Chapter 1

Review of Energy Dispersion Relations in Solids

References:

- Ashcroft and Mermin, *Solid State Physics*, Holt, Rinehart and Winston, 1976, Chapters 8, 9, 10, 11.
- Bassani and Parravicini, *Electronic States and Optical Transitions in Solids*, Pergamon, 1975, Chapter 3.
- Kittel, *Introduction to Solid State Physics*, Wiley, 1986, pp. 228-239.
- Mott & Jones – *The Theory of the Properties of Metals and Alloys*, Dover, 1958 pp. 56–85.
- Omar, *Elementary Solid State Physics*, Addison–Wesley, 1975, pp. 189–210.
- Ziman, *Principles of the Theory of Solids*, Cambridge, 1972, Chapter 3.

1.1 Introduction

The transport properties of solids are closely related to the energy dispersion relations $E(\vec{k})$ in these materials and in particular to the behavior of $E(\vec{k})$ near the Fermi level. Conversely, the analysis of transport measurements provides a great deal of information on $E(\vec{k})$. Although transport measurements do not generally provide the most sensitive tool for studying $E(\vec{k})$, such measurements are fundamental to solid state physics because they can be carried out on nearly all materials and therefore provide a valuable tool for characterizing materials. To provide the necessary background for the discussion of transport properties, we give here a brief review of the energy dispersion relations $E(\vec{k})$ in solids. In this connection, we consider in Chapter 1 the two limiting cases of weak and tight binding. In Chapter 2 we will discuss $E(\vec{k})$ for real solids including prototype metals, semiconductors, semimetals and insulators.

1.2 One Electron $E(\vec{k})$ in Solids

1.2.1 Weak Binding or Nearly Free Electron Approximation

In the weak binding approximation, we assume that the periodic potential $V(\vec{r}) = V(\vec{r}+\vec{R_n})$ is sufficiently weak so that the electrons behave almost as if they were free and the effect of the periodic potential can be handled in perturbation theory (see Appendix A). In this formulation $V(\vec{r})$ can be an *arbitrary* periodic potential. The weak binding approximation has achieved some success in describing the valence electrons in metals. For the core electrons, however, the potential energy is comparable with the kinetic energy so that core electrons are tightly bound and the weak binding approximation is not applicable. In the weak binding approximation we solve the Schrödinger equation in the limit of a very weak periodic potential

$$\mathcal{H}\psi = E\psi. \tag{1.1}$$

Using time–independent perturbation theory (see Appendix A) we write

$$E(\vec{k}) = E^{(0)}(\vec{k}) + E^{(1)}(\vec{k}) + E^{(2)}(\vec{k}) + ... \tag{1.2}$$

and take the unperturbed solution to correspond to $V(\vec{r}) = 0$ so that $E^{(0)}(\vec{k})$ is the plane wave solution

$$E^{(0)}(\vec{k}) = \frac{\hbar^2 k^2}{2m}. \tag{1.3}$$

The corresponding normalized eigenfunctions are the plane wave states

$$\psi_{\vec{k}}^{(0)}(\vec{r}) = \frac{e^{i\vec{k}\cdot\vec{r}}}{\Omega^{1/2}} \tag{1.4}$$

in which Ω is the volume of the crystal.

The first order correction to the energy $E^{(1)}(\vec{k})$ is the diagonal matrix element of the perturbation potential taken between the unperturbed states:

$$\begin{aligned} E^{(1)}(\vec{k}) &= \langle \psi_{\vec{k}}^{(0)} \mid V(\vec{r}) \mid \psi_{\vec{k}}^{(0)} \rangle = \tfrac{1}{\Omega}\textstyle\int_\Omega e^{-i\vec{k}\cdot\vec{r}} V(\vec{r}) e^{i\vec{k}\cdot\vec{r}} d^3r \\ &= \tfrac{1}{\Omega_0}\textstyle\int_{\Omega_0} V(\vec{r}) d^3r = \overline{V(\vec{r})} \end{aligned} \tag{1.5}$$

where $\overline{V(\vec{r})}$ is independent of $\vec{k}$, and Ω_0 is the volume of the unit cell. Thus, in first order perturbation theory, we merely add a constant energy $\overline{V(\vec{r})}$ to the free particle energy, and that constant term is exactly the mean potential energy seen by the electron, averaged over the unit cell. The terms of interest arise in second order perturbation theory and are

$$E^{(2)}(\vec{k}) = \sum_{\vec{k}'}{}' \frac{|\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle|^2}{E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}')} \tag{1.6}$$

where the prime on the summation indicates that $\vec{k}' \neq \vec{k}$. We next compute the matrix element $\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle$ as follows:

$$\begin{aligned} \langle \vec{k}'|V(\vec{r})|\vec{k}\rangle &= \textstyle\int_\Omega \psi_{\vec{k}'}^{(0)*} V(\vec{r}) \psi_{\vec{k}}^{(0)} d^3r \\ &= \tfrac{1}{\Omega}\textstyle\int_\Omega e^{-i(\vec{k}'-\vec{k})\cdot\vec{r}} V(\vec{r}) d^3r \\ &= \tfrac{1}{\Omega}\textstyle\int_\Omega e^{i\vec{q}\cdot\vec{r}} V(\vec{r}) d^3r \end{aligned} \tag{1.7}$$

where $\vec{q}$ is the difference wave vector $\vec{q} = \vec{k} - \vec{k}'$ and the integration is over the whole crystal. We now exploit the periodicity of $V(\vec{r})$. Let $\vec{r} = \vec{r}\,' + \vec{R}_n$ where $\vec{r}\,'$ is an arbitrary vector in a unit cell and $\vec{R}_n$ is a lattice vector. Then because of the periodicity $V(\vec{r}) = V(\vec{r}\,')$

$$\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle = \frac{1}{\Omega}\sum_n \int_{\Omega_0} e^{i\vec{q}\cdot(\vec{r}\,'+\vec{R}_n)} V(\vec{r}\,')d^3r' \tag{1.8}$$

where the sum is over unit cells and the integration is over the volume of one unit cell. Then

$$\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle = \frac{1}{\Omega}\sum_n e^{i\vec{q}\cdot\vec{R}_n} \int_{\Omega_0} e^{i\vec{q}\cdot\vec{r}\,'} V(\vec{r}\,')d^3r'. \tag{1.9}$$

Writing the following expressions for the lattice vectors $\vec{R}_n$ and for the wave vector $\vec{q}$

$$\begin{aligned} \vec{R}_n &= \textstyle\sum_{j=1}^3 n_j \vec{a}_j \\ \vec{q} &= \textstyle\sum_{j=1}^3 \alpha_j \vec{b}_j \end{aligned} \tag{1.10}$$

where n_j is an integer, then the lattice sum $\sum_n e^{i\vec{q}\cdot\vec{R}_n}$ can be carried out exactly to yield

$$\sum_n e^{i\vec{q}\cdot\vec{R}_n} = \left[\prod_{j=1}^3 \frac{1 - e^{2\pi i N_j \alpha_j}}{1 - e^{2\pi i \alpha_j}}\right] \tag{1.11}$$

where $N = N_1N_2N_3$ is the total number of unit cells in the crystal and α_j is a real number. This sum fluctuates wildly as $\vec{q}$ varies and averages to zero. The sum is appreciable only if

$$\vec{q} = \sum_{j=1}^3 m_j \vec{b}_j \tag{1.12}$$

where m_j is an integer and $\vec{b}_j$ is a primitive vector in reciprocal space, so that $\vec{q}$ must be a reciprocal lattice vector. Hence we have

$$\sum_n e^{i\vec{q}\cdot\vec{R}_n} = N\delta_{\vec{q},\vec{G}} \tag{1.13}$$

since $\vec{b}_j \cdot \vec{R}_n = 2\pi l_{jn}$ where l_{jn} is an integer.

This discussion shows that the matrix element $\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle$ is only important when $\vec{q} = \vec{G}$ is a reciprocal lattice vector $= \vec{k} - \vec{k}'$ from which we conclude that the periodic potential $V(\vec{r})$ only connects wave vectors $\vec{k}$ and $\vec{k}'$ separated by a reciprocal lattice vector. We note that this is the same relation that determines the Brillouin zone boundary. The matrix element is then

$$\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle = \frac{N}{\Omega}\int_{\Omega_0} e^{i\vec{G}\cdot\vec{r}\,'} V(\vec{r}\,')d^3r\,'\delta_{\vec{k}'-\vec{k},\vec{G}} \tag{1.14}$$

where

$$\frac{N}{\Omega} = \frac{1}{\Omega_0} \tag{1.15}$$

and the integration is over the unit cell. We introduce $V_{\vec{G}}$ = Fourier coefficient of $V(\vec{r})$ where

$$V_{\vec{G}} = \frac{1}{\Omega_0}\int_{\Omega_0} e^{i\vec{G}\cdot\vec{r}\,'} V(\vec{r}\,')d^3r\,' \tag{1.16}$$

so that

$$\langle \vec{k}'|V(\vec{r})|\vec{k}\rangle = \delta_{\vec{k}-\vec{k}',\vec{G}}\, V_{\vec{G}}. \tag{1.17}$$

We can now use this matrix element to calculate the 2^{nd} order change in the energy based on perturbation theory (see Appendix A)

$$E^{(2)}(\vec{k}) = \sum_{\vec{G}} \frac{|V_{\vec{G}}|^2}{k^2-(k')^2}\left(\frac{2m}{\hbar^2}\right) = \frac{2m}{\hbar^2}\sum_{\vec{G}} \frac{|V_{\vec{G}}|^2}{k^2-(\vec{G}+\vec{k})^2}. \tag{1.18}$$

We observe that when $k^2 = (\vec{G}+\vec{k})^2$ the denominator in Eq. 1.18 vanishes and $E^{(2)}(\vec{k})$ can become very large. This condition is identical with the Laue diffraction condition. Thus, at a Brillouin zone boundary, the weak perturbing potential has a very large effect and therefore non–degenerate perturbation theory will not work in this case.

For $\vec{k}$ values near a Brillouin zone boundary, we must then use degenerate perturbation theory (see Appendix A). Since the matrix elements coupling the plane wave states $\vec{k}$ and $\vec{k}+\vec{G}$ do not vanish, *first-order degenerate* perturbation theory is sufficient and leads to the determinantal equation

$$\begin{vmatrix} E^{(0)}(\vec{k}) + E^{(1)}(\vec{k}) - E & \langle \vec{k}+\vec{G}|V(\vec{r})|\vec{k}\rangle \\ \langle \vec{k}|V(\vec{r})|\vec{k}+\vec{G}\rangle & E^{(0)}(\vec{k}+\vec{G}) + E^{(1)}(\vec{k}+\vec{G}) - E \end{vmatrix} = 0 \tag{1.19}$$

in which

$$\begin{aligned} E^{(0)}(\vec{k}) &= \tfrac{\hbar^2 k^2}{2m} \\ E^{(0)}(\vec{k}+\vec{G}) &= \tfrac{\hbar^2(\vec{k}+\vec{G})^2}{2m} \end{aligned} \tag{1.20}$$

and

$$\begin{aligned} E^{(1)}(\vec{k}) &= \langle \vec{k}|V(\vec{r})|\vec{k}\rangle = \overline{V(\vec{r})} = V_0 \\ E^{(1)}(\vec{k}+\vec{G}) &= \langle \vec{k}+\vec{G}|V(\vec{r})|\vec{k}+\vec{G}\rangle = V_0. \end{aligned} \tag{1.21}$$

Solution of this determinantal equation (Eq. 1.19) yields:

$$[E - V_0 - E^{(0)}(\vec{k})][E - V_0 - E^{(0)}(\vec{k}+\vec{G})] - |V_{\vec{G}}|^2 = 0, \tag{1.22}$$

or equivalently

$$E^2 - E[2V_0 + E^{(0)}(\vec{k}) + E^{(0)}(\vec{k}+\vec{G})] + [V_0 + E^{(0)}(\vec{k})][V_0 + E^{(0)}(\vec{k}+\vec{G})] - |V_{\vec{G}}|^2 = 0. \tag{1.23}$$

Solution of the quadratic equation (Eq. 1.23) yields

$$E^{\pm} = V_0 + \frac{1}{2}[E^{(0)}(\vec{k}) + E^{(0)}(\vec{k}+\vec{G})] \pm \sqrt{\frac{1}{4}[E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})]^2 + |V_{\vec{G}}|^2} \tag{1.24}$$

and we come out with two solutions for the two strongly coupled states. It is of interest to look at these two solutions in two limiting cases:

case (i) $|V_{\vec{G}}| \ll \frac{1}{2}|[E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})]|$

In this case we can expand the square root expression in Eq. 1.24 for small $|V_{\vec{G}}|$ to obtain:

$$\begin{aligned} E(\vec{k}) &= V_0 + \tfrac{1}{2}[E^{(0)}(\vec{k}) + E^{(0)}(\vec{k}+\vec{G})] \\ &\pm \tfrac{1}{2}[E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})] \cdot [1 + \tfrac{2|V_{\vec{G}}|^2}{[E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})]^2} + \ldots] \end{aligned} \tag{1.25}$$

which simplifies to the two solutions:

$$E^-(\vec{k}) = V_0 + E^{(0)}(\vec{k}) + \frac{|V_{\vec{G}}|^2}{E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})} \tag{1.26}$$

$$E^+(\vec{k}) = V_0 + E^{(0)}(\vec{k}+\vec{G}) + \frac{|V_{\vec{G}}|^2}{E^{(0)}(\vec{k}+\vec{G}) - E^{(0)}(\vec{k})} \tag{1.27}$$

and we recover the result Eq. 1.18 obtained before using non–degenerate perturbation theory. This result in Eq. 1.18 is valid far from the Brillouin zone boundary, but near the zone boundary the more complete expression of Eq. 1.24 must be used.

case (ii) $|V_{\vec{G}}| \gg \frac{1}{2}|[E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})]|$

Sufficiently close to the Brillouin zone boundary

$$|E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})| \ll |V_{\vec{G}}| \tag{1.28}$$

so that we can expand $E(\vec{k})$ as given by Eq. 1.24 to obtain

$$E^{\pm}(\vec{k}) = \frac{1}{2}[E^{(0)}(\vec{k}) + E^{(0)}(\vec{k}+\vec{G})] + V_0 \pm \left[|V_{\vec{G}}| + \frac{1}{8}\frac{[E^{(0)}(\vec{k}) - E^{(0)}(\vec{k}+\vec{G})]^2}{|V_{\vec{G}}|} + \ldots\right] \tag{1.29}$$

$$\cong \frac{1}{2}[E^{(0)}(\vec{k}) + E^{(0)}(\vec{k}+\vec{G})] + V_0 \pm |V_{\vec{G}}|, \tag{1.30}$$

so that at the Brillouin zone boundary $E^+(\vec{k})$ is elevated by $|V_{\vec{G}}|$, while $E^-(\vec{k})$ is depressed by $|V_{\vec{G}}|$ and the band gap that is formed is $2|V_{\vec{G}}|$, where $\vec{G}$ is the reciprocal lattice vector for which $E(\vec{k}_{B.Z.}) = E(\vec{k}_{B.Z.} + \vec{G})$ and

$$V_{\vec{G}} = \frac{1}{\Omega_0}\int_{\Omega_0} e^{i\vec{G}\cdot\vec{r}} V(\vec{r}) d^3r. \tag{1.31}$$

From this discussion it is clear that every Fourier component of the periodic potential gives rise to a specific band gap. We see further that the *band gap* represents a range of energy values for which there is no solution to the eigenvalue problem of Eq. 1.19 for real k (see Fig. 1.1). In the band gap we assign an imaginary value to the wave vector which can be interpreted as a highly damped and non–propagating wave.

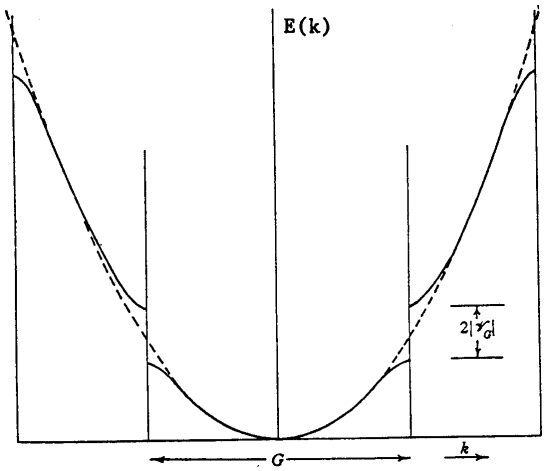


Figure 1.1: One dimensional electron energy bands for the nearly free electron model shown in the extended Brillouin zone scheme. The dashed curve corresponds to the case of free electrons and the solid curves to the case where a weak periodic potential is present. The band gaps at the zone boundaries are $2|V_{\vec{G}}|$.

We note that the larger the value of $\vec{G}$, the smaller the value of $V_{\vec{G}}$, so that higher Fourier components give rise to smaller band gaps. Near these energy discontinuities, the wave functions become linear combinations of the unperturbed states

$$\begin{aligned} \psi_{\vec{k}} &= \alpha_1 \psi^{(0)}_{\vec{k}} + \beta_1 \psi^{(0)}_{\vec{k}+\vec{G}} \\ \psi_{\vec{k}+\vec{G}} &= \alpha_2 \psi^{(0)}_{\vec{k}} + \beta_2 \psi^{(0)}_{\vec{k}+\vec{G}} \end{aligned} \tag{1.32}$$

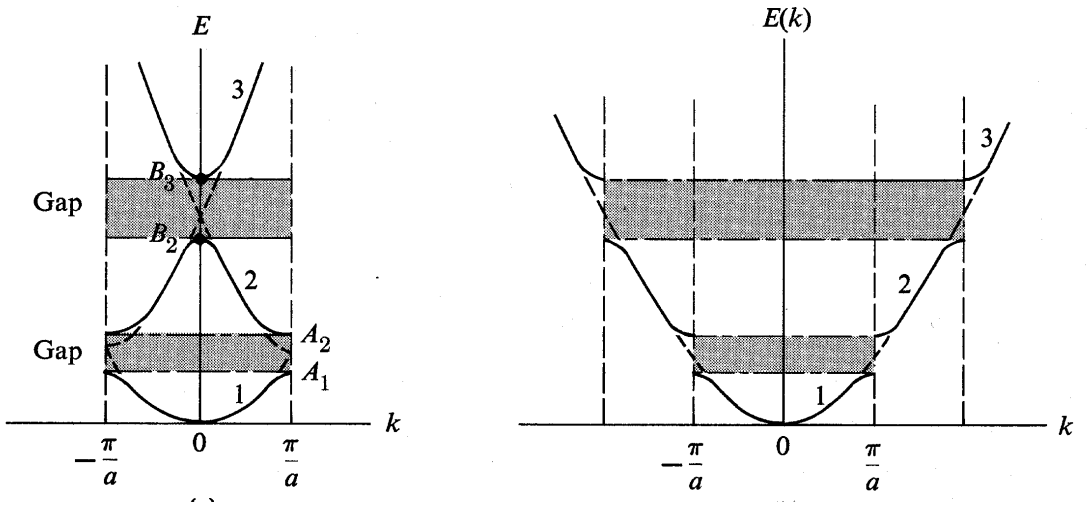
and at the zone boundary itself, instead of traveling waves $e^{i\vec{k}\cdot\vec{r}}$, the wave functions become standing waves $\cos\vec{k}\cdot\vec{r}$ and $\sin\vec{k}\cdot\vec{r}$. We note that the $\cos(\vec{k}\cdot\vec{r})$ solution corresponds to a maximum in the charge density at the lattice sites and therefore corresponds to an energy minimum (the lower level). Likewise, the $\sin(\vec{k}\cdot\vec{r})$ solution corresponds to a minimum in the charge density and therefore corresponds to a maximum in the energy, thus forming the upper level.

In constructing $E(\vec{k})$ for the reduced zone scheme we make use of the periodicity of $E(\vec{k})$ in reciprocal space

$$E(\vec{k}+\vec{G}) = E(\vec{k}). \tag{1.33}$$

The reduced zone scheme more clearly illustrates the formation of energy bands (labeled (1) and (2) in Fig. 1.2), band gaps E_g and band widths (defined in Fig. 1.2 as the range of energy between E_{min} and E_{max} for a given energy band).

We now discuss the connection between the $E(\vec{k})$ relations shown above and the transport properties of solids, which can be illustrated by considering the case of a semiconductor. An intrinsic semiconductor at temperature $T = 0$ has no carriers so that the Fermi level runs right through the band gap. On the diagram of Fig. 1.2, this would mean that the



(b) (a)

Figure 1.2: (a) One dimensional electron energy bands for the nearly free electron model shown in the extended Brillouin zone scheme for the three bands of lowest energy. (b) The same $E(\vec{k})$ as in (a) but now shown on the reduced zone scheme. The shaded areas denote the band gaps between bands n and $n+1$ and the white areas the band states.

Fermi level might run between bands (1) and (2), so that band (1) is completely occupied and band (2) is completely empty. One further property of the semiconductor is that the band gap E_g be small enough so that at some temperature (e.g., room temperature) there will be reasonable numbers of thermally excited carriers, perhaps $10^{15}/\text{cm}^3$. The doping with donor (electron donating) impurities will raise the Fermi level and doping with acceptor (electron extracting) impurities will lower the Fermi level. Neglecting for the moment the effect of impurities on the $E(\vec{k})$ relations for the perfectly periodic crystal, let us consider what happens when we raise the Fermi level into the bands. If we know the shape of the $E(\vec{k})$ curve, we are in a position to estimate the velocity of the electrons and also the so–called *effective mass* of the electrons. From the diagram in Fig. 1.2 we see that the conduction bands tend to fill up electron states starting at their energy extrema.

Since the energy bands have zero slope about their extrema, we can write $E(\vec{k})$ as a quadratic form in $\vec{k}$. It is convenient to write the proportionality in terms of the quantity called the effective mass m^*

$$E(\vec{k}) = E(0) + \frac{\hbar^2 k^2}{2m^*} \tag{1.34}$$

so that m^* is defined by

$$\frac{1}{m^*} \equiv \frac{\partial^2 E(\vec{k})}{\hbar^2 \partial k^2} \tag{1.35}$$

and we can say in some approximate way that an electron in a solid moves as if it were a free electron but with an effective mass m^* rather than a free electron mass. The larger the band curvature, the smaller the effective mass. The mean velocity of the electron is also

found from $E(\vec{k})$, according to the relation

$$\vec{v}_k = \frac{1}{\hbar}\frac{\partial E(\vec{k})}{\partial \vec{k}}. \tag{1.36}$$

For this reason the energy dispersion relations $E(\vec{k})$ are very important in the determination of the transport properties for carriers in solids.

1.2.2 Tight Binding Approximation

In the tight binding approximation a number of assumptions are made and these are different from the assumptions that are made for the weak binding approximation. The assumptions for the tight binding approximation are:

1. The energy eigenvalues and eigenfunctions are known for an electron in an isolated atom.

2. When the atoms are brought together to form a solid they remain sufficiently far apart so that each electron can be assigned to a particular atomic site. This assumption is not valid for valence electrons in metals and for this reason, these valence electrons are best treated by the weak binding approximation.

3. The periodic potential is approximated by a superposition of atomic potentials.

4. Perturbation theory can be used to treat the difference between the actual potential and the atomic potential.

Thus both the weak and tight binding approximations are based on perturbation theory. For the weak binding approximation the unperturbed state is the free electron plane–wave state, while for the tight binding approximation, the unperturbed state is the atomic state. In the case of the weak binding approximation, the perturbation Hamiltonian is the weak periodic potential itself, while for the tight binding case, the perturbation is the *difference* between the periodic potential and the atomic potential around which the electron is localized.

We review here the major features of the tight binding approximation. Let $\phi(\vec{r}-\vec{R}_n)$ represent the atomic wave function for an atom at a lattice position denoted by $\vec{R}_n$, which is measured with respect to the origin. The Schrödinger equation for an electron in an isolated atom is then:

$$\Big[-\frac{\hbar^2}{2m}\nabla^2 + U(\vec{r}-\vec{R}_n) - E^{(0)}\Big]\phi(\vec{r}-\vec{R}_n) = 0 \tag{1.37}$$

where $U(\vec{r}-\vec{R}_n)$ is the atomic potential and $E^{(0)}$ is the atomic eigenvalue (see Fig. 1.3). We now assume that the atoms are brought together to form the crystal for which $V(\vec{r})$ is the periodic potential, and $\psi(\vec{r})$ and $E(\vec{k})$ are, respectively, the wave function and energy eigenvalue for the electron in the crystal:

$$\Big[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) - E\Big]\psi(\vec{r}) = 0. \tag{1.38}$$

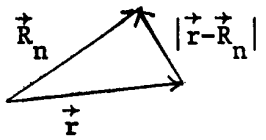


Figure 1.3: Definition of the vectors used in the tight binding approximation.

In the tight binding approximation we write $V(\vec{r})$ as a sum of atomic potentials:

$$V(\vec{r}) \simeq \sum_n U(\vec{r} - \vec{R}_n). \tag{1.39}$$

If the interaction between neighboring atoms is ignored, then each state has a degeneracy of N = number of atoms in the crystal. However, the interaction between the atoms lifts this degeneracy.

The energy eigenvalues $E(\vec{k})$ in the tight binding approximation for a non–degenerate s–state is simply given by

$$E(\vec{k}) = \frac{\langle \vec{k}|\mathcal{H}|\vec{k}\rangle}{\langle \vec{k}|\vec{k}\rangle}. \tag{1.40}$$

The normalization factor in the denominator $\langle \vec{k}|\vec{k}\rangle$ is inserted because the wave functions $\psi_{\vec{k}}(\vec{r})$ in the tight binding approximation are usually not normalized. The Hamiltonian in the tight binding approximation is written as

$$\mathcal{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) = \Big\{ -\frac{\hbar^2}{2m}\nabla^2 + [V(\vec{r}) - U(\vec{r} - \vec{R}_n)] + U(\vec{r} - \vec{R}_n) \Big\} \tag{1.41}$$

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}' \tag{1.42}$$

in which $\mathcal{H}_0$ is the atomic Hamiltonian at site n

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m}\nabla^2 + U(\vec{r} - \vec{R}_n) \tag{1.43}$$

and $\mathcal{H}'$ is the difference between the actual periodic potential and the atomic potential at lattice site n

$$\mathcal{H}' = V(\vec{r}) - U(\vec{r} - \vec{R}_n). \tag{1.44}$$

We construct the wave functions for the unperturbed problem as a linear combination of atomic functions $\phi_j(\vec{r} - \vec{R}_n)$ labeled by quantum number j

$$\psi_j(\vec{r}) = \sum_{n=1}^{N} C_{j,n}\phi_j(\vec{r} - \vec{R}_n) \tag{1.45}$$

and so that $\psi_j(\vec{r})$ is an eigenstate of a Hamiltonian satisfying the periodic potential of the lattice. In this treatment we assume that the tight binding wave–functions ψ_j can be

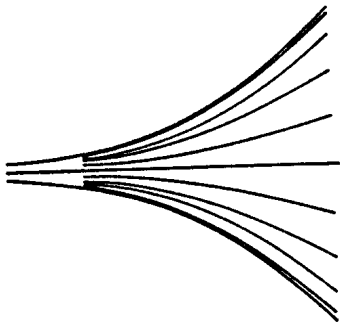


Figure 1.4: The relation between atomic states and the broadening due to the presence of neighboring atoms. As the interatomic distance decreases (going to the right in the diagram), the level broadening increases so that a band of levels occurs at atomic separations characteristic of solids.

identified with a *single* atomic state ϕ_j; this approximation must be relaxed in dealing with degenerate levels. According to Bloch's theorem, $\psi_j(\vec{r})$ in the solid must satisfy the relation:

$$\psi_j(\vec{r}+\vec{R}_m) = e^{i\vec{k}\cdot\vec{R}_m}\psi_j(\vec{r}) \tag{1.46}$$

where $\vec{R}_m$ is an arbitrary lattice vector. This restriction imposes a special form on the coefficients $C_{j,n}$.

Substitution of the expansion in atomic functions $\psi_j(\vec{r})$ from Eq. 1.45 into the left side of Eq. 1.46 yields:

$$\begin{aligned}\psi_j(\vec{r}+\vec{R}_m) &= \textstyle\sum_n C_{j,n}\ \phi_j(\vec{r}-\vec{R}_n+\vec{R}_m) \\ &= \textstyle\sum_Q C_{j,Q+m}\ \phi_j(\vec{r}-\vec{R}_Q) \\ &= \textstyle\sum_n C_{j,n+m}\ \phi(\vec{r}-\vec{R}_n)\end{aligned} \tag{1.47}$$

where we have utilized the substitution $\vec{R}_Q = \vec{R}_n - \vec{R}_m$ and the fact that Q is a dummy index. Now for the right side of the Bloch theorem (Eq. 1.46) we have

$$e^{i\vec{k}\cdot\vec{R}_m}\psi_j(\vec{r}) = \sum_n C_{j,n}e^{i\vec{k}\cdot\vec{R}_m}\phi_j(\vec{r}-\vec{R}_n). \tag{1.48}$$

The coefficients $C_{j,n}$ which relate the actual wave function $\psi_j(\vec{r})$ to the atomic functions $\phi_j(\vec{r}-\vec{R}_n)$ are therefore not arbitrary but must thus satisfy:

$$C_{j,n+m} = e^{i\vec{k}\cdot\vec{R}_m}C_{j,n} \tag{1.49}$$

which can be accomplished by setting:

$$C_{j,n} = \xi_j e^{i\vec{k}\cdot\vec{R}_n} \tag{1.50}$$

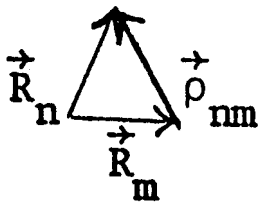


Figure 1.5: Definition of $\vec{\rho}_{nm}$ denoting the distance between atoms at $\vec{R}_m$ and $\vec{R}_n$.

where the new coefficient ξ_j is independent of n. We therefore obtain:

$$\psi_{j,\vec{k}}(\vec{r}) = \xi_j \sum_n e^{i\vec{k}\cdot\vec{R}_n} \phi_j(\vec{r} - \vec{R}_n) \tag{1.51}$$

where j is an index labeling the particular atomic state of degeneracy N and $\vec{k}$ is the quantum number for the translation operator and labels the Bloch state $\psi_{j,\vec{k}}(\vec{r})$.

For simplicity, we will limit the present discussion of the tight binding approximation to s–bands (non–degenerate atomic states) and therefore we can suppress the j index on the wave functions. (The treatment for p–bands is similar to what we will do here, but more complicated because of the degeneracy of the atomic states.) To find matrix elements of the Hamiltonian we write

$$\langle\vec{k}'|\mathcal{H}|\vec{k}\rangle = |\xi|^2 \sum_{n,m} e^{i(\vec{k}\cdot\vec{R}_n - \vec{k}'\cdot\vec{R}_m)} \int_\Omega \phi^*(\vec{r} - \vec{R}_m)\mathcal{H}\phi(\vec{r} - \vec{R}_n) d^3r \tag{1.52}$$

in which the integration is carried out throughout the volume of the crystal. Since $\mathcal{H}$ is a function which is periodic in the lattice, the only significant distance (see Fig. 1.5) is

$$(\vec{R}_n - \vec{R}_m) = \vec{\rho}_{nm}. \tag{1.53}$$

We then write the integral in Eq. 1.52 as:

$$\langle\vec{k}'|\mathcal{H}|\vec{k}\rangle = |\xi|^2 \sum_{\vec{R}_m} e^{i(\vec{k}-\vec{k}')\cdot\vec{R}_m} \sum_{\vec{\rho}_{nm}} e^{i\vec{k}\vec{\rho}_{nm}} \mathcal{H}_{mn}(\vec{\rho}_{nm}) \tag{1.54}$$

where we have written the matrix element $\mathcal{H}_{mn}(\vec{\rho}_{nm})$ as

$$\mathcal{H}_{mn}(\vec{\rho}_{nm}) = \int_\Omega \phi^*(\vec{r} - \vec{R}_m)\mathcal{H}\phi(\vec{r} - \vec{R}_m - \vec{\rho}_{nm}) d^3r = \int_\Omega \phi^*(\vec{r}\,')\mathcal{H}\phi(\vec{r}\,' - \vec{\rho}_{nm}) d^3r'. \tag{1.55}$$

We note here that the integral in Eq. 1.55 depends only on $\vec{\rho}_{nm}$ and not on $\vec{R}_m$. According to Eq. 1.13, the first sum in Eq. 1.54 is

$$\sum_{\vec{R}_m} e^{i(\vec{k}-\vec{k}')\cdot\vec{R}_m} = \delta_{\vec{k}',\vec{k}+\vec{G}} N \tag{1.56}$$

where $\vec{G}$ is a reciprocal lattice vector. It is convenient to restrict the $\vec{k}$ vectors to lie within the first Brillouin zone (i.e., we limit ourselves to reduced wave vectors). This is consistent with the manner of counting states for a crystal with periodic boundary conditions of length d on a side

$$k_i d = 2\pi m_i \qquad \text{for each direction } i \tag{1.57}$$

where m_i is an integer in the range $1 \leq m_i < N_i$ where $N_i \approx N^{1/3}$ and N is the total number of unit cells in the crystal. From Eq. 1.57 we have

$$k_i = \frac{2\pi m_i}{d}. \tag{1.58}$$

The maximum value that a particular m_i can assume is N_i and the maximum value for k_i is $2\pi/a$ at the Brillouin zone boundary since $N_i/d = 1/a$. With this restriction, $\vec{k}$ and $\vec{k}'$ must both lie within the 1^{st} B.Z. and thus cannot differ by any reciprocal lattice vector other than $\vec{G} = 0$. We thus obtain the following form for the matrix element of $\mathcal{H}$ (and also the corresponding forms for the matrix elements of $\mathcal{H}_0$ and $\mathcal{H}'$):

$$\langle \vec{k}'|\mathcal{H}|\vec{k}\rangle = |\xi|^2 N \delta_{\vec{k},\vec{k}'} \sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}} \mathcal{H}_{mn}(\vec{\rho}_{nm}) \tag{1.59}$$

yielding the result

$$E(\vec{k}) = \frac{\langle \vec{k}|\mathcal{H}|\vec{k}\rangle}{\langle \vec{k}|\vec{k}\rangle} = \frac{\sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}} \mathcal{H}_{mn}(\vec{\rho}_{nm})}{\sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}} \mathcal{S}_{mn}(\vec{\rho}_{nm})} \tag{1.60}$$

in which

$$\langle \vec{k}'|\vec{k}\rangle = |\xi|^2 \delta_{\vec{k},\vec{k}'} N \sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}} \mathcal{S}_{mn}(\vec{\rho}_{nm}) \tag{1.61}$$

where the matrix element $\mathcal{S}_{mn}(\vec{\rho}_{nm})$ measures the overlap of atomic functions on different sites

$$\mathcal{S}_{mn}(\vec{\rho}_{nm}) = \int_{\Omega} \phi^*(\vec{r})\phi(\vec{r} - \vec{\rho}_{nm}) d^3r. \tag{1.62}$$

The overlap integral $\mathcal{S}_{mn}(\vec{\rho}_{nm})$ will be nearly 1 when $\vec{\rho}_{nm} = 0$ and will fall off rapidly as $\vec{\rho}_{nm}$ increases, which exemplifies the spirit of the tight binding approximation. By selecting $\vec{k}$ vectors to lie within the first Brillouin zone, the orthogonality condition on the wave function $\psi_{\vec{k}}(\vec{r})$ is automatically satisfied. Writing $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}'$ yields:

$$\begin{aligned} \mathcal{H}_{mn} \quad &= \int_{\Omega} \phi^*(\vec{r} - \vec{R}_m) \left[-\tfrac{\hbar^2}{2m}\nabla^2 + U(\vec{r} - \vec{R}_n) \right] \phi(\vec{r} - \vec{R}_n) d^3r \\ &+ \int_{\Omega} \phi^*(\vec{r} - \vec{R}_m)[V(\vec{r}) - U(\vec{r} - \vec{R}_n)]\phi(\vec{r} - \vec{R}_n) d^3r \end{aligned} \tag{1.63}$$

or

$$\mathcal{H}_{mn} = E^{(0)} \mathcal{S}_{mn}(\vec{\rho}_{nm}) + \mathcal{H}'_{mn}(\vec{\rho}_{nm}) \tag{1.64}$$

which results in the general expression for the tight binding approximation:

$$E(\vec{k}) = E^{(0)} + \frac{\sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}} \mathcal{H}'_{mn}(\vec{\rho}_{nm})}{\sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}} \mathcal{S}_{mn}(\vec{\rho}_{nm})}. \tag{1.65}$$

In the spirit of the tight binding approximation, the second term in Eq. 1.65 is assumed to be small, which is a good approximation if the overlap of the atomic wave functions is small. We classify the sum over $\vec{\rho}_{nm}$ according to the distance between site m and site n: (i) zero distance, (ii) the nearest neighbor distance, (iii) the next nearest neighbor distance, etc.

$$\sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}}\mathcal{H}'_{mn}(\vec{\rho}_{nm}) = \mathcal{H}'_{nn}(0) + \sum_{\vec{\rho}_1} e^{i\vec{k}\cdot\vec{\rho}_{nm}}\mathcal{H}'_{mn}(\vec{\rho}_{nm}) + ... \tag{1.66}$$

The zeroth neighbor term $\mathcal{H}'_{nn}(0)$ in Eq. 1.66 results in a constant additive energy, independent of $\vec{k}$. The sum over nearest neighbor distances $\vec{\rho}_1$ gives rise to a $\vec{k}$–dependent perturbation, and hence is of particular interest in calculating the band structure. The terms $\mathcal{H}'_{nn}(0)$ and the sum over the nearest neighbor terms in Eq. 1.66 are of comparable magnitude, as can be seen by the following argument. In the integral

$$\mathcal{H}'_{nn}(0) = \int \phi^*(\vec{r}-\vec{R}_n)[V - U(\vec{r}-\vec{R}_n)]\phi(\vec{r}-\vec{R}_n)d^3r \tag{1.67}$$

we note that $|\phi(\vec{r}-\vec{R}_n)|^2$ has an appreciable amplitude only in the vicinity of the site $\vec{R}_n$. But at site $\vec{R}_n$, the potential energy term $[V - U(\vec{r}-\vec{R}_n)] = \mathcal{H}'$ is a small term, so that $\mathcal{H}'_{nn}(0)$ represents the product of a small term times a large term. On the other hand, the integral $\mathcal{H}'_{mn}(\vec{\rho}_{nm})$ taken over nearest neighbor distances has a factor $[V - U(\vec{r}-\vec{R}_n)]$ which is large near the mth site; however, in this case the wave functions $\phi^*(\vec{r}-\vec{R}_m)$ and $\phi(\vec{r}-\vec{R}_n)$ are on different atomic sites and have only a small overlap on nearest neighbor sites. Therefore $\mathcal{H}'_{mn}(\vec{\rho}_{nm})$ over nearest neighbor sites also results in the product of a large quantity times a small quantity.

In treating the denominator in the perturbation term of Eq. 1.65, we must sum

$$\sum_{\vec{\rho}_{nm}} e^{i\vec{k}\cdot\vec{\rho}_{nm}}\mathcal{S}_{mn}(\vec{\rho}_{nm}) = \mathcal{S}_{nn}(0) + \sum_{\vec{\rho}_1} e^{i\vec{k}\cdot\vec{\rho}_{nm}}\mathcal{S}_{mn}(\vec{\rho}_{nm}) + ... \tag{1.68}$$

In this case the leading term $\mathcal{S}_{nn}(0)$ is approximately unity and the overlap integral $\mathcal{S}_{mn}(\vec{\rho}_{nm})$ over nearest neighbor sites is small, and can be neglected to lowest order in comparison with unity. The nearest neighbor term in Eq. 1.68 is of comparable relative magnitude to the next nearest neighbor terms arising from $\mathcal{H}_{mn}(\vec{\rho}_{nm})$ in Eq. 1.66.

We will here make *several explicit evaluations* of $E(\vec{k})$ in the tight–binding limit to show how this method incorporates the crystal symmetry. For illustrative purposes we will give results for the simple cubic lattice (SC), the body centered cubic (BCC) and face centered cubic lattice (FCC). We shall assume here that the overlap of atomic potentials on neighboring sites is sufficiently weak so that only nearest neighbor terms need be considered in the sum on $\mathcal{H}'_{mn}$ and only the leading term need be considered in the sum of $\mathcal{S}_{mn}$.

For the simple cubic structure there are 6 terms in the nearest neighbor sum on $\mathcal{H}'_{mn}$ in Eq. 1.65 with $\vec{\rho}_1$ vectors given by:

$$\vec{\rho}_1 = a(\pm 1, 0, 0),\quad a(0, \pm 1, 0),\quad a(0, 0, \pm 1). \tag{1.69}$$

By symmetry, $\mathcal{H}'_{mn}(\vec{\rho}_1)$ is the same for all of the $\vec{\rho}_1$ vectors so that

$$E(\vec{k}) = E^{(0)} + \mathcal{H}'_{nn}(0) + 2\mathcal{H}'_{mn}(\vec{\rho}_1)[\cos k_x a + \cos k_y a + \cos k_z a] + ... \tag{1.70}$$

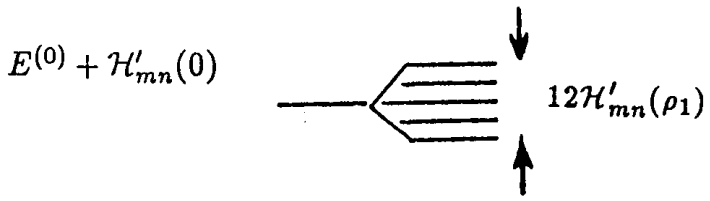


Figure 1.6: The relation between the atomic levels and the broadened level in the tight binding approximation.

where $\vec{\rho}_1$ = the nearest neighbor separation and k_x, k_y, k_z are components of the wave vector $\vec{k}$ in the first Brillouin zone.

This dispersion relation $E(\vec{k})$ clearly satisfies three properties which characterize the energy eigenvalues in typical periodic structures:

1. Periodicity in $\vec{k}$ space under translation by a reciprocal lattice vector $\vec{k} \rightarrow \vec{k} + \vec{G}$,
2. $E(\vec{k})$ is an even function of $\vec{k}$ (i.e., $E(k) = E(-k)$)
3. $\partial E/\partial k = 0$ at the Brillouin zone boundary

In the above expression (Eq. 1.70) for $E(\vec{k})$, the maximum value for the term in brackets is $\pm$ 3. Therefore for a simple cubic lattice in the tight binding approximation we obtain a bandwidth of 12 $\mathcal{H}'_{mn}(\rho_1)$ from nearest neighbor interactions as shown in Fig. 1.6.

Because of the different locations of the nearest neighbor atoms in the case of the BCC and FCC lattices, the expression for $E(\vec{k})$ will be different for the various cubic lattices. Thus the form of the tight binding approximation explicitly takes account of the crystal structure. The results for the simple cubic, body centered cubic and face centered cubic lattices are summarized below.

simple cubic

$$E(\vec{k}) = \text{ const} + 2\mathcal{H}'_{mn}(\vec{\rho}_1)[\cos k_x a + \cos k_y a + \cos k_z a] + ... \tag{1.71}$$

body centered cubic

The eight $\vec{\rho}_1$ vectors for the nearest neighbor distances in the BCC structure are $(\pm a/2, \pm a/2, \pm a/2)$ so that there are 8 exponential terms which combine in pairs such as:

$$\left[\exp\frac{ik_x a}{2}\exp\frac{ik_y a}{2}\exp\frac{ik_z a}{2} + \exp\frac{-ik_x a}{2}\exp\frac{ik_y a}{2}\exp\frac{ik_z a}{2}\right] \tag{1.72}$$

to yield

$$2\cos(\frac{k_x a}{2})\exp\frac{ik_y a}{2}\exp\frac{ik_z a}{2}. \tag{1.73}$$

We thus obtain for the BCC structure:

$$E(\vec{k}) = \text{ const} + 8\mathcal{H}'_{mn}(\vec{\rho}_1)\cos(\frac{k_x a}{2})\cos(\frac{k_y a}{2})\cos(\frac{k_z a}{2}) + ... \tag{1.74}$$

where $\mathcal{H}'_{mn}(\vec{\rho}_1)$ is the matrix element of the perturbation Hamiltonian taken between nearest neighbor atomic orbitals.

face centered cubic

For the FCC structure there are 12 nearest neighbor distances $\vec{\rho}_1$: $(0, \pm\frac{a}{2}, \pm\frac{a}{2})$, $(\pm\frac{a}{2}, \pm\frac{a}{2}, 0)$, $(\pm\frac{a}{2}, 0, \pm\frac{a}{2})$, so that the twelve exponential terms combine in groups of 4 to yield:

$$\exp\tfrac{ik_xa}{2}\exp\tfrac{ik_ya}{2}+\exp\tfrac{ik_xa}{2}\exp\tfrac{-ik_ya}{2}+\exp\tfrac{-ik_xa}{2}\exp\tfrac{ik_ya}{2}+\exp\tfrac{-ik_xa}{2}\exp\tfrac{-ik_ya}{2}= 4\cos(\tfrac{k_xa}{2})\cos(\tfrac{k_ya}{2}), \tag{1.75}$$

thus resulting in the energy dispersion relation

$$E(\vec{k}) = \text{ const}+4\mathcal{H}'_{mn}(\vec{\rho}_1)\left[\cos(\frac{k_ya}{2})\cos(\frac{k_za}{2})+\cos(\frac{k_xa}{2})\cos(\frac{k_za}{2})+\cos(\frac{k_xa}{2})\cos(\frac{k_ya}{2})\right]+... \tag{1.76}$$

We note that $E(\vec{k})$ for the FCC is different from that for the SC or BCC structures. The tight-binding approximation has symmetry considerations built into its formulation through the symmetrical arrangement of the atoms in the lattice. The situation is quite different in the weak binding approximation where symmetry enters into the form of $V(\vec{r})$ and determines which Fourier components $V_{\vec{G}}$ will be important in creating band gaps.

1.2.3 Weak and Tight Binding Approximations

We will now make some general statements about bandwidths and forbidden band gaps which follow from either the tight binding or weak binding (nearly free electron) approximations. With increasing energy, the bandwidth tends to increase. On the tight–binding picture, the higher energy atomic states are less closely bound to the nucleus, and the resulting increased overlap of the wave functions results in a larger value for $\mathcal{H}'_{mn}(\vec{\rho}_1)$ in the case of the higher atomic states: that is, for silicon, which has 4 valence electrons in the $n = 3$ shell, the overlap integral $\mathcal{H}'_{mn}(\vec{\rho}_1)$ will be smaller than for germanium which is isoelectronic to silicon but has instead 4 valence electrons in the $n = 4$ atomic shell. On the weak–binding picture, the same result follows, since for higher energies, the electrons are more nearly free; therefore, there are more allowed energy ranges available, or equivalently, the energy range of the forbidden states is smaller. Also in the weak–binding approximation the band gap of $2|V_{\vec{G}}|$ tends to decrease as $\vec{G}$ increases, because of the oscillatory character of $e^{-i\vec{G}\cdot\vec{r}}$ in

$$V_{\vec{G}} = \frac{1}{\Omega_0}\int_{\Omega_0} e^{-i\vec{G}\cdot\vec{r}}V(\vec{r})d^3r. \tag{1.77}$$

From the point of view of the tight–binding approximation, the increasing bandwidth with increasing energy (see Fig. 1.7) is also equivalent to a decrease in the forbidden band gap. At the same time, the atomic states at higher energies become more closely spaced, so that the increased bandwidth eventually results in band overlaps. When band overlaps occur, the tight-binding approximation as given above must be generalized to treat coupled or interacting bands using degenerate perturbation theory (see Appendix A).

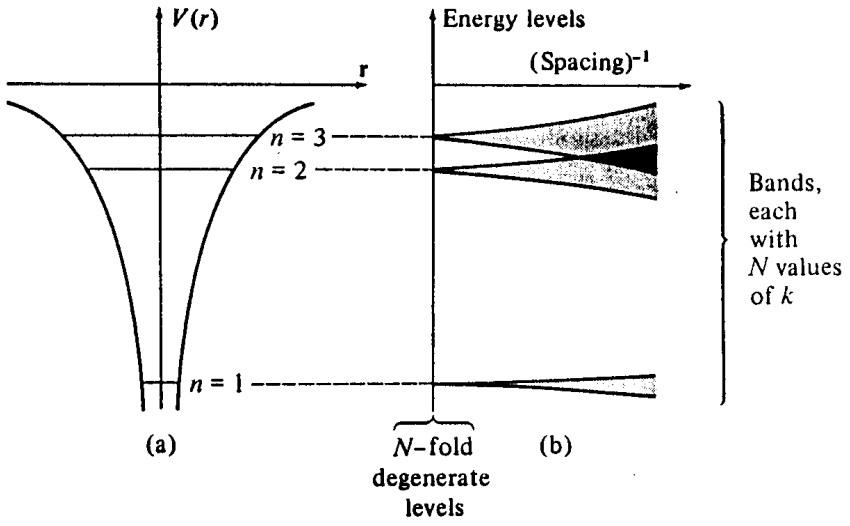


Figure 1.7: Schematic diagram of (a) the quantized energy levels and (b) the increased bandwidth and decreased band gap in the tight binding approximation as the interatomic separation decreases.

1.2.4 Tight Binding Approximation with 2 Atoms/Unit Cell

We present here a simple example of the tight binding approximation for a simplified version of polyacetylene which has two carbon atoms (with their appended hydrogens) per unit cell. In Fig. 1.8 we show, within the box defined by the dotted lines, the unit cell for *trans*-polyacetylene $(CH)_x$. This unit cell of an infinite one-dimensional chain contains two inequivalent carbon atoms, A and B. There is one π-electron per carbon atom, thus giving rise to two π-energy bands in the first Brillouin zone. These two bands are called bonding π-bands for the valence band, and anti-bonding π-bands for the conduction band.

The lattice unit vector and the reciprocal lattice unit vector of this one-dimensional polyacetylene chain are given by $\vec{a}_1 = (a, 0, 0)$ and $\vec{b}_1 = (2\pi/a, 0, 0)$, respectively. The Brillouin zone in 1D is the line segment $-\pi/a < k < \pi/a$ and the Brillouin zone boundary is at $k = \pm\pi/a$. The Bloch orbitals consisting of A and B atoms are given by

$$\psi_j(r) = \frac{1}{\sqrt{N}} \sum_{R_\alpha} e^{ikR_\alpha} \phi_j(r - R_\alpha), \quad (\alpha = \mathrm{A}, \mathrm{B}) \tag{1.78}$$

where the summation is taken over the atom site coordinate R_α for the A or B carbon atoms in the solid.

To solve for the energy eigenvalues and wavefunctions we need to solve the general equation:

$$\mathcal{H}\psi = E\mathcal{S}\psi \tag{1.79}$$

where $\mathcal{H}$ is the $n \times n$ tight binding matrix Hamiltonian for the n coupled bands ($n = 2$ in the case of polyacetylene) and $\mathcal{S}$ is the corresponding $n \times n$ overlap integral matrix. To obtain a solution to this matrix equation, we require that the determinant $|\mathcal{H} - E\mathcal{S}|$ vanish.

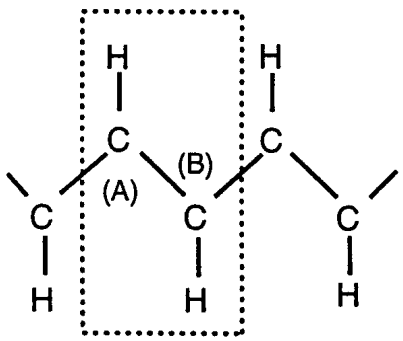


Figure 1.8: The unit cell of *trans*-polyacetylene bounded by a box defined by the dotted lines, and showing two inequivalent carbon atoms, A and B, in the unit cell.

This approach is easily generalized to periodic structures with more than 2 atoms per unit cell.

The (2×2) matrix Hamiltonian, $\mathcal{H}_{\alpha\beta}$, $(\alpha, \beta = A, B)$ is obtained by substituting Eq. (1.78) into

$$\mathcal{H}_{jj'}(\vec{k}) = \langle\psi_j|\mathcal{H}|\psi_{j'}\rangle, \quad \mathcal{S}_{jj'}(\vec{k}) = \langle\psi_j|\psi_{j'}\rangle \qquad (j, j' = 1, 2), \tag{1.80}$$

where the integrals over the Bloch orbitals, $\mathcal{H}_{jj'}(\vec{k})$ and $\mathcal{S}_{jj'}(\vec{k})$, are called transfer integral matrices and overlap integral matrices, respectively. When $\alpha = \beta =$ A, we obtain the diagonal matrix element

$$\begin{aligned}
\mathcal{H}_{AA}(r) &= \frac{1}{N}\sum_{R,R'} e^{ik(R-R')}\langle\phi_A(r-R')|\mathcal{H}|\phi_A(r-R)\rangle \\
&= \frac{1}{N}\sum_{R'=R} E_{2p} + \frac{1}{N}\sum_{R'=R\pm a} e^{\pm ika}\langle\phi_A(r-R')|\mathcal{H}|\phi_A(r-R)\rangle \\
&\quad +(\text{terms equal to or more distant than } R' = R \pm 2a) \\
&= E_{2p} + (\text{terms equal to or more distant than } R' = R \pm a).
\end{aligned} \tag{1.81}$$

In Eq. (1.81) the main contribution to the matrix element $\mathcal{H}_{AA}$ comes from $R' = R$, and this gives the orbital energy of the $2p$ level, E_{2p}. We note that E_{2p} is not simply the atomic energy value for the free atom, because the Hamiltonian $\mathcal{H}$ also includes a crystal potential contribution. The next order contribution to $\mathcal{H}_{AA}$ in Eq. (1.81) comes from terms in $R' = R \pm a$, which are here neglected for simplicity. Similarly, $\mathcal{H}_{BB}$ also gives E_{2p} to the same order of approximation.

Next let us consider the off-diagonal matrix element $\mathcal{H}_{AB}(r)$ which explicitly couples the A unit to the B unit. The largest contribution to $\mathcal{H}_{AB}(r)$ arises when atoms A and B are nearest neighbors. Thus in the summation over R', we only consider the terms with

$R' = R \pm a/2$ as a first approximation and neglect more distant terms to obtain

$$\begin{aligned} \mathcal{H}_{AB}(r) &= \frac{1}{N}\sum_R \Big\{ e^{-ika/2}\langle\phi_A(r-R)|\mathcal{H}|\phi_B(r-R-a/2)\rangle \\ &\quad + e^{ika/2}\langle\phi_A(r-R)|\mathcal{H}|\phi_B(r-R+a/2)\rangle \Big\} \\ &= 2t\cos(ka/2) \end{aligned} \tag{1.82}$$

where t is the transfer integral appearing in Eq. (1.82) and is denoted by

$$t = \langle\phi_A(r-R)|\mathcal{H}|\phi_B(r-R\pm a/2)\rangle. \tag{1.83}$$

Here we have assumed that all the π bonding orbitals are of equal length (1.5Å bonds). In the real $(CH)_x$ compound, bond alternation occurs, in which the bonding between adjacent carbon atoms alternates between single bonds (1.7Å) and double bonds (1.3Å). With this bond alternation, the two matrix elements between atomic wavefunctions in Eq. (1.82) are not equal. Although the distortion of the lattice lowers the total energy, the electronic energy always decreases more than the lattice energy in a one-dimensional material. This distortion deforms the lattice by a process called the Peierls instability. This instability arises for example when a distortion is introduced into a system containing a previously degenerate system with 2 equivalent atoms per unit cell. The distortion making the atoms inequivalent increases the unit cell by a factor of 2 and decreases the reciprocal lattice by a factor of 2. If the energy band was formally half filled, a band gap is introduced by the Peierls instability at the Fermi level, which lowers the total energy of the system. It is stressed here that t has a negative value which means that t is an attractive potential that bonds atoms together to form a condensed state of matter. The matrix element $\mathcal{H}_{BA}(r)$ is obtained from $\mathcal{H}_{AB}(r)$ through the Hermitian conjugation relation $\mathcal{H}_{BA} = \mathcal{H}^*_{AB}$, but since $\mathcal{H}_{AB}$ is real in this case, we obtain $\mathcal{H}_{BA} = \mathcal{H}_{AB}$.

The overlap matrix $\mathcal{S}_{ij}$ can be calculated by a similar method as was used for $\mathcal{H}_{ij}$, except that the intra-atomic integral $\mathcal{S}_{ij}$ yields a unit matrix in the limit of large interatomic distances, if we assume that the atomic wavefunction is normalized so that $\mathcal{S}_{AA} = \mathcal{S}_{BB} = 1$. It is assumed that for polyacetylene, the $\mathcal{S}_{AA}$ and $\mathcal{S}_{BB}$ matrix elements are still approximately unity. For the off-diagonal matrix element for polyacetylene we have $\mathcal{S}_{AB} = \mathcal{S}_{BA} = 2s\cos(ka/2)$, where s is an overlap integral between the nearest A and B atoms,

$$s = \langle\phi_A(r-R)|\phi_B(r-R\pm a/2)\rangle. \tag{1.84}$$

The secular equation for the $2p_z$ orbital of CH_x is obtained by setting the determinant of $|\mathcal{H} - E\mathcal{S}|$ to zero to obtain

$$\begin{aligned} &\begin{vmatrix} E_{2p}-E & 2(t-sE)\cos(ka/2) \\ 2(t-sE)\cos(ka/2) & E_{2p}-E \end{vmatrix} \\ &= (E_{2p}-E)^2 - 4(t-sE)^2\cos^2(ka/2) \\ &= 0 \end{aligned} \tag{1.85}$$

yielding the eigenvalues of the energy dispersion relations of Eq. (1.85)

$$E_\pm(\vec{k}) = \frac{E_{2p} \pm 2t\cos(ka/2)}{1 \pm 2s\cos(ka/2)}, \quad (-\frac{\pi}{a} < k < \frac{\pi}{a}) \tag{1.86}$$

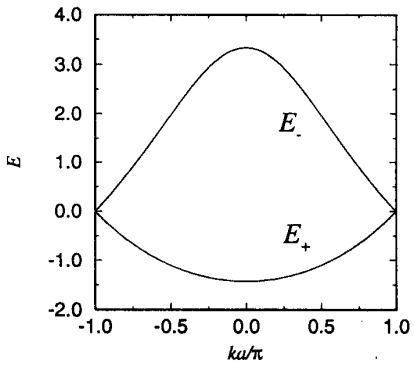


Figure 1.9: The energy dispersion relation $E_{\pm}(\vec{k})$ for polyacetylene $[(\mathrm{CH})_x]$, given by Eq. (1.86) with values for the parameters $t = -1$ and $s = 0.2$. Curves $E_{+}(\vec{k})$ and $E_{-}(\vec{k})$ are called bonding π and antibonding π^* energy bands, respectively, and the energy is plotted in units of t.

in which the + sign is associated with the bonding π-band and the $-$ sign is associated with the antibonding π^*-band, as shown in Fig. 1.9. Here it is noted that by setting E_{2p} to zero (thereby defining the origin of the energy), the levels E_{+} and E_{-} are degenerate at $ka = \pm\pi$. Figure 1.9 is constructed for $t < 0$ and $s > 0$. Since there are two π electrons per unit cell, each with a different spin orientation, both electrons occupy the bonding π energy band. The effect of the inter-atomic bonding is to lower the total energy below E_{2p}.

Chapter 2

Examples of Energy Bands in Solids

References

- J.C. Slater - Quantum Theory of Atoms and Molecules, Chapter 10.
- R.E. Peierls - Quantum Theory of Solids, Chapter 4
- F. Bassani and G. Pastori Paravicini - Electronic States and Optical Transitions in Solids, Chapter 4

2.1 General Issues

We present here some examples of energy bands which are representative of metals, semiconductors and insulators, and we point out some of the characteristic features in each case. Figure 2.1 distinguishes in a schematic way between insulators (a), metals (b), semimetals (c), a thermally excited semiconductor (d) for which at $T = 0$ all states in the valence band are occupied and all states in the conduction band are unoccupied, assuming no impurities or crystal defects. Finally in Fig. 2.1(e), we see a p-doped semiconductor which is deficient in electrons, not having sufficient electrons to fill the valence band completely as in (d). The semiconductor (e) will have a non-zero carrier density at $T = 0$ while for semiconductor (d) the carrier density will be zero at $T = 0$.

Figure 2.2 shows a schematic view of the electron dispersion relations for an insulator (a), while (c) shows dispersion relations for a metal. In the case of Fig. 2.2(b), we have a semimetal if the number of electrons equals the number of holes, but a metal otherwise.

In this chapter we examine a number of representative $E(\vec{k})$ diagrams for illustrative materials. For each of the $E(\vec{k})$ diagrams we consider the following questions:

1. Is the material a metal, a semiconductor (direct or indirect gap), semimetal or insulator?
2. To which atomic (molecular) levels do the bands on the band diagram correspond? Which bands are important in determining the electronic structure? What are the bandwidths, bandgaps?

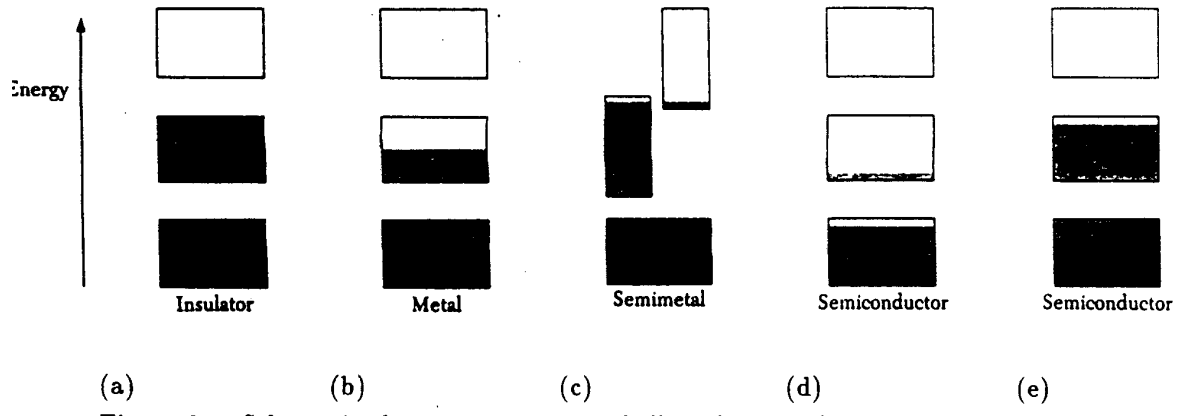


Figure 2.1: Schematic electron occupancy of allowed energy bands for an insulator, metal, semimetal and semiconductor. The vertical extent of the boxes indicates the allowed energy regions: the shaded areas indicate the regions filled with electrons. In a semimetal (such as bismuth) one band is almost filled and another band is nearly empty at a temperature of absolute zero. A pure semiconductor (such as silicon) becomes an insulator at $T = 0$. Panel (d) shows an intrinsic semiconductor at a finite temperature, with carriers that are thermally excited. Panel (e) shows a p-doped semiconductor that is electron-deficient, as, for example, because of the introduction of acceptor impurities.

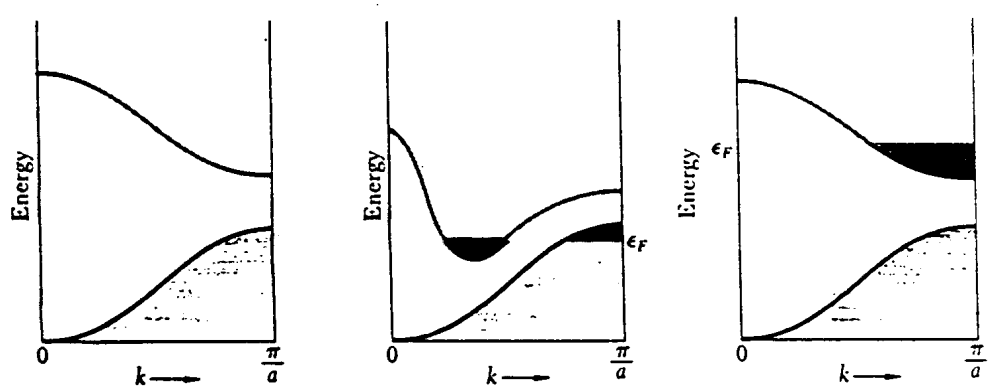


Figure 2.2: Occupied states and band structures giving (a) an insulator, (b) a metal or a semimetal because of band overlap, and (c) a metal because of partial occupation of an electron band. In (b) the band overlap for a 3D solid need not occur along the same direction of the wave vector in the Brillouin zone.

3. What information does $E(\vec{k})$ diagram provide concerning the following questions:

   (a) Where are the carriers in the Brillouin zone?

   (b) Are the carriers electrons or holes?

   (c) Are there many or few carriers?

   (d) How many carrier pockets of each type are there in the Brillouin zone?

   (e) What is the shape of the Fermi surface?

   (f) Are the carrier velocities high or low?

   (g) Are the carrier mobilities for each carrier pocket high or low?

4. What information is provided concerning the optical properties?

   (a) Where in the Brillouin Zone is the threshold for optical transitions?

   (b) At what photon energy does the optical threshold occur?

   (c) For semiconductors, does the threshold correspond to a direct gap or an indirect gap (phonon-assisted) transition?

2.2 Metals

2.2.1 Alkali Metals–e.g., Sodium

For the alkali metals the valence electrons are nearly free and the weak binding approximation describes these electrons quite well. The Fermi surface is nearly spherical and the band gaps are small. The crystal structure for the alkali metals is body centered cubic (BCC) and the $E(\vec{k})$ diagram is drawn starting with the bottom of the half–filled conduction band. For example, the $E(\vec{k})$ diagram in Fig. 2.3 for sodium begins at ~ -0.6 Rydberg and represents the $3s$ conduction band. The filled valence bands lie much lower in energy and are not shown in Fig. 2.3.

For the case of sodium, the $3s$ conduction band is very nearly free electron–like and the $E(\vec{k})$ relations are closely isotropic. Thus the $E(\vec{k})$ relations along the $\Delta(100)$, $\Sigma(110)$ and $\Lambda(111)$ directions [see Fig. 2.3(b)] are essentially coincident and can be so plotted, as shown in Fig. 2.3(a). For these metals, the Fermi level is determined so that the $3s$ band is exactly half–occupied, since the Brillouin zone is large enough to accommodate 2 electrons per unit cell. Thus the radius of the Fermi surface k_F satisfies the relation

$$\frac{4}{3}\pi k_F^3 = \frac{1}{2}V_{\text{B.Z.}} = \frac{1}{2}(2)(\frac{2\pi}{a})^3, \text{ or } \quad \frac{k_F\, a}{2\pi} \sim 0.63, \tag{2.1}$$

where $V_{\text{B.Z.}}$ and a are, respectively, the volume of the Brillouin zone and the lattice constant. For the alkali metals, the effective mass m^* is nearly equal to the free electron mass m_0 and the Fermi surface is nearly spherical and never comes close to the Brillouin zone boundary. The zone boundary for the Σ, Λ and Δ directions are indicated in the $E(\vec{k})$ diagram of Fig. 2.3 by vertical lines. For the alkali metals, the band gaps are very small compared to the band widths and the $E(\vec{k})$ relations are parabolic ($E = \hbar^2k^2/2m^*$) almost up to the Brillouin zone boundaries. By comparing $E(\vec{k})$ for Na with the BCC empty lattice bands (see Fig. 2.4) for which the potential $V(r) = 0$, we can see the effect of the very weak periodic

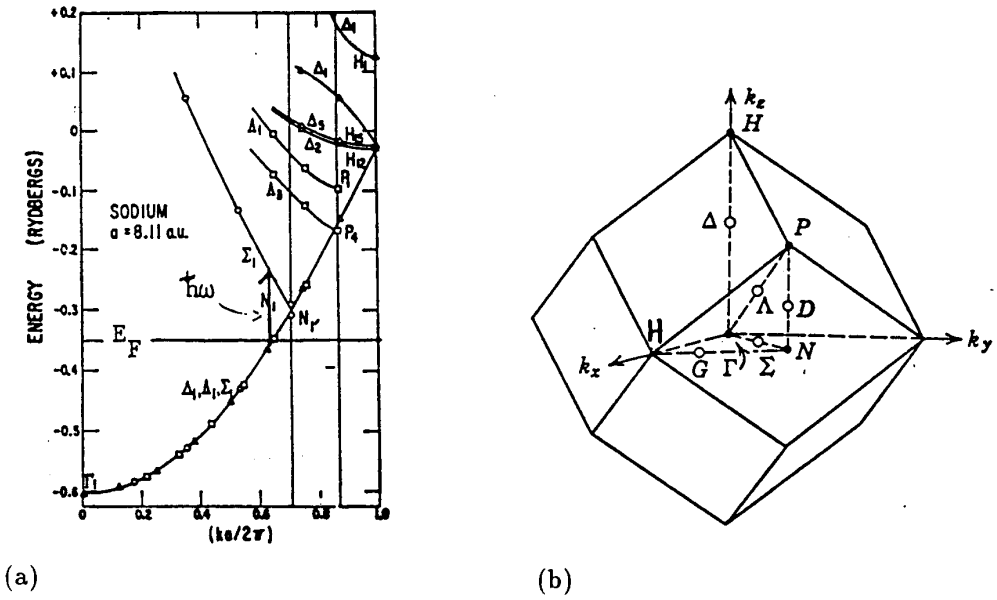


Figure 2.3: (a) Energy dispersion relations for the nearly free electron metal sodium which has an atomic configuration $1s^2 2s^2 2p^6 3s$. (b) The Brillouin zone for the BCC lattice showing the high symmetry points and axes. Sodium can be considered as a prototype alkali metal crystalline solid for discussing the dispersion relations for nearly free electron metals.

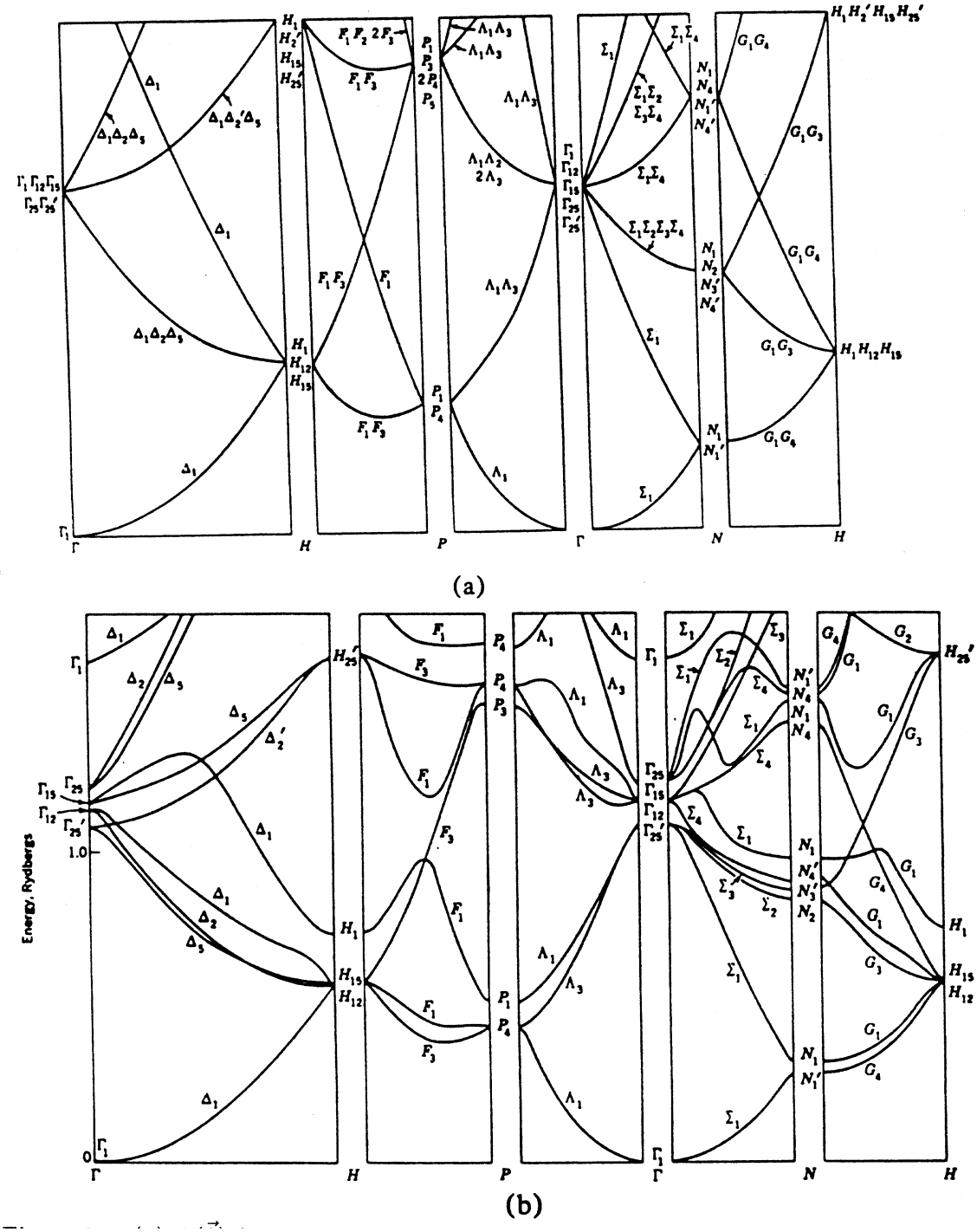


Figure 2.4: (a) $E(\vec{k})$ for a BCC lattice in the empty lattice approximation, $V \equiv 0$. (b) $E(\vec{k})$ for sodium, showing the effect of a weak periodic potential in lifting accidental band degeneracies at $k = 0$ and at the zone boundaries (high symmetry points) in the Brillouin zone. Note that the splittings are quite different for the various bands and at different high symmetry points.

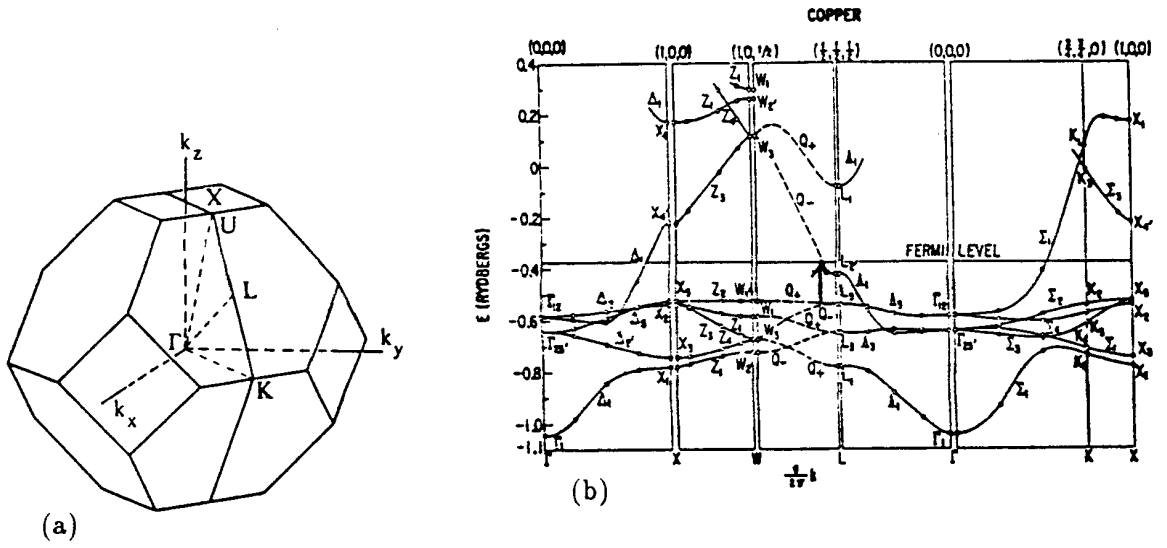


Figure 2.5: (a) Brillouin zone for a FCC lattice showing high symmetry points. (b) The calculated energy bands for copper along the various symmetry axes of the FCC Brillouin zone shown in (a).

potential in partially lifting the band degeneracy at the various high symmetry points in the Brillouin zone. The threshold for optical transitions corresponds to photons having sufficient energy to take an electron from an occupied state at k_F to an unoccupied state at k_F, since the wave vector for photons is very small compared with the Fermi wave vector k_F and since wave vector conservation (i.e., crystal momentum conservation) is required for optical transitions. The threshold for optical transitions is indicated by $\hbar\omega$ and a vertical arrow in Fig. 2.3. Because of the low density of initial and final states for a given energy separation, we would expect optical interband transitions for alkali metals to be very weak and this is in agreement with experimental observations for all the alkali metals. The notation a.u. in Fig. 2.3 stands for atomic units and expresses lattice constants in units of Bohr radii. The electron energy is given in Rydbergs where 1 Rydberg = 13.6 eV, the ionization energy of a hydrogen atom.

2.2.2 Noble Metals

The noble metals are copper, silver and gold and they crystallize in a face centered cubic (FCC) structure; the usual notation for the high symmetry points in the FCC Brillouin zone are shown on the diagram in Fig. 2.5(a). As in the case of the alkali metals, the noble metals have one valence electron/atom and therefore one electron per primitive unit cell. However, the free electron picture does not work so well for the noble metals, as you can see by looking at the energy band diagram for copper given in Fig. 2.5(b).

In the case of copper, the bands near the Fermi level are derived from the $4s$ and $3d$ atomic levels. The so-called $4s$ and $3d$ bands accommodate a total of 12 electrons, while the number of available electrons is 11. Therefore the Fermi level must cross these bands.

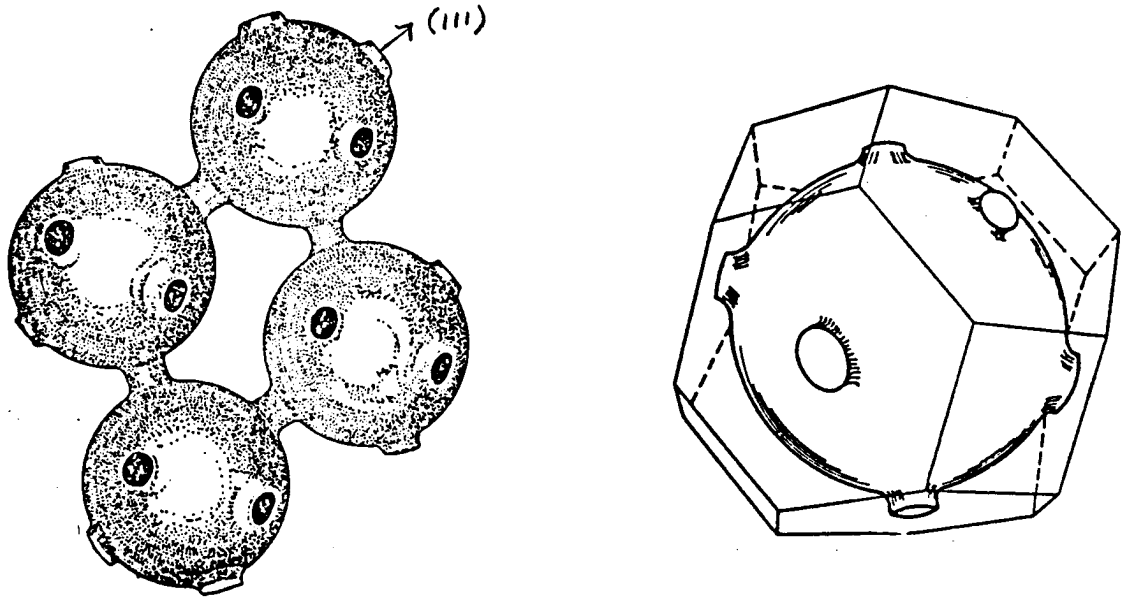


Figure 2.6: (a) The copper Fermi surface in the extended zone scheme. (b) A sketch of the Fermi surface of copper inscribed within the FCC Brillouin zone.

Consequently copper is metallic. In Fig. 2.5(b) we see that the $3d$ bands are relatively flat and show little dependence on wave vector $\vec{k}$. We can trace the $3d$ bands by starting at $\vec{k} = 0$ with the $\Gamma_{25'}$ and Γ_{12} levels. On the other hand, the $4s$ band has a strong k–dependence and a large curvature. This band can be traced by starting at $\vec{k} = 0$ with the Γ_1 level. About halfway between Γ and X, the $4s$ level approaches the $3d$ levels and *mixing* or *hybridization* occurs. As we further approach the X–point, we can again pick up the $4s$ band (beyond where the interaction with the $3d$ bands occurs) because of its high curvature. This $4s$ band eventually crosses the Fermi level before reaching the Brillouin Zone boundary at the X point. A similar mixing or hybridization between $4s$ and $3d$ bands occurs in going from Γ to L, except that in this case the $4s$ band reaches the Brillouin Zone boundary *before* crossing the Fermi level.

Of particular significance for the transport properties of copper is the band gap that opens up at the L–point. In this case, the band gap is between the $L_{2'}$ level below the Fermi level E_F and the L_1 level above E_F. Since this bandgap is comparable with the typical bandwidths in copper, we cannot expect the Fermi surface to be free electron–like. By looking at the energy bands $E(\vec{k})$ along the major high symmetry directions such as the (100), (110) and (111) directions, we can readily trace the origin of the copper Fermi surface [see Fig. 2.6(a)]. Here we see basically a spherical Fermi surface with necks pulled out in the (111) directions and making contact with the Brillouin zone boundary through these necks, thereby linking the Fermi surface in one zone to that in the next zone in the extended zone scheme. In the (100) direction, the cross section of the Fermi surface is nearly circular, indicative of the nearly parabolic $E(\vec{k})$ relation of the $4s$ band at the Fermi

level in going from Γ to X. In contrast, in going from Γ to L, the $4s$ band never crosses the Fermi level. Instead the $4s$ level is depressed from the free electron parabolic curve as the Brillouin zone boundary is reached, thereby producing a higher density of states. Thus, near the zone boundary, more electrons can be accommodated per unit energy range, or to say this another way, there will be increasingly more $\vec{k}$ vectors with approximately the same energy. This causes the constant energy surfaces to be pulled out in the direction of the Brillouin zone boundary [see Fig. 2.6(b)]. This "pulling out" effect follows both from the weak binding and tight binding approximations and the effect is more pronounced as the strength of the periodic potential (or $V_{\vec{G}}$) increases.

If the periodic potential is sufficiently strong so that the resulting bandgap at the zone boundary straddles the Fermi level, as occurs at the L–point in copper, the Fermi surface makes contact with the Brillouin zone boundary. The resulting Fermi surfaces are called *open surfaces* because the Fermi surfaces between neighboring Brillouin zones are connected, as seen in Fig. 2.6(a). The electrons associated with the necks are contained in the electron pocket shown in the $E(\vec{k})$ diagram away from the L–point in the LW direction which is $\perp$ to the {111} direction. The copper Fermi surface shown in Fig. 2.6(a) bounds *electron* states. Hole pockets are formed in copper [see Fig. 2.6(a)] in the extended zone and constitute the unoccupied space between the electron surfaces. Direct evidence for hole pockets is provided by Fermi surface measurements to be described later in this course.

From the $E(\vec{k})$ diagram for copper [Fig. 2.5(b)] we see that the threshold for optical interband transitions occurs for photon energies sufficient to take an electron at constant $\vec{k}$–vector from a filled $3d$ level to an unoccupied state above the Fermi level. Such interband transitions can be made near the L–point in the Brillouin zone [as shown by the vertical arrow on Fig. 2.5(b)]. Because of the high density of initial states in the d–band, these transitions will be quite intense. The occurrence of these interband transitions at ~ 2 eV gives rise to a large absorption of electromagnetic energy in this photon energy region. The reddish color of copper metal is thus due to a higher reflectivity for photons in the red (below the threshold for interband transitions) than for photons in the blue (above this threshold).

2.2.3 Polyvalent Metals

The simplest example of a polyvalent metal is aluminum with 3 electrons/atom and having a $3s^2 3p$ electronic configuration for the valence electrons. (As far as the number of electrons/atom is concerned, two electrons/atom completely fill a non-degenerate band—one for spin up, the other for spin down.) Because of the partial filling of the $3s^2 3p^6$ bands, aluminum is a metal. Aluminum crystallizes in the FCC structure so we can use the same notation as for the Brillouin zone in Fig. 2.5(a). The energy bands for aluminum (see Fig. 2.7) are very free electron–like. This follows from the small magnitudes of the band gaps relative to the band widths on the energy band diagram shown in Fig. 2.7. The lowest valence band shown in Fig. 2.7 is the $3s$ band which can be traced by starting at zero energy at the Γ point ($\vec{k} = 0$) and going out to X_4 at the X–point, to W_3 at the W–point, to L_2' at the L–point and back to Γ_1 at the Γ point ($\vec{k} = 0$). Since this band always lies below the Fermi level, it is completely filled, containing 2 electrons. The third valence electron partially occupies the second and third p–bands (which are more accurately described as hybridized $3p$–bands with some admixture of the $3s$ bands with which they interact). From

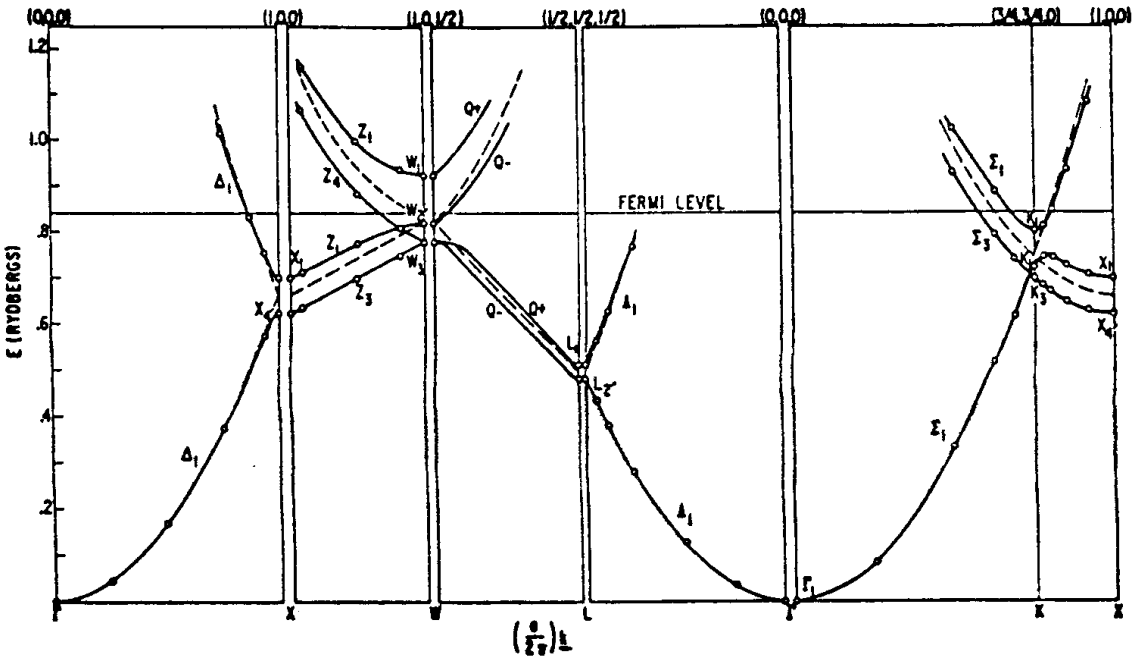


Figure 2.7: Electronic energy band diagram for aluminum which crystallizes in a FCC structure. The dashed lines correspond to the free electron model and the solid curves include the effect of the periodic potential $V(\vec{r})$.

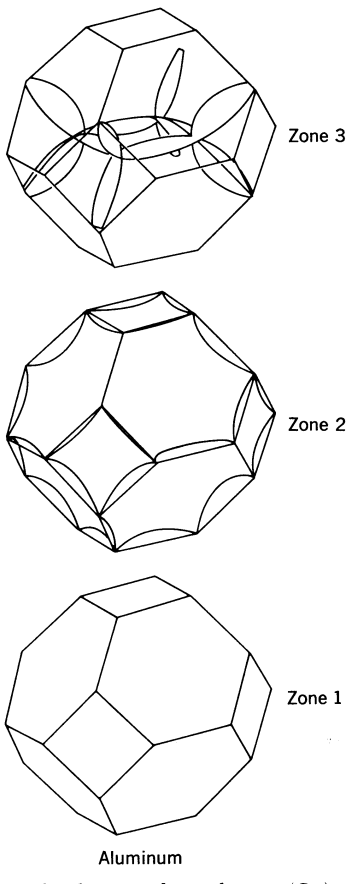


Figure 2.8: The three valence electrons for aluminum occupy three Brillouin zones. Zone 1 is completely occupied. Zone 2 in nearly filled with electrons and is best described as a hole surface, where the holes occupy the interior portion of the second zone, shown in the figure. Zone 3 is a complex electron structure with occupied electron states near the Brillouin zone boundaries, and the occupied states are shown only in part for clarity.

Fig. 2.7 we can see that the second band is partly filled; the occupied states extend from the Brillouin zone boundary inward toward the center of the zone; this can be seen in going from the X point to Γ, on the curve labeled Δ_1. Since the second band states near the center of the Brillouin zone remain unoccupied, the volume enclosed by the Fermi surface in the second band is a hole pocket. The aluminum Fermi surface showing the Zone 2 holes is presented in Fig. 2.8. Because $E(\vec{k})$ for the second band in the vicinity of E_F is free electron–like, the masses for the holes are approximately equal to the free electron mass.

The 3^{rd} zone electron pockets are small and are found around the K– and W–points as can be seen in Fig. 2.7. These electron pockets are $\vec{k}$ space volumes that enclose electron states (see Fig. 2.8), and because of the large curvature of $E(\vec{k})$, these electrons have relatively small masses. Figure 2.7 gives no evidence for any 4^{th} zone pieces of Fermi surface, and for this reason we can conclude that all the electrons are either in the second band or in the third band as shown in Fig. 2.8. The total electron concentration is sufficient to exactly

fill a half of the volume of the Brillouin zone V_{BZ}:

$$V_{e,2} + V_{e,3} = \frac{V_{BZ}}{2}. \tag{2.2}$$

With regard to the second zone, it is partially filled with electrons and the rest of the zone is empty (since holes correspond to the unfilled states):

$$V_{h,2} + V_{e,2} = V_{BZ}, \tag{2.3}$$

with the volume that is empty slightly exceeding the volume that is occupied. Therefore we focus attention on the more dominant second zone holes. Substitution of Eq. 2.2 into Eq. 2.3 then yields for the second zone holes and the third zone electrons

$$V_{h,2} - V_{e,3} = \frac{V_{BZ}}{2} \tag{2.4}$$

where the subscripts e, h on the volumes in $\vec{k}$ space refer to electrons and holes and the Brillouin zone (B.Z.) index is given for each of the carrier pockets. Because of the small masses and high mobility of the 3^{rd} zone electrons, they play a more important role in the transport properties of aluminum than would be expected from their small numbers.

From the $E(\vec{k})$ diagram in Fig. 2.7 we see that at the same $\vec{k}$–points (near the K– and W–points in the Brillouin zone) there are occupied $3s$ levels and unoccupied $3p$ levels separated by $\sim$ 1eV. From this we conclude that optical interband transitions should be observable in the 1eV photon energy range. Such interband transitions are in fact observed experimentally and are responsible for the departures from nearly perfect reflectivity of aluminum mirrors in the vicinity of 1 eV.

2.3 Semiconductors

Assume that we have a semiconductor at $T = 0$ K with no impurities. The Fermi level will then lie within a band gap. Under these conditions, there are no carriers, and no Fermi surface. We now illustrate the energy band structure for several representative semiconductors in the limit of $T = 0$K and no impurities. Semiconductors having no impurities or defects are called *intrinsic* semiconductors.

2.3.1 PbTe

In Fig. 2.9 we illustrate the energy bands for PbTe. This direct gap semiconductor [see Fig. 2.10(a)] is chosen initially for illustrative purposes because the energy bands in the valence and conduction bands that are of most importance to determine the physical properties of PbTe are non-degenerate. Therefore, the energy states in PbTe near E_F are simpler to understand than for the more common semiconductors silicon and germanium, and for many of the III–V and II–VI compound semiconductors.

In Fig. 2.9, we show the position of E_F for the idealized conditions of the intrinsic (no carriers at $T = 0$) semiconductor PbTe. From a diagram like this, we can obtain a great deal of information which could be useful for making semiconductor devices. For example, we can calculate effective masses from the band curvatures, and electron velocities from the slopes of the $E(\vec{k})$ dispersion relations shown in Fig. 2.9.

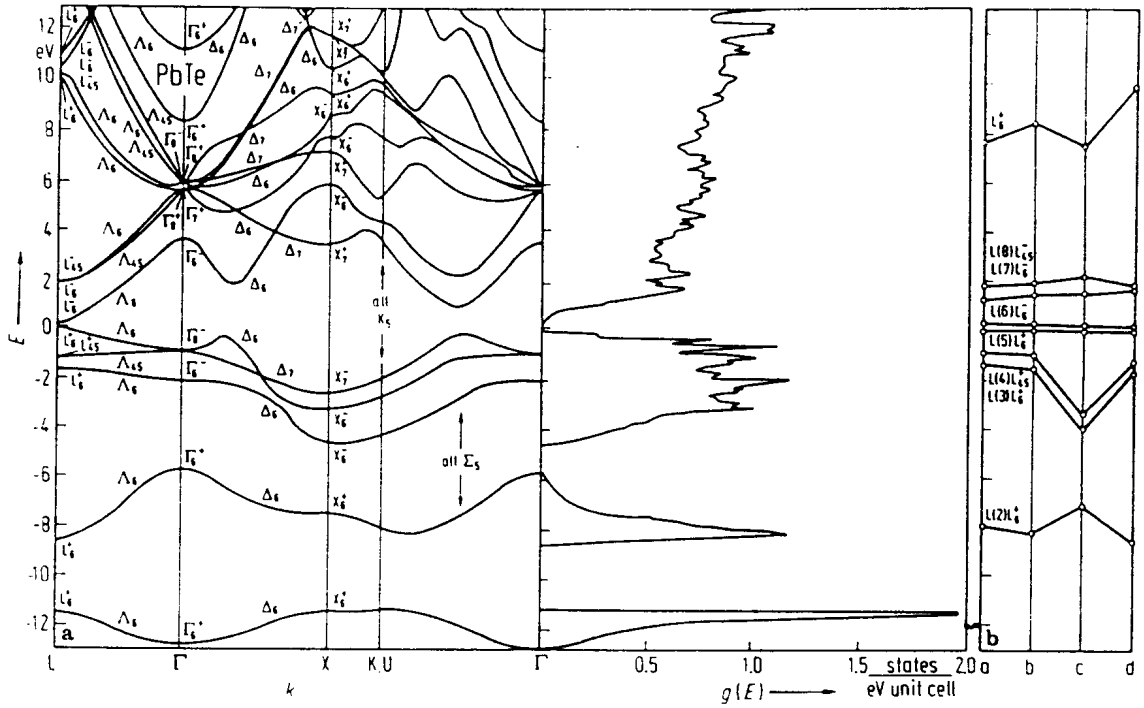


Figure 2.9: (a) Energy band structure and density of states for PbTe obtained from an empirical pseudopotential calculation. (b) Theoretical values for the L point bands calculated by different models (labeled a, b, c, d on the x-axis) (Ref. Landolt and Bornstein).

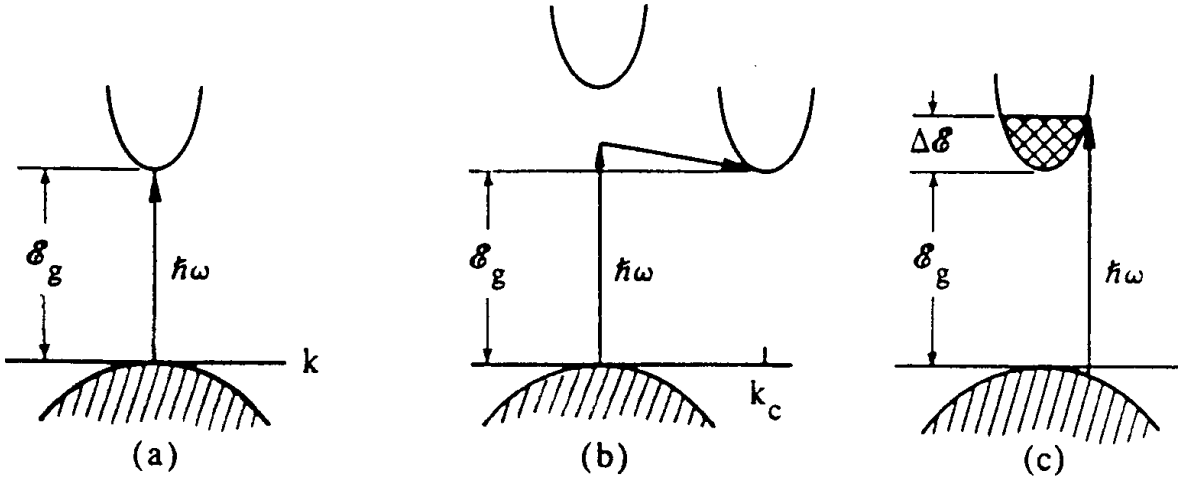


Figure 2.10: Optical absorption processes for (a) a direct band gap semiconductor, (b) an indirect band gap semiconductor, and (c) a direct band gap semiconductor with the conduction band filled to the level shown.

Suppose we add impurities (e.g., donor impurities) to PbTe. The donor impurities will raise the Fermi level and an electron pocket will eventually be formed in the L_6^- conduction band about the L–point. This electron pocket will have an ellipsoidal Fermi surface because the band curvature is different as we move away from the L point in the $L\Gamma$ direction as compared with the band curvature as we move away from L on the Brillouin zone boundary containing the L point (e.g., LW direction). Figure 2.9 shows $E(\vec{k})$ from L to Γ corresponding to the (111) direction. Since the effective masses

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2}\frac{\partial^2 E(\vec{k})}{\partial k_i \partial k_j} \tag{2.5}$$

for both the valence and conduction bands in the longitudinal $L\Gamma$ direction are heavier than in the LK and LW directions, the ellipsoids of revolution describing the carrier pockets are prolate for both holes and electrons. The L and Σ point room temperature band gaps are 0.311 eV and 0.360 eV, respectively. For the electrons, the effective mass parameters are $m_\perp = 0.053m_e$ and $m_\parallel = 0.620m_e$. The experimental hole effective masses at the L point are $m_\perp = 0.0246m_e$ and $m_\parallel = 0.236m_e$ and at the Σ point, the hole effective mass values are $m_\perp = 0.124m_e$ and $m_\parallel = 1.24m_e$. Thus for the L-point carrier pockets, the semi-major axis of the constant energy surface along $L\Gamma$ will be longer than along LK. From the $E(\vec{k})$ diagram for PbTe in Fig. 2.9 one would expect that hole carriers could be thermally excited to a second band at the Σ point, which is indicated on the $E(\vec{k})$ diagram. At room temperature, these Σ point hole carriers contribute significantly to the transport properties.

Because of the small gap (0.311 eV) in PbTe at the L–point, the threshold for interband transitions will occur at infrared frequencies. PbTe crystals can be prepared either p–type or n–type, but never perfectly stoichiometrically (i.e., intrinsic PbTe has not been prepared). Therefore, at room temperature the Fermi level E_F often lies in either the valence or conduction band for actual PbTe crystals. Since optical transitions conserve wavevector, the interband transitions will occur at k_F [see Fig. 2.10(c)] and at a higher photon energy than the direct band gap. This increase in the threshold energy for interband transitions in *degenerate* semiconductors (where E_F lies within either the valence or conduction bands) is called the *Burstein shift*.

2.3.2 Germanium

We will next look at the $E(\vec{k})$ relations for: (1) the group IV semiconductors which crystallize in the diamond structure and (2) the closely related III–V compound semiconductors which crystallize in the zincblende structure (see Fig. 2.11 for a schematic diagram for this class of semiconductors). These semiconductors have degenerate valence bands at $\vec{k} = 0$ [see Fig. 2.11(d)] and for this reason have more complicated $E(\vec{k})$ relations for carriers than is the case for the lead salts discussed in §B.1.1. The $E(\vec{k})$ diagram for germanium is shown in Fig. 2.12. Ge is a semiconductor with a bandgap occurring between the top of the valence band at $\Gamma_{25'}$, and the bottom of the lowest conduction band at L_1. Since the valence and conduction band extrema occur at *different* points in the Brillouin zone, Ge is an *indirect gap* semiconductor [see Fig. 2.10(b)]. Using the same arguments as were given in §B.1.1 for the Fermi surface of PbTe, we see that the constant energy surfaces for electrons in germanium are ellipsoids of revolution [see Fig. 2.11(c)]. As for the case of PbTe, the ellipsoids of revolution are elongated along ΓL which is the heavy mass direction in this case.

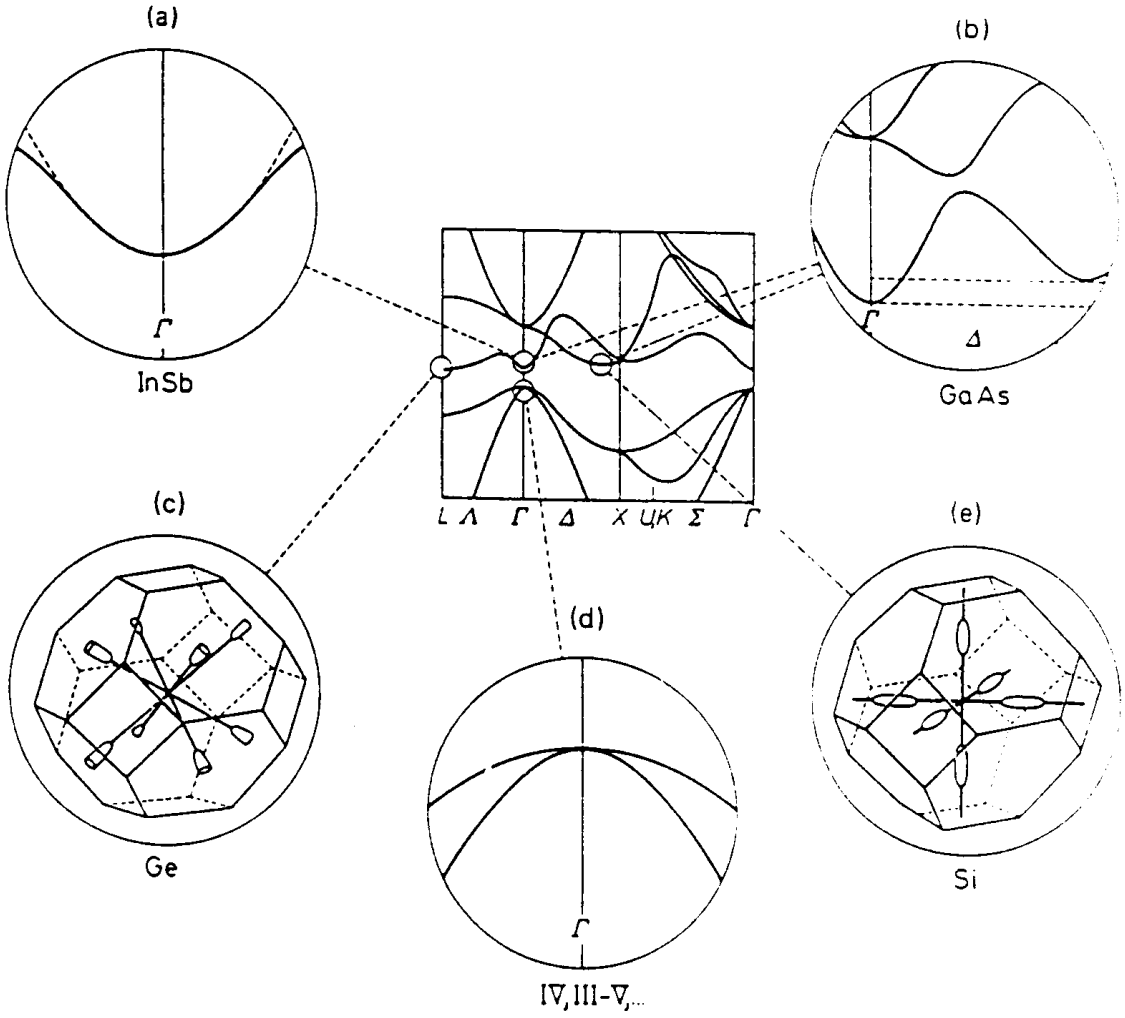


Figure 2.11: Important details of the band structure of typical group IV and III–V semiconductors.

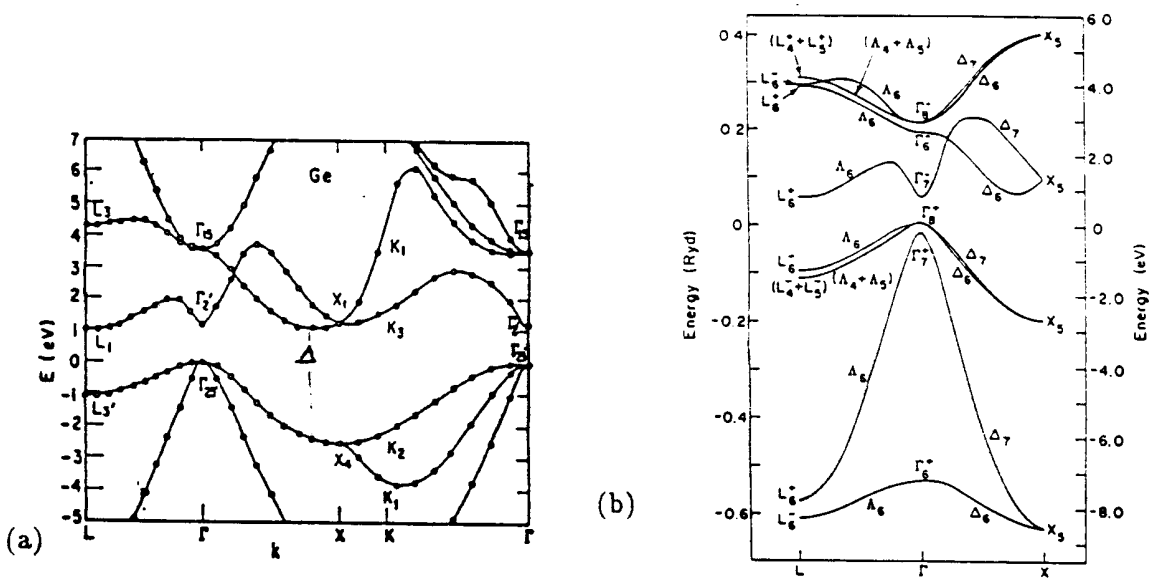


Figure 2.12: Electronic energy band structure of Ge (a) without spin-orbit interaction. (b) The electronic energy bands near $k = 0$ when the spin-orbit interaction is included.

Since the multiplicity of L–points is 8, we have 8 half–ellipsoids of this kind within the first Brillouin zone, just as for the case of PbTe. By translation of these half–ellipsoids by a reciprocal lattice vector we can form 4 full–ellipsoids. The $E(\vec{k})$ diagram for germanium (see Fig. 2.12) further shows that the next highest conduction band above the L point minimum is at the Γ–point ($\vec{k}$=0) and after that along the ΓX axis at a point commonly labeled as a Δ–point. Because of the degeneracy of the highest valence band, the Fermi surface for holes in germanium is more complicated than for electrons. The lowest direct band gap in germanium is at $\vec{k} = 0$ between the $\Gamma_{25'}$ valence band and the $\Gamma_{2'}$ conduction band. From the $E(\vec{k})$ diagram we note that the electron effective mass for the $\Gamma_{2'}$ conduction band is very small because of the high curvature of the $\Gamma_{2'}$ band about $\vec{k} = 0$, and this effective mass is isotropic so that the constant energy surfaces are spheres.

The optical properties for germanium show a very weak optical absorption for photon energies corresponding to the indirect gap (see Fig. 2.13). Since the valence and conduction band extrema occur at a different $\vec{k}$–point in the Brillouin zone, the indirect gap excitation requires a phonon to conserve crystal momentum. Hence the threshold for this indirect transition is

$$(\hbar\omega)_{\text{threshold}} = E_{L_1} - E_{\Gamma_{25'}} - E_{\text{phonon}}. \tag{2.6}$$

The optical absorption for germanium increases rapidly above the photon energy corresponding to the direct band gap $E_{\Gamma_{2'}} - E_{\Gamma_{25'}}$, because of the higher probability for the direct optical excitation process. However, the absorption here remains low compared with the absorption at yet higher photon energies because of the low density of states for the Γ–point transition, as seen from the $E(\vec{k})$ diagram. Very high optical absorption, however, occurs for photon energies corresponding to the energy separation between the $L_{3'}$ and L_1

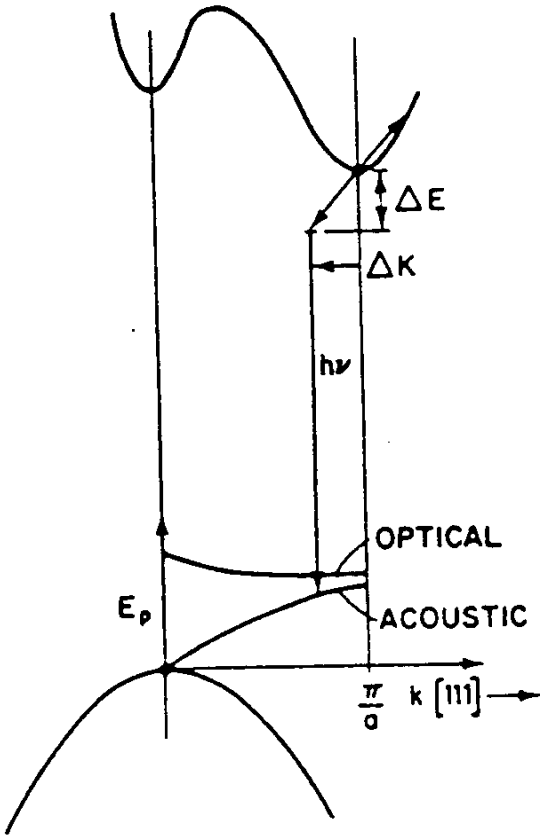


Figure 2.13: Illustration of the indirect emission of light due to carriers and phonons in Ge. [$h\nu$ is the photon energy; ΔE is the energy delivered to an electron; E_p is the energy delivered to the lattice (phonon energy)].

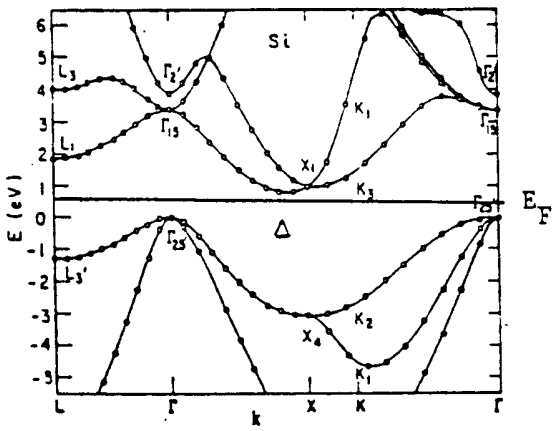


Figure 2.14: Electronic energy band structure of Si.

bands which is approximately the same for a large range of $\vec{k}$ values, thereby giving rise to a very large joint density of states (the number of states with constant energy separation per unit energy range). A large *joint density of states* arising from the *tracking* of conduction and valence bands is found for germanium, silicon and the III–V compound semiconductors, and for this reason these materials tend to have high dielectric constants (to be discussed in Part II of this course which focuses on optical properties).

2.3.3 Silicon

From the energy band diagram for silicon shown in Fig. 2.14, we see that the energy bands of Si are quite similar to those for germanium. They do, however, differ in detail. For example, in the case of silicon, the electron pockets are formed around a Δ point located along the ΓX (100) direction. For silicon there are 6 electron pockets within the first Brillouin zone instead of the 8 half–pockets which occur in germanium. The constant energy surfaces are again ellipsoids of revolution with a heavy longitudinal mass and a light transverse effective mass [see Fig. 2.11(e)]. The second type of electron pocket that is energetically favored is about the L_1 point, but to fill electrons there, we would need to raise the Fermi energy by $\sim 1\,\text{eV}$.

Silicon is of course the most important semiconductor for device applications and is at the heart of semiconductor technology for transistors, integrated circuits, and many electronic devices. The optical properties of silicon also have many similarities to those in germanium, but show differences in detail. For Si, the indirect gap [see Fig. 2.10(b)] occurs at $\sim 1\,\text{eV}$ and is between the $\Gamma_{25'}$ valence band and the Δ conduction band extrema. Just as in the case for germanium, strong optical absorption occurs for large volumes of the Brillouin zone at energies comparable to the $L_{3'} \rightarrow L_1$ energy separation, because of the "tracking" of the valence and conduction bands. The density of electron states for Si covering a wide energy range is shown in Fig. 2.15 where the corresponding energy band

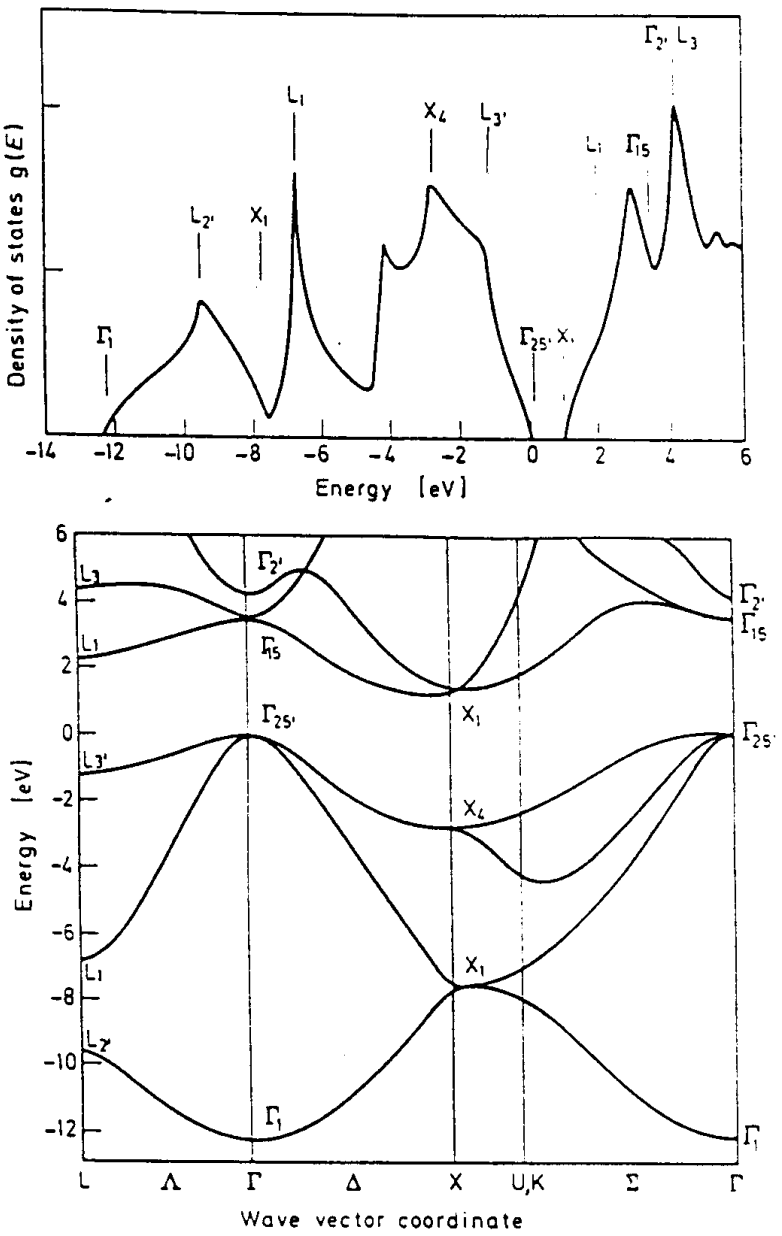


Figure 2.15: (a) Density of states in the valence and conduction bands of silicon, and (b) the corresponding $E(\vec{k})$ curves showing the symbols of the high symmetry points of the band structure.

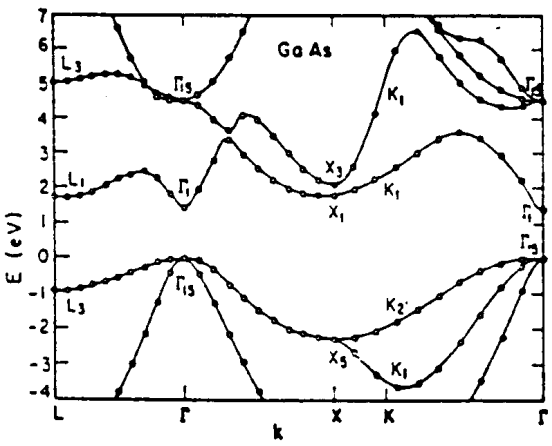


Figure 2.16: Electronic energy band structure of the III-V compound GaAs.

diagram is also shown. Most of the features in the density of states can be identified with the band model.

2.3.4 III–V Compound Semiconductors

Another important class of semiconductors is the III–V compound semiconductors which crystallize in the zincblende structure; this structure is like the diamond structure except that the two atoms/unit cell are of a different chemical species. The III–V compounds also have many practical applications, such as semiconductor lasers for fast electronics and communications, GaAs in light emitting diodes, and InSb for infrared detectors. In Fig. 2.16 the $E(\vec{k})$ diagram for GaAs is shown and we see that the electronic levels are very similar to those of Si and Ge. One exception is that the lowest conduction band for GaAs is at $\vec{k} = 0$ so that both valence and conduction band extrema are at $\vec{k}$=0. Thus GaAs is a *direct* gap semiconductor [see Fig. 2.10(a)], and for this reason, GaAs shows a stronger and more sharply defined optical absorption threshold than Si or Ge. Figure 2.11(b) shows a schematic of the conduction bands for GaAs. Here we see that the lowest conduction band for GaAs has high curvature and therefore a small effective mass. This mass is isotropic so that the constant energy surface for electrons in GaAs is a sphere and there is just one such sphere in the Brillouin zone. The next lowest conduction band is at a Δ point and a significant carrier density can be excited into this Δ point pocket at high temperatures.

The constant energy surface for electrons in the direct gap semiconductor InSb shown in Fig. 2.17 is likewise a sphere, because InSb is also a direct gap semiconductor. InSb differs from GaAs in having a very small band gap ($\sim$ 0.2 eV), occurring in the infrared. Both direct and indirect band gap materials are found in the III–V compound semiconductor family. Except for optical phenomena close to the band gap, these compound semiconductors all exhibit very similar optical properties which are associated with the band-tracking phenomena discussed in §2.3.2.

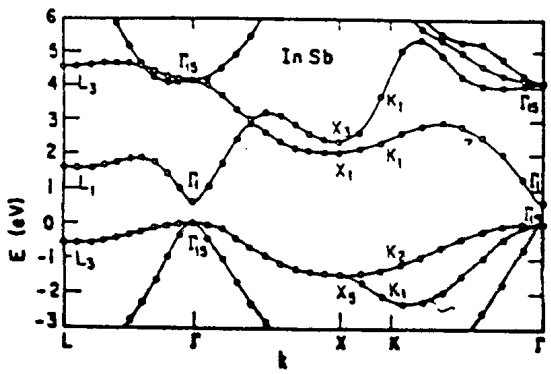


Figure 2.17: Electronic energy band structure of the III-V compound InSb.

2.3.5 "Zero Gap" Semiconductors – Gray Tin

It is also possible to have "zero gap" semiconductors. An example of such a material is gray tin which also crystallizes in the diamond structure. The energy band model for gray tin without spin–orbit interaction is shown in Fig. 2.18(a). On this diagram the zero gap occurs between the $\Gamma_{25'}$ valence band and the $\Gamma_{2'}$ conduction band, and the Fermi level runs right through this degeneracy point between these bands. Spin–orbit interaction (to be discussed later in this course) is very important for gray tin in the region of the $\vec{k} = 0$ band degeneracy, and a detailed diagram of the energy bands near $\vec{k} = 0$ and including the effect of spin–orbit interaction is shown in Fig. 2.18(b). In gray tin the effective mass for the conduction band is much lighter than for the valence band, as can be seen by the band curvatures shown in Fig. 2.18(b).

Optical transitions in Fig. 2.18(b) labeled B occur in the far infrared spectral region from the upper valence band to the conduction band. In the near infrared, interband transitions labeled A are induced from the Γ_7^- valence band to the Γ_8^+ conduction band. We note that gray tin is classified as a zero gap semiconductor rather than a semimetal (see §2.4) because there are no band overlaps in a zero-gap semiconductor anywhere in the Brillouin zone. Because of the zero band gap in grey tin, impurities play a major role in determining the position of the Fermi level. Gray tin is normally prepared n-type which means that there are some electrons present in the conduction band (for example, a typical electron concentration would be $10^{15}/\text{cm}^3$ which amounts to less than 1 carrier/10^7 atoms).

2.3.6 Molecular Semiconductors – Fullerenes

Other examples of semiconductors are molecular solids such as C_{60} (see Fig. 2.19). For the case of solid C_{60}, we show in Fig. 2.19(a) a C_{60} molecule, which crystallizes in a FCC structure with four C_{60} molecules per conventional simple cubic unit cell. A small distortion of the bonds, lengthening the C–C bond lengths of the single bonds to 1.46Å and shortening the double bonds to 1.40Å, stabilizes a band gap of ∼1.5eV [see Fig. 2.19(b)]. In this semi-

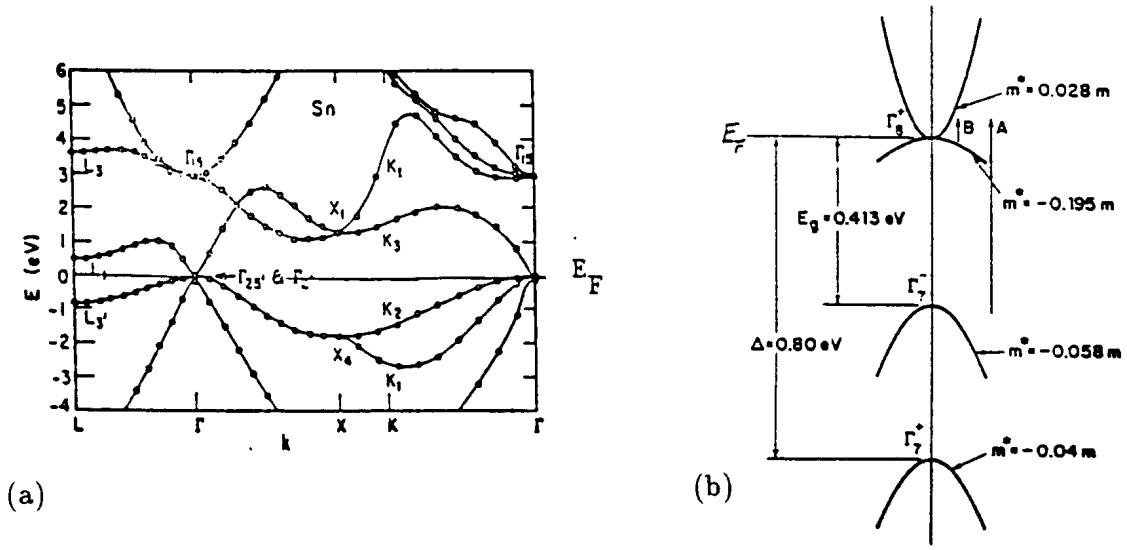


Figure 2.18: (a) Electronic energy band structure of gray Sn, neglecting the spin-orbit interaction. (b) Detailed diagram of the energy bands of gray tin near $k = 0$, including the spin-orbit interaction. The Fermi level goes through the degenerate point between the filled valence band and the empty conduction band in the idealized model for gray tin at $T = 0$. The Γ_7^- hole band has the same symmetry as the conduction band for Ge when spin–orbit interaction is included, as shown in Fig. 2.12(b). The Γ_7^+ hole band has the same symmetry as the "split–off" valence band for Ge when spin–orbit interaction is included.

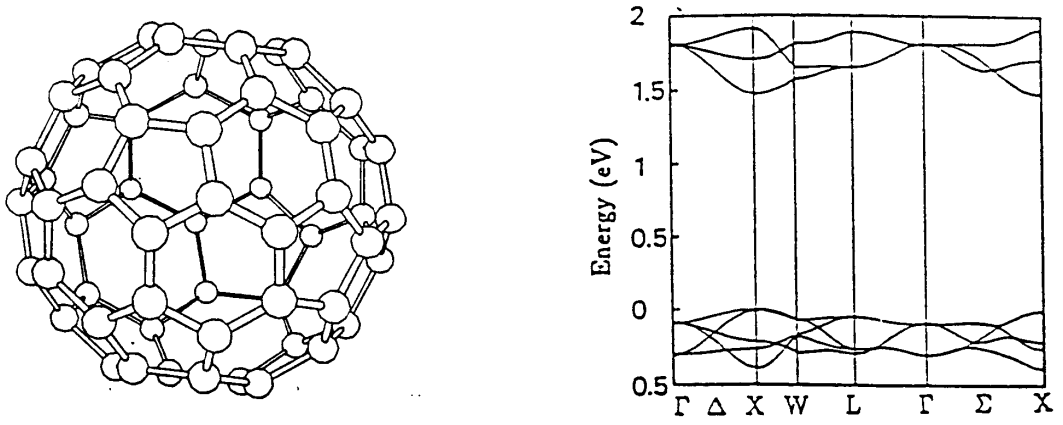


Figure 2.19: (a)Structure of the icosahedral C_{60} molecule, and (b) the calculated one-electron electronic energy band structure of FCC solid C_{60}. The Fermi energy lies between the occupied valence levels and the empty conduction levels.

conductor, the energy bandwidths are very small compared with the band gaps, so that this material can be considered as an organic molecular semiconductor. The transport properties of C_{60} differ markedly from those for conventional group IV or III–V semiconductors.

2.4 Semimetals

Another type of material that commonly occurs in nature is the *semimetal*. Semimetals have exactly the correct number of electrons to completely fill an integral number of Brillouin zones. Nevertheless, in a semimetal the highest occupied Brillouin zone is not filled up completely, since some of the electrons find lower energy states in "higher" zones (see Fig. 2.2). For semimetals the number of electrons that spill over into a higher Brillouin zone is exactly equal to the number of holes that are left behind. This is illustrated schematically in Fig. 2.20(a) where a two-dimensional Brillouin zone is shown and a circular Fermi surface of equal area is inscribed. Here we can easily see the electrons in the second zone at the zone edges and the holes at the zone corners that are left behind in the first zone. Translation by a reciprocal lattice vector brings two pieces of the electron surface together to form a surface in the shape of a lens, and the 4 pieces at the zone corners form a rosette shaped hole pocket. Typical examples of semimetals are bismuth and graphite. For these semimetals the carrier density is on the order of one carrier/10^6 atoms.

The carrier density of a semimetal is thus not very different from that which occurs in doped semiconductors, but the behavior of the conductivity $\sigma(T)$ as a function of temperature is very different. For intrinsic semiconductors, the carriers which are excited thermally contribute significantly to conduction. Consequently, the conductivity tends to rise rapidly with increasing temperature. For a semimetal, the carrier concentration does not change significantly with temperature because the carrier density is determined by the band overlap. Since the electron scattering by lattice vibrations increases with increasing temperature, the conductivity of semimetals tends to fall as the temperature increases.

A schematic diagram of the energy bands of the semimetal bismuth is shown in Fig. 2.20(b). Electron and hole carriers exist in equal numbers but at different locations in the Brillouin zone. For Bi, electrons are at the L–point, and holes at the T–point [see Fig. 2.20(b)]. The crystal structure for Bi can be understood from the NaCl structure by considering a very small displacement of the Na FCC structure relative to the Cl FCC structure along one of the body diagonals and an elongation of that body diagonal relative to the other 3 body diagonals. The special {111} direction corresponds to $\Gamma - T$ in the Brillouin zone, while the other three {111} directions are labeled as $\Gamma - L$.

Instead of a band gap between valence and conduction bands (as occurs for semiconductors), semimetals are characterized by a band overlap in the millivolt range. In bismuth, a small band gap also occurs at the L–point between the conduction band and a lower filled valence band. Because the coupling between these L-point valence and conduction bands is strong, some of the effective mass components for the electrons in bismuth are anomalously small. As far as the optical properties of bismuth are concerned, bismuth behaves much like a metal with a high reflectivity at low frequencies due to the presence of free carriers.

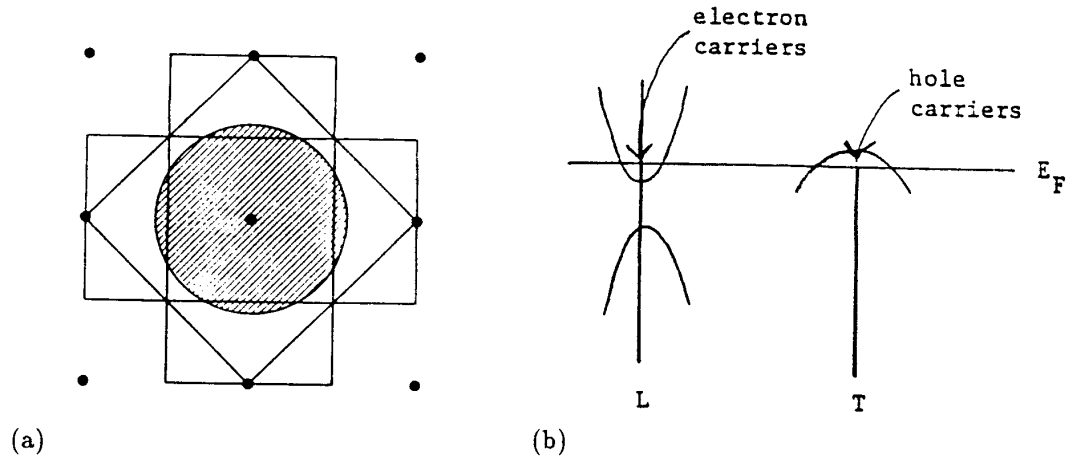


Figure 2.20: (a) Schematic diagram of a semimetal in two dimensions. (b) Schematic diagram of the energy bands $E(\vec{k})$ of bismuth showing electron pockets at the L point and a hole pocket at the T point. The T point is the point at the Brillouin zone boundary in the $\{111\}$ direction along which a stretching distortion occurs in real space, and the L points refer to the 3 other equivalent $\{1\bar{1}\bar{1}\}$, $\{\bar{1}1\bar{1}\}$, and $\{\bar{1}\bar{1}1\}$ directions.

2.5 Insulators

The electronic structure of insulators is similar to that of semiconductors, in that both insulators and semiconductors have a band gap separating the valence and conduction bands. However, in the case of insulators, the band gap is so large that thermal energies are not sufficient to excite a significant number of carriers.

The simplest insulator is a solid formed of rare gas atoms. An example of a rare gas insulator is solid argon which crystallizes in the FCC structure with one Ar atom/primitive unit cell. With an atomic configuration $3s^2 3p^6$, argon has filled $3s$ and $3p$ bands which are easily identified in the energy band diagram in Fig. 2.21. These occupied bands have very narrow band widths compared to their band gaps and are therefore well described by the tight binding approximation. This figure shows that the higher energy states forming the conduction bands (the hybridized $4s$ and $3d$ bands) show more dispersion than the more tightly bound valence bands. The band diagram shows argon to have a direct band gap at the Γ point of about 1 Rydberg or 13.6 eV. Although the $4s$ and $3d$ bands have similar energies, identification with the atomic levels can easily be made near $k = 0$ where the lower lying $4s$-band has considerably more band curvature than the $3d$ levels which are easily identified because of their degeneracies [the so called three-fold t_g ($\Gamma_{25'}$) and the two-fold e_g (Γ_{12}) crystal field levels for d-bands in a cubic crystal].

Another example of an insulator formed from a closed shell configuration is found in Fig. 2.22. Here the closed shell configuration results from charge transfer, as occurs in all ionic crystals. For example in the ionic crystal LiF (or in other alkali halide compounds), the valence band is identified with the filled anion orbitals (fluorine p–orbitals in this case) and at much higher energy the empty cation conduction band levels will lie (lithium s–

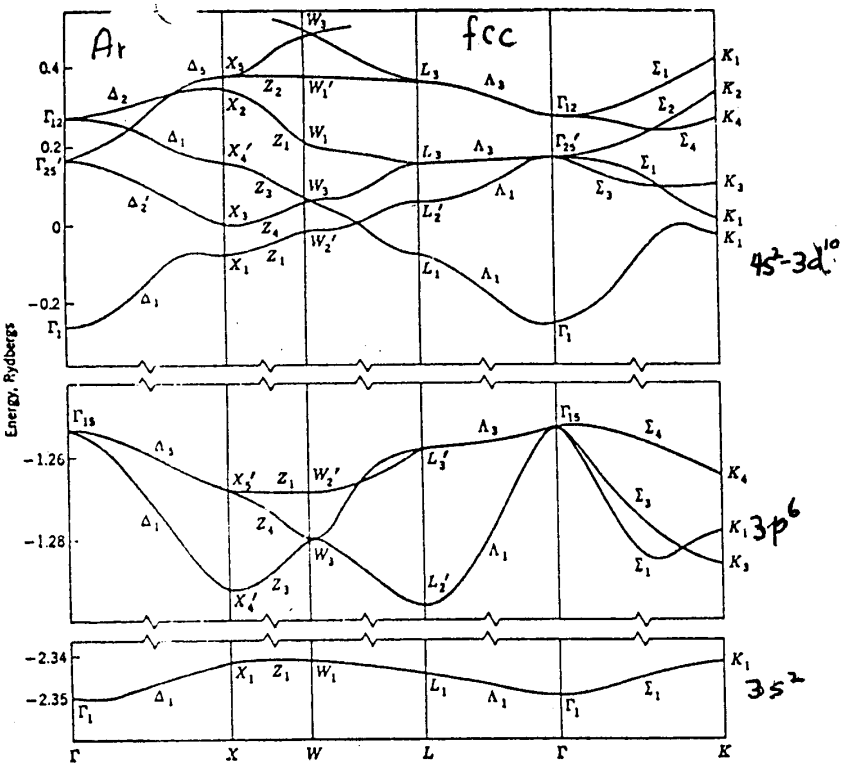


Figure 2.21: Electronic energy band structure of Argon.

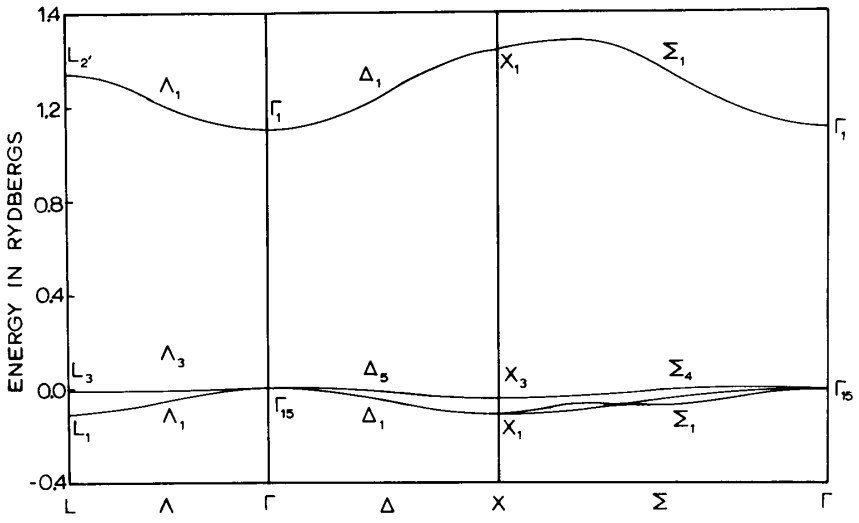


Figure 2.22: Band structure of the alkali halide insulator LiF. This ionic crystal is used extensively for UV optical components because of its large band gap.

orbitals in this case). Because of the wide band gap separation in the alkali halides between the valence and conduction bands, such materials are transparent at optical frequencies.

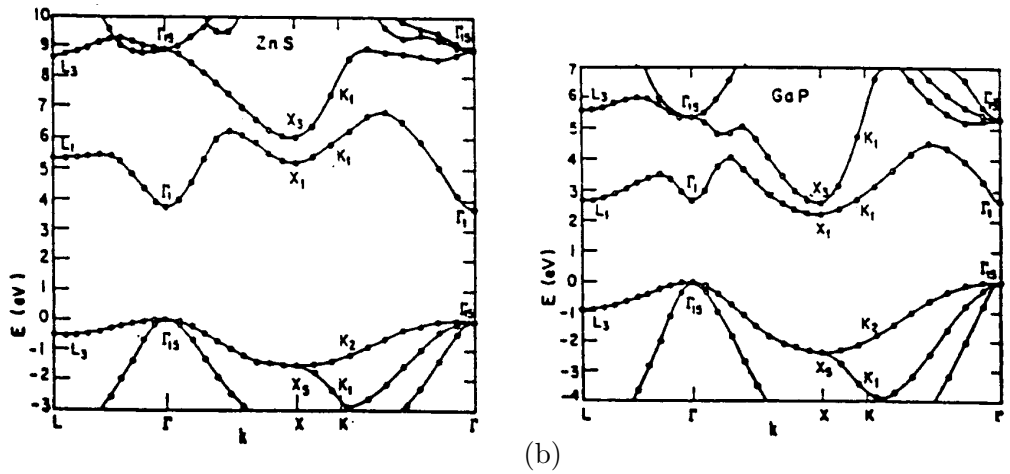
Insulating behavior can also occur for wide bandgap semiconductors with covalent bonding, such as diamond, ZnS and GaP (see Fig. 2.23). The $E(\vec{k})$ diagrams for these materials are very similar to the dispersion relations for typical III–V semiconducting compounds and for the group IV semiconductors silicon and germanium; the main difference, however, is the large band gap separating valence and conduction bands.

Even in insulators there is often a measurable electrical conductivity. For these materials the band electronic transport processes become less important relative to charge hopping from one atom to another by over-coming a potential barrier. Ionic conduction can also occur in insulating ionic crystals. From a practical point of view, one of the most important applications of insulators is for the control of electrical breakdown phenomena.

The principal experimental methods for studying the electronic energy bands depend on the nature of the solid. For insulators, the optical properties are the most important, while for semiconductors both optical and transport studies are important. For metals, optical properties are less important and Fermi surface studies become more important.

In the case of insulators, electrical conductivity can arise through the motion of lattice ions as they move from one lattice vacancy to another, or from one interstitial site to another. Ionic conduction therefore occurs through the presence of lattice defects, and is promoted in materials with open crystal structures. In ionic crystals there are relatively few mobile electrons or holes even at high temperature, so that conduction in these materials is predominantly due to the motions of ions.

Ionic conductivity (σ_{ionic}) is proportional both to the density of lattice defects (vacancies



(a) (b)

Figure 2.23: Electronic energy band structure of (a) cubic ZnS, a direct gap semi–insulating II–VI semiconductor, and (b) cubic GaP, an indirect gap semi–insulating III–V semiconductor. These wide bandgap semiconductors are of current interest for their optical properties.

and interstitials) and to the diffusion rate, so that we can write

$$\sigma_{\text{ionic}} \sim e^{-(E+E_0)/k_B T} \tag{2.7}$$

where E_0 is the activation energy for ionic motion and E is the energy for formation of a defect (a vacancy, a vacancy pair, or an interstitial). Being an activated process, ionic conduction is enhanced at elevated temperatures. Since defects in ionic crystals can be observed visibly as the migration of color through the crystal, ionic conductivity can be distinguished from electronic conductivity by comparing the transport of charge with the transport of mass, as can, for example, be measured by the material plated out on electrodes in contact with the ionic crystal.

In this course, we will spend a good deal of time studying optical and transport properties of solids. In connection with topics on magnetism, we will also study Fermi surface measurements which are closely connected to issues relevant to transport properties. Since Fermi surface studies, for the most part, are resonance experiments and involve the use of a magnetic field, it is pedagogically more convenient to discuss these topics in Part III of the course, devoted to Magnetism. We have presented this review of the electronic energy bands of solids because the $E(\vec{k})$ relations are closely connected with a large number of common measurements in the laboratory, and because a knowledge of the $E(\vec{k})$ relations forms the basis for many device applications.

Chapter 3

Effective Mass Theory

Reference

- Smith, Janak and Adler, *Electron Conduction in Solids*, McGraw-Hill, 1967, Chapter 6.

3.1 Wavepackets in Crystals and Group Velocity of Electrons in Solids

In a crystal lattice, the electronic motion which is induced by an applied field is conveniently described by a wavepacket composed of eigenstates of the unperturbed crystal. These eigenstates are Bloch functions

$$\psi_{nk}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}} u_{nk}(\vec{r}) \tag{3.1}$$

and are associated with band n. These wavepackets are solutions of the time-dependent Schrödinger equation

$$\mathcal{H}_0 \psi_n(\vec{r}, t) = i\hbar \frac{\partial \psi_n(\vec{r}, t)}{\partial t} \tag{3.2}$$

where the time independent part of the Hamiltonian can be written as

$$\mathcal{H}_0 = \frac{p^2}{2m} + V(\vec{r}), \tag{3.3}$$

where $V(\vec{r}) = V(\vec{r} + \vec{R}_n)$ is the periodic potential. The wave packets $\psi_n(\vec{r}, t)$ can be written in terms of the Bloch states $\psi_{nk}(\vec{r})$ as

$$\psi_n(\vec{r}, t) = \sum_k A_{n,k}(t) \psi_{nk}(\vec{r}) = \int d^3k A_{n,k}(t) \psi_{nk}(\vec{r}) \tag{3.4}$$

where we have replaced the sum by an integration over the Brillouin zone, since permissible $\vec{k}$ values for a macroscopic solid are *very* closely spaced. If the Hamiltonian $\mathcal{H}_0$ is time-independent as is often the case, we can write

$$A_{n,k}(t) = A_{n,k} e^{-i\omega_n(\vec{k})t} \tag{3.5}$$

where

$$\hbar\omega_n(\vec{k}) = E_n(\vec{k}) \tag{3.6}$$

and thereby obtain

$$\psi_n(\vec{r},t) = \int d^3k A_{n,k} u_{nk}(\vec{r}) e^{i[\vec{k}\cdot\vec{r}-\omega_n(\vec{k})t]}. \tag{3.7}$$

We can localize the wavepacket in $\vec{k}$-space by requiring that the coefficients $A_{n,k}$ be large only in a confined region of $\vec{k}$-space centered at $\vec{k} = \vec{k}_0$. If we now expand the band energy in a Taylor series around $\vec{k} = \vec{k}_0$ we obtain:

$$E_n(\vec{k}) = E_n(\vec{k}_0) + (\vec{k} - \vec{k}_0) \cdot \left.\frac{\partial E_n(\vec{k})}{\partial \vec{k}}\right|_{\vec{k}=\vec{k}_0} + ..., \tag{3.8}$$

where we have written $\vec{k}$ as

$$\vec{k} = \vec{k}_0 + (\vec{k} - \vec{k}_0). \tag{3.9}$$

Since $|\vec{k} - \vec{k}_0|$ is assumed to be small compared with Brillouin zone dimensions, we are justified in retaining only the first two terms of the Taylor expansion in Eq. 3.8 given above. Substitution into Eq. 3.4 for the wave packet yields:

$$\psi_n(\vec{r},t) \simeq e^{i(\vec{k}_0\cdot\vec{r}-\omega_n(\vec{k}_0)t)} \int d^3k A_{n,k}\ u_{nk}(\vec{r}) e^{i(\vec{k}-\vec{k}_0)\cdot[\vec{r}-\frac{\partial\omega_n(\vec{k})}{\partial\vec{k}}t]} \tag{3.10}$$

where

$$\hbar\omega_n(\vec{k}_0) = E_n(\vec{k}_0) \tag{3.11}$$

and

$$\hbar\frac{\partial\omega_n(\vec{k})}{\partial\vec{k}} = \frac{\partial E_n(\vec{k})}{\partial\vec{k}} \tag{3.12}$$

and the derivative $\partial\omega_n(\vec{k})/\partial\vec{k}$ which appears in the phase factor of Eq. 3.10 is evaluated at $\vec{k} = \vec{k}_0$. Except for the periodic function $u_{nk}(\vec{r})$, Eq. 3.10 is in the standard form for a wavepacket moving with *"group velocity"* $\vec{v}_g$

$$\vec{v}_g \equiv \frac{\partial\omega_n(\vec{k})}{\partial\vec{k}} \tag{3.13}$$

so that

$$\vec{v}_g = \frac{1}{\hbar}\frac{\partial E_n(\vec{k})}{\partial\vec{k}}, \tag{3.14}$$

while the phase velocity

$$\vec{v}_p = \frac{\omega_n(\vec{k})}{\vec{k}} = \frac{\partial E_n(\vec{k})}{\hbar\partial\vec{k}}. \tag{3.15}$$

In the limit of free electrons the group velocity becomes

$$\vec{v}_g = \frac{\vec{p}}{m} = \frac{\hbar\vec{k}}{m} \tag{3.16}$$

and $\vec{v}_g = \vec{v}_p$ in this limit. This result also follows from the above discussion using

$$E_n(\vec{k}) = \frac{\hbar^2k^2}{2m} \tag{3.17}$$

$$\frac{\partial E_n(\vec{k})}{\hbar \partial \vec{k}} = \frac{\hbar \vec{k}}{m}. \tag{3.18}$$

We shall show later that the electron wavepacket moves through the crystal very much like a free electron provided that the wavepacket remains localized in k space during the time interval of interest in the particular problem under consideration. Because of the uncertainty principle, the localization of a wavepacket in reciprocal space implies a delocalization of the wavepacket in real space.

We use wavepackets to describe electronic states in a solid when the crystal is perturbed in some way (e.g., by an applied electric or magnetic field). We make frequent applications of wavepackets to transport theory (e.g., electrical conductivity). In many practical applications of transport theory, use is made of the Effective-Mass Theorem, which is the most important result of transport theory.

We note that the above discussion for the wavepacket is given in terms of the **perfect crystal**. In our discussion of the Effective-Mass Theorem we will see that these wavepackets are also of use in describing situations where the Hamiltonian which enters Schrödinger's equation contains both the unperturbed Hamiltonian of the perfect crystal $\mathcal{H}_0$ and the perturbation Hamiltonian $\mathcal{H}'$ arising from an external perturbation. Common perturbations are applied electric or magnetic fields, or a lattice defect or an impurity atom.

3.2 The Effective Mass Theorem

We shall now present the Effective Mass theorem, which is central to the consideration of the electrical and optical properties of solids. An elementary proof of the theorem will be given here for a simple but important case, namely the non–degenerate band which can be identified with the corresponding atomic state. The theorem will be discussed from a more advanced point of view which considers also the case of degenerate bands in the following courses in the physics of solids sequence.

For many practical situations we find a solid in the presence of some perturbing field (e.g., an externally applied electric field, or the perturbation created by an impurity atom or a crystal defect). The perturbation may be either time–dependent or time–independent. We will show here that under many common circumstances this perturbation can be treated in the effective mass approximation whereby the periodic potential is replaced by an effective Hamiltonian based on the $E(\vec{k})$ relations for the perfect crystal.

To derive the effective mass theorem, we start with the time–dependent Schrödinger equation

$$(\mathcal{H}_0 + \mathcal{H}')\psi_n(\vec{r}, t) = i\hbar \frac{\partial \psi_n(\vec{r}, t)}{\partial t}. \tag{3.19}$$

We then substitute the expansion for the wave packet

$$\psi_n(\vec{r}, t) = \int d^3k A_{nk}(t) e^{i\vec{k}\cdot\vec{r}} u_{nk}(\vec{r}) \tag{3.20}$$

into Schrödinger's equation and make use of the Bloch solution

$$\mathcal{H}_0 e^{i\vec{k}\cdot\vec{r}} u_{nk}(\vec{r}) = E_n(\vec{k}) e^{i\vec{k}\cdot\vec{r}} u_{nk}(\vec{r}) \tag{3.21}$$

to obtain:

$$(\mathcal{H}_0+\mathcal{H}')\psi_n(\vec{r},t)=\int d^3k[E_n(\vec{k})+\mathcal{H}']A_{nk}(t)e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r}) = i\hbar(\partial\psi_n(\vec{r},t)/\partial t)$$
$$=i\hbar\int d^3k\dot{A}_{nk}(t)e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r}). \tag{3.22}$$

It follows from Bloch's theorem that $E_n(\vec{k})$ is a periodic function in the reciprocal lattice. We can therefore expand $E_n(\vec{k})$ in a Fourier series in the direct lattice

$$E_n(\vec{k}) = \sum_{\vec{R}_\ell} E_{n\ell}\, e^{i\vec{k}\cdot\vec{R}_\ell} \tag{3.23}$$

where the $\vec{R}_\ell$ are lattice vectors. Now consider the differential operator $E_n(-i\vec{\nabla})$ formed by replacing $\vec{k}$ by $-i\vec{\nabla}$

$$E_n(-i\vec{\nabla}) = \sum_{\vec{R}_\ell} E_{n\ell}\, e^{\vec{R}_\ell\cdot\vec{\nabla}}. \tag{3.24}$$

Consider the effect of $E_n(-i\vec{\nabla})$ on an arbitrary function $f(\vec{r})$. Since $e^{\vec{R}_\ell\cdot\vec{\nabla}}$ can be expanded in a Taylor series, we obtain

$$e^{\vec{R}_\ell\cdot\vec{\nabla}}f(\vec{r})=[1+\vec{R}_\ell\cdot\vec{\nabla}+\tfrac{1}{2}(\vec{R}_\ell\cdot\vec{\nabla})(\vec{R}_\ell\cdot\vec{\nabla})+...]f(\vec{r})$$
$$=f(\vec{r})+\vec{R}_\ell\cdot\vec{\nabla}f(\vec{r})+\tfrac{1}{2!}R_{\ell,\alpha}R_{\ell,\beta}\frac{\partial^2}{\partial r_\alpha\partial r_\beta}f(\vec{r})+... \tag{3.25}$$
$$=f(\vec{r}+\vec{R}_\ell).$$

Thus the effect of $E_n(-i\vec{\nabla})$ on a Bloch state is to produce $E_n(k)$ because

$$E_n(-i\vec{\nabla})\psi_{nk}(\vec{r}) = \sum_{\vec{R}_\ell} E_{n\ell}\psi_{nk}(\vec{r}+\vec{R}_\ell) = \sum_{\vec{R}_\ell} E_{n\ell}e^{i\vec{k}\cdot\vec{R}_\ell}e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r}) = E_n(\vec{k})\psi_{nk}(\vec{r}), \tag{3.26}$$

since from Bloch's theorem

$$\psi_{nk}(\vec{r}+\vec{R}_\ell) = e^{i\vec{k}\cdot\vec{R}_\ell}\Big[e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r})\Big]. \tag{3.27}$$

Substitution of

$$E_n(-i\vec{\nabla})\psi_{nk}(\vec{r}) = E_n(\vec{k})\psi_{nk}(\vec{r}) \tag{3.28}$$

from Eq. 3.26 into Schrödinger's equation (Eq. 3.22) yields:

$$\int d^3k\,\Big[E_n(-i\vec{\nabla})+\mathcal{H}'\Big]A_{nk}(t)e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r}) =\Big[E_n(-i\vec{\nabla})+\mathcal{H}'\Big]\int d^3k\; A_{nk}(t)e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r}) \tag{3.29}$$

so that

$$\Big[E_n(-i\vec{\nabla})+\mathcal{H}'\Big]\psi_n(\vec{r},t) = i\hbar\frac{\partial\psi_n(\vec{r},t)}{\partial t}. \tag{3.30}$$

The result of Eg. 3.30 is called the *effective mass theorem*. We observe that the original crystal Hamiltonian $p^2/2m+V(\vec{r})$ does not appear in this equation. It has instead been

replaced by an effective Hamiltonian which is an operator formed from the solution $E(\vec{k})$ for the perfect crystal in which we replace $\vec{k}$ by $-i\vec{\nabla}$. For example, for the free electron ($V(\vec{r}) \equiv 0$)

$$E_n(-i\vec{\nabla}) \rightarrow -\frac{\hbar^2\nabla^2}{2m}. \tag{3.31}$$

In applying the effective mass theorem, we assume that $E(\vec{k})$ is known either from the results of a theoretical calculation or from the analysis of experimental results. What is important here is that once $E(\vec{k})$ is known, the effect of various perturbations on the ideal crystal can be treated in terms of the solution to the energy levels of the perfect crystal, without recourse to consideration of the full Hamiltonian. In practical cases, the solution to the effective mass equation is much easier to carry out than the solution to the original Schrödinger equation.

According to the above discussion, we have assumed that $E(\vec{k})$ is specified throughout the Brillouin zone. For many practical applications, the region of $\vec{k}$-space which is of importance is confined to a small portion of the Brillouin zone. In such cases it is only necessary to specify $E(\vec{k})$ in a local region (or regions) and to localize our wavepacket solutions to these local regions of $\vec{k}$-space. Suppose that we localize the wavepacket around $\vec{k} = \vec{k}_0$, and correspondingly expand our Bloch functions around $\vec{k}_0$,

$$\psi_{nk}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}u_{nk}(\vec{r}) \simeq e^{i\vec{k}\cdot\vec{r}}u_{nk_0}(\vec{r}) = e^{i(\vec{k}-\vec{k}_0)\cdot\vec{r}}\psi_{nk_0}(\vec{r}) \tag{3.32}$$

where we have noted that $u_{nk}(\vec{r}) \simeq u_{nk_0}(\vec{r})$ has only a weak dependence on $\vec{k}$. Then our wavepacket can be written as

$$\psi_n(\vec{r},t) = \int d^3k\ A_{nk}(t)e^{i(\vec{k}-\vec{k}_0)\cdot\vec{r}}\psi_{nk_0}(\vec{r}) = F(\vec{r},t)\psi_{nk_0}(\vec{r}) \tag{3.33}$$

where $F(\vec{r},t)$ is called the *amplitude* or *envelope* function and is defined by

$$F(\vec{r},t) = \int d^3k\ A_{nk}(t)e^{i(\vec{k}-\vec{k}_0)\cdot\vec{r}}. \tag{3.34}$$

Since the time dependent Fourier coefficients $A_{nk}(t)$ are assumed here to be large only near $\vec{k} = \vec{k}_0$, then $F(\vec{r},t)$ will be a slowly varying function of $\vec{r}$, because in this case

$$e^{i(\vec{k}-\vec{k}_0)\vec{r}} \simeq 1 + i(\vec{k}-\vec{k}_0)\cdot\vec{r} + ... \tag{3.35}$$

It can be shown that the envelope function also satisfies the effective mass equation

$$\left[E_n(-i\vec{\nabla}) + \mathcal{H}'\right]F(\vec{r},t) = i\hbar\frac{\partial F(\vec{r},t)}{\partial t} \tag{3.36}$$

where we now replace $\vec{k} - \vec{k}_0$ in $E_n(\vec{k})$ by $-i\vec{\nabla}$. This form of the effective mass equation is useful for treating the problem of donor and acceptor impurity states in semiconductors, and $\vec{k}_0$ is taken as the band extremum.

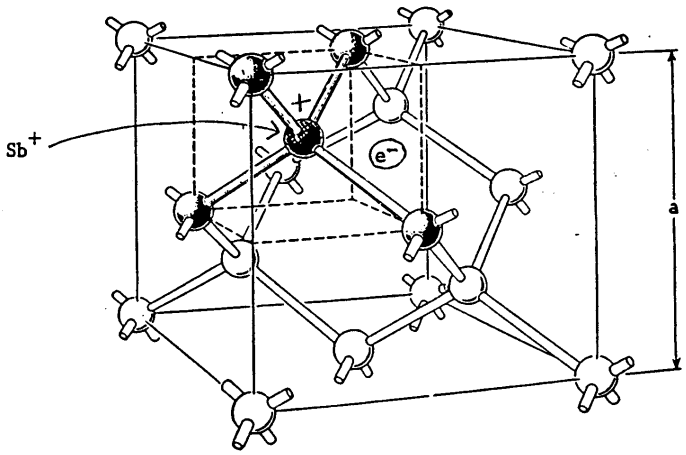


Figure 3.1: Crystal structure of diamond, showing the tetrahedral bond arrangement with an Sb^+ ion on one of the lattice sites and a free donor electron available for conduction.

3.3 Application of the Effective Mass Theorem to Donor Impurity Levels in a Semiconductor

Suppose that we add an impurity from column V in the Periodic Table to a semiconductor such as silicon or germanium, which are both members of column IV of the periodic table. This impurity atom will have one more electron than is needed to satisfy the valency requirements for the tetrahedral bonds which the germanium or silicon atoms form with their 4 valence electrons (see Fig. 3.1).

This extra electron from the impurity atom will be free to wander through the lattice, subject of course to the coulomb attraction of the ion core which will have one unit of positive charge. We will consider here the case where we add just a small number of these impurity atoms so that we may focus our attention on a single, isolated substitutional impurity atom in an otherwise perfect lattice. In the course of this discussion we will define more carefully what the limits on the impurity concentration must be so that the treatment given here is applicable.

Let us also assume that the conduction band of the host semiconductor in the vicinity of the band "minimum" at $\vec{k}_0$ has the simple analytic form

$$E_c(\vec{k}) \simeq E_c(\vec{k}_0) + \frac{\hbar^2(k-k_0)^2}{2m^*}. \tag{3.37}$$

We can consider this expression for the conduction band level $E_c(\vec{k})$ as a special case of the Taylor expansion of $E(\vec{k})$ about an energy band minimum at $\vec{k} = \vec{k}_0$. For the present discussion, $E(\vec{k})$ is assumed to be isotropic in $\vec{k}$; this typically occurs in cubic semiconductors with band extrema at $\vec{k} = 0$. The quantity m^* in this equation is the *effective mass* for the electrons. We will see that the energy levels corresponding to the donor electron will lie in the band gap below the conduction band minimum as indicated in the diagram in Fig. 3.2. To solve for the impurity levels explicitly, we may use the time-independent form of the

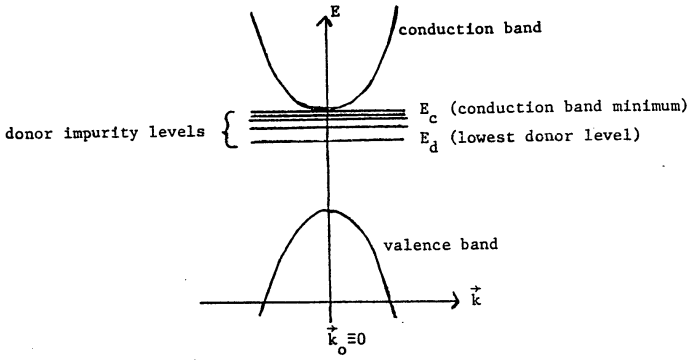


Figure 3.2: Schematic band diagram showing donor levels in a semiconductor.

effective mass theorem derived from Eq. 3.36

$$\left[E_n(-i\vec{\nabla}) + \mathcal{H}'\right] F(\vec{r}) = (E - E_c) F(\vec{r}). \tag{3.38}$$

Equation 3.38 is applicable to the impurity problem in a semiconductor provided that the amplitude function $F(\vec{r})$ is sufficiently slowly varying over a unit cell. In the course of this discussion, we will see that the donor electron in a column IV (or III-V or II-VI compound semiconductor) will wander over many lattice sites and therefore this approximation on $F(\vec{r})$ will be justified.

For a singly ionized donor impurity (such as arsenic in germanium), the perturbing potential $\mathcal{H}'$ can be represented as a Coulomb potential

$$\mathcal{H}' = -\frac{e^2}{\varepsilon r} \tag{3.39}$$

where ε is an average dielectric constant of the crystal medium which the donor electron sees as it wanders through the crystal. Experimental data on donor impurity states indicate that ε is very closely equal to the low frequency limit of the electronic dielectric constant $\varepsilon_1(\omega)|_{\omega=0}$, which we will discuss extensively in treating the optical properties of solids (Part II of this course). The above discussion involving an isotropic $E(\vec{k})$ is appropriate for semiconductors with conduction band minima at $\vec{k}_0 = 0$. The Effective Mass equation for the unperturbed crystal is

$$E_n(-i\vec{\nabla}) = -\frac{\hbar^2}{2m^*}\nabla^2 \tag{3.40}$$

in which we have replaced $\vec{k}$ by $-i\vec{\nabla}$.

The donor impurity problem in the effective mass approximation thus becomes

$$\left[-\frac{\hbar^2}{2m^*}\nabla^2 - \frac{e^2}{\varepsilon r}\right] F(\vec{r}) = (E - E_c) F(\vec{r}) \tag{3.41}$$

where all energies are measured with respect to the bottom of the conduction band E_c. If we replace m^* by m and e^2/ε by e^2, we immediately recognize this equation as Schrödinger's

equation for a hydrogen atom under the identification of the energy eigenvalues with

$$E_n = \frac{e^2}{2n^2 a_0} = \frac{me^4}{2n^2\hbar^2} \tag{3.42}$$

where a_0 is the Bohr radius $a_0 = \hbar^2/me^2$. This identification immediately allows us to write E_ℓ for the donor energy levels as

$$E_\ell = E_c - \frac{m^* e^4}{2\varepsilon^2 \ell^2 \hbar^2} \tag{3.43}$$

where $\ell = 1, 2, 3, \ldots$ is an integer denoting the donor level quantum numbers and we identify the bottom of the conduction band E_c as the ionization energy for this effective hydrogenic problem. Physically, this means that the donor levels correspond to bound (localized) states while the band states above E_c correspond to delocalized nearly–free electron–like states. The lowest or "ground-state" donor energy level is then written as

$$E_d = E_{\ell=1} = E_c - \frac{m^* e^4}{2\varepsilon^2\hbar^2}. \tag{3.44}$$

It is convenient to identify the "effective" first Bohr radius for the donor level as

$$a_0^* = \frac{\varepsilon\hbar^2}{m^* e^2} \tag{3.45}$$

and to recognize that the wave function for the ground state donor level will be of the form

$$F(\vec{r}) = \mathcal{C} e^{-r/a_0^*} \tag{3.46}$$

where $\mathcal{C}$ is the normalization constant. Thus the solutions to Eq. 3.41 for a semiconductor are hydrogenic energy levels with the substitutions $m \to m^*$, $e^2 \to (e^2/\varepsilon)$ and the ionization energy, usually taken as the zero of energy for the hydrogen atom, now becomes E_c, the conduction band extremum.

For a semiconductor like germanium we have a very large dielectric constant, $\varepsilon \simeq 16$. The value for the effective mass is somewhat more difficult to specify in germanium since the constant energy surfaces for germanium are located about the L–points in the Brillouin zone (see §2.3.2) and are ellipsoids of revolution. Since the constant energy surfaces for such semiconductors are non-spherical, the effective mass tensor is anisotropic. However we will write down an average effective mass value $m^*/m \simeq 0.12$ (Kittel ISSP) so that we can estimate pertinent magnitudes for the donor levels in a typical semiconductor. With these values for ε and m^* we obtain:

$$E_c - E_d \simeq 0.007\text{eV} \tag{3.47}$$

and the effective Bohr radius

$$a_0^* \simeq 70\text{Å}. \tag{3.48}$$

These values are to be compared with the ionization energy of 13.6 eV for the hydrogen atom and with the hydrogenic Bohr orbit of $a_0 = \hbar^2/me^2 = 0.5$Å.

Thus we see that a_0^* is indeed large enough to satisfy the requirement that $F(\vec{r})$ be slowly varying over a unit cell. On the other hand, if a_0^* were to be comparable to a lattice unit cell dimension, then $F(\vec{r})$ could not be considered as a slowly varying function of $\vec{r}$

and generalizations of the above treatment would have to be made. Such generalizations involve: (1) treating $E(\vec{k})$ for a wider region of $\vec{k}$-space, and (2) relaxing the condition that impurity levels are to be associated with a single band. From the uncertainty principle, the localization in momentum space for the impurity state requires a delocalization in real space; and likewise, the converse is true, that a localized impurity in real space corresponds to a delocalized description in $\vec{k}$-space. Thus "shallow" hydrogenic donor levels (close in energy to the band extremum) can be attributed to a specific band at a specific energy extremum at $\vec{k}_0$ in the Brillouin zone. On the other hand, "deep" donor levels (far in energy from the band extremum) are not hydrogenic and have a more complicated energy level structure. Deep donor levels cannot be readily associated with a specific band or a specific $\vec{k}$ point in the Brillouin zone.

In dealing with this impurity problem, it is very tempting to discuss the donor levels in silicon and germanium. For example in silicon where the conduction band extrema are at the Δ point (see §2.3.3), the effective mass theorem requires us to replace $E(-i\vec{\nabla})$ by

$$E_n(-i\vec{\nabla}) \rightarrow -\frac{\hbar^2}{2m_\ell^*}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m_t^*}\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \tag{3.49}$$

and the resulting Schrödinger equation can no longer be solved analytically. Although this is a very interesting problem from a practical point of view, numerical solutions are needed in this case.

3.4 Quasi-Classical Electron Dynamics

According to the "Correspondence Principle" of Quantum Mechanics, wavepacket solutions of Schrödinger's equation (see §3.1) follow the trajectories of classical particles and satisfy Newton's laws. One can give a Correspondence Principle argument for the form which is assumed by the velocity and acceleration of a wavepacket. According to the Correspondence Principle, the connection between the classical Hamiltonian and the quantum mechanical Hamiltonian is made by the identification of $\vec{p} \rightarrow (\hbar/i)\vec{\nabla}$. Thus

$$E_n(-i\vec{\nabla}) + \mathcal{H}'(\vec{r}) \leftrightarrow E_n(\vec{p}/\hbar) + \mathcal{H}'(\vec{r}) = \mathcal{H}_{\text{classical}}(\vec{p},\vec{r}). \tag{3.50}$$

In classical mechanics, Hamilton's equations give the velocity according to :

$$\dot{\vec{r}} = \frac{\partial \mathcal{H}}{\partial \vec{p}} = \nabla_p \mathcal{H} = \frac{\partial E(\vec{k})}{\hbar \partial \vec{k}} \tag{3.51}$$

in agreement with the group velocity for a wavepacket given by Eq. 3.13. Hamilton's equation for the acceleration is given by:

$$\dot{\vec{p}} = -\frac{\partial \mathcal{H}}{\partial \vec{r}} = -\frac{\partial \mathcal{H}'(\vec{r})}{\partial \vec{r}}. \tag{3.52}$$

For example, in the case of an applied electric field $\vec{E}$ the perturbation Hamiltonian is

$$\mathcal{H}'(\vec{r}) = -e\vec{r}\cdot\vec{E} \tag{3.53}$$

so that

$$\dot{\vec{p}} = \hbar\dot{\vec{k}} = e\vec{E}. \tag{3.54}$$

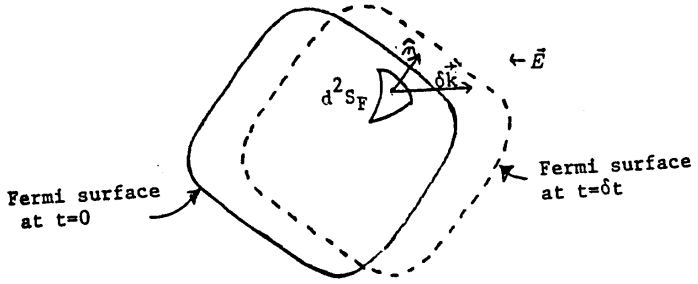


Figure 3.3: Displaced Fermi surface at $t = \delta t$ under the action of an electric field $\vec{E}$.

In this equation $e\vec{E}$ is the classical Coulomb force on an electric charge due to an applied field $\vec{E}$. It can be shown (to be derived rigorously in the advanced course) that in the presence of a magnetic field $\vec{B}$, the acceleration theorem follows the Lorentz force equation

$$\dot{\vec{p}} = \hbar\dot{\vec{k}} = e[\vec{E} + (1/c)\vec{v} \times \vec{B}] \tag{3.55}$$

where

$$\vec{v} = \frac{\partial E(\vec{k})}{\hbar \partial \vec{k}}. \tag{3.56}$$

In the crystal, the crystal momentum $\hbar\vec{k}$ for the wavepacket plays the role of the momentum for a classical particle.

3.5 Quasi-Classical Theory of Electrical Conductivity – Ohm's Law

We will now apply the idea of the quasi-classical electron dynamics in a solid to the problem of the electrical conductivity for a metal with an arbitrary Fermi surface and band structure. The electron is treated here as a wavepacket with momentum $\hbar\vec{k}$ moving in an external electric field $\vec{E}$ in compliance with Newton's laws. Because of the acceleration theorem, we can think of the electric field as creating a "displacement" of the electron distribution in $\vec{k}$-space. We remember that the Fermi surface encloses the region of occupied states within the Brillouin zone. The effect of the electric field is to change the wave vector $\vec{k}$ of an electron by

$$\delta\vec{k} = \frac{e}{\hbar}\vec{E}\delta t \tag{3.57}$$

(where we note that the charge on the electron e is a negative number). We picture the displacement $\delta\vec{k}$ of Eq. 3.57 by the displacement of the Fermi surface in time shown in Fig. 3.3. From this diagram we see that the incremental volume of $\vec{k}$-space $\delta^3 V_{\vec{k}}$ which is "swept out" in the time δt due to the presence of the field $\vec{E}$ is

$$\delta^3 V_{\vec{k}} = d^2 S_F \; \hat{n} \cdot \delta\vec{k} = d^2 S_F \; \hat{n} \cdot (\frac{e}{\hbar}\vec{E}\delta t) \tag{3.58}$$

and the electron density is found from

$$n = \frac{2}{(2\pi)^3} \int_{E \le E_F} d^3k \tag{3.59}$$

where d^2S_F is the element of area on the Fermi surface and $\hat{n}$ is a unit vector normal to this element of area and $\delta^3 V_{\vec{k}} \rightarrow d^3k$ both denote elements of volume in $\vec{k}$–space. The definition of the electrical current density is the current flowing through a unit area in real space and is given by the product of the [number of electrons per unit volume] with the [charge per electron] and with the [group velocity] so that the current density $\delta\vec{j}$ created by applying the electric field $\vec{E}$ for a time interval δt is given by

$$\delta\vec{j} = \int [2/(2\pi)^3] \ \cdot \ [\delta^3 V_{\vec{k}}] \cdot [e] \cdot [\vec{v}_g] \tag{3.60}$$

where $\vec{v}_g$ is the group velocity for electron wavepacket and $2/(2\pi)^3$ is the density of electronic states in $\vec{k}$-space (including the spin degeneracy of two) because we can put 2 electrons in each phase space state. Substitution for $\delta^3 V_{\vec{k}}$ in Eq. 3.60 by Eq. 3.58 yields the instantaneous rate of change of the current density averaged over the Fermi surface

$$\frac{\partial\vec{j}}{\partial t} = \frac{e^2}{4\pi^3\hbar} \oint\oint \vec{v}_g \ \hat{n} \cdot \vec{E} \ d^2S_F = \frac{e^2}{4\pi^3\hbar} \oint\oint \vec{v}_g \left(\frac{\vec{v}_g \cdot \vec{E}}{|v_g|}\right) (d^2S_F) \tag{3.61}$$

since the group velocity given by Eq. 3.13 is directed normal to the Fermi surface. In a real solid, the electrons will not be accelerated indefinitely, but will eventually collide with an impurity, or a lattice defect or a lattice vibration (phonon).

These collisions will serve to maintain the displacement of the Fermi surface at some steady state value, depending on τ, the average time between collisions. We can introduce this relaxation time through the expression

$$n(t) = n(0)e^{-t/\tau} \tag{3.62}$$

where $n(t)$ is the number of electrons that have not made a collision at time t, assuming that the last collision had been made at time $t = 0$. The relaxation time is the average collision time

$$\langle t \rangle = \frac{1}{\tau} \int_0^\infty t e^{-t/\tau} dt = \tau. \tag{3.63}$$

If in Eq. 3.61, we set $\langle \delta t \rangle = \tau$ and write the average current density as $\vec{j} = \langle \delta\vec{j} \rangle$, then we obtain

$$\vec{j} = \frac{e^2\tau}{4\pi^3\hbar} \int \vec{v}_g \frac{\vec{v}_g \cdot \vec{E}}{|v_g|} \ (d^2S_F). \tag{3.64}$$

We define the conductivity tensor as $\vec{j} = \overleftrightarrow{\sigma} \cdot \vec{E}$, so that Eq. 3.64 provides an explicit expression for the tensor $\overleftrightarrow{\sigma}$.

$$\overleftrightarrow{\sigma} = \frac{e^2\tau}{4\pi^3\hbar} \int \frac{\vec{v}_g \vec{v}_g}{|v_g|} \ (d^2S_F). \tag{3.65}$$

In the free electron limit $\overleftrightarrow{\sigma}$ becomes a scalar (isotropic conduction) and is given by the Drude formula which we derive below from Eq. 3.65. Using the equations for the free

electron limit

$$E=\hbar^2k^2/2m$$

$$E_F=\hbar^2k_F^2/2m \tag{3.66}$$

$$\vec{v}_g=\hbar\vec{k}_F/m.$$

We then obtain

$$\vec{v}_g\vec{v}_g \rightarrow v_x^2 = v_y^2 = v_z^2 = v^2/3 \tag{3.67}$$

$$\int d^2S_F = 4\pi k_F^2, \tag{3.68}$$

so that the number of electrons/unit volume can be written as:

$$n = \frac{1}{4\pi^3}\frac{4\pi}{3}\,k_F^3. \tag{3.69}$$

Therefore

$$\vec{j} = \frac{e^2\tau}{4\pi^3\hbar}\left(\frac{\hbar k_F}{m}\right)\frac{1}{3}\,\vec{E}(4\pi k_F^2) = \frac{ne^2\tau}{m}\,\vec{E}. \tag{3.70}$$

Thus the free electron limit gives Ohm's law in the familiar form

$$\sigma = \frac{ne^2\tau}{m} = ne\mu, \tag{3.71}$$

showing that the electrical conductivity in the diffusion regime where scattering is important depends on both the carrier density n and the carrier mobility μ. For low dimensional systems that are important on the nano-scale, in nanoscience and nanotechnology, ballistic transport is dominant, and in this regime the carriers can go from the anode to the cathode without scattering. This regime will be considered in a later lecture.

A slightly modified form of Ohm's law is also applicable to conduction in a material for which the energy dispersion relations have a simple parabolic form and m has been replaced by the effective mass m^*, $E(\vec{k}) = \hbar^2k^2/2m^*$. In this case σ is given by

$$\sigma = ne^2\tau/m^* \tag{3.72}$$

where the effective mass is found from the band curvature $1/m^* = \partial^2E/\hbar^2\partial k^2$. The generalization of Ohm's law can also be made to deal with solids for which the effective mass tensor is *anisotropic* and this will be discussed later in this course.

Chapter 4

Transport Phenomena

References:

- Ziman, *Principles of the Theory of Solids*, Cambridge Univ. Press, 1972, Chapters 7 and 9.
- Ashcroft and Mermin, *Solid State Physics*, Holt, Rinehart and Winston, 1976, Chapters 13.
- Smith, Janak and Adler, *Electronic Conduction in Solids*, McGraw-Hill, 1967, Chapters 7, 8, and 9.

4.1 Introduction

In this section we study some of the transport properties for metals and semiconductors. An intrinsic semiconductor at $T = 0$ has no carriers and therefore there is no transport of carriers under the influence of external fields. However at finite temperatures there are thermally generated carriers. Impurities also can serve to generate carriers and transport properties. For insulators, there is very little charge transport and in this case, the defects and the ions themselves can participate in charge transport under the influence of external applied fields. Metals make use of the Fermi-Dirac distribution function but are otherwise similar to semiconductors, for which the Maxwell–Boltzmann distribution function is usually applicable.

At finite fields, the electrical conductivity will depend on the product of the carrier density and the carrier mobility. For a one carrier system, the Hall effect gives the carrier density and the magnetoresistance gives the mobility, the key parameters governing the transport properties of a semiconductor. From the standpoint of device applications, the carrier density and the carrier mobility are the parameters of greatest importance.

To the extent that electrons can be considered as particles, the electrical conductivity, the electronic contribution to the thermal conductivity and the magnetoresistance are all found by solving the Boltzmann equation. For the case of nano-scale systems, where the wave aspects of the electron must be considered (called mesoscopic physics), more sophisticated approaches to the transport properties must be considered. To review the standard

procedures for classical electrons, we briefly review the Boltzmann equation and its solution in the next section.

4.2 The Boltzmann Equation

The Boltzmann transport equation is a statement that in the steady state, there is no net change in the distribution function $f(\vec{r},\vec{k},t)$ which determines the probability of finding an electron at position $\vec{r}$, crystal momentum $\vec{k}$ and time t. Therefore we get a zero sum for the changes in $f(\vec{r},\vec{k},t)$ due to the 3 processes of diffusion, the effect of forces and fields, and collisions:

$$\left.\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right|_{\text{diffusion}} + \left.\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right|_{\text{fields}} + \left.\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right|_{\text{collisions}} = 0. \tag{4.1}$$

It is customary to substitute the following differential form for the diffusion process

$$\left.\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right|_{\text{diffusion}} = -\vec{v}(\vec{k}) \cdot \frac{\partial f(\vec{r},\vec{k},t)}{\partial \vec{r}} \tag{4.2}$$

which expresses the continuity equation in real space in the absence of forces, fields and collisions. For the forces and fields, we write correspondingly

$$\left.\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right|_{\text{fields}} = -\frac{\partial \vec{k}}{\partial t} \cdot \frac{\partial f(\vec{r},\vec{k},t)}{\partial \vec{k}} \tag{4.3}$$

and by combining Eqs. 4.1, 4.2, and 4.3, we obtain the Boltzmann equation:

$$\frac{\partial f(\vec{r},\vec{k},t)}{\partial t} + \vec{v}(\vec{k}) \cdot \frac{\partial f(\vec{r},\vec{k},t)}{\partial \vec{r}} + \frac{\partial \vec{k}}{\partial t} \cdot \frac{\partial f(\vec{r},\vec{k},t)}{\partial \vec{k}} = \left.\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right|_{\text{collisions}} \tag{4.4}$$

which includes derivatives for all the variables of the distribution function on the left hand side of the equation and the collision terms appear on the right hand side of Eq. 4.4. The first term in Eq. 4.4 gives the explicit time dependence of the distribution function and is needed for the solution of ac driving forces or for impulse perturbations. Boltzmann's equation is usually solved using two approximations:

1. The perturbation due to external fields and forces is assumed to be small so that the distribution function can be linearized and written as:

$$f(\vec{r},\vec{k}) = f_0(E) + f_1(\vec{r},\vec{k}) \tag{4.5}$$

   where $f_0(E)$ is the equilibrium distribution function (the Fermi function) which depends only on the energy E, while $f_1(\vec{r},\vec{k})$ is the perturbation term giving the departure from equilibrium.

2. The collision term in the Boltzmann equation is written in the **relaxation time approximation** so that the system returns to equilibrium uniformly:

$$\left.\frac{\partial f}{\partial t}\right|_{\text{collisions}} = -\frac{(f - f_0)}{\tau} = -\frac{f_1}{\tau} \tag{4.6}$$

where τ denotes the relaxation time and in general is a function of crystal momentum, i.e., $\tau = \tau(\vec{k})$. The physical interpretation of the relaxation time is the time associated with the rate of return to the equilibrium distribution when the external fields or thermal gradients are switched off. Solution to Eq. 4.6 when the fields are switched off at $t = 0$ leads to

$$\frac{\partial f}{\partial t} = -\frac{(f - f_0)}{\tau} \tag{4.7}$$

which has solutions

$$f(t) = f_0 + \Big[f(0) - f_0\Big] e^{-t/\tau} \tag{4.8}$$

where f_0 is the equilibrium distribution and $f(0)$ is the distribution function at time $t = 0$. The relaxation previously described by Eq. 4.8 follows a Poisson distribution, indicating that collisions relax the distribution function exponentially to f_0 with a time constant τ.

With these approximations, the Boltzmann equation is solved to find the distribution function which in turn determines the number density and current density. The current density $\vec{j}(\vec{r}, t)$ is given by

$$\vec{j}(\vec{r}, t) = \frac{e}{4\pi^3} \int \vec{v}(\vec{k}) f(\vec{r}, \vec{k}, t) d^3k \tag{4.9}$$

in which the crystal momentum $\hbar\vec{k}$ plays the role of the momentum $\vec{p}$ in specifying a volume in phase space. Every element of size h (Planck's constant) in phase space can accommodate one spin $\uparrow$ and one spin $\downarrow$ electron. The carrier density $n(\vec{r}, t)$ is thus simply given by integration of the distribution function over k-space

$$n(\vec{r}, t) = \frac{1}{4\pi^3} \int f(\vec{r}, \vec{k}, t) d^3k \tag{4.10}$$

where d^3k is an element of 3D wavevector space. The velocity of a carrier with crystal momentum $\hbar\vec{k}$ is related to the $E(\vec{k})$ dispersion expression by

$$\vec{v}(\vec{k}) = \frac{1}{\hbar} \frac{\partial E(\vec{k})}{\partial \vec{k}} \tag{4.11}$$

and $f_0(E)$ is the Fermi distribution function

$$f_0(E) = \frac{1}{1 + e^{(E - E_F)/k_B T}} \tag{4.12}$$

which defines the equilibrium state in which E_F is the Fermi energy and k_B is the Boltzmann constant.

4.3 Electrical Conductivity

To calculate the static electrical conductivity, we consider an applied electric field $\vec{E}$ which for convenience we will take to be along the x-direction. We will assume for the present that there is no magnetic field and that there are no thermal gradients present. The electrical

conductivity is expressed in terms of the conductivity tensor $\overleftrightarrow{\sigma}$ which is evaluated explicitly from the relation

$$\vec{j} = \overleftrightarrow{\sigma} \cdot \vec{E}, \tag{4.13}$$

from solution of Eq. 4.9, using $\vec{v}(\vec{k})$ from Eq. 4.11 and the distribution function $f(\vec{r}, \vec{k}, t)$ from solution of the Boltzmann equation represented by Eq. 4.4. The first term in Eq. 4.4 vanishes since the dc applied field $\vec{E}$ has no time dependence.

For the second term in the Boltzmann equation Eq. 4.4, $\vec{v}(\vec{k}) \cdot \partial f(\vec{r}, \vec{k}, t)/\partial \vec{r}$, we note that

$$\frac{\partial f}{\partial \vec{r}} \simeq \frac{\partial f_0}{\partial \vec{r}} = \frac{\partial f_0}{\partial T} \frac{\partial T}{\partial \vec{r}}. \tag{4.14}$$

Since there are no thermal gradients present in the simplest calculation of the electrical conductivity given in this section, this term does not contribute to Eq. 4.4. For the third term in Eq. 4.4, which we write as

$$\dot{\vec{k}} \cdot \frac{\partial f(\vec{r}, \vec{k}, t)}{\partial \vec{k}} = \sum_{\alpha} \dot{k}_{\alpha} \frac{\partial f(\vec{r}, \vec{k}, t)}{\partial k_{\alpha}} \tag{4.15}$$

where the right hand side shows the summation over the vector components, we do get a contribution, since the equations of motion ($F = ma$) give

$$\hbar \dot{\vec{k}} = e\vec{E} \tag{4.16}$$

and

$$\frac{\partial f(\vec{r}, \vec{k}, t)}{\partial \vec{k}} = \frac{\partial (f_0 + f_1)}{\partial \vec{k}} = \frac{\partial f_0}{\partial E} \frac{\partial E}{\partial \vec{k}} + \frac{\partial f_1}{\partial \vec{k}}. \tag{4.17}$$

In considering the linearized Boltzmann equation, we retain only the leading terms in the perturbing electric field, so that $(\partial f_1/\partial \vec{k})$ can be neglected and only the term $(\partial f_0/\partial E)\hbar \vec{v}(\vec{k})$ need be retained. We thus obtain the linearized Boltzmann equation for the case of an applied static electric field and no thermal gradients:

$$\dot{\vec{k}} \cdot \frac{\partial f(\vec{r}, \vec{k}, t)}{\partial \vec{k}} = \frac{\phi}{\tau} \frac{\partial f_0}{\partial E} = -\frac{f_1}{\tau} \tag{4.18}$$

where it is convenient to write:

$$f_1 = -\phi \left(\frac{\partial f_0}{\partial E} \right) \tag{4.19}$$

in order to show the $(\partial f_0/\partial E)$ dependence explicitly. Substitution of Eqs. 4.16 and 4.17 into Eq. 4.18 yields

$$\left[\frac{e\vec{E}}{\hbar} \left(\frac{\partial f_0}{\partial E} \right) \right] \cdot [\hbar \vec{v}(\vec{k})] = \frac{\phi(\vec{k})}{\tau} \left(\frac{\partial f_0}{\partial E} \right) \tag{4.20}$$

so that

$$\phi(\vec{k}) = e\tau \vec{E} \cdot \vec{v}(\vec{k}). \tag{4.21}$$

Thus we can relate $\phi(\vec{k})$ to $f_1(\vec{k})$ by

$$f_1(\vec{k}) = -\phi(\vec{k}) \frac{\partial f_0(E)}{\partial E} = -e\tau \vec{E} \cdot \vec{v}(\vec{k}) \frac{\partial f_0(E)}{\partial E}. \tag{4.22}$$

The current density is then found from the distribution function $f(\vec{k})$ by calculation of the average value of $\langle ne\vec{v}\rangle$ over all k-space

$$\vec{j} = \frac{1}{4\pi^3}\int e\vec{v}(\vec{k})f(\vec{k})d^3k = \frac{1}{4\pi^3}\int e\vec{v}(\vec{k})f_1(\vec{k})d^3k \tag{4.23}$$

since

$$\int e\vec{v}(\vec{k})f_0(\vec{k})d^3k = 0. \tag{4.24}$$

Equation 4.24 states that no net current flows in the absence of an applied electric field, another statement of the equilibrium condition. Substitution for $f_1(\vec{k})$ given by Eq. 4.22 into Eq. 4.23 for $\vec{j}$ yields

$$\vec{j} = -\frac{e^2\vec{E}}{4\pi^3}\cdot\int \tau\vec{v}\vec{v}\frac{\partial f_0}{\partial E}d^3k \tag{4.25}$$

where in general $\tau = \tau(\vec{k})$ and $\vec{v}$ is given by Eq. 4.11. A comparison of Eqs. 4.25 and 4.13 thus yields the desired result for the conductivity tensor $\overleftrightarrow{\sigma}$

$$\overleftrightarrow{\sigma} = -\frac{e^2}{4\pi^3}\int \tau\vec{v}\vec{v}\,\frac{\partial f_0}{\partial E}d^3k \tag{4.26}$$

where $\overleftrightarrow{\sigma}$ is a symmetric second rank tensor ($\sigma_{ij} = \sigma_{ji}$). The evaluation of the integral in Eq. 4.26 over all k-space depends on the $E(\vec{k})$ relations through the $\vec{v}\vec{v}$ terms and the temperature dependence comes through the $\partial f_0/\partial E$ term. We will in §4.4 evaluate Eq. 4.26 for a simple example of a metal, and in §4.5 do the same for an intrinsic semiconductor.

4.4 Electrical Conductivity of Metals

To exploit the energy dependence of $(\partial f_0/\partial E)$ in applying Eq. 4.26 to metals, it is more convenient to evaluate $\overleftrightarrow{\sigma}$ if we replace $\int d^3k$ with an integral over the constant energy surfaces

$$\int d^3k = \int d^2S dk_\perp \equiv \int d^2S dE/|\partial E/\partial\vec{k}|. \tag{4.27}$$

Thus Eq. 4.26 is written as

$$\overleftrightarrow{\sigma} = -\frac{e^2}{4\pi^3}\int \frac{\tau\vec{v}\vec{v}}{|\partial E/\partial\vec{k}|}\,\frac{\partial f_0}{\partial E}d^2S\,dE. \tag{4.28}$$

From the Fermi-Dirac distribution function $f_0(E)$ shown in Fig. 4.1, we see that the derivative $(-\partial f_0/\partial E)$ can approximately be replaced by a δ-function for the case of a metal, so that Eq. 4.28 can be written as

$$\overleftrightarrow{\sigma} = \frac{e^2}{4\pi^3\hbar}\int_{\text{Fermi surface}} \tau\vec{v}\vec{v}\,\frac{d^2S}{v}. \tag{4.29}$$

For a cubic crystal, $[v_x v_x] = v^2/3$ and thus $\overleftrightarrow{\sigma}$ has only diagonal components σ that are all equal to each other:

$$\sigma = \frac{e^2}{4\pi^3\hbar}\int_{\text{Fermi surface}} \tau v\,\frac{d^2S}{3} = \frac{ne^2\tau}{m^*} \tag{4.30}$$

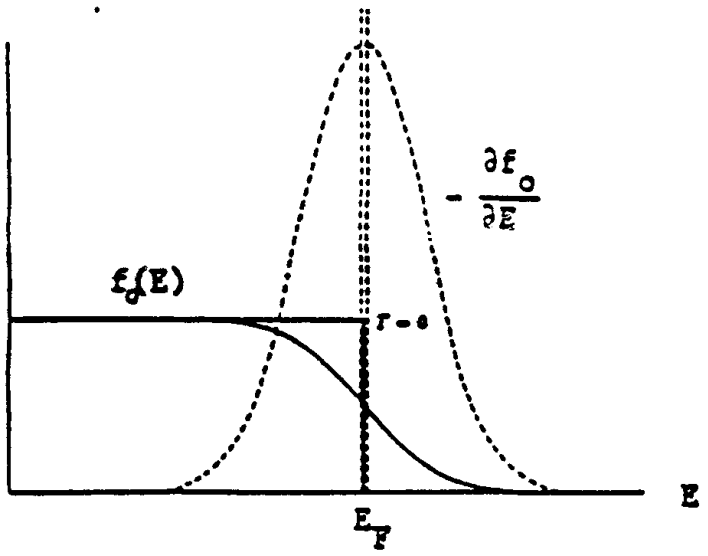


Figure 4.1: Schematic plot of $f_0(E)$ and $-\partial f_0(E)/\partial E$ for a metal showing the δ-function like behavior near the Fermi level E_F for the derivative.

since

$$n = (1/4\pi^3)(4\pi/3)k_F^3 \tag{4.31}$$

and

$$v_F = \hbar k_F/m. \tag{4.32}$$

The result

$$\sigma = ne^2\tau/m^* \tag{4.33}$$

is called the Drude formula for the dc electrical conductivity. Generalization of this methodology to metals with anisotropic Fermi surfaces or with more than one type of carrier can be done directly and requires numerical calculations in most cases.

4.5 Electrical Conductivity of Semiconductors

We show in this section that the simple Drude model $\sigma = ne^2\tau/m^*$ can also be recovered for a semiconductor from the general relation given by Eq. 4.26, using a simple parabolic band model and a constant relaxation time approximation. When a more complete theory is used, departures from the simple Drude model will result.

In deriving the Drude model for a semiconductor we make three approximations:

- Approximation #1
  In the case of electron states in intrinsic semiconductors having no donor or acceptor impurities, we have the condition $(E - E_F) \gg k_BT$ since E_F is in the band gap and E is the energy of an electron in the conduction band, as shown in Fig. 4.2.

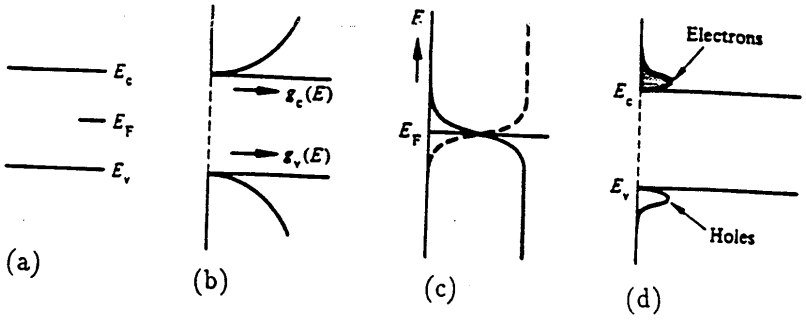


Figure 4.2: Electron and hole states in the conduction and valence bands of an intrinsic semiconductor. (a) Location of E_F in an intrinsic semiconductor. (b) The corresponding density of states for electrons and holes. (c) The Fermi functions for electrons (solid curve) and holes (dashed curve). (d) The occupation of electron and hole states in an intrinsic semiconductor.

Thus, the first approximation is equivalent to writing

$$f_0(E) = \frac{1}{1 + \exp[(E - E_F)/k_B T]} \simeq \exp[-(E - E_F)/k_B T] \tag{4.34}$$

which is equivalent to using the Maxwell–Boltzmann distribution in place of the full Fermi-Dirac distribution. Since E is usually measured with respect to the bottom of the conduction band, E_F is a negative energy and it is therefore convenient to write $f_0(E)$ as

$$f_0(E) \simeq e^{-|E_F|/k_B T} e^{-E/k_B T} \tag{4.35}$$

so that the derivative of the Fermi function becomes

$$\frac{\partial f_0(E)}{\partial E} = -\frac{e^{-|E_F|/k_B T}}{k_B T} e^{-E/k_B T}. \tag{4.36}$$

- Approximation #2
  For simplicity we assume a constant relaxation time τ that is independent of $\vec{k}$ and E. This approximation is made for simplicity and may not be valid for specific cases. Some common scattering mechanisms yield an energy-dependent relaxation time, such as acoustic deformation potential scattering or ionized impurity scattering, where $r = -1/2$ and $r = +3/2$, respectively, in the relation $\tau = \tau_0 (E/k_B T)^r$.

- Approximation #3
  To illustrate the explicit evaluation of the integral in Eq. 4.26, we consider the simplest case, assuming an isotropic, parabolic band $E = \hbar^2 k^2 / 2m^*$ for the evaluation of $\vec{v} = \partial E / \hbar \partial \vec{k}$ about the conduction band extremum.

Using this third approximation we can write

$$\begin{aligned}
\vec{v}\vec{v}&=\tfrac{1}{3}v^2\ \overleftrightarrow{1}\\
k^2&=2m^*E/\hbar^2\\
2kdk&=2m^*dE/\hbar^2\\
v^2&=2E/m^*\\
v&=\hbar k/m^*
\end{aligned}\tag{4.37}$$

where $\overleftrightarrow{1}$ is the unit second rank tensor. We next convert Eq. 4.26 to an integration over energy and write

$$d^3k = 4\pi k^2 dk = 4\pi\sqrt{2}(m^*/\hbar^2)^{3/2}\sqrt{E}dE \tag{4.38}$$

so that Eq. 4.26 becomes

$$\sigma = \frac{e^2\tau}{4\pi^3}\left(\frac{8\sqrt{2}\pi\sqrt{m^*}}{3\hbar^3 k_B T}\right)e^{-|E_F/k_BT}\int_0^\infty E^{3/2}dEe^{-E/k_BT} \tag{4.39}$$

in which the integral over energy E is extended to ∞ because there is negligible contribution for large E and because the definite integral

$$\int_0^\infty x^p dx e^{-x} = \Gamma(p+1) \tag{4.40}$$

can be evaluated exactly, $\Gamma(p)$ being the Γ function which has the property

$$\begin{aligned}
\Gamma(p+1)&=p\Gamma(p)\\
\Gamma(1/2)&=\sqrt{\pi}.
\end{aligned}\tag{4.41}$$

Substitution into Eq. 4.39 thus yields

$$\sigma = \frac{2e^2\tau}{m^*}\left(\frac{m^*k_BT}{2\pi\hbar^2}\right)^{3/2}e^{-|E_F|/k_BT} \tag{4.42}$$

which gives the temperature dependence of σ. Now the carrier density calculated using the same approximations becomes

$$\begin{aligned}
n&=(4\pi^3)^{-1}\ \int f_0(E)\ d^3k\\
&=(4\pi^3)^{-1}\ e^{-|E_F|/k_BT}\int e^{-E/k_BT}4\pi\ k^2dk\\
&=(\sqrt{2}/\pi^2)\left(m^*/\hbar^2\right)^{3/2}e^{-|E_F|/k_BT}\int_0^\infty\sqrt{E}dEe^{-E/k_BT}
\end{aligned}\tag{4.43}$$

where

$$\int_0^\infty \sqrt{E}dEe^{-E/k_BT} = \frac{\sqrt{\pi}}{2}(k_BT)^{3/2} \tag{4.44}$$

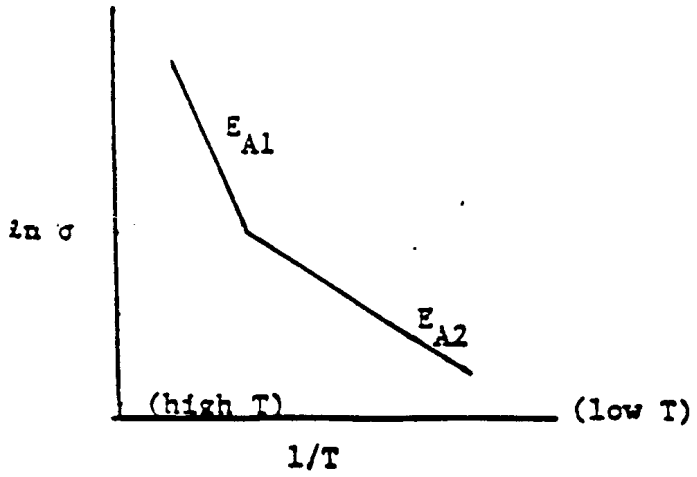


Figure 4.3: Schematic diagram of an Arrhenius plot of $\ln \sigma$ vs $1/T$ showing two carrier types with different activation energies.

which gives the final result for the temperature dependence of the carrier density

$$n = 2\left(\frac{m^* k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-|E_F|/k_B T} \tag{4.45}$$

so that by substitution into Eq. 4.42, the Drude formula is recovered

$$\sigma = \frac{ne^2\tau}{m^*} \tag{4.46}$$

for a semiconductor with constant τ and isotropic, parabolic dispersion relations.

To find σ for a semiconductor with more than one spherical carrier pocket, the conductivities per carrier pocket are added

$$\sigma = \sum_i \sigma_i \tag{4.47}$$

where i is the carrier pocket index. We use these simple formulae to make rough estimates for the carrier density and conductivity of semiconductors. For more quantitative analysis, the details of the $E(\vec{k})$ relation must be considered, as well as an energy dependent τ and use of the complete Fermi function.

The electrical conductivity and carrier density of a semiconductor with one carrier type exhibits an exponential temperature dependence so that the slope of $\ln \sigma$ vs $1/T$ yields an activation energy (see Fig. 4.3). The plot of $\ln \sigma$ vs $1/T$ is called an "Arrhenius plot". If a plot of $\ln \sigma$ vs $1/T$ exhibits one temperature range with activation energy E_{A1} and a second temperature range with activation energy E_{A2}, then two carrier behavior is suggested. Also in such cases, the activation energies can be extracted from an Arrhenius plot as shown in the schematic of Fig. 4.3.

4.5.1 Ellipsoidal Carrier Pockets

The conductivity results given above for a spherical Fermi surface can easily be generalized to an ellipsoidal Fermi surface which is commonly found in degenerate semiconductors. Semiconductors are degenerate at $T = 0$ when the Fermi level is in the valence or conduction band rather than in the energy band gap.

For an ellipsoidal Fermi surface, we write

$$E(\vec{k}) = \frac{\hbar^2 k_x^2}{2m_{xx}} + \frac{\hbar^2 k_y^2}{2m_{yy}} + \frac{\hbar^2 k_z^2}{2m_{zz}} \tag{4.48}$$

where the effective mass components m_{xx}, m_{yy} and m_{zz} are appropriate to the band curvatures in the x, y, z directions, respectively. Substitution of

$$k'_\alpha = k_\alpha \sqrt{m_0/m_\alpha} \tag{4.49}$$

for $\alpha = x, y, z$ brings Eq. 4.48 into spherical form

$$E(\vec{k}') = \frac{\hbar^2 k'^2}{2m_0} \tag{4.50}$$

where $k'^2 = k'^2_x + k'^2_y + k'^2_z$. For the volume element d^3k in Eq. 4.26 we have

$$d^3k = \sqrt{m_{xx}m_{yy}m_{zz}/m_0^3}\, d^3k' \tag{4.51}$$

and the carrier density associated with a single carrier pocket becomes

$$n_i = 2\sqrt{m_{xx}m_{yy}m_{zz}} \left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-|E_F|/k_B T}. \tag{4.52}$$

For an ellipsoidal constant energy surface (see Fig. 4.4), the directions of the electric field, electron velocity and electron acceleration will in general be different. Let (x, y, z) be the coordinate system for the major axes of the constant energy ellipsoid and (X, Y, Z) be the laboratory coordinate system. Then in the laboratory system the current density $\vec{j}$ and electric field $\vec{E}$ are related by

$$\begin{pmatrix} j_X \\ j_Y \\ j_Z \end{pmatrix} = \begin{pmatrix} \sigma_{XX} & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{YX} & \sigma_{YY} & \sigma_{YZ} \\ \sigma_{ZX} & \sigma_{ZY} & \sigma_{ZZ} \end{pmatrix} \begin{pmatrix} E_X \\ E_Y \\ E_Z \end{pmatrix} \tag{4.53}$$

As an example, suppose that the electric field is applied in the XY plane along the X axis at an angle θ with respect to the x axis of the constant energy ellipsoid (see Fig. 4.4). The conductivity tensor is easily written in the xyz crystal coordinate system where the xyz axes are along the principal axes of the ellipsoid:

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = ne^2\tau \begin{pmatrix} 1/m_{xx} & 0 & 0 \\ 0 & 1/m_{yy} & 0 \\ 0 & 0 & 1/m_{zz} \end{pmatrix} \begin{pmatrix} E\cos\theta \\ E\sin\theta \\ 0 \end{pmatrix} \tag{4.54}$$

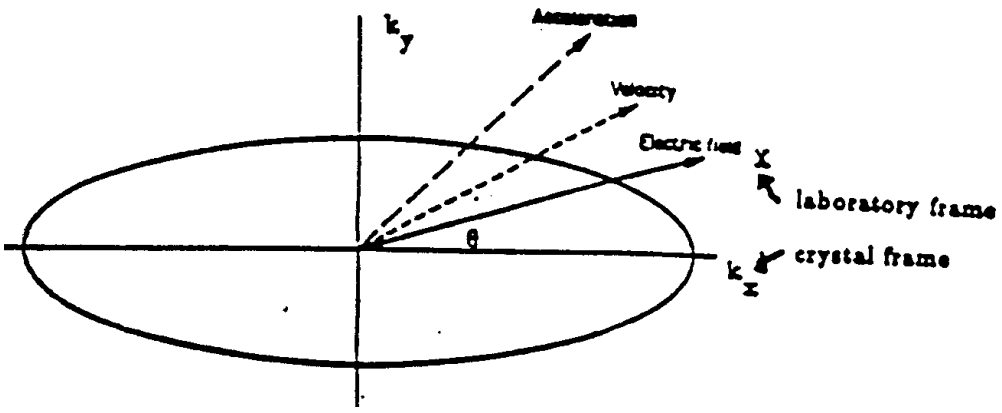


Figure 4.4: Schematic diagram of an ellipsoidal constant energy surface.

A coordinate transformation from the crystal axes to the laboratory frame allows us to relate $\overleftrightarrow{\sigma}_{\text{crystal}}$ which we have written easily by Eq. 4.54 to $\overleftrightarrow{\sigma}_{\text{Lab}}$ which we measure by Eq. 4.53. In general

$$\overleftrightarrow{\sigma}_{\text{Lab}} = R \, \overleftrightarrow{\sigma}_{\text{crystal}} \, R^{-1} \tag{4.55}$$

where

$$R = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.56}$$

and

$$R^{-1} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.57}$$

so that the conductivity tensor $\overleftrightarrow{\sigma}_{\text{Lab}}$ in the lab frame becomes:

$$\overleftrightarrow{\sigma}_{\text{Lab}} = ne^2\tau \begin{pmatrix} \cos^2\theta/m_{xx} + \sin^2\theta/m_{yy} & \cos\theta\sin\theta(1/m_{yy} - 1/m_{xx}) & 0 \\ \cos\theta\sin\theta(1/m_{yy} - 1/m_{xx}) & \sin^2\theta/m_{xx} + \cos^2\theta/m_{yy} & 0 \\ 0 & 0 & 1/m_{zz} \end{pmatrix} \tag{4.58}$$

Semiconductors with ellipsoidal Fermi surfaces usually have several such surfaces located in crystallographically equivalent locations. In the case of cubic symmetry, the sum of the conductivity components results in an isotropic conductivity even though the contribution from each ellipsoid is anisotropic. Thus measurement of the electrical conductivity provides

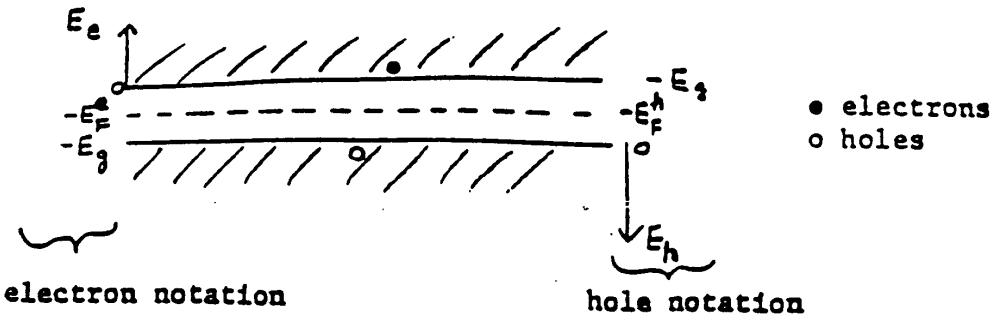


Figure 4.5: Schematic diagram of the band gap in a semiconductor showing the symmetry of electrons and holes.

no information on the anisotropy of the Fermi surfaces of cubic materials. However, measurement of the magnetoresistance does provide such information, since the application of a magnetic field gives special importance to the magnetic field direction, thereby lowering the effective crystal symmetry.

4.6 Electrons and Holes in Intrinsic Semiconductors

Intrinsic semiconductors refer to semiconductors with no doping and no departures from perfect stoichiometry. In this section we consider the symmetry between electron and holes and we show how the Fermi energy is found for such semiconductors. In §4.7, we consider the corresponding issues for doped semiconductors containing impurities or departures from ideal stoichiometry.

In the absence of doping, carriers are generated by thermal or optical excitations. Thus at $T = 0$, all valence band states are occupied and all conduction band states are empty. Thus, for each electron that is excited into the conduction band, a hole is left behind. For intrinsic semiconductors, conduction is by both holes and electrons. The Fermi level is thus determined by the condition that the number of electrons is equal to the number of holes. Writing $g_v(E_h)$ and $g_c(E_e)$ as the density of hole states in the valence band and electron states in the conduction band, respectively, we obtain

$$n_h = \int_0^\infty g_v(E_h)\hat{f}_0(E_h + E_F^h)dE_h = \int_0^\infty g_c(E_e)\hat{f}_0(E_e + E_F^e)dE_e = n_e \tag{4.59}$$

where the notation we have used is shown in Fig. 4.5. Here the energy gap E_g is written as the sum of the Fermi energies for electrons and holes, both taken as positive numbers

$$E_F^e + E_F^h = E_g, \tag{4.60}$$

and the Fermi functions $\hat{f}_0$ are written so as to include explicitly the Fermi energy. The condition $n_e = n_h$ for intrinsic semiconductors is used to determine the position of the Fermi levels for electrons and holes within the band gap. If the band curvatures of the valence and conduction bands are the same, then their effective masses are the same magnitude

and E_F lies at midgap. We also derive in this section the general result for the placement of E_F when $m_e^* \neq m_h^*$.

On the basis of this interpretation, the holes obey Fermi statistics as do the electrons, only we must measure the hole energies downward, while electron energies are measured upwards, as indicated in Fig. 4.5. This approach clearly builds on the symmetry relation between electrons and holes. It is convenient to measure electron energies E_e with respect to the bottom of the conduction band E_c so that $E_e = E - E_c$ and to measure hole energies E_h with respect to the top of the valence band E_v so that $E_h = -(E - E_v)$. The Fermi level for the electrons is $-E_F^e$ (measured from the bottom of the conduction band which is taken as $E = 0$) and for holes it is $-E_F^h$ (measured from the top of the valence band which is taken as $E = 0$ for holes), so that E_F^e and E_F^h have positive values. Referring to Eq. 4.59, $\hat{f}_0(E_e + E_F^e)$ denotes the Fermi function for electrons where $E - E_F$ is written explicitly

$$\hat{f}_0(E_e + E_F^e) = \frac{1}{1 + \exp[(E_e + E_F^e)/k_B T]} \tag{4.61}$$

and is consistent with the definitions given above. A similar expression to Eq. 4.61 follows for $\hat{f}_0(E_h + E_F^h)$.

In an intrinsic semiconductor, the magnitudes of the energies E_F^e and E_F^h are both much greater than thermal energies, i.e., $|E_F^e| \gg k_B T$ and $|E_F^h| \gg k_B T$, where $k_B T$ at room temperature is $\sim$25 meV. Thus the distribution functions can be approximated by the Boltzmann form

$$\begin{aligned} \hat{f}_0(E_e + E_F^e) &\simeq e^{-(E_e + E_F^e)/k_B T} \\ \hat{f}_0(E_h + E_F^h) &\simeq e^{-(E_h + E_F^h)/k_B T}. \end{aligned} \tag{4.62}$$

If m_e and m_h are, respectively, the electron and hole effective masses and if we write the dispersion relations around the valence and conduction band extrema as

$$\begin{aligned} E_e &= \hbar^2 k^2/(2m_e) \\ E_h &= \hbar^2 k^2/(2m_h) \end{aligned} \tag{4.63}$$

then the density of states for electrons at the bottom of the conduction band and for holes at the top of the valence band can be written in their respective nearly free electron forms (see Eq. 4.68)

$$\begin{aligned} g_c(E_e) &= \tfrac{1}{2\pi^2}\left(2m_e/\hbar^2\right)^{3/2} E_e^{1/2} \\ g_v(E_h) &= \tfrac{1}{2\pi^2}\left(2m_h/\hbar^2\right)^{3/2} E_h^{1/2}. \end{aligned} \tag{4.64}$$

These expressions follow from

$$n = \frac{1}{4\pi^3}\frac{4\pi}{3}k^3 \tag{4.65}$$

and substitution of k via the simple parabolic relation

$$E = \frac{\hbar^2 k^2}{2m^*} \tag{4.66}$$

so that

$$n = \frac{1}{3\pi^2}\left(\frac{2m^*E}{\hbar^2}\right)^{3/2} \tag{4.67}$$

and

$$g(E) = \frac{dn}{dE} = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E^{1/2}. \tag{4.68}$$

Substitution of this density of states expression into Eq. 4.45 results in a carrier density

$$n_e = 2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_F^e/k_B T}. \tag{4.69}$$

Likewise for holes we obtain

$$n_h = 2\left(\frac{m_h k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_F^h/k_B T}. \tag{4.70}$$

Thus the famous product rule is obtained

$$n_e n_h = 4\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3} (m_e m_h)^{3/2} e^{-E_g/k_B T} \tag{4.71}$$

where $E_g = E_F^e + E_F^h$. But for an intrinsic semiconductor $n_e = n_h$. Thus by taking the square root of the above expression, we obtain both n_e and n_h

$$n_e = n_h = 2\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} (m_e m_h)^{3/4} e^{-E_g/2k_B T}. \tag{4.72}$$

Comparison with the expressions given in Eqs. 4.69 and 4.70 for n_e and n_h allows us to solve for the Fermi levels E_F^e and E_F^h

$$n_e = 2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_F^e/k_B T} = 2\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} (m_e m_h)^{3/4} e^{-E_g/2k_B T} \tag{4.73}$$

so that

$$\exp(-E_F^e/k_B T) = (m_h/m_e)^{3/4} \exp(-E_g/2k_B T) \tag{4.74}$$

and

$$E_F^e = \frac{E_g}{2} - \frac{3}{4} k_B T \ln(m_h/m_e). \tag{4.75}$$

If $m_e = m_h$, we obtain the simple result that $E_F^e = E_g/2$ which says that the Fermi level lies in the middle of the energy gap. However, if the masses are not equal, E_F will lie closer to the band edge with higher curvature, thereby enhancing the Boltzmann factor term in the thermal excitation process, to compensate for the lower density of states for the higher curvature band.

If however $m_e \ll m_h$, the Fermi level approaches the conduction band edge and the full Fermi functions have to be considered. In this case

$$n_e = \frac{1}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2} \int_{E_c}^{\infty} \frac{(E - E_c)^{1/2} dE}{\exp[(E - E_F^e)/k_B T] + 1} \equiv N_e F_{1/2}\left(\frac{E_F^e - E_c}{k_B T}\right) \tag{4.76}$$

where E_c is the bottom of the conduction band, E_F^e is the Fermi energy for electrons which here is allowed the possibility of moving up into the conduction band (and therefore its sign cannot be predetermined), and N_e is the "effective electron density" which, in accordance with Eq. 4.69, is given by

$$N_e = 2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2}. \tag{4.77}$$

The Fermi integral in Eq. 4.76 is written in standard form as

$$F_j(\eta) = \frac{1}{j!}\int_0^\infty \frac{x^j dx}{\exp(x-\eta)+1}. \tag{4.78}$$

We can take $F_j(\eta)$ from the tables in Blakemore, "Semiconductor Physics" (Appendix B). For the semiconductor limit ($\eta < -4$), then $F_j(\eta) \to \exp(\eta)$. Clearly, when the full version of $F_j(\eta)$ is required to describe the carrier density, then $F_j(\eta)$ is also needed to describe the conductivity. These refinements are important for a detailed solution of the transport properties of semiconductors over the entire temperature range of interest, and eventually in the presence of doping.

4.7 Donor and Acceptor Doping of Semiconductors

In general a semiconductor has electron and hole carriers due to the presence of impurities as well as from thermal excitation processes. For many applications, impurities are intentionally introduced to generate carriers: donor impurities to generate electrons in n-type semiconductors and acceptor impurities to generate holes in p-type semiconductors. Assuming for the moment that each donor contributes one electron to the conduction band, then the donors can contribute an excess carrier concentration up to N_d, where N_d is the donor impurity concentration. Similarly, if every acceptor contributes one hole to the valence band, then the excess hole concentration will be N_a, where N_a is the acceptor impurity concentration. In general, the semiconductor is **partly compensated**, which means that both donor and acceptor impurities are present, thereby giving a partial cancellation to the net carrier concentration. Furthermore, at finite temperatures, the donor and acceptor levels will be partially occupied, so that somewhat less than the maximum charge will be released as **mobile** charge into the conduction and valence bands. The density of electrons bound to a donor site n_d is found from the grand canonical ensemble in statistical mechanics as

$$\frac{n_d}{N_d} = \frac{\sum_j N_j e^{-(E_j-\mu N_j)/k_B T}}{\sum_j e^{-(E_j-\mu N_j)/k_B T}} \tag{4.79}$$

where E_j and N_j are, respectively, the energy and number of electrons that can be placed in state j, and μ is the chemical potential (Fermi energy). Referring to Table 4.1, the system can be found in one of three states: one where no electrons are present in state j (hence no contribution is made to the energy), and two states where one electron is present (one with spin $\uparrow$, the other with spin $\downarrow$) corresponding to the donor energy E_d, where the sign is included directly so that E_d has a positive energy value. Placing two electrons in the same energy state would result in a very high energy because of the Coulomb repulsion between the two electrons; therefore this possibility is neglected in practical calculations. Writing

Table 4.1: Occupation of impurity states in the grand canonical ensemble.

states	N_j	spin	E_j
1	0	-	0
2	1	$\uparrow$	$-E_d$
3	1	$\downarrow$	$-E_d$
4	2	$\uparrow\downarrow$	$-E_d + E_{\text{Coulomb}}$

either $N_j = 0, 1$ for the 3 states of importance, we obtain for the relative ion concentration of occupied donor sites

$$\frac{n_d}{N_d} = \frac{2e^{-(\varepsilon_d - \mu)/k_BT}}{1 + 2e^{-(\varepsilon_d - \mu)/k_BT}} = \frac{1}{1 + \frac{1}{2}e^{(\varepsilon_d - \mu)/k_BT}} = \frac{1}{1 + \frac{1}{2}e^{-(E_d - E_F^e)/k_BT}} \tag{4.80}$$

in which E_d and E_F^e are positive numbers, but lie below the zero of energy which is taken to be at the bottom of the conduction band. The energy ε_d denotes the energy for the donor level and is a negative number relative to the zero of energy.

Consequently, the concentration of electrons thermally ionized into the conduction band will be

$$N_d - n_d = \frac{N_d}{1 + 2e^{(E_d - E_F^e)/k_BT}} = n_e - n_h \tag{4.81}$$

where n_e and n_h are the mobile electron and hole concentrations. At low temperatures, where $E_d \sim k_BT$, almost all of the carriers in the conduction band will be generated by the ionized donors, so that $n_h \ll n_e$ and $(N_d - n_d) \simeq n_e$. The Fermi level will then adjust itself so that $N_d - n_d \simeq n_e$. From Eqs. 4.45 and 4.81 the following equation determines E_F^e:

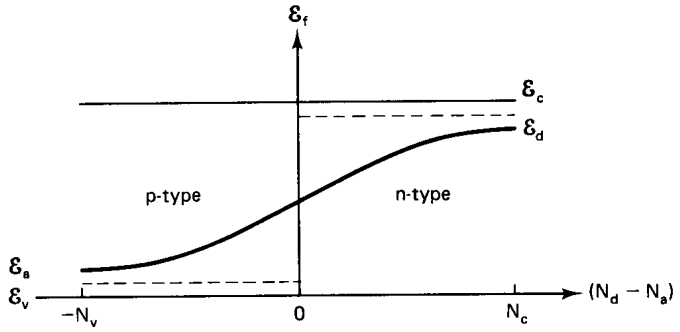
$$n_e = 2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-E_F^e/k_BT} \simeq \frac{N_d}{1 + 2e^{(E_d - E_F^e)/k_BT}}. \tag{4.82}$$

Solution of Eq. 4.82 shows that the presence of the ionized donor carriers moves the Fermi level up above the middle of the band gap and close to the bottom of the conduction band. For the donor impurity problem, the Fermi level will be close to the position of the donor level E_d, as shown in Fig. 4.6. The position of the Fermi level also varies with temperature. Figure 4.6 assumes that almost all the donor electrons (or acceptor holes) are ionized and are in the conduction band, which is typical of temperatures where the n-type (or p-type) semiconductor would be used for electrons.

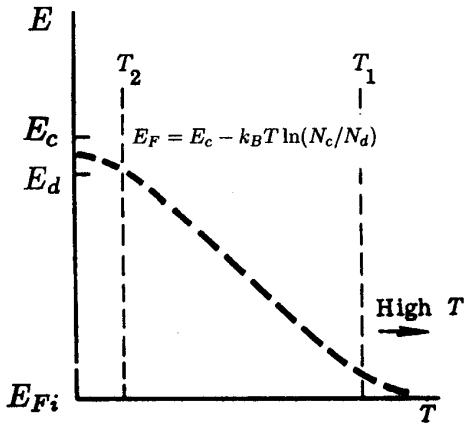
Figure 4.7(a) shows the dependence of the Fermi level on temperature. Here T_1 denotes the temperature at which the thermal excitation of intrinsic electrons and holes become important, and T_1 is normally a high temperature. In contrast, T_2 is normally a very low temperature and denotes the temperature below which donor-generated electrons begin to freeze out in impurity level bound states and no longer contribute to conduction. This carrier freeze-out is illustrated in Fig. 4.7(b). In the temperature range $T_2 < T < T_1$, the Fermi level in Fig. 4.7(a) falls as T increases according to

$$E_F = E_c - k_BT \ln(N_c/N_d) \tag{4.83}$$

Figure 4.6: Variation of the Fermi energy ($E_F \equiv \mathcal{E}_f$) with donor and acceptor concentrations. For a heavily doped n-type semiconductor E_F is close to the donor level E_d, while for a heavily doped p-type semiconductor E_F is close to E_a. This plot is made assuming almost all the donor and acceptor states are ionized.



(a)



(b)

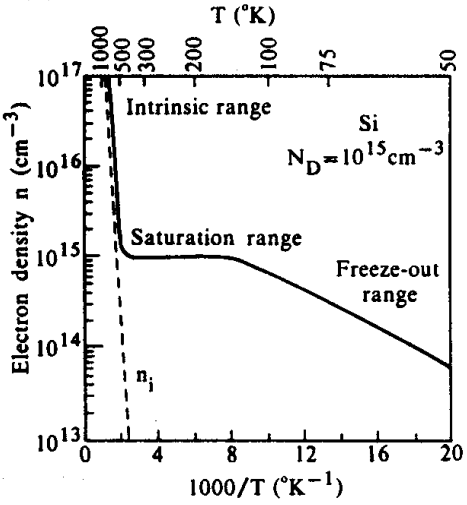


Figure 4.7: (a) Temperature dependence of the Fermi energy for an n-type doped semiconductor. See the text for definitions of T_1 and T_2. Here E_{Fi} denotes the position of the Fermi level in the high temperature limit where the thermal excitation of carriers far exceeds the electron density contributed by the donor impurities. (b) Temperature dependence of the electron density for Si doped with 10^{15} cm^{-3} donors.

where $N_c = 2m_e k_B T/(2\pi\hbar^2)$. In Fig. 4.7(b) we see the temperature dependence of the carrier concentration in the intrinsic range ($T > T_1$), the saturation range ($T_2 < T < T_1$), and finally the low temperature range ($T < T_2$) where carriers freeze out into bound states in the impurity band at E_d. The plot of the electron density n in Fig. 4.7(b) is presented as a function of $(1000/T)$ and the corresponding temperature values are shown on the upper scale of the figure.

For the case of acceptor impurities, an ionized acceptor level releases a hole into the valence band, or alternatively, an electron from the valence band gets excited into an acceptor level, leaving a hole behind. At very low temperature, the acceptor levels are filled with holes under freeze-out conditions. Because of hole-hole Coulomb repulsion, we can place no more than one hole in each acceptor level. A singly occupied hole can have either spin up or spin down. Thus for the acceptor levels, a formula analogous to Eq. 4.80 for donors is obtained for the occupation of an acceptor level

$$\frac{n_a}{N_a} = \frac{1}{1 + \frac{1}{2}e^{-(E_a - E_F^h)/k_B T}} \tag{4.84}$$

so that the essential symmetry between holes and electrons is maintained. To obtain the hole concentration in the valence band, we use a formula analogous to Eq. 4.81.

A situation which commonly arises for the acceptor levels relates to the degeneracy of the valence bands for group IV and III-V compound semiconductors. We will illustrate the degenerate valence band in the case where spin-orbit interaction is considered (which is usually the situation that is relevant for opto-electronic applications). Under strong spin-orbit interaction we have a degenerate heavy and light hole band and a lower lying split-off band. The two degenerate bands are only weakly coupled, so that we can approximate the impurity acceptor levels by hydrogenic acceptor levels for the heavy hole $\varepsilon_{a,h}$ and light hole $\varepsilon_{a,l}$ bands. In this case the split-off band does not contribute significantly because it lies much lower in energy. The density of holes bound to both types of acceptor sites is given by

$$\frac{n_a}{N_a} = \frac{\sum_j N_j e^{-(E_j - \mu N_j)/k_B T}}{\sum_j e^{-(E_j - \mu N_j)/k_B T}}, \tag{4.85}$$

following Eq. 4.79, where we note that the heavy hole and light hole bands can each accommodate one spin up and one spin down electron for each wavevector k. Using the same arguments as above, we obtain:

$$\frac{n_a}{N_a} = \frac{2e^{-(\varepsilon_{a,l} - \mu)/k_B T} + 2e^{-(\varepsilon_{a,h} - \mu)/k_B T}}{1 + 2e^{-(\varepsilon_{a,l} - \mu)/k_B T} + 2e^{-(\varepsilon_{a,h} - \mu)/k_B T}} \tag{4.86}$$

so that

$$\frac{n_a}{N_a} = \frac{1 + e^{-(\varepsilon_{a,h} - \varepsilon_{a,l})/k_B T}}{1 + \frac{1}{2}e^{(\varepsilon_{a,l} - \mu)/k_B T} + e^{-(\varepsilon_{a,h} - \varepsilon_{a,l})/k_B T}}. \tag{4.87}$$

If the thermal energy is large in comparison to the difference between the acceptor levels for the heavy and light hole bands, then

$$[(\varepsilon_{a,h} - \varepsilon_{a,l})/k_B T] \ll 1 \tag{4.88}$$

and

$$\exp[-(\varepsilon_{a,h} - \varepsilon_{a,l})/k_B T] \simeq 1 \tag{4.89}$$

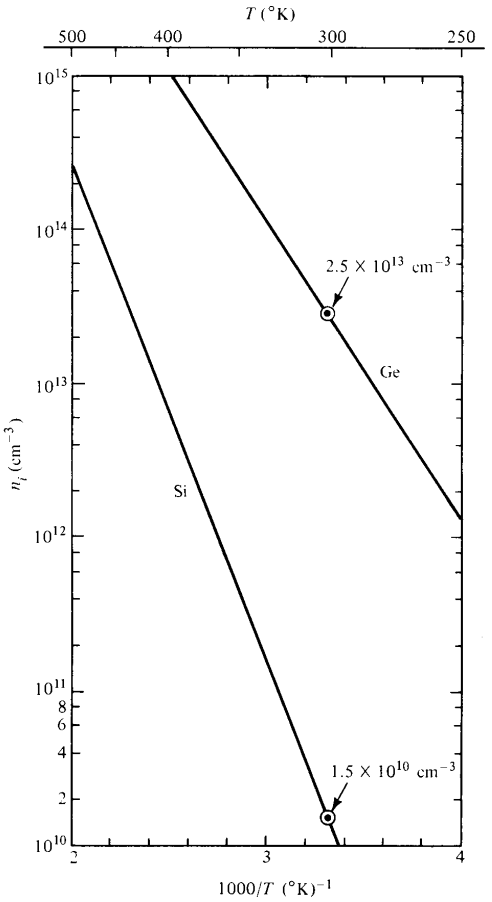


Figure 4.8: Temperature dependence of the electron concentration for intrinsic Si and Ge in the range $250 < T < 500\,\mathrm{K}$. Circles indicate the doping levels that must be exceeded to have extrinsic carriers dominate over thermally excited carriers at 300 K.

so that the density of holes bound to acceptor sites becomes

$$\frac{n_a}{N_a} \simeq \frac{1}{1+\frac{1}{4}e^{-(\varepsilon_{a,l}-\mu)/k_BT}} = \frac{1}{1+\frac{1}{4}e^{(E_a-E_F^h)/k_BT}} \tag{4.90}$$

where E_a and E_F^h are positive values corresponding to $\varepsilon_{a,l}$ and μ, respectively. From Eqs. 4.81 and 4.90, the temperature dependence of E_F can be calculated for the case of doped semiconductors considering doping by donor and acceptor impurities, either separately or at the same time. Figure 4.6 shows the doping dependence of E_F for p-doped semiconductors as well as n-doped semiconductors.

4.8 Characterization of Semiconductors

In describing the electrical conductivity of semiconductors, it is customary to write the conductivity as

$$\sigma = n_e|e|\mu_e + n_h|e|\mu_h \tag{4.91}$$

Table 4.2: Mobilities for some typical semiconductors at room temperature in units of cm^2/V-sec.

Crystal	Electrons	Holes	Crystal	Electrons	Holes
Diamond	1800	1200	GaAs	8000	300
Si	1350	480	GaSb	5000	1000
Ge	3600	1800	PbS	550	600
InSb	77,000	750	PbSe	1020	930
InAs	30,000	460	PbTe	2500	1000
InP	4600	100	AgCl	50	–
AlAs	280	–	KBr (100K)	100	–
AlSb	900	400	SiC	100	10–20

in which n_e and n_h are the carrier densities for the carriers, and μ_e and μ_h are their **mobilities**. We have shown in Eq. 4.46 that for cubic materials the static conductivity can under certain approximations be written as

$$\sigma = \frac{ne^2\tau}{m^*} \tag{4.92}$$

for each carrier type, so that the mobilities and effective masses are related by

$$\mu_e = \frac{|e|\langle\tau_e\rangle}{m_e} \tag{4.93}$$

and

$$\mu_h = \frac{|e|\langle\tau_h\rangle}{m_h} \tag{4.94}$$

which show that materials with small effective masses have high mobilities. By writing the electrical conductivity as a product of the carrier density with the mobility, it is easy to contrast the temperature dependence of σ for metals and semiconductors. For metals, the carrier density n is essentially independent of T, while μ is temperature dependent. In contrast, n for semiconductors is highly temperature dependent in the intrinsic regime [see Fig. 4.7(b)] and μ is relatively less temperature dependent. Figure 4.8 shows the carrier concentration for intrinsic Si and Ge in the neighborhood of room temperature ($250 < T < 500\,\mathrm{K}$), demonstrating the rapid increase of the carrier concentration. These values of n indicate the doping levels necessary to exceed the intrinsic carrier level at a given temperature. Figure 4.9 shows the mobility for n-type Si samples with various impurity levels. The observed temperature dependence can be explained by the different temperature dependences of the impurity scattering and phonon scattering mechanisms (see Fig. 4.10). This is further discussed in Chapter 6.

A table of typical mobilities for semiconductors is given in Table 4.2. By way of comparison, μ for copper at room temperature is $35\,\mathrm{cm}^2$/volt-sec. When using conductivity formulae in esu units, remember that the mobility is expressed in cm^2/statvolt-sec and that all the numbers in Table 4.2 have to be multiplied by 300 to match the units given in the notes.

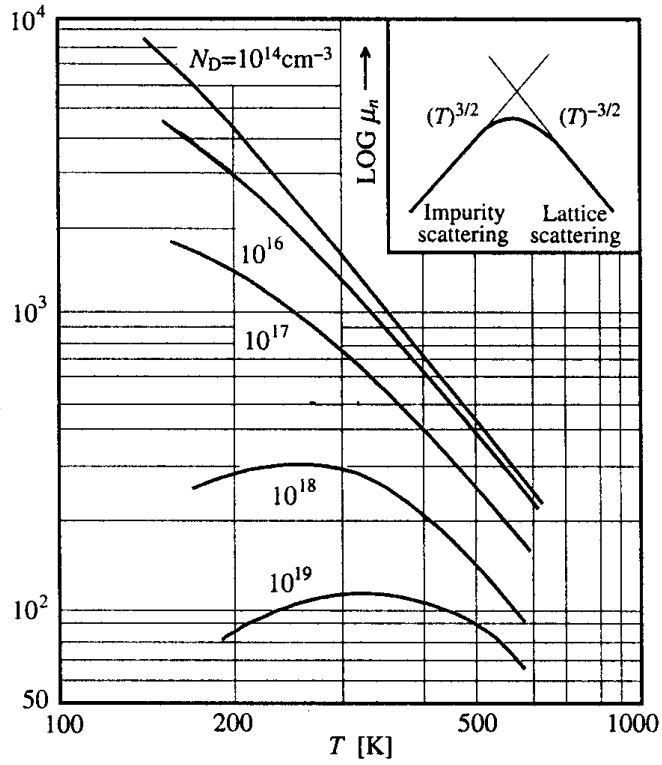


Figure 4.9: Temperature dependence of the mobility for n-type Si for a series of samples with different impurity concentrations. Note that the mobility is not as strong a function of temperature as is the carrier density shown in Fig. 4.8. At low temperature impurity scattering by the donor impurity ions becomes important as shown in the inset. The different temperature dependences of impurity and electron-phonon scattering allows one to identify the important scattering mechanisms experimentally.

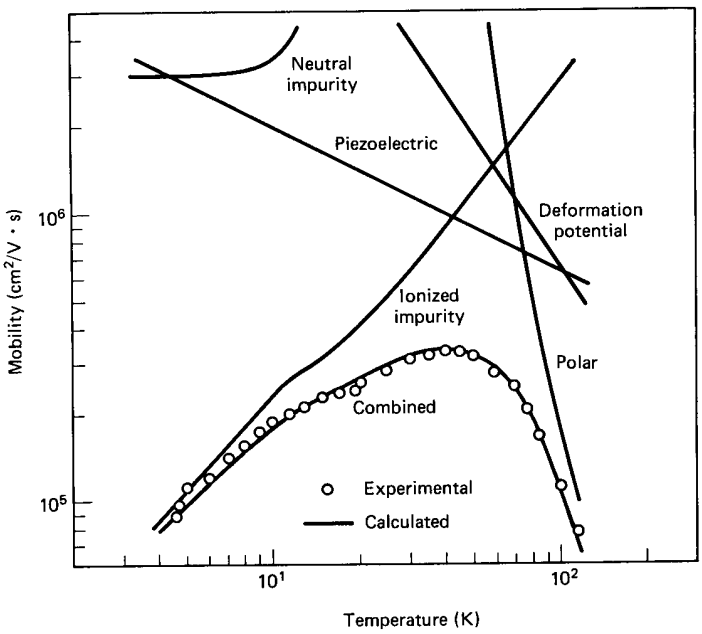


Figure 4.10: Temperature dependence of the mobility for n-type GaAs showing the separate and combined scattering processes.

In the characterization of a semiconductor for device applications, researchers are expected to provide information on the carrier density and mobility, preferably as a function of temperature. Such plots are shown in Figs. 4.8 and 4.9. When presenting characterization data in condensed form, the carrier density and mobility of semiconductors are traditionally given at 300 K and 77 K. Other information of values in semiconductor physics are values of the effective masses (Table 4.3) and of the energy gaps (Table 4.4).

Table 4.3: Semiconductor effective masses of electrons and holes in direct gap semiconductors.

Crystal	Electron	Holes			Spin-orbit
	m_e/m_0	m_{hh}/m_0	m_{lh}/m_0	m_{soh}/m_0	Δ(eV)
InSb	0.015	0.39	0.021	(0.11)	0.82
InAs	0.026	0.41	0.025	0.08	0.43
InP	0.073	0.4	(0.078)	(0.15)	0.11
GaSb	0.047	0.3	0.06	(0.14)	0.80
GaAs	0.066	0.5	0.082	0.17	0.34

Table 4.4: Semiconductor energy gaps between the valence and conduction bands.

		E_g, eV				E_g, eV	
Crystal	Gap[a]	0 K	300 K	Crystal	Gap[a]	0 K	300 K
Diamond	i	5.4	–	HgTe[b]	d	–	–
Si	i	1.17	1.11	PbS	d	0.286	0.34-0.37
Ge	i	0.744	0.66	PbSe	i	0.165	0.27
αSn	d	0.00	0.00	PbTe	i	0.190	0.29
InSb	d	0.23	0.17	CdS	d	2.582	2.42
InAs	d	0.43	0.36	CdSe	d	1.840	1.74
InP	d	1.42	1.27	CdTe	d	1.607	1.44
GaP	i	2.32	2.25	ZnO	–	3.436	3.2
GaAs	d	1.52	1.43	ZnS	–	3.91	3.6
GaSb	d	0.81	0.68	SnTe	d	0.3	0.18
AlSb	i	1.65	1.6	AgCl	–	–	3.2
SiC(hex)	i	3.0	–	AgI	–	–	2.8
Te	d	0.33	–	Cu_2O	d	2.172	–
ZnSb	–	0.56	0.56	TiO_2	–	3.03	–

[a]The indirect gap is labeled by i, and the direct gap is labeled by d.
[b]HgTe is a zero gap semiconductor, and because of non-ideal stoichiometry, the Fermi level may be in the valence or conduction band.

Chapter 5

Thermal Transport

References:

- Ziman, *Principles of the Theory of Solids*, Cambridge Univ. Press, 1972, Chapters 7.
- Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, 1965, pp. 393-397.
- Wolfe, Holonyak and Stillman, *Physical Properties of Semiconductors*, Prentice Hall, 1989, chapter 5.

5.1 Thermal Transport

The electrons in solids not only conduct electricity but also conduct heat, as they transfer energy from a hot junction to a cold junction. Just as the electrical conductivity characterizes the response of a material to an applied voltage, the thermal conductivity likewise characterizes the material with regard to heat flow in the presence of a temperature gradient. In fact the electrical conductivity and thermal conductivity are coupled, since thermal conduction also transports charge and electrical conduction also transports energy. This coupling between electrical and thermal transport gives rise to thermo-electricity. In this chapter, we discuss first the thermal conductivity for metals, semiconductors and insulators and then consider the coupling between electrical and thermal transport which gives rise to thermoelectric phenomena. In Chapter 6, we discuss scattering mechanisms for electrons and phonons.

5.2 Thermal Conductivity

5.2.1 General Considerations

Thermal transport, like electrical transport follows from the Boltzmann equation. We will first derive a general expression for the electronic contribution to the thermal conductivity using Boltzmann's equation. We will then apply this general expression to find the thermal conductivity for metals and then for semiconductors. The total thermal conductivity $\overleftrightarrow{\kappa}$ of

any material is, of course, the superposition of the electronic part $\overleftrightarrow{\kappa}_e$ with the lattice part $\overleftrightarrow{\kappa}_L$:

$$\overleftrightarrow{\kappa} = \overleftrightarrow{\kappa}_e + \overleftrightarrow{\kappa}_L \ . \tag{5.1}$$

We now consider the calculation of the electronic contribution to the thermal conductivity. The contribution of the phonons to the thermal conductivity is considered in §5.2.4. The application of a temperature gradient to a solid gives rise to a flow of heat. We define $\vec{U}$ as the thermal current that is driven by the heat energy $E - E_F$, which in turn is the excess energy of an electron above the equilibrium energy E_F. Neglecting time dependent effects, we define $\vec{U}$ as

$$\vec{U} = \frac{1}{4\pi^3} \int \vec{v}\,(E - E_F)\ f(\vec{r},\vec{k}) d^3k \tag{5.2}$$

where the distribution function $f(\vec{r},\vec{k})$ is related to the Fermi function f_0 by $f = f_0 + f_1$. Under equilibrium conditions there is no thermal current density

$$\int \vec{v}(E - E_F) f_0 \ d^3k = 0 \tag{5.3}$$

so that the thermal current is driven by the thermal gradient which causes a departure from the equilibrium distribution:

$$\vec{U} = \frac{1}{4\pi^3} \int \vec{v}(E - E_F) f_1 \ d^3k \tag{5.4}$$

where the electronic contribution to the thermal conductivity tensor $\overleftrightarrow{\kappa}_e$ is defined by the relation

$$\vec{U} = - \overleftrightarrow{\kappa}_e \cdot \frac{\partial T}{\partial \vec{r}}. \tag{5.5}$$

Assuming no explicit time dependence for the distribution function, the function f_1 representing the departure of the distribution from equilibrium is found from solution of Boltzmann's equation

$$\vec{v} \cdot \frac{\partial f}{\partial \vec{r}} + \dot{\vec{k}} \cdot \frac{\partial f}{\partial \vec{k}} = -\frac{f_1}{\tau} \tag{5.6}$$

for a time independent temperature gradient. In the absence of an electric field, $\dot{\vec{k}} = 0$ and the drift velocity $\vec{v}$ is found from the equation

$$\vec{v} \cdot \frac{\partial f}{\partial \vec{r}} = -\frac{f_1}{\tau}. \tag{5.7}$$

Using the linear approximation for the term $\partial f/\partial \vec{r}$ in the Boltzmann equation, we obtain

$$\begin{aligned} \tfrac{\partial f}{\partial \vec{r}} \cong \tfrac{\partial f_0}{\partial \vec{r}} &= \tfrac{\partial}{\partial \vec{r}}\left[\tfrac{1}{1+e^{(E-E_F)/k_BT}}\right] \\ &= \left\{-\tfrac{e^{(E-E_F)/k_BT}}{[1+e^{(E-E_F)/k_BT}]^2}\right\}\left\{-\tfrac{1}{k_BT}\tfrac{\partial E_F}{\partial \vec{r}} - \tfrac{(E-E_F)}{k_BT^2}\tfrac{\partial T}{\partial \vec{r}}\right\} \\ &= \{k_BT\tfrac{\partial f_0}{\partial E}\}\{-\tfrac{1}{k_BT}\tfrac{\partial T}{\partial \vec{r}}\}\{\tfrac{\partial E_F}{\partial T} + \tfrac{(E-E_F)}{T}\} = -\tfrac{\partial f_0}{\partial E}\{\tfrac{\partial E_F}{\partial T} + \tfrac{(E-E_F)}{T}\}\tfrac{\partial T}{\partial \vec{r}}. \end{aligned} \tag{5.8}$$

We will now give some typical values for these two terms for semiconductors and metals. For semiconductors, we evaluate the expression in Eq. 5.8 by referring to Eq. 4.75

$$E_F = \frac{1}{2}E_g - \frac{3}{4}k_BT\ \ell n(m_h/m_e) \tag{5.9}$$

from which

$$\frac{\partial E_F}{\partial T} \sim \frac{3}{4} k_B \, \ell n(m_h/m_e) \tag{5.10}$$

showing that the temperature dependence of E_F arises from the inequality of the valence and conduction band effective masses. If $m_h = m_e$, which would be the case of strongly coupled "mirror" bands, then $\partial E_F/\partial T$ would vanish. For a significant mass difference such as m_h/m_e=2, we obtain $\partial E_f/\partial T \sim 0.5 k_B$ from Eq. 5.10. For a band gap of 0.5 eV and the Fermi level in the middle of the gap, we obtain for the other term in Eq. 5.8

$$[(E - E_F)/T] \approx [0.5/(1/40)]k_B = 20 k_B \tag{5.11}$$

where $k_B T \approx 1/40$ eV at room temperature. Thus for a semiconductor, the term $(E - E_F)/T$ is much larger than the term $(\partial E_F/\partial T)$.

For a metal with a spherical Fermi surface, the following relation

$$E_F = E_F^0 - \frac{\pi^2}{12} \frac{(k_B T)^2}{E_F^0} \tag{5.12}$$

is derived in standard textbooks on statistical mechanics, so that at room temperature and assuming that for a typical metal $E_F^0 = 5$ eV, we obtain from Eq. 5.12

$$\left| \frac{\partial E_F}{\partial T} \right| = \frac{\pi^2}{6} \frac{(k_B T)}{E_F^0} k_B \approx \frac{10}{6} \frac{(\frac{1}{40})}{5} k_B \approx 8 \times 10^{-3} k_B. \tag{5.13}$$

Thus, for both semiconductors and metals, the term $(E - E_F)/T$ tends to dominate over $(\partial E_F/\partial T)$, though there can be situations where the term $(\partial E_F/\partial T)$ cannot be neglected. In this presentation, we will temporarily neglect the term ∇E_F in Eq. 5.8 in calculation of the electronic contribution to the thermal conductivity, but we will include this term formally in our derivation of thermoelectric effects in §5.3.

Typically, the electron energies of importance in any transport problem are those within $k_B T$ of the Fermi energy so that for many applications for metals, we can make the rough approximation,

$$\frac{E - E_F}{T} \approx k_B \tag{5.14}$$

though the results given in this section are derived without the above approximation for $(E - E_F)/T$. Rather, all integrations are carried out in terms of the variable $(E - E_F)/T$.

We return now to the solution of the Boltzmann equation in the relaxation time approximation

$$\frac{\partial f}{\partial \vec{r}} = \left(-\frac{\partial f_0}{\partial E} \right) \left(\frac{E - E_F}{T} \right) \left(\frac{\partial T}{\partial \vec{r}} \right). \tag{5.15}$$

Solution of the Boltzmann equation yields

$$f_1 = -\tau \vec{v} \cdot \left(\frac{\partial f}{\partial \vec{r}} \right) = \tau \vec{v} \left(\frac{\partial f_0}{\partial E} \right) \left(\frac{E - E_F}{T} \right) \frac{\partial T}{\partial \vec{r}}. \tag{5.16}$$

Substitution of f_1 in the equation for the thermal current

$$\vec{U} = \frac{1}{4\pi^3} \int \vec{v} (E - F_F) f_1 d^3 k \tag{5.17}$$

then results in

$$\vec{U} = \frac{1}{4\pi^3 T}\left(\frac{\partial T}{\partial \vec{r}}\right) \cdot \int \tau \vec{v}\vec{v}(E - E_F)^2 \left(\frac{\partial f_0}{\partial E}\right) d^3k. \tag{5.18}$$

Using the definition of the thermal conductivity tensor $\overleftrightarrow{\kappa}_e$ given by Eq. 5.5 we write the electronic contribution to the thermal conductivity $\overleftrightarrow{\kappa}_e$ as

$$\overleftrightarrow{\kappa}_e = \frac{-1}{4\pi^3 T} \int \tau \vec{v}\vec{v}(E - E_F)^2 \left(\frac{\partial f_0}{\partial E}\right) d^3k \tag{5.19}$$

where $d^3k = d^2S\ dk_\perp = d^2S\ dE/|\partial E/\partial \vec{k}| = d^2S dE/(\hbar v)$ is used to exploit our knowledge of the dependence of the distribution function on the energy as discussed below.

5.2.2 Thermal Conductivity for Metals

In the case of a metal, the integral for $\overleftrightarrow{\kappa}_e$ given by Eq. 5.19 can be evaluated easily by converting the integral over phase space $\int d^3k$ to an integral over $\int dE\ d^2S_F$ in order to exploit the δ-function property of $-(\partial f_0/\partial E)$. We then make use of the following result that you will show for homework and can be found in any standard statistical mechanics text (see for example, Reif)

$$\int G(E)\left(-\frac{\partial f_0}{\partial E}\right) dE = G(E_F) + \frac{\pi^2}{6}(k_B T)^2 \left[\frac{\partial^2 G}{\partial E^2}\right]_{E_F} + \ldots \tag{5.20}$$

It is necessary to consider the expansion given in Eq. 5.20 in solving Eq. 5.19 since $G(E_F)$ vanishes at $E = E_F$ for the integral defined in Eq. 5.19 for $\overleftrightarrow{\kappa}_e$. To solve the integral equation of Eq. 5.19 we make the identification of

$$G(E) = g(E)(E - E_F)^2 \tag{5.21}$$

where

$$g(E) = \frac{1}{4\pi^3} \int \tau \vec{v}\vec{v} d^2S/\hbar v \tag{5.22}$$

so that $G(E_F) = 0$ and $(\partial G/\partial E)|_{E_F} = 0$ while

$$\left[\frac{\partial^2 G}{\partial E^2}\right]_{E_F} = G''(E_F) = 2g(E_F). \tag{5.23}$$

These relations will be used again in connection with the calculation of the thermopower in §5.3.1. For the case of the thermal conductivity for a metal we then obtain

$$\overleftrightarrow{\kappa}_e = \frac{\pi^2}{3}(k_B T)^2 g(E_F) = \frac{(k_B T)^2}{12\pi\hbar} \int \tau \vec{v}\vec{v} \frac{d^2S_F}{v} \tag{5.24}$$

where the integration is over the Fermi surface. We immediately recognize that the integral appearing in Eq. 5.24 is the same as that for the electrical conductivity (see Eqs. 4.26 and 4.29)

$$\overleftrightarrow{\sigma} = \frac{e^2}{4\pi^3\hbar} \int \tau \vec{v}\vec{v} \frac{d^2S_F}{v} \tag{5.25}$$

so that the electronic contribution to the thermal conductivity and the electrical conductivity tensors are proportional to each other

$$\overleftrightarrow{\kappa}_e = \overleftrightarrow{\sigma}\, T \left(\frac{\pi^2 k_B^2}{3e^2} \right) \tag{5.26}$$

and Eq. 5.26 is known as the Wiedemann–Franz Law. The physical basis for this relation is that in electrical conduction each electron carries a charge e and experiences an electrical force $e\vec{E}$ so that the electrical current per unit field is e^2. In thermal conduction, each electron carries a unit of thermal energy $k_B T$ and experiences a thermal force $k_B \partial T/\partial \vec{r}$ so that the heat current per unit thermal gradient is proportional to $k_B^2 T$. Therefore the ratio of $|\kappa_e|/|\sigma|$ must be on the order of $(k_B^2 T/e^2)$. The Wiedemann–Franz law suggests that the ratio $\kappa_e/(\sigma T)$ should be a constant (called the Lorenz number), independent of materials properties

$$\left| \frac{\kappa_e}{\sigma T} \right| = \frac{\pi^2}{3} \left(\frac{k_B}{e} \right)^2 = 2.45 \times 10^{-8} \quad \text{watt ohm/deg}^2. \tag{5.27}$$

The ratio $(\kappa_e/\sigma T)$ is approximately constant for all metals at high temperatures $T > \Theta_D$ and at very low temperatures $T \ll \Theta_D$, where Θ_D is the Debye temperature. The derivation of the Wiedemann–Franz Law depends on the relaxation time approximation, which is valid at high temperatures $T > \Theta_D$ where the electron scattering is dominated by the quasi-elastic phonon scattering process and is valid also at very low temperatures $T \ll \Theta_D$ where phonon scattering is unimportant and the dominant scattering mechanism is impurity and lattice defect scattering, both of which tend to be elastic scattering processes. These scattering processes are discussed in Chapter 6 where we discuss in more detail the temperature dependence for κ. When specific scattering processes are considered in detail, the value of the Lorenz number may change.

The temperature dependence of the thermal conductivity of a metal is given in Fig. 5.1. From Eq. 5.27 we can write the following relation for the electronic contribution to the thermal conductivity κ_e when the Wiedemann–Franz law is satisfied

$$\kappa_e = \left(\frac{ne^2\tau}{m^*} \right) T \frac{\pi^2}{3} \left(\frac{k_B}{e} \right)^2. \tag{5.28}$$

At very low temperatures where scattering by impurities, defects, crystal boundaries is dominant, σ is independent of T and therefore from the Wiedemann–Franz law, $\kappa_e \sim T$. At somewhat higher temperatures, but still in the regime $T \ll \Theta_D$, electron-phonon scattering starts to dominate and κ_e starts to decrease. In this regime, the electrical conductivity exhibits a T^{-5} dependence. However only small q phonons participate in this regime. Thus it is only the phonon density which increases as T^3 that is relevant to the phonon-electron scattering, thereby yielding an electrical resistivity with a T^3 dependence and a conductivity with a T^{-3} dependence. Using Eq. 5.28, we thus find that in the low T range, where only low q phonons participate in thermal transport, κ_e should show a T^{-2} dependence, in agreement with Fig. 5.1. At high T where all the phonons contribute to thermal transport, we have $\sigma \sim 1/T$ so that κ_e becomes independent of T. Since $\Theta_D \sim 300$ K for Cu, this temperature range far exceeds the upper limit of Fig. 5.1.

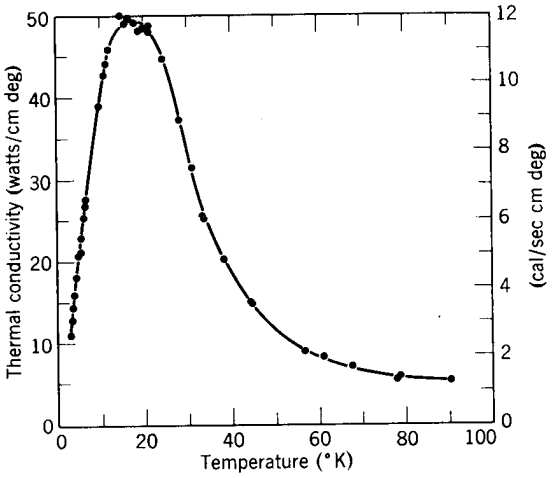


Figure 5.1: The temperature dependence of the thermal conductivity of copper. Note that both κ and T are plotted on linear scales. At low temperatures where the phonon density is low, the thermal transport is by electrons predominantly, while at high temperatures, thermal transport by phonons becomes more important.

5.2.3 Thermal Conductivity for Semiconductors

For the case of non-degenerate semiconductors, the integral for $\overleftrightarrow{\kappa}_e$ in Eq. 5.19 is evaluated by replacing $(E - E_F) \rightarrow E$, since in a semiconductor the electrons that can conduct heat must be in the conduction band, and the lowest energy an electron can have in the thermal conduction process is at the conduction band minimum which is taken as the zero of energy in these calculations. Then the thermal conductivity for a non-degenerate semiconductor can be written as

$$\overleftrightarrow{\kappa}_e = \frac{1}{4\pi^3 T} \int \tau \vec{v}\vec{v}\, E^2 \Big(- \frac{\partial f_0}{\partial E} \Big) d^3k. \tag{5.29}$$

For intrinsic semiconductors, the Fermi distribution function can normally be approximated by the Maxwell–Boltzmann distribution so that

$$(\partial f_0/\partial E) \rightarrow -(1/k_B T) e^{(-|E_F^e|/k_B T)} e^{(-E/k_B T)}, \tag{5.30}$$

assuming that the doping level is not high enough to push E_F close to the band edge (band extremum) or into the conduction band to produce a degenerate semiconductor. For a parabolic band we have $E = \hbar^2 k^2 / 2m^*$, so that the volume element in reciprocal space can be written as

$$\int d^3k = \int 4\pi k^2 dk = \int_0^\infty 2\pi (2m^*/\hbar^2)^{3/2} E^{1/2} dE, \tag{5.31}$$

and $\vec{v} = (1/\hbar)(\partial E/\partial \vec{k}) = \hbar\vec{k}/m$. Assuming a constant relaxation time, we then substitute all these terms into Eq. 5.29 for $\overleftrightarrow{\kappa}_e$ and integrate to obtain

$$\begin{aligned}\kappa_{exx} &= (1/4\pi^3 T) \int \tau v_x^2 E^2 \left[(k_B T)^{-1} e^{-|E_F^e|(/k_B T)}\ e^{-E/(k_B T)}\right] 2\pi (2m^*/\hbar^2)^{3/2} E^{1/2} dE \\ &= \left[k_B (k_B T)\tau/(3\pi^2 m^*)\right] (2m^* k_B T/\hbar^2)^{3/2}\ e^{-|E_F^e|/(k_B T)} \int_0^\infty x^{7/2} e^{-x} dx\end{aligned} \tag{5.32}$$

where $\int_0^\infty x^{7/2} e^{-x} dx = 105\sqrt{\pi}/8$, from which it follows that κ_{exx} has a temperature dependence of the form

$$T^{5/2} e^{-|E_F^e|/(k_B T)} \tag{5.33}$$

in which the exponential term is dominant for temperatures of physical interest, where $k_B T \ll |E_F^e|$. We note from Eq. 4.42 that for a semiconductor, the temperature dependence of the electrical conductivity is given by

$$\sigma_{xx} = \frac{2e^2\tau}{m^*} \left(\frac{m^* k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-|E_F^e|/(k_B T)}. \tag{5.34}$$

Assuming cubic symmetry, we can write the conductivity tensor as

$$\overleftrightarrow{\sigma} = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{xx} & 0 \\ 0 & 0 & \sigma_{xx} \end{pmatrix} \tag{5.35}$$

so that the electronic contribution to the thermal conductivity of a semiconductor can be written as

$$\kappa_{exx} = \left(\frac{35}{2}\right)\left(\frac{k_B^2}{e^2}\right) \sigma_{xx} T \tag{5.36}$$

where

$$\sigma_{xx} = ne^2\tau/m_{xx} = ne\mu_{xx} \tag{5.37}$$

and we note that the coefficient (35/2) for this calculation for semiconductors is different from the corresponding coefficient ($\pi^2/3$) for metals (see Eq. 5.27). Except for numerical constants, the formal results relating the electronic contribution to the thermal conductivity κ_{exx} and σ_{xx} are similar for metals and semiconductors, with the electronic thermal conductivity and electrical conductivity being proportional.

A major difference between semiconductors and metals is the magnitude of the electrical conductivity and hence of the electronic contribution to the thermal conductivity. Since σ_{xx} is much smaller for semiconductors than for metals, κ_e for semiconductors is relatively unimportant and the thermal conductivity tends to be dominated by the lattice contribution κ_L.

5.2.4 Thermal Conductivity for Insulators

In the case of insulators, heat is only carried by phonons (lattice vibrations). The thermal conductivity in insulators therefore depends on phonon scattering mechanisms (see

Chapter 6). The lattice thermal conductivity is calculated from kinetic theory and is given by

$$\kappa_L = \frac{C_p \overline{v}_q \Lambda_{\rm ph}}{3} \tag{5.38}$$

where C_p is the heat capacity, $\overline{v}_q$ is the average phonon velocity and $\Lambda_{\rm ph}$ is the phonon mean free path. As discussed above, the total thermal conductivity of a solid is given as the sum of the lattice contribution κ_L and the electronic contribution κ_e. For metals the electronic contribution dominates, while for insulators and semiconductors the phonon contribution dominates. Let us now consider the temperature dependence of κ_{exx} (see Fig. 5.2) for heat conduction by phonons. At very low T in the defect scattering range, the heat capacity has a dependence $C_p \propto T^3$ while $\overline{v}_q$ and $\Lambda_{\rm ph}$ are almost independent of T. As T increases and we enter the phonon-phonon scattering regime due to normal scattering processes and involving only low q phonons, C_p is still increasing with T but the increase is slower than T^3, while $\overline{v}_q$ remains independent of T and $\Lambda_{\rm ph}$. As T increases further, the thermal conductivity increases more and more gradually and eventually starts to decrease because of phonon-phonon scattering events, for which the density of phonons available for scattering depends on the Bose–Einstein factor $[\exp(\hbar\omega/k_BT) - 1]$. This causes a peak in $\kappa_L(T)$. The decrease in $\kappa_L(T)$ becomes more pronounced as C_p becomes independent of T and $\Lambda_{\rm ph}$ continues to be proportional to $[\exp(\hbar\bar{\omega}/k_BT) - 1]$ where $\bar{\omega}$ is a typical phonon frequency (see §6.4.1). As T increases further, we eventually enter the $T \gg \Theta_D$ regime, where Θ_D is the Debye temperature. In this regime, the temperature dependence of $\Lambda_{\rm ph}$ simply becomes $\Lambda_{\rm ph} \sim (1/T)$. Referring to Fig. 5.2 for $\kappa(T)$ for NaF we see that the peak in κ occurs at about 18 K where the complete Bose–Einstein factor $[\exp(\hbar\bar{\omega}/k_BT) - 1]$ must be used to describe the T dependence of κ_L. For much of the temperature range in Fig. 5.2, only low q phonons participate in the thermal conduction process. At higher temperatures where larger q phonons contribute to thermal conduction, umklapp processes become important in the phonon scattering process, as discussed in §6.3.1. The discussion in this section also applies to the lattice contribution to the thermal conductivity for metals, semimetals and semiconductors.

5.3 Thermoelectric Phenomena

In many metals and semiconductors there exists a coupling between the electrical current and the thermal current. This coupling can be appreciated by observing that when electrons carry thermal current, they are also transporting charge and therefore generating electric fields. This coupling between the charge transport and heat transport gives rise to thermoelectric phenomena. In our discussion of thermoelectric phenomena we start with a general derivation of the coupled equations for the electrical current density $\vec{j}$ and the thermal current density $\vec{U}$:

$$\vec{j} = \frac{e}{4\pi^3} \int \vec{v} f_1 d^3k \tag{5.39}$$

$$\vec{U} = \frac{1}{4\pi^3} \int \vec{v}(E - E_F) f_1 d^3k \tag{5.40}$$

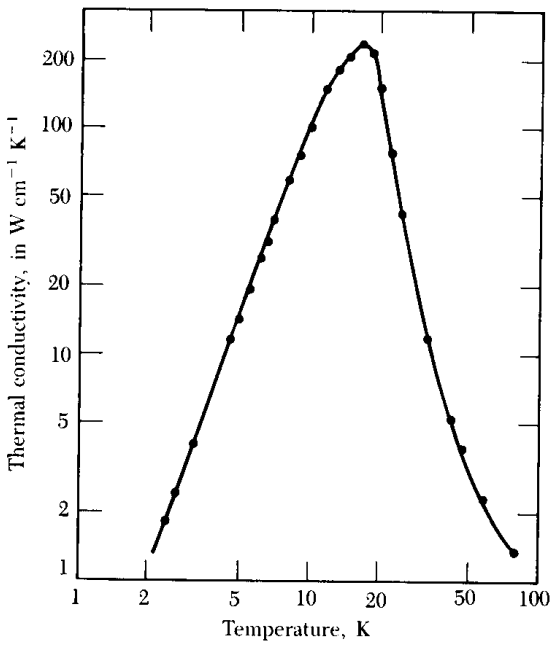


Figure 5.2: Temperature dependence of the thermal conductivity of a highly purified insulating crystal of NaF. Note that both κ and T are plotted on a log scale, and that the peak in κ occurs at quite a low temperature (~ 17 K). The temperature dependence of κ is further discussed in §6.4.

and the perturbation to the distribution function f_1 is found from solution of Boltzmann's equation in the relaxation time approximation:

$$\vec{v} \cdot \frac{\partial f}{\partial \vec{r}} + \dot{\vec{k}} \cdot \frac{\partial f}{\partial \vec{k}} = -\frac{(f - f_0)}{\tau}, \tag{5.41}$$

which is written here for the case of time independent forces and fields. Substituting for $(\partial f/\partial \vec{r})$ from Eqs. 5.8 and 5.15, for $\partial f/\partial \vec{k}$ from Eq. 4.17, and for $\partial f/\partial \vec{k} = (\partial f_0/\partial E)(\partial E/\partial \vec{k})$ yields

$$\vec{v} \cdot (\partial f_0/\partial E)\Big[(e\vec{E} - \vec{\nabla} E_F) - \frac{(E - E_F)}{T}(\vec{\nabla} T)\Big] = -\frac{f_1}{\tau}, \tag{5.42}$$

so that the solution to the Boltzmann equation in the presence of an electric field and a temperature gradient is

$$f_1 = \vec{v}\tau \cdot (\partial f_0/\partial E)\Big\{[(E - E_F)/T]\vec{\nabla} T - e\vec{E} + \vec{\nabla} E_F\Big\}, \tag{5.43}$$

in which e is negative for electrons and positive for holes. The electrical and thermal currents in the presence of both an applied electric field and a temperature gradient can thus be obtained by substituting f_1 into $\vec{j}$ and $\vec{U}$ in Eqs. 5.39 and 5.40 to yield expressions of the form:

$$\vec{j} = e^2 \overleftrightarrow{\kappa}_0 \cdot (\vec{E} - \frac{1}{e}\vec{\nabla} E_F) - (e/T) \overleftrightarrow{\kappa}_1 \cdot \vec{\nabla} T \tag{5.44}$$

and

$$\vec{U} = e \overleftrightarrow{\kappa}_1 \cdot (\vec{E} - \frac{1}{e}\vec{\nabla} E_F) - (1/T) \overleftrightarrow{\kappa}_2 \cdot \vec{\nabla} T \tag{5.45}$$

where $\overleftrightarrow{\kappa}_0$ is related to the conductivity tensor $\overleftrightarrow{\sigma}$ by

$$\overleftrightarrow{\kappa}_0 = \frac{1}{4\pi^3} \int \tau \vec{v}\vec{v}(-\partial f_0/\partial E) d^3k = \frac{\overleftrightarrow{\sigma}}{e^2}, \tag{5.46}$$

and

$$\overleftrightarrow{\kappa}_1 = \frac{1}{4\pi^3} \int \tau \vec{v}\vec{v}(E - E_F)\Big(-\frac{\partial f_0}{\partial E}\Big) d^3k, \tag{5.47}$$

and $\overleftrightarrow{\kappa}_2$ is related to the thermal conductivity tensor $\overleftrightarrow{\kappa}_e$ by

$$\overleftrightarrow{\kappa}_2 = \frac{1}{4\pi^3} \int \tau \vec{v}\vec{v}(E - E_F)^2(-\partial f_0/\partial E) d^3k = T \overleftrightarrow{\kappa}_e \, . \tag{5.48}$$

Note that the integrands for $\overleftrightarrow{\kappa}_1$ and $\overleftrightarrow{\kappa}_2$ are both related to that for $\overleftrightarrow{\kappa}_0$ by introducing factors of $(E - E_F)$ and $(E - E_F)^2$, respectively. Note also that the same integral $\overleftrightarrow{\kappa}_1$ occurs in the expression for the electric current $\vec{j}$ induced by a thermal gradient $\vec{\nabla} T$ and in the expression for the thermal current $\vec{U}$ induced by an electric field $\vec{E}$. The motion of charged carriers across a temperature gradient results in a flow of electric current expressed by the term $-(e/T)(\overleftrightarrow{\kappa}_1) \cdot \vec{\nabla} T$. This term is the origin of thermoelectric effects.

The discussion up to this point has been general. If specific boundary conditions are imposed, we obtain a variety of thermoelectric effects such as the Seebeck effect, the Peltier effect and the Thomson effect. We now define the conditions under which each of these thermoelectric effects occur.

We define the thermopower $\overleftrightarrow{\mathcal{S}}$ (Seebeck coefficient) and the Thomson coefficient $\mathcal{T}_b$ under conditions of zero current flow. Then referring to Eq. 5.44, we obtain under open circuit conditions

$$\vec{j} = 0 = e^2 \overleftrightarrow{\kappa}_0 \cdot(\vec{E} - \frac{1}{e}\vec{\nabla}E_F) - (e/T) \overleftrightarrow{\kappa}_1 \cdot\vec{\nabla}T \tag{5.49}$$

so that the Seebeck coefficient $\overleftrightarrow{\mathcal{S}}$ is defined by

$$\vec{E} - \frac{1}{e}\vec{\nabla}E_F = (1/eT)\overleftrightarrow{\kappa}_0^{-1} \cdot \overleftrightarrow{\kappa}_1 \cdot \vec{\nabla}T \equiv \overleftrightarrow{\mathcal{S}} \cdot\vec{\nabla}T, \tag{5.50}$$

and $\mathcal{S}$ is sometimes called the thermopower. Using the relation $\vec{\nabla}E_F = \frac{\partial E_F}{\partial T}\vec{\nabla}T$ we obtain the definition for the Thomson coefficient $\mathcal{T}_b$

$$\vec{E} = \left(\frac{1}{e}\frac{\partial E_F}{\partial T} + \overleftrightarrow{\mathcal{S}}\right)\vec{\nabla}T \equiv \overleftrightarrow{\mathcal{T}}_b \cdot\vec{\nabla}T \tag{5.51}$$

where

$$\overleftrightarrow{\mathcal{T}}_b = T\frac{\partial}{\partial T} \overleftrightarrow{\mathcal{S}} \,. \tag{5.52}$$

For many thermoelectric systems of interest, $\overleftrightarrow{\mathcal{S}}$ has a linear temperature dependence, and in this case it follows from Eq. 5.52 that $\overleftrightarrow{\mathcal{T}}_b$ and $\overleftrightarrow{\mathcal{S}}$ for such systems are almost equivalent for practical purposes. Therefore the Seebeck and Thomson coefficients are used almost interchangeably in the literature.

From Eq. 5.50 and neglecting the term in ∇E_F, as is usually done, we have

$$\overleftrightarrow{\mathcal{S}} = (1/eT) \overleftrightarrow{\kappa}_0^{-1} \cdot \overleftrightarrow{\kappa}_1 \tag{5.53}$$

which is simplified by assuming an isotropic medium, yielding the scalar quantities

$$\mathcal{S} = (1/eT)(\kappa_1/\kappa_0). \tag{5.54}$$

However in an anisotropic medium, the tensor components of $\overleftrightarrow{\mathcal{S}}$ are found from

$$\mathcal{S}_{ij} = (1/eT)(\kappa_0^{-1})_{i\alpha}(\kappa_1)_{\alpha j}, \tag{5.55}$$

where the Einstein summation convention is assumed. Figure 5.3 shows a schematic diagram for measuring the thermopower or Seebeck effect in an n-type semiconductor. At the hot junction the Fermi level is higher than at the cold junction. Electrons will move from the hot junction to the cold junction in an attempt to equalize the Fermi level, thereby creating an electric field which can be measured in terms of the open circuit voltage V shown in Fig. 5.3.

Another important thermoelectric coefficient is the Peltier coefficient $\overleftrightarrow{\Pi}$, defined as the proportionality between $\vec{U}$ and $\vec{j}$

$$\vec{U} \equiv \overleftrightarrow{\Pi} \cdot\vec{j} \tag{5.56}$$

in the absence of a thermal gradient. For $\vec{\nabla}T = 0$, Eqs. 5.44 and 5.45 become

$$\vec{j} = e^2 \overleftrightarrow{\kappa}_0 \cdot\left(\vec{E} - \frac{1}{e}\vec{\nabla}E_F\right) \tag{5.57}$$

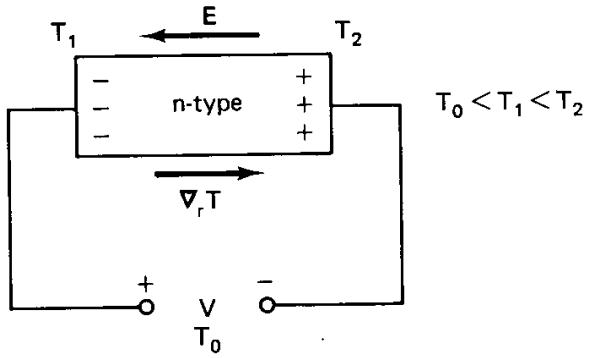


Figure 5.3: Determination of the Seebeck effect for an n-type semiconductor. In the presence of a temperature gradient, electrons will move from the hot junction to the cold junction, thereby creating an electric field and a voltage V across the semiconductor.

$$\vec{U} = e \overleftrightarrow{\kappa}_1 \cdot \left(\vec{E} - \frac{1}{e} \vec{\nabla} E_F \right) \tag{5.58}$$

so that

$$\vec{U} = \ (1/e) \overleftrightarrow{\kappa}_1 \cdot (\overleftrightarrow{\kappa}_0)^{-1} \cdot \vec{j} = \overleftrightarrow{\Pi} \cdot \vec{j} \tag{5.59}$$

where

$$\overleftrightarrow{\Pi} = (1/e) \overleftrightarrow{\kappa}_1 \cdot (\overleftrightarrow{\kappa}_0)^{-1}. \tag{5.60}$$

Comparing Eqs. 5.53 and 5.60 we see that $\overleftrightarrow{\Pi}$ and $\overleftrightarrow{\mathcal{S}}$ are related by

$$\overleftrightarrow{\Pi} = T \overleftrightarrow{\mathcal{S}}, \tag{5.61}$$

where T is the temperature. For isotropic materials the Peltier coefficient thus becomes a scalar, and is proportional to the thermopower $\mathcal{S}$:

$$\Pi = \frac{1}{e}(\kappa_1/\kappa_0) = T\mathcal{S}, \tag{5.62}$$

while for anisotropic materials the tensor components of $\overleftrightarrow{\Pi}$ can be found in analogy with Eq. 5.55. We note that both $\overleftrightarrow{\mathcal{S}}$ and $\overleftrightarrow{\Pi}$ exhibit a linear dependence on e and therefore depend explicitly on the **sign of the carrier**, and measurements of $\overleftrightarrow{\mathcal{S}}$ or $\overleftrightarrow{\Pi}$ can be used to determine whether transport is dominated by electrons or holes.

We have already considered the evaluation of $\overleftrightarrow{\kappa}_0$ in treating the electrical conductivity and $\overleftrightarrow{\kappa}_2$ in treating the thermal conductivity. To treat thermoelectric phenomena we need now to evaluate $\overleftrightarrow{\kappa}_1$

$$\overleftrightarrow{\kappa}_1 = \frac{1}{4\pi^3} \int \tau \vec{v} \vec{v} (E - E_F)(-\partial f_0/\partial E) d^3k. \tag{5.63}$$

In §5.3.1 we evaluate $\overleftrightarrow{\kappa}_1$ for the case of a metal and in §5.3.2 we evaluate $\overleftrightarrow{\kappa}_1$ for the case of the electrons in an intrinsic semiconductor. In practice, the thermopower is of interest for heavily doped semiconductors, which are either degenerate with the Fermi level in the conduction or valence band or very close to these band edges. A thermoelectric device has both n-type and p-type legs or constituents.

5.3.1 Thermoelectric Phenomena in Metals

All thermoelectric effects in metals depend on the tensor $\overleftrightarrow{\kappa}_1$ which we evaluate below for the case of a metal. We can then obtain the thermopower

$$\overleftrightarrow{\mathcal{S}} = \frac{1}{eT}(\overleftrightarrow{\kappa}_1 \cdot \overleftrightarrow{\kappa}_0^{-1}) \tag{5.64}$$

or the Peltier coefficient

$$\overleftrightarrow{\Pi} = \frac{1}{e}(\overleftrightarrow{\kappa}_1 \cdot \overleftrightarrow{\kappa}_0^{-1}) \tag{5.65}$$

or the Thomson coefficient

$$\overleftrightarrow{\mathcal{T}}_b = T\frac{\partial}{\partial T}\overleftrightarrow{\mathcal{S}}\,. \tag{5.66}$$

To evaluate $\overleftrightarrow{\kappa}_1$ for metals we wish to exploit the δ-function behavior of $(-\partial f_0/\partial E)$. This is accomplished by converting the integration d^3k to an integration over dE and over a constant energy surface, $d^3k = d^2S\ dE/\hbar v$. From Fermi statistics we have the general relation (see Eq. 5.20)

$$\int G(E)\Big(-\frac{\partial f_0}{\partial E}\Big)dE = G(E_F) + \frac{\pi^2}{6}(k_BT)^2\left[\frac{\partial^2 G}{\partial E^2}\right]_{E_F} + \cdots. \tag{5.67}$$

For the integral in Eq. 5.63 which defines $\overleftrightarrow{\kappa}_1$, we can write

$$G(E) = g(E)(E - E_F) \tag{5.68}$$

where

$$g(E) = \frac{1}{4\pi^3}\int \tau\vec{v}\vec{v}\ d^2S/v \tag{5.69}$$

and the integration in Eq. 5.69 is carried out over a constant energy surface at energy E. Differentiation of $G(E)$ then yields

$$\begin{aligned} G'(E) &= g'(E)(E - E_F) + g(E) \\ G''(E) &= g''(E)(E - E_F) + 2g'(E). \end{aligned} \tag{5.70}$$

Evaluation at $E = E_F$ yields

$$\begin{aligned} G(E_F) &= 0 \\ G''(E_F) &= 2g'(E_F). \end{aligned} \tag{5.71}$$

We therefore obtain

$$\overleftrightarrow{\kappa}_1 = \frac{\pi^2}{3}(k_BT)^2 g'(E_F). \tag{5.72}$$

We interpret $g'(E_F)$ in Eq. 5.72 to mean that the same integral $\overleftrightarrow{\kappa}_0$ which determines the conductivity tensor is evaluated on a constant energy surface E, and $g'(E_F)$ is the energy derivative of that integral evaluated at the Fermi energy E_F. The temperature dependence of $g'(E_F)$ is related to the temperature dependence of τ, since v is essentially temperature independent. For example, we will see in Chapter 6 that acoustic phonon scattering in the

high temperature limit $T \gg \Theta_D$ yields a temperature dependence $\tau \sim T^{-1}$ so that $\overleftrightarrow{\kappa}_1$ in this important case for metals will be proportional to T.

For a spherical constant energy surface $E = \hbar^2 k^2/2m^*$ and assuming a relaxation time τ that is independent of energy, we can readily evaluate Eq. 5.72 to obtain

$$g(E) = \frac{\tau}{3\pi^2 m^*}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E^{3/2} \tag{5.73}$$

$$g'(E_F) = \frac{\tau}{2\pi^2 m^*}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E_F^{1/2} \tag{5.74}$$

and

$$\kappa_1 = \frac{\tau}{6m^*}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E_F^{1/2}(k_B T)^2. \tag{5.75}$$

Using the same approximations, we can write for κ_0:

$$\kappa_0 = \frac{\tau}{3\pi^2 m^*}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E_F^{3/2} \tag{5.76}$$

so that from Eq. 5.64 we have for the Seebeck coefficient

$$\mathcal{S} = \frac{\kappa_1}{\kappa_0 eT} = \frac{\pi^2 k_B}{2e}\,\frac{k_B T}{E_F}. \tag{5.77}$$

From Eq. 5.77 we see that $\mathcal{S}$ exhibits a linear dependence on T and a sensitivity to the sign of the carriers. We note from Eq. 5.64 that a low carrier density implies a large $\mathcal{S}$ value. Thus degenerate (heavily doped with $n \sim 10^{18}$–$10^{19}/\text{cm}^3$) semiconductors tend to have higher thermopowers than metals. The derivation given here works as a good approximation for these very heavily doped semiconductors.

5.3.2 Thermopower for Intrinsic Semiconductors

In this section we evaluate $\overleftrightarrow{\kappa}_1$ for electrons in an intrinsic or lightly doped semiconductor for illustrative purposes. Intrinsic semiconductors are not important for practical thermoelectric devices since the contributions of electrons and holes to $\overleftrightarrow{\kappa}_1$ are of opposite signs and tend to cancel. Thus it is only heavily doped semiconductors with a single carrier type that are important for thermoelectric applications.

The evaluation of the general expression for the integral $\overleftrightarrow{\kappa}_1$

$$\overleftrightarrow{\kappa}_1 = \frac{1}{4\pi^3}\int \tau \vec{v}\vec{v}(E - E_F)\Big(-\frac{\partial f_0}{\partial E}\Big)d^3k \tag{5.78}$$

is different for semiconductors and metals. Referring to Fig. 5.4 for an intrinsic semiconductor we need to make the substitution $(E - E_F) \to E$ in Eq. 5.78, since only conduction electrons can carry heat. The equilibrium distribution function for an intrinsic semiconductor can be written as

$$f_0 = e^{-E/(k_B T)}\; e^{-|E_F^e|/(k_B T)} \tag{5.79}$$

so that

$$\frac{\partial f_0}{\partial E} = -\frac{1}{k_B T}\; e^{-E/(k_B T)}\; e^{-|E_F^e|/(k_B T)}. \tag{5.80}$$

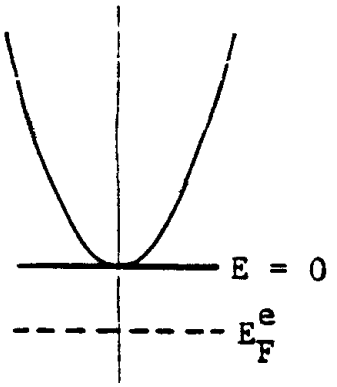


Figure 5.4: Schematic E vs k diagram, showing that $E = 0$ is the lowest electronic energy for heat conduction.

To evaluate d^3k we need to assume a model for $E(\vec{k})$. For simplicity, assume a simple parabolic band

$$E = \hbar^2 k^2 / 2m^* \tag{5.81}$$

$$d^3k = 4\pi k^2 dk \tag{5.82}$$

so that

$$d^3k = 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2} dE \tag{5.83}$$

and also

$$\vec{v} = \frac{1}{\hbar}(\partial E / \partial \vec{k}) = \hbar \vec{k} / m^*. \tag{5.84}$$

Substitution into the equation for $\overleftrightarrow{\kappa}_1$ for a semiconductor with the simple $E = \hbar^2 k^2 / 2m^*$ dispersion relation then yields upon integration

$$\kappa_{1xx} = \frac{5\tau k_B T}{m^*} (m^* k_B T / 2\pi\hbar^2)^{3/2} e^{-|E_F^e|/(k_B T)}. \tag{5.85}$$

This expression is valid for a semiconductor for which the Fermi level is far from the band edge $(E - E_F) \gg k_B T$. The thermopower is then found by substitution

$$\mathcal{S} = \frac{1}{eT} \left(\kappa_{1xx} / \kappa_{0xx} \right) \tag{5.86}$$

where the expression for $\kappa_{0xx} = \sigma_{xx}/e^2$ is given by Eq. 5.46. We thus obtain the result

$$\mathcal{S} = \frac{5}{2} \frac{k_B}{e} \tag{5.87}$$

which is a constant independent of temperature, independent of the band structure, but sensitive to the sign of the carriers. The calculation in this section was for the contribution of electrons. In an actual intrinsic semiconductor, the contribution of both electrons and holes to κ_1 must be found. Likewise the calculation for κ_{0xx} would also include contributions

from both electrons and holes. Since the contribution to $(1/e)\kappa_{1xx}$ for holes and electrons are of opposite sign, we can from Eq. 5.87 expect that $\mathcal{S}$ for holes will cancel $\mathcal{S}$ for electrons for an intrinsic semiconductor, while the κ_0 for holes and electrons will add.

Materials with a high thermopower or Seebeck coefficient are heavily doped degenerate semiconductors for which the Fermi level is close to the band edge and the complete Fermi function must be used. Since $\mathcal{S}$ depends on the sign of the charge carriers, thermoelectric materials are doped either heavily doped n-type or heavily doped p-type semiconductors to prevent cancellation of the contribution from electrons and holes, as occurs in intrinsic semiconductors which because of thermal excitations of carriers have equal concentrations of electrons and holes.

5.3.3 Effect of Thermoelectricity on the Thermal Conductivity

From the coupled equations given by Eqs. 5.44 and 5.45 it is seen that the proportionality between the thermal current $\vec{U}$ and the temperature gradient $\vec{\nabla}T$ in the absence of electrical current ($\vec{j}=0$) contains terms related to $\overleftrightarrow{\kappa}_1$. We now solve Eqs. 5.44 and 5.45 to find the contribution of the thermoelectric terms to the electronic thermal conductivity. When $\vec{j}=0$, Eq. 5.44 becomes

$$(\vec{E}-\frac{1}{e}\vec{\nabla}E_F)=\frac{1}{eT}\,\overleftrightarrow{\kappa^{-1}}_0\cdot\overleftrightarrow{\kappa}_1\cdot\vec{\nabla}T \tag{5.88}$$

so that

$$\vec{U}=-(1/T)\left[\overleftrightarrow{\kappa}_2-\overleftrightarrow{\kappa}_1\cdot\overleftrightarrow{\kappa^{-1}}_0\cdot\overleftrightarrow{\kappa}_1\right]\cdot\vec{\nabla}T \tag{5.89}$$

where $\overleftrightarrow{\kappa}_0$, $\overleftrightarrow{\kappa}_1$, and $\overleftrightarrow{\kappa}_2$ are given by Eqs. 5.46, 5.47, and 5.48, respectively, or

$$\overleftrightarrow{\kappa}_0=\frac{1}{4\pi^3\hbar}\int\tau\vec{v}\vec{v}\frac{d^2S_F}{v}, \tag{5.90}$$

$$\overleftrightarrow{\kappa}_1=\frac{\pi^2}{3}(k_BT)^2\left(\frac{\partial\,\overleftrightarrow{\kappa}_0}{\partial E}\right)_{E_F}, \tag{5.91}$$

and

$$\overleftrightarrow{\kappa}_2=\frac{(k_BT)^2}{12\pi\hbar}\int\tau\vec{v}\vec{v}\frac{d^2S_F}{v}. \tag{5.92}$$

We now evaluate the contribution to the thermal conductivity from the thermoelectric coupling effects for the case of a metal having a simple dispersion relation

$$E=\hbar^2k^2/2m^*. \tag{5.93}$$

In this case where τ is considered to be independent of E, Eqs. 5.91 and 5.90, respectively, provide expressions for κ_1 and κ_0 from which

$$\frac{1}{T}\kappa_1\kappa_0^{-1}\kappa_1=\frac{\pi^4n\tau}{4m^*}k_B^2T\left(\frac{k_BT}{E_F}\right)^2 \tag{5.94}$$

so that from Eqs. 5.28 and 5.94 the total electronic thermal conductivity for the metal becomes

$$\kappa_e=\frac{\pi^2n\tau}{3m^*}k_B^2T\left[1-\frac{3\pi^2}{4}\left(\frac{k_BT}{E_F}\right)^2\right]. \tag{5.95}$$

For typical metals $(T/T_F)\sim(1/30)$ at room temperature so that the thermoelectric correction term is less than 1%. For highly degenerate semiconductors as are of interest for thermoelectric applications, the complete Fermi function must be considered.

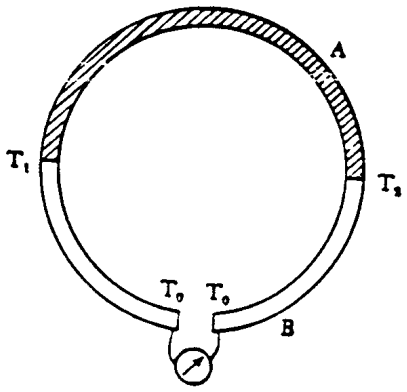


Figure 5.5: Thermopower between two different metals showing the principle of operation of a thermocouple under open circuit conditions (i.e., $j = 0$).

5.4 Thermoelectric Measurements

5.4.1 Seebeck Effect (Thermopower)

The thermopower $\mathcal{S}$ as defined in Eq. 5.50 and is the characteristic coefficient in the Seebeck effect, where a metal subjected to a thermal gradient $\vec{\nabla}T$ exhibits an electric field $\vec{E} = \mathcal{S}\vec{\nabla}T$. The measurements are made under an open-circuit voltage V and under conditions of no current flow.

In the application of the Seebeck effect to thermocouple operation, we usually measure the difference in thermopower $\mathcal{S}_A - \mathcal{S}_B$ between two different metals A and B by measuring the open circuit voltage V_{AB} as shown in Fig. 5.5. This voltage can be calculated from

$$\begin{aligned} V_{AB} &= -\oint \vec{E} \cdot d\vec{r} = -\oint \mathcal{S} \tfrac{\partial T}{\partial \vec{r}}\, d\vec{r} \\ &= \int_{T_0}^{T_1} \mathcal{S}_B dT + \int_{T_1}^{T_2} \mathcal{S}_A dT + \int_{T_2}^{T_0} \mathcal{S}_B dT \\ &= \int_{T_1}^{T_2} (\mathcal{S}_A - \mathcal{S}_B) dT. \end{aligned} \tag{5.96}$$

With $T_1 \neq T_2$, an open-circuit potential difference V_{AB} can be measured and Eq. 5.96 shows that V_{AB} is independent of the temperature T_0. Thus if T_1 is known and V_{AB} is measured, then temperature T_2 can be found from the calibration table of the thermocouple. From the simple expression of Eq. 5.77

$$\mathcal{S} = \frac{\pi^2 k_B}{2e} \frac{k_B T}{E_F} \tag{5.97}$$

a linear dependence of $\mathcal{S}$ on T is predicted for simple metals. For actual thermocouples used for temperature measurements, the $\mathcal{S}(T)$ dependence is approximately linear, but is given by an accurate calibration table to account for small deviations from this linear relation. Thermocouples are calibrated at several fixed temperatures and the calibration table comes from a fit of these thermal data to a polynomial function that is approximately linear in T.

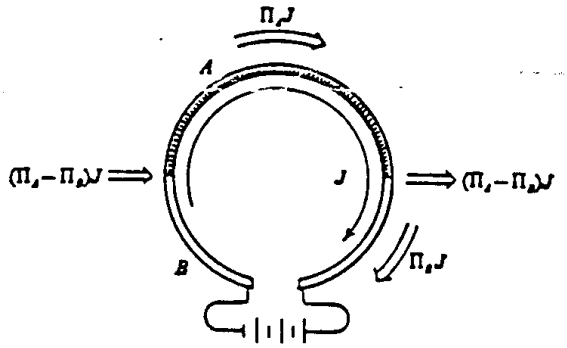


Figure 5.6: A heat engine based on the Peltier Effect with heat $(\Pi_A - \Pi_B)j$ introduced at one junction and extracted at another under the conditions of no temperature gradient ($\nabla T = 0$).

5.4.2 Peltier Effect

The Peltier effect is the observation of a thermal current $\vec{U} = \overleftrightarrow{\Pi} \cdot \vec{j}$ in the presence of an electric current $\vec{j}$ with no thermal gradient ($\vec{\nabla} T = 0$) so that

$$\overleftrightarrow{\Pi} = T \overleftrightarrow{\mathcal{S}} \, . \tag{5.98}$$

The Peltier effect measures the heat generated (or absorbed) at the junction of two dissimilar metals held at constant temperature, when an electric current passes through the junction. Sending electric current around a circuit of two dissimilar metals cools one junction and heats another and is the basis of the operation of thermoelectric coolers. This thermoelectric effect is represented schematically in Fig. 5.6. Because of the similarities between the Peltier coefficient and the Seebeck coefficient, materials exhibiting a large Seebeck effect also show a large Peltier effect. Since both $\overleftrightarrow{\mathcal{S}}$ and $\overleftrightarrow{\Pi}$ are proportional to $(1/e)$, the sign of $\overleftrightarrow{\mathcal{S}}$ and $\overleftrightarrow{\Pi}$ is negative for electrons and positive for holes in the case of degenerate semiconductors. Reversing the direction of $\vec{j}$, will interchange the junctions where heat is generated (absorbed).

5.4.3 Thomson Effect

Assume that we have an electric circuit consisting of a single metal conductor. The power generated in a sample, such as an n-type semiconductor, as shown in Fig. 5.7, is

$$P = \vec{j} \cdot \vec{E} \tag{5.99}$$

where the electric field can be obtained from Eqs. 5.49 and 5.51 as

$$\vec{E} = (\overleftrightarrow{\sigma^{-1}}) \cdot \vec{j} - \overleftrightarrow{\mathcal{T}}_b \cdot \vec{\nabla} T \tag{5.100}$$

where $\overleftrightarrow{\mathcal{T}}_b$ is the Thomson coefficient defined in Eq. 5.51 and is related to the Seebeck coefficient $\overleftrightarrow{\mathcal{S}}$ as discussed in §5.3 and §5.3.1. Substitution of Eq. 5.100 into Eq. 5.99 yields the

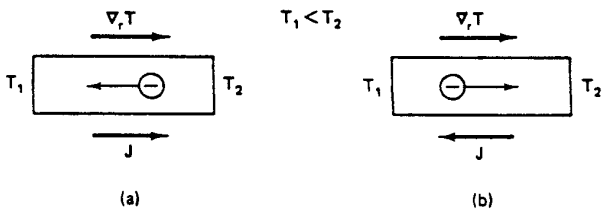


Figure 5.7: The Thomson term in an n-type semiconductor produces (a) heating when $\vec{j}$ and $\vec{\nabla}T$ are in the same direction and (b) cooling when $\vec{j}$ and $\vec{\nabla}T$ are in opposite directions.

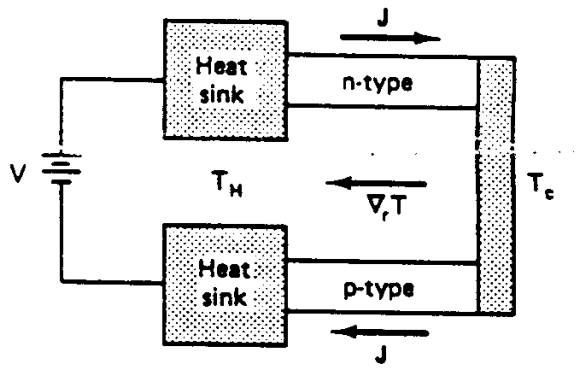


Figure 5.8: Schematic diagram of a thermoelectric cooler. The heat sinks and cold junctions are metals that form ohmic contacts to the active thermoelectric n-type and p-type semiconductors.

total power dissipation

$$P = \vec{j} \cdot (\overleftrightarrow{\sigma^{-1}}) \cdot \vec{j} - \vec{j} \cdot \overleftrightarrow{\mathcal{T}}_b \cdot \vec{\nabla}T. \tag{5.101}$$

The first term in Eq. 5.101 is the conventional joule heating term while the second term is the contribution from the Thomson effect. For an n-type semiconductor $\overleftrightarrow{\mathcal{T}}_b$ is negative. Thus when $\vec{j}$ and $\vec{\nabla}T$ are parallel, heating will result, as in Fig. 5.7(a). However if $\vec{j}$ and $\vec{\nabla}T$ are antiparallel, as in Fig. 5.7(b), cooling will occur. Thus reversal of the direction of $\vec{j}$ without changing the direction of $\vec{\nabla}T$ will reverse the sign of the Thomson contribution. Likewise, a reversal in the direction of $\vec{\nabla}T$ keeping the direction of $\vec{j}$ unchanged will also reverse the sign of the Thomson contribution.

Thus, if either (but not both) the direction of the electric current or the direction of the thermal gradient is reversed, an absorption of heat from the surroundings will take place. The Thomson effect is utilized in thermoelectric refrigerators which are useful as practical low temperature laboratory coolers. Referring to Fig. 5.8, we see a schematic diagram explaining the operation of a thermoelectric cooler. We see that for a degenerate n-type semiconductor where $\overleftrightarrow{\mathcal{T}}_b$, $\overleftrightarrow{\mathcal{S}}$, and $\overleftrightarrow{\Pi}$ are all negative, when $\vec{j}$ and $\vec{\nabla}T$ are antiparallel

then cooling occurs and heat is extracted from the cold junction and transferred to the heat sink at temperature T_H. For the p-type leg, all the thermoelectric coefficients are positive, so Eq. 5.101 shows that cooling occurs when $\vec{j}$ and $\vec{\nabla}T$ are parallel. Thus both the n-type and p-type legs in a thermoelectric element contribute to cooling in a thermoelectric cooler.

5.4.4 The Kelvin Relations

The three thermoelectric effects are related and these relations were first derived by Lord Kelvin after he became a Lord and changed his name from Thomson to Kelvin. The Kelvin relations are based on arguments of irreversible thermodynamics and relate Π, $\mathcal{S}$, and $\mathcal{T}_b$.

If we define the thermopower $\mathcal{S}_{AB} = \mathcal{S}_B - \mathcal{S}_A$ and the Peltier coefficient similarly $\Pi_{AB} = \Pi_A - \Pi_B$ for material A joined to material B, then we obtain the first Kelvin relation:

$$\mathcal{S}_{AB} = \frac{\Pi_{AB}}{T}. \tag{5.102}$$

The Thomson coefficient $\mathcal{T}_b$ is defined by

$$\mathcal{T}_b = T\frac{\partial \mathcal{S}}{\partial T} \tag{5.103}$$

which allows determination of the Seebeck coefficient at temperature T_0 by integration of the above equation

$$\mathcal{S}(T_0) = \int_0^{T_0} \left[\frac{\mathcal{T}_b(T)}{T}\right] dT. \tag{5.104}$$

Furthermore, from the above definitions, we deduce the second Kelvin relation

$$\mathcal{T}_{b,A} - \mathcal{T}_{b,B} = T\frac{\partial \mathcal{S}_{AB}}{\partial T} = T\frac{\partial \mathcal{S}_A}{\partial T} - T\frac{\partial \mathcal{S}_B}{\partial T} \tag{5.105}$$

from which we obtain an expression relating all three thermoelectric coefficients

$$\mathcal{T}_{b,A} = T\frac{\partial \mathcal{S}_A}{\partial T} = T\frac{\partial(\Pi_A/T)}{\partial T} = \frac{\partial \Pi_A}{\partial T} - \frac{\Pi_A}{T} = \frac{\partial \Pi_A}{\partial T} - \mathcal{S}_A. \tag{5.106}$$

5.4.5 Thermoelectric Figure of Merit

A good thermoelectric material for cooling applications must have a high thermoelectric figure of merit, Z, which is defined by

$$Z = \frac{\mathcal{S}^2\sigma}{\kappa} \tag{5.107}$$

where $\mathcal{S}$ is the thermoelectric power (Seebeck coefficient), σ is the electrical conductivity, and κ is the thermal conductivity. In order to achieve a high Z, one requires a high thermoelectric power $\mathcal{S}$, a high electrical conductivity σ to maintain high carrier mobility, and a low thermal conductivity κ to retain the applied thermal gradient. In general, it is difficult in practical systems to increase Z for the following reasons: increasing $\mathcal{S}$ for simple materials also leads to a simultaneous decrease in σ, and an increase in σ leads to a comparable increase in the electronic contribution to κ because of the Wiedemann–Franz law. So with known conventional solids, a limit is rapidly obtained where a modification to

any one of the three parameters $\mathcal{S}$, σ, or κ adversely affects the other transport coefficients, so that the resulting Z does not vary significantly. Currently, the commercially available materials with the highest Z are Bi_2Te_3 alloys such as $Bi_{0.5}Sb_{1.5}Te_3$ with $ZT \sim 1$ at 300K.

Only small increases in Z have been achieved in the last two to three decades. Research on thermoelectric materials has therefore been at a low level since about 1960. Since 1994, new interest has been revived in thermoelectricity with the discovery of new materials: skutterudites – $CeFe_{4-x}Co_xSb_{12}$ or $LaFe_{4-x}Co_xSb_{12}$ for $0 < x < 4$, which offer promise for higher Z values in bulk materials, and low dimensional systems (quantum wells, quantum wires) which offer promise for enhanced Z relative to bulk Z values in the same material. Thus thermoelectricity has again become an active research field.

5.5 Phonon Drag Effect

For a simple metal such as an alkali metal one would expect the thermopower $\mathcal{S}$ to be given by the simple expression in Eq. 5.77, and to be negative since the carriers are electrons. This is true at room temperature for all of the alkali metals except Li. Furthermore, $\mathcal{S}$ is positive for the noble metals Au, Ag and Cu. The anomalous sign of $\mathcal{S}$ in these metals can be understood by recalling the complex Fermi surfaces for these metals (see Fig. 2.6), where we note that copper in fact exhibits hole orbits in the extended zone. In general, with multiple carrier types as occur in semiconductors, the interpretation of thermopower data can become complicated.

Another complication which must also be considered, especially at low temperatures, is the *phonon drag effect*. In the presence of a thermal gradient, the phonons will diffuse and "drag" the electrons along with them because of the electron-phonon interaction. For a simple explanation of phonon drag, consider a gas of phonons with an average energy density $E_{\rm ph}/V$ where V is the volume. Using kinetic theory, we find that the phonon gas exerts a pressure

$$P = \frac{1}{3}\left(\frac{E_{\rm ph}}{V}\right) \tag{5.108}$$

on the electron gas. In the presence of a thermal gradient, the electrons are subject to a force density

$$F_x/V = -dP/dx = -\frac{1}{3V}\left(\frac{dE_{\rm ph}}{dT}\right)\frac{dT}{dx}. \tag{5.109}$$

To prevent the flow of current, this force must be balanced by the electric force. Thus, for an electron density n, we obtain

$$-neE_x + F_x/V = 0 \tag{5.110}$$

giving a phonon-drag contribution to the thermopower. Using the definition of the Seebeck coefficient for an open circuit system, we can write

$$\mathcal{S}_{\rm ph} = \frac{E_x}{(dT/dx)} \approx -\left(\frac{1}{3enV}\right)\frac{dE_{\rm ph}}{dT} = \frac{C_{\rm ph}}{3en} \tag{5.111}$$

where $C_{\rm ph}$ is the phonon heat capacity per unit volume. Although this is only a rough approximate derivation, it predicts the correct temperature dependence, in that the phonon-drag contribution is important at temperatures where the phonon specific heat is large.

The total thermopower is a sum of the diffusion contribution (considered in §5.4.1) and the phonon drag term $\mathcal{S}_{\mathrm{ph}}$.

The phonon drag effect depends on the electron-phonon coupling; at higher temperatures where the phonon-phonon coupling (Umklapp processes) becomes more important than the electron-phonon coupling, phonon drag effects become less important (see §6.4.4).

Chapter 6

Electron and Phonon Scattering

References:

- Kittel, *Introduction to Solid State Physics*, 6^{th} Ed., Wiley, 1986, Appendix C.
- Ashcroft and Mermin, *Solid State Physics*, Holt, Rinehart and Winston, 1976, Chapters 16 and 26.
- Hang, *Theoretical Solid State Physics*, Volume 2, Pergamon 1972, Chapter 4.

6.1 Electron Scattering

The transport properties of solids depend on the availability of carriers and on their scattering rates. In the previous chapters, we focused on the carriers and their generation. In this chapter, we focus on the scattering mechanisms.

Electron scattering brings an electronic system which has been subjected to external perturbations back to equilibrium. Collisions also alter the momentum of the carriers as the electrons are brought back into equilibrium. Electron collisions can occur through a variety of mechanisms such as electron-phonon, electron-impurity, electron-defect, and electron-electron scattering processes. Electron scattering is handles by the collision term in the Boltzmann equation.

In principle, the collision rates can be calculated from scattering theory. To do this, we introduce a transition probability $S(\vec{k}, \vec{k'})$ for scattering from a state $\vec{k}$ to a state $\vec{k'}$. Since electrons obey the Pauli principle, scattering will occur from an occupied to an unoccupied state. The process of scattering from $\vec{k}$ to $\vec{k'}$ decreases the distribution function $f(\vec{r}, \vec{k}, t)$ and depends on the probability that $\vec{k}$ is occupied and that $\vec{k'}$ is unoccupied. The process of scattering from $\vec{k'}$ to $\vec{k}$ increases the distribution function $f(\vec{r}, \vec{k}, t)$ and depends on the probability that state $\vec{k'}$ is occupied and state $\vec{k}$ is unoccupied. We will use the following notation for describing the scattering process:

- f_k is the probability that an electron occupies a state $\vec{k}$
- $[1 - f_k]$ is the probability that state $\vec{k}$ is unoccupied
- $S(\vec{k}, \vec{k'})$ is the probability per unit time that an electron in state $\vec{k}$ will be scattered to state $\vec{k'}$

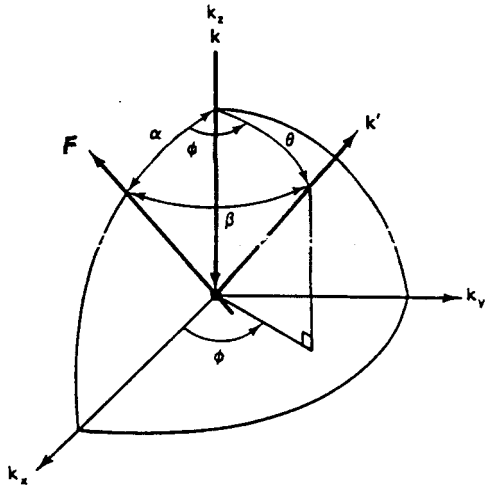


Figure 6.1: Spherical coordinate system in reciprocal space for an electron with wave vector $\vec{k}$ (along the k_z axis) scattering into a state with wavevector $\vec{k}'$ in an arbitrary force field $\vec{F}$. The scattering center is at the origin. For simplicity the event is rotated so that $\vec{F}$ has no k_y component.

- $S(\vec{k}', \vec{k})$ is the probability per unit time that an electron in state $\vec{k}'$ will be scattered back into state $\vec{k}$.

Using these definitions, the rate of change in the distribution function in the Boltzmann equation (see Eq. 4.4) due to collisions can be written as:

$$\left.\frac{\partial f(\vec{r}, \vec{k}, t)}{\partial t}\right|_{\text{collisions}} = \int d^3k' [f_{k'}(1 - f_k)S(\vec{k}', \vec{k}) - f_k(1 - f_{k'})S(\vec{k}, \vec{k}')] \tag{6.1}$$

where d^3k' is a volume element in $\vec{k}'$ space. The integration in Eq. 6.1 is over k space and the spherical coordinate system is shown in Fig. 6.1, together with the arbitrary force $\vec{F}$ responsible for the scattering event that introduces a perturbation described by

$$f_{\vec{k}} = f_{0\vec{k}} + \frac{\partial f_{0\vec{k}}}{\partial E}\frac{\hbar}{m^*}\vec{k} \cdot \vec{F} + \ldots \tag{6.2}$$

Using Fermi's Golden Rule for the transition probability per unit time between states $\vec{k}$ and $\vec{k}'$ we can write

$$S(\vec{k}, \vec{k}') \simeq \frac{2\pi}{\hbar}|\mathcal{H}_{\vec{k}\vec{k}'}|^2\{\delta[E(\vec{k})] - \delta[E(\vec{k}')]\} \tag{6.3}$$

where the matrix element of the Hamiltonian coupling states $\vec{k}$ and $\vec{k}'$ is

$$\mathcal{H}_{\vec{k}\vec{k}'} = \frac{1}{N}\int_V \psi^*_{\vec{k}}(\vec{r})\nabla V \psi_{\vec{k}'}(\vec{r}) d^3r \tag{6.4}$$

in which N is the number of unit cells in the sample and ∇V is the perturbation Hamiltonian responsible for the scattering event associated with the force $\vec{F}$.

At equilibrium $f_k = f_0(E)$ and the principle of detailed balance applies

$$S(\vec{k'},\vec{k}) f_0(E')[1 - f_0(E)] = S(\vec{k},\vec{k'}) f_0(E)[1 - f_0(E')] \tag{6.5}$$

so that the distribution function does not experience a net change via collisions when in the equilibrium state

$$(\partial f(\vec{r},\vec{k},t)/\partial t)|_{\text{collisions}} = 0. \tag{6.6}$$

We define collisions as elastic when $E(\vec{k'}) = E(\vec{k})$ and in this case $f_0(E') = f_0(E)$ so that $S(\vec{k'},\vec{k}) = S(\vec{k},\vec{k'})$. Collisions for which $E(\vec{k'}) \neq E(\vec{k})$ are termed inelastic. The term quasi-elastic is used to characterize collisions where the percentage change in energy is small. For our purposes here, we shall consider $S(\vec{k},\vec{k'})$ as a known function which can be calculated quantum mechanically by a detailed consideration of the scattering mechanisms which are important for a given practical case; this statement is true in principle, but in practice $S(\vec{k},\vec{k'})$ is usually specified in an approximate way.

The return to equilibrium depends on the frequency of collisions and the effectiveness of a scattering event in randomizing the motion of the electrons. Thus, small angle scattering is not as effective in restoring a system to equilibrium as large angle scattering. For this reason we distinguish between τ_D, the time for the system to be restored to equilibrium, and τ_c, the time between collisions. These times are related by

$$\tau_D = \frac{\tau_c}{1 - \cos\theta} \tag{6.7}$$

where θ is the mean change of angle of the electron velocity on collision (see Fig. 6.1). The time τ_D is the quantity which enters into Boltzmann's equation while $1/\tau_c$ determines the actual scattering rate.

The mean free time between collisions, τ_c, is related to several other quantities of interest: the mean free path ℓ_f, the scattering cross section σ_d, and the concentration of scattering centers N_c by

$$\tau_c = \frac{1}{N_c \sigma_d v} \tag{6.8}$$

where v is the drift velocity given by

$$v = \frac{\ell_f}{\tau_c} = \frac{1}{N_c \sigma_d \tau_c} \tag{6.9}$$

and v is in the direction of the electron transport. From Eq. 6.9, we see that $\ell_f = 1/N_c\sigma_d$. The drift velocity is of course very much smaller in magnitude than the instantaneous velocity of the electron at the Fermi level, which is typically of magnitude $v_F \sim 10^8$ cm/sec. Electron scattering centers include phonons, impurities, dislocations, the crystal surface, etc.

The most important electron scattering mechanism for both metals and semiconductors is electron-phonon scattering (scattering of electrons by the thermal motion of the lattice), though the scattering process for metals differs in detail from that in semiconductors. In the case of metals, much of the Brillouin zone is occupied by electrons, while in the case of semiconductors, most of the Brillouin zone is unoccupied and represents states into which electrons can be scattered. In the case of metals, electrons are scattered from one point on the Fermi surface to another point, and a large change in momentum occurs, corresponding

to a large change in $\vec{k}$. In the case of semiconductors, changes in wave vector from $\vec{k}$ to $-\vec{k}$ normally correspond to a very small change in wave vector, and thus changes from $\vec{k}$ to $-\vec{k}$ can be accomplished much more easily in the case of semiconductors. By the same token, small angle scattering (which is not so efficient for returning the system to equilibrium) is especially important for semiconductors where the change in wavevector is small. Since the scattering processes in semiconductors and metals are quite different, they will be discussed separately below.

Scattering probabilities for more than one scattering process are taken to be additive and therefore so are the reciprocal scattering time or scattering rates:

$$(\tau^{-1})_{\text{total}} = \sum_i \tau_i^{-1} \tag{6.10}$$

since $1/\tau$ is proportional to the scattering probability. Metals have large Fermi wavevectors k_F, and therefore large momentum transfers Δk can occur as a result of electronic collisions. In contrast, for semiconductors, k_F is small and so also is Δk on collision.

6.2 Scattering Processes in Semiconductors

6.2.1 Electron-Phonon Scattering

Electron-phonon scattering is the dominant scattering mechanism in crystalline semiconductors except at very low temperatures. Conservation of energy in the scattering process, which creates or absorbs a phonon of energy $\hbar\omega(\vec{q})$, is written as:

$$E_i - E_f = \pm\hbar\omega(\vec{q}) = \frac{\hbar^2}{2m^*}(k_i^2 - k_f^2), \tag{6.11}$$

where E_i is the initial energy, E_f is the final energy, k_i the initial wavevector, and k_f the final wavevector. Here, the "+" sign corresponds to the creation of phonons (the phonon emission process), while the "–" sign corresponds to the absorption of phonons. Conservation of momentum in the scattering by a phonon of wavevector $\vec{q}$ yields

$$\vec{k_i} - \vec{k_f} = \pm\vec{q}. \tag{6.12}$$

For semiconductors, the electrons involved in the scattering event generally remain in the vicinity of a single band extremum and involve only a small change in $\vec{k}$ and hence only low phonon $\vec{q}$ vectors participate. The probability that an electron makes a transition from an initial state i to a final state f is proportional to:

(a) the availability of final states for electrons,

(b) the probability of absorbing or emitting a phonon,

(c) the strength of the electron-phonon coupling or interaction.

The first factor, the availability of final states, is proportional to the density of final electron states $\rho(E_f)$ times the probability that the final state is unoccupied. (This occupation probability for a semiconductor is assumed to be unity since the conduction band is essentially empty.) For a simple parabolic band $\rho(E_f)$ is (from Eq. 4.64):

$$\rho(E_f) = \frac{(2m^*)^{3/2}E_f^{1/2}}{2\pi^2\hbar^3} = (2m^*)^{3/2}\frac{[E_i \pm \hbar\omega(\vec{q})]^{1/2}}{2\pi^2\hbar^3}, \tag{6.13}$$

where Eq. 6.11 has been employed and the "+" sign corresponds to absorption of a phonon and the "–" sign corresponds to phonon emission.

The probability of absorbing or emitting a phonon is proportional to the electron-phonon coupling $G(\vec{q})$ and to the phonon density $n(\vec{q})$ for absorption, and $[1 + n(\vec{q})]$ for emission, where $n(\vec{q})$ is given by the Bose-Einstein factor

$$n(\vec{q}) = \frac{1}{e^{\frac{\hbar\omega(\vec{q})}{k_B T}} - 1}. \tag{6.14}$$

Combining the terms in Eqs. 6.13 and 6.14 gives a scattering probability (or $1/\tau_c$) proportional to a sum over final states

$$\frac{1}{\tau_c} \sim \frac{(2m^*)^{3/2}}{2\pi^2\hbar^3} \sum_{\vec{q}} G(\vec{q}) \left[\frac{[E_i + \hbar\omega(\vec{q})]^{1/2}}{e^{\frac{\hbar\omega(\vec{q})}{k_B T}} - 1} + \frac{[E_i - \hbar\omega(\vec{q})]^{1/2}}{1 - e^{\frac{-\hbar\omega(\vec{q})}{k_B T}}} \right] \tag{6.15}$$

where the first term in the big bracket of Eq. 6.15 corresponds to phonon absorption and the second term to phonon emission. If $E_i < \hbar\omega(\vec{q})$, only the phonon absorption process is energetically allowed.

The electron-phonon coupling coefficient $G(\vec{q})$ in Eq. 6.15 depends on the electron-phonon coupling mechanism. There are three important coupling mechanisms in semiconductors which we briefly describe below: electromagnetic coupling, piezoelectric coupling, and deformation-potential coupling.

Electromagnetic coupling is important only for semiconductors where the charge distribution has different signs on neighboring ion sites. In this case, the oscillatory electric field can give rise to oscillating dipole moments associated with the motion of neighboring ion sites in the optical modes (see Fig. 6.2). The electromagnetic coupling mechanism is important in coupling electrons to optical phonon modes in III-V and II-VI compound semiconductors, but does not contribute in the case of silicon. To describe the optical modes we can use the Einstein approximation, since $\omega(\vec{q})$ is only weakly dependent on $\vec{q}$ for the optical modes of frequency ω_0. In this case $\hbar\omega_0 \gg k_B T$ and $\hbar\omega_0 \gg E$ where E is the electron energy, so that from Eq. 6.15 the collision rate is proportional to

$$\frac{1}{\tau_c} \sim \frac{m^{*3/2}(\hbar\omega_0)^{1/2}}{e^{\hbar\omega_0/k_B T} - 1}. \tag{6.16}$$

Thus, the collision rate depends on the temperature T, the optical phonon frequency ω_0 and the electron effective mass m^*. The corresponding mobility for the optical phonon scattering is

$$\mu = \frac{e\langle\tau\rangle}{m^*} \sim \frac{e(e^{\hbar\omega_0/k_B T} - 1)}{m^{*5/2}(\hbar\omega_0)^{1/2}} \tag{6.17}$$

Thus for optical phonon scattering, the mobility μ is independent of the electron energy E and decreases with increasing temperature.

As in the case of electromagnetic coupling, piezoelectric coupling is important in semiconductors which are ionic or partly ionic. If these crystals lack inversion symmetry, then *acoustic* mode vibrations generate regions of compression and rarefaction in a crystal which lead to electric fields (see Fig. 6.3). The piezoelectric scattering mechanism is thus associated with the coupling between electrons and phonons arising from these electromagnetic fields.

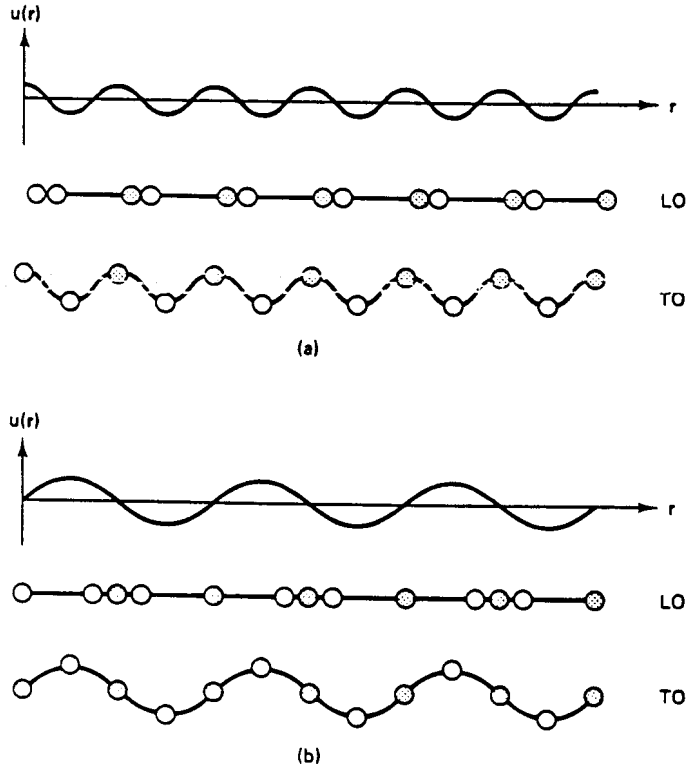


Figure 6.2: Displacements $\vec{u}(\vec{r})$ of a diatomic chain of atoms for longitudinal optical (LO) and transverse optical (TO) phonons at (a) the center and (b) the edge of the Brillouin zone. The lighter mass atoms are indicated by open circles. For zone edge optical phonons, only the lighter atoms are displaced.

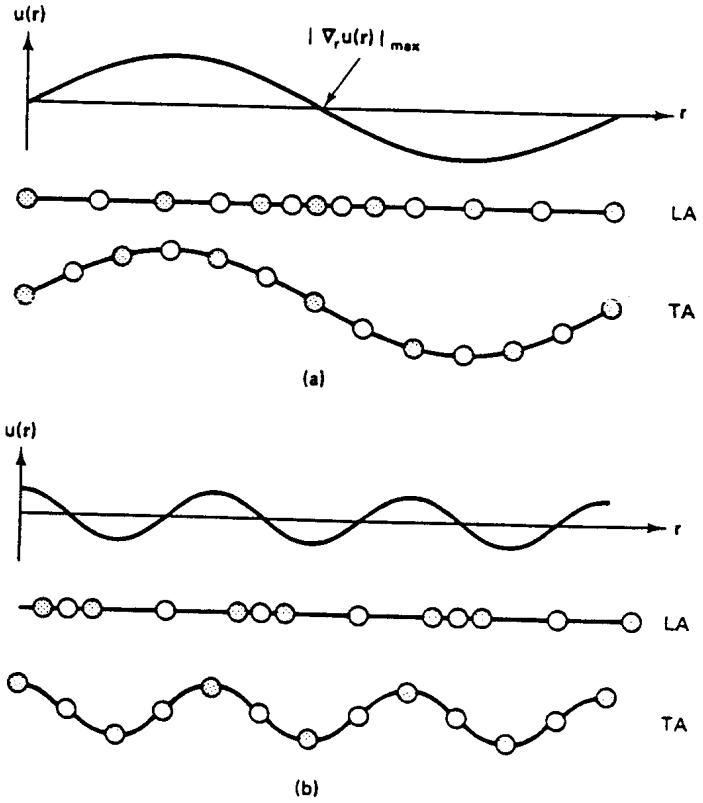


Figure 6.3: Displacements $\vec{u}(\vec{r})$ of a diatomic chain of atoms for longitudinal acoustic (LA) and transverse acoustic (TA) phonons at (a) the center and (b) the edge of the Brillouin zone. The lighter mass atoms are indicated by open circles. For zone edge acoustic phonons, only the heavier atoms are displaced.

The zincblende structure of the III–V compounds (e.g., GaAs) lacks inversion symmetry. In this case the perturbation potential is given by

$$\Delta V(\vec{r},t) = \frac{-ie\varepsilon_{pz}}{\epsilon_0 q}\vec{\nabla}\cdot\vec{u}(\vec{r},t) \tag{6.18}$$

where ε_{pz} is the piezoelectric coefficient and $\vec{u}(\vec{r},t) = u\exp(i\vec{q}\cdot\vec{r}-\omega t)$ is the displacement during a normal mode oscillation. Note that the phase of $\Delta V(\vec{r},t)$ in piezoelectric coupling is shifted by $\pi/2$ relative to the case of electromagnetic coupling.

The deformation-potential coupling mechanism is associated with energy shifts of the energy band extrema caused by the compression and rarefaction of crystals during acoustic mode vibrations. The deformation potential scattering mechanism is important in crystals like silicon which have inversion symmetry (and hence no piezoelectric scattering coupling) and have the same species on each site (and hence no electromagnetic coupling). The longitudinal acoustic modes are important for phonon coupling in n-type Si and Ge where the conduction band minima occur away from $\vec{k}=0$.

For deformation potential coupling, it is the LA acoustical phonons that are most important, though contributions by LO optical phonons still make some contribution. For the acoustic phonons, we have the condition $\hbar\omega \ll k_BT$ and $\hbar\omega \ll E$, while for the optical phonons it is usually the case that $\hbar\omega \gg k_BT$ at room temperature. For the range of acoustic phonon modes of interest, $G(\vec{q}) \sim q$, where q is the phonon wave vector and $\omega \sim q$ for acoustic phonons. Furthermore for the LA phonon branch the phonon absorption process will depend on $n(q)$ in accordance with the Bose factor

$$\frac{1}{e^{\hbar\omega/k_BT}-1} \simeq \frac{1}{[1+\frac{\hbar\omega}{k_BT}+\cdots]-1} \sim \frac{k_BT}{\hbar\omega} \sim \frac{k_BT}{q}, \tag{6.19}$$

while for phonon emission

$$\frac{1}{1-e^{-\hbar\omega/k_BT}} \simeq \frac{1}{1-[1-\frac{\hbar\omega}{k_BT}+\cdots]} \sim \frac{k_BT}{\hbar\omega} \sim \frac{k_BT}{q}. \tag{6.20}$$

Therefore, in considering both phonon absorption and phonon emission, the factors

$$G(\vec{q})[e^{\hbar\omega/k_BT}-1]^{-1}$$

and

$$G(\vec{q})[1-e^{-\hbar\omega/k_BT}]^{-1}$$

are independent of q for the LA branch. Consequently for the acoustic phonon scattering process, the carrier mobility μ decreases with increasing T according to (see Eq. 6.15)

$$\mu = \frac{e\langle\tau\rangle}{m^*} \sim m^{*-5/2}E^{-1/2}(k_BT)^{-1}. \tag{6.21}$$

For the optical LO contribution, we have $G(\vec{q})$ independent of $\vec{q}$ but an $E^{1/2}$ factor is introduced by Eq. 6.15 for both phonon absorption and emission, leading to the same basic dependence as given by Eq. 6.21. Thus, we find that the temperature and energy dependence of the mobility μ is different for the various electron-phonon coupling mechanisms. These differences in the E and T dependences can thus be used to identify which scattering mechanism is dominant in specific semiconducting samples. Furthermore, when explicit account is taken of the energy dependence of τ, departures from the strict Drude model $\sigma = ne^2\tau/m^*$ can be expected.

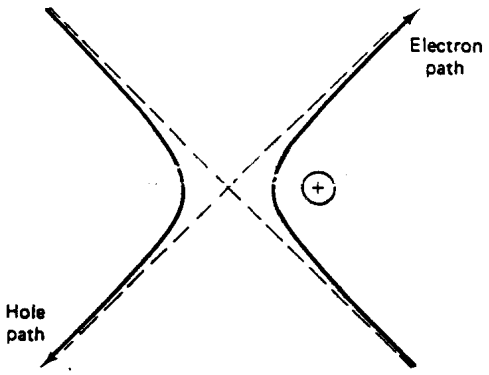


Figure 6.4: Trajectories of electrons and holes in ionized impurity scattering. The scattering center is at the origin.

6.2.2 Ionized Impurity Scattering

As the temperature is reduced, phonon scattering becomes less important so that in this regime, ionized impurity scattering and other defect scattering mechanisms can become dominant. Ionized impurity scattering can also be important in heavily doped semiconductors over a wider temperature range because of the larger defect density. This scattering mechanism involves the deflection of an electron with velocity v by the Coulomb field of an ion with charge Ze, as modified by the dielectric constant ε of the medium and by the screening of the impurity ion by free electrons (see Fig. 6.4). Most electrons are scattered through small angles as they are scattered by ionized impurities. The perturbation potential is given by

$$\Delta V(\vec{r}) = \frac{\pm Ze^2}{4\pi\epsilon_0 r} \tag{6.22}$$

and the $\pm$ signs denote the different scattering trajectories for electrons and holes (see Fig. 6.4). In Eq. 6.22 the screening of the electron by the semiconductor environment is handled by the static dielectric constant of the semiconductor ϵ_0. Because of the long-range nature of the Coulomb interaction, screening by other free carriers and by other ionized impurities could be important. Such screening effects are further discussed in §6.2.4.

The scattering rate $1/\tau_I$ due to ionized impurity scattering is given to a good approximation by the Conwell–Weisskopf formula

$$\frac{1}{\tau_I} \sim \frac{Z^2 N_I}{m^{*1/2} E^{3/2}} \; \ell n \left\{ 1 + \left[\frac{4\pi\varepsilon E}{Ze^2 N_I^{1/3}} \right]^2 \right\} \tag{6.23}$$

in which N_I is the ionized charged impurity density. The Conwell–Weisskopf formula works quite well for heavily doped semiconductors. We note here that $\tau_I \sim E^{3/2}$, so that it is the low energy electrons that are most affected by ionized impurity scattering (see Fig. 4.10).

Neutral impurities also introduce a scattering potential, but it is much weaker than that for the ionized impurity. Free carriers can polarize a neutral impurity and interact with the

resulting dipole moment, or can undergo an exchange interaction. In the case of neutral impurity scattering, the perturbation potential is given by

$$\Delta V(\vec{r}) \simeq \frac{\hbar^2}{m^*}\left(\frac{r_B}{r^5}\right)^{1/2} \tag{6.24}$$

where r_B is the ground state Bohr radius of the electron in a doped semiconductor and r is the distance of the electron to the impurity scattering center.

6.2.3 Other Scattering Mechanisms

Other scattering mechanisms in semiconductors include:

(a) neutral impurity centers — these make contributions at very low temperatures, and are mentioned in §6.2.2.

(b) dislocations — these defects give rise to anisotropic scattering at low temperatures.

(c) boundary scattering by crystal surfaces — this scattering becomes increasingly important the smaller the crystal size. Boundary scattering can become a dominant scattering mechanism in nanostructures (e.g., quantum wells, quantum wires and quantum dots), when the sample size in the confinement direction is smaller that the bulk mean free path.

(d) intervalley scattering from one equivalent conduction band minimum to another. This scattering process requires a phonon with large q and consequently results in a relatively large energy transfer.

(e) electron-electron scattering – similar to charged impurity scattering in being dominated by a Coulomb scattering mechanism, except that spin effects become important. This mechanism can be important in distributing energy and momentum among the electrons in the solid and thus can act in conjunction with other scattering mechanisms in establishing equilibrium.

(f) electron-hole scattering — depends on having both electrons and holes present. Because the electron and hole motions induced by an applied electric field are in opposite directions, electron-hole scattering tends to reverse the direction of the incident electrons and holes. Radiative recombination, i.e., electron-hole recombination with the emission of a photon, must also be considered.

6.2.4 Screening Effects in Semiconductors

In the vicinity of a charged impurity or an acoustic phonon, charge carriers are accumulated or depleted by the scattering potential, giving rise to a charge density

$$\rho(\vec{r}) = e[n(\vec{r}) - p(\vec{r}) + N_a^-(\vec{r}) - N_d^+(\vec{r})] = en^*(\vec{r}) \tag{6.25}$$

where $n(\vec{r})$, $p(\vec{r})$, $N_a^-(\vec{r})$, $N_d^+(\vec{r})$, and $n^*(\vec{r})$ are, respectively, the electron, hole, ionized acceptor, ionized donor, and effective total carrier concentrations as a function of distance r

to the scatterer. We can then write expressions for these quantities in terms of their excess charge above the uniform potential in the absence of the charge perturbation

$$\begin{aligned} n(\vec{r}) &= n + \delta n(\vec{r}) \\ N_d^+(\vec{r}) &= N_d^+ + \delta N_d^+(\vec{r}), \end{aligned} \tag{6.26}$$

and similarly for the holes and acceptors. The space charge $\rho(\vec{r})$ is related to the perturbing potential by Poisson's equation

$$\nabla^2 \phi(\vec{r}) = -\frac{\rho(\vec{r})}{\epsilon_0}. \tag{6.27}$$

Approximate relations for the excess concentrations are

$$\begin{aligned} \delta n(\vec{r})/n &\simeq -e\phi(\vec{r})/(k_B T) \\ \delta N_d^+(\vec{r})/N_d^+ &\simeq e\phi(\vec{r})/(k_B T) \end{aligned} \tag{6.28}$$

and similar relations for the holes. Substitution of Eq. 6.25 into Eqs. 6.26 and 6.28 yield

$$\nabla^2 \phi(\vec{r}) = -\frac{n^* e^2}{\epsilon_0 k_B T} \phi(\vec{r}). \tag{6.29}$$

We define an effective Debye screening length λ such that

$$\lambda^2 = \frac{\epsilon_0 k_B T}{n^* e^2}. \tag{6.30}$$

For a spherically symmetric potential Eq. 6.29 becomes

$$\frac{d^2}{dr^2}\Big(r\phi(r)\Big) = \frac{r\phi(r)}{\lambda^2} \tag{6.31}$$

which yields a solution

$$\phi(r) = \frac{Ze^2}{4\pi\epsilon_0 r} e^{-r/\lambda}. \tag{6.32}$$

Thus, the screening effect produces an exponential decay of the scattering potential $\phi(r)$ with a characteristic length λ that depends through Eq. 6.30 on the effective electron concentration. When the concentration gets large, λ decreases and screening becomes more effective.

When applying screening effects to the ionized impurity scattering problem, we Fourier expand the scattering potential to take advantage of the overall periodicity of the lattice

$$\Delta V(\vec{r}) = \sum_G A_G \exp(i\vec{G} \cdot \vec{r}) \tag{6.33}$$

where the Fourier coefficients are given by

$$A_G = \frac{1}{V} \int_V \nabla V(\vec{r}) \exp(-i\vec{G} \cdot \vec{r}) d^3 r \tag{6.34}$$

and the matrix element of the perturbation Hamiltonian in Eq. 6.4 becomes

$$\mathcal{H}_{\vec{k},\vec{k}'} = \frac{1}{N}\sum_G \int_V e^{-i\vec{k}\cdot\vec{r}} u_k^*(r) A_G e^{-i\vec{G}\cdot\vec{r}} e^{i\vec{k}'\cdot\vec{r}} u_{k'}(r) d^3r. \tag{6.35}$$

We note that the integral in Eq. 6.35 vanishes unless $\vec{k} - \vec{k}' = \vec{G}$ so that

$$\mathcal{H}_{\vec{k},\vec{k}'} = \frac{A_G}{N}\int_V u_k^*(r) u_{k'}(r) d^3r \tag{6.36}$$

within the first Brillouin zone so that for parabolic bands $u_k(\vec{r}) = u_{k'}(\vec{r})$ and

$$\mathcal{H}_{\vec{k},\vec{k}'} = A_{\vec{k}-\vec{k}'}. \tag{6.37}$$

Now substituting for the scattering potential in Eq. 6.34 we obtain

$$A_G = \frac{Ze^2}{4\pi\epsilon_0 V}\int_V \exp(-i\vec{G}\cdot\vec{r}) d^3r \tag{6.38}$$

where $d^3r = r^2 \sin\theta d\theta d\phi dr$ so that, for $\phi(r)$ depending only on r, the angular integration gives 4π and the spatial integration gives

$$A_G = \frac{Ze^2}{\epsilon_0 V|\vec{G}|^2} \tag{6.39}$$

and

$$\mathcal{H}_{\vec{k},\vec{k}'} = \frac{Ze^2}{\epsilon_0 V|\vec{k}-\vec{k}'|^2}. \tag{6.40}$$

Equations 6.39 and 6.40 are for the scattering potential without screening. When screening is included in considering the ionized impurity scattering mechanism, the integration becomes

$$A_G = \frac{Ze^2}{4\pi\epsilon_0 V}\int_V e^{-r/\lambda} e^{-i\vec{G}\cdot\vec{r}} d^3r = \frac{Ze^2}{\epsilon_0 V[|\vec{G}|^2 + |1/\lambda|^2]} \tag{6.41}$$

and

$$\mathcal{H}_{\vec{k},\vec{k}'} = \frac{Ze^2}{\epsilon_0 V[|\vec{k}-\vec{k}'|^2 + |1/\lambda|^2]} \tag{6.42}$$

so that screening clearly reduces the scattering due to ionized impurity scattering. The discussion given here also extends to the case of scattering in metals, which is treated below.

Combining the various scattering mechanisms discussed above for semiconductors, the picture given by Fig. 6.5 emerges. Here we see the temperature dependence of each of the important scattering mechansims and the effect of each of these processes on the carrier mobility. Here it is seen that screening effects are important for carrier mobilities at low temperature.

6.3 Electron Scattering in Metals

Basically the same scattering mechanisms are present in metals as in semiconductors, but because of the large number of occupied states in the conduction bands of metals, the temperature dependences of the various scattering mechanisms are quite different.

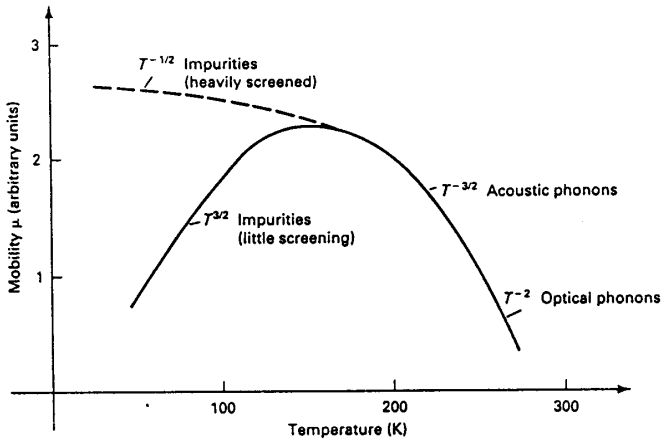


Figure 6.5: Typical temperature dependence of the carrier mobility in semiconductors, showing the effect of the dominant scattering mechanisms and the temperature dependence of each.

6.3.1 Electron-Phonon Scattering

For a review of phonons in the harmonic oscillator approximation see Appendix C.

In metals as in semiconductors, the dominant scattering mechanism is usually electron-phonon scattering. In the case of metals, electron scattering is mainly associated with an electromagnetic interaction of ions with nearby electrons, the longer range interactions being **screened** by the numerous mobile electrons. For metals, we must therefore consider explicitly the probability that a state $\vec{k}$ is occupied $f_0(\vec{k})$ or unoccupied $[1 - f_0(\vec{k})]$. The scattering rate is found by explicit consideration of the scattering rate into a state $\vec{k}$ and the scattering out of that state. Using the same arguments as in §6.2.1, the collision term in Boltzmann's equation is given by

$$\left.\frac{\partial f}{\partial t}\right|_{\text{collisions}} \sim \frac{1}{\tau} \simeq$$

$$\sum_q G(\vec{q}) \left\{ \overbrace{[1 - f_0(\vec{k})]\Big[\underbrace{f_0(\vec{k} - \vec{q})n(\vec{q})}_{\text{phonon absorption}} + \underbrace{f_0(\vec{k} + \vec{q})[1 + n(\vec{q})]}_{\text{phonon emission}} \Big]}^{\text{scattering into } \vec{k}} \right. \tag{6.43}$$

$$\left. - \overbrace{[f_0(\vec{k})]\Big[\underbrace{[1 - f_0(\vec{k} + \vec{q})]n(\vec{q})}_{\text{phonon absorption}} + \underbrace{[1 - f_0(\vec{k} - \vec{q})][1 + n(\vec{q})]}_{\text{phonon emission}} \Big]}^{\text{scattering out of } \vec{k}} \right\}$$

Here the first term in Eq. 6.43 is associated with scattering electrons into an element of phase space at $\vec{k}$ with a probability given by $[1 - f_0(\vec{k})]$ that state $\vec{k}$ is unoccupied and has contributions from both phonon absorption processes and phonon emission processes. The second term arises from electrons scattered out of state $\vec{k}$ and here, too, there are contributions from both phonon absorption processes and phonon emission processes. The equilibrium distribution function $f_0(\vec{k})$ for the electron is the Fermi distribution function

Symbol	Metal	Θ_D(K)
$\oplus$	Au	175
$\circ$	Na	202
$\triangle$	Cu	333
$\square$	Al	395
$\bullet$	Ni	472

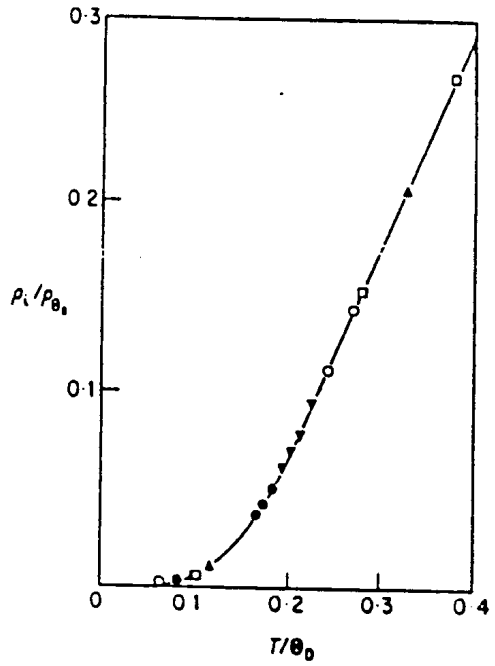


Figure 6.6: Universal curve of the temperature dependence of the ideal resistivity of various metals normalized to the value at the Debye temperature as a function of the dimensionless temperature T/Θ_D.

while the function $n(\vec{q})$ for the phonons is the Bose distribution function (Eq. 6.14). Phonon absorption depends on the phonon density $n(\vec{q})$, while phonon emission depends on the factor $\{1+n(\vec{q})\}$. These factors arise from the properties of the creation and annihilation operators for phonons (to be reviewed in recitation). The density of final states for metals is the density of states at the Fermi level which is consequently approximately independent of energy and temperature. The condition that, in metals, electron scattering takes place to states near the Fermi level implies that the largest phonon wave vector in an electron collision is $2k_F$ where k_F is the electron wave vector at the Fermi surface.

Of particular interest is the temperature dependence of the phonon scattering mechanism in the limit of low and high temperatures. Experimentally, the temperature dependence of the resistivity of metals can be plotted on a universal curve (see Fig. 6.6) in terms of ρ_T/ρ_{Θ_D} vs. T/Θ_D where Θ_D is the Debye temperature. This plot includes data for several

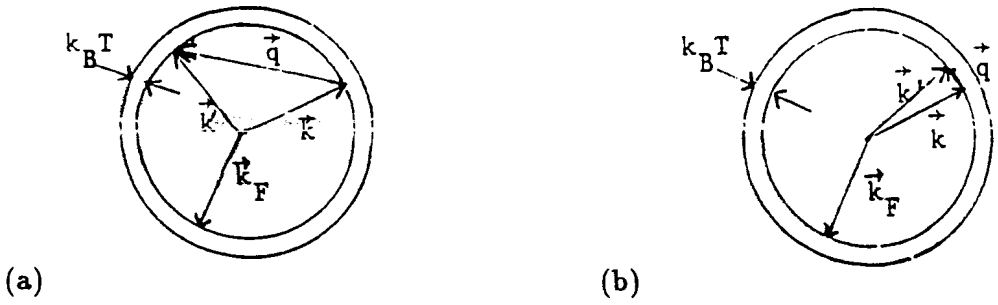


Figure 6.7: (a) Scattering of electrons on the Fermi surface of a metal. Large angle scattering dominates at high temperature ($T > \Theta_D$) and this regime is called the "quasi-elastic" limit. (b) Small angle scattering is important at low temperature ($T < \Theta_D$) and is in general an inelastic scattering process.

metals, and values for the Debye temperature of these metals are given with the figure.

In accordance with the plot in Fig. 6.6, $T \ll \Theta_D$ defines the low temperature limit and $T \gg \Theta_D$ the high temperature limit. Except for the very low temperature defect scattering limit, the electron-phonon scattering mechanism dominates, and the temperature dependence of the scattering rate depends on the product of the density of phonon states and the phonon occupation, since the electron-phonon coupling coefficient is essentially independent of T. The phonon concentration in the high temperature limit becomes

$$n(\vec{q}) = \frac{1}{\exp(\hbar\omega/k_B T) - 1} \approx \frac{k_B T}{\hbar\omega} \tag{6.44}$$

since $(\hbar\omega/k_B T) \ll 1$, so that from Eq. 6.44 we have $1/\tau \sim T$ and $\sigma = ne\mu \sim T^{-1}$. In this high temperature limit, the scattering is quasi-elastic and involves large-angle scattering, since phonon wave vectors up to the Debye wave vector q_D are involved in the electron scattering, where q_D is related to the Debye frequency ω_D and to the Debye temperature Θ_D according to

$$\hbar\omega_D = k_B\Theta_D = \hbar q_D v_q \tag{6.45}$$

where v_q is the velocity of sound.

We can interpret q_D as the radius of a Debye sphere in $\vec{k}$-space which defines the range of accessible $\vec{q}$ vectors for scattering, i.e., $0 < q < q_D$. The magnitude of q_D is comparable to the Brillouin zone dimensions but the energy change of an electron (ΔE) on scattering by a phonon will be less than $k_B\Theta_D \simeq 1/40eV$ so that the restriction of $(\Delta E)_{max} \simeq k_B\Theta_D$ implies that the maximum electronic energy change on scattering will be small compared with the Fermi energy E_F. We thus obtain that for $T > \Theta_D$ (the high temperature regime), $\Delta E < k_B T$ and the scattering will be quasi-elastic as illustrated in Fig. 6.7a. In the opposite limit, $T \ll \Theta_D$, we have $\hbar\omega_q \simeq k_B T$ (because only low frequency acoustic phonons are available for scattering) and in the low temperature limit there is the possibility that $\Delta E > k_B T$, which implies inelastic scattering. In the low temperature limit, $T \ll \Theta_D$, the scattering is also small-angle scattering, since only low energy (low q wave vector) phonons

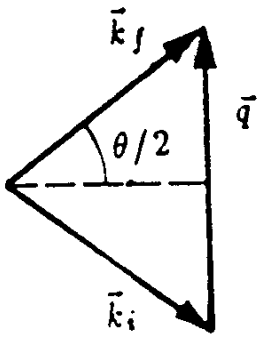


Figure 6.8: Geometry of the scattering process, where θ is the scattering angle between the incident and scattered electron wave vectors $\vec{k}_i$ and $\vec{k}_f$, respectively, and q is the phonon wave vector.

are available for scattering (as illustrated in Fig. 6.7b). At low temperature, the phonon density contributes a factor of T^3 to the scattering rate (Eq. 6.43) when the sum over phonon states is converted to an integral and q^2dq is written in terms of the dimensionless variable $\hbar\omega_q/k_BT$ with $\omega = v_q q$. Since small momentum transfer gives rise to small angle scattering, the diagram in Fig. 6.8 involves Eq. 6.7. Because of the small energy transfer we can write,

$$|\vec{k_i} - \vec{k_f}| \sim k_f(1 - \cos\theta) \approx \frac{1}{2}k_f\theta^2 \approx \frac{1}{2}k_f(q/k_f)^2 \tag{6.46}$$

so that another factor of q^2 appears in the integration over $\vec{q}$ when calculating $(1/\tau_D)$. Thus, the electron scattering rate at low temperature is predicted to be proportional to T^5 so that $\sigma \sim T^{-5}$ (Bloch–Grüneisen formula). Thus, when phonon scattering is the dominant scattering mechanism in metals, the following results are obtained:

$$\sigma \sim \Theta_D/T \qquad T \gg \Theta_D \tag{6.47}$$

$$\sigma \sim (\Theta_D/T)^5 \qquad T \ll \Theta_D \tag{6.48}$$

In practice, the resistivity of metals at very low temperatures is dominated by other scattering mechanisms, such as impurities, boundary scattering, etc., and at very low T electron-phonon scattering (see Eq. 6.48) is relatively unimportant.

The possibility of umklapp processes further increases the range of phonon modes that can contribute to electron scattering in electron-phonon scattering processes. In an umklapp process, a non-vanishing reciprocal lattice vector can be involved in the momentum conservation relation, as shown in the schematic diagram of Fig. 6.9.

In this diagram, the relation between the wave vectors for the phonon and for the incident and scattered electrons $\vec{G} = \vec{k} + \vec{q} + \vec{k'}$ is shown when crystal momentum is conserved for a non-vanishing reciprocal lattice vector $\vec{G}$. Thus, phonons involved in an umklapp process have large wave vectors with magnitudes of about 1/3 of the Brillouin zone dimensions. Therefore, substantial energies can be transferred on collision through an umklapp process. At low temperatures, normal scattering processes (i.e., normal as distinguished from umklapp processes) play an important part in completing the return to equilibrium of an

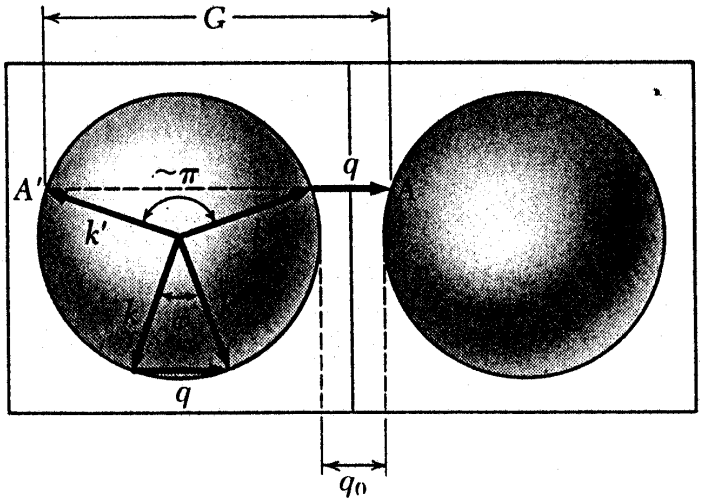


Figure 6.9: Schematic diagram showing the relation between the phonon wave vector $\vec{q}$ and the electron wave vectors $\vec{k}$ and $\vec{k}'$ in two Brillouin zones separated by the reciprocal lattice vector $\vec{G}$ (umklapp process).

excited electron in a metal, while at high temperatures, umklapp processes become more important.

The discussion presented up to this point is applicable to the creation or absorption of a single phonon in a particular scattering event. Since the restoring forces for lattice vibrations in solids are not strictly harmonic, anharmonic corrections to the restoring forces give rise to multiphonon processes where more than one phonon can be created or annihilated in a single scattering event. Experimental evidence for multiphonon processes is provided in both optical and transport studies. In some cases, more than one phonon at the same frequency can be created (harmonics), while in other cases, multiple phonons at different frequencies (overtones) are involved.

6.3.2 Other Scattering Mechanisms in Metals

At very low temperatures where phonon scattering is of less importance, other scattering mechanisms become important, and we can write

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i} \tag{6.49}$$

where the sum is over all the scattering processes.

(a) Charged impurity scattering — The effect of charged impurity scattering (Z being the difference in the charge on the impurity site as compared with the charge on a regular lattice site) is of less importance in metals than in semiconductors, because of the strong screening effects by the free electrons in metals.

(b) Neutral impurities — This process pertains to scattering centers having the same charge as the host. Such scattering has less effect on the transport properties than scattering by charged impurity sites, because of the much weaker scattering potential.

(c) Vacancies, interstitials, dislocations, size-dependent effects — the effects for these defects on the transport properties are similar to those for semiconductors. Boundary scattering can become very important in metal nanostructures when the sample length in some direction becomes less than the mean free path in the corresponding bulk crystal.

For most metals, phonon scattering is relatively unimportant at liquid helium temperatures, so that resistivity measurements at 4K provide a method for the detection of impurities and crystal defects. In fact, in characterizing the quality of a high purity metal sample, it is customary to specify the resistivity ratio $\rho(300\text{K})/\rho(4\text{K})$. This quantity is usually called the residual resistivity ratio (RRR), or the residual *resistance* ratio. In contrast, a typical semiconductor is characterized by its conductivity and Hall coefficient at room temperature and at 77 K.

6.4 Phonon Scattering

Whereas electron scattering is important in electronic transport properties, phonon scattering is important in thermal transport, particularly for the case of insulators where heat is carried mainly by phonons. The major scattering mechanisms for phonons are phonon-phonon scattering, phonon-boundary scattering, defect-phonon scattering, and phonon-electron scattering which are briefly discussed in the following subsections.

6.4.1 Phonon-phonon scattering

The dominant phonon scattering process in crystalline materials is usually phonon-phonon scattering. Phonons are scattered by other phonons because of anharmonic terms in the restoring potential. This scattering process permits:

- two phonons to combine to form a third phonon or
- one phonon to break up into two phonons.

In these anharmonic processes, energy and wavevector conservation apply:

$$\vec{q}_1 + \vec{q}_2 = \vec{q}_3 \qquad \text{normal processes} \tag{6.50}$$

or

$$\vec{q}_1 + \vec{q}_2 = \vec{q}_3 + \vec{Q} \qquad \text{umklapp processes} \tag{6.51}$$

where $\vec{Q}$ corresponds to a phonon wave vector of magnitude equal to a non-zero reciprocal lattice vector. Umklapp processes are important when q_1 or q_2 are large, i.e., comparable to a reciprocal lattice vector (see Fig. 6.10). vectors. When umklapp processes (see Fig. 6.10) are present, the scattered phonon wavevector $\vec{q}_3$ can be in a direction opposite to the energy flow, thereby giving rise to thermal resistance. Because of the high momentum transfer

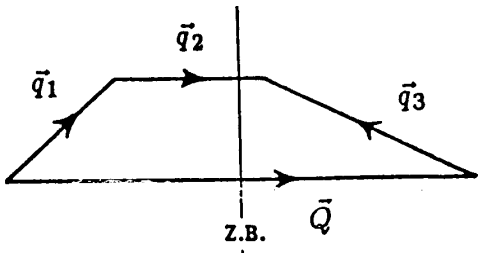


Figure 6.10: Phonon-phonon umklapp processes. Here $\vec{Q}$ is a non-zero reciprocal lattice vector, and $\vec{q}_1$ and $\vec{q}_2$ are the incident phonon wavevectors involved in the scattering process, and $\vec{q}_3$ is the wavevector of the scattered phonon.

and the large phonon energies that are involved, umklapp processes dominate the thermal conductivity at high T.

The phonon density is proportional to the Bose factor so that the scattering rate is proportional to

$$\frac{1}{\tau_{\rm ph}} \sim \frac{1}{(e^{\hbar\omega/(k_B T)} - 1)}. \tag{6.52}$$

At high temperatures $T \gg \Theta_D$, the scattering time thus varies as T^{-1} since

$$\tau_{\rm ph} \sim (e^{\hbar\omega/k_B T} - 1) \sim \hbar\omega/k_B T \tag{6.53}$$

while at low temperatures $T \sim \Theta_D$, an exponential temperature dependence for $\tau_{\rm ph}$ is found

$$\tau_{\rm ph} \sim e^{\hbar\omega/k_B T} - 1. \tag{6.54}$$

These temperature dependences are important in considering the lattice contribution to the thermal conductivity (see §5.2.4).

6.4.2 Phonon-Boundary Scattering

Phonon-boundary scattering is important at low temperatures where the phonon density is low. In this regime, the scattering time is independent of T. The thermal conductivity in this range is proportional to the phonon density which is in turn proportional to T^3. Phonon-boundary scattering is also very important for low dimensional systems where the sample size in some dimension is less than the corresponding phonon mean free path in the bulk 3D crystal. Phonon-boundary scattering combined with phonon-phonon scattering results in a thermal conductivity κ for insulators with the general shape shown in Fig. 6.11 (see §5.2.4). The lattice thermal conductivity follows the relation

$$\kappa_L = C_p v_q \Lambda_{\rm ph}/3 \tag{6.55}$$

where the phonon mean free path $\Lambda_{\rm ph}$ is related to the phonon scattering probability $(1/\tau_{\rm ph})$ by

$$\tau_{\rm ph} = \Lambda_{\rm ph}/v_q \tag{6.56}$$

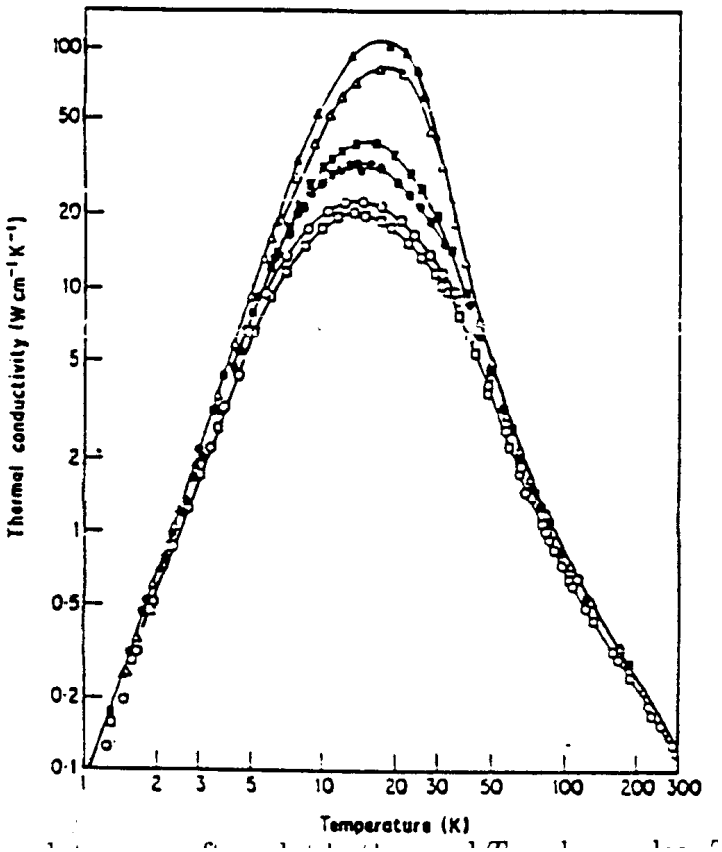


Figure 6.11: For insulators, we often plot both the thermal conductivity κ and the temperature T on log scales. The various curves are for LiF with different concentrations of Li isotopes ^{6}Li and ^{7}Li. For highly perfect crystals, it is possible to observe the scattering effects due to Li ions of different masses, which act as lattice defects.

in which v_q is the velocity of sound and C_p is the heat capacity at constant pressure. Phonon-boundary scattering becomes more important as the crystallite size decreases. The scattering conditions at the boundary can be specular (where after scattering only $q_\perp$ is reversed and $q_\parallel$ is unchanged) for a very smooth surface or diffuse (where after scattering the q is randomized) for a rough surface. Periodic corrugations on a surface can also give rise to interesting scattering effects for both electrons and phonons.

6.4.3 Defect-Phonon Scattering

Defect-phonon scattering includes a variety of crystal defects, charged and uncharged impurities and different isotopes of the host constituents. The thermal conductivity curves in Fig. 6.11 show the scattering effects due to different isotopes of Li. The low mass of Li makes it possible to see such effects clearly. Isotope effects are also important in graphite and diamond which have the highest thermal conductivity of any solid.

6.4.4 Electron-Phonon Scattering

If electrons scatter from phonons, the reverse process also occurs. When phonons impart momentum to electrons, the electron distribution is affected. Thus, the electrons will also carry energy as they are dragged also by the stream of phonons. This phenomenon is called phonon drag. In the case of phonon drag we must simultaneously solve the Boltzmann equations for the electron and phonon distributions which are coupled by the phonon drag term.

6.5 Temperature Dependence of the Electrical and Thermal Conductivity

For the electrical conductivity, at very low temperatures, impurity, defect, and boundary scattering dominate. In this regime σ is independent of temperature. At somewhat higher temperatures but still far below Θ_D the electrical conductivity for metals exhibits a strong temperature dependence (see Eq. 6.48)

$$\sigma \propto (\Theta_D/T)^5 \qquad T \ll \Theta_D. \tag{6.57}$$

At higher temperatures where $T \gg \Theta_D$, scattering by phonons with any q vector is possible and the formula

$$\sigma \sim (\Theta_D/T) \qquad T \gg \Theta_D \tag{6.58}$$

applies. We now summarize the corresponding temperature ranges for the thermal conductivity.

Although the thermal conductivity was formally discussed in §5.2, a meaningful discussion of the temperature dependence of κ depends on scattering processes. The total thermal conductivity κ in general depends on the lattice and electronic contributions, κ_L and κ_e, respectively. The temperature dependence of the lattice contribution is discussed in §5.2.4 with regard to the various phonon scattering processes and their temperature dependence. For the electronic contribution, we must consider the temperature dependence of the electron scattering processes discussed in §5.2 and §6.2.

At very low temperatures, in the impurity/defect/boundary scattering range, σ is independent of T and the same scattering processes apply for both the electronic thermal conductivity and the electrical conductivity so that $\kappa_e \propto T$ in the impurity scattering regime where $\sigma \sim$ constant and the Wiedemann–Franz law is applicable. From Fig. 5.1 we see that for copper, defect and boundary scattering are dominant below ~ 20 K, while phonon scattering becomes important at higher T.

At low temperatures $T \ll \Theta_D$, but with T in a regime where phonon scattering has already become the dominant scattering mechanism, the thermal transport depends on the electron-phonon collision rate which in turn is proportional to the phonon density. At low temperatures the phonon density is proportional to T^3. This follows from the proportionality of the phonon density of states arising from the integration of $\int q^2 dq$, and from the dispersion relation for the acoustic phonons $\omega = qv_q$

$$q = \omega/v_q = xkT/\hbar v_q \tag{6.59}$$

where $x = \hbar\omega/k_BT$. Thus in the low temperature range of phonon scattering where $T \ll \Theta_D$ and the Wiedemann–Franz law is no longer satisfied, the temperature dependence of τ is found from the product $T(T^{-3})$ so that $\kappa_e \propto T^{-2}$. One reason why the Wiedemann–Franz law is not satisfied in this temperature regime is that κ_e depends on the collision rate τ_c while σ depends on the time to reach thermal equilibrium, τ_D. At low temperatures where only low q phonons participate in scattering events the times τ_c and τ_D are not the same.

At high T where $T \gg \Theta_D$ and the Wiedemann–Franz law applies, κ_e approaches a constant value corresponding to the regime where σ is proportional to $1/T$. This occurs at temperatures much higher than those shown in Fig. 5.1. The decrease in κ above the peak value at ∼17 K follows a $1/T^2$ dependence quite well.

In addition to the electronic thermal conductivity, heat can be carried by the lattice vibrations or phonons. The phonon thermal conductivity mechanism is in fact the principal mechanism operative in semiconductors and insulators, since the electronic contribution in this case is negligibly small. Since κ_L contributes also to metals, the total measured thermal conductivity for metals should exceed the electronic contribution $(\pi^2 k_B^2 T\sigma)/(3e^2)$. In good metallic conductors of high purity, the electronic thermal conductivity dominates and the phonon contribution tends to be small. On the other hand, in conductors where the thermal conductivity due to phonons makes a significant contribution to the total thermal conductivity, it is necessary to separate the electronic and lattice contributions before applying the Wiedemann–Franz law to the total κ.

With regard to the lattice contribution, κ_L at very low temperatures is dominated by defect and boundary scattering processes. From the relation

$$\kappa_L = \frac{1}{3} C_p v_q \Lambda_{\rm ph} \tag{6.60}$$

we can determine the temperature dependence of κ_L, since $C_p \sim T^3$ at low T while the sound velocity v_q and phonon mean free path $\Lambda_{\rm ph}$ at very low T are independent of T. In this regime the number of scatterers is independent of T.

In the regime where only low q phonons contribute to transport and to scattering, only normal scattering processes contribute. In this regime C_p is still increasing as T^3, v_q is independent of T, but $1/\Lambda_{\rm ph}$ increases in proportion to the phonon density of states. With increasing T, the temperature dependence of C_p becomes less pronounced and that for $\Lambda_{\rm ph}$ becomes more pronounced as more scatters participate, leading eventually to a decrease in κ_L. We note that it is only the inelastic collisions that contribute to the decrease in $\Lambda_{\rm ph}$, since elastic phonon-phonon scattering has a similar effect as impurity scattering for phonons. The inelastic collisions are of course due to anharmonic forces.

Eventually phonons with wavevectors large enough to support umklapp processes are thermally activated. Umklapp processes give rise to thermal resistance and in this regime κ_L decreases as $\exp(-\Theta_D/T)$. In the high temperature limit $T \gg \Theta_D$, the heat capacity and phonon velocity are both independent of T. The $\kappa_L \sim 1/T$ dependence arises from the $1/T$ dependence of the mean free path, since in this limit the scattering rate becomes proportional to k_BT.