

TOPIC 10: UNIFORM ELECTRONIC SEMICONDUCTORS IN EQUILIBRIUM**READING ASSIGNMENT**

"Uniform Electronic Semiconductors in Equilibrium", from *Solid State and Semiconductor Physics*, J.P. McKelvey

LECTURE PROGRAM

1. Semiconductors
 - a. Band model
 - b. Chemical bond model
2. Intrinsic Semiconductors and Impurity Semiconductors
 - a. Density of states for an intrinsic semiconductor
 - b. Distribution function, Fermi level, occupancy for an intrinsic semiconductor
 - c. N-type impurity semiconductor
 - d. P-type impurity semiconductor
3. Statistics of Holes and Electrons
 - a. Boltzmann approximation
 - b. Calculation of n_0 , p_0
 - c. $n_0 p_0 = n_i^2(T)$
 - d. np product for an impurity semiconductor
 - e. Intrinsic Fermi level
4. Ionization Energy of Impurity Centers
 - a. Calculation of ionization energy using the Bohr model
 - b. Energy band diagram
5. Statistics of Impurity Semiconductors
6. Case of Incomplete Ionization of Impurity Levels
7. Conductivity

CHAPTER 9

UNIFORM ELECTRONIC SEMICONDUCTORS IN EQUILIBRIUM

9.1 SEMICONDUCTORS

In the preceding chapters our main concern has been to understand the electrical and thermal properties of metallic conductors and of insulators, and to understand why some substances are extremely good electrical conductors and others are extremely poor ones. These objectives have now been accomplished in a general way, and from this point on we shall turn our attention toward a rather detailed examination of the properties of semiconductors. There are three reasons why this is a useful thing to do. First, it is usually possible to use Maxwell-Boltzmann statistics in discussing the statistical mechanics of charge carriers in semiconductors. This means that it is possible to obtain exact analytical solutions for many problems in semiconductors which can be solved only by approximate or numerical methods for metals, where Fermi-Dirac statistics must be used. An understanding of such effects in semiconductors may furnish valuable physical insight into their counterparts in metallic substances. In short, it is possible to understand electronic processes more easily and more deeply in semiconductors than in any other class of crystals. Second, the technology of crystal growth has made it possible to produce crystals of certain semiconducting substances (notably germanium and silicon) of fantastic purity and crystalline perfection, far in excess of what can be obtained at present for metals and insulators. The existence of these nearly perfect crystals makes it possible for the experimentalist to observe electronic transport properties and thermoelectric and galvanomagnetic effects with ease and precision, secure in the knowledge that what he is trying to observe will not be obscured by effects caused by impurities or structural imperfections in the crystal lattice. We may thus easily observe and interpret phenomena in semiconductors which would be difficult or impossible either to measure accurately or to explain quantitatively in metallic substances. Finally, semiconductors have acquired over the last two decades an enormous technological importance. They have come to be used in making all sorts of electronic devices, including rectifiers, transistors, photocells, voltage regulators, parametric amplifiers, and switching devices. An appreciation of the fundamental electronic transport properties in these substances is of direct application in understanding and analyzing the operation of all the device structures of contemporary technical importance.

A semiconductor is a crystalline substance having an energy band structure in which a band of electronic states, completely filled at zero temperature, is separated from one which is completely empty at absolute zero by a narrow region of forbidden

energies. This band structure is shown schematically in Figure 9.1(a). At absolute zero the semiconductor is a perfect insulator, since there are no partially filled bands. At higher temperatures, however, a few electrons from the valence band may acquire enough random thermal energy to be excited across the forbidden band to become conduction electrons in the heretofore empty *conduction* band. The empty states left behind in the lower or *valence* band also contribute to the conductivity, behaving like

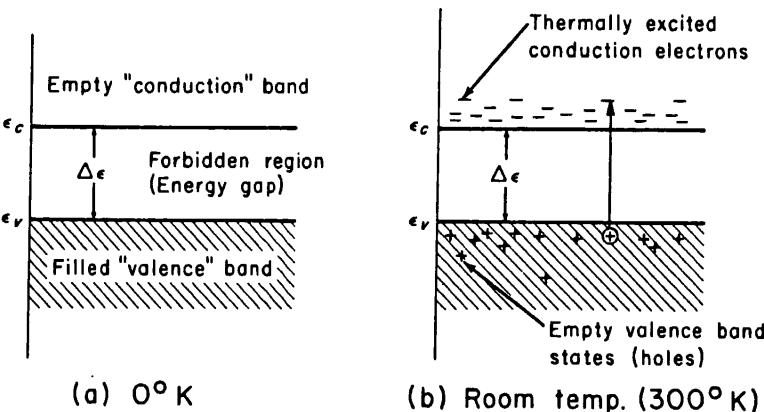


FIGURE 9.1. Conduction and valence bands of a pure semiconductor (a) at absolute zero, (b) at room temperature, showing thermally excited electrons and holes.

positively charged holes. It is clear that the number of conduction electrons and the number of holes must rise with increasing temperature, and hence the electrical conductivity likewise increases with temperature.

The physical mechanism of electron and hole conduction in *covalent* semiconductors such as carbon (diamond), germanium, and silicon, which form crystals having the diamond structure of Figure 1.9 can be understood from Figure 9.2. This figure shows the calculated energy band structure of diamond, plotted against the interatomic spacing, as in Figure 8.10. The corresponding diagrams for silicon and germanium are quite similar. When N isolated atoms are assembled into a crystal, the $2s$ and $2p$ atomic levels first broaden into energy bands; as the interatomic spacing decreases these bands broaden further, and eventually overlap. As the interatomic distance becomes still smaller, the continuum of what were originally $2s$ and $2p$ states splits once more into two bands, each of which, however, now contain precisely $4N$ states. At the equilibrium interatomic distance a , these bands are separated by an "energy gap" or forbidden region of width $\Delta\epsilon$. Since the electronic structure of carbon is $1s^2)2s^22p^2$, there are $4N$ valence electrons available, which exactly suffice to fill the lower of the two bands, and form the valence band of the crystal. The upper band becomes the conduction band. The forbidden energy region under normal conditions has a width of about 7 eV in diamond, 1.2 eV in silicon and 0.7 eV in germanium. Due to thermal expansion of the lattice, this "energy gap" has a weak dependence on temperature; it is clear from Figure 9.2 that the energy gap will decrease as the crystal expands. It is also evident from this diagram that $\Delta\epsilon$ will be a function of pressure, becoming larger as the interatomic spacing is reduced by the application of pressure by hydrostatic or other means. Beyond the point A in Figure 9.2, where the $2p$ and $2s$ bands overlap, the s and p character of the electronic states is lost; in the valence

band of the semiconductor, therefore, the electron wave functions are a mixture of s and p atomic wave functions.

The electrons in the valence band are the electrons which form the tetrahedrally disposed electron-pair covalent bonds between the atoms in Figure 1.9. The thermal excitation of an electron from the valence band to the conduction band corresponds physically to the removal of an electron from a covalent pair bond by thermal agitation.

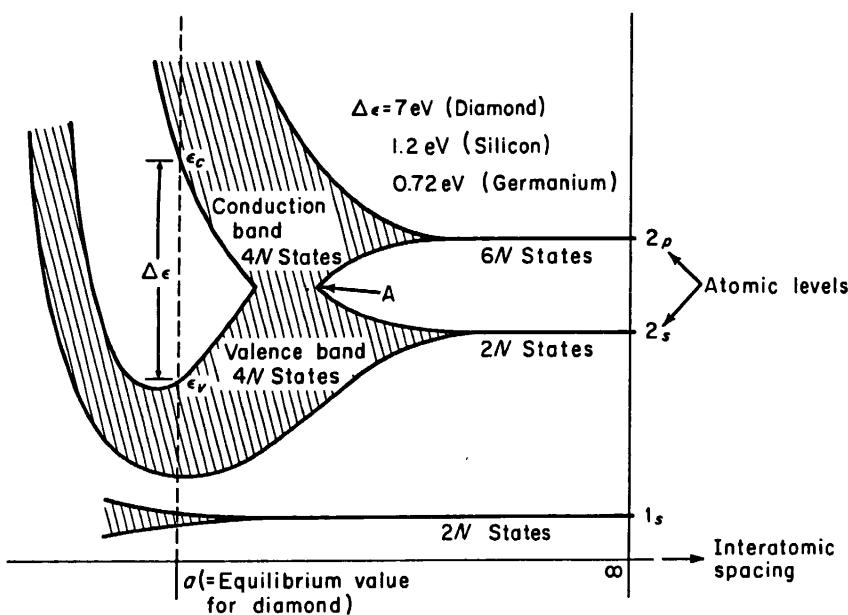


FIGURE 9.2. The bands arising from the $1s$, $2s$, and $2p$ atomic states of carbon (diamond) as a function of interatomic distance. [After W. Shockley, *Electrons and Holes in Semiconductors*, D. van Nostrand Co., Inc., New York (1950).]

of the lattice. The electron then becomes a free electron, outside the covalent lattice bond scheme, and is available to function as a charge carrier to conduct current through the crystal. This process is depicted schematically in Figure 9.3(a), though it must be remembered that the bond scheme shown there is a two-dimensional rendering of what in actuality is a three-dimensional network of tetrahedral bonds. The excitation of an electron leaves behind a localized defect in the covalent bond structure of the crystal, which can be identified as a valence band state which could be occupied by an electron but is actually empty. This defect is simply the "hole" of Section 8.5. Both the free electron and the hole are migratory; the free electron may wander about within the crystal in a random fashion, impelled by thermal energy it may acquire from the lattice. The hole, likewise, may migrate by virtue of the fact that an electron in a covalent bond pair adjacent to the hole may very easily move into the hole, completing the bond pair at the former hole site, but transferring the location of the hole to the site whence the electron came. This hole migration may also be caused by thermal agitation of the lattice. Free electrons and holes will also move in response to an electric field, and may then give rise to a macroscopic current flow through the crystal. This situation is illustrated in Figure 9.3(b). Here all the electrons in the crystal, in both valence band and conduction band, are subjected to a force $-eE$, acting to the right in the diagram. The free electrons move to the right, causing a

conventional current flow to the left in view of their negative charge. In addition to this, an electron in a covalent bond pair adjacent to a hole may move to the right into the empty electron site associated therewith, the hole then moving to the left, to the site whence this electron came; this process may now be repeated, the net result being a transfer of an electron to the right, accompanied by the motion of the hole to the

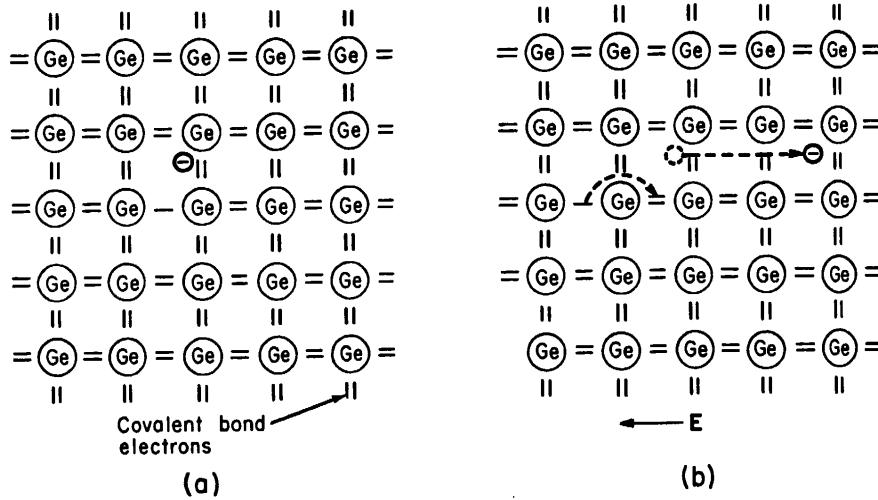


FIGURE 9.3. (a) A free electron and a hole arising from thermal ionization of an electron originally in a covalent bond. (b) Motion of free electron (to the right) and hole (to the left) when an electric field is applied as shown.

left, hence in the direction in which a *positively charged particle* would move under the influence of the applied field. The net electron current to the right again causes a conventional current flow to the left, which might as well be represented as a current of *positive holes* in that direction. Current flow may thus arise both from the movement of free electrons and from the migration of valence electrons into and out of empty states in the valence band, which we prefer to think of as the migration of positive holes.

9.2 INTRINSIC SEMICONDUCTORS AND IMPURITY SEMICONDUCTORS

A semiconductor in which holes and electrons are created solely by thermal excitation across the energy gap is called an *intrinsic* semiconductor. Holes and electrons which are created in this manner are often referred to as *intrinsic charge carriers* and the conductivity arising from such carriers is sometimes termed *intrinsic conductivity*. In an intrinsic semiconductor the concentrations of electrons and holes must always be *the same*, since the thermal excitation of an electron inevitably creates one and only one hole.

The population of holes and electrons in an intrinsic semiconductor is described statistically in terms of the Fermi-Dirac distribution function and the density of states functions for the valence and conduction bands. Since the bottom of the conduction

band and the top of the valence band exhibit an essentially parabolic dependence of ϵ on k , the behavior of electrons and holes in these regions is essentially that of free particles, with appropriate effective mass factors. Electrons and holes are only very rarely excited into regions of the conduction and valence bands where their properties may depart from free-particle behavior at physically realizable temperatures, and so the effects of such excitations may for practical purposes be neglected. The density of states functions to be used, then, are essentially those for free particles, the density of states in the conduction band being given by

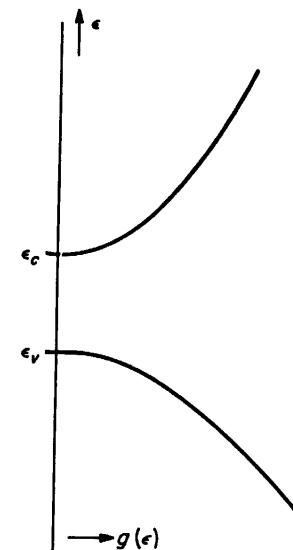
$$g_c(\epsilon) d\epsilon = \frac{8\sqrt{2\pi}}{h^3} m_n^{*3/2} \sqrt{\epsilon - \epsilon_c} d\epsilon \quad (\epsilon > \epsilon_c) \quad (9.2-1)$$

and the density of states in the valence band by

$$g_v(\epsilon) d\epsilon = \frac{8\sqrt{2\pi}}{h^3} m_p^{*3/2} \sqrt{\epsilon_v - \epsilon} d\epsilon, \quad (\epsilon < \epsilon_v) \quad (9.2-2)$$

where m_n^* is the effective mass for electrons in the conduction band, and m_p^* the effective mass for holes in the valence band.¹ The density of states in the forbidden region $\epsilon_v < \epsilon < \epsilon_c$ is, of course, zero. A plot of the density of states curve for an intrinsic semiconductor is shown in Figure 9.4. If m_p^* and m_n^* are precisely equal, the

FIGURE 9.4. Density of states function for an intrinsic semiconductor.



Fermi energy must lie exactly in the center of the forbidden region. This is true because otherwise the population of electrons in the conduction band and holes in the valence band, obtained by integrating the product of the density of states function and the probability factor $f_0(\epsilon)$ for electrons in the conduction band or $1 - f_0(\epsilon)$ for holes in the valence band would not be equal. The situation is represented diagrammatically in Figure 9.5(a). If m_p^* and m_n^* are not equal (and this is usually the case), the

¹ The subscripts n and p refer to *negative* and *positive* charge carriers. This convention will be used consistently.

Fermi energy must be adjusted upwards or downwards a bit, away from the precise center of the energy gap, to equalize the population integrals, and hence must lie *near* but not *at* the center of the forbidden region. This state of affairs is shown in Figure 9.5(b). We shall investigate the statistics of holes and electrons in some detail in a later section.

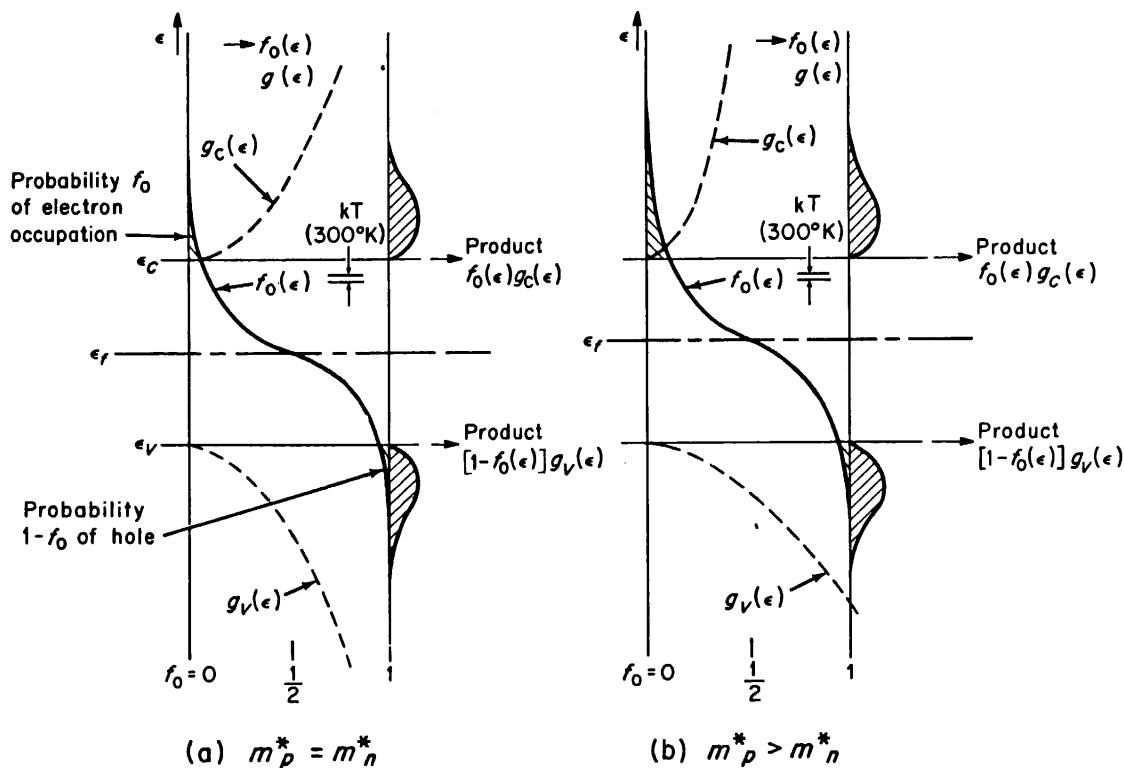


FIGURE 9.5. Distribution function, Fermi level, density of states function and electron and hole populations for an intrinsic semiconductor (a) where $m_p^* = m_n^*$ and (b) where $m_p^* > m_n^*$. The "spread" of the Fermi distribution is exaggerated here for the sake of illustrative clarity; in reality at 300°K the Fermi distribution would be much more like the "step" function which is found for $T=0$.

It is quite easy to introduce very small amounts of substances such as arsenic, antimony or other elements belonging to group V of the periodic table into otherwise pure crystals of silicon or germanium as substitutional impurities, that is, as impurity atoms which occupy lattice sites which would normally be occupied by atoms of the covalent semiconductor. The group V atoms have five valence electrons apiece. Four of these are used in forming covalent electron pair bonds with neighboring semiconductor atoms; the fifth is bound to the impurity atom only by electrostatic forces which are quite weak, and it can therefore be easily ionized by thermal agitation of the lattice at ordinary temperatures to provide an *extra* conduction electron. The impurity atom which is left behind then becomes a positive ion, which, however, is *immobile* in view of the fact that it is strongly bound to four neighboring atoms by the usual covalent bonds. The situation is illustrated in Figure 9.6(a). In crystals containing this type of impurity there are more electrons than holes (although *some* holes are still present because hole-electron pairs are still created thermally on occasion). Such crystals are termed *n*-type semiconductors, the terminology arising from the fact that

most of the charge carriers are *negative* electrons. The component of electrical conductivity arising from the impurity atoms is called *impurity conductivity*, and a substance most of whose charge carriers originate from impurity atoms is called an *impurity semiconductor*. The substitutional group V atoms are often called *donor* atoms, because they each donate an additional free electron to the crystal. It takes so little energy (as we shall soon see) to ionize a typical group V donor impurity atom in Si and Ge that essentially *all* group V impurities in these materials are ionized at temperatures above about 20°K.

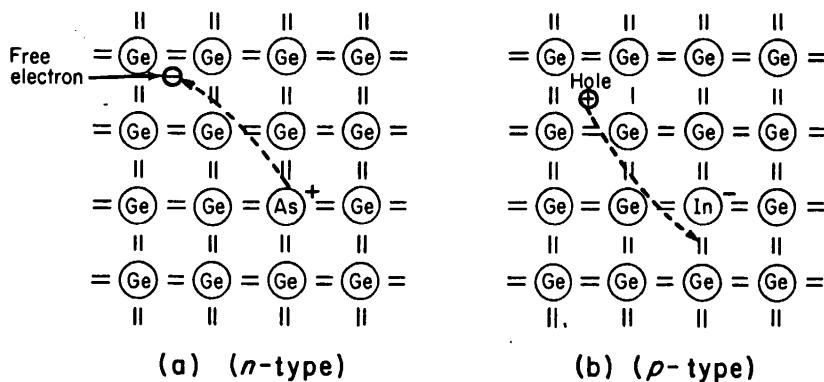


FIGURE 9.6. (a) Free electron arising from ionization of a substitutional arsenic impurity atom; (b) free hole arising from ionization of a substitutional indium impurity atom.

If, instead of group V atoms, group III impurity atoms (Al, Ga, In, and so forth) are introduced substitutionally into the lattice, a rather different effect is observed. These atoms have only three valence electrons, which are used in forming covalent pair bonds to three neighboring atoms, but the fourth bond of necessity lacks an electron. There is, in effect, an extra *hole* built into the covalent bond structure at the impurity atom. This hole can easily migrate away from the impurity site, because an extra electron from a neighboring covalent bond can migrate to the impurity site and fill in the fourth electron pair bond there (which, of course, puts a *negative* charge on the impurity atom); the hole is then associated with a neighboring atom, indistinguishable from a thermally created hole. This state of affairs is illustrated in Figure 9.6(b). The energy required for the migration of the hole away from the impurity site is of the order of the energy required to remove the extra electron from a donor atom. Therefore, except at very low temperatures, all the holes will be migratory and all the group III impurity atoms will have the character of immobile negative ions. In crystals containing predominantly this type of impurity there are more holes than electrons, although there will always be some electrons arising from thermal excitation. Crystals of this sort are called *p*-type semiconductors, because the charge carriers are for the most part *positive*. The substitutional group III atoms are usually termed *acceptor* atoms, because they readily accept an electron from the covalent bond structure, freeing a mobile hole. When impurities of either type are present in a semiconductor crystal, the conductivity is invariably *greater* than that associated with a pure or intrinsic semiconductor at the same temperature, due to the extra charge carriers contributed by the impurity atoms, and, in general, the greater the impurity concentration, the greater the conductivity.

The statistical picture of *n*-type and *p*-type semiconductors is characterized by the presence of the Fermi level above (for *n*-type) or below (for *p*-type) the position associated with the pure or intrinsic crystal. In an *n*-type crystal, for example, there cannot be more electrons than holes unless the Fermi level is adjusted upward from the

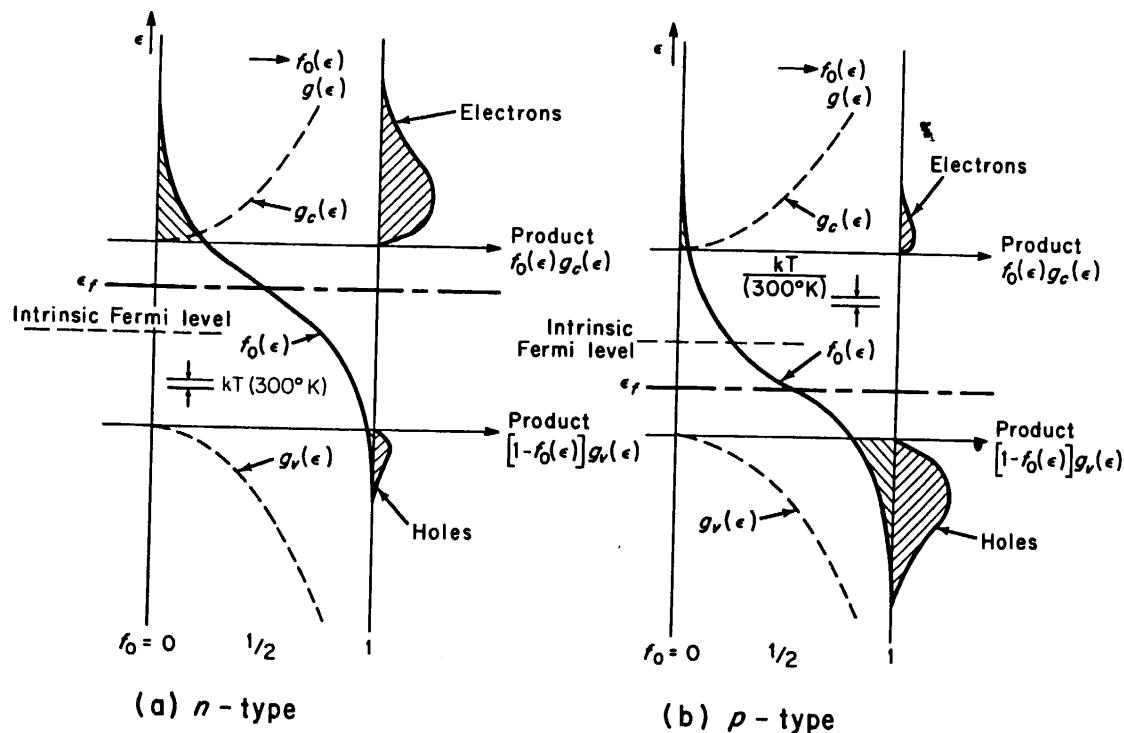


FIGURE 9.7. Distribution function, Fermi level, and electron and hole populations for (a) an *n*-type impurity semiconductor and (b) a *p*-type impurity semiconductor. Again, as in Figure 9.5, the "spread" of the Fermi distribution is exaggerated for illustrative purposes.

intrinsic position, and vice-versa for *p*-type, as shown in Figure 9.7. As the temperature and impurity concentration are varied, the position of the Fermi level changes in a rather complex way, about which we shall have more to say later.

9.3 STATISTICS OF HOLES AND ELECTRONS—THE CASE OF THE INTRINSIC SEMICONDUCTOR

Under conditions of thermal equilibrium, the number of electrons dn_0 per unit volume having energy in a range $d\epsilon$ about ϵ in the conduction band of *any* semiconductor, intrinsic or impurity, is, according to the results of Chapter 5

$$dn_0 = f_0(\epsilon)g_c(\epsilon) d\epsilon = \frac{8\sqrt{2\pi}}{h^3} m_n^{*3/2} \frac{\sqrt{\epsilon - \epsilon_c} d\epsilon}{1 + e^{\frac{\epsilon - \epsilon_f}{kT}}} \quad (9.3-1)$$

where $f_0(\epsilon)$ represents the equilibrium Fermi distribution function (5.5-20) and $g_c(\epsilon)$ is the density of states factor (9.2-1). Now $\Delta\epsilon$, the width of the forbidden energy

region is typically of the order of 1 eV, and kT at room temperature (300°K) is only about 1/40 eV. If the Fermi energy is *within the forbidden energy region, several kT units away from the edge of the conduction band* (that is, if $\epsilon_c - \epsilon_f \gg kT$), then for all energies belonging to the conduction band, the exponential factor in the denominator of the Fermi distribution function is much greater than unity. We may then represent the Fermi distribution function in the conduction band as

$$f_0(\epsilon) \cong e^{-(\epsilon - \epsilon_f)/kT} \quad (9.3-2)$$

which is a distribution function of essentially Maxwellian form. The situation here is substantially that which was discussed in Chapter 5 in connection with Equation (5.5-27). Since we have already shown in Figure 9.5 that the Fermi energy for an intrinsic semiconductor must lie well within the forbidden region of energy, and since $kT \ll \Delta\epsilon$ at all accessible temperatures as pointed out above, the condition $\epsilon_c - \epsilon_f \gg kT$ will always be satisfied in practice, and we shall always be able to use (9.3-2) for intrinsic semiconductors rather than the more complicated expression (5.5-20). As a matter of fact, in most cases involving impurity semiconductors we shall find that this condition is also satisfied, allowing us to use (9.3-2) also for impurity semiconductors. It is clear, for example, that this simplification would be valid for the impurity semiconductors of Figure 9.7. In both Figures 9.5 and 9.7 it is easily seen that the exponential "tail" of the Fermi distribution function is the only part of that function which overlaps the conduction and valence bands. The part of the Fermi function in the neighborhood of the Fermi energy, which is invariably the part which causes mathematical difficulties, coincides in all these cases with a region where the density of available electronic states is zero. We shall refer to the simplifying approximation (9.3-2) as the *Boltzmann approximation*.

Using this approximation in (9.3-1) and integrating to get the total number of electrons per unit volume in the conduction band, we find

$$n_0 = \int_{\epsilon_c}^{\infty} dn_0 = \frac{8\sqrt{2\pi}}{h^3} m_n^{*3/2} e^{\epsilon_f/kT} \int_{\epsilon_c}^{\infty} \sqrt{\epsilon - \epsilon_c} e^{-\epsilon/kT} d\epsilon. \quad (9.3-3)$$

Actually, the limits of integration should be from ϵ_c to the top of the conduction band, wherever that may be; however, since $f_0(\epsilon)$ falls off very rapidly with increasing ϵ (at least for reasonable values of temperature) most of the electrons are concentrated in states near the bottom of the conduction band, and so long as the top of the band is many kT units higher in energy than ϵ_c , it makes little difference whether we integrate to the top of the band or to infinity. Making the substitution

$$x_c = \frac{\epsilon - \epsilon_c}{kT} \quad (9.3-4)$$

the integral (9.3-3) can be expressed in the form

$$n_0 = \frac{8\sqrt{2\pi}}{h^3} (m_n^* kT)^{3/2} e^{\frac{\epsilon_f - \epsilon_c}{kT}} \int_0^{\infty} x_c^{1/2} e^{-x_c} dx_c. \quad (9.3-5)$$

According to (5.4-4), the value of the integral above is $\Gamma(3/2) = \sqrt{\pi}/2$. Substituting

this value for the integral, the final result may be written

$$n_0 = U_c e^{-\frac{\epsilon_c - \epsilon_f}{kT}} \quad (9.3-6)$$

where

$$U_c = 2(2\pi m_n^* k T / h^2)^{3/2}. \quad (9.3-7)$$

The numerical value of the quantity U_c can be written (using cgs units) as

$$U_c = (4.83 \times 10^{15})(m_n^*/m_0)^{3/2} T^{3/2} \text{ cm}^{-3}, \quad (9.3-8)$$

where m_0 is the inertial mass of the electron. For $m_n^* = m_0$ and $T = 300^\circ\text{K}$, this gives $U_c = 2.51 \times 10^{19} \text{ cm}^{-3}$.

The number of holes p_0 per unit volume in the valence band of a semiconductor can be found in a similar manner. In this case we may write as dp_0 , the number of holes in the valence band in a range of energy $d\varepsilon$ about ε , the expression

$$dp_0 = f_{p0}(\varepsilon) g_v(\varepsilon) d\varepsilon = [1 - f_0(\varepsilon)] g_v(\varepsilon) d\varepsilon. \quad (9.3-9)$$

The quantity $f_{p0}(\varepsilon)$ is the probability that a hole will be associated with a quantum state of energy ε , in other words, that such a state will be unoccupied. This probability is just one minus the probability of occupation $f_0(\varepsilon)$. Since $f_0(\varepsilon)$ is given by (5.5-20) it is easily seen that

$$f_{p0}(\varepsilon) = 1 - f_0(\varepsilon) = 1 - \frac{1}{1 + e^{(\varepsilon - \epsilon_f)/kT}} = \frac{1}{1 + e^{(\epsilon_f - \varepsilon)/kT}}. \quad (9.3-10)$$

Again, if the Fermi energy is several kT units above the edge of the valence band, that is, if $\epsilon_f - \epsilon_v \gg kT$, the exponential factor in the denominator of the expression (9.3-10) will be much larger than unity for all values of ε in the valence band, whereby one may set

$$f_{p0}(\varepsilon) \cong e^{-(\epsilon_f - \varepsilon)/kT}. \quad (9.3-11)$$

This is the Boltzmann approximation for holes in the valence band. Using (9.3-11) and the density of states function (9.2-2) in (9.3-9) and integrating over the valence band we obtain

$$p_0 = \frac{8\sqrt{2\pi}}{h^3} m_p^{*3/2} e^{-\epsilon_f/kT} \int_{-\infty}^{\epsilon_v} \sqrt{\epsilon_v - \varepsilon} e^{\varepsilon/kT} d\varepsilon. \quad (9.3-12)$$

Again, the lower limit of the integral has been taken as $-\infty$ rather than the bottom of the valence band; this simplification can be justified by essentially the same argument which was made in connection with (9.3-3) for the conduction band. Making the substitution

$$x_v = \frac{\epsilon_v - \varepsilon}{kT} \quad (9.3-13)$$

the integral can be expressed in the form

$$p_0 = \frac{8\sqrt{2}\pi}{h^3} (m_p^* kT)^{3/2} e^{-(\epsilon_f - \epsilon_v)kT} \int_0^\infty x_v^{1/2} e^{-x_v} dx_v, \quad (9.3-14)$$

which can be evaluated as a Γ -function as above to give

$$p_0 = U_v e^{-(\epsilon_f - \epsilon_v)/kT} \quad (9.3-15)$$

with

$$U_v = 2(2\pi m_p^* kT/h^2)^{3/2} \quad (9.3-16)$$

The numerical value of U_v may be obtained from (9.3-8), substituting m_p^* in the formula for m_n^* .

The product $n_0 p_0$ can easily be shown to be a function *only* of the energy gap $\Delta\epsilon$, the effective masses, and the temperature, *independent* of the Fermi level or impurity content. Thus, multiplying the expressions (9.3-6) and (9.3-15) for n_0 and p_0 , one may obtain

$$n_0 p_0 = U_c U_v e^{-(\epsilon_c - \epsilon_v)/kT} = U_c U_v e^{-\Delta\epsilon/kT}. \quad (9.3-17)$$

Using (9.3-7) and (9.3-16) this can be written in the form

$$n_0 p_0 = 4 \left(\frac{2\pi(m_p^* m_n^*)^{1/2} kT}{h^2} \right)^3 e^{-\Delta\epsilon/kT}. \quad (9.3-18)$$

For a given semiconducting substance, the effective masses and the energy gap $\Delta\epsilon$ are fixed; hence the product $n_0 p_0$ in a given material must be a function only of temperature. This is essentially a mass-action law governing the relative concentrations of holes and electrons in a given material. If the semiconductor is in the pure or intrinsic state, then the concentrations of holes and electrons must be equal, as we have already seen, because the only holes and electrons which can be present are those which are generated in pairs by direct thermal excitation of valence band electrons. In such a material, then

$$p_0 = n_0 = n_i(T) \quad (9.3.19)$$

where $n_i(T)$ stands simply for the number of holes or electrons per unit volume in an *intrinsic* sample of the semiconductor in question at temperature T . In a sample of the same substance which is *not* intrinsic but rather an *impurity* semiconductor, the number of holes and electrons is no longer equal; nevertheless, according to (9.3-17) *the product $n_0 p_0$ must be the same for this material as for the intrinsic substance*. We may then write, from (9.3-19)

$$n_0 p_0 = n_i^2(T) \quad (9.3-20)$$

with

$$n_i(T) = 2 \left(\frac{2\pi(m_p^* m_n^*)^{1/2} kT}{h^2} \right)^{3/2} e^{-\Delta\epsilon/2kT}, \quad (9.3-21)$$

SEC. 9.4

as given by (9.3-18). According to these results, the concentration of carriers in an intrinsic semiconductor is a strong function of temperature, increasing rapidly as the temperature rises, and is likewise strongly dependent upon the energy gap $\Delta\epsilon$, decreasing rapidly as $\Delta\epsilon$ increases.

The mass action relation (9.3-17) or (9.3-20) is strongly reminiscent of the laws governing the relative concentrations of ions in substances which are weakly ionized, for example the H^+ and OH^- ions in aqueous systems. As a matter of fact, semiconductors can be understood as an example of the classical Arrhenius theory of weakly ionized electrolytes in terms of a system in which a covalent bond dissociates into an electron and a hole with dissociation energy $\Delta\epsilon$.

In a purely intrinsic sample, where the hole and electron concentrations are equal, we may equate expressions (9.3-6) and (9.3-15), take the logarithm of both sides of the resulting equation, and solve for the Fermi energy, obtaining

$$\epsilon_f = \frac{1}{2}(\epsilon_v + \epsilon_c) + kT \ln \sqrt{U_v/U_c} \quad (\text{intrinsic}). \quad (9.3-22)$$

Since U_v and U_c are given by (9.3-16) and (9.3-7), we may express this result in the form

$$\epsilon_{fi} = \frac{1}{2}(\epsilon_v + \epsilon_c) + kT \ln(m_p^*/m_n^*)^{3/4} \quad (\text{intrinsic}), \quad (9.3-23)$$

using the symbol ϵ_{fi} to denote the Fermi energy of an intrinsic semiconductor. It is clear from this that in an intrinsic semiconductor the Fermi energy is displaced from the center of the energy gap by only a relatively small amount of energy as given by the second term on the right-hand side of (9.3-23). If $m_p^* = m_n^*$, the Fermi level is exactly at the midpoint of the forbidden region, as indicated previously.

69
he
n-
ly

9.4 IONIZATION ENERGY OF IMPURITY CENTERS

A donor center in a semiconductor consists of a fixed ion of charge $+e$ to which an electron is loosely bound. If the binding energy is small enough, the "orbit" of the electron will be quite large—so large, indeed, compared to the interatomic spacing that many atoms will be encompassed by the path of the electron about the donor ion.² Under these circumstances it will be approximately correct to regard the electron as being immersed in a uniform polarizable medium whose dielectric constant κ is the macroscopic dielectric constant of the semiconductor crystal, since the electrostatic force between electron and donor ion will then be modified, on the average, by the polarization of many intervening atoms. The picture is similar to that of a hydrogen atom in a uniform continuous medium of dielectric constant κ , if one can envision such a situation. This picture will be valid as long as κ is of sufficient magnitude so that the Bohr orbits are large compared with the interatomic spacing.

We may now work out the size of the orbits and associated energy levels according

² Although descriptive of the situation, this is not very good wave mechanical language; it would be more accurate to say that the wave function of the electron extends over many interatomic distances, so that many atoms are included within the region where the probability density is large.

to the Bohr theory discussed in Section 4.5. It would really be more accurate to use the wave mechanics of the hydrogen atom for these calculations, but since we know that the Bohr theory gives the correct energy levels, and since the Bohr radii and the wave mechanical average electron-nuclear distances are almost the same, we are confident that the simpler Bohr theory will yield the right answers for this very similar situation. The quantization condition for the orbital angular momentum holds in both cases, so that Equation (4.5-2) is still correct provided the electron mass is replaced by m_n^* . On the other hand, due to the polarization of the crystal, the electrostatic force between the donor ion and the bound electron is reduced by a factor κ , becoming $e^2/\kappa r_n^2$, so that Equation (4.5-3) should now be written

$$e^2/\kappa r_n^2 = m_n^* v_n^2/r_n = m_n^* r_n \omega_n^2. \quad (9.4-1)$$

Equations (9.4-1) and (4.5-2) may now be solved as simultaneous equations for r_n and ω_n giving [in analogy to (4.5-4) and (4.5-5)],

$$r_n = \frac{n^2 \hbar^2 \kappa}{m_n^* e^2} \quad (9.4-2)$$

and

$$\omega_n = \frac{m_n^* e^4}{n^3 \hbar^3 \kappa^2} \quad (9.4-3)$$

The kinetic and potential energies may now be evaluated as in (4.5-6) and (4.5-7), recalling, however, that the potential energy is now given by $-e^2/\kappa r_n$, whereby the total energy is

$$\epsilon_n = \epsilon_k + \epsilon_p = -\frac{m_n^* e^4}{2 n^2 \kappa^2 \hbar^2}. \quad (9.4-4)$$

The Bohr theory for the hydrogen atom is expressed by the above equations with $m_n^* = m_0$ and $\kappa = 1$. It is evident from (9.4-2) that the Bohr orbits for the donor center are larger than the hydrogen orbits by a factor κ and also by a factor m_0/m_n^* . For n -type germanium, $\kappa = 16$ and $m_n^* \cong m_0/4$, so that the radius of the first Bohr orbit is 64 times the radius of the first Bohr orbit of hydrogen. The value for hydrogen is 0.528 Å, so that the value for the donor atom in germanium is about 34 Å. This is indeed considerably greater than the interatomic distance (which is about 2.44 Å for germanium), hence this picture of the donor center is in this instance to a large extent physically justified.

The ionization energy of the donor center is, from (9.4-4), smaller than that of the hydrogen atom by a factor $1/\kappa^2$ and also by a factor m_n^*/m_0 . For germanium this amounts to $(1/256) \cdot (1/4) = 1/1024$. The energy of ionization for the ground state of a donor atom would then be expected to be about 1000 times less than that of the ground state of the hydrogen atom (13.6 eV). We arrive thus at the figure 0.013 eV for the energy of ionization of a donor atom in germanium. This value is in good agreement with the experimentally determined values obtained from measurements of carrier concentration as a function of temperature and from infrared optical absorption as given in Table 9.1. The higher values of ionization energy observed for silicon are accounted for partially by the smaller dielectric constant ($\kappa = 12$) for

that material, and partly by the higher effective masses for electrons in silicon. The characteristics of acceptor centers may be understood along the same lines of reasoning, using a picture involving a negatively charged acceptor ion and a positively charged hole in a uniform dielectric medium.

TABLE 9.1.
Ionization Energies for Donor and
Acceptor Centers in Ge and Si

Impurity	Ge	Si
P	0.012 eV	0.045 eV
As	0.0127	0.05
Sb	0.010	0.039
B	0.0104	0.045
Al	0.0102	0.06
Ga	0.0108	0.07
In	0.0112	0.16

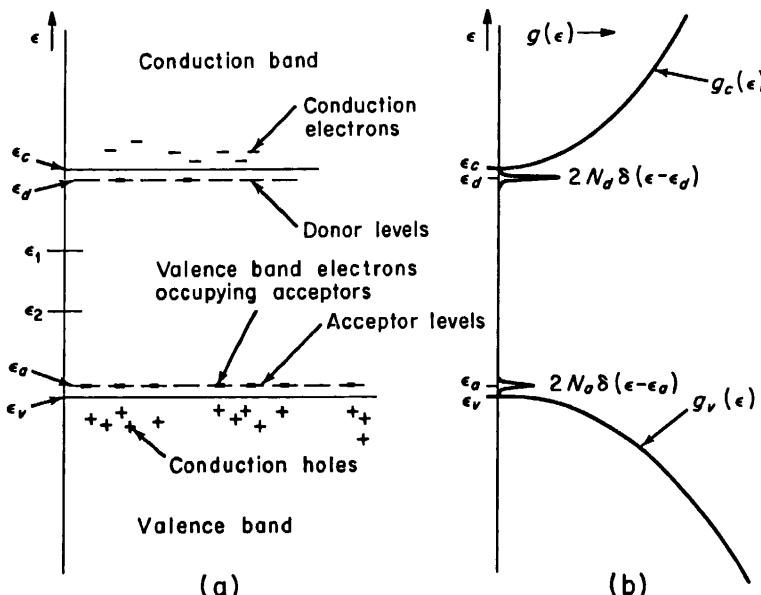


FIGURE 9.8. (a) Energy band diagram of an impurity semiconductor, showing donor and acceptor levels, (b) corresponding density of states curve.

The electrons contributed by donor atoms can thus be construed to originate from localized *donor states* which lie within the forbidden energy gap, a few hundredths of an electron volt *below* the conduction band. Likewise the holes contributed by acceptors can be visualized as being created when electrons which would normally occupy states near the top of the valence band are promoted into initially empty *acceptor levels* which lie in energy a few hundredths of an electron volt *above* the valence band. This scheme of *donor and acceptor levels* is shown in Figure 9.8(a), and the related density of states picture is shown in Figure 9.8(b).

9.5 STATISTICS OF IMPURITY SEMICONDUCTORS

Before discussing the actual statistical mechanics of impurity semiconductors it is necessary to examine a subtle feature of the occupation statistics of donor and acceptor levels. Confining our discussion for the moment to donor levels, one might be tempted to conclude that the occupation probability associated with such levels would be given simply by the equilibrium Fermi function (5.5-20) with $\epsilon = \epsilon_d$. This is not quite true, however, because of the spin degeneracy of the donor levels. There are really two quantum states associated with each donor impurity level, corresponding to the two allowed spin orientations of the electron on the donor atom. But as soon as one of these states is occupied occupancy of the other one is precluded, since the valency requirements of the donor ion are satisfied by one and only one electron. This situation alters the statistical problem which leads to the distribution function.

Consider a system having energy levels of this type, as shown in Figure 9.9. We

					---		---	
Energy level no.	1	2	3	4	---	<i>i</i>	---	<i>n</i>
Energy	ϵ_1	ϵ_2	ϵ_3	ϵ_4	---	ϵ_i	---	ϵ_n
Degeneracy	g_1	g_2	g_3	g_4	---	g_i	---	g_n
No. of electrons	N_1	N_2	N_3	N_4	---	N_i	---	N_n

FIGURE 9.9. Energy levels and notation used in the calculations of Section 9.5.

shall suppose that the electrons occupying the levels are indistinguishable particles, and that there are N_i electrons in the *i*th level, whose degeneracy is g_i . The number of ways of inserting the first electron into the quantum states belonging to the *i*th level is g_i ; the number of ways of inserting the second is $g_i - 2$, since the occupancy of the first state by the first electron precludes the occupancy of the state of opposite spin; the number of ways of inserting the third is $g_i - 4$, and so on. For the *i*th the number will be $g_i - 2N_i + 2$. The total number of ways of arranging N_i indistinguishable electrons in g_i states under these circumstances is then

$$\frac{g_i(g_i - 2)(g_i - 4) \cdots (g_i - 2N_i + 2)}{N_i!} = \frac{2^{g_i/2}(g_i/2)!}{\frac{g_i}{2} - N_i N_i! \left(\frac{g_i}{2} - N_i\right)!}. \quad (9.5-1)$$

The factor $N_i!$ in the denominator is included because those distributions which are identical except for the permutation of the electrons among themselves must not be counted as distinct distributions, since the electrons are indistinguishable one from another. The total number of ways of realizing a distribution wherein there are $N_1, N_2 \cdots N_n$ electrons in levels 1, 2, ... *n* is just the product of factors of the form (9.5-1) over the levels of the system. This is just the quantity $Q(N_1 N_2 \cdots N_n)$ of Chapter 5 [see, for example, (5.5-3)]. For this system, then, we have

$$Q_d(N_1, N_2 \cdots N_n) = \prod_{i=1}^n \frac{2^{N_i(g_i/2)!}}{N_i! \left(\frac{g_i}{2} - N_i\right)!}. \quad (9.5-2)$$

The resulting distribution function is found by maximizing $\ln Q_d$ as before, using the method of Lagrangean multipliers developed in Chapter 5, the result being

$$N_j = \frac{\frac{1}{2}g_j}{1 + \frac{1}{2}e^{(\epsilon_j - \epsilon_f)/kT}}, \quad (9.5-3)$$

where ϵ_f is the Fermi energy, defined in the usual way. For donor levels, the number of donor atoms is just half the number of spin states, or $\frac{1}{2}g_j$, so that for this system, we may write

$$n_d = \frac{N_d}{1 + \frac{1}{2}e^{(\epsilon_d - \epsilon_f)/kT}} \quad (9.5-4)$$

where N_d is the concentration of donor impurity atoms, ϵ_d is the energy of the donor levels, and n_d is the number of electrons per unit volume occupying the donor levels (thus the concentration of *unionized* donors). The Fermi function is clearly modified by the presence of the factor 1/2 before the exponential term in the denominator.

In a generally similar way one may show that if N_a is the concentration of acceptor impurity atoms and ϵ_a the energy of the acceptor levels, then the concentration of holes associated with acceptor atoms (thus the density of unionized acceptor sites) p_a will be

$$p_a = \frac{N_a}{1 + \frac{1}{2}e^{(\epsilon_f - \epsilon_a)/kT}}. \quad (9.5-5)$$

The density of states associated with the donor and acceptor levels may be represented in terms of Dirac δ -functions by

$$g_d(\epsilon) = 2N_d\delta(\epsilon - \epsilon_d) \quad (9.5-6)$$

and

$$g_a(\epsilon) = 2N_a\delta(\epsilon - \epsilon_a). \quad (9.5-7)$$

These equations express the fact that there are no donor or acceptor levels at energies other than ϵ_d and ϵ_a , and that the total number of donor and acceptor quantum states per unit volume is $2N_d$ and $2N_a$, respectively.

If the Boltzmann approximation is valid for *both* the conduction band and the donor levels (that is, if the Fermi level is several kT units below ϵ_d), the exponential factor in the denominator of the expression in (9.5-4) is much greater than unity, and hence the latter may be neglected; Equation (9.5-4) may then be written as

$$n_d = 2N_d e^{-(\epsilon_d - \epsilon_f)/kT}. \quad (9.5-8)$$

Since the concentration of electrons in the conduction band is given by (9.3-6) we may write the ratio of n_d , the number of electrons associated with unionized donors, to $n_0 + n_d$, the total number of free and loosely bound electrons as

$$\frac{n_d}{n_0 + n_d} = \frac{1}{1 + \frac{U_c}{2N_d} e^{-(\epsilon_c - \epsilon_d)/kT}}. \quad (9.5-9)$$

Now $\varepsilon_c - \varepsilon_d$ is the donor ionization energy, which is ordinarily of the order of or smaller than kT ; the exponential factor above is therefore of the order of unity. If $N_d \ll \frac{1}{2}U_c$ the ratio of the number of electrons on unionized donors to the total number will be very small. Since at 300°K U_c is of the order of 10^{19} cm^{-3} , this condition is satisfied at that temperature for all donor impurity concentrations which are small compared to this figure. In such instances, it is clear that the donors are *almost completely ionized*, and it is usually convenient and accurate to proceed on the assumption that their ionization is complete, setting n_d equal to zero. At very low temperatures the exponential factor in (9.5-9) may be appreciably smaller than unity, and the criterion for complete donor ionization must be written in the more general form

$$N_d \ll \frac{1}{2}U_c e^{-(\varepsilon_c - \varepsilon_d)/kT}. \quad (9.5-10)$$

In a similar fashion, provided the Boltzmann approximation for holes is valid both in the valence band and at the acceptor levels, one may show that the acceptor levels are practically completely ionized whenever

$$N_a \ll \frac{1}{2}U_v e^{-(\varepsilon_a - \varepsilon_v)/kT}. \quad (9.5-11)$$

Unless the temperature is very low, this amounts to the requirement $N_a \ll \frac{1}{2}U_v$. In most circumstances, the Fermi level will be within the forbidden gap, many kT units away from both acceptor and donor levels, and both acceptor and donor impurities which may be present will be almost completely ionized.

In a semiconductor at equilibrium there must be either a thermal hole or a positively charged donor ion for every free electron, and a thermal electron or a negatively charged acceptor ion for every free hole. The entire crystal must thus be electrically neutral. This electrical neutrality condition can be expressed by equating the algebraic sum of all negative and positive charges to zero, whereby (recalling that the concentration of *ionized* donors is $N_d - n_d$ and of ionized acceptors $N_a - p_a$),

$$\cancel{N_d - n_d + N_a - p_a} = 0 \quad p_0 - n_0 + N_d - N_a + p_a - n_d = 0. \quad (9.5-12)$$

For all but the lowest temperatures and the highest values of impurity concentration, the Boltzmann approximation will be valid for both conduction and valence bands and for both donor and acceptor levels, whereby the concentrations of unionized donors and acceptors, p_a and n_d , may be neglected in (9.5-12), giving

$$p_0 - n_0 + N_d - N_a = 0 \quad (9.5-13)$$

or, using (9.3-6) and (9.3-15),

$$U_v e^{-(\varepsilon_f - \varepsilon_v)/kT} - U_c e^{-(\varepsilon_c - \varepsilon_f)/kT} + (N_d - N_a) = 0. \quad (9.5-14)$$

If we let

$$\alpha = e^{\varepsilon_f/kT}, \quad \beta_c = e^{-\varepsilon_c/kT} \quad \text{and} \quad \beta_v = e^{\varepsilon_v/kT}, \quad (9.5-15)$$

then (9.5-14) can be written in the form

$$\alpha^2 - \frac{N_d - N_a}{U_c \beta_c} \alpha - \frac{U_v \beta_v}{U_c \beta_c} = 0. \quad (9.5-16)$$

Solving this quadratic equation for α , and taking the logarithm of the result, one may obtain

$$\ln \alpha = \frac{\varepsilon_f}{kT} = \ln \left[\frac{N_d - N_a}{2U_c \beta_c} + \sqrt{\left(\frac{N_d - N_a}{2U_c \beta_c} \right)^2 + \frac{U_v \beta_v}{U_c \beta_c}} \right]. \quad (9.5-17)$$

We must choose the positive sign of the radical in (9.5-17) since when $N_d = N_a = 0$, $e^{\varepsilon_f/kT}$ must be a *positive* quantity. This can be put in a somewhat more satisfying form by noting that

$$\ln(a + \sqrt{a^2 + x^2}) = \ln x + \sinh^{-1} \frac{a}{x}. \quad (9.5-18)$$

Using this relation, and substituting the values given by (9.5-15) for β_v and β_c , (9.5-17) can be expressed as

$$\varepsilon_f = \frac{1}{2}(\varepsilon_v + \varepsilon_c) + kT \ln \left(\frac{m_p^*}{m_n^*} \right)^{\frac{1}{2}} + kT \sinh^{-1} \left(\frac{N_d - N_a}{2\sqrt{U_c U_v} e^{-\Delta\varepsilon/2kT}} \right). \quad (9.5-19)$$

The first two terms on the right-hand side of this equation are seen from (9.3-23) to represent the Fermi level ε_{fi} for an intrinsic semiconductor, and since from (9.3-17) and (9.3-20) $n_i = \sqrt{U_c U_v} e^{-\Delta\varepsilon/2kT}$, we may write (9.5-19) finally as

$$\varepsilon_f = \varepsilon_{fi} + kT \sinh^{-1} \left(\frac{N_d - N_a}{2n_i} \right). \quad (9.5-20)$$

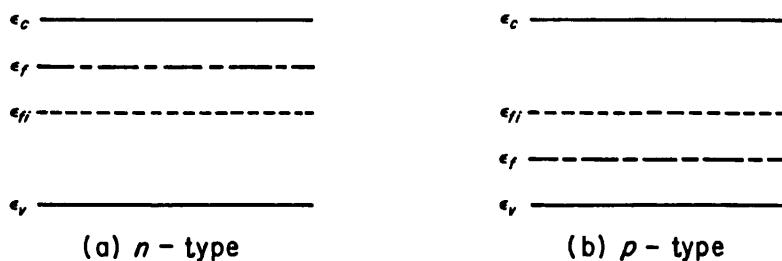


FIGURE 9.10. Relative positions of band edges, intrinsic Fermi level, and actual Fermi level in (a) *n*-type and (b) *p*-type semiconductors.

This expression gives the Fermi level for an impurity semiconductor in the range over which the Boltzmann approximation is satisfied and the donors and acceptors can be regarded as completely ionized. The quantities n_0 , p_0 , n_d and p_a can easily be obtained from (9.3-6), (9.3-15), (9.5-4), and (9.5-5) once the Fermi level has been found.

Since $\sinh^{-1} x$ is positive for $x > 0$ and negative for $x < 0$, it is easily seen that $\varepsilon_f > \varepsilon_{fi}$ for *n*-type semiconductors ($N_d - N_a > 0$), and $\varepsilon_f < \varepsilon_{fi}$ for *p*-type semiconductors ($N_d - N_a < 0$) as illustrated in Figure 9.10. If the number of donors and

acceptors are equal, the inverse hyperbolic sine function vanishes and the material behaves exactly like an intrinsic semiconductor so far as electron and hole populations are concerned. Under these conditions the *n*- and *p*-type impurities are said to be fully compensated. If the net impurity density $|N_d - N_a|$ is much larger than n_i , the number of thermally excited carriers will be small compared with the total number; in this case the argument of the inverse hyperbolic sine function in (9.5-20) is very large. Since $\sinh^{-1} x \cong \pm \ln |2x|$ for large values of x , it is clear that under these circumstances (9.5-20) becomes

$$\varepsilon_f = \varepsilon_{fi} \pm kT \ln \frac{|N_d - N_a|}{n_i}, \quad (9.5-21)$$

the plus sign being used for *n*-type material ($N_d > N_a$), the minus sign for *p*-type ($N_a > N_d$). A semiconducting material of this type is said to be a strongly *extrinsic* semiconductor, and the range of applicability of (9.5-21) is often referred to as the *extrinsic range*.

The density of carriers in the conduction and valence bands can, of course, be found by substituting the expression (9.5-17) for the Fermi level into (9.3-6) and (9.3-15). These quantities may also be determined from the electrical neutrality condition (9.5-13), with the help of (9.3-20). For example, substituting $p_0 = n_i^2/n_0$ into (9.5-13), we obtain

$$n_0^2 - (N_d - N_a)n_0 - n_i^2 = 0 \quad (9.5-22)$$

which may be solved for n_0 , giving

$$n_0 = \frac{1}{2}(N_d - N_a) + \sqrt{\frac{1}{4}(N_d - N_a)^2 + n_i^2}. \quad (9.5-23)$$

In solving the quadratic equation, it is necessary to choose the positive sign for the radical, since the result must reduce to $+n_i$ for $N_d = N_a = 0$. Likewise, substituting $n_0 = n_i^2/p_0$ into the charge neutrality equation, we can show that

$$p_0 = -\frac{1}{2}(N_d - N_a) + \sqrt{\frac{1}{4}(N_d - N_a)^2 + n_i^2}. \quad (9.5-24)$$

From these formulas, it is clear that when $N_d - N_a = 0$, $n_0 = p_0 = n_i$. This is the case of complete compensation (or, if $N_d = N_a = 0$, of an intrinsic semiconductor). If $N_d - N_a > 0$, $n_0 > p_0$ as expected for an *n*-type impurity semiconductor; if $N_d - N_a < 0$, $p_0 > n_0$ as required for a *p*-type semiconductor. For strongly *extrinsic* materials these formulas become much simpler. For example, if $N_d - N_a \gg n_i$, as it is for a strongly *n*-type sample, then

$$n_0 \cong N_d - N_a \quad \text{and} \quad p_0 \cong \frac{n_i^2}{N_d - N_a} \quad (9.5-25)$$

while for a strongly *p*-type material where $N_d - N_a$ is large and negative,

$$p_0 \cong N_a - N_d \quad \text{and} \quad n_0 \cong \frac{n_i^2}{N_a - N_d}. \quad (9.5-26)$$

For the sake of complete generality we have assumed from the beginning that *both* donor and acceptor impurities are present. In many cases, however, especially those involving crystals into which impurities of one type have been intentionally incorporated, the concentration of the *other* impurity species may be completely neglected and indeed set equal to zero in all the formulas of this section.

9.6

CASE OF INCOMPLETE IONIZATION OF IMPURITY LEVELS (VERY LOW TEMPERATURE)

At very low temperatures there may not be sufficient thermal energy available to maintain complete ionization of donor and acceptor impurity levels, as determined by the criteria (9.5-10) and (9.5-11). At zero absolute temperature, for example, in an *n*-type semiconductor, the donor levels must be completely occupied and the conduction band completely devoid of electrons. Both these conditions cannot be realized unless the Fermi level lies *between* the donor levels and the conduction band, as depicted in Figure 9.11. In this case the Boltzmann approximation can no longer be used to

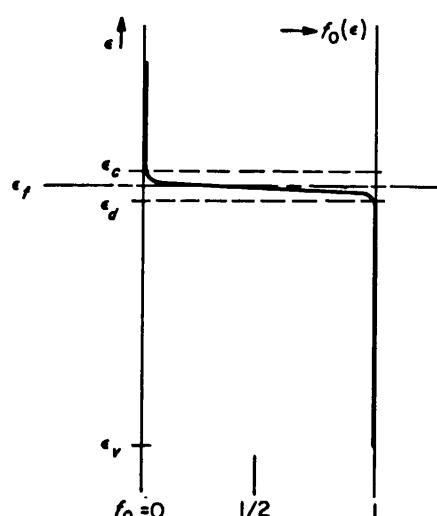


FIGURE 9.11. Distribution function in an *n*-type semiconductor at a very low temperature.

describe the occupation probability associated with the donor levels, since the Fermi level is now above rather than below the donor energy levels. The Boltzmann approximation can, however, still be used for the electrons in the conduction band, because the Fermi level will remain many kT units below the bottom of the conduction band; it should be recalled in this connection that kT is only about 0.001 eV at 10°K.

We shall treat in detail the case of a purely *n*-type semiconductor where $N_a = 0$, and we shall assume that T is so low that the concentration of thermally created holes in the valence band is negligible. Since $N_a = 0$, of course, p_a must also vanish. Setting $N_a = p_a = p_0 = 0$ in (9.5-12), one may write

$$n_0 - N_d + n_d = 0. \quad (9.6-1)$$

Using (9.3-6) and (9.5-4) to express n_0 and n_d in this equation, we obtain

$$U_c e^{-(\varepsilon_c - \varepsilon_f)/kT} - N_d \left(1 - \frac{1}{1 + \frac{1}{2} e^{(\varepsilon_d - \varepsilon_f)/kT}} \right) = 0. \quad (9.6-2)$$

However, since

$$1 - \frac{1}{1 + \frac{1}{2} e^{(\varepsilon_d - \varepsilon_f)/kT}} = \frac{1}{1 + 2e^{-(\varepsilon_d - \varepsilon_f)/kT}}, \quad (9.6-3)$$

(9.6-2) can be written as

$$U_c e^{-(\varepsilon_c - \varepsilon_f)/kT} + 2U_c e^{-(\varepsilon_c + \varepsilon_d - 2\varepsilon_f)/kT} - N_d = 0. \quad (9.6-4)$$

Letting

$$\beta_c = e^{-\varepsilon_c/kT}, \quad \beta_d = e^{-\varepsilon_d/kT} \quad \text{and} \quad \alpha = e^{\varepsilon_f/kT}, \quad (9.6-5)$$

(9.6-4) becomes

$$\alpha^2 + \frac{\alpha}{2\beta_d} - \frac{N_d}{2U_c \beta_c \beta_d} = 0. \quad (9.6-6)$$

This equation can be solved for α to yield

$$\alpha = -\frac{1}{4\beta_d} + \sqrt{\frac{1}{16\beta_d^2} + \frac{N_d}{2U_c \beta_c \beta_d}}. \quad (9.6-7)$$

The plus sign must be chosen in order that $\alpha = e^{\varepsilon_f/kT}$ be positive and thus ε_f be real.
Now, by definition

$$\ln(x + \sqrt{x^2 + a^2}) = \ln a + \sinh^{-1} \frac{x}{a}. \quad (9.6-8)$$

Taking the logarithm of both sides of (9.6-7), using (9.6-8) and reverting to original notation with (9.6-5), one may finally obtain

$$\varepsilon_f = \frac{1}{2}(\varepsilon_c + \varepsilon_d) + \frac{kT}{2} \ln \frac{N_d}{2U_c} - kT \sinh^{-1} \left(\sqrt{\frac{U_c}{8N_d}} e^{-\Delta\varepsilon_i/2kT} \right), \quad (9.6-9)$$

where $\Delta\varepsilon_i = \varepsilon_c - \varepsilon_d$ is the donor ionization energy. The density of free electrons in the conduction band may now be calculated from (9.3-6).

At absolute zero, (9.6-9) gives $\varepsilon_f = \frac{1}{2}(\varepsilon_c + \varepsilon_d)$, which means that the Fermi level is midway between the donor levels and the conduction band. As the temperature increases, the Fermi level first increases slightly (remaining nevertheless below ε_c) and then decreases, moving down through the donor levels toward the center of the gap. For temperatures such that the Fermi level is several kT units below the donor energy ε_d at which point the donors are almost completely ionized, Equation (9.6-9) above

C. 9.6

SEC. 9.7

UNIFORM ELECTRONIC SEMICONDUCTORS

277

6-2)

6-3)

6-4)

5-5)

i-6)

-7)

al.

-8)

ial

9)

in
el
re
id
p.
gy
ve

reduces, approximately, to (9.5-21). The proof of this statement is assigned as an exercise. It can be shown in a similar way that for a *p*-type semiconductor under these same circumstances, the Fermi level at zero temperature lies midway between the acceptor levels and the valence band, and with increasing temperature moves first downward a bit, then upward, approaching the value given by (9.5-21) as the acceptor levels become fully ionized. In both *n*-type and *p*-type semiconductors the Fermi level approaches the *intrinsic* value ϵ_{fi} as given by (9.3-23) for sufficiently high temperatures, since at some high temperature the number n_i of thermally excited carriers will be far in excess of the number $|N_d - N_a|$ contributed by donor and acceptor impurities. Beyond this point the behavior of the material will in every way approximate that of an intrinsic semiconductor. In some instances, however, this transition temperature may exceed the melting point of the semiconductor. The Fermi level, as determined by (9.6-9) in the low temperature range and by (9.5-20) at higher temperatures is shown as a function of temperature for both *n*-type and *p*-type materials in Figure 9.12.

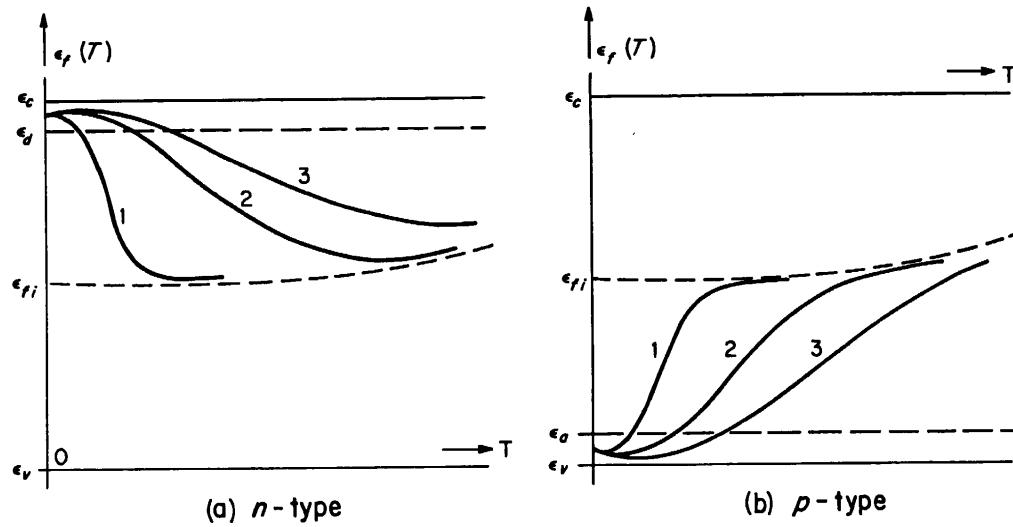


FIGURE 9.12. Variation of Fermi energy with temperature for (a) *n*-type and (b) *p*-type semiconductors, of various impurity densities. The curves marked "1" correspond to relatively low impurity densities, those marked "2" to intermediate densities, and those marked "3" to relatively high impurity densities.

9.7 CONDUCTIVITY

The current density carried by holes and by electrons is equal to the product of their respective charge densities and velocities. Thus, the current density, or the charge density per unit time transported by holes and electrons is

$$I_n = \rho_n v_n = -env_n \quad (9.7-1)$$

and

$$I_p = \rho_p v_p = e p v_p \quad (9.7-2)$$

where ρ_n and ρ_p refer to the charge densities associated with the electron and hole densities n and p , and v_n and v_p are the average vector velocities of electrons and holes, respectively. Since the average electron velocity is $-\mu_n E$ and the average hole velocity is $\mu_p E$, where μ_n and μ_p are the electron and hole mobilities and E is the field, we may write (9.7-1) and (9.7-2) as

$$\mathbf{I}_n = ne\mu_n \mathbf{E} \quad (9.7-3)$$

and

$$\mathbf{I}_p = pe\mu_p \mathbf{E}, \quad (9.7-4)$$

whereby the total electrical current density \mathbf{I} may be expressed as

$$\mathbf{I} = e(n\mu_n + p\mu_p) \mathbf{E} = \sigma \mathbf{E}. \quad (9.7-5)$$

The electrical conductivity σ is then given by

$$\sigma = e(n\mu_n + p\mu_p). \quad (9.7-6)$$

The mobilities μ_n and μ_p in the above formulas are defined as

$$\mu_n = \frac{e\bar{\tau}_n}{m_n^*} \quad \text{and} \quad \mu_p = \frac{e\bar{\tau}_p}{m_p^*} \quad (9.7-7)$$

where $\bar{\tau}_n$ and $\bar{\tau}_p$ are weighted averages of the relaxation times τ_n for electrons and τ_p for holes over the Maxwell-Boltzmann distribution, calculated as prescribed by Equation (7.3-12).

It should be noted also that n and p are the actual instantaneous values of hole and electron concentration, which are not necessarily identical with the *equilibrium* hole and electron densities n_0 and p_0 . It is, however, so far assumed that no *gradients* of carrier density (which would give rise to *diffusion currents*) are present. If the carrier densities are those associated with the equilibrium state, of course, we shall find that

$$\sigma_0 = e(n_0\mu_n + p_0\mu_p) \quad (9.7-8)$$

is the corresponding value of conductivity. It is necessary to make this distinction because it is quite possible to create carrier densities in excess of the equilibrium values in semiconductors, as we shall see later.

For an intrinsic semiconductor with carrier densities equal to the equilibrium values, $n_0 = p_0 = n_i$ and (9.7-8) becomes

$$\sigma_0 = en_i(\mu_n + \mu_p) = en_i\mu_p(b + 1) \quad (9.7-9)$$

where b is defined as the ratio of electron to hole mobility, that is,

$$b = \mu_n/\mu_p. \quad (9.7-10)$$

EC. 9.7

SEC. 9.7

UNIFORM ELECTRONIC SEMICONDUCTORS

279

hole
holes,
locity
may

9.7-3)

9.7-4)

1.7-5)

1.7-6)

.7-7)

d τ_p
l byhole
rium
ients
rrier
that

7-8)

tion
ium

ium

7-9)

.10)

Using (9.3-21), (9.7-9) can be written in the form

$$\sigma_0 = 2e\mu_p(b + 1) \left(\frac{2\pi(m_p^* m_n^*)^{1/2} kT}{h^2} \right)^{1/2} e^{-\Delta\varepsilon/2kT}. \quad (9.7-11)$$

Since the mobility μ_p usually has a temperature dependence which largely cancels the $T^{1/2}$ temperature variation of the term preceding the exponential factor, and since b is not strongly temperature dependent, the above variation of σ_0 as a function of $1/T$ is essentially exponential. A semilogarithmic plot of σ_0 versus $1/T$ thus yields a straight line of slope $\Delta\varepsilon/2k$, and the value of the energy gap $\Delta\varepsilon$ may be rather accurately determined by determining the slope of an experimentally measured plot of $\ln \sigma_0$ versus $1/T$ in the intrinsic range.

For a semiconductor which is not necessarily intrinsic, but whose hole and electron densities correspond to the equilibrium values n_0 , p_0 , the conductivity is given by (9.7-8). Using (9.7-10) to eliminate μ_n , and then substituting the expressions (9.5-23) and (9.5-24) for n_0 and p_0 , we may write the conductivity as

$$\sigma_0 = e\mu_p (bn_0 + p_0) = e\mu_p \left[\frac{1}{2} (b - 1)(N_d - N_a) + (b + 1) \sqrt{\frac{1}{4}(N_d - N_a)^2 + n_i^2} \right]. \quad (9.7-12)$$

Consider for the moment an *n*-type semiconductor, for which $N_d - N_a$ is positive. At low temperatures, $n_i \ll N_d - N_a$, whereupon, according to (9.7-12), $\sigma_0 \cong (N_d - N_a)b\mu_p = (N_d - N_a)e\mu_n$, which is independent of temperature except insofar as μ_n may depend on T . The semiconductor is then a strongly extrinsic material, most of the carriers being contributed by impurity atoms. As the temperature is increased, n_i will increase rapidly, according to (9.3-21), and at length n_i will equal, then exceed, $N_d - N_a$. For temperatures such that $n_i \gg N_d - N_a$, (9.7-12) gives $\sigma_0 \cong e\mu_p n_i(b + 1)$, the same value given by (9.7-9) for an intrinsic semiconductor, and the conductivity then rises rapidly with increasing temperature due to the rapid increase of n_i . The sample is now an essentially intrinsic semiconductor. The transition between these two extremes of behavior occurs at the temperature where $n_i = N_d - N_a$. As the impurity density $N_d - N_a$ increases this transition temperature increases, and for different semiconducting substances of equal impurity density the transition temperature increases as the band gap $\Delta\varepsilon$ increases, because the temperature required to generate a given density of intrinsic carriers increases with $\Delta\varepsilon$. The same qualitative behavior may be shown to follow from (9.7-12) for *p*-type materials, wherein $N_d - N_a$ is negative. A semilogarithmic plot of conductivity versus $1/T$ as predicted by (9.7-12) for several samples of varying impurity content is shown in Figure 9.13.

At extremely low temperatures (less than about 30°K) there is not sufficient thermal energy available to maintain the ionization of donor and acceptor levels. In this temperature range the conductivity again falls rapidly with decreasing temperature as free electrons and holes are "frozen out" of the conduction and valence bands into donor and acceptor levels. In crystals of quite high impurity density the wave functions of electrons on neighboring donor atoms or of holes on neighboring acceptor atoms may overlap appreciably. The result of this is to broaden the donor or acceptor level into a narrow energy band, just as atomic energy levels broaden into a band, in the tight binding approximation, as the electronic wave functions of neighboring atoms

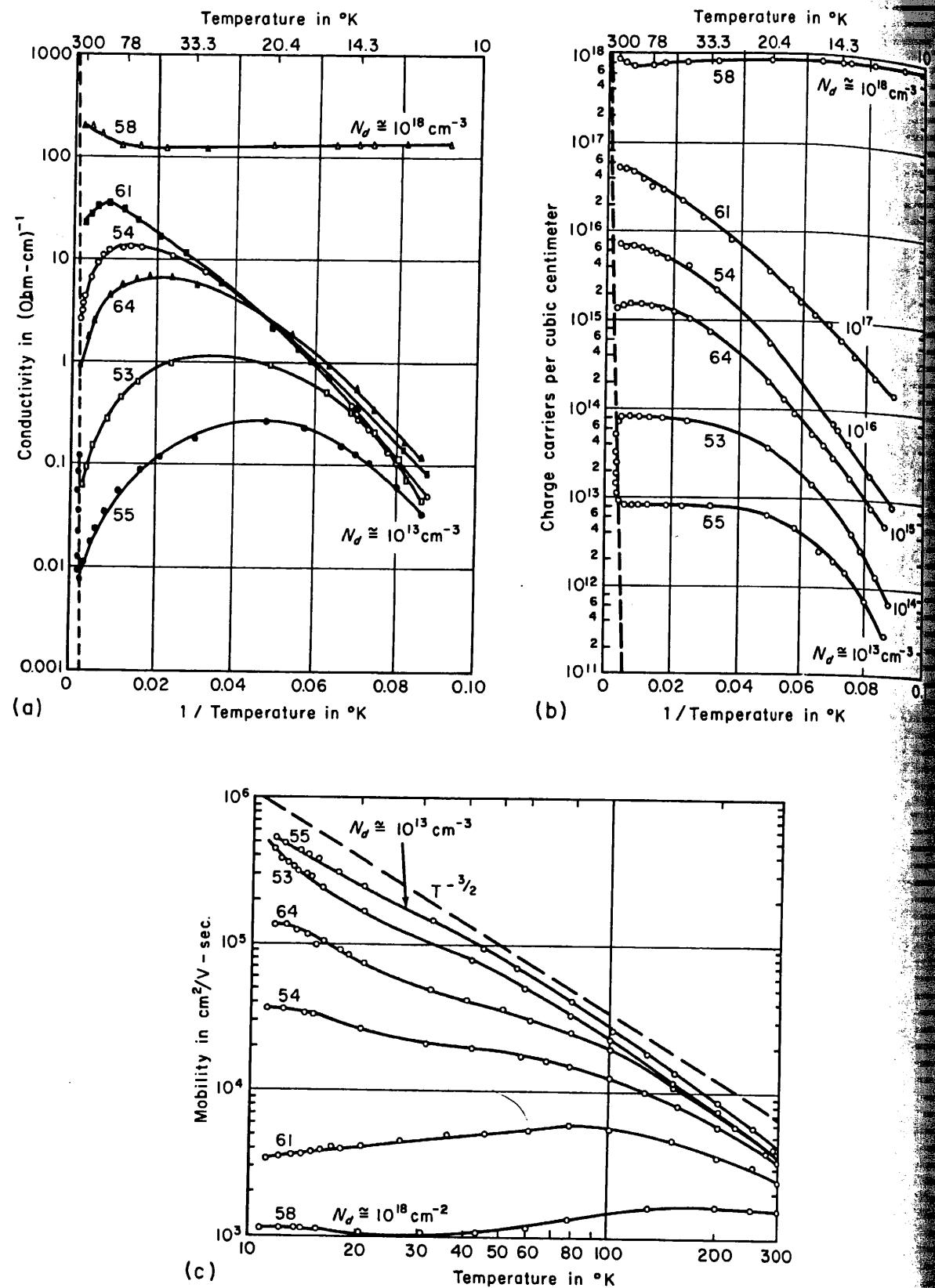


FIGURE 9.13. (a) Electrical conductivity, (b) carrier concentration and (c) electron mobility for a series of *n*-type germanium samples covering a wide range of donor impurity content. [After E. M. Conwell, *Proc. I.R.E.*, 40: 1327 (1952).]

begin to overlap. Since there are twice as many states in this band as there are electrons (or holes), a small residual *impurity band conductivity* may be observed in the partially filled impurity band in the extreme low temperature region where electrons and holes are "frozen out" of the conduction and valence bands into the donor and acceptor levels.

9.8 THE HALL EFFECT AND MAGNETORESISTANCE

In Section 7.6 the Hall effect in free-electron gases was discussed for the case of electrons or holes all having the same x -component of velocity. This analysis leads to the expression

$$E_y = RI_0B_0 \quad (9.8-1)$$

for the transverse Hall field, where I_0 is the current density, B_0 is the magnetic induction and R is a constant of proportionality called the Hall coefficient. The orientation of the sample and of the current and fields is as shown in Figure 7.3. In the simple case discussed in Section 7.6, where the charge carriers are assumed to be electrons, the expression

$$R = -\frac{1}{n_0ec} \quad (\text{electrons}) \quad (9.8-2)$$

was obtained for the Hall coefficient. It is easy to see from the analysis of Section 7.6 that if the charge carriers are positive holes whose concentration is p_0 , the expression for the Hall coefficient would be

$$R = \frac{1}{p_0ec} \quad (\text{holes}). \quad (9.8-3)$$

We have already pointed out that the effect of the *distribution* of velocities has been neglected in obtaining these results. We must now attempt to include the effect of the velocity distribution in redetermining the Hall coefficient for a semiconductor. In addition, we must account for the fact that in a semiconductor positive and negative charge carriers may be present *simultaneously*.

Beyond these effects which are of immediate interest, there are two other complications with which we shall eventually be concerned. Up to this point, in all the transport and carrier population calculations we have made, we have always gone under the assumption that the surfaces of constant energy in momentum space or k -space were spherical. This assumption arises naturally when the minimum value of energy in the reduced Brillouin zone occurs at the central point $k = 0$. But there are substances, including germanium and silicon (conduction bands only), in which the energy extremum occurs not at the origin, but at a set of crystallographically equivalent points elsewhere in the Brillouin zone. In such instances, the constant energy surfaces in the neighborhood of the energy minima may be ellipsoidal, rather than spherical. The effect of this is to introduce an anisotropy into the effective mass

associated with each ellipsoid; the carriers, although behaving as free particles in the sense that the variation of ϵ with \mathbf{k} is parabolic for any direction of motion, have a different effective mass along each such direction. Although the anisotropy in the transport properties vanishes when one sums over the energy ellipsoids associated with all the energy minimum points, the procedure for computing density of states factors and averaging transport properties over velocities must be modified to account for this rather different configuration of constant energy surfaces.

In addition to this, the *valence* bands of germanium and silicon really consist of two distinct overlapping bands with each of which is associated a different effective mass. Although the constant energy surfaces for both of these bands are roughly spherical near the maximum point at $k = 0$, there are actually two species of holes ("light" holes and "heavy" holes corresponding to the two effective masses) present simultaneously.

Any theory of transport processes which seriously attempts to explain all the very abundant experimental data in this field relating to germanium and silicon must of necessity take into account both these added complications. But to begin with, we shall have enough to do in extending the picture of the Hall effect and related phenomena to cover electrons and holes simultaneously and to include the effect of the distribution of velocities. We shall therefore continue to assume that we are dealing with a simple semiconductor with spherical energy surfaces in both conduction and valence band, each having only one variety of electrons or holes. We shall later indicate how our results may be extended to allow for the two additional difficulties discussed above.

We shall begin by considering only a single type of carrier (we shall choose holes to work with) and incorporating the effect of the velocity distribution. The most obvious way of proceeding is from the Boltzmann equation with the Lorentz force term (7.6-2). It is simpler, however, and quite instructive to use an alternative method³ which begins with the equation of motion for a hole in the presence of both a steady electric field \mathbf{E} and a steady magnetic induction \mathbf{B}_0 . This equation can be written

$$\mathbf{F} = m_p^* \frac{d\mathbf{v}}{dt} = e\mathbf{E} + \frac{e}{c} \mathbf{v} \times \mathbf{B}_0. \quad (9.8-4)$$

If we assume that the sample, current, and magnetic field are oriented as shown in Figure 7.3, then, clearly \mathbf{B}_0 has only a z -component, and (9.8-4) reduces to two equations for the x - and y -velocity components, which have the form

$$dv_x/dt = \frac{eE_x}{m_p^*} + \omega_0 v_y \quad (9.8-5)$$

$$dv_y/dt = \frac{eE_y}{m_p^*} - \omega_0 v_x, \quad (9.8-6)$$

where

$$\omega_0 = eB_0/m_p^*c. \quad (9.8-7)$$

³ See, for example, H. Brooks, *Advances in Electronics and Electron Physics*. New York: Academic Press, Inc. (1955), Vol. VII, p. 127 ff.

-9.8

SEC. 9.8

UNIFORM ELECTRONIC SEMICONDUCTORS

283

This set of equations can be most easily solved by multiplying (9.8-6) by the imaginary number i and adding the result to (9.8-5). One obtains thereby the single equation

$$\frac{dV}{dt} + i\omega_0 V = \frac{e\mathcal{E}}{m_p^*} \quad (9.8-8)$$

where the complex quantities V and \mathcal{E} are defined by

$$V = v_x + iv_y \quad \text{and} \quad \mathcal{E} = E_x + iE_y. \quad (9.8-9)$$

Equation (9.8-8) is easily solved by multiplying both sides by $e^{i\omega_0 t}$. The left-hand side then can be written as the time derivative of $Ve^{i\omega_0 t}$, and the resulting equation integrated with respect to time to give

$$Ve^{i\omega_0 t} = \frac{e\mathcal{E}}{i\omega_0 m_p^*} e^{i\omega_0 t} + C \quad (9.8-10)$$

where C is a constant of integration. If we let $V_0 = v_{x0} + iv_{y0}$ be the value of V at $t = 0$, C can be evaluated in terms of V_0 from this equation, allowing us to write (9.8-10) as

$$V = V_0 e^{-i\omega_0 t} + \frac{e\mathcal{E}}{i\omega_0 m_p^*} (1 - e^{-i\omega_0 t}). \quad (9.8-11)$$

This expression must first be averaged over the exponential distribution of relaxation times, whereby, in analogy with (7.2-13)

$$\langle V \rangle_{\text{Av over } t} = \frac{\int_0^\infty V(t) e^{-t/\tau_p} dt}{\int_0^\infty e^{-t/\tau_p} dt} = \frac{1}{1 + i\omega_0 \tau_p} \left[V_0 + \frac{e\mathcal{E}\tau_p}{m_p^*} \right], \quad (9.8-12)$$

where τ_p represents the relaxation time for holes. We must now average this expression over a Maxwell-Boltzmann distribution of velocities. If we assume, as usual, that τ_p is a function of the magnitude v , but that it is independent of *direction*, i.e., isotropic, then the averages associated with the first term in (9.8-12) will always amount to an average of a function only of the magnitude of v times a component v_{x0} ($= v_0 \sin \theta \cos \phi$) or v_{y0} ($= v_0 \sin \theta \sin \phi$). Such an average is always zero, because the integral of the resulting expression over the azimuthal coordinate ϕ always vanishes. We may thus drop this term from our consideration altogether. If we separate the second term above into real and imaginary parts, we may identify the real part with v_x and the imaginary part with v_y , according to (9.8-9). We obtain thus

$$\begin{aligned} \langle V \rangle_{\text{Av over } t} &= \frac{e}{m_p^*} \left[\left(\tau_p - \frac{\omega_0^2 \tau_p^3}{1 + \omega_0^2 \tau_p^2} \right) E_x + \frac{\omega_0 \tau_p^2}{1 + \omega_0^2 \tau_p^2} E_y \right] \\ &\quad + i \frac{e}{m_p^*} \left[\frac{\tau_p}{1 + \omega_0^2 \tau_p^2} E_y - \frac{\omega_0 \tau_p^2}{1 + \omega_0^2 \tau_p^2} E_x \right], \end{aligned} \quad (9.8-13)$$

and it is this expression which must be averaged over the Boltzmann distribution to obtain average values for v_x and v_y . We saw in Section 7.3, using the Boltzmann equation, that if $\alpha(v)$ is the quantity to be averaged, the proper average to take is

$$\bar{\alpha} = \frac{\langle v^2 \alpha(v) \rangle}{\langle v^2 \rangle}. \quad (9.8-14)$$

In the expression (9.8-13), if there is no magnetic field then $\omega_0 = 0$ and the equation gives simply the velocities v_x and v_y in terms of the electric fields E_x and E_y . In this limit the equation should agree with (7.3-11), and indeed it does, provided that τ_p is averaged according to the prescription (9.8-14). As a matter of fact, in the general case all the averages must be computed in this way, leading to

$$\bar{V} = \bar{v}_x + i\bar{v}_y \quad (9.8-15)$$

where

$$\bar{v}_x = \text{Re}(\bar{V}) = \frac{e}{m_p^*} \left[\bar{\tau}_p E_x + \omega_0 \left(\frac{\tau_p^2}{1 + \omega_0^2 \tau_p^2} \right) E_y - \omega_0^2 \left(\frac{\tau_p^3}{1 + \omega_0^2 \tau_p^2} \right) E_x \right] \quad (9.8-16)$$

and

$$\bar{v}_y = \text{Im}(\bar{V}) = \frac{e}{m_p^*} \left[\left(\frac{\tau_p}{1 + \omega_0^2 \tau_p^2} \right) E_y - \omega_0 \left(\frac{\tau_p^2}{1 + \omega_0^2 \tau_p^2} \right) E_x \right]. \quad (9.8-17)$$

The x - and y -components of the current density may, in the customary way, be expressed as

$$I_x = p_0 e \bar{v}_x \quad \text{and} \quad I_y = p_0 e \bar{v}_y. \quad (9.8-18)$$

In the usual experimental arrangement, as shown in Figure 7.3, the y -component of current is zero, whereby, from (9.8-17)

$$E_y = \omega_0 \frac{\tau_p^2 / (1 + \omega_0^2 \tau_p^2)}{\tau_p / (1 + \omega_0^2 \tau_p^2)} E_x. \quad (9.8-19)$$

Substituting this in (9.8-16) one may obtain finally

$$I_x = \frac{p_0 e^2}{m_p^*} \left[\bar{\tau}_p + \omega_0^2 \left(\frac{[\tau_p^2 / (1 + \omega_0^2 \tau_p^2)]^2}{\tau_p / (1 + \omega_0^2 \tau_p^2)} - \frac{[\tau_p^3 / (1 + \omega_0^2 \tau_p^2)]}{\tau_p / (1 + \omega_0^2 \tau_p^2)} \right) \right] E_x. \quad (9.8-20)$$

If τ_p is independent of v , then $\overline{f(\tau_p)} = f(\tau_p)$ and the above equations reduce to the very simple and intuitive results

$$I_x = \sigma_0 E_x = \frac{p_0 e^2 \tau_p}{m_p^*} E_x \quad (9.8-21)$$

$$E_y = \omega_0 \tau_p E_x = \frac{e B_0 \tau_p}{m_p^* c} \frac{I_x}{\sigma_0} = \frac{1}{p_0 e c} B_0 I_x \quad (9.8-22)$$

which are in agreement with (7.3-14) and with (9.8-1) and (9.8-3). It is clear that the simple expression for the Hall coefficient obtained previously is correct even when there is a Boltzmann distribution of velocities, provided that τ is independent of velocity. It is also evident from (9.8-21) that in this case the electrical conductivity is independent of the magnetic field; there is no *magnetoresistance*.

Usually, however, τ_p does depend on velocity, and then these results are no longer correct. Under such circumstances it is most convenient to make the simplifying assumption that $\omega_0\tau_p \ll 1$. This condition will be satisfied best for small values of the magnetic induction and small values of τ_p , which are generally associated with higher temperatures. It is quite generally satisfied for most substances under commonly encountered experimental conditions down to the liquid nitrogen temperature range, although it is also usually quite possible to violate it, if that is what is desired. For example, if $B_0 = 10,000$ oersteds, $\tau_p = 10^{-12}$ sec (corresponding to a mobility of the order of 2000) and $m_p^* = m_0$, then $\omega_0\tau_p \approx 0.2$. The requirement would then be marginally satisfied, but higher fields or larger relaxation times would result in its violation. If the condition is assumed to hold, the quantity $\omega_0^2\tau_p^2$ in (9.8-20) can be neglected throughout in comparison with unity. This equation may then be written [expressing ω_0 by (9.8-7)] as

$$I_x = \sigma_0 E_x \left[1 - \frac{e^2 B_0^2}{m_p^{*2} c^2} \frac{\bar{\tau}_p^3 \tau_p - (\bar{\tau}_p^2)^2}{(\bar{\tau}_p)^2} \right] \quad (9.8-23)$$

where $\sigma_0 = p_0 e^2 \bar{\tau}_p / m_p^*$. It is apparent now that the electrical conductivity is dependent on the magnetic field; this effect is called *magnetoresistance*. It can be shown that the second term in brackets above is always positive, so that the conductivity is invariably *reduced* by the magnetic field. Equation (9.8-23) is more commonly written in a slightly different way; solving for E_x in terms of I_x and noting that $(1 - \omega_0^2\tau_p^2)^{-1} \approx 1 + \omega_0^2\tau_p^2$ for $\omega_0\tau_p \ll 1$, we may obtain

$$E_x = \rho_0 I_x \left[1 + \frac{e^2 B_0^2}{m_p^{*2} c^2} \frac{\bar{\tau}_p^3 \tau_p - (\bar{\tau}_p^2)^2}{(\bar{\tau}_p)^2} \right] = (\rho_0 + \Delta\rho) I_x, \quad (9.8-24)$$

where $\rho_0 = 1/\sigma_0$ is the zero-field resistivity. The *magnetoresistance coefficient* is then defined as

$$\frac{\Delta\rho}{\rho_0 B_0^2} = \frac{e^2}{m_p^{*2} c^2} \frac{\bar{\tau}_p^3 \tau_p - (\bar{\tau}_p^2)^2}{(\bar{\tau}_p)^2}. \quad (9.8-25)$$

It is clear from (9.8-24) that the small-field longitudinal magnetoresistance predicted by these equations is proportional to the *square* of the magnetic induction B_0 . Physically, the magnetoresistance effect is due to the fact that the magnetic field deflects the particles upward or downward along the y -axis in a direction normal to the direction of the current vector. The resulting trajectories are curved rather than straight, and the average drift distance along the current flow direction between successive collisions is thereby reduced.

The Hall field E_y , using the same approximation in (9.8-19), becomes

$$E_y = \omega_0 \frac{\bar{\tau}_p^2}{\tau_p} E_x. \quad (9.8-26)$$

The field E_x in this equation may be expressed with good accuracy as I_x/σ_0 , the magnetoresistive term in (9.8-24) being small compared to the zero-field term for $\omega_0\tau_p \ll 1$. Substituting this value, along with the expressions for σ_0 and ω_0 into (9.8-26), we find that

$$E_y = RB_0I_x \quad (9.8-27)$$

where [recalling (9.8-14)] the Hall coefficient R is given by

$$R = \frac{1}{p_0ec} \frac{\bar{\tau}_p^2}{(\bar{\tau}_p)^2} = \frac{1}{p_0ec} \frac{\langle v^2 \rangle \langle v^2 \tau_p^2 \rangle}{\langle v^2 \tau_p \rangle^2}. \quad (9.8-28)$$

The Hall coefficient is modified by the factor $(\bar{\tau}_p^2)/(\bar{\tau}_p)^2$ from the value given by the simple theory of Chapter 7. It is evident also that the value given by (9.8-27) is a small-field value, good only so long as $\omega_0\tau_p \ll 1$; should $\omega_0\tau_p$ equal or exceed unity, the magnetoresistance term in (9.8-24) would become comparable in magnitude to the zero-field term and we could no longer replace E_x by I_x/σ_0 in (9.8-26). We should have to express E_x as I_x/σ , where σ has a strong magnetic field dependence. The Hall coefficient would then exhibit a significant dependence upon the magnetic field. The extension of these analyses to the case where $\omega_0\tau$ is no longer small compared with unity involves extensive numerical calculations and will not be attempted here. In the above analysis, we have assumed that the charge carriers are holes. The result for an n -type specimen, wherein the charge carriers are electrons, can be obtained by replacing e with $-e$ throughout. It should be noted that when this is done the characteristic frequency (9.8-7) becomes $-eB_0/m_n^*c$, and therefore one must also replace ω_0 with $-\omega_0$ throughout. In this manner, one may readily ascertain that when the carriers are electrons, the sign of the Hall coefficient is changed, but the form of the conductivity and magnetoresistance coefficients are unaltered.

If the dominant scattering process is acoustical mode lattice scattering, then, according to the results of Section 7.5, the mean free path will be essentially independent of velocity, whereby $\tau_p(v) = \lambda_p/v$. It is then clear that $\langle v^2 \tau_p^2 \rangle = \langle \lambda_p^2 \rangle = \lambda_p^2$; also, from (7.3-12) and (7.3-18) we see that $\langle v^2 \rangle = 3kT/m$ and $\langle v^2 \tau_p \rangle = \lambda_p \langle v \rangle = \lambda_p \bar{c}$, where \bar{c} is given by (5.4-20). Substituting these values into (9.8-28), we obtain $\bar{\tau}_p^2/(\bar{\tau}_p)^2 = 3\pi/8$, and the Hall coefficient then becomes

$$\begin{aligned} R &= \frac{3\pi}{8} \frac{1}{p_0ec} \quad (p\text{-type}) \\ &= -\frac{3\pi}{8} \frac{1}{n_0ec} \quad (n\text{-type}). \end{aligned} \quad (9.8-29)$$

The magnetoresistance coefficient (9.8-25) may be evaluated in this case using the same general approach. One may in fact show that when acoustical mode phonon scattering is the dominant mechanism, (9.8-23) becomes

$$I_x = \sigma_0 E_x \left[1 - \frac{e^2 B_0^2}{m_p^{*2} c^2} \frac{\lambda_p^2}{kT} \left(\frac{4-\pi}{8} \right) \right].$$

The details of deriving this result are assigned as an exercise for the student.

It would seem, from this expression, that the magnetoresistance might become

SEC. 9.8

UNIFORM ELECTRONIC SEMICONDUCTORS

287

very large at low temperatures, and this is indeed possible. Under these circumstances, however, the contribution of small amounts of impurity scattering, which causes a magnetoresistance effect proportional to T^3 becomes important and in fact represents the limiting factor as the absolute temperature approaches zero.

Consider now a semiconductor crystal which is not strongly extrinsic, but which contains both free electrons and holes in significant concentrations. In this case, the holes and electrons can be considered separately by exactly the same methods used before in deriving equations (9.8-16) and (9.8-17). The hole and electron velocity components are then expressed as equations having the form of (9.8-16) and (9.8-17), the velocity components \bar{v}_{px} and \bar{v}_{py} for the holes being identical with those equations, while the electron velocity components \bar{v}_{nx} and \bar{v}_{ny} are given by equations wherein e is replaced by $-e$, m_p^* by m_n^* , τ_p by τ_n (the latter symbol referring to the relaxation time for electrons in the conduction band), and eB_0/m_p^*c by $-eB_0/m_n^*c$. If one makes the further assumption that $\omega_0\tau \ll 1$ for both holes and electrons so that the term $\omega_0^2\tau^2$ may be neglected in the denominators of the averages, then one may write for the current components the expressions

$$I_x = -n_0e\bar{v}_{nx} + p_0e\bar{v}_{px} = \frac{n_0e^2}{m_n^*} [E_x\bar{\tau}_n - E_y\omega_{on}\bar{\tau}_n^2 - E_x\omega_{on}^2\bar{\tau}_n^3] \\ + \frac{p_0e^2}{m_p^*} [E_x\bar{\tau}_p + E_y\omega_{op}\bar{\tau}_p^2 - E_x\omega_{op}^2\bar{\tau}_p^3], \quad (9.8-30)$$

$$\text{and } I_y = -n_0e\bar{v}_{ny} + p_0e\bar{v}_{py} = \frac{n_0e^2}{m_n^*} (E_y\bar{\tau}_n + E_x\omega_{on}\bar{\tau}_n^2) + \frac{p_0e^2}{m_p^*} (E_y\bar{\tau}_p - E_x\omega_{op}\bar{\tau}_p^2), \quad (9.8-31)$$

$$\text{where } \omega_{on} = \frac{eB_0}{m_n^*c} \quad \text{and} \quad \omega_{op} = \frac{eB_0}{m_p^*c}. \quad (9.8-32)$$

Using (9.7-7) to eliminate the effective masses and rearranging the terms, the above equations may be put in the form

$$I_x = n_0e\mu_n [E_x - E_y\omega_{on}(\bar{\tau}_n^2/\bar{\tau}_n) - E_x\omega_{on}^2(\bar{\tau}_n^3/\bar{\tau}_n)] \\ + p_0e\mu_p [E_x + E_y\omega_{op}(\bar{\tau}_p^2/\bar{\tau}_p) - E_x\omega_{op}^2(\bar{\tau}_p^3/\bar{\tau}_p)], \quad (9.8-33)$$

$$\text{and } I_y = E_x((n_0e\mu_n\omega_{on}(\bar{\tau}_n^2/\bar{\tau}_n) - p_0e\mu_p\omega_{op}(\bar{\tau}_p^2/\bar{\tau}_p)) + E_y(n_0e\mu_n + p_0e\mu_p)). \quad (9.8-34)$$

Since the Hall voltage is measured under open-circuit conditions, $I_y = 0$, as before, whereby (9.8-34) reduces to

$$E_y = \frac{p_0\mu_p\omega_{op}\frac{\bar{\tau}_p^2}{\bar{\tau}_p} - n_0\mu_n\omega_{on}\frac{\bar{\tau}_n^2}{\bar{\tau}_n}}{n_0\mu_n + p_0\mu_p} \cdot E_x \quad (9.8-35)$$

By substituting this expression back into (9.8-33), we may after some algebra obtain

$$I_x = \sigma_0 E_x \left[1 - \frac{n_0 \mu_n \omega_{on}^2 \frac{\tau_n^3}{\bar{\tau}_n} + p_0 \mu_p \omega_{op}^2 \frac{\tau_p^3}{\bar{\tau}_p}}{n_0 \mu_n + p_0 \mu_p} - \left(\frac{n_0 \mu_n \omega_{on} \frac{\tau_n^2}{\bar{\tau}_n} - p_0 \mu_p \omega_{op} \frac{\tau_p^2}{\bar{\tau}_p}}{n_0 \mu_n + p_0 \mu_p} \right)^2 \right]. \quad (9.8-36)$$

where σ_0 , as usual, is given by (9.7-8). The magnetoresistance term is now much more complex, but is nevertheless small compared to the zero-field value provided that $\omega_{on} \tau_n \ll 1$ and $\omega_{op} \tau_p \ll 1$. The Hall field is given by (9.8-35), and the small-field Hall coefficient can be obtained by replacing E_x by I_x/σ_0 , as before, the difference between σ and σ_0 as represented by the magnetoresistance term being negligible under these conditions. Expressing ω_{on} and ω_{op} by (9.8-32) and eliminating the effective masses in favor of the mobilities and relaxation times by (9.7-7), we obtain

$$E_y = RB_0 I_x = \frac{1}{ec} \frac{p_0 \frac{\tau_p^2}{(\bar{\tau}_p)^2} - b^2 n_0 \frac{\tau_n^2}{(\bar{\tau}_n)^2}}{(bn_0 + p_0)^2} B_0 I_x. \quad (9.8-37)$$

where, as always, $b = \mu_n/\mu_p$.

This expression gives the Hall coefficient in a nonextrinsic semiconductor. It is seen to reduce to (9.8-29) when $p_0 \gg n_0$ or $n_0 \gg p_0$. If the dominant scattering process for both electrons and holes is acoustical mode phonon scattering⁴ then $\tau_n^2/(\bar{\tau}_n)^2 = \tau_p^2/(\bar{\tau}_p)^2 = 3\pi/8$ as shown previously and (9.8-37) then yields

$$R = \frac{3\pi}{8ec} \frac{p_0 - b^2 n_0}{(bn_0 + p_0)^2}. \quad (9.8-38)$$

For an intrinsic semiconductor $p_0 = n_0 = n_i$, and (9.8-38) reduces to

$$R = \frac{3\pi}{8} \frac{1}{n_i ec} \frac{1 - b}{1 + b} \quad (\text{Intrinsic}). \quad (9.8-39)$$

For $p_0 \gg n_0$, the Hall coefficient, as given by (9.8-37), is positive; for $n_0 \gg p_0$ it is negative. A reversal of the sign of the Hall coefficient takes place for values of p_0 and n_0 such that the numerator of (9.8-37) vanishes. If the conditions under which (9.8-38) is valid are satisfied, this reversal of the sign of the Hall field occurs when $p_0 = b^2 n_0$. Since $n_0 p_0 = n_i^2$, this amounts simply to $p_0 = bn_i$ or $n_0 = n_i/b$. If $b \neq 1$, the reversal point will not coincide exactly with the intrinsic point. For example, for $b > 1$, there will be a range of carrier concentrations wherein the sign of the Hall coefficient is negative despite the fact that the sample is *p*-type!

The Hall effect, aside from any fundamental interest it may excite, is important because it affords a practical way of telling whether a sample is *n*-type or *p*-type and determining the concentration of carriers. When combined with measurements of conductivity, the Hall effect can also be used to determine the mobility of the charge

⁴ Despite the fact that formula (9.8-38) is very frequently used to analyze Hall measurements in nonextrinsic materials, these conditions are often not strictly fulfilled—usually because of the presence of optical mode phonon scattering.

carriers. In samples which are highly *extrinsic* and in which the predominant scattering mechanism is acoustical mode phonon scattering, it is clear from (9.8-29) that a measurement of R can be used directly to obtain n_0 or p_0 , and the conductivity type follows from the sign of R . If other scattering mechanisms are important, then, of course, in order to obtain an absolute measurement of carrier concentration, the quantity $\bar{\tau}^2/(\bar{\tau})^2$ associated with the dominant scattering mechanism must be known, and in order to evaluate this one must know how τ varies as a function of velocity. However, the factor $\bar{\tau}^2/(\bar{\tau})^2$ is not much greater nor less than unity for any known scattering process which occurs in actual practice, so that a measurement of carrier concentration which is accurate within about a factor of two can always be obtained whatever the scattering mechanisms are, and this is often all that is required.

The mobility of the charge carriers in extrinsic samples can be obtained by eliminating the carrier concentration between the expression (9.8-28) and the equation $\sigma_0 = p_0 e \mu_p$, which results from (9.7-8) when the concentration of the minority carriers (electrons in this example) is neglected. The result, for a *p*-type semiconductor, is

$$\mu_p = c R \sigma_0 \frac{(\bar{\tau}_p)^2}{\tau_p^2}. \quad (9.8-40)$$

For an *n*-type extrinsic semiconductor, one may likewise show that

$$\mu_n = -c R \sigma_0 \frac{(\bar{\tau}_n)^2}{\tau_n^2}. \quad (9.8-41)$$

The product $\pm c R \sigma_0$ is often defined as the *Hall Mobility*, thus

$$\begin{aligned} \mu_{Hn} &= -c R \sigma_0 && (\text{i}n\text{-t}y\text{p}e) \\ \mu_{Hp} &= c R \sigma_0 && (\text{p}-\text{t}y\text{p}e). \end{aligned} \quad (9.8-42)$$

Equations (9.8-40) and (9.8-41) can then be written

$$\mu_p = \mu_{Hp} ((\bar{\tau}_p)^2 / \tau_p^2) \quad (9.8-43)$$

and

$$\mu_n = \mu_{Hn} ((\bar{\tau}_n)^2 / \tau_n^2). \quad (9.8-44)$$

The quantities μ_p and μ_n are the true *drift mobilities* or *conductivity mobilities*; the Hall mobility on the other hand is simply a quantity with the dimensions of mobility which is easily obtained from quantities which can be quite simply measured experimentally. The factor $(\bar{\tau})^2/\tau^2$ must be known to convert Hall mobility into the true drift mobility. Since this factor is usually not far from unity, however, the Hall mobility gives a rough indication of the true mobility in most cases which arise in connection with experiment. Experimental measurements of the carrier concentration and electron Hall mobility for *n*-type germanium samples as a function of temperature are illustrated in Figure 9.13 for a wide range of donor impurity concentrations.

For nonextrinsic samples, one proceeds in much the same way. From (9.8-37) or (9.8-38) one may eliminate either n_0 or p_0 by means of the mass action relation $n_0 p_0 = n_i^2$. The resulting equation, expressing n_0 or p_0 as a function of R is of the

fourth degree and cannot be treated analytically in a simple way, but may be solved in each instance by approximate or numerical methods. In addition, there is an ambiguity in that there are two real roots rather than one. This can be seen from Figure 9.14, where the Hall coefficient is plotted as a function of p_0 ; for a given value of R there are always *two* possible values of p_0 . In practice, however, it is usually fairly easy to choose which root is the right one by making a series of Hall effect

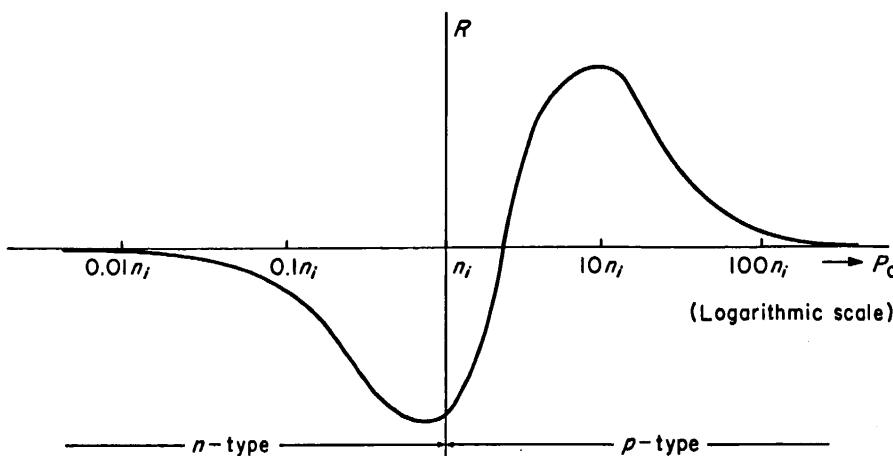


FIGURE 9.14. Hall coefficient of a semiconductor crystal as a function of hole concentration.

measurements over a wide range of temperatures. In addition, in order to obtain the carrier concentrations by the above procedure, it is necessary to know the mobility ratio b . This may be ascertained by making separate measurements of Hall mobility in highly extrinsic *n*- and *p*-type samples, or from the results of excess carrier drift measurements, which will be discussed later. If n_0 and p_0 are ascertained from the Hall coefficient in this way, and if the ratio $\mu_n/\mu_p = b$ is known, the mobilities μ_n and μ_p may be determined from conductivity data, by means of (9.7-8).

9.9

CYCLOTRON RESONANCE AND ELLIPSOIDAL ENERGY SURFACES

The electrons and holes in semiconductors may be made to exhibit a resonance in the presence of a large constant magnetic field (which we shall assume to lie along the z -axis) and a small transverse oscillating radiofrequency or microwave electromagnetic field whose electric vector lies in the xy -plane. The resulting hole or electron orbits are circular at resonance, and energy is absorbed from the rf-field every half cycle, as in a cyclotron. The situation is, in fact, precisely analogous to what happens to nuclear charged particles in a cyclotron, and for this reason the phenomenon is termed *cyclotron resonance*.

Let us initially investigate the case of spherical energy surfaces and a single scalar effective mass. The Lorentz force on a hole (and we shall consider holes in a *p*-type sample in this specific calculation) is given by (9.8-4). If the effective mass m_p^* is

independent of direction, and if the electric vector of the oscillating field, whose frequency is ω , is assumed to vibrate along the x -direction, then

$$E_x = E_0 e^{i\omega t}; \quad E_y = E_z = 0 \quad (9.9-1)$$

$$\text{while} \quad B_z = B_0; \quad B_x = B_y = 0, \quad (9.9-2)$$

and the components of the Lorentz force give for the equations of motion

$$F_x = m_p^* \frac{d^2 x}{dt^2} = \frac{e}{c} v_y B_0 + e E_0 e^{i\omega t} \quad (9.9-3)$$

$$F_y = m_p^* \frac{d^2 y}{dt^2} = -\frac{e}{c} v_x B_0 \quad (9.9-4)$$

$$F_z = m_p^* \frac{d^2 z}{dt^2} = 0. \quad (9.9-5)$$

In writing these equations it has been assumed that the magnetic vector of the rf-field is negligible compared with the constant field B_0 . The z -component equation (9.9-5) tells us merely that the particle moves with constant velocity along the z -direction; this is of no particular interest and we need trouble ourselves no further about it. The x - and y -component equations may be written

$$\frac{d^2 x}{dt^2} = \omega_0 \frac{dy}{dt} + \frac{e E_0}{m_p^*} e^{i\omega t} \quad (9.9-6)$$

$$\frac{d^2 y}{dt^2} = -\omega_0 \frac{dx}{dt} \quad (9.9-7)$$

where ω_0 is the very same frequency defined by (9.8-7) in connection with the Hall effect.

If oscillatory solutions of the form $x = x_0 e^{i\omega t}$ and $y = y_0 e^{i\omega t}$ are assumed these equations reduce to

$$i\omega\omega_0 y_0 + \omega^2 x_0 = -e E_0 / m_p^* \quad (9.9-8)$$

$$\text{and} \quad -\omega^2 y_0 + i\omega\omega_0 x_0 = 0 \quad (9.9-9)$$

which can be solved for the amplitudes x_0 and y_0 to give

$$\dot{x}_0 = \frac{e E_0 / m_p^*}{\omega_0^2 - \omega^2} \quad (9.9-10)$$

$$y_0 = \frac{i\omega_0}{\omega} x_0 = \frac{i\omega_0 e E_0 / m_p^*}{\omega(\omega_0^2 - \omega^2)}. \quad (9.9-11)$$

When $\omega = \omega_0 = e B_0 / m_p^* c$ resonance occurs and the amplitudes become very large.

The resonance frequency ω_0 is often referred to as the *cyclotron frequency*. Clearly, if ω_0 is measured experimentally and if the magnetic induction B_0 is accurately known, the effective mass m_p^* can be determined. In practice, the cyclotron resonance effect affords one of the best methods of measuring effective masses. From (9.9-10) and (9.9-11), it can be seen that as $\omega \rightarrow \omega_0$

$$y_0 \cong ix_0 = x_0 e^{i\pi/2} \quad (9.9-12)$$

whereby

$$x(t) = x_0 e^{i\omega_0 t} \quad (9.9-13)$$

$$y(t) = y_0 e^{i\omega_0 t} = x_0 e^{i(\omega_0 t + \pi/2)}. \quad (9.9-14)$$

At resonance, then, $x(t)$ and $y(t)$ are orthogonal harmonic vibrations of equal amplitude which differ in phase by 90° . It is easily seen that the resultant particle trajectory, representing the hole orbit, is circular.

In practice, the resonance is usually observed by measuring the *Q-factor* of a microwave resonant cavity as the magnetic field B_0 (and hence the frequency ω_0) is varied. When ω_0 equals the microwave excitation frequency, a sharp decrease in the *Q* of the cavity is obtained, because at resonance a great deal of electromagnetic field energy is absorbed when carriers are excited to large resonance amplitudes and then are scattered by impurities or lattice vibrations. It is worth noting that in this experiment one wishes to make the quantity $\omega_0 \tau$ much *larger* than unity, so that the carriers can be excited through several complete orbits before being scattered. Under these conditions they are capable of absorbing a great deal of microwave energy and transforming it, when scattering takes place, into vibrations of the crystal lattice. The experiment is usually performed at low temperature, so as to obtain a large value of τ , and with as large a static field B_0 as is possible consistent with the requirement that ω_0 be an experimentally accessible microwave frequency. In the case of Hall effect measurements, on the other hand, the objective is often to make $\omega_0 \tau$ much *less* than unity so as to simplify the interpretation of the data.

When the cyclotron resonance effect in silicon and germanium was studied experimentally, it was found for both substances that there were *more* than two resonances (one for electrons, one for holes) as predicted by the simple theory discussed above.^{5,6} Moreover, it was observed that some of the cyclotron resonance frequencies underwent significant changes as the orientation of the crystal with respect to the static magnetic field was altered, while others were practically independent of sample orientation. These effects were finally explained using a model in which the constant energy surfaces for electrons in k -space are ellipsoidal rather than spherical, and in which the presence of three separate valence bands, two of which are degenerate at $k = 0$, are considered. The constant energy surfaces for holes in these valence bands are roughly spherical, although since $\partial^2 \epsilon / \partial k^2$ differs for the different valence bands there is more than one possible effective mass for holes. The various features of this model have been verified not only by the agreement with experimental data which has been obtained, but also by quantum-mechanical calculations of the energy band structure which begin, more or less, from first principles.^{7,8,9}

⁵ G. Dresselhaus, A. F. Kip and C. Kittel, *Phys. Rev.*, **92**, 827 (1953).

⁶ R. N. Dexter, H. J. Zeiger and B. Lax, *Phys. Rev.*, **104**, 637 (1956).

⁷ F. Herman, *Phys. Rev.*, **93**, 1214 (1954); *Proc. Inst. Radio Engrs.*, **43**, 1703 (1955).

⁸ D. P. Jenkins, *Physica*, **20**, 967 (1954).

⁹ E. M. Conwell, *Proc. Inst. Radio Engrs.*, **46**, 1281 (1958).

We shall first consider the matter of ellipsoidal energy surfaces, and investigate how the results of the cyclotron resonance experiment are modified when the energy surfaces have this form. Up to this point we have always assumed that the minimum energy point in the reduced Brillouin zone occurs at $k = 0$, in the center of the zone, as illustrated in Figure 8.6. In this case an electron near the energy minimum exhibits the dynamical behavior of a free electron, leading to constant energy surfaces which, within this region of \mathbf{k} -space, are approximately spherical, as shown in Figure 8.11(a).

It has been found, however, that the minimum value of energy within the reduced zone need not necessarily be at $k = 0$, but may lie elsewhere within the zone. This is indeed the observed state of affairs for electrons in the conduction band of germanium and silicon. Since the diamond lattice is really two interpenetrating face-centered cubic lattices separated along the cube diagonal by a distance $a\sqrt{3}/4$ (where a is the cube edge), the primitive unit cell for this lattice is the same as that shown in Figure 1.5, except that there is a second atom within the cell. The reciprocal lattice is thus the same as that for the f.c.c. structure, and hence the Brillouin zone has the same shape as the f.c.c. zone shown in Figure 8.14(c). A plot of ϵ versus k along the (100) direction in this Brillouin zone for electrons in the conduction band of silicon is shown in Figure 9.15(a). The smallest value of ϵ is reached at a value of k which is about

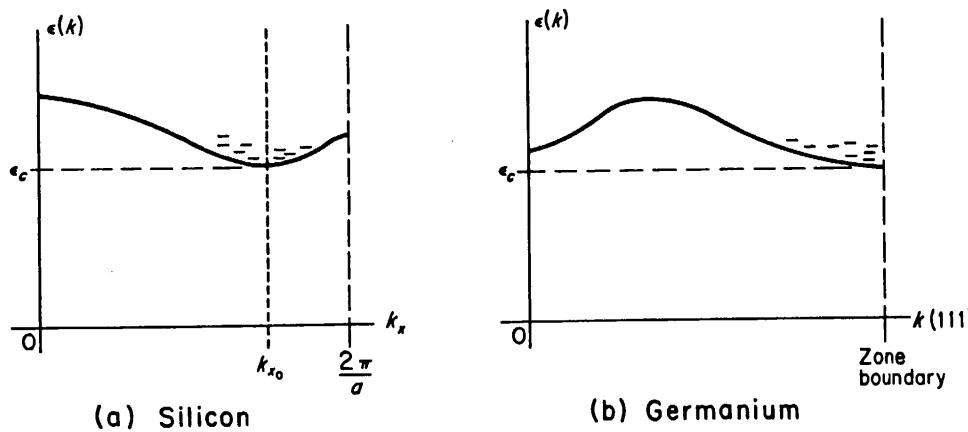
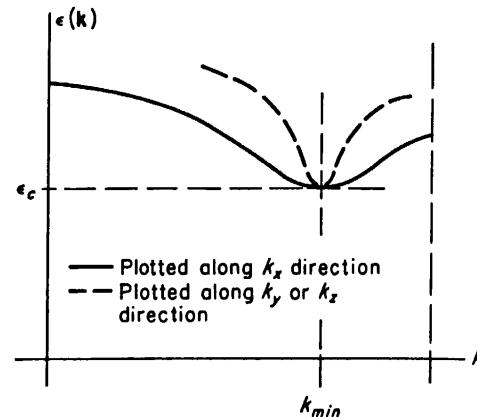


FIGURE 9.15. The ϵ versus k relation for (a) the conduction band of silicon, plotted along the k_x -direction, (b) the conduction band of germanium, plotted along a (111) direction in \mathbf{k} -space.

$0.8(2\pi/a)$. This is the absolute minimum value of ϵ within the Brillouin zone; curves of ϵ versus k along other paths may exhibit minimum values of ϵ , but none so low as the value ϵ_c in the figure. For energies greater than ϵ_c , one may construct constant energy surfaces about the minimum energy point. It is clear that the energy versus crystal momentum curve shown in Figure 9.15(a) is roughly parabolic about the minimum point, corresponding to free electron behavior with an appropriate effective mass which we shall call m_{\parallel}^* . If one were to plot the ϵ versus k curve along some path passing through the minimum point, but perpendicular to the k_x axis (for example a line parallel to the k_y axis) one would also obtain a curve which would be parabolic about the minimum point, but there is now no aspect of crystal symmetry which would require that the curvature in this direction be the same as it is along the k_x direction. (This situation is illustrated in Figure 9.16.) As a result the effective mass related to changes in the momentum component along this direction has a value which is

different from the value discussed previously, and which we shall denote as m_{\perp}^* . It turns out, as we shall soon see, that requirements of crystal symmetry demand that the effective mass have the same value m_{\perp}^* for any direction in k-space normal to the k_x -direction for this example.

FIGURE 9.16. The ϵ versus k relation for the conduction band of silicon plotted along the k_x -direction (solid curve) and along a line normal to the k_x -axis which passes through the point k_{\min} (dashed curve). The different curvatures at the minimum point reflect the anisotropy of the effective mass.



The energy in excess of ϵ_c represents the kinetic energy of a free electron in the conduction band, whereby

$$\epsilon - \epsilon_c = \frac{(p_x - p_{x0})^2}{2m_{\parallel}^*} + \frac{(p_y^2 + p_z^2)}{2m_{\perp}^*} \quad (9.9-15)$$

where $p_{x0} = \hbar k_{x0}$ represents the value of p_x at the bottom of the band, as shown in Figure 9.15(a). According to this the surface of constant energy for an electron of energy ϵ slightly greater than ϵ_c is seen to be an *ellipsoid of revolution* whose center is at $k_x = k_{x0}$, $k_y = k_z = 0$. Since, however, the six $\langle 100 \rangle$ directions of a cubic lattice are crystallographically equivalent, the ϵ versus k curve along any of these six directions must be the same. There must then be six equivalent energy minima along the six $\langle 100 \rangle$ directions, and associated with each of these minima there must be a family of ellipsoidal constant energy surfaces similar to those described by (9.9-15). This situation is illustrated in Figure 9.17(a). It is now clear why the ellipsoids have to be ellipsoids of revolution; as we saw in the previous chapter, the constant energy surfaces must have all the symmetry properties of the Brillouin zone, which in turn has all the symmetry properties of a cubic crystal. If the ellipsoids were not ellipsoids of revolution, the family of constant energy surfaces would no longer be invariant under all the operations for which a cubic crystal remains invariant.

The situation for electrons in the conduction band of germanium is somewhat similar, except that there are a set of eight equivalent minimum energy points which lie along the $\langle 111 \rangle$ directions in k-space at the intersection of those directions with the surface of the Brillouin zone. These minima thus lie at the centers of the hexagonal faces of the Brillouin zone. The constant energy surfaces are again ellipsoidal in form, but since the energy minima are at the zone boundary, and since there is a rather large forbidden energy region which must be surmounted before electrons can be excited outside the Brillouin zone in these directions, there are eight half-ellipsoids which extend into the interior of the Brillouin zone along the $\langle 111 \rangle$ directions, as shown in Figure 9.17(b). For many purposes these may be considered to be equivalent to four full ellipsoids. In both cases there is an effective mass *tensor* associated with each

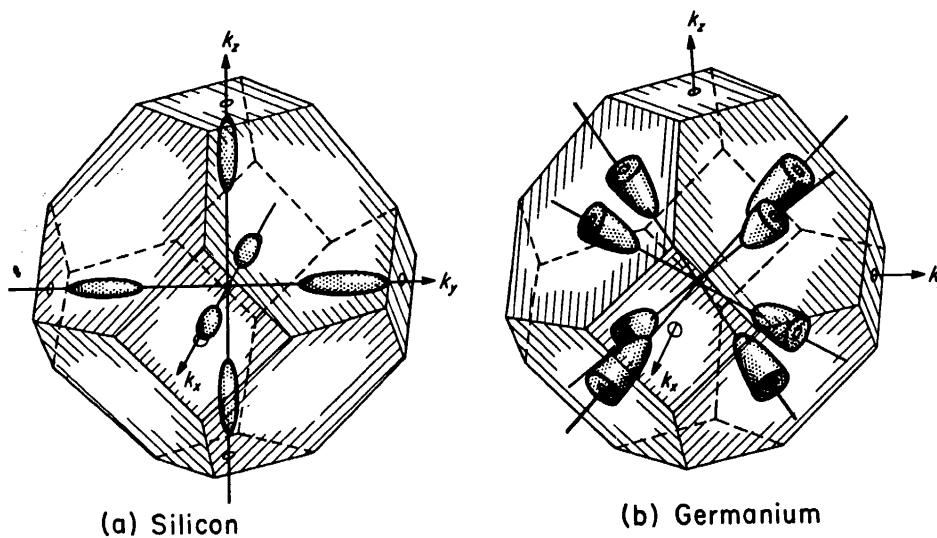


FIGURE 9.17. Ellipsoidal constant energy surfaces in \mathbf{k} -space for (a) silicon and (b) germanium. In silicon the major axes of the ellipsoids are along $\{100\}$ directions, while in germanium, since the energy minimum is on the zone boundary, the constant energy surfaces form eight half-ellipsoids whose major axes lie along $\{111\}$ directions.

ellipsoid whose elements are described by (8.8-7). In calculating the transport properties and other physical characteristics of crystals in which the energy surfaces are ellipsoidal, one usually calculates the effect due to one ellipsoid (which is generally anisotropic) and then sums over all ellipsoids. In cubic crystals, the transport properties are found to be isotropic after the summation is carried out, despite the fact that anisotropies are associated with individual ellipsoids. In both germanium and silicon the longitudinal effect mass m_{\parallel}^* is much larger than the transverse mass m_{\perp}^* , the ellipsoids thus being quite long and thin. The actual values, as determined by cyclotron resonance experiments,⁵ are given in Table 9.2.

TABLE 9.2.
Effective Masses of Electrons in Germanium
and Silicon

	m_{\parallel}^*	m_{\perp}^*	"mass ratio" $m_{\parallel}^*/m_{\perp}^*$
Germanium	$1.64 m_0$	$0.0819 m_0$	20.0
Silicon	$0.98 m_0$	$0.19 m_0$	5.2

Let us now investigate the cyclotron resonance effect in materials where the energy surfaces are ellipsoidal. To begin with we shall calculate the resonance frequency associated with a *single* ellipsoid, which we shall assume to be symmetric about the k_z -axis and to be centered on the point k_{z0} , as shown in Figure 9.18. The magnetic induction \mathbf{B}_0 makes an angle θ with the major axis of the ellipsoid, and for convenience we shall assume that the oscillating electric field vector has only a y -component. The choice of this coordinate system is made only for the purposes of the present

calculation, and we shall see when we are finished that the only coordinate of physical importance is the angle θ between the field and the longitudinal axis of the ellipsoid.

Under these circumstances, the equation of motion (for a particle of positive charge e) may be written in terms of the tensor (8.8-6) which represents the reciprocal of the effective mass as

$$\frac{d\mathbf{v}}{dt} = \left(\frac{1}{m^*} \right) \cdot \mathbf{F} = \left(\frac{1}{m^*} \right) \cdot (e\mathbf{E} + \frac{e}{c} \mathbf{v} \times \mathbf{B}). \quad (9.9-16)$$

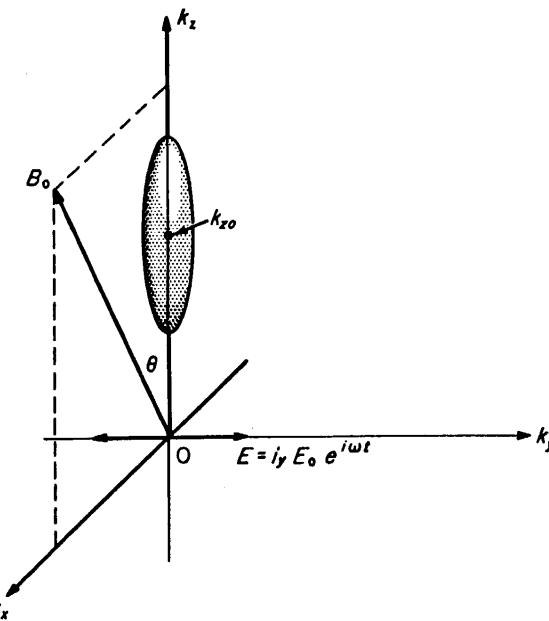


FIGURE 9.18. Vector geometry used for the calculation of Section 9.9.

The relation between ϵ and \mathbf{k} for the case of the ellipsoidal energy surface shown in Figure 9.18 must, according to (9.9-15) be

$$\epsilon - \epsilon_c = \hbar^2 \left[\frac{k_x^2 + k_y^2}{2m_{\perp}^*} + \frac{(k_z - k_{z0})^2}{2m_{\parallel}^*} \right]. \quad (9.9-17)$$

The tensor components $(1/m^*)_{\alpha\beta}$ may now be found according to (8.8-7). In this case, $\partial^2\epsilon/\partial k_{\alpha}\partial k_{\beta} = 0$ for $\alpha \neq \beta$, so that all the off-diagonal elements of the tensor are zero. The diagonal elements are easily evaluated, the result being

$$(1/m^*)_{xx} = (1/m^*)_{yy} = 1/m_{\perp}^* \quad \text{and} \quad (1/m^*)_{zz} = 1/m_{\parallel}^*. \quad (9.9-18)$$

The effective mass tensor then has the form

$$\left(\frac{1}{m^*} \right) = \begin{bmatrix} \frac{1}{m_{\perp}^*} & 0 & 0 \\ 0 & \frac{1}{m_{\perp}^*} & 0 \\ 0 & 0 & \frac{1}{m_{\parallel}^*} \end{bmatrix} \quad (9.9-19)$$

and

$$\left(\frac{1}{m^*}\right) \cdot \mathbf{F} = \begin{bmatrix} \frac{1}{m_\perp^*} & 0 & 0 \\ 0 & \frac{1}{m_\perp^*} & 0 \\ 0 & 0 & \frac{1}{m_\parallel^*} \end{bmatrix} \cdot \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \begin{bmatrix} F_x/m_\perp^* \\ F_y/m_\perp^* \\ F_z/m_\parallel^* \end{bmatrix}. \quad (9.9-20)$$

Inserting the Lorentz force components for F_x , F_y , and F_z and equating to the components of dv/dt as directed by (9.9-16), the equation of motion can be written

$$m_\perp^* \frac{dv_x}{dt} = \frac{e}{c} (v_y B_z - v_z B_y) + e E_x$$

$$m_\perp^* \frac{dv_y}{dt} = \frac{e}{c} (v_z B_x - v_x B_z) + e E_y \quad (9.9-21)$$

$$m_\parallel^* \frac{dv_z}{dt} = \frac{e}{c} (v_x B_y - v_y B_x) + e E_z.$$

Noting that $B_x = B_0 \sin \theta$, $B_z = B_0 \cos \theta$, $B_y = 0$, $E_y = E_0 e^{i\omega t}$, $E_x = E_z = 0$, these equations may be stated as

$$dv_x/dt = \omega_\perp v_y \cos \theta$$

$$dv_y/dt = \omega_\perp v_z \sin \theta - \omega_\perp v_x \cos \theta + \frac{e E_0}{m_\perp^*} e^{i\omega t} \quad (9.9-22)$$

$$dv_z/dt = -\omega_\parallel v_y \sin \theta$$

where

$$\omega_\perp = \frac{eB_0}{m_\perp^* c} \quad \text{and} \quad \omega_\parallel = \frac{eB_0}{m_\parallel^* c}. \quad (9.9-23)$$

As before, one may assume oscillatory solutions of the form $x(t) = x_0 e^{i\omega t}$, $y(t) = y_0 e^{i\omega t}$, and $z(t) = z_0 e^{i\omega t}$, substitute these into the equations of motion (9.9-22) and solve for the amplitudes to obtain

$$\begin{aligned} x_0 &= \frac{e E_0}{m_\perp^*} \frac{i \omega_\perp \cos \theta}{\omega(\omega^2 - \omega_\perp^2 \cos^2 \theta - \omega_\perp \omega_\parallel \sin^2 \theta)} \\ y_0 &= -\frac{e E_0}{m_\perp^*} \frac{1}{\omega(\omega^2 - \omega_\perp^2 \cos^2 \theta - \omega_\perp \omega_\parallel \sin^2 \theta)} \quad (9.9-24) \\ z_0 &= -\frac{e E_0}{m_\perp^*} \frac{i \omega_\parallel \sin \theta}{\omega(\omega^2 - \omega_\perp^2 \cos^2 \theta - \omega_\perp \omega_\parallel \sin^2 \theta)} \end{aligned}$$

These amplitudes become very large when

$$\omega = \sqrt{\omega_{\perp}^2 \cos^2 \theta + \omega_{\parallel}^2 \sin^2 \theta}, \quad (9.9-25)$$

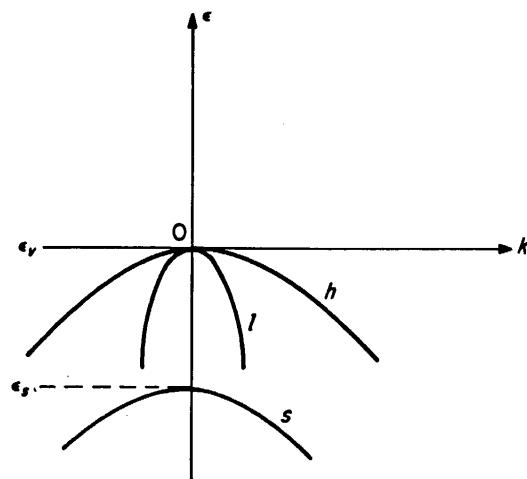
and this is accordingly the cyclotron resonance frequency for the particular ellipsoid which we are considering. The resonance frequency depends only upon the magnetic field, the effective masses and the angle θ between the field \mathbf{B}_0 and the major axis of the ellipsoid. For ellipsoidal energy surfaces, it is clear that the resonance frequency is a *strong function of the orientation of the sample relative to the magnetic field*. At the resonant frequency the particle orbits can be shown from (9.9-24) to be elliptical in form and to lie in a plane perpendicular to the field \mathbf{B}_0 . The above results have been derived for positive particles of charge e , but for electrons we need only replace e by $-e$ (in which case we must also replace ω_{\parallel} by $-\omega_{\parallel}$ and ω_{\perp} by $-\omega_{\perp}$). The resonant frequency is obviously unaltered by these substitutions; the only difference is that electrons traverse their orbits in the opposite sense from that found for positive particles.

In the case of silicon, if the field \mathbf{B}_0 is oriented along the z -direction, then $\theta = 0$ for the two ellipsoids whose major axes are along the k_z direction and $\theta = 90^\circ$ for the other four ellipsoids. There will then, according to (9.9-25), be two superimposed resonance peaks at $\omega = \omega_{\perp}$ from the first two ellipsoids and four superimposed resonances at $\omega = \sqrt{\omega_{\perp}\omega_{\parallel}}$ from the other four. Assuming that the number of electrons belonging to each ellipsoid is equal, which should be true in equilibrium, the second resonance, involving four ellipsoids should have twice the absorptive strength of the first, which involves only two. If the field \mathbf{B}_0 is otherwise oriented, there may be as many as three resonance peaks, corresponding to the three possible angles between \mathbf{B}_0 and the ellipsoid axes. This theory of cyclotron resonance with ellipsoidal energy surfaces serves to account for all the experimentally observed resonances which are strongly affected by sample orientation, and allows a determination of both m_{\perp}^* and m_{\parallel}^* from the experimentally determined cyclotron resonance frequencies, once a correct model for the orientation of the ellipsoids is proposed.

The cyclotron resonance data for holes in the valence bands of silicon and germanium do not exhibit the features characteristic of the effect with ellipsoidal energy surfaces. Although there are multiple resonance peaks for holes in both materials, the characteristic changes of resonance frequency with sample orientation which are observed for ellipsoidal energy surfaces are either absent or much less pronounced. The picture of the valence band which seems to fit these (and other) data best is that of two separate bands which are degenerate at the energy maximum at $k = 0$, but which have different effective masses. There is also a third "split-off" band with a maximum energy (at $k = 0$) somewhat below the top of the valence band. This state of affairs is shown in Figure 9.19. There are simultaneously present light and heavy holes (belonging to the l and h bands in the figure), and a different cyclotron resonance effect for each type can be observed. Since the energy surfaces associated with these bands are approximately spherical, the orientation dependence of the cyclotron resonance effect is much less pronounced than that which is associated with the ellipsoidal surfaces of the conduction band. Nevertheless, the energy surfaces are only roughly spherical; they are somewhat distorted from a truly spherical form, even for small values of k . This "warping" must be taken into account to explain all the details of the cyclotron resonance data, especially in the case of the heavy hole band, where the departure from sphericity is most pronounced.

The energy difference between the maximum point of the lower-lying "split-off" band and the maximum of the valence band ($\epsilon_v - \epsilon_s$ in Figure 9.19) amounts to about 0.28 eV in germanium and 0.035 eV in silicon. This energy difference, in the case of germanium, is sufficiently great that under ordinary conditions the population of holes in the lower band will be negligible. In silicon, however, except at low temperatures, a significant number of holes will be found in this band. The splitting of the bands is caused by the interaction of the electron spins of the valence electrons with

FIGURE 9.19. The three distinct branches of the valence bands of silicon and germanium.



the magnetic moment arising from their orbital motion. If it were not for this *spin-orbit coupling* the energy difference $\epsilon_v - \epsilon_s$ would be zero and all three bands would coalesce at the point $k = 0$. The effective masses associated with the three varieties of holes are given in Table 9.3.

TABLE 9.3.
Effective Masses of Holes in Germanium
and Silicon

	m_l^*	m_h^*	m_s^*
Germanium	$0.044 m_0$	$0.28 m_0$	$0.077 m_0$
Silicon	$0.16 m_0$	$0.49 m_0$	$0.245 m_0$

The existence of a hole as one particular species is in general limited to a single mean free time. For example, a hole may easily be scattered from the *l* band to the *h*-band or vice versa by any of the scattering processes in which it may normally be involved. For this reason it is not possible to detect the existence of the separate varieties of holes in any measurement (such as conductivity or carrier drift) which involves only the average behavior of the carrier over relatively long distances or times, within which many collisions may occur. The results of such a measurement will necessarily only depend upon the properties of the carrier *averaged* over a long time interval during which the identities of holes in the three bands may be frequently interchanged. In the cyclotron resonance experiment, where one measures a resonance which is excited in a time *short* compared to the mean free time ($\omega_0\tau \gg 1$), the presence

of the several varieties of holes is easily exhibited. The same remark can be made with reference to the electrons in the various equivalent energy minima of the conduction band; the normal scattering processes may readily transfer electrons from one ellipsoid to another (*intervalley* scattering), so that in a time long compared to the relaxation time, the identity of an electron as belonging to one particular ellipsoid is irretrievably lost.

9.10 DENSITY OF STATES, CONDUCTIVITY AND HALL EFFECT WITH COMPLEX ENERGY SURFACES

In a situation where there are two parabolic bands, degenerate at $k = 0$, with spherical energy surfaces, giving rise to two different species of carriers, the relative populations of the two species and the density of states factors can be easily calculated. We shall discuss the case of a valence band of this sort, in which there are present simultaneously light holes (effective mass m_l^*) and heavy holes (effective mass m_h^*). In this instance we may write the total hole concentration p_0 as the sum of the light hole concentration p_l and the heavy hole concentration p_h , whereby, from (9.3-15) and (9.3-16),

$$\begin{aligned} p_0 &= p_l + p_h = U_{vl} e^{-(\epsilon_f - \epsilon_v)/kT} + U_{vh} e^{-(\epsilon_f - \epsilon_v)/kT} \\ &= 2 \left(\frac{2\pi kT}{h^2} \right)^{3/2} \left(m_l^{*3/2} + m_h^{*3/2} \right) e^{-(\epsilon_f - \epsilon_v)/kT} \end{aligned} \quad (9.10-1)$$

where U_{vl} and U_{vh} are factors of the form of (9.3-16) for light and heavy holes. This can be written in the form

$$p_0 = 2 \left(\frac{2\pi m_{ds}^* kT}{h^2} \right)^{3/2} e^{-(\epsilon_f - \epsilon_v)/kT}, \quad (9.10-2)$$

as if there were a single species of carrier, if we define the *density of states equivalent effective mass* m_{ds}^* as

$$m_{ds}^{*3/2} = m_l^{*3/2} + m_h^{*3/2}. \quad (9.10-3)$$

This is an approximation to the actual state of affairs in the valence band of silicon and germanium, although it is a rather rough one, since the effect of the "split-off" band is entirely neglected and since the actual energy surfaces are by no means perfectly spherical. The conductivity follows from (9.10-2) directly, by the use of (9.7-8).

For a set of ellipsoidal energy surfaces such as those discussed in connection with the conduction band of germanium or silicon, one may calculate the density of states factor $g(\epsilon)$ by an extension of the procedure developed in Section 5.2. The equation (9.9-17) for one of the ellipsoids may be reduced by the transformation

$$p'_x = p_x \sqrt{m_0/m_{\perp}^*} = \hbar k_x \sqrt{m_0/m_{\perp}^*}$$

$$p'_y = p_y \sqrt{m_0/m_{\perp}^*} = \hbar k_y \sqrt{m_0/m_{\perp}^*}$$

$$p'_z = (p_z - p_{z0}) \sqrt{m_0/m_{\parallel}^*} = \hbar(k_z - k_{z0}) \sqrt{m_0/m_{\parallel}^*} \quad (9.10-4)$$

to that of a sphere in \mathbf{p}' -space.

$$\epsilon - \epsilon_c = \frac{p'^2_x + p'^2_y + p'^2_z}{2m_0} = \frac{p'^2}{2m_0}. \quad (9.10-5)$$

The volume of \mathbf{p}' -space in the spherical shell bounded by the radii p' and $p' + dp'$ for a single ellipsoid may now be calculated exactly as in Section 5.2 for spherical energy surfaces, but since there is more than one ellipsoidal constant energy surface, the total volume of \mathbf{p}' -space corresponding to energies in the range ϵ to $\epsilon + d\epsilon$ will be obtained by multiplying the volume found for a single surface by the number of equivalent ellipsoids. If this number is v , then the corresponding volume $dV_{p'}$ of \mathbf{p}' -space is

$$dV_{p'} = 4\sqrt{2\pi v m_0^{3/2}} \sqrt{\epsilon - \epsilon_c} d\epsilon. \quad (9.10-6)$$

But, from (9.10-4) the volume elements $dV_p = dp_x dp_y dp_z$ and $dV_{p'} = dp'_x dp'_y dp'_z$ are related by

$$dV_{p'} = \frac{m_0^{3/2}}{(m_{\perp}^{*2} m_{\parallel}^*)^{1/2}} dV_p, \quad (9.10-7)$$

whereby the volume dV_p of \mathbf{p} -space within the energy range ϵ to $\epsilon + d\epsilon$ becomes

$$dV_p = 4\sqrt{2\pi v (m_{\perp}^{*2} m_{\parallel}^*)^{1/2}} \sqrt{\epsilon - \epsilon_c} d\epsilon. \quad (9.10-8)$$

Dividing this by the volume $h^3/2$ assigned to a single quantum state gives the number of quantum states in this energy range which, by definition, is $g(\epsilon) d\epsilon$, whereby

$$g(\epsilon) d\epsilon = \frac{8\sqrt{2\pi v (m_{\perp}^{*2} m_{\parallel}^*)^{1/2}}}{h^3} \sqrt{\epsilon - \epsilon_c} d\epsilon. \quad (9.10-9)$$

If, once more, we define a *density of states equivalent effective mass* m_{ds}^* as

$$m_{ds}^{*3/2} = (m_{\perp}^{*2} m_{\parallel}^*)^{1/2}, \quad (9.10-10)$$

Equation (9.10-9) can be written (except for the factor v) in the same form (5.2-22) as the usual density of states expression for the case where the energy surfaces are spherical. By a straightforward application of the procedures by which (9.3-6) and (9.3-7) were obtained, it is easily seen that we must obtain

$$n_0 = 2v \left(\frac{2\pi m_{ds}^* k T}{h^2} \right)^{3/2} e^{-(\epsilon_c - \epsilon_f)/kT} \quad (9.10-11)$$

with m_{ds}^* as given by (9.10-10).

In calculating the conductivity of a semiconductor in which the constant energy surfaces are ellipsoidal, we must first calculate the current or conductivity contribution from a single ellipsoid and then sum over all the ellipsoids to obtain the total conductivity. Since the effective mass associated with each individual ellipsoid is a tensor quantity, so also is the conductivity. In general, the equations relating the current density and electric field may be written as tensor relations of the form

$$\mathbf{I} = \underline{\sigma} \cdot \mathbf{E} \quad (9.10-12)$$

and

$$\mathbf{E} = \underline{\rho} \cdot \mathbf{I} \quad (9.10-13)$$

where $\underline{\sigma}$ is the *conductivity tensor* and $\underline{\rho}$ is a resistivity tensor. Since, clearly, from (9.10-12) and (9.10-13),

$$\underline{\rho} \cdot \mathbf{I} = \underline{\rho} \cdot \underline{\sigma} \cdot \mathbf{E} = \mathbf{E}, \quad (9.10-14)$$

we must have

$$\underline{\rho} \cdot \underline{\sigma} = \underline{1}, \quad (9.10-15)$$

where $\underline{1}$ is the unit tensor, whose elements are $\delta_{\alpha\beta}$. The resistivity tensor is thus the inverse of the conductivity tensor, that is,

$$\underline{\rho} = \underline{\sigma}^{-1}. \quad (9.10-16)$$

The elements of the conductivity tensor $\underline{\sigma}^{(i)}$ related to the i th energy ellipsoid may in analogy with (7.3-15) be written

$$\sigma_{\alpha\beta}^{(i)} = n^{(i)} e^2 \bar{\tau} \left(\frac{1}{m^*} \right)_{\alpha\beta} \quad (9.10-17)$$

where $n^{(i)}$ is the density of charge carriers associated with that ellipsoid. The total conductivity may then be expressed as

$$\underline{\sigma} = \sum_i \underline{\sigma}^{(i)}, \quad (9.10-18)$$

the sum being taken over all ellipsoids. Using (8.8-7) to express the tensor components, (9.10-17) can be written as

$$\underline{\sigma}^{(i)} = \frac{n^{(i)} e^2 \bar{\tau}}{\hbar^2} \begin{bmatrix} \frac{\partial^2 \epsilon^{(i)}}{\partial k_x^2} & \frac{\partial^2 \epsilon^{(i)}}{\partial k_x \partial k_y} & \frac{\partial^2 \epsilon^{(i)}}{\partial k_x \partial k_z} \\ \frac{\partial^2 \epsilon^{(i)}}{\partial k_y \partial k_x} & \frac{\partial^2 \epsilon^{(i)}}{\partial k_y^2} & \frac{\partial^2 \epsilon^{(i)}}{\partial k_y \partial k_z} \\ \frac{\partial^2 \epsilon^{(i)}}{\partial k_z \partial k_x} & \frac{\partial^2 \epsilon^{(i)}}{\partial k_z \partial k_y} & \frac{\partial^2 \epsilon^{(i)}}{\partial k_z^2} \end{bmatrix} \quad (9.10-19)$$

where $\epsilon^{(i)}(\mathbf{k})$ is the energy expressed as a function of \mathbf{k} for the i th ellipsoid. This treatment assumes, of course, that the relaxation time τ is the same for all possible directions in \mathbf{k} -space, which may not be true in the most general case.

In the case of silicon, the major axes of the ellipsoids lie along the coordinate axes of \mathbf{k} -space; for the two ellipsoids whose major axes lie along the k_z -axis, in the [001] and [00̄1] directions the function $\epsilon^{(001)}(\mathbf{k})$ is given by (9.9-17), leading to tensor components (9.9-18) for the effective mass tensor. The conductivity tensor for these ellipsoids then becomes

$$\underline{\sigma}^{(001)} = \underline{\sigma}^{(00\bar{1})} = n^{(001)} e^2 \bar{\tau} \begin{bmatrix} \frac{1}{m_{\perp}^*} & 0 & 0 \\ 0 & \frac{1}{m_{\perp}^*} & 0 \\ 0 & 0 & \frac{1}{m_{\parallel}^*} \end{bmatrix} \quad (9.10-20)$$

The equations for the other ellipsoids may be written and the tensor components obtained in the same way. For the two ellipsoids whose major axes lie along the k_x -axis, then

$$\underline{\sigma}^{(100)} = \underline{\sigma}^{(\bar{1}00)} = n^{(100)} e^2 \bar{\tau} \begin{bmatrix} \frac{1}{m_{\parallel}^*} & 0 & 0 \\ 0 & \frac{1}{m_{\perp}^*} & 0 \\ 0 & 0 & \frac{1}{m_{\perp}^*} \end{bmatrix}, \quad (9.10-21)$$

while for the two ellipsoids whose major axes lie along the k_y -axis,

$$\underline{\sigma}^{(010)} = \underline{\sigma}^{(0\bar{1}0)} = n^{(010)} e^2 \bar{\tau} \begin{bmatrix} \frac{1}{m_{\perp}^*} & 0 & 0 \\ 0 & \frac{1}{m_{\parallel}^*} & 0 \\ 0 & 0 & \frac{1}{m_{\perp}^*} \end{bmatrix} \quad (9.10-22)$$

Since all six ellipsoids are equivalent energy minima, $n^{(100)} = n^{(\bar{1}00)} = n^{(010)} = n^{(0\bar{1}0)} = n^{(001)} = n^{(00\bar{1})} = n_0/6$. Performing the summation indicated in (9.10-18) and utilizing this fact, one may obtain

$$\begin{aligned} \underline{\sigma}_0 &= n_0 e^2 \bar{\tau} \begin{bmatrix} \frac{1}{3} \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right) & 0 & 0 \\ 0 & \frac{1}{3} \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right) & 0 \\ 0 & 0 & \frac{1}{3} \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right) \end{bmatrix} \\ &= n_0 e^2 \bar{\tau} \frac{1}{3} \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right) \underline{\mathbf{1}} \end{aligned} \quad (9.10-23)$$

for the total conductivity. Since $\underline{\sigma}$ is a scalar multiple of the unit tensor $\underline{1}$, the *total* conductivity is *isotropic* (albeit the conductivity associated with a single ellipsoid is not). The scalar magnitude of the conductivity can be represented as

$$\sigma_0 = \frac{n_0 e^2 \bar{\tau}}{m_c^*} \quad (9.10-24)$$

where m_c^* is a *conductivity effective mass* defined by

$$\frac{1}{m_c^*} = \frac{1}{3} \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right). \quad (9.10-25)$$

It can be shown that exactly the same result will be obtained for *any* set of ellipsoidal energy surfaces whatever, so long as the configuration is invariant under all of the symmetry operations under which a cubic crystal is invariant. In fact, it can be proved that the conductivity of all cubic crystals must be isotropic, whatever specific form the surfaces of constant energy take. In particular, Equations (9.10-23) through (9.10-25) are valid *also* for the conduction electrons in germanium, where the constant energy ellipsoids lie along the $\langle 111 \rangle$ directions. It is instructive to work out the conductivity tensor for germanium; this problem is assigned as an exercise for the reader.

The resistivity tensor $\underline{\rho}$ is a matrix of components inverse to the conductivity matrix. The elements of the matrix which is the inverse of a matrix \underline{A} with elements $a_{\alpha\beta}$ are given by the formula

$$a_{\alpha\beta}^{-1} = \frac{\text{cof}(a_{\beta\alpha})}{\Delta(a)}, \quad (9.10-26)$$

where $\text{cof}(a_{\beta\alpha})$ is the cofactor¹⁰ of the element $a_{\beta\alpha}$ and $\Delta(a)$ is the determinant of the coefficients of \underline{A} . The elements of $\underline{\rho}$ are thus readily seen to be

$$\underline{\rho} = \underline{\sigma}^{-1} = \frac{1}{\Delta(\sigma)} \begin{bmatrix} \sigma_{yy}\sigma_{zz} - \sigma_{yz}^2 & \sigma_{xz}\sigma_{yz} - \sigma_{xy}\sigma_{zz} & \sigma_{xy}\sigma_{yz} - \sigma_{xz}\sigma_{yy} \\ \sigma_{yz}\sigma_{xz} - \sigma_{xy}\sigma_{zz} & \sigma_{xx}\sigma_{zz} - \sigma_{xz}^2 & \sigma_{xz}\sigma_{xy} - \sigma_{xx}\sigma_{yz} \\ \sigma_{xy}\sigma_{yz} - \sigma_{xz}\sigma_{yy} & \sigma_{xz}\sigma_{xy} - \sigma_{xx}\sigma_{yz} & \sigma_{xx}\sigma_{yy} - \sigma_{xy}^2 \end{bmatrix} \quad (9.10-27)$$

where

$$\begin{aligned} \Delta(\sigma) = & \sigma_{xx}(\sigma_{yy}\sigma_{zz} - \sigma_{yz}^2) - \sigma_{xy}(\sigma_{xy}\sigma_{zz} - \sigma_{yz}\sigma_{xz}) \\ & + \sigma_{xz}(\sigma_{xy}\sigma_{yz} - \sigma_{yy}\sigma_{xz}). \end{aligned} \quad (9.10-28)$$

It is readily seen that if $\underline{\sigma}$ has the form (9.10-23), the resistivity tensor is isotropic and can be represented by a scalar quantity whose magnitude is simply the reciprocal of the conductivity (9.10-24).

We shall illustrate the computation of the Hall effect in a material with ellipsoidal energy surfaces by considering in detail the Hall effect in a material where the major axes of the ellipsoids extend along the k_x -, k_y - and k_z -axes (as, for example, in silicon),

¹⁰ See, for example, L. A. Pipes, *Applied Mathematics for Engineers and Physicists*. New York: McGraw-Hill (1946), p. 71.

in the special case where the magnetic field is in the z -direction and the current is flowing in the x -direction. We shall also assume that only one species of carrier is present. Under these conditions the inverse effective mass tensors for the various ellipsoids have the form shown in Equations (9.10-20, 21, and 22), and in particular, the tensor equation of motion for carriers of charge e in the (100) and ($\bar{1}00$) ellipsoids can be written, in analogy with (9.8-4), as

$$\left(\frac{1}{m^*}\right) \cdot \underline{\underline{F}} = \begin{bmatrix} \frac{1}{m_{||}^*} & 0 & 0 \\ 0 & \frac{1}{m_{\perp}^*} & 0 \\ 0 & 0 & \frac{1}{m_{\perp}^*} \end{bmatrix} \begin{bmatrix} eE_x + \frac{e}{c}(v_yB_z - v_zB_y) \\ eE_y + \frac{e}{c}(v_zB_x - v_xB_z) \\ eE_z + \frac{e}{c}(v_xB_y - v_yB_x) \end{bmatrix} = \begin{bmatrix} dv_x/dt \\ dv_y/dt \\ dv_z/dt \end{bmatrix}. \quad (9.10-29)$$

Performing the indicated matrix multiplication, writing out the component equations of motion, and noting that $B_x = B_y = 0$, while $B_z = B_0$, we may obtain

$$\begin{aligned} \frac{dv_x}{dt} &= \frac{eE_x}{m_{||}^*} + \omega_{||}v_y \\ \frac{dv_y}{dt} &= \frac{eE_y}{m_{\perp}^*} - \omega_{\perp}v_x \\ \frac{dv_z}{dt} &= \frac{eE_z}{m_{\perp}^*}, \end{aligned} \quad (9.10-30)$$

where $\omega_{||}$ and ω_{\perp} are as given by (9.9-23).

If the first of these equations is multiplied by $\sqrt{\omega_{\perp}}$ and added to $i\sqrt{\omega_{||}}$ times the second, the resulting equation can be expressed in the form

$$\frac{dZ}{dt} + i\sqrt{\omega_{||}\omega_{\perp}}Z = e\left(\frac{\sqrt{\omega_{\perp}}}{m_{||}^*}E_x + \frac{i\sqrt{\omega_{||}}}{m_{\perp}^*}E_y\right) \quad (9.10-31)$$

where Z is the complex quantity

$$Z = v_x\sqrt{\omega_{\perp}} + iv_y\sqrt{\omega_{||}}. \quad (9.10-32)$$

Equation (9.10-31) is clearly seen to have exactly the same form as (9.8-8). All the steps involving the solution of this equation and the averaging over path lengths and velocity distribution, leading to Equations (9.8-16) and (9.8-17) may now be repeated, starting with (9.10-31). The physical justification of each step is the same as before. In this manner we may show that

$$\begin{aligned} \bar{v}_x^{(100)} &= \bar{v}_x^{(\bar{1}00)} = \frac{1}{\sqrt{\omega_{\perp}}} \operatorname{Re}(\bar{Z}) \\ &= e \left[\frac{\bar{\tau}E_x}{m_{||}^*} + \frac{\omega_{||}}{m_{\perp}^*} \left(\frac{\tau^2}{1 + \omega_{||}\omega_{\perp}\tau^2} \right) E_y - \frac{\omega_{||}\omega_{\perp}}{m_{||}^*} E_x \left(\frac{\tau^3}{1 + \omega_{||}\omega_{\perp}\tau^2} \right) \right] \end{aligned} \quad (9.10-33)$$

while $\bar{v}_y^{(100)} = \bar{v}_y^{(T00)} = \frac{1}{\sqrt{\omega_{\parallel}}} \text{Im}(\bar{Z})$

$$= e \left[\frac{1}{m_{\perp}^*} \left(\frac{\tau}{1 + \omega_{\parallel} \omega_{\perp} \tau^2} \right) E_y - \frac{\omega_{\perp}}{m_{\parallel}^*} \left(\frac{\tau^2}{1 + \omega_{\parallel} \omega_{\perp} \tau^2} \right) E_x \right]. \quad (9.10-34)$$

If we confine our results to the small field Hall coefficient, we may assume $\omega_{\parallel} \omega_{\perp} \tau^2 \ll 1$ and neglect the $\omega_{\parallel} \omega_{\perp} \tau^2$ terms in the denominators in Equations (9.10-33) and (9.10-34). Now the current densities may be obtained from the velocities as before, whence, for the i th ellipsoid,

$$I_x^{(i)} = p_0^{(i)} e \bar{v}_x^{(i)} \quad \text{and} \quad I_y^{(i)} = p_0^{(i)} e \bar{v}_y^{(i)}, \quad (9.10-35)$$

while, summing over all ellipsoids, the total current density components are

$$I_x = \sum_i p_0^{(i)} e \bar{v}_x^{(i)} \quad \text{and} \quad I_y = \sum_i p_0^{(i)} e \bar{v}_y^{(i)}. \quad (9.10-36)$$

In the steady state, there can be no current in the y -direction, so that from (9.10-36) we must have

$$\sum_i p_0^{(i)} e \bar{v}_y^{(i)} = 0. \quad (9.10-37)$$

As mentioned previously the concentrations $p_0^{(i)}$ for all six ellipsoids are the same and equal to $p_0/6$. The expression for $\bar{v}_y^{(100)}$ and $\bar{v}_y^{(T00)}$ is given above as (9.10-34). For the (010) and (01̄0) ellipsoids, it is easily established by using the appropriate mass tensor in (9.10-29) that the equations of motion for the x - and y -components of velocity are the same as those given in (9.10-30) except that m_{\parallel}^* and m_{\perp}^* , and ω_{\parallel} and ω_{\perp} are interchanged. The expression for $\bar{v}_y^{(010)}$ and $\bar{v}_y^{(01̄0)}$ is then just (9.10-34) with these quantities interchanged. In the case of the (001) and (001̄) ellipsoids, it is clear, following the same procedure, that the equations of motion for the x - and y -components of velocity are the same as those in (9.10-30) except that m_{\parallel}^* and ω_{\parallel} do not appear at all; they are replaced by m_{\perp}^* and ω_{\perp} in all instances. The expression for $\bar{v}_y^{(001)}$ and $\bar{v}_y^{(001̄)}$ is then obtained from (9.10-34) by replacing m_{\parallel}^* and ω_{\parallel} by m_{\perp}^* and ω_{\perp} . Making all these substitutions, performing the summation indicated in (9.10-37) and solving for E_y , in terms of E_x , we may obtain

$$E_y = \frac{eB_0}{c} \left(\frac{\tau^2}{\bar{\tau}} \right) \frac{2m_{\perp}^* + m_{\parallel}^*}{m_{\perp}^{*2} m_{\parallel}^* \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right)} E_x. \quad (9.10-38)$$

Neglecting magnetoresistance effects, which are not important for low magnetic field strengths, we may express E_x in terms of I_x , according to (9.10-24), as

$$E_x = I_x / \sigma_0 = \frac{3I_x}{p_0 e^2 \bar{\tau} \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right)}. \quad (9.10-39)$$

Substituting this into (9.10-38) we may write, finally

$$E_y = RB_0I_x \quad (9.10-40)$$

with

$$R = \frac{1}{p_0ec} \frac{\bar{\tau}^2}{(\bar{\tau})^2} \frac{3(2m_{\perp}^* + m_{\parallel}^*)}{m_{\perp}^{*2}m_{\parallel}^* \left(\frac{2}{m_{\perp}^*} + \frac{1}{m_{\parallel}^*} \right)^2}. \quad (9.10-41)$$

This is often written in a slightly different form; if one defines K as the ratio of the longitudinal to transverse mass, whereby

$$K = m_{\parallel}^*/m_{\perp}^*, \quad (9.10-42)$$

then (9.10-41) can be expressed as

$$R = \frac{1}{p_0ec} \frac{\bar{\tau}^2}{(\bar{\tau})^2} \frac{3K(K+2)}{(2K+1)^2}. \quad (9.10-43)$$

Although we assumed particularly simple directions for the sample current and for the magnetic field, had we taken the trouble to work out the result for arbitrary direction of current flow or field orientation, we should have found the small-field Hall coefficient (9.10-41) or (9.10-43) to be independent of the orientation of current and field direction with respect to the energy ellipsoids. The small field Hall coefficient may thus be shown to be *isotropic*, although this is not true if the magnetic field is so large that $\omega_{\parallel}\omega_{\perp}\tau^2$ is no longer small compared to unity. The above results were obtained also for a particularly simple set of energy ellipsoids, but they can be shown to hold for *any* set of ellipsoidal energy surfaces so long as the set has cubic symmetry. These results were, of course, worked out for charge carriers having a charge $+e$. For *n*-type semiconductors (and *n*-type Si and Ge are the most common examples of substances with ellipsoidal energy surfaces) one must replace e by $-e$ and p_0 by n_0 in (9.10-41) and (9.10-43). If Equation (9.10-43) is multiplied by σ_0 , and if the quantity $cR\sigma_0$ is, as before, defined as the "Hall mobility" μ_H , it is clear that the ratio of the Hall mobility to the true drift mobility μ is given by

$$\frac{\mu_H}{\mu} = \frac{\bar{\tau}^2}{(\bar{\tau})^2} \frac{3K(K+2)}{(2K+1)^2} \quad (9.10-44)$$

for a semiconductor with ellipsoidal constant energy surfaces.

A complete discussion of other galvanomagnetic effects and of magnetoresistance for semiconductors having ellipsoidal energy surfaces is somewhat beyond the scope of this work and will be omitted. It should be noted particularly, however, that the magnetoresistance is not isotropic under these conditions, and that measurements of magnetoresistance as a function of sample orientation and magnetic field direction can provide a rich source of experimental data against which any proposed model of the energy surfaces may be checked in detail.

9.11 SCATTERING MECHANISMS AND MOBILITY OF CHARGE CARRIERS

In semiconductors, as in metals, the most important scattering processes involve interactions of electrons (or holes) with lattice vibrations and with impurity atoms. In relatively pure crystals, or at relatively high temperatures, the former interaction is ordinarily predominant, while at low temperatures or in impure specimens, the latter may be more important. There is a broad range of intermediate conditions where both processes are significant.

The effect of any particular scattering interaction i may be evaluated by calculating the mean free time τ_i associated with this process. The overall free time τ may then be expressed by

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i} \quad (9.11-1)$$

as shown in Section 7.5. It is clear from this that the scattering process which leads to the shortest free time τ_i is the dominant one. Once the mean free time τ is known, the mobility may easily be expressed as $e\bar{\tau}/m^*$.

Unfortunately, the calculation of the free time associated with any of the scattering processes which are of importance in semiconductors is a lengthy and complex problem, the mathematical details of which are in some respects beyond the level which has been set for the present treatment. We must, therefore, limit ourselves to a rather crude and qualitative treatment of lattice scattering, although we shall attempt to give a fairly complete treatment of impurity scattering along the lines which were first taken by Conwell and Weisskopf.¹¹

At low temperatures the thermal energy available to excite optical-mode lattice vibrations is quite limited, and the most important lattice scattering process is the scattering of charge carriers by acoustical mode lattice vibrations. We shall consider only the scattering of carriers by longitudinal mode vibrations; the reason why the effect of transverse modes is negligible will become clear in the course of the discussion.

The passage of a longitudinal vibration through a crystal gives rise to alternate regions of compression and extension of the crystal lattice. When the crystal is compressed, the positions of the energy bands are altered in such a manner that the width of the forbidden gap is increased (see Figure 9.2); in the condition of extension the forbidden gap width decreases. This effect causes a local variation in the energy associated with the conduction band edge, as shown in Figure 9.20(a), which, to simplify the calculation, may be replaced by the stepwise variation shown in Figure 9.20(b) without introducing much error. There is a similar variation in the energy of the valence band edge. One may then calculate the reflection probability for an electron incident upon a single "step" of Figure 9.20(b).

To do this, an electron incident upon such a step of height $\delta\epsilon_c$ is represented by a free particle wave function of positive momentum of the form

$$\psi_1(x) = Ae^{ik_0x}. \quad (9.11-2)$$

¹¹ E. Conwell and V. F. Weisskopf, *Phys. Rev.*, 77, 388 (1950).

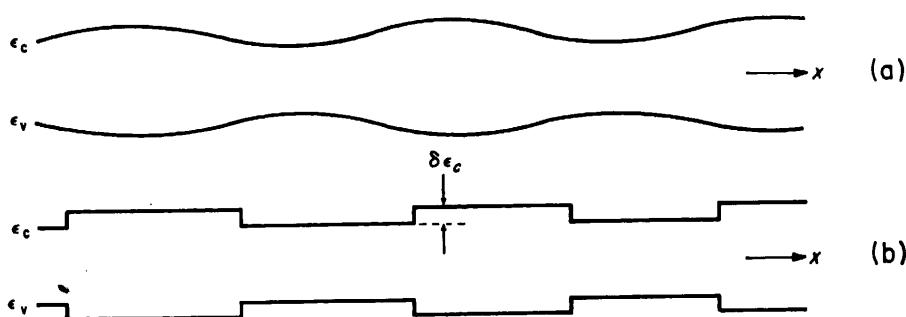


FIGURE 9.20. The sinusoidal variation of conduction and valence band energies in a semiconductor crystal caused by the compressional and extensional forces associated with a longitudinal thermal vibration. (b) A "square wave" variation which approximates the one shown in (a) and which simplifies the calculation of relaxation time.

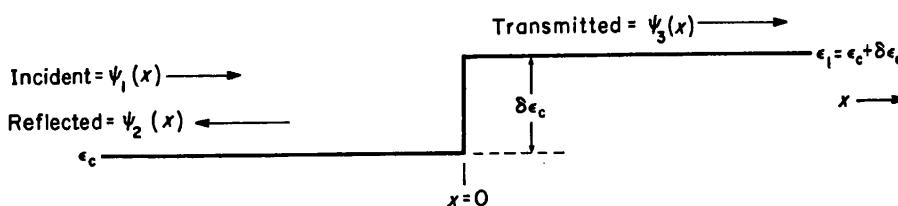


FIGURE 9.21. The "reflection" of an electron from a potential barrier of height $\delta\epsilon_c$.

The electron may be "reflected" at the step, or it may be transmitted, as illustrated in Figure 9.21. Accordingly, there may be a component of the wave function associated with a negative momentum in the region ($x < 0$), representing the reflected electron, which must have the form

$$\psi_2(x) = Be^{-ik_0x}. \quad (9.11-3)$$

In the region beyond the barrier ($x > 0$) the wave function representing an electron transmitted over the barrier is

$$\psi_3(x) = Ce^{ik_1x}. \quad (9.11-4)$$

These are all free particle wave functions satisfying Schrödinger's equation for constant potential. Since the momentum $\hbar k_0$ for the reflected component is assumed to be the same as the incident particle momentum, the collision interaction is assumed to be perfectly elastic. Actually, however, since the lattice wave as represented by the energy barrier is moving with the velocity of sound, there is a Doppler effect which renders the reflected momentum different from the incident momentum. Since the velocity of sound is small compared to the average thermal velocity of the particle the momentum shift is only a small fraction of the incident particle momentum and may safely be neglected in this calculation. From (4.9-8) it is clear that if the energy of the incident electron is ϵ_0 and that of the "transmitted" electron is ϵ_1 , then

$$\epsilon_0 = \hbar^2 k_0^2 / 2m_n^* \quad \text{and} \quad \epsilon_1 = \epsilon_0 - \delta\epsilon_c = \hbar^2 k_1^2 / 2m_n^* \quad (9.11-5)$$

whereby

$$-\delta\epsilon_c = \epsilon_1 - \epsilon_0 = \frac{\hbar^2}{2m_n^*} (k_1^2 - k_0^2). \quad (9.11-6)$$

Since the wave function and its derivative must be continuous at $x = 0$, we may write

$$\psi_1(0) + \psi_2(0) = \psi_3(0) \quad (9.11-7)$$

and

$$\psi'_1(0) + \psi'_2(0) = \psi'_3(0).$$

Substituting the values given by (9.11-2, 3 and 4) into these equations, it is easy to show that

$$\frac{B}{A} = \frac{k_0 - k_1}{k_0 + k_1} \quad \text{and} \quad \frac{C}{A} = \frac{2k_0}{k_0 + k_1}. \quad (9.11-8)$$

The transmission and reflection probabilities R and T are given by

$$R = \frac{\psi_2^* \psi_2}{\psi_1^* \psi_1} = \left(\frac{k_0 - k_1}{k_0 + k_1} \right)^2 \quad (9.11-9)$$

$$T = 1 - R = \frac{4k_0 k_1}{(k_0 + k_1)^2} \quad (9.11-10)$$

If the step height $\delta\epsilon_c$ is small, which we shall assume is the case, then, according to (9.11-6), $k_1 \approx k_0$, and (9.11-9) can, with the help of this approximation and (9.11-6), be written as

$$R \approx \left(\frac{m_n^* \delta\epsilon_c}{2\hbar^2 k_0^2} \right)^2. \quad (9.11-11)$$

The quantity $\delta\epsilon_c$ can be related to the compressional or extensional strain $\delta V/V_0$ to a first approximation by a linear relation of the form

$$-\delta\epsilon_c = \Xi_c \cdot \frac{\delta V}{V_0}, \quad (9.11-12)$$

where the *deformation potential constant* Ξ_c represents the shift of the conduction band edge per unit dilatational strain.

Since the shear distortions associated with transverse vibrational modes produce no first-order change in volume, no change in ϵ_c results from these vibrations, and their effect in scattering charge carriers is negligible. The passage of a longitudinal elastic wave through the crystal, however, is accompanied by coherent regions of compression or extension which are *of the order of $l/2$ in linear extent*, where l is the wavelength of the disturbance. If a volume element V_0 of this linear extent is subjected to a dilatational stress producing a maximum pressure δp and a volume change δV , it is easily seen that the stored strain energy is $-\frac{1}{2}\delta p \delta V$, and if the source of the strain energy is thermal, its magnitude must be proportional to kT ; in other words,

$$\delta\epsilon = -\frac{1}{2}\delta p \delta V = ckT \quad (9.11-13)$$

SEC. 9.11

UNIFORM ELECTRONIC SEMICONDUCTORS

311

where c is a constant.¹² But since the compressibility β is defined as

$$\beta = \frac{1}{V_0} \frac{\delta V}{\delta p}, \quad (9.11-14)$$

we may express δp in terms of β by this equation and substitute the result into (9.11-13) to obtain

$$\delta \epsilon = \frac{1}{2} \frac{(\delta V)^2}{\beta V_0} = ckT \quad (9.11-15)$$

whereby

$$\frac{(\delta V)^2}{V_0^2} = \frac{2c\beta kT}{V_0}. \quad (9.11-16)$$

Substituting this result into (9.11-12), and using the resulting expression for $\delta \epsilon_c^2$ in (9.11-11), R may be written as

$$R = \frac{m_n^{*2}}{2\hbar^4 k_0^4} \frac{c\beta kT}{V_0} \Xi_c^2. \quad (9.11-17)$$

Now, in a distance δx , the probability of scattering (reflection) is $\delta x/\lambda_n$ where λ_n is the mean free path. In going a distance $l/2$ (the linear dimension of the volume V_0) the probability of reflection is approximately the value given as R above. Then $l/2\lambda_n$ may be equated to R , and one may solve for λ_n to obtain

$$\lambda_n = \frac{\hbar^4 k_0^4 l V_0}{m_n^{*2} c \beta k T \Xi_c^2} = \frac{\hbar^4}{8 m_n^{*2} c \beta k T \Xi_c^2}. \quad (9.11-18)$$

The second expression follows from the first by noting that $V_0 = l^3/8$ and $k_0 = 2\pi/l$. Making the assumption that λ_n is independent of velocity, the approximate validity of which has been discussed in Section 7.5, λ_n and the mean free time are related by (7.3-18). Using this equation, and Equation (5.4-20), $\bar{\tau}_n$ can be expressed in the form

$$\bar{\tau}_n = \frac{\hbar^4}{3\sqrt{8\pi m_n^{*3/2} c \beta (kT)^{3/2} \Xi_c^2}}. \quad (9.11-19)$$

This calculation is clearly quite crude; in particular, no attempt has been made to evaluate c . It clearly illustrates, nevertheless, the principles involved and also exhibits the dependence of $\bar{\tau}_n$ and thus the mobility on temperature, effective mass and deformation potential constant. A full quantum mechanical treatment¹³ yields the result

$$\bar{\tau}_n = \frac{\sqrt{8\pi}}{3} \frac{\hbar^4 c_{ll}}{m_n^{*3/2} (kT)^{3/2} \Xi_c^2}, \quad (9.11-20)$$

¹² This constant has to do primarily with the fraction of all the possible normal modes of vibration which contribute to a dilatational strain of the type discussed here. It is clearly independent of temperature, effective mass and deformation potential constant.

¹³ See, for example, W. Shockley, *Electrons and Holes in Semiconductors*. New York: D. van Nostrand, Inc. (1950), p. 539.

where c_{ll} is the elastic constant for a longitudinal extension in the [110] direction. The corresponding mobility is

$$\mu_n = \frac{e\bar{\tau}}{m_n^*} = \frac{\sqrt{8\pi}}{3} \frac{e\hbar^4 c_{ll}}{m_n^{*5/2} (kT)^{3/2} \Xi_c^2}. \quad (9.11-21)$$

If the constant energy surfaces are ellipsoidal rather than spherical, it is clear [from (9.11-5)] that the effective mass m_n^* in (9.11-17, 19 and 20) should be replaced by the density of states effective mass (9.10-10). The additional factor of m_n^* which appears in (9.11-21) must naturally in this case be the conductivity effective mass (9.10-25), so that the factor $m_n^{*5/2}$ in the denominator of (9.11-21) will then become $m_{ds}^{*3/2} m_c^*$. The above calculations refer specifically to electrons in the conduction band, but the procedure for holes in the valence band is essentially the same, and the results are the same as those given above, with the electron masses and conduction band deformation potential constant replaced by their respective valence band analogues.

According to (9.11-20) and (9.11-21), the mobility should vary as $T^{-3/2}$ and as $m_n^{*-5/2}$ when acoustical mode lattice scattering is the dominant scattering interaction. If the ratio of the electron to hole mobility as expressed by (9.11-21) and its analogue for holes is calculated, one may easily show that

$$\frac{\mu_n}{\mu_p} = (m_n^*/m_p^*)^{-5/2} (\Xi_c/\Xi_v)^{-2}. \quad (9.11-22)$$

Since Ξ_c and Ξ_v are usually about the same, it is often quoted as a rough rule of thumb that the ratio of mobilities is the inverse ratio of the effective masses to the 5/2 power.

Unfortunately, the experimental results are only in rough agreement with the predictions of (9.11-21). The temperature variation of the mobility in the range of temperatures where lattice scattering predominates, in particular, is usually rather stronger than the $T^{-3/2}$ variation found by this analysis. Average values of the measured mobility at 300°K (where lattice scattering is dominant in all but very impure samples) and the measured temperature variation of mobility for holes and electrons in germanium and silicon are shown in Table 9.4. In materials with energy surfaces which are essentially spherical, such as *p*-type Si and Ge, the effect of scattering by *optical* mode lattice vibrations is important, especially at higher temperatures.¹⁴

TABLE 9.4.
Measured Mobility Data for Silicon and Germanium

	μ_n	Temp. var. of μ_n	μ_p	Temp. var. of μ_p
Si	1600	$T^{-2.5}$	500	$T^{-2.3}$
Ge	3800	$T^{-1.66}$	1900	$T^{-2.3}$

In cases where the energy surfaces are ellipsoidal, for example, *n*-type Ge and Si, effects arising from the scattering of electrons from one ellipsoid to another (inter-valley scattering), which have not been considered in the above calculation are also of

¹⁴ H. Ehrenreich and A. W. Overhauser, *Phys. Rev.*, **104**, 649 (1956).

importance. In this process when an electron is transferred from one ellipsoid to another, conservation of momentum requires the generation or absorption of a phonon whose energy is of the same order of that of the electron itself, and hence these scattering events cannot be regarded as even approximately elastic. A detailed consideration of this type of scattering by Herring¹⁵ has shown that the observed temperature dependence of mobility in both *n*-type Ge and *n*-type Si can be explained on the basis of intervalley scattering, although, of course, the optical mode contribution is no doubt significant also. In addition, the effect of transverse vibrational modes is much greater when the band structure is of the "many valley" form,¹⁶ and hence cannot be entirely neglected for *n*-type Ge and Si. For a more detailed discussion of these matters, the excellent articles by Herring¹⁵ and Blatt¹⁷ are recommended.

The calculation of the relaxation time for impurity scattering is based upon the theory of scattering of charged particles by the Coulomb potential of nuclei which was originally developed by Rutherford to explain the scattering of α -particles.^{18,19} In a semiconductor crystal which contains substitutional impurities, the Coulomb potential due to the charged donor and acceptor ions serve to deflect the paths of electrons and holes much as the potential of a heavy nucleus will deflect an α -particle in the Rutherford scattering experiment.

If the differential cross section $\sigma(\theta)$ is defined such that $\sigma(\theta) d\Omega$ is the differential area of the incident particle beam which is scattered through an angle θ into an element of solid angle $d\Omega$, then it may be shown by either classical or quantum mechanics¹⁹ that

$$\sigma(\theta) d\Omega = \left(\frac{Ze^2}{2\kappa m^* v_0^2} \right)^2 \frac{1}{\sin^4 \frac{\theta}{2}} d\Omega \quad (9.11-23)$$

and that

$$\tan \frac{\theta}{2} = \frac{Ze^2}{\kappa am^* v_0^2}. \quad (9.11-24)$$

The scattering geometry is shown in Figure 9.22 for a repulsive Coulomb interaction (hole and donor ion) and in Figure 9.23 for an attractive Coulomb interaction (electron and donor ion); the above equations are applicable to both cases. The quantity κ is the dielectric constant of the substance, v_0 is the initial (and final) velocity and a is the "impact parameter," or the perpendicular distance between the scattering center and the projection of the initial line of approach of the particle.

For a single incident particle, the number of collisions which it may make per unit time into a solid angle $d\Omega$ is readily seen to be $Nv_0\sigma(\theta) d\Omega$, where N is the number of scattering impurities per unit volume. Since energy is conserved in the scattering interaction, the magnitudes of the initial and final momenta of the scattered particle are the same; since the direction is changed by the scattering angle θ , however,

¹⁵ C. Herring, *Bell Syst. Tech. J.*, 34, 237 (1955).

¹⁶ C. S. Smith, *Phys. Rev.*, 94, 42 (1954).

¹⁷ F. J. Blatt, *Theory of Mobility of Electrons in Solids in Solid State Physics, Advances in Research and Applications*. New York: Academic Press (1957), Vol. 4, p. 199. See particularly p. 332 ff.

¹⁸ E. Rutherford, *Phil. Mag.*, 21, 669 (1911).

¹⁹ R. B. Leighton, *Principles of Modern Physics*. New York: McGraw-Hill (1959), p. 485 ff.

there is an accompanying change in the forward x -component of momentum of

$$\Delta p_x = p_0(1 - \cos \theta). \quad (9.11-25)$$

Now, from (9.11-23), it is clear that most of the collisions produce only a *small* angle of deflection. The relaxation time, however, is the time required, on the average, for the forward velocity of the particle to be reduced to zero, and this clearly is much

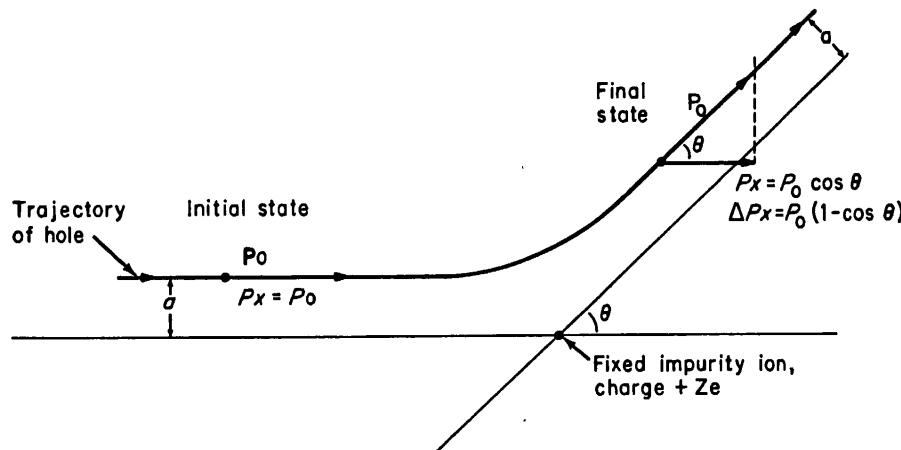


FIGURE 9.22. The trajectory of a hole which is scattered by Coulomb interaction with a fixed scattering center of charge $+Ze$.

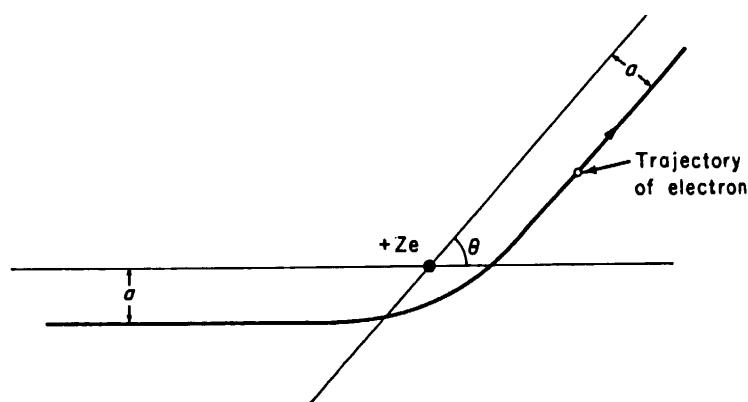


FIGURE 9.23. The trajectory of an electron scattered by a fixed scattering center of charge $+Ze$.

longer than the average time between such scattering events. As a matter of fact, from (9.11-25), it can be seen that the *effective* number of velocity destroying collisions is *less* than the actual total number of scatterings by a factor $(1 - \cos \theta)$. The equivalent number of velocity destroying collisions per unit time involving an angle of deflection θ into a solid angle $d\Omega$ is thus not $Nv_0\sigma(\theta) d\Omega$ but rather

$$dn_{\text{eff}} = Nv_0\sigma(\theta)(1 - \cos \theta) d\Omega = d(1/\tau). \quad (9.11-26)$$

This quantity is just the reciprocal of the relaxation time for particles of this specific

initial velocity undergoing collisions involving this particular scattering angle. The total relaxation time may be calculated by integrating over the solid angle and averaging over velocities.

Since the scattering is independent of the azimuthal angle ϕ about the polar axis, $d\Omega = 2\pi \sin \theta d\theta$, and integrating (9.11-26) over the solid angle, one may obtain

$$\begin{aligned} \frac{1}{\tau(v_0)} &= 2\pi N v_0 \int_{\theta_0}^{\pi} \sigma(\theta)(1 - \cos \theta) \sin \theta d\theta = 2\pi N v_0 \left(\frac{Ze^2}{2km^*v_0^2} \right)^2 \int_{\theta_0}^{\pi} \frac{(1 - \cos \theta) \sin \theta}{\sin^4 \frac{\theta}{2}} d\theta \\ &= -16\pi N v_0 \left(\frac{Ze^2}{2km^*v_0^2} \right)^2 \ln \sin \frac{\theta_0}{2} = 8\pi N v_0 \left(\frac{Ze^2}{2km^*v_0^2} \right)^2 \ln \left[1 + \left(\frac{km^*v_0^2}{2Ze^2 N^{1/3}} \right)^2 \right]. \end{aligned} \quad (9.11-27)$$

In performing the integration it is best to express all trigonometric functions in terms of the argument $\theta/2$. The upper limit of the integral must be $\theta = \pi$, which corresponds to a "direct hit" which reflects the incident particle back upon its original path. For such a collision, the impact parameter a is zero. The lower limit would be $\theta = 0$ for an isolated scattering center (which would give rise to an infinite integrated cross-section for scattering); but in the crystal the influence of an individual scattering center extends outwards only to a distance of the order of the mean spacing between the scattering ions. It is therefore reasonable to "cut off" the scattering from a particular ion when the impact parameter a equals $d/2$, where $d = N^{-1/3}$ is the mean distance between impurity atoms. The lower limit of the integral is then at

$$a = \frac{1}{2N^{1/3}} \quad (9.11-28)$$

which, according to (9.11-24), corresponds to

$$\theta = \theta_0 = 2 \tan^{-1} \left(\frac{2Ze^2 N^{1/3}}{km^*v_0^2} \right). \quad (9.11-29)$$

The above calculations permit one to express $\tau(v_0)$ in the form

$$\tau(v_0) = \frac{\kappa^2 m^{*2} v_0^3}{2\pi Z^2 e^4 N} \frac{1}{\ln \left[1 + \left(\frac{km^*v_0^2}{2Ze^2 N^{1/3}} \right)^2 \right]}. \quad (9.11-30)$$

It is now necessary to calculate $\bar{\tau}$ by the usual averaging process, whereby

$$\bar{\tau} = \frac{\langle v_0^2 \tau(v_0) \rangle}{\langle v_0^2 \rangle} = \frac{\kappa^2 m^{*2}}{2\pi Z^2 e^4 N} \frac{\int_0^\infty \frac{v_0^7 e^{-m^*v_0^2/2kT} dv_0}{\ln [1 + (km^*v_0^2/2Ze^2 N^{1/3})^2]} \int_0^\infty v_0^4 e^{-m^*v_0^2/2kT} dv_0}{\int_0^\infty v_0^6 e^{-m^*v_0^2/2kT} dv_0}. \quad (9.11-31)$$

This integral cannot be exactly evaluated by analytical means, but if one observes that

the logarithmic term in the denominator varies only very slowly with respect to v_0 as compared with the multiplying factor $v_0^7 e^{-m^*v_0^2/2kT}$, an approximate value can be obtained by assuming that the logarithmic term has a constant value which is equal to the value it attains when the multiplying factor is a maximum. It is easily ascertained that this maximum value is reached for $v_0 = \sqrt{7kT/m^*}$. Setting v_0 equal to this quantity in the logarithmic term, removing it from the integral, and evaluating the remaining integrals by referring to Table 5.1, we may obtain

$$\bar{\tau} = \frac{8\kappa^2(kT)^{\frac{3}{2}}(2m^*)^{\frac{1}{2}}}{\pi^{\frac{3}{2}}Z^2e^4N \ln [1 + (7\kappa kT/2Ze^2N^{\frac{1}{3}})^2]} \quad (9.11-32)$$

whereby the mobility will be given by

$$\mu = \frac{e\bar{\tau}}{m^*} = \frac{8\sqrt{2}\kappa^2(kT)^{\frac{3}{2}}}{\pi^{\frac{3}{2}}Z^2e^3m^{*\frac{1}{2}}N \ln [1 + (7\kappa kT/2Ze^2N^{\frac{1}{3}})^2]}. \quad (9.11-33)$$

This method of calculation was developed originally by Conwell and Weisskopf^{11,20} and (9.11-33) is often called the Conwell-Weisskopf formula. The same calculation was done in a somewhat more refined way by Brooks and Herring;²¹ their result is similar, except that the slowly varying logarithmic term is somewhat different. The temperature and concentration dependences are the same in both cases. In the above calculation, the averaging was done assuming spherical energy surfaces. The computation for ellipsoidal surfaces is much more complex and will not be considered here; the magnitude of the relaxation time will be affected but not the temperature and concentration dependences. We have assumed throughout that the impurity atoms were completely ionized; at very low temperatures this is no longer true. Under these circumstances the quantity N in the above equations must be taken to represent the concentration of *ionized* impurity centers only, as given by (9.5-4) or (9.5-8). This introduces an additional temperature dependence, but it is significant only at very low temperatures.

According to Equation (9.11-33), when ionized impurity scattering is the dominant scattering process, the mobility should be found to be proportional to $T^{3/2}$ and to exhibit an inverse dependence on the concentration of impurities. Although it is rather difficult to obtain specimens in which impurity scattering alone determines the mobility, these predictions have been verified at least approximately by experiment. A plot of mobility *versus* impurity concentration for germanium is shown in Figure 9.24. For low impurity concentrations, the dominant scattering interaction is lattice scattering, which is independent of impurity content, and the mobility [as given, for example, by (9.11-21)] is likewise independent of impurity density. At high concentrations of impurity atoms, the ionized impurity scattering becomes the most important

²⁰ There appears to be an error in the original Conwell-Weisskopf article which has persisted in most of the discussions of this subject by other authors. Conwell and Weisskopf state in their original paper [their Equation (18)] that the maximum of the function multiplying the logarithmic term in the integrand in the numerator of (9.11-31) occurs at $v^2 = 6kT/m$, while a simple calculation suffices to show that it really occurs at $7kT/m$. This accounts for the factor $7/2$ in the argument of the logarithmic term of (9.11-33) in place of the factor 3 which is in the original Conwell-Weisskopf formula. The correction to the original result is clearly quite small, and does not affect its general validity.

²¹ H. Brooks, *Phys. Rev.*, **83**, 879 (1951).

scattering process and the mobility decreases with increasing impurity density as predicted by the Conwell-Weisskopf formula (9.11-33). In the intermediate range, both scattering interactions are important and the overall mobility μ must be expressed as $\mu^{-1} = \mu_I^{-1} + \mu_l^{-1}$ where μ_I is the mobility attributable to ionized impurity scattering alone and μ_l that which would be found were lattice scattering the only important process.

In addition to lattice scattering and impurity scattering, certain other scattering interactions are of occasional importance. Neutral impurity centers as well as ionized impurity centers may give rise to appreciable scattering if present in sufficient numbers.

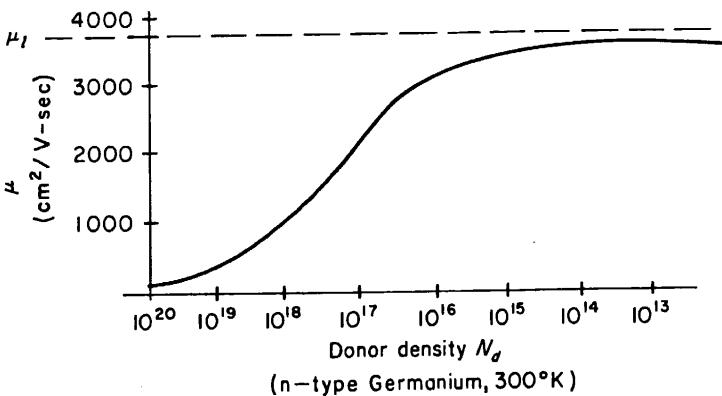


FIGURE 9.24. The effect of impurity concentration upon the room-temperature mobility of electrons in *n*-type germanium, according to the Conwell-Weisskopf theory of impurity scattering. [After E. M. Conwell, *Proc. I.R.E.*, **40**: 1327 (1952).]

A treatment of this subject has been given by Erginsoy.²² Neutral impurity scattering is found to be independent of temperature, and hence may be of importance at very low temperatures, when lattice scattering is negligible and most of the impurities are unionized. Furthermore, vacancies, interstitial atoms, dislocations, grain boundaries, and sample surfaces can scatter holes and electrons, although in many cases the scattering attributable to these agencies is negligible in comparison with lattice or impurity scattering.

The account of the behavior of bulk semiconductors given in this chapter is by no means complete. Indeed, we have not really touched upon thermal or optical properties at all, and for lack of space we have had to omit an account of the very interesting effects which may be observed in semiconductors under hydrostatic or uniaxial stresses.²³ It is hoped, however, that the present treatment has illuminated some of the very basic and important areas and has developed methods of general usefulness in investigating some of the subjects which have not been discussed in detail.

²² C. Erginsoy, *Phys. Rev.*, **79**, 1013 (1950).

²³ R. W. Keyes, "The Effects of Elastic Deformation on the Electrical Conductivity of Semiconductors," in *Solid State Physics, Advances in Research and Applications*. New York: Academic Press (1960), Vol. 11, p. 149.

EXERCISES

1. Make plots, on semilogarithmic graph paper, of the intrinsic carrier concentration n_i as a function of $1/T$ for (a) Germanium (b) Silicon (c) InSb ($\Delta\epsilon = 0.22$ eV, $m_n^* = 0.013 m_0$, $m_p^* = 0.18 m_0$) (d) GaAs ($\Delta\epsilon = 1.5$ eV, $m_n^* = 0.1 m_0$, $m_p^* = 0.4 m_0$). Be sure your plots show clearly the variation of n_i with T over the range from 10–500°K.

2. Consider a two-dimensional intrinsic semiconductor for which, in analogy with two-dimensional metallic systems, the density of states functions are

$$g_n(\epsilon) d\epsilon = (4\pi m_n^*/h^2) d\epsilon \quad \epsilon > \epsilon_c$$

$$g(\epsilon) = 0 \quad \epsilon_v < \epsilon < \epsilon_c$$

$$g_p(\epsilon) d\epsilon = (4\pi m_p^*/h^2) d\epsilon \quad \epsilon < \epsilon_v$$

Show, using Fermi statistics only, *without* the use of the Boltzmann approximation that for the case where $m_p^*/m_n^* = 2$ the position of Fermi level may be expressed as

$$\epsilon_c - \epsilon_f = \frac{1}{2} \Delta\epsilon - \frac{kT}{2} \ln \frac{8}{3} - kT \ln \cos \frac{\Phi}{3},$$

where

$$\Phi = \tan^{-1} \left[\frac{32}{27} (e^{4\epsilon_f/kT}) - 1 \right]^{1/2}$$

3. Make plots of the conductivity of both *p*-type and *n*-type germanium as a function of net impurity concentration $|N_d - N_a|$ for values of $|N_d - N_a|$ ranging from zero to 10^{17} cm $^{-3}$, at temperatures of 80°K, 300°K and 500°K. Assume that μ_n varies as $T^{-1.65}$ and μ_p as $T^{-2.3}$. Use logarithmic graph paper.

4. For what value of carrier concentrations n_0, p_0 is the conductivity of a semiconductor a *minimum*? What value of net impurity content $|N_d - N_a|$ does this imply?

5. For a semiconductor with spherical energy surfaces in both conduction and valence bands, but with *two* species of holes, show that the small field Hall coefficient can be written as

$$R = \frac{p_{01}\mu_{p1}^2 \frac{\tau_{p1}^2}{(\tau_{p1})^2} + p_{02}\mu_{p2}^2 \frac{\tau_{p2}^2}{(\tau_{p2})^2} - n_0\mu_n^2 \frac{\tau_n^2}{(\tau_n)^2}}{ec(n_0\mu_n + p_{01}\mu_{p1} + p_{02}\mu_{p2})^2}$$

where p_{01}, p_{02} are the concentrations, μ_{p1}, μ_{p2} the mobilities, and τ_{p1}, τ_{p2} the relaxation times for the two species of holes.

6. Find, for each of the constant energy surfaces of germanium, the individual conductivity tensors $\sigma_{ab}^{(i)}$ referred to the cube axes as coordinate axes. In germanium the constant-energy surfaces are four ellipsoids whose major axes lies along the four $\langle 111 \rangle$ directions. Show that the result (9.10-23) is obtained by summing over all ellipsoids.

7. Show that if the static field is in the [100] direction, only a single cyclotron resonance peak at the frequency $\omega = (eB/c)[(m_\parallel^* + 2m_\perp^*)/3m_\parallel^*m_\perp^{*2}]^{1/2}$ will be observed for germanium. What resonances would be observed with the static field in the [111] direction?

8. Using the laws of classical mechanics, derive the Rutherford scattering formula (9.11-23) and the associated relation (9.11-24).

9. A semiconductor specimen is determined to have a resistivity of 12.5 ohm-cm. The specimen is 1 cm long by 5 mm wide by 1 mm thick. A current of 1 ma. is made to flow

9.11

SEC. 9.11

UNIFORM ELECTRONIC SEMICONDUCTORS

319

along the longest dimension of the sample and a Hall voltage of 5 mV is measured across the 5 mm width of the sample, when a magnetic field of 2000 gauss is used. What is the carrier concentration and the Hall mobility of the carriers in the sample? You may assume that the sample is an extrinsic one, that the energy surfaces are spherical, and that acoustical mode lattice scattering is the dominant scattering process.

10. Fill in the details involving the maximization of expression (9.5-2) by the method of Lagrangean multipliers, thus completing the derivation of (9.5-3).

11. Show that (9.3-6) and (9.3-15) can be transformed so as to read

$$n_0 = n_i e^{(e_f - e_{f1})/kT}$$

$$p_0 = n_i e^{(e_{f1} - e_f)/kT}.$$

12. Explain *physically* the principal features of the experimental curves shown in Figure 9.13, and discuss the relationships between the three separate plots.

13. Show that in the case where acoustical mode lattice scattering is the dominant scattering mechanism, Equation (9.8-23) may be expressed in the form shown at the bottom of page 286.

14. In the case where ionized impurity scattering is the dominant mechanism, it is evident from Equation (9.11-30) that the dependence of the relaxation time on velocity can be expressed as $\tau(v) = \alpha v^3$ where α is constant, provided that the weak velocity dependence of the logarithmic term is neglected. Show, in this instance, that the longitudinal magnetoresistance Equation (9.8-23) may be written in the form

$$I_x = \sigma_0 E_x \left[1 - \frac{e^2 B_0^2}{m_p^{*2} c^2} \left(\frac{2kT}{m_p^*} \right) \left(120 - \frac{\pi(315)^2}{(64)^2} \right) \right].$$

GENERAL REFERENCES

- A. C. Beer, "Galvanomagnetic Effects in Semiconductors"; Supplement 4 in *Solid State Physics, Advances in Research and Applications*, Academic Press, Inc., New York (1963).
- F. J. Blatt, "Theory of Mobility of Electrons in Solids," in *Solid State Physics, Advances in Research and Applications*, Vol. 4, Academic Press, Inc., New York (1957).
- H. Brooks, "Theory of the Electrical Properties of Germanium and Silicon," in *Advances in Electronics and Electron Physics*, Vol. 7, Academic Press, Inc., New York (1955).
- E. Burstein and P. H. Egli, "The Physics of Semiconductor Materials," in *Advances in Electronics and Electron Physics*, Vol. 7, Academic Press, Inc., New York (1955).
- H. Y. Fan, "Valence Semiconductors, Germanium and Silicon," in *Solid State Physics, Advances in Research and Applications*, Vol. 1, Academic Press, Inc., New York (1955).
- R. W. Keyes, "The Effect of Elastic Deformation on the Electrical Conductivity of Semiconductors," in *Solid State Physics, Advances in Research and Applications*, Vol. 11, Academic Press, Inc., New York (1960).
- W. Shockley, *Electrons and Holes in Semiconductors*, D. van Nostrand Co., Inc., New York (1950).
- R. A. Smith, *Semiconductors*, Cambridge University Press, New York (1961).