

Chapter 3

SEMICONDUCTORS DOMINATED BY IMPURITY LEVELS

IN Section 2.4, an allusion was made to the importance of impurity levels, in the sense that consideration was given to the relation of n_0 and p_0 when these were unequal and the Fermi level departed from the intrinsic position. It is now necessary to consider in detail just how electrons are distributed in thermal equilibrium between the bands and the various sets of levels provided by impurities.

3.1 OCCUPANCY FACTOR FOR IMPURITY LEVELS

Appendix A proves that under thermal equilibrium conditions, the probability that any state of energy E will be occupied by an electron is given by the Fermi–Dirac factor

$$f(E) = \frac{1}{1 + \exp \left[\frac{E - \phi}{kT} \right]} \quad (310.1)$$

In the last chapter we have seen that the application of Eq. (310.1) to the distribution of states in the permitted bands leads to expressions involving the Fermi–Dirac integrals $\mathcal{F}_j(\eta)$. It is similarly necessary to find out what fraction of impurity atoms retain their outermost electron (or acquire an additional one) when the Fermi level is at any specified energy.

Following Wilson's (1931:2) theory of the extrinsic semiconductive process, a rather simple approach to the problem of center occupancy

was widely used during the two following decades. This approach is typified by the following argument. Consider a semiconductor containing N_d monovalent donor impurity atoms per unit volume, such that in the ground state each atom has an electron trapped at an energy E_d below the base of the conduction band (see part (a) of Fig. 13.8). It would seemingly appear that the number of states within any small range of energy centered on $(E_c - E_d)$ must be

$$\int_{E_c - E_d - \Delta}^{E_c - E_d + \Delta} g(E) dE = N_d \quad (310.2)$$

A donor which has an electron trapped at this energy is electrically neutral; the density of these may be indicated as N_{dn} . Similarly, $N_{di} = (N_d - N_{dn})$ denotes the density of ionized donor atoms. From the formulation of Fermi-Dirac statistics it is to be expected that

$$N_{di} : N_{dn} = 1 : \exp\left[\frac{\phi + E_d - E_c}{kT}\right] \quad (310.3)$$

so that the density of *neutral* donors would be

$$N_{dn} = \frac{N_d}{1 + \exp\left[\frac{E_c - E_d - \phi}{kT}\right]} \quad (310.4)$$

This result looks very simple and appealing, but unfortunately represents an incorrect simplification of the matter. The factor overlooked is that of impurity level spin degeneracy, which we now consider.

3.1.1 IMPURITY LEVEL SPIN DEGENERACY

The important consideration overlooked in the preceding discussion is that each impurity atom will usually offer *more* than one state for either the neutral or ionized configuration. This has been recognized by a few workers in the field for a considerable time (e.g. 1948:2), but there has been a more general awareness since the appearance of contributions by Landsberg (1952:4, 1953:5) and by Guggenheim

(1953:6). Using a free energy approach and the grand canonical ensemble, respectively, these authors show how spin and orbital degeneracy of bound states have an effect on the number of electrons retained by a set of impurities.

As the simplest example of this, consider a set of simple monovalent donor impurities for which all electrons save the least tightly bound are in paired valence bonds. The wave-function of the outermost electron is of purely *s* character, and since this is an unpaired electron, it can be trapped in two ways, with spin either up or down. Note that a donor can not trap *two* electrons, since once one electron is trapped, electrostatic forces raise the remaining spin possibility to a very high energy; nevertheless the neutral state of the impurity has a statistical weight of two compared with the ionized state. Accordingly we should replace Eq. (310.3) by

$$N_{di} : N_{dn} = 1 : 2 \exp \left[\frac{\phi + E_d - E_c}{kT} \right] \quad (311.1)$$

Then the density of *neutral* donors is

$$N_{dn} = \frac{N_d}{1 + \frac{1}{2} \exp \left[\frac{E_c - E_d - \phi}{kT} \right]} \quad (311.2)$$

For other kinds of impurity level, the coefficient of the exponential in Eq. (311.2) will differ from $\frac{1}{2}$. We may say in general that for a level at energy E_r , the probability that this level will contain an electron is

$$P(E_r) = \frac{1}{1 + \beta_r \exp \left[\frac{E_r - \phi}{kT} \right]} \quad (311.3)$$

In the current literature, the term *impurity level spin degeneracy* is applied rather indiscriminately to *either* β_r or β_r^{-1} , whichever is larger than unity. This may appear a little confusing, but should not cause undue difficulty if it is borne in mind that Eq. (311.3) refers always to the probability that a level contains an *electron* at the specified energy, *not* necessarily that the impurity is electrically neutral. For our illustrative example of simple donors, $\beta^{-1} = 2$ is the spin degeneracy, since the level can be occupied by an *electron* in two ways. When the impurity in question is an acceptor center which requires an extra

electron to complete a set of paired bonds, $\beta = 2$ is the spin degeneracy—since the *absence* of this electron can be described in two ways.

When an impurity level is created by splitting off states from a conduction or valence band with multiple or degenerate extrema, β or β^{-1} will be larger than two. Thus for acceptor levels introduced by Group III impurities into a semiconductor such as germanium (with two valence bands degenerate at $\mathbf{k} = 0$), Kohn (1957:10) has shown that β should equal four, and this is in accord with experimental observations (e.g. 1959:6).

Moreover, when the impurity ground state wave-function is not spherically symmetrical, orbital degeneracy affects the statistical weights of the neutral and ionized conditions. The effect of this can be included in the quantity β , which is still loosely referred to as a “spin” degeneracy factor. For multivalent impurities which can donate or accept several electrons from successively deeper states, there will be a β_r characteristic of each state of ionization.

3.2 SEMICONDUCTORS CONTROLLED BY A SINGLE MONOVALENT DONOR SPECIES

In Chapter 2 we traced the connection between the Fermi level and the conduction band free electron density at any temperature. The previous section has further established the connection between the Fermi level and the fraction of impurity sites which retain an electron. Combining these two pieces of information, we can now see how electrons are *distributed* between impurity levels and the conduction band.

It is equally simple to study the models of either n -type or p -type semiconductors, though this section will in practice discuss primarily n -type behavior, concentrating on the conduction band. Except when especially noted, it will be assumed that the hole density p_0 is negligibly small compared with n_0 and with the densities of impurity states (i.e. we shall not discuss the transition towards intrinsic conduction at high temperatures). This is not as restrictive an assumption as it may seem, for provided that the majority of donor levels lie inside the upper half of the intrinsic gap (at least two or three kT above ϕ_i) all these impurities will become ionized before temperatures are reached at which intrinsic excitation is at all appreciable. This means that the intrinsic transition takes place from a region of temperature-independent N_r , as described in Section 2.4.

3.2.1 TEMPERATURE DEPENDENCE OF n_0 AND ϕ FOR A SET OF SIMPLE UNCOMPENSATED DONORS

The model to be considered first is a purely academic and unrealistic one. For it is impossible to prepare a semiconductor containing only one set of impurities, with no compensating centers at all. Grave errors can result from applying this model over the complete temperature range, though it can be a reasonable approximation to the truth at temperatures which are not too low. This earliest described model of an extrinsic semiconductor (1931:2) is used simply because it helps to demonstrate some features which are useful in discussing more realistic models.

For this preliminary discussion, then, the model to be used is the simplified one of Fig. 32.1. There are N_d donor levels of spin degeneracy

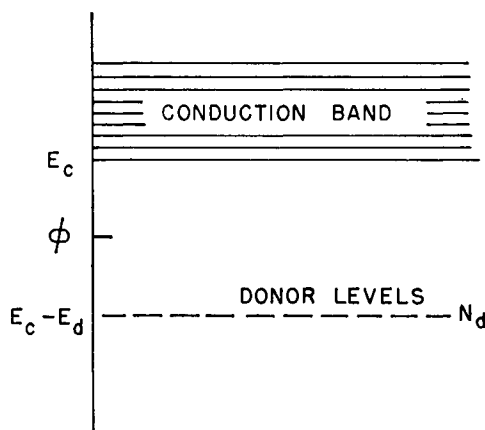


FIG. 32.1. The simplest model of an extrinsic n -type semiconductor.

β^{-1} per unit volume at an energy E_d below the base of the conduction band. The latter is characterized by a mass m_c and through this by the quantity $N_c = 2(2\pi m_c kT/h^2)^{3/2}$. The Fermi level ϕ may be above or below either E_c or $(E_c - E_d)$.

Now the number of electrons excited in the conduction band is known to be

$$n_0 = N_c \mathcal{F}_{1/2} \left(\frac{\phi - E_c}{kT} \right) \quad (321.1)$$

while from the formalism of Eq. (311.3) it is known that the number of electrons remaining in the donor levels is

$$\mathcal{N}_{dn} = \frac{\mathcal{N}_d}{1 + \beta \exp\left[\frac{E_c - E_d - \phi}{kT}\right]} \quad (321.2)$$

Since electrons in the conduction band can have come only from the donor levels, the sum of the densities (321.1) and (321.2) must be \mathcal{N}_d itself. Hence we have that

$$n_0 = \mathcal{N}_c \mathcal{F}_{1/2}\left[\frac{\phi - E_c}{kT}\right] = \frac{\mathcal{N}_d}{1 + \beta^{-1} \exp\left[\frac{E_d + \phi - E_c}{kT}\right]} \quad (321.3)$$

Adopting the usual dimensionless notation, and in addition denoting (E_d/kT) as ϵ_d , the previous equation can be written

$$n_0 = \mathcal{N}_c \mathcal{F}_{1/2}(\eta) = \mathcal{N}_d [1 + \beta^{-1} \exp(\epsilon_d + \eta)]^{-1} \quad (321.4)$$

When the donors are not too numerous, and of reasonably large ionization energy, conditions will tend to remain non-degenerate (ϕ several kT below E_c) at all temperatures. They will certainly be non-degenerate at the very lowest temperatures, for then all the electrons are withdrawn to the donor levels. As temperature rises and some electrons become excited into the conduction band, η rises; but it passes through a maximum and falls again at the high temperatures for which impurity ionization approaches completion.

When conditions are non-degenerate over the entire temperature range, the solution of Eq. (321.4) is conveniently simple. The approximation $\mathcal{F}_{1/2}(\eta) \approx \exp(\eta)$ can then be used at all temperatures. Thus substituting (n_0/\mathcal{N}_c) for $\exp(\eta)$ on the right side of Eq. (321.4), there results a simple quadratic in n_0 ,

$$n_0 \left[1 + \frac{n_0}{\beta \mathcal{N}_c} \exp(\epsilon_d) \right] = \mathcal{N}_d \quad (321.5)$$

with the solution

$$n_0 = \frac{2\mathcal{N}_d}{1 + \sqrt{1 + (4\mathcal{N}_d/\beta \mathcal{N}_c) \exp(\epsilon_d)}} \quad (321.6)$$

Fig. 32.2 illustrates a typical numerical example of the temperature dependence of n_0 prescribed by Eq. (321.6). Following the usual custom, this figure plots $\ln(n_0)$ vs. $1/T$. This is done for the benefit of the low temperature "reserve" region, when few electrons are excited and Eq. (326.6) approximates to

$$n_0 \approx (\beta N_c N_d)^{1/2} \exp(-E_d/2kT), \quad \text{small } T \quad (321.7)$$

The exponential contributes most but not all of the temperature dependence in this range, as can be seen in the figure from a comparison

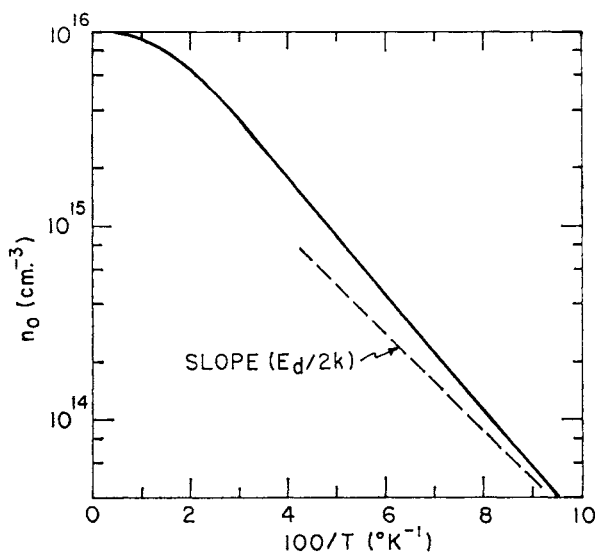


FIG. 32.2. Semilogarithmic plot of n_0 vs. $1/T$ for a set of uncompensated donors when conditions are non-degenerate at all temperatures. Calculated for $N_d = 10^{16} \text{ cm}^{-3}$, $E_d = 0.01 \text{ eV}$, $m_c = 0.25 m_0$, $\beta = \frac{1}{2}$.

of the straight line of slope $(E_d/2k)$ with the curve of a calculated carrier density. The discrepancy in slopes results from the $T^{3/2}$ temperature dependence of N_c .

Eq. (321.7) was originally derived by Wilson (1931:2) and was for a long time adopted uncritically in the interpretation of experimental

data. We should note that the equation can properly be applied only if

- (a) conditions are completely non-degenerate;
- (b) the number of carriers n_0 excited to the conduction band is very small compared with N_d ; and furthermore
- (c) the number of electrons lost to compensating impurities is very small compared with n_0 , i.e. $N_a \ll n_0 \ll N_d$.

We have deliberately assumed *zero* compensation for the present, and thus are safe on the last score. As will be revealed in a moment, the numerical values adopted in Fig. 32.2 are consistent with non-degeneracy at all temperatures. Before proceeding to this topic, it may be noted in passing that the carrier density of Eq. (321.6) tends to N_d at sufficiently high temperatures, producing what have previously been described as exhaustion conditions. Since at sufficiently low temperatures almost *none* of the donors are ionized, while at high temperatures they are almost *all* ionized, it is evident that ϕ tends from a position above $(E_c - E_d)$ to one below this energy on warming.

When conditions are non-degenerate, the Fermi level is always related to n_0 through

$$(\phi - E_c) = kT\eta = kT \ln(n_0/N_c) \quad (321.8)$$

Substituting for n_0 from Eq. (321.6),

$$\begin{aligned} (\phi - E_c) &= kT\eta \\ &= -kT \ln\{(N_c/2N_d) + \sqrt{[(N_c/2N_d)^2 + (N_c/\beta N_d) \exp(E_d/kT)]}\} \end{aligned} \quad (321.9)$$

Fig. 32.3 (based on the same numerical parameters as Fig. 32.2) shows how n_0 , η and ϕ vary with temperature in accordance with Eq. (321.6) and (321.9).

In connection with this figure, note first the behavior of $(\phi - E_c)$, shown in part (c). At high temperatures, when essentially all the impurities are ionized, conditions approach those discussed in Section 2.2; for when T is large enough to make $\exp(E_d/kT) \ll (N_c/N_d)$, the Fermi level of Eq. (321.9) approximates to

$$\phi \approx E_c - kT \ln(N_c/N_d), \quad \text{large } T \quad (321.10)$$

This would be true for the example of Fig. 32.3 at temperatures $> 100^\circ\text{K}$, when at least 90% of the carriers have been excited from the donor levels.

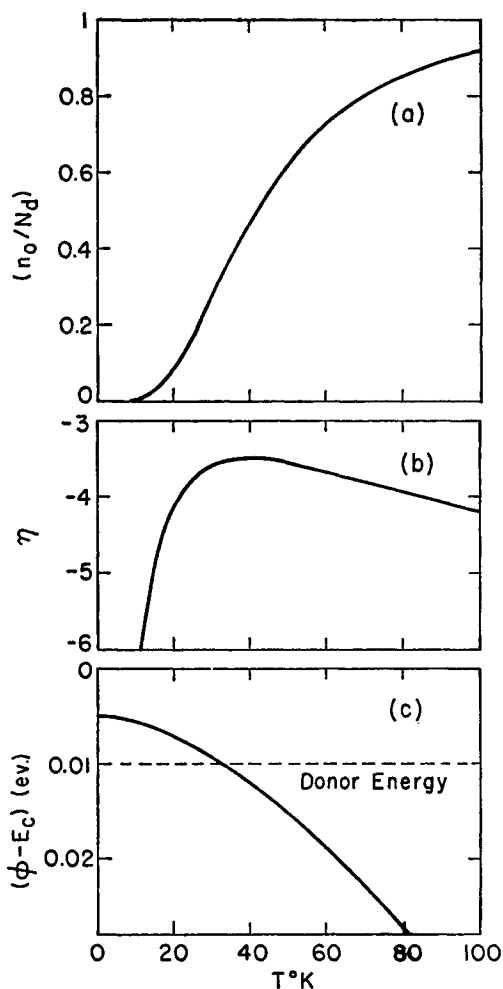


FIG. 32.3. The temperature dependence of (n_0/N_D) , η and ϕ for the uncompensated donor model of Fig. 32.2.

On proceeding to lower temperatures, a progressively smaller fraction of the carriers is excited, and ϕ moves closer to the conduction band. In conformity with (321.2) ϕ must coincide with $(E_c - E_d)$ at the temperature for which $(1 + \beta)^{-1}$ of the donors are neutral. In the present example, this temperature is 32°K.

When only a very small fraction of the donors are ionized, ϕ approaches the energy $(E_c - \frac{1}{2}E_d)$, but the approach is not a monotonic function of temperature. This occurs because the low temperature approximation to Eq. (321.9) is

$$\phi \approx (E_c - \frac{1}{2}E_d) - \frac{1}{2}kT \ln(N_c/\beta N_d), \quad \text{small } T \quad (321.11)$$

At some sufficiently small temperature T_1 , N_c will be equal to βN_d ; then ϕ will be higher than $(E_c - \frac{1}{2}E_d)$ for the temperature range $0 < T < T_1$. This is not readily apparent in Fig. 32.3, and so the data from 10°K downwards are shown on an expanded scale in Fig. 32.4. For the numerical example considered in these figures, $T_1 = 4.1^\circ\text{K}$.

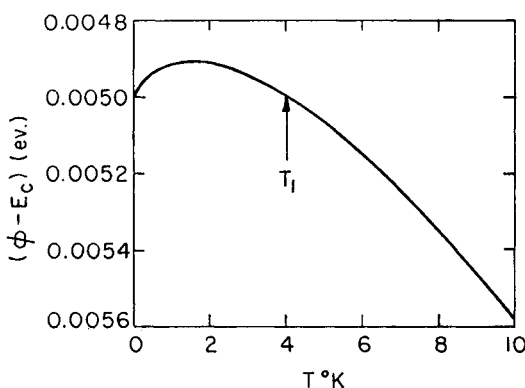


FIG. 32.4. An amplification of the low-temperature region in part (c) of Fig. 32.3, showing ϕ rising above $(E_c - \frac{1}{2}E_d)$ when the temperature is small enough to make $N_c < \beta N_d$.

Part (b) of Fig. 32.3 shows how the reduced Fermi level η varies with temperature. As already predicted, this variable passes through a maximum η_m at a temperature for which an appreciable fraction of the donors is ionized. The value of this maximum is sufficiently negative to confirm that n_0 was correctly calculated from the non-degenerate equation (321.6). It is obviously desirable to have a general procedure for determining η_{\max} from parameters such as N_d , β , m_c and E_d , so that it is known in advance whether Eq. (321.6) is applicable or whether Eq. (321.4) must be solved in more general form to yield n_0 and η .

Taking η from Eq. (321.9), differentiating with respect to temperature and setting the result equal to zero, it can be shown as a result

of some tedious but elementary manipulation that η will have a maximum at a temperature T_m for which $\epsilon_{dm} = (E_d/kT_m)$ satisfies the condition

$$(2\epsilon_{dm} - 3)^2 \exp(\epsilon_{dm}) = 6\beta N_c \epsilon_{dm} / N_d \quad (321.12)$$

To employ this condition, we should like to collect everything involving temperature-dependence on one side. Since N_c varies as $T^{3/2}$, we can write (321.12) as

$$\mathcal{Z} = 12(2\pi m_0/h^2)^{3/2} \epsilon_{dm}^{-1/2} (2\epsilon_{dm} - 3)^{-2} \exp(-\epsilon_{dm}) \quad (321.13)$$

where the quantity

$$\mathcal{Z} = N_d \beta^{-1} (m_0/m_c E_d) \quad (321.14)$$

depends on the density and character of the impurities but not on temperature. When N_d is expressed in cm^{-3} and E_d in eV, then

$$\mathcal{Z} = 3.6 \times 10^{22} \epsilon_{dm}^{-1/2} (2\epsilon_{dm} - 3)^{-2} \exp(-\epsilon_{dm}) \text{ cm}^{-3} \text{ eV}^{-3/2} \quad (321.15)$$

It is a simple matter to determine the corresponding values of \mathcal{Z} and ϵ_{dm} , and their relationship is shown in Fig. 32.5. Knowledge of ϵ_{dm}

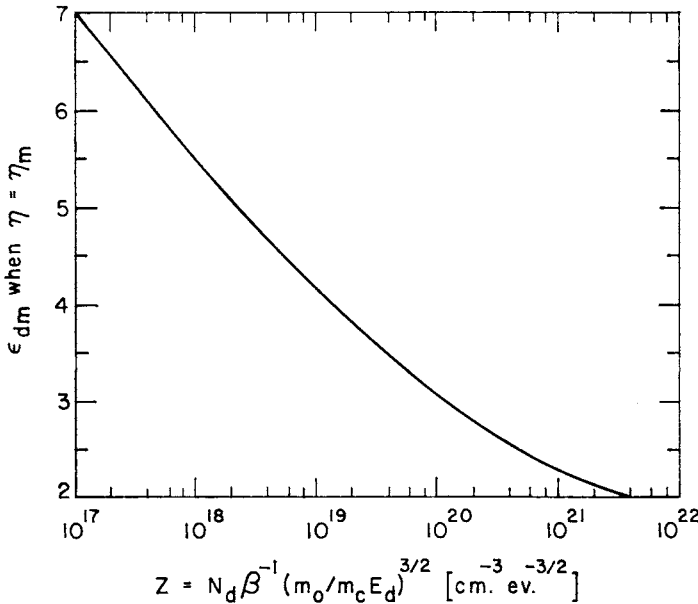


FIG. 32.5. Determination of the temperature at which η reaches its maximum value η_m , for a set of uncompensated donor levels. The ordinate is the parameter $\epsilon_{dm} = (E_d/kT_m)$ which satisfies Eq. (321.15).

means that $T_m = (E_d/k\epsilon_{dm})$ is obtained, and η_m itself can then be determined.

When Eq. (321.4) is differentiated and the condition $(d\eta/dT) = 0$ imposed, a simple relationship between η_m and ϵ_{dm} is obtained,

$$\eta_m = -\epsilon_{dm} + \ln \left[\frac{3\beta}{2\epsilon_{dm} - 3} \right] \quad (321.16)$$

This is valid whether conditions be degenerate or not. Combination of Eq. (321.16) with the root of Eq. (321.15) leads to the graphical relationship of η_m to Z shown in Fig. 32.6. Strictly speaking we should

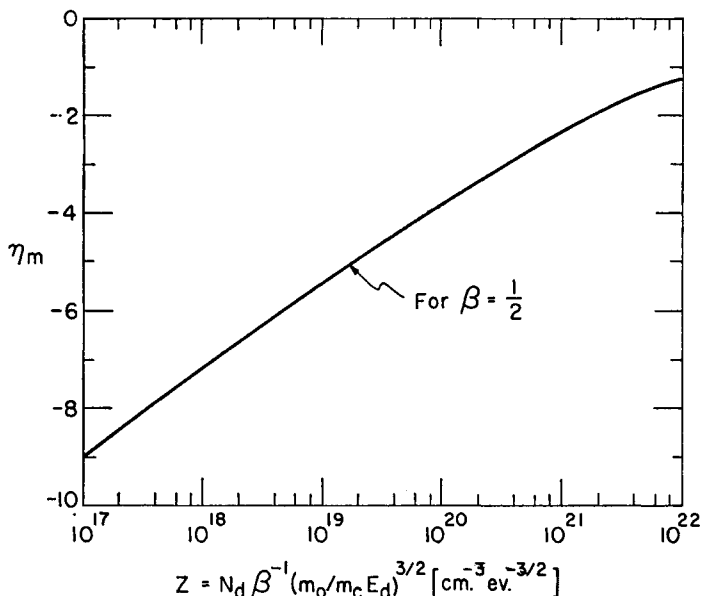


FIG. 32.6. The maximum value reached by η as a function of the density, ionization energy, etc., of a set of uncompensated donor centers. The curve shown is for $\beta = \frac{1}{2}$. With any other value, the curve should be raised by the amount $\ln(2\beta)$.

draw a family of parallel curves for various values of β ; the one shown is for $\beta = \frac{1}{2}$ so that the curve for any other β would be displaced vertically by $\ln(2\beta)$.

The solution of Eq. (321.4) is less simple than we have described if the donor density is large enough to produce some conduction band

degeneracy over part of the temperature range. One of the earliest and most striking discussions of the carrier statistics problem for a degenerate extrinsic semiconductor is that of Shifrin (1944:1). No great complication is involved provided that the maximum value of η is not much greater than unity, for then $\mathcal{F}_{1/2}(\eta)$ can be approximated by an expression of the form $[C + \exp(-\eta)]^{-1}$. As discussed in Appendix C, this kind of expression with $C \approx 0.27$ is suitable whenever $\eta < 1.3$. With such an approximation to the behavior of $\mathcal{F}_{1/2}(\eta)$, Eq. (321.4) becomes a quadratic in $\exp(\eta)$:

$$N_c \exp(2\eta + \epsilon_a) + \beta(N_c - CN_a) \exp(\eta) = \beta N_a, \quad \eta < 1.3 \quad (321.17)$$

For the reduced Fermi level, the solution is

$$\exp(\eta) = \frac{2N_a}{(N_c - CN_a) + \sqrt{[(N_c - CN_a)^2 + 4\beta^{-1}N_aN_c \exp(\epsilon_a)]}}, \quad \eta < 1.3 \quad (321.18)$$

while the corresponding free electron density is

$$n_0 = \frac{2N_aN_c}{(N_c + CN_a) + \sqrt{[(N_c - CN_a)^2 + 4\beta^{-1}N_aN_c \exp(\epsilon_a)]}}, \quad \eta < 1.3 \quad (321.19)$$

Considerably more difficulty is encountered if the donor density is so large that η goes through a maximum value rather larger than unity. While Appendix C does discuss approximate expressions for $\mathcal{F}_{1/2}(\eta)$ in the degenerate domain, none of these has a form which would permit a simple analytic solution for Eq. (321.4). However, it is always possible to find the mutually consistent values of ϵ_a and η . We can write (321.4) as

$$\begin{aligned} [1 + \beta^{-1} \exp(\epsilon_a + \eta)]^{-1} &= N_c N_a^{-1} \mathcal{F}_{1/2}(\eta) \\ &= A \epsilon_a^{-3/2} \mathcal{F}_{1/2}(\eta) \end{aligned} \quad (321.20)$$

where A is a dimensionless quantity characteristic of the density and type of donor centers,

$$\begin{aligned} A &= 2(2\pi m_c E_a / h^2)^{3/2} N_a^{-1} \\ &= 6 \times 10^{21} (E_a m_c / m_0)^{3/2} N_a^{-1} \quad \text{for} \quad \begin{cases} N_a \text{ in cm}^{-3} \\ E_a \text{ in eV} \end{cases} \end{aligned} \quad (321.21)$$

Now if we denote

$$\left. \begin{aligned} x_1 &= [1 + \beta^{-1} \exp(\epsilon_a + \eta)]^{-1} \\ x_2 &= [A \epsilon_a^{-3/2} \mathcal{F}_{1/2}(\eta)] \end{aligned} \right\} \quad (321.22)$$

it will always be possible by numerical or graphical methods to find the value of η which permits the condition $x_1 = x_2$ to be satisfied for any value of ϵ_a (i.e. for any temperature). Such procedures are certainly tedious, but there is no other alternative for an exact solution.

In order to bring home the change wrought by conduction band degeneracy on the temperature dependence of n_0 , Fig. 32.7 shows how n_0 and η vary with $1/T$ for a typical numerical example. The ionization energy, spin factor and effective mass ratio are all the same as for the previous example of Fig. 32.2; the only parameter changed is that N_a is now assumed to be 10^{19} cm^{-3} instead of 10^{16} cm^{-3} . It is interesting to compare the shape of the curve in Fig. 32.2 with that of part (a) in Fig. 32.7. In the former case the slope of $\ln(n_0)$ vs. $1/T$ was not very different from $(E_a/2k)$ when donor ionization was less than 10%. This is certainly not true for the example of Fig. 32.7, because the Fermi level is well above E_c until the temperature is low enough to make $n_0 < 0.01 N_a$. At *sufficiently* small temperatures and ionized densities, the slope does eventually approach $(E_a/2k)$, but Fig. 32.7 demonstrates how dangerous it can be to infer an impurity ionization energy from an experimental curve unless all the factors are known.

Of course, the situation for a real solid is complicated by a number of other factors as well, which must now be considered in turn.

3.2.2 THE REALISTIC CASE—PARTLY COMPENSATED IMPURITIES

The most important respect in which the discussion of Sub-section 3.2.1 fails to represent the properties of any real extrinsic semiconductor is that semiconductors inevitably contain traces of *several kinds* of impurity. Sophistication in purification techniques may *reduce* the influence of "unwelcome" impurities, but no process can completely eradicate them.

There are several kinds of complexity which have to be considered. Thus an n -type semiconductor will usually contain significant quantities of more than one donor species. It will always contain acceptor impurities of various kinds. Some impurities may be able to provide or accept more than one electron. Moreover, electrons can be trapped by an impurity not only in ground states but also in excited states, and the spin and orbital degeneracy of any bound state may be modified by splitting occasioned by a magnetic field or by anisotropic elastic strain.

Thus the general problem for an extrinsic semiconductor is quite complicated. It is possible, as Landsberg (1958:10) has shown very powerfully (by the method of the grand canonical ensemble), to write

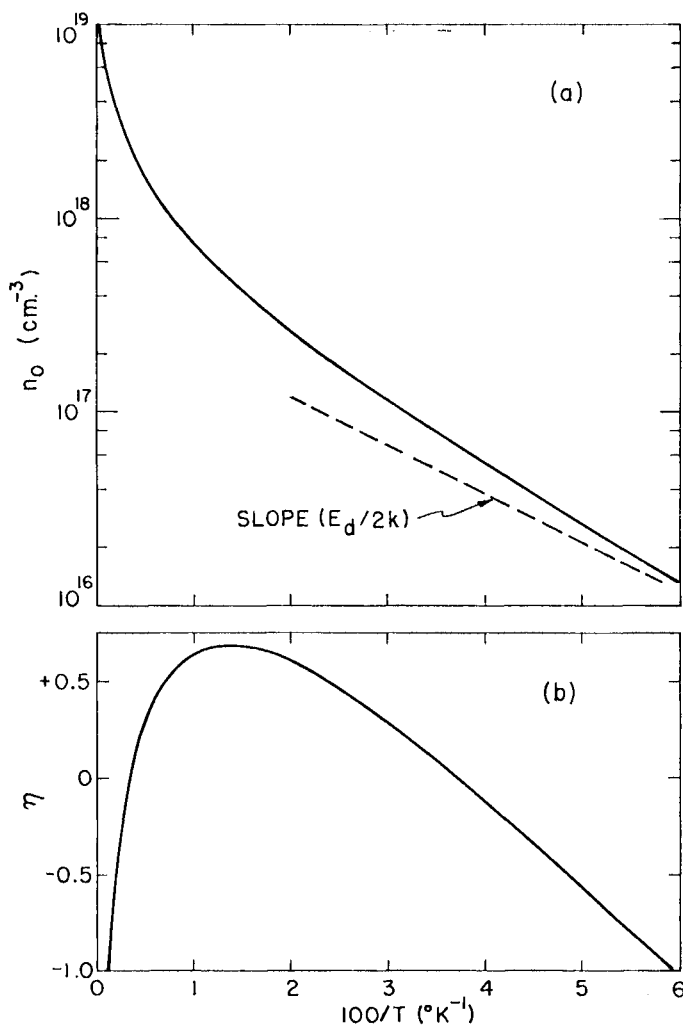


FIG. 32.7. Variation of (a) n_0 and (b) η with reciprocal temperature for a set of uncompensated donors when conditions are partially degenerate over much of the temperature range. Calculated for $N_d = 10^{19} \text{ cm}^{-3}$, $E_d = 0.01 \text{ eV}$, $m_e = 0.25 m_0$, $\beta = \frac{1}{2}$.

down a general expression for the electronic equilibrium in a semiconductor containing any impurity configuration; but this is *not* the same thing as a complete solution. Such a complex problem must be attacked piecemeal, by considering which complications can be relinquished in each individual case.

Accordingly, the complications of multivalent impurities are deferred to Section 3.4. Section 3.3 deals with the problem of an *n*-type semiconductor containing two species of donors at appreciably different energies in the upper half of the gap—and it will be seen at that time that each species will dominate the behavior over a different temperature range. The influence of occupied excited bound states is evaluated

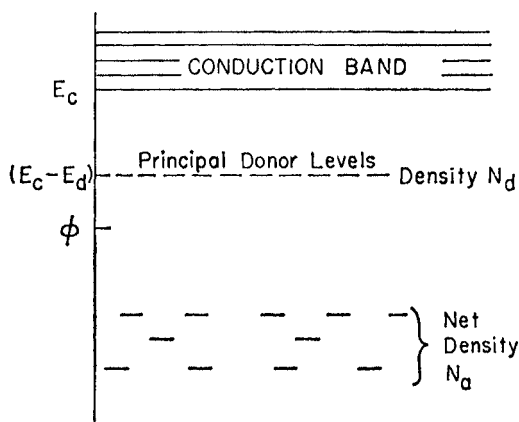


FIG. 32.8. Model of a partially compensated extrinsic *n*-type semiconductor.

in 3.2.3, but first it is fitting to come to grips with the most basic of all these problems; that an extrinsic semiconductor is always partially compensated (1935:1, 1939:1).

Consider then the situation idealized in Fig. 32.8. In addition to the principal set of donor levels, we suppose that there are appreciable concentrations of several other impurity species, whose levels lie further from the conduction band. The Fermi level ϕ will presumably lie well above the energies of all these other kinds of impurity state for all temperatures of interest.

If this is the case, all *donor* states among them will retain their electrons; that is, they will be electrically neutral. Thus their density

will have no bearing on the solution of our problem. All *acceptors*, however, will acquire extra electrons as consistent with their position below the Fermi level. How many acceptors there are of each type is unimportant; the only thing which matters is the *total* density of electrons required to satisfy the needs of these acceptors, a density which will be denoted as \mathcal{N}_a .

Now since these \mathcal{N}_a electrons occupying compensating states have had to come from the principal donors, there remains only a density $(\mathcal{N}_d - \mathcal{N}_a)$ of electrons which can be distributed between the donor levels at $(E_c - E_d)$ and the conduction band. The fraction of ionized donors now varies from $(\mathcal{N}_a/\mathcal{N}_d)$ at low temperatures to unity at high temperatures, and the free electron density can not become larger than $(\mathcal{N}_d - \mathcal{N}_a)$ until the falling Fermi level reaches (a) another set of impurity states, or (b) its intrinsic position.

Since the number of ionized donors is equal to $(n_0 + \mathcal{N}_a)$ and also to $(\mathcal{N}_d - \mathcal{N}_{adn})$ [see Eq. (321.2)], we have for a partially compensated semiconductor that

$$n_0 + \mathcal{N}_a = \mathcal{N}_d [1 + \beta^{-1} \exp(\epsilon_d + \eta)]^{-1} \quad (322.1)$$

When the degree of degeneracy is not known, and n_0 can safely be related to η only by using $n_0 = \mathcal{N}_c \mathcal{F}_{1/2}(\eta)$, we can still write

$$[1 + \beta^{-1} \exp(\epsilon_d + \eta)]^{-1} = [A \epsilon_d^{-3/2} \cdot \mathcal{F}_{1/2}(\eta) - (\mathcal{N}_a/\mathcal{N}_d)] \quad (322.2)$$

where $A = [2(2\pi m_c E_d/h^2)^{3/2}/\mathcal{N}_d]$ is the dimensionless quantity previously introduced in Eq. (321.20). It was noted for the uncompensated situations of Sub-section 3.2.1 that pairs of values for η and ϵ_d could always be found which would make the two sides of Eq. (321.20) equal. Obviously the same kind of procedure will hold good for the partially compensated case of Eq. (322.2).

Eq. (322.1) reduces to a quadratic in $\exp(\eta)$ provided that $\eta \leq +1.3$, so that $(n_0/\mathcal{N}_c) = \mathcal{F}_{1/2}(\eta)$ can be approximated by $[C + \exp(-\eta)]^{-1}$, as described in Appendix C.

When this expression is substituted into Eq. (322.1), some elementary manipulation leads to

$$\begin{aligned} \exp(2\eta) + \exp(\eta) \left\{ \frac{\mathcal{N}_a + \beta \exp(-\epsilon_d) [\mathcal{N}_c - C(\mathcal{N}_d - \mathcal{N}_a)]}{\mathcal{N}_c + C\mathcal{N}_a} \right\} \\ = \frac{\beta(\mathcal{N}_d - \mathcal{N}_a) \exp(-\epsilon_d)}{(\mathcal{N}_c + C\mathcal{N}_a)}, \quad \eta \leq +1.3 \end{aligned} \quad (322.3)$$

This has the solution

$$\exp(\eta) = \frac{2(N_d - N_a)}{[N_c - C(N_d - N_a) + \beta^{-1}N_a \exp(\epsilon_d)] + \sqrt{\{[N_c - C(N_d - N_a) + \beta^{-1}N_a \exp(\epsilon_d)]^2 + 4\beta^{-1}(N_c + CN_a)(N_d - N_a) \exp(\epsilon_d)\}}} \quad (322.4)$$

for $\exp(\eta)$ and

$$n_0 = \frac{2N_c(N_d - N_a)}{[N_c + C(N_d - N_a) + \beta^{-1}N_a \exp(\epsilon_d)] + \sqrt{\{[N_c + C(N_d - N_a) + \beta^{-1}N_a \exp(\epsilon_d)]^2 + 4\beta^{-1}(N_c + CN_a)(N_d - N_a) \exp(\epsilon_d)\}}} \quad (322.5)$$

for the corresponding free electron density.

These expressions are considerably simplified when the Fermi level is several kT below the conduction band, so that $\mathcal{F}_{1/2}(\eta) \approx \exp(\eta)$ and

$$n_0 = \frac{2(N_d - N_a)}{[1 + (N_a/\beta N_c) \exp(\epsilon_d)] + \sqrt{\{[1 + (N_a/\beta N_c) \exp(\epsilon_d)]^2 + (4/\beta N_c)(N_d - N_a) \exp(\epsilon_d)\}}} \quad (322.6)$$

The value of $\exp(\eta)$ is then of course (n_0/N_c) .

A comparison with the equations of Sub-section 3.2.1 shows that Eqs. (322.4), (322.5) and (322.6) reduce to the forms of Eqs. (321.18), (321.19) and (321.6) when the density of compensating centers tends towards zero. But at low temperatures even a very small acceptor density has a considerable effect on n_0 and η . This effect is of the same general character for any degree of conduction band degeneracy, thus the remarks we shall make about the non-degenerate case would apply with little modification to more strongly doped material.

Compare then Eqs. (322.6) and (321.6). The denominator of each is equal to 2 at high temperatures when the donor ionization is complete. On considering progressively lower temperatures, the quantities $(4/\beta N_c)(N_d - N_a) \exp(\epsilon_d)$ and $(N_a/\beta N_c) \exp(\epsilon_d)$ become comparable with and then larger than unity. When the degree of compensation is very small ($N_a \ll N_d$), it is possible to find a temperature range for which $N_a \ll n_0 \ll N_d$. This condition is equivalent to the statement $(N_a/\beta N_c) \exp(\epsilon_d) \ll 1 \ll (4N_d/\beta N_c) \exp(\epsilon_d)$; when such a tempera-

ture range exists, the carrier density approximates to

$$n_0 \approx (\beta N_c N_a)^{1/2} \exp(-E_d/2kT), \quad N_a \ll n_0 \ll N_d \quad (322.7)$$

which is of course just the same as Eq. (321.7). A moment's reflection will confirm that it is entirely reasonable for n_0 not to depend in any material fashion on the density of compensating impurities when this density is much smaller than n_0 .

Such a situation cannot be maintained indefinitely on progressive cooling. Eventually n_0 must approach, and then become smaller than, the density N_a . This happens when the two terms $(N_a/\beta N_c) \exp(\epsilon_d)$ assume the dominant role in the denominator of Eq. (322.6). At sufficiently low temperatures then,

$$n_0 \approx \beta N_c \left[\frac{N_d - N_a}{N_a} \right] \exp(-E_d/kT), \quad n_0 \ll N_a < N_d \quad (322.8)$$

The dependence of $\ln(n_0)$ on $1/T$ is twice as large for this temperature range as in the range of Eq. (322.7).

When the degree of compensation is rather large, there is no range of temperature for which Eq. (322.7) has any validity; on cooling from the exhaustion range, the carrier density decreases rather rapidly and soon conforms with Eq. (322.8).

Fig. 32.9 provides a semilogarithmic plot of n_0 vs. $1/T$ for a semiconductor containing donors and rather fewer acceptors. As with the previous example of Figs. 32.2–32.4, a model is adopted of donors 0.01 eV below a conduction band, when $\beta = \frac{1}{2}$ and $m_c = 0.25 m_0$. The four curves of Fig. 32.9 correspond with different degrees of compensation, but for simplicity it is assumed that $(N_d - N_a)$ is 10^{16} cm^{-3} for each case. Curve (i) corresponds with $N_a = 0$, and is identical with the curve of Fig. 32.2. Curve (ii) is for $N_a = 10^{14} \text{ cm}^{-3}$ (about 1% compensation), curve (iii) for $N_a = 10^{15} \text{ cm}^{-3}$ (about 9% compensation) and curve (iv) for $N_a = 10^{16} \text{ cm}^{-3}$ (50% compensation). In the cases of curves (ii) and (iii), it will be seen that n_0 is reasonably close to the zero-compensation curve until the temperature is low enough for N_a to become comparable with n_0 . The slope then increases and reaches a limiting value of $\sim (E_d/k)$ at the lower temperatures, when Eq. (322.8) applies. The compensation is so large for curve (iv) that

$$n_0 \approx \frac{(N_d - N_a)}{1 + (N_a/\beta N_c) \exp(\epsilon_d)} \quad \text{heavy compensation} \quad (322.9)$$

is a reasonable approximation to the truth at all temperatures.

The Fermi level corresponding with Eq. (322.6):

$$\begin{aligned}
 (\phi - E_c) &= kT\eta = kT \ln(n_0/N_c) \\
 &= kT \ln \left\{ \frac{2\beta(N_d - N_a)}{[\beta N_c + N_a \exp(\epsilon_d)] + \sqrt{\{[\beta N_c + N_a \exp(\epsilon_d)]^2 + 4\beta N_c(N_d - N_a) \exp(\epsilon_d)\}}} \right\}
 \end{aligned}
 \tag{322.10}$$

is not given by a particularly simple expression. From what has been said about the behavior of n_0 , Eq. (322.10) obviously must become

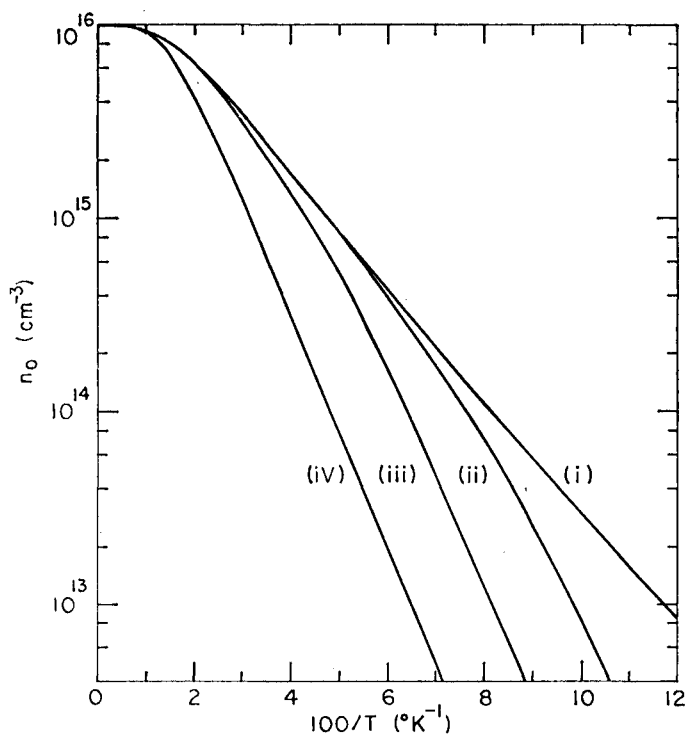


FIG. 32.9. Semilogarithmic plot of n_0 vs. $1/T$ for a set of partially compensated donors under non-degenerate conditions. For simplicity it is assumed that $(N_d - N_a) = 10^{16} \text{ cm}^{-3}$ in each case, with $E_d = 0.01 \text{ eV}$, $m_c = 0.25m_0$, $\beta = \frac{1}{2}$. (i) For $N_a = 0$, i.e. zero compensation. This is identical with the model of Figs. 32.2–32.4. (ii) For $N_a = 10^{14} \text{ cm}^{-3}$. (iii) $N_a = 10^{15} \text{ cm}^{-3}$. (iv) $N_a = 10^{16} \text{ cm}^{-3}$.

essentially equivalent to (321.9) whenever $n_0 \gg N_a$. When the temperature is low enough and the compensation severe enough for N_a to be larger than n_0 , Eq. (322.10) reduces to

$$\phi = E_c - E_a + kT \ln \left[\frac{\beta(N_d - N_a)}{N_a} \right], \quad n_0 \ll N_a \quad (322.11)$$

Such behavior is evident in curves (ii), (iii) and (iv) in part (b) of Fig. 32.10. It will be noted that the Fermi level starts from the

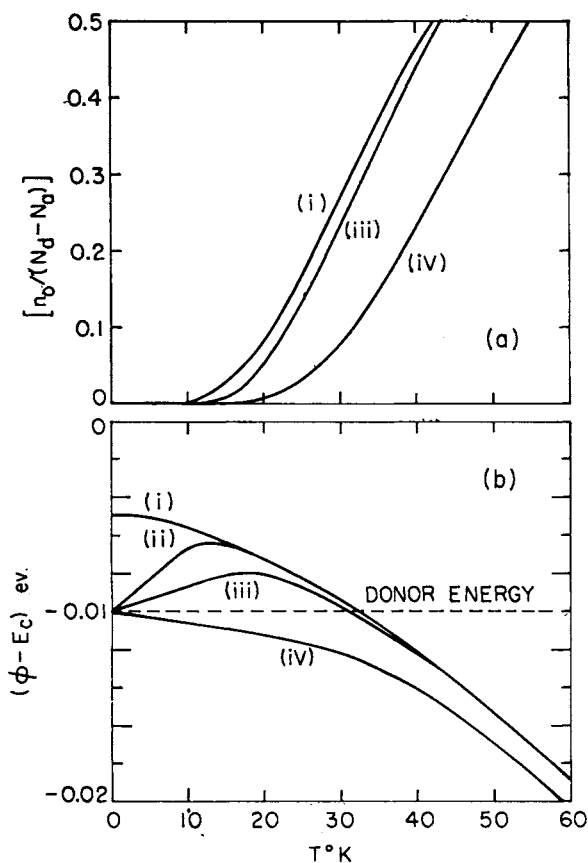


FIG. 32.10. The temperature dependence of $[n_0/(N_d - N_a)]$ and ϕ for the four numerical cases of Fig. 32.9. The curve for (ii) is not shown in part (a) since it lies extremely close to that of (i).

donor energy at $T = 0^\circ\text{K}$; this must always happen no matter how small the fractional compensation, since if there are *any* compensating impurities *at all*, the principal donors are less than 100% occupied at $T = 0$. This means that the Fermi level must differ from the donor level by a finite multiple of kT as $T \rightarrow 0$, and must coincide with $(E_c - E_a)$ at absolute zero itself.

Whether the Fermi level rises or falls as T increases from absolute zero depends on whether $[\beta(\mathcal{N}_a - \mathcal{N}_a)/\mathcal{N}_a]$ is larger or smaller than unity. In Fig. 32.10, the former is true of curves (ii) and (iii) and the latter of curve (iv). As T increases further, the Fermi level curves for samples of the same $(\mathcal{N}_a - \mathcal{N}_a)$ but different \mathcal{N}_a eventually coincide. This process has reached completion for the situations characterized by curves (i), (ii) and (iii) in Fig. 32.10, but the Fermi level for the most heavily compensated sample will not join with the others until a temperature of some 150°K is reached.

It is customary to plot $\ln(n_0)$ vs. $1/T$ in discussing the temperature dependence of carrier density, but other forms of presentation are sometimes useful. As an example, we may note from the curves in part (a) of Fig. 32.10 that the addition of compensating centers moves the curve of n_0 vs. T bodily to the right, by an amount which is approximately proportional to $[\mathcal{N}_a/(\mathcal{N}_a - \mathcal{N}_a)]$. The curve corresponding with case (ii) could not be drawn since it is almost indistinguishable from the zero-compensation curve (i) at all temperatures.

The reader may perhaps be wondering at this point how all the preceding information can be used in the analysis of experimental data, when the given information consists of values of n_0 for various temperatures. One of the most popular approaches starts from Eq. (322.1), rearranged to read

$$\frac{\exp(\eta)(n_0 + \mathcal{N}_a)}{(\mathcal{N}_a - \mathcal{N}_a - n_0)} = \beta \exp(-\epsilon_a) \quad (322.12)$$

When n_0 is not large enough at any temperature to provoke the complications of degeneracy, $\exp(\eta) = n_0/\mathcal{N}_c$, and

$$\frac{n_0(n_0 + \mathcal{N}_a)}{(\mathcal{N}_a - \mathcal{N}_a - n_0)} = \beta \mathcal{N}_c \exp(-E_a/kT) \quad (322.13)$$

The task now is to find values for \mathcal{N}_a and \mathcal{N}_a which will permit the two sides of (322.13) to be equal at all temperatures.

This does not require many experimental points if E_a , β , and m_c are known, but this is not generally the case. It is not possible to isolate β and m_c , only the combination $\beta m_c^{3/2}$ from use of (322.13). Thus if m_c is known, β can be determined, and vice versa. Suppose for instance that the effective mass for the conduction band is known. Referring to the curves of Fig. 32.9, $(N_a - N_a)$ can be estimated from the limit of n_0 in the exhaustion range. A reasonable approximation to E_a can be obtained from the slope of $\ln(n_0)$ vs. $1/T$ in the low temperature range, and (N_a/β) from the value of n_0 at any temperature in this range [using Eq. (322.8)]. The method of least squares can then be applied to Eq. (322.13) to find the values of the parameters which give the best fit over the entire temperature range.

As with all the preceding discussion, methods applicable to the relationships of donors to the conduction band can be adopted in their entirety for *p*-type semiconductors with majority acceptors above the valence band. The counterpart of Eq. (322.13) for a non-degenerate *p*-type semiconductor is

$$\frac{p_0(p_0 + N_a)}{(N_a - N_a - p_0)} = \left(\frac{N_v}{\beta} \right) \exp(-E_a/kT) \quad (322.14)$$

where the spin degeneracy β of the acceptors is in general larger than unity.

Fig. 32.11 provides an example of experimental results (derived from Hall coefficient measurements) fitted by means of Eq. (322.14). The sample is one of a series the present author once investigated (1959:6), composed of germanium doped with indium acceptors and partially compensated with antimony. For the combination of the two valence bands of germanium, the density-of-states effective mass is known (1955:9) to be $m_v = 0.37 m_0$. The results could be fitted in the most satisfactory fashion by using the numerical values noted in the figure caption for N_a , N_a , E_a and β . It will be observed that this fit is obtained with β equal to 4. The more conventional choice for simple acceptors of $\beta = 2$ (combined with a different value of N_a to satisfy the low temperature data) led to an appreciably less satisfactory fit over the range 20–50°K. That an acceptor should offer four states above the double valence band of germanium was in fact predicted by Kohn (1957:10).

Before congratulating ourselves prematurely that experimental data can be reliably handled using Eqs. (322.13) or (322.14), it must be

observed that these equations are not entirely correct. Next we have to see how they can be generalized to take account of excited states.

3.2.3 THE INFLUENCE OF EXCITED STATES

An isolated hydrogen atom consists of an electron moving under the influence of a proton. The ground state of this system, the $1s$ state, has a spin degeneracy of 2, and an energy of some 13.6 eV is required to ionize the atom. But there are also many possible excited states of this

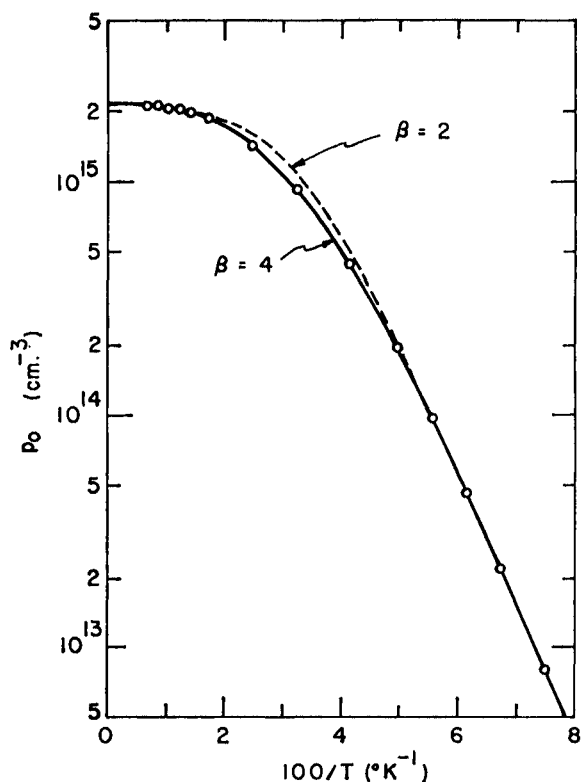


FIG. 32.11. The variation of p_0 with $1/T$ for germanium doped with indium and compensated with antimony (1959:6), fitted to satisfy Eq. (322.14) when an effective mass $m_e = 0.37 m_0$ is used for the valence band. The optimum fit is for $N_a = 2.66 \times 10^{15} \text{ cm}^{-3}$, $N_d = 0.48 \times 10^{15} \text{ cm}^{-3}$, $E_a = 0.01025 \text{ eV}$ and $\beta = 4$.

atom, the eight $2s$ and $2p$ states, the eighteen $3s$, $3p$ and $3d$ states, and so on. It is dangerous to press too far for the analogy between a hydrogen atom and a monovalent impurity center, but such analogies are useful in reminding us that a donor impurity is electrically neutral whether it has an electron bound in the ground state at $(E_c - E_d)$ or in an excited state rather closer to the conduction band. The importance of excited states was remarked by Shifrin (1944:1), but allowance for them in discussions of impurity level—conduction band equilibrium has only become popular in recent years (1955:20, 1956:14).

Rather than the model of Fig. 32.8, we should then accept Fig. 32.12 as the picture of a partly compensated n -type semiconductor. We

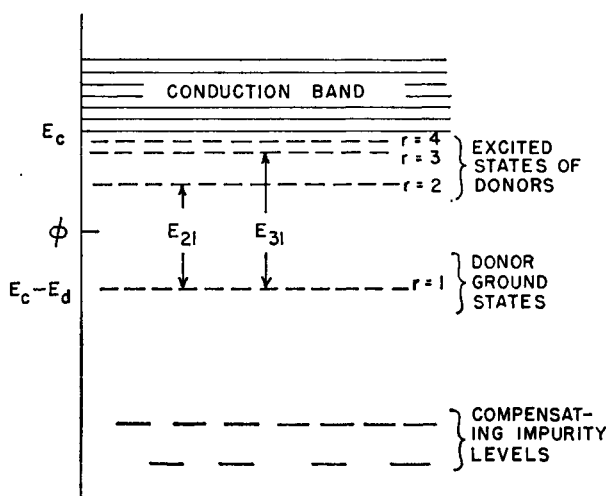


FIG. 32.12. Model of a partly compensated n -type semiconductor including the excited states of the principal donors.

denote by the symbol 1 the lowest, or ground level of the principal donors; this state has a spin degeneracy of β_1^{-1} . Groups of excited states in general have different degeneracy factors, say β_r^{-1} , and lie at energies $E_{r1} = kT\epsilon_{r1}$ above the ground states.

A donor is capable of binding an electron in one of the β_r^{-1} states at energy $(E_c - E_d + E_{r1})$ only if it does not have an electron already bound in any other state, either at the same level or a different one. If we denote the density of donors with electrons trapped at the r th level

by N_{dnr} , then

$$N_{dnr} = \frac{\{N_d - \sum_{s \neq r} N_{dns}\}}{1 + \beta_r \exp(-\eta - \epsilon_d + \epsilon_{r1})} \quad (323.1)$$

The summation extends over all levels except the r th. Now the numerator on the right can be written

$$\begin{aligned} \{N_d - \sum_{s \neq r} N_{dns}\} &= N_d + N_{dnr} - \sum_{\text{all } s} N_{dns} \\ &= (N_d - N_{dn}) + N_{dnr} \\ &= (N_{di} + N_{dnr}) \end{aligned} \quad (323.2)$$

Placing this expression in the numerator of Eq. (323.1), we arrive at a relationship which might have been expected all along for the ratio of N_{dnr} to the ionized donor density:

$$\frac{N_{dnr}}{N_{di}} = \beta_r^{-1} \exp(\eta + \epsilon_d - \epsilon_{r1}) \quad (323.3)$$

This is of course exactly the kind of relationship laid down in Eq. (311.1) when only the ground states were being considered.

The total density of neutral donors is given by summing N_{dnr} over all r :

$$N_{dn} = N_{di} \sum_{r=1}^{\infty} \beta_r^{-1} \exp(\eta + \epsilon_d - \epsilon_{r1}) \quad (323.4)$$

Since $(N_{dn} + N_{di})$ is just N_d itself, (Eq. 323.4) leads to a pair of equations

$$\left. \begin{aligned} N_{dn} &= \frac{N_d \sum_{r=1}^{\infty} \beta_r^{-1} \exp(\eta + \epsilon_d - \epsilon_{r1})}{1 + \sum_{r=1}^{\infty} \beta_r^{-1} \exp(\eta + \epsilon_d - \epsilon_{r1})} \\ N_{di} &= \frac{N_d}{1 + \sum_{r=1}^{\infty} \beta_r^{-1} \exp(\eta + \epsilon_d - \epsilon_{r1})} \end{aligned} \right\} \quad (323.5)$$

which are of completely general validity. Whatever the energies and spin degeneracies of excited states, their influence can be calculated.

Let us return now to consider Fig. 32.12, and the distribution of electrons between the conduction band, all states of the donors, and any compensating impurities. N_{di} must (in the absence of any free valence band holes) be equal to the sum of n_0 and N_a . Then

$$n_0 = N_c \mathcal{F}_{1/2}(\eta) = \frac{N_d}{1 + \sum_{r=1}^{\infty} \beta_r^{-1} \exp(\eta + \epsilon_d - \epsilon_{r1})} - N_a \quad (323.6)$$

When free electrons are not unduly numerous, so that the Fermi level lies below the conduction band, the non-degenerate approximation $\exp(\eta) \approx (n_0/N_c)$ can be substituted into the denominator of Eq. (323.6):

$$n_0 + N_a = \frac{N_d}{1 + (n_0/N_c) \sum_{r=1}^{\infty} \beta_r^{-1} \exp(\epsilon_d - \epsilon_{r1})} \quad (323.7)$$

Rearranging Eq. (323.7) in the manner employed previously, the quadratic equation for n_0 can be given in the form

$$\begin{aligned} \frac{n_0(n_0 + N_a)}{(N_d - N_a - n_0)} &= \frac{N_c}{\sum_{r=1}^{\infty} \beta_r^{-1} \exp(\epsilon_d - \epsilon_{r1})} \\ &= \frac{(\beta_1 N_c) \exp(-\epsilon_d)}{1 + \sum_{r=2}^{\infty} (\beta_1/\beta_r) \exp(-\epsilon_{r1})} \end{aligned} \quad (323.8)$$

As with Eq. (322.13), from which Eq. (323.8) differs only in the additional presence of the summation over excited states, the right-hand side depends on the character of the impurities and on the temperature, but not on the donor or acceptor densities. This makes the form of Eqs. (322.13) and (323.8) convenient for the treatment of experimental data when it is known that the conduction band density is small enough at all temperatures for $N_c \exp(\eta)$ to be a good approximation for $\mathcal{F}_{1/2}(\eta)$.

For compactness, we may denote

$$\sum_{r=2}^{\infty} (\beta_1/\beta_r) \exp(-\epsilon_{r1})$$

by the symbol F , so that

$$\frac{n_0(n_0 + \mathcal{N}_a)}{(\mathcal{N}_a - \mathcal{N}_a - n_0)} = \frac{\beta_1 \mathcal{N}_c \exp(-\epsilon_a)}{1 + F} \quad (323.9)$$

Since F is a function of temperature, the temperature dependence of n_0 will be slightly different from that predicted by (322.13), and if the data are analyzed on the basis of Eq. (322.13) rather than Eq. (323.9), it may look as though the energy E_a were slightly temperature-dependent (1957:16). It will depend to a major extent on the distribution of excited states in energy whether their occupancy can affect the equilibrium of Eq. (323.9). When there are some excited states with energies rather close to that of the ground states, F may become comparable with or even larger than unity in certain temperature ranges.

Certainly at low enough temperatures, $F \ll 1$, and Eqs. (322.13) and (323.9) will coincide. All neutral donors then hold an electron in the lowest possible energy state. Similarly at very high temperatures the Fermi level is well below all states of the donors, and $n_0 \approx (\mathcal{N}_a - \mathcal{N}_a)$. It is for intermediate temperatures that the value of F can have a maximum effect on n_0 .

Let us suppose initially that an electron bound by a donor ion can be described perfectly by Eq. (145.4), the effective mass equation. Temporarily neglecting all the complications which have been superimposed on the simple Bohr model of a hydrogen atom, we may suppose that the ground and excited states of a donor resemble those of the original Bohr atom, renormalized by the effective mass and dielectric constant. This is the model in which s, p, d , etc., states of a given shell all have the same energy. Then $(\beta_1/\beta_r) = r^2$ and $E_{r1} = E_a(1 - r^{-2})$. In terms of such a simple model,

$$F = 4 \exp(-3\epsilon_a/4) + 9 \exp(-8\epsilon_a/9) + 16 \exp(-15\epsilon_a/16) + \dots \quad (323.10)$$

As Shifrin (1944:1) remarks, it is only necessary to consider the first few terms of such a series since for a finite donor density the wave-functions for the higher excited states will overlap quite strongly and these states will form part of the conduction band.

As an illustrative numerical example of the effect produced by excited state occupancy when F has the form (323.10), consider a semiconductor with donor levels $E_a = 0.01$ eV below the conduction band when $(\mathcal{N}_a - \mathcal{N}_a) = 10\mathcal{N}_a = 10^{16}$ cm⁻³ and $m_c = 0.25 m_0$. This was the example previously used as curve (iii) in Figs. 32.9 and 32.10. The

broken curve in Fig. 32.13 is identical with curve (iii) of Fig. 32.9, and the solid line shows the variation of n_0 with $1/T$ when excited states are taken into account by using Eqs. (323.9) and (323.10). For this example, the excited states have quite a considerable effect between 20°K and the exhaustion range. It would be possible crudely to fit the behavior of the solid curve in Fig. 32.13 by using Eq. (322.13), but only with seriously erroneous values of β and N_a .

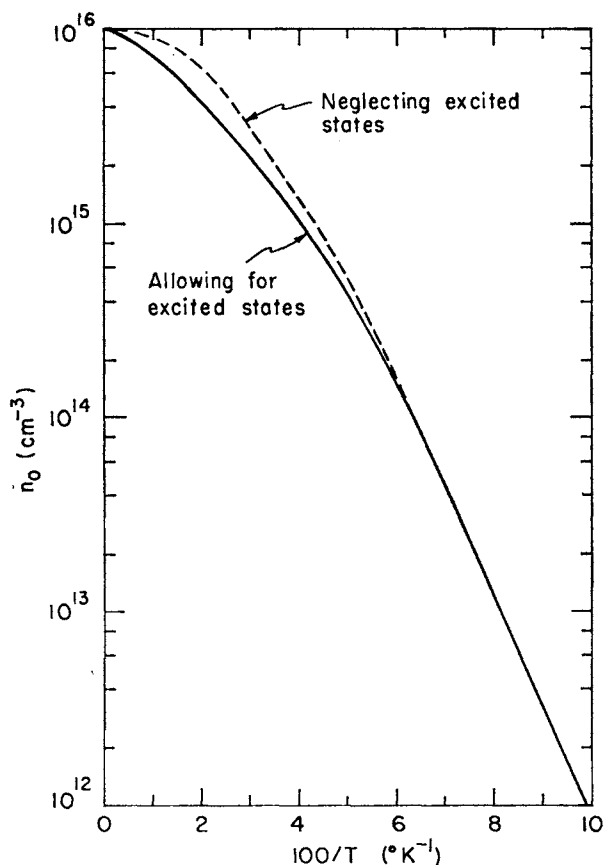


FIG. 32.13. Semilogarithmic plot of n_0 vs. $1/T$ for an n -type model with $N_d = 1.1 \times 10^{16} \text{ cm}^{-3}$, $N_a = 10^{15} \text{ cm}^{-3}$, $E_d = 0.01 \text{ eV}$, $m_e = 0.25m_0$. The broken curve is identical with curve (iii) of Fig. 32.9. The solid curve shows the change resulting from taking into account occupancy of hydrogenic excited states.

We are then immediately faced with the question: how trustworthy is the analysis applied to the experimental data in Fig. 32.11? For excited states of the indium acceptors were not allowed for in those calculations. This is the kind of question which must be asked on an individual basis for each type of impurity in each semiconductor, since the effective mass equation has to be modified in accordance with the band structure peculiarities of a semiconductor. The energy spacings and degeneracies of excited states usually bear little relation to a hydrogenic spectrum. In the case of Group III acceptors in germanium, the calculations of Schechter (1955:23, 1956:20) suggest that the $2p$ states are split into several components, most of which are too close to the valence band to affect the carrier density-temperature relationship to any extent. Incidentally, the designation of these states as $2p$ is rather confusing, since there is no direct connection with the hydrogen energy scheme.

At any rate, for p -type germanium with Group III acceptors, it would seem to be justifiable to ignore the excited states. Similar simplifications are permissible for silicon containing these same types of acceptor. Infrared absorption measurements (e.g. 1956:17, 1958:20) and the calculations of Kohn and Luttinger (1955:22, 1957:10) indicate that the excited states are not close enough to the ground state to have much effect. In the extreme case of indium acceptors in silicon, the optical measurements of Newman (1956:19) indicate that the ground state binding energy is ten times larger than that of the first excited state.

For semiconductors such as germanium and silicon, description of acceptor states in terms of an effective mass equation is complicated by the existence of two valence bands, the light- and heavy-hole bands, degenerate at $\mathbf{k} = 0$.

Complexities of a different kind occur in considering donor states for these semiconductors, because the conduction band has multiple minima and the impurity wave functions are appropriate combinations of contributions from each minimum (1954:13, 1955:17, 1957:10). One possibility which emerges from this theoretical work is that the $1s$ states—the deepest lying states of the system—can themselves be split into two factions. This is the topic of the next sub-section.

3.2.4 IMPURITY GROUND STATE SPLIT IN THE CRYSTAL FIELD

It is not surprising that theoretical studies of donor states in semiconductors should have concentrated on silicon and germanium, since

the conduction band shapes of these semiconductors are known so well. Work on silicon is particularly profitable because the energies of optical transitions fall in a convenient spectral range. The left column in Fig. 32.14 shows the spacings of levels for phosphorus donors determined

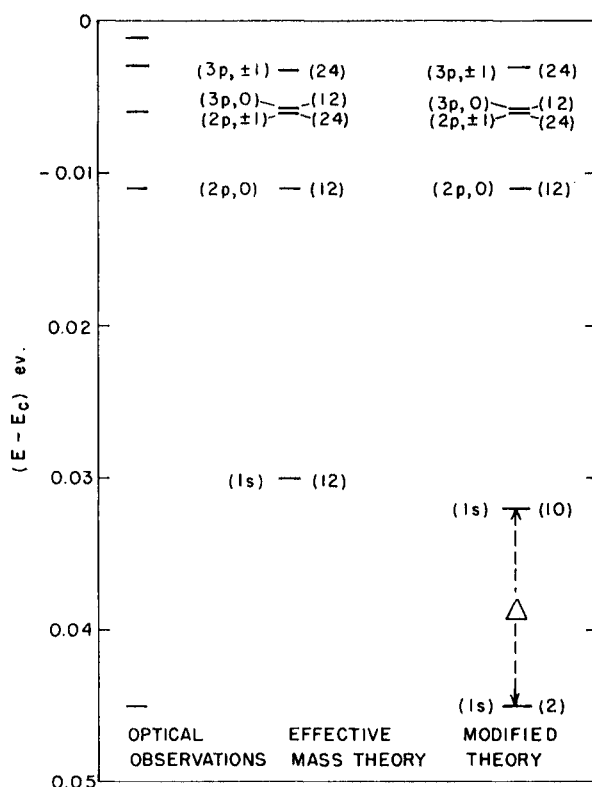


FIG. 32.14. Comparison of optical absorption experimental results (1956:18) with theoretical term schemes for phosphorus donors in silicon. The figure in parentheses at the right of each term is β^{-1} , the total state degeneracy (including spin).

optically by Picus *et al.* (1956:18). Comparison with the center column shows that the excited states agree well with the term scheme obtained from effective mass theory. But theory predicts a ground state much shallower than the energy found in practice for any Group V donor.

This discrepancy is now believed to be a result of the composition of the ground state wave-functions, influenced by the six equivalent conduction band minima. The $1s$ level is compounded of six states, each with a spin degeneracy of 2, or twelve states in all. Two of these states have wave-functions which do not vanish at the donor nucleus (completely symmetrical states) whereas the wave-functions of the other ten do vanish at this point. The two completely symmetrical states of the representation A_1 will be affected much more by a breakdown of the effective mass equation (145.4) than the others, and accordingly lie at a considerably lower energy than the remaining $1s$ states. This revised theoretical picture is shown at the right in Fig. 32.14. (Actually, group theory shows that the upper ten $1s$ states may be split again into $4+6$ states, but such splitting will presumably be rather small compared with the energy difference Δ between the two lowest $1s$ states and their erstwhile companions.)

Optical identification of the upper $1s$ states will be very difficult. For transitions between the two sets of $1s$ states are obviously forbidden. At the low temperatures required to obtain narrow optical line widths, it will be almost impossible to maintain enough electrons in the upper levels to make a transition into a $2p$ or $3p$ state detectable.

For this reason, Long and Myers (1959:13) used an electrical method in measuring the energy difference Δ . As illustrated in Fig. 32.15, they measured the temperature dependence of free electron density (by deduction from the Hall coefficient), and found the value of Δ which permitted the optimum fit to

$$\frac{n_0(n_0 + N_a)}{(N_a - N_a - n_0)} = \frac{N_c \exp(-E_d/kT)}{2 + 10 \exp(-\Delta/kT) + \sum_{r=2}^{\infty} \beta_r^{-1} \exp(-E_{r1}/kT)} \quad (324.1)$$

which is an obvious adaptation of Eq. (323.8). For phosphorus donors in silicon they determined Δ as (0.010 ± 0.0002) eV. Similar procedures should be applicable to other types of donor.

By now it should be discouragingly clear that the determination of densities and energies of bound states is by no means a simple matter. Unless a great deal of information is available about the band structure and other features of a semiconductor, we can not know whether a simple analysis is permissible. Yet if a simple analysis is carried out—say on the lines of Eq. (322.13)—when it is inappropriate, the numerical values deduced may be seriously in error. With most semiconductors,

our state of ignorance at the time of writing is so profound that it would be unwarrantably pretentious to attempt more than a simple semi-quantitative analysis. Accordingly, while we have considered excited

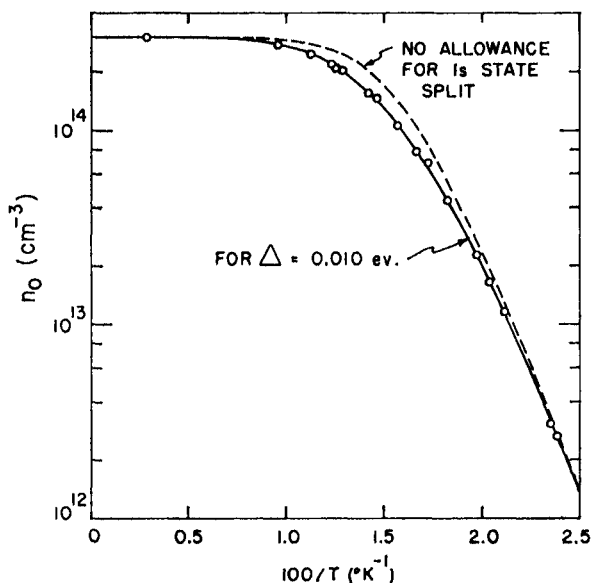


FIG. 32.15. Variation of n_0 with $1/T$ for silicon containing partly compensated phosphorus donors in silicon, according to measurements of Long and Myers (1959:13). Calculated curves are for $N_d = 6.9 \times 10^{14} \text{ cm}^{-3}$, $N_a = 3.8 \times 10^{14} \text{ cm}^{-3}$, $E_d = 0.0435 \text{ eV}$.

states and split ground states with due seriousness in these two subsections, it will not be fruitful to consider them in connection with all the remaining models of this chapter.

3.2.5 IMPURITY STATES SPLIT BY ANISOTROPIC ELASTIC STRAIN

The complexities of impurity wave-functions and energy levels considered so far result only from the complicated form of the periodic potential in any real solid. It should now be noted—rather briefly—that deformation of the crystal lattice will have further effects on impurity states.

A simple isotropic elastic strain should not affect the structure of impurity states in most materials, for this should be equivalent to the effects of thermal expansion and contraction on the lattice. (Even this will not be true for an anisotropic semiconductor such as tellurium, which has a negative compressibility along the c -axis (1954:14), so that hydrostatic pressure alters the bond angles.) But in any semiconductor, an anisotropic elastic strain will perturb the edges of the bands (1950:7). This in turn will modify impurity state wave-functions. That impurity states would be affected by strain was first pointed out

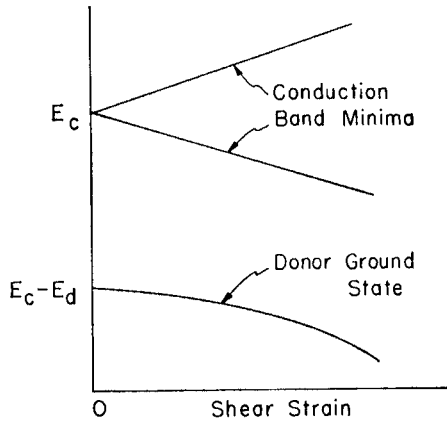


FIG. 32.16. Suggested variation with shear strain of the conduction band edges and donor ground state energy for Group V donors in silicon, according to Kohn (1957:10).

by Price (1956:21) and the implications for donor and acceptor states in Ge and Si are discussed by Kohn (1957:10).

Kohn remarks that a shear strain will split each conduction band in Si into two branches, with energy linearly dependent on the magnitude of the shear. On the other hand, the first-order shift of the donor ground state vanishes, and the main effect is the second-order one, as sketched in Fig. 32.16.

Measurements of the conductivity of n -type germanium at low temperatures under uniaxial strain have been reported by Fritzsche (1959:14) (1960:6). With antimony doped germanium, the principal effect is the inhibition of impurity conduction parallel to the strain. Strikingly different behavior is exhibited with arsenic donors (1961:2) due to the much larger valley-orbit splitting.

3.2.6 EFFECT OF A MAGNETIC FIELD ON IMPURITY STATES

In previous chapters it has been seen that the application of a magnetic field splits the conduction and valence bands into series of sub-bands founded on Landau levels. By analogy, it may be expected that a magnetic field will also influence the energies of bound impurity states.

Yafet *et al.* (1956:22) have calculated this influence for the model of a hydrogenic ground state, when the effective mass m_c is isotropic.

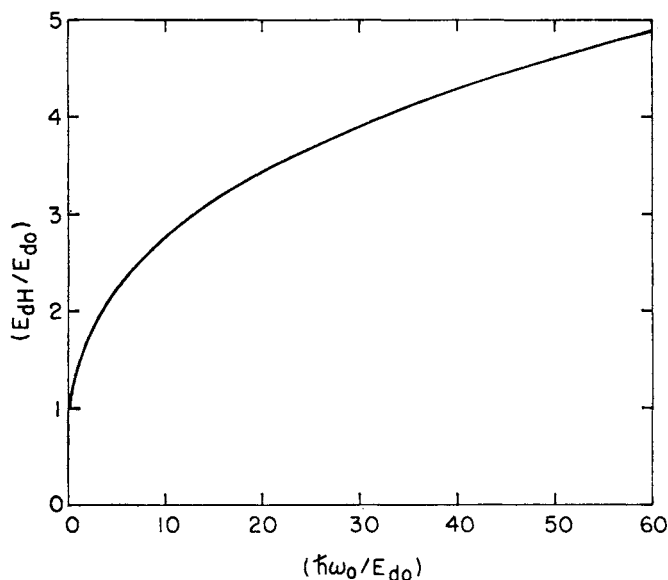


FIG. 32.17. Variation of donor ionization energy with magnetic field strength, after the calculations of Yafet *et al.* (1956:22).

As we already know, the lowest conduction states are at an energy $\frac{1}{2}\hbar\omega_0$ above E_c in the presence of a magnetic field (see Sub-sections 1.4.6 and 2.2.2). The donor ground state also moves towards higher energy with magnetic field, but at a rather slower rate, so that the effective donor ionization energy E_{dH} is larger than its value E_{d0} for zero field. The essential result of the variational calculation of Yafet *et al.* is shown in Fig. 32.17. It will be noted that E_{dH} changes with field most rapidly

when the dimensionless parameter $(\hbar\omega_0/E_{d0})$ is still small; but appreciable increases in E_d still correspond with rather large magnetic fields for most materials. As an example, with group V donors in germanium, $\hbar\omega_0$ does not exceed E_{d0} for magnetic fields less than 10^5 oersteds.

The model used by Yafet *et al.* indicates that the impurity wave function becomes restricted to a smaller volume when a magnetic field is applied. This shrinking occurs mainly in directions normal to the field, but also to some extent along the field direction. The overall effect is that of transforming a rather large spherically symmetrical wave-function into a very small cigar shape. Such constriction of the orbit for a bound electron necessarily means that the effective mass approximation must become less tenable as the field increases; thus the high field region of Fig. 32.17 should be regarded as highly tentative.

Experiments are reported by Keyes and Sladek (1956:23) on the variation of n_0 with magnetic field for n -type InSb, in support of the calculations just mentioned. The carrier density certainly decreased with increasing H , though this was probably due in large measure to the magnetic break-up of a nascent impurity band (Section 3.5).

3.2.7 SOME COMMENTS IN SUMMARY

At this point we should take stock of the degree of complexity likely in an actual semiconductor. Except for material under rather severe strain or in a very large magnetic field, the complications of Sub-sections 3.2.5 and 3.2.6 can usually be neglected. Whether the statistical description is complicated by the influence of excited states or a split ground level as in Sub-sections 3.2.3 and 3.2.4 depends very much on the individual impurity and semiconductor; however, there is reason to believe that many systems will be essentially free from such complications.

This brings us to the discussion of Sub-sections 3.2.1 and 3.2.2. The apparently unrealistic case of Sub-section 3.2.1 was discussed in some detail for several reasons. One good reason is that it provides the simplest introduction to the ideas involved. Another is that some of the information can be applied to partly compensated semiconductors.

For instance, Figs. 32.5 and 32.6 indicate the maximum value of η and the temperature for which it occurs as functions of the parameter ζ . Now in a partly compensated semiconductor, the behavior of η is

quite different at very low temperatures; but by the time ionization has proceeded far enough to take η through a maximum, the Fermi level for a given $(N_d - N_a)$ is almost independent of the actual degree of compensation (see Fig. 32.10). Thus even for a partly compensated semiconductor, η_m and T_m can be found with reasonable accuracy using Figs. 32.5 and 32.6, substituting $(N_d - N_a)$ into Z . This gives a good indication of the range of η which will be encountered and shows whether a solution must be effected with Eq. (322.2) or with the simpler non-degenerate forms of Eq. (322.6) and (322.12).

The uncertain effects of compensation discourage the use of other simple graphical aids for determining n_0 and ϕ as functions of temperature when N_d , E_d , etc., are known. In a previous discussion of the interrelationships of carrier populations in bands and impurity levels (1952:3) the present author used a number of graphical aids, based on the uses to which Eq. (321.4) had been put by Landsberg *et al.* (1951:5) and by Lehovc and Kedesdy (1951:15). But such curves must be replaced by families of curves for different degrees of compensation, and by families of families if excited states are to be allowed for. This removes the essential basis of graphical methods, that of simplifying the task of calculation.

3.3 SEMICONDUCTORS DOMINATED BY SEVERAL LOCALIZED LEVELS

Throughout Section 3.2 we considered the dynamics of situations dominated by a single species of monovalent donor. Other impurity levels were acknowledged only for their effect on compensation. Such a viewpoint is justifiable when the Fermi level is either far above or far below all other kinds of impurity level; but these conditions cannot necessarily be maintained over a wide range of temperature. Now it is necessary to consider more general models.

3.3.1 SEVERAL INDEPENDENT TYPES OF MONOVALENT DONOR

The simplest kind of generalization is indicated by the model of Fig. 33.1. Two entirely independent species of monovalent donor are supposed, which present levels for electron occupancy at characteristic energies below the conduction band. There must also be some assorted

compensating acceptor levels, which accommodate N_a trapped electrons per unit volume. Only $(N_{d1} + N_{d2} - N_a)$ electrons remain for distribution between the two sets of donors and the conduction band.

Two completely different forms of behavior will exist depending on whether N_a is larger or smaller than N_{d1} . If $N_a > N_{d1}$, the upper donor levels will be completely denuded of electrons at 0°K and the Fermi level will be locked to the energy $(E_c - E_{d2})$. As temperature rises some electrons will be excited to the conduction band and ϕ will readjust itself in an appropriate manner. At sufficiently high temperatures, ϕ will be several kT lower than $(E_c - E_{d2})$, and n_0 will approach

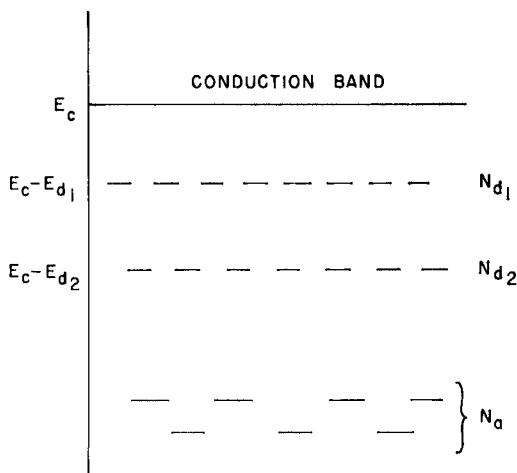


FIG. 33.1. Model of an n -type semiconductor containing two independent kinds of donors at different energies, and assorted low lying compensating centers. The temperature variation of n_0 depends on whether N_a is larger than or smaller than N_{d1} .

its limiting value of $(N_{d1} + N_{d2} - N_a)$. Note that unless the two donor binding energies are very similar, the Fermi level will always be many kT below the upper donor levels, and these will remain empty at all temperatures. This situation is really entirely equivalent to the one we have considered throughout Section 3.2, for a semiconductor dominated by the *single species* of N_{d2} donors with an *effective* compensating density of $(N_a - N_{d1})$ states.

On the other hand, if $N_a < N_{d1}$, at absolute zero the lower donors will be completely full and the upper donors will contain $(N_{d1} - N_a)$

electrons. The Fermi level at 0°K then coincides with the energy $(E_c - E_{d1})$. As temperatures rises, the upper donors offer their electrons to the conduction band, and ϕ must move below $(E_c - E_{d1})$ as this supply becomes exhausted. On further warming, no appreciable increase of n_0 is possible until ϕ has dropped almost to the energy $(E_c - E_{d2})$; the second set of donors then begins to lose electrons and at

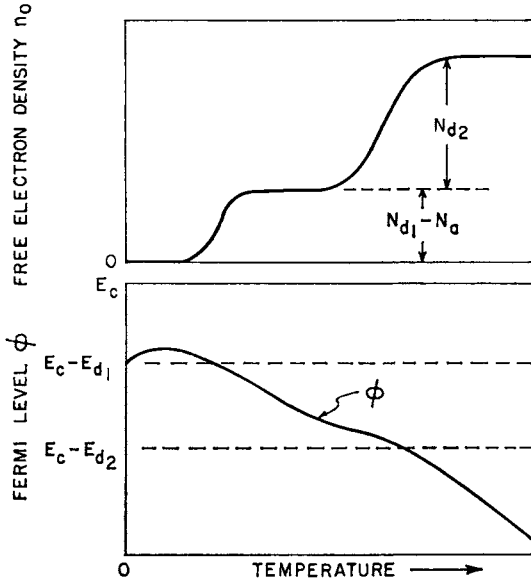


FIG. 33.2. Schematic change of n_0 and ϕ with temperature when the upper set of two donor species is incompletely compensated.

sufficiently high temperatures n_0 saturates at $(N_{d1} + N_{d2} - N_a)$ while ϕ drops more rapidly in accordance with

$$\phi = E_c - kT \ln \left[\frac{N_c}{N_{d1} + N_{d2} - N_a} \right] \quad \text{high } T \quad (331.1)$$

This kind of behavior is shown schematically in Fig. 33.2. The extension to three or more independent kinds of impurity is obvious.

The preceding description can be readily expressed in mathematical form. Suppose that there are M species of monovalent donor impurities, and that the j th class is of density N_{dj} , ground state binding energy

E_{aj} , etc. The number of ionized donors in this class may, from Eq. (323.5), be written as

$$(N_{aj})_{\text{ion}} = \frac{N_{aj}}{1 + \beta_j^{-1}[1 + F_j] \exp(\eta + \epsilon_{aj})} \quad (331.2)$$

where the quantity F_j indicates the influence of any excited states, as previously discussed. The sum of all ionized donors is equal to the conduction band electron density plus all electrons in compensating acceptors. Thus

$$[N_a + n_0] = [N_a + N_c \mathcal{F}_{1/2}(\eta)] = \sum_{j=1}^M \frac{N_{aj}}{1 + \beta_j^{-1}[1 + F_j] \exp(\eta + \epsilon_{aj})} \quad (331.3)$$

When the Fermi level is at least several kT below the conduction band, so that $\mathcal{F}_{1/2}(\eta)$ can be replaced by $\exp(\eta)$, Eq. (331.3) is an equation of order $(M+1)$ in either n_0 or $\exp(\eta)$.

Such an equation is by no means simple to solve in the general case, but there is one saving feature. For any temperature we choose to specify, the ground state energies of *most* classes of impurity will be either *well above* or *well below* the Fermi level; in the former case $(N_{aj})_{\text{ion}} = N_{aj}$ and in the latter $(N_{aj})_{\text{ion}} = 0$. Thus there will not usually be more than one or two classes of donor for which $(N_{aj})_{\text{ion}}$ lies between its extreme values, and Eq. (331.3) will be either a quadratic or a cubic in $\exp(\eta)$. For the example we sketched in Fig. 33.2, the two levels were so far apart in energy that ionization of one set of levels reached completion before it became at all appreciable in the other set. Separate quadratic equations can then be obtained from Eq. (331.3) to cover the various temperature ranges. Solution of a cubic equation is only necessary when two energy states are rather close in energy, and even then only for the temperature range in which ϕ progresses through those levels.

3.3.2 ELECTRON DISTRIBUTION OVER A SET OF MULTIVALENT FLAWS

It was remarked in Section 1.6 that flaws often have the capacity to donate or accept more than one electron. When a set of multivalent centers is the dominant one in determining the electronic equilibrium, it is necessary to give a generalized discussion of the distribution over available states.

This subject has received attention in recent years from Landsberg (1956:14, 1958:10), and from Shockley and Last (1957:21), while Champness (1956:8) has discussed the particular case of divalent donors. The equations can be formulated from a direct statistical approach, by the use of the grand canonical ensemble, or from free energy considerations. The reader is referred to the literature references just cited for elaborations of the various points of view.

Some of the attributes of this problem are indicated for divalent donors by Fig. 33.3. When the Fermi level is sufficiently high, as in

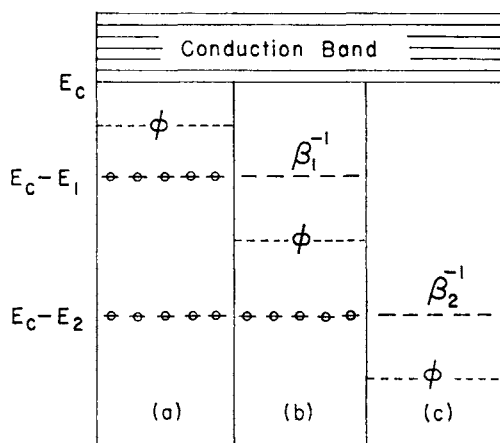


FIG. 33.3. The conventional way of indicating the energy states for divalent donors, indicating (a) the neutral condition, (b) the singly ionized condition, (c) the doubly ionized condition.

part (a) of the figure, each donor is electrically neutral, and holds two fairly weakly bound electrons with a *total* binding energy of $(E_1 + E_2)$. Now the two electrons will probably have rather similar orbits and binding energies; the significance of marking the two separate energies E_1 and E_2 is that the *center* must receive an energy E_1 to ionize it singly, and in the singly ionized state the second electron is held with energy E_2 . This is the equilibrium condition when $E_1 < (E_c - \phi) < E_2$ as in part (b) of Fig. 33.3. A singly ionized divalent donor offers empty states at $(E_c - E_1)$ with a statistical weight of β_1^{-1} .

When the Fermi level is lower again, the donors are doubly ionized and β_2^{-1} states per donor are offered at the energy $(E_c - E_2)$. This

situation is shown in part (c) of Fig. 33.3. Note that no available states are now shown at the upper level, since these states do not come into existence until a single electron has been recaptured. As soon as this first electron is captured, a second electron can be received only into one of the upper states which develop at this point.

In general, for each stage of ionization there will be not only a ground state but also a series of excited states into which an electron can be received. However, as Shockley and Last (1957:21) point out, it is not proper to construe the placement of an electron at the energy $(E_c - E_1)$ as a valid excited state of the singly ionized configuration. This is a meaningless state of affairs; the excited states for the condition of single ionization will usually be quite different.

Landsberg (1956:14, 1958:10) has provided a very compact derivation for the total number of electrons *retained* by a set of N centers, through differentiation of the grand partition function for the system, including allowance for excited states (which we shall not consider further). For our purposes, mindful of the amphoteric centers to be discussed in the next sub-section, we prefer to consider the numbers N_j of centers each of which has already *lost* j electrons. Thus N_0 centers are neutral, N_1 have lost one electron and present β_1^{-1} levels at energy $(E_c - E_1)$, and so on. If the centers are of valency M with respect to the crystal, then

$$\frac{N_j}{N} = \frac{\beta_j \exp(-\epsilon_j - j\eta)}{1 + \sum_{k=1}^M \beta_k \exp(-\epsilon_k - k\eta)} \quad (332.1)$$

The total number of electrons lost from the impurities is

$$\sum_{j=1}^M j N_j$$

and this must equal the number in the conduction band and in any compensating levels. Thus

$$n_0 = N_c \mathcal{F}_{1/2}(\eta) = -N_a + \sum_{j=1}^M j N_j \quad (332.2)$$

When conditions are non-degenerate, Eqs. (332.1) and (332.2) lead to an equation of order $(M+1)$ for n_0 or $\exp(\eta)$. Thus Champness

(1956:8) obtained a cubic equation for the divalent case, which he solved for a numerical example of zero compensation.

This problem is not too dissimilar from that of several kinds of independent donors present in identical amounts [see Eq. (331.3)] when the energy levels are well separated. Shockley and Last (1957:21) have discussed the problem of two separate donors treated as a composite flaw.

As with any other of the systems considered so far, the extent of compensation will determine which level of a multivalent donor will house the Fermi level at low temperatures. For divalent donors, if the amount

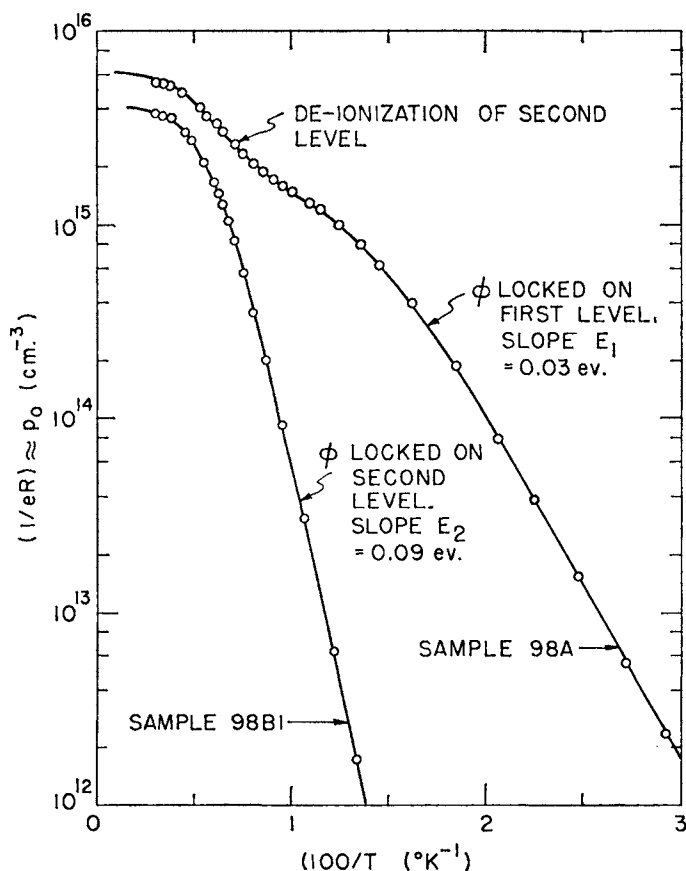


FIG. 33.4. Variation of free hole density with temperature for zinc-doped germanium crystals of Woodbury and Tyler (1956:24).

of compensation is rather small, the Fermi level will lock on to $(E_c - E_1)$ at low temperatures and

$$n_0 \approx \beta_1 N_c \left(\frac{N - N_a}{N_a} \right) \exp(-E_1/kT) \quad \begin{cases} \text{low temp.} \\ N_a < N \end{cases} \quad (332.3)$$

As temperature rises, the first stage of ionization becomes complete, and $n_0 \approx (N - N_a)$. Further heating drives the Fermi level down towards and beyond the second ionization energy, and an additional N electrons are added to the conduction band in the course of this process.

On the other hand, if $N < N_a < 2N$, the saturation free electron density is $(2N - N_a)$, which is smaller than N_a . A drop in n_0 occurs as soon as the temperature is low enough for ϕ to approach $(E_c - E_2)$, and on further cooling the Fermi level locks to this energy and provides an $\exp(-E_2/kT)$ temperature dependence for n_0 .

These two forms of behavior are illustrated in Fig. 33.4, which is based on experimental data of Tyler and Woodbury (1956:24). The impurities in these samples are actually acceptors (zinc in germanium), but as has been emphasized previously, donor and acceptor situations are always formally equivalent. For sample 98A, cooling first changes all the zinc acceptors from double to single ionization, then the Fermi level approaches and locks on to the first ionization level. The second sample evidently has a donor density slightly in excess of the zinc concentration, and the low-temperature distribution still has many doubly charged zinc atoms.

3.3.3 AMPHOTERIC IMPURITIES

Some unusual kinds of impurity can act as either donor or acceptor, depending on the Fermi level. [This is true, for instance, of gold in germanium (1956:25).] By an amphoteric impurity we mean one which is most likely to be electrically neutral when the Fermi level lies at some energy E_0 within the intrinsic gap. At some lower Fermi energy the impurity is likely to lose an electron—the attribute of a donor. Perhaps more stages of donor ionization will occur before ϕ drops as low as $(E_c - E_i)$. But also, the center is likely to *add* one or more electrons when ϕ has an energy higher than E_0 yet still below E_c ; this constitutes single or multiple acceptor action.

For gold in germanium, a single donor level is rather close to the valence band and there are three successive stages of acceptor ionization

before the Fermi level enters the conduction band. Fig. 33.5 indicates the intrinsic gap width of germanium at room temperature and the positions of the four levels, after Dunlap (1957:28). The right-hand

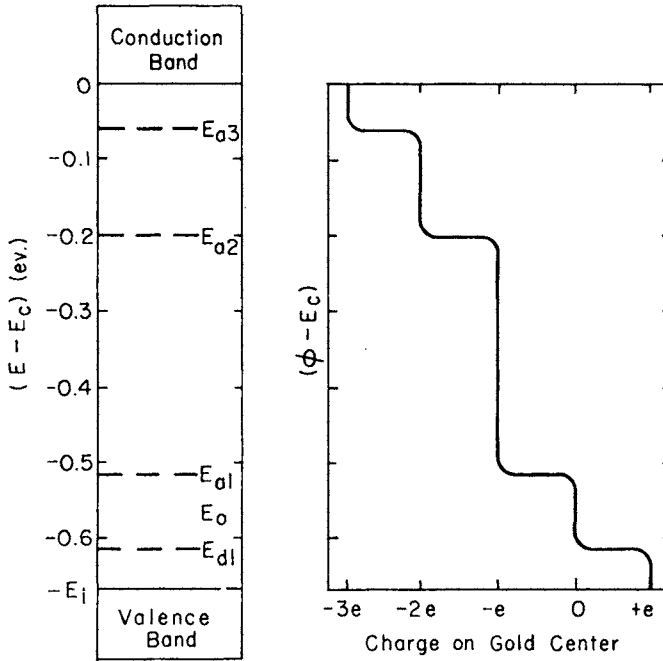


FIG. 33.5. The energy levels of the amphoteric impurity gold in germanium showing how the charge per atom will depend on the Fermi level at low temperatures.

portion of this figure shows how the charge per gold center should vary with the Fermi level at a rather low temperature (such that kT is small compared with the spacing of levels).

The statistical approach to amphoteric impurities is the same as for more conventional multivalent impurities, since it is still legitimate to use the result

$$\frac{N_{r-1}}{N_r} = \frac{\beta_r}{\beta_{r-1}} \exp(\eta + \epsilon_r) \quad (333.1)$$

[compare for instance with Eq. (311.1)] where now r may be either positive, zero, or negative.

3.4 THE INFLUENCE OF LATTICE DEFECTS

In addition to the impurity levels provided by foreign atoms, there are often localized levels provided in semiconductors by lattice defects such as vacancies and interstitials. This section comments very briefly on two facets of the problem.

3.4.1 NON-STOICHIOMETRIC COMPOUNDS

When a semiconducting compound is maintained at a high temperature, there will be a tendency for the ratio of the elements present to change from that of the pure stoichiometric compound. The amount by which it changes depends on a number of parameters, including the pressure at which vapor of the most volatile component is maintained around the sample, and the temperature of the treatment. It was indicated in 1.6.2 that a very complete description of this electronic-ionic equilibration is given by Kröger and Vink (1956:27). See also the discussion by Brebrick (1961:3).

At any rate, when a compound is rapidly cooled to a much lower temperature, a number of lattice defects are "frozen-in" (1939:1) and their density will have an influence on the low-temperature electronic equilibrium. As one of the two principal choices, we may suppose these defects to be lattice vacancies and that their habit is of acceptors.

It should perhaps be remarked that vacancies can also be introduced into an elemental semiconductor such as germanium by strenuous heat treatment, and frozen-in by sudden cooling. The true vacancy effect in germanium is very small and is often masked by the influence of chemical contaminants such as copper (1952:5, 1952:6); but by careful exclusion of contaminants, effects which are probably due to vacancy acceptors can be seen (1953:11, 1953:12). The vacancy concentrations which can be produced in compounds such as oxides or sulphides are frequently several orders of magnitude larger.

If the semiconductor contains N_d monovalent chemical donors at an energy $(E_c - E_d)$, and is now confronted with N_A vacancy acceptors at energy $(E_c - E_A)$, the electronic equilibrium is described by

$$[n_0 - p_0] = (N_d)_{\text{ion}} - (N_A)_{\text{ion}} \quad (341.1)$$

The left side of the equation includes p_0 since the vacancy levels may well be closer to the valence band than the conduction band. Assuming

that the Fermi level will not be closer than $2kT$ from either band, we have

$$[N_c \exp(\eta) - N_v \exp(-\eta - \epsilon_i)] = \frac{N_d}{1 + \beta_d^{-1} \exp(\eta + \epsilon_d)} - \frac{N_A}{1 + \beta_A \exp(-\eta - \epsilon_A)} \quad (341.2)$$

This is a quartic in $\exp(\eta)$, but it can always be reduced to an equation of lower order. As was seen in the previous discussions of compensation, if the donors lie in the upper half of the gap, p_0 can be dropped and

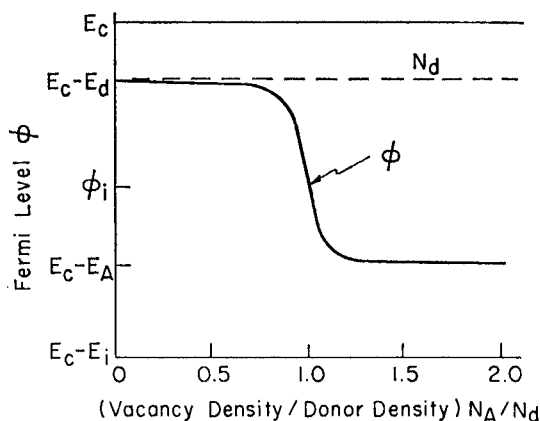


FIG. 34.1. Variation of low temperature Fermi level with the density of "frozen-in" vacancy acceptors if the semiconductor originally has N_d chemical donors.

$(N_A)_{ion}$ simplified to just N_A when the vacancy density is not as large as N_d . When the non-stoichiometry is sufficient to make $N_A > N_d$, the low temperature Fermi level switches to $(E_c - E_A)$ and other terms can be simplified. Fig. 34.1 shows how ϕ will vary with the ratio (N_A/N_d) for a rather low temperature. The semiconductor is essentially intrinsic when $N_A = N_d$ if the two sets of levels are several kT above and below ϕ_i respectively.

For vacancy acceptors which can accept more than one electron, the course of ϕ with the ratio (N_A/N_d) will be more or less as indicated

in Fig. 34.2. Judgment of a high temperature equilibration process must be quite careful if enough vacancies are to be introduced to compensate the donors yet not so many that the Fermi level locks to the lower acceptor level.

3.4.2 IRRADIATION EFFECTS

Exposure of a semiconductor to thermal neutrons leads to transmutation of some of the host atoms, often providing new kinds of chemical donors and acceptors. This is *not* the problem of immediate concern

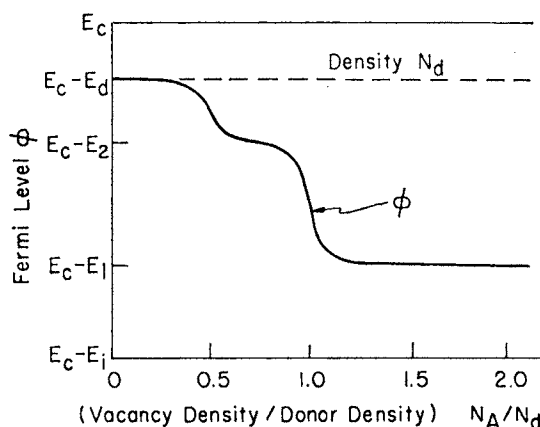


FIG. 34.2. The variation of ϕ with (N_A/N_D) at low temperatures if each vacancy can accept two electrons, at energies $(E_c - E_1)$ and $(E_c - E_2)$.

here. Instead, consider a semiconductor already containing densities of shallow donor and acceptor states N_d , N_a , which are almost completely ionized at room temperature. (This is a reasonable assumption for semiconductors such as germanium, silicon, tellurium, InSb, PbS, and many more.) When such a semiconductor is bombarded with *fast* nucleons (energy more than say 20 eV and perhaps up to several MeV), Frenkel defects are introduced, equal densities of interstitials and vacancies. It is reasonable to suppose that irradiation has produced N donor levels at some energy E_D below the conduction band and an *equal number* of acceptor levels at energy $(E_c - E_A)$. The electronic

equilibrium is now given by

$$[n_0 - p_0] = (N_d - N_a) + \frac{N}{1 + \beta_D^{-1} \exp(\eta + \epsilon_D)} - \frac{N}{1 + \beta_A \exp(-\eta - \epsilon_A)} \quad (342.1)$$

The model characterized by this equation has been discussed by James and Lehman (1951:14) and has some rather interesting features. Consider a semiconductor which is *n*-type to start with, $N_d > N_a$, and suppose that both sets of defect levels are well below the conduction band. Then (342.1) simplifies to

$$n_0 \approx (N_d - N_a) - N \quad (342.2)$$

for which situation

$$\phi \approx E_c - kT \ln \left[\frac{N_c}{N_d - N_a - N} \right] \quad (342.3)$$

The fall of ϕ with increasing N is at first very gradual, but as N approaches $(N_d - N_a)$ this fall becomes cataclysmic. This can be appreciated most readily from viewing Fig. 34.3, which is based on a model James and Lehman use for germanium.

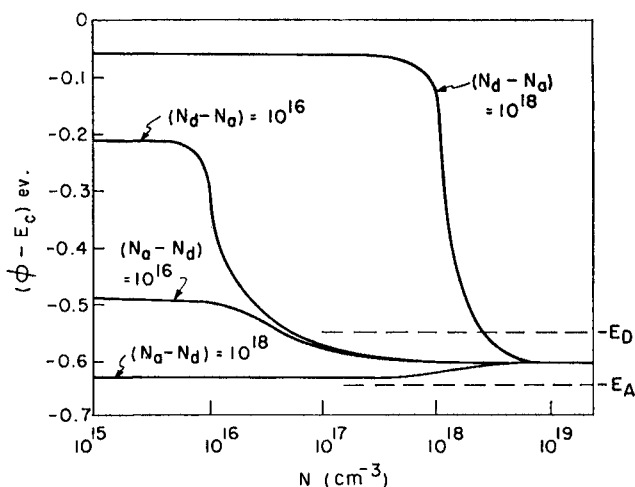


FIG. 34.3. Schematic variation of the Fermi level with Frenkel pair density for a semiconductor similar to germanium. From the model of James and Lehman (1951:14).

When N becomes large compared with $|N_a - N_d|$, the last two terms of Eq. (342.1) are essentially equal, and the Fermi level adopts an energy intermediate between $(E_c - E_D)$ and $(E_c - E_A)$. By equating these two terms,

$$\phi = E_c - \frac{1}{2}(E_D + E_A) + \frac{1}{2}kT \ln(\beta_D \beta_A) \quad (342.4)$$

The change of Fermi level is less dramatic for material which was originally p -type, but a break in the behavior of ϕ still occurs when N exceeds $(N_a - N_d)$. Whether ϕ rises or falls as a result of heavy irradiation depends on the energies E_D , E_A , and on the original value of $(N_a - N_d)$.

For a material in which the defect levels straddle the intrinsic Fermi level ϕ_i , irradiation will tend to make either n -type or p -type material virtually intrinsic, a situation which may be the case for silicon (1951:13).

As experimental evidence on irradiation effects is accumulated (e.g. 1957:27), it appears likely that the simple model of Fig. 34.3 is inadequate for any semiconductor. Vacancies and interstitials themselves are likely to present more than one level of ionization, and in compounds there are four or more possible types of defect. It is fruitful to consider the more elaborate models which can be envisaged only in relation to experimental data on actual materials.

3.5 IMPURITY BANDS AND THE BEHAVIOR OF AN IMPURITY METAL

When impurity levels are separated by a finite energy from the conduction or valence bands, it is impossible to maintain a finite carrier density in either of the principal bands down to 0°K. However, when impurity atoms are very numerous—and hence rather closely spaced—they tend to interact with each other. As the impurity density is increased, the impurity ionization energy E_d or E_a becomes smaller, and above a critical density—characteristic of the host lattice and the impurity type—disappears completely. Shifrin (1944:1) forecast that this should always occur. Experimental evidence of the attenuation and final disappearance of the ionization energy is well exemplified by the classic paper of Pearson and Bardeen (1949:4) on silicon doped with phosphorus or boron. Their experimental result for boron

acceptors is indicated in Fig. 35.1, and the curve joining the points for samples with different acceptor densities is given by

$$E_a = [0.084 - 4.3 \times 10^{-8} N_a^{1/3}] \text{ eV} \quad (350.1)$$

Pearson and Bardeen noted that if the ionization energy is reduced by overlap between impurity wave-functions, this should be inversely proportional to the average distance between an ionized impurity and

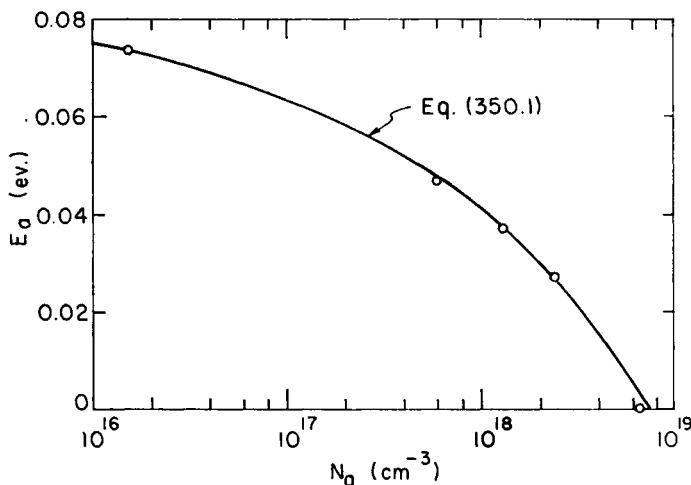


FIG. 35.1. Variation of boron acceptor ionization energy with acceptor concentration, according to the measurements of Pearson and Bardeen (1949:4).

a free carrier. This at first sight seems to support the presence of a term proportional to $N_a^{1/3}$ in Eq. (350.1). Further reflection destroys this illusion: if Coulomb interactions provide the seat for the reduction of ionization energy, the reduction should depend not on the *total* impurity density but only on those which are *ionized* in the temperature range of interest. Taking this more rigorous approach, Castellan and Seitz (1951:6) were unable to account for the effects seen in silicon. More recently, Debye and Conwell (1954:2) have been able to fit the ionization energy for arsenic donors in germanium by an expression.

$$E_d = [0.0125 - 2.35 \times 10^{-8} (N_{di})^{1/3}] \text{ eV} \quad (350.2)$$

(see Fig. 35.2) where N_{di} denotes the density of donors which are ionized at the low temperatures for which E_d can be measured. Not much reliable experimental information is available for impurities in other semiconductors.

Our theoretical understanding of the manner in which impurity ionization energies depend on density is still incomplete. Pincherle (1951:7) has shown that part of the effect can come from the screening

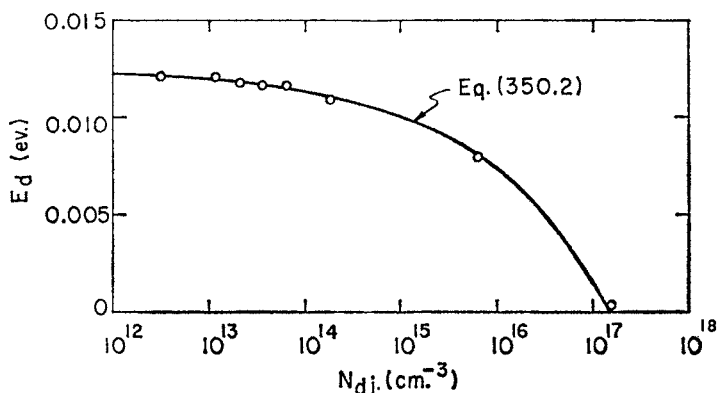


FIG. 35.2. The ionization energy for arsenic donors in Ge, as a function of the average density N_{di} of ionized donors in the temperature range used for fitting. The experimental points are the data of Debye and Conwell (1954:2).

of the field around an ion by free carriers. Mott (1956:3) has presented some arguments from which one would expect an *abrupt* change from finite to zero E_d at a critical impurity density. Landsberg (1960:1) makes semi-quantitative estimates of how such a transition might be affected by statistical spatial fluctuations in impurity density.

At any rate, when donor impurities become sufficiently numerous, they can no longer be described in terms of discrete localized levels below the conduction band. Instead they form an *impurity band* of non-localized states situated at the energy E_c , the base of the conduction band. (For low lying acceptor levels the corresponding impurity band is of course at the very top of the valence band.) The system of an impurity band contiguous to one of the principal bands gives us *impurity-*

metallic behavior. The term impurity metal is used since the number of carriers in non-localized states is then constant throughout the extrinsic temperature range, just as a metal has a temperature-independent electron density.

From a strict point of view, impurity states should not be regarded as entirely additional to the Bloch states of conduction and valence bands. As Brooks (1955:5) remarks, impurity levels usually represent states split off from the main bands. Thus when co-operation between donors leads to the formation of an impurity band joined to the conduction band, the conduction band is perhaps no more than getting its own states back again!

There is, however, more to the picture than this. Certainly if the donor density is *very* large (say large compared with 10^{20} cm^{-3}), the density of states function Eq. (144.4) will be appreciably affected by the provision of donor states, and the impurity band will be spread over a considerable range of energy. However, if the donor density is only slightly larger than the critical value required for impurity-metallic behavior (say $\sim 10^{18} \text{ cm}^{-3}$), $g(E)$ for the conduction band will be but little affected, whereas the impurity band will consist essentially of $\beta^{-1}N_a$ non-localized states all in the immediate neighborhood of the energy E_c . It is this concept of a "weak" impurity metal which is considered first.

3.5.1 WEAK IMPURITY METALS

It will be noted that the number of states in the impurity band is described as $\beta^{-1}N_a$, not just as N_a . We have to make an adjustment to a slight change of conception with regard to impurity spin degeneracy. For previous sections have considered separate non-interacting impurities; each had β^{-1} wave-functions describing a possible ground state of the system, but only one of these could be occupied at a time. This restriction is no longer true for an impurity *band*; the wave-functions are sufficiently non-localized so that more than one electron can occupy states derived from a given impurity atom. (The simple matter of electron supply will ensure, however, that the $N_a\beta^{-1}$ impurity states contain *less than* N_a electrons.)

Suppose that in addition to the donors, there is a smaller density N_a of assorted acceptor impurities, each of which can (and will) take the opportunity to acquire an extra electron. All compensating acceptors

will be ionized, since their levels fall well below the anticipated Fermi energy. Then there are $(N_d - N_a)$ electrons per unit volume to be distributed over the conduction band states and the $\beta^{-1}N_d$ impurity band states. If the latter are all at E_c , the equation expressing neutrality is

$$(N_d - N_a) = N_c \mathcal{F}_{1/2} \left(\frac{\phi - E_c}{kT} \right) + \beta^{-1} N_d f(E_c) \quad (351.1)$$

where $f(E_c)$ denotes the usual Fermi occupancy factor $\{1 + \exp(\epsilon - \eta)\}^{-1}$ or $\{1 + \exp[(E - \phi)/kT]\}^{-1}$ for the energy E_c .

Since the effective density of conduction band states varies as $T^{3/2}$, the term in Eq. (351.1) containing N_c must become negligibly small at a sufficiently low temperature. This means that all the available electron supply is in the impurity band and none in the conduction band at $T = 0$. At temperatures sufficiently small for this to be essentially the case, Eq. (351.1) is linear in $\exp(\eta)$, viz.

$$(N_d - N_a)[1 + \exp(\eta)] = \beta^{-1} N_d \exp(\eta), \quad T \rightarrow 0 \quad (351.2)$$

Thus the Fermi energy and its reduced version tend at low temperatures towards

$$\left. \begin{aligned} \eta &= \ln \left[\frac{N_d - N_a}{N_d(\beta^{-1} - 1) + N_a} \right] \\ \phi &= E_c + kT \ln \left[\frac{N_d - N_a}{N_d(\beta^{-1} - 1) + N_a} \right] \end{aligned} \right\} T \rightarrow 0 \quad (351.3)$$

In the particular case of no compensation and simple donors (for which $\beta^{-1} = 2$), the logarithmic term in Eq. (351.3) is zero, and ϕ is at the energy E_c for any temperature low enough to keep almost all the electrons out of the "conventional" conduction band states. When β^{-1} is larger, or when there is any appreciable compensation, $(d\phi/dT)$ has a finite negative value even at the lowest temperatures. Thus for $\beta^{-1} = 2$ and $(N_a/N_d) = 0.2$, we have $(d\phi/dT) = -0.4k$ at 0°K .

In any event, ϕ will fall more rapidly at high temperatures than at low, since an appreciable number of the ordinary conduction band states become accessible for occupancy. The Fermi level must then be found by solving Eq. (351.1). Note that since η will necessarily be negative, one of the approximate expressions developed in Appendix C can be substituted for $\mathcal{F}_{1/2}(\eta)$ without serious loss of accuracy. In order

to make Eq. (351.1) no higher than quadratic in $\exp(\eta)$, we may use

$$\mathcal{F}_{1/2}(\eta) \approx \frac{\exp(\eta)}{1 + C \exp(\eta)} \quad (351.4)$$

knowing that the error will be less than $\pm 3\%$ for any negative value of η if $C = 0.27$.

Employing the form (351.4), we can write Eq. (351.1) as the quadratic

$$\begin{aligned} \exp(2\eta)[N_c + CN_d(\beta^{-1} - 1) + CN_a] + \\ + \exp(\eta)[N_c + N_d(\beta^{-1} - 1 - C) + N_a(1 + C)] - (N_d - N_a) = 0 \end{aligned} \quad (351.5)$$

whose solution for $\exp(\eta)$ is

$$\exp(\eta) = \frac{[N_c + N_d(\beta^{-1} + 1 - C) - N_a(1 - C)]^2 - 4\beta^{-1}N_d \times \\ \times (N_d - N_a)(1 - C)^{1/2} - [N_c + N_d(\beta^{-1} - 1 - C) + N_a(1 + C)]}{2[N_c + CN_d(\beta^{-1} - 1) + CN_a]} \quad (351.6)$$

At sufficiently *low* temperatures, this of course asymptotically approaches the form required by Eq. (351.3); while (not surprisingly) $\exp(\eta)$ tends towards $(N_d - N_a)/N_c$ at sufficiently *high* temperatures. At intermediate temperatures, when the populations of the impurity band and the conduction band are comparable, Eq. (351.6) is not exactly a simple expression to visualize.

Fig. 35.3 shows how ϕ and η vary with temperature when the impurity metal is derived from a conduction band with $m_c = m_0$ (i.e. $N_c = 4.83 \times 10^{15} T^{3/2} \text{ cm}^{-3}$) and a set of 10^{18} cm^{-3} donors for which $\beta^{-1} = 2$. The uppermost curve in each part of the figure relates to a condition of zero compensation, $N_a = 0$. As has been remarked in earlier sections, compensation can never be completely absent in any real material; but the curves would not be appreciably different from those shown provided that N_a were smaller than, say, $0.01N_d$. As noted in connection with Eq. (351.3), the reduced Fermi level is zero at 0°K for $\beta^{-1} = 2$ and $N_a = 0$. The two pairs of curves for $N_a = 2 \times 10^{17} \text{ cm}^{-3}$ (20% compensation) and $N_a = 4 \times 10^{17} \text{ cm}^{-3}$ (40% compensation) illustrate the marked effect of compensation on η and ϕ , particularly at low temperatures.

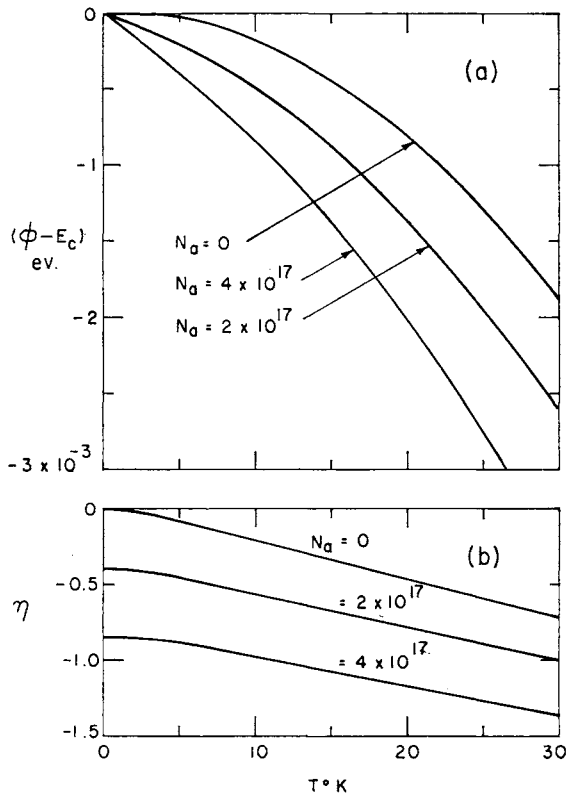


FIG. 35.3. Variation of (a) Fermi level, and (b) reduced Fermi level with temperature for a weak impurity metal. Conduction band supposed with $m_c = m_0$, and 10^{18} cm^{-3} donors for which $\beta^{-1} = 2$. Curves show the effect of 40%, 20%, or zero compensation.

3.5.2 STRONG IMPURITY METALS

It has been supposed in the last sub-section that the density of impurity atoms was large enough to create an impurity band, but not much more. In such a context it was reasonable to suppose that the impurity band would be very narrow in energy, situated just about at E_c . It was also reasonable to suppose that $g(E)$ for the conduction band would be not appreciably affected by the subtraction of states needed for the impurity band.

Such assumptions cease to be permissible if the impurity density is very large, say 10^{20} cm^{-3} or more. The situation then becomes what we choose to call that of a "strong" impurity metal. Fig. 35.4 indicates qualitatively the features of a weak impurity metal and what might be expected for a strong impurity metal. In the latter case, overlap between impurity wave-functions is so strong that the impurity band must cover an appreciable range of energy. The center of gravity for

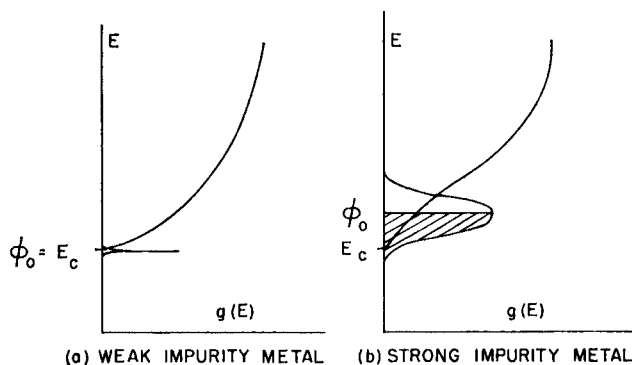


FIG. 35.4. Possible variation of the density of states with energy (a) for a weak impurity metal in which the impurity band is very narrow, situated at about E_c on the energy scale, and not appreciably affecting the density of states function for the conduction band; (b) for a strong impurity metal in which the impurity band is rather broad, possibly centered on an energy higher than E_c , and such that $g(E)$ for the conduction band is seriously perturbed.

this distribution may occur at an energy above E_c . The presence of this large impurity band must necessarily have a considerable effect on $g(E)$ for the original conduction band. This effect should be most pronounced on the lower energy states of the conduction band, making that band non-parabolic as suggested in the figure.

For a strong impurity band compounded from donor states with $\beta^{-1} = 2$, the Fermi level will be a little below the midpoint of the impurity band at low temperatures (the exact position depending on the number of accessible conduction band states and the extent of compensation). This Fermi level will decrease very slowly on warming and will remain above E_c up to fairly high temperatures.

A theoretical understanding of strong impurity bands and their effect on the conduction band parameters is blocked by formidable mathematical problems. Thus we are not in a position to suggest analytic forms for $g(E)$ in either of the bands illustrated by Fig. 35.4(b), and do not propose to enter into any more detailed discussion. At least the preceding remarks do give some qualitative indications of the expected behavior for a strong impurity metal. There is considerable interest in these problems, for such materials occur in important semiconducting devices. Thus the degree of doping is probably enough to move the Fermi level into either the conduction or valence band in the emitter region of many transistors, and in tunnel diodes (1958:5).

3.5.3 OCCUPANCY OF WEAKLY INTERACTING IMPURITIES

When the spacing of impurity centers is moderately large, the overlap of impurity wave-functions is not strong enough to cause the complete breakdown of localized states in the manner of an impurity metal. Erginsoy (1950:8) suggested that while the impurity ground states would be localized, excited states might still form impurity bands, some of which could be separated by small energies from the conduction or valence bands. It is difficult to judge realistically what the effect of excited state bands might be, since such bands would presumably be separated from the Fermi level by many kT at low temperatures; however Koshino (1956:28) has suggested that they provide the key to some results of Fritzsche (1955:29) with gallium doped germanium.

Certainly when majority impurity densities are of the order of 10^{16} cm^{-3} or less, the overlap between impurity wave-functions will be very small and it is not proper to refer to transition of an electron from one impurity to another as constituting impurity *band* conduction. At one time, theoretical discussions of this subject were entangled in the concept that each impurity atom provides two states (of opposite spin) and that a half full impurity band would be realized by the ground states. This concept is false, since the Bloch collective electron treatment is not applicable to a set of very weakly interacting centers; the Heitler-London approach is more appropriate.†

† As is also the case in materials for which the spacing between the atoms of the ordinary lattice is rather large. We have commented on this previously in connection with NiO and related compounds.

For a system of impurity centers sparsely distributed through a semiconductor, charge transfer between the impurities is possible only if some centers have been ionized through the presence of compensating centers. This has been described by Conwell (1956:29), Mott (1956:3), and Price (1957:29, 1958:21). It is interesting to consider the basis of Price's model for a set of partly compensated monovalent donor impurities.

Price points out that if there are $N_a \ll N_d$ compensating acceptors present, every one will be ionized, leaving $(N_d - N_a)$ electrons on the N_d donor levels. The distribution of these electrons will not be entirely random. For donor centers which happen to be reasonably close to an ionized acceptor will be rather *less* likely (as a result of Coulomb repulsion) to contain an electron than donors which are far from any ionized acceptors. Price suggests that each acceptor will be surrounded by a small group of r donors which he calls "trapping sites". An electron must have an extra energy E_t to exist on a trapping site† when it is the first site in the group to be occupied, although subsequent sites in the group can be occupied more cheaply, or even free of any penalty.

For simplicity, we consider here the use of *one* trapping site per compensating acceptor. Then there are simply $N_f = (N_d - N_a)$ "free" donor sites at energy $(E_c - E_d)$, and $N_t = N_a$ "trapping" sites at energy $(E_c - E_d + E_t)$. If we denote the density of unoccupied free sites as p_f and the density of unoccupied trapping sites as p_t , then

$$p_t = \frac{N_a}{1 + \beta^{-1} \exp(\eta + \epsilon_d - \epsilon_t)} \quad (353.1)$$

$$p_f = \frac{N_d - N_a}{1 + \beta^{-1} \exp(\eta + \epsilon_d)} \quad (353.2)$$

Now $(p_f + p_t)$ must equal the sum of N_a and any free electrons left in the conduction band, n_0 . Using this fact, we have from Eq. (353.1) and (353.2) that

$$\begin{aligned} \left[\frac{N_d - N_a - p_f}{p_f} \right] &= \beta^{-1} \exp(\epsilon_d + \eta) \\ &= \exp(\epsilon_t) \left[\frac{N_a - p_t}{p_t} \right] \\ &= \exp(\epsilon_t) \left[\frac{p_f - n_0}{N_a + n_0 - p_f} \right] \end{aligned} \quad (353.3)$$

† In other words, a "hole" is trapped by energy E_t on a trapping site.

This can be rearranged as a quadratic equation for p_f :

$$p_f^2(z-1) + p_f[N_d - n_0(z-1)] = (N_d - N_a)(N_a + n_0) \quad (353.4)$$

where z denotes $\exp(\epsilon_t)$. The solution for the density of holes among the free sites is

$$p_f = \frac{2(N_d - N_a)(N_a + n_0)}{\{[N_d - n_0(z-1)]^2 + 4(z-1)(N_d - N_a)(N_a + n_0)\}^{1/2} + [N_d - n_0(z-1)]} \quad (353.5)$$

At high temperatures (such that $E_t \ll kT$ and $z \approx 1$), this reduces to

$$p_f \approx (N_a + n_0)(1 - N_a/N_d), \quad E_t \ll kT \quad (353.6)$$

which indicates as expected that at a reasonably high temperature there will be $(N_a + n_0)$ vacant donor sites distributed uniformly between the "free" states and the "trapping" states.

At low temperatures, ($E_t \gg kT$, $z \gg 1$), it is reasonable to expect that n_0 will become extremely small. Then Eq. (353.5) reduces to

$$p_f = [N_a(N_d - N_a)]^{1/2} \exp(-E_t/2kT), \quad E_t \gg kT \quad (353.7)$$

Experimental results for the temperature dependence of electrical conductivity in doped semiconductors such as germanium and silicon support the general form of Eq. (353.7). Correlation with electrical conductivity of course requires a model for the mobility of holes on "free" donor sites. The theory of Mott and Twose (1959:12) suggests that the mobility for hopping between impurity atoms may not depend on temperature. A similar conclusion is reached by Miller and Abrahams (1960:23). Some temperature dependence of mobility is however expected from the alternative theories of Kasuya and Koide (1958:22) and Conwell (1956:29).

Most experimental papers have been content to report the apparent trapping energy E_t from an essentially exponential dependence of conductivity on reciprocal temperature [e.g. (1955:29, 1957:30, 1960:9)]. This author has, however, found (1959:6) that experimental results for indium-doped germanium fit the changing temperature dependence of Eq. (353.5) quite well. More elaborate models could presumably give an even more satisfactory description, and Price (1958:21) suggests the form such elaborations might take.