DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITY OF MICHIGAN
ANN ARBOR

THEORY OF SEMICONDUCTOR DEVICES

J. A. BECKER AND J. N. SHIVE
BELL TELEPHONE LABORATORIES

LECTURES GIVEN AT

SUMMER ELECTRONICS SYMPOSIUM

AUGUST 7 TO AUGUST 18
1950

W. G. DOW DIRECTOR

COPYRIGHT 1950 BY
BELL TELEPHONE LABORATORIES INC.
NEW YORK, NEW YORK

—enm
VMR0291

Lectures on Theory of Semiconductor Devices

University of Michigan

August 7 to 18, 1950

J. A. Becker and J. N. Shive
Bell Telephone Laboratories

Lecture 1 -

Theory of Processes Occurring Inside Semiconductors

Lecture 2 -

Theory of Thermistors (Covered in Bell Lab. Reprint)

Lecture 3 -

Fermi Level, Its Importance, and How It Is Determined

Lecture 4 -

Hall Effect and Thermoelectric Effect

Lecture 5 -

Structure of Barrier Layer

Lecture 6 -

Electron and Hole Currents in Barriers

Lecture 7 -

Optical Properties and the Inner Photoelectric Effects

Lecture 8 -

Theory of Transistor Action (Covered in Bell Lab. Reprint)

Engin
Ulrich
2-28-52
78136

LECTURE 1

THEORY OF PROCESSES OCCURRING INSIDE SEMICONDUCTORS

GENERAL

In the last two decades there have been developed a number of semiconductor devices which have already found large scale application in engineering; these include copper oxide and selenium rectifiers, modulators, silicon and germanium detectors and mixers, thermistors for time-delay, protective and regulating devices, and the transistor amplifier and oscillator. Such devices and their uses have been described in the technical journals. In eight lectures we hope to present a coherent theory of semiconductor devices and to explain semiconductor properties in terms of the modern theory of the solid state. Our course will take us into a consideration of the crystalline structure of solids and into a detailed study of the behavior of electrons inside crystals. One of the most useful tools now available for such studies is the concept of electron energies. We shall use it frequently and it is our task now to present the fundamentals of this concept.

BAND THEORY

If one plots the potential in the vicinity of an isolated hydrogren atom nucleus, one obtains the picture shown in Figure. 1-1. In this figure we have presented the convention that the zero of potential is the horizontal dashed line corresponding to the energy of the system when the electron is far away from the nucleus. As the electron is allowed to approach the nucleus, since the two are of equal but opposite charges, work is done by the system and the potential of the configuration in practical volts is equal to $-\frac{e}{r} \times 9 \times 10^{11}$, where e is the electronic charge in coulombs and r is the distance between the electron and the nucleus in cms.

If we now inquire about the various ways in which the electron can reside permanently in this "potential well" so as to produce a stable hydrogen atom we come face to face with certain restrictions which arise from the obedience of the system to the laws of quantum mechanics. These laws say that the electron may reside on any one of a limited number of discrete total energy levels within this

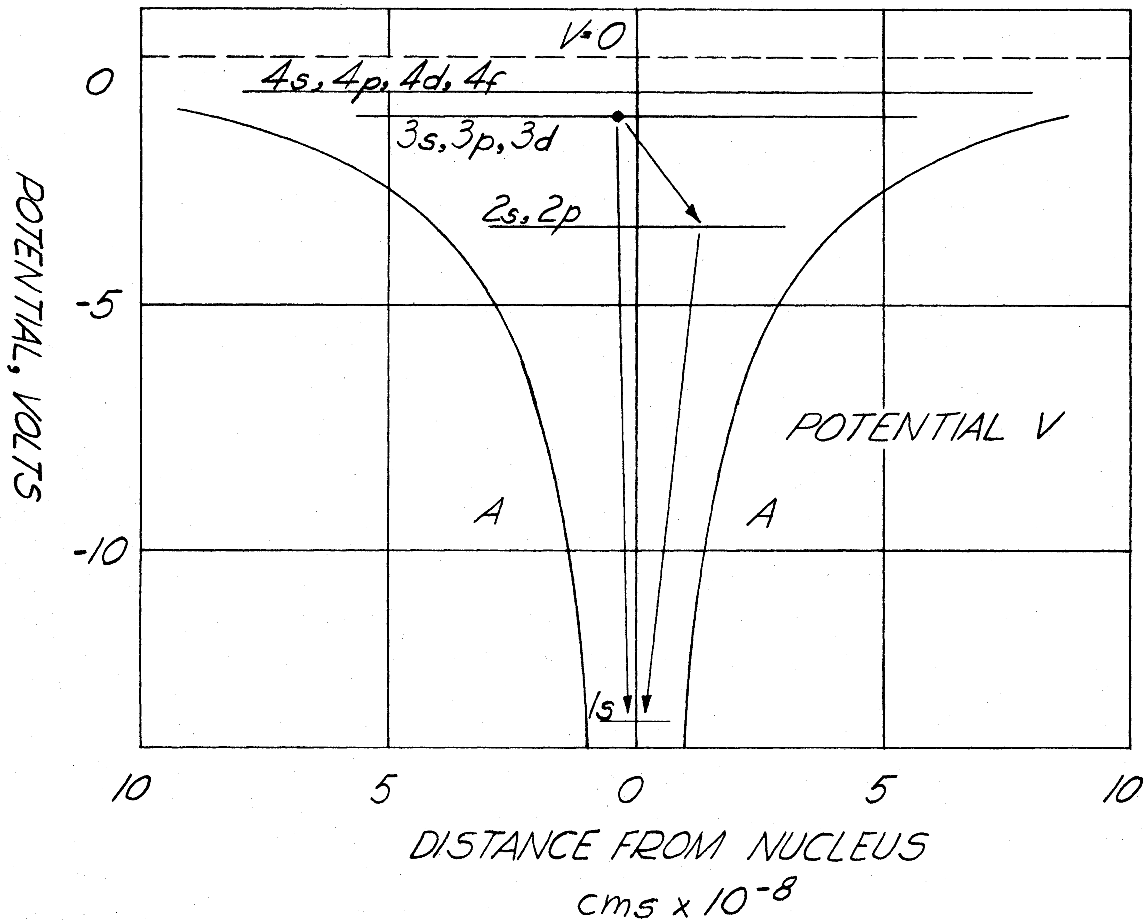


Figure 1-1 Plot of the Potential Energy of an Electron in the Vicinity of a Hydrogen Nucleus. Horizontal lines inside the "well" represent the allowed energy levels of the atom.

potential well, and nowhere else. These discrete energy levels are represented in Figure 1-1 by the solid horizontal lines designated by their respective spectroscopic term notations. When the electron occupies one of these discrete energy levels as drawn, the picture then represents one of the several possible configurations of the hydrogen atom. The binding energy of the configuration is given by the vertical distance from the occupied level to the dashed line V = 0 at the top of the diagram.

Note that the horizontal lines represent possible total energies of the system, potential plus kinetic. Since the sides of the well are given by the potential energy alone, it follows that the electron cannot classically be found outside the well in the regions marked A, since these regions correspond to negative kinetic energy.

One must remember that, once an electron actually occupies one of the energy levels, the picture changes. If now a second electron approaches the neutral atom it sees an entirely different energy scheme. The deep potential well would not exist for the second electron beyond the orbit of the first, since the positive charge of the nucleus has been practically neutralized by the negative charge of the first electron. While hydrogen nuclei sometimes do exist in stable linkage with two electrons simultaneously in hydride ions, the second electron is very weakly held and the lifetime of the configuration, once it is formed, is usually short.

In Figure 1-1 , the electron is pictured as occupying the third lowest energy level. The configuration represented is that of a hydrogen atom in one of its several possible energy states. It is one of the laws of nature that a system left alone by itself will spontaneously attempt to rearrange itself in the direction of lower energy. After a short time in the energy level depicted in Figure 1-1 the electron will fall to some lower-lying level, and will eventually come to residence in the lowest available level, either by a single jump or by a succession of smaller jumps to intermediate levels on the way. At every such jump ("transition," to use the spectroscopists' language) a quantum of radiation is emitted by the atom, and the energy of the quantum is exactly equal to the difference between the energies of the atom in its initial and final configurations. Possible spontaneous jumps of the electron to lower energy levels accompanied by the emission of radiation are represented by the arrows in Figure 1-1 . By the reverse of the above process, an electron in the lowest energy level can be raised to one of the higher levels by absorbing the energy of a quantum of radiation incident upon the atom from without. If the quantum thus incident has energy greater than the vertical distance from the lowest level to the dashed line $V = 0$, it may completely remove the electron from the vicinity of the nucleus, which then becomes a hydrogen ion. The liberated electron will have kinetic energy E_k equal to the difference between the energy brought by the quantum and the binding energy of the level in which the electron was initially resident: $E_k = h\nu - eV$, where h is Planck's constant and ν is the frequency of the radiation. The electron thus liberated wanders about freely in space until it finds something else with a potential well to fall into.

Let us look now at the energy level scheme for a more complicated atom. A sodium nucleus contains an excess positive charge equal to 11 proton charges. Consequently a neutral sodium atom has 11 electrons. The energy profile of the atom may be as shown in Figure 1-2 . In complicated atoms such as this the discrete energy levels are not, in general, arranged with any obvious order. If we take a sodium nucleus and start, in our imagination, adding the 11 electrons one by one

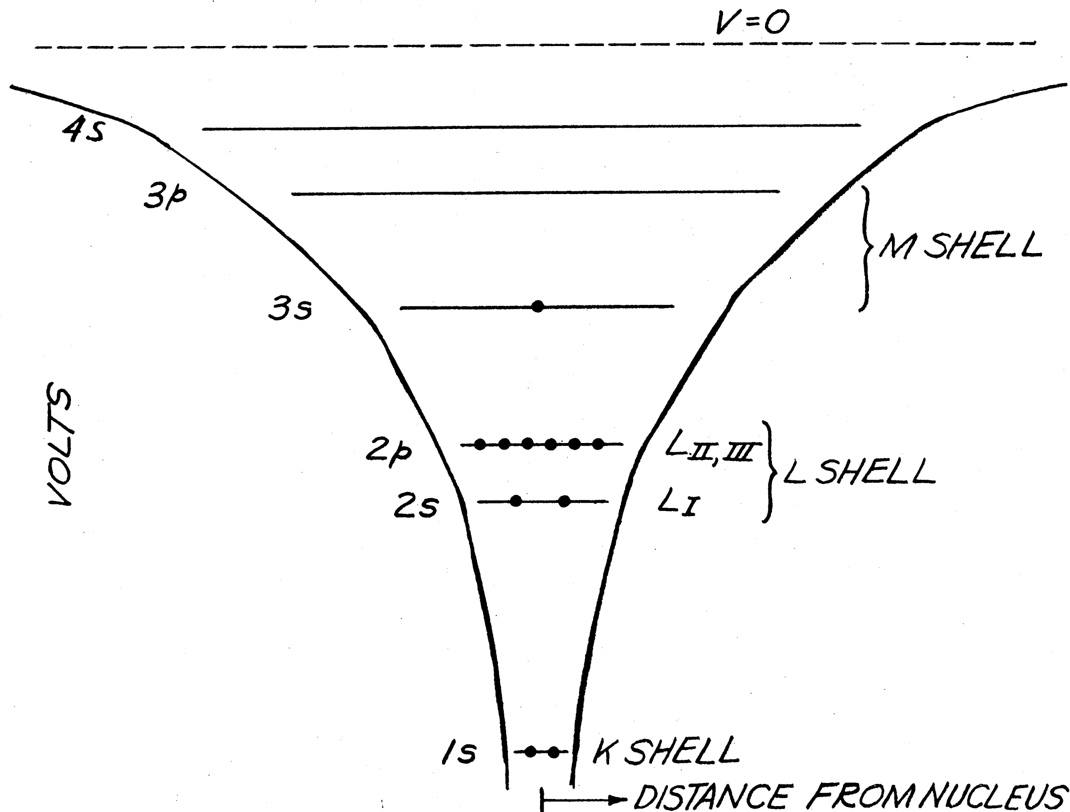


Figure 1-2 Plot of the Energy Level Scheme for a Sodium Atom

the first two will shortly come to residence on the lowest possible energy level (K electron shell, to use the X-ray designation). With the addition of the third electron another quantum mechanical law comes into effect. This law forbids more than two electrons* to occupy the same energy level in a single system. The name of "exclusion principle" has been applied to this law. This third electron will therefore occupy the second lowest energy level (L_I shell), as will the fourth. The fifth electron will occupy the third lowest level (L_{II} shell), and so on. This third lowest level shows a population of six electrons: it may be considered as a coincidence of three energy levels which are not separated enough for atoms of low atomic number to show as separate lines on the diagram. With the addition of each electron the shape of the profile of Figure 1-2 will change, as will the positions of all the energy levels contained by it, since the addition of another electron changes the net charge of the system. The energy levels have been drawn to represent the state of affairs just as the 11th electron is added to complete the neutral atom, and the potential has been to take into account the screening effect of the various inner electron shells. A twelfth electron, looking at this atom from

*These two electrons must have opposite spin.

nearby, would see no potential well at all, and would consequently experience no electrical urge to add itself to the atom. In fact, should it attempt to do so, its additional charge would so distort the profile of the atom that it would be repelled. This statement is not generally true for all atoms. Some electronegative atoms, for example, oxygen, can take on one or more extra electrons in addition to their normal complement. In order to present information which would otherwise be lost, license has been taken in not drawing Figure 1-2 to accurate scale. Compared with the separations shown between uppermost occupied levels, the distance down to the lowest level should really be several feet off the bottom of the page. An idea of the positions of the various levels in different atoms can be obtained experimentally by observation of their optical and X-ray spectra. It has been found that the binding energy of an electron in the lowest level of a hydrogen atom is about 13 electron volts. For sodium it is about 1100 electron volts, while for the most complicated atom of all, uranium, it is of the order of 100,000 electron volts.

Note that the uppermost occupied level in sodium contains one electron. This is the valence electron upon whose behavior depend most of the chemical and optical properties of the atom.

Suppose we now bring a large number of sodium atoms together to form a small sample of a sodium crystal. We might be justified from analogy in drawing the energy profile along a line of atoms something like that shown in Figure 1-3 .

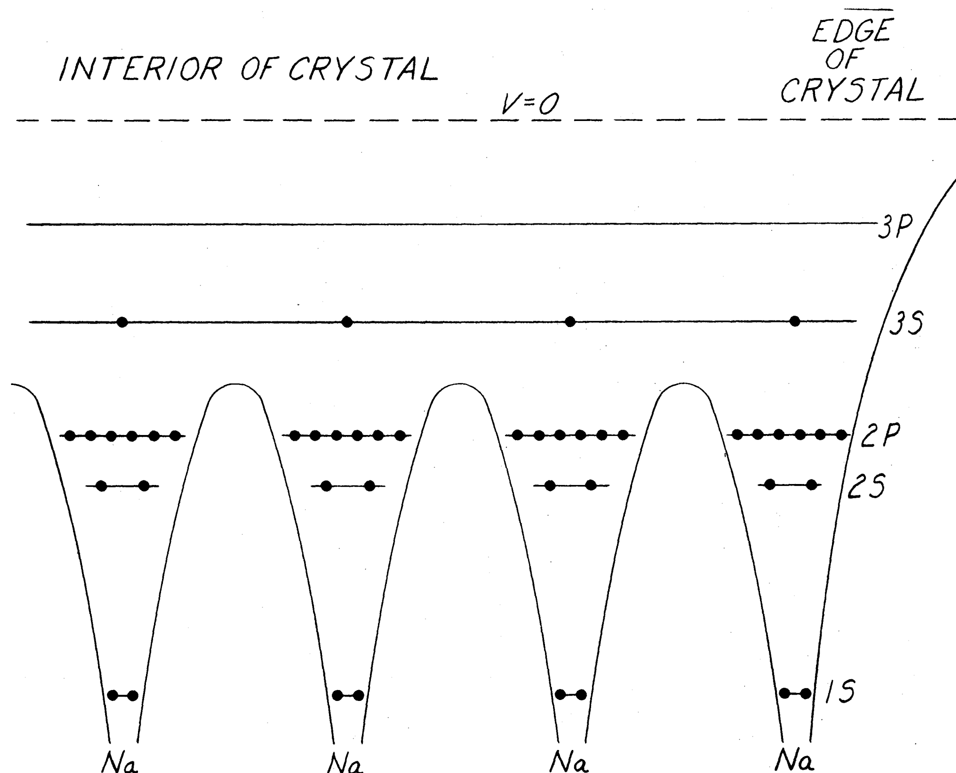


Figure 1-3 Energy Level Scheme for a Sodium Crystal as it Might be Inferred (Naively) from the Energy Level Scheme of an Isolated Sodium Atom

Because the atoms are now close together the potential between them cannot rise to V = 0 as it does outside the surface of the crystal. As the figure is drawn the potential hills between the atoms do not rise even as far as the uppermost energy level which is occupied by the single valence electron in each atom. There is thus nothing to prevent the valence electron originally belonging to one atom from leaving its parent atom and wandering freely through the crystal. The only requirement on the exercise of such freedom by the valence electrons would be that nowhere in the crystal would there there be at any time many more or many fewer such electrons than would correspond to reasonable over-all neutrality of net charge.

The picture of 1-3 , however justified, is somewhat over-simple, and the actual situation of the electrons in the atoms of a crystal requires further elaboration. Observe that the valence electrons of the various atoms all have the same energy in Figure 1-3 . Since they no longer belong to any particular atoms they must be considered as belonging to the crystal at large, with the crystal itself considered as a single system. Having all these electrons with the same energy in a single system would be a violation of the exclusion principle. Actually, when atoms come together to form a crystal, the discrete energy levels normally occupied by the electrons in the individual atoms split up into bands, as shown in Figure 1-4. Each band consists, itself, of a large number of closely spaced discrete energy levels. There are as many such levels in each band as there are atoms in the crystal specimen, and there are as many such bands in the crystal as there are energy levels in an isolated atom of the same substance. There is now no trouble with the exclusion principle, since each level within each band can accommodate two electrons.

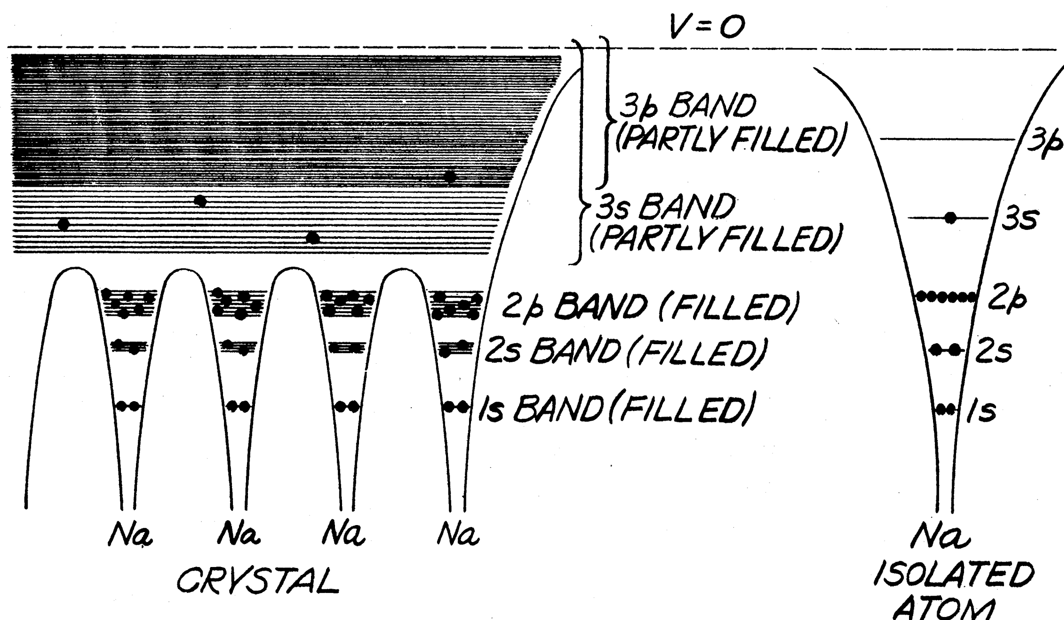


Figure 1-4 Correct Energy Level Scheme for a Sodium Crystal, Showing how the Bands of the Crystal Arise from the Discrete Levels of the Isolated Atom

In the case of sodium the splitting-up of the higher-lying bands is such that the 3s band and the 3p band overlap each other. In a crystal specimen with n monovalent atoms, there are n valence electrons. The valence electron band in the crystal has n closely spaced levels, with accommodations for 2n electrons. The band should be half filled, just as in the isolated atom of sodium the valence electron level is half filled with the one valence electron. Actually, in sodium the bottom of the overlapping 3p band comes at a lower energy than the halfway filling point for the 3s valence band. Consequently some of the original 3s electrons spill over into the 3p band, and from that point upward both bands fill equally to the limit of the number of valence electrons available.

There are a number of higher-lying spectroscopic levels in the isolated atom above those shown in Figure 1-2 . These give rise to additional higher-lying, overlapping bands in Figure 1-4 which have been omitted in the interest of keeping the presentation as simple as possible.

It may seem artificial to say that the discrete energy levels of individual atoms split up into bands when the atoms are brought together to form a crystal, merely in order to avoid violation of the exclusion principle. The complete explanation lies in the realm of quantum physics, but a part-way satisfactory insight into the mechanics of the splitting-up can be had by considering a mechanical analogy. Consider a row of n mechanical oscillators, all identical, all independent and all oscillating at the same frequency. Now suppose that a mechanical coupling is imposed between each oscillator and its neighbors in the line. It will now be found that the previously independent oscillators form a single system whose vibration frequency may have any one of n different values in a spectrum of n closely spaced frequencies grouped about the original single frequency of the isolated oscillator. The greater the coupling between the oscillators the broader will be the band and the farther apart in the band will be the n frequencies available to the system.

Similarly, in the crystal, a quantum-mechanical coupling exists between the electrons of one atom and those of its neighbors. This coupling is strongest in the higher lying electron levels. Consequently the splitting-up of the energy levels in the higher-lying bands of the crystal is greatest, and these bands are broadest from top to bottom. Because of the potential hills between the atoms in the crystal, the quantum mechanical coupling between electrons of lower-lying levels from one atom to the next is much smaller, and the lower-lying bands are narrower. For the lowest band of all, for the innermost shell of electrons in the atoms, the coupling is almost completely nonexistent, since these electrons are shielded from the outer world by the outer shells of electrons. Consequently, the lowest band is very narrw - hardly any wider than the discrete level for an isolated atom. An attempt has been made in drawing Figure 1-4 to show the relative breadths of these various bands.

Figure 1-5 shows the valence band scheme for a more complicated material, copper,[1] and illustrates how the bands develop from the energy levels of the

---

H. M. Krutter; Physical Review, Vol. 48, p.664, 1935.

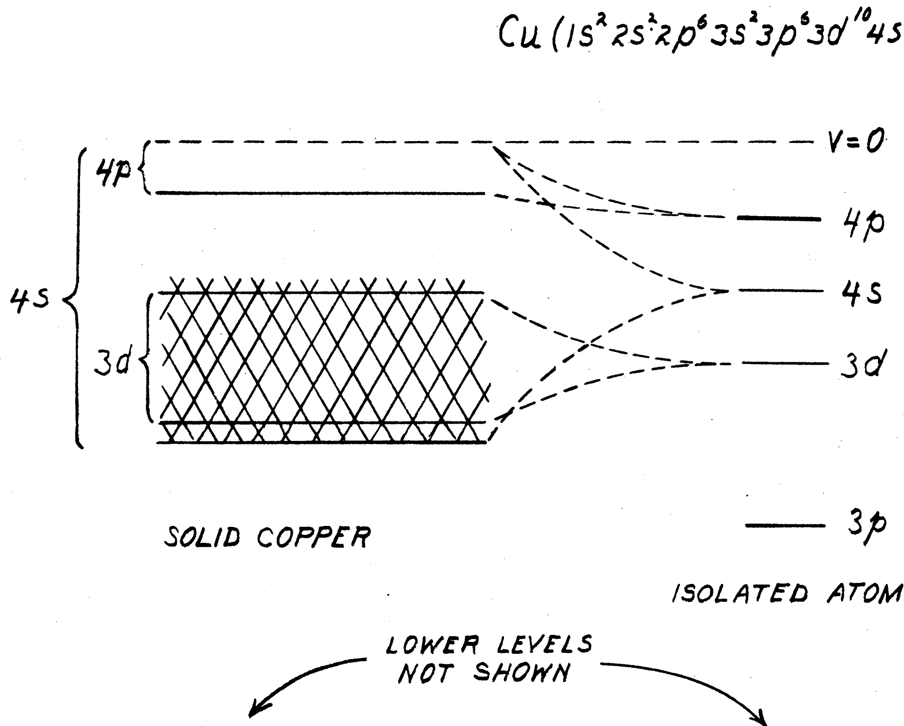


Figure 1-5 Band Scheme for Copper

isolated copper atom. The 10 electrons in the 3d level and the one electron in the 4s level of the isolated atoms cause a filling of the corresponding bands in the solid to a point slightly above the top of the 3d band as shown.

In our further consideration of band pictures we shall simplify the presentation in the figures. Since the deeper-lying electrons belonging to the inner shells of the atoms are quite tightly bound and do not contribute to the ordinary electrical behavior of the substance, we shall no longer include the lower-lying bands in the pictures we draw. Also we shall omit drawing separately the closely spaced energy levels in the bands and merely draw horizontal lines to denote the top and bottom edges of the bands. And rather than show the individual electrons by separate dots, we shall merely indicate by cross-hatching how far the bands are filled with electrons. This presentation has already been given in Figure 1-5 .

Figure 1-6 shows a band scheme in cuprous oxide, Cu_2O. In this compound the copper atoms are farther apart than in solid copper, hence the appearance of the copper bands is different from that of Figure 1-5 . The copper atoms are so far apart that the 3d and 4s bands do not overlap. The oxygen 2p band may lie below the 3d band of copper as shown, and has accommodations for six electrons per atom. The electron structure of oxygen gives normally only four electrons per atom in this band. However, if each of two copper atoms contributes a 4s electron to each oxygen atom in the 2p band of the latter, the structure of Figure 1-6 results, in which the

oxygen 2p and copper 3d bands are completely full and the copper 4s band is completely empty. Observe that in the cases of metallic copper and sodium there are empty electron levels immediately above the top of the electron distribution, while in Cu_2O there is an energy gap between the top of the filled 3d band and the bottom of the empty 4s band. As we shall see in the next section, the differences between metals and nonmetals can be explained in terms of the arrangement and degree of filling of the energy bands in the solid. The picture presented in Figure 1-6 is that for stoichiometrically pure cuprous oxide. The presence of impurities, departures from stoichiometry, or lattice defects complicates the simple picture in a way which will be discussed later.

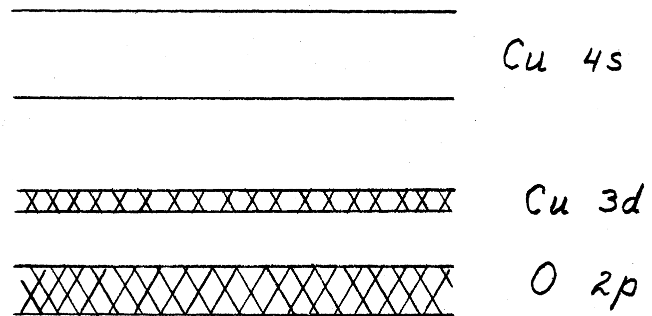


Figure 1-6 Tentative Band Scheme for Cuprous Oxide

BEHAVIOR OF ELECTRONS IN BANDS, ELECTRICAL CONDUCTIVITY IN METALS, INSULATORS, AND SEMICONDUCTORS

Having seen how the bands of allowed electron energies come about, let us investigate the behavior of electrons in these bands and show how the electrical conductivity properties of various types of material can be accounted for.

For reasons of fulfilling the minimum energy requirement, the electrons in a partly filled band occupy the lowest levels in the band and fill the band, as far as their numbers permit, from the bottom upward. At the absolute zero of temperature the top of this electron distribution will be quite sharp in the sense that all levels, up to the limit of the electrons available, are full, and all higher levels are completely empty. However, at temperatures above absolute zero the top of the electron distribution is "fuzzed out," in the sense that on a statistical basis some of the levels below the top of the distribution will be empty, while some levels above the top of the distribution will be occupied. The extent of this fuzzing out increases as the temperature increases, and the fuzzed out region has a breadth in energy of the order of kT, where k is Boltzmann's constant, and T is absolute temperature. For 300°K, kT is about .025 electron volt.

The electrons are constantly moving about, helter skelter through the crystal lattice, the energy, momentum, and velocity of each electron being specified by the quantum mechanical requirements associated with the particular level the electron occupies. On the average, and in the absence of electric fields, there are as many electrons going one way as another at any instant, and the net current due to

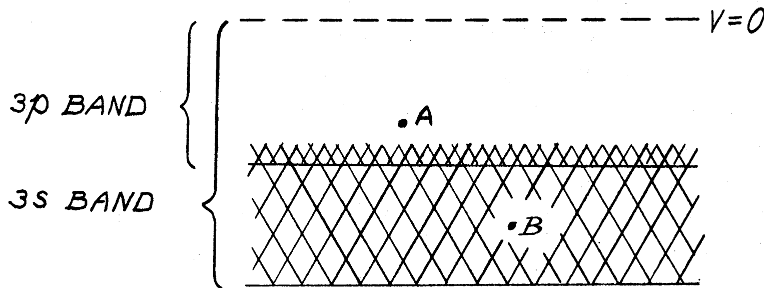


Figure 1-7 Band Scheme for Sodium Showing Mechanism of Electrical Conduction in a Metal

their motion is zero. The electrons are continually colliding with the atoms of the lattice and undergoing scattering and reflection in the process, as at the surfaces of the specimen as well.

Now consider an electron which by chance may occupy an energy level somewhat above the top of the distribution, as at A in Figure 1-7 . When an electric field is impressed across the specimen the electron experiences an urge to accelerate. Acceleration means a change of velocity, momentum, and usually direction as well. Hence, it means transferring the electron from its present level to a succession of other energy levels corresponding to increased momentum in the direction of the field. Since such energy levels are readily available in the neighborhood of A, the electron can be accelerated by the field and contribute to a net current in the sample. Consider, however, another electron at B, deep down in the distribution of completely filled levels. All the other levels in the immediate neighborhood of B to which this electron might be transferred by acceleration in the electric field are already filled by other electrons. This electron, therefore, cannot be accelerated unless the field is strong enough to transfer to it enough energy in one mean free path to lift it to some empty level near the top of the distribution. (If B is more than a few hundredths electron volt below the top of the distribution it would require unattainably high fields to do this). Such electrons, therefore, cannot contribute to the net current in the sample. We are left with the observation that it is only those electrons in and near the fuzzed-out top of the distribution that can participate in carrying a net current. These electrons ordinarily comprise around one per cent of the total population of the valence band. The modern theory of conductivity in solids thus differs from the classical picture, in which all valence electrons were supposed to participate in conduction. Since the same conduction properties are accounted for on the modern view by only a small fraction of the electrons, it follows that the electrons must have mean free paths much larger than classically supposed. This is indeed true, mean free path lengths for conduction electrons in metals being the order of hundreds of atom diameters. For semiconductors mean free paths may be as long as this, or may be orders of magnitude shorter.

The positive temperature coefficient of resistance of typical metallic conductors can easily be explained in terms of the band picture. The behavior of an electron in any given energy level is specified quantum mechanically by the wave function associated with the level. It is one of the properties of waves in periodic structures that if the structure is geometrically perfect the waves suffer no reflection or scattering. If the structure is imperfect, as for the interior of a crystal whose ions are displaced from their zero positions by thermal agitation, the waves are reflected and scattered. Such reflection and scattering accounts for the finite length of the mean free path of the conduction electrons. When the temperature increases, the mean free paths become shorter, thus increasing the probability that an accelerated electron will in unit time lose the energy it has acquired from the field and fall back down to the top of the electron distribution. The conduction electrons thus lose their acquired energy more frequently, and on the average their increment of velocity in the direction of the field is smaller.

Another type of lattice imperfection which acts as an agent for shortening the electron mean free paths comes from the presence of impurities. An impurity atom or ion interrupts the periodicity of the lattice structure and promotes increased scattering of electron waves.

From the foregoing remarks on the band structure and electron behavior in a metallic conductor it is a simple step to explain the electrical properties of insulators. An insulator is a substance in which there is an energy gap between the top of the valence band and the bottom of the next higher band, and in which there are exactly enough valence electrons to fill completely the valence band, Figure 1-8. It is evident that here there are no conduction electrons, for all the energy levels in the valence band are filled, and there are no unoccupied levels of higher energy or momentum to which electrons can be transferred by acceleration in an electric field. (If fields of the order of 100 million volts per centimeter could be realized, then some valence electrons could acquire enough energy from the field on one mean free path to be transferred to the higher unoccupied band where they could carry a net current, and the sample would become conducting. However, such fields cannot be realized short of breakdown.) The electrons in the valence band are all free in the sense that they are not bound to particular atoms in the crystal, but

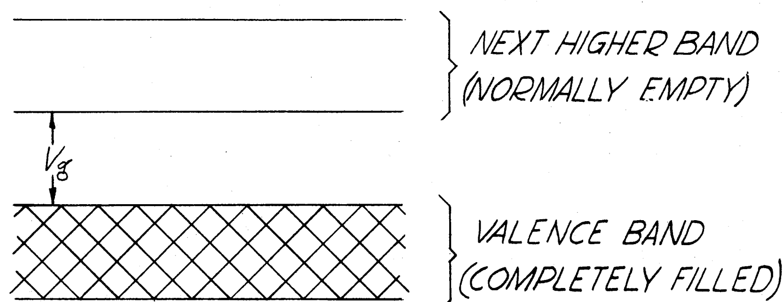


Figure 1-8 Band Scheme for an Insulator, With no Conduction Electrons

their motions are all specified and the application of a field produces no change in the net motion of the electron population.

If the temperature of the insulator of Figure 1-8 should be raised, then a few electrons from the filled valence band will be transferred by thermal excitation to the bottom levels of the upper empty band, where they can transport net current. The number n_e of electrons thus promoted is given in statistical equilibrium by $n_e \sim \exp\left(\frac{-eV_g}{2kT}\right)$ where e is the electronic charge, V_g the separation of the filled and empty bands in electron volts, k is Boltzmann's constant, and T the absolute temperature. The slight conductivity imparted to insulators by raising their temperature thus increases exponentially with temperature. By heating to a sufficiently high temperature, an insulator can be turned into a semiconductor. A semiconductor which owes its conductivity to thermal activation of electrons from the valence band to the upper empty band is called an "intrinsic semiconductor." This term is used to distinguish such semiconductors from impurity semiconductors which we shall discuss soon.

For every electron in an intrinsic semiconductor which is transferred to the empty band, a vacant space or "hole" is left in the electronic population of the otherwise filled valence band. These holes behave like particles having charge equal to the electronic charge but positive in sign, and with mass approximately equal to an electron's mass. They are usually termed "positive holes." A positive hole in an otherwise filled band can be accelerated by an electric field in about the same way an electron in an otherwise empty band can be accelerated to produce a net current. It is evident, therefore, that the presence of positive holes contributes to the conductivity of intrinsic semiconductors. Since the holes are equal in number to the electrons in the empty band about half the current in an intrinsic semiconductor is carried by positive holes.

Many room temperature semiconductors owe their conductivity to the presence of impurities. To see how this comes about let us consider a substance which would ordinarily be an insulator, but in which are dispersed a few impurity atoms. Figure 1-9 illustrates such a case for silicon with a phosphorus impurity atom substituted for one of the silicon atoms in the lattice. Silicon is tetravalent, and each atom is held in its place by four covalent electron-pair bonds to its four nearest neighbors. These valence bond electrons form the valence band in the band picture, and the band is

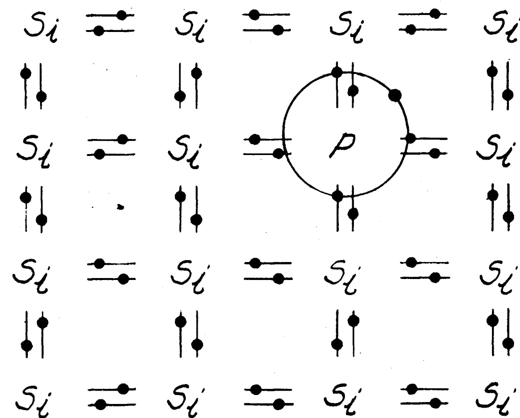


Figure 1-9 Diagram of a Silicon Crystal Lattice with an Atom of Phosphorus Impurity

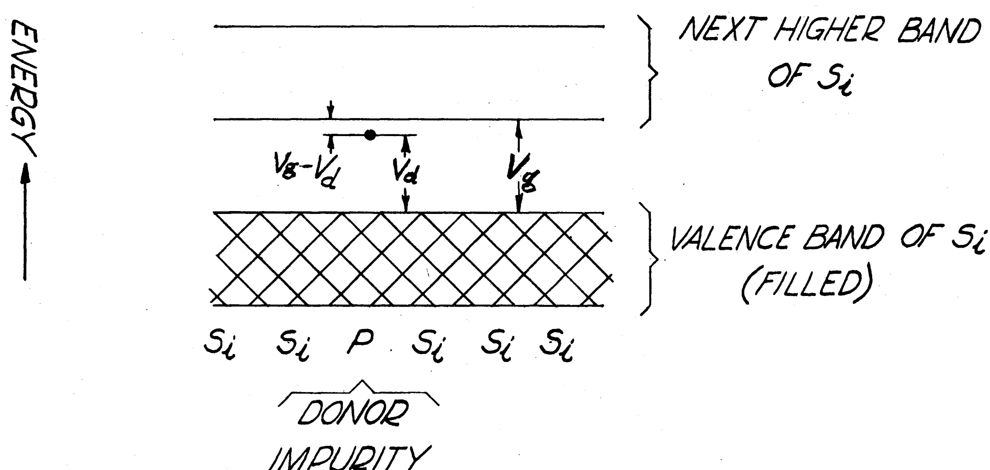


Figure 1-10 Band Scheme of Silicon, Showing Localized Electron Level in the Vicinity of the Phosphorus Impurity Atom
This level may easily give up its electron to the upper band of the crystal, thus endowing the crystal with conductivity.

completely filled by the contribution of four valence electrons from each atom, Figure 1-10 . The phosphorus atom, being pentavalent, has one electron more than is necessary to complete the bonding requirements of the crystal lattice. This extra electron is loosely held in the vicinity of the phosphorus atom by coulomb forces. If the temperature is high enough this electron can be released from the neighborhood of the impurity atom by thermal activation to the bottom of the empty band of the silicon lattice. In the band scheme of 1-10 , the extra electron of the phosphorus impurity is supposed to reside on a localized phosphorus energy level a short distance below the bottom of the empty band of the silicon lattice. Only a small amount of thermal energy is required to transfer it to the empty band, and such an impurity semiconductor will have considerable conductivity at room temperature. The number n_e of such activated electrons existing in statistical equilibrium depends on the density N_d of phosphorus impurities, the energy $e(V_g-V_d)$ for liberation, and the absolute temperature T according to the relationship $n_e \sim \sqrt{N_d} \exp\left(-\frac{e(V_g-V_d)}{2kT}\right)$, where k is Boltzmann's constant. Because of their nature and function the phosphorus atoms are called donor impurities. Semiconductors which owe their conductivity predominantly to the transport of electrons in an otherwise empty band are called n type semiconductors (n for negative carriers).

Figure 1-11 shows an aluminum atom present as an impurity in a silicon crystal lattice. Aluminum, being trivalent, has one electron too few to satisfy the valence bonding requirements of the lattice. Under proper conditions of thermal activation the aluminum atom may rob from one of its neighbors the electron needed to complete its own bonding. That neighbor, in turn, robs one of its neighbors, and the electron deficit can be handed along from silicon to silicon. In the band picture the valence bond which lacks an electron is thought of as a localized energy

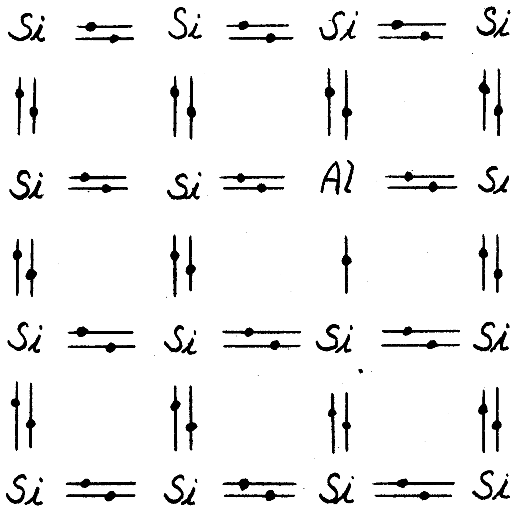


Figure 1-11 Diagram of a Silicon Crystal Lattice with an Atom of Aluminum Impurity

level associated with the aluminum impurity, situated a short distance V_a above the top of the filled band, (Figure 1-12). If the temperature is high enough an electron from the filled band may be thermally activated to occupy this energy level, leaving behind a hole in the filled band. Such a hole can act as a conductor of net current. An impurity which can thus receive electrons from a filled band is called an acceptor, and a semiconductor which owes its conductivity predominantly to the presence of holes in the filled band is called a p-type semiconductor (p for positive carriers).

Some substances can be either p type or n type, depending on which impurity predominates. In fact, both silicon and germanium can be prepared at will to have either desired type of conductivity. Cuprous oxide is always found as a p-type semiconductor. In this case the acceptor impurities are believed to be vacant copper lattice sites. Selenium, too, is always p type. Here the acceptors are believed to be the unsaturated ends of polymeric chains of selenium atoms in the crystal.

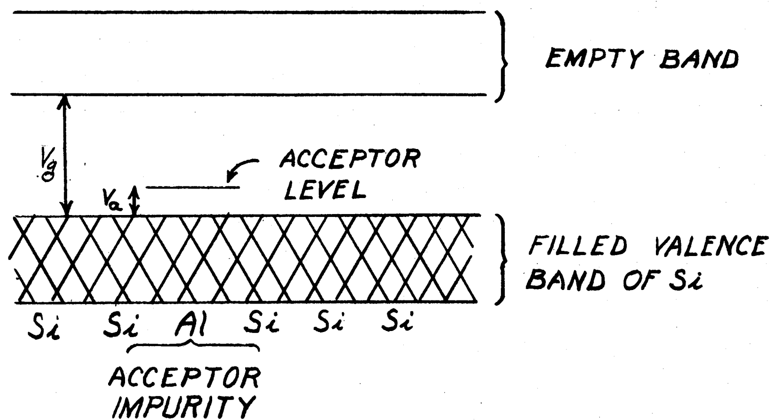


Figure 1-12 Band Scheme of Silicon, Showing Localized Electron Level in the Vicinity of the Aluminum Impurity Atom
This level is normally empty, but it can easily accept an electron from the valence band of the crystal, leaving a conductivity hole.

The picture just developed explains the following experimental facts which otherwise are difficult to interpret:

1. N-type oxides, such as ZnO, when heated in a neutral or slightly reducing atmosphere become good conductors, presumably because they contain excess zinc which can donate electrons. If they then are heated in atmospheres which are increasingly more oxidizing, their conductivity decreases until eventually they are intrinsic semiconductors or insulators.
2. P-type oxides, such as NiO, when heat treated in strongly oxidizing atmospheres are good conductors. Very likely they contain oxygen in excess of the stoichiometric proportion, and this oxygen acts as an acceptor impurity. When these semiconductors are heated in less oxidizing or neutral atmospheres they become poorer conductors, semiconductors, or insulators.
3. When a p-type oxide is sintered with another p-type oxide, the conductivity increases. Similarly for two n-type oxides. But when a p type is added to an n type the conductivity decreases.
4. If a metal forms several oxides the one in which the metal exerts its highest valence is n type, while the one in which it exerts its lowest valence will be p type.

On the basis of the classical theory of conduction in which the current was carried by free negatively charged electrons, the predicted sign of the Hall constant, R_H and the thermoelectric power, Q, was negative. The frequently observed positive values of R_H and Q were considered anomalous, and no satisfactory explanation could be given. It was also difficult to explain the small observed values of R_H and Q for metals and for intrinsic semiconductors. Both of these difficulties were removed by the band theory which introduced the concept of positively charged hole carriers. In p-type semiconductors where the holes predominate, R_H and Q are positive. In intrinsic semiconductors in which electron and hole carriers exist in nearly equal concentrations, R_H and Q are small because the effects of the holes cancel those of the electrons. This cancellation is not complete because the effective masses and the mobilities of the holes are not ordinarily quite equal to those of the electrons. For a more complete description of the band theory the reader is referred to Torrey and Whitmer, "Crystal Rectifiers;" F. Seitz, "The Modern Theory of Solids;" A. H. Wilson, "The Theory of Metals," and "Metals and Semiconductors;" and N. F. Mott and R. W. Gurney, "Electronic Processes in Ionic Crystals."

Figure 1-13 shows once more the plots of log conductivity versus 1/T for several samples of germanium having different amounts of deliberately-added antimony. Antimony acts as a donor, imparting n-type conductivity to the germanium. From the foregoing developments we can explain the qualitative shapes of these curves. Since the number of current carriers, and hence the conductivity σ, for both intrinsic and impurity semiconductors, depends exponentially on the reciprocal

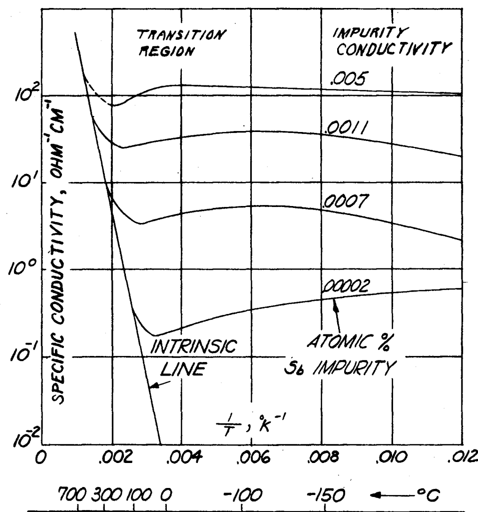


Figure 1-13 Log Conductivity versus 1/T for Four Germanium Specimens with Different Concentrations of Antimony Impurity

of the absolute temperature, it is evident why these curves are plotted on co-ordinates of log σ versus 1/T. At very low temperatures the conductivity is entirely due to electrons thermally activated from the donor impurity levels. As the temperature increases, the number of such electrons increases exponentially, giving the nearly straight line section marked "impurity" conductivity in the figure. As the temperature is increased still more the donors become exhausted, and the conductivity decreases because the increasing thermal disorder of the lattice reduces the mean free path of the conduction electrons. This section of the conductivity curve is designated "transition" region. Finally, at still higher temperatures, electrons become thermally activated in increasing numbers from the filled band, and the conductivity rises again along another straight line section marked "intrinsic line." In such semiconductors, the impurity conductivity dominates at low temperatures, while the intrinsic conductivity domi temperatures, with the log σ versus 1/T curve approximately straight regions. For either of these regions we may write:

$$\sigma = \sigma_\infty \, \varepsilon^{\frac{-B}{T}}$$

or

$$\log_{10} \sigma = \log_{10} \sigma_\infty - \frac{B}{2.3} T \quad ,$$

where log σ_∞ is the intercept and -B/2.3 is the slope of the straight lines. We now proceed to relate these two empirical constants with theory to see what can be deduced from them.

For an <u>intrinsic</u> semiconductor for which V_g is independent of T, statistical theory gives:

$$\sigma = 2UT^{\frac{3}{2}} e(v_e + v_h) \, \varepsilon^{\frac{-V_g}{2} \frac{e}{k} \frac{1}{T}} \quad , \qquad (1\text{-}1)$$

where U is a universal constant whose value is 2.4×10^{15}; $e = 1.6 \times 10^{-19}$ coulombs; e/k = 11600 degrees C per volt; and v_e and v_h are the mobilities of electrons and holes. Consequently, should we plot $\log_{10} (\sigma/T^{3/2})$ versus 1/T, then if $(v_e + v_h)$ is

independent of T or varies very slowly with T, we should obtain a straight line whose slope $= -\frac{1}{2.3}\frac{V_g}{2}\frac{e}{k}$ and whose intercept = log $[2Ue\ (v_e + v_h)]$. We should thus be able to deduce values of V_g and of $(v_e + v_h)$ from σ versus T data. When this is done, it is found that values of V_g agree with values deduced from other types of experiments but values of $(v_e + v_h)$ differ by factors of 10 or 100 or more from those deduced from other experiments.

We, therefore, aim to modify the theory to get agreement. There are theoretical and experimental reasons for thinking that V_g depends on T. Suppose we assume that

$$V_g = V_{go} + c\ T \qquad (1\text{-}2)$$

where c is considered a constant whose dimensions are volts/°C. Substitute equation 1-2 in 1-1

$$\sigma = 2UT^{\frac{3}{2}}\ e(v_e + v_h)\ \varepsilon^{-\frac{c}{2}\frac{e}{k} - \frac{V_{go}}{2}\frac{e}{k}\frac{1}{T}}$$

or

$$\log_{10}\left(\frac{\sigma}{T^{\frac{3}{2}}}\right) = \log_{10}\ [2U\ e(v_e + v_h)] - \frac{c}{4.6}\frac{e}{k} - \left(\frac{V_{go}}{4.6}\frac{e}{k}\right)\frac{1}{T}\ . \qquad (1\text{-}3)$$

On these assumptions a plot of log $(\sigma/T^{\frac{3}{2}})$ versus 1/T should be a straight line whose slope is

$$\frac{-V_{go}}{4.6}\frac{e}{k}$$

and whose intercept is

$$\log\ 2U\ e(v_e + v_h) - \frac{c}{4.6}\frac{e}{k}\ .$$

Since this contains two unknowns, namely $(v_e + v_h)$ and c we cannot evaluate either of them. However, if we know one of them from another experiment we can then calculate the other unknown. Thus if we can deduce values of $(v_e + v_h)$ from Hall effect experiments we can then deduce values of c. We can thus determine how V_g varies with T.

To see how important the new term $-\frac{c}{4.6}$ 11600 may be let us suppose that $c = 10^{-3}$ volts/°C. (This means that when T = 300, V_g is 0.3 volt smaller than when T = 0). Then $\frac{c}{4.6}\ 11600 = \frac{11.6}{4.6} = 2.5$. This means that the intercept is 2.5 decades or a factor of 700 higher than it would have been if V_g had been independent of T.

In the impurity region at sufficiently low temperatures the log σ versus 1/T lines are also approximately straight. In this region, theory again predicts a simple equation relating the parameters of the plot lines to the physical constants of the semiconductor. For an n-type semiconductor

$$\sigma = (2UT^{\frac{3}{2}} N_d)^{\frac{1}{2}} e\, v_e\, \varepsilon^{\frac{-(V_g-V_d)e}{2kT}}$$

or

$$\log_{10} \sigma = \frac{1}{2} \log_{10} (2UT^{\frac{3}{2}} N_d) + \log_{10} (e\, v_e) \frac{-(V_g-V_d)e}{2kT} \frac{1}{2.3}, \quad (1\text{-}4)$$

where $(V_g-V_d)e$ is the activation required to promote a donor electron to the bottom of the conduction band (Figure 1-10) and N_d is the number of such donor impurity levels per cm^3. This equation holds only if there are no acceptor impurities active to complicate the conductivity picture, and if the temperature is so low that no electrons are activated from the filled band.

The intercept now depends on $(N_d)^{\frac{1}{2}}$ and on v_e. Hence, if v_e can be ascertained from the intrinsic region or from independent experiments, N_d can be determined. The slope of the line determines V_d.

A similar equation holds if there are normally filled acceptor levels above the filled band but no donor levels (Figure 1-12).

Table 1-1 summarizes the values of V_{go}, the width of the unallowed gap in volts for 0°K, for a few semiconductors.

Values of V_{go}, width of unallowed gap for 0°K, expressed in volts.

Si	Ge	Cu_2O	ZnO	Fe_2O_3	$Ni_1 Mn_4 O_{6.4}$
1.2	.88	1.4 or 1.8	2.2	2.3	.68

Table 1-1

In Figures 1-10 and 1-12 , the donor and acceptor impurity levels were represented as discrete, localized levels like those of an isolated atom. If the impurity concentration in a particular semiconductor specimen is "reasonably" large (one per cent or more), the impurity atoms will be close enough together in the lattice so that a small amount of wave-mechanical coupling may exist between neighboring impurity atoms. The isolated impurity levels will then broaden out into narrow bands. Because the impurities are still comparatively far apart, however, electrons can jump from one impurity atom to another only with difficulty, and the

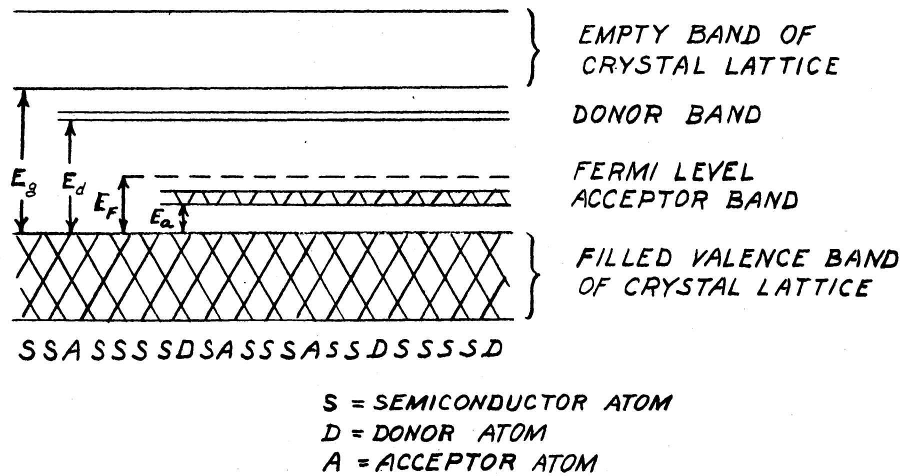


Figure 1-14 Band Scheme for A Semiconductor Containing Both Donor and Acceptor Impurities
The impurities are supposed present in high enough concentrations so that the donor and acceptor levels widen out into narrow bands.

mobility of electrons in these impurity bands is many times lower than in the bands of the matrix crystal. To the best of the author's knowledge no cases are known where conductivity within such impurity bands has ever been observed and identified in the presence of the higher conductivity in the bands of the crystal.

Figure 1-14 shows a complete band picture for a semiconductor (S) containing atoms of both acceptor (A) and donor (D) type impurities. E_g is the energy gap between the filled and empty band; E_d is the energy of the donors above the top of the filled band; E_a is the energy of the acceptors above the top of the filled band. E_g, E_d, and E_a are energies in joules; they are customarily expressed in equivalent volts, V_g, V_d, V_a by which is meant that $E_g = V_g e$; $E_d = V_d e$; $E_a = V_a e$; $E_F = V_F e$. N_d = number of donors/cm^3; N_a = number of acceptors/cm^3. The figure also shows a line for E_F, the energy of the Fermi level. This Fermi level plays an important part in the theory of the solid state which will be discussed in the following chapter.

LECTURE 2

THEORY OF THERMISTORS

See — "Properties and Uses of Thermistors", J. A. Becker, C. B. Green, and G. L. Pearson. Published in the Bell System Technical Journal, Vol. 26, pp. 170-212, January, 1947; and in the Electrical Engineering Vol. 65, Transactions, pp. 711-725, November, 1946. Or see — Bell Telephone System Technical Publications, Monograph B-1443.

LECTURE 3

THE FERMI LEVEL, ITS IMPORTANCE AND HOW IT IS DETERMINED *

It is a fundamental law of nature that a system adjusts itself in the direction of lowest total energy. Thus water is found in the bottom of a container, charge flows along a conductor to a region of lower potential, and a hot body cools down to the temperature of its surroundings. In response to the same general law the electrons in the band system of a solid try to occupy the lowest energy levels available to them. We have already stated that in a band having fewer electrons than are required to fill it completely the electrons will fill the band from the bottom upward as far as their numbers permit, with a slight fuzzing out of the top of the distribution. We wish now to inquire about the detailed mechanics of the filling of energy states.

On the basis of quantum mechanics, Fermi showed that in any system of atoms, ions, and electrons, the probability, P, of finding an electron in an energy state of total energy E is given by the equation

$$P = \frac{1}{1+ \varepsilon^{(E-E_F)/kT}} = \frac{1}{1+ \varepsilon^{e(V-V_F)/kT}} \qquad (3\text{-}1)$$

in which k is Boltzman's constant (1.38×10^{-23} joules/°C) and T is the absolute temperature of the system. E_F is that value of energy, and V_F its equivalent voltage, at which the probability of occupancy is 1/2. When $T = 0$, all energy states below V_F are occupied by electrons; all states above V_F are vacant. Since V and V_F occur only as $(V - V_F)$ it makes no difference where the zero of energy is taken. E_F or V_F is called the Fermi level; it is also called the thermodynamic potential; it has also been called the electrochemical potential. **Fig. 3-1** shows a plot of P vs. $V - V_F$ for $T = 0°$,

* The authors gratefully acknowledge the assistance of John Bardeen in formulating the concepts in this chapter.

300°, 1000° and 1500°K. From this plot and the form of equation 3-1, it is apparent that P is symmetrical about $V - V_F = 0$; this means that the value of P for any value of $V - V_F$ is equal to the value of 1-P for $-(V - V_F)$. The quantity 1 - P is the probability that an energy level will be unoccupied, which means that this level contains a hole.

There are a number of reasons why the Fermi Level is important: (1) When a metal and a semiconductor are electrically connected and if initally they were at different electrical potentials, charges will flow from one to another. This changes the potential of one with respect to the other; the process continues until the potential energy of the carriers in the metal is the same as the potential energy of the carriers in the semiconductor. They are then in equilibrium and the Fermi levels are at the same potential. (2) We shall see below that the concentration of electrons in the empty band or of the holes in the filled band are rather simply related to the Fermi level. (See equations (3-2c) and (3-2d) below.) (3) In P type semiconductors the thermoelectric power Q is V_F/T; in N type semiconductors $Q = (V_g - V_F)/T$.* (4) The Fermi level plays an important part in the theory of rectifiers between a metal and a semiconductor.

For these and other reasons it is important to present a method for locating the Fermi level in a semiconductor which has one or more filled bands, an empty but allowed band, an unallowed band, a very narrow band of donors, and a very narrow band of acceptors. Such a band picture is shown in Fig. 3-2. The method for locating the Fermi level in such a system is based on the law of the conservation of charge. This means that the total number of electrons in the system at any temperature T does not change with T and is equal to the number of electrons contributed to the system by the donors and by the filled band.

To express this conservation law in the form of an equation, let us define symbols for the number of electrons which exist at any temperature in the various permitted energy states shown in Fig. 3-2. Let

n_e = number of electrons/cm^3 in the nominally empty or conduction band.

n_D = number of electrons/cm^3 in the donor level or narrow band

n_A = number of electrons/cm^3 in the acceptor level or narrow band

---

* See Fig. 3-2 for definition of V_g.

n_{fb} = number of electrons/cm^3 in the filled band
n_h = number of holes/cm^3 in the filled band
N_d = number of donors/cm^3
N_a = number of acceptors/cm^3
N_{fb} = number of energy states/cm^3 in the filled band.

The conservation law then states that

$$n_e + n_D + n_A + n_{fb} = N_d + N_{fb}$$

Since $N_{fb} - n_{fb} = n_h$ (this follows from the definition of n_h)

$$n_e + n_D + n_A = N_d + n_h \tag{3-2}$$

It now is necessary to express n_D, n_A, n_e and n_h as functions of the Fermi level V_F. From Eq. (3-1) it follows that:

$$n_D = \frac{N_d}{1 + \varepsilon^{e(V_d - V_F)/kT}} \tag{3-2a}$$

and

$$n_A = \frac{N_a}{1 + \varepsilon^{e(V_a - V_F)/kT}} \tag{3-2b}$$

To obtain an expression for n_e, we must take an interval of voltage dV in the empty band, multiply the number of states in this interval by the probability P given by Eq. (3-1), and integrate for all intervals dV in the empty band. This is done in standard texts on Statistics or Quantum Mechanics. (See e.g. Seitz's book) The result is

$$n_e = 2\left(\frac{2\pi m_e k}{h^2}\right)^{3/2} T^{3/2} \varepsilon^{-e(V_g - V_F)/kT}$$

$$= 2\, U_e\, T^{3/2}\, \varepsilon^{-e(V_g - V_F)/kT} \tag{3-2c}$$

and similarly for holes in the filled band:

$$n_h = 2\,U_h\,T^{3/2}\ \varepsilon^{-eV_F/kT} \qquad (3\text{-}2d)$$

where

$$U_e = \left(\frac{2\pi m_e k}{h^2}\right)^{3/2}$$

$$U_h = \left(\frac{2\pi m_h k}{h^2}\right)^{3/2}$$

m_e = effective mass of electrons in empty band

m_h = effective mass of holes in filled band

k = Boltzman's constant

h = Plank's constant

If m_e and m_h are assumed equal to the normal mass of the electron or 9.11×10^{-28} grams

$$U_e = U_h = 2.42 \times 10^{15}\ \mathrm{cm}^{-3}\ {}^{\circ}\mathrm{K}$$

If there are more than one kind of donors or of acceptors, the terms involving N_d or N_a must be summed.

Depending upon the nature of the donor or acceptor atoms, n_D and n_A in equations (3-2a) and (3-2b) may have to be "weighted". This means that the exponential term in the denominator of Eq. (3-2a) may have to be multiplied by 2 and the corresponding exponential term in Eq. (3-2b) may have to be multiplied by 1/2. For simplicity we have neglected these factors.

In the derivation of the expressions for n_e and n_h, (Eqs. 3-2c and 3-2d) it has been assumed that the values of n_e and n_h are appreciably less than $4.84 \times 10^{15}\ T^{3/2}$. This means that classical statistics were assumed in the derivation, or that the 1 in the denominator of Eq. (3-1) was negligible. Hence Eq. (3-2) is applicable only to cases in which this assumption is fulfilled. For M type and for I type semiconductors with high values of n_e and n_h more complicated expressions must be used.

Next we consider the utility of Eq. (3-2). In a system in which T, N_d, N_a, V_d, V_a and V_g are known, V is the only unknown and hence in such

a system V_F will adjust itself until Eq. (3-2) is satisfied. To solve for V_F explicitly is a difficult matter, but in certain special cases the solution is relatively simple. We will consider a number of such cases.

CASE 1. There are no donors or acceptors.

Hence

$$N_d = 0 \text{ and } N_a = 0 \text{ so that}$$

$$n_e = n_h \quad \text{or} \quad U_e \, \varepsilon^{-e(V_g - V_F)/kT} = U_h \, \varepsilon^{-e V_F/kT}$$

and

$$\varepsilon^{2eV_F/kT} = \frac{U_h}{U_e} \, \varepsilon^{eV_g/kT}$$

or

$$V_F = \frac{V_g}{2} + \frac{3}{2} \cdot \frac{2.3}{2} \cdot \frac{kT}{e} \log_{10} \frac{m_h}{m_e} \tag{3-3}$$

If

$$m_h = m_e = m$$

$$V_F = V_g/2 \tag{3-3a}$$

This says that the Fermi level will be at the middle of the unallowed band for any T. V_F may still vary slowly with T if V_g depends slightly on T.

It is now easy to solve for n_e and for n_h by substitution in Eqs. (3-2c) and (3-2d).

$$n_e = n_h = 2U \; T^{3/2} \, \varepsilon^{-e V_g/2kT} \tag{3-4}$$

CASE 2. There are no donors and T is low enough so that there are practically no electrons in the empty band.

Hence

$$N_d = 0 \text{ and } n_e = 0$$

Equation (3-2) then reduces to

$$\frac{N_a}{1 + \varepsilon^{e(V_a - V_F)/kT}} = n_h = 2U_h T^{3/2} \, \varepsilon^{-eV_F/kT}$$

$$N_a = n_h \left(1 + \frac{2U_h T^{3/2} \varepsilon^{-e V_F/kT}}{2U_h T^{3/2} \varepsilon^{-e V_a/kT}}\right) = n_h \left(1 + \frac{n_h}{n_a}\right) \qquad (3\text{-}5)$$

where

n_a is defined by $n_a \equiv 2U_h T^{3/2} \varepsilon^{-eV_a/kT}$ (3-5a)

Here n_a has the physical significance of being the number of holes per cc there would be in the filled band if the Fermi level came exactly at the acceptor level. At any temperature $n_a = n_h$ when $V_F = V_a$. It is a known quantity since U_h, T and V_a are known.

Write Eq. (3-5) as

$$\frac{n_h^2}{n_a} + n_h - N_a = 0 \qquad (3\text{-}5b)$$

whose solution is

$$n_h = \frac{-1 \pm (1 + 4\, N_a/n_a)^{1/2}}{(2/n_a)} = \frac{n_a}{2}\left[-1 \pm (1 + 4\, N_a/n_a)^{1/2}\right]$$

or

$$\frac{n_h}{n_a} = (1/2)\left[-1 + \left(1 + 4\,\frac{N_a}{n_a}\right)^{1/2}\right] \qquad (3\text{-}6)$$

Only the + sign in $\pm$ is significant since n_h must be a positive quantity. Since n_a can be calculated for any T, n_h can also be determined.

The solution of (3-6) takes on a simpler form in two limiting conditions:

(1) when $(N_a/n_a) > 100$ because N_a is large or T is small. In this case

$$n_h = (N_a n_a)^{1/2} = (2U)^{1/2}\, T^{3/4}\, (N_a)^{1/2}\, \varepsilon^{-eV_a/2kT} \qquad (3\text{-}6a)$$

In this case only a small fraction of acceptors are filled.

(2) when $(N_a/n_a) < 0.1$ because N_a is small or T is large. In this case

$$n_h = \frac{n_a}{2}\left(-1 + 1 + 2\frac{N_a}{n_a} - 2\left(\frac{N_a}{n_a}\right)^2 + \ldots\right)$$

$$= N_a\left(1 - \frac{N_a}{n_a} + \ldots\right) \qquad (3\text{-}6b)$$

In this case nearly all the acceptors are filled. They are said to be saturated.

Since n_h can be calculated from (3-6) it is then simple to evaluate $\varepsilon^{-eV_F/kT}$ or V_F from Eq. (3-2d). It is also possible to derive analytical expressions for these functions.

$$\varepsilon^{-eV_F/kT} = (1/2)\, \varepsilon^{eV_a/kT}\left[-1 + (1 + 4\, N_a/n_a)^{1/2}\right]. \qquad (3\text{-}7)$$

When $\frac{N_a}{n_a} > 100$,

$$\varepsilon^{-eV_F/kT} = (N_a/2UT^{3/2})^{1/2} \varepsilon^{-eV_a/2kT}$$

and

$$V_F = \frac{V_a}{2} + \frac{2.3\ kT}{2\ e} \ (\log 2U + (3/2) \log T - \log N_a). \qquad (3\text{-}7a)$$

When $(N_a/n_a) < 0.1$,

$$V_F = 2.3 \frac{kT}{e}[\log 2U + (3/2) \log T - \log N_a - \log (1 - N_a/n_a)] \qquad (3\text{-}7b)$$

We now apply these equations to solve the problem: given a system in which the concentration of acceptors N_a, and their activation energy V_a are specified, compute n_h vs T and then V_F vs T. It will be assumed that the unallowed gap V_g is so large that n_e will be negligible compared to n_h at all temperatures. (Case 5 will remove this restriction.) Specifically take $N_a = 10^{18}$ acceptors/cm^3; V_a = .2 volts; and V_g 3.0 volts. The following is recommended as an expeditious procedure.

Prepare a table such as Table 3-1. Column 1 lists the temperature T for which values of n_h and V_F are to be computed. Column 2 gives values of n_a calculated from Eq. 3-5a or read off from a graph to be described below. (Fig. 3-4). Column 3 lists N_a/n_a. Column 4 lists n_h/n_a calculated from Eq. (3-6) or read off from Fig. 3-3 which is a plot of this equation. Column 5 lists n_h. Column 6 lists 1/T. Column 7 lists V_F calculated from Eq. (3-2d) or read off from a graph which will now be described.

In this and similar problems one deals with an equation of the form

$$n = 2UT^{3/2} \varepsilon^{-\phi e/kT} = 2UT^{3/2}\ 10^{-\phi e/2.3kT} \qquad \text{(General Equation)}$$

in which there are three variables n, T and ϕ; and the constants U, e and k. Two of the variables are specified and the third is to be evaluated. This process is greatly facilitated by plotting n on a logarithmic scale vs ϕ with T as a parameter. Such a plot is shown in Fig. 3-4a and its extension Fig. 3-4b. In these plots n might be n_a if ϕ is V_a; n might be n_e if ϕ is $V_g - V_F$; n might be n_h if ϕ is V_F; etc.

In Table 3-1, column 2 for n_a, is obtained by setting $\phi = V_a = .2$ and reading off values of $n = n_a$ for various values of T. Again for obtaining values of V_F in column 7, the figure is very convenient. By now T and n_h are known. This determines a point on the graph. The value of ϕ for this point can be read off on the x axis and is the value of V_F that satisfies Eq. 3-2d.

In Fig. 3-5, the solid line curve marked $N_a = 10^{18}$ is a plot of log n_h vs 1/T for the data in Table 3-1. The other solid lines in the figure are similarly calculated for other values of N_a. The lines marked n_h and n_e near the intrinsic line n_i will be dealt with under Case 5. The long dashed line is for n_a.

In Fig. 3-6, the solid line curve marked $N_a = 10^{18}$ is a plot of V_F vs T from the data in Table 3-1. The other solid lines are for other values of N_a. The dashed lines at high temperatures show how V_F varies with T if V_g is not indefinitely large. They will be discussed under Case 5.

Case 3. There are no acceptors and no holes in the filled band.

This case is similar to Case 2 and the formulas are similar. Eq. 3-5a becomes $n_d \equiv 2U_e T^{3/2} \varepsilon^{-e(V_g - V_d)/kT}$ and Eq. 3-6 becomes

$$\frac{n_e}{n_d} = 1/2 \left[-1 + \left(1 + 4\frac{N_d}{n_d}\right)^{1/2}\right].$$

Case 4. Both donors and acceptors are present but $N_a > N_d$ and T is low enough so that $n_e = 0$.

Eq. 3-2 becomes

$$\frac{N_d}{1 + \varepsilon^{e(V_d - V_F)/kT}} + \frac{N_a}{1 + \varepsilon^{e(V_a - V_F)/kT}} = N_d + n_h \qquad (3\text{-}8)$$

The first term can be neglected compared to N_d because the Fermi level must be in the lower half of the unallowed band. Hence $e(V_d - V_F)kT$ will be large and the denominator in the first term will be a large number.

Introducing the quantity n_a and proceeding as in Case 2, one obtains:

$$(n_h)^2 + (N_d + n_a)\, n_h - n_a\,(N_a - N_d) = 0 \qquad (3\text{-}9)$$

whose solution is

$$n_h = \frac{N_d + n_a}{2}\left[-1 + \left(1 + \frac{4n_a\,(N_a - N_d)}{(N_d + n_a)^2}\right)^{1/2}\right] \qquad (3\text{-}10)$$

Here again various simplifications can be made depending on whether n_a is greater than N_a, is between N_a and N_d, or is less than N_d. However, these approximations hold over such short temperature ranges that they give a misleading impression.

A better procedure is to calculate n_a vs T, or use Fig. 3-4, put in values in Eq. (3-10), assign values to N_a and N_d, and calculate n_h. Then use Eq. (3-2d) to calculate V_F, or use Fig. 3-4 to obtain V_F. In this case the small number of donors fill up an equal number of acceptors. As the concentration of donors increases, the conductivity decreases. Another interesting feature of this case is that for sufficiently low temperatures, $V_F = V_a$ instead of $V_a/2$. This effect extends over a longer temperature range if N_d is nearly equal to $N_a/2$.

CASE 5 No donors but no restribtions on T.

At low temperatures this case reduces to Case 2; at very high temperatures it approaches Case 1. At intermediate temperatures, where the conductivity is about to become intrinsic, electrons are present in the empty band, and Eq. (3-2) becomes

$$n_e + \frac{N_a}{1 + \varepsilon^{e(V_a - V_F)/kT}} = n_h \tag{3-11}$$

which can be written as

$$\frac{n_h^2}{n_a} + n_h - N_a = n_e \tag{3-12}$$

where as before

$$n_a \equiv 2UT^{3/2} \varepsilon^{-eV_a/kT} \tag{3-12a}$$

A good procedure to evaluate n_h as a function of T will be illustrated by considering a specific case for which $N_a = 10^{16}$ $E_a = 0.2 \times 1.6 \times 10^{-19}$ joules or $V_a = 0.2$ volts

$E_g = 2.0 \times 1.6 \times 10^{-19}$ joules

or $V_g = 2.0$ volts

Make a plot of log y vs 1/T where y may be n_a, N_a, n_h or n_e. Such a plot is shown in Fig. 3-5. The lines for N_a and n_a can be drawn directly since N_a is $= 10^{16}$ and n_a is given by Eq. (3-12a). At very high Ts, in the intrinsic range,

$$n_e = n_h = 2UT^{3/2} \varepsilon^{-\frac{2.0}{2} \times \frac{11600}{T}} = n_i$$

At lower Ts, n_e will always be below this line and n_h always above it.

Below some T, (800° K in this case) n_e will be very small compared to n_h^2/n_a, N_a and n_h. Eq. (3-12) reduces to Eq. (3-5b).

For this case the curve for n_h can be quickly plotted by noting the following:

At very low temperatures n_h is small compared with N_a, so that Eq. (3-12) reduces to $n_h^2 = n_a N_a$ or $n_h = (n_a N_a)^{1/2}$. This means that log n_h vs T is a line

which passes through the intersection of n_a with N_a and has a slope of half of that for n_a. On a log plot this means that at any T, n_h is midway between N_a and n_a.

At the T for which $n_h = n_a$, Eq. (3-12) becomes $2n_h = N_a$ or $n_h = N_a/2$.

At a T at which $n_a = N_a$, $V_F = V_a$ and

$$n_h = \frac{N_a}{2}(-1+\sqrt{5}) = \frac{1.24}{2} N_a = .62\ N_a$$

At somewhat higher temperatures, n_h approaches N_a.

At still higher T where n_e is appreciable compared with n_h we must use the complete Eq. 3-12, namely:

$$\frac{n_h^2}{n_a} + n_h - N_a = n_e \qquad 3\text{-}12)$$

or

$$n_h \left(1 + \frac{n_h}{n_a}\right) - N_a -$$

Now

$$\frac{n_h}{n_a} = \frac{2UT^{3/2}\,\varepsilon^{-V_F\ e/kT}}{2UT^{3}\,\varepsilon^{-V_a\ e/kT}} = \varepsilon^{-(V_F - V_a)\ e/kT} = 10^{-(V_F - V_a)\ 11600/2.3T}$$

$$= 10^{-(V_F - V_a)\ 5050/T}$$

Since T will be about 1000 and $V_F - V_a$ will be about 0.3, $\frac{n_h}{n_a}$ = about $10^{-1.5}$ = about $\frac{1}{30}$ or smaller. In general this is negligible compared to 1 and hence

$$n_h - N_a - n_e = 0 \qquad 3\text{-}13)$$

Define $n_i \equiv 2UT^{3/2}\,\varepsilon^{-\frac{V_g}{2}\ e/kT}$ (3-14)

n_i is the value of n_h or n_e that would prevail in an intrinsic semiconductor with an energy gap V_g volts.

Then

$$\frac{n_h}{n_i} - \frac{N_a}{n_i} - \frac{n_e}{n_i} = 0 \qquad 3\text{-}15)$$

Now

$$\frac{n_h}{n_i} = \frac{2UT^{3/2}\,\varepsilon^{-V_F e/kT}}{2UT^{3/2}\,\varepsilon^{-V_g e/2kT}} = \varepsilon^{\left(\frac{V_g}{2} - V_F\right)e/kT} \qquad (3\text{-}16)$$

and

$$\frac{n_e}{n_i} = \frac{2UT^{3/2}\,\varepsilon^{-(V_g - V_F)e/kT}}{2UT^{3/2}\,\varepsilon^{-V_g e/2kT}} = \varepsilon^{-\left(\frac{V_g}{2} - V_F\right)e/kT} \qquad (3\text{-}17)$$

Hence
$$n_h/n_i = n_i/n_e \tag{3-18}$$

If y is defined as n_h/n_i,

then
$$y^2 - \frac{N_a}{n_i} y - 1 = 0$$

$$y = \frac{\frac{N_a}{n_i} + \left[\left(\frac{N_a}{n_i}\right)^2 + 4\right]^{1/2}}{2} = \frac{N_a}{2n_i}\left[1 + \left[1+4\left(\frac{n_i}{N_a}\right)^2\right]^{1/2}\right]$$

or
$$\frac{n_h}{N_a} = \frac{1}{2} + \frac{1}{2}\left[1 + 4\left(\frac{n_i}{N_a}\right)^2\right]^{1/2} \tag{3-19}$$

For very low T where $\frac{n_i}{N_a}$ is very small, $n_h = N_a$

For very high T where $\frac{n_i}{N_a}$ is very large, $n_h = n_i$

For intermediate values of T, values of $\frac{n_h}{N_a}$ can be read off from Fig. 3-7 which is a plot of Eq. (3-19).

To evaluate n_h, V_F, and n_e in the range of T where n_e is appreciable compared to n_h we proceed as follows: Prepare a table like Table 3-2. Column 1 lists values of T in this range. The lowest value of T must be such that n_h/n_a is less than .10 as determined in Table 3-1, in this case 900° K. The second column lists 1/T. The third column lists values of n_i read off from Fig. 3-4. The fourth column lists n_i/N_a. The fifth column lists n_h/N_a read off from Fig. 3-7 (Eq. 3-19). The sixth column lists n_h. The seventh column lists V_F read off from Fig. 3-4. The last column lists n_e calculated from Eq. (3-13).

In Figs. 3-5 and 3-6 these quantities or similar ones for other values of N_a have been plotted.

CASE 6

There are two kinds of acceptors: N_a and N_a' with energy levels at V_a and V_a'; V_g is very large compared with V_a or V_a'.

Eq. 3-2 becomes

$$\frac{N_a}{1 + \varepsilon^{e(V_a - V_F)/kT}} + \frac{N_a'}{1 + \varepsilon^{e(V_a' - V_F)/kT}} = n_h \tag{3-20}$$

$$\frac{N_a}{1 + \frac{n_h}{n_a}} + \frac{N_a'}{1 + \frac{n_h}{n_a'}} = n_h \tag{3-21}$$

where $$n_a' = 2UT^{3/2}\ \varepsilon^{-eV_a'/kT} \tag{3-21a}$$

The general solution of Eq. 3-21 is complicated. A simplification results if T is low enough so that the 1 in the denominators can be neglected. Then

$$N_a n_a + N_a' n_a' = n_h^2$$

If T is very large: $n_h = N_a + N_a'$
For low enough T and if $N_a = N_a'$ then Eq. (3-20) becomes
$n_h = (N_a n_a)^{1/2} (1 + \frac{n_a'}{n_a})^{1/2} = (N_a n_a)^{1/2} (1 + \varepsilon^{-e(V_a' - V_a)/kT})^{1/2}$

An especially interesting case to consider is that in which $N_a' \gg N_a$ and $V_a' > V_a$. The contribution to n_h from N_a' and N_a are quite distinct. At very low T the log n_h vs 1/T curve has a slope of $\frac{-V_a}{4.6}\ \frac{e}{k}$; at slightly higher T, n_h saturates at N_a; at somewhat higher T the contribution to n_h from N_a' predominates and log n_h rises rapidly again with a slope of $\frac{-V_a'}{4.6}\ \frac{e}{k}$; at still higher T, log n_h saturates at a value of n_a'.

Still another interesting and important case is to imagine that the acceptor levels form a wide band whose lower edge is V_a above the filled band. Divide the band into sub-bands each ΔV_a wide and containing ΔN_a levels. Then sum up or integrate the contributions due to all the sublevels.

Table 3-1

$N_a = 10^{18}$ $V_a = .2$ $V_g > 3.0$

T	n_a	N_a/n_a	n_h/n_a	n_h	$(1/T)x10^3$	V_F
100	$4x10^8$	$2.5x10^9$	$4.5x10^4$	$1.8x10^{13}$	10	.107
200	10^{14}	10^4	$1.0x10^2$	$1.0x10^{16}$	5	.120
300	10^{16}	10^2	9.5	$.95x10^{17}$	3.33	.141
400	$1.1x10^{17}$	9.1	2.54	$2.8x10^{17}$	2.50	.168
500	$5.2x10^{17}$	1.92	.960	$5.0x10^{17}$	2.00	.202
600	$1.4x10^{18}$	.715	.478	$6.7x10^{17}$	1.67	.238
700	$3.3x10^{18}$	.302	.245	$8.1x10^{17}$	1.43	.283
800	$6.0x10^{18}$	.167	.145	$8.7x10^{17}$	1.25	.330
900	$9.6x10^{18}$	.104	.095	$9.12x10^{17}$	1.11	.384
1000	$1.45x10^{19}$	$6.9x10^{-2}$	.069	$1.0x10^{18}$	1.00	.430
1100	$2.2x10^{19}$	$4.55x10^{-2}$	--	$1.0x10^{18}$	.90	.490
1200	$2.8x10^{19}$	$3.57x10^{-2}$	--	$1.0x10^{18}$	.834	.547
1300	$3.8x10^{19}$	$2.62x10^{-2}$	--	$1.0x10^{18}$	.770	.606
1400	$4.8x10^{19}$	$2.08x10^{-2}$	--	$1.0x10^{18}$	.714	.666
1500	$5.8x10^{19}$	$1.73x10^{-2}$	--	$1.0x10^{18}$	.667	.725

Table 3-2

$N_a = 10^{13}$ $V_a = .2$ $V_g = 1.0$

T	1/Tx10^3	n_i	n_i/N_a	n_h/N_a	n_h	V_F	n_e
900	1.11	2.0x10^{17}	2.0x10^{-1}	1.04	1.04x10^{18}	.372	.04x10^{18}
1000	1.00	4.4x10^{17}	4.4x10^{-1}	1.19	1.19x10^{18}	.413	.19 "
1100	.90	8.8x10^{17}	8.8x10^{-1}	1.52	1.52x10^{18}	.450	.52 "
1200	.834	1.58x10^{18}	1.58x10^{-1}	2.15	2.15x10^{18}	.470	1.15 "
1300	.770	2.60x10^{18}	2.60x10^{-1}	3.08	3.08x10^{18}	.480	2.08 "
1400	.714	4.0x10^{18}	4.00x10^{-1}	4.50	4.50x10^{18}	.485	3.50 "
1500	.667	5.7x10^{18}	5.7x10^{-1}	6.30	6.30x10^{18}	.490	5.30 "

CAPTIONS FOR FIGURES

Fig. 3-1	Plot of the Fermi Probability P as a Function of $V-V_F$ for Four Different Temperatures.
Fig. 3-2	Energy Band Structure of a Typical Semiconductor.
Fig. 3-3	Plot of n_h/N_a vs N_a/n_a From Equation (3-6) and Table 3-1.
Fig. 3-4 (a and b)	Universal Plot of n vs ϕ for the General Equation for Several Values of the Parameter T.
Fig. 3-5	Calculated Values of n_h, n_e, n_a, and n_i vs $\frac{1}{T}$ for the Cases Described in the Figure.
Fig. 3-6	Plots of Fermi Level Voltage V_F vs Temperature for Three Different Values of Acceptor Density.
Fig. 3-7	Plot of n_h/N_a vs n_i/n_a from Equation (3-19).

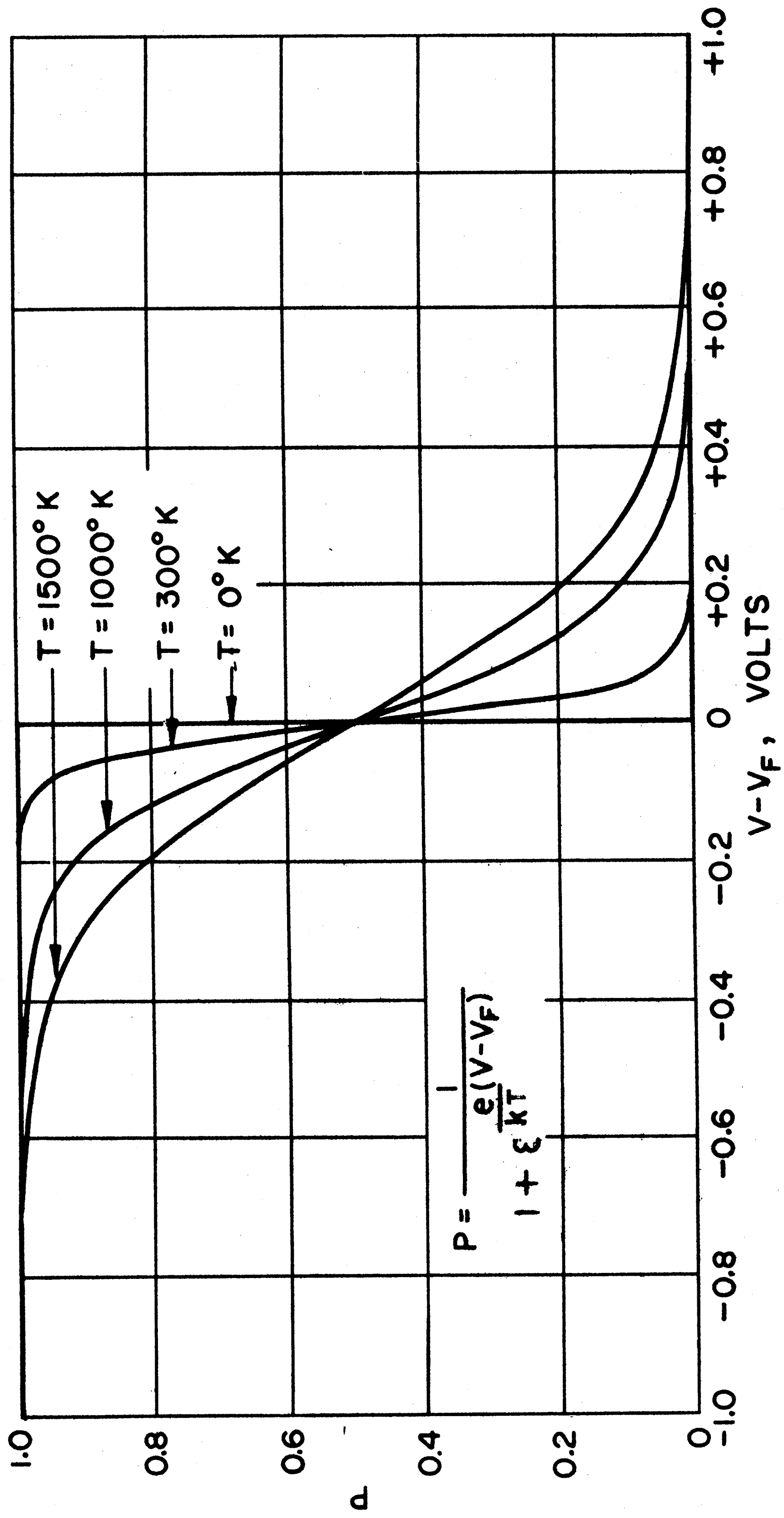
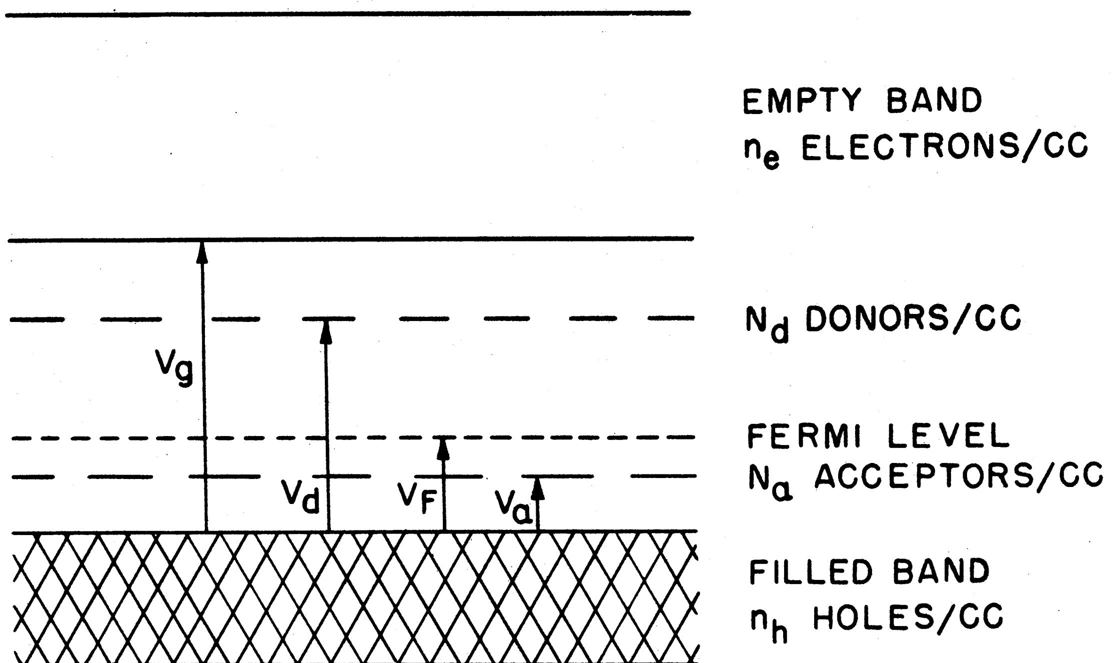


FIG. 3-1



THE POSITION OF THE FERMI LEVEL
DEPENDS ON TEMPERATURE

FIG. 3-2

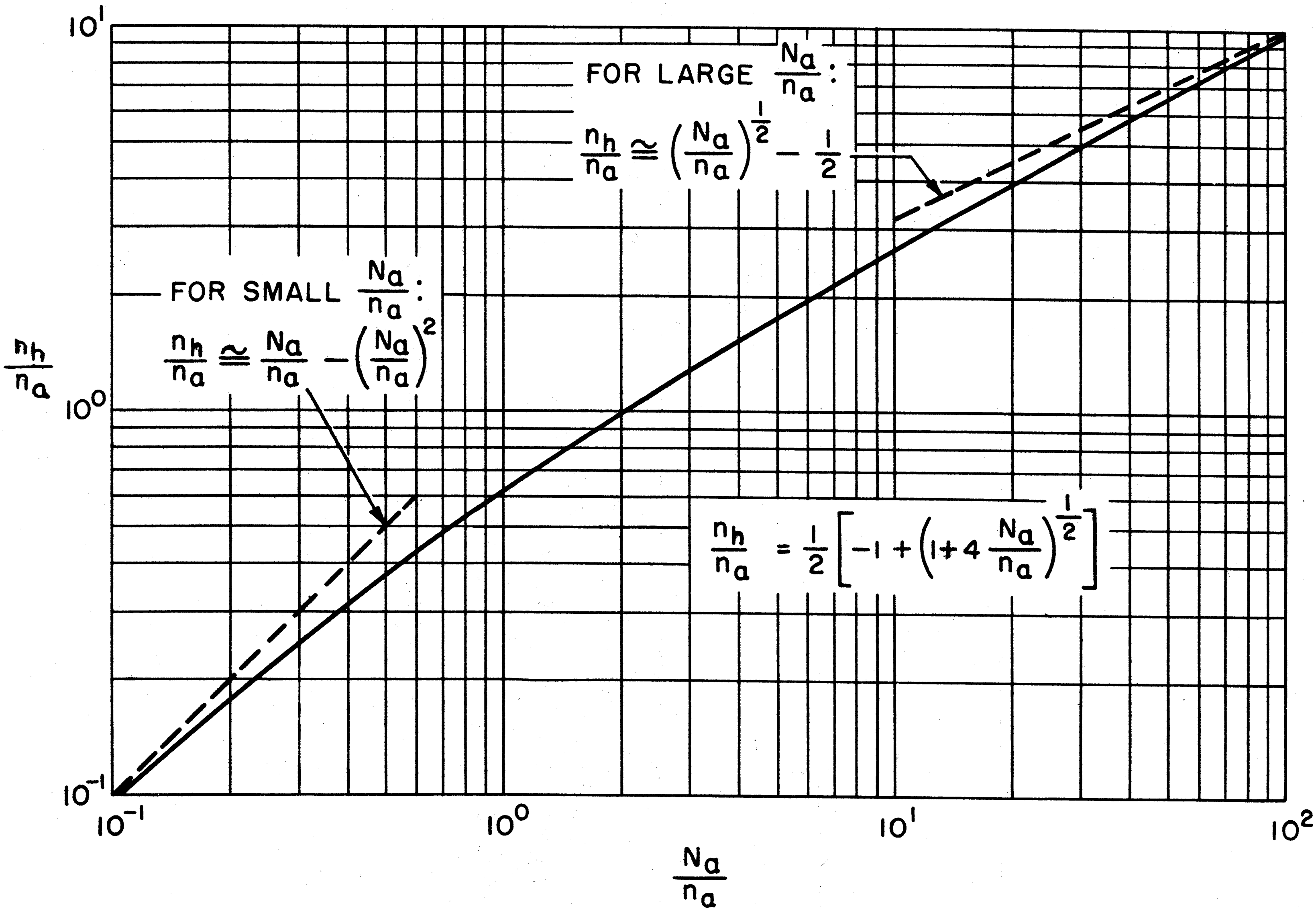
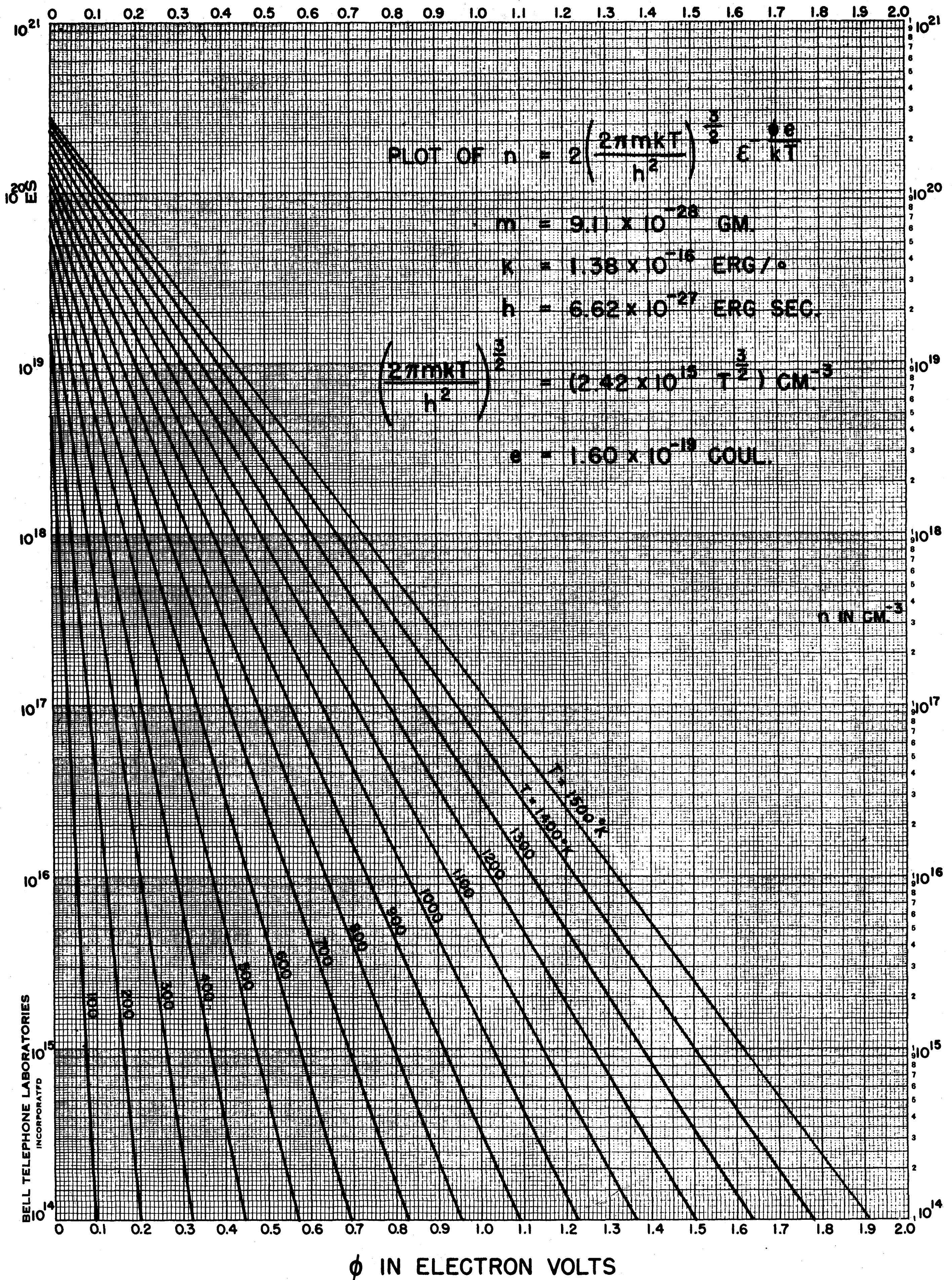
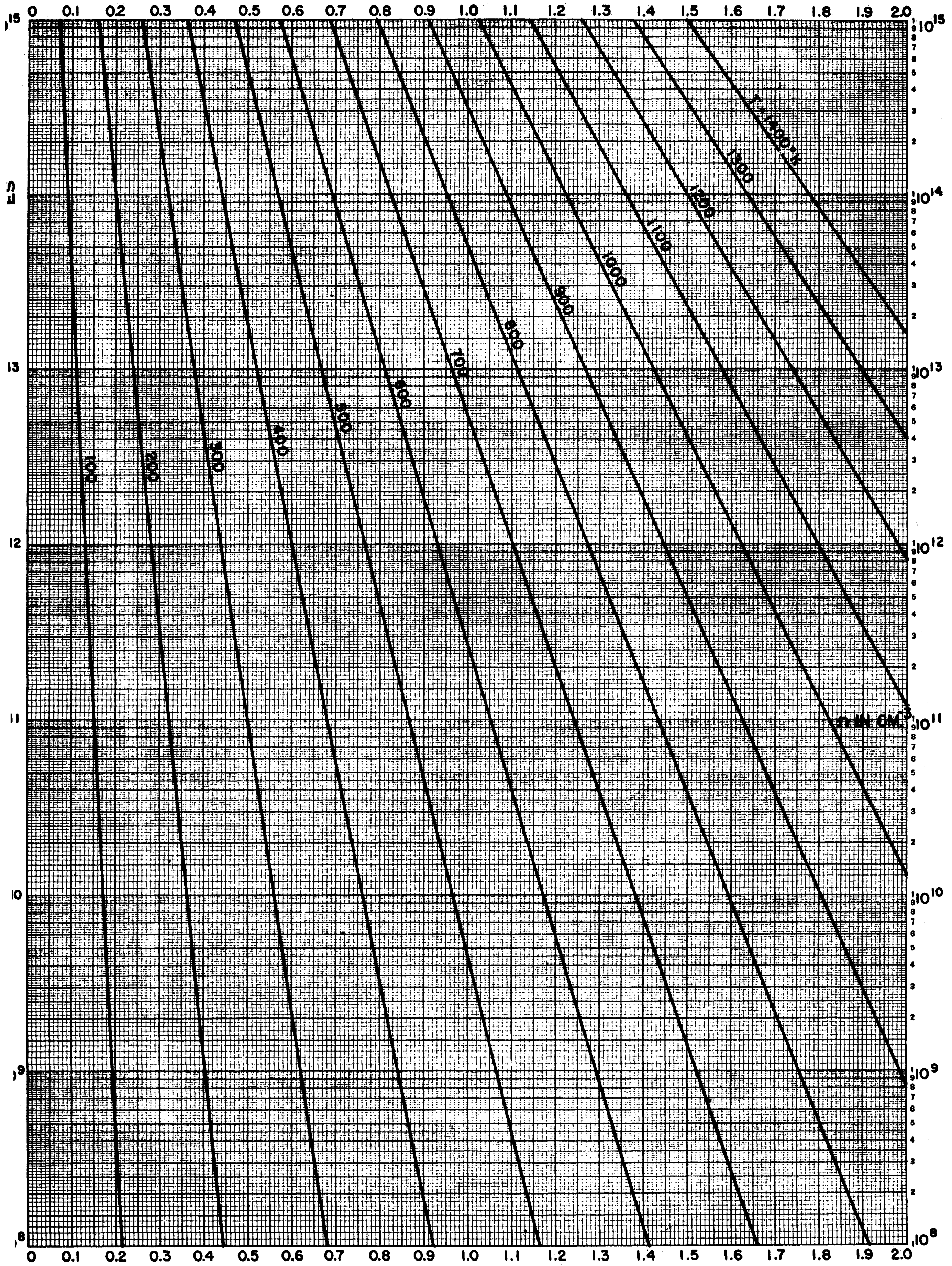


FIG. 3-3



ϕ IN ELECTRON VOLTS

FIG. 3-4A



ϕ IN ELECTRON VOLTS

FIG. 3-4B

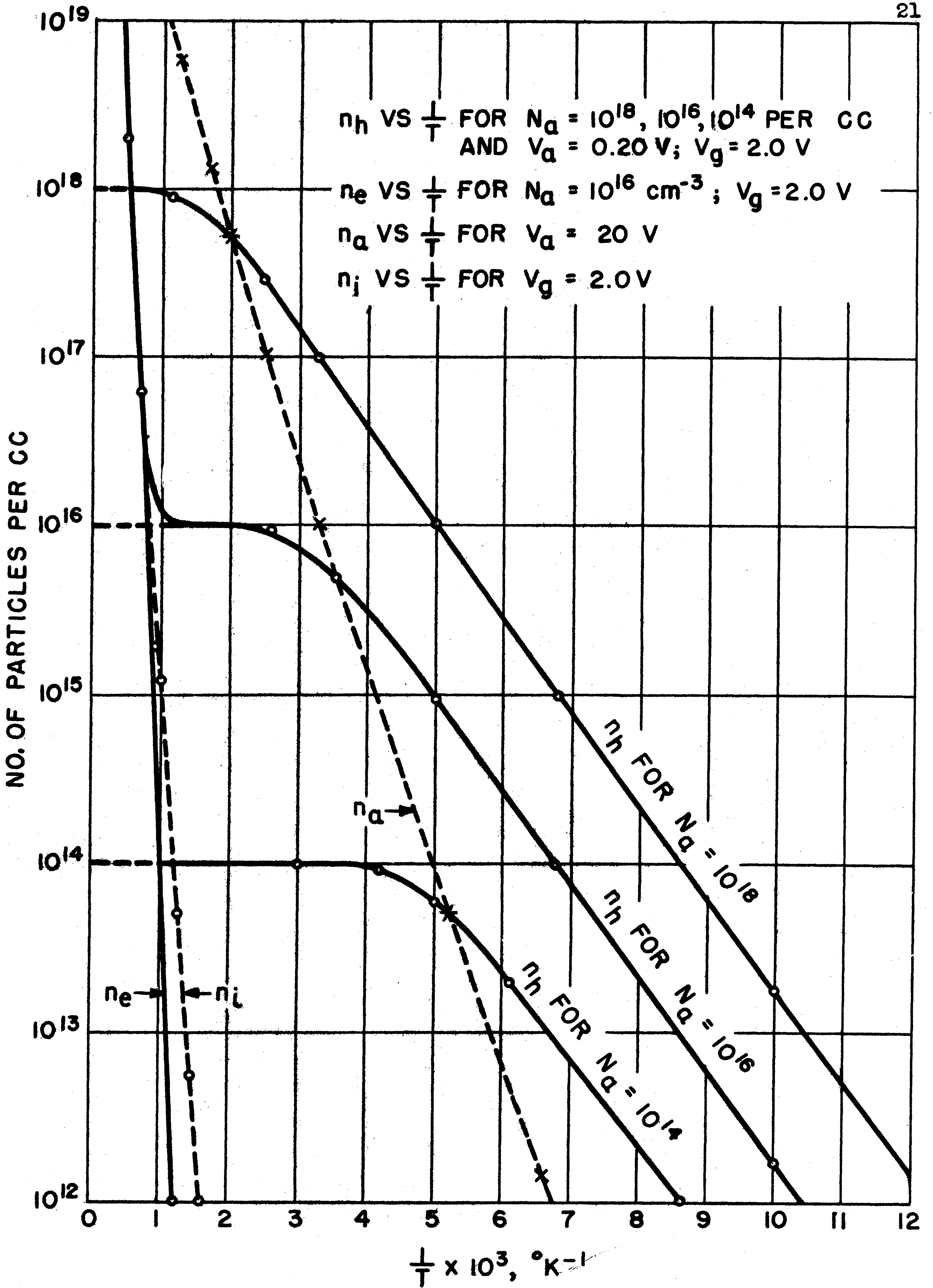


FIG. 3-5

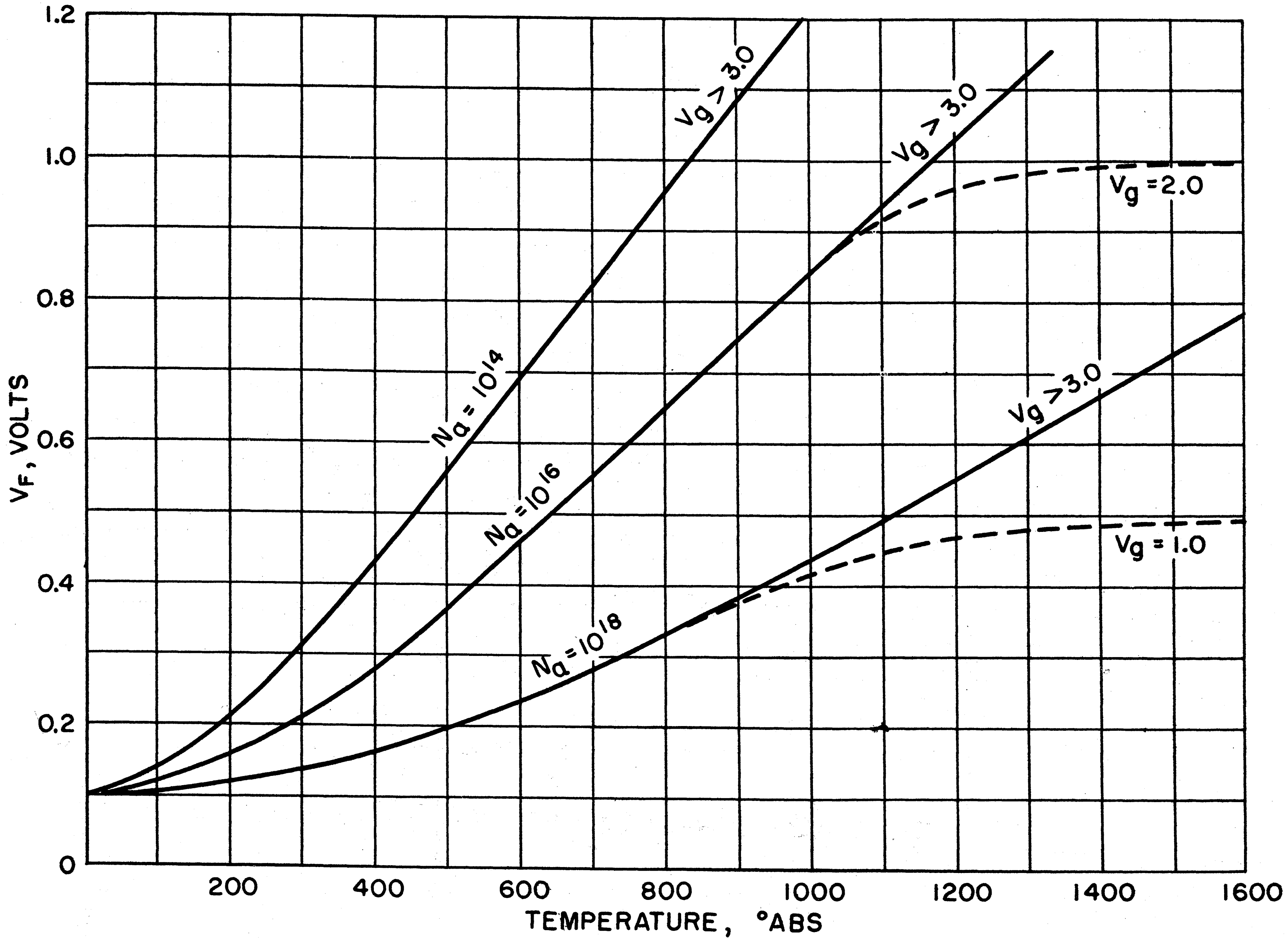


FIG. 3-6

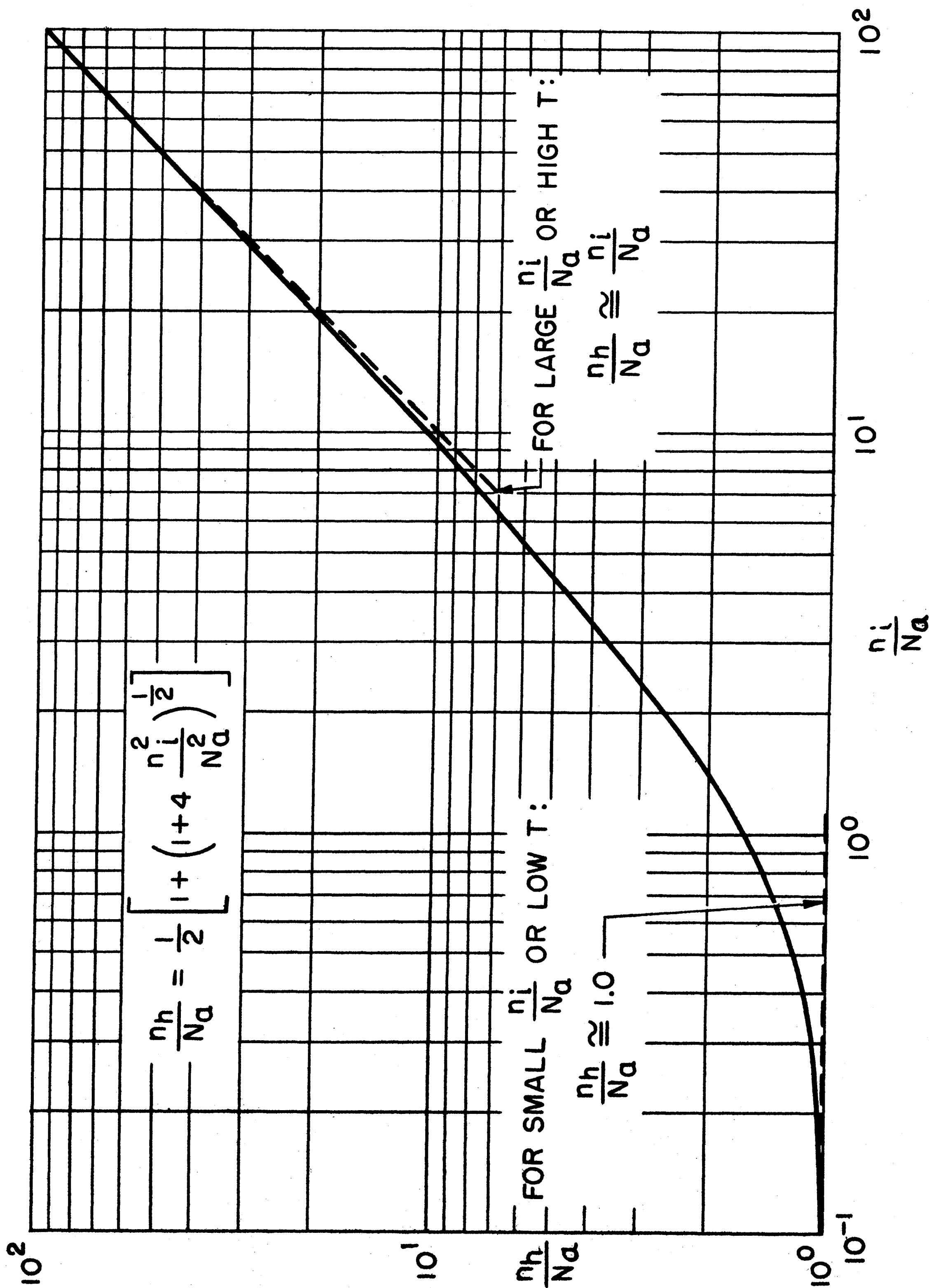


FIG. 3-7

LECTURE 4

HALL EFFECT AND THERMOELECTRIC EFFECT

HALL EFFECT

The Hall effect consists in the development of a transverse voltage across a specimen of material carrying a current at right angles to a magnetic field (Figure 4-1). In terms of the symbols of the figure the Hall constant R_H is defined by $V_H = 10^{-8} R_H I_H/d$, when V_H is in volts, I in amperes, H in gausses, d in centimeters, and R_H in centimeters cubed per coulomb.

The number n of free electrons or holes per cubic centimeter in semiconductors can be deduced from Hall measurements: For a simple case in which carriers of only one sign are responsible for the conductivity, theory states that

$$n = \pm \frac{3\pi}{8e\ R_H}, \qquad (4\text{-}1)$$

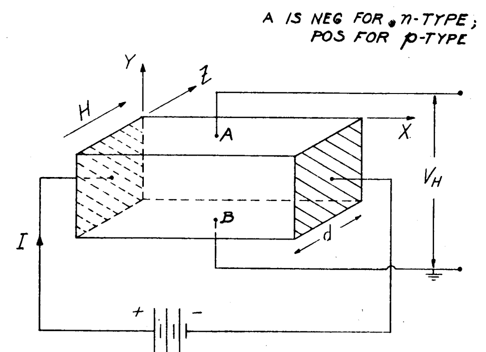


Figure 4-1 Arrangement of Specimen for Hall Effect Measurements

where R_H is the Hall constant in centimeters cubed per coulomb, e is the electronic charge (1.59×10^{-19} coulomb), the minus sign is for electrons, and the plus sign for holes. Putting in the values of the constants, we find that n is equal to 7.4×10^{18} divided by R_H. Hall measurements on two p-type silicon samples C and D at room temperature gave R_H values of +6.5 and +0.5 centimeter cubed per coulomb, respectively, which indicates 1.1×10^{18} and 1.5×10^{19} holes per cubic centimeter. Although the impurity of sample C is unknown, that in sample D is 3 boron atoms for each 10,000 silicon atoms. As there are 5.2×10^{22} silicon atoms per cubic centimeter, this means that 1.7×10^{19} atoms of boron are present in this sample and that each boron produces approximately one hole.

The conductivity σ of an electrical conductor is given by

$$\sigma = nev , \qquad (4\text{-}2)$$

where n is the number of current carriers per cubic centimeter, e is their charge, and v their mobility in centimeters per second per volt per centimeter. Combining equations 4-1 and 4-2 we see that the mobility of the holes or electrons in a simple semiconductor is

$$v = \frac{8 R_H \sigma}{3\pi} = 0.85 R_H \sigma . \qquad (4\text{-}3)$$

The mobilities of the holes in silicon samples C and D at room temperature are, respectively, 55 and 33 centimeters per second per volt per centimeter.

In intrinsic semiconductors where electrons and holes are both present at the same time, the Hall coefficient is

$$R_H = - \frac{3\pi}{8e} \frac{(n_1 c^2 - n_2)}{(n_1 c + n_2)^2} , \qquad (4\text{-}4)$$

where n_1 is the electron density, n_2 the hole density, and c the ratio of the electron mobility to the hole mobility. As c is generally somewhat greater than one, intrinsic semiconductors usually show the negative sign.

The value of the mean free path of the current carriers is shown from the simple theory to be

$$\ell = \frac{2 \times 10^{-7} R_H (2\pi mkT)^{1/2}}{\pi e} \sigma , \qquad (4\text{-}5)$$

where ℓ is the mean free path in centimeters, m is the mass of the carrier ($m = 9 \times 10^{-28}$ grams for electrons and is assumed to be the same for holes), and the remainder of the constants are as previously defined. Substituting in these values, we see that

$$\ell = 3.5 \times 10^{-10} T^{1/2} R_H \sigma . \qquad (4\text{-}6)$$

At room temperature the mean free paths of silicon samples C and D are, respectively, 4×10^{-7} and 2.3×10^{-7} centimeter. These are approximately ten times greater than the lattice constant of silicon which is 5.4×10^{-8} centimeter.

An additional relationship, between ℓ and v, can be obtained by combining equation 4-6 with 4-3 .

$$\ell = 4.2 \times 10^{-10} T^{1/2} v . \qquad (4\text{-}7)$$

Direct measurement of the mobilities of electrons and holes in germanium[1] at room temperature give

[1] Shockley, Pearson, and Haynes, 28, No. 3, Bell System Technical Journal (July, 1949), p. 344.

$$v_e = 2500,$$

$$v_h = 1700 \text{ cms}^2/\text{volt sec.}$$

These mobility values thus correspond by equation 4-7 to mean free paths of 1.8×10^{-5} cms for electrons and 1.2×10^{-5} cms for holes, in germanium at 300° absolute. These mean free path lengths are more than a hundred times the inter-atomic distance in the germanium crystal.

It is obvious from equation 4-1 that if R_H is determined over a range of temperatures, it is possible to compute values of n vs T. From the sign of R_H it is possible to deduce whether these carriers are electrons or holes.

If both R_H and σ are determined for a range of T, then from equation 9.1-3, the mobility can be computed as a function of T.

Figure 4-2 [2] shows a plot of log n_h vs 1/T for four samples of p-type silicon containing the following amounts of boron expressed in atomic per cent

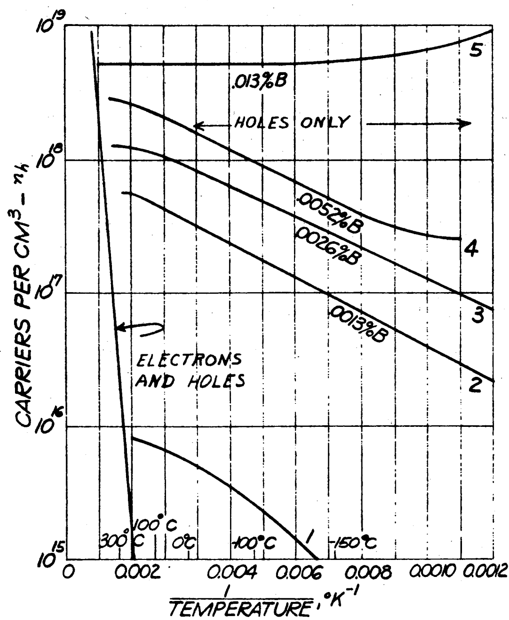


Figure 4-2 Plots of Carrier Density vs Reciprocal Temperature for Silicon Specimens Having Different Concentrations of Boron (p-type) Impurity

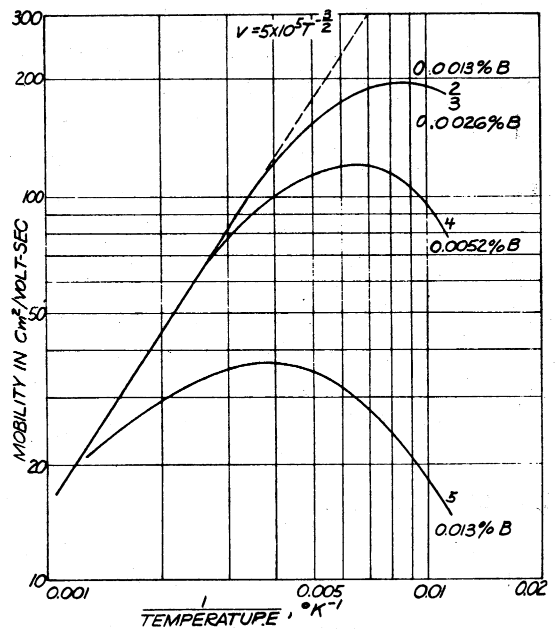


Figure 4-3 Plots of Carrier Mobility vs Reciprocal Temperature for the Same Silicon Samples as in Figure 4-2

---

[2] Bardeen and Pearson, 75, Physical Review (1949), p. 865.

Sample number	2	3	4	5
Atomic % B in Si	.0013	.0026	.0052	.013

Figure 4-3 [2] shows a plot of log v vs 1/T for these same samples. Note that in the impurity region, the mobility decreases as the impurity increases. For a given sample, as T increases, v increases, comes to a maximum, and then decreases. In the intrinsic region the mobility is nearly the same for all samples and decreases as T increases according to the equation,

$$v_h = 5 \times 10^5 \ T^{-3/2} \ . \qquad (4\text{-}8)$$

At the same time the mean free path decreases. Combining equation 4-7 with 4-8 , we obtain

$$\ell_h = 2.1 \times 10^{-4}/T \ . \qquad (4\text{-}9)$$

A simple explanation of the shapes of these mobility curves can be given as follows. At low temperatures the mobility and mean free path are determined chiefly by impurity scattering, hence the lower mobilities at these temperatures for the samples having the higher impurity concentrations. As the temperature increases and holes move faster, the collision cross section of a hole with an impurity atom decreases, since the path of the particle will be deviated less during a near collision at high speed than at low speed. Hence the mobility increases with increasing temperature in this range. For still higher temperatures, however, the increasing scattering effect of lattice vibrations causes the mobility to decrease again.

For electrons in n-type silicon, experiments have shown mobilities in the intrinsic range to be about three times the mobility of holes:

$$v_e = 15 \times 10^5 \ T^{-3/2} \ , \qquad (4\text{-}10)$$

accordingly,

$$\ell_e = 6.3 \times 10^{-4}/T \ .$$

9.2 THERMOELECTRIC EFFECT

The physical phenomenon of the thermoelectric effect is treated in standard texts. We shall now investigate what can be learned, from thermoelectric measurements, about the fundamental properties of semiconductors. In particular, we shall discuss how the Zeebeck voltage (Figure 4-4) is related to the thermoelectric power and the Peltier and Thomson voltages, and how these in turn are related to the Fermi level.

[2] Bardeen and Pearson, 75, Physical Review (1949), p. 865.

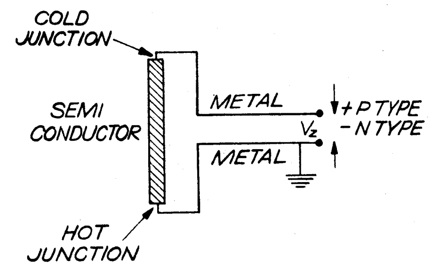


Figure 4-4 Arrangement of Specimen for Thermoelectric Voltage Measurement

Consider a rod of semiconductor with platinum contacts at its two ends, such as shown in the upper part of Figure 4-5 . Let us suppose this semiconductor to be of p-type. If both ends of the rod are at temperature T_1, the Fermi level will be a certain distance V_{F_1} above the filled band of the semiconductor throughout its length, and all the bands and levels of the semiconductor will be horizontal. It has been shown in Chapter 8 that the Fermi level in the metal at either end must be continuous with the Fermi level in the semiconductor at that junction. If the temperature of the right-hand junction is raised to $T_2 = T_1 + \Delta T$, the bands and levels assume the slanted positions shown in the potential diagram in the bottom part of Figure 4-5 . In this diagram the slant is greatly exaggerated in order to illustrate the arguments now to be presented. The temperature difference ΔT is supposed small enough so that all changes occurring in the positions of the various bands and levels are linear with temperature, and hence linear with horizontal distance on the diagram. Let the platinum contacts be connected to platinum wires whose free ends are brought around, outside the diagram, to complete the circuit through the potentiometer which measures the Zeebeck voltage Z. The terminals of the wires at this instrument must be at the same temperature.

The magnitude of the thermoelectric voltage Z is proportional to the difference in temperature between the hot and the cold junction if ΔT is small. The quotient Q between Z and ΔT is called the thermoelectric "power." It is not a power in the sense in which we now use the word power: its physical dimensions are volts per degree. If both hot and cold junctions are raised or lowered in temperature, keeping ΔT constant, it is found that Q depends on the mean temperature. Q as a function of T can equally well be measured by keeping the cold junction at a fixed temperature T_1 and plotting the thermoelectric voltage Z against the variable temperature T_2 of the hot junction. The slope of this plot at any value of T gives Q for that temperature.

The Zeebeck voltage is made up of four separate contributions, whose origins we shall now describe in terms of the band picture of Figure 4-5 **for a P type semiconductor.**

It was observed by Peltier in 1834 that when current is driven across a junction between two different materials, a heating or cooling occurs at the junction, depending on the direction of current flow. This heating or cooling is over and above ordinary joule heating. The Peltier heat coefficient π is expressed as joules of heat liberated or absorbed per coulomb of charge crossing the junction. It therefore has the physical dimensions of a voltage associated with the contact. We now know that the Peltier effect is due to the fact that when current carriers cross a contact their average energy undergoes a change. The mechanism of this change can be seen by the following argument. Consider a current of positive holes in the metal near the left-hand junction moving toward the right. Only those holes which are below the lower dashed horizontal line can cross the junction into the semiconductor. **Each hole that does cross from the metal into the semiconductor gains**

a potential energy equal to eV_F. In addition the average kinetic energy transported per hole is $2kT_1$.* The voltage equivalents of these two quantities are shown on the diagram, and their sum, designated by π_1, is the Peltier voltage at this junction.

$$\pi_1 = V_{F_1} + \frac{2kT_1}{e} .$$

The metal must continually furnish energy to other holes in an attempt to restore the hole distribution-in-energy which has been upset by the selective passing on of the high energy holes into the semiconductor. This restoration energy comes from the thermal energy of the metal crystal lattice, which consequently experiences a cooling effect. It can be seen by applying the same argument in reverse that current flowing in the opposite direction produces a heating at this junction.

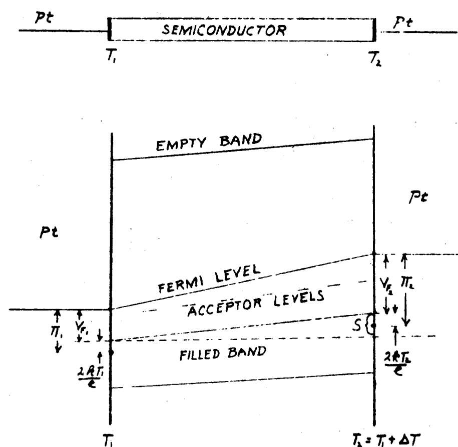


Figure 4-5 Top: Semiconductor With Platinum Electrodes in a Thermoelectric Voltage Experiment

Bottom: Diagram of the Energy Bands in a P type Semiconductor

Possible rectifying barriers at the two metal-semiconductor contacts are not indicated in Figure 4-5 : their presence or absence would have no effect on the argument just presented. Observe that the magnitude of π is different at the two junctions. This is partly because the position of the Fermi level in the semiconductor, with respect to the top of the filled band, depends on temperature, and partly because the kinetic energy transported by a carrier particle depends on temperature. The Peltier voltages at the two junctions constitute two of the four contributions to the Zeebeck voltage.

In 1853 it was deduced by W. Thomson from thermodynamic reasoning that a conductor carrying a current along a temperature gradient should experience a heating or cooling, in addition to and independent of joule heating, depending on

*In dealing with transport phenomena of this sort it can be shown that the average kinetic energy transported per particle is $2kT_1$ rather than the $3/2$ kT ordinarily associated with the equilibrium average kinetic energy of the particle.

the direction of the current. This effect was experimentally verified and shown to correspond to an emf contributing to the Zeebeck voltage. It is our task now to examine the origin of this Thomson heating or cooling in terms of the band picture.

The number of holes per cubic centimeter in the filled band of the semiconductor at any temperature is given by

$$n_h = 2UT^{3/2} \varepsilon^{-\frac{eV_F}{kT}} .$$

It was shown in the last **lecture** that this equation corresponds, at all temperatures below the range in which the acceptors are saturated, to a greater equilibrium hole density at the hot end than at the cold end. Consequently holes will diffuse preferentially from the hot end to the cold end, partly because of this concentration gradient and partly because of the higher kinetic energy and consequently higher diffusion velocities at the hot end. These processes charge the cold end of the semiconductor positively with respect to the hot end and set up an electrostatic field in the semiconductor which, when steady state is reached, causes a return current of holes equal in magnitude and opposite in direction to the diffusion currents. This steady-state field causes the tilting of the energy bands and levels indicated in Figure 4-5 , and produces the electrostatic potential difference S between the two ends of the semiconductor.

Consider now a current of holes moving in the semiconductor from the left-hand end to the right-hand end. As they move, their potential energy decreases by the amount eS per hole, while their kinetic energy transport increases by $2k\Delta T$ per hole because the holes are moving into a region of higher temperature. The net energy change per hole, $-eS + 2k\Delta T$, is abstracted from or delivered to the crystal lattice of the semiconductor. The heat thus absorbed or delivered in joules per coulomb, of charge transported per degree of temperature difference, is called the Thomson coefficient σ: it is defined as positive if positive current flowing in the direction of positive temperature gradient increases the average total energy of the carriers (cools the substance). The Thomson coefficient has the physical dimensions of a voltage per degree: consequently the net effect of the processes described is to establish an emf $\sigma\Delta T$ between the two ends of the semiconductor, where

$$\left.\begin{aligned} \sigma\Delta T &= -S + \frac{2k\Delta T}{e} , \\ \text{or} \quad \int_{T_1}^{T_2} \sigma dT &= -S\Big|_{T_1}^{T_2} + 2\,\frac{k}{e}\,(T_2 - T_1) . \end{aligned}\right\} \qquad (4\text{-}11)$$

The sign of σ for any semiconductor may be positive or negative, depending on the separate magnitudes of the terms in equation 4-11 . Direct calculation of σ is difficult: it involves a knowledge of the behavior of the Fermi level with

temperature and a knowledge of how charged particles diffuse in response to density and temperature gradients. It will be shown below that for a p-type semiconductor, σ is negative at low temperature and slightly positive at higher temperatures.

The total circuit emf measured by the potentiometer is now the difference between the Peltier emfs at the two junctions, plus the difference between the Thomson emfs in the metal and the semiconductor:

$$Z = \pi_2 - \pi_1 + \sigma_m \Delta T - \sigma_s \Delta T,$$

or

$$Z = \pi_2 - \pi_1 + \int_{T_1}^{T_2} \sigma_m dT - \int_{T_1}^{T_2} \sigma_s dT, \qquad (4\text{-}12)$$

where the subscripts m and s denote metal and semiconductor, respectively. Because σ is not constant for all values of T_1, the terms $\sigma\Delta T$ must be replaced by integral forms $\int_{T_1}^{T_2} \sigma dT$ if ΔT is large. For most metals σ_m is nearly always small compared with σ_s. Moreover for platinum, σ_m is nearly zero over a wide temperature range. The third term of equation 4-12 may therefore be omitted from further consideration, and the Zeebeck voltage Z is numerically equal to the vertical difference between the Fermi levels of the metal contacts at the two ends of the diagram in Figure 4-5 .

If we complete the thermoelectric circuit through a high resistance instead of through a potentiometer, a small current will flow continuously. Its value is Z divided by the sum of the internal and external resistances. The direction of this current in the diagram of Figure 4-5 is with holes flowing from hot to cold in the semiconductor and with electrons flowing from hot to cold in the external circuit. The value of the current must be the same everywhere in the circuit: let its magnitude be i amperes. As a result of its flowing, heat is gained and lost by various parts of the circuit. As holes flow into the semiconductor at the hot junction, heat is absorbed at the rate of $i\pi_2$ watts: this heat must be replenished at the same rate by the heat reservoir which maintains the hot junction at temperature T_2. In the semiconductor heat is delivered to or abstracted from the crystal lattice, depending on whether $-\sigma + k/e$ is plus or minus. This heat delivery rate is $i\sigma\Delta T$ watts: the heat must be furnished or dissipated by the surroundings or by conduction from or to the heat reservoirs at the junctions. Finally, heat is delivered by the current at the cold junction at the rate $i\pi_1$ watts, and heat must be taken away at this rate in order to maintain this junction at temperature T_1. The difference between heat flowing into the system and heat flowing out per second is $i(\pi_2 - \sigma\Delta T - \pi_1)$ watts. It is this heat per second absorbed by the system which furnishes the power necessary to maintain the current. This power is eventually dissipated as joule heat in the external and internal circuit.

If the external resistance is replaced by a battery whose emf exceeds Z and whose polarity is such as to send current through the circuit in a direction opposite to that just discussed, the heating effects will be reversed. If the entire circuit is initially at the same temperature, the current flow will cause parts of it to heat up and other parts to cool: in particular, junction 1 will become cooled while junction 2 will become warmed.

It is to be understood that the entire argument presented up to this point holds only for a p-type semiconductor at temperatures below the intrinsic temperature range. By transforming the ideas and arguments to fit the behavior of electrons instead of holes, the situation for an n-type semiconductor can be described. At high temperatures where intrinsic behavior is assumed by the semiconductor, new arguments involving the behavior of both holes and electrons together must be employed. In this temperature region the thermoelectric quantities become complicated functions of electron and hole properties which are beyond the purpose of this text to describe.

By applying the laws of thermodynamics of a complete thermoelectric circuit, W. Thomson was able to deduce certain useful relationships between the thermoelectric quantities.*

$$\left.\begin{aligned} \pi &= Q\,T \\ \sigma_s - \sigma_m \cong \sigma_s &= \frac{TdQ}{dT} \end{aligned}\right\} \qquad (4\text{-}13)$$

By using these relationships the following equalities can be written:

$$Q \equiv \frac{dZ}{dT} = \frac{\pi}{T} = \int_0^T \frac{\sigma_s}{T}\,dT \quad . \qquad (4\text{-}14)$$

In this equation Q, Z, π, and σ are all functions of T. In particular, if we can obtain experimentally a plot of any of these quantities against T, we can determine from equation 4-7 plots of all the other quantities against T. We have already discussed how Q vs T can be obtained from a Z vs T curve obtained experimentally. Figure 4-6 shows such a plot of Q vs T for a p-type semiconductor. The reason for plotting this function Q is that on such a plot the other functions have simple geometrical interpretations. Thus from equation 4-13

$$Z = \int_{T_1}^{T_2} Q\,dT \ ,$$

*See Starling, Electricity and Magnetism, Ch. VIII; or Bridgman, Thermodynamics of Electrical Phenomena in Metals, Ch. II.

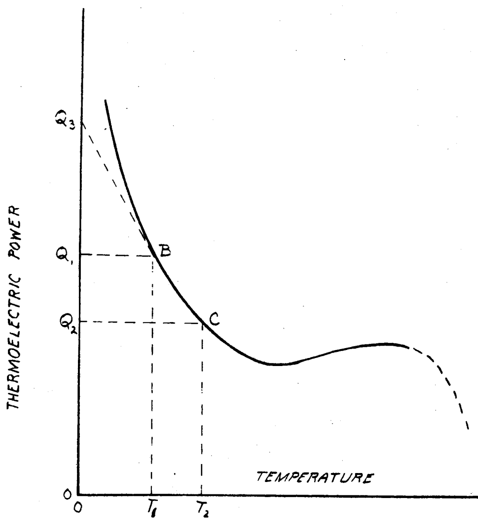


Figure 4-6 Experimental Plot of Q vs T for a p-type Semiconductor

and on the figure the value of Z between temperature T_1 and T_2 is simply the area $T_1BCT_2T_1$.

Since from equation 4-13, $\pi = QT$, it follows that on the figure:

$$\pi_1 = \text{area } OQ_1BT_1O$$

and $$\pi_2 = \text{area } OQ_2CT_2O \ .$$

Again, since from equation 4-13 $\sigma_s = T\frac{dQ}{dT}$, it follows that σ_s for any T can be had from the plot by multiplying its slope at any T by that T. Numerically this is equal to $Q_1 - Q_3$ in Figure 4-6. It is seen that σ is negative in the low temperature region of the plot, that it becomes positive at higher temperatures, and negative again at still higher temperatures.

In Figure 4-6 the Q vs T curve is shown as a dashed line in the high temperature region. In this region the Fermi level approaches the middle of the gap between the filled band and the empty band, and the semiconductor becomes intrinsic. For this condition π is no longer simply equal to $V_F + \frac{2kT}{e}$ but becomes a more complicated function which will be derived below. The result is that in this temperature region π decreases with increasing temperature and may even become negative if the electron mobility exceeds the hole mobility. Since $\pi = \frac{Q}{T}$, Q will change sign when π changes sign. At very high temperatures toward the upper end of the intrinsic range $m_e \cong m_h$, both electron and hole mobilities decrease, and π and Q approach zero.

Instead of starting from an experimentally determined Q vs T curve and deducing the other thermoelectric functions, as was indicated above, one may start with a theoretically calculated curve of V_F vs T and from it deduce curves for π, Q, Z, σ, and S. This we shall now do for the case of a p-type semiconductor having acceptor density $N_a = 10^{17}$ cm^{-3} acceptor activation energy $V_a = 0.2$ volt, and forbidden energy gap $V_g = 2.0$ volts. To calculate V_F we use the formulas derived in Lecture 3, Case 2. Figure 4-7 shows V_F vs T calculated in this way. It shows also a plot of $\frac{2kT}{e}$ vs T. The sum of these two curves gives the curve of π vs T in this figure.

Figure 4-8 gives a Q vs T curve calculated from the π curve of the last figure by means of equation 4-13. The dashed line extension of the Q curve at high temperatures in the intrinsic region shows qualitatively how π might be expected to behave in this range.

The thermoelectric Zeebeck emf Z for the semiconductor thermocouple between any two temperatures T_1 and T_2 can be obtained by integrating the Q vs T curve

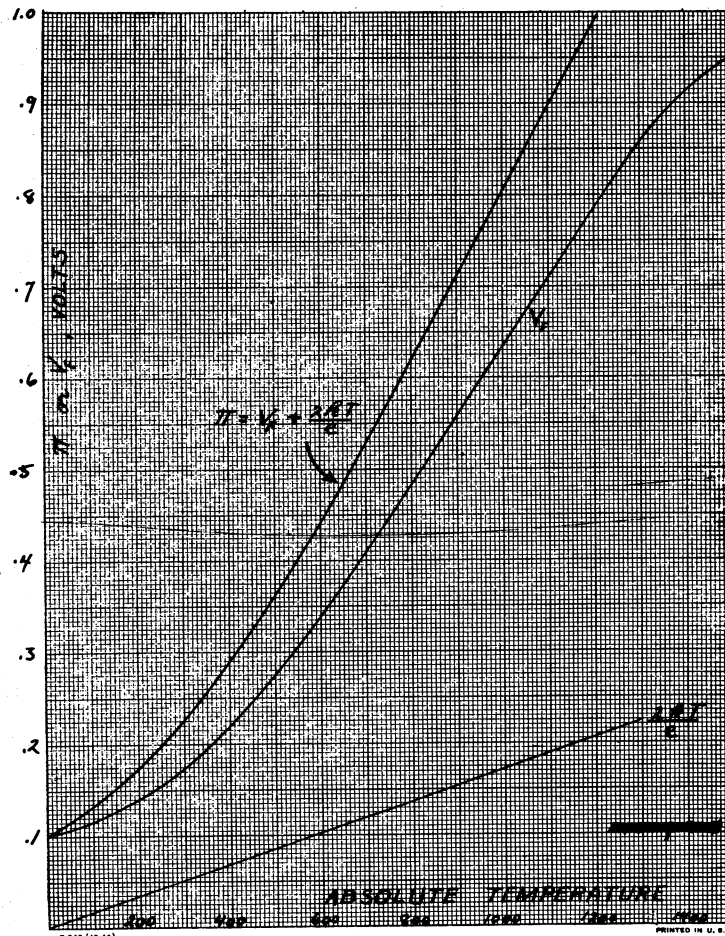


Figure 4-7 Theoretical Curve of V_F vs T for a p-type Semiconductor Having $N_a = 10^{17}$ cm^{-3}, $V_a = 0.2$ Volt, and $V_g = 2.0$ Volts. The straight line plot is $\frac{2kT}{e}$ vs T. The top curve for π is the sum of the lower two.

between these temperatures as justified by equation 4-14, $Z\Big|_{T_1}^{T_2} = \int_{T_1}^{T_2} Q dT$.

Figure 4-9 shows a Z vs T_2 curve for the case of the semiconductor under consideration when the cold junction is held at $T_1 = 100°$ K and T_2 is varied.

Figure 4-10 presents a σ_s vs T curve calculated from the Q vs T curve of Figure 4-8 with the aid of equation 4-13 .

From equation 4-11 the quantity S for this semiconductor whose ends are at temperatures T_1 and T_2 can be obtained by integrating the σ curve between these temperatures, changing the sign of the integral, and adding the quantity $\frac{2k}{e}(T_2-T_1)$. Figure 4-12 shows these various quantities for the semiconductor thermocouple with its cold end at 100° K and its hot end at various higher tempertures T_2. The plot of $-\int_{100}^{T_2} \sigma dT$ was obtained by graphical integration of Figure 4-10 .

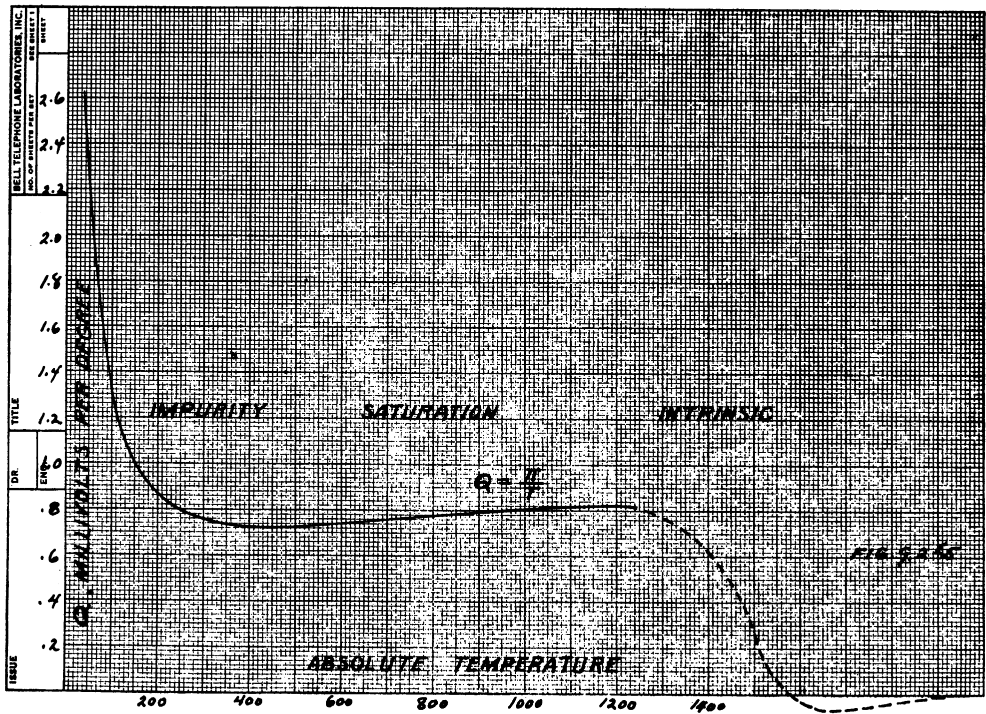


Figure 4-8 Theoretical Curve for Q vs T Obtained From the π Curve of Figure 9.2-4

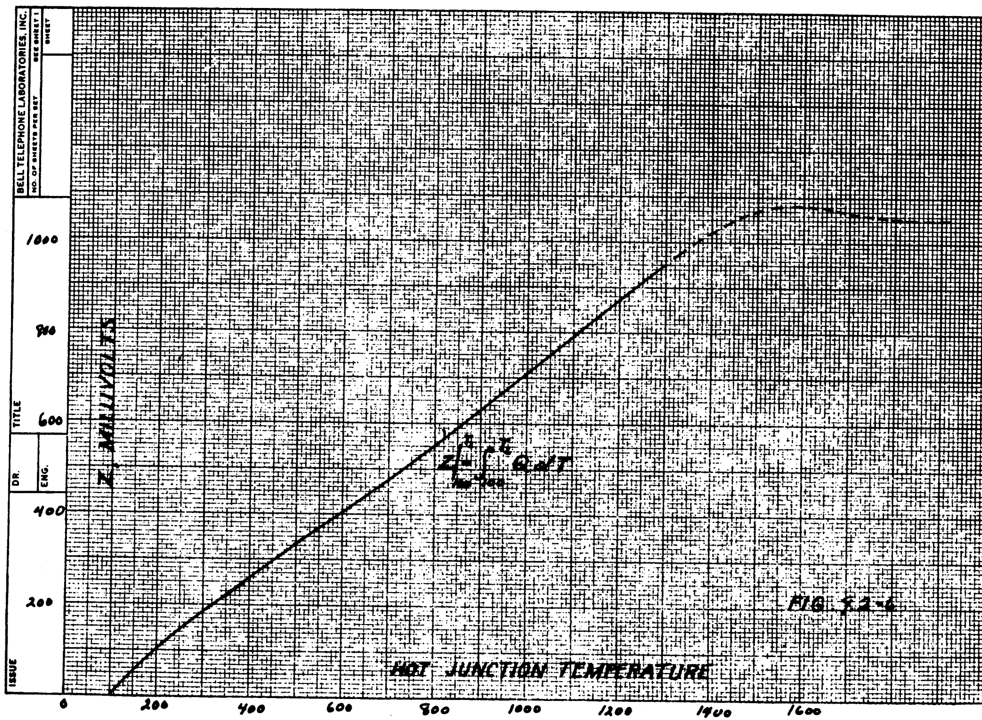


Figure 4-9 Z vs T Curve Obtained by Integration of the Q vs T Curve of Figure 9.2-5. The lower limit of integration is 100°K.

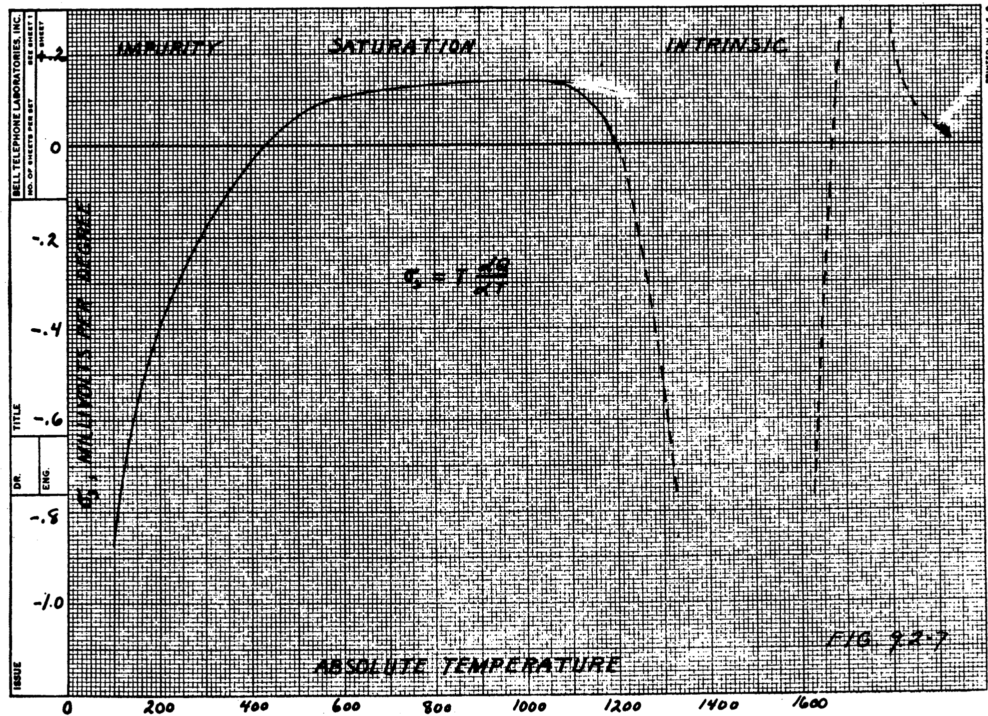


Figure 4-10 Theoretical Curve for σ vs T Obtained From the Q Curve of Figure 9.2-5

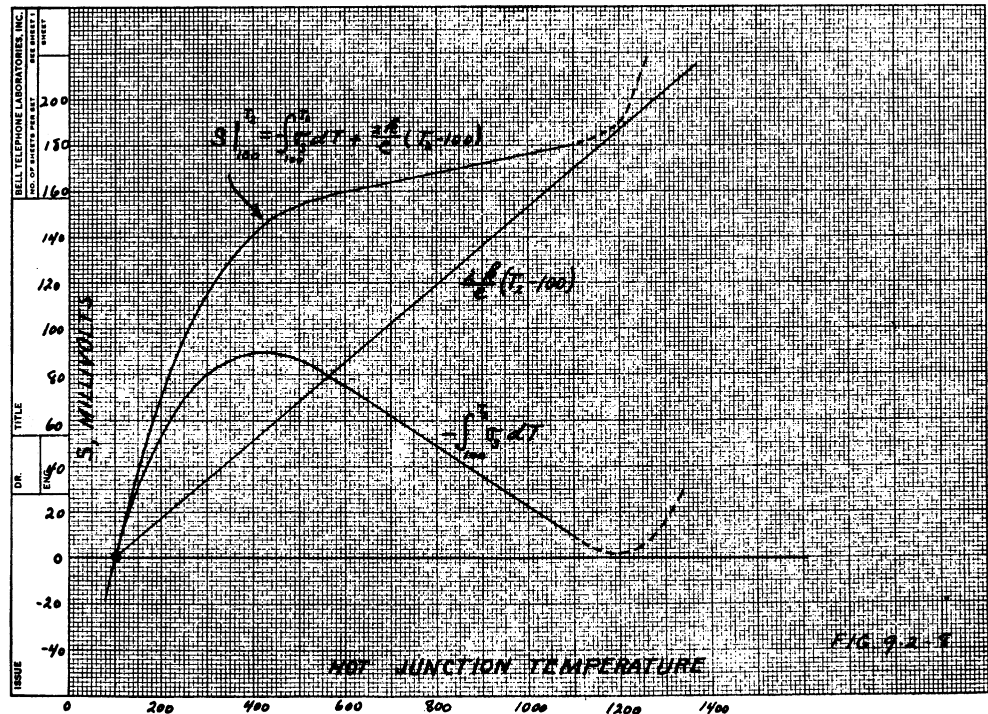


Figure 4-11 Theoretical Curve for S vs T Obtained From the σ Curve of Figure 9.2-7. The top curve for S is obtained by adding the two lower curves.

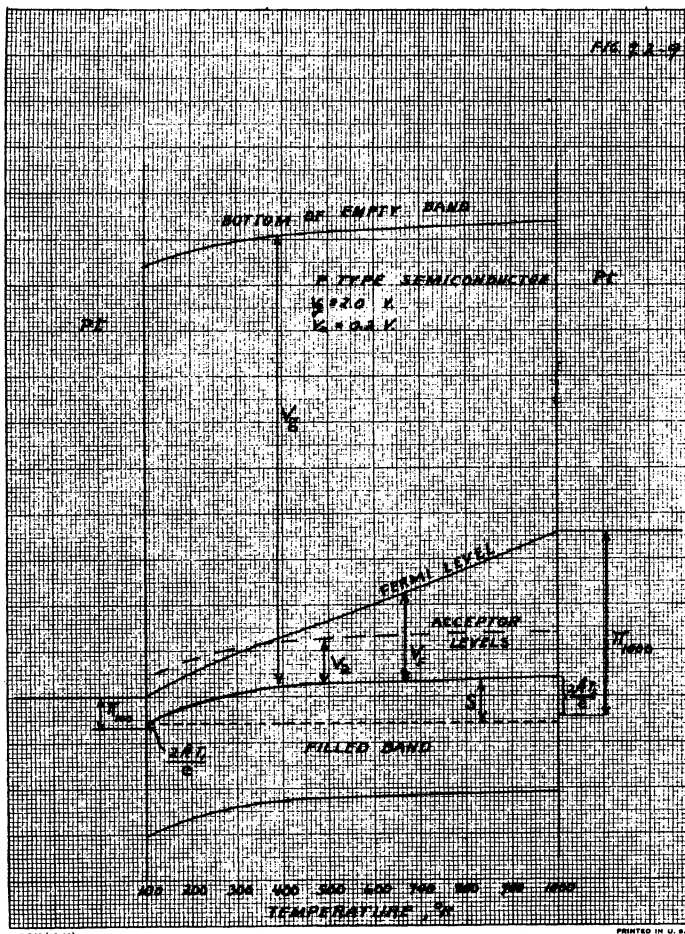


Figure 4-12 Energy Level Diagram for a p-type Semiconductor in a Thermocouple Circuit

Finally, Figure 4-12 shows the energy level diagram for a semiconductor-platinum thermocouple like that of Figure 4-5 for which the cold end is at 100° K and the hot end at 1000° K. We assume a temperature distribution which is linear between the two ends, but now the temperature difference is so large that the positions of the Fermi level and the edges of the bands are not linear with temperature or distance. This figure combines the information presented separately in Figures 4-7 to 4-11 . The semiconductor has $N_a = 10^{17}$ cm^{-3}, $V_a = 0.2$ volt, and $V_g = 2.0$ volts.

Since Q is related to the position of the Fermi level, it is possible from thermoelectric data to determine the density n of charge carriers. For p- or n-type semiconductors in which the carriers are all of one sign the theory of Lecture 3 gives for n:

$$n = 2UT^{3/2} \varepsilon^{-\frac{eV_F}{kT}} . \qquad (4-15)$$

But $$Q = \frac{\pi}{T} = \frac{1}{T}\,(V_F + \frac{2kT}{e}) \;,$$

from which $$\frac{V_F}{T} = Q - \frac{2k}{e} \;.$$

Hence $$n = 2UT^{3/2}\;\varepsilon^{-\frac{Qe}{k} + 2} \qquad (4\text{-}16)$$

where $$UT^{3/2} = 2.42 \times 10^{15}\; T^{3/2}\; cm^{-3}$$

and $\frac{e}{k} = 11600°$ C per volt.

Equation 4-16 can be rewritten:

$$\log_{10} n = 15.68 + \tfrac{3}{2}\log_{10} T - 5050\, Q + 0.87 \;. \qquad (4\text{-}17)$$

Equations 4-15, 4-16, and 4-17 are valid only for p-type or n-type semiconductors, and only if n is less than about one tenth of $4.84 \times 10^{15}\; T^{3/2}$. In these equations, the absolute value of Q should be substituted; the sign of Q determines whether n is for holes or electron carriers.

Derivations of the Relationship Between the Fermi Level V_F and the Peltier Voltage π*

Consider a semiconductor with an unallowed gap V_g. At any temperature T the Fermi level V_F will lie somewhere in this gap. We will treat only the case of nondegeneracy, i.e., n_e or n_h is always less than $2U_eT^{3/2}$ or $2 \times (\frac{2\pi\, m_e\, kT}{h^2})^{3/2}$. ($m_e$ = effective mass of electron in semiconductor; m_h = effective mass of hole.) We can derive the relationship between π and V_F since the thermoelectric power Q or π/T is the entropy per unit current flowing across a metal-semiconductor junction at temperature T. Hence we must express the entropy and currents in terms of the Fermi level and the number and mobility of the holes and electrons.

The entropy per electron in the conduction band = energy above the Fermi level/T = $[(V_g - V_F)e + 2kT]/T$. The 2kT comes in because the average electron that crosses the junction carries with it a kinetic energy 2kT. The term $(V_g - V_F)e$ is the increase in potential energy between an electron in the metal and an electron in the conduction band of the semiconductor. Similarly the entropy of a hole in the filled band = $(V_F\, e + 2kT)/T$.

*We are indebted to our colleague Conyers Herring for this derivation.

Entropy flux in a field E = electron flux x entropy per electron + hole flux x entropy per hole =

$$\frac{E\, n_h\, ev_h}{T}\left(V_F + \frac{2kT}{e}\right) - \frac{E\, n_e ev_e}{T}\left(V_g - V_F + \frac{2kT}{e}\right),$$

in which n_e = number of electrons/cm³ in the semiconductor conduction band and v_e their mobility; n_h = holes/cm³ in the semiconductor filled band and v_h their mobility.

Current density in field $E = E\, n_e ev_e + E\, n_h ev_h$. Hence,

$$Q = \frac{\pi}{T} = \frac{1}{T}\,\frac{n_h v_h\left(V_F + \frac{2kT}{e}\right) - n_e v_e\left(V_g - V_F + \frac{2kT}{e}\right)}{n_e v_e + n_h v_h}. \quad (4\text{-}18)$$

nce

$$n_e = 2U_e T^{3/2}\,\varepsilon^{-(V_g - V_F)e/kT} \quad \text{and} \quad n_h = 2U_h T^{3/2}\,\varepsilon^{-V_F e/kT},$$

$$\pi = \frac{m_h^{*3/2}\, v_h\, \varepsilon^{-V_F e/kT}\left(V_F + \frac{2kT}{e}\right) - m_e^{*3/2}\, v_e\, \varepsilon^{-(V_g - V_F)e/kT}\left(V_g - V_F + \frac{2kT}{e}\right)}{m_e^{*3/2}\, v_e\, \varepsilon^{-(V_g - V_F)e/kT} + m_h^{*3/2}\, v_h\, \varepsilon^{-V_F e/kT}}. \quad (4\text{-}19)$$

If we make the approximations that $m_e^* = m_h^*$ and $v_e = v_h$, consider a case where $n_h \gg n_e$; then $\pi = V_F + \frac{2kT}{e}$. (4-20)

In this derivation it is implicitly assumed that the mean free paths of electrons and holes are independent of T or of their kinetic energy.

RECAPITULATION OF FORMULAS

For convenience we will now give the formulas, applying to various types of semiconductors, which relate Hall effect, thermoelectric effect, and conductivity measurements to the carrier density, mobility, and mean free path.

1. For pure n- or for pure p-type:

$$\sigma = n\, e\, v \quad (4\text{-}21)$$

$$n = \pm\frac{3\pi}{8eR_H} = \pm\frac{7.4 \times 10^{18}}{R_H} \quad (4\text{-}22)$$

$$v = \pm\, 0.85\, R_H\, \sigma \quad (4\text{-}23)$$

$$n = 2UT^{3/2}\,\varepsilon^{-Q\, e/k\, +\, 2} = 4.84 \times 10^{15}\, T^{3/2}\, 10^{-5050Q\, +\, 0.87}. \quad (4\text{-}24)$$

2. For I- or intrinsic type:

$$\sigma = n_e\, e\, v_e + n_h\, e\, v_h \tag{4-25}$$

$$R_H = \frac{3\pi}{8e}\,(n_e\, c^2 - n_h)/(n_e\, c + n_h)^2 \tag{4-26}$$

where $c = v_e/v_h$, the ratio of electron to hole mobility.

$$Q = \frac{1}{T}\,\frac{n_e v_e (V_g - V_F + 2\,\frac{kT}{e}) - n_h v_h\,(V_F + 2\,\frac{kT}{e})}{n_e v_e + n_h v_h}\,, \tag{4-27}$$

where $(V_g - V_F)$ and V_F are given by

$$n_e = 4.84 \times 10^{15}\, T^{3/2}\, \varepsilon^{-(V_g - V_F)e/kT}\,,$$

$$n_h = 4.84 \times 10^{15}\, T^{3/2}\, \varepsilon^{-V_F\, e/kT}\,.$$

V_g, V_F are in volts; Q in volts/°C.

(This assumes that the effective masses of electrons and holes are both equal to that of the free electron = 9.03×10^{28} gms.) If $n_e = n_h = n$:

$$\sigma = n\, e(v_e + v_h) = n\, e\,\frac{c + 1}{v_h} \tag{4-28}$$

or

$$n = \frac{-7.4 \times 10^{18}}{R_H}\;\frac{c - 1}{c + 1} \tag{4-29}$$

$$n = 2UT^{3/2}\;\varepsilon^{-(Q\frac{c+1}{c-1}\frac{e}{k} - 2)} = 4.84 \times 10^{15}\, T^{3/2}\, \varepsilon^{-(11600\, Q\frac{c+1}{c-1} - 2)}\,. \tag{4-30}$$

From equations 9.3-9 and 9.3-10 the quantity $\frac{c+1}{c-1}$, and hence c, can be evaluated. By substitution in equation 9.3-9, n can be evaluated. By substituting values for n and c in equation 9.3-8, v_h can be evaluated. Since $c = v_e/v_h$, v_e can be evaluated. Hence even though v_e differs from v_h, they can be determined if R_H, Q, and σ can be experimentally determined at any T in the intrinsic range.

For n-, p-, and I-types, the mean free path, ℓ, can be determined from the equation

$$v_e = (4/3)\, e\ell\, (2\pi m_e kT)^{1/2}\ \text{(e.s.u.)}\,,$$

where m_e = effective mass of the electron. If m_e is taken equal to the mass of the free electron

$$\ell = 4.12 \times 10^{-10} \, v \, T^{1/2} \text{ (cm)} . \qquad (4\text{-}31)$$

This applies to either holes or electrons if their effective masses are taken equal to that of the free electron.

3. <u>For semiconductors in which n is greater than</u>

$$4.84 \times 10^{15} \, T^{3/2} ,$$

$$\sigma = n \, e \, v, \qquad (4\text{-}32)$$

$$R_H = \pm \frac{1}{en} = \pm \frac{6.25 \times 10^{18}}{n} , \qquad (4\text{-}33)$$

or

$$n = \pm \frac{6.25 \times 10^{18}}{R_H} , \qquad (4\text{-}34)$$

$$v = \pm \sigma R_H . \qquad (4\text{-}35)$$

9.4 RELATIVE MERITS OF HALL AND THERMOELECTRIC DATA

We have seen that the experimental data for R_H and Q supplement each other if the mobilities of holes and electrons are different. Even if they are the same it is desirable to obtain both, if possible. In certain cases where the resistivity is high, it is impossible to measure the Hall voltage because the Johnson noise between the Hall voltage contacts exceeds the Hall voltage. For such samples it is still possible to measure Q. This and σ will then determine n and v. Furthermore the equipment needed to determine Q is easier to make than that to determine R_H. On the other hand, the interpretation of R_H is simpler and the precisions obtainable from R_H are greater.

LECTURE 5

STRUCTURE OF BARRIER LAYER AND HEIGHT OF THE POTENTIAL BARRIER IN A SEMICONDUCTOR IN CONTACT WITH A METAL

In order to determine how the currents in a metal-semiconductor rectifer vary with the applied potential, with temperature, and with the nature of the semiconductor, it is necessary to investigate the nature of the barrier layer in the semiconductor near the metal. The structure of the barrier layer also determines the high frequency behavior and other properties of the rectifier.

When a metal and uniform N type semiconductor such as Ge are brought into contact and if the electron affinity of the metal exceeds that of the Ge, electrons will be transferred to the metal surface. As a result the Ge near the metal has very few electrons to balance the positive charge on the ionized donor atoms. Let N_d be the number of donor atoms/cm^3. At room temperature practically all of these will be ionized. Let D be the depth or distance x in the Ge to the point where the positive charge due to N_d is neutralized by conduction electrons. Then $N_d \times D$ = number of electrons/cm^2 of surface transferred to the metal. As a result of this positive space charge and the negative charge on the metal, the potential ϕ for electrons in the bottom of the empty band rises from ϕ_e for $x > D$ to ϕ_o at $x = 0$. The zero of the potential is at the Fermi Level in the metal.

Fig. 5-1a shows schematically how ϕ will vary with x when no potential is applied between the semiconductor and the metal. It also indicates: ϕ_g the potential difference of the gap or unallowed band; ϕ_h the potential of the top of the filled band below the Fermi Level; and the space charge distance D in which the ionized donors are not balanced by conduction electrons.

Fig. 5-1b shows the situation when a positive potential, V, of about 3 volts is applied to the semiconductor. This would be the reverse direction for an n type semiconductor. Since Fig. 5-1b represents potential energy of electrons which have a negative charge, this means that an electron at the Fermi Level in the semiconductor at distances greater than D has a potential of -V electron volts or that the Fermi Level in the semiconductor is lowered with respect to the Fermi Level in the metal by V volts. The distance D is increased by a factor of about 2, because more donors have transferred their electron to the metal; the negative charge on the surface of the metal has increased by a

factor of about 2. The field of $\frac{d\phi}{dx}$ near the metal has increased by a factor of about 2. ϕ_o has not changed; however, the height of the potential hill which an electron sees when it tries to go from the metal into the semiconductor has been reduced to ϕ_m or by an amount $\Delta\phi = \phi_o - \phi_m$.

Our problem is to derive expressions for ϕ as a function of x, and how this will depend on the applied potential V, and on the donor density N_d. In the solution of this problem we will also obtain expressions for D, $\frac{d\phi}{dx}$, and $\Delta\phi$ in terms of V and N_d. The solution of the problem is based on the fact that ϕ must satisfy Poisson's equation, namely

$$\frac{d^2\phi'}{dx^2} = + \frac{4\pi}{K} N_d e' \qquad (5\text{-}1)$$

(The + sign is used because we are using potentials for electrons instead of for a positive test charge.)

e' = charge on electron in e.s.u. = 4.8×10^{-10}

ϕ' = electrostatic potential = volts/300

K = dielectric constant = 16 for Ge

ϕ' and e' are primed to indicate that ϕ and e are in e.s.u. Simi
ϕ_o, ϕ_e and V will be primed if the quantities they represent are to be i
Later if these quantities are to be in practical units, the prime will b
This equation is subject to the following boundary conditions:

1) When x = 0 $\qquad \phi' = \phi'_o$ a constant (The subscript o designa that x = 0)

2) When x = D $\qquad \phi' = + V' + \phi'_e$ and $\frac{d\phi}{dx} = 0$

All potentials are referred to the Fermi Level in the metal. V' is the electron potential in e.s.u. applied to the semiconductor. V' is a positive number in the conducting direction; it is negative in the reverse direction, such as in Fig. 5-1b.

The solution of Eq. (5-1) is of the form

$$\phi' = (\phi'_e + V') + C(D - x)^2 \qquad (5\text{-}2)$$

where C is a constant, independent of x, whose value is to be determined.
By differentiating Eq. (5-2) we get

$$\frac{d\phi'}{dx} = + C\ 2(D - x)\ (-1) = -\ 2C(D - x) \qquad (5\text{-}3)$$

When

$$x = 0, \quad \left.\frac{d\phi'}{dx}\right|_{x=o} \text{ or } \left.\frac{d\phi'}{dx}\right|_{o} = -\ 2CD$$

By differentiating Eq. (5-3) we get

$$\frac{d^2\phi'}{dx^2} = -2C(-1) = 2C \qquad (5\text{-}4)$$

From Eqs. (5-1) and (5-4)

$$2C = \frac{4\pi N_d e'}{K}$$

or

$$C = \frac{2\pi N_d e'}{K} \qquad (5\text{-}5)$$

When x = 0

$$\phi' = \phi'_o = \phi'_e + V' + CD^2$$

Hence

$$D = \left\{\frac{\phi'_o - V' - \phi'_e}{C}\right\}^{1/2} = \left\{\frac{(\phi'_o - V'_e - \phi'_e)K}{2\pi N_d e'}\right\}^{1/2} \qquad (5\text{-}6)$$

When x = D

$$\phi' = \phi'_e + V'$$

and

$$\frac{d\phi'}{dx} = -2C(0) = 0$$

Hence Eq. (5-2) satisfies Poisson's equation (5-1) and the boundary conditions if the constants C and D have the values given in Eqs. (5-5) and (5-6).
If the potentials are expressed in volts and electron charge in coulombs, we shall use unprimed symbols. Then equation (5-2), (5-3), (5-5), and (5-6) become

$$\phi = (\phi_e + V) + 300C(D - x)^2 = (\phi_e + V) + (\phi_o - \phi_e - V)\left(1 - \frac{x}{D}\right)^2 \qquad (5\text{-}2a)$$

$$\frac{d\phi}{dx} = F = -2 \times 300C\,(D - x) = -600 \times 1.89 \times 10^{-10}\, N_d(D - x)$$
$$= -1.13 \times 10^{-7}\, N_d\,(D - x) \qquad (5\text{-}3a)$$

$$C = \frac{2\pi N_d}{K}\, e' = \frac{2\pi \times 4.8 \times 10^{-10}}{16}\, N_d = 1.89 \times 10^{-10}\, N_d \qquad (5\text{-}5a)$$

$$D = \left\{\frac{\phi_o - V - \phi_e}{300\, C}\right\}^{1/2}$$
$$= 4.2 \times 10^3 \left\{\frac{\phi_o - \phi_e - V}{N_d}\right\}^{1/2} \qquad (5\text{-}6a)$$

$$\frac{d\phi}{dx} = -4.75 \times 10^{-4}\,(N_d)^{1/2}\,(\phi_o - V - \phi_e)^{1/2} + 1.13 \times 10^{-7}\, N_d x \qquad (5\text{-}4a)$$

In these equations K = 16, the value of the dielectric constant for Ge.

It is instructive to calculate values for D, $\frac{d\phi}{dx}$, and ϕ for various values of N_d, V, and x. To do this we must assign values to ϕ_o and ϕ_e appropriate for Ge. Subsequent analysis will show that ϕ_o is nearly equal to the width of the unallowed gap which is about .7 electron volts. Hence we will put ϕ_o = .70 volts. ϕ_e will depend on N_d and T. For N_d less than about 2×10^{19} at T = 300, all the donors will have donated electrons to the conduction band so that $n_e = N_d$. Then n_e will be related to ϕ_e by

$$n_e = N_d = 2UT^{3/2} E^{-\phi_e e/kT} \qquad (5\text{-}7a)$$

where U - universal constant = $\left(\frac{2\pi \;\; mk}{h^2}\right)^{3/2} = 2.42 \times 10^{15}$

Hence we calculate ϕ_e vs N_d at room T or T = 300°K

Table 5-1

N_d =	10^{15}	10^{16}	10^{17}	10^{18}	10^{19}
ϕ_e in Volts =	.26	.20	.14	.08	.04
ϕ_h in Volts =	.44	.50	.56	.62	.66

For N_d near 10^{19} a more complicated equations than (5-7a) must be used. This yields a value of about .04 for ϕ_e instead of .024. Then for $N_d = 10^{18}$ (or about .0025% donor impurity) (There are 4.5×10^{22} atoms of Ge per cm^3)

$$D = \frac{4.2 \times 10^3}{10^9} (.7 - .08 - V)^{1/2} = 4.2 \times 10^{-6} (.62 - V)^{1/2} \qquad (5\text{-}6b)$$

V =	.22	0	-1	-3	-10	-30	-100	-200
10^6 D =	2.56	3.3	5.34	8.0	13.7	23.2	42.2	59.4

Fig. 5-2 shows a plot of D vs V for $N_d = 10^{18}$. This same plot can be used for any other N_d provided the D scale is multiplied by $(10^{18}/N_d)^{1/2}$. (For really precise values the curve must also be shifted slightly along the V axis to allow for the small variation of ϕ_e with N_d as given in the above table.)

$\frac{d\phi}{dx}$ or the field will have the following values at x = 0 and will decrease linearly to 0 at x = D

$$-\frac{d\phi}{dx}\Big|_o = +4.75 \times 10^{-4} \times 10^9 (.7 - .08 - V)^{1/2} = 4.75 \times 10^5 (.62 - V)^{1/2}$$

$V =$	.22	0	-1	-3	-10	-30	-100	-200
$-10^{-6} \frac{d\phi}{dx}\Big\|_{x=o} =$	.30	.38	.60	.90	1.55	2.64	4.76	6.73

Thus the field near the metal surface in the Ge will be about .4 million volts/cm at zero applied field and will be about 5 million volts/cm for 100 volts in the reverse direction ($V = -100$). We shall see below that such high fields will have appreciable effects in modifying the current voltage curves.

Fig. 5-3 shows how the field or $\frac{d\phi}{dx}$ varies with x, the distance in the semiconductor from the metal surface, for a series of values of V. This figure is for $N_d = 10^{18}$. It can be applied to any other values of N_d provided the field scale is multiplied by $(N_d/10^{18})^{1/2}$ and the x scale is multiplied by $(10^{18}/N_d)^{1/2}$. Thus if N_d were 10^{16}, the field for any value of V at $x = 0$ would be reduced by a factor of 10 and the distance at which the field would be reduced to zero (i.e., the distance D) would be increased by a factor of 10.

To obtain ϕ as a function of x we combine equations (5-3a) and (5-6a)

$$\phi = (\phi_e + V) + (\phi_o - \phi_e - V) \left[1 - \frac{x}{4.2 \times 10^3 \left(\frac{\phi_o - \phi_e - V}{N_d}\right)^{1/2}}\right]^2 \qquad (5\text{-}2b)$$

To obtain data for a plot of ϕ vs x at any V, we prepare a table. The last row is calculated from Eq. (5-2a).

$\frac{x}{D}$	= 0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.
$(1-\frac{x}{D})^2$	= 1	.81	.64	.49	.36	.25	.16	.09	.04	.01	0

For $V = -100$;

$10^6 x$	= 0	4.22	8.44	12.6	16.9	21.1	25.3	29.5	33.7	38	42.
	= .7	-18.4	-35.4	-50.5	-63.7	-74.7	-83.8	-90.8	-95.9	-98.9	-99.

Fig. 5-4 is a plot of ϕ vs x for a series of values of V. ϕ is the work in equivalent volts required to take an electron from the Fermi Level in the metal and place it in the bottom of the conduction band at a distance x in the

semiconductor. This figure is for $N_d = 10^{18}$. It can be applied to any other value of N_d provided the x scale is multiplied by $(10^{18}/N_d)^{1/2}$.

Effect of Image Forces

Now consider the work an electron must do to escape from the metal into the semiconductor. At any distance x from the metal surface, the force on the electron is given by the image law.

$$\text{Force} = \frac{e^2}{K(2x)^2} = \frac{e^2}{4Kx^2} \qquad (5\text{-}8)$$

The work per charge on the electron or the pseudopotential P is given by

$$P = \frac{e}{4K}\left(\frac{1}{x_o} - \frac{1}{x}\right) = P_o - \frac{e}{4K}\,\frac{1}{x} \qquad (5\text{-}9)$$

Where x_o is the maximum excursion from the surface for an average electron rotating in its orbit at 0°K. The excursion is measured from a plane passing through the nuclei of the outermost layer of atoms.
Potentials tending to take the electron from the Gc toward the metal are considered positive. This equation is valid only from x_o to ∞.

When a field of F' e.s. volts/cm is applied or exists in the Ge near the metal, the P vs x curve will be lowered by F' x or if F' depends on x by $\int_{x_o}^{x} F'dx$.

At some critical distance $x = x_c$, $\frac{dP'}{dx} = \frac{e}{4Kx_c^2} = F$ (5-10)

so that $x_c = \left(\frac{e}{4KF}\right)^{1/2}$

At $x = x_c$ the (P - F x) vs x curve will have a maximum. Beyond x_c the (P - F x) vs x curve will have a negative slope. Hence any electron which has enough energy to reach the maximum will go over it. This means that the work function ϕ_o' is reduced by an amount $\Delta\phi'$ given by

$$\Delta\phi' = -\,2F'\,x_c = \left(\frac{F'e}{K}\right)^{1/2} \qquad (5\text{-}11)$$

(Proof: ϕ_o is reduced for 2 reasons (1) The electron has had work done on it in getting to x_c equal to $F\,x_c = \left(\frac{Fe}{4K}\right)^{1/2}$ and (2) the electron no longer needs to do work in going from x_c to ∞: this amounts to $\frac{e}{4Kx_c} = \left(\frac{Fe}{4K}\right)^{1/2}$.

Hence these two reductions are equal and the total reduction is $2F\,x_c = (\frac{Fe}{K})^{1/2}$).

Now the field F is given by Eq. (5-3) or (5-3a) and if x_c is small compared to D we can use Eq. (5-3a) for the field at x = o. Hence F = -2CD and from Eq. (5-11)

$$\Delta\phi' = (\frac{e}{K})^{1/2} (2CD)^{1/2} = (\frac{e}{K})^{1/2} 2^{1/2} C^{1/4} (\phi_o' - \phi_e' - V')^{1/4}$$

Putting in expressions for C and D from Eqs. (5-5) and (5-6)

$$\Delta\phi' = (\frac{e}{K})^{1/2} 2^{1/2} \left\{\frac{2\pi N_d e}{K}\right\}^{1/4} (\phi_o' - \phi_e' - V')^{1/4}$$

$$= 2^{1/2}\, 2^{1/4}\, \pi^{1/4}\, e^{3/4} \frac{N_d^{1/4}}{K^{3/4}} (\phi_o' - \phi_e' - V')^{1/4}$$

$$= 2.3 \times 10^{-7} \frac{N_d^{1/4}}{K^{3/4}} (\phi_o' - \phi_e' - V')^{1/4} \qquad (5\text{-}12)$$

In equation (5-12), $\Delta\phi'$, ϕ_o', and V' are in e.s.u. If $\Delta\phi$, ϕ_o, and V are in volts, then the numerical constant becomes

$$\frac{300}{(300)^{1/4}} \times 2.3 \times 10^{-7} = 1.66 \times 10^{-5}$$

and

$$\Delta\phi = 1.66 \times 10^{-5} \frac{N_d^{1/4}}{K^{3/4}} (\phi_o - \phi_e - V)^{1/4} \qquad (5\text{-}13)$$

For Ge K = 16 (Brattain + Briggs)
(Torrey + Wittmer give 13)
Hence, $\Delta\phi = 2.08 \times 10^{-6} (N_d)^{1/4} (\phi_o - \phi_e - V)^{1/4}$ (5-14)

If for a particular sample of Ge, $N_d = 10^{18}$

Then $$\Delta\phi = \frac{1.66 \times 10^{-5} (10^{18})^{1/4}}{16^{3/4}} (\phi_o - \phi_e - V)^{1/4}$$

$$= .066 (\phi_o - \phi_e - V)^{1/4} \qquad (5\text{-}14a)$$

For V = - 10 volts $\qquad \Delta\phi = .066\ (.6 + 10)^{1/4} = .119$ volts

So that ϕ_o is reduced from .70 to .581 volts. As we shall see later this will permit 100 times as many electrons to go from the metal to the semiconductor as would go if $\Delta\phi$ were zero.

For V = - 100 volts $\qquad \Delta\phi = .209$ volts

CAPTIONS FOR FIGURES

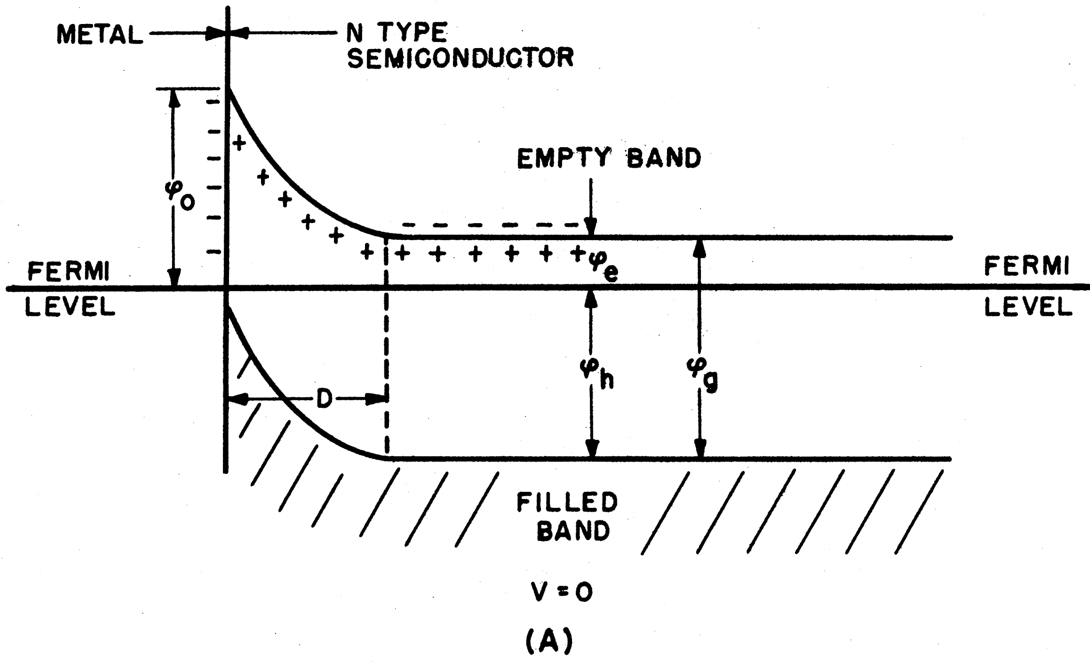
Fig. 5-1 (a) Energy Level Diagram for a Rectifying Metal-to-N-Type Semiconductor Boundary. No voltage applied.

(b) Same, but with +3 volts applied to the semiconductor.

Fig. 5-2 Plot of Barrier Thickness vs. Applied Voltage for the Barrier of Fig. 5-1 for $N_d = 10^{18}$ cm^{-3}; K = 16.

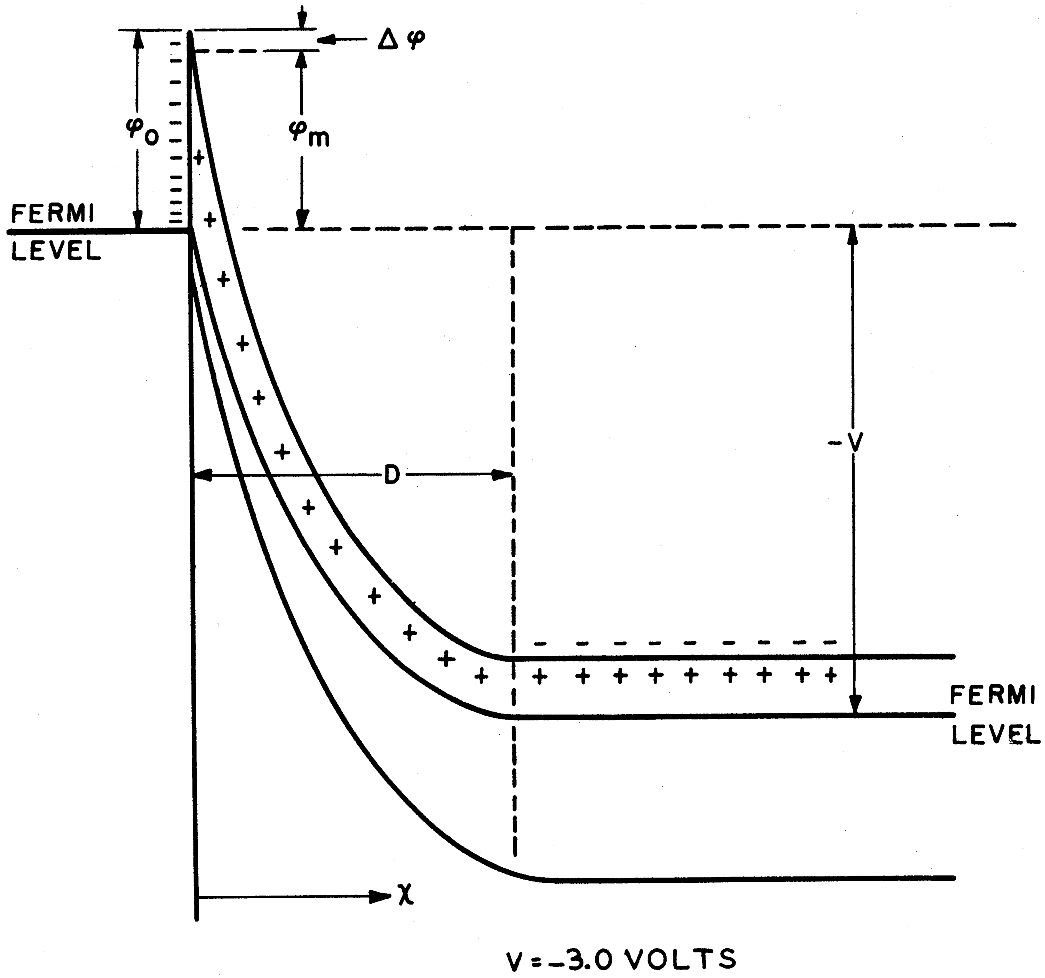
Fig. 5-3 Plot of Barrier Field vs. Distance from the Metal Surface for the Barrier of Fig. 5-1 for $N_d = 10^{18}$ cm^{-3}; K = 16; and various values of V

vs. x for th
ied Voltages
uivalent vol
spect to the



V = 0

(A)



V = −3.0 VOLTS

(B)

FIG. 5-1

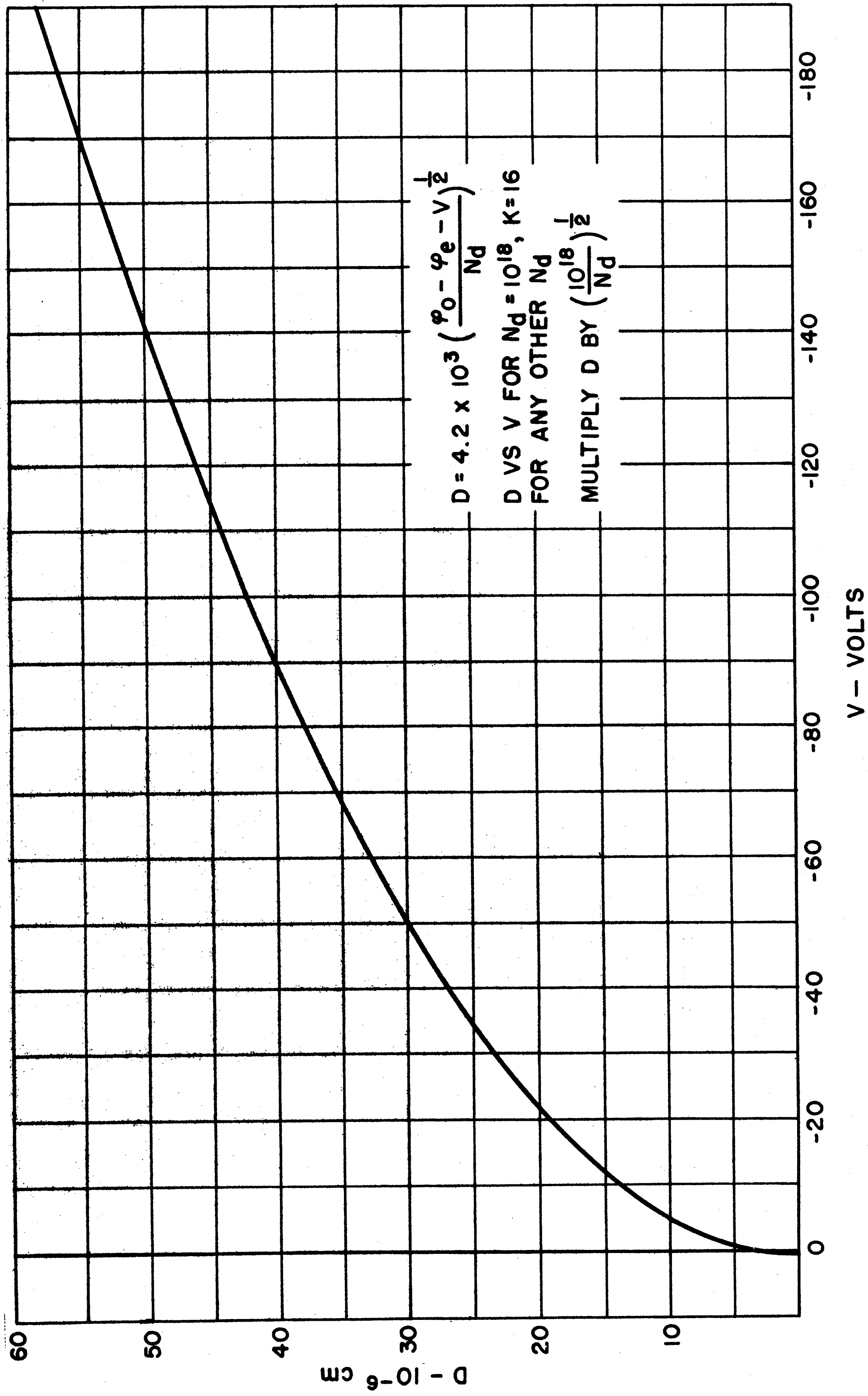


FIG. 5-2

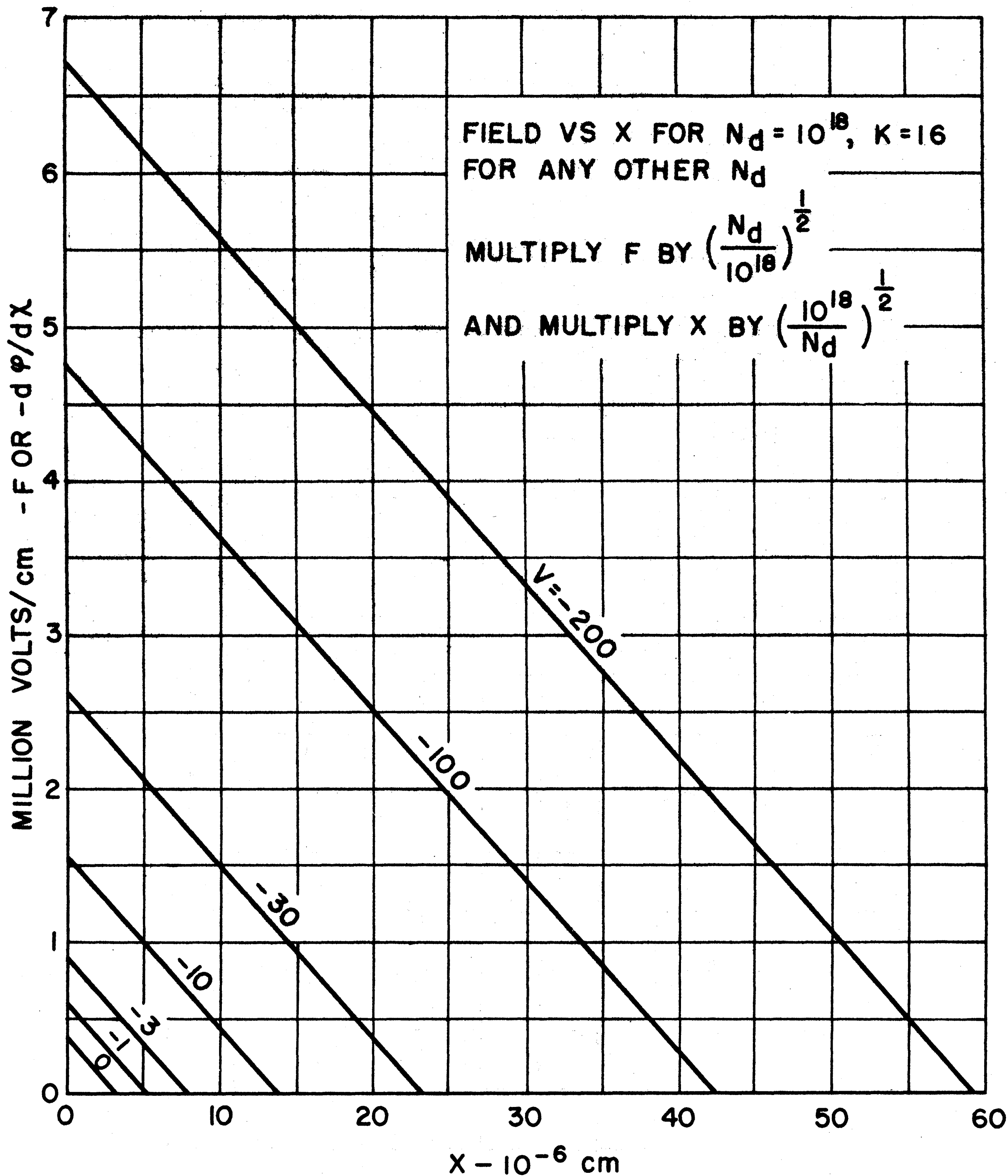


FIG. 5-3

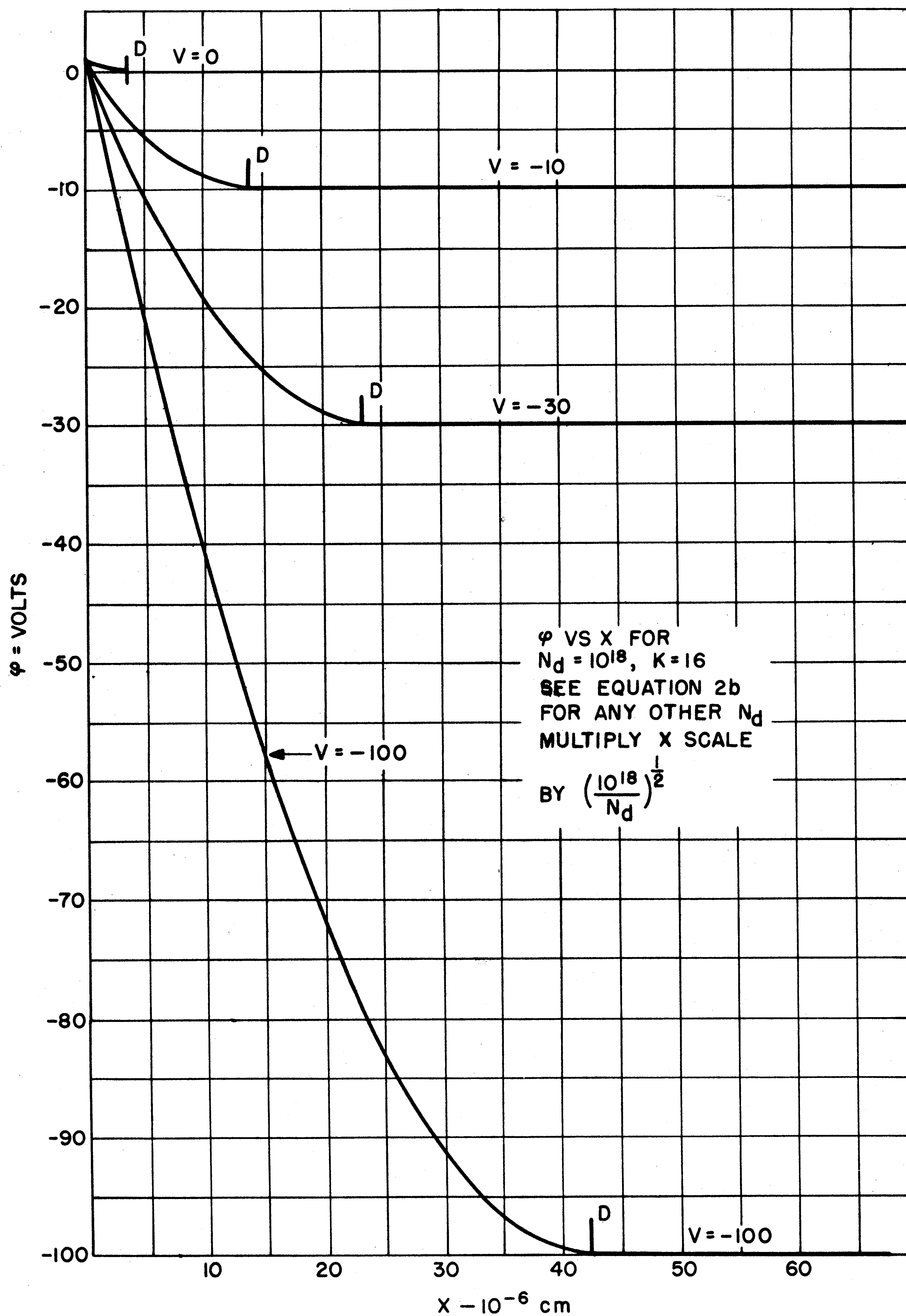


FIG. 5-4

LECTURE 6

ON ELECTRON AND HOLE CURRENTS IN BARRIERS

Consider a barrier layer in a semiconductor in contact with a metal at zero applied potential ($V = 0$). An electron potential vs distance curve for such a barrier is shown in Fig. (6-1). Then consider what electron and hole currents can flow in such a barrier. In the semiconductor beyond the distance x_c there are some electrons that have energies greater than ϕ_m equivalent volts. Of these, those that are directed toward the metal and originate in a distance greater than x_c and less than $x_c + \lambda$ (λ = mean free path), will constitute an electron current from the semiconductor to the metal. Call this j_2, the current in amperes per cm^2. Similarly in the metal there are electrons having energies greater than ϕ_m equivalent volts. These produce an electron current j_1 from the metal into the semiconductor. Similarly in the metal there are holes that have energies greater than ϕ_h electron volts, are directed away from the metal surface and have a mean free path greater than D. These result in a hole current j_3 from the metal into the semiconductor. In the semiconductor beyond D, there are holes with energies greater than ϕ_h electron volts. Those that originate in a region D to D + λ and are directed toward the metal will cross the plane at D and will be pulled to the metal by the field which exists in the region D. These constitute a hole current j_4 from the semiconductor to the metal. Finally there are hole-electron pairs generated in the region D by thermal agitation. Any electron thus generated will be pulled toward the semiconductor by the high fields in D and the hole will be pulled toward the metal. Let j_5 be the current resulting from the electron-hole pairs generated in D. There must also be a current j_6 resulting from electrons diffusing from the semiconductor into the region D and recombining with holes which have diffused from the metal into D. In equilibrium $j_1 = j_2$; $j_3 = j_4$; and $j_5 = j_6$ so that the net current is zero.

When the electron potential of the semiconductor is <u>raised</u> with respect to the metal (V is positive*) $j_2 > j_1$; $j_3 > j_4$; and $j_6 > j_5$ and a net current flows in the forward direction. Similarly when the electron potential of the semiconductor is lowered with respect to the metal (V is negative) $j_1 > j_2$;

---

*An electron potential has the opposite sign from the conventional potential which is defined in terms of a unit positive test charge.

$j_4 > j_3$; and $j_5 > j_6$. There will then be a net current in the reverse direction.

Our problem then is to deduce quantitative expressions for these six currents. We will start with j_1. In the metal the number of electrons per cm^3 having energies greater than ϕ_m is given by

$$N_e = 2UT^{3/2}\varepsilon^{-\phi_m e/kT} = 2\left(\frac{2\pi mk}{h^2}\right)^{3/2} T^{3/2}\varepsilon^{-\phi_m e/kT} \qquad (6\text{-}1)$$

The number of electrons crossing one cm^2 in any plane in one direction is equal to $N_e\bar{v}$ where $\bar{v}$ is the root mean square velocity of the electrons. If the electrons could move only in the direction x normal to the plane, then $\bar{v}$ would be given by

$$\frac{1}{2}m(\bar{v}_x)^2 = \frac{1}{2}kT \text{ or } \bar{v}_x = \left(\frac{kT}{m}\right)^{1/2} \qquad (6\text{-}2)$$

If the components of velocity in the y and z directions are taken into account

$$\bar{v} = \left(\frac{kT}{2\pi m}\right)^{1/2} \qquad (6\text{-}3)$$

Hence

$$j_1 = \frac{4\pi mk^2e}{h^3}T^2\varepsilon^{-\phi_m e/kT}$$

$$= 120T^2\varepsilon^{-\phi_m e/kT} \qquad (6\text{-}4)$$

where 120 amperes/cm^2 $°K^2$ is the numerical value of the coefficient. This equation is the same as the well known Richardson-Dushman equation for thermionic emission.

Next consider j_2. If the applied potential were zero, $j_1 = j_2$. If an electron potential V is applied to the semiconductor and V is positive, the Fermi Level in the semiconductor is raised and the electrons in it which approach the metal see a potential hill of $\phi_m - V$. This is true only as long as $\phi_m > V$. If V is negative, the Fermi Level in the semiconductor is lowered and the potential hill becomes $\phi_m + V$. Hence we can write

$$j_2 = 120\ T^2\varepsilon^{-(\phi_m - V)e/kT} \qquad (6\text{-}5)$$

in which V is positive for the conducting direction and negative for the reverse direction.

Similarly

$$j_3 = 120\ T^2\ \varepsilon^{-(\phi_h - V)e/kT} \qquad (6\text{-}6)$$

and

$$j_4 = 120\ T^2\ \varepsilon^{-\phi_h e/kT} \qquad (6\text{-}7)$$

In Eq. (6-5) and (6-6) it is assumed that the mean free path λ is comparable with the distance D. Hence we must examine whether this assumption is justified in the case of Ge to which we will later wish to apply these equations. While λ cannot be measured directly, it is related to the mobility which can be measured. This relationship is given by

$$\lambda = \frac{3}{4}\frac{b}{e}(2\pi mkT)^{1/2} * \qquad (6\text{-}8)$$

$e = 4.8 \times 10^{-10}$ e.s.u.; m = mass of electron $= 9 \times 10^{-28}$ gms;
k = Boltzmann's constant $= 1.37 \times 10^{-6}$ ergs/°K
b = mobility (= 1500 cm/sec/volt/cm) $= 4.5 \times 10^{5}$ cm/sec/e.s.volts/c

(For impure Ge it is slightly less); at T = 300°K.

Putting in these values we find $\lambda = 1.07 \times 10^{-5}$ cm., which is comparable with D in those cases in which we will want to use Eqs. (6-5) and (6-6). However, Bethe and Herzfeld** have pointed out that λ need not be as large as D but need be equal to or larger than the distance in which the potential char $\frac{kT}{e}$ volts or about 1/40 volt. This requirement is likely to be satisfiec samples of Ge.

The net electron current $j_e \equiv j_2 - j_1$ is given by

$$j_e = 120\, T^2 \varepsilon^{-\phi_m e/kT} \left(\varepsilon^{+Ve/kT} - 1\right) \qquad (6\text{-}9)$$

The net hole current $j_h \equiv j_3 - j_4$ is given by

$$j_h = 120\, T^2 \varepsilon^{-\phi_h e/kT} \left(\varepsilon^{+Ve/kT} - 1\right) \qquad (6\text{-}10)$$

From Eqs. (6-9) and (6-10) it follows that if $\phi_m > \phi_h$, $j_h > j_e$ and the current through the barrier will be predominantly hole current. This is found by experiment to be the case for some metal point contacts on N type Ge. If, on the other hand, $\phi_h > \phi_m$, $j_e > j_h$ and most of the current would be electron current.

In Eq. (6-10), V appears only in the term $\varepsilon^{+Ve/kT}$ since ϕ_h does not depend on V. However in Eq. (6-9) ϕ_m is an implicit function of V. For Ge with a dielectric constant of 16, $\Delta\phi$ in volts is given by Eq. (5-14) of the previous section, namely

$$\Delta\phi = 2.08 \times 10^{-6} (N_d)^{1/4} (\phi_o - \phi_e - V)^{1/4}$$

- -

* See Torrey + Whittmer - Crystal Rectifier p. 54

** See Torrey-Whittmer p. 82

Hence Eq. (6-9) becomes

$$j_e = 120\ T^2 \varepsilon^{-\phi_o e/kT}\ \varepsilon^{+2.08 \times 10^{-6}(N_d)^{1/4}(\phi_o - \phi_e - V)^{1/4}(e/kT)} (\varepsilon^{+Ve/kT} - 1) \qquad (6\text{-}11)$$

Before values of j_h and j_e can be computed for various voltages with N_d as a parameter, it is necessary to compute valuesfor ϕ_h, ϕ_e, and ϕ_o for various values of N_d. The results of such computations for a particular case are given in Table 6-1 below. For any N_d, ϕ_e can be computed as a function of T by methods given in the section on the Fermi Level. In particular ϕ_e will depend on V_d, the equivalent voltage required to take an electron from a neutral donor and place it in the bottom of the conduction band. For V_d we have rather arbitrarily chosen .2 volts because a plot of V_F vs T with N_a as a parameter had previously been computed. It is given in Fig. (3-6) in the section on the Fermi Level. From this figure and for T = 300°K we have taken the values of ϕ_e given in the table. Since $\phi_h = \phi_g - \phi_e$, ϕ_h was computed using ϕ_g = .75 volts for Ge. ϕ_o is probably dependent on the work function of the metal points used to make contact with the Ge and on the kind and amount of atoms absorbed on the surface of the Ge. Lacking any better procedure, we have arbitrarily assumed that $\phi_o = \phi_h$.

Table 6-1 for T = 300°K

N_d =	10^{15}	10^{16}	10^{17}	10^{18}	10^{19}
ϕ_e =	.26	.22	.18	.15	.12
ϕ_{go} =	.75	.75	.75	.75	.75
ϕ_h =	.49	.53	.57	.61	.63
ϕ_o =	"	"	"	"	"
$\phi_o - \phi_e$ =	.23	.31	.39	.46	.51

Fig. (6-2) shows plots of log V vs log j_h for Eq. (6-10) for a series of values of N_d, the number of donors/cm^3. While ϕ_h does not depend on V, it does depend on N_d. For each curve there is a forward direction branch for positive values of V and a reverse direction branch for negative values of V. For very small values of V these two branches merge together in a 45° line which means that Ohm's law is obeyed. For negative values of V (reverse voltages), j_h becomes independent of V for V greater than about .1 volt.

Fig. 6-3 shows log V vs log j_e for Eq. (6-11), for a series of values of N_d. In this case, V appears in two exponential factors and ϕ_e is an implicit function of N_d. The current in the reverse direction no longer saturates with V but increases at a faster and faster rate as V increases. The reverse voltage at which the current begins to markedly deviate from its saturation value decreases as N_d increases. Hence from the shape of the reverse current curve one can surmise whether one is dealing with predominantly hole or electron current and get some idea of the density of donors.

Significance of Ohmic Resistance at Low Voltages

Let us now examine what information can be deduced from the ohmic resistance at low applied voltages. Consider first the simplest case, namely, one for which $\phi_m > \phi_h$. We then use Eq. (6-10). When V is so small that $Ve/kT < 1$,

$$\varepsilon^{Ve/kT} = 1 + Ve/kt$$

and

$$j_h/V = 120\ T^2\ \varepsilon^{-\phi_h e/kT}\ (e/kT) \qquad (6\text{-}12)$$

or

$$R_o \equiv V/j_h = (kT/e)\ \varepsilon^{+\phi_h e/kT} / 120\ T^2 \qquad (6\text{-}13)$$

Since ϕ_h is the only unknown, it can be determined from an experimental value of R_o at temperature T. By experimentally determining R_o as a function of T, one can calculate ϕ_h as a function of T. Since $\phi_h + \phi_e = \phi_g$ = about .75 volts for Ge, one can compute ϕ_e as a function of T (caution: ϕ_g may decrease slightly as T increases). Since the variation of ϕ_e with T depends on N_d one can deduce values of N_d. To do this, one must know the contact area between the metal and the semiconductor since R_o is the resistance per unit area of contact.

A more complicated case is that for which $\phi_h > \phi_m$ for which j_e is given by Eq. (6-11). For low enough values of V this can be written

$$j_e = 120\ T^2\ \varepsilon^{-\phi_o e/kT} \left[1 + 2.08 \times 10^{-6} (N_d)^{1/4} (\phi_o - \phi_e - V) e/kT\right]\ Ve/kT \qquad (6\text{-}14)$$

For $Nd = 10^{16}$, the bracket has a value of 1.4 and

$$R_o \equiv \frac{V}{j_e} = \frac{kT}{e}\ \varepsilon^{\phi_o e/kT} / (120\ T^2 \times 1.4) \qquad (6\text{-}14a)$$

Hence ϕ_o can be evaluated from an experimental value of R_o. If ϕ_o and N_d are both unknown it is not possible to evaluate either from such an experiment.

However if N_d is known from any subsidiary experiment, ϕ_e is determined, and ϕ_o can be uniquely calculated.

If ϕ_m and ϕ_h are nearly equal, then j_e and j_h are comparable with one another. In this case it is difficult to uniquely determine ϕ_m and ϕ_h. However if either one can be determined from an independent experiment, the other can be calculated from the value of R_o.

Current due to Thermal Generation of Hole Electron Pairs

Besides the electron and hole currents j_1 to j_4 which enter the space charge region D from the metal and semiconductor sides, there must be an additional current due to the thermal generation of hole electron pairs in the region D. In this section we derive expressions for this current as a function of the applied voltage and the density of donors N_d. It turns out that this current is comparatively small for small values of N_d and V; but becomes the dominant current at high reverse voltages. The larger the values of N_d, the smaller is the value of -V at which the current j_5 exceeds the other currents.

First we need an expression for the number of hole electron pairs generated per cm^3 per sec. Call this quantity G. In the appendix we prove that G is given by*

$$G = \left(\frac{3k}{m}\right)^{1/2} 4 \left(\frac{2\pi mk}{h^2}\right)^3 S\, T^{7/2} \varepsilon^{-\phi_g e/kT}$$

$$= 1.58 \times 10^{37} S\, T^{7/2} \varepsilon^{-\phi_g e/kT} \qquad (6\text{-}15)$$

in which S is the collision cross-section for a hole and an electron. In Ge the value of $S = 2.7 \times 10^{-18}$ cm^2. Hence in Ge

$$G = 4.27 \times 10^{19}\, T^{7/2} \varepsilon^{-\phi_g e/kT} \qquad (6\text{-}15a)$$

For T = 300°K and ϕ_g = .75 volts

$$G = 5 \times 10^{15} \text{ electron hole pairs/cm}^3\text{/sec} \qquad (6\text{-}15b)$$

For T = 300 and ϕ_g = .70 volts

$$G = 3.5 \times 10^{16} \text{ pairs/cm}^3\text{/sec.} \qquad (6\text{-}15c)$$

Such values of G would yield values of j_5 of 5.6×10^{-8} amperes/cm^2. However as higher voltages are applied in the reverse direction and the fields in region D approach values comparable with a million volts/cm, the work required to transfer an electron from the filled to the conduction band decreases. This

---

*We are grateful to Dr. J. Bardeen for pointing out to us the method for deriving the expression for G.

means that ϕ_g decreases and G and j_5 increase. Hence we must consider how ϕ_g decreases with the field strength F.

For this we consider that the electrons in the filled band are in the ground state or in their normal orbit about the Ge nucleus or kernel; and that the electrons in the conduction band are in an excited state or in an excited orbit. Let δ be the distance an electron must go to get from the normal state to the excited state. In a field F work is done on the electron equal to $Fe\delta$ so that

$$\phi_g = \phi_{go} - F\delta \qquad (6\text{-}16)$$

where $\phi_{go} = \phi_g$ for $F = 0$

For a Ge atom in free space δ might be 2×10^{-8} cm. For solid Ge would be increased by the dielectric constant whose value is about 16, so that δ might be about 30×10^{-8} cm. For a field of 10^6 volts/cm, ϕ_g would be reduced by .30 volts; G and j_5 would be increased by a factor of 10^5 which makes j_5 comparable with the other currents.

To derive an equation for j_5 proceed as follows: Consider a lamina in Fig. 6-1 in the region D at distance x and width dx. Let Δj_5 be the contribution to j_5 due to hole electron pairs generated in this lamina.

Then

$$\Delta j_5 = 4.27 \times 10^{19}\, e\, T^{3.5}\, \varepsilon^{-\phi_{go} e/kT}\, \varepsilon^{F\delta e/kT}\, \Delta x \qquad (6\text{-}17)$$

In Eq. (5-4a) it was shown that

$$-F = F_o - (F_o/D)\, x \qquad (6\text{-}18)$$

where

$$F_o = 4.75 \times 10^{-4}\, (N_d)^{1/2}\, (\phi_o - \phi - V)^{1/2} \qquad (6\text{-}18a)$$

and

$$(F_o/D) = 1.13 \times 10^{-7}\, N_d \qquad (6\text{-}18b)$$

Hence

$$\Delta j_5 = J_5\, \varepsilon^{F_o \delta e/kT}\, \varepsilon^{-(F_o/D)\delta ex/kT}\, \Delta x \qquad (6\text{-}19)$$

and

$$j_5 = J_5\, \varepsilon^{F_o \delta e/kT} \int_0^D \varepsilon^{-(F_o/D)\delta ex/kT}\, dx \qquad (6\text{-}20)$$

in which $J_5 = 4.27 \times 10^{19} e\, T^{3.5}\, \varepsilon^{-\phi_{go} e/kT}$

Let $a = (F_o/D)\, \delta\, e/kT$

Then the integral becomes

$$\int_0^D \varepsilon^{-ax} dx = \left. \frac{\varepsilon^{-ax}}{-a} \right]_0^D = \frac{1}{a}\, (1 - \varepsilon^{-aD})$$

and $j_5 = J_5\ (\varepsilon^{aD}/a)\ (1-\varepsilon^{-aD}) = (J_5/a)(\varepsilon^{aD}-1)$

or $j_5 = \frac{J_5}{F_o\,\delta\, e/kTD}\ (\varepsilon^{F_o \delta\, e/kT}-1)$

From (6-18a) and (6-18b)

$$j_5 = \frac{5.22 \times 10^3 T^{4.5}\, \varepsilon^{-11600\phi_{go}/T}}{\delta\, N_d}\ (\varepsilon^{5.51\delta\, (N_d)^{1/2}(\phi_o - \phi_e - V)^{1/2}/T} - 1) \qquad (6\text{-}21)$$

For T = 300°K and ϕ_{go} = .75 volts

$$j_5 = \frac{183}{\delta\, N_d}\ (\varepsilon^{1.84 \times 10^{-2}\delta\,(N_d)^{1/2}\ (\phi_o - \phi_e - V)^{1/2}} - 1) \qquad (6\text{-}21a)$$

For T = 300°K and ϕ_{go} = .70 volts

$$j_5 = \frac{1300}{\delta\, N_d}\ (\varepsilon^{1.84 \times 10^{-2}\delta\,(N_d)^{1/2}\ (\phi_o - \phi_e - V)^{1/2}} - 1) \qquad (6\text{-}21b)$$

For T = 300 $\qquad \phi_{go} = .75 \qquad \delta = 30 \times 10^{-8}$ cm

$N_d = 10^{17}$ donors/cm^3 (From table 6-1 it follows that $\phi_o - \phi_e = .39$ volts)

$$j_5 = 6.1 \times 10^{-9}\, \varepsilon^{1.74(.39-V)^{1/2}} = 6.1 \times 10^{-9}\, 10^{.76(.39-V)^{1/2}} \qquad (6\text{-}21c)$$

Beyond V = -100 volts, j_5 exceeds j_e and j_h

Fig. 6-4 is a plot of log V vs log j where j stands for any of the four quantities j_h, j_e, j_5, or j_s

and $j_s = j_h + j_e + j_5$

The parameters used in calculating the curves are:

$\delta = 30 \times 10^{-8}$ $\quad N_d = 10^{17}$ $\quad T = 300°K$ $\quad \phi_{go} = .75$ volts $\quad \phi_o - \phi_e = .39$ volts

Both forward and reverse branches are shown. For all values of +V and for values of -V less than 60 volts, $j_e > j_h \gg j_5$ and the shape of the j_s curve is largely determined by j_e. Between -60 and -100 volts j_5 and j_e are nearly equal. Beyond -100 volts, $j_5 > j_e > j_h$ and the sum current is dominated by j_5.

It will be apparent that if other values had been assigned to ϕ_o, the

relative contributions of j_e and j_h to the sum curve at low or medium values of -V would have been quite different. If the reverse current between .1 and 10 volts saturates very well, it means that most of the current is carried by holes. If the reverse current between .1 and 10 volts increases by a factor of 1.5 to 60, it means that most of the current is carried by electrons; the greater the factor, the larger the values of N_d, the concentration of donor impurities. If the reverse current increases very rapidly with voltage (factors of 1000 for doubling the voltage) the current is mainly due to hole-electron pair generation; the lower the voltage at which this rapid increase of current begins, the greater will be the value of N_d.

Fig. 6-5 shows computed curves for log V vs log j_s for a family of values of N_d. For these curves the following values have been assigned to the constants: $T = 300°K$, $\phi_{go} = .75$ volts, $\phi_o = \phi_h$, (ϕ_h depends upon N_d, T, and ϕ_{go}) $\delta = 30 \times 10^{-8}$ cm.

Figs. 6-6, 6-7, and 6-8 are log V vs log j_5 for a family of values of N_d. Values of δ are respectively 30, 20, and 10 times 10^{-8} cm. For all of these figures $T = 300°K$, $\phi_{go} = .75$ volts, and $\phi_o = \phi_h$.

Fig. 6-9 shows some experimental curves of log V vs log j_s for the reverse direction only. The data were taken by K. M. Olsen at Bell Telephone Laboratories. They are for Ge samples and phosphor bronze points. It is estimated that the area of contact was 10^{-5} cm^2. The Ge samples were cut from the top of ingots in which varying percentages of Antimony, Sb, had been added. These values of p have been converted to rough estimated values of N_d based on Pearson's work with radioactive tracers and Hall effect measurements from which N_d values can be deduced*. Qualitatively the shapes of these curves agree with those in Fig. 6-5 and they vary with N_d as predicted. Note that for the two highest values of p or N_d, (curves 4 + 5) the reverse current shows no tendency to saturate and the reverse current increases rapidly with voltage even for voltages near 1 volt. From this one might deduce that at concentrations of donor impurities near 1 per cent or N_d of 10^{20}, the space charge fields even at 0 applied voltage are in the neighborhood of 10^6 volts per cm, and the currents at low voltages are already predominately due to hole-electron pair generation; in other words $j_5 > j_e$ or j_h.

- -

*Pearson, Struthers, and Theurer, Physical Review, 75, 1949, p. 344.

CAPTIONS FOR FIGURES

Fig. 6-1 Energy level Diagram of a Rectifying Barrier Showing the Various Currents Considered in the Text.

Fig. 6-2 Log-Voltage vs. Log-Current-Density Plots (log V vs. log j_h) for ϕ_h = 0.75 volts; T = 300° K; Several Values of N_d.

Fig. 6-3 Log-Voltage vs. Log-Current-Density Plots (log V vs. log j_e) for $\phi_e = \phi_h$; T = 300° K; Several Values of N_d. (See Table 5-1)

Fig. 6-4 Plots of Log V vs. Log j_h, j_e, j_5 and j_s, Showing How the Complete Characteristic is the Sum of the Several Contributions.

Fig. 6-5 Calculated Plots of Log V vs. Log j_s for the Constants Given, with $\delta = 30 \times 10^{-8}$ cm.

Fig. 6-6 Calculated Plots of Log V vs. Log j_5 for the Constants Given, with $\delta = 30 \times 10^{-8}$ cm.

Fig. 6-7 Calculated Plots of Log V vs. Log j_5 for the Constants Given, with $\delta = 30 \times 10^{-8}$ cm.

Fig. 6-8 Calculated Plots of Log V vs. Log j_5 for the Constants Given, with $\delta = 30 \times 10^{-8}$ cm.

Fig. 6-9 Experimental Plots of Log V vs. Log j_s for Phosphor Bronze + Germanium Rectifying Contact. Reverse Direction Only. After K.M. Olsen.

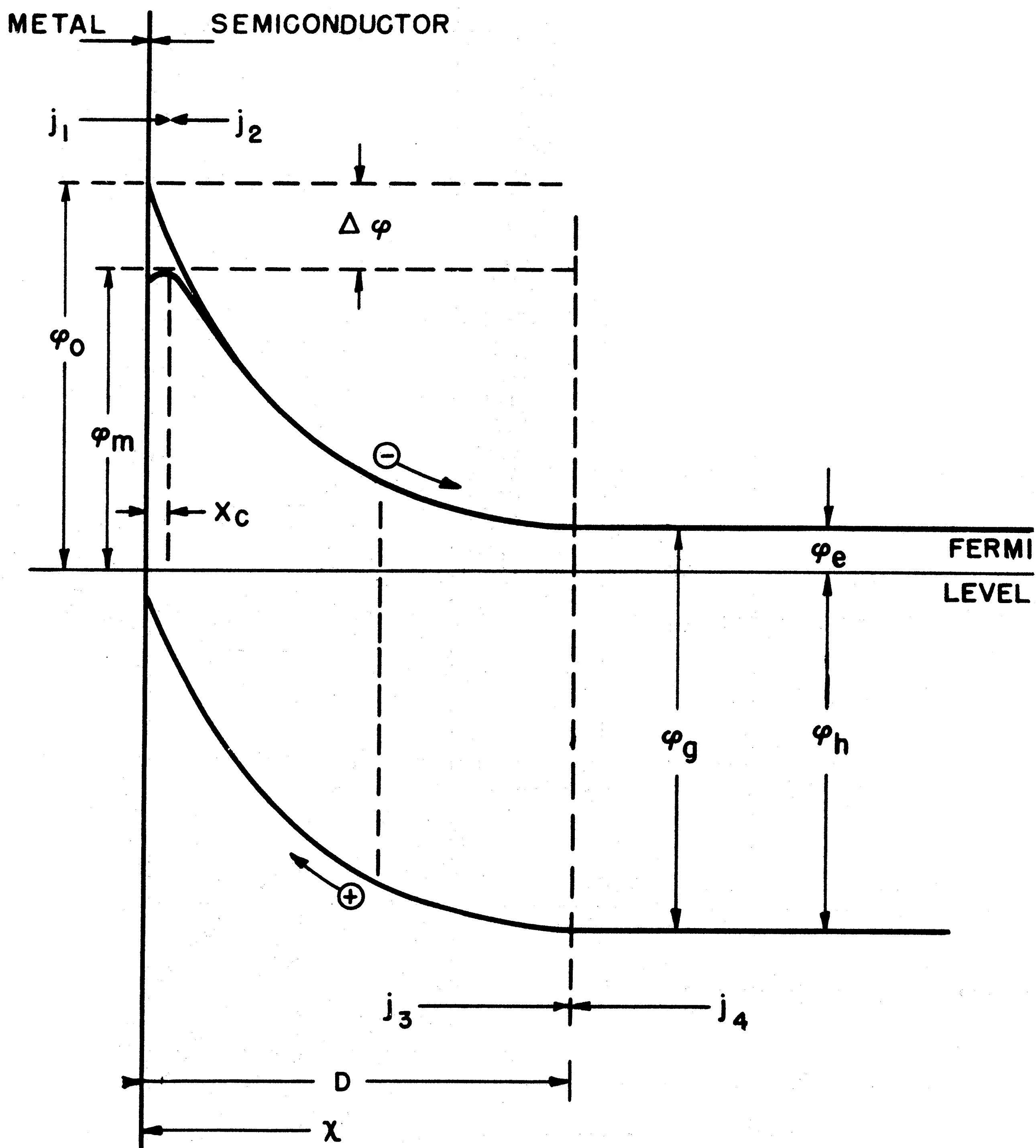


FIG. 6-1

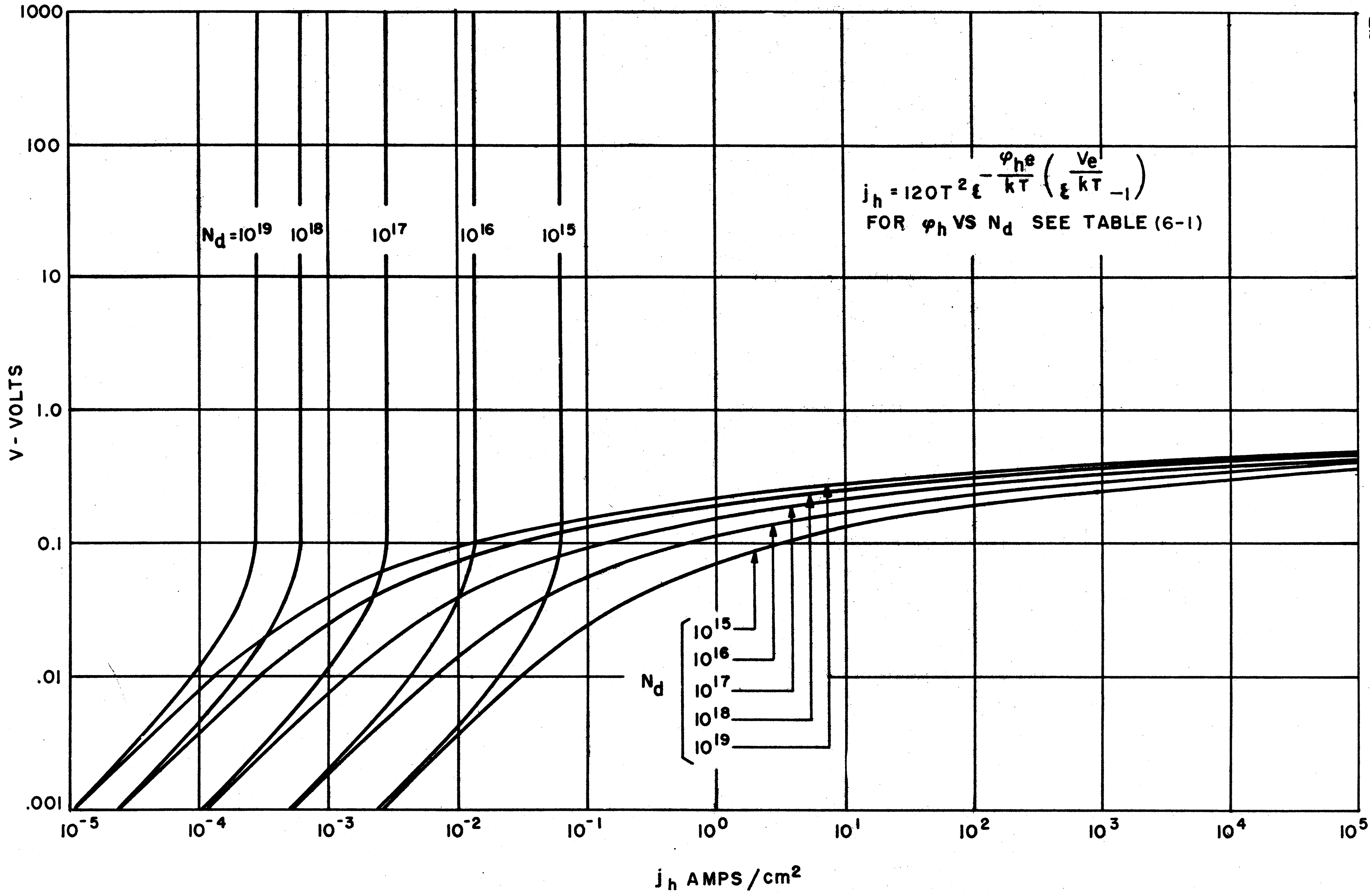


FIG. 6-2

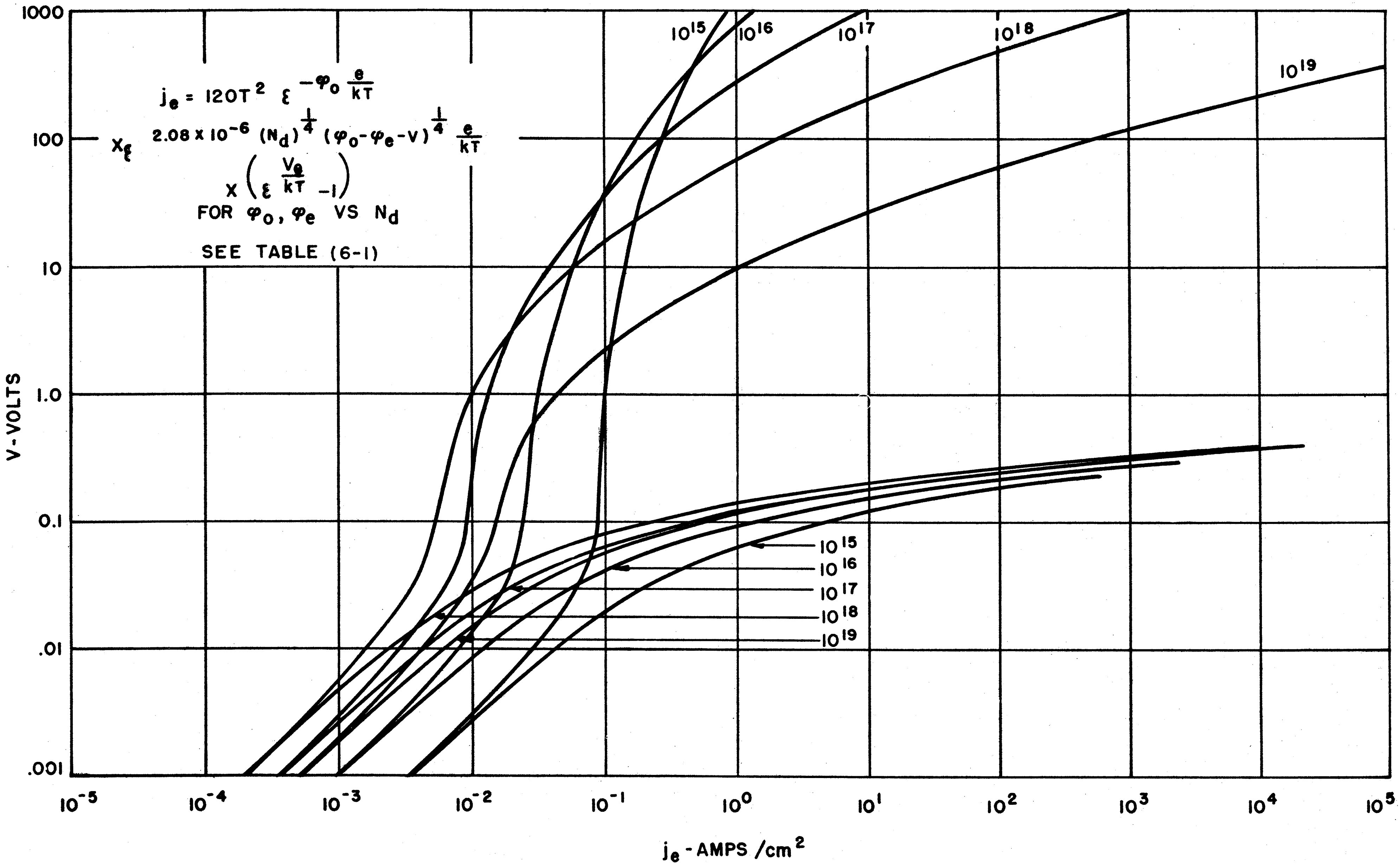


FIG. 6-3

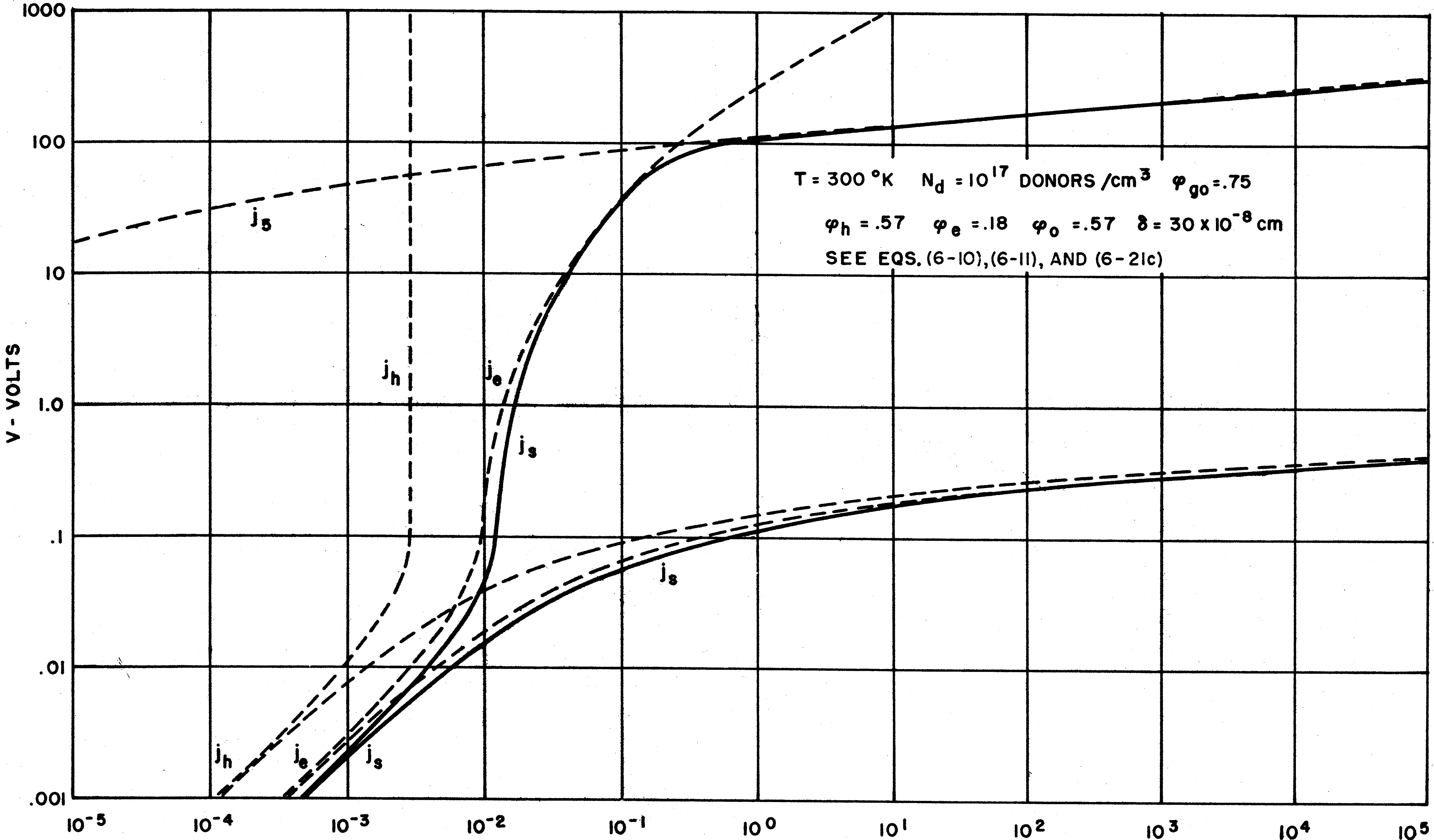


FIG. 6-4

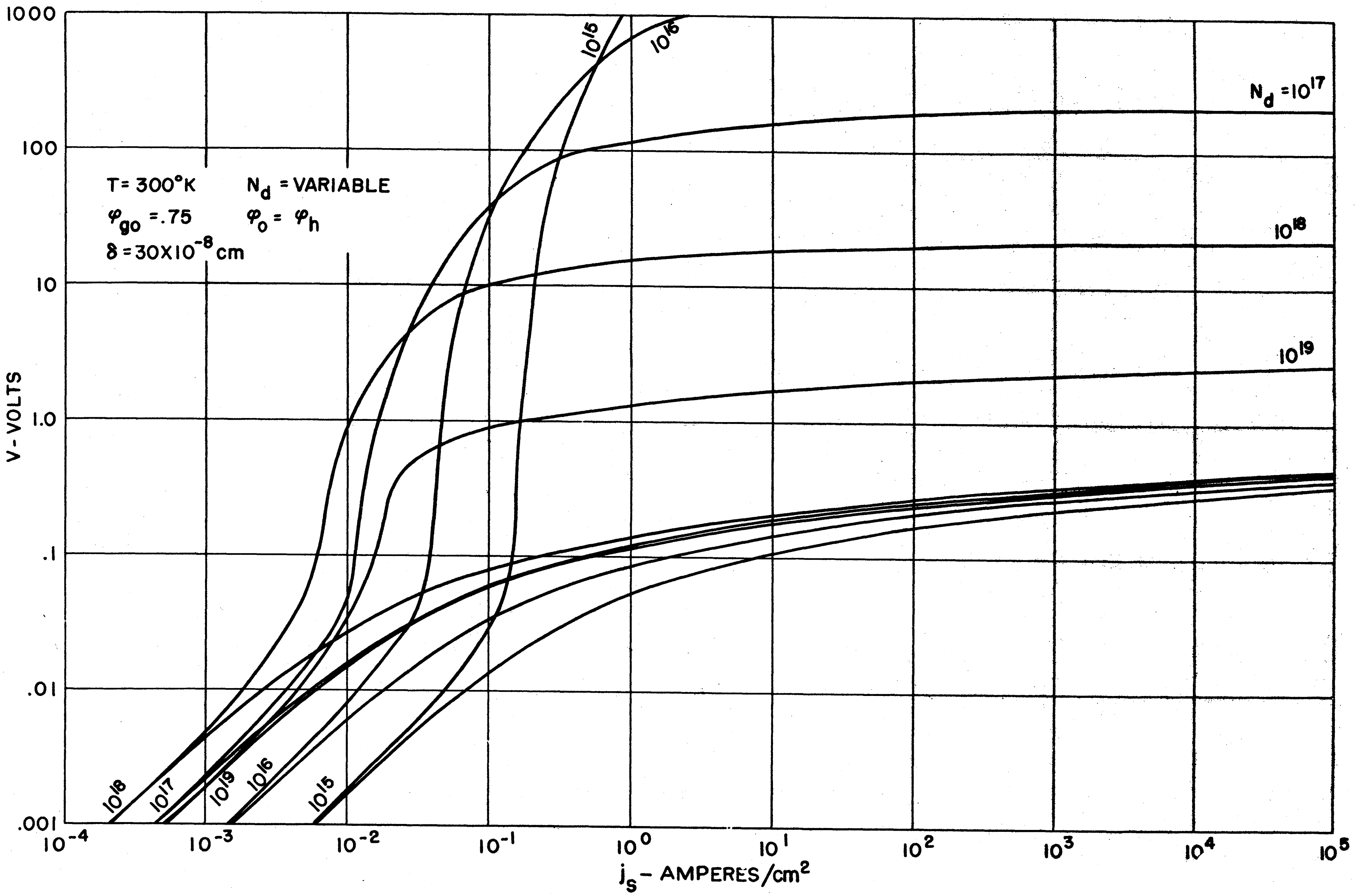


FIG. 6-5

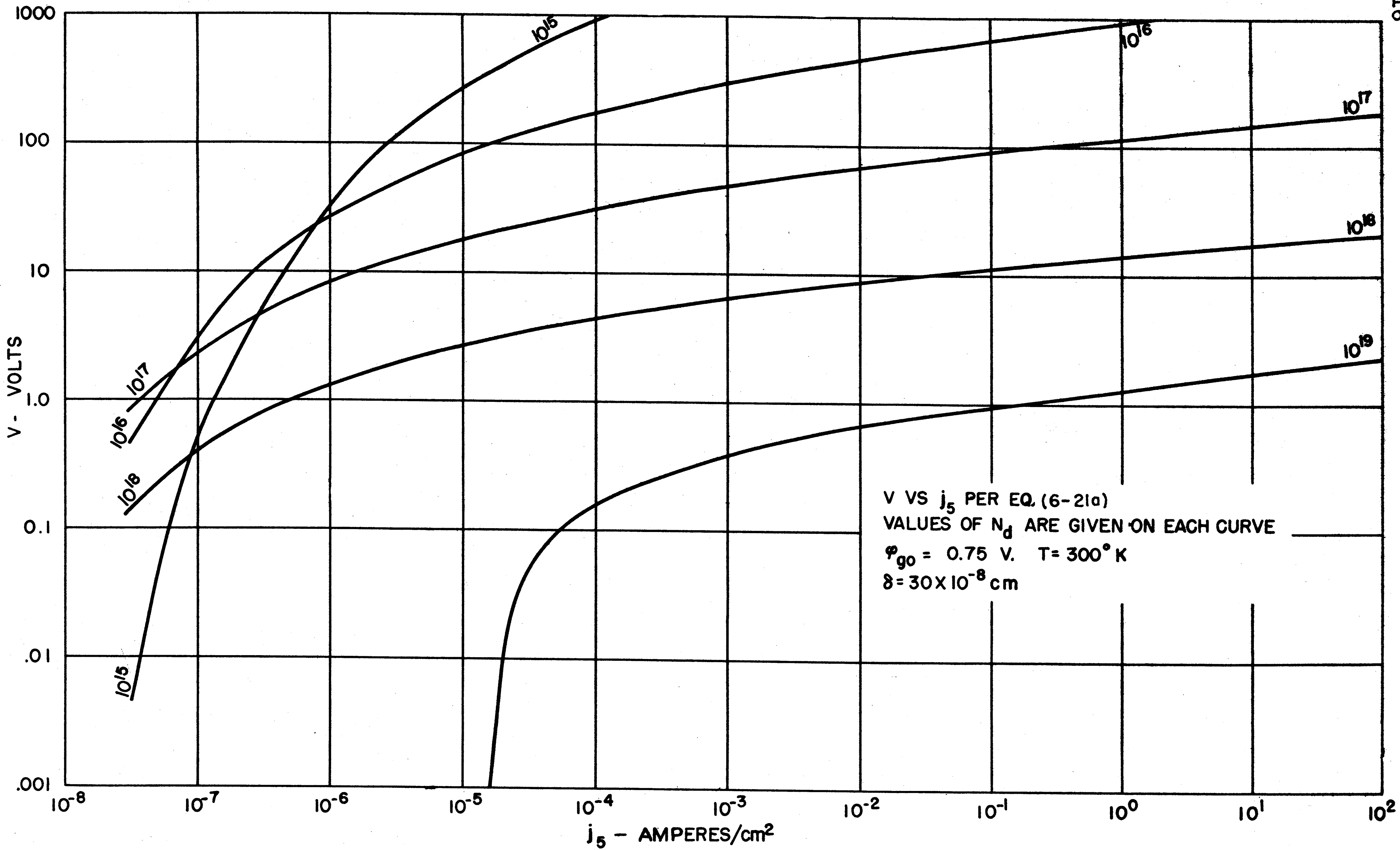


FIG. 6-6

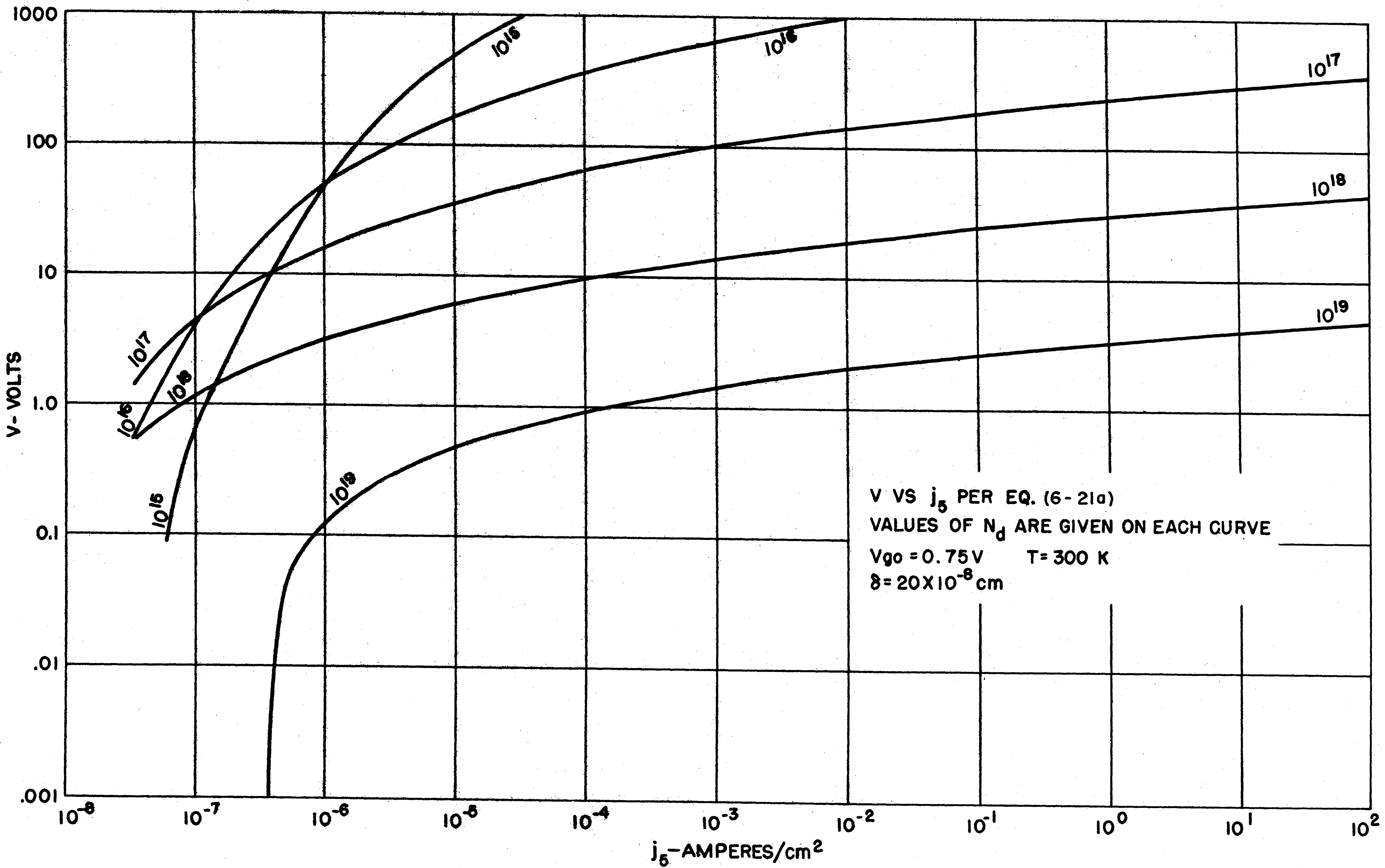


FIG. 6-7

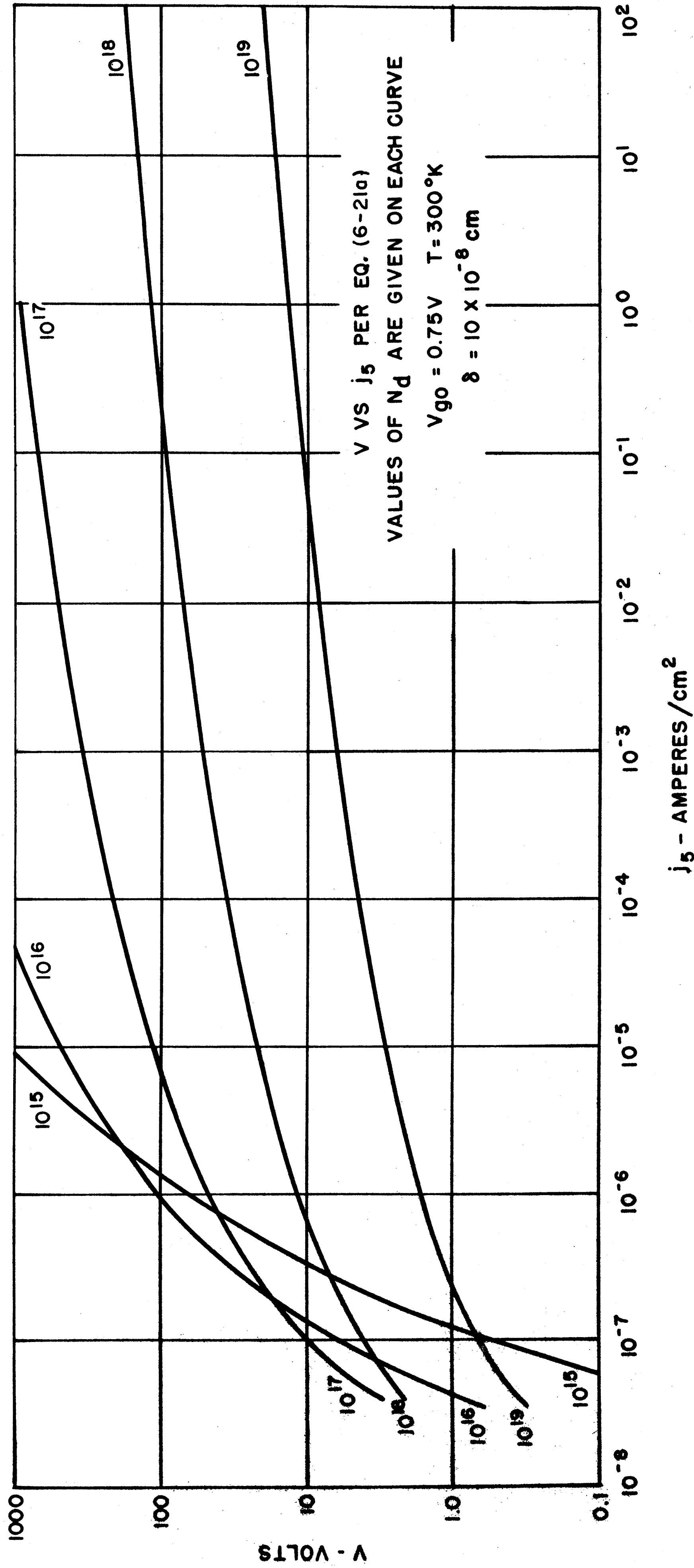


FIG. 6-8

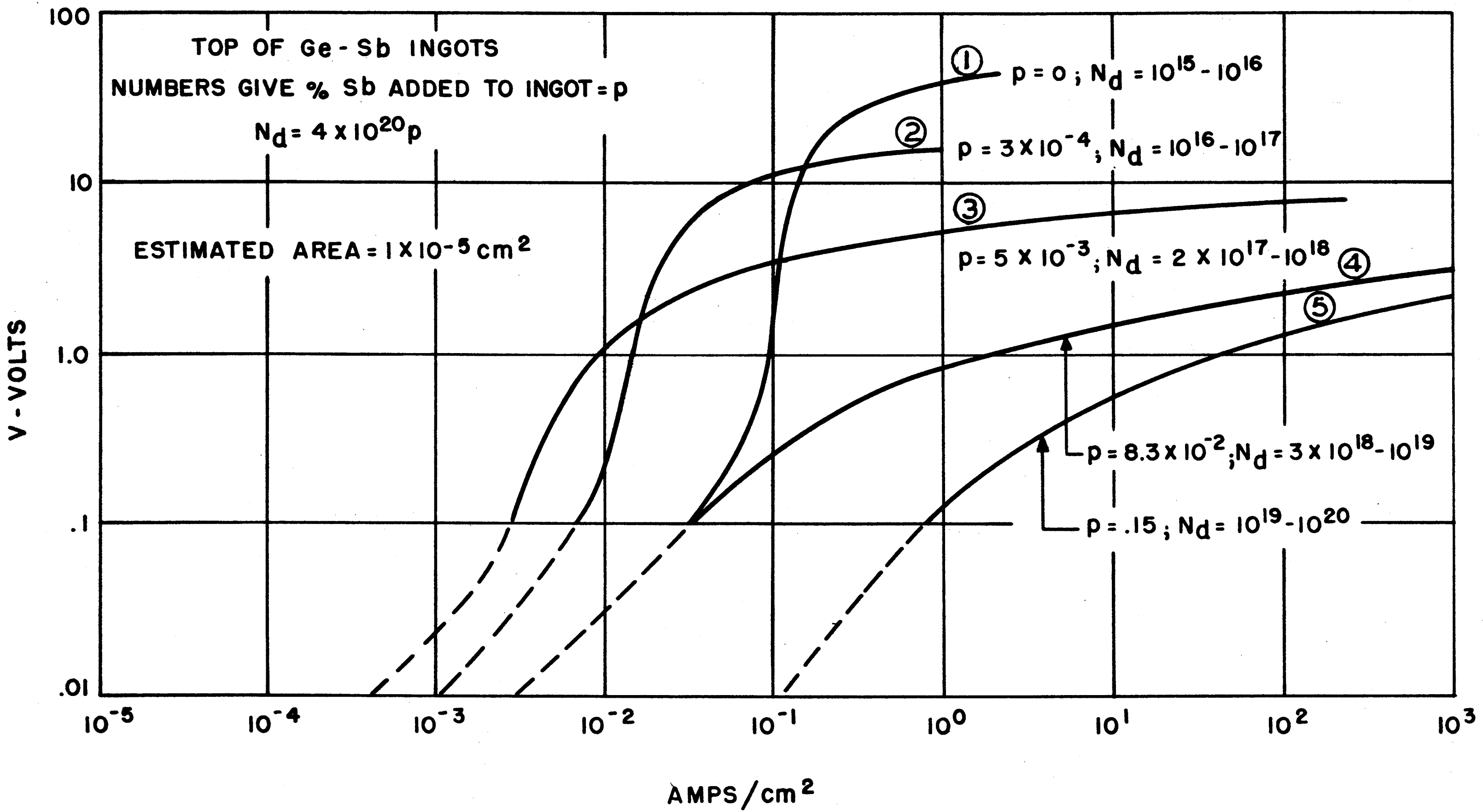


FIG. 6-9

LECTURE 7

OPTICAL PROPERTIES AND THE INNER PHOTOELECTRIC EFFECTS

GENERAL

It is well known that many oxides are highly colored, that some of them increase their conductivity when illuminated, and that some oxide rectifiers show a photovoltaic effect. While the theories of these effects are only in an early stage of development and apply only in a few simple systems, they promise to be as significant in revealing the distribution of energy states and the mechanism of the production of free carriers of current in as great a detail as do thermal effects.

The basis of the modern theory of these photoeffects is the Einstein equation:

$$h\nu = E = Ve \quad . \tag{7-1}$$

Where h is Planck's constant (6.62×10^{-34} Joule x sec); ν is the frequency of the light; E is the energy in Joules which the quantum h can impart to an electron, V is the voltage equivalent of E; and $e = 1.6 \times 10^{-19}$ coulombs. The wave length λ in cm is related to the velocity of light c(3×10^{10} cm/sec) and to ν by

$$\lambda\nu = c \quad . \tag{7-2}$$

Hence, $\lambda V = h\,c/e = 1.23 \times 10^{-4}$ volt cm. If λ is expressed in microns, (μ or 10^{-4} cm), then $\lambda V = 1.23$ volt microns. Table 7-1 gives the wave length range in microns and the corresponding voltage range for four colors:

Table 7-1

	Red	-	-	Yellow	-	Green	-	Blue	-
Microns		.77	.615		.575		.515		.41
Equivalent Electron Volts		1.6	2.0		2.15		2.4		3.0

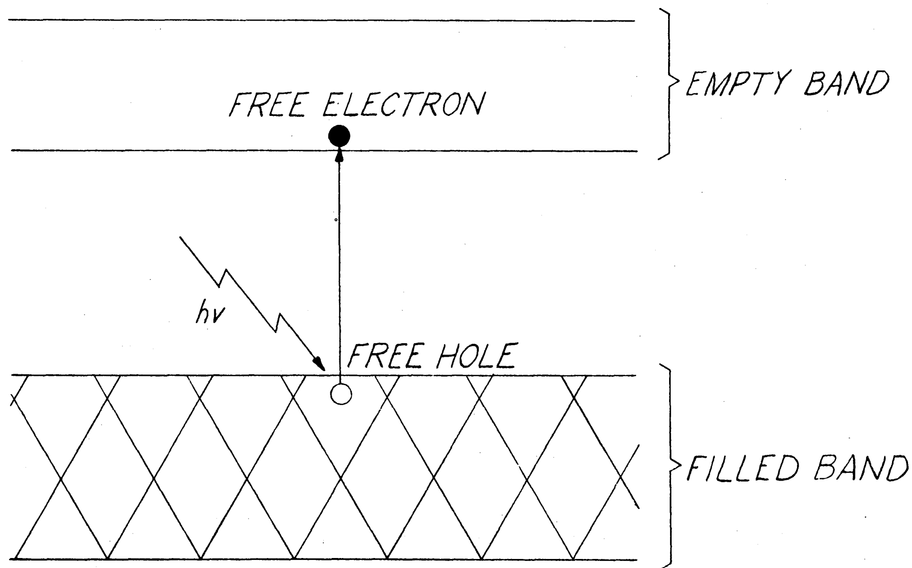


Figure 7-1 Energy Band Diagram Illustrating the Mechanism of the Photoconductivity Effect in an Intrinsic Semiconductor

The photoelectric properties of semiconductors can similarly be elucidated in terms of the band picture. Consider the case of an intrinsic semiconductor shown in Figure 7-1 . When such a material is exposed to light having a quantum energy $h\nu$ equal to or greater than the energy gap eV_g between the top of the filled band and the bottom of the empty band, then the absorption of such quanta will result in the creation of electron-hole pairs and the appearance of photoconductivity in the sample. The number of such carriers builds up rapidly until their rate of disappearance by recombination equals their rate of production by photoabsorption. In the primary photoelectric process described above equilibrium is established in a small fraction of a microsecond when the light is turned on or off.

Some substances exhibit the phenomenon of phosphorescence, in which the sample glows in the dark for some time after having been exposed to excitation by either radiation or particle bombardment. Such substances are believed to contain impurities of two kinds: (1) donor atoms which have normally occupied localized energy levels too far below the bottom of the empty band to act as donors at ordinary temperatures, and (2) acceptor atoms which have normally empty localized energy levels a short distance below the bottom of the empty band. We shall call the former deep donors and the latter high acceptors. Figure 7-2 shows the arrangement of these levels in a phosphor. When the sample is irradiated some of the deep donor electrons are raised to the empty band by the absorption of energy equal to or greater than eV_d. From the empty band they fall into the high acceptors. When the excitation is removed the high acceptors give up the electrons to the empty band, slowly, by thermal activation. From the empty band they then fall back down to the deep donor levels with the emission of phosphorescence radiation of wave length

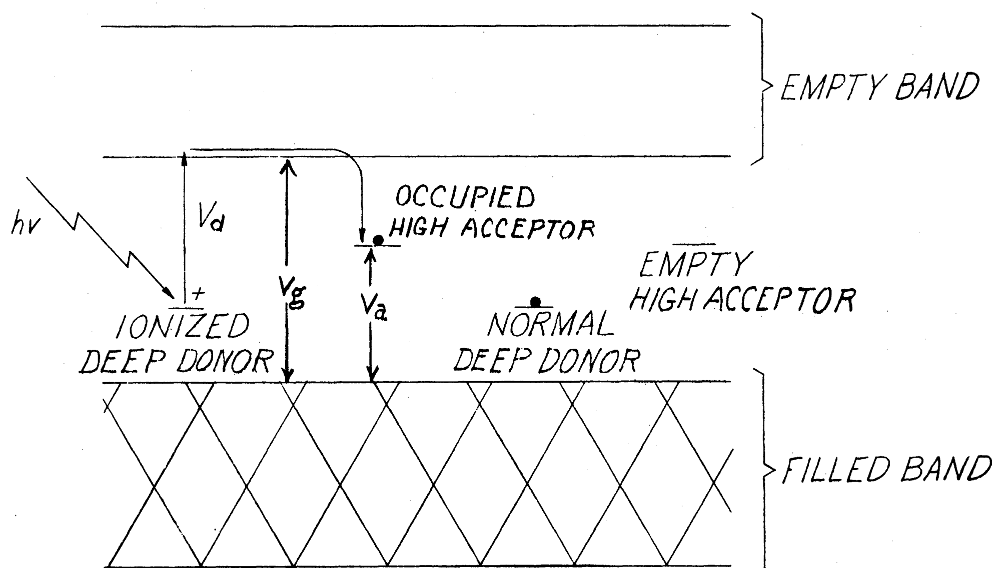


Figure 7-2 Energy Diagram for a Phosphor

given by $\lambda = \frac{hc}{eV_d}$. Phosphorescence is thus seen to be fluorescence which has been delayed by the trapping process involving the high acceptors.

The energy level scheme of Figure 7-2 suggests a number of additional possible processes:

1. Activation of an electron from the filled band to a high acceptor level by the absorption of a light quantum $\lambda = \frac{hc}{eV_a}$. This process leaves a positive hole in the filled band.

2. Such a positive hole may recombine with an electron in the empty band with the emission of a quantum $\lambda = \frac{hc}{eV_g}$. The emitted quantum has higher energy than the quantum which activated the process. At first sight, this might appear to present a violation of the energy conservation law. However, the high energy of the emitted quantum is the result of the addition of the energy of the exciting quantum and the energy furnished by thermal vibrations of the crystal to build up the system to the initial state from which the quantum $\lambda = \frac{hc}{eV_g}$ is radiated. Such a quantum would be called "anti-Stokes" radiation. In 1848 G. Stokes stated his theorem that in fluorescence processes the emitted radiation may have an energy only equal to or less than that of the exciting radiation. Another example of anti-Stokes radiation is found in the Raman spectra of molecules. In fact, process (2) may be considered as a kind of Raman effect in solids.

3. Recombination of the electron on a high acceptor level with a positive hole in the filled band with emission of a quantum $\lambda = \frac{hc}{eV_a}$. This process is the reverse of (1).

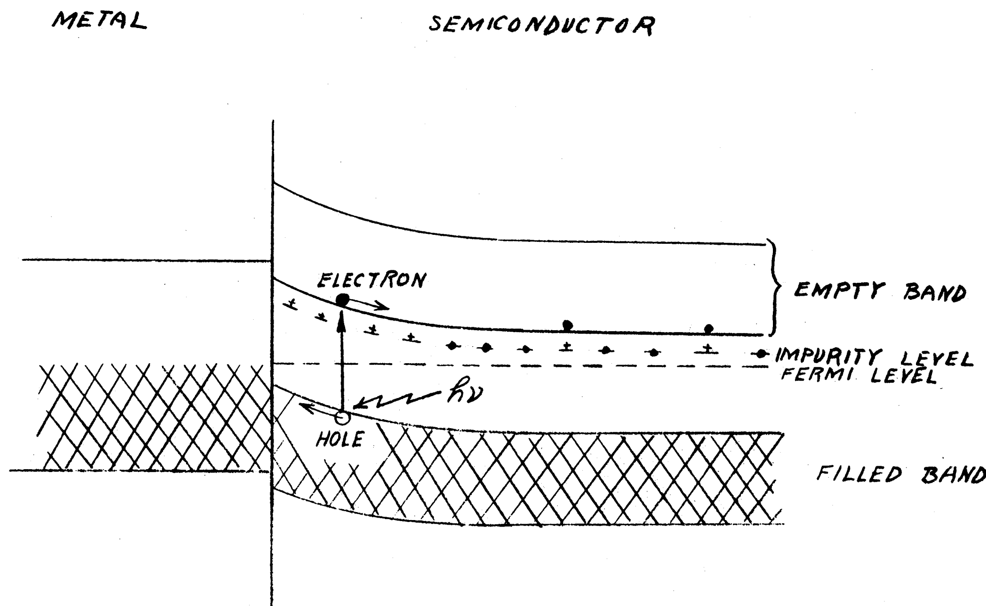


Figure 7-3 Energy Band Diagram Illustrating the Photovoltaic Effect Mechanism at a Metal-Semiconductor Boundary

4. Combination of an electron in a deep donor level with a positive hole in the filled band with the emission of a quantum $\lambda = \frac{hc}{e(V_g - V_d)}$.

5. Activation of an electron in the filled band to a deep donor level previously emptied by light. This process is the reverse of (4).

Some of these processes appear less probable to occur than others. For example, (5) might seem least probable and (1) most probable. However, these first-sight probabilities are subject to the operation of selection rules which may completely forbid some of the processes. Experiments on phosphor absorption and emission spectra have not yielded much information about these secondary effects.

The photoemf effect can be explained by reference to the band picture of a rectifying barrier (Figure 7-3). If an electron-hole pair is produced by photoabsorption inside such a barrier, the natural barrier field separates the charges before recombination can occur. The electrons flow into the interior of the semiconductor, while the holes flow into the metal. This process charges the semiconductor negative with respect to the metal, raises the Fermi level in the semiconductor above that in the metal, and produces an externally observable emf. As far as the rectifier barrier is concerned, it now acts as though an applied potential difference were being impressed across it, in the forward direction. Forward current, of magnitude given by the rectifier equation (6-9 or 6-10) flows across the barrier in the direction opposite to that of the photocurrent. A new steady state is established for any incident light flux when the photoemf produces a rectifier current exactly equal and opposite to the photocurrent.

The width V_g of the forbidden gap can be determined photoelectrically by observing the longest wave length (smallest quantum energy) which can induce photoconductivity. Such determinations generally agree satisfactorily with those obtained by measuring the slope of the log ρ versus 1/T plot in the intrinsic temperature range. For some substance, however, (as Na Cl, BaO)[1] the photoelectric V_g is smaller than that measured thermally. In a few such cases there is the additional complication that the crystal exhibits absorption of light at larger wave lengths than those required to induce photoconductivity.

These nonconformities can be explained by supposing the production of excitons.[2] An exciton is an entity composed of an electron and the hole it has just left. The two particles are not completely separated, but are held trapped in each other's electrostatic field. The exciton may be thought of as something like a hydrogen atom embedded in the crystal. The hole behaves like the hydrogen nucleus, supporting a hydrogen-like set of discrete localized energy levels, one or more of which may lie somewhere in the forbidden gap between the bands of the crystal. Figure 7-4 shows a diagram of an exciton. It is evident that an exciton can be created by light of longer wave length than that necessary to lift the electron clear up to the bottom of the conduction band and produce photoconductivity.

If an exciton level is only a short distance below the bottom of the conduction band, then an electron already raised to the exciton level by photon absorption may be thermally elevated the rest of the way up to the conduction band. In this way photoconductivity can be brought about by light of smaller quantum energy than that corresponding to V_g.

If the exciton levels are too far below the conduction band to permit thermal promotion of electrons from the exciton levels, then the substance will absorb light beginning at a threshold corresponding to a quantum energy V_{exc}, but

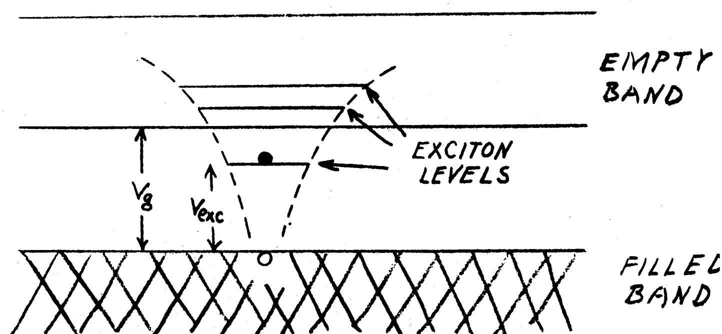


Figure 7-4 Energy Band Diagram for a Semiconductor Showing Energy Levels for an Exciton

loyd Smith; Report No. Scientific Administration 21/49,10 June, 1949.
ott and Gurney; "Electronic Processes in Ionic Crystals," (Oxford) Ch. III.

it will not become photoconducting until the quantum energy reaches or exceeds V_g. Excitons cannot of themselves transport current, as they are electrically neutral and hence unresponsive to an electric field. The exciton mechanism thus explains how such a crystal can absorb light without becoming photoconducting.

For most semiconductor substances, however, it appears that the lowest exciton level is above the bottom of the conduction band of the crystal. In these cases true photoconductivity will be induced at lower quantum energies than will exciton production. Even when an exciton is produced in such cases, the electron jumps from the exciton level into the conduction band of the crystal, thus dissolving the exciton and becoming a conduction electron. The possible complications due to excitons will be disregarded in the development of the rest of this chapter.

COLORS OF SEMICONDUCTORS

For a pure oxide with an energy gap of V_g volts, one would expect that wave lengths longer than $1.23/V_g$ microns would not have sufficient energy to raise an electron from the filled to the empty band. The oxide should thus be transparent to such wave lengths. On the other hand, wave lengths shorter than $1.23/V_g$ should be absorbed and should produce holes in the filled band and electrons in the empty band. These holes and electrons should make the oxide photoconductive and should render the material opaque to the wave lengths in question. If the energy gap corresponds to 3 volts or more, the oxide should be transparent and colorless to all visible light and should not be photoconducting in the visible spectral range. Apparently BaO, MgO, TiO_2, and SiO_2 are such oxides. An oxide for which $V_g = 2.15$ volts should transmit $\lambda > 0.575$ microns and absorb wave lengths shorter than this. It should thus transmit yellow and red light and should be orange colored; it should be photoconducting for green and blue light provided that the carriers generated by the light are appreciable in comparison with the carriers produced by thermal means. Nearly pure cadmium oxide and copper oxide may be such substances. From the fact that nearly pure Cu_2O is red, one can conclude that V_g is about 1.8 volts. The value of V_g deduced from the intrinsic conductivity is 2 x 0.7 or 1.4 volts. For these oxides, the valence electrons in the metal atom are in "s" shells or orbitals. In the solid, the conduction band is probably the widened s orbital level of the free atom.

For many other oxides, the mechanism of absorption must be more complicated. This is particularly true of such oxides like MnO_2, Mn_2O_3, MnO, Fe_2O_3, NiO, and CuO in which the metal ion has partially filled "d" shells. These oxides do not fit the simple band picture. In them the light can raise electrons to a higher energy state but the electrons and holes are either bound or remain free for only a very short time. Hence they do not contribute much to the conductivity. Even though they have high resistivities, they can absorb visible light, are highly colored, but have not been observed to be photoconductive. The exciton mechanism just described offers a possible explanation of this behavior.

Other oxides, which are colorless when pure, become colored when comparatively small amounts of impurities are added. Concomitantly, their conductivity

increases. While pure Al_2O_3 is colorless, it becomes highly colored when certain impurities are added. BaO is white if pure, but is red if it contains about one per cent excess Ba. Pure ZnO is white; as more excess Zn is added the color changes to green, then yellow, and then gray; the conductivity steadily increases from 10^{-4} to 2 $ohms^{-1}$ cm^{-1}; the carriers are electrons so that such ZnO is n type. Apparently as the concentration of excess Zn increases, the donor band widens out, longer wave lengths can be absorbed, and the energy required to produce conduction electrons decreases. We should thus expect that a family of log σ versus 1/T plots for increasing amounts of excess Zn should show successively smaller slopes. This is indeed the case. In Cu_2O, a deficiency of Cu can be produced by medium-temperature heat treatment in oxygen. These vacant Cu sites act like acceptors so that the oxide is p type. As the amount of Cu deficiency increases, the conductivity increases and the color of the oxide changes to a deeper red and eventually to black. This color change again may indicate that the acceptor band widens out with increasing concentration and as a consequence the work to raise an electron from the filled band to the acceptor band decreases. In NiO, as the deficiency in Ni increases, the color changes from pale green to dark green and then to gray, while the conductivity increases from 10^{-8} to 10^{-3} ohm^{-1} cm^{-1}. While the energy level band structure in NiO may be complicated, this color change may indicate a widening band structure with increasing concentration of Ni deficiency. Various examples of color changes due to excess or deficiency of metal are summarized in Table 7-1 .

In the above discussion we have used "metal deficiency" rather than excess oxygen because X-ray studies on such oxides have shown that the lattice is determined chiefly by the size of oxygen ion. The metal ions fit into the spaces between the closely packed oxygen ions. The matrix can maintain itself even though some of the metal sites are not filled; but it does not seem probable that additional large oxygen ions could be fitted into the already closely packed oxygen matrix.

Some oxides show pronounced color changes with temperature. Thus Mn_3O_4 is reddish brown at 20°C and turns progressively darker until at 200°C it is black. This does not seem to be due to any change in crystal structure or change of phase, since it occurs reversibly and without any appreciable time delay. This temperature color effect may indicate that the donor and acceptor bands can widen out with temperature or else shift with respect to the empty or full band. This would result in a temperature dependence of the activation energy. If this dependence is very pronounced and nonlinear, the log σ versus 1/T plot should show marked curvature and should have unusual values for the intercept.

With increasing conductivity, whether due to more donors and acceptors or to higher temperatures, any normally transparent semiconductor should eventually become opaque to all wave lengths because of the light absorbing effect of free electrons.

Table 7-1 * correlates the colors of a number of oxides with the number and kind of electrons in the metal ion. From this we can make a number of

---

*We are indebted to F. J. Morin for this table.

interesting and suggestive observations:

1. When the number of electrons in the d shell is 0, or 10, the oxides are colorless. This suggests that the energy from any filled band such as the 2p oxygen band to the nearest vacant s level is greater than 3 volts.
2. When the d shell contains 1, 3, 5 or 7 electrons the oxides transmit only a part of the visible spectrum and are brightly colored; the transmitted colors correspond roughly to 2.8, 2.3, and 1.8 volts for 1, 3, 5 d electrons. This probably means that the d bands are neither very narrow nor very wide and that the empty level into which they can be raised has about the same width; this width is about 0.5 volts.
3. When the d shell contains an even number of electrons, the oxide absorbs all visible light and is black.
4. These "d" energy bands must still be localized near each metal ion; the electrons in these d bands cannot pass from atom to atom without being raised to a conduction level.

THE BAND STRUCTURE OF Cu_2O AS DEDUCED FROM PHOTOCONDUCTIVITY

From the spectral distribution of the photoconductive effect it is possible to obtain considerable information about the relative location of the allowed and unallowed energies and thus to deduce the band structure. Such information is more detailed than that deduced from conductivity versus T data since photoconductivity permits one to observe the effects of a particular quantum of energy while conductivity data results from a whole range of energies. For cuprous oxide (Cu_2O) Schoenwald[3] has reported how the photocurrent near a wave length λ depends upon λ. Since at any λ, the current is proportional to the watts of incident radiation and is also proportional to the electric field strength, it is logical to plot i_p versus λ where i_p is the photocurrent per watt of incident energy for a field strength of 1 volt/cm. Such a plot is shown in Figure 7-5 . The wave lengths have been converted to equivalent volts V from the equation: $\lambda V = 1.23$ micron volts.

Three curves, A, B, and C are shown in the figure for three samples of Cu_2O that illustrate three types. For each sample the value of the specific resistance at room temperature is given. In any given type the values of i_p might vary over a range of 100 so that no significance should be attached to the fact that the scales in A, B, and C differ by factors of 10. They might be significant if the treatment for each individual sample could be specified. Of a total of 52 samples examined by Schoenwald, 6 were like A, 11 were like B, and 35 were like C.

Before discussing the significance of Figure 7-5 , it is desirable to describe the method of preparation for the samples, and the techniques used in measuring the photocurrent, and to state some conclusions reached by Schoenwald and by Engelhard.[4]

---

[3] B. Schoenwald; Annalen der Physik (5) 15, p. 395, 1932.

[4] Ref. 4, page 10

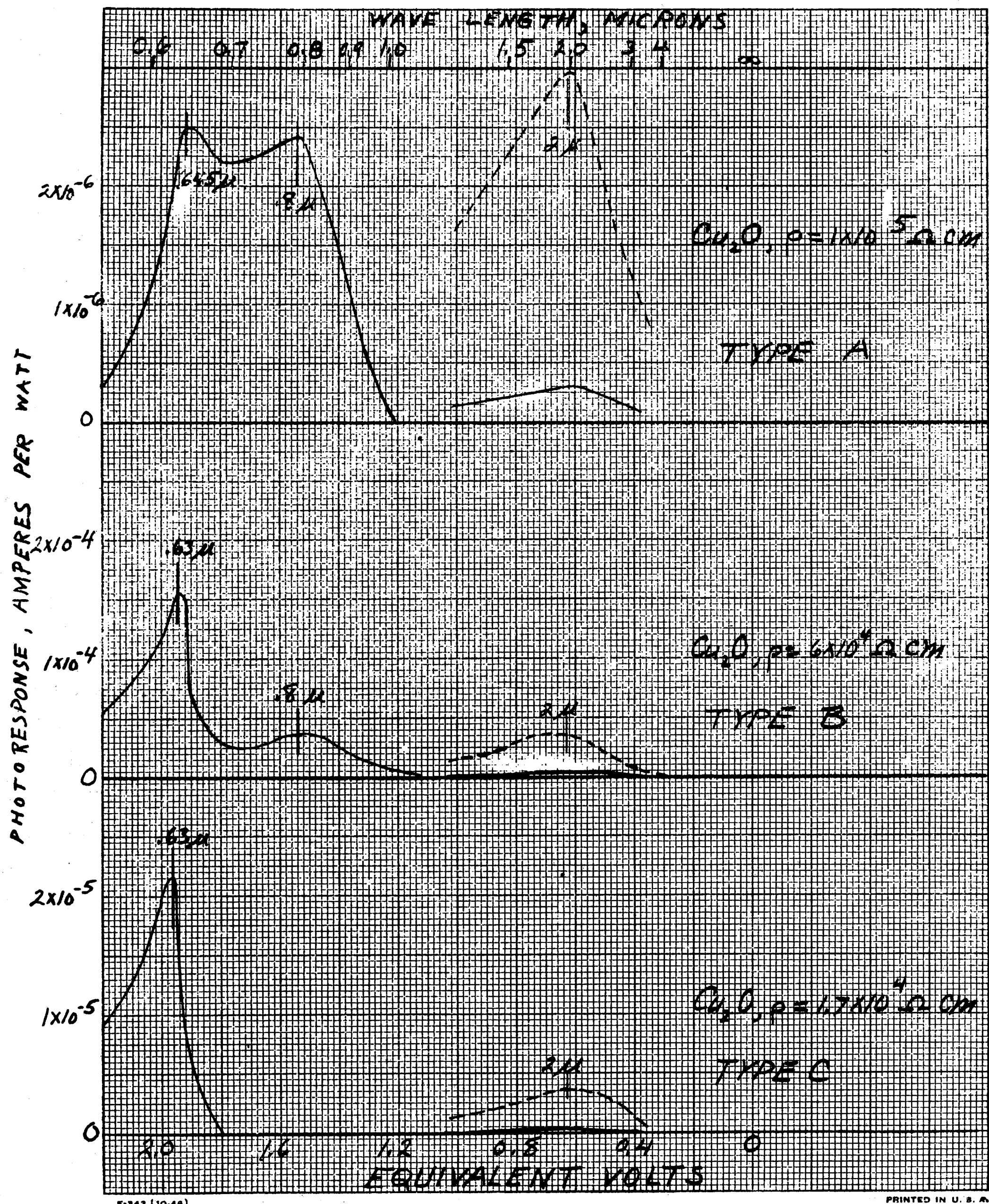


Figure 7-5 Photoresponse versus Wave Length (Upper Scale) and Equivalent Electron Volts (Lower Scale) for the Photoconductivity in Three Specimens of Cuprous Oxide (After Schoenwald)

The samples of Cu_2O were made by oxidizing thin sheets of copper at 950 to 1000°C in air until the copper was completely oxidized. They were then divided into four groups heat treated in four ways:

1. In vacuum at 900 to 1000°C.
2. In vacuum at 500 to 600°C.

3. In oxygen at 900 to 1000°C.

4. In oxygen at 500 to 600°C.

The heat treatment took place in alundum tubes and lasted 10 to 20 hours. The tubes were rapidly cooled in an air stream or quenched in water. The vacuum treated units were coated with a thin film of reduced copper. The oxygen treated units were coated with a thin film of black CuO. These films were chemically removed. The surfaces of the Cu_2O samples were rough and scattered light to a different degree from one sample to the next. This may account in part for the variability in photocurrent discussed above. The samples were translucent to red light. The approximate dimensions of the samples were: length 1.5 cm; width 1.0 cm; thickness 0.1 cm. Silver contacts were evaporated over the ends and the currents were carried lengthwise. The samples were illuminated thicknesswise.

Even for intense illumination through a spectrometer, the photocurrent is very small compared to the "dark" current. Hence the photocurrents were measured by chopping the light with a rotating disc at a frequency of 1500 cps; feeding the small a-c current, which is superposed on a large d-c current, into an amplifier which will pass 1500 cps but will not pass direct current; rectifying the amplified alternating current and measuring it with a d-c meter.

Engelhard[4] showed that the specific resistance of Cu_2O at room temperature varied from 10^6 to $10^{2.5}$ ohms cm. The mobilities of the carriers were computed from the Hall effect and varied from 100 to 1.0 cm/sec per volt/cm. Engelhard states that the mobilities of the photocurrent carriers were the same as those for the dark current carriers. From the way in which the photocurrent decreased as the frequency of chopping the light increased, Schoenwald deduced that the mean life of a hole was 2×10^{-5} sec. (No data are given how this varied from sample to sample.) Hence, in a sample in which the mobility is 100 cm^2/volt sec. and the field is 50 volts/cm an average carrier should drift a distance of $100 \times 50 \times 2 \times 10^{-5} = 10^{-1}$ cm or 1 mm.

Engelhard states that when a sample is heated in vacuum at 1000°C the resistance is greatly increased (100 to 1000X) while the Hall constant R_H changes only slightly. This means that the product of $R_H\,\sigma$, which is practically equal to the mobility, decreases about as much as does σ; or that the concentration of carriers is changed only slightly while their mobility is greatly decreased. He ascribes this to metallic Cu particles which can be seen in a microscope. When such samples are subsequently heated in oxygen, $R_H\,\sigma$ takes on larger values.

We will now return to consider what can be deduced from Figure 7-5 . The simplest type is that shown in C and the largest effect is produced for light whose voltage equivalent is greater than 1.8. This is interpreted to be the energy necessary to raise an electron from the top of the filled band to the bottom of the allowed empty band. From the fact that the current does not increase abruptly at

---

[4]E. Engelhard; Annalen der Physik (5) 17, p. 501, 1933.

1.8 volts, it can be deduced that the edges of the unallowed gap are not sharp or that the concentration of allowed states does not change abruptly from 0 to some large value at the edges of the gap. For values of $V > 2.0$ volts, the current per watt decreases. One reason for this is that the number of quanta per watt and hence the photocurrent, should decrease as $1/V$.

In C there is another photoconductivity region extending from about 0.3 volt to 1.1 volts with a peak near 0.6 volt. Since the intensity of this conductivity is comparatively small, the response in this region has been enlarged ten fold and redrawn in the dashed curve. From the existence and location of this response, one can deduce that there must exist an acceptor level 0.3 volt above the top of the filled band to which electrons can be raised. This acceptor level band might have a finite width and might extend say from 0.3 to 0.6 volt. From the fact that wave lengths of 0.3 to 1.1 volts can produce a photocurrent, it follows that the sum of the width of the acceptor level plus the width of the region available near the top of the filled band must equal 0.8 volt or more. We say "more" because the probability of removing an electron from a small interval of energy in the filled band may decrease rapidly with the depth below the top of the filled band. From the fact that a watt near 0.6 volt produces only about one fiftieth as much current as does a watt near 2.0 volts, one can conclude that the density of states in the acceptor level is about 200 times less than the density of states in the empty band. From the fact that the Hall constant for the photocurrent has the same sign and value as that for the dark current, one concludes that the carriers are predominantly holes and not electrons. Hence, it follows that the electrons which are raised to the acceptor level or band, do not partake in the current; they are immobile and must be bound to a particular region in the lattice. We shall see below that the acceptor levels are probably lattice sites from which a metal atom is missing and that the electron is accepted by the oxygen atoms surrounding this vacant metal lattice site.

In A and B there is an additional voltage (or wave length) region which can produce photocurrents. This begins near 1.2 volts, comes to maximum near 1.5 volts, and then decreases. In type A, this current is as large as that for the band that starts at 1.8 volts; in type B, the current is only about one fourth as large. Apparently, in some samples of Cu_2O, there are additional ways by which carriers can be produced by radiation of this equivalent voltage. The presence of this additional mechanism apparently varies considerably from one sample to the next presumably due to differences in heat treatment. We are inclined to say that curves A and B are really the same type but differ only in the degree to which this additional mechanism is present. In type C, it is completely absent. We would like to propose that this additional conductivity effect is to be ascribed to the presence of interstitial Cu. The number per cm^3 of such interstitial Cu atoms or ions would be expected to vary greatly with heat treatment: it should be large for vacuum units quenched from a high temperature; it should be small or absent for units treated in oxygen and cooled slowly. One might expect that an interstitial atom would readily give up its electron to the conduction band or to the acceptor levels produced by vacant metal lattice sites, if these are present. One would therefore expect that such interstitial Cu would give rise to a donor level or band near the upper part of the energy

gap and that this donor band normally would be empty if an equal or greater number of vacant lattice sites were present.

There are then two ways by which the photoresponse near 1.5 volts could be explained: (1) the light could raise electrons from the filled band to the donor band, in which case the bottom of the donor band should be 1.2 volts above the top of the filled band; or (2) the top of the acceptor band (which is partially filled by the electrons which have dropped from the donor to the acceptor band) is 1.2 volts below the bottom of the empty band. If the second explanation is the correct one, then the width of the acceptor band should be about 0.3 volt.

The choice between these two explanations could be decided by measuring the sign of the Hall coefficient for illumination with light of 1.5 volts or 0.8 micron. If this sign were positive it would support the first explanation; if negative, the second one. The existence of donor levels as well as acceptor levels has been deduced by Brattain and Bardeen from the shape of log σ versus $1/T$ plots. They conclude that N_a (number of acceptor levels/cm^3) exceeds N_d (number of donor levels/cm^3) in all samples examined. The ratio N_d/N_a varies considerably; from perhaps .05 to 0.80.

It is natural to ask how the acceptor levels and the donor levels are produced. The acceptor levels are thought to be due to vacant metal lattice sites. In an occupied metal site the metal gives up its valence electron to the surrounding oxygen atoms. When the site is vacant the surrounding oxygen atoms are anxious to accept electrons from any available source. Vacant metal sites can be produced (1) by heat treatment in oxygen which results in Cu_2O with a Cu deficiency; or (2) by heat treatment at high temperatures in which some of the Cu atoms (or ions) are transferred from their normal sites to interstitial positions. Such treatment must be followed by rapid cooling to freeze the Cu atoms in interstitial positions.

Donor levels are thought to be due to interstitial Cu atoms or ions. Such a copper atom should give up its electron quite readily and become a Cu^+ ion or an empty donor level. Such interstitial Cu might be produced either by the second mechanism discussed above or by heat treatment in a vacuum as discussed by Engelhard.

The presence of interstitial Cu^+ ions and vacant lattice sites should show itself in still another way. Both of these should act as scattering centers for the quantum mechanical waves that represent the passage of electrons or holes when current is carried in a semiconductor. Hence, they should decrease the mobility and thus decrease the conductivity.

As a result of all these considerations we propose the following energy level diagram for Cu_2O, Figure 7-6 . This differs from similar diagrams proposed by others only in ascribing an appreciable width to the acceptor and the donor bands. The magnitude of this width will vary from sample to sample depending on the concentration of whatever is responsible for the acceptor and the donor levels. There are no data on which to base an estimate of the width of the donor band.

The acceptor band is due to vacant metal lattice sites. Let N_a be the number of such vacant sites per cm^3.

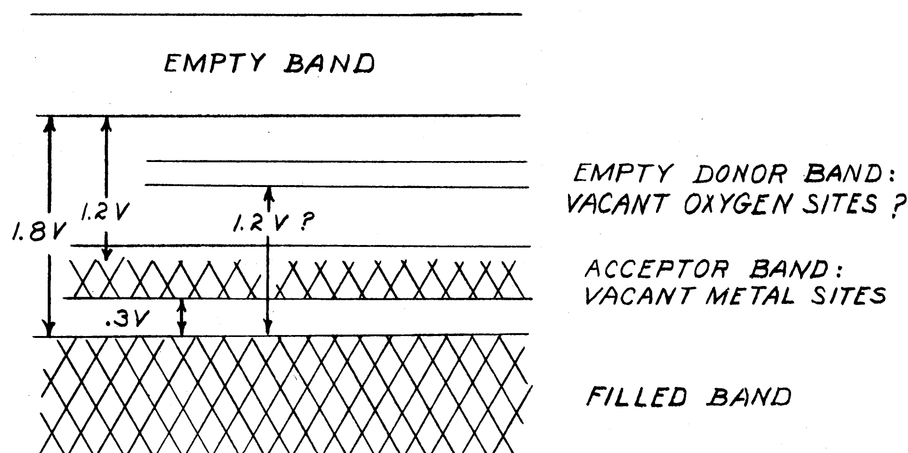


Figure 7-6 Proposed Energy Band Scheme for Cuprous Oxide Suggested By Results of Photoelectric and Optical Measurements

The donor band may be due to interstitial Cu or to vacant oxygen sites. Let N_d be the number of such sites.

If $N_d > N_a$, then all acceptor levels should be filled and the donor levels should be partly filled. If $N_d = N_a$, then all donor levels should be empty and all acceptor levels filled. If $N_d < N_a$, the donor levels should all be empty and the acceptor levels should be only partly filled. Most Cu_2O samples are like the last case.

The acceptor levels are localized in space near the vacant metal lattice sites. When such a level is filled it constitutes an immobile negative ion which does not add to the conductivity. Similarly, the donor levels are localized in space. When such a level donates an electron it becomes an immobile positive ion. The electron will be donated to a lower lying empty acceptor level if any such exist even at low temperature; after all the acceptor levels are filled by electrons from the donor levels, the remaining filled donor levels may donate an electron to the conduction band if the temperature is high enough.

EVIDENCE FOR THE TEMPERATURE DEPENDENCE OF THE ENERGY GAP, V_g

Pfund[5] measured the change in resistance of a Cu_2O sample when it was irradiated with various wave lengths of monochromatic light. He investigated only the region from 0.2 to 0.8 micron (6.1 to 1.54 volts). However, he did measure the effect at two different temperatures, namely, -125°C and 19°C. His results are shown in Figure 7-7 .

Note that at the higher temperature the curve near the long wave length limit is shifted towards longer wave lengths. Note also that there is no sharp

[5] A. H. Pfund; Physical Review, Vol. 7, p. 295. 1916.

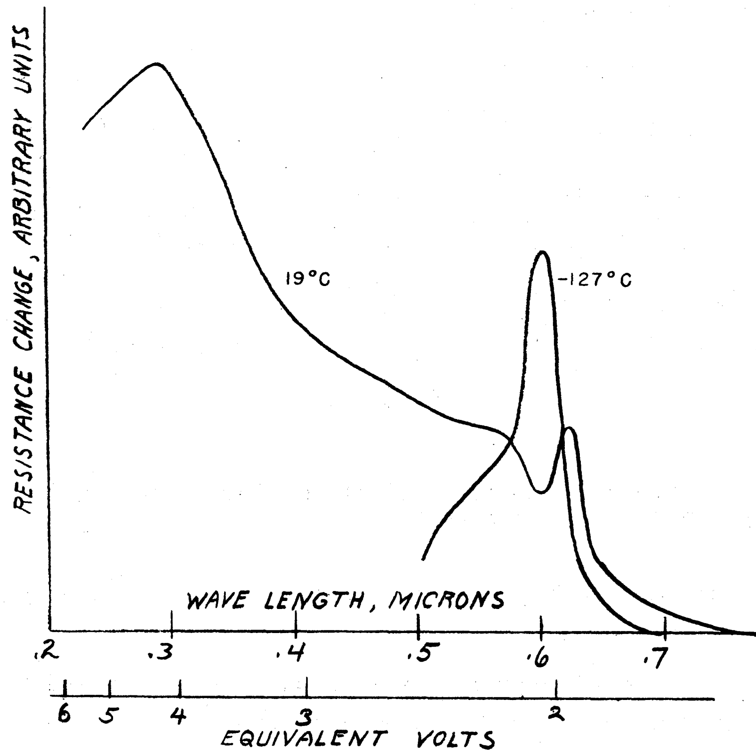


Figure 7-7 Photoconductivity versus Wave Length and Equivalent Electron Volts for Cuprous Oxide at Two Different Temperatures (After Pfund)

cutoff but a gradual tailing off. This "tail" is more pronounced at the higher T. If the curves are extrapolated to the X axis and the wave lengths converted to values of V_g it is found that V_g decreases from 1.8 to 1.6 volts as T increases from -127°C to 19°C. Hence, the temperature coefficient of V_g, or c, is -1.4×10^{-3} volts per degree centigrade for Cu_2O.

Additional evidence can be deduced from data on silicon. Pearson and Bardeen[6] have shown that, in the intrinsic range, the mobility v is given by:

$$v_e = 3\ v_h = 15 \times 10^5\ T^{-\frac{3}{2}} \frac{cm^2}{sec}\ \text{volts}$$

so that

$$(v_e + v_h) = 20 \times 10^5\ T^{-\frac{3}{2}}$$

They also give log σ versus 1/T curves from which one can calculate that $\log \sigma_\infty = 4.3$. From equation 1-3 it follows that

[6] G. L. Pearson and J. Bardeen; Physical Review, Vol. 75, p. 865, 1949.

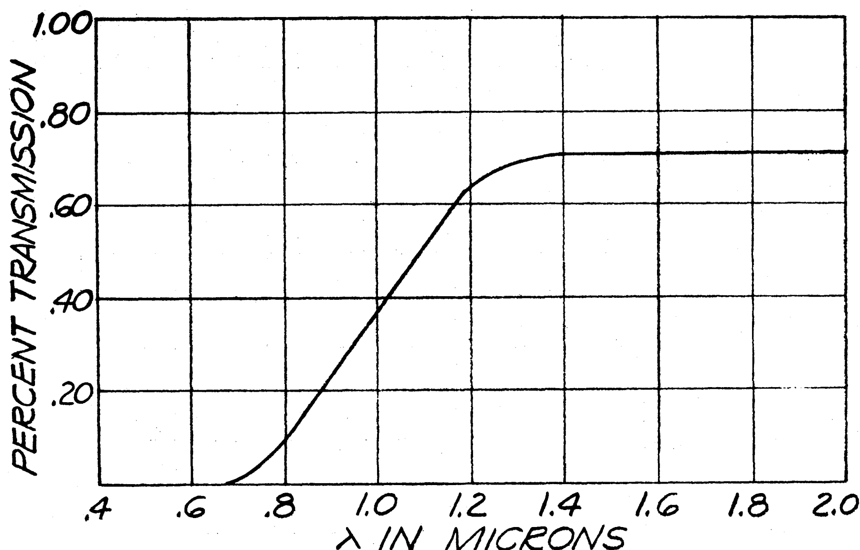


Figure 7-8 Per Cent Transmission versus Wave Length for Silicon (After Pearson)

$$\log \sigma_{\infty} = 4.3 = \log \left(2 \cdot UT^{\frac{3}{2}} \, 20 \times 10^{15} \, T^{-\frac{3}{2}}\right) - \frac{c}{4.6} \frac{e}{k} \quad .$$

Hence,

$$c = -5 \times 10^{-4} \, \frac{\text{volts}}{°C} \quad .$$

From the slope of their curve we deduce the V_{go} = 1.12 volts. Hence V_g at 300°K (room T) = 1.12 -5 x 10^{-4} x 300 or 0.97 volts. Hence, the long wave length limit for absorption and for photoconductivity should be at 1.27 microns rather than 1.1 micron if c were 0. For Si the optical transmission versus λ has been measured by Pearson (Figure 7-8)[7]. Teal[8] and co-workers have measured the ratio of the change in resistance to the dark resistance versus λ. The results are shown in Figure 7-9 .

It is to be noted that beyond about 1.3 microns, the transmission is constant and equal to 67 per cent. Apparently 33 per cent of the incident light is reflected or scattered at the surfaces. Starting at 1.3 microns, absorption sets in and increases rapidly at 1.2 microns. This agrees quite well with the predicted value of 1.27 for the absorption limit.

The extrapolation of the steeper portion of the photoresistance curve hits the axis near 1.20 microns, while the tail goes out to 1.35 microns. This agrees

---

[7] Unpublished.

[8] G. K. Teal, J. R. Fisher, and A. W. Treptow: Physical Review 69, p.686, 1946.

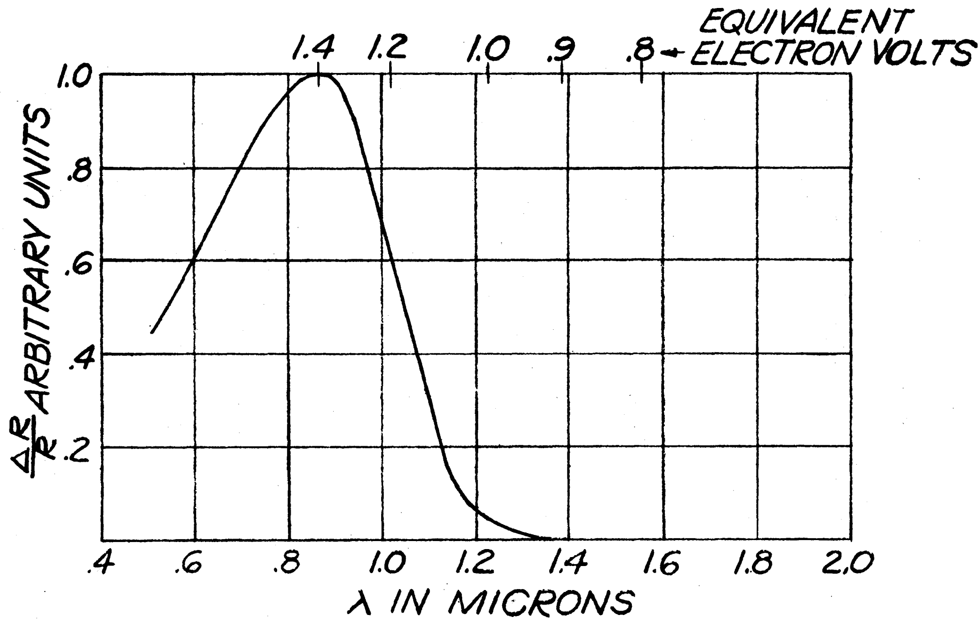


Figure 7-9 Photoconductivity versus Wave Length for Silicon (After Teal, etal)

well with 1.27. The long wave limit is definitely greater than 1.1 microns. As before, there is a pronounced tail towards longer wave lengths.

This evidence from the slope and intercept of conductivity curves, from the mobility data, and from absorption and photoconductivity data is self consistent and leads to the conclusion that the unallowed energy gap, V_g, is temperature dependent. For Cu_2O and Si, V_g decreases about 1.4 and 0.5 millivolts per degress C, respectively.

Table 7-1 Colors in Semiconductors Due to Deviations from Stoichiometric Values

Oxide or Semiconductor	Stoichiometric Color	Color with Excess Metal, N-type	Ratio of Excess	Conductivity
KCl	Colorless	Violet	.0003	ohm^{-1} cm^{-1}
NaCl	Colorless	Yellow		
BaO	White	Red	.01	
ZnO	White	Green Yellow Gray	.002	10^{-4} to 2
TiO_2	Pale Yellow	Dark Blue to Black	0	10^{10}
			.001	10^{1}
			.01	1
			0.12	10^{-2}
		Color with Excess Oxygen, P-type		
Cu_2O	Red	Black		
NiO	Pale Green	Dark Green to Gray	.05	10^{-8} to 10^{-3}
UO_2				
CoO				
Mn_3O_4	Reddish Brown	Black		

Mn_3O_4 is reddish brown at room temperature - turns black at about 200°C.

Metal	Atomic Distance	Compounds	Cationic Distance	Color	3d Electrons	Compounds	Cationic Distance	Color	3d Electrons
Sc		Sc_2O_3		Colorless	0				
Ti	2.95	Ti_2O_3		Violet	1				
V	2.63	V_2O_3		Black	2				
Cr	2.50	Cr_2O_3		Green	3				
Mn	2.58	Mn_2O_3		Black	4	MnO	3.14	Gray Green	5
Fe	2.48	Fe_2O_3	2.89	Red	5	FeO	3.06	Black	6
Co	2.50					CeO	3.00	Green Brown	7
Ni	2.49					NiO	2.95	Green	8
Cu	2.55	Cu_2O		Red		CuO		Black	9
Zn	2.66					ZnO		Colorless	10

Metal	Atomic Distance	Compounds	Cationic Distance	Color	3d Electrons	Compounds	Cationic Distance	Color	3d Electrons
Y	3.60	Y_2O_3		Colorless	0				
Zr	3.13	Zr_2O_3							
Cb	2.85	Cb_2O_3				Cb_2O_5		Colorless	0
Mo	2.72	Mo_2O_3		Yellow	3				
Ld	3.73	Ld_2O_3		Colorless					
Hf	3.14	Hf_2O_3							
Ta	2.85	Ta_2O_3				Ta_2O_5		Colorless	0
W	2.74	W							
Th	3.59	Th_2O_3							
U	2.75								

UNIVERSITY OF MICHIGAN
3 9015 02514 8118