Silicon The crystal has the diamond structure, so the first Brillouin zone is the truncated octahedron appropriate to a face-centered cubic Bravais lattice. The conduction band has six symmetry-related minima at points in the ⟨100⟩ directions, about 80 percent of the way to the zone boundary (Figure 28.5). By symmetry each of the six ellipsoids must be an ellipsoid of revolution about a cube axis. They are quite cigar-shaped, being elongated along the cube axis. In terms of the free electron mass m, the effective mass along the axis (the longitudinal effective mass) is $m_L \approx 1.0m$ while the effective masses perpendicular to the axis (the transverse effective mass) are $m_T \approx 0.2m$. There are two degenerate valence band maxima, both located at $\mathbf{k} = \mathbf{0}$, which are spherically symmetric to the extent that the ellipsoidal expansion is valid, with masses of $0.49m$ and $0.16m$ (Figure 28.6).

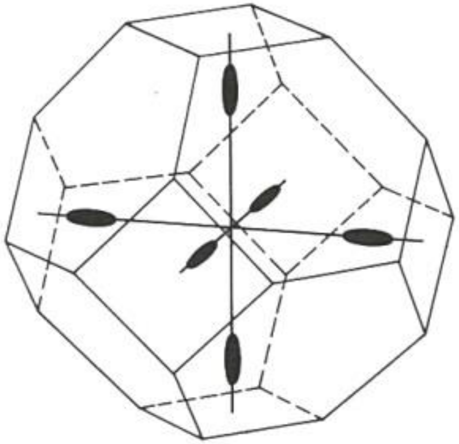


Figure 28.5
Constant-energy surfaces near the conduction band minima in silicon. There are six symmetry-related ellipsoidal pockets. The long axes are directed along ⟨100⟩ directions.

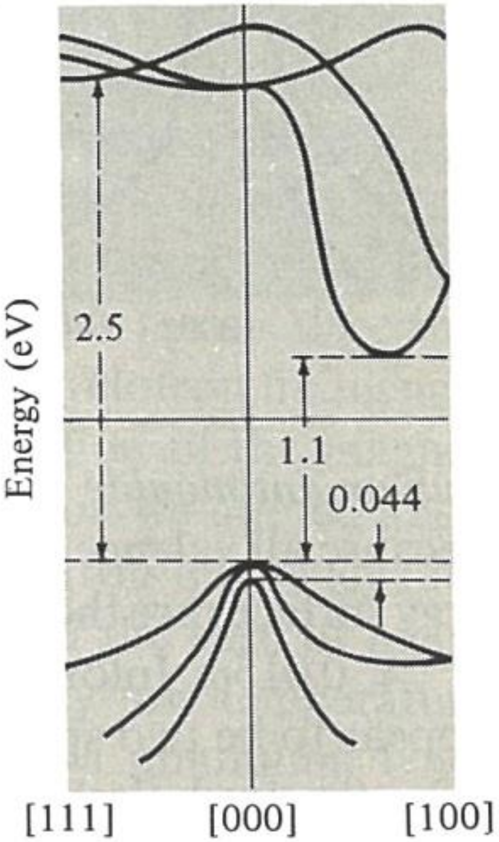


Figure 28.6
Energy bands in silicon. Note the conduction band minimum along [100] that gives rise to the ellipsoids of Figure 28.5. The valence band maximum occurs at $\mathbf{k} = 0$, where two degenerate bands with different curvatures meet, giving rise to "light holes" and "heavy holes." Note also, the third band, only 0.044 eV below the valence band maximum. This band is separated from the other two only by spin-orbit coupling. At temperatures on the order of room temperature ($k_B T = 0.025$ eV) it too may be a significant source of carriers. (From C. A. Hogarth, ed., *Materials Used in Semiconductor Devices*, Interscience, New York, 1965.)

Germanium The crystal structure and Brillouin zone are as in silicon. However, the conduction band minima now occur at the zone boundaries in the $\langle 111 \rangle$ directions. Minima on parallel hexagonal faces of the zone represent the same physical levels, so there are four symmetry-related conduction band minima. The ellipsoidal constant energy surfaces are ellipsoids of revolution elongated along the $\langle 111 \rangle$ directions, with effective masses $m_L \approx 1.6m$, and $m_T \approx 0.08m$ (Figure 28.7). There are again two degenerate valence bands, both with maxima at $\mathbf{k} = \mathbf{0}$, which are spherically symmetric in the quadratic approximation with effective masses of $0.28m$ and $0.044m$ (Figure 28.8).

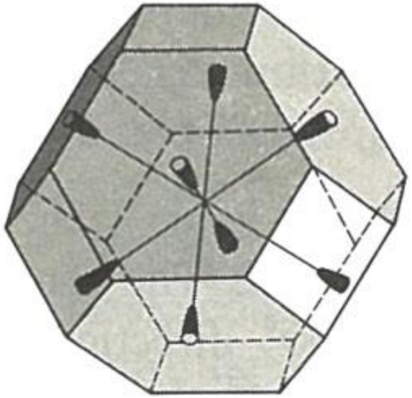


Figure 28.7
Constant-energy surfaces near the conduction band minima in germanium. There are eight symmetry-related half ellipsoids with long axes along $\langle 111 \rangle$ directions centered on the midpoints of the hexagonal zone faces. With a suitable choice of primitive cell in k-space these can be represented as four ellipsoids, the half ellipsoids on opposite faces being joined together by translations through suitable reciprocal lattice vectors.

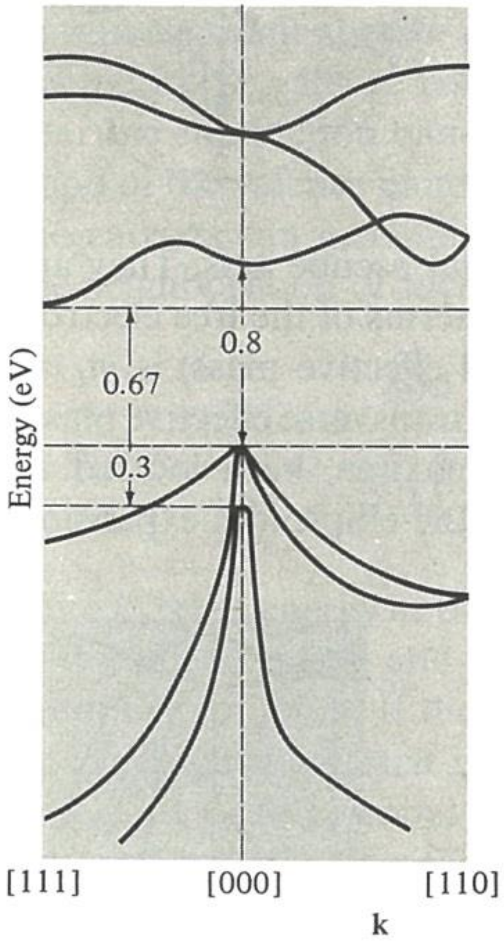


Figure 28.8
Energy bands in germanium. Note the conduction band minimum along [111] at the zone boundary that gives rise to the four ellipsoidal pockets of Figure 28.7. The valence band maximum, as in silicon, is at $\mathbf{k} = \mathbf{0}$, where two degenerate bands with different curvatures meet, giving rise to two pockets of holes with distinct effective masses. (From C. A. Hogarth, ed., *Materials Used in Semiconductor Devices*, Interscience, New York, 1965.)

Indium antimonide This compound, which has the zincblende structure, is interesting because all valence band maxima and conduction band minima are at $\mathbf{k} = 0$. The energy surfaces are therefore spherical. The conduction band effective mass is very small, $m^* \approx 0.015m$. Information on the valence band masses is less unambiguous, but there appear to be two spherical pockets about $\mathbf{k} = 0$, one with an effective mass of about $0.2m$ (heavy holes) and another with effective mass of about $0.015m$ (light holes).

CYCLOTRON RESONANCE

The effective masses discussed above are measured by the technique of cyclotron resonance. Consider an electron close enough to the bottom of the conduction band (or top of the valence band) for the quadratic expansion (28.2) to be valid. In the presence of a magnetic field **H** the semiclassical equations of motion (12.32) and (12.33) imply that the velocity **v**(**k**) obeys the single set of equations

$$\mathsf{M}\,\frac{d\mathbf{v}}{dt} = \mp\frac{e}{c}\,\mathbf{v}\times\mathbf{H}. \tag{28.4}$$

In a constant uniform field (taken along the z-axis) it is not difficult to show (Problem 1) that (28.4) has an oscillatory solution

$$\mathbf{v} = \operatorname{Re}\,\mathbf{v}_0 e^{-i\omega t}, \tag{28.5}$$

provided that

$$\boxed{\omega = \frac{eH}{m^*c},} \tag{28.6}$$

where m^*, the "cyclotron effective mass," is given by

$$m^* = \left(\frac{\det\mathsf{M}}{M_{zz}}\right)^{1/2}. \tag{28.7}$$

This result can also be written in terms of the eigenvalues and principal axes of the mass tensor as (Problem 1):

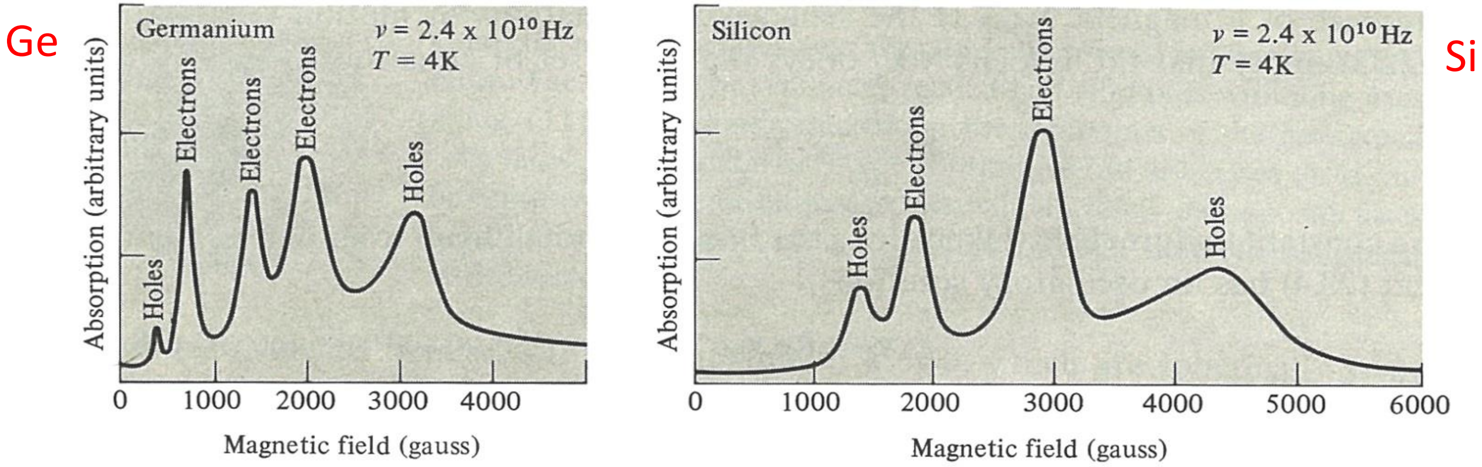
$$\boxed{m^* = \sqrt{\frac{m_1 m_2 m_3}{\hat{H}_1{}^2 m_1 + \hat{H}_2{}^2 m_2 + \hat{H}_3{}^2 m_3}},} \tag{28.8}$$

where the $\hat{H}_i$ are the components along the three principal axes of a unit vector parallel to the field.

For a given ω, m* varies with H

Note that the cyclotron frequency depends, for a given ellipsoid, on the orientation of the magnetic field with respect to that ellipsoid, but not on the initial wave vector or energy of the electron. Thus for a given orientation of the crystal with respect to the field, all electrons in a given ellipsoidal pocket of conduction band electrons (and, by the same token, all holes in a given ellipsoidal pocket of valence band holes) precess at a frequency entirely determined by the effective mass tensor describing that pocket. There will therefore be a small number of distinct cyclotron frequencies. By noting how these resonant frequencies shift as the orientation of the magnetic field is varied, one can extract from (28.8) the kind of information we quoted above.

To observe cyclotron resonance it is essential that the cyclotron frequency (28.6) be larger than or comparable to the collision frequency. As in the case of metals, this generally requires working with very pure samples at very low temperatures, to reduce both impurity scattering and phonon scattering to a minimum. Under such conditions the electrical conductivity of a semiconductor will be so small that (in contrast to the case of a metal (page 278)) the driving electromagnetic field can penetrate far enough into the sample to excite the resonance without any difficulties associated with a skin depth. On the other hand, under such conditions of low temperatures and purity the number of carriers available in thermal equilibrium to participate in the resonance may well be so small that carriers will have to be created by other means—such as photoexcitation. Some typical cyclotron resonance data are shown in Figure 28.9.



(a)

(b)

Figure 28.9

Typical cyclotron resonance signals in (a) germanium and (b) silicon. The field lies in a (110) plane and makes an angle with the [001] axis of 60° (Ge) and 30° (Si). (From G. Dresselhaus et al., *Phys. Rev.* **98**, 368 (1955).)

Ashroft & Mermin, p. 575-511.

Extrinsic Case: Some General Features

If impurities contribute a significant fraction of the conduction band electrons and/or valence band holes, one speaks of an "extrinsic semiconductor." Because of these

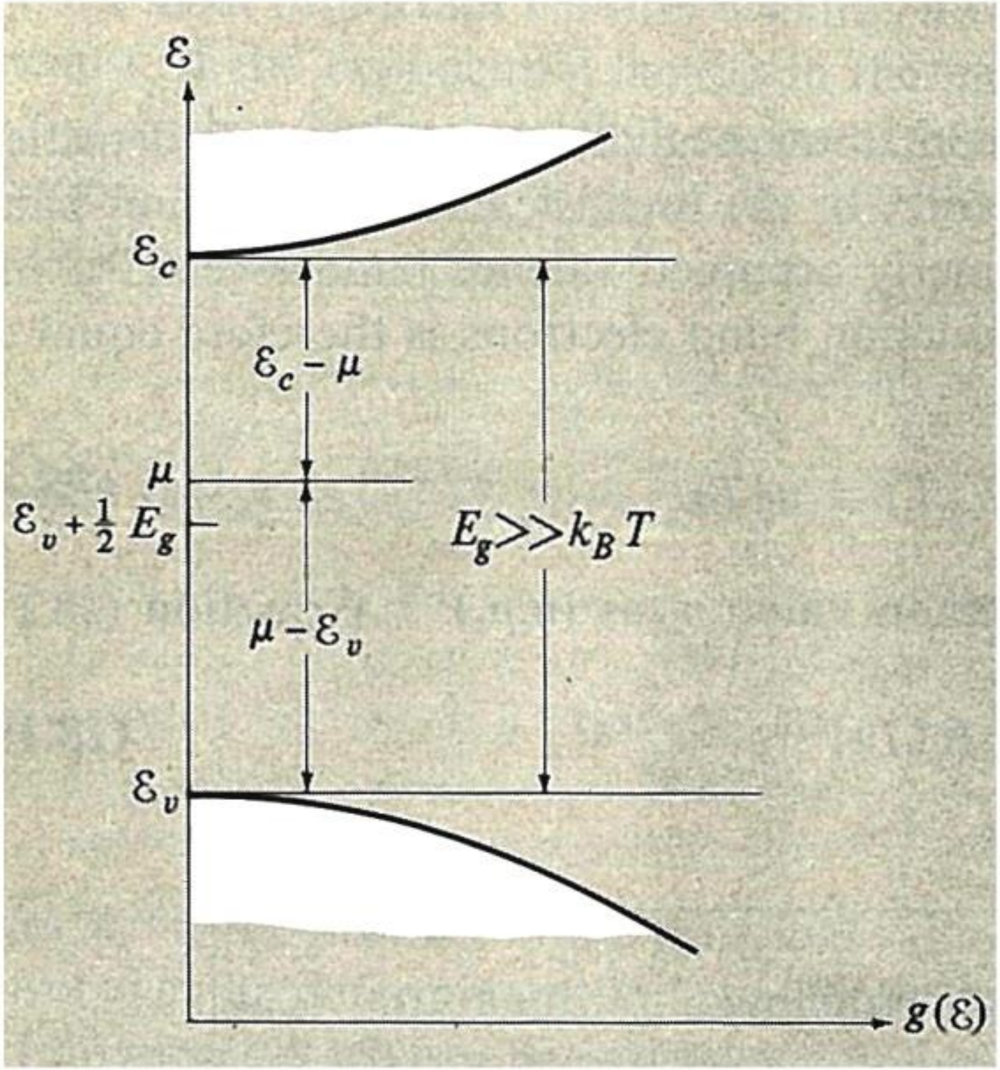


Figure 28.10
In an intrinsic semiconductor with an energy gap E_g large compared with k_BT, the chemical potential μ lies within order k_BT of the center of the energy gap, and is therefore far compared with k_BT from both boundaries of the gap at $\mathcal{E}_c$ and $\mathcal{E}_v$.

added sources of carriers the density of conduction band electrons need no longer be equal to the density of valence band holes:

$$n_c - p_v = \Delta n \neq 0. \tag{28.23}$$

Since the law of mass action Eq. (28.17) holds regardless of the importance of impurities, we can use the definition (28.19) of $n_i(T)$ to write quite generally,

$$n_c p_v = n_i^2. \tag{28.24}$$

Equations (28.24) and (28.23) permit one to express the carrier densities in the extrinsic case in terms of their intrinsic values n_i and the deviation Δn from intrinsic behavior:

$$\boxed{\left\{\begin{matrix} n_c \\ p_v \end{matrix}\right\} = \frac{1}{2}\left[(\Delta n)^2 + 4n_i^2\right]^{1/2} \pm \tfrac{1}{2}\Delta n.} \tag{28.25}$$

The quantity $\Delta n/n_i$, which measures the importance of the impurities as a source of carriers, can be given a particularly simple expression as a function of chemical potential μ, if we note that Eqs. (28.12) have the form[14]

$$n_c = e^{\beta(\mu - \mu_i)} n_i; \quad p_v = e^{-\beta(\mu - \mu_i)} n_i. \tag{28.26}$$

Therefore

$$\boxed{\frac{\Delta n}{n_i} = 2 \sinh \beta(\mu - \mu_i).} \tag{28.27}$$

We have noted that if the energy gap E_g is large compared with k_BT, then the intrinsic chemical potential μ_i will satisfy the assumption (28.10) of nondegeneracy. But Eq. (28.27) requires that if μ_i is far from $\mathcal{E}_c$ or $\mathcal{E}_v$ on the scale of k_BT, then μ must be as well, unless Δn is many orders of magnitude larger than the intrinsic carrier density n_i. Thus the nondegeneracy assumption underlying the derivation of (28.27) is valid when $E_g \gg k_BT$, unless we are in a region of extreme extrinsic behavior.

Note also that when Δn is large compared with n_i, then Eq. (28.25) asserts that the density of one carrier type is essentially equal to Δn, while that of the other type is smaller by a factor of order $(n_i/\Delta n)^2$. Thus when impurities do provide the major source of carriers, one of the two carrier types will be dominant. An extrinsic semiconductor is called "*n*-type" or "*p*-type" according to whether the dominant carriers are electrons or holes.

To complete the specification of the carrier densities in extrinsic semiconductors one must determine Δn or μ. To do this we must examine the nature of the electronic levels introduced by the impurities and the statistical mechanics of the occupation of these levels in thermal equilibrium.

Ashroft & Mermin, p. 581-584.

POPULATION OF IMPURITY LEVELS IN THERMAL EQUILIBRIUM

To assess the extent to which carriers can be thermally excited from impurity levels, we must compute the mean number of electrons in the levels at a given temperature and chemical potential. We assume that the density of impurities is low enough that the interaction of electrons (or holes) bound at different impurity sites is negligible. We may then calculate the number density of electrons n_d (or holes p_a) bound to donor (or acceptor) sites by simply multiplying by the density of donors N_d (or acceptors N_a) the mean number of electrons (or holes) there would be if there were only a single impurity. For simplicity we assume that the impurity introduces only a single one-electron orbital level.[21] We calculate its mean occupancy as follows:

Donor Level If we ignored electron-electron interactions the level could either be empty, could contain one electron of either spin, or two electrons of opposite spins. However, the Coulomb repulsion of two localized electrons raises the energy of the doubly occupied level so high that double occupation is essentially prohibited. Quite generally, the mean number of electrons in a system in thermal equilibrium is given by:

$n_d = \langle N \rangle N_d$

$$\langle n \rangle = \frac{\sum N_j e^{-\beta(E_j - \mu N_j)}}{\sum e^{-\beta(E_j - \mu N_j)}}, \qquad (28.30)$$

where the sum is over all states of the system, E_j and N_j, are the energy and number of electrons in state j, and μ is the chemical potential. In the present case the system is a single impurity with just three states: one with no electrons present which makes no contribution to the energy, and two with a single electron present of energy ε_d. Therefore (28.30) gives

$$\langle n \rangle = \frac{2e^{-\beta(\varepsilon_d-\mu)}}{1 + 2e^{-\beta(\varepsilon_d-\mu)}} = \frac{1}{\frac{1}{2}e^{\beta(\varepsilon_d-\mu)} + 1}, \tag{28.31}$$

so that[22]

$n_d = <N> N_d$

$$\boxed{n_d = \frac{N_d}{\frac{1}{2}e^{\beta(\varepsilon_d-\mu)} + 1}.} \tag{28.32}$$

Acceptor Level In contrast to a donor level, an acceptor level, when viewed as an electronic level, can be singly or doubly occupied, but not empty. This is easily seen from the hole point of view. An acceptor impurity can be regarded as a fixed, negatively charged attractive center superimposed on an unaltered host atom. This additional charge $-e$ can weakly bind one hole (corresponding to one electron being in the acceptor level). The binding energy of the hole is $\varepsilon_a - \varepsilon_v$, and when the hole is "ionized" an additional electron moves into the acceptor level. However, the configuration in which no electrons are in the acceptor level corresponds to two holes being localized in the presence of the acceptor impurity, which has a very high energy due to the mutual Coulomb repulsion of the holes.[23]

Bearing this in mind, we can calculate the mean number of electrons at an acceptor level from (28.30) by noting that the state with no electrons is now prohibited, while the two-electron state has an energy that is ε_a higher than the two one-electron states. Therefore

$$\langle n \rangle = \frac{2e^{\beta\mu} + 2e^{-\beta(\varepsilon_a - 2\mu)}}{2e^{\beta\mu} + e^{-\beta(\varepsilon_a - 2\mu)}} = \frac{e^{\beta(\mu - \varepsilon_a)} + 1}{\frac{1}{2}e^{\beta(\mu - \varepsilon_a)} + 1}. \tag{28.33}$$

The mean number of holes in the acceptor level is the difference between the maximum number of electrons the level can hold (two) and the actual mean number of electrons in the level ($\langle n \rangle$): $\langle p \rangle = 2 - \langle n \rangle$, and therefore $p_a = N_a \langle p \rangle$ is given by

$$\boxed{p_a = \frac{N_a}{\frac{1}{2}e^{\beta(\mu - \varepsilon_a)} + 1}.} \tag{28.34}$$

THERMAL EQUILIBRIUM CARRIER DENSITIES OF IMPURE SEMICONDUCTORS

Consider a semiconductor doped with N_d donor impurities and N_a acceptor impurities per unit volume. To determine the carrier densities we must generalize the constraint $n_c = p_v$ (Eq. (28.18)) that enabled us to find these densities in the intrinsic (pure) case. We can do this by first considering the electronic configuration at $T = 0$. Suppose $N_d \geqslant N_a$. (The case $N_d < N_a$ is equally straightforward and leads to the same result (28.35).) Then in a unit volume of semiconductor N_a of the N_d electrons supplied by the donor impurities can drop from the donor levels into the acceptor levels.[24] This gives a ground-state electronic configuration in which the valence band and acceptor levels are filled, $N_d - N_a$ of the donor levels are filled, and the conduction band levels are empty. In thermal equilibrium at temperature T the electrons will be redistributed among these levels, but since their total number remains the same, the number of electrons in conduction band or donor levels, $n_c + n_d$, must exceed its value at $T = 0$, $N_d - N_a$, by precisely the number of empty levels (i.e., holes), $p_v + p_a$, in the valence band and acceptor levels:

$$\boxed{n_c + n_d = N_d - N_a + p_v + p_a.} \tag{28.35}$$

This equation, together with the explicit forms we have found for n_c, p_v, n_d, and n_a as functions of μ and T, permits one to find μ as a function of T, and therefore to find the thermal equilibrium carrier densities at any temperature. A general analysis is rather complicated, and we consider here only a particularly simple and important case:

Suppose that

$$\begin{aligned} \mathcal{E}_d - \mu &\gg k_B T, \\ \mu - \mathcal{E}_a &\gg k_B T. \end{aligned} \tag{28.36}$$

Since $\mathcal{E}_d$ and $\mathcal{E}_a$ are close to the edges of the gap, this is only slightly more restrictive than the nondegeneracy assumption (28.10). Condition (28.36) and the expressions (28.32) and (28.34) for n_d and p_a insure that thermal excitation fully "ionizes" the impurities, leaving only a negligible fraction with bound electrons or holes: $\underline{n_d \ll N_d,}$ $\underline{p_a \ll N_a.}$ Equation (28.35) therefore becomes

$$\boxed{\Delta n = n_c - p_v = N_d - N_a,} \tag{28.37}$$

so Eqs. (28.25) and (28.27) now give the carrier densities and chemical potential as explicit functions of the temperature alone:

$$\begin{Bmatrix} n_c \\ p_v \end{Bmatrix} = \frac{1}{2}[(N_d - N_a)^2 + 4n_i^2]^{1/2} \pm \frac{1}{2}[N_d - N_a] \tag{28.38}$$

$$\frac{N_d - N_a}{n_i} = 2 \sinh \beta(\mu - \mu_i). \tag{28.39}$$

If the gap is large compared with $k_B T$, the assumption (28.36) we began with should remain valid unless μ is quite far from μ_i on the scale of $k_B T$. According to Eq. (28.39), this will only happen when $|N_d - N_a|$ is several orders of magnitude greater than the intrinsic carrier density n_i. Therefore Eq. (28.38) correctly describes the transition from predominantly intrinsic behavior ($n_i \gg |N_d - N_a|$) well into the region of predominantly extrinsic behavior ($n_i \ll |N_d - N_a|$). Expanding (28.38), we find that at low impurity concentrations the corrections to the purely intrinsic carrier densities are

Intrinsic

$$\boxed{\left\{\begin{matrix} n_c \\ p_v \end{matrix}\right\} \approx n_i \pm \tfrac{1}{2}(N_d - N_a),} \tag{28.40}$$

while for a considerable range of carrier concentrations in the extrinsic regime,

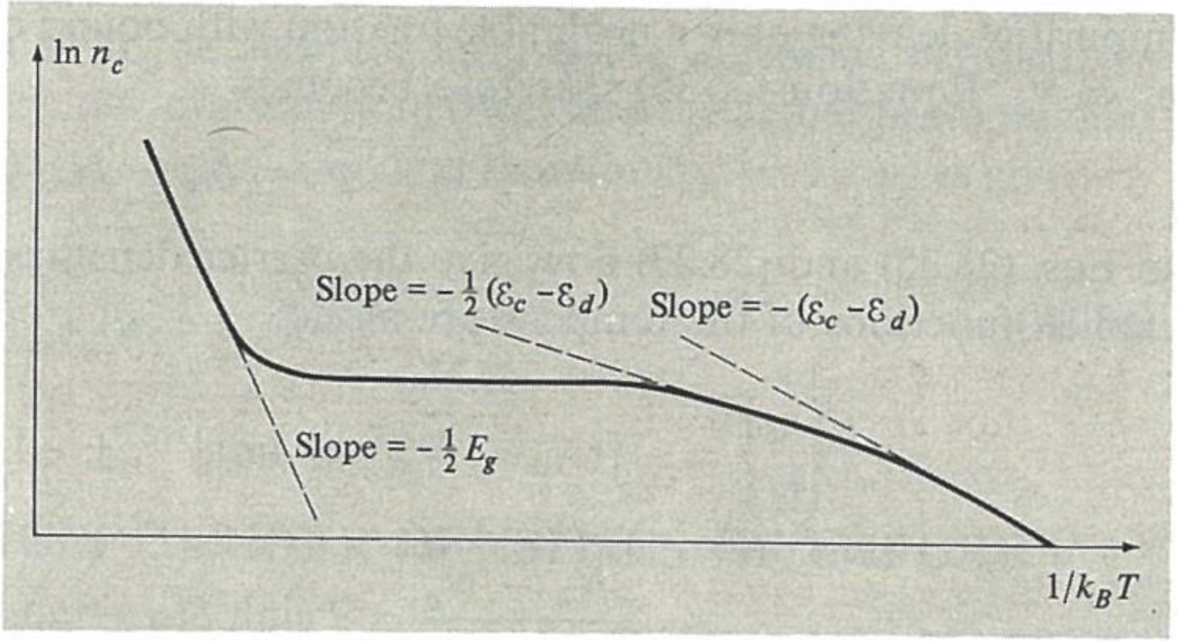
Extrinsic

$$\boxed{\begin{aligned} &\left.\begin{aligned} n_c &\approx N_d - N_a \\ p_v &\approx \frac{{n_i}^2}{N_d - N_a} \end{aligned}\right\} \quad N_d > N_a; \\ &\left.\begin{aligned} n_c &\approx \frac{{n_i}^2}{N_a - N_d} \\ p_v &\approx N_a - N_d \end{aligned}\right\} \quad N_a > N_d. \end{aligned}} \tag{28.41}$$

Equation (28.41) is quite important in the theory of semiconducting devices (Chapter 29). It asserts that the net excess of electrons (or holes) $N_d - N_a$ introduced by the impurities is almost entirely donated to the conduction (or valence) band; the other band has the very much smaller carrier density $n_i^2/(N_d - N_a)$, as required by the law of mass action, (28.24).

If the temperature is too low (or the impurity concentration too high), condition (28.36) eventually fails to hold, and either n_d/N_d or p_a/N_a (but not both) ceases to be negligible, i.e., one of the impurity types is no longer fully ionized by thermal excitation. As a result, the dominant carrier density declines with decreasing temperature (Figure 28.13).[25]

Figure 28.13
Temperature dependence of the majority carrier density (for the case $N_d > N_a$). The two high-temperature regimes are discussed in the text; the very low-temperature behavior is described in Problem 6.



IMPURITY BAND CONDUCTION

As the temperature approaches zero, so does the fraction of ionized impurities, and therefore also the density of carriers in the conduction or valence bands. Nevertheless, some small residual conductivity is observed even at the lowest temperatures. This is because the wave function of an electron (or hole) bound to an impurity site has considerable spatial extent, and therefore the overlap of wave functions at different impurity sites is possible even at fairly low concentrations. When this overlap is not negligible, it is possible for an electron to tunnel from one site to another. The resulting transport of charge is known as "impurity band conduction."

The use of the term "band" in this context is based on an analogy with the tight-binding method (Chapter 10), which shows that a set of atomic levels with a single energy can broaden into a band of energies, when wave function overlap is taken into account. The impurities, however, are usually not situated at the sites of a Bravais lattice, and one must therefore be cautious in attributing to the impurity "bands" features associated with electronic bands in *periodic* potentials.[26]