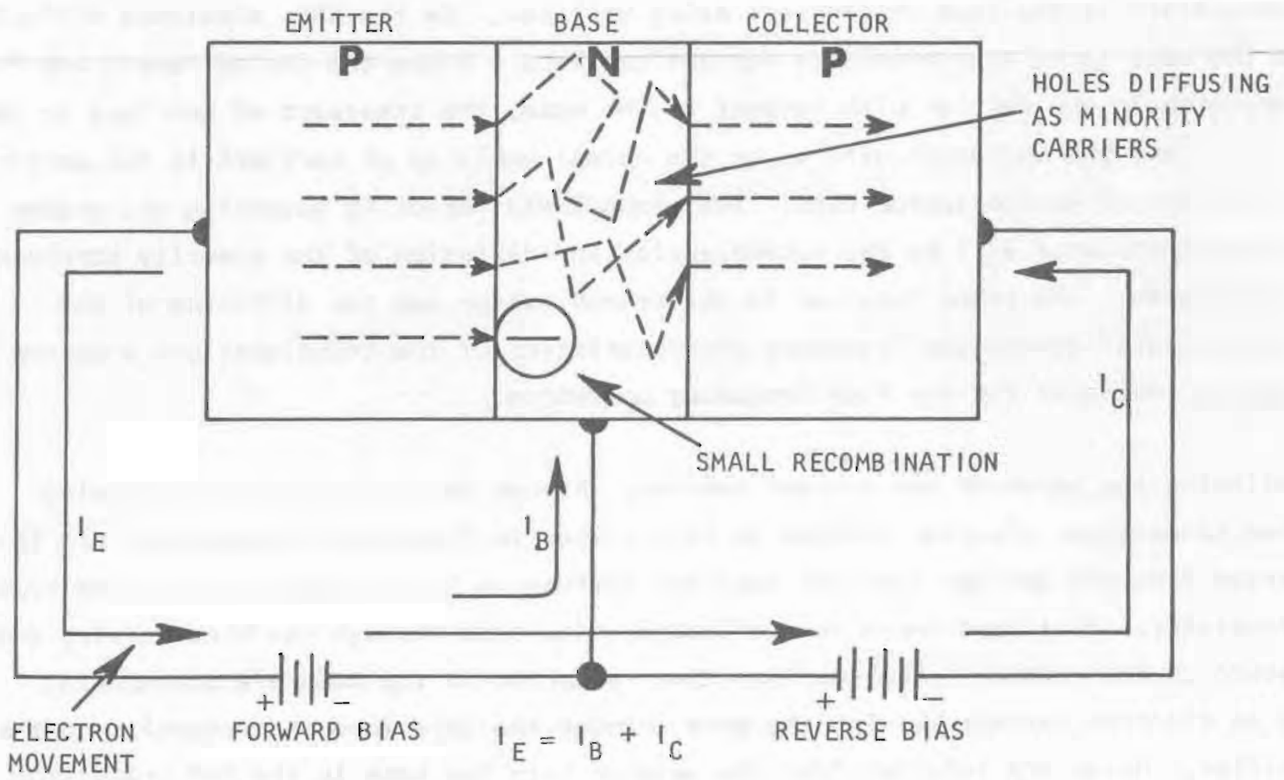


FIGURE 16-3

the emitter into the base. The emitter injects electrons into the base in the NPN transistor. Some of the injected electrons recombine while the rest diffuse to the collector. The emitter injects holes into the base in the PNP transistor where some recombine and the remainder diffuse to the collector. The speed of the carriers in the base will depend on the diffusion constant for the particular type of carrier and intrinsic material that was used to form the base.

For identical intrinsic material and construction, the diffusion of electrons will occur at a faster speed than the diffusion of holes. This is evident from the electron and hole diffusion constants that were discussed in the first section. This, of course, will have to be taken into account when dealing with frequency characteristics of an NPN versus a PNP transistor. The thing to keep in mind is that the basic operation of the transistor is the same whether it is a PNP or an NPN. The controlling circuitry is the emitter-base junction. The emitter injects carriers into the base, while some recombine and the remainder diffuse to the collector. The collector current is dependent on the number of minority carriers in the base, which in turn is dependent on the temperature and the number of minority carriers injected from the emitter. The major difference in the two transistors is the type of carriers being utilized. In the NPN, electrons diffusing in the base serve as predominate current carriers. Since the doping levels are very high in the emitter with respect to the base, the transport of carriers in the emitter and the collector will be by the normal mobility of carriers in the particular type of semiconductor used. The predominate factor in governing the action of the transistor will be the recombination and diffusion of the minority carriers in the base. The times involved in the recombination and the diffusion of the carriers will govern the frequency characteristics of the transistor and a narrow base is indicated for any high frequency operation.

Following the paths of the current carriers through the transistor and relating them to external electron current as illustrated in Figure 16-3, electrons are injected from the emitter into the base and diffuse to the collector in the NPN type transistor. Electrons leave the collector, flow back through the bias supply, and return to the emitter. The carriers that recombine in the base are accompanied by an electron current leaving the base through the base lead and returning to the emitter. Holes are injected from the emitter into the base in the PNP transistor. Some recombination occurs, and the rest diffuse through the base to the collector.

When a hole is injected, or formed, at the ohmic contact at the emitter, an electron leaves the emitter and flows toward the positive side of the bias source. When holes move from the ohmic contact at the emitter toward the base terminal from the negative side of the bias source. Diffusing holes that reach the collector become majority carriers once accelerated across the collector junction. Holes are majority carriers in P type material and once the diffusing carriers enter the collector, they are again majority carriers. The movement of holes in the collector toward the ohmic contact results in electrons leaving the negative side of the reverse bias collector source and flowing toward the collector. As shown in Figure 16-3, the transport of holes from the emitter through the base to the collector results in electron current in an opposite direction in the external circuit. Of course, if conventional current flow is being used, the hole movement can be related directly to the direction of conventional current flow. Conventional current will flow in the same direction as the hole movement through the transistor in the external circuit. It is best to keep in mind that in these publications, unless otherwise stated, the current in the external circuit is <u>electron current</u>.

The d-c emitter current is given the symbol I_E, the d-c base current (the term base current indicates the current in the base lead) is given the symbol I_B, and the current in the collector lead is given the symbol I_C. As shown in Figure 16-3, the emitter current (assuming no leakage current) is the sum of the recombination current in the base (base lead current), and the diffusion current to the collector (collector lead current). In other words, the sum of the d-c base current and the d-c collector current is equal to emitter current. For example, if the transistors in Figure 16-3 have an emitter current of 21ma and a base current of 1ma, the collector current is 20ma. As another example, if the base current is 2ma and the collector current is 48ma, the emitter current is 50ma. This is assuming no leakage current and that the base lead current is the recombination current, while collector current is the diffusion current in the base.

Figure 17.3 illustrates current dealt with in percentage terms. In this case consider emitter current 100%, base current 4% and collector current 96%. 96% of the emitter current flows in the collector, while 4% flows in the base lead.

The d-c current transfer or gain from <u>emitter</u> to collector is 96%. Assign 0.96 as a factor for the transistor in figure 17-3.

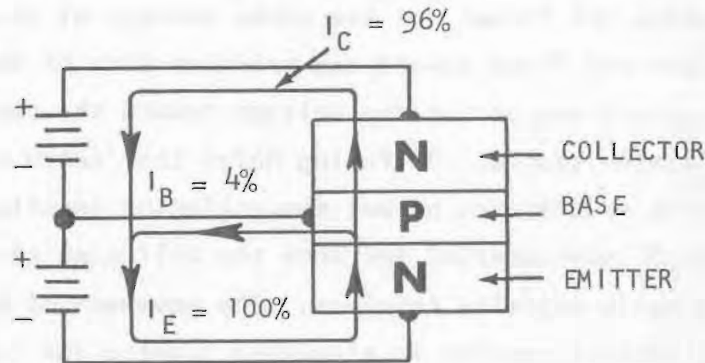


FIGURE 17-3

If the factor from emitter to collector is 0.92 for a particular transistor, and the emitter current is 5ma, the collector current is the product of 0.92 and 5ma, or 4.6ma. There will always be less current in the collector circuit than in the emitter. Therefore, it can be said that the d-c current gain (or transfer) from the emitter to the collector will always be less than 1, or less than unity, and that the collector current will always be less than the emitter current. The percentage of the emitter current that flows in the collector with reference to the emitter is termed d-c alpha. The symbol for d-c alpha is α_{d-c} or h_{FB}, depending on the particular equivalent circuit or model that is being used for the transistor. When h_{FB} is being used, the h in the symbol stands for hybrid and simply indicates that the model or equivalent circuit that is being referred to is the hybrid model. This will be discussed at length later. The F in the symbol indicates forward current gain (or transfer) and the B indicates that the configuration is a common base configuration. When the base is common to both input and output, the measurement must be from emitter to collector. In other words, a common point on the transistor is the base. The current transfer is then between the emitter and the collector. Since the collector current is always less than the emitter current (unless the transistor is an early point contact transistor), the d-c alpha will always be less than unity.

Note in Figure 18-3 that 2% of the emitter current flows in the base, while 98% of the emitter current flows in the collector. In this case, the current transfer between the base and the collector is a factor of 49. In other words, there is 49 times as much d-c current in the collector as in the base lead. To formulate the

factor 49, simply divide 2% into 98% and get a current transfer of 49. Since the collector current is generally much greater than the base current, the current transfer or gain parameter from base to collector is greater than unity.

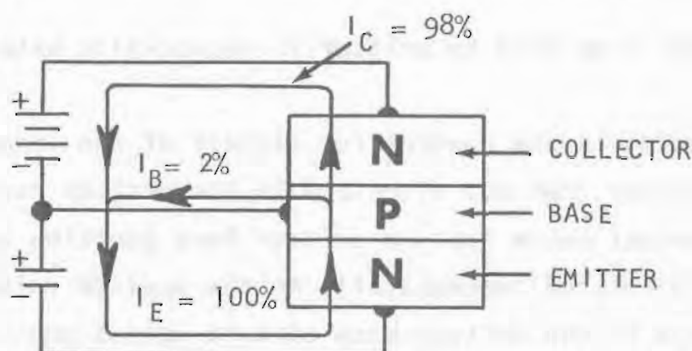


FIGURE 18-3

When dealing with a current transfer from base to collector, the common element is the emitter. The current transfer when the emitter is common is greater than one. The d-c current transfer or gain between base and collector is referred to as d-c beta and given the symbol β_{d-c}, or h_{FE}. h_{FE}, when taken as a factor with base current, will give the value of collector current. In formula form:

$$I_B \times h_{FE} = I_C$$

The h indicates hybrid, while the F indicates forward current gain, and the E indicates that the emitter is common. When the emitter is indicated as being common, the gain is from base to collector. As an example, a transistor with a d-c beta of 60 and a d-c base current of 15ma has a d-c collector current of 900ma. The d-c beta can be determined by dividing the collector current by the base current, or in formula form:

$$h_{FE} = \frac{I_C}{I_B}$$

Since it is possible to place the transistor in a common collector configuration, the current gain from base to emitter must be considered. The collector is common to both the input and output circuits and the concern is with the current gain from base to emitter. Since the d-c current gain from base to collector is referred to as beta, the d-c current gain from base to emitter is equal to beta plus one. This is arrived at by considering that the emitter current is the sum of base current and collector current. If collector current is equal to the product of beta and

base current, then the emitter current is the product of beta and base current plus base current, or simply the product of beta plus one and base current, or:

$$(h_{FE} + 1)\ I_B = I_E$$

The current gain factor from base to emitter is essentially beta plus one (h_{FE} + 1).

The emitter-base junction is the controlling circuit of the transistor and can be related to the PN junction that was discussed in the earlier parts of this volume. The voltage versus current curve for the emitter base junction is an exponential in that the current is related exponentially to the applied voltage and temperature. The current and voltage in the emitter-base circuit have a non-linear relationship in that the current will not change linearly as the voltage from emitter to base changes. When dealing with a-c signals and the transistor, the concern is with obtaining a linear relationship between the currents in the input circuit and the currents in the output circuit. If the current in the collector circuit of the transistor is dependent on the current in the emitter-base circuit of the transistor, there will be a linear relationship. If an attempt is made to vary the collector current by changing the emitter-base voltage, the relationship will be non-linear unless the change is very incremental. The relationship between current and voltage in the transistor is exponential rather than linear. This is illustrated in Figure 19-3.

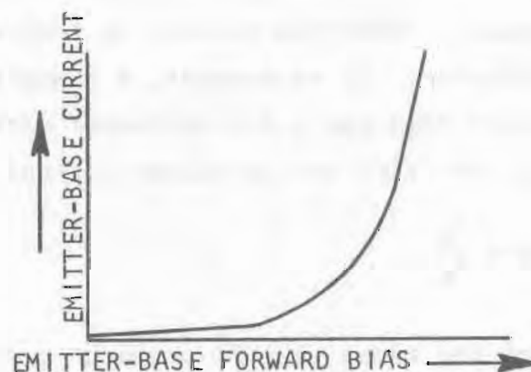


FIGURE 19-3

The current varies exponentially as the emitter-base forward voltage varies. An attempt to control current in the transistor with a varying voltage will result in an exponential relationship. This is shown in Figure 20-3. For the change in

voltage given in Figure 20-3, a non-linear change in current results.

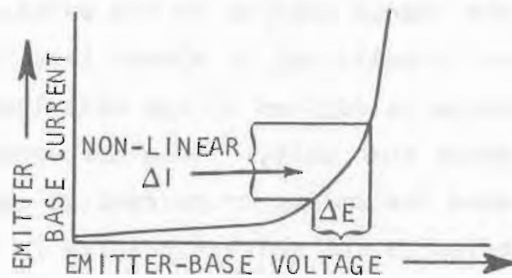


FIGURE 20-3

Since the current crossing the emitter-base junction is the total current in the transistor, both the base current and collector current will show a non-linear change with respect to an applied voltage change on the emitter-base junction.

Due to the non-linear characteristics of the input of the transistor, it is pretty common to find a constant current source driving a transistor when linear results are required. It has been discussed that a constant current source is a high impedance source with respect to the circuit it is driving. It is also common to find a constant current d-c biasing source used with a transistor. The emitter-base junction of a transistor is essentially biased and driven with current rather than voltage for the most linear results.

Amplification with a transistor can be accomplished by varying either the emitter or the base lead current which will be accompanied by a change in the collector current. Since the emitter-base junction is a forward biased junction, it has a relatively low impedance. The same or a greater current change occuring in the collector junction which has a high impedance can result in voltage amplification.

It is possible to place a large load impedance in the collector circuit. (It is reverse biased and has a high impedance.) This allows voltage gain to be accomplished. If the input signal is applied to the base, current gain can also be accomplished.

Assume that exactly the same magnitude of current change is applied in the emitter lead and then in the base lead. Applying a current change in the emitter lead will result in a change in the collector current, but the change in collector current

will be somewhat less than the change applied to the emitter lead. Remember that the current gain from emitter to collector is always less than unity. If the exact same magnitude of current change is applied to the base lead, the current gain from base to collector can be greater than unity. In other words, increasing the base current by one unit might cause the collector current to change by several units. The same one unit change applied to the emitter results in less than one unit change in the collector current. To illustrate this, consider a transistor that has just had the emitter current increased by ten units. If nine of these units diffuse to the collector and one of these units recombines and becomes base lead current, the current gain emitter to collector is less than unity or, in this case, 0.9. If the same transistor has a current change applied in the base lead of one unit, ten added units of current will be injected from the emitter into the base. Nine of these injected units will diffuse to the collector, while one unit recombines and becomes base lead current. The current gain from base to collector is nine, while the same transistor has a current gain emitter to collector of 0.9. From this, it can be seen that the a-c and the d-c current gain from emitter to collector will be less than unity, while the a-c and d-c current gain from base to collector can be greater than unity. By the same example, note that a change of one unit in the base current increased the emitter current by ten units. The current gain from base to emitter can be greater than unity and in the case just stated is one unit greater than the current gain from base to collector. As with the d-c current gain base to emitter, the a-c current gain base to emitter is beta plus one. The beta referred to is signal or a-c beta, while the beta previously referred to was the d-c beta. The symbol for a-c beta is the symbol β, or h_{fe}.

h_{fe} can be found by applying an incremental change in base current to the transistor, noting the change in collector current. h_{fe} can be determined by the formula:

$$h_{fe} = \frac{\Delta I_c}{\Delta I_b}\bigg|_{\Delta V_{CE} = 0}$$

By the same approach, the current gain from emitter to collector, or a-c alpha, which is given the symbol α, or h_{fb}, can be found by applying an incremental change in emitter current and noting the change in collector current.
h_{fb} or α can be found using the formula:

$$h_{fb} = \frac{\Delta I_c}{\Delta I_E}\bigg|_{\Delta V_{CB} = 0}$$

Since the collector voltage does have a small affect on collector current, the collector voltage must be held constant during the measurements in both cases. This might be done by placing a large capacitor across the collector of the transistor and insuring that the collector voltage doesn't change. These characteristics of the transistor are covered more thoroughly in Table 8&9. Table 8&9 is not necessary for the coverage in this volume, but is included for the learner that wants to pursue this further.

COLLECTOR JUNCTION THERMAL CURRENT:

There will be some current in the collector circuit when it is reverse biased, even if the emitter is not injecting carriers into the base. The reverse biased collector junction is dependent on minority carriers for current, and there will be some minority carriers present due to thermal energy regardless of whether the emitter is injecting carriers into the base. This is illustrated in Figure 21-3.

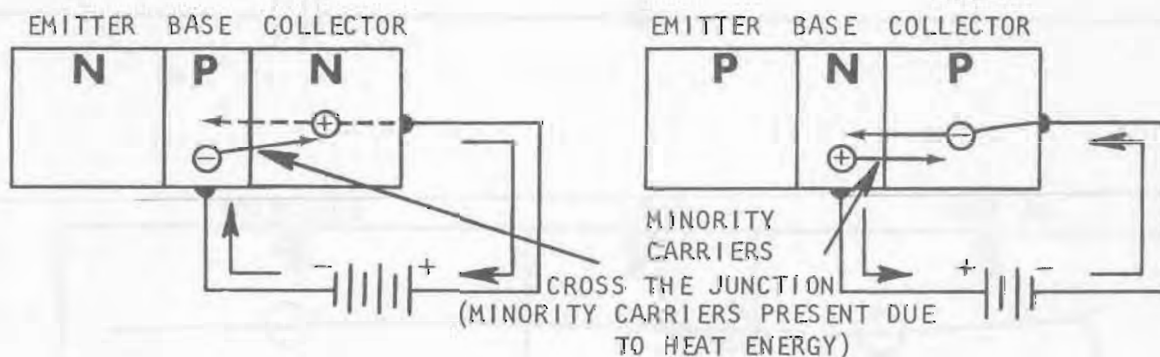
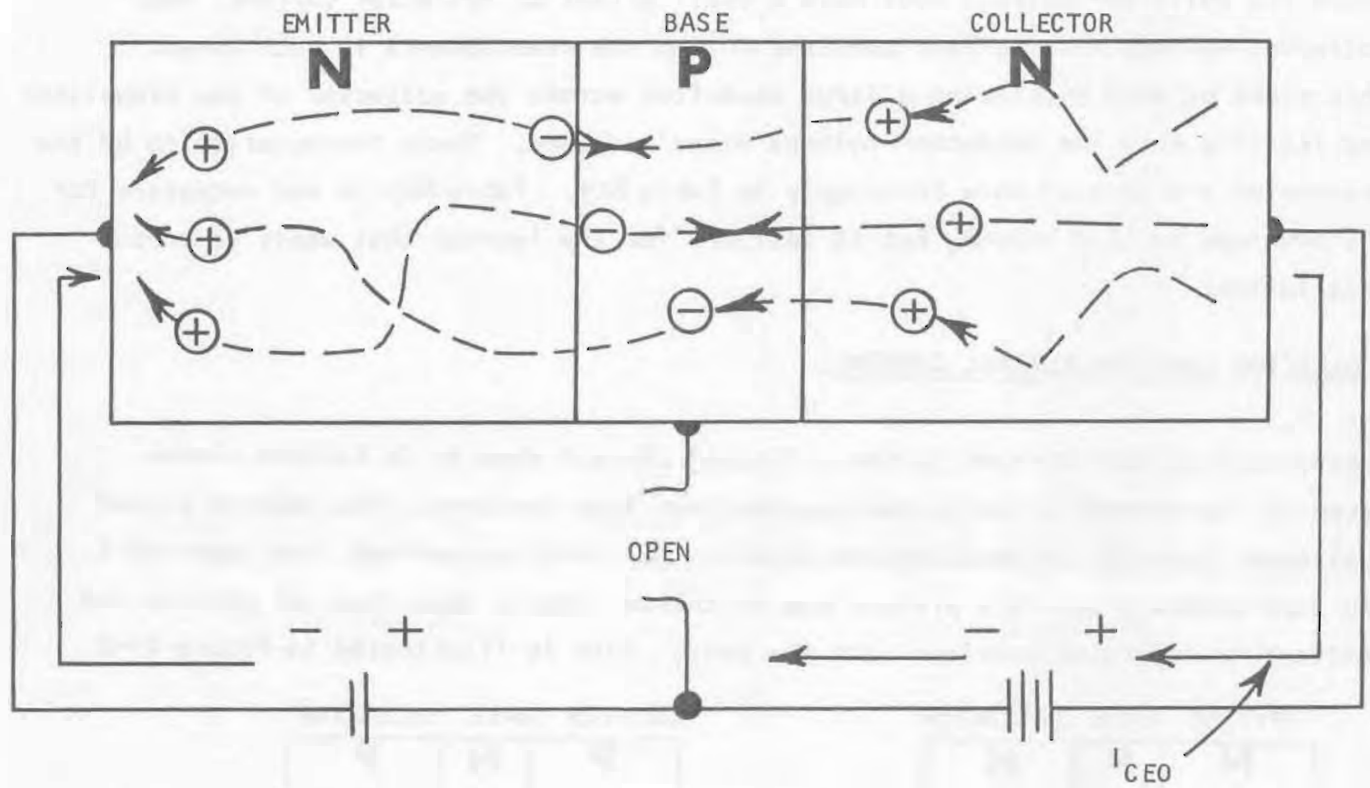


FIGURE 21-3

The holes in N type semiconductor and the electrons in P type that are present as a result of thermal energy forming electron pairs will provide carriers which will flow across the reverse biased collector junction. The amount of current will be related exponentially to the amount of reverse voltage applied and the temperature. (Note that this current will flow whether the transistor is NPN or PNP.) The current that flows when the emitter lead is essentially open circuited (emitter not injecting carriers into the base) and the collector reverse biased is referred to as I_{CBO}. I_{CBO} is the current between collector and base with the emitter open circuited. The current that flows under this condition will be limited by the number of minority carriers in the base as a result of thermal energy.



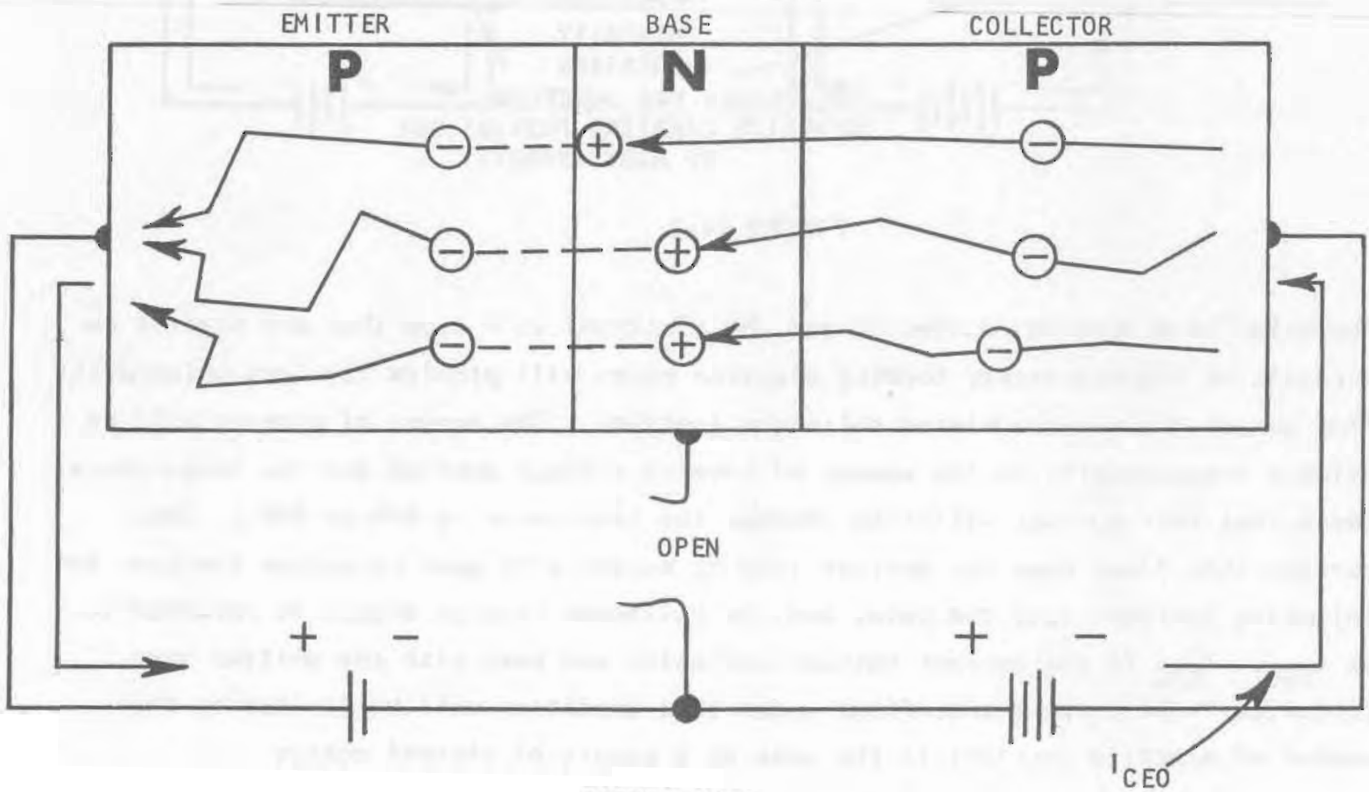


FIGURE 22-3

Since the current I_{CBO} is dependent on the total number of minority carriers available, I_{CBO} will be lower in a silicon transistor at a given temperature than a germanium transistor. It is well to note at this point that it might be necessary to actually reverse bias the emitter-base junction to have the transistor completely non-conducting. Since the current I_{CBO} will flow even if the emitter is essentially open-circuited, an opposing bias current must be applied to completely turn off the transistor. A silicon transistor has a lower typical I_{CBO} when compared to a germanium transistor.

Consider an attempt to turn off a transistor by opening the base lead of a properly biased transistor. When the base lead is open, the external bias voltage is removed from the emitter-base junction. There is, however, a complete circuit from the collector through the collector bias source, the the emitter bias source, and back to the emitter, as illustrated in Figure 22-3.

The minority carriers crossing the collector junction are accompanied by added carriers injected from the emitter into the base. These added carriers appear to the collector as a continuous source of minority carriers, and a magnified current flows. The amount of current flowing with the base lead open is essentially the product of d-c beta of the transistor and I_{CBO}. Current between collector and emitter with the base lead open circuited (I_{CEO}) is essentially the product of h_{FE} and I_{CBO}. When the minority carriers are not allowed to flow in the base lead, they will take the path of the emitter-base junction. Any d-c current crossing the emitter-base junction is magnified by h_{FE} of the transistor. Therefore, the current that flows with the base lead open circuited will be the collector junction saturation current magnified by the d-c beta of the transistor. If the base lead is not open circuited but a high resistance is placed in the base lead, the current that flows will have a magnitude between I_{CBO} and I_{CEO}.

SYMBOLS:

The diagram in Figure 23-3 shows the symbols for an NPN and a PNP transistor. A line that intersects at right angles in a semiconductor symbol is assumed to be a non-rectifying or ohmic contact, while a line that intersects at an angle is assumed to be a rectifying contact or a junction. The collector-base junction in

Figure 23-3 is indicated as a rectifying contact. The emitter-base junction in Figure 23-3 is also indicated as a rectifying contact. The emitter-base and collector-base junctions are both rectifiers. The lead making a connection to the base is shown at right angles to the base and indicates a non-rectifying connection. The connection to the base is ohmic or a non-rectifying contact.

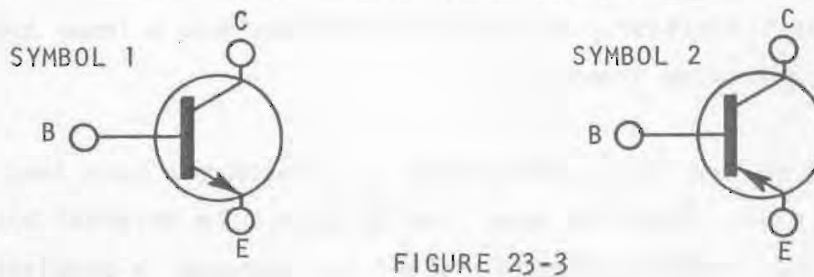


FIGURE 23-3

In the diode symbol discussed in the first part of this volume, it was observed that electron current moves against the arrow, while conventional current flows with the arrow in the symbol. Hole movement in semiconductors is with the arrow in a symbol. The symbol on the left in Figure 23-3 has an arrow on the line indicating the emitter that points away from the base. An NPN transistor has the emitter injecting electrons into the base when conducting, and electrons move against the arrow in the symbol. The symbol on the right in Figure 23-3 has the emitter arrow point in towards the base. This indicates that holes are moving from the emitter into the base and this must be a symbol for a PNP transistor. This is illustrated in Figure 24-3.

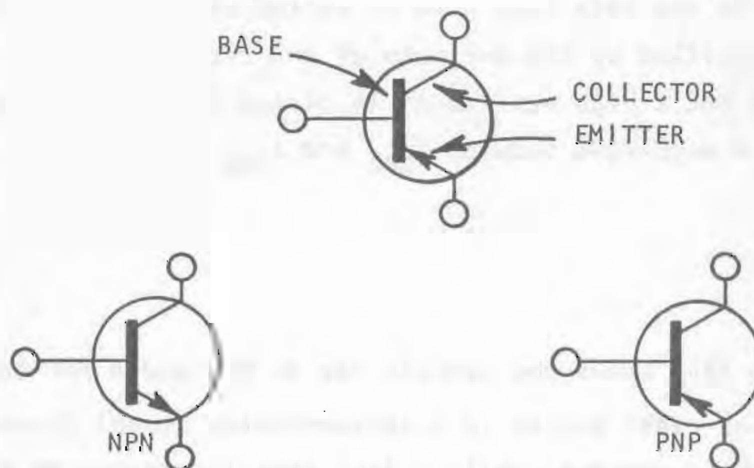


FIGURE 24-3

Figure 25-3 shows the direction and symbols for the d-c currents and voltages associated with the transistor. (The external lead currents are shown as electron currents in this diagram.) The currents and voltages associated with a transistor will have the same symbols when dealing with an NPN or a PNP transistor, and only the polarities will change.

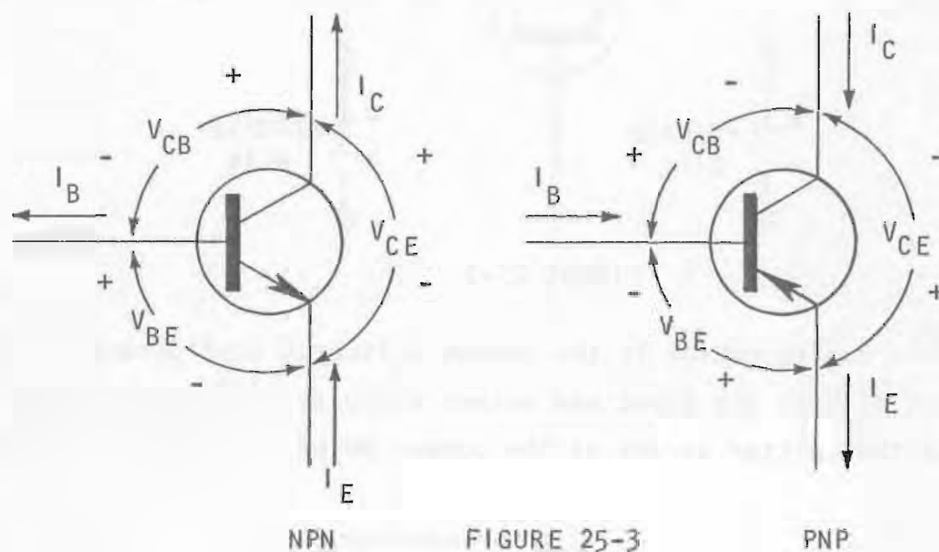


FIGURE 25-3

POSSIBLE CONFIGURATIONS:

Since the transistor has three leads, there are three possible orientations. The transistor can be oriented in a common emitter, common base, or common collector configuration.

The emitter can be made common to both the input and output circuit and when this is done, the base serves as the input point, while the collector serves as the output point. This is illustrated in Figure 26-3.

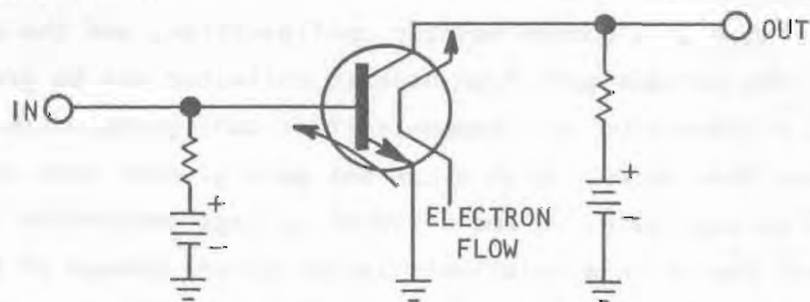


FIGURE 26-3

The transistor can also be placed in a common base configuration. In this case, the base is made common to both the input and output circuit. The emitter circuit serves as the input point and the collector serves as the output point. This is illustrated in Figure 27-3.

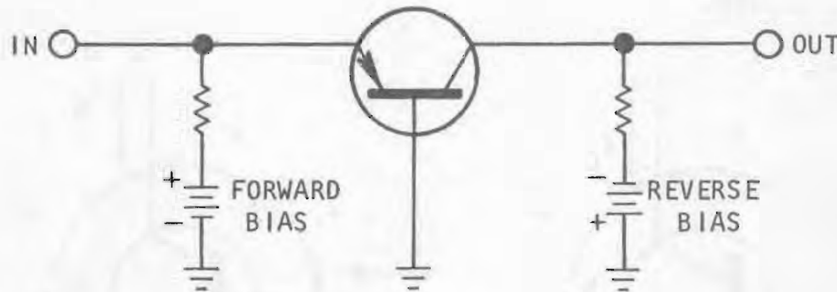


FIGURE 27-3

The third possible configuration is the common collector configuration. The collector is made common to both the input and output circuits. The base serves as the input point, and the emitter serves as the output point. This is illustrated in Figure 28-3.

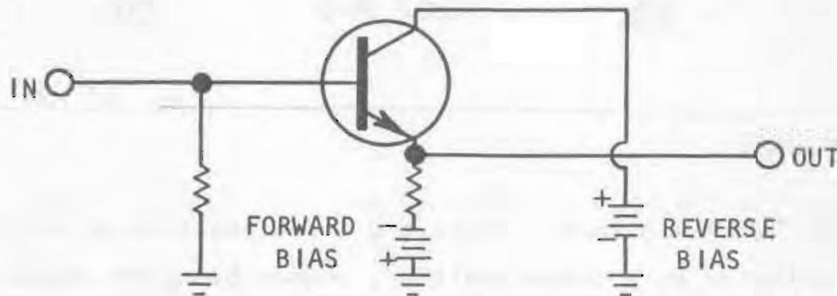


FIGURE 28-3

The characteristics of the transistors already discussed will give insight into some of the possibilities for gain with the transistor in the three configurations. First, consider the common emitter configuration. The input signal is applied to the base of a transistor in a common emitter configuration, and the output is taken at the collector. The current gain from base to collector can be greater than unity. Therefore, a transistor in a common emitter configuration can offer a current gain greater than unity. With a current gain greater than unity, there is the possibility of voltage gain. A small signal voltage applied to the base-emitter junction can result in a relatively large current change in the output or collector circuit. The collector circuit is reverse biased and can have a higher impedance than the emitter-base junction. Therefore, a larger signal voltage can

be developed in the collector. The common emitter configuration has the possibility of both current and voltage gain and a corresponding high power gain.

The common base configuration, however, has a current gain from emitter to collector that is less than unity. The current gain of a common-base configuration is less than unity. However, a current change in the emitter-base circuit can be accomplished with a small voltage, while this same current change can flow through a large impedance in the reverse biased collector junction. This makes possible voltage gain. Since voltage gain can be accomplished, power gain is possible with the common base configuration.

A common collector configuration can have a larger output than input current. Since the emitter carries the largest current in the transistor and it serves as the output point while the base serves as the input point, current gain can be accomplised with the common collector configuration. Figure 29-3 illustrates the total input circuit loop of a common collector configuration.

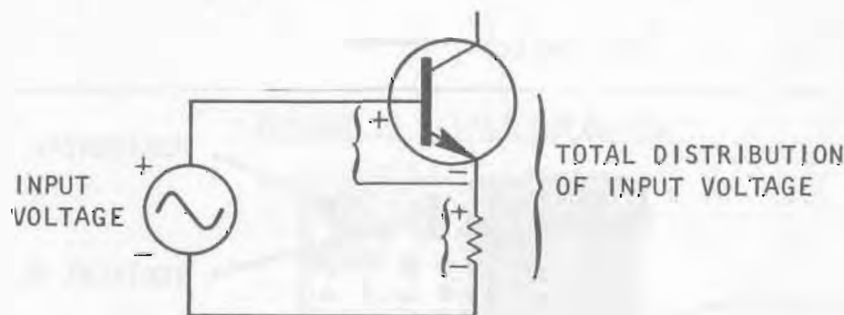
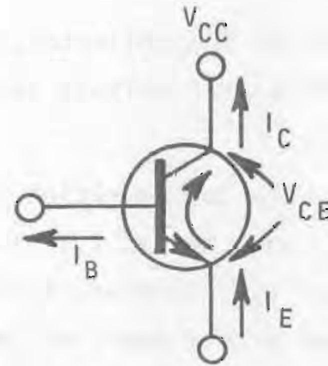
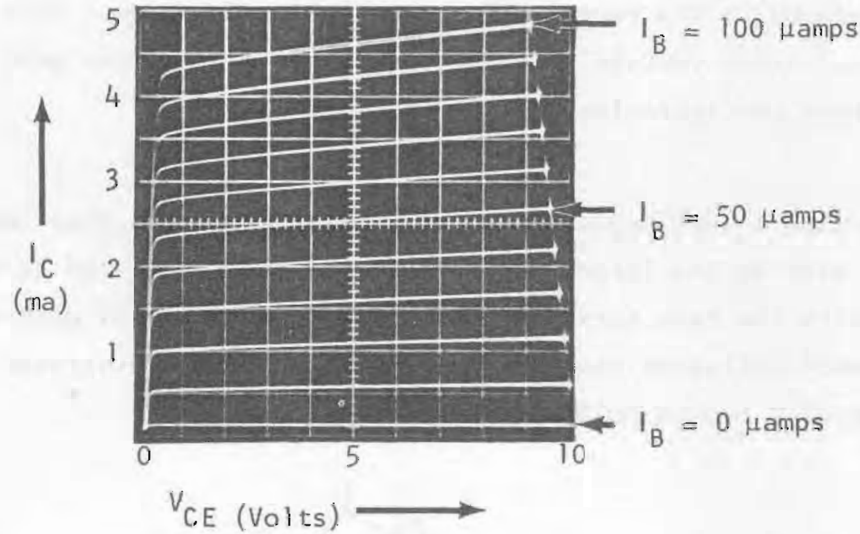


FIGURE 29-3

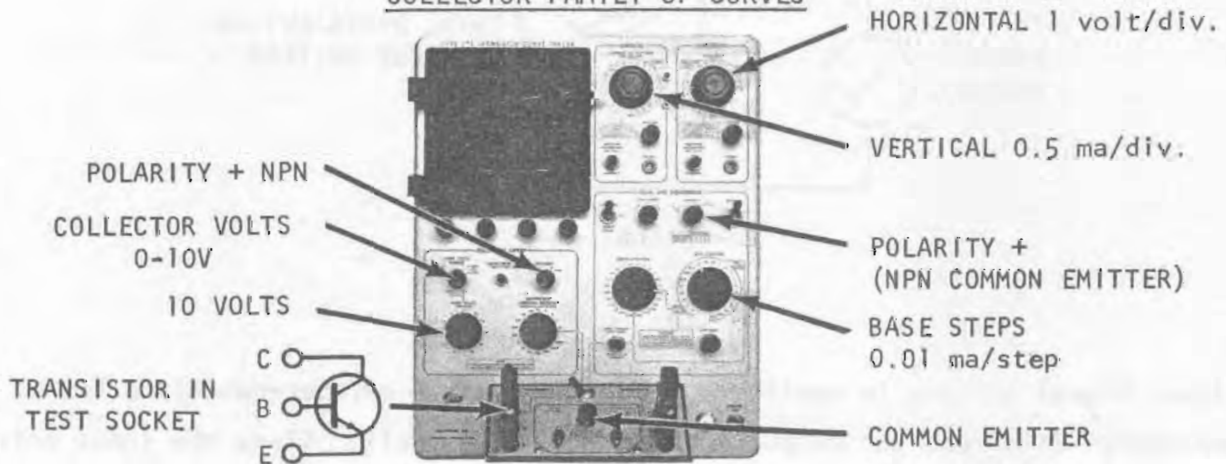
An input signal voltage is applied not only across the emitter-base junction of the transistor, but across the output load resistance as well. Since the input voltage is distributed across both the load resistor and the emitter-base junction of the transistor, the total voltage across the output load resistor can never be as large as the input voltage. The voltage gain of a common collector configuration is less than one. Since the current gain can be greater than unity, power gain can be realized.

V_{CC} = COLLECTOR SUPPLY VOLTAGE
I_B = d-c BASE CURRENT
I_C = d-c COLLECTOR CURRENT
I_E = d-c EMITTER CURRENT
V_{CE} = d-c VOLTAGE, COLLECTOR TO EMITTER





COLLECTOR FAMILY OF CURVES



TEST SETUP

TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER
TOP PHOTO TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA

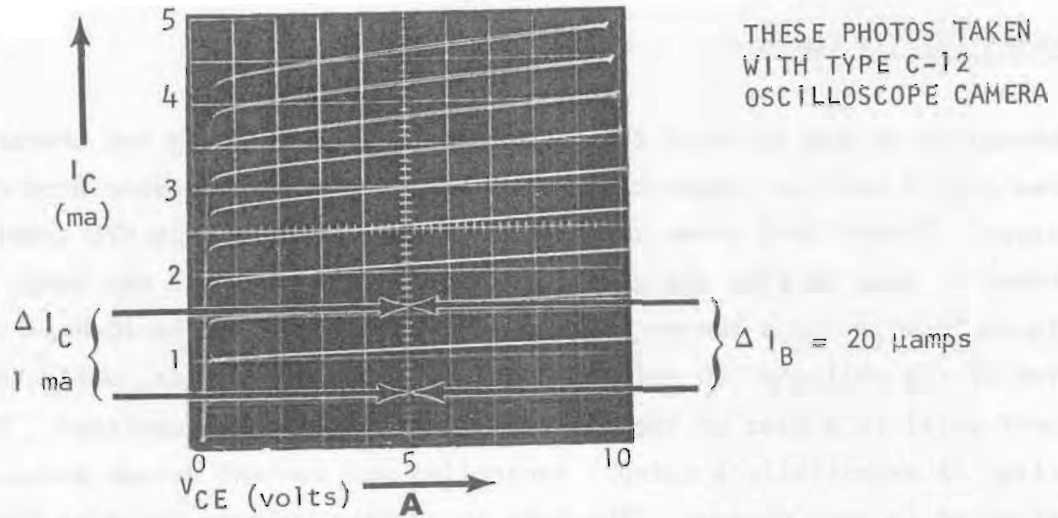
FIGURE 30-3

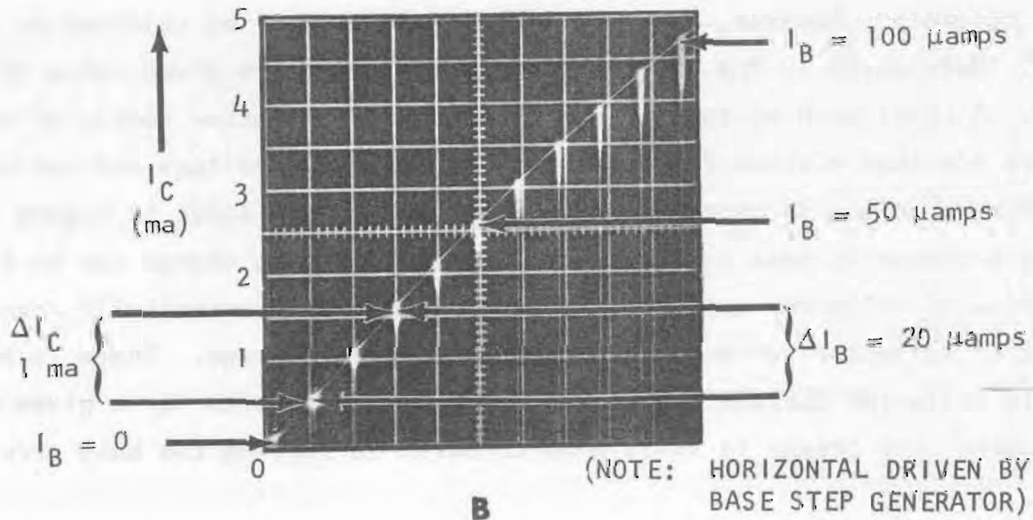
CHARACTERISTIC CURVES:

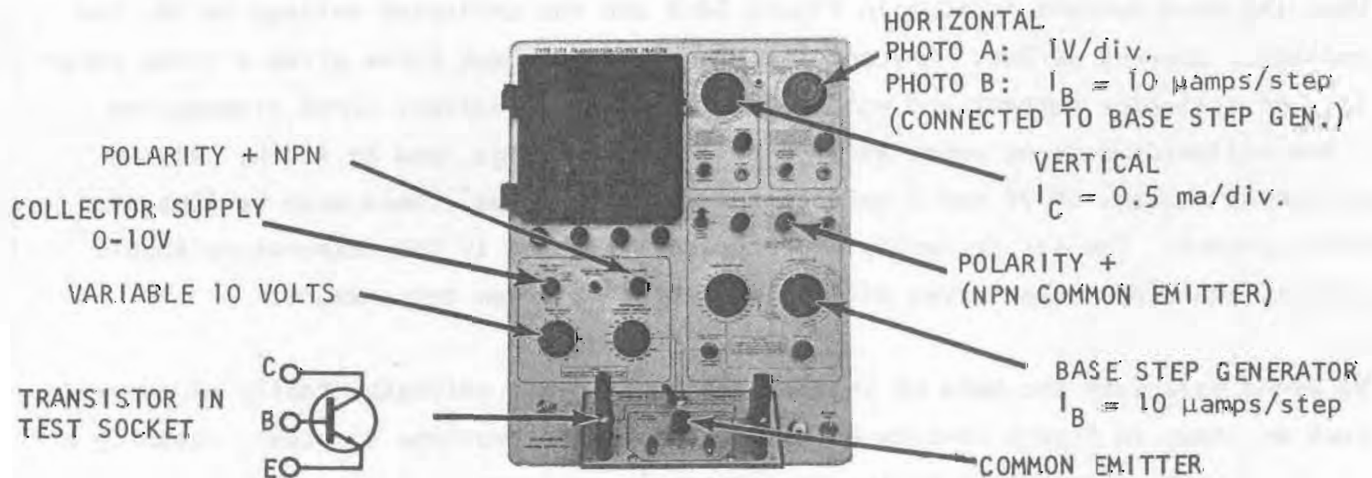
One method of dealing with the transistor and determining its characteristics is to construct a chart or graph of the voltages and currents associated with the transistor. Figure 30-3 shows an example. The Tektronix Type 575 Transistor-Curve Tracer is used to plot the graph on the face of a cathode ray tube. The graph in Figure 30-3 is for a common emitter configuration. The horizontal (or X axis) is a plot of the collector to emitter voltage of the transistor, while the vertical (or Y axis) is a plot of the collector current of the transistor. Since the transistor is essentially a current controlled and current driven device, the running parameter is base current. The base to emitter voltage could be plotted as the running parameter; however, the most linear results will be obtained by using base current. Each curve in the chart in Figure 30-3 is for a given value of base current. A chart such as this is referred to as a collector family of curves. The chart provides a cross reference between collector voltage and collector current for different values of base current. Referring to the chart in Figure 30-3 and assuming a change in base current, the collector current change can be found for a given value of collector voltage. An increase in base current will result in an increase in collector current for a given collector voltage. There is a small change in collector current over a range of collector volts for a given base current; however, the change is small when compared to varying the base current.

When the base current is 60μa in Figure 30-3 and the collector voltage is 5V, the collector current is 3ma. Any point along a base current curve gives a cross reference of collector current and voltage. The 30μa base current curve crosses the 1.6ma collector current point at 5V of collector voltage, and so forth. With a collector voltage of 7V and a collector current of 3.5ma, there must be 70μa of base current. The set or family of curves will change if the temperature should change; therefore, the curves will be plotted at a given temperature.

We might calculate the beta of the transistor from the collector family of curves such as shown in Figure 30-3 by holding the collector voltage constant, assuming a change in base current and noting the change in collector current. This is illustrated in Figure 31-3. To hold the collector voltage constant, simply make the measurement on the 5V collector voltage line. To assume a change of 20μa in base current, make the measurement between the 10 and 30μa base current curves. This







TEST SETUP

TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER

FIGURE 31-3

gives a collector current change of 1ma as shown in Figure 31-3. Dividing the 20μa (ΔI_B) into the 1ma change (ΔI_C), the beta of the transistor on this portion of the collector family of curves can be determined. Measuring beta at another point on the curves will give a different beta than measured at this particular point. Therefore, were this measured value of beta stated, the point on the curves at which the measurement was made would have to be indicated along with the temperature at which the measurement was made.

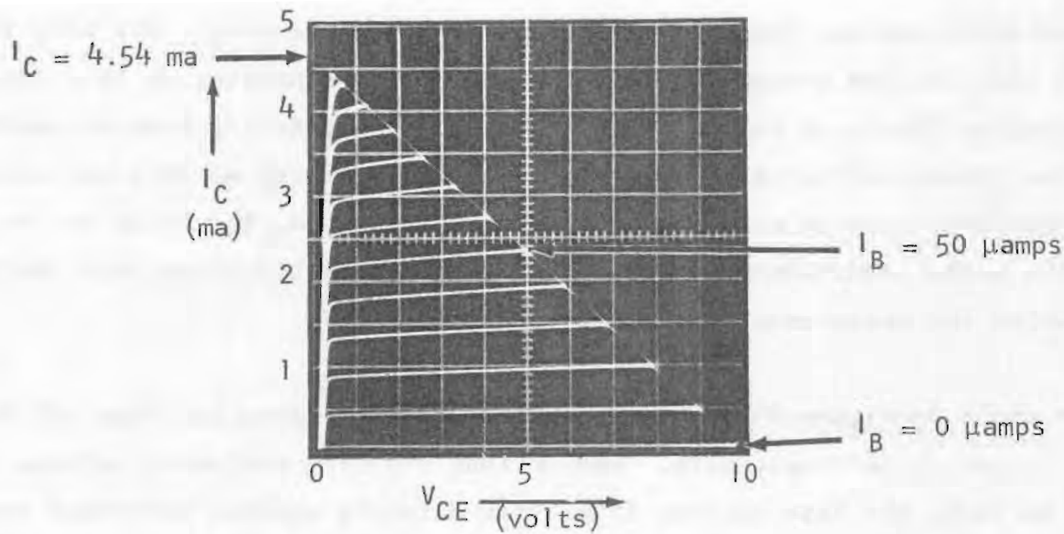
The center photo in Figure 31-3 shows another method of using the Type 575 Transistor-Curve Tracer to calculate beta. Rather than plotting collector voltage versus collector current, the base current is plotted directly against collector current and the slope of the line can be used to calculate the beta of the transistor. Selecting the same area of the curves as was used in the measurement of beta on the collector family of curves should give a similar result. Making the measurement in Figure 31-3 between the two points of 10 and 30μa of base current, the collector current change is once again 1ma, the same value of beta for this transistor as was obtained from the measurement on the collector family of curves. Either approach is valid as long as the portion of the transistor's characteristics at which beta was measured is stated. For both of the measurements made in Figure 31-3, the beta or h_{fe} measured is:

$$h_{fe} = \frac{\Delta I_C}{\Delta I_B} = \frac{\Delta 1ma}{\Delta 20\mu a} = 50$$

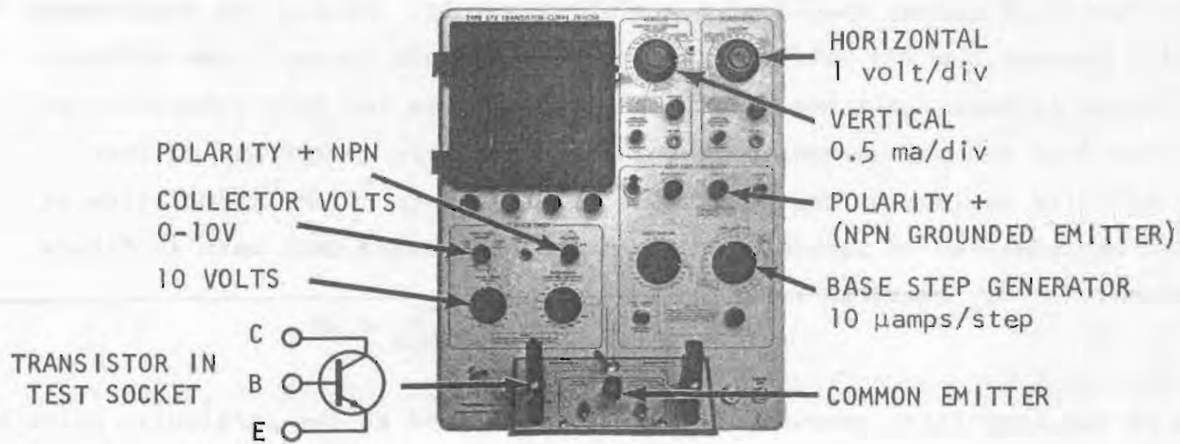
The beta of the transistor measured in Figure 31-3 is 50 at the particular point of operation at which the measurements were made. We might say that the beta of the transistor is 50 with a static operating point of 1ma of collector current. The temperature should also be specified. The measurement in Figure 31-3 was made at 25°C, or room temperature.

LOAD LINES:

To gain information on transistor operation in a circuit, a load line may be constructed on the characteristic curves of the transistor. The term load line indicates the line through the characteristic curves that will indicate the operation of the transistor in the circuit. This is determined by the resistance or the impedance in the circuit, depending on whether d-c or a-c characteristics are being dealt with. For d-c characteristics, the slope of the line constructed on the



COLLECTOR FAMILY OF CURVES
WITH 2.2K LOAD LINE



TEST SETUP

TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER
TOP PHOTO TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA

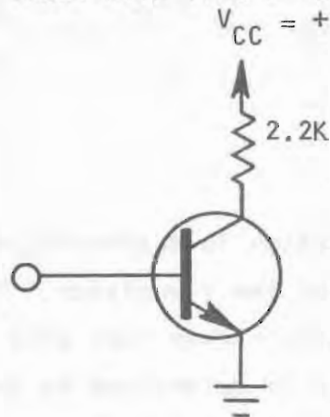
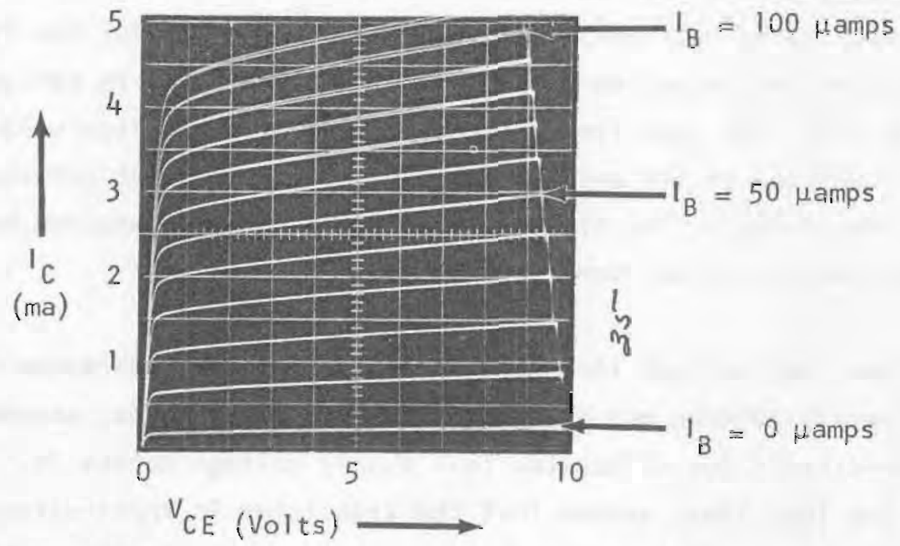
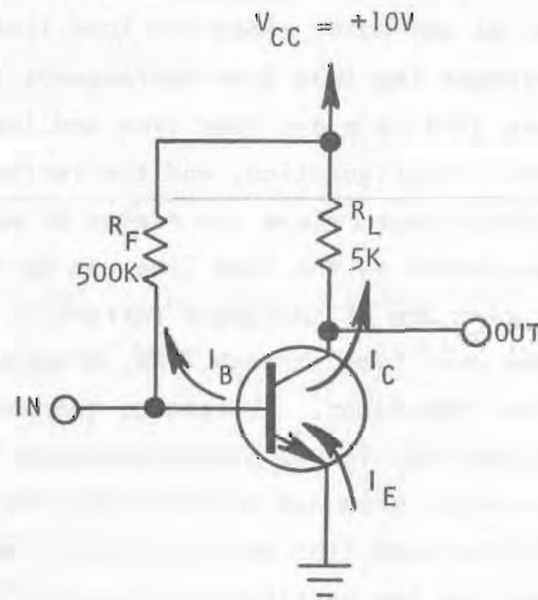


FIGURE 32-3

characteristic curves will be determined by the resistance in series with the transistor. This is illustrated in Figure 32-3. In this case, the transistor is in a common emitter configuration and it has a total of 2.2KΩ in series. The supply voltage is 10V. The load line for the configuration in Figure 32-3 extends between the supply voltage of 10V and the point of maximum current governed by the resistance in the circuit. The slope of the load line is determined by the resistance in series with the transistor.

To calculate the d-c load line for a transistor on its characteristic curves, assume that the supply voltage and V_{CE} are equal. In other words, assume the transistor is an open-circuit and it has the full supply voltage across it. To gain another point on the load line, assume that the transistor is short circuited and the total circuit current is limited by the series resistance. In Figure 32-3, 10V divided by the 2.2KΩ gives 4.54ma of maximum current. Only two points are required to construct a load line. Connecting these two points with a straight line indicates the operation of the transistor at any point along the load line. The point at which the base current of 50μa crosses the load line corresponds to a collector voltage of 5V. The load line in Figure 32-3 is a d-c load line and indicates the d-c operation of the transistor in a circuit configuration, and the series resistance and collector supply voltage in the circuit determines the region of operation on the characteristic curves. Any two points on the load line can be selected to establish it. As an example, assume that 2ma of collector current is flowing in the transistor in Figure 32-3. The 2ma must flow through 2.2K, dropping 4.4V. The remainder of the voltage is across the transistor. Therefore, the point where 2ma crosses 5.6V (4.4V across the load resistor leaves 5.6V across the transistor) is one point on the d-c load line. This point extended to 10V gives the d-c load line for the transistor. To construct a d-c load line on a collector family of curves, simply find two points of operation for the particular series resistance and supply voltage in the transistor circuit configuration. (We will find that the same is true for an a-c signal load line; however, we will have to deal with the impedance in series with the transistor, rather than simply the d-c resistance.) The slope of the load line such as shown in Figure 32-3 will be the same when the series resistance is 2.2K regardless of the value of supply voltage. In other words, if the supply voltage is 5V rather than 10V, the load line has exactly the same slope, but is placed differently on the collector family of curves. To illustrate this, assume that the collector supply voltage is 5V in Figure 32-3 and construct a load line. If you





TEST SETUP

TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER
TOP PHOTO TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA

R_L = COLLECTOR LOAD RESISTOR
R_F = BASE BIASING RESISTOR
V_{CC} = COLLECTOR SUPPLY VOLTAGE

FIGURE 33-3

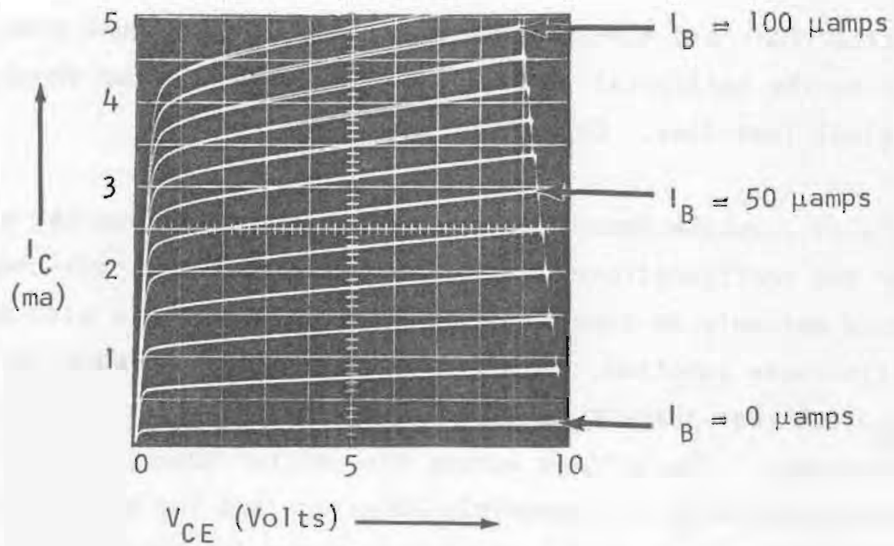
constructed your load line properly, you should have found that it extended between 5V on the horizontal to 2.27ma on the vertical, and should run parallel to the original load line. Check your load line and see.

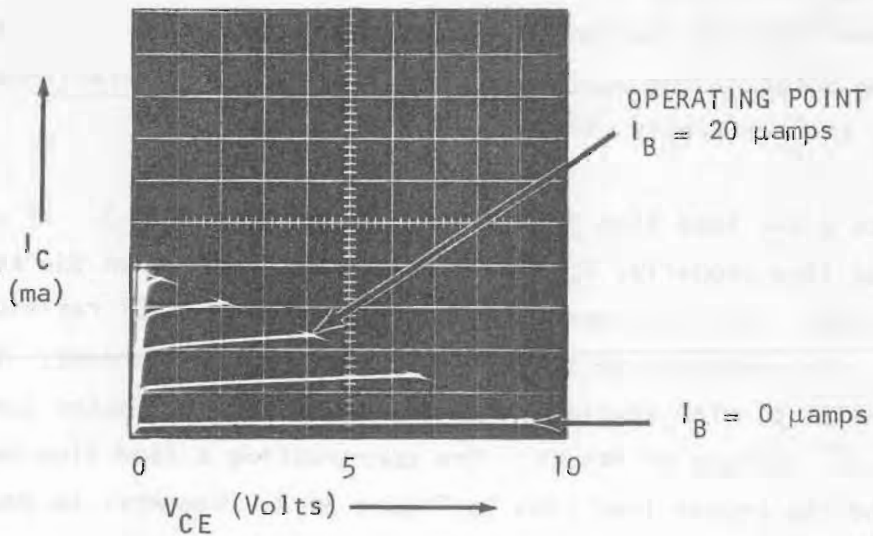
Biasing with a synthesized constant current source is typical with transistors. Consider the configurations shown in Figure 33-3. The collector supply voltage is being used not only to supply the collector bias, but is also being used to bias the emitter-base junction. Using a high resistance of 500K for the resistor designated R_F, the base current magnitude is essentially established by the value of this resistance. The voltage across the emitter-base junction is a very small value with respect to 10V (possibly 300mv or less for a germanium transistor). Neglecting the voltage across the emitter-base junction, there will be approximately 10V across 500KΩ of resistance. This sets the current at approximately 20μa. The operating point on the collector family of curves is established by this biasing resistor and the supply voltage.

Construct a d-c load line for the circuit in Figure 33-3. If you have constructed your load line properly, it should extend between 10V on the horizontal and 2ma on the vertical. With the operating point set by the 500K resistor (R_F) in the base circuit, the operating point is set at 20μa of base current. This is illustrated in Figure 34-3. The static operating is set at a collector current of ≈1.2ma and a collector voltage of ≈3.5V. Try constructing a load line on Figure 35-3. You will find the proper load line in Figure 36-3. Remember to determine the operating point as well.

TRANSISTOR SATURATION:

As discussed earlier in this volume, when both junctions of a transistor become forward biased, the transistor is termed in saturation. It is possible in the common emitter configuration, for instance, to cause sufficient transistor current to flow, that the collector voltage falls below the base voltage and the collector junction becomes forward biased. When this happens, both the collector and the emitter start to inject carriers into the base. Since the current in the collector is the algebraic sum of the two diffusion currents, the collector current is decreased. This is one extreme of the transistor's operation. Of course, the other





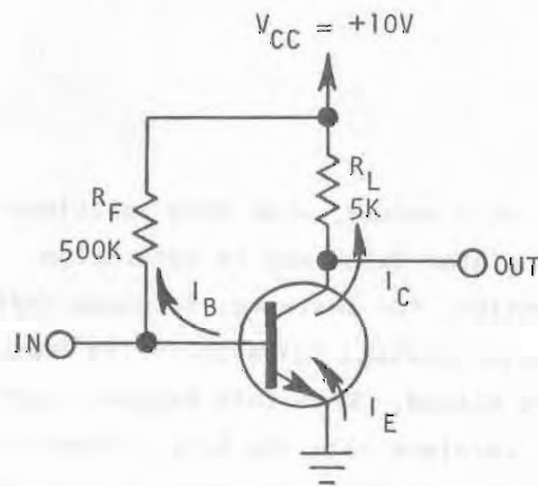
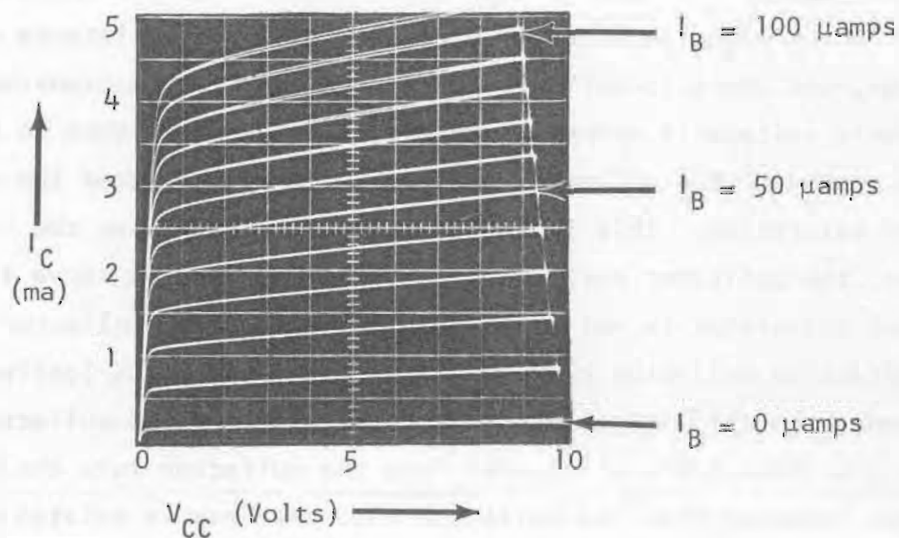
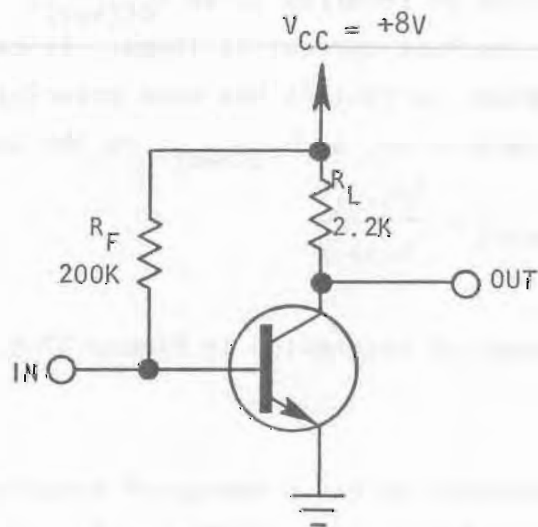


FIGURE 34-3





R_L = COLLECTOR LOAD RESISTOR

R_F = BASE BIASING RESISTOR

V_{CC} = COLLECTOR SUPPLY VOLTAGE

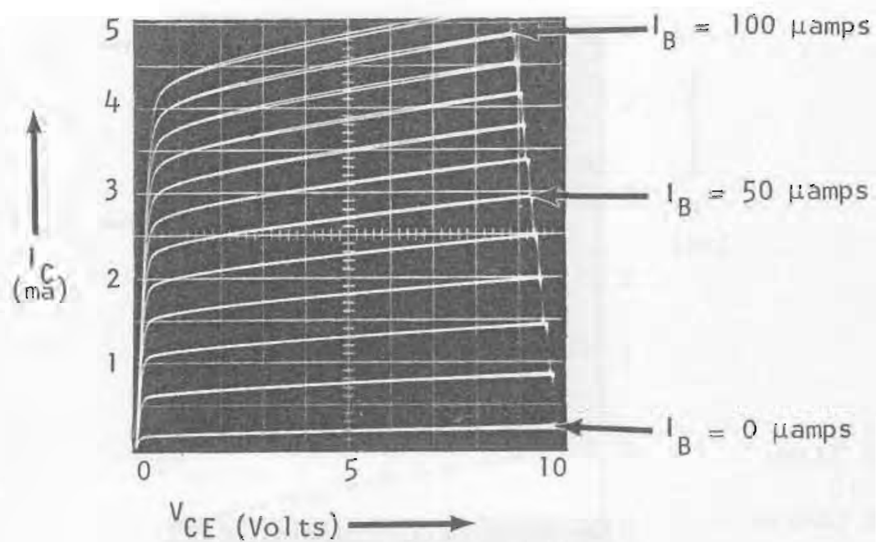
FIGURE 35-3

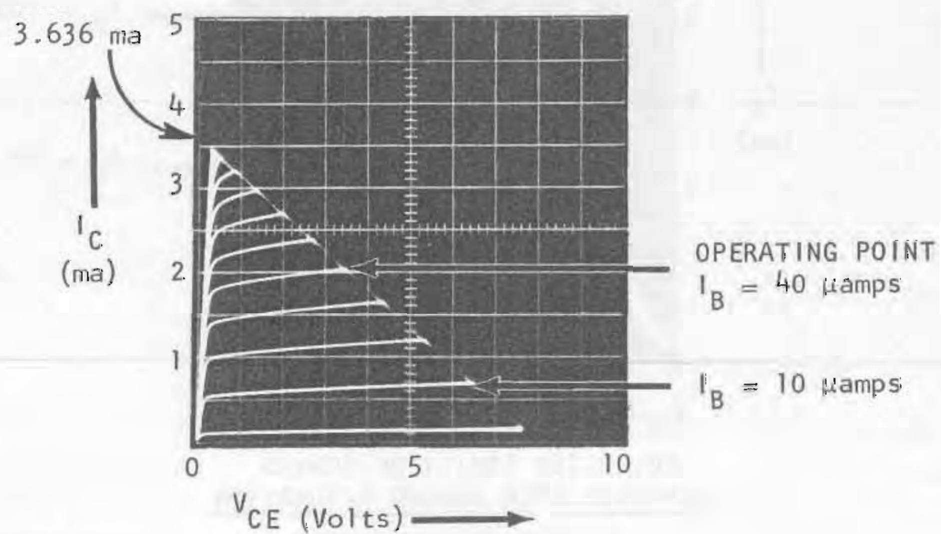
extreme is when the transistor is in cut-off, and the only current flowing is a result of the hole electron pairs present due to thermal energy and perhaps some surface leakage. In the off-state, the transistor offers a very high resistance (near open circuit), and the voltage across the transistor approaches the supply voltage. In the opposite extreme, or saturation, the resistance of the transistor is very low, and there is only a small voltage across the transistor. The remainder of the supply voltage is dropped across the series resistance in the circuit. The left hand region of the collector family of curves indicates the region around transistor saturation. This is shown in Figure 37-3. When the transistor enters saturation, the collector current on a given base current curve is reduced. As long as the transistor is not in saturation, changes in collector voltage have only a small affect on collector current. Once into saturation (collector becoming forward biased), the collector voltage has a large affect on collector current. This is a result of the carriers injected from the collector into the base. The number of carriers injected from the collector into the base is related exponentially to the forward voltage collector to base. Table 8 pursues this further for the learner who wishes to approach the subject in a deeper fashion. Table 8 is not necessary for the coverage in the volume. The collector to emitter voltage at which a transistor enters saturation is referred to as $V_{CE(sat)}$. $V_{CE(sat)}$ in Figure 37-3 is approximately 0.3V when the base current is 100μa. It can also be said that when the transistor's collector current is 4ma upon entering saturation, the collector voltage is 0.3V. Beta at saturation, or $h_{FE(sat)}$, is the current gain of the transistor at saturation.

$$h_{FE(sat)} = \frac{I_{C(sat)}}{I_{B(sat)}}$$

Assuming 100μa of base current at saturation in Figure 37-3, $h_{FE(sat)}$ is equal to 40.

Any transistor in forward conduction has a charge of minority carriers, effectively stored in the base. The minority carriers diffusing from the emitter to the collector represent a stored charge. In order to turn the transistor on, the stored charge must be established in the base. The stored charge must be removed to turn off a conducting transistor. When the transistor enters saturation, the collector injecting carriers into the base increases the stored charge in the base. It will take a longer period of time to remove the stored charge when attempting to turn off a transistor that is in saturation. It will also take a longer period





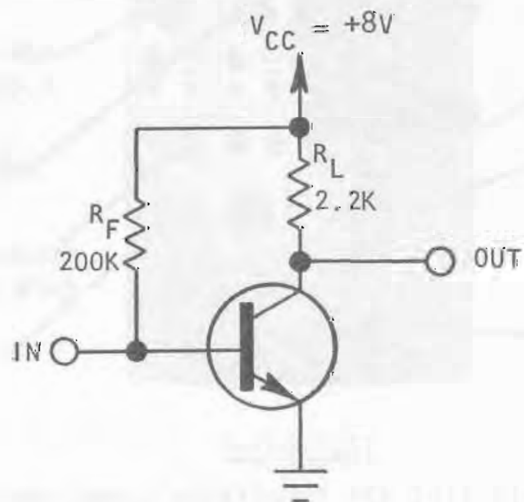
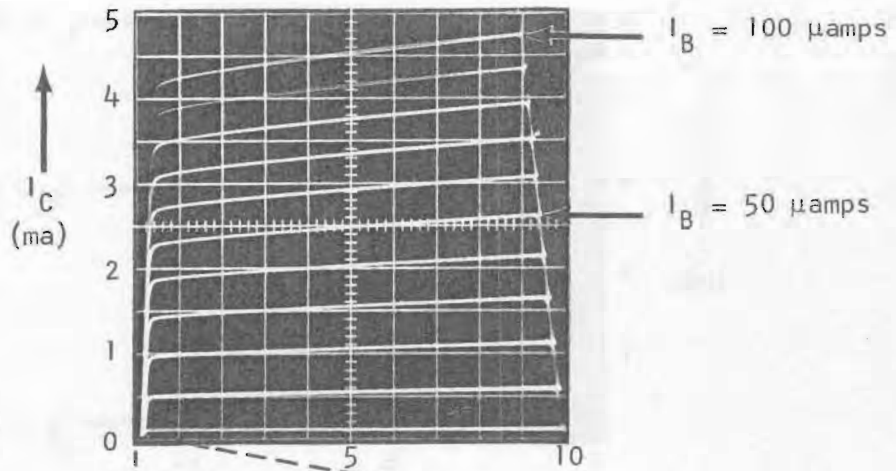
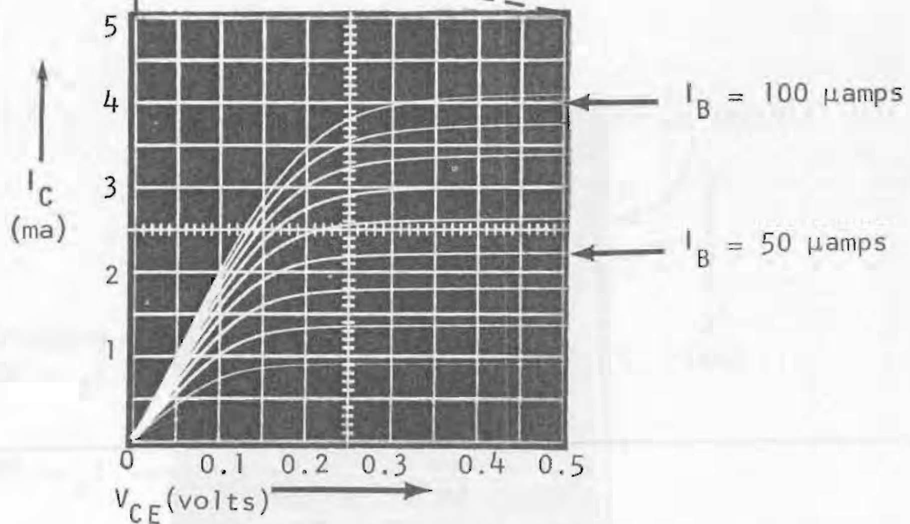


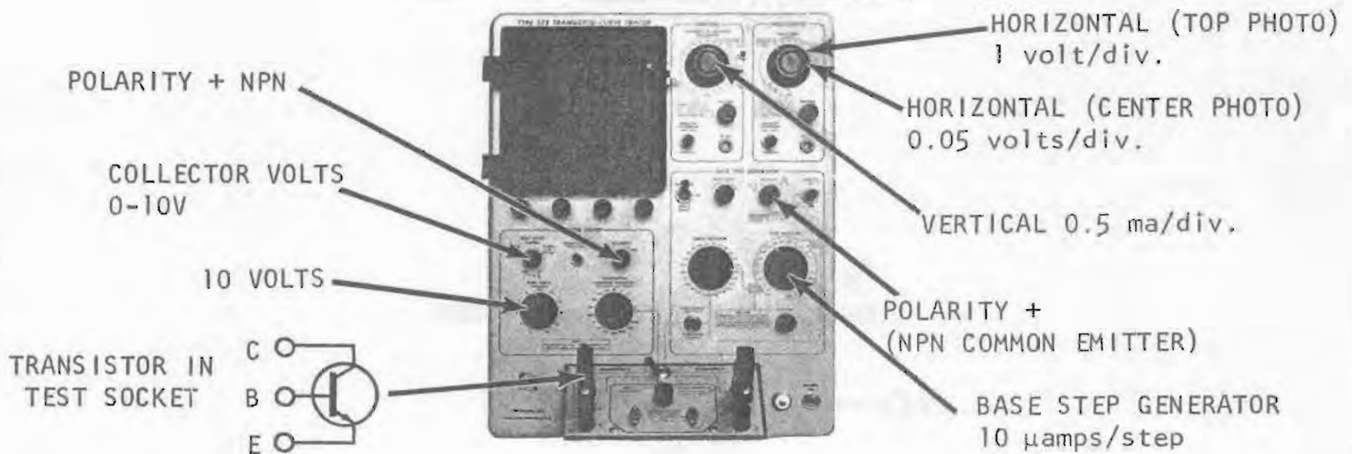
FIGURE 36-3



THESE PHOTOS TAKEN
WITH TYPE C-12
OSCILLOSCOPE CAMERA



COLLECTOR FAMILY OF CURVES
EXPANDED AREA AROUND SATURATION



TEST SETUP

TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER

FIGURE 37-3

of time to establish the stored charge when turning on a transistor to saturation. The deeper the transistor is driven into saturation, the more stored charge can be expected in the base and the more time it will take to establish and remove this stored charge in an attempt to turn the transistor on and off. This is illustrated in Figure 38-3. Figure 38-3 shows the stored charge distribution for a forward conducting transistor and for a transistor which has the collector junction reverse biased, and the combination of the two representing the stored charge of a forward conducting transistor that is in saturation. Notice in Figure 38-3 that the transistor on entering saturation has the stored charge in the base increased greatly over the stored charge for a forward conducting transistor not in saturation. For a given impedance, it will take a longer period of time to establish this greater stored charge when turning the transistor on to saturation and also a greater period of time to remove the stored charge when an attempt is made to turn off the conducting transistor.

The increase in the stored charge in the base of the transistor, when the transistor enters saturation, will greatly effect the switching characteristics of the transistor.

REVIEW:

The transistor configuration and characteristic curves in Figure 39-3 will be used in this example. We will construct a load line to determine the operating point base current, collector voltage and collector current for the circuit in Figure 39-3. Then, with an applied signal current of 10μa peak-to-peak, we will determine what signal current is flowing in the output circuit and what is the circuit current gain. One extreme of the transistor's operating range in Figure 39-3 is the supply voltage. The d-c load line will intersect the 0 collector current line at 9V of collector voltage in Figure 39-3. The second extreme of the transistor's operation is determined by the series resistance in the paths of the emitter and collector current. The load line will intersect the 0 collector voltage line at 1.8ma of collector current. This is determined by dividing the collector supply voltage of 9V by the series resistance which is 5KΩ, and the result is the maximum current that can flow, or 1.8ma. Connecting these two points with a line is the d-c load line for the circuit configuration in Figure 39-3. The operating point

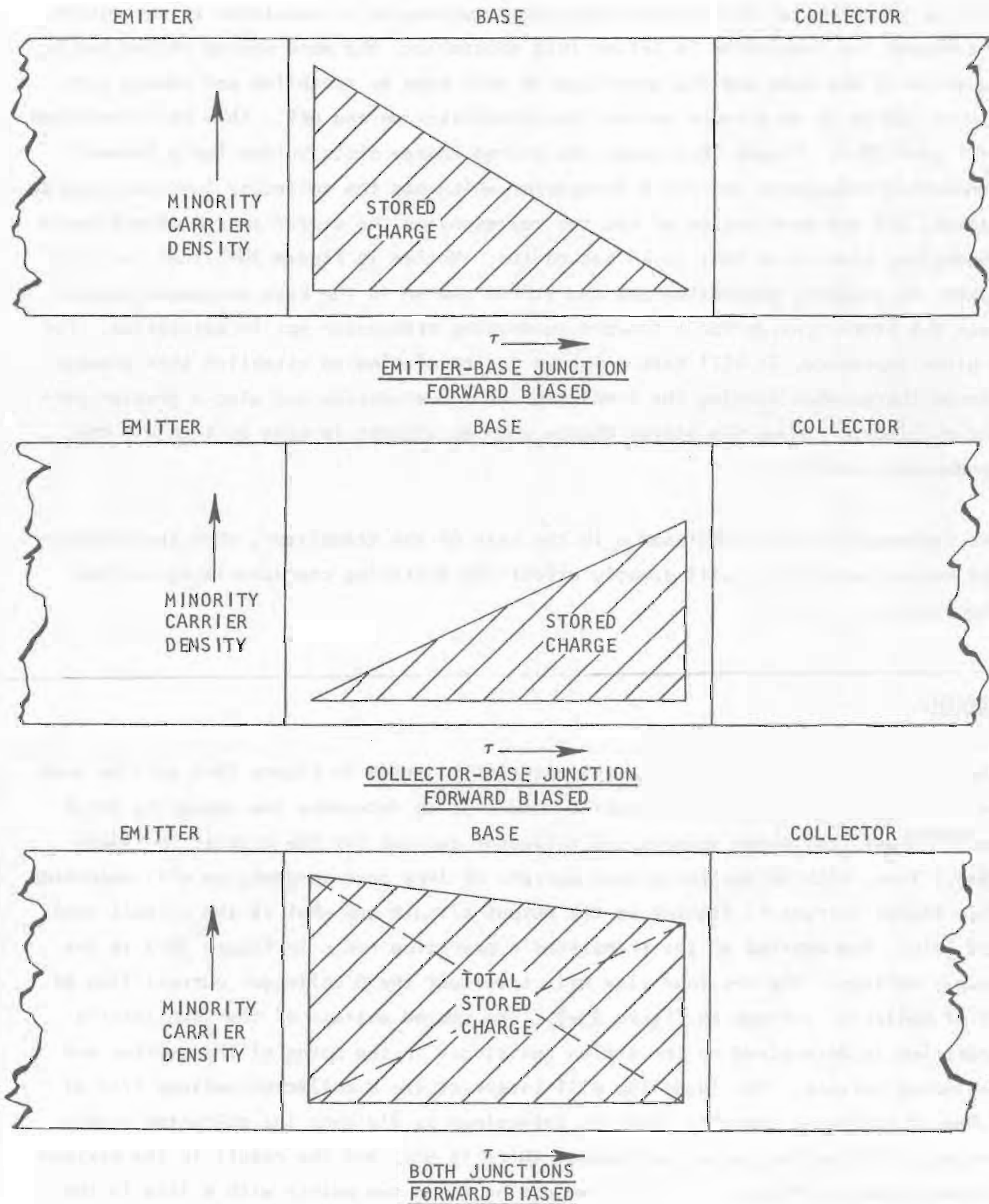


FIGURE 38-3

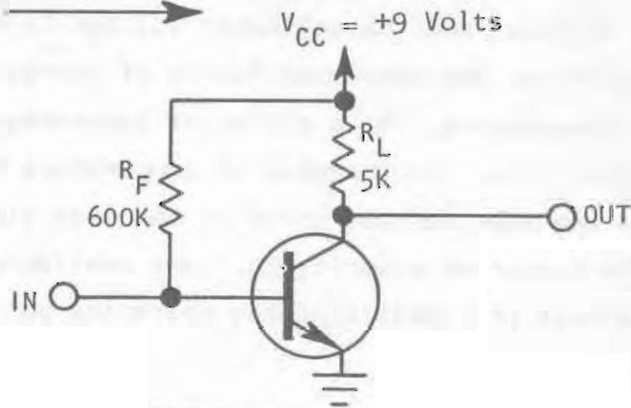


FIGURE 39-3

base current is determined by the base biasing resistor, R_F.

The operating point base current in Figure 39-3 is found by dividing 9V by the 600KΩ of R_F. Notice that R_F is nothing more than synthesizing a constant current source to supply base current. The 9V in series with the 600KΩ R_F is a high impedance source to supply base current to the transistor in Figure 39-3. This is illustrated in Figure 40-3.

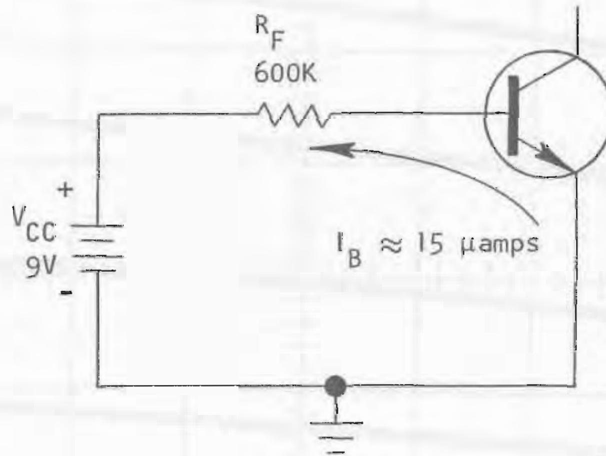


FIGURE 40-3

The 600K resistor in series with the emitter-base junction of the transistor, and the 9V source, represents a very high resistance biasing source. The current in the closed loop is essentially governed by the value of the resistor R_F. An input generator, applied at the terminal marked IN in Figure 39-3, will vary the current in the emitter-base junction that will, in turn, vary the current in the collector circuit. The operating point of 15μa, when plotted on the load line in Figure 39-3, gives a collector current of approximately 0.74ma and a collector voltage of 5.4V. In other words, at the static operating point, the collector current is 0.74ma, the base current is 15μa, and the collector voltage is 5.4V. This is assuming that the temperature at which the collector family of curves was measured in Figure 39-3 is the operating temperature. At a different temperature, the operating point will move on the d-c load line. An increase in temperature results in the operating point existing at a higher current point on the load line. An increase in temperaturn increases the number of minority carriers available and increases the quiescent current. Methods of stabilizing the operating point will be discussed in a later volume.

40 μa

2 ma

LOAD LINE

20 μamps

1 ma

$\Delta I_C = 0.48$ ma

0.74 ma

$\Delta I_B = 10$ μamps

I_C

5.4V

0 μamps

0 5V 10V

V_{CE} ⟶

A_i = CURRENT GAIN

$$A_i = \frac{\Delta I_C}{\Delta I_B} \text{ (ALONG THE LOAD LINE)}$$

$$A_i = \frac{0.48 \text{ ma}}{10 \text{ μamps}}$$

$$A_i = 48$$

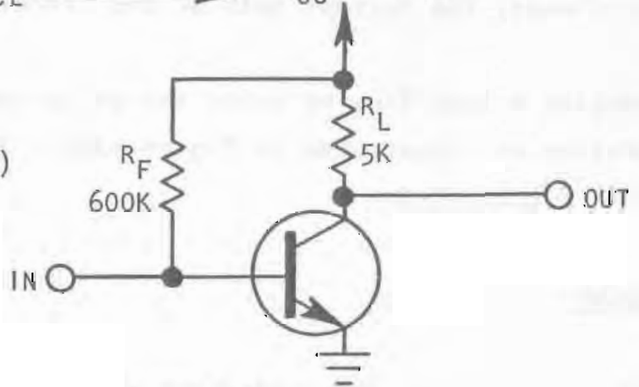


FIGURE 41-3

The beta of the transistor will give the current gain of the transistor itself, but does not take into account the effects of the circuit. Making a current gain measurement along the load line, however, gives a measurement of the current gain of the circuit, and not just the transistor.

Assuming an input signal current of 10μa peak-to-peak, the base current will vary 5ma above and below the operating point in Figure 39-3. It will vary from 15μa down to 10μa and up to 20μa over the full swing of the input signal. Plotting this change along the load line in Figure 39-3 and noting the change in collector current that occurs along the load line, the needed values can be obtained to calculate the current gain of the circuit. The collector current change that occurs along the load line in Figure 39-3 is 0.48ma for a peak-to-peak change of 10μa in base current. Current gain in this configuration is the change in collector current divided by the change in base current. In the case of the transistor in Figure 39-3:

$$A_i = \frac{\Delta I_c}{\Delta I_b} = \frac{0.48\text{ma}}{10\mu\text{amp}} = 48$$

The circuit current gain in Figure 39-3 is 48. Beta of the transistor in Figure 39-3 is over 50; however, circuit components bring the circuit gain down to 48. The load line and measurements that we have been discussing are illustrated in Figure 41-3. The slope of the d-c load line is established by the value of d-c resistance in series with the transistor and its placement is determined by the value of supply voltage. The base current is determined by the value of the constant current synthesizing resistor, R_F. This determines the quiescent operating point of the transistor. If the applied change in base current is assumed to take place along the load line, the change in circuit collector current can be found. If the change in circuit or output collector current is divided by the applied input base current change, the current gain of the circuit can be found.

Practice constructing a load line by using the collector family of curves and the circuit configuration as illustrated in Figure 42-3. The load line and operating point are shown in Figure 43-3.

COLLECTOR BREAKDOWN:

In normal amplifier operation, the transistor's collector junction is reverse biased

40 μamps

2 ma

20 μamps

1 ma

I_C

0 μamps

0 5V 10V

V_{CE}

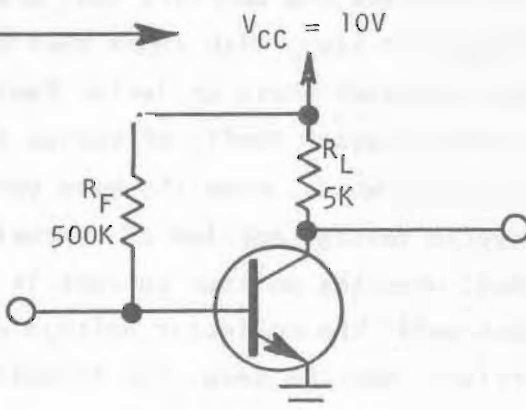


FIGURE 42-3

except when driven into saturation. A large reverse voltage can result in the collector entering an avalanche breakdown mode. Avalanche breakdown in transistors can be related to avalanche breakdown as discussed for diodes.

There are minority carriers present in all parts of the transistor at normal operating temperatures. Since the collector junction is reverse biased, any minority carriers that enter the transition region at the collector junction will cross the junction. A high value of reverse voltage will accelerate the carriers to great speeds. The accelerated carriers will strike other atoms in the structure, freeing more carriers which are also accelerated by the high reverse voltage. This multiplication of carriers at high values of reverse voltage results in the collector entering a reverse breakdown or an avalanche condition. Avalanche breakdown occurring is illustrated in Figure 44-3. The top photo in Figure 44-3 shows a collector family of curves for a transistor, as plotted on the Tektronix Type 575 Transistor-Curve Tracer. The second photo is a collector family of curves for the same transistor, but includes the area around collector avalanche breakdown. When the collector breakdown voltage is exceeded, the collector current increases greatly as shown in the second photo in Figure 44-3. For the condition indicated for zero base current (I_B) in Figure 44-3, avalanche breakdown occurs at approximately 20V, V_{CE}.

Avalanche breakdown in the collector of the transistor occurs as a result of the same mechanism as discussed for the basic diode. The orientation of the transistor, however, can effect the voltage point at which breakdown occurs as a result of the transistor's gain influencing the breakdown point. If the emitter is injecting carriers into the base, there will be many more minority carriers in the base than if the emitter is not injecting carriers into the base. The high value of reverse voltage will tend to accelerate the carriers that are present in the base as a result of the emitter injection along with those that are present due to thermal agitation. Consider the transistor whose collector family of curves is shown in photo one in Figure 45-3. The collector family of curves shows the transistor breaking down with zero I_B. In other words, when the base current is zero, the transistor breaks down with a reverse voltage applied of approximately 20V. Notice in photo two of Figure 45-3 that, when the emitter current is reduced to zero, the transistor does not break down until the collector voltage exceeds 25V. Without the emitter injecting carriers into the base, the transistor will have a higher reverse breakdown voltage rating. The reverse breakdown voltage with the emitter current

40 μamps
2 ma
LOAD LINE
OPERATING POINT
20 μamps
1 ma
I_C
0 μamps
0
5V
10V
V_{CE}

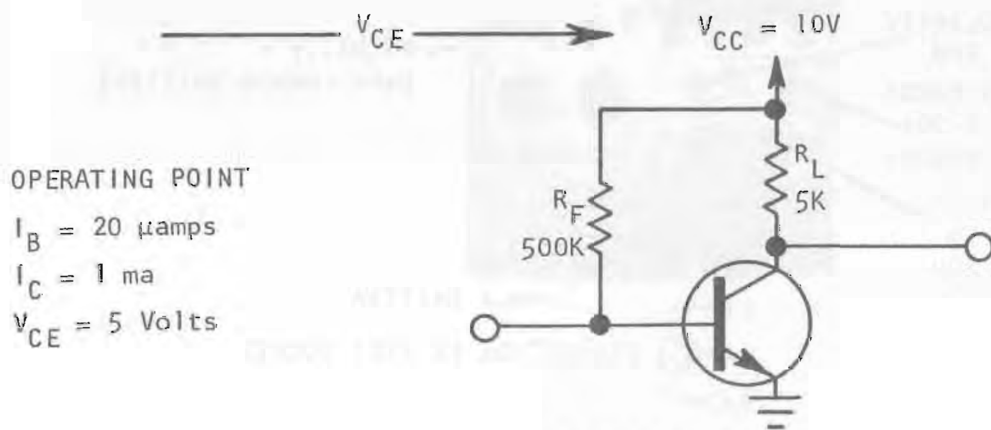
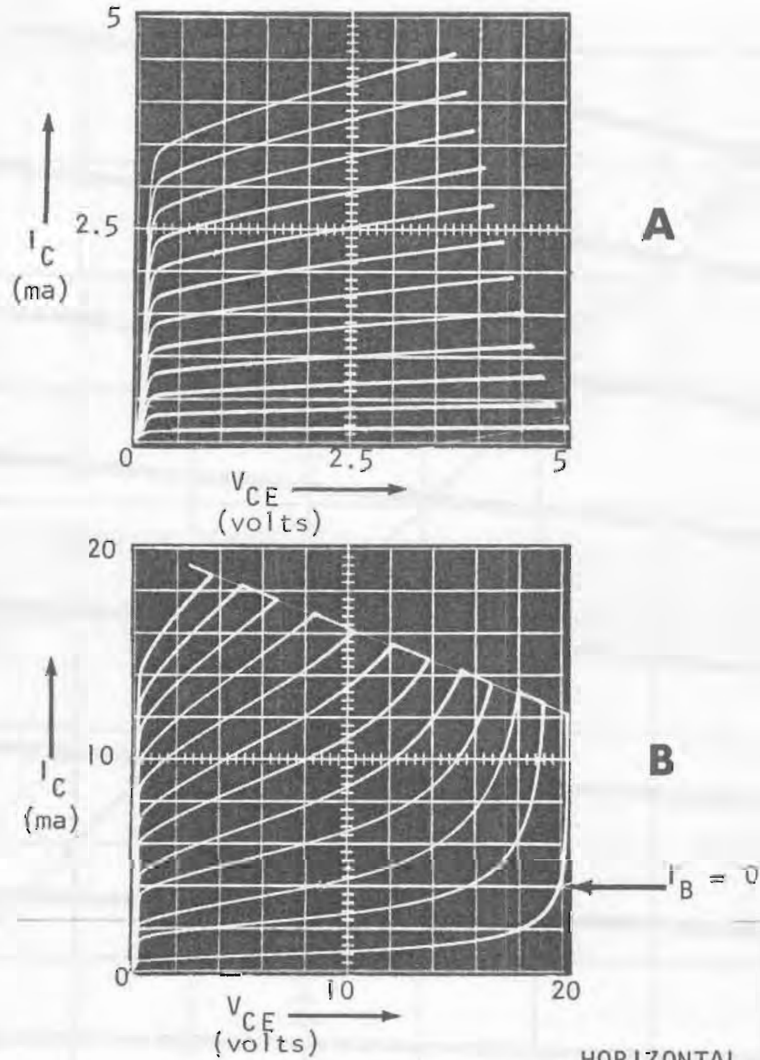


FIGURE 43-3



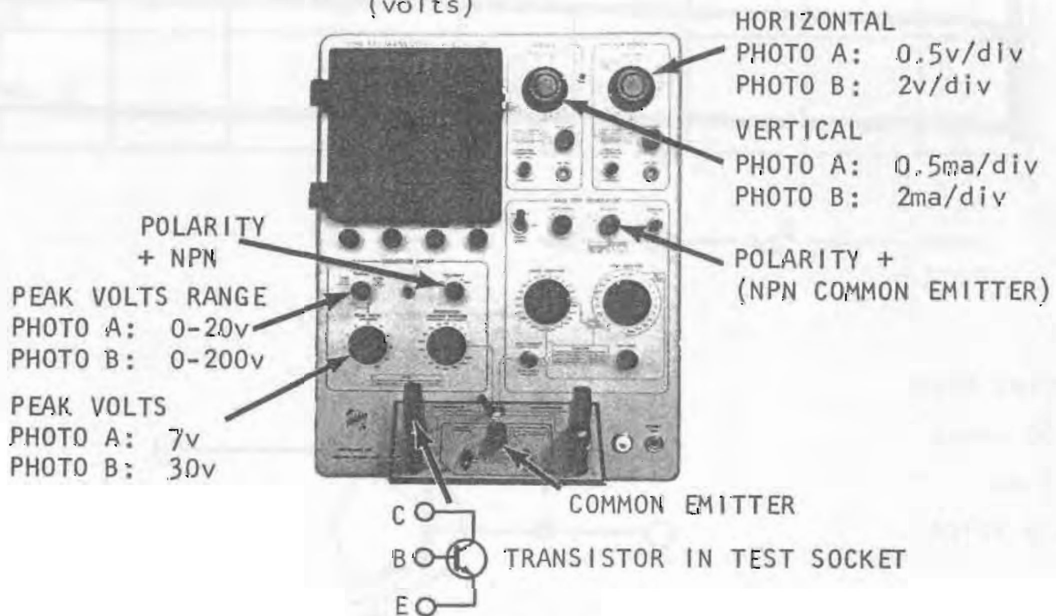
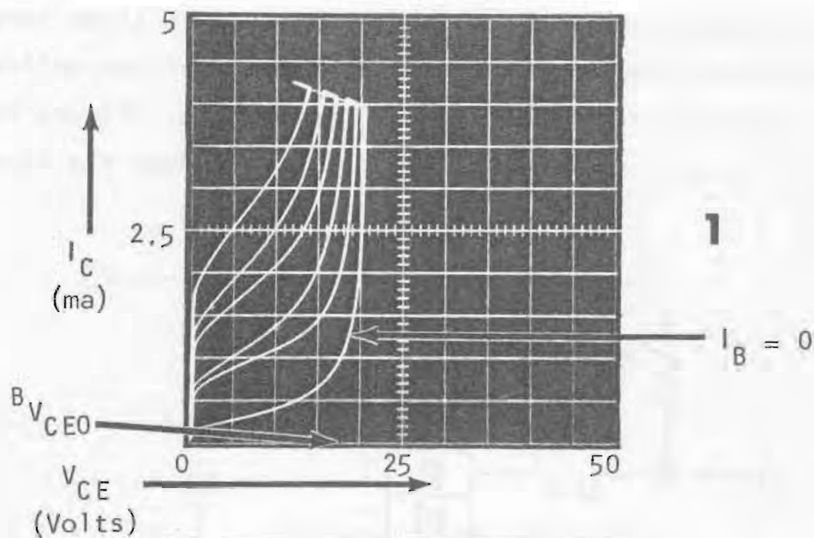
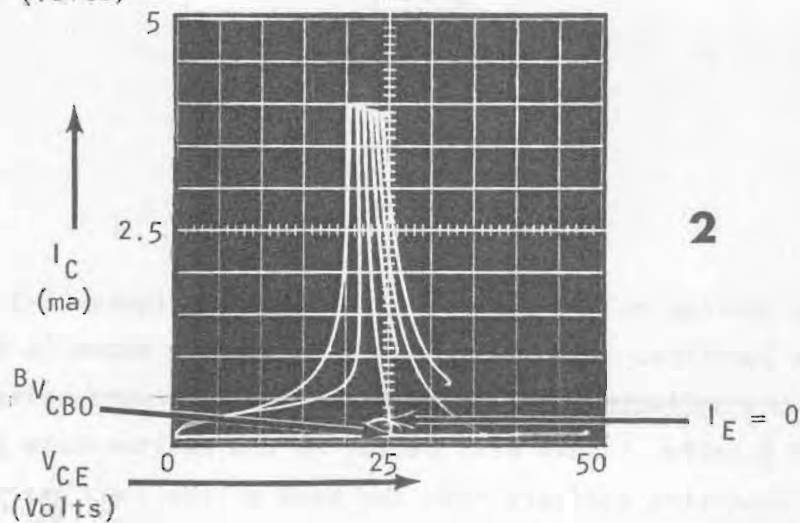
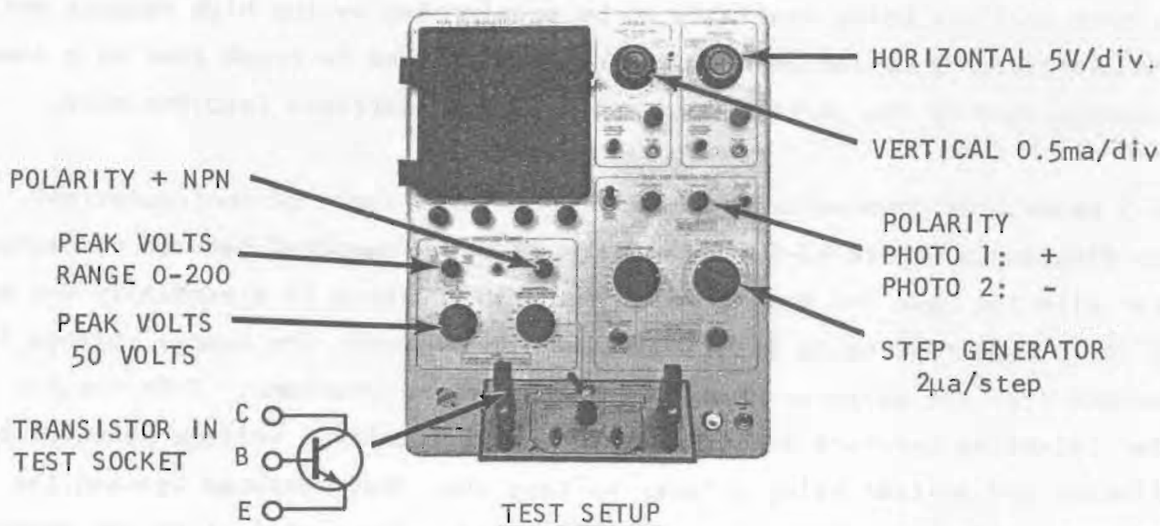


FIGURE 44-3







TEST SETUP

TEKTRONIX TYPE 575 TRANSISTOR-CURVE TRACER
PHOTOS 1 AND 2 TAKEN WITH TYPE C-12 OSCILLOSCOPE CAMERA
FIGURE 45-3

equaling zero, is the normal avalanche breakdown point at a given temperature for the collector base diode. The reverse breakdown voltage of the collector base diode is reduced when the emitter injects carriers into the base. Figure 46-3 illustrates how the emitter can inject carriers into the base even though the base lead is d-c open circuited.

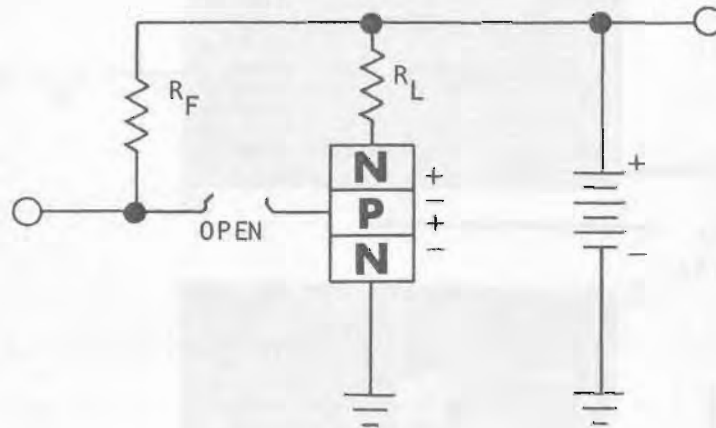
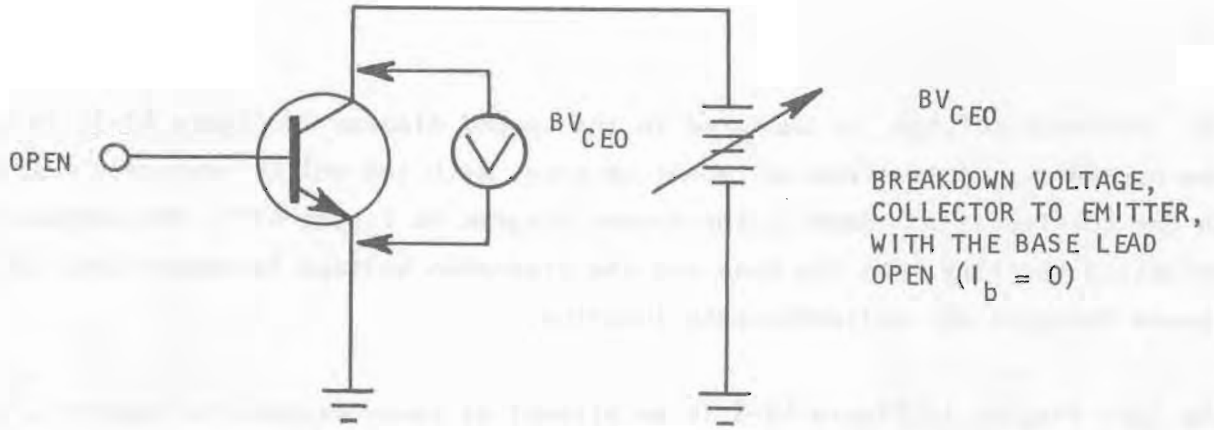
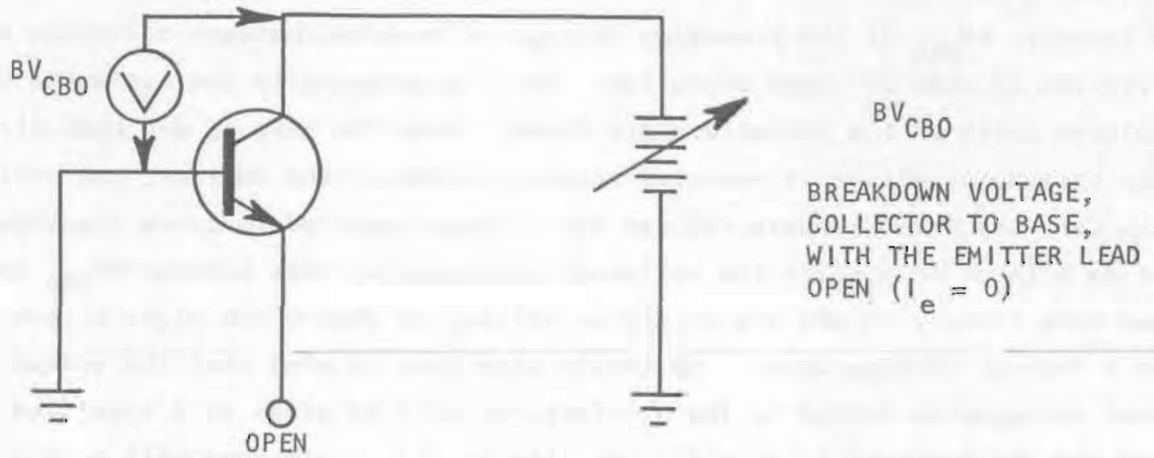


FIGURE 46-3

The reverse voltage applied to the collector junction in Figure 46-3 is also applied to the emitter-base junction. Although the supply voltage shown is a reverse voltage as applied to the collector-base junction, it is a forward voltage as applied to the emitter-base junction. This will result in the emitter-base junction being forward biased and injecting carriers into the base of the transistor. This will result in more carriers being available to be accelerated by the high reverse voltage at the collector junction and the collector will tend to break down at a lower applied voltage than if the emitter were not injecting carriers into the base.

Figure 47-3 shows breakdown voltages as given for three types of configurations. In the top diagram in Figure 47-3, the voltage is being measured between collector and emitter with the base d-c open circuited. When the base is essentially d-c open circuited (or a high resistance is placed in the base lead), the supply voltage is applied across both the collector-base and emitter-base junctions. This results in the emitter injecting carriers into the base and the breakdown voltage measured between collector and emitter being a lower voltage than that measured between the collector and base, with the emitter d-c open circuited. The symbol given the voltage as measured in the top diagram in Figure 47-3 is BV_{CEO}. This indicates the breakdown voltage collector to emitter, with the 0 indicating that the third terminal, the





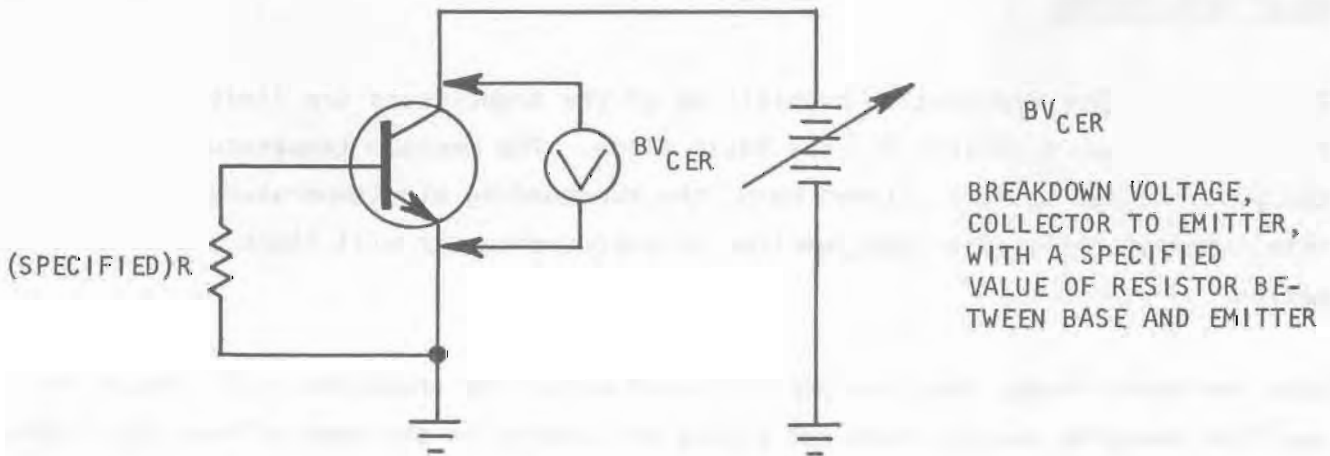


FIGURE 47-3

base, is d-c open circuited.

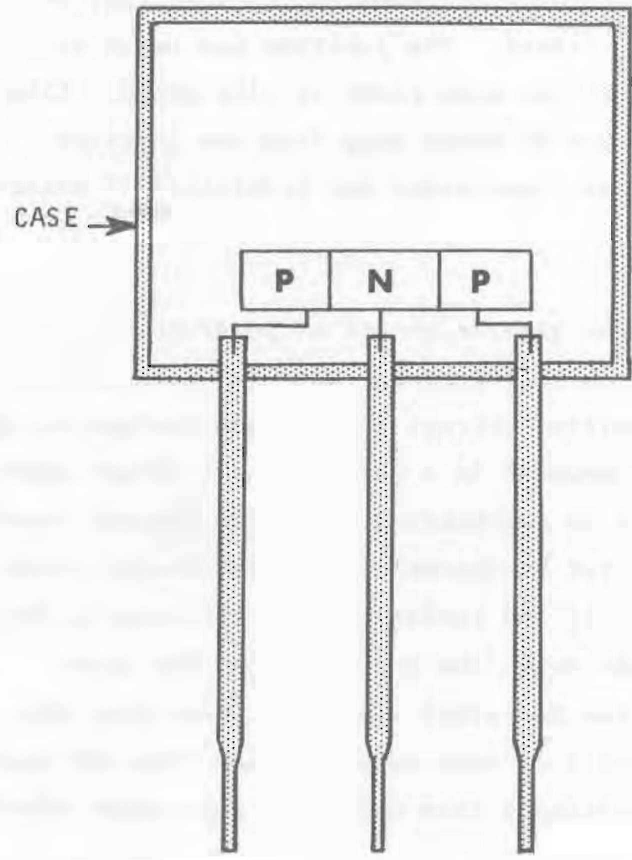
The breakdown voltage, as measured in the second diagram in Figure 47-3, is BV_{CBO}, the breakdown voltage from collector to base, with the emitter d-c open circuited. In the configuration shown in the second diagram in Figure 47-3, the emitter is not injecting carriers into the base and the breakdown voltage is essentially the avalanche point of the collector-base junction.

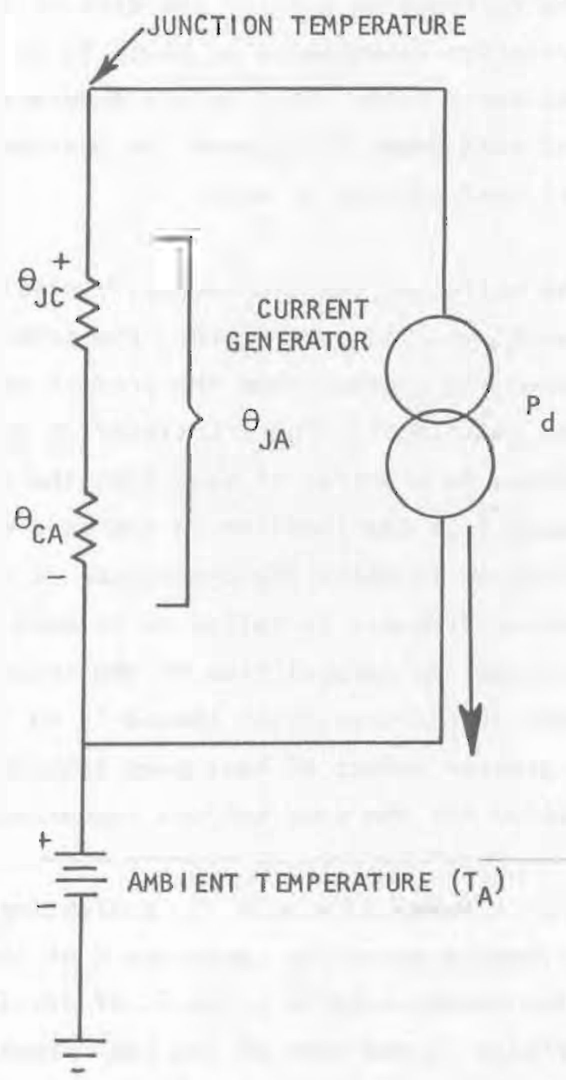
The last diagram in Figure 47-3 is an attempt by manufacturers to specify a breakdown voltage that will more closely relate to an actual circuit configuration. In the final diagram in Figure 47-3, the breakdown voltage is between collector and emitter with a specified value of resistance between the base and the emitter. As a summary, BV_{CBO} is the breakdown voltage as measured between collector and base with the emitter d-c open circuited. This is essentially the avalanche breakdown voltage point of the collector-base diode. When the base is d-c open circuited and the breakdown voltage is measured between collector and emitter, the emitter injecting carriers into the base reduces the voltage required to cause breakdown. BV_{CER} is an attempt to specify the collector breakdown voltage between BV_{CBO} and BV_{CEO}, and more closely relate the breakdown voltage to that which might be encountered in a typical configuration. We should also keep in mind that the values of breakdown voltages as stated by the manufacturer will be given at a specified temperature and, if the temperature is different, the breakdown voltages will be different.

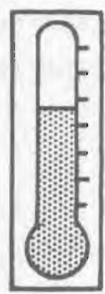
POWER DISSIPATION:

The maximum power dissipation capbilities of the transistors are limited by the same factors as those discussed for the basic diode. The maximum temperature at which the junction can operate without harm, the surrounding air temperature, and the total thermal resistance from junction to surrounding air will limit the power dissipation.

Like the basic diode, the transport of carriers in the transistor will result in carriers changing energy bands and giving off energy in the form of heat and light. The primary concern at the transistor junction is the energy given off in the form of heat. The more power dissipated at the junction, the more heat is given off.







THERMAL		ELECTRICAL EQUIVALENT
AMBIENT TEMPERATURE	=	VOLTAGE
THERMAL RESISTANCE	=	RESISTANCE
POWER DISSIPATION	=	CURRENT
JUNCTION TEMPERATURE	=	VOLTAGE

THERMAL TO ELECTRICAL ANALOGY

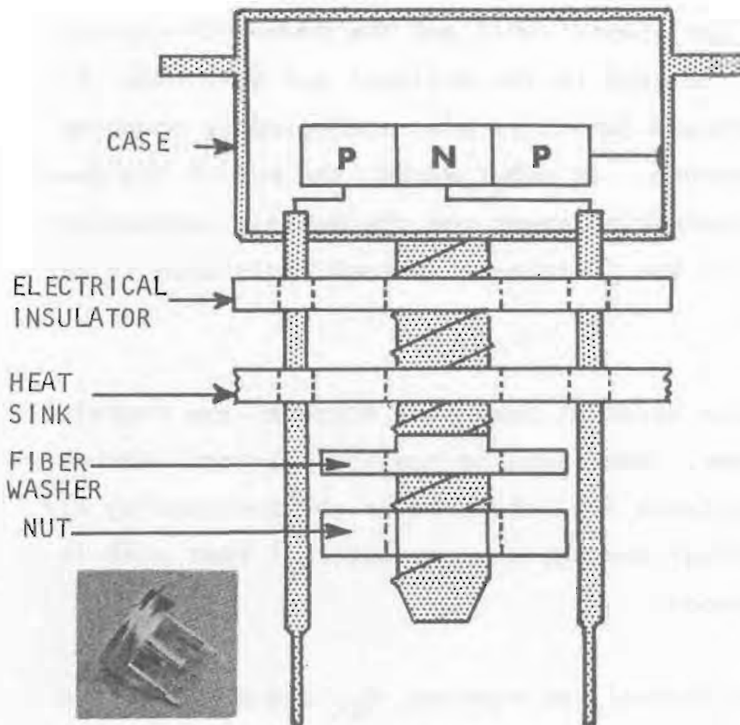
FIGURE 48-3

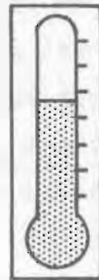
The heat generated at the junction raises the junction temperature above that of the surrounding air (or the case of the transistor). The junction can reach an operating temperature at which it is harmed if too much power is dissipated. Like the basic diode, some of the generated heat can be moved away from the junction and more power dissipated for the same ambient temperature and transistor if external heat sinking is used.

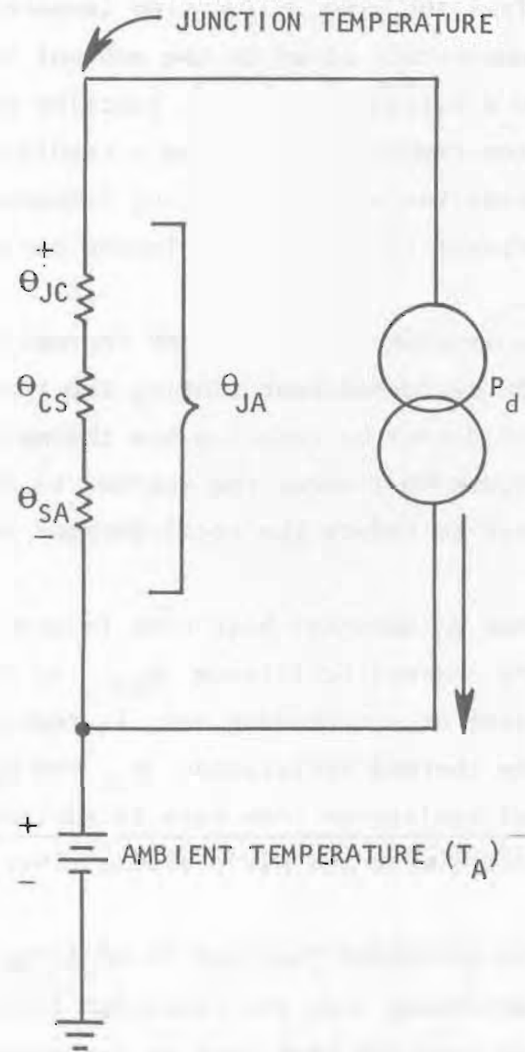
The collector junction normally dissipates the greater amount of power of the two junctions. In other words, the product of the collector current and voltage is generally greater than the product of the emitter current and voltage (referring to the junctions). The transistor is normally mounted in a case and this offers opposition to transfer of heat from the junction to surrounding air. The thermal resistance from the junction to the case will be set by thermal connection to the encapsulation in which the transistor is placed. If the radiation from the case to the surrounding air is relied on to move the heat away from the junction, the power dissipation capabilities of the transistor can be rather limited. Connecting the case of the transistor thermally to the chassis or some external heat sink can move a greater amount of heat away from the transistor's case and allow more power dissipation for the same ambient temperature.

The difference between the ambient or surrounding air temperature and the maximum allowable operating temperature of the junction, gives the allowable rise in junction temperature as a result of dissipating power at the junction. The opposition offered in the path of the heat transfer from the junction to the surrounding air is termed thermal resistance. The higher the thermal resistance, the less heat will be transfered (other considerations being equal).

Figure 48-3 shows the thermal to electrical analogy as applied to the transistor. This analogy is very similar to the analogy as used with the basic diode and the formulas are also very similar. In the majority of cases, the thermal resistance of interest is between the collector junction and the case, and between the case and the surrounding air as illustrated in Figure 48-3. The collector junction has the higher power dissipation than the emitter junction typically. Once the thermal to electrical analogy is formulated, such as shown in Figure 48-3, Ohm's Law, Kirchoff's Law, and the basic theorems may be applied to solve the thermal considerations of the transistor as was done for the basic diode earlier in this volume.







NOTE: When a heat sink is used, Θ_{CS} and Θ_{SA} replace Θ_{CA}. The sum of Θ_{CS} and Θ_{SA} is always much lower than Θ_{CA}.

JUNCTION POWER DISSIPATION = CURRENT
OPPOSITION TO HEAT TRANSFER = RESISTANCE
AMBIENT TEMPERATURE = VOLTAGE
JUNCTION TEMPERATURE = VOLTAGE

INSULATING WASHER	TYPICAL THERMAL RESISTANCE, (Θ_{CS}) IN °C/W	
	DRY	W/SILICON LUBRICANT
NONE	0.2	0.1
TEFLON	1.45	.8
MICA	0.8	0.4
ANODIZED ALUMINUM	0.4	0.35

THERMAL TO ELECTRICAL ANALOGY WHEN USING A SEPARATE HEAT SINK

FIGURE 49-3

In the analogy, the product of thermal resistance total and the power dissipation gives the rise in junction temperature (voltage in the analogy) and this rise in temperature added to the ambient temperature (which is also analogically compared to a voltage) gives the junction temperature. In other words, the sum of the junction temperature rise as a result of dissipating power and the ambient temperature gives the actual operating temperature of the junction. Thermal resistance is expressed in degrees centigrade per watt.

Connecting the transistor thermally to an external heat sink (such as the chassis) can be termed heat sinking the transistor. Heat sinking reduces the total thermal resistance by reducing the thermal resistance from the case to the surrounding air. Figure 49-3 shows the thermal to electrical analogy when an external heat sink is used to reduce the total thermal resistance.

When an external heat sink is used, two thermal resistances, Θ_{CS} and Θ_{SA}, replace the thermal resistance, Θ_{CA}. In other words, the thermal resistance, case to ambient or surrounding air, is replaced by two other thermal resistances. The sum of the thermal resistances, Θ_{CS} and Θ_{SA}, will always be much lower than Θ_{CA}, the thermal resistance from case to ambient. (These considerations are the same as was discussed with the basic diode. They are simply being applied to the transistor.)

The collector junction is of prime interest when we are dealing with the thermal resistance from the collector junction to the case, from the case to the heat sink, and from the heat sink to the surrounding air. Higher power handling transistors generally have the collector connected directly to the case to aid in the heat transfer. In most cases, this requires the case to be insulated electrically from the heat sink. The electrical insulator used must have good heat conductivity and still electrically insulate the transistor from the chassis. In some instances, the transistor manufacturer connects the emitter thermally to the case, and the transistor is referred to as a reverse polarity device.

To make a good thermal contact with the transistor, the insulating washer must have no ridges or gaps between the insulating washer and the transistor, and between the insulating washer and the heat sink. The gaps and so forth can be minimized by highly polishing the insulating washer, the transistor, and heat sink so that a good contact is made. If the polishing that must be done to make a good thermal

contact is prohibitive, silicon lubricant can be added to the insulating washer to fill in the gaps. Silicon lubricant will establish a good thermal contact between the transistor and the insulating washer and the insulating washer and the heat sink without polishing. If no heat sink is used, silicon lubricant can still be used to form a good thermal contact between the transistor's case and the heat sink. The chart at the bottom of Figure 49-3 shows the typical thermal resistances with and without silicon lubricant for several types of insulating washers.

Let's apply the thermal to electrical analogy as shown in Figure 49-3 to a typical power transister:

Given: $\Theta_{SA} = 3.2°C/watt$
$\Theta_{JC} = 1.2°C/watt$
$\Theta_{CS} = 0.6°C/watt$
Maximum junction operating temperature (T_{Jmax}) = 150°C
Ambient temperature (T_A) = 50°C

The sum of the individual thermal resistances as listed gives the total thermal resistance. In the case that we have just listed, $\Theta_{JC} + \Theta_{CS} + \Theta_{SA} = \Theta_{JA} = 5°C/watt$. The total thermal resistance for the transistor we are discussing is 5°C/watt. We have stated that the expected ambient temperature is 50°C and the maximum allowable junction temperature is 150°C. From this we can calculate the allowable rise in junction temperature as a result of power dissipation by subtracting the ambient temperature from the maximum operating temperature of the junction. In the case stated, we can allow a temperature rise at the junction of 100°C as a result of dissipating power: ($\Delta T_{Jallowable} = T_{Jmax} - T_A = 150°C - 50°C = 100°C$) We can also state that the product of power dissipation and the total thermal resistance must not exceed the maximum allowable rise in junction temperature. We can arrange this to solve for the maximum power dissipation that the transistor can tolerate. We can state in formula form that: $P_{Dmax}\, \Theta_{JA} = T_{Jmax} - T_A$

where P_{Dmax} equals the maximum power dissipation of the transistor, Θ_{JA} is the total thermal resistance, T_{Jmax} is the maximum operating temperature of the junction, and T_A is the ambient temperature. Rearranging this formula, we can say that:

$$P_{Dmax} = \frac{T_{Jmax} - T_A}{\theta_{JA}}$$

Substituting the values for the transistor that we have previously stated, we find that: $P_{Dmax} = \frac{150°C - 50°C}{5°C/watt}$.

To find the maximum allowable power dissipation for a given ambient temperature, divide the thermal resistance into the allowable change in junction temperature, which in this case is $\frac{100°C}{5°C/watt}$, which gives us a maximum allowable power dissipation of 20 watts.

Let's try this again with the following characteristics given:

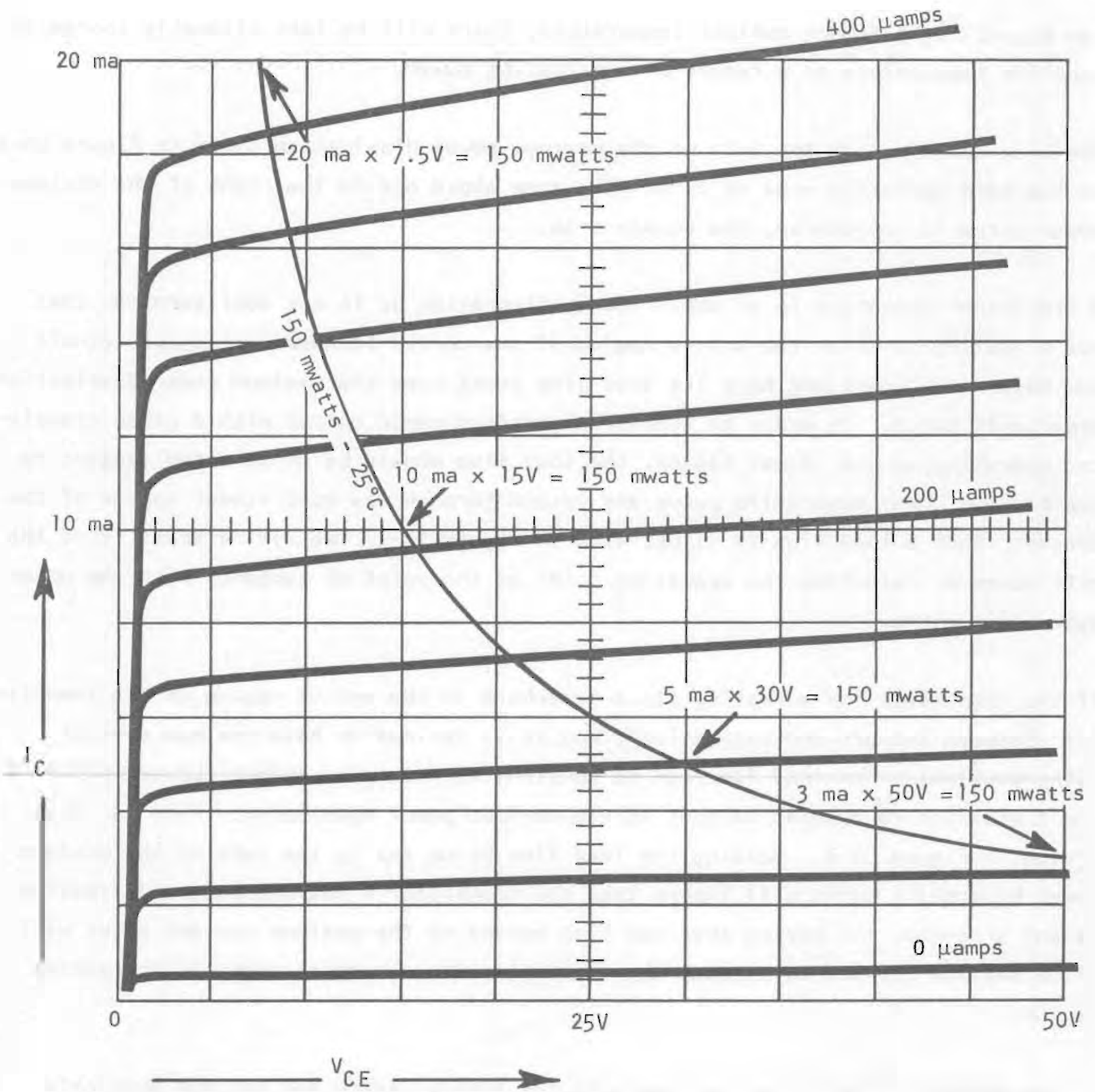
$$\theta_{JC} = 2.1°C/watt,\ \theta_{CS} = 0.8°C/watt,\ \theta_{SA} = 5.1°C/watt,\ T_A = 70°C,\ T_{Jmax} = 150°C$$

Let's solve the maximum power dissipation (P_{Dmax}) of the transistor with these conditions. With an ambient temperature expected of 70°C and a maximum allowable junction temperature of 150°C, we have an allowable temperature rise at the junction as a result of dissipating power of 80°C. The total thermal resistance is the sum of all the thermal resistances or, in this case, 8°C/watt. An allowable change in junction temperature of 80°C divided by 8°C/watt gives a maximum power dissipation of 10 watts. Showing that in formula form:

$$P_{Dmax} = \frac{T_{Jmax} - T_A}{\theta_{JA}} = \frac{150°C - 70°C}{8°C/watt} = 10 \text{ watts}$$

PLOTTING MAXIMUM POWER DISSIPATION ON THE COLLECTOR FAMILY OF CURVES:

When the maximum power dissipation of the transistor is stated by the manufacturer or calculated using thermal resistances as we have just been doing, it can be plotted on the collector family of curves as shown in Figure 50-3. This is done simply by connecting all points of maximum power dissipation with a line. Notice that the line will show a hyperbolic curve when plotted in this fashion on the collector family of curves. For the transistor in Figure 50-3, the maximum power dissipation is 150mwatts. This is a fairly low power transistor and probably doesn't incorporate an external heat sink. Connecting all points on the collector family of curves that equal 150mwatts of power dissipation gives the curve as shown in Figure 50-3. Notice that this power dissipation is given at 25°C. This same transistor at a higher temperature will have a lower maximum allowable power dissipation. The reason is evident from the calculations that we have just accomplished in the past



COLLECTOR FAMILY OF CURVES WITH A MAXIMUM POWER CURVE - 150 mwatts - 25°C

FIGURE 50-3

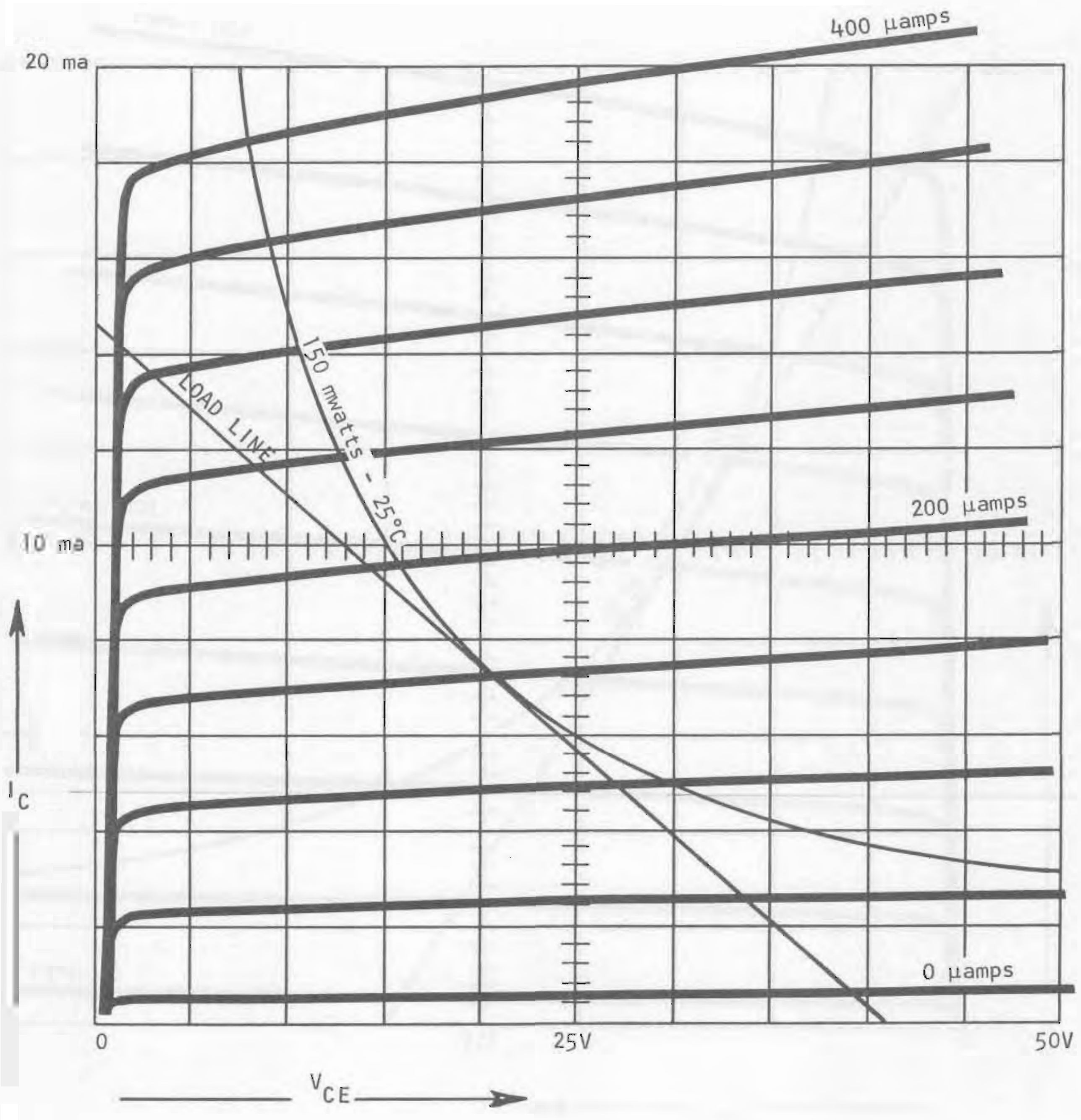
few pages. At a higher ambient temperature, there will be less allowable change in junction temperature as a result of dissipating power.

The area below and to the left of the maximum power dissipation curve in Figure 50-3 is the safe operating area at 25°C. The area above and to the right of the maximum power curve is, of course, the unsafe area.

A transistor operating in an amplifier configuration or in any configuration that has a resting point in the active region of the curves (somewhere between cut-off and saturation) must not have its load line cross over the maximum power dissipation hyperbolic curve. In order to accomplish maximum power output with a given transistor operating in the linear region, the load line should be constructed tangent to the maximum power hyperbolic curve and extend through the most linear region of the curves. Such a load line is illustrated in Figure 51-3. We will discuss later the attributes of selecting the operating point at the point of tangency with the power hyperbolic curve.

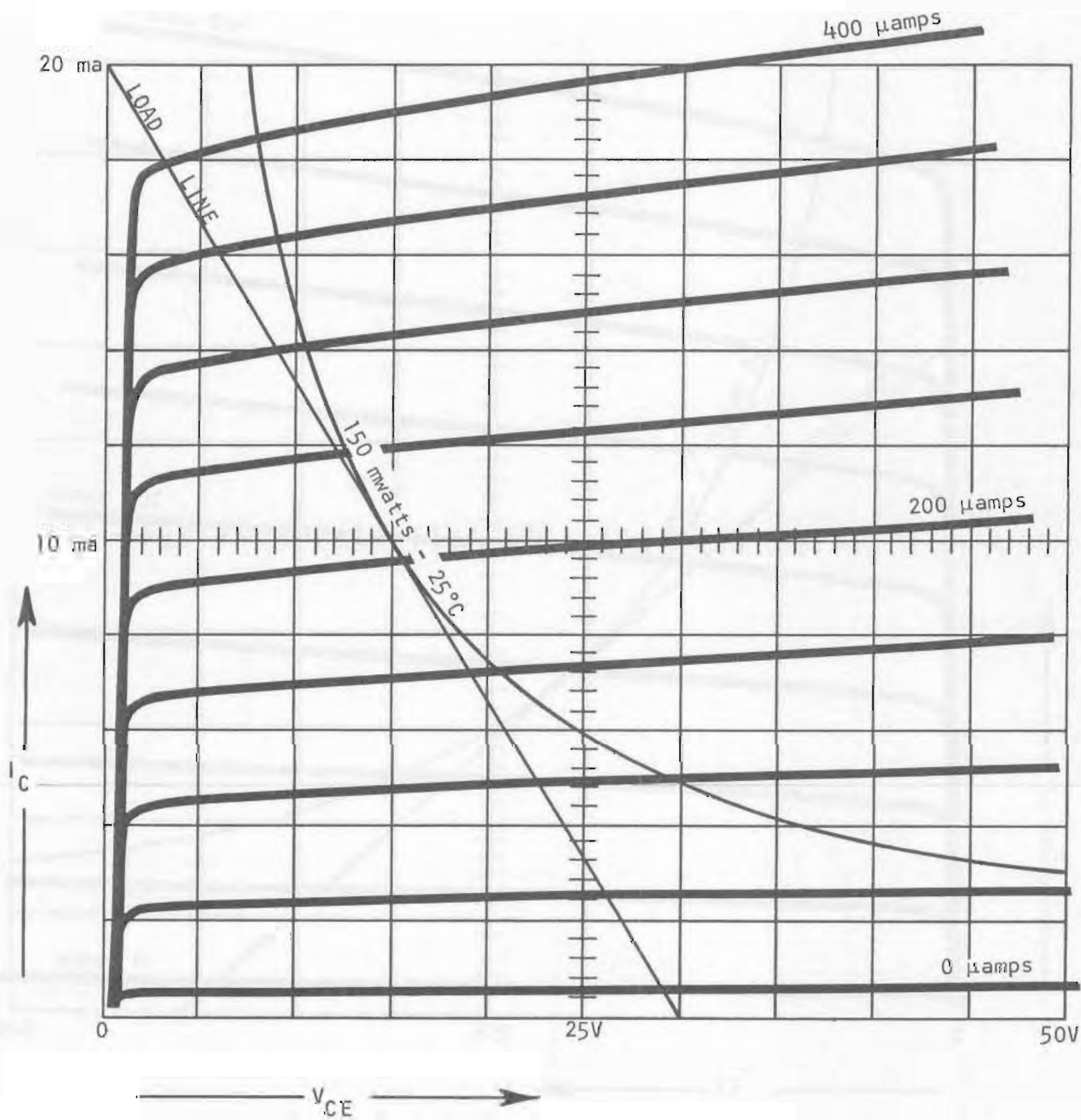
If the transistor has a resting state somewhere in the active region of the transistor (between cut-off and saturation), and it is desired to have maximum current swing available, the load line can be constructed from the maximum current point of the transistor to a point tangent to the maximum power hyperbolic. This is illustrated in Figure 52-3. Holding the load line below and to the left of the maximum power hyperbolic curve will insure that the transistor's maximum power dissipation is not exceeded, and having the load line extend to the maximum current point will allow maximum current swing when the transistor has a resting state in the active region.

If the transistor has a resting state in the active region and maximum available voltage swing is desired, the load line can be constructed from the maximum collector voltage point to a point tangent to the maximum power hyperbolic. This is illustrated in Figure 53-3. Once again, keeping the load line below and to the left of the maximum power hyperbolic curve insures that the transistor will not exceed its maximum power dissipation, and having the load line extend to the maximum voltage point insures that the maximum available voltage swing can be accomplished. The maximum power dissipation curve that we have discussed to this point is true for the transistor that must rest in the active region between saturation and cut-



COLLECTOR FAMILY OF CURVES WITH MAXIMUM POWER CURVE AND LOAD LINE FOR MAXIMUM LINEAR POWER

FIGURE 51-3



COLLECTOR FAMILY OF CURVES WITH MAXIMUM POWER CURVE AND LOAD LINE FOR MAXIMUM POWER WITH MAXIMUM POSSIBLE CURRENT SWING

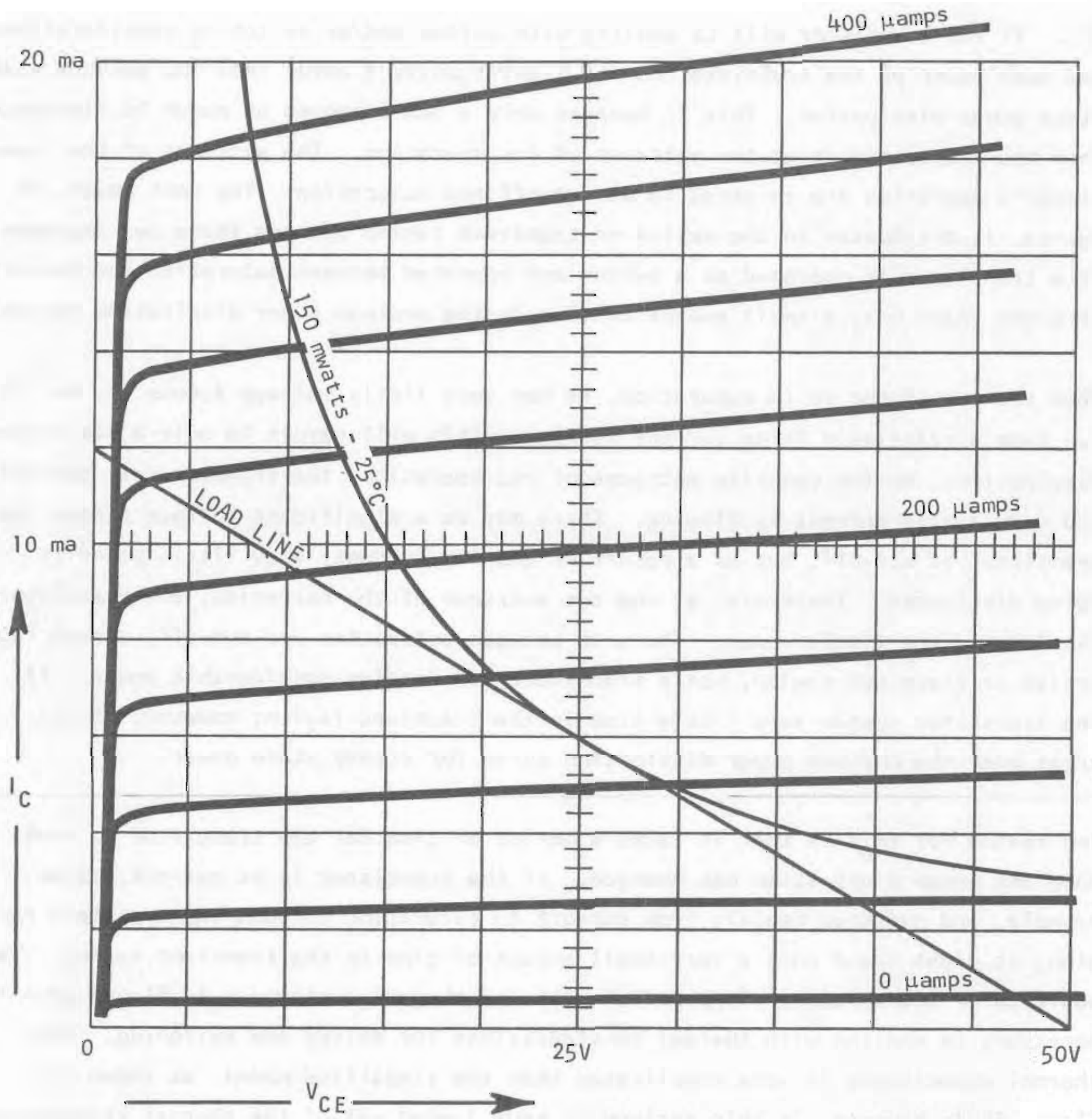
FIGURE 52-3

off. If the transistor will be dealing with pulses and/or switching considerations, the peak power of the transistor can be significantly greater than the maximum steady state power dissipation. This is because only a small amount of power is dissipated when the transistor is at the extremes of its operation. The extremes of the transistor's operation are referred to as cut-off and saturation. The most power, of course, is dissipated in the active or transient region between these two extremes. If a transistor is operated as a switch and operated between saturation and cut-off, it might spend only a small amount of time in the maximum power dissipation region.

When the transistor is in saturation, it has very little voltage across it, but it can have a relatively large current flowing. This will result in only a small power dissipation. As the opposite extremes of its operation, the transistor is cut off and very little current is flowing. There may be a significant voltage across the transistor at cut-off, but as a result of the low current, very little power is being dissipated. Therefore, at the two extremes of the operation, the transistor dissipates very little power. The area between saturation and cut-off, termed the active or transient region, has a transistor dissipating considerable power. If the transistor spends very little time in the transient region; however, it can cross over the maximum power dissipation curve for steady state power.

The reason for this is that it takes a period of time for the transistor to heat once the power dissipation has changed. If the transistor is at cut-off, as an example, and switched rapidly from cut-off to saturation through the transient region, it might spend only a very small amount of time in the transient region. The addition of the capacitor in parallel with the thermal resistance in Figure 54-3 is necessary in dealing with thermal considerations for pulses and switching. The thermal capacitance is more complicated than the simplified model, as shown in Figure 54-3; however, in this analogy we have lumped all of the thermal capacitance for the transistor and its associated encironment into one representative capacitor. The product of the thermal resistance and the thermal capacitance will result in a thermal time constant. It will require five thermal time constants after a change in power dissipation before the junction will reach its final temperature.

As an example, consider a transistor that is operating quiescently in its off condition. Applying a pulse to the transistor and driving it to saturation, the transistor will move through the transient or active region and then rest in the satur-



COLLECTOR FAMILY OF CURVES WITH A MAXIMUM POWER CURVE AND A LOAD LINE FOR MAXIMUM POWER WITH MAXIMUM POSSIBLE VOLTAGE SWING

FIGURE 53-3

CASE

P N P

JUNCTION TEMPERATURE

θ_{JC} +

CURRENT GENERATOR

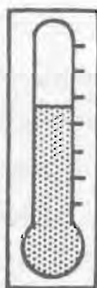
C_{JA}

θ_{JA}

P_d

θ_{CA} −

\+ − AMBIENT TEMPERATURE (T_A)



AMBIENT TEMPERATURE (T_A)

(NOTE: C_{JA} is in watt-seconds /°C)

THERMAL		ELECTRICAL EQUIVALENT
AMBIENT TEMPERATURE	=	VOLTAGE
THERMAL RESISTANCE	=	RESISTANCE
POWER DISSIPATION	=	CURRENT
JUNCTION TEMPERATURE	=	VOLTAGE

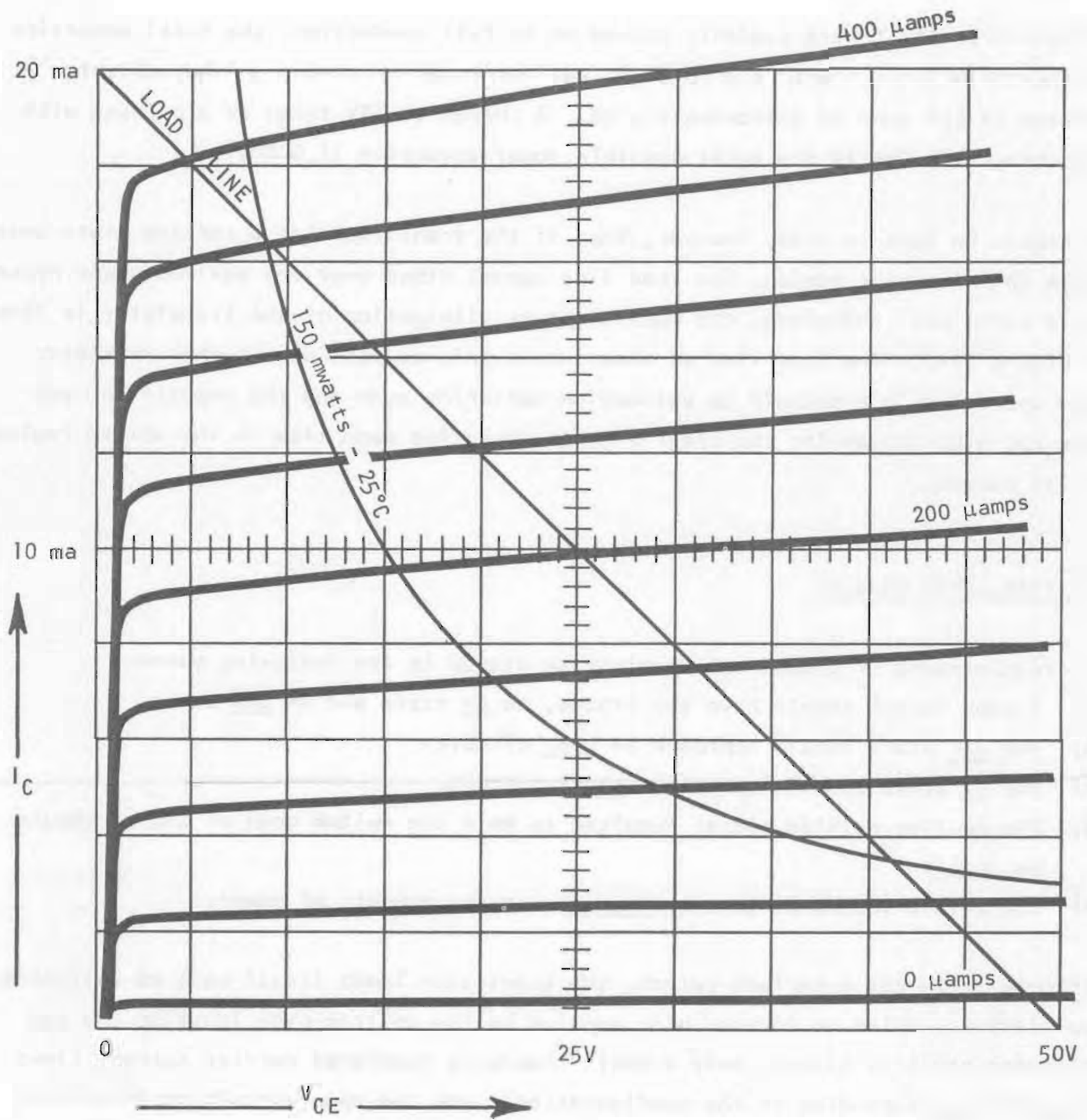
THERMAL TO ELECTRICAL ANALOGY

FIGURE 54-3

ation condition. If the transistor spends less than five thermal time constants in the active region, it will never reach the temperature to which the peak power would indicate that it would reach. In other words, the product of peak power and thermal resistance will give a temperature to which the junction will start to change. It will take five thermal time constants to change to this temperature and if the transistor does not remain in the active region for five thermal time constants, the junction will never reach this temperature. Therefore, if the transistor is operated between its extremes and spends very little time in the active region where it will be dissipating maximum power, it is possible to have the load line cross over the maximum steady state power dissipation curve. This is assuming that the repetition rate (the number of times the transistor enters the active region per second) is not high enough that the transistor spends too much time in the area of maximum power dissipation. The repetition rate will be limited for a transistor acting as a switch with its load line crossing over the maximum power dissipation curve.

If the transistor does not spend too much time in the active region, however, its excursions can enter the active region beyond the maximum steady state power dissipation curve on its collector family of curves. This is illustrated by the load line as constructed in Figure 55-3. The load line in Figure 55-3 is constructed between the maximum voltage point and the maximum current point for the particular transistor represented. The hyperbolic curve constructed on the collector family of curves in Figure 55-3 limits the transistor's maximum steady state power dissipation at 25°C to 150mw. The load line shown intersects the center of the collector family of curves and at that particular point, there is 10ma flowing and 25V across the transistor for a collector power dissipation of 250mw. The transistor during its excursion through the active region can dissipate as much as 250mw. If the transistor does not spend sufficient time in the active region to allow the junction to heat to the temperature that would result if it were dissipating 250mw continuously, the transistor can safely make an excursion through the active region without harm.

Assuming a total possible current swing in Figure 55-3 of 0 to 20ma, and a total possible voltage swing of 0 to 50V, we might estimate the total possible power swing as being the product of these two excursions or 1,000mw. In other words, suppose the transistor was sitting in its cut-off state with approximately 50V across the



at 25°C $\begin{cases} B_{VCEO} = 50V \text{ (maximum voltage)} \\ I_C(max) = 20ma \text{ (maximum forward current)} \end{cases}$

TOTAL POSSIBLE POWER SWING = 1,000 mwatts

$I_C(max) \times V_{CE}(max) = 20ma \times 50V = 1,000$ mwatts

FIGURE 55-3

transistor. If it were suddenly turned on to full conduction, the total excursion in the collector current is 0 to 20ma, and the total excursion in the collector voltage is 50V down to approximately 0V. A change of 50V taken as a product with the change of 20ma is the total possible power excursion (1,000mw).

It should be kept in mind, however, that if the transistor has a resting state somewhere in the active region, its load line cannot cross over the maximum power hyperbolic curve and, therefore, the maximum power dissipation of the transistor is 150mw in Figure 55-3. The load line as shown would only be possible if the transistor were operating in a cut-off to saturation switching mode and the repetition rate were not high enough for the transistor to spend too much time in the active region of its curves.

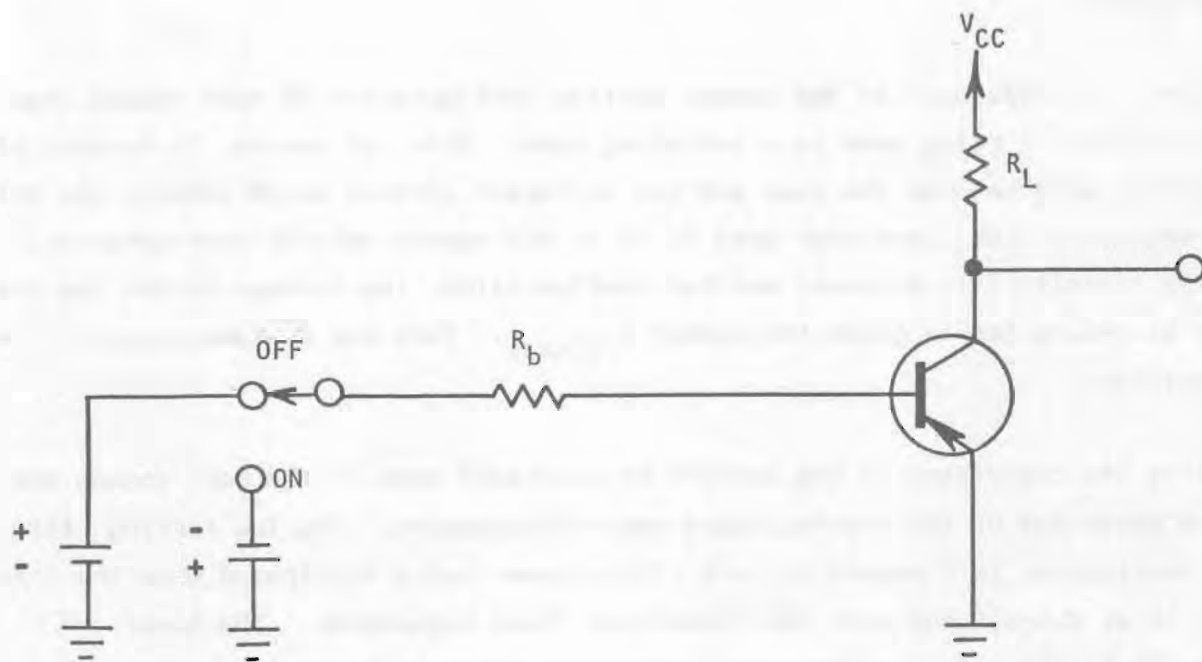
THE TRANSISTOR SWITCH:

The requirements of a good switch might be stated in the following manner:

(a) A good switch should have two states, an on state and an off state.
(b) The off state should approach an open circuit.
(c) The on state should approach a short circuit.
(d) The applied driving signal required to hold the switch open or closed should be small.
(e) The switch should be able to handle adequate amounts of power.

Although it is not a perfect switch, the transistor lends itself well to switching applications. With no forward bias applied to the emitter-base junction and the collector properly biased, only a small thermally generated carrier current flows (I_{CBO} or I_{CEO} depending on the configuration), and the open circuit resistance of the transistor will be on the order of 100KΩ to serveral megohms. The transistor can also have a reverse bias applied to the emitter-base junction to reduce the off-state current.

The transistor can be turned on and driven to saturation if sufficient base current is provided. Its collector to emitter resistance can be reduced in the range of 0.05Ω to 50Ω when the transistor is turned on to saturation. This can result in the voltage across the transistor being reduced to perhaps one-tenth of a volt in



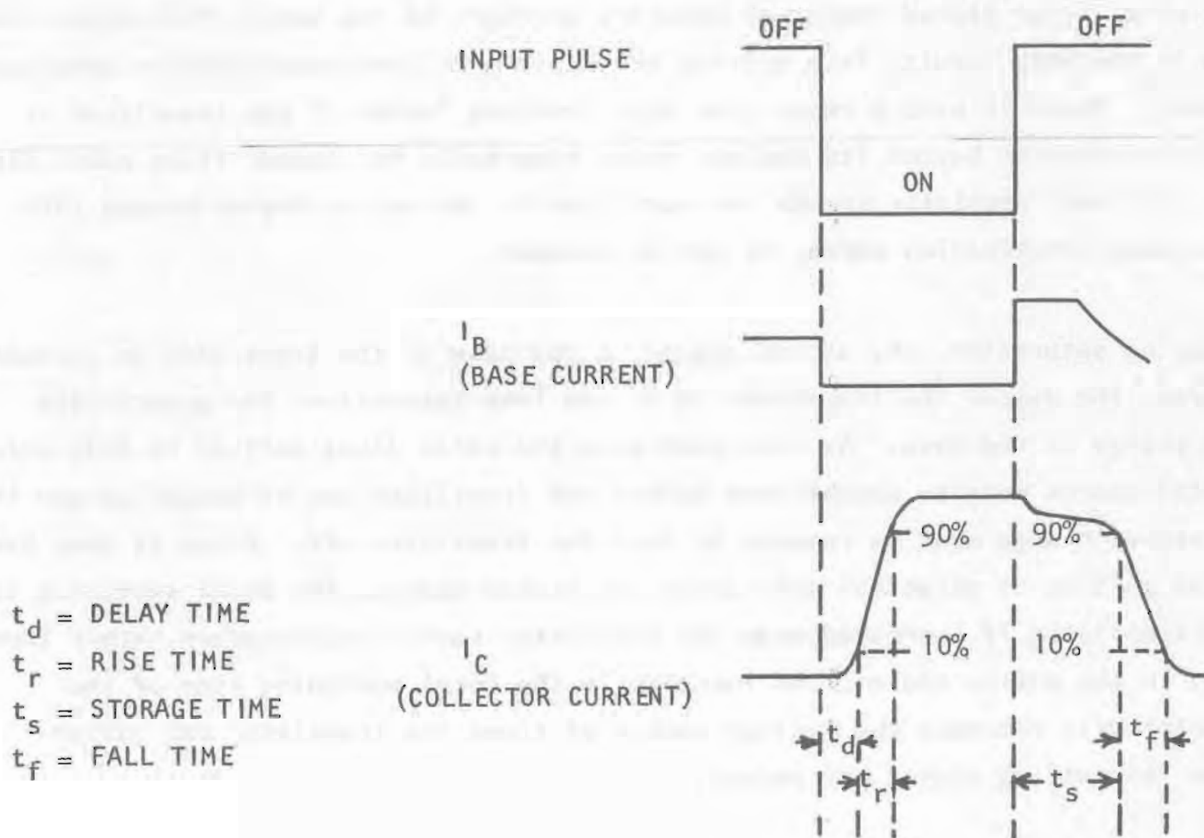


FIGURE 56-3

the on state.

Operating the transistor in the common emitter configuration is most common when the transistor is being used in a switching mode. This, of course, is because of the current gain between the base and the collector circuit which reduces the drive requirements for the transistor when it is in the common emitter configuration. With the transistor in a common emitter configuration, the voltage across the transistor at saturation is given the symbol $V_{CE(sat)}$. This was discussed earlier in this volume.

Operating the transistor in the cut-off to saturated mode is the most common and has the advantage of low resting state power dissipation. The low resting state power dissipation is a result of very little power being dissipated when the transistor is at cut-off and when the transistor is at saturation. The power, of course, is dissipated when the transistor moves through the active region. This switching mode has the disadvantage, however, of limiting the repetition rate as a result of a larger stored charge of minority carriers in the base. The larger stored charge in the base results from driving the transistor into saturation as previously discussed. There is also a repetition rate limiting factor if the transistor is making an excursion beyond its maximum power hyperbolic for steady state power dissipation. If the transistor spends too much time in the active region beyond this maximum power dissipation curve, it can be damaged.

On entering saturation, the stored charge in the base or the transistor is increased. Of course, the deeper the transistor is driven into saturation, the greater the stored charge in the base. As discussed with the basic diode earlier in this volume, the total charge must be established before the transistor can be turned on and the total stored charge must be removed to turn the transistor off. Since it does take a period of time to establish and remove the stored charge, the total switching time of the transistor is increased when the transistor rests in saturation rather than resting in the active region. An increase in the total switching time of the transistor will decrease the maximum number of times the transistor can switch between its resting states per second.

Figure 56-3 illustrates a simple circuit for measuring the switching characteristics of a transistor. The collector might be monitored with an oscilloscope. Care would have to be taken to insure that the oscilloscope could measure the times in-

volved in the measurement circuit.

When the transistor is turned on, it will take a period of time for the stored charge to be established and for the recombination and diffusion to reach an equilibrium rate. In other words, it will take a period of time for the injected carriers to move across the base to the collector and it will take a period of time for the carriers that find recombination traps to reach an equilibrium recombination rate. Therefore, when a change is applied at the emitter-base junction, there is some time before the change occurs in the collector. In Figure 56-3, the time between the application of the input change and the point that the collector current is changed 10% of its total excursion is termed the delay time of the transistor.

The time between the points on the leading edge of the collector current waveform where the current has changed 10% and where the current has changed 90% of its total excursion is termed the risetime. Risetime is measured between the 10 and 90% on the collector current waveform. The risetime is determined by the charge control time constant in the base which is governed by the minority carrier lifetime. Table 8 covers this in a deeper manner dealing with a physical model. Table 8 is not necessary for this coverage; however, it is included for those who want to pursue this a bit further with a physical model. The sum of delay time and risetime is the transistor's turn-on time.

Referring to Figure 56-3 again, the time between the removal of the input pulse and the point that the collector current is reduced to 90% of its maximum value is termed the storage time. Storage time is governed by the period of time it takes to remove the stored charge in the base of the conducting transistor. Storage time is, of course, increased when the forward bias collector injects carriers into the base when the transistor enters saturation. Storage time will be minimum when the transistor does not enter saturation and will increase as the transistor is driven deeper into saturation. Holding the transistor out of saturation will reduce the storage time. The time between the 90% and the 10% on the falling portion of the collector current waveform in Figure 56-3 is termed the fall time. The sum of storage time and fall time gives the transistor's turn-off time.

Turn-off time = storage time + fall time, and turn-on time = delay time + risetime. The total transistor switching time is the sum of all of the switching times as

measured in Figure 56-3.

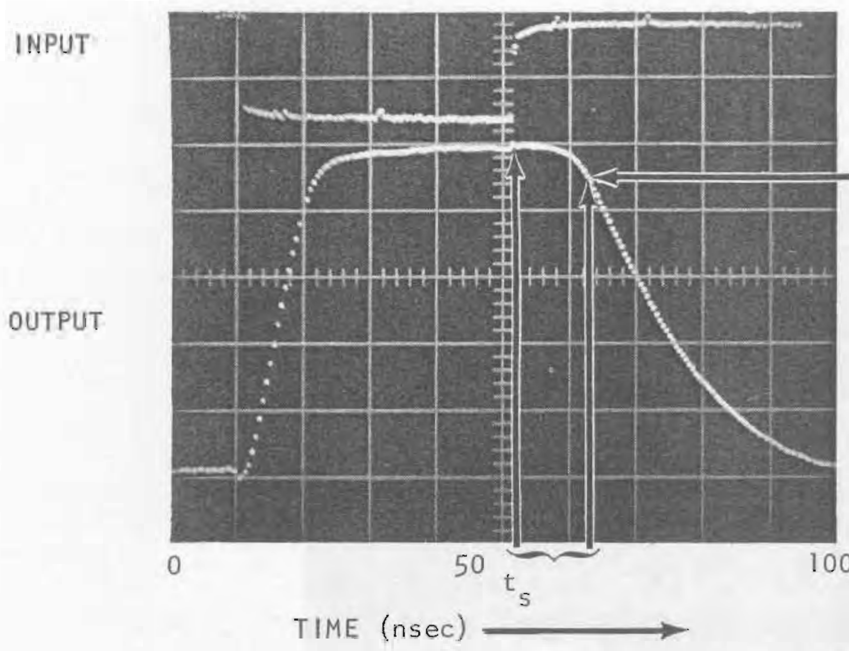
The manual switch, as used in Figure 56-3, is not fast enough to give accurate readings of switching times. In practice, an electronic switch is used to drive the transistor when switching time measurements are to be made. Figure 57-3 shows a measurement set up to measure fast transistor switching times. The Type 290 Switching Time Tester provides a circuit, bias voltages, input and output jacks along with facilities to drive the input of the transistor. The output is monitored with a sampling oscilloscope. In the case of Figure 57-3, a Tektronix Type 661 Sampling Oscilloscope is used. This oscilloscope is a dual trace unit and, therefore, both the input driving signal and the transistor's output may be monitored.

The input pulse and the transistor's output are monitored and measurements taken of switching times exactly as outlined in Figure 56-3. However, faster electronic switching is accomplished by driving the transistor with a Type 109 Fast Rise Pulse Generator. These measurements are similar to the diode switching time measurements that were made previously in this volume. Therefore, it should be remembered that risetime considerations of the measurement system must be taken into account when making measurements on the fast transistor.

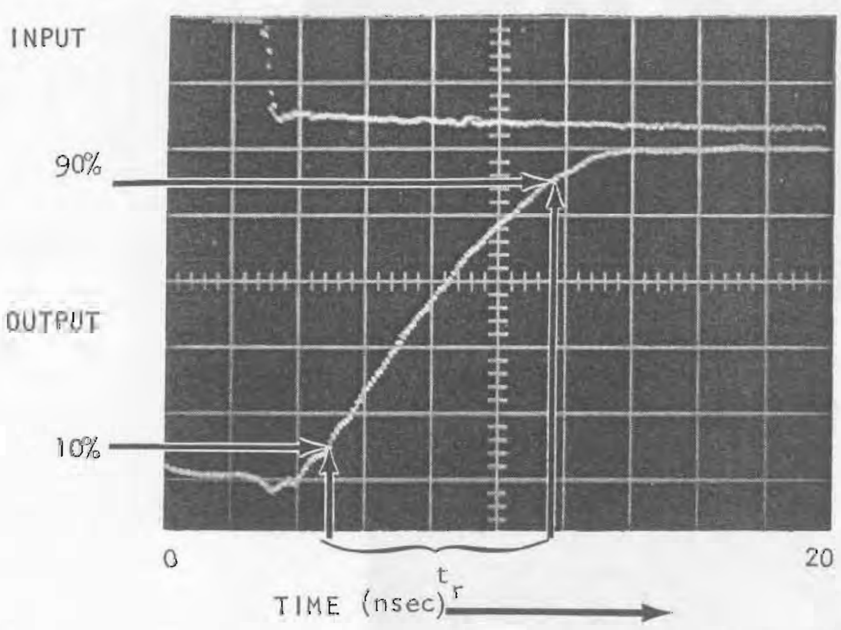
The risetime, as measured in Figure 57-3, is approximately 7nsec, and the storage time measured is approximately 11nsec. The measurement system in Figure 57-3 can be used to make measurements down to fractional nanoseconds switching times.

The Type 290 Transistor Switching Time Tester, as shown in Figure 57-3, can be somewhat limited in that the transistor must be placed in the configuration as built into the Type 290. More flexibility can be accomplished by using the Type 292 Semiconductor Tester as shown in Figure 58-3. The Type 292 Semiconductor Tester allows the circuits for the transistor to be selected and built on the circuit boards provided while maintaining the transistor in an optimum measurement environment. This does not limit the measurement to a specified circuit, but allows the measurement circuit to be fabricated on special test fixtures provided. The Type 292 offers the added flexibility that diode measurements, etc., can be accomplished along with transistor switching time measurements.

The measurement approaches that we have been discussing are fine if only a limited

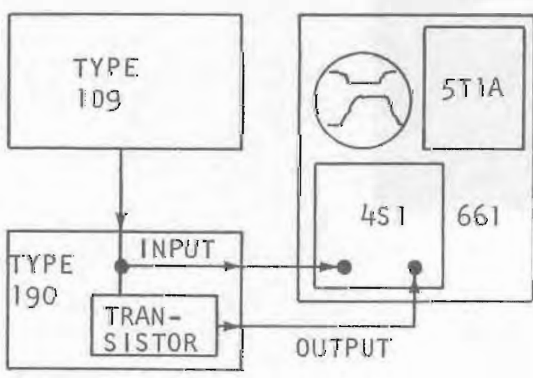


TOP PHOTOS TAKEN WITH A TYPE C-12 OSCILLOSCOPE CAMERA



TEST SET-UP

TEKTRONIX TYPE 661 SAMPLING OSCILLOSCOPE, 4S1 DUAL TRACE SAMPLING VERTICAL, 5T1A TIMING UNIT, TYPE 109 PULSE GENERATOR, AND TYPE 290 TRANSISTOR SWITCHING TIME TESTER



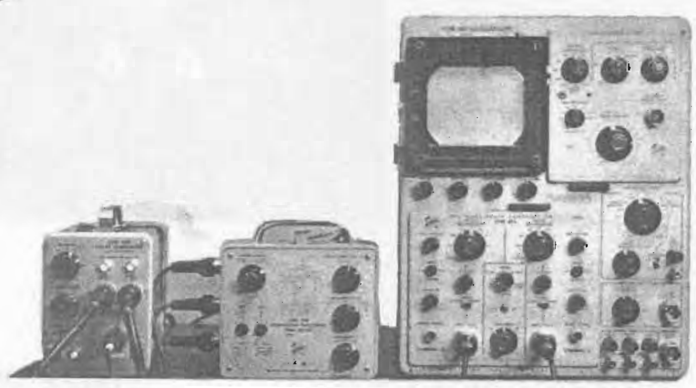
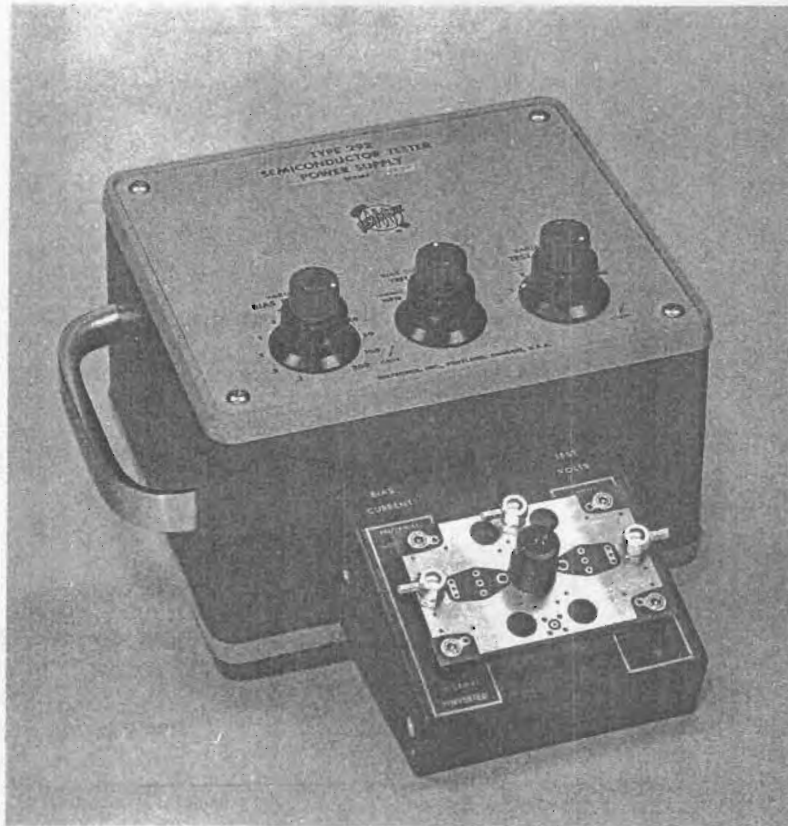


FIGURE 57-3



TEKTRONIX TYPE 292 SEMICONDUCTOR TESTER
AND ACCESSORIES

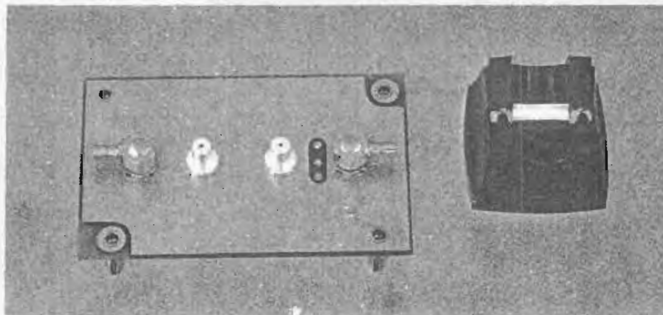


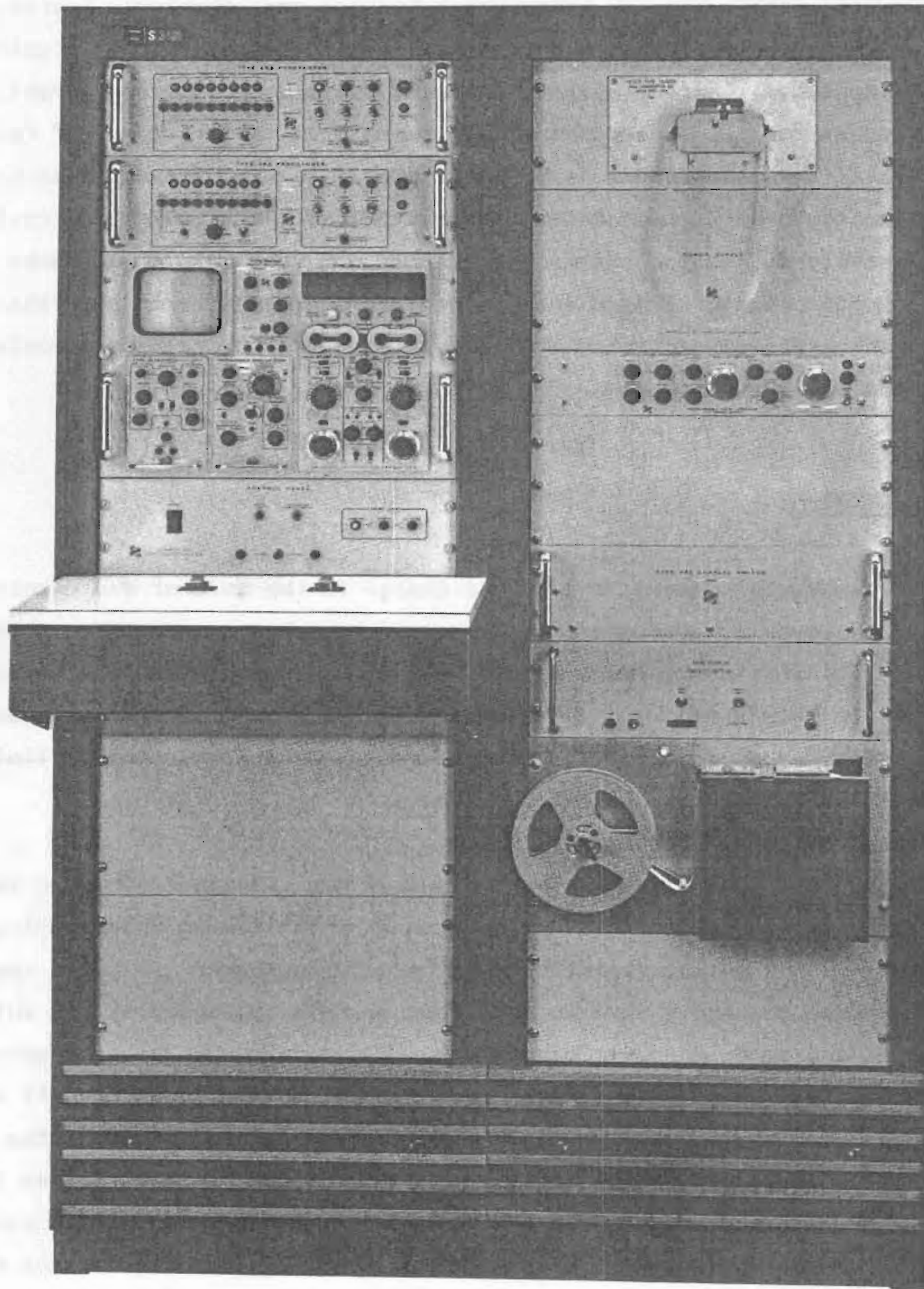
FIGURE 58-3

number of measurements in a period of time are anticipated. For transistor manufacturers, however, the system of measurement becomes more complex. A great number of measurements must be made in a short period of time if the system is going to offer any advantages. Figure 59-3 illustrates a computer type, programmable system that can be used for rapid sequencial measurements of switching times of fast transistors. This is the Tektronix Type S-3401 system. This system can also be used for measurements on other semiconductor devices to include integrated circuits. This is primarily designed for the manufacturer who must make a great number of checks in a short period of time and is used as an example here to show the complexity of the measurement system that can come about for checking semiconductor switching times.

SPEED-UP CAPACITOR:

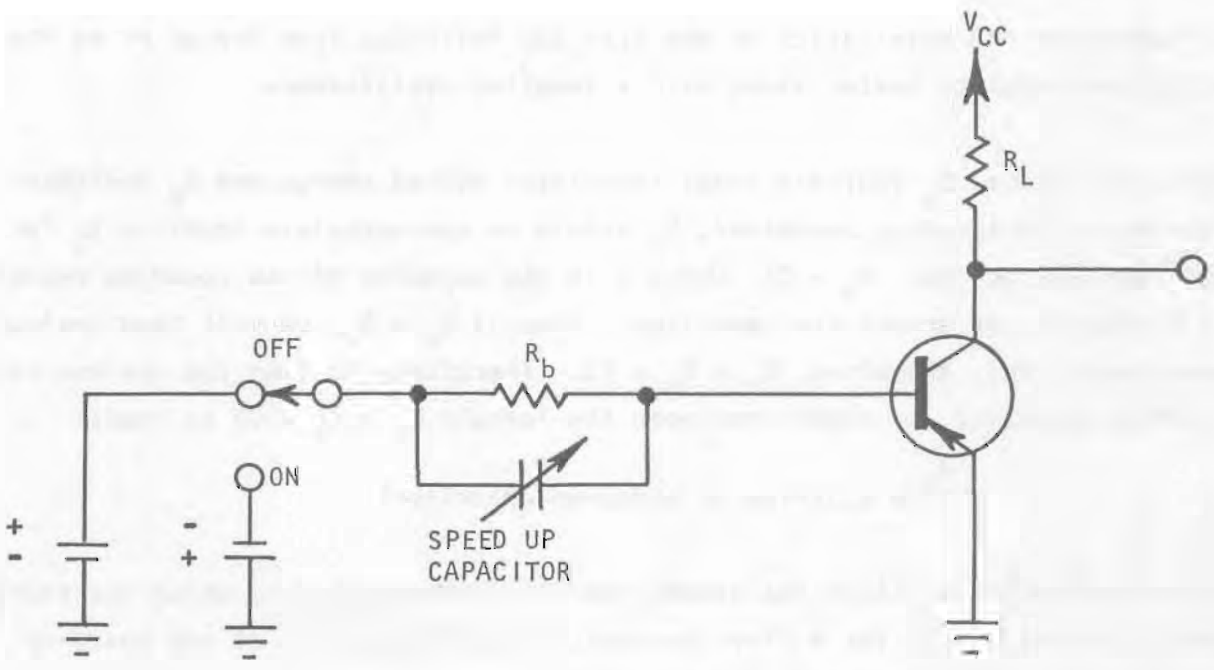
As we have previously discussed, the stored charge in the base of the transistor must be established to turn on the transistor and must be removed to turn the transistor off. Sufficient current must be supplied to establish and remove the storage charge if this is to be accomplished. Time constant of the resistance of the driving source and the effective capacitance in the input loop will limit the establishment and removal of the stored charge.

There is normally resistance in the input stage of the transistor for d-c considerations and this will increase the time involved in establishing and removing the stored charge. If the output resistance of the driving source is low, a capacitor can be added across the resistance in the input circuit and speed up the switching time. Since it speeds up the switching time, it is referred to as a speed-up capacitor. Actually, the speed-up capacitor is compensating the input circuit so that the time constant of the input circuit is equal to the time constant of the charge control circuit in the base of the transistor. The speed-up capacitor and its effect on the switching time is shown in Figure 60-3. When the speed-up capacitor, as shown in Figure 60-3, is selected for optimum speed-up action (when the time constant of R_B and the speed-up capacitor is equal to the time constant of the charge control or effective RC time constant in the base), the charge on the speed-up capacitor speeds up the establishment and removal of the stored charge in the transistor base. We might adjust the speed-up capacitor while observing the tran-



TEKTRONIX TYPE S-3401 SYSTEM

FIGURE 59-3



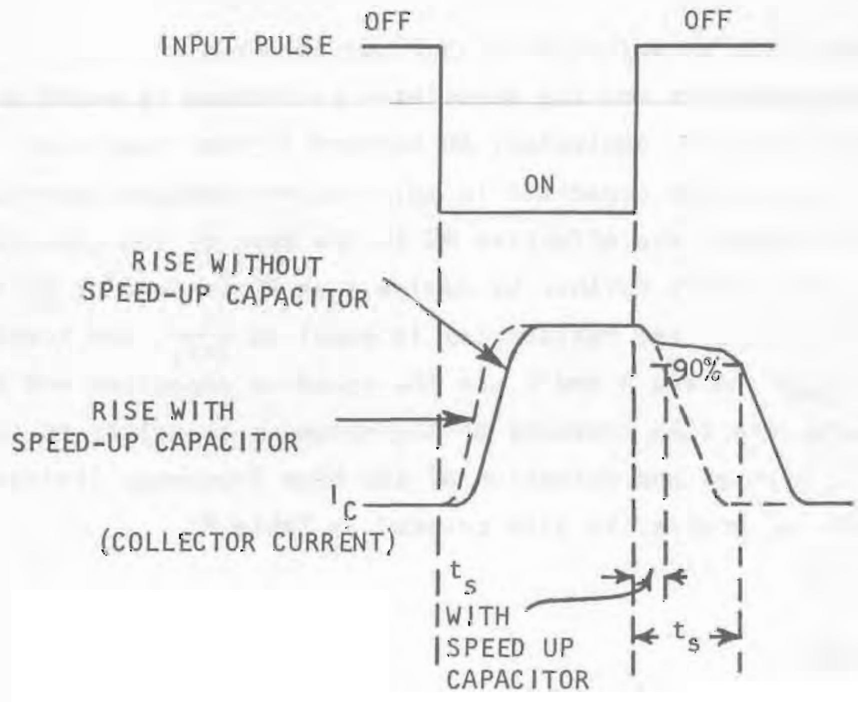


FIGURE 60-3

sistor switching characteristics on the Type 290 Switching Time Tester or on the Type 292 Semiconductor Tester along with a sampling oscilloscope.

If we let the symbol Q_s indicate total transistor stored charge and Q_c indicate the charge on the speed-up capacitor, Q_c should be approximately equal to Q_s for optimum speed-up action. $Q_c = CE$, where C is the capacity of the speed-up capacitor and E is the voltage across the capacitor. Then if $Q_c = Q_s$, we will have optimum speed-up action and, therefore, $Q_c = Q_s = CE$. Therefore, to find the optimum value of speed-up capacitor, we might transpose the formula $Q_c = Q_s = CE$ to read:

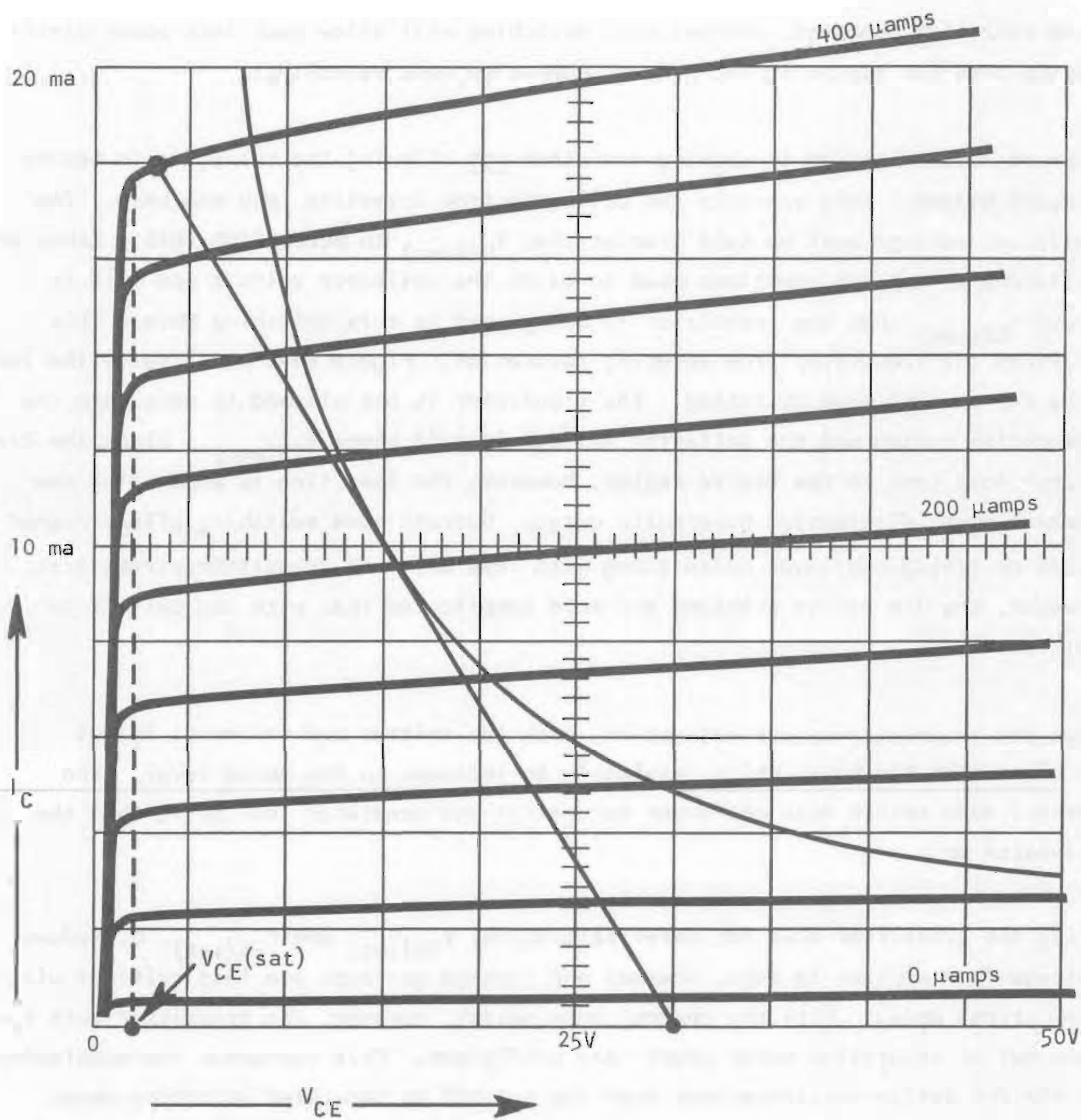
$$\frac{Q_s}{E} = C \text{ (value of speed-up capacitor)}$$

If the manufacturer specifies the stored charge or some method by which the stored charge can be calculated for a given current, the optimum value of the speed-up capacitor can be calculated from the foregoing formula. It is evident that the charge on the speed-up capacitor should equal the transistor stored charge for optimum speed-up action.

When the speed-up capacitor is adjusted for optimum speed-up action, the time constant of the speed-up capacitor and its associated resistance is equal to the time constant of the charge control, equivalent RC network in the transistor. We might then say that when the speed-up capacitor is adjusted for optimum speed-up action, that $RC = \frac{1}{\omega t}$ (since RC equals the effective RC in the base of the transistor). We might investigate this a bit further by saying that RC (referring to the speed-up capacitance and its associated resistance) is equal to $\frac{1}{2\pi f_t}$, and transposing we might say that $f_t = \frac{1}{2\pi RC}$, where R and C are the speed-up capacitor and its associated resistance. Therefore, the time constant of the speed-up capacitor in its associated resistance will give an approximation of the high frequency limitations of the transistor. This, of course, is also covered in Table 8.

CURRENT MODE SWITCHING:

An obvious method of reducing the saturated mode storage time is to keep the transistor out of saturation. This type of switching mode is referred to as current mode switching, and the transistor operates between cut-off and the active region. Current mode switching will allow a faster repetition rate than cut-off to saturated



CURRENT MODE SWITCHING
CUT-OFF TO ACTIVE REGION

FIGURE 61-3

mode switching; however, current mode switching will allow much less power dissipation than low rep rate, cut-off to saturation mode switching.

Current mode switching is nothing more than not allowing the collector to become forward biased. This prevents the collector from injecting into the base. The collector voltage must be held greater than $V_{CE(sat)}$ to accomplish this. Zener or avalanche diodes are sometimes used to catch the collector voltage and hold it above $V_{CE(sat)}$ when the transistor is being used in this switching mode. This prevents the transistor from entering saturation. Figure 61-3 illustrates the load line for current mode switching. The transistor is not allowed to move into the saturation region and the collector voltage is held above $V_{CE(sat)}$. Since the transistor must rest in the active region, however, the load line is kept below the maximum power dissipation hyperbolic curve. Current mode switching offers higher speed switching and lower noise along with less critical transistor parameters; however, the d-c design problems are more complicated than with the cut-off to saturated mode.

When the transistor enters saturation, both the emitter and collector inject carriers into the base, which results in an increase in the noise level. The current mode switch does not enter saturation and generates less noise than the saturated mode.

Since the transistor does not enter saturation, $V_{CE(sat)}$ and $h_{FE(sat)}$, breakdown voltage from emitter to base, thermal and leakage currents are less critical with the current mode. With the current mode switch, however, the transistor must remain out of saturation under worst case conditions. This increases the complexity of the d-c design considerations over the cut-off to saturated switching mode. Also, since the current mode switch has a resting state in the active or transient region, care must be taken that the maximum power of the transistor is not exceeded.

We might summarize this by saying the advantages of the current mode switch over the cut-off to saturated mode switch are faster switching speed, less critical transistor parameters, and the disadvantages are higher resting state power dissipation and the more complicated d-c design considerations than the off to saturated mode.

AVALANCHE SWITCHING MODE:

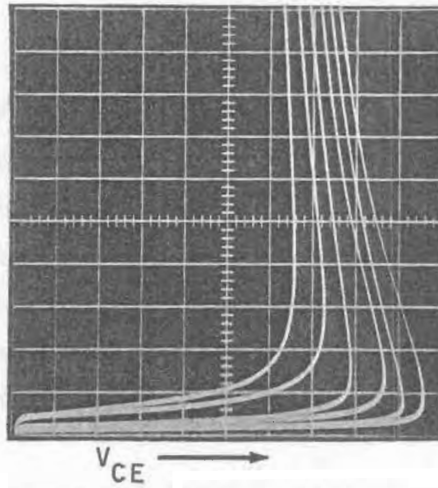
The avalanche switching mode has the transistor switching between cut-off and a point in the avalanche breakdown region. The switching time can be on the order of fractional nanoseconds with this switching mode. One resting state of the avalanche switching transistor is just out of the avalanche breakdown condition, and the other resting state is in the avalanche breakdown region. This is illustrated in Figure 62-3. The application of external energy to cause the transistor to switch states, is termed triggering and the applied energy to cause switching is termed a trigger.

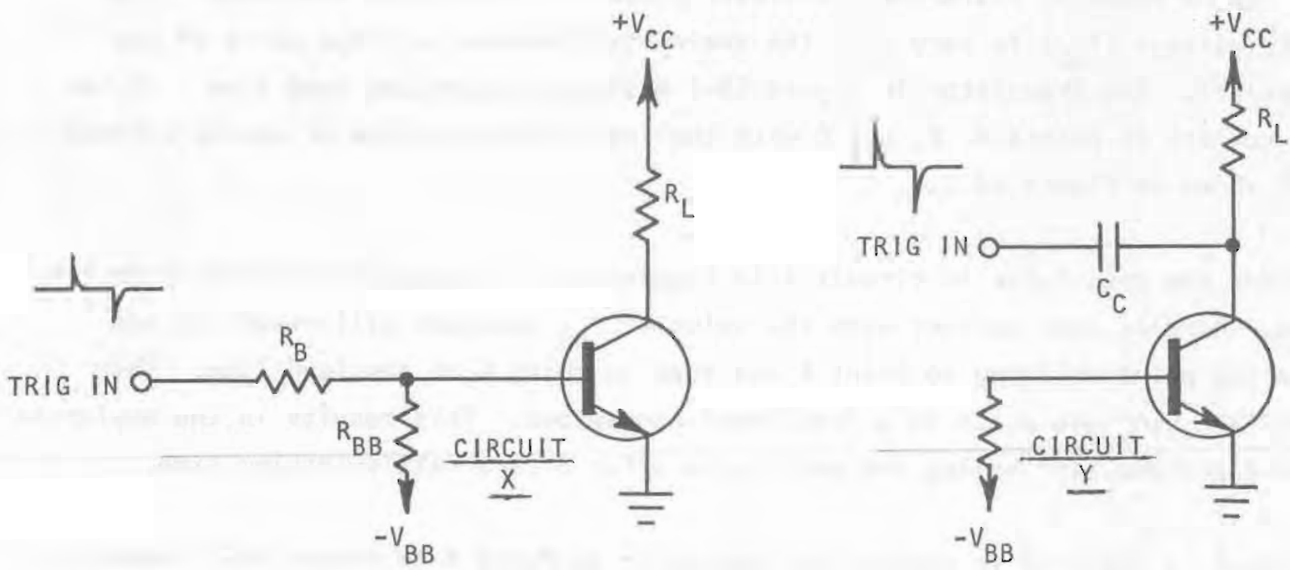
Figure 62-3 shows an avalanche transistor biased for avalanche switching. The supply voltage (V_{CC}) is very near the avalanche breakdown voltage point of the transistor. The transistor in Figure 62-3 must operate on the load line. It can only operate at points A, B, and C with the load line and value of supply voltage (V_{CC}) shown in Figure 62-3.

Consider the transistor in circuit X in Figure 62-3 is operating at Point A on the curve. Forcing more current with the value of V_{CC} constant will result in the operating point shifting to Point B and then to Point C on the load line. This transition can take place in a fractional nanosecond. This results in the avalanche switching transistor having the ability to offer a very fast switching time.

A trigger is required to return the transistor to Point A in Figure 62-3, once it has been switched to Point C. The avalanche switching mode shown in Figure 62-3 is bi-stable, meaning that the transistor has two stable states.

Circuit Y in Figure 62-3 has a trigger applied to the collector. The trigger will increase or decrease the collector voltage, depending on its polarity. A positive going trigger will increase the applied collector voltage and the load line will essentially move to Point B. At that instant, the only point on the load line that the transistor can exist is at Point E. Once triggered to Point E and the trigger removed, the transistor in circuit Y will rest at Point C, when the trigger is no longer present. For the transistor to move back to Point A on the load line from Point C, the collector must be triggered in the opposite polarity. We can say that circuit Y in Figure 62-3 is bi-stable and switches between Points A and C.





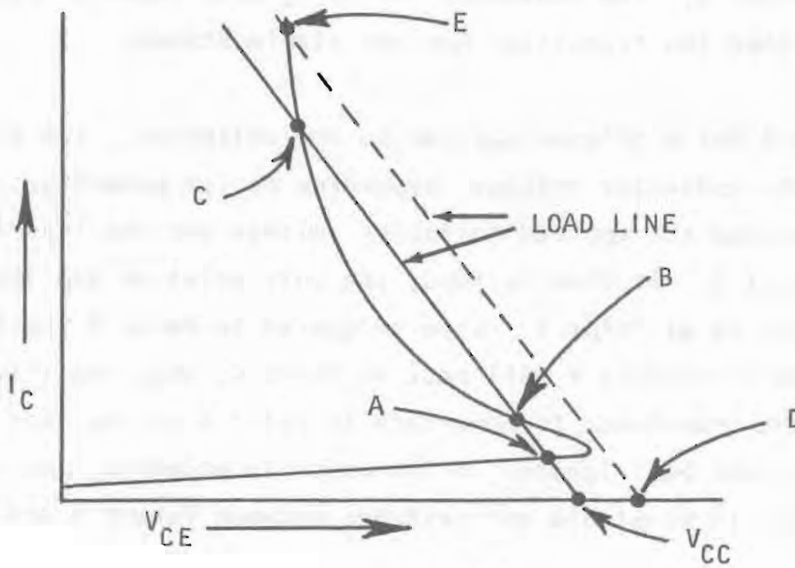


FIGURE 62-3

Figure 63-3 shows an avalanche transistor in a typical application. The avalanche transistor is being used as a fast rising pulse generator.

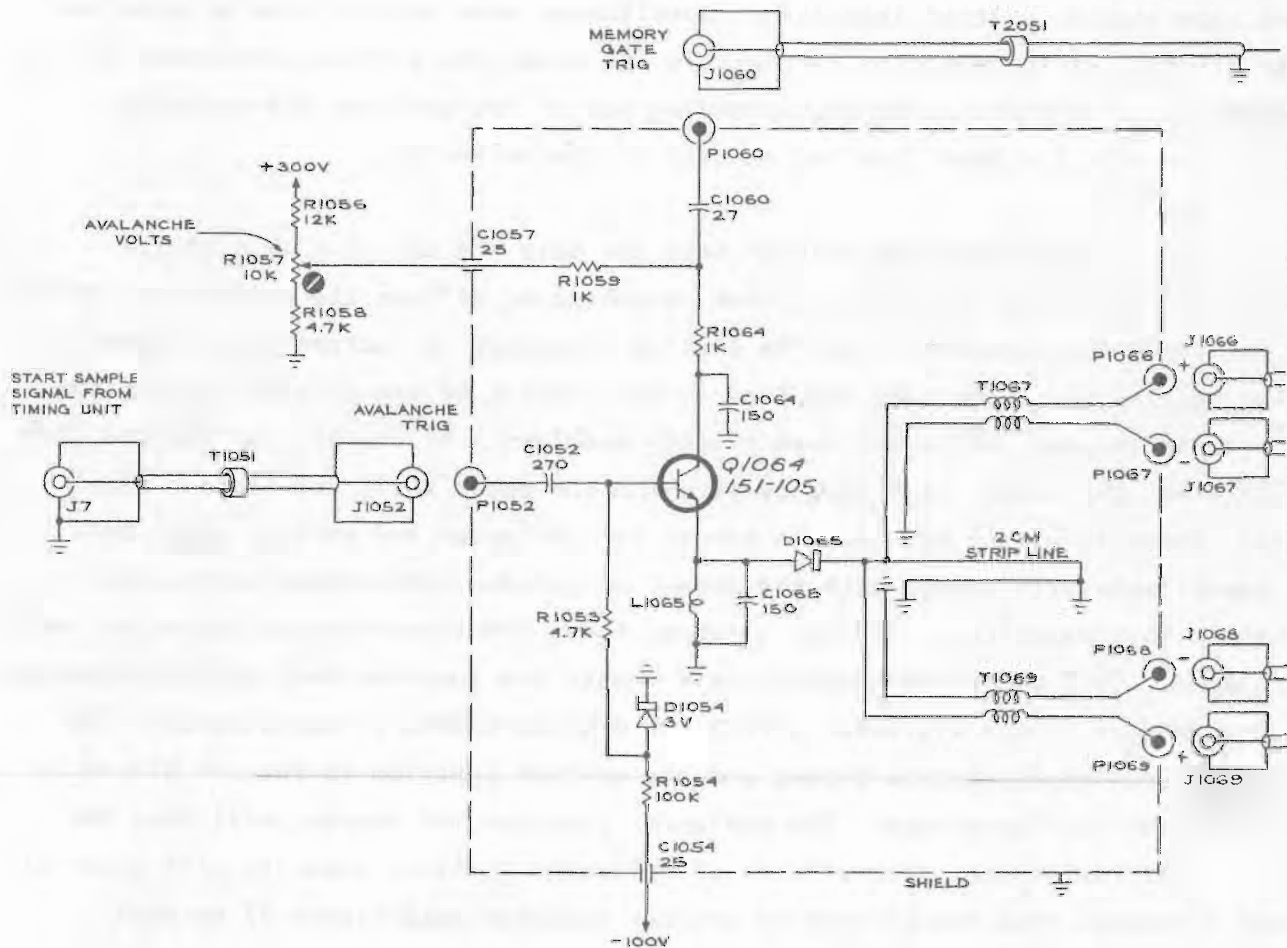


FIGURE 63-3

The avalanche volts adjustment (R1057) allows the quiescent operating point of the avalanche transistor to be adjusted just below the avalanche breakdown voltage. An incoming pulse applied to the base of the transistor through C1052 will cause the transistor to switch to a point in the avalanche breakdown region. The pulse is then shaped by a shorted transmission line, referred to on the schematic as a 2cm strip line. Note the use of the zener diode (D1054) as a voltage reference device. D1065 is being used as a coupling diode. We might anticipate that D1065 must be a rather fast switching diode in order to accomplish the transfer of energy that it must accomplish in this configuration.

HIGH FREQUENCY TRANSISTORS:

Among the high frequency limiting factors in the transistor are the collector transition capacitance, emitter transition capacitance, base transit time as governed by the lifetime of the minority carriers in the base, the diffusion constant of the material used in the base, and the spreading out of the carriers during their diffusion through the base from the emitter to the collector.

The carriers injected from the emitter into the base are existing as minority carriers and diffusing until they either recombine or diffuse the entire base thickness and reach the collector. As the applied frequency is increased, a frequency will be reached where the time involved in the transit of the carriers across the base results in phase shift and loss of gain problems. Of course, the greater the transit time, the lower the frequency at which the phase shift and loss of gain occurs. There is also a capacitance across the collector and emitter junctions. This capacitance will change with the amount of current and voltage across the junction. This capacitance is also referred to as the transition or depletion layer capacitance. This capacitance effectively shunts the junction and, as the frequency is increased, will have a greater effect for a given amount of capacitance. The collector junction is reverse biased and the emitter junction is forward biased in most amplifier configurations. The collector junction, of course, will have the higher signal resistance. The effects of collector junction capacity will occur at a lower frequency than the effects of emitter junction capacitance if we deal strictly with the RC time constant involved.

Since the carriers in their transit from the emitter to the collector are transported by diffusion, they take an erratic path in the base. Figure 64-3 shows a typical path that a carrier might take in its movement across the base.

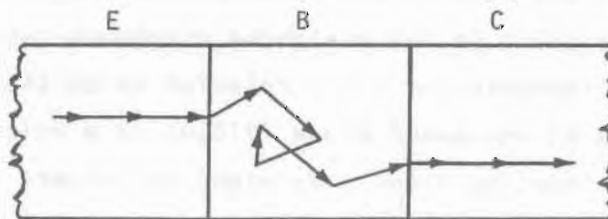
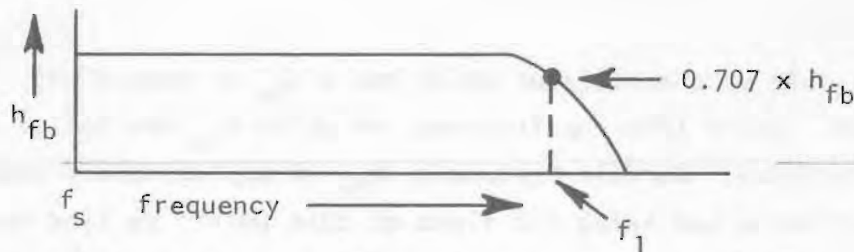


FIGURE 64-3

Considering the diagram in Figure 64-3, several carriers entering the base at the same time could very well arrive at the collector at different times. The carriers might all take different paths in their transit through the base. The wider the active region of the base, the more opportunity the carriers will have to spread out and arrive at the collector at different times.

Manufacturers specify cut-off frequencies for transistors in different configurations. The symbol $f_{h_{fb}}$ or f_{α_b} is the symbol given the frequency at which low frequency h_{fb} has decreased to 0.707 of its low frequency value. This decrease in the h_{fb} or α of the transistor with increasinq frequency is the result of factors alreach discussed. The frequency $f_{h_{fb}}$ or f_{α_b} is primarily for use with the common base configuration. This is referred to as the cut-off frequency of the transistor in a common base configuration. Figure 65-3 shows a plot of applied frequency versus h_{fb}.



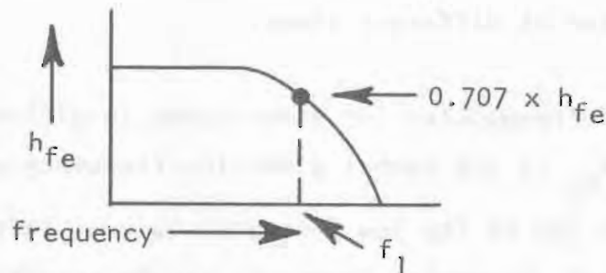
(NOTE: f_s = standard frequency, usually 1kc)

FIGURE 65-3

f_s indicates the standard frequency at which the low frequency parameters were measured, generally 1kc. As the frequency is increased, a point is reached at which there is a significant decrease in h_{fb} with an increase in frequency. f_1 indicated in Figure 65-3 is the frequency at which h_{fb} has been reduced to 0.707 of its low frequency value. This is the frequency that is given the symbol $f_{h_{fb}}$ or f_{α_b}.

$f_{h_{fe}}$ or f_{α_e} is the symbol given the cut-off frequency of the transistor in a common emitter configuration. This indicates the frequency at which h_{fe} has diminished to 0.707 of its low frequency value. For a given transistor, this frequency will be a much lower frequency than $f_{h_{fb}}$. The diagram in Figure 66-3 shows a plot of frequency

versus h_{fe}. Once again, the starting frequency is the standard frequency, usually 1kc, and this is the frequency at which the low frequency parameters were measured. f_1 is the frequency at which h_{fe} has diminished to 0.707 of its low frequency value and this is considered the cut-off frequency for the common emitter configuration.



(NOTE: f_s = frequency standard, usually 1kc)
FIGURE 66-3

The relationship between h_{fe} and h_{fb} is shown by the formula:

$$h_{fb} = \frac{h_{fe}}{1 + h_{fe}}$$

Let's take an example of a transistor which has a h_{fe} or beta of 49. This means that h_{fb} is -0.98. Let's take the frequency at which h_{fe} has fallen to 0.707 of its low frequency value. At this frequency, h_{fe} is approximately equal to 34.6. If we apply the formula and solve for alpha at this point, we find that alpha or $h_{fb} = \frac{h_{fe}}{1 + h_{fe}} = \frac{34.6}{1 + 34.6}$ or approximately -.097. In other words, when beta has diminished to 0.707 of its low frequency value, alpha has only diminished a very little bit.

Now, let's take the example of the frequency at which h_{fb} has diminished to 0.707 of its low frequency value. In transposing the formula for h_{fb}, we will find that:

$$h_{fe} = \frac{h_{fb}}{1 + h_{fb}}$$ (Note: h_{fb} is <u>always a negative</u> quantity.)

When alpha or h_{fb} has decreased to 0.707 of its low frequency value for the transistor already specified, alpha equals: $h_{fb} = 0.707 \times -0.98 = -0.69$

Therefore, beta of this same transistor at the frequency where alpha is diminished to 0.707 of its low frequency value is equal to: $h_{fe} = \frac{h_{fb}}{1 + h_{fb}} = \frac{0.69}{1 - 0.69} = 2.2$

In other words, when alpha has diminished to 0.707 of its low frequency value, beta has diminished from 49 down to 2.2. Therefore, we can see that the cut-off frequency of the common emitter configuration is going to occur at a much lower frequency than the cut-off frequency of the common base configuration. The cut-off frequency $f_{h_{fe}}$ is related to the cut-off frequency $f_{h_{fb}}$ by the formula:

$$f_{h_{fe}} = f_{h_{fb}} (1 + h_{fb})$$

(h_{fb} is always given as a negative quantity)

In the formula, $(1 + h_{fb})$ ends up being a number somewhat less than one. Let's take the example of the transistor we have been discussing whose h_{fb} is -0.98 and h_{fe} is 49. If the transistor has a $f_{h_{fb}}$ of 10mc, h_{fe} will be reduced to 0.707 of its low frequency value at $f_{h_{fe}} = f_{h_{fb}} (1 + h_{fb}) = 10mc\ (1 + -0.98) = 200KC$. With the parameters as listed for the transistor, a cut-off frequency for the transistor in a common base configuration of 10mc will only yield a cut-off frequency in the common emitter configuration of 200kc.

We can also transpose the formula to solve for the $f_{h_{fb}}$ if we know $f_{h_{fe}}$;

$$f_{h_{fb}} = \frac{f_{h_{fe}}}{1 + h_{fb}} \quad (h_{fb} \text{ is always a negative quantity})$$

Consider a transistor with a h_{fb} of -0.96 and a $f_{h_{fe}}$ of 500KC. Solving for the $f_{h_{fb}}$ by the formula;

$$f_{h_{fb}} = \frac{f_{h_{fe}}}{1 + h_{fb}} = 500 \times \frac{10^3}{0.04} = \Omega 2.5mc$$

Therefore, the transistor whose alpha or $h_{fb} = -0.96$ and $f_{h_{fe}} = 500KC$ will have a common base cut-off frequency of 12.5mc.

f_T (GAIN BANDWIDTH PRODUCT):

The frequency at which the common emitter current gain h_{fe} is reduced to unity is sometimes given for the transistor. This is essentially the gain bandwidth product of the transistor (referring to current gain). Consider the diagram in

Figure 67-3. Note that when the low frequency beta is reduced to one or unity, the frequency at that point is given the symbol f_T.

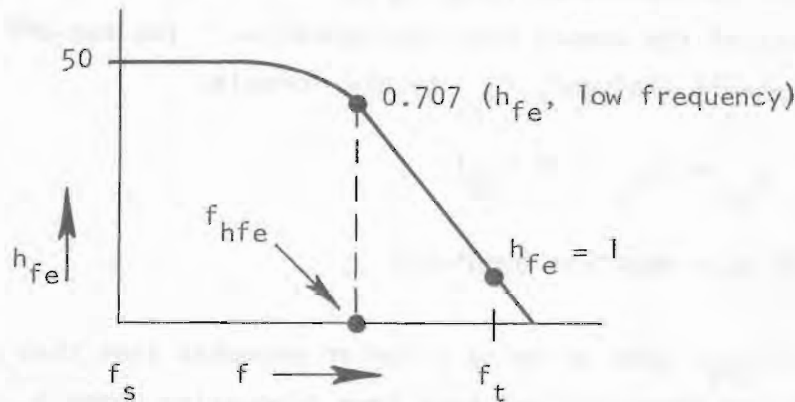


FIGURE 67-3

When f_T is known, the gain at any frequency between $f_{h_{fe}}$ and f_T can be found. Also, if the gain is known between these two frequency points, the frequency can be found.

As an example, consider a transistor whose f_T is stated as 50mc. We can say that gain x bandwidth = 50mc; therefore, 1 (gain at f_T) x 50mc (frequency, f_T) = 50mc. Now, for this same transistor, let's substitute a gain of 2 and determine the possible bandwidth. Substituting a gain of 2, the product of 2 and some number must equal 50mc because 50mc is the gain bandwidth product. Transposing this, we can determine the frequency at which the gain will be 2. 50mc divided by 2 gives a frequency of 25mc. When h_{fe} of the transistor is equal to 2, the bandwidth is 25mc.

Suppose the low frequency small signal, h_{fe}, is stated as being 70. h_{fe} at the 3db .707 point is 0.707 x 70 or approximately 50. We can determine the 3db down frequency using the gain bandwidth product or f_T. If the gain is 50 at the 3db point, 50 into 50mc gives 1mc or, in other words, $f_{h_{fe}}$ of the transistor is 1mc.

The reason for listing a frequency such as f_T is that the transistor in the common emitter configuration is definitely useful beyond the 3db down or $f_{h_{fe}}$ frequency. In regenerative circuits, the gain might not have to be much more than unity. To allow a more useful parameter, f_T can be used with the low frequency beta to determine the gain and the bandwidth if one or the other is known.

TRANSISTOR OUTPUT CAPACITANCE:

C_{ob} is the symbol given the output capacitance of the transistor in the common base orientation. C_{oe} is the symbol given the output capacitance for the transistor in the common emitter orientation. In the symbol C_{ob}, the C indicates capacitance, the o indicates output, and the b indicates that the configuration is common base. In the symbol C_{oe}, the C indicates capacitance, the o indicates output, and the e indicates the orientation is a common emitter. When the transistor is in a common base orientation, the output capacitance is essentially the capacitance of the collector junction and any strays. When the transistor is in the common emitter orientation, the output capacitance becomes a little more complex since the output is taken between the collector and the emitter, rather than the collector and base. The capacitance between the base and emitter, between the collector and emitter, and between the collector and base, all enter into the output capacity when the transistor is in the common emitter configuration. There is a relationship between the two, and that is the current gain parameter of the transistor. We can find the capacitance C_{ob} when C_{oe} is given, by the formula:

$$C_{ob} = C_{oe}\ (1 + h_{fb})$$ (Note: h_{fb} is always a negative quantity)

Consider a transistor with a $h_{fb} = -0.98$ and a $C_{oe} = 200pf$. Using the formula;

$$C_{ob} = 200pf\ (1 - 0.98) = 200pf \times 0.02 = 4pf$$

A transistor with the parameters listed and an output capacity for a common emitter configuration of approximately 200pf will have an output capacity in the common base configuration of about 4pf. If C_{ob} is given and the transistor is to take on a common emitter orientation, we might want to solve for C_{oe}. This becomes a simple matter of transposing the formula: $C_{ob} = C_{oe}\ (1 + h_{fb})$ to read: $C_{oe} = \frac{C_{ob}}{1 + h_{fb}}$.

Consider a transistor whose $h_{fb} = -0.97$ and whose C_{ob} is listed as 9pf. Substituting these values into the formula:

$$C_{oe} = \frac{C_{ob}}{1 + h_{fb}} = \frac{9pF}{0.03} = 300\ pF$$

A transistor with an alpha of -0.97 and a common base output capacitance of 9pf will exhibit an output capacitance of 300 pf when placed in a common emitter configuration.

The capacitance of the transistor will be given under a stated set of operating conditions and the capacitance will vary with the operating point. A change in the collector voltage will change the depletion layer at the collector junction and essentially change the dialectric of the capacitor. The capacitance of the emitter-base junction will also vary with the width of depletion region.

It should be evident from the preceding discussion that high frequency transistors will call for a construction with a narrow but uniform base width (referring to the active region of the base of the transistor) and a low collector and emitter junction capacitance.

The narrower the base, the less the transit time and the less time that the carriers have to spread out during their diffusion in the base. This is assuming a given minority carrier lifetime, for a given minority carrier lifetime in the base, making the base very narrow, increases the frequency at which phase shift becomes a problem. The thin base, however, should have a uniform thickness.

A transistor such as that shown in Figure 68-3 has some high frequency problems built into the transistor.

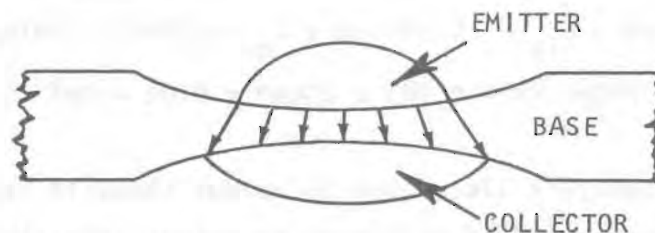
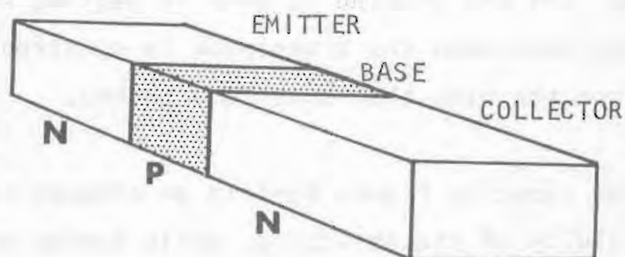


FIGURE 68-3

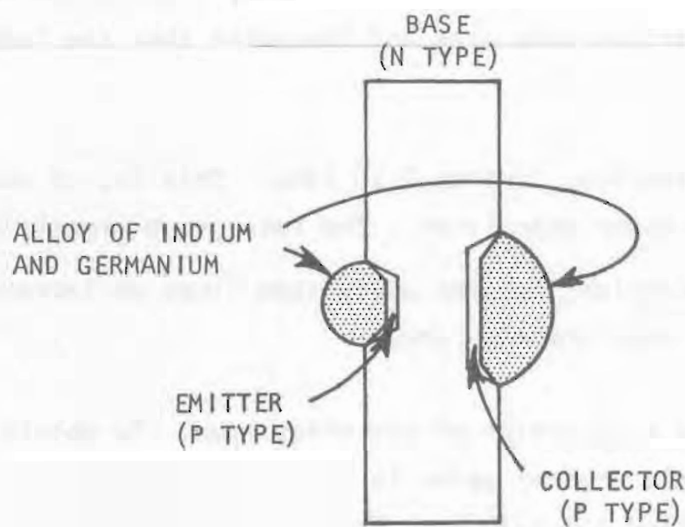
Notice that some of the carriers will have to spend more time in the base than others, due to the fact that the base is not the same thickness throughout the entire active area. This indicates that along with a very thin base, the base should have a uniform active width.

Consider the transistor types as shown in Figure 69-3.

The rate grown transistor is manufactured by pulling a seed crystal from a container of molten materials. The crystal is said to be grown in the process.



RATE GROWN



ALLOY

FIGURE 69-3

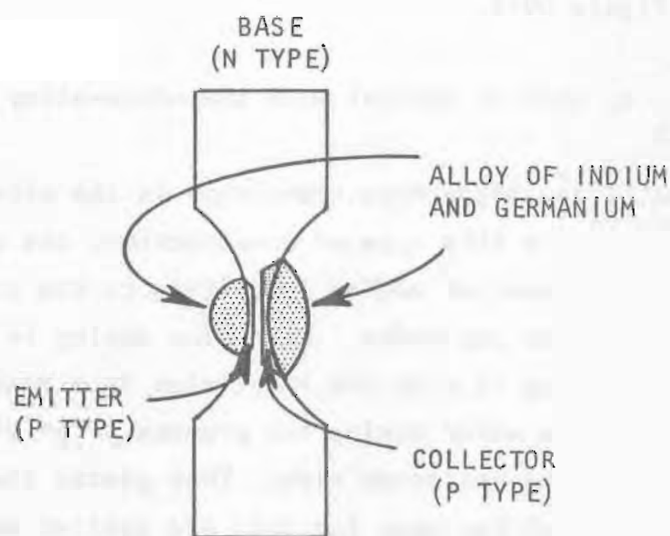
Changing the rate at which the crystal is grown is one method of forming the N and P type regions. Different amounts of dopent impurities enter the crystals when the growing rate is changed. The rate grown crystal is cut up and makes a number of transistors. The base could be made thinner for high frequency considerations in the rate grown process, but the problem is both in getting a uniform width and attaching a wire to the base when the transistor is constructed. The base should actually be thinner than the wire that must be attached.

The alloy transistor as shown in Figure 69-3 is an attempt to gain a narrow base and still maintain rigidity of the structure, while having something to which a wire can be connected to make an electrical contact to the base. The alloy transistor is constructed by alloying two dots of material to the sides of a very thin crystal wafer. The dot should give an impurity concentration opposite of the thin wafer. As shown in Figure 69-3, the dots become the emitter and collector contacts and the crystal wafer becomes the base. The alloying process makes possible a much thinner base between the emitter and collector with the ability to connect a wire to the crystal wafer for an electrical connection to the base. It is difficult to gain uniformity if the base is made too thin in the alloy transistor. Also, the bulk resistance between the active base area and the point that the lead is attached to comes into play.

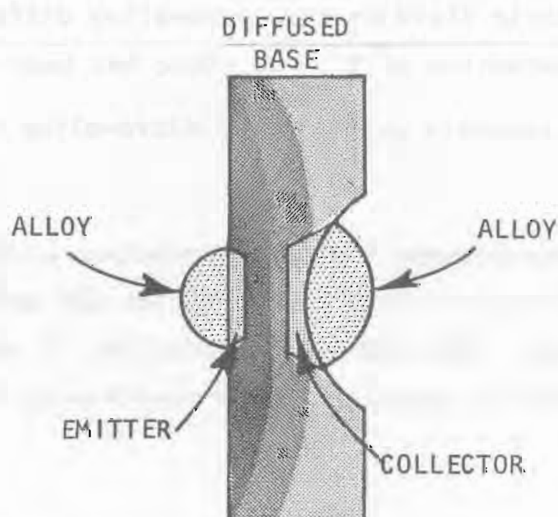
Typical $f_{h_{fb}}$ for the alloy transistor is from 1 to 10mc. This is, of course, an increase in $f_{h_{fb}}$ over the rate grown transistor. The rate grown transistor is rugged and useful at the lower frequencies, but the alloy type gives an increase in the transistor cut-off frequency over the rate grown.

The micro alloy transistor is a variation of the alloy type. To obtain a very thin uniform base, the center of the crystal wafer is etched very thin with streams of acid before the dots are applied and alloyed. This gives the characteristics of an alloy transistor with a much thinner base.

During the etching process, the concentration of acid is applied to both sides of the crystal wafer which removes some of the material. The outer edges are not etched and give the device strength and rigidity. The dots are applied to the thin portion of the wafer and when alloyed, result in a very thin base. The dots, of



MICRO-ALLOY



MICRO-ALLOY DIFFUSED

FIGURE 70-3

course, should give an impurity concentration that is the opposite of the etched wafer as shown in Figure 69-3.

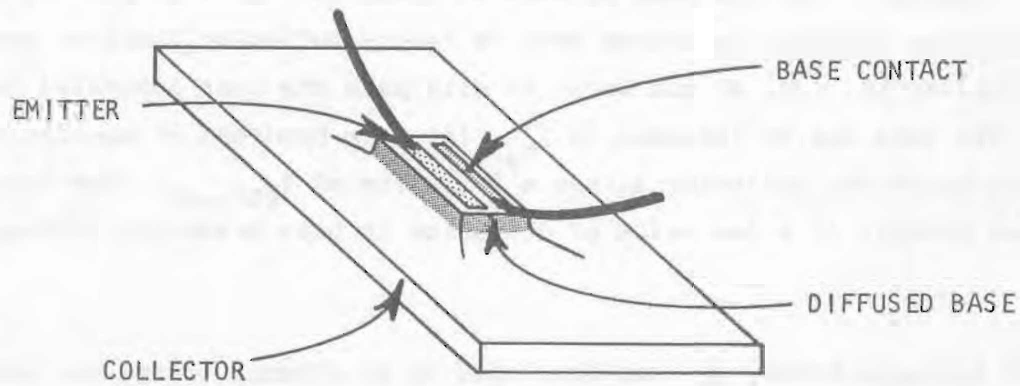
An increase in $f_{h_{fb}}$ to 40mc is typical with the micro-alloy transistor.

A further variation of the alloy type transistor is the micro-alloy diffused or drift field transistor. In this type of construction, the dopents are diffused (diffused to mean the process of adding impurities to the crystal and not the diffusion of carriers) into the wafer. Diffusion doping is the process of heating the wafer while surrounding it with the impurities in a gaseous form. The impurities diffuse or enter the wafer during the process. The diffused wafer is then etched predominantly on the collector side. This places the heavier doped portion of the base nearest the emitter when the dots are applied and alloyed. The doping will appear to be graduated heavy to light from the emitter to the collector.

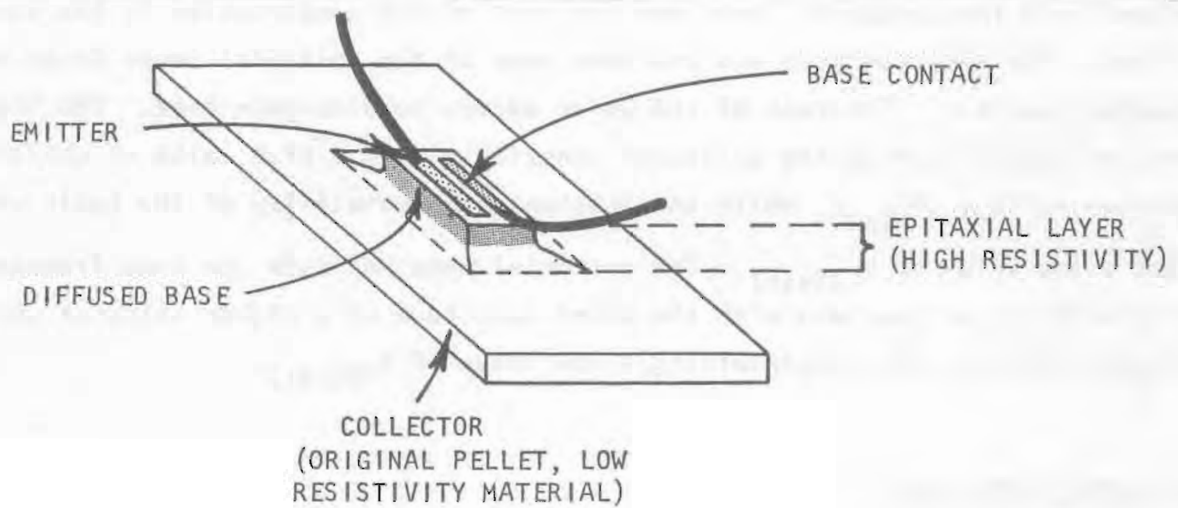
This graduated doping in the base gives the effect of an electric accelerating field in the base. The diffusing carriers are accelerated, reducing both transit time and the spreading out of the carriers in the base during their diffusion. Of course, this effective electric field in the micro-alloy diffused base increases the frequency limits. An extension of $f_{h_{fe}}$ to 200mc has been obtained with this construction. Another name commonly used for the micro-alloy diffused is the drift field transistor.

To reduce the bulk resistance between the base connecting wire and the active area of the base, and the capacitance of both the collector and emitter junctions, the mesa transistor was developed. The aim and construction of the mesa transistor is a very narrow base, low junction capacity, while maintaining the collector large to dissipate heat.

The mesa construction starts with a thin layer of dopents being diffused into the basic wafer. The diffusion impurities in the doping process have opposite characteristics to the dopents in the basic wafer. The thin diffused layer and the basic wafer form a junction. This is shown in Figure 71-3. The thin diffused layer becomes the base. A non-rectifying or ohmic contact is connected to the layer for electrical connection. Since the diffused layer is the base, the basic wafer should be doped heavier than the diffused layer. The emitter can either be alloyed



MESA



EPITAXIAL MESA

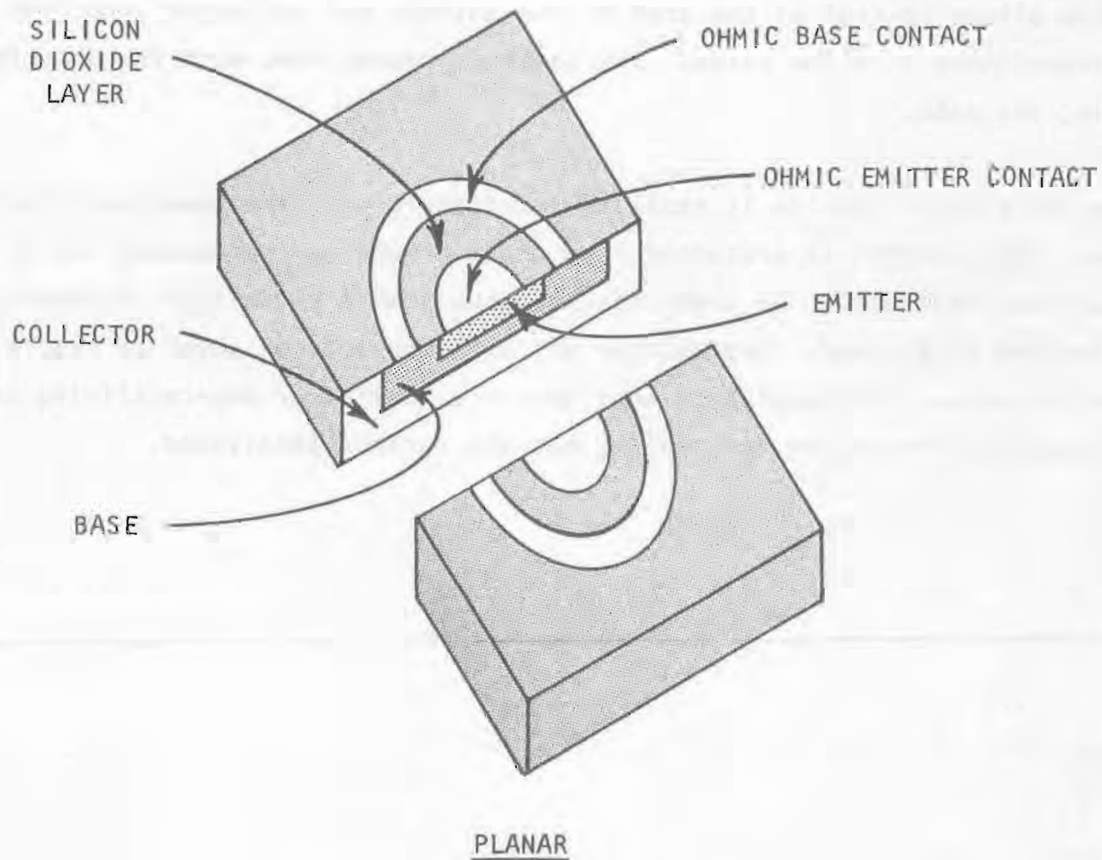
FIGURE 71-3

into the base or an aluminum strip used to construct the emitter. The alloyed emitter or aluminum strip and the diffused layer form the second junction. The top of the wafer is etched away except that portion containing the base lead and the emitter. This leaves a flat top peak or mesa as shown in Figure 70-3. The area around the collector junction is etched away to reduce collector junction capacitance, while leaving the rest of the wafer to dissipate the heat generated by power dissipation. The mesa has an increase in $f_{h_{fb}}$ into the hundreds of megacycles. The low resistance of the collector allows a low value of $V_{CE(sat)}$. The low resistance collector results in a low value of collector to base breakdown voltage, however.

A variation of the mesa transistor was developed in an attempt to try and maintain a low $V_{CE(sat)}$ and still have a high value of collector breakdown voltage (BV_{CBO}). This variation of the mesa is referred to as an epitaxial mesa. The epitaxial mesa is constructed by passing a gas over the structure until the gas has deposited a layer of solidified material on top of the original structure. This is what is referred to as an epitaxial layer on the top of the basic wafer. By controlling the impurity content of the gas, we can control the impurity content in the epitaxial layer. This layer in the epitaxial mesa is a high resistivity layer. The base is diffused into the epitaxial layer and the rest of the construction is the same as the mesa. The diffused base and the remainder of the epitaxial layer forms the collector junction. The rest of the wafer serves to dissipate heat. The high resistance epitaxial layer at the collector junction allows a high value of collector breakdown voltage (BV_{CBO}), while the inherent low resistivity of the basic wafer allows a low value of $V_{CE(sat)}$. The epitaxial mesa has much the same frequency characteristics as the mesa with the added advantage of a higher value of collector breakdown voltage, while maintaining a low value of $V_{CE(sat)}$.

THE PLANAR TRANSISTOR:

The planar transistor is constructed by diffusing both the emitter and the base into the basic crystal wafer. When this is done, the planar is referred to as a double diffused transistor. A layer of silicon dioxide is grown on the surface of the basic crystal wafer. Diffusion of impurities cannot occur through the silicon dioxide layer. Holes can be cut through the silicon dioxide, however, and allow



PLANAR

FIGURE 72-3

well controlled diffusion of the emitter and base as shown in Figure 71-3. The original pellet becomes the collector. The silicon dioxide masks the remainder of the collector area and keeps the junction confined to a selected area. This maintains a low junction capacitance and the mesa or flat-topped peak is not necessary. In other words, we do not need to etch the planar since the masking using silicon dioxide allows control of the area of the emitter and collector junctions and keep the capacitance at a low value. The masking process does much the same job as etching the mesa.

Since the silicon dioxide is replaced immediately over the junction after it is formed, the junction is protected from the surrounding environment and is said to be surface passivated. We might say that the planar transistor is double diffused and surface passivated. The emitter and base contacts as shown in Figure 72-3 are ohmic contacts. This indicates that they are passive or non-rectifying contacts. The junctions are active internally, but are surface passivated.

QUESTIONS FOR SECTION 3

Read each question carefully, studying any diagrams provided and select the most correct answer.

1. Forward biasing a PN junction results in majority carriers crossing the junction and becoming minority carriers. The average time the carriers exist after crossing the junction and before recombining is termed ________________.
   a. half life of the material
   b. minority carrier lifetime
   c. transit time
   d. recombination time

2. Applying forward bias to a PN junction with one side dopes much lighter than the other, results in some carriers failing to recombine and diffusing through the lightly doped side as ______________ carriers.
   a. heat
   b. bias
   c. majority
   d. minority

3. Doping one side of a PN junction much lighter and making it very narrow results in much of the forward bias current in the external circuit being the result of ______________ current in the diode.
   a. diffusion
   b. majority carrier
   c. photon
   d. phonon

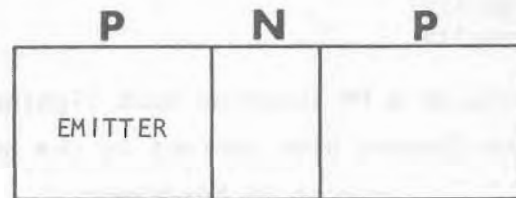
4. Forward bias current transported by diffusion of carriers results when the ____________ side of the junction is doped much lighter than the opposite side.
   a. N
   b. P
   c. either N or P
   d. none of the above

5. When two junctions are formed in a single piece of semiconductor, recombination occurs at both junctions until a state of ______________ exists at both junctions and carrier movement across the junctions ______________.
   a. pandimonium, starts
   b. balance, stops
   c. equilibrium, starts
   d. reverse bias, reverses

6. Junction transistors are constructed by doping two junctions in a single piece of semiconductor with the two ends __________ doped and the center __________ doped.
   a. lightly, heavily
   b. heavily, heavily
   c. heavily, lightly
   d. lightly, lightly

7. The center portion of a junction transistor is made narrow to enhance the transport of carriers by __________ in the center portion.
   a. photon
   b. majority carrier
   c. only electron
   d. diffusion

8. The center portion in the diagram is termed the __________ and the right hand portion the __________.



   a. base, collector
   b. collector, base
   c. gate, collector
   d. gate, anode

9. Injection of carriers from the emitter into the base is accomplished by application of __________ __________ to the emitter-base junction.
   a. thermal energy
   b. reverse bias
   c. forward bias
   d. series resistance

10. The collector junction is normally __________ biased and depends on __________ carriers in the base for current.
    a. reverse, minority
    b. forward, majority
    c. forward, minority
    d. reverse, majority

11. The emitter current is the sum of __________ and __________ current in the __________.

    a. recombination, diffusion, base
    b. collector, diffusion, base
    c. base, collector, transistor
    d. either a or c
    e. base, recombination, collector
    f. either b or e

12. h_{FB} is the symbol for __________ current gain from the __________ to the __________ of the transistor.

    a. small signal, emitter. collector
    b. d-c, base, collector
    c. small signal, base, collector
    d. d-c, emitter, collector

13. The product of d-c current gain, __________(symbol), and the d-c base current will give the value of __________ current.

    a. h_{FE}, collector
    b. h_{FB}, collector
    c. h_{FE}, emitter
    d. h_{FB}, emitter

14. A variation in base or emitter current will be accompanied by a change in collector current. For a given change in base or emitter current, a change in __________ current will result in the greatest change in collector current.

    a. base
    b. emitter

15. I_{CBO} indicates current between __________ and __________, with the __________ open circuited.

    a. emitter, collector, base
    b. collector, emitter, base
    c. collector, base, emitter
    d. emitter, base, collector

16. I_{CBO} is made up of carriers present as a result of __________ and will vary with a change in __________.

    a. thermal energy, forward bias
    b. reverse bias, thermal resistance
    c. forward bias, temperature
    d. thermal energy, temperature
    e. reverse bias, capacitance

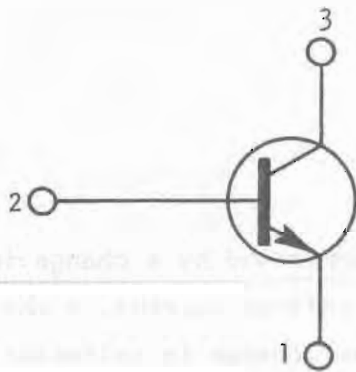
17. I_{CEO} is the current between __________ and __________ with the __________ open circuited.

   a. collector, emitter, base
   b. collector, base, emitter
   c. emitter, base, collector
   d. collector, emitter, emitter

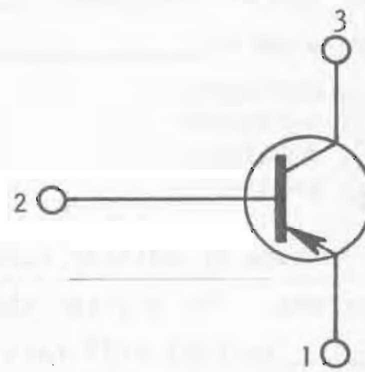
18. $I_{CEO} \approx$ __________ $\times I_{CBO}$.

   a. h_{FB}
   b. h_{FE}
   c. ambient temperature
   d. thermal resistance

19.



SYMBOL A



SYMBOL B

Symbol B is a/an __________ transistor and point 2 indicates the __________ lead.

   a. NPN, base
   b. PNP, base
   c. PNP, emitter
   d. NPN, emitter

20. A transistor in a common emitter configuration has a greater possible current gain than __________.

   a. a transistor in a common base configuration
   b. a transistor in a common collector configuration
   c. unity
   d. any of the above
   e. either a or c

21. A transistor in a common __________ configuration can yield the highest possible power gain of the three configurations.

    a. emitter
    b. base
    c. collector

22. A transistor in a common __________ configuration cannot offer a current gain greater than unity.

    a. collector
    b. emitter
    c. base

23. A transistor in a common __________ configuration cannot offer a voltage gain greater than unity.

    a. collector
    b. emitter
    c. base

24. h_{fe} is the __________, a-c current gain of the transistor in a common __________ configuration.

    a. direct current, base
    b. small signal, emitter
    c. direct current, emitter
    d. small signal, base

25. The small signal, a-c, current gain of the transistor in a common base configuration is given the symbol ________ and ________ include circuit effects.

    a. h_{fb}, does not
    b. h_{fc}, does not
    c. h_{fb}, does
    d. h_{fe}, does
    e. h_{fc}, does not

26. Power gain is possible with the common base configuration, although __________ gain is less than unity, since __________ gain greater than unity is possible.

    a. voltage, current
    b. current, voltage

27. Power gain is possible with the common collector configuration, although ___________ gain is less than unity, since ___________ gain greater than unity is possible.

    a. current, voltage
    b. voltage, current

28. Emitter follower is another name given a transistor in a common ___________ configuration.

    a. emitter
    b. collector
    c. base

29. h_{fc} is the small signal, low frequency current gain of the transistor in a common ___________ configuration and is approximately equal to ___________.

    a. base, $h_{fb} + 1$
    b. emitter, $h_{fe} + 1$
    c. collector, $h_{fb} + 1$
    d. collector, $h_{fe} + 1$

30. The emitter follower offers a ___________ input resistance and a ___________ output resistance.

    a. low, low
    b. high, high
    c. high, low
    d. low, high

31. A transistor in a common ___________ configuration has a very low output resistance is able to supply large output ___________.

    a. collector, currents
    b. emitter, voltages
    c. base, voltages
    d. base, currents

32. The slope of a d-c load line constructed on a collector family of curves is determined by the ___________ in ___________ the transistor.

    a. reactance, series with
    b. resistance, series with
    c. current, the base of
    d. current, the collector of

33. The d-c load line is constructed between the maximum __________ point and the maximum __________ point on the collector family of curves set by the series resistance and the total supply __________.

    a. voltage, current, current
    b. voltage, current, voltage
    c. voltage, current, resistance
    d. resistance, current, voltage

34. The transistor is termed in saturation when the collector becomes __________ __________. The collector voltage at this point is given the symbol __________.

    a. cut-off, V_{CC}
    b. forward biased, BV_{CBO}
    c. reversed biased, $V_{CE(sat)}$
    d. forward biased, $V_{CE(sat)}$
    e. fully saturated, $BV_{CBO(sat)}$

35. Saturation in a transistor results in an increase in the __________ __________ in the base as a result of the collector junction becoming __________ biased.

    a. stored charge, reverse
    b. depletion region, forward
    c. depletion region, reverse
    d. stored charge, forward

36. Avalanche breakdown occurs at high __________ collector voltages due to the __________ of carriers at the collector junction.

    a. forward, acceleration
    b. reverse, recombination
    c. reverse, multiplication
    d. forward, multiplication

37. Changing the transistor's configuration will change the collector voltage point at which avalanche breakdown occurs. BV_{CBO} indicates the breakdown voltage, __________ to __________ with the __________ d-c open circuited.

    a, collector, base, emitter
    b. collector, emitter base
    c. emitter base, collector
    d. collector, ground, emitter

38. BV_{CEO} indicates the breakdown voltage, __________ to __________ with the __________ d-c open circuited.

a. base, emitter, collector
b. collector, emitter, base
c. collector, base, emitter
d. collector, ground, emitter

39. BV_{CER} is the breakdown voltage, __________ to __________ with a specified value of resistance between __________ and __________.

a. collector, emitter, collector, base
b. collector, base, base, emitter
c. collector, emitter, base, emitter
d. collector, emitter, collector, emitter

40. Maximum steady state power dissipation in a transistor is limited by the maximum allowable junction __________, surrounding air __________, and the total __________ __________, junction to ambient.

a. temperature, temperature, thermal resistance
b. current, temperature, temperature gradient
c. voltage, temperature, thermal resistance
d. temperature, temperature, temperature gradient

41. Θ_{JA} is the symbol for thermal resistance, __________ to __________ and is measured in __________.

a. jacket, ambient, °C/watt
b. junction, area, watts/°C
c. junction, ambient, watts/°C
d. junction, ambient, °C/watt

42. $\Theta_{JC} + \Theta_{CS} +$ __________ $= \Theta_{JA}$ when a separate __________ sink is used with a transistor.

a. Θ_{SC}, heat
b. Θ_{CA}, conductance
c. Θ_{CA}, heat
d. Θ_{SA}, heat

43. __________ lubricant reduces thermal resistance, case to __________ and as a result, reduces total thermal resistance.

a. Germanium, heat sink
b. Silicone, heat sink
c. Gallium, heat sink
d. Gallium, ambient

44. $\Theta_{JC} = 1.8°C/watt$
$\Theta_{SA} = 3.6°C/watt$
$\Theta_{CS} = 0.6°C/watt$
$T_A = 60°C$
$T_{Jmax} = 150°C$

$$P_{max} = \frac{T_{J(max)} - T_A}{\Theta_{JA}}$$

The maximum steady state power the transistor can dissipate with the thermal characteristics listed above is ___________.

a. 1.5 watts
b. 30 watts
c. 15 watts
d. 3 watts

45. For maximum linear power operating in the active region, the load line should be placed ___________________________________.

a. tangent to the maximum power curve in the most linear region of curves.
b. from BV_{CBO} to a point tangent to the maximum power curve
c. from I_C(max) to a point tangent to the maximum power curve
d. I_C(max) to V_{CE}(max)

46. Answer d in question 45 will allow _____________ for a given transistor.

a. maximum linear power
b. maximum power with maximum current swing in the linear region
c. maximum power for low repetition rate switching
d. maximum power for high repetition rate switching

47. Little power is dissipated at cut-off and __________ while most of the power is dissipated in the _______________ region of the transistor's characteristics.

a. transient, active
b. depletion, transient
c. saturation, active
d. saturation, depletion

48. _______________ power can be dissipated during switching with the transistor is operated as a cut-off to saturated, low rep-rate switch and the load line can ________________ the maximum power curve on a collector family of curves.

a. less, not cross over
b. more, be tangent to
c. more, be in the safe area below
d. more, cross over

49. The off to saturated switching mode offers ____________ resting state power, but limited repetition rate due to increased __________.

a. low, storage time
b. high, storage time
c. maximum, transit time
d. low, transit time

50. A transistor switch turn on time (t_{on}) is equal to the sum of ______________ and ______________, and the transistors turn off time is equal to the sum of ___________ and _____________ when using the switching parameters.

a. rise time, transit time fall time, response time
b. rise time, transit time storage time, fall time
c. rise time, delay time storage time, fall time
d. rise time, interface time storage time, recovery time

51. A capacitor used to reduce switching time in a transistor switching configuration, by aiding in the establishment and removal of the stored charge is referred to as a ____________ ___________ capacitor.

a. switching filter
b. speed up
c. charge chopper
d. time chopper

52. The charge on the capacitor mentioned in question 51 should be ____________ the transistors stored charge for an effective increase in switching time.

a. equal to
b. less than
c. ten times
d. one tenth of

53. __________ __________ switching offers faster switching times than the saturated mode, but sacrifices ______________.

a. cut-off, flexibility
b. current mode, less critical transistor specification
c. voltage mode, power
d. current mode, power capabilities

54. A/an ____________ switching transistor has the transistor switching in the collector breakdown region and can offer rise times in the fractional ________ second region.

a. avalanche, nano
b. avalanche, micro
c. current mode, nano
d. current mode, micro
e. avalanche, milli

55. The three main high frequency limiting factors in transistors are __________, __________, and __________.

a. doping ratio, junction capacity, storage time
b. junction capacity, base transit time, thermal resistance
c. base transit time, spreading of carriers in the base, junction capacity
d. emitter transit time, spreading of carriers in the base, junction capacity

56. __________ is the symbol for the cut-off frequency of a transistor in a common base configuration and indicates the frequency at which __________ falls to 0.707 of its low frequency value.

a. f_{hfe}, h_{fe}
b. f_t, h_{fb}
c. f_{hfb}, h_{fb}
d. f_{α_b}, h_{fe}

57. C_{ob} is the symbol for the output capacitance of a transistor in a common __________ orientation. C_{oe} is the symbol for the output capacitance of a transistor in a common __________ orientation.

a. base, emitter
b. emitter, base

58. A __________ base and __________ junction capacity is the aim in high frequency transistors.

a. heavily doped, low
b. lightly doped, diffused
c. narrow, low
d. narrow, variable

59. f_t is the symbol for the frequency at which __________ falls to __________.

a. h_{fe}, unity
b. h_{fb}, 0.707 of low frequency value
c. h_{fe}, 0.707 of low frequency value
d. h_{fb}, unity

60. f_{hfe} indicates the cut-off (__________) of the transistor in a common ________________ configuration.

a. h_{fe} = unity, emitter
b. low frequency h_{fe} x 0.707, base
c. h_{fe} = unity, base
d. low frequency h_{fe} x 0.707, emitter

ANSWER SHEET

QUESTIONS FOR SECTION 1

1. d
2. b
3. a
4. c
5. d
6. a
7. d
8. b
9. d
10. e
11. d
12. c
13. e
14. b
15. d
16. a
17. f
18. c
19. a
20. d
21. d
22. a
23. e
24. a
25. b
26. e
27. c
28. a
29. d
30. d
31. d
32. e
33. b
34. c
35. a
36. b
37. d
38. c
39. a
40. b

ANSWER SHEET

QUESTIONS FOR SECTION 2

1. b
2. c
3. a
4. a
5. d
6. b
7. d
8. c
9. a
10. b
11. d
12. a
13. b
14. c
15. c
16. c
17. a
18. b
19. a
20. d
21. a
22. b
23. c
24. d
25. b
26. b
27. a
28. c
29. b
30. b
31. a
32. b
33. c
34. c
35. c
36. b
37. a
38. c
39. a
40. d
41. b
42. c
43. a
44. d
45. c
46. a
47. d
48. b
49. a
50. d
51. e
52. a
53. c
54. a
55. b
56. d
57. b
58. a
59. c
60. b
61. e
62. d

ANSWER SHEET
QUESTIONS FOR SECTION 3

1. b
2. d
3. a
4. d
5. b
6. c
7. d
8. a
9. c
10. a
11. d
12. d
13. a
14. a
15. c
16. d
17. a
18. b
19. b
20. e
21. a
22. c
23. a
24. b
25. a
26. b
27. b
28. b
29. d
30. c
31. a
32. b
33. b
34. d
35. d
36. c
37. a
38. b
39. c
40. a
41. d
42. d
43. b
44. c
45. a
46. c
47. c
48. d
49. a
50. c
51. b
52. a
53. d
54. a
55. c
56. c
57. a
58. c
59. a
60. d

PERIODIC TABLE OF THE ELEMENTS

Max. No. Electrons Per Shell	Valence / No. Shells	I	II	III	IV	V	VI	VII		VIII		0
2	1	1-H-1 Hydrogen										2-He-4 Helium
8	2	3-Li-7 Lithium	4-Be-9 Beryllium	5-B-11 Boron 2-3	6-C-12 Carbon 2-4	7-N-14 Nitrogen	8-0-16 Oxygen	9-F-19 Fluorine				10-N3-20 Neon
18	3	11-Na-23 Sodium	12-Mg-24 Magnesium	13-Al-27 Aluminum 2-8-3	14-Si-28 Silicon 2-8-4	15-P-31 Phosphorus	16-S-32 Sulfur	17-Cl-35 Chlorine				18-A-40 Argon
32	4	19-K-39 Potassium	20-Ca-40 Calcium	21-Sc-45 Scadium	22-Ti-48 Titanium	23-V-51 Vanadium	24-Cr-52 Chromium	25-Mn-55 Manganese	26-Fe-56 Iron	27-Co-59 Cobalt	28-Ni-58 Nickel	
		29-Cu-63 Copper	30-Zn-64 Zinc 2-8-18-2	31-Ga-69 Gallium 2-8-18-3	32-Ge-74 Germanium 2-8-18-4	33-As-75 Arsenic 2-8-18-5	34-Se-80 Selenium 2-8-18-6	35-Br-79 Bromine				36-Kr-84 Krypton
32	5	37-Rb-85 Rubidium	38-Sr-88 Strontium	39-Y-89 Yttrium	40-Zr-90 Zirconium	41-Nb-93 Niobium	42-Mo-09 Molybdenum	43-Tc-99 Technetium	44-Ru-102 Ruthenium	45-Rh-103 Rhodium	46-Pd-106 Palladium	
		47-Ag-107 Silver	48-Cd-114 Cadmium	49-In-115 Indium 2-8-18-18-3	50-Sn-120 Tin	51-Sb-121 Antimony 2-8-18-18-5	52-Te-130 Tellurium	53-I-127 Iodine				54-X3-132 Xenon
	6	55-Cs-133 Cesium	56-Ba-138 Barium	57-La-139 Lanthanum	72-Hf-180 Hafnium	73-Ta-181 Tantalum	74-W-184 Wolfram (Tungsten)	75-Re-187 Rhenium	76-Os-192 Osmium	77-Ir-193 Iridium	78-Pt-195 Platinum	
		79-Au-197 Gold	80-Hg-202 Mercury	81-Ti-205 Tallium	82-Pb-206 Lead	83-Bi-209 Bismuth	84-Po-210 Polonium	85-At-210 Astatine				86-Rn-222 Radon
	7	87-Fr-223 Francium	88-Ra-226 Radium	89-Ac-227 Actinium								