Chapter 24

Fermi's Golden Rule

24.1 Introduction

In this chapter, we derive a very useful result for estimating transition rates between quantum states due to time-dependent perturbation. The results will be used heavily in subsequent chapters to understand the optical and electronic transport properties of semiconductors.

24.2 Fermi's Golden Rule

Consider an unperturbed quantum system in state $|\Psi_{t_0}\rangle$ at time $t = t_0$. It evolves to the state $|\Psi_t\rangle$ at a future instant t. The time evolution of the state vector is governed by the unperturbed Hamiltonian H_0 according to the time-dependent Schrodinger equation

$$i\hbar\frac{\partial}{\partial t}|\Psi_t\rangle = H_0|\Psi_t\rangle. \tag{24.1}$$

If the system was in an eigenstate $|\Psi_{t_0}\rangle = |0\rangle$ of energy E_0 at time t_0, then the state at a future time differs from the initial state by a phase factor

$$H_0|\Psi_{t_0}\rangle = E_0|\Psi_{t_0}\rangle \implies |\Psi_t\rangle = e^{-i\frac{E_0}{\hbar}(t-t_0)}|\Psi_{t_0}\rangle. \tag{24.2}$$

This is a *stationary* state; if the quantum state started in an eigentstate, it remains in that eigenstate as long as there is no perturbation. But the eigen-state vector still 'rotates' in time with frequency $\omega_0 = E_0/\hbar$ in the Hilbert space as indicated schematically

in Figure 24.1. It is called stationary because physical observables of the eigenstate will require not the amplitude, but the inner product, which is $\langle\Psi_t|\Psi_t\rangle = \langle\Psi_{t_0}|\Psi_{t_0}\rangle$. This is manifestly stationary in time.

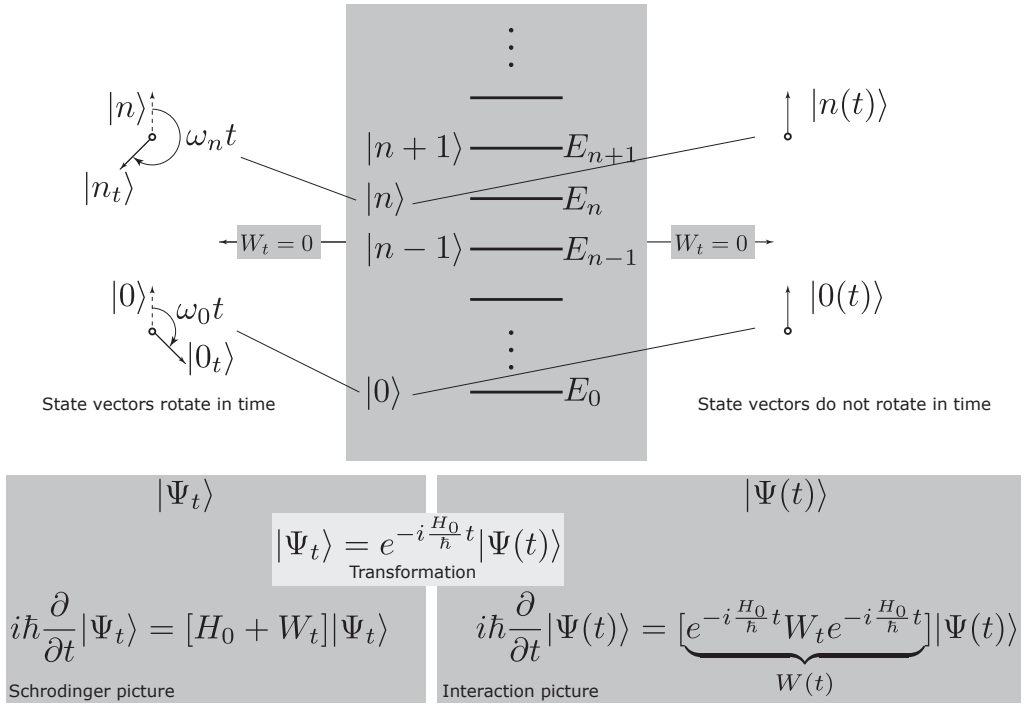


FIGURE 24.1: Schrodinger vs. Interaction pictures of time-evolution of quantum state.

Now let us perturb the system with a *time-dependent* term W_t. This perturbation can be due to a voltage applied on a semiconductor device, or electromagnetic waves (photons) incident on a semiconductor. The new Schrodinger equation for the time evolution of the state is

$$i\hbar\frac{\partial}{\partial t}|\Psi_t\rangle = [H_0 + W_t]|\Psi_t\rangle. \tag{24.3}$$

In principle, solving this equation will yield the complete future quantum states. In practice, this equation is unsolvable, even for the simplest of perturbations. Physically, the perturbation will 'scatter' a particle that was, say in state $|0\rangle$ to state $|n\rangle$. However, we had noted that even in the *absence* of perturbations, the eigen-state vectors were already evolving with time in the Hilbert space. For example, state vector $|0\rangle$ was rotating at an angular frequency ω_0, and state vector $|n\rangle$ at ω_n. This is shown schematically in the left of Figure 24.1. It would be nice to work with *unperturbed* state vectors that do not change in time, as in the right of Figure 24.1. This calls for a transformation to a vector space that 'freezes' the time evolution of the unperturbed eigen state-vectors. Such a transformation is achieved by the relation

$$|\Psi_t\rangle = e^{-i\frac{H_0}{\hbar}t}|\Psi(t)\rangle, \tag{24.4}$$

where H_0 is the Hamiltonian *operator*. Note that the operator now sits in the exponential, but it should not worry us much. We will see that it is rather useful to have it up there. The reason for this non-obvious transformation is because when we put this into the Schrodinger equation in Equation 24.3, we get

$$i\hbar\left(-\frac{i}{\hbar}H_0 e^{-i\frac{H_0}{\hbar}t}|\Psi(t)\rangle + e^{-i\frac{H_0}{\hbar}t}\frac{\partial}{\partial t}|\Psi(t)\rangle\right) = [H_0 + W_t]e^{-i\frac{H_0}{\hbar}t}|\Psi(t)\rangle, \tag{24.5}$$

and there is a crucial cancellation, leaving us with

$$\boxed{i\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = [e^{+i\frac{H_0}{\hbar}t}W_t e^{-i\frac{H_0}{\hbar}t}]|\Psi(t)\rangle = W(t)|\Psi(t)\rangle} \tag{24.6}$$

where $W(t) = e^{+i\frac{H_0}{\hbar}t}W_t e^{-i\frac{H_0}{\hbar}t}$. Can we take the *operator* $e^{-i\frac{H_0}{\hbar}t}$ from the left to the right side as $e^{+i\frac{H_0}{\hbar}t}$? Yes we can, because $e^{+i\frac{H_0}{\hbar}t}\cdot e^{-i\frac{H_0}{\hbar}t} = I$, the identity operator.

The boxed form of the time-evolution is called the *interaction* picture, as opposed to the conventional form of Equation 24.3, which is called the 'Schrodinger' picture. Note that if there is no perturbation, $W_t = 0 \implies W(t) = 0 \implies i\hbar\frac{\partial|\Psi(t)\rangle}{\partial t} = 0$. Then, $|\Psi(t)\rangle = |\Psi(t_0)\rangle$, and we have managed to find the state vector representation in which the unperturbed eigenvectors are indeed frozen in time.

Now lets turn the perturbation W_t on. Formally, the state vector at time t in the interaction representation is obtained by integrating both sides:

$$|\Psi(t)\rangle = |\Psi(t_0)\rangle + \frac{1}{i\hbar}\int_{t_0}^{t} dt' W(t')|\Psi(t')\rangle, \tag{24.7}$$

and it looks as if we have solved the problem. However, there is a catch - the unknown state vector $|\Psi(t)\rangle$ appears also on the right side - inside the integral. This is also a recursive relation! It reminds of the Brilloiun-Wigner form of non-degenerate perturbation theory. Let's try to iterate the formal solution once:

$$|\Psi(t)\rangle = |\Psi(t_0)\rangle + \frac{1}{i\hbar}\int_{t_0}^{t} dt' W(t')\left[|\Psi(t_0)\rangle + \frac{1}{i\hbar}\int_{t_0}^{t'} dt'' W(t'')|\Psi(t'')\rangle\right], \tag{24.8}$$

and then keep going:

$$|\Psi(t)\rangle = \underbrace{|\Psi(t_0)\rangle}_{\sim W^0} + \underbrace{\frac{1}{i\hbar}\int_{t_0}^{t} dt' W(t')|\Psi(t_0)\rangle}_{\sim W^1} + \underbrace{\frac{1}{(i\hbar)^2}\int_{t_0}^{t} dt' W(t')\int_{t_0}^{t'} dt'' W(t'')|\Psi(t_0)\rangle}_{\sim W^2} + ... \tag{24.9}$$

We thus obtain a formal perturbation series to many orders. The hope is that the series converges rapidly if the perturbation is 'small', because successive terms increase as a power law, which for a small number gets even smaller. Let's accept that weak argument now at face value, and we return later to address, justify, and where possible, fix this cavalier approximation.

Let $|\Psi(t_0)\rangle = |0\rangle$ be the initial state of the quantum system. The perturbation is turned on at time t_0. The probability *amplitude* for the system to be found in state $|n\rangle$ at time $t(> t_0)$ is $\langle n|\Psi_t\rangle$. Note the Schrodinger representation! But the transformation from Schrodinger to interaction picture helps: $\langle n|\Psi_t\rangle = \langle n|e^{-i\frac{H_0}{\hbar}t}\Psi(t)\rangle = e^{-i\frac{E_n}{\hbar}t}\langle n|\Psi(t)\rangle$. This implies $|\langle n|\Psi_t\rangle|^2 = |\langle n|\Psi(t)\rangle|^2$ - for all eigenstates $|n\rangle$. Let us make an approximation in this section and retain only the first order term in the perturbation series. We will return later and discuss the higher order terms that capture multiple-scattering events. Retaining only the terms of Eq. 24.9 to first order in the perturbation W gives

$$\langle n|\Psi(t)\rangle \approx \underbrace{\langle n|0\rangle}_{=0} + \frac{1}{i\hbar}\int_{t_0}^{t} dt' \langle n|W(t')|0\rangle = \frac{1}{i\hbar}\int_{t_0}^{t} dt' \langle n|e^{+i\frac{H_0}{\hbar}t'} W_{t'} e^{-i\frac{H_0}{\hbar}t'}|0\rangle. \tag{24.10}$$

Let us assume the perturbation to be of the form $W_t = e^{\eta t}W$ representing a 'slow turn on', with $\eta = 0^+$, and $W = W(r)$ a function that depends only on space. If $\eta = 0$, then the perturbation is time-independent. But if $\eta = 0^+$, then $e^{\eta t_0} \to 0$ as $t_0 \to -\infty$. This construction thus effectively kills the perturbation far in the distant past, but slowly turns it on to full strength at $t = 0$. We will discuss more of the physics buried inside η later. For now, we accept it as a mathematical construction, with the understanding to take the limit $\eta \to 0$ at the end. Then, the amplitude in state $|n\rangle$ simplifies:

$$\langle n|\Psi(t)\rangle \approx \frac{1}{i\hbar}\int_{t_0}^{t} dt' \underbrace{\boxed{\langle n|e^{+i\frac{H_0}{\hbar}t'}}}_{e^{+i\frac{E_n}{\hbar}t'}\langle n|} e^{\eta t'} W \underbrace{\boxed{e^{-i\frac{H_0}{\hbar}t'}|0\rangle}}_{e^{-i\frac{E_0}{\hbar}t'}|0\rangle} = \frac{\langle n|W|0\rangle}{i\hbar}\int_{t_0}^{t} dt' e^{i\left(\frac{E_n - E_0}{\hbar}\right)t'} e^{\eta t'}, \tag{24.11}$$

and the integral over time may be evaluated exactly to yield

$$\int_{t_0}^{t} dt' e^{i\left(\frac{E_n-E_0}{\hbar}\right)t'} e^{\eta t'} = \frac{e^{i\left(\frac{E_n-E_0}{\hbar}\right)t} e^{\eta t} - e^{i\left(\frac{E_n-E_0}{\hbar}\right)t_0} e^{\eta t_0}}{i\left(\frac{E_n-E_0}{\hbar}\right)+\eta} \underbrace{=}_{t_0\to-\infty} \frac{e^{i\left(\frac{E_n-E_0}{\hbar}\right)t} e^{\eta t}}{i\left(\frac{E_n-E_0}{\hbar}\right)+\eta}. \quad (24.12)$$

The amplitude then is

$$\langle n|\Psi(t)\rangle \approx \frac{\langle n|W|0\rangle}{i\hbar} \cdot \frac{e^{i\left(\frac{E_n-E_0}{\hbar}\right)t} e^{\eta t}}{i\left(\frac{E_n-E_0}{\hbar}\right)+\eta} = \langle n|W|0\rangle \cdot \frac{e^{i\left(\frac{E_n-E_0}{\hbar}\right)t} e^{\eta t}}{(E_0-E_n)+i\hbar\eta}. \quad (24.13)$$

The *probability* of the state making a transition from $|0\rangle$ to $|n\rangle$ at time t is

$$|\langle n|\Psi_t\rangle|^2 = |\langle n|\Psi(t)\rangle|^2 \approx |\langle n|W|0\rangle|^2 \frac{e^{2\eta t}}{(E_0-E_n)^2+(\hbar\eta)^2}. \quad (24.14)$$

The *rate* of transitions from state $|0\rangle \to |n\rangle$ is

$$\frac{1}{\tau_{|0\rangle\to|n\rangle}} = \frac{d}{dt}|\langle n|\Psi(t)\rangle|^2 \approx |\langle n|W|0\rangle|^2 \left(\frac{2\eta}{(E_0-E_n)^2+(\hbar\eta)^2}\right) e^{2\eta t}. \quad (24.15)$$

Now we take $\eta \to 0^+$. The third term $e^{2\eta t} \to 1$, but we must be careful with the quantity in the bracket. When $\eta \to 0$, this quantity is 0, *except* when the term $E_0 - E_n = 0$; then the term seems indeterminate. By making a plot of this function, we can convince ourselves that it approaches a Dirac delta function in the variable $E_0 - E_n$. The mathematical identity $\lim_{\eta\to0^+} \frac{2\eta}{x^2+\eta^2} = \lim_{\eta\to0^+} \frac{1}{i}[\frac{1}{x-i\eta} - \frac{1}{x+i\eta}] = 2\pi\delta(x)$, where $\delta(...)$ confirms this: in the limit, the term indeed becomes the Dirac-delta function.

Then, using $\delta(ax) = \delta(x)/|a|$, the rate of transitions is given by

$$\boxed{\frac{1}{\tau_{|0\rangle\to|n\rangle}} \approx \frac{2\pi}{\hbar}|\langle n|W|0\rangle|^2\delta(E_0-E_n),} \quad (24.16)$$

which is the Fermi's golden rule. The general form is $2\pi/\hbar$ times the transition matrix element squared, times a Dirac-delta function as a statement of energy conservation.

24.3 Perturbations oscillating in time

Now suppose the perturbation potential was oscillating in time. We will encounter such perturbations frequently, in the form of electron-photon, or electron-phonon interactions. The mathematical nature of such perturbations with a slow turn-on is

$$W_t = 2We^{\eta t}\cos(\omega t) = e^{\eta t}W(e^{i\omega t} + e^{-i\omega t}) \tag{24.17}$$

which leads to a $|0\rangle \to |n\rangle$ transition amplitude

$$\langle n|\Psi(t)\rangle \approx \frac{\langle n|W|0\rangle}{i\hbar}\left(\int_{t_0}^{t} dt' e^{i\left(\frac{E_n - E_0 + \hbar\omega}{\hbar}\right)t'} e^{\eta t'} + \int_{t_0}^{t} dt' e^{i\left(\frac{E_n - E_0 - \hbar\omega}{\hbar}\right)t'} e^{\eta t'}\right), \tag{24.18}$$

Similar to Equations 24.12 and 24.13, evaluating the integral with $t_0 \to -\infty$, we get the amplitude for transitions

$$\langle n|\Psi(t)\rangle \approx \langle n|W|0\rangle \cdot \left(\frac{e^{i\left(\frac{E_n - E_0 + \hbar\omega}{\hbar}\right)t} e^{\eta t}}{(E_0 - E_n + \hbar\omega) + i\hbar\eta} + \frac{e^{i\left(\frac{E_n - E_0 - \hbar\omega}{\hbar}\right)t} e^{\eta t}}{(E_0 - E_n - \hbar\omega) + i\hbar\eta}\right). \tag{24.19}$$

The probability is then

$$\begin{aligned}|\langle n|\Psi(t)\rangle|^2 \approx |\langle n|W|0\rangle|^2 \cdot [&\frac{e^{2\eta t}}{(E_0 - E_n + \hbar\omega)^2 + (\hbar\eta)^2} + \frac{e^{2\eta t}}{(E_0 - E_n - \hbar\omega)^2 + (\hbar\eta)^2} + \\ &\frac{e^{2i\omega t} e^{2\eta t}}{(E_0 - E_n + \hbar\omega + i\hbar\eta)(E_0 - E_n - \hbar\omega - i\hbar\eta)} + \\ &\frac{e^{-2i\omega t} e^{2\eta t}}{(E_0 - E_n + \hbar\omega - i\hbar\eta)(E_0 - E_n - \hbar\omega + i\hbar\eta)}]\end{aligned} \tag{24.20}$$

The rate of transition is then

$$\begin{aligned}\frac{d}{dt}|\langle n|\Psi(t)\rangle|^2 \approx |\langle n|W|0\rangle|^2 \cdot [&\frac{2\eta e^{2\eta t}}{(E_0 - E_n + \hbar\omega)^2 + (\hbar\eta)^2} + \frac{2\eta e^{2\eta t}}{(E_0 - E_n - \hbar\omega)^2 + (\hbar\eta)^2} + \\ &\frac{2(\eta + i\omega) e^{2i\omega t} e^{2\eta t}}{(E_0 - E_n + \hbar\omega + i\hbar\eta)(E_0 - E_n - \hbar\omega - i\hbar\eta)} + \\ &\frac{2(\eta - i\omega) e^{-2i\omega t} e^{2\eta t}}{(E_0 - E_n + \hbar\omega - i\hbar\eta)(E_0 - E_n - \hbar\omega + i\hbar\eta)}].\end{aligned} \tag{24.21}$$

Notice that the last two (interference) terms are a complex conjugate pair, which they must be, because the rate of transition is real. The sum is then $2\times$ the real part of either term. After some manipulations, one obtains

$$\frac{d}{dt}|\langle n|\Psi(t)\rangle|^2 \approx \langle n|W|0\rangle|^2 e^{2\eta t}\cdot\left(\frac{2\eta}{(E_0-E_n+\hbar\omega)^2+(\hbar\eta)^2}+\frac{2\eta}{(E_0-E_n-\hbar\omega)^2+(\hbar\eta)^2}\right)[1-\cos(2\omega t)]+ 2\sin(2\omega t)\left(\frac{E_0-E_n+\hbar\omega}{(E_0-E_n+\hbar\omega)^2+(\hbar\eta)^2}-\frac{E_0-E_n-\hbar\omega}{(E_0-E_n-\hbar\omega)^2+(\hbar\eta)^2}\right). \tag{24.22}$$

Note that the rate has a part that does not oscillate, and another which does, with *twice* the frequency of the perturbing potential. If we average over a few periods of the oscillation, $\langle\cos(2\omega t)\rangle_t = \langle\sin(2\omega t)\rangle_t = 0$. Then, by taking the limit $\eta \to 0^+$ in the same fashion as in Equation 24.16, we obtain the Fermi's golden rule for oscillating perturbations:

$$\boxed{\frac{1}{\tau_{|0\rangle\to|n\rangle}} \approx \frac{2\pi}{\hbar}\times|\langle n|W|0\rangle|^2\times[\underbrace{\delta(E_0-E_n+\hbar\omega)}_{\text{absorption}}+\underbrace{\delta(E_0-E_n-\hbar\omega)}_{\text{emission}}].} \tag{24.23}$$

The Dirac-delta functions now indicate that the exchange of energy between the quantum system and the perturbing field is through *quanta* of energy: either by absorption, leading to $E_n = E_0 + \hbar\omega$, or emission, leading to $E_n = E_0 - \hbar\omega$. The rates of each individual processes are the same. Which process (emission or absorption) dominates depends on the occupation functions of the quantum states.

24.4 Transitions to a continuum of states

The Fermi golden rule results in Equation 24.16 and 24.23 are in a form suitable for tracking transitions between *discrete*, or individual states $|0\rangle$ and $|n\rangle$. For many situations encountered in semiconductors, these transitions will be between states within, or between energy bands, where a *continuum* of states exist. In those cases, the *net* transition rate will be obtained by summing over all relevant states. Even the transition between manifestly discrete states - for example the electron ground state of hydrogen atom to the first excited state - by the absorption of a photon - occurs by the interaction between the discrete electron states and the states of the electromagnetic spectrum, which forms a continuum.

As an example, consider the transitions between electron states in the conduction band due to a point scatterer in a 3D semiconductor. Let us say the point scatterer potential is $W(r) = V_0\delta(\mathbf{r})$, with V_0 in units of eV·m^3. This is not an oscillating potential, so we use the golden rule result of Equation 24.16. We first find the matrix element between electron states $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$:

$$\langle \mathbf{k}'|V_0\delta(\mathbf{r})|\mathbf{k}\rangle = \int d^3\mathbf{r} \left(\frac{e^{-i\mathbf{k}'\cdot\mathbf{r}}}{\sqrt{V}}\right) V_0\delta(\mathbf{r}) \left(\frac{e^{+i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{V}}\right) = \frac{V_0}{V}, \tag{24.24}$$

where we have used the property that the Fourier transform of a Dirac-delta function is equal to 1. Then, the transition (or scattering) rate to any state $|\mathbf{k}'\rangle$ is

$$\frac{1}{\tau(|\mathbf{k}\rangle \rightarrow |\mathbf{k}'\rangle)} = \frac{2\pi}{\hbar}\left(\frac{V_0}{V}\right)^2 \delta(E_\mathbf{k} - E_{\mathbf{k}'}). \tag{24.25}$$

The net scattering 'out' of state $|\mathbf{k}\rangle$ into the continuum of states $|\mathbf{k}'\rangle$ is then given by

$$\frac{1}{\tau(|\mathbf{k}\rangle)} = \sum_{\mathbf{k}'} \frac{1}{\tau(|\mathbf{k}\rangle \rightarrow |\mathbf{k}'\rangle)} = \frac{2\pi}{\hbar}\left(\frac{V_0}{V}\right)^2 \underbrace{\sum_{\mathbf{k}'} \delta(E_\mathbf{k} - E_{\mathbf{k}'})}_{D(E_\mathbf{k})}, \tag{24.26}$$

where we note that the sum over final states of the Dirac-delta function is the density of states $D(E_\mathbf{k})$ in units eV^{-1} of the electron at energy $E_\mathbf{k}$. This procedure illustrates an important result - the scattering rate for continuum of states is in general proportional to a density of states relevant to the problem. The strength of scattering increases as the square of the scattering potential. The occurrence of the (volume)^2 term in the denominator may be disconcerting at first. However, the macroscopic volume (or area, or length) terms will for most cases cancel out because of purely physical reasons. For example, for the problem illustrated here, if instead of just *one* point scatterer, we had N, the *density* of scatterers is $n_{sc} = N/V$. Together with the conversion process $\sum_{\mathbf{k}'} \rightarrow V\int d^3\mathbf{k}'/(2\pi)^3$, we obtain

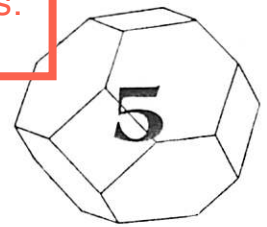
$$\frac{1}{\tau(E_\mathbf{k})} = \frac{2\pi}{\hbar}\left(\frac{V_0}{V}\right)^2 n_{sc}V \int \frac{d^3\mathbf{k}'}{\frac{(2\pi)^3}{V}} \delta(E_\mathbf{k} - E_{\mathbf{k}'}) = \frac{2\pi}{\hbar}V_0^2 n_{sc} g(E_\mathbf{k}). \tag{24.27}$$

Here the density of states $g(E_\mathbf{k})$ is per unit volume, in units $1/(\text{eV.m}^3)$, as is standard in semiconductor physics. The scattering rate is linearly proportional to the density of scatterers. What is not immediately clear is how can we capture the effect of N scatterers by just multiplying the individual scatterer rate by N. This can be done if

the scatterers are uncorrelated, as will be discussed in the transport chapters. For now, note that the macroscopic volume has canceled out, as promised.

(a) What is the concentration of acceptors? (Assume that $N_d = 0$.)
(b) How many donors ($g_d = 2$, $\Delta\mathcal{E}_d = 6$ meV) must be introduced to reduce the hole concentration to 1×10^{17} cm^{-3}?
(c) Consider a sample having the impurity concentrations calculated in parts (a) and (b). If the temperature increases slightly, will the hole concentration increase or decrease? Why?

4.13. A semiconductor sample is made quite thin in its Z dimension so that k_z may assume very few discrete values. What is the concentration of electrons in the conduction band at equilibrium (electrons per unit area)? Assume that the conduction band is spherically symmetric, the material is nondegenerate, and the effective mass approximation is valid. If the Fermi energy is $4kT$ below the conduction band energy, what is the concentration of conduction band electrons per unit area in GaAs at 300 K? ($m_e = 0.067m$.)



Transport Properties

When electrons are in thermal equilibrium with the lattice, they are distributed among possible energy levels in a manner given by the Fermi–Dirac function of (4.41). Under these conditions no net transport of charge or energy occurs since the probability that a state with wavevector $\mathbf{k}$ is occupied is the same as that for a state with wavevector $-\mathbf{k}$. That is, the equilibrium distribution function, f_0, is symmetrical about the origin in $\mathbf{k}$-space.

When external forces or temperature gradients are applied to the material, however, this is no longer true. Under these conditions we can, in principle, determine the nonequilibrium distribution function, f, in a manner similar to that used for f_0 in Section 4.1. In the nonequilibrium case, however, we would have to maximize (4.10) for the most probable distribution subject to the additional constraints that a steady flow of charge and energy be maintained. That is,

$$\mathbf{J} = -q \sum_k n_k \mathbf{v}_k \tag{5.1}$$

$$\mathbf{W} = \sum_k e_k n_k \mathbf{v}_k \tag{5.2}$$

where $\mathbf{J}$ is the electrical current density, $\mathbf{W}$ is the heat flow density, and e_k is the heat content per electron. These additional constraints produce an asymmetry in the nonequilibrium distribution function which shifts its center away from the origin in k-space. In this chapter we examine this nonequilibrium distribution of electrons and use it to determine the transport of charge and energy in semiconductors.

5.1 BOLTZMANN'S EQUATION

The approach we take in determining the transport properties of semiconductors is to first construct an electron wave packet from plane wave solutions to the time-dependent Schrödinger equation. This is the same method as that used in Section 2.8. Then, from the correspondence principle, we can use a semiclassical approach.

On this basis, let $f(\mathbf{k}, \mathbf{r}, t)$ be the probability that a state with wavevector $\mathbf{k}$ is occupied by an electron with position $\mathbf{r}$ at time t. The electrons are continually changing their position according to (2.109) and, under the influence of forces $\mathbf{F}_t$ (other than the periodic crystal forces), are continually changing their wavevector according to (2.111). $\mathbf{F}_t$ includes applied forces $\mathbf{F}$ and forces $\mathbf{F}_c$ due to electron collisions with lattice vibrations and other imperfections in the crystal. Therefore, at time $t + dt$ the probability that a state with wavevector $\mathbf{k} + d\mathbf{k}$ is occupied by an electron with position $\mathbf{r} + d\mathbf{r}$ is given by

$$f\left(\mathbf{k} + \frac{1}{\hbar}\mathbf{F}_t\, dt, \mathbf{r} + \mathbf{v}\, dt, t + dt\right)$$

The total rate of change in the distribution function in the region of the point $\mathbf{r}$ is then

$$\frac{df}{dt} = \frac{1}{\hbar}\mathbf{F}_t \cdot \nabla_k f + \mathbf{v} \cdot \nabla_r f + \frac{\partial f}{\partial t} \tag{5.3}$$

On the right side of (5.3), the first term takes into account changes in the distribution due to forces, the second term accounts for changes due to concentration gradients, and the last term is the local change in the distribution at the point $\mathbf{r}$. Equation (5.3) is referred to as Boltzmann's transport equation.

Since the total number of states in the crystal is constant, the total rate of change of the distribution function must be zero (Liouville's theorem), and

$$\frac{\partial f}{\partial t} = -\frac{1}{\hbar}\mathbf{F}_t \cdot \nabla_k f - \mathbf{v} \cdot \nabla_r f = \left.\frac{\partial f}{\partial t}\right|_c - \frac{1}{\hbar}\mathbf{F} \cdot \nabla_k f - \mathbf{v} \cdot \nabla_r f \tag{5.4}$$

Because of the difficulty of finding a value for $\mathbf{F}_c$, we separate the collision forces from the applied forces by defining a local change in the distribution due to collisions only as

$$\left.\frac{\partial f}{\partial t}\right|_c \equiv -\frac{1}{\hbar}\mathbf{F}_c \cdot \nabla_k f \tag{5.5}$$

Let us examine this collision term. The action of applied forces and gradients tends to disturb the distribution function f from its equilibrium value f_0. If this disturbance is removed, the scattering processes will tend to restore equilibrium. When the change in the distribution is not large compared to its initial value, it is reasonable to assume that

$$\frac{\partial f}{\partial t} = \left.\frac{\partial f}{\partial t}\right|_c = \frac{-(f - f_0)}{\tau_m} \tag{5.6}$$

where τ_m is a constant of proportionality called the *momentum relaxation time*. In general, τ_m depends on the electron energy and is different for different scattering mechanisms. We examine τ_m for various scattering processes in some detail in Chapter 6. In the meantime, integrating (5.6), we find that

$$f(t) - f_0 = [f(0) - f_0] \exp\left(-\frac{t}{\tau_m}\right) \tag{5.7}$$

That is, the momentum relaxation time τ_m characterizes an exponential relaxation of the distribution function f to its equilibrium value f_0.

In the steady state $\partial f/\partial t = 0$ and using this and (5.6) in (5.4), we obtain the steady-state Boltzmann equation in the relaxation time approximation,

$$f = f_0 - \frac{\tau_m}{\hbar}\mathbf{F} \cdot \nabla_k f - \tau_m \mathbf{v} \cdot \nabla_r f \tag{5.8}$$

Since

$$\nabla_k f = \frac{\partial f}{\partial \mathscr{E}} \nabla_k \mathscr{E} \tag{5.9}$$

the Boltzmann equation can be put in the form

$$f = f_0 - \frac{\tau_m}{\hbar}\frac{\partial f}{\partial \mathscr{E}}\mathbf{F} \cdot \nabla_k \mathscr{E} - \tau_m \mathbf{v} \cdot \nabla_r f \tag{5.10}$$

From (2.109)

$$\mathbf{v} = \frac{1}{\hbar}\nabla_k \mathscr{E}$$

so that (5.10) is finally

$$f = f_0 - \tau_m \mathbf{v} \cdot \left(\frac{\partial f}{\partial \mathscr{E}}\mathbf{F} + \nabla_r f\right) \tag{5.11}$$

or

$$f = f_0 - \frac{\tau_m}{\hbar}\nabla_k \mathscr{E} \cdot \left(\frac{\partial f}{\partial \mathscr{E}}\mathbf{F} + \nabla_r f\right) \tag{5.12}$$

The Boltzmann equation in the form of (5.12) tells us that the nonequilibrium distribution of electrons depends on the scattering processes through the

term τ_m, on the band structure through $\nabla_k \mathscr{E}$, on applied forces through $(\partial f/\partial \mathscr{E})\mathbf{F}$, and on concentration gradients through $\nabla_r f$. Therefore, we have, in general, a rather difficult partial differential equation to solve for f, the nonequilibrium distribution function.

5.2 DISTRIBUTION FUNCTION

Before looking at a more general solution for f, let us look at the simplest possible case. We will assume that the only applied force is a small electric field $\mathbf{E}$ and that there are no concentration or temperature gradients. Under these conditions (5.11) becomes

$$f = f_0 + q\tau_m \frac{\partial f}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{E} \tag{5.13}$$

Equation (5.13) can be integrated to obtain an analytical expression for f provided that the energy dependence of τ_m is known. The solution, however, is nonlinear in $\mathbf{E}$. Under the relaxation time assumption that the change in distribution function is not large, we can also make the approximation that

$$\frac{\partial f}{\partial \mathscr{E}} \simeq \frac{\partial f_0}{\partial \mathscr{E}} \tag{5.14}$$

so that

$$f = f_0 + q\tau_m \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{E} \tag{5.15}$$

This retains only a linear term in E, which is consistent with our initial assumption of a small electric field.

Let us now look at a more general situation, where we include a small electric field $\mathbf{E}$ and an arbitrary magnetic field $\mathbf{B}$ in the force term and retain the term for concentration and temperature gradients. Under these conditions (5.8) is

$$\frac{f - f_0}{\tau_m} = \frac{+q}{\hbar} (\mathbf{E} + \mathbf{v} \times \mathbf{B})\cdot\nabla_k f - \mathbf{v}\cdot\nabla_r f \tag{5.16}$$

We will assume that the solution for (5.16) has the form of (5.15),

$$f = f_0 + \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{G} \tag{5.17}$$

and then solve for the unknown vector $\mathbf{G}$. Inserting (5.17) into (5.16) the term on the left-hand side of (5.16) is simply

$$\frac{f - f_0}{\tau_m} = \frac{1}{\tau_m} \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{G} \tag{5.18}$$

The terms on the right-hand side of (5.16) are more difficult to evaluate.

Ignoring $+q/\hbar$ for the moment, the first term on the right-side of (5.16) is

$$(\mathbf{E} + \mathbf{v} \times \mathbf{B})\cdot\nabla_k f = \mathbf{E}\cdot\nabla_k f_0 + (\mathbf{v} \times \mathbf{B})\cdot\nabla_k f_0 + \mathbf{E}\cdot\nabla_k \left(\frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{G}\right) + (\mathbf{v} \times \mathbf{B})\cdot\nabla_k \left(\frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{G}\right) \tag{5.19}$$

In (5.19) the third term on the right has both $\mathbf{E}$ and $\mathbf{G}$ and is thus a second-order term in $\mathbf{E}$. Neglecting this third term and making the substitution

$$\nabla_k f_0 = \frac{\partial f_0}{\partial \mathscr{E}} \hbar\mathbf{v} \tag{5.20}$$

in the first and second term on the right, (5.19) becomes

$$(\mathbf{E} + \mathbf{v} \times \mathbf{B})\cdot\nabla_k f = \hbar \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{E} + \hbar \frac{\partial f_0}{\partial \mathscr{E}} (\mathbf{v} \times \mathbf{B})\cdot\mathbf{v} + (\mathbf{v} \times \mathbf{B})\cdot\nabla_k \left(\frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{G}\right) \tag{5.21}$$

Since $(\mathbf{v} \times \mathbf{B})\cdot\mathbf{v}$ is identically zero, the second term on the right in (5.21) is zero. When we perform the gradient operation in the third term, (5.21) is

$$(\mathbf{E} + \mathbf{v} \times \mathbf{B})\cdot\nabla_k f = \hbar \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{E} + \frac{\partial f_0}{\partial \mathscr{E}} (\mathbf{v} \times \mathbf{B})\cdot\nabla_k(\mathbf{v}\cdot\mathbf{G}) + (\mathbf{v}\cdot\mathbf{G})(\mathbf{v} \times \mathbf{B})\cdot\nabla_k \frac{\partial f_0}{\partial \mathscr{E}} \tag{5.22}$$

From (5.20) we see that

$$\nabla_k \frac{\partial f_0}{\partial \mathscr{E}} = \frac{\partial^2 f_0}{\partial \mathscr{E}^2} \hbar\mathbf{v}$$

and the third term on the right in (5.22) vanishes because of $(\mathbf{v} \times \mathbf{B})\cdot\mathbf{v}$. The second term on the right in (5.22) can be resolved into Cartesian components and rearranged to obtain the result,

$$\frac{-q}{\hbar} (\mathbf{E} + \mathbf{v} \times \mathbf{B})\cdot\nabla_k f = -q \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{E} - \frac{q}{\hbar^2} \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot[\mathbf{B} \times (\mathbf{G}\cdot\nabla_k)\nabla_k \mathscr{E}] \tag{5.23}$$

This is the desired form for the first term on the right in (5.16).

Let us now examine the second term on the right in (5.16),

$$\mathbf{v}\cdot\nabla_r f = \mathbf{v}\cdot\nabla_r f_0 + \mathbf{v}\cdot\nabla_r\left(\frac{\partial f_0}{\partial \mathscr{E}}\,\mathbf{v}\cdot\mathbf{G}\right) \tag{5.24}$$

For our purposes we can assume that the spatial dependence of $\mathbf{G}$ is small and consider only the first term in (5.24). We then have

$$\begin{aligned}\mathbf{v}\cdot\nabla_r f &= \frac{\partial f_0}{\partial((\mathscr{E}-\mu)/kT)}\,\mathbf{v}\cdot\nabla_r\left(\frac{\mathscr{E}-\mu}{kT}\right)\\ &= kT\frac{\partial f_0}{\partial \mathscr{E}}\,\mathbf{v}\cdot\nabla_r\left(\frac{\mathscr{E}-\mu}{kT}\right)\end{aligned} \tag{5.25}$$

where μ is the chemical potential.

Using (5.18), (5.23), and (5.25), Boltzmann's equation is now

$$\begin{aligned}\frac{1}{\tau_m}\frac{\partial f_0}{\partial \mathscr{E}}\,\mathbf{v}\cdot\mathbf{G} = {} & +q\frac{\partial f_0}{\partial \mathscr{E}}\,\mathbf{v}\cdot\mathbf{E} + \frac{q}{\hbar^2}\frac{\partial f_0}{\partial \mathscr{E}}\,\mathbf{v}\cdot[\mathbf{B}\times(\mathbf{G}\cdot\nabla_k)\nabla_k\mathscr{E}]\\ & - kT\frac{\partial f_0}{\partial \mathscr{E}}\,\mathbf{v}\cdot\nabla_r\left(\frac{\mathscr{E}-\mu}{kT}\right)\end{aligned} \tag{5.26}$$

Since each term in (5.26) has a common factor $(\partial f_0/\partial\mathscr{E})\mathbf{v}$ on the left, it can be eliminated to obtain

$$\frac{1}{\tau_m}G = +q\mathbf{E} - kT\,\nabla_r\left(\frac{\mathscr{E}-\mu}{kT}\right) + \frac{q}{\hbar^2}[\mathbf{B}\times(\mathbf{G}\cdot\nabla_k)\nabla_k\mathscr{E}] \tag{5.27}$$

Defining an electrothermal field for electrons, $\mathscr{F}$, by

$$q\mathscr{F} = +q\mathbf{E} - T\nabla_r\left(\frac{\mathscr{E}-\mu}{T}\right) \tag{5.28}$$

(5.27) has the form

$$\mathbf{G} = q\tau_m\mathscr{F} + \frac{q\tau_m}{\hbar^2}[\mathbf{B}\times(\mathbf{G}\cdot\nabla_k)\nabla_k\mathscr{E}] \tag{5.29}$$

Equation (5.29) can be solved for $\mathbf{G}$ by using an explicit expression for the conduction band minima. For this purpose we will assume ellipsoidal minima with a quadratic dispersion relationship as given by (2.117). In vector notation we have

$$\mathscr{E} = \mathscr{E}_c + \tfrac{1}{2}\hbar^2\mathbf{k}\cdot\mathbf{M}\cdot\mathbf{k} \tag{5.30}$$

where

$$\mathbf{M} = \begin{bmatrix}\dfrac{1}{m_1^*} & 0 & 0\\ 0 & \dfrac{1}{m_2^*} & 0\\ 0 & 0 & \dfrac{1}{m_3^*}\end{bmatrix} \tag{5.31}$$

is the effective mass tensor, the $\mathbf{k}$ on its left is a row vector, and the $\mathbf{k}$ on its right is a column vector. From (5.30),

$$(\mathbf{G}\cdot\nabla_k)\nabla_k\mathscr{E} = \hbar^2\mathbf{M}\cdot\mathbf{G} \tag{5.32}$$

and (5.29) takes the form

$$\mathbf{G} = q\tau_m\mathscr{F} + q\tau_m\mathbf{B}\times(\mathbf{M}\cdot\mathbf{G}) \tag{5.33}$$

Reducing this equation to its components and solving for $\mathbf{G}$, we have, finally,

$$\mathbf{G} = q\tau_m\left[\frac{\mathscr{F} - q\tau_m\mathbf{M}\cdot(\mathscr{F}\times\mathbf{B}) + (q\tau_m)^2(\det\mathbf{M})(\mathscr{F}\cdot\mathbf{B})(\mathbf{M}^{-1}\cdot\mathbf{B})}{1 + (q\tau_m)^2(\det\mathbf{M})(\mathbf{M}^{-1}\cdot\mathbf{B})\cdot\mathbf{B}}\right] \tag{5.34}$$

The nonequilibrium distribution function for electrons in ellipsoidal conduction band minima is obtained by using $\mathbf{G}$ from (5.34) in (5.17). For spherical minima the distribution function is

$$f = f_0 + \frac{\partial f_0}{\partial \mathscr{E}}\,q\tau_m\mathbf{v}\cdot\left[\frac{\mathscr{F} - (q\tau_m/m^*)(\mathscr{F}\times\mathbf{B}) + (q\tau_m/m^*)^2(\mathscr{F}\cdot\mathbf{B})\mathbf{B}}{1 + (q\tau_m/m^*)^2\mathbf{B}\cdot\mathbf{B}}\right] \tag{5.35}$$

We can see that there are four components to the distribution function. The first is simply the equilibrium function f_0 given by (4.41), which does not contribute to the transport of charge and energy. The term that involves $\mathscr{F}$ is the ohmic contribution to the transport properties. This term accounts for electrical and thermal conductivity as well as the Seebeck, Peltier, and Thomson effects. The term with $\mathscr{F}\times\mathbf{B}$ is the Hall contribution to transport and accounts for the Hall, Ettinghausen, Nernst, and Righi–Leduc effects. The $\mathbf{B}^2$ terms in the numerator and denominator of (5.35) account for magnetoresistive effects. We discuss these various effects in more detail later.

The distribution function derived above for electrons can also be used for holes when the appropriate parameters of q, m^*, and τ_m are substituted in the equations.

From (5.1) and (5.2) we can determine the current density $\mathbf{J}$ and heat flow density $\mathbf{W}$ by summing (or integrating) $n_k\mathbf{v}_k$ and $e_k n_k\mathbf{v}_k$, respectively, over the first Brillouin zone. In (4.68), however, we have already obtained

an expression for n, the number of electrons in the conduction band minima. For this reason we can approach the problem from a different point of view. That is, the current density can be determined by

$$\mathbf{J} = -qn\langle \mathbf{v} \rangle \tag{5.36}$$

where $\langle \mathbf{v} \rangle$ is the average velocity of the n electrons in the nonequilibrium distribution. In a similar manner the heat flow density can be obtained from

$$\mathbf{W} = n\langle e\mathbf{v} \rangle \tag{5.37}$$

where the heat content per electron and the velocity are averaged over the distribution. The problem is to determine how this averaging should be performed.

5.3 CHARGE TRANSPORT

For this purpose let us examine the current density for spherical conduction band minima in a small electric field. The average velocity is obtained by summing the velocities of all the electrons in the distribution and normalizing the result. That is,

$$\langle \mathbf{v} \rangle = \frac{\int_{-\infty}^{\infty} \mathbf{v} f \, d\mathbf{v}}{\int_{-\infty}^{\infty} f \, d\mathbf{v}} \tag{5.38}$$

where f is given by (5.15). Inserting (5.15) in (5.38), we have

$$\langle \mathbf{v} \rangle = \frac{\int_{-\infty}^{\infty} \mathbf{v} f_0 \, d\mathbf{v} + q \int_{-\infty}^{\infty} \tau_m (\partial f_0 / \partial \mathscr{E}) \mathbf{v}(\mathbf{v} \cdot \mathbf{E}) \, d\mathbf{v}}{\int_{-\infty}^{\infty} f_0 \, d\mathbf{v} + q \int_{-\infty}^{\infty} \tau_m (\partial f_0 / \partial \mathscr{E})(\mathbf{v} \cdot \mathbf{E}) \, d\mathbf{v}} \tag{5.39}$$

The term on the left in the numerator of this equation is an average over the equilibrium distribution. Since there is no transport of charge in equilibrium, this term is zero. The term on the right in the denominator provides for additional nonequilibrium carriers over the equilibrium concentration. We will take this term to be zero as well. Equation (5.39) is therefore

$$\langle \mathbf{v} \rangle = \frac{+q \int_{-\infty}^{\infty} \tau_m (\partial f_0 / \partial \mathscr{E}) \mathbf{v}(\mathbf{v} \cdot \mathbf{E}) \, d\mathbf{v}}{\int_{-\infty}^{\infty} f_0 \, d\mathbf{v}} \tag{5.40}$$

For spherical conduction band minima we can replace the integrals over three-dimensional velocity space by integrals over energy with relative ease. From (2.109) and (2.118) the relationship between $\mathscr{E}$ and $\mathbf{v}$ is

$$\mathscr{E} - \mathscr{E}_c = \tfrac{1}{2} m^* \mathbf{v}^2 = \tfrac{1}{2} m^* v^2 \tag{5.41}$$

and the differential volume in velocity space is

$$d\mathbf{v} = 4\pi v^2 \, dv \tag{5.42}$$

With (5.41) and (5.42), (5.40) becomes

$$\langle \mathbf{v} \rangle = \frac{+q \int_{\mathscr{E}_c}^{\infty} \tau_m (\partial f_0 / \partial \mathscr{E}) \mathbf{v}(\mathbf{v} \cdot \mathbf{E})(\mathscr{E} - \mathscr{E}_c)^{1/2} \, d\mathscr{E}}{\int_{\mathscr{E}_c}^{\infty} f_0 (\mathscr{E} - \mathscr{E}_c)^{1/2} \, d\mathscr{E}} \tag{5.43}$$

If we consider an electric field in the x direction, the term

$$\mathbf{v}(\mathbf{v} \cdot \mathbf{E}) = +v_x^2 E_x \tag{5.44}$$

Assuming equipartition of energy, each degree of freedom has the same average kinetic energy and

$$v^2 = v_x^2 + v_y^2 + v_z^2 = 3v_x^2 \tag{5.45}$$

Using (5.44) and (5.45), (5.43) is then

$$\langle v_x \rangle = \frac{2qE_x}{3m^*} \frac{\int_{\mathscr{E}_c}^{\infty} \tau_m (\partial f_0 / \partial \mathscr{E})(\mathscr{E} - \mathscr{E}_c)^{3/2} \, d\mathscr{E}}{\int_{\mathscr{E}_c}^{\infty} f_0 (\mathscr{E} - \mathscr{E}_c)^{1/2} \, d\mathscr{E}} \tag{5.46}$$

Defining a drift velocity $\mathbf{v}_d$ as the average velocity of the carriers over the distribution, and introducing the dimensionless variables of (4.65), (5.46) becomes

$$v_{dx} = \frac{-qE_x}{m^*} \langle \tau_m \rangle \tag{5.47}$$

where

$$\langle \tau_m \rangle = \frac{2}{3} \frac{\int_0^{\infty} \tau_m (-\partial f_0 / \partial x) x^{3/2} \, dx}{\int_0^{\infty} f_0 x^{1/2} \, dx} \tag{5.48}$$

Equation (5.48) gives the proper form for the averaging procedure over the distribution of electrons. By evaluating the average momentum relaxation time in the manner proscribed, we can determine the drift velocity from (5.47). Equation (5.47) tells us that for small electric fields, the drift velocity is directly proportional to the field. The constant of proportionality

is called the *conductivity mobility*, μ_c. Thus (5.47) can be put in the form

$$v_{dx} = -\mu_c E_x \tag{5.49}$$

where

$$\mu_c = \frac{q\langle\tau_m\rangle}{m^*} \tag{5.50}$$

From (5.36) and (5.47) the current density in the x direction is

$$J_x = \frac{q^2 n\langle\tau_m\rangle}{m^*} E_x \tag{5.51}$$

Since the constant of proportionality between current density and electric field is referred to as the conductivity, we have

$$\sigma = \frac{q^2 n\langle\tau_m\rangle}{m^*} \tag{5.52}$$

and

$$\sigma = qn\mu_c \tag{5.53}$$

Thus, for the simple case of a small applied electric field, we can define all the transport parameters in terms of the average momentum relaxation time, $\langle\tau_m\rangle$. Once $\langle\tau_m\rangle$ has been obtained, the transport problem is solved.

In general, however, the quantity to be averaged is more complex. For example, to determine the energy transport from (5.37), an extra energy term is included in the average. Also, from (5.34) the vector **G** depends on multiple powers of τ_m and depends on energy through both τ_m and $\mathcal{F}$. Thus the quantity that must be averaged over the electron distribution in more complex transport problems has the form $\tau_m^s x^t$, where s and t are to be determined. Equation (5.48) shows that the averaging procedure for this quantity is

$$\langle\tau_m^s x^t\rangle = \frac{2}{3}\frac{\int_0^\infty \tau_m^s(-\partial f_0/\partial x)x^{t+3/2}\,dx}{\int_0^\infty f_0 x^{1/2}\,dx} \tag{5.54}$$

Equation (5.54) can be evaluated if we know the dependence of τ_m on electron energy. In Chapter 6 we will find that τ_m can be represented as having a simple power dependence on energy for most scattering mechanisms. Therefore, let us take the momentum relaxation time as having the form

$$\tau_m = \tau_0 x^r \tag{5.55}$$

where τ_0 is independent of energy. Equation (5.54) is then

$$\langle\tau_m^s x^t\rangle = \frac{2}{3}\tau_0^s \frac{\int_0^\infty (-\partial f_0/\partial x)x^{sr+t+3/2}\,dx}{\int_0^\infty f_0 x^{1/2}\,dx} \tag{5.56}$$

This equation can be solved by integrating by parts.

Let $u = x^{sr+t+3/2}$ and $dv = (-\partial f_0/\partial x)\,dx$. Then $du = (sr + t + \frac{3}{2})x^{sr+t+1/2}\,dx$ and $v = -f_0$. Using these expressions in the numerator of (5.56) gives us

$$\langle\tau_m^s x^t\rangle = \frac{2}{3}\tau_0^s \frac{-[f_0 x^{sr+t+3/2}]_0^\infty + (sr + t + \frac{3}{2})\int_0^\infty f_0 x^{sr+t+1/2}\,dx}{\int_0^\infty f_0 x^{1/2}\,dx} \tag{5.57}$$

or

$$\langle\tau_m^s x^t\rangle = \frac{2}{3}\left(sr + t + \frac{3}{2}\right)\tau_0^s \frac{\int_0^\infty f_0 x^{sr+t+1/2}\,dx}{\int_0^\infty f_0 x^{1/2}\,dx} \tag{5.58}$$

The integral in the numerator is a Fermi–Dirac integral of order j given by

$$F_j(\eta) = \frac{1}{j!}\int_0^\infty f_0 x^j\,dx \tag{5.59}$$

The integral in the denominator is simply a Fermi–Dirac integral of order $\frac{1}{2}$ which we saw before in (4.67). Values for these integrals are tabulated in Appendix B. Using (5.59) and (4.67), we obtain

$$\langle\tau_m^s x^t\rangle = \frac{4}{3\sqrt{\pi}}\left(sr + t + \frac{3}{2}\right)!\,\tau_0^s \frac{F_{sr+t+1/2}(\eta)}{F_{1/2}(\eta)} \tag{5.60}$$

which is the final form for the average.

To determine transport parameters, we will use expressions of the form $\langle\tau_m^s x^t\rangle$, which can then be evaluated with (5.60). As an example, the conductivity mobility in (5.50) and the conductivity in (5.52) can be obtained from

$$\langle\tau_m\rangle = \frac{4(r + \frac{3}{2})!\,\tau_0}{3\sqrt{\pi}}\frac{F_{r+1/2}(\eta)}{F_{1/2}(\eta)} \tag{5.61}$$

when the value of r for the appropriate scattering mechanism is known.

Let us now look at the transport of electrons in both electric and magnetic fields. Assuming no concentration or temperature gradients, (5.34)

reduces to

$$\mathbf{G} = +q\tau_m \left[\frac{\mathbf{E} - (q\tau_m/m^*)(\mathbf{E} \times \mathbf{B}) + (q\tau_m/m^*)^2 \mathbf{B}(\mathbf{E}\cdot\mathbf{B})}{1 + (q\tau_m/m^*)^2 \mathbf{B}\cdot\mathbf{B}} \right] \tag{5.62}$$

From (5.47) and (5.48) the *drift velocity* is

$$\mathbf{v}_d = \frac{2}{3m^*} \frac{\int_0^\infty \mathbf{G}(-\partial f_0/\partial x) x^{3/2}\, dx}{\int_0^\infty f_0 x^{1/2}\, dx} \tag{5.63}$$

and the current density is given by (5.36). Because of (5.58), it is not necessary to average each term of $\mathbf{G}$ to obtain the current density. We simply have, by inspection,

$$\mathbf{J} = \frac{q^2 n}{m^*} \left\langle \frac{\tau_m}{1 + (\omega_c \tau_m)^2} \right\rangle \mathbf{E} - \frac{q^3 n}{m^{*2}} \left\langle \frac{\tau_m^2}{1 + (\omega_c \tau_m)^2} \right\rangle (\mathbf{E} \times \mathbf{B}) + \frac{q^4 n}{m^{*3}} \left\langle \frac{\tau_m^3}{1 + (\omega_c \tau_m)^2} \right\rangle \mathbf{B}(\mathbf{E}\cdot\mathbf{B}) \tag{5.64}$$

where we have introduced the cyclotron frequency

$$\omega_c \equiv \frac{q\,|\,\mathbf{B}\,|}{m^*} \tag{5.65}$$

The first term in (5.64) is the ohmic term. The factor $1 + (\omega_c \tau_m)^2$ in the denominator of the average in this term reflects the magnetoresistance or reduction in conductivity due to the magnetic field. The second term reflects the Hall effect; it also has magnetoresistance associated with it. The third term is an additional magnetoresistance term.

Let us look at (5.64) for small magnetic fields. Under this condition the second-order terms in $\mathbf{B}$, which produce the magnetoresistance, are small and

$$\mathbf{J} = \frac{q^2 n}{m^*} \langle \tau_m \rangle \mathbf{E} - \frac{q^3 n}{m^{*2}} \langle \tau_m^2 \rangle (\mathbf{E} \times \mathbf{B}) \tag{5.66}$$

If we take $\mathbf{B} = \hat{z} B_z$, (5.66) becomes

$$J_x = \frac{q^2 n}{m^*} \langle \tau_m \rangle E_x - \frac{q^3 n}{m^{*2}} \langle \tau_m^2 \rangle E_y B_z \tag{5.67}$$

$$J_y = \frac{q^2 n}{m^*} \langle \tau_m \rangle E_y + \frac{q^3 n}{m^{*2}} \langle \tau_m^2 \rangle E_x B_z \tag{5.68}$$

$$J_z = \frac{q^2 n}{m^*} \langle \tau_m \rangle E_z \tag{5.69}$$

When $J_y = 0$, (5.68) gives us

$$E_x = -\frac{m^*}{qB_z} \frac{\langle \tau_m \rangle}{\langle \tau_m^2 \rangle} E_y \tag{5.70}$$

Using (5.70) in (5.67) and neglecting a second-order term in B_z, we have

$$J_x = \frac{-qn}{B_z} \frac{\langle \tau_m \rangle^2}{\langle \tau_m^2 \rangle} E_y \tag{5.71}$$

That is, J_x and B_z induce a field E_y. This is the Hall effect. These geometric constraints are obtained experimentally by applying a magnetic field in the z direction, a current in the x direction, and measuring the voltage in the y direction with a high-impedance voltmeter, so that the current in the y direction is negligible.

The *Hall constant* is defined as

$$R_H \equiv \frac{E_y}{J_x B_z} = -\frac{1}{qn} \frac{\langle \tau_m^2 \rangle}{\langle \tau_m \rangle^2} \tag{5.72}$$

From (5.72) we can see that the concentration of electrons in the conduction band can be obtained from an experimental determination of the Hall constant. If the charge carriers are holes in the valence bands, the negative q's in (5.28), (5.34), and (5.36) are replaced by positive q's. The resulting Hall constant is

$$R_H = \frac{1}{qp} \frac{\langle \tau_m^2 \rangle}{\langle \tau_m \rangle^2} \tag{5.73}$$

Thus the sign of the Hall constant (and Hall field) indicates the sign of the charge carriers and Hall measurements can be used to distinguish between n- and p-type material. With (5.53) for the conductivity, we can define a *Hall mobility* as

$$\mu_H = R_H \sigma = \mu_c \frac{\langle \tau_m^2 \rangle}{\langle \tau_m \rangle^2} \tag{5.74}$$

This mobility differs from the conductivity mobility μ_c by the factor

$$r_H \equiv \frac{\langle \tau_m^2 \rangle}{\langle \tau_m \rangle^2} \tag{5.75}$$

which is referred to as the *Hall factor*. For a nondegenerate semiconductor, we find, from (5.60),

$$r_H = \frac{3\sqrt{\pi}}{4} \frac{(2r + \frac{3}{2})!}{[(r + \frac{3}{2})!]^2} \tag{5.76}$$

(In the analysis of experimental data, the Hall factor is often assumed to be

1. Depending on the relevant scattering mechanisms and temperature, this can produce about an 80% error in the carrier concentration.)

Let us next examine the flow of charge for small electric fields in the presence of electron concentration gradients. We will assume there are no magnetic fields or temperature gradients. Under these conditions the electrothermal field (5.28) for electrons is

$$\mathcal{F} = +\mathbf{E} + \frac{1}{q}\nabla_r\mu = +\frac{1}{q}\nabla_r\zeta \tag{5.77}$$

and from (5.34),

$$\mathbf{G} = q\tau_m\mathbf{E} + \tau_m\nabla_r\mu = +\tau_m\nabla_r\zeta \tag{5.78}$$

Equations (5.36), (5.47), and (5.34) tell us that the electron current density is given by

$$\mathbf{J} = \frac{+qn}{m^*}\langle\mathbf{G}\rangle \tag{5.79}$$

or

$$\mathbf{J} = \frac{q^2n}{m^*}\langle\tau_m\rangle\,\mathbf{E} + \frac{qn}{m^*}\langle\tau_m\rangle\,\nabla_r\mu \tag{5.80}$$

Using (5.50) for the conductivity mobility, (5.80) becomes

$$\mathbf{J} = qn\mu_n\mathbf{E} + n\mu_n\,\nabla_r\mu \tag{5.81}$$

where μ_n indicates conductivity mobility for electrons.

The gradient of the chemical potential can be written in terms of a concentration gradient as

$$\nabla_r\mu = \frac{\partial\eta}{\partial n}\frac{\partial\mu}{\partial\eta}\nabla_r n \tag{5.82}$$

Since

$$\frac{d}{d\eta}F_j(\eta) = F_{j-1}(\eta) \tag{5.83}$$

[J. McDougall and E. C. Stoner, *Philos. Trans. R. Soc. London 237*, 67 (1938)], (4.68) gives us

$$\frac{\partial n}{\partial\eta} = N_cF_{-1/2}(\eta) \tag{5.84}$$

Also, from (4.65),

$$\frac{\partial\eta}{\partial\mu} = \frac{1}{kT} \tag{5.85}$$

Using (5.84), (5.85), and (4.68) the chemical potential gradient is

$$\nabla_r\mu = \frac{kT}{N_cF_{1/2}(\eta)}\nabla_r n \tag{5.86}$$

and the electron current density is

$$\mathbf{J} = qn\mu_n\mathbf{E} + kT\mu_n\frac{F_{1/2}(\eta)}{F_{-1/2}(\eta)}\nabla_r n \tag{5.87}$$

Equation (5.87) shows that in the presence of an electric field and a concentration gradient, the electron current density consists of two components: The first component is proportional to the electric field and is called the *drift* term. The second component is directly related to the concentration gradient and is referred to as the *diffusion* term. Notice that the electron current density is in the same direction as the concentration gradient, which is in the direction of increasing concentration.

The diffusion component of current is usually obtained from Fick's first law as $qD_n\,\nabla_r n$, where D_n is the diffusion constant of the electrons. In comparison with (5.87), we find that

$$D_n = \frac{kT}{q}\mu_n\frac{F_{1/2}(\eta)}{F_{-1/2}(\eta)} \tag{5.88}$$

Equation (5.88) is the Einstein relationship between the diffusion coefficient and the conductivity mobility. Although this relationship is easily derived for an equilibrium condition where the total current density is zero, the approach we have taken shows that Einstein's relation is also valid under nonequilibrium conditions.

Following similar arguments for valence band holes, the current density is

$$\mathbf{J} = qp\mu_p\mathbf{E} - kT\mu_p\frac{F_{1/2}(\eta)}{F_{-1/2}(\eta)}\nabla_r p \tag{5.89}$$

where μ_p is the conductivity mobility for holes. Notice that the diffusion component of hole current is opposite to the direction of the hole gradient. From Fick's law the diffusion coefficient for holes is also of the form of (5.88). When the Fermi energy is in the energy gap at least $4kT$ removed from either band edge, (5.88) reduces to

$$D_n = \frac{kT}{q}\mu_n \tag{5.90}$$

The equations we have derived for the mobility (5.50), conductivity (5.52), and Hall constant (5.72) are applicable for electrons in spherical conduction band minima. When the electrons transport charge in an ellipsoidal minimum, the situation is somewhat more complicated. Consider one ellip-

soidal conduction band minimum at Γ given by (5.30). If the x direction is taken as one of the axes of the constant energy ellipsoids, an electric field in the x direction will produce a current in the x direction,

$$J_x = \frac{q^2 n \langle \tau_m \rangle}{m_1^*} E_x \tag{5.91}$$

Similar expressions containing m_2^* and m_3^* are obtained in the y and z directions. The total current density is therefore

$$\mathbf{J} = q^2 n \langle \tau_m \rangle \mathbf{M} \cdot \mathbf{E} \tag{5.92}$$

where the effective mass tensor $\mathbf{M}$ is given by (5.31). This can also be put in the form

$$\mathbf{J} = \boldsymbol{\sigma} \cdot \mathbf{E} \tag{5.93}$$

where $\boldsymbol{\sigma}$ is a conductivity tensor and $\mathbf{E}$ is a column vector. Thus, for an ellipsoidal minimum at Γ, the current density is not necessarily in the same direction as the applied electric field.

When there are g_c equivalent ellipsoidal conduction band minima, it is necessary to account for the fact that the concentration of electrons in each minimum is n/g_c. In this case the current density is obtained by summing the concentration of electrons in each minimum, while allowing for the effective mass each minimum has in the direction of the current. For a semiconductor with conduction band minima in the direction of X, such as Si, this is relatively easy. As shown in Fig. 5.1, when the current is in the x direction, the two minima along the k_x axis each contribute $n/6$ electrons with effective mass m_1^*, while the two minima along the k_y axis contribute $n/6$ electrons each with effective mass m_2^*. In the third dimension, the two minima along the k_z axis also contribute $n/6$ electrons of mass m_3^*. The total

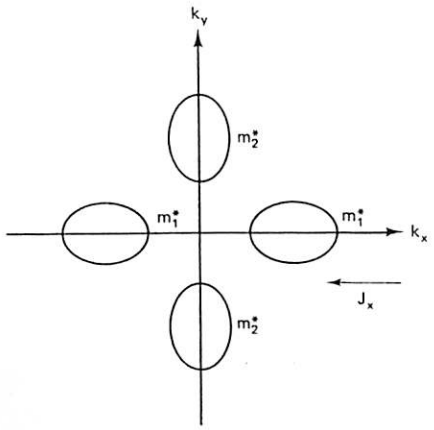


Figure 5.1 Diagram showing how equivalent ellipsoidal minima contribute to conduction along one of their principal axes.

current density in the x direction is therefore

$$J_x = \frac{q^2 n \langle \tau_m \rangle}{6} \left(\frac{2}{m_1^*} + \frac{2}{m_2^*} + \frac{2}{m_3^*} \right) E_x \tag{5.94}$$

Similar expressions are obtained for the components of current in the y and z directions. Thus the total current can be put in the form

$$\mathbf{J} = \frac{q^2 n \langle \tau_m \rangle}{m_c} \mathbf{E} \tag{5.95}$$

where

$$\frac{1}{m_c} \equiv \frac{1}{3} \left(\frac{1}{m_1^*} + \frac{1}{m_2^*} + \frac{1}{m_3^*} \right) \tag{5.96}$$

defines the *conductivity effective mass*.

Equation (5.96) is also valid for semiconductors, such as Ge, which have equivalent minima in directions other than X. Notice that the conductivity effective mass defined by (5.96) is a scalar, so that the current density and the electric field are always in the same direction. Comparing (5.95) with (5.92), we see that equivalent ellipsoidal minima result in isotropic conductivity, while an ellipsoidal minimum at Γ produces anisotropic conductivity. This difference due to the position of ellipsoidal minima is reflected in the conductivity of sphalerite crystals with indirect bandgaps as compared to wurtzite crystals with direct bandgaps.

When the Hall effect in ellipsoidal minima is examined, similar results are obtained. That is, an ellipsoidal conduction band minimum at Γ produces an anisotropic Hall effect, while equivalent minima produce an isotropic Hall effect. In the latter case, for simplicity, a *Hall effective mass* can be defined to reduce the expression for the Hall mobility to the form of (5.74).

5.4 CHARGE AND ENERGY TRANSPORT

To determine the heat flow density, (5.37) tells that we must find the average

$$\langle e\mathbf{v} \rangle = \frac{\int_{-\infty}^{\infty} e\mathbf{v} f \, d\mathbf{v}}{\int_{-\infty}^{\infty} f \, d\mathbf{v}} \tag{5.97}$$

where e is heat content per electron. This equation has the same form as (5.38) except that, in this case, we must include e in the average since e depends on the electron energy. We have already solved this problem with (5.60), so in the same way we obtained (5.79) we can simply write

$$\mathbf{W} = \frac{-n}{m^*} \langle e\mathbf{G} \rangle \tag{5.98}$$

where the minus sign indicates that **W** is opposite in direction to **G** and thus **J** for electrons. It is now necessary to obtain e in terms of $\mathscr{E}$, so that the average in (5.98) can be determined.

Since heat is that portion of the total electron energy which can be added or removed in disordered form, the heat content of all the electrons in the system is given by the entropy term in Euler's equation (4.28). Thus the heat content per electron is

$$e = \left[\frac{\partial}{\partial n'}(TS)\right]_{T,V} \tag{5.99}$$

From (4.29) for the Helmholtz function,

$$TS = \mathscr{E}' - F \tag{5.100}$$

where $\mathscr{E}'$ is the total energy of all the electrons. Using (5.100) in (5.99), we have

$$e = \left(\frac{\partial \mathscr{E}'}{\partial n'}\right)_{T,V} - \left(\frac{\partial F}{\partial n'}\right)_{T,V} \tag{5.101}$$

The first term in (5.101) is simply the energy per electron, $\mathscr{E}$. From (4.39) the second term is simply the chemical potential, μ. The heat content per electron is therefore

$$e = \mathscr{E} - \mu \tag{5.102}$$

Let us now obtain the equations that determine the transport of charge and energy under the following conditions. We will assume that the electric and magnetic fields are small and allow for temperature and concentration gradients. Also, for simplicity we will assume spherical energy band extrema. From (5.79) and (5.34) the equation for the electron current density is

$$\mathbf{J} = \frac{+q^2 n}{m^*}\left[\langle \tau_m \mathscr{F}\rangle - \frac{q}{m^*}\langle \tau_m^2(\mathscr{F} \times \mathbf{B})\rangle\right] \tag{5.103}$$

Using (5.98), (5.102), and (5.34), the electron heat current density is

$$\mathbf{W} = \frac{-qn}{m^*}\left[\langle \tau_m e\mathscr{F}\rangle - \frac{q}{m^*}\langle \tau_m^2 e(\mathscr{F} \times \mathbf{B})\rangle\right] \tag{5.104}$$

where from (5.28),

$$\mathscr{F} = +\mathbf{E} - \frac{1}{q}\nabla_r e + \frac{e}{qT}\nabla_r T \tag{5.105}$$

We will use (5.103), (5.104), and (5.105) to examine various effects that involve the transport of charge and energy in semiconductors.

5.4.1 Thermal Conductivity

One of the more important thermal transport parameters is the thermal conductivity. Although in lightly doped semiconductors most of the heat is carried by lattice vibrations or phonons, in heavily doped semiconductors a substantial proportion is carried by electrons. The thermal conductivity, κ, is defined as the proportionality factor between the heat current density and the temperature gradient,

$$\mathbf{W} = -\kappa \nabla_r T \tag{5.106}$$

The minus sign is required because the heat flows from higher to lower temperatures. To determine the thermal conductivity, we examine a sample under open-circuit conditions ($\mathbf{J} = 0$) with no magnetic field. For small temperature gradients, (5.103) is

$$0 = +\langle \tau_m\rangle\left(\mathbf{E} + \frac{1}{q}\nabla_r \mu\right) + \frac{1}{qT}\langle \tau_m e\rangle \nabla_r T \tag{5.107}$$

and (5.104) is

$$\mathbf{W} = \frac{-qn}{m^*}\left[\langle \tau_m e\rangle\left(\mathbf{E} + \frac{1}{q}\nabla_r \mu\right) + \frac{1}{qT}\langle \tau_m e^2\rangle \nabla_r T\right] \tag{5.108}$$

Using (5.107) in (5.108) yields

$$\mathbf{W} = \frac{-n}{m^* T}\left[\langle \tau_m e^2\rangle - \frac{\langle \tau_m e\rangle^2}{\langle \tau_m\rangle}\right]\nabla_r T \tag{5.109}$$

and the thermal conductivity due to electrons is

$$\kappa = \frac{n}{m^* T}\left[\langle \tau_m e^2\rangle - \frac{\langle \tau_m e\rangle^2}{\langle \tau_m\rangle}\right] \tag{5.110}$$

Numerical values for the averages in (5.110) can be obtained with (5.60).

5.4.2 Thermoelectric Effects

We can see from (5.107) that under open-circuit conditions, the electrons diffuse down the temperature gradient and set up an electric field that opposes the motion of electrons due to the gradient. The production of an electric field by a temperature gradient is referred to as the *Seebeck* or *thermoelectric* effect. In the steady state the electric field is given by (5.107) as

$$\mathbf{E} = -\frac{1}{q}\nabla_r \mu - \frac{1}{q\langle \tau_m\rangle T}\langle \tau_m e\rangle \nabla_r T \tag{5.111}$$

Using

$$\nabla_r \mu = \frac{\partial \mu}{\partial T} \nabla_r T \tag{5.112}$$

(5.111) gives us

$$\mathbf{E} = -\frac{1}{qT}\left[T\frac{\partial \mu}{\partial T} + \frac{\langle \tau_m e \rangle}{\langle \tau_m \rangle}\right] \nabla_r T \tag{5.113}$$

or

$$\mathbf{E} = T\frac{d}{dT}\left[\frac{\langle \tau_m e \rangle}{qT\langle \tau_m \rangle}\right] \nabla_r T \tag{5.114}$$

Thus the electric field is related to the temperature gradient by means of the equation

$$\mathbf{E} = \mathcal{T} \nabla_r T \tag{5.115}$$

where

$$\mathcal{T} \equiv T\frac{d}{dT}\left[\frac{\langle \tau_m e \rangle}{qT\langle \tau_m \rangle}\right] \tag{5.116}$$

is the *Thomson coefficient*.

As indicated in Fig. 5.2, the Seebeck effect can be examined by measuring the voltage across a semiconductor in a temperature gradient. The voltage is given by

$$\begin{aligned} V &= -\oint \mathbf{E}\cdot d\mathbf{r} = \oint \mathcal{T} \nabla_r T \cdot d\mathbf{r} \\ V &= \int_{T_0}^{T_1} \mathcal{T}_m \, dT + \int_{T_1}^{T_2} \mathcal{T}_s \, dT + \int_{T_2}^{T_0} \mathcal{T}_m \, dT \\ V &= \int_{T_1}^{T_2} (\mathcal{T}_s - \mathcal{T}_m) \, dT \end{aligned} \tag{5.117}$$

where $\mathcal{T}_s$ and $\mathcal{T}_m$ are the Thomson coefficients for the semiconductor and

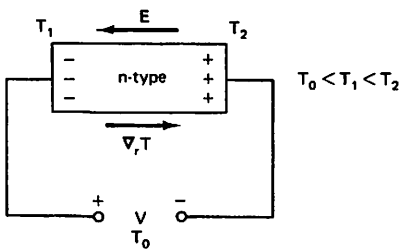


Figure 5.2 Determination of Seebeck effect for *n*-type semiconductor. *V* is negative.

metal, respectively. If the metal is chosen to have $\mathcal{T}_m = 0$, the Seebeck voltage is directly related to the Thomson coefficient of the semiconductor and the temperature difference across the sample. Notice that the Thomson coefficient for electrons is negative, as indicated by (5.115). For holes under the same conditions, the Thomson coefficient is positive, so that the field and temperature gradients are in the same direction.

A transport parameter closely related to the Thomson coefficient is the absolute *thermoelectric power*, $\mathcal{P}$. The relationship is

$$\mathcal{T} = -T\frac{d}{dT}\mathcal{P} \tag{5.118}$$

or from (5.116),

$$\mathcal{P} \equiv -\frac{\langle \tau_m e \rangle}{qT\langle \tau_m \rangle} \tag{5.119}$$

Notice that the thermoelectric power for electrons and holes have opposing signs, due to the dependence on q. Because of this the sign of the thermoelectric power can be used to determine whether a material exhibits *n*- or *p*-type conductivity.

When the electric current density is not constrained to be zero, it adds an additional component to the heat current density. Under these conditions (5.103) can be written in the form

$$\mathbf{E} + \frac{1}{q}\nabla_r \mu = \frac{m^*}{q^2 n\langle \tau_m \rangle}\mathbf{J} - \frac{1}{qT}\frac{\langle \tau_m e \rangle}{\langle \tau_m \rangle}\nabla_r T \tag{5.120}$$

Substituting this into (5.108), the heat current density is

$$\mathbf{W} = \frac{-\langle \tau_m e \rangle}{q\langle \tau_m \rangle}\mathbf{J} - \frac{n}{m^* T}\left[\langle \tau_m e^2 \rangle - \frac{\langle \tau_m e \rangle^2}{\langle \tau_m \rangle}\right]\nabla_r T \tag{5.121}$$

Using (5.110) and (5.119), this is simply

$$\mathbf{W} = T\mathcal{P}\mathbf{J} - \kappa \nabla_r T \tag{5.122}$$

Thus the electric current density carries heat in addition to that transported by the temperature gradient. This is referred to as the *Peltier effect*. The constant of proportionality between heat current density and electric current density is the Peltier coefficient, Π, where

$$\Pi \equiv T\mathcal{P} \tag{5.123}$$

$$\Pi = -\frac{\langle \tau_m e \rangle}{q\langle \tau_m \rangle} \tag{5.124}$$

Because of the dependence on q, the Peltier coefficient is negative for electrons and positive for holes.

In addition to transporting heat, the electric current density also generates heat. The net rate at which heat is generated per unit volume is equal to the rate generated per unit volume minus the rate at which it is transported away, or

$$P = \mathbf{J}\cdot\mathbf{E} - \nabla_r\cdot\mathbf{W} \tag{5.125}$$

From (5.120), (5.52), and (5.115),

$$\mathbf{E} = \frac{1}{\sigma}\mathbf{J} + \mathscr{T}\,\nabla_r T \tag{5.126}$$

Also, from (5.122) and (5.123),

$$\mathbf{W} = \Pi\mathbf{J} - \kappa\,\nabla_r T \tag{5.127}$$

Using (5.126) and (5.127) in (5.125), we have

$$P = \frac{\mathbf{J}\cdot\mathbf{J}}{\sigma} + \mathscr{T}\mathbf{J}\cdot\nabla_r T - \Pi\,\nabla_r\cdot\mathbf{J} + \kappa\,\nabla_r\cdot\nabla_r T \tag{5.128}$$

Thus the net rate at which heat is generated per unit volume, P, has several components. The first term in (5.128) is simply the Joule heat. The second term is referred to as the *Thomson heat*. The third term, which involves the divergence of the electric current, allows for the generation or recombination of electrons in the unit volume and will not be considered further. Finally, the last term provides for the transport of heat out of the volume by thermal conduction.

Notice that the Thomson heat term in (5.128) changes sign when either the current density or the temperature gradient is reversed. Since the Thomson coefficient is negative for electrons, in n-type material heating is produced when $\mathbf{J}$ and $\nabla_r T$ are in the same direction. That is, the electrons going from a higher to a lower temperature have to give heat to the lattice. When $\mathbf{J}$ and $\nabla_r T$ are in opposite directions, the electrons produce cooling since they take heat from the lattice in going from a lower to a higher temperature. These effects are indicated schematically in Fig. 5.3. Since the Thomson coefficient for holes is positive, cooling is produced when $\mathbf{J}$ and $\nabla_r T$ are in

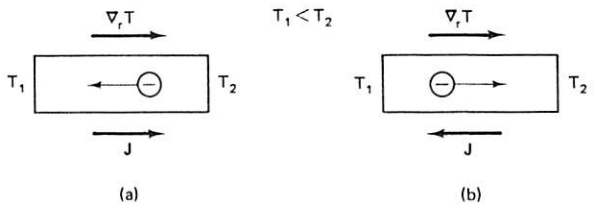


Figure 5.3 The Thomson term in an n-type semiconductor produces (a) heating when $\mathbf{J}$ and $\nabla_r T$ are in the same direction and (b) cooling when they are in opposite directions.

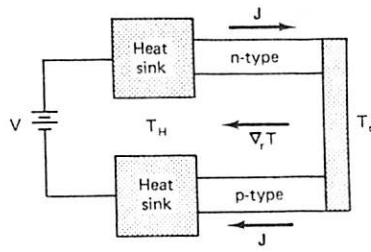


Figure 5.4 Schematic of a thermoelectric cooler. The heat sinks and cold junctions are metals that form ohmic contacts.

the same direction. These results can be used to construct a thermoelectric cooler in the manner shown in Fig. 5.4.

Similar effects are obtained for electrons in concentration gradients. In the absence of a temperature gradient, the electric field for n-type material is given from (5.87) as

$$\mathbf{E} = \frac{\mathbf{J}}{\sigma} - \frac{kT}{qN_c F_{-1/2}(\eta)}\nabla_r n \tag{5.129}$$

In this case, (5.125) is

$$P = \frac{\mathbf{J}\cdot\mathbf{J}}{\sigma} - \frac{kT}{qN_c F_{-1/2}(\eta)}\mathbf{J}\cdot\nabla_r n - \Pi\,\nabla_r\cdot\mathbf{J} \tag{5.130}$$

Thus, when $\mathbf{J}$ and $\nabla_r n$ are in the same direction, the electrons take heat from the lattice as they go from higher to lower concentrations and cooling is produced.

5.4.3 Thermomagnetic Effects

When we allow for a small magnetic field, in addition to small electric fields and temperature gradients, the transport of heat produces several thermomagnetic effects. We examine these in the Hall configuration shown in Fig. 5.5. In Section 5.3 we examined the Hall effect assuming that no tem-

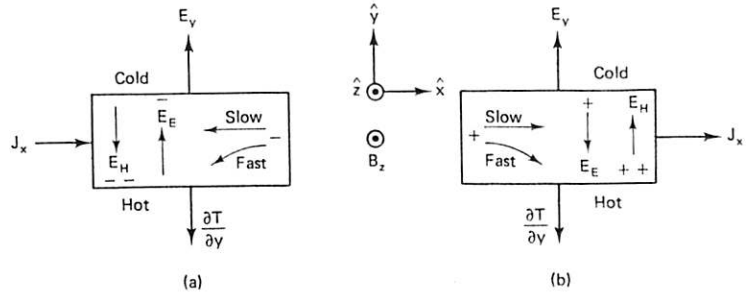


Figure 5.5 Hall and Ettinghausen effects ($0 \le r$) for (a) electrons and (b) holes.

perature gradients were present (isothermal conditions) and that there was no electric current in the $\hat{y}$ direction. That is, $J_y = \partial T/\partial x = \partial T/\partial y = 0$. Under these conditions, the electrons were deflected by the magnetic field in the direction shown in Fig. 5.5(a) and an electric field, E_H, was induced in the negative $\hat{y}$ direction to balance the Lorentz force. However, this Hall field can only exactly balance the Lorentz force on electrons with average velocity. If we assume that the momentum relaxation time increases with energy ($\tau_m = \tau_0 x^r$, where $0 \leq r$), the faster or hotter electrons are deflected more and the slower ones less by the magnetic field. As a result, the side of the sample where the faster carriers are deflected becomes warmer and the opposite side cooler, inducing a temperature gradient. In a manner similar to the thermoelectric effect, the warmer electrons tend to diffuse to the cooler surface, where they set up an electric field, as in (5.115), to oppose the diffusion. The mechanism that produces this electric field is referred to as the *Ettinghausen effect*.

The Ettinghausen coefficient is defined under conditions such that no heat current is transferred to the surroundings (adiabatic conditions). For $J_y = \partial T/\partial x = W_y = 0$, this coefficient is

$$P_E \equiv \frac{\partial T/\partial y}{J_x B_z} \tag{5.131}$$

Applying these conditions to (5.103) and (5.104), we can eliminate the electric field and the chemical potential gradient,

$$P_E = \frac{\mu_c}{q\kappa}\left[\frac{\langle \tau_m^2 \mathscr{E}\rangle}{\langle \tau_m\rangle^2} - \frac{\langle \tau_m^2\rangle\langle \tau_m \mathscr{E}\rangle}{\langle \tau_m\rangle^3}\right] \tag{5.132}$$

From this equation we expect the Ettinghausen field to change sign when the sign of the charge carrier is reversed. Figure 5.5(b) for holes shows that this is, indeed, what occurs. In either case, however, the direction of the Ettinghausen field depends on the energy dependence of the momentum relaxation time. It can be verified by (5.132) that when r in (5.55) is less than zero, the slower carriers are deflected more than the faster ones and the direction of the Ettinghausen field in Fig. 5.5 are reversed. Thus the direction of the Ettinghausen field depends on the sign of the carrier and the scattering mechanism.

Two other thermomagnetic effects we will mention are the *Nernst* and *Righi–Leduc effects*. As indicated in Fig. 5.6, these effects are the thermal analogues of the Hall and Ettinghausen effects, respectively. The Nernst coefficient is defined under isothermal conditions ($J_x = J_y = \partial T/\partial y = 0$) as

$$Q_N \equiv \frac{E_y}{(\partial T/\partial x)B_z} \tag{5.133}$$

Thus we see that the Nernst effect is a process whereby a transverse electric

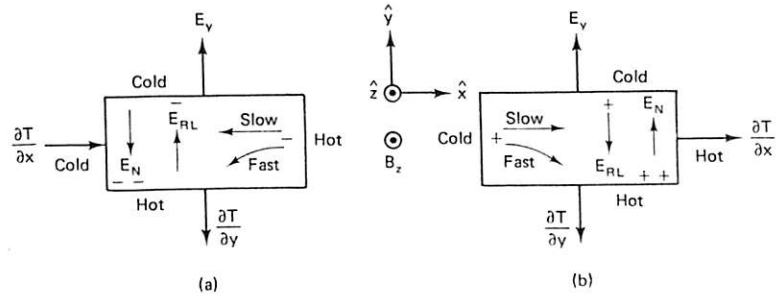


Figure 5.6 Nernst and Righi–Leduc effects ($0 \leq r$) for (a) electrons and (b) holes.

field is produced by a temperature gradient. This equation can be compared to (5.72) for the Hall constant. Under these conditions (5.103) and (5.104) give us

$$Q_N = \frac{\mu_c}{qT}\left[\frac{\langle \tau_m^2 \mathscr{E}\rangle}{\langle \tau_m\rangle^2} - \frac{\langle \tau_m^2\rangle\langle \tau_m \mathscr{E}\rangle}{\langle \tau_m\rangle^3}\right] \tag{5.134}$$

or

$$Q_N = \frac{\kappa}{T} P_E \tag{5.135}$$

The Righi–Leduc coefficient is defined under adiabatic conditions ($J_x = J_y = W_y = 0$) as

$$S_{RL} \equiv \frac{\partial T/\partial y}{(\partial T/\partial x)B_z} \tag{5.136}$$

which gives

$$S_{RL} = \frac{n\mu_c^2}{qT\kappa}\left[\frac{\langle \tau_m^2 \mathscr{E}^2\rangle}{\langle \tau_m^2\rangle} + \frac{\langle \tau_m^2\rangle\langle \tau_m \mathscr{E}\rangle^2}{\langle \tau_m\rangle^4} - \frac{2\langle \tau_m^2 \mathscr{E}\rangle\langle \tau_m \mathscr{E}\rangle}{\langle \tau_m\rangle^3}\right] \tag{5.137}$$

Arguments regarding the sign of these coefficients are similar to those made for the sign of the Ettinghausen coefficient.

5.5 HIGH-FREQUENCY TRANSPORT

The dc theory of charge and energy transport developed in this chapter can be applied to transport at high frequencies with only slight modification. Let us apply a small sinusoidal electric field,

$$\mathbf{E} = \mathbf{E}_0 \exp(-i\omega t) \tag{5.138}$$

to a semiconductor sample where ω is the angular frequency of the field.

Assuming no other applied forces and no temperature or concentration gradients, we can examine the form of the nonequilibrium distribution function. In the same manner in which (5.15) was obtained, we have

$$f = f_0 + q\tau_m \frac{\partial f_0}{\partial \mathscr{E}} \mathbf{v}\cdot\mathbf{E}_0 \exp(-i\omega t) \tag{5.139}$$

The time-dependent Boltzmann equation in the relaxation time approximation from (5.4), (5.5), and (5.9) is

$$\frac{\partial f}{\partial t} = -\mathbf{v}\cdot\left(\frac{\partial f}{\partial \mathscr{E}}\mathbf{F} + \nabla_r f\right) - \frac{f - f_0}{\tau_m} \tag{5.140}$$

Taking the time derivative of (5.139), we have

$$\frac{\partial f}{\partial t} = +q\tau_m \frac{\partial f_0}{\partial \mathscr{E}}\left(-i\omega\mathbf{v} + \frac{d\mathbf{v}}{dt}\right)\cdot\mathbf{E}_0 \exp(-i\omega t) \tag{5.141}$$

Since

$$\frac{d\mathbf{v}}{dt} = -q\mathbf{M}\cdot\mathbf{E}_0 \exp(-i\omega t) \tag{5.142}$$

the last term in (5.141) is second order in E and can be neglected. We now have

$$\frac{\partial f}{\partial t} = -i\omega(f - f_0) \tag{5.143}$$

and the Boltzmann equation is

$$f - f_0 = \frac{-\tau_m}{(1 - i\omega\tau_m)}\mathbf{v}\cdot\left(\frac{\partial f}{\partial \mathscr{E}}\mathbf{F} + \nabla_r f\right) \tag{5.144}$$

Except for the term $(1 - i\omega\tau_m)$, (5.144) has the same form as (5.11). Because of this, all of the dc transport equations can be used at high frequency if we replace τ_m by τ_m^* where

$$\tau_m^* \equiv \frac{\tau_m}{1 - i\omega\tau_m} \tag{5.145}$$

The general solution of the Boltzmann equation (5.144) for ellipsoidal minima is

$$f = f_0 + \frac{\partial f_0}{\partial \mathscr{E}} q\tau_m^* \mathbf{v}\cdot\left[\frac{\mathscr{F} - q\tau_m^*\mathbf{M}\cdot(\mathscr{F}\times\mathbf{B}) + (q\tau_m^*)^2(\det\mathbf{M})(\mathscr{F}\cdot\mathbf{B})(\mathbf{M}^{-1}\cdot\mathbf{B})}{1 + (q\tau_m^*)^2(\det\mathbf{M})(\mathbf{M}^{-1}\cdot\mathbf{B})\cdot\mathbf{B}}\right] \tag{5.146}$$

and this is used to determine the transport parameters. Since τ_m^* is a complex

number, all the terms in the distribution function (except for f_0) are complex. Thus, in general, the high-frequency components of electric current density and heat current density are not in phase with the applied forces.

A useful effect can be observed when we examine the denominator of the last term in (5.146). If we define a cyclotron frequency as in (5.65) by

$$\omega_c^2 \equiv q^2(\det\mathbf{M})(\mathbf{M}^{-1}\cdot\mathbf{B})\cdot\mathbf{B} \tag{5.147}$$

the denominator becomes $1 + (\omega_c\tau_m^*)^2$. Using (5.145) gives

$$1 + (\omega_c\tau_m^*)^2 = 1 + \frac{(\omega_c\tau_m)^2}{(1 - i\omega\tau_m)^2} \tag{5.148}$$

which can be put in the form

$$1 + (\omega_c\tau_m^*)^2 = \frac{(\omega_c^2 - \omega^2)\tau_m^2 - 2i\omega\tau_m + 1}{(1 - i\omega\tau_m)^2} \tag{5.149}$$

Equation (5.149) tells us that for conditions such that $\omega\tau_m$ is much greater than 1, $1 + (\omega_c\tau_m^*)^2$ exhibits a sharp minimum when $\omega = \omega_c$. Thus the transport properties, such as charge current density, J, will exhibit a resonant peak when the applied frequency is equal to the cyclotron frequency. This effect is referred to as *cyclotron resonance*. From (5.65) or (5.147) the frequency at which cyclotron resonance occurs can be used to determine effective mass.

5.6 HIGH ELECTRIC FIELD EFFECTS

Up to now we have limited our analysis of transport properties to small electric fields. Under these conditions the energy the carrier distribution gains from the electric field is lost to the lattice through collisions with low-energy acoustic phonons or impurities. The average energy of the electrons, therefore, remains close to the thermal equilibrium value, $\frac{3}{2}kT$, and the drift velocity of the distribution is linearly related to the electric field. However, because the average electron energy in semiconductors is so small, it is relatively easy to obtain significant deviations from this ohmic behavior. For moderate electric fields the collisions with acoustic phonons and impurities, which serve to maintain the electron distribution and the lattice at the same temperature, become less effective and the electrons gain energy from the field faster than they can lose it to the lattice. In this situation, the electron distribution can be characterized by an effective temperature, T_e, which is "hotter" than the lattice temperature, T. The relationship between the drift velocity and electric field is no longer linear and nonohmic electrical behavior is observed.

When the electrons have gained sufficient energy from the field, they can transfer energy to the lattice by the generation of high energy optical

phonons. Since this process is an efficient energy loss mechanism for the electrons, the drift velocity of the distribution reaches a limiting value where it no longer increases with the electric field. This value is referred to as the *saturated drift velocity* and is obtained for electron energies of the order of the optical phonon frequencies given in Table 3.2. Figure 5.7(a) and (b) show experimental drift velocity–electric field characteristics of electrons and holes, respectively, for Ge, Si, and GaAs up to the saturated regions. As can be seen, all of these characteristics, except for electrons in GaAs, exhibit, qualitatively, the nonohmic behavior described above. The negative resistance region in the GaAs $\mathbf{v}_d$ versus $\mathbf{E}$ curve is due to the transfer of electrons from the Γ to L conduction band minima. This forms the basis of the Gunn effect.

Since these *hot electron effects* play an important role in the operation of several semiconductor devices, let us examine them in more detail. Although there are several approaches to the problem of high electric field transport in semiconductors [Esther M. Conwell, *High Field Transport in Semiconductors* (New York: Academic Press, 1967)], we will use an approach with which analytical results can readily be obtained. Let us assume that the electron distribution can be described by

$$f = \frac{1}{1 + \exp\left[(\mathscr{E} - \zeta)/kT_e\right]} \tag{5.150}$$

where T_e is the effective temperature of the electron distribution. Notice that this distribution function has the form of the equilibrium distribution

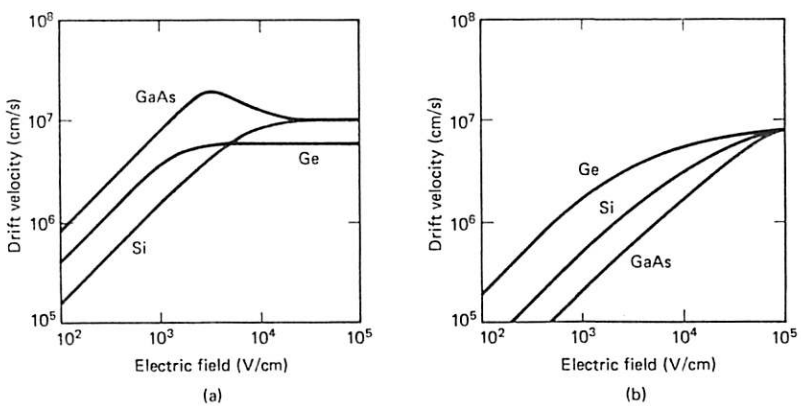


Figure 5.7 Experimental drift velocity–electric field characteristics for (a) electrons and (b) holes in several semiconductors. [After C. B. Norris and J. F. Gibbons, *IEEE Trans. Electron Devices ED-14*, 38 (1967); C. Y. Duk and J. L. Moll, *IEEE Trans. Electron Devices ED-14*, 46 (1967); T. E. Seidel and D. L. Scharfetter, *J. Phys. Chem. Solids 28*, 2563 (1967); J. G. Ruch and G. S. Kino, *Appl. Phys. Lett. 10*, 40 (1967); V. L. Dalal, *Appl. Phys. Lett. 16*, 489 (1970).]

function, f_0, derived in (4.41), except for T_e, which is a function of electric field, $\mathbf{E}$. We will also maintain the form of the low-field relationship between the drift velocity and the electric field (5.49).

$$\mathbf{v}_d = -\mu(T_e)\mathbf{E} \tag{5.151}$$

where the mobility, μ, is now a function of T_e or electric field. In a similar manner,

$$\mu(T_e) = \frac{q\langle\tau_m(T_e)\rangle}{m^*} \tag{5.152}$$

where

$$\langle\tau_m(T_e)\rangle = \frac{\int_{-\infty}^{\infty} \tau_m(T_e) f \, d\mathbf{v}}{\int_{-\infty}^{\infty} f \, d\mathbf{v}} \tag{5.153}$$

and f is given by (5.150). Under these assumptions, we can proceed with our examination of high-field transport.

The average over velocity space of the momentum relaxation time in (5.153) can be simplified to an average over energy

$$\langle\tau_m(T_e)\rangle = \tau_0(T_e) \frac{\int_0^{\infty} f x_e^{r+1/2} \, dx_e}{\int_0^{\infty} f x_e^{1/2} \, dx_e} \tag{5.154}$$

or, using (5.59),

$$\langle\tau_m(T_e)\rangle = \frac{2}{\sqrt{\pi}} \left(r + \frac{1}{2}\right)! \, \tau_0(T_e) \frac{F_{r+1/2}(\eta_e)}{F_{1/2}(\eta_e)} \tag{5.155}$$

where

$$x_e \equiv \frac{\mathscr{E} - \mathscr{E}_c}{kT_e} \quad \text{and} \quad \eta_e \equiv \frac{\zeta - \mathscr{E}_c}{kT_e} \tag{5.156}$$

Equation (5.155) tells us that, in general, the appropriate average is

$$\langle\tau_m^s(T_e)x_e^t\rangle = \frac{2}{\sqrt{\pi}} \left(sr + t + \frac{1}{2}\right)! \, \tau_0^s(T_e) \frac{F_{sr+t+1/2}(\eta_e)}{F_{1/2}(\eta_e)} \tag{5.157}$$

To proceed further we need to know the dependence of τ_0 on T_e. This dependence, of course, depends on the particular scattering mechanism. For several scattering mechanisms, the low-field momentum relaxation time can be put in the form

$$\tau_m = \tau_0 x^r = CT^u x^r \tag{5.158}$$

where C is independent of temperature. If we assume that the high-field momentum relaxation time has the same form,

$$\tau_m(T_e) = \tau_0(T_e)x_e^r = CT_e^u x_e^r \tag{5.159}$$

then

$$\tau_m(T_e) = \tau_0\left(\frac{T_e}{T}\right)^u x_e^r \tag{5.160}$$

From (5.152), (5.155), and (5.160) the field-dependent mobility is

$$\mu(T_e) = \frac{q\tau_0}{m^*}\left(\frac{T_e}{T}\right)^u \frac{2}{\sqrt{\pi}}\left(r + \frac{1}{2}\right)! \frac{F_{r+1/2}(\eta_e)}{F_{1/2}(\eta_e)} \tag{5.161}$$

or simply

$$\mu(T_e) = \mu_0\left(\frac{T_e}{T}\right)^u \tag{5.162}$$

Thus if we can determine the dependence of T_e on $\mathbf{E}$, we can obtain the dependence of μ and, from (5.151), the dependence of $\mathbf{v}_d$ on $\mathbf{E}$.

The effective electron temperature can be determined from conservation of energy. From (5.4) and (5.6) the time-dependent Boltzmann equation is

$$\frac{\partial f}{\partial t} = \frac{+q}{\hbar}\mathbf{E}\cdot\nabla_k f - \mathbf{v}\cdot\nabla_r f - \frac{f - f_0}{\tau_m} \tag{5.163}$$

If we multiply each term of this equation by the electron energy, $\mathscr{E}$, and average it over the electron distribution, we obtain

$$\frac{\int_{-\infty}^{\infty}\mathscr{E}\frac{\partial f}{\partial t}\,d\mathbf{v}}{\int_{-\infty}^{\infty} f\,d\mathbf{v}} = \frac{\frac{+q}{\hbar}\mathbf{E}\cdot\int_{-\infty}^{\infty}\mathscr{E}\,\nabla_k f\,d\mathbf{v}}{\int_{-\infty}^{\infty} f\,d\mathbf{v}} - \frac{\int_{-\infty}^{\infty}\mathscr{E}\mathbf{v}\cdot\nabla_r f\,d\mathbf{v}}{\int_{-\infty}^{\infty} f\,d\mathbf{v}} - \frac{\int_{-\infty}^{\infty}\mathscr{E}\frac{f - f_0}{\tau_m}\,d\mathbf{v}}{\int_{-\infty}^{\infty} f\,d\mathbf{v}} \tag{5.164}$$

With (2.109) and (5.38), the first term on the right-hand side can be reduced to $q\mathbf{v}_d\cdot\mathbf{E}$. Neglecting the diffusion term, (5.164) then reduces to the form

$$\frac{d}{dt}\langle\mathscr{E}\rangle = q\mathbf{v}_d\cdot\mathbf{E} - \frac{\langle\mathscr{E}\rangle - \langle\mathscr{E}_0\rangle}{\tau_e} \tag{5.165}$$

where we have defined an *energy relaxation time*, τ_e, by

$$\frac{\langle\mathscr{E}\rangle - \langle\mathscr{E}_0\rangle}{\tau_e} \equiv \frac{\int_{-\infty}^{\infty}\mathscr{E}\left(\frac{f - f_0}{\tau_m}\right)d\mathbf{v}}{\int_{-\infty}^{\infty} f\,d\mathbf{v}} \tag{5.166}$$

Equation (5.165) is the energy balance equation for the hot electron distribution. It tells us that the net energy gained per unit time is equal to the power supplied by the electric field minus the energy lost to collisions. Under steady-state conditions the electron distribution attains a temperature T_e and

$$\frac{\langle\mathscr{E}\rangle - \langle\mathscr{E}_0\rangle}{\tau_e} = q\mathbf{v}_d\cdot\mathbf{E} \tag{5.167}$$

Notice from (5.165) that τ_e characterizes the relaxation of the hot electron distribution to its average thermal equilibrium energy $\langle\mathscr{E}_0\rangle$ when the electric field is turned off. We see from (5.166) that, since τ_m is in general a function of $\mathscr{E}$, τ_m and τ_e are not equal. That is, the times that characterize the relaxation of energy and momentum are different.

From (5.156) and (5.157) the average energy of the hot electron distribution is

$$\langle\mathscr{E}\rangle = kT_e\langle x_e\rangle = \frac{3}{2}kT_e\frac{F_{3/2}(\eta_e)}{F_{1/2}(\eta_e)} \tag{5.168}$$

In a similar manner the average energy of the equilibrium distribution is

$$\langle\mathscr{E}_0\rangle = kT\langle x\rangle = \frac{3}{2}kT\frac{F_{3/2}(\eta)}{F_{1/2}(\eta)} \tag{5.169}$$

Using (5.156) and (5.166) yields

$$\frac{\langle\mathscr{E}\rangle}{\tau_e} \equiv \left\langle\frac{\mathscr{E}}{\tau_m}\right\rangle = kT_e\left\langle\frac{x_e}{\tau_m}\right\rangle \tag{5.170}$$

The average in this equation can be evaluated with (5.157) and (5.160) as

$$\frac{\langle\mathscr{E}\rangle}{\tau_e} = \frac{2}{\sqrt{\pi}}\left(\frac{3}{2} - r\right)!\frac{kT_e}{\tau_0}\left(\frac{T}{T_e}\right)^u\frac{F_{3/2-r}(\eta_e)}{F_{1/2}(\eta_e)} \tag{5.171}$$

and

$$\tau_e = \frac{3\sqrt{\pi}}{4}\frac{\tau_0}{(3/2 - r)!}\left(\frac{T_e}{T}\right)^u\frac{F_{3/2}(\eta_e)}{F_{3/2-r}(\eta_e)} \tag{5.172}$$

Under nondegenerate conditions, we can readily obtain an expression

relating T_e to E. Using (5.151), (5.161), and (5.171) in (5.167), we have

$$\left(\frac{3}{2} - r\right)! \frac{k}{\tau_0} \left(\frac{T}{T_e}\right)^u (T_e - T) = \left(r + \frac{1}{2}\right)! \frac{q^2\tau_0}{m^*} \left(\frac{T_e}{T}\right)^u E^2 \quad (5.173)$$

or

$$\left(\frac{T_e}{T} - 1\right) \left(\frac{T}{T_e}\right)^{2u} = \frac{q^2\tau_0^2}{m^*kT} \frac{(r + \frac{1}{2})!}{(\frac{3}{2} - r)!} E^2$$

$$= (\beta E)^2 \quad (5.174)$$

For hot electrons T_e is much greater than T and (5.174) can be solved for T_e, with the result

$$\frac{T_e}{T} = (\beta E)^{-2/(2u-1)} \quad (5.175)$$

From (5.162) the mobility is

$$\mu(T_e) = \mu_0 \left(\frac{T_e}{T}\right)^u$$

or

$$\mu(\mathbf{E}) = \mu_0(\beta \mathbf{E})^{-2u/(2u-1)} \quad (5.176)$$

With (5.176) we can examine the dependence of the mobility on electric field for scattering processes in which the relaxation time approximation can be used.

The temperature dependence of the momentum relaxation time for acoustic phonon scattering is $u = -\frac{3}{2}$. Using this value in (5.176), the mobility is

$$\mu(T_e) = \mu_0 \left(\frac{T}{T_e}\right)^{3/2}$$
$$\mu(\mathbf{E}) = \mu_0 \left(\frac{1}{\beta E}\right)^{3/4} \quad (5.177)$$

Thus the mobility for this scattering mechanism decreases as the electric field increases. From (5.151) and (5.177) the drift velocity,

$$\mathbf{v}_d = -\mu_0 \beta^{-3/4} \mathbf{E}^{1/4} \quad (5.178)$$

does not saturate at high electric fields. For ionized impurity scattering $u = +\frac{3}{2}$ and the mobility is

$$\mu(T_e) = \mu_0 \left(\frac{T_e}{T}\right)^{3/2}$$
$$\mu(\mathbf{E}) = \mu_0 \left(\frac{1}{\beta E}\right)^{3/2} \quad (5.179)$$

Thus, for this scattering mechanism the mobility also decreases with electric field, and the drift velocity does not saturate. From (5.176) we see that u must be infinite to obtain a completely saturated drift velocity. This corresponds to scattering by optical phonons, where the energy changes are so large that the use of a momentum relaxation time is a poor approximation.

Under these conditions a reasonable approximation to the saturated drift velocity can be obtained by assuming that the energy of the hot electron distribution is dissipated in the generation of longitudinal optical phonons

$$\frac{\langle \mathscr{E} \rangle - \langle \mathscr{E}_0 \rangle}{\tau_e} \simeq \frac{\hbar\omega_{LO}}{\tau_e} \quad (5.180)$$

and that the proportionality factor between drift velocity and electric field is

$$\mu = \frac{q\tau_e}{m^*} \quad (5.181)$$

Using (5.180) and (5.181) in the energy balance equation (5.167), we find that

$$\tau_e = \frac{(\hbar\omega_{LO}m^*)^{1/2}}{qE} \quad (5.182)$$

and the drift velocity is saturated at the value

$$v_d = \left(\frac{\hbar\omega_{LO}}{m^*}\right)^{1/2} \quad (5.183)$$

Values for the longitudinal optical phonon frequencies from Table 3.2 can be used in (5.183) to obtain saturated drift velocities which are in reasonable agreement with experimental results.

If the electric field is increased further in the saturated region, at some point the charge carriers will have sufficient energy to generate an electron–hole pair in a collision with the lattice. This process is known as *impact ionization*. When each of these electron–hole pairs creates an additional pair by impact ionization, an unstable situation is obtained where, in principle, the number of charge carriers increases without limit. This situation is referred to as *avalanche breakdown* and is readily observed in the reverse-bias current–voltage characteristics of *p-n* junctions. These phenomena are examined in more detail in Chapter 9.

PROBLEMS

5.1. Find the nonequilibrium distribution function to second order in $\mathbf{E} = \hat{x}E_x$ for a parabolic band. Assume a Maxwellian equilibrium distribution and that $\tau_m = \tau_0(\mathscr{E}/kT)^{-1/2}$.

5.2. Derive an expression for the Hall factor assuming Fermi statistics and $\tau_m = \tau_0 x^r$. Plot r_H versus $\eta(-4 \leq \eta \leq 10)$ for $r = -\frac{1}{2}$ and $r = \frac{3}{2}$. Explain the dependence on η.

5.3. Determine the diffusion coefficient (D_n) for electrons in a fully *degenerate* n-type semiconductor, specifically for Ge with $\mu_n = 300\ \text{cm}^2/\text{V}\cdot\text{s}$, $n = 10^{19}\ \text{cm}^{-3}$, and $m_n = 0.2m_0$.

5.4. Show how all but one of the following secondary effects can be canceled in a Hall experiment by taking four measurements with B and J reversals: Ettinghausen, Nernst, Righi–Leduc, thermoelectric, and IR drop (probe misalignment).

5.5. Assume that a semiconductor has spherical energy surfaces in both the conduction and the valence bands, but with two species of holes. Find an expression for the Hall coefficient in terms of appropriate averages of τ_{p1}, τ_{p2}, τ_n, μ_{p1}, μ_{p2}, μ_n; p_1, p_2, n.

5.6. Consider an n-type semiconductor material as shown in Fig. P5.6 of mobility $10^4\ \text{cm}^2/\text{V}\cdot\text{s}$ which is to be used as a fast microwave switch. The device, a conductive bar, is switched from the more conductive to the less conductive state, with an applied electric field, by "heating" electrons from the direct to the indirect conduction band minima. Estimate the applied field needed.

$$m^*(\Gamma) = 0.07m, \qquad \frac{N_c(X)}{N_c(\Gamma)} \simeq 60$$

5.7. Consider the Righi–Leduc effect in a uniformly doped nondegenerate n-type semiconductor with spherical energy surfaces. A thermal gradient is maintained in the x direction and a magnetic field in the z direction, with $J_x = J_y = W_y = 0$. If $\nabla_y T = S_{\text{RL}} B_z \nabla_x T$:

(a) Find S_{RL} in terms of appropriate averages of τ^s.

(b) For spherical energy surfaces and $\tau = \tau_0 \mathscr{E}^p$, express S_{RL} in terms of gamma functions.

(c) For $p = -\frac{1}{2}$, show that $S_{\text{RL}} = -(21\pi/32)(k^2 T/q\kappa) n\mu_n^2$, where κ is the thermal conductivity.

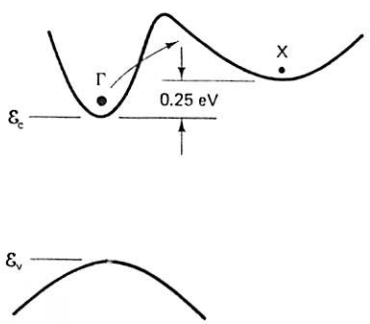


Figure P5.6

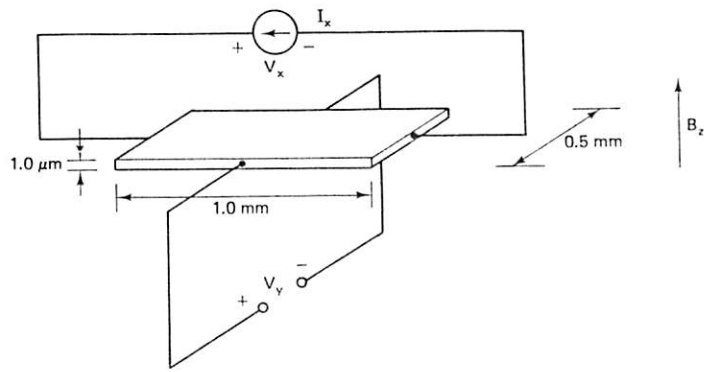


Figure P5.8

5.8. Hall meaurements are performed on a sample in the configuration shown in Fig. P5.8 with $I_x = 5$ mA, $B_z = 0.1$ T, and $T = 77$ K. Measured values are $V_x = 10$ V and $V_y = -5$ V. Assuming that $m_e = 0.1m$, $m_h = 0.5m$, and $\tau_m = \tau_0 x^{3/2}$ (ionized impurities), determine the following:

(a) Type of material
(b) Hall constant
(c) Carrier concentration
(d) Conductivity
(e) Conductivity mobility
(f) Hall mobility

5.9. Calculate the root-mean-square z-directed velocity of conduction band electrons in an isotropic material at equilibrium. Assume that nondegenerate statistics are valid and that the material exhibits spherically symmetric constant energy surfaces about Γ.

5.10. Consider a collection of electrons in an isotropic material which are constrained to possess a nonzero z-directed net velocity (e.g., through the application of an electric field $\mathbf{E} = -E_z\hat{z}$, $\mathbf{J} = \sigma\mathbf{E}$). Show that the distribution function has the form

$$f = \frac{1}{1 + \exp\left[\beta(\mathscr{E} - \alpha - \gamma v_z)\right]}$$

where v_z is the z-directed velocity.

5.11. The steady-state distribution function for a nondegenerate material under certain conditions is

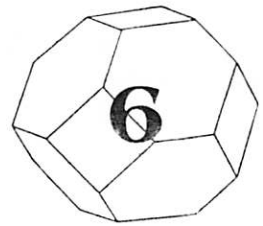
$$f = \exp\left[-\beta(\mathscr{E} - \zeta)\right] \exp\left(\frac{\hbar^2\beta}{m^*}\gamma k_z\right)$$

where $\beta = 1/kT$, ζ and γ are constants. The conduction band states near Γ

are described by

$$\mathscr{E} = \mathscr{E}_c + \frac{\hbar^2 k^2}{2m^*}$$

Determine the concentration of electrons in the conduction band and their mean z-directed velocity. For the conditions $\gamma = 10^7\ \text{m}^{-1}$, $T = 300$ K, $m^* = 0.0665m$, and $n = 10^{16}\ \text{cm}^{-3}$, what are the z-directed velocity and current density?



Scattering Processes

In Chapter 5 we used the relaxation time approximation to examine the transport of charge and energy in electric, magnetic, and thermal fields. There it was assumed that a momentum relaxation time, τ_m, could be defined for various carrier scattering processes, such that

$$\tau_m = \tau_0 x^r \tag{5.55}$$

where x is the electron kinetic energy in units of kT,

$$x \equiv \frac{\mathscr{E} - \mathscr{E}_c}{kT} \tag{4.65}$$

and τ_0 and the exponent r are independent of energy. In this chapter we examine this assumption and discuss the physics of the more important scattering processes. Rather than being all-inclusive, we will derive momentum relaxation times for ionized and neutral impurity scattering, as examples, and then show how these can be combined with values for phonon scattering to model and predict experimental mobility.

6.1 SCATTERING POTENTIALS

As discussed in Chapter 2, an electron moving in a perfect periodic crystal potential with no applied force has a constant velocity and is not scattered by the atoms of the crystal. When a force is applied to an electron, its acceleration can be described by a modified Newton's law where the perfect

periodic crystal potential is taken into account by an effective mass. Therefore, to describe the deceleration or scattering of an electron by a crystal defect, it is convenient to examine the perturbation that the defect produces on the perfect crystal potential. This perturbation is referred to as a *scattering potential*, $\Delta U(\mathbf{r})$, and has units of energy. In the following section we examine the scattering processes associated with impurities and phonons and derive their scattering potentials.

6.1.1 Impurities

For an ionized impurity the scattering process is dominated by the electrical interaction between its charge and the charge of the free carrier. For an ion with charge Zq the perturbation on the perfect crystal potential is simply the Coulomb energy,

$$\Delta U(\mathbf{r}) = \frac{\pm Zq^2}{4\pi\epsilon(0)\mathbf{r}} \tag{6.1}$$

where $\mathbf{r}$ is the distance between the ion and the charge carrier. The plus sign is valid when the charges on the ion and the carrier have the same polarity, and the minus sign is for charges of opposite polarity. As indicated in Fig. 6.1, the scattering trajectories of the free carriers are described by a hyperbola with the ion at a focal point.

In (6.1) the screening of the Coulomb potential by atomic and ionic polarization of the constituent atoms is described by the use of the static permittivity of the material, $\epsilon(0)$. Because of the long-range nature of the Coulomb potential, it is also necessary to consider the screening of (6.1) by other free carriers and ionized impurities. We look at this in Section 6.2.

A nonionized or neutral impurity has a scattering potential which is much weaker but more complex than for an ionized impurity. For a hydrogen-like neutral impurity, Coulombic scattering with the ground-state ($1s$) electron cloud occurs. The free carriers, however, can also interact with a neutral impurity by polarizing it or by changing places with a bound electron

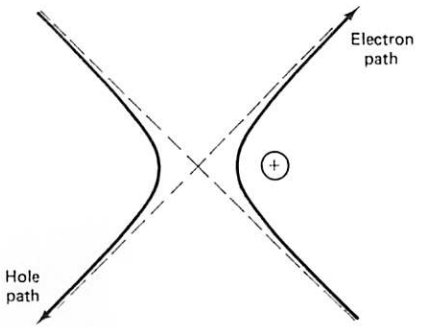


Figure 6.1 Trajectories of electrons and holes in ionized impurity scattering.

or hole. This raises the possibility of a combination of Coulombic, dipole, and exchange scattering which is not easy to analyze.

An empirical analysis [C. Erginsoy, *Phys. Rev. 79*, 1013 (1950)] of electron scattering by hydrogen atoms indicates that neutral impurity scattering can be approximated by a *differential scattering cross section*,

$$\sigma = \frac{20r_B}{k} \tag{6.2}$$

where r_B is the ground-state Bohr radius of the impurity, and k is the wavevector of the free electron. It can be shown (see Sections 6.3 and 6.4) that this scattering cross section is approximately equivalent to a scattering potential,

$$\Delta U(\mathbf{r}) \simeq \frac{\hbar^2}{m^*}\left(\frac{r_B}{r^5}\right)^{1/2} \tag{6.3}$$

where $\mathbf{r}$ is the distance between the neutral impurity and the free carrier. It is interesting to note that this $r^{-5/2}$ dependence is longer ranged than the r^{-4} dependence expected for dipole scattering.

6.1.2 Acoustic Phonons

The acoustic phonons in a crystal can scatter carriers by two different and independent processes. These are called *deformation potential scattering* and *piezoelectric scattering*. These scattering mechanisms can be examined, qualitatively, by means of Fig. 6.2, where the displacements, $u(\mathbf{r})$, of a chain of atoms from their Bravais lattice sites are shown for the longitudinal (LA) and transverse (TA) components of (a) zone center and (b) zone edge acoustic phonons.

As can be seen, the distance between adjacent atoms (the size of the unit cell) is strongly affected by the LA phonons, and little affected by the TA phonons. From the tight-binding model of the energy gap variation with lattice constant (see Fig. 2.14), we see that these LA phonons will produce a modulation of the conduction and valence band edges, $\mathcal{E}_c$ and $\mathcal{E}_v$. This modulation in space and time disturbs the periodicity of the crystal potential and produces the so-called deformation potential scattering of the electrons and holes.

For the long-wavelength acoustic phonons, it is convenient to treat the material as an elastic continuum. Then we can see in Fig. 6.2 that the maximum expansion and contraction of the unit cell produced by LA phonons occurs in regions where the divergence of the displacement vector (gradient of the displacement amplitude), or the *strain*, is maximum. Therefore, the scattering potential for deformation potential scattering must be proportional to the strain. Consider the displacement produced by an acoustic phonon

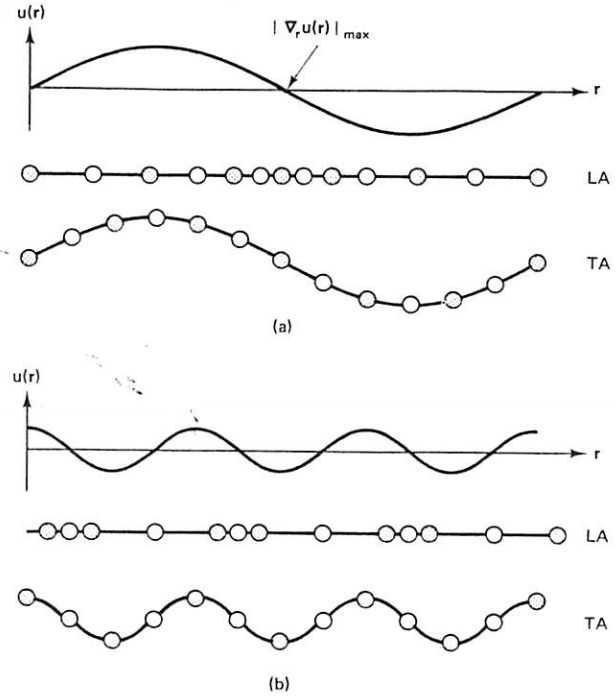


Figure 6.2 Displacements of a diatomic chain for LA and TA phonons at (a) the center and (b) the edge of the Brillouin zone. The lighter mass atoms are indicated by open circles. For zone edge acoustic phonons only the heavier atoms are displaced.

of frequency, ω_s, and wavevector, $\mathbf{q}_s$,

$$\mathbf{u}(\mathbf{r}, t) = \mathbf{a}u(\mathbf{r}, t) \tag{6.4}$$

where

$$u(\mathbf{r}, t) = u \exp [i(\mathbf{q}_s \cdot \mathbf{r} - \omega_s t)] \tag{6.5}$$

In these equations $\mathbf{a}$ is the displacement direction, and u is the amplitude. The strain associated with the displacement is

$$\nabla \cdot \mathbf{u}(\mathbf{r}, t) = \mathbf{a} \cdot \nabla u(\mathbf{r}, t) \tag{6.6}$$

$$\nabla \cdot \mathbf{u}(\mathbf{r}, t) = i\mathbf{q}_s \cdot \mathbf{a}u(\mathbf{r}, t) \tag{6.7}$$

Equation (6.7) indicates that for the transverse components of a phonon where the displacement and the wavevector are orthogonal, $\mathbf{q}_s \cdot \mathbf{a} = 0$, and no strain is produced. The scattering potential for the longitudinal component

is, therefore,

$$\Delta U(\mathbf{r}, t) = \mathcal{E}_A \nabla \cdot \mathbf{u}(\mathbf{r}, t) \tag{6.8}$$

where the *deformation potential*, $\mathcal{E}_A$, in units of energy, is defined as the proportionality constant between the scattering potential (units of energy) and the strain.

For some semiconductors with two or more atoms per unit cell, there is no crystal inversion symmetry. In these crystals the strain, caused predominantly by the LA phonons, polarizes the ions and produces internal electric fields that vary with time and space. The carrier scattering caused by these electric fields is called *piezoelectric scattering*. The scattering potential for electrons is simply

$$\Delta U(\mathbf{r}, t) = -q\psi(\mathbf{r}, t) \tag{6.9}$$

where $\psi(\mathbf{r}, t)$ is the electrostatic potential associated with the internal fields,

$$\psi(\mathbf{r}, t) = -\int \mathbf{E}(\mathbf{r}, t) \cdot d\mathbf{r} \tag{6.10}$$

To evaluate (6.9), it is thus necessary to determine the fields produced by the piezoelectric interaction.

At a given frequency, ω, the relationships among the electric flux density, $\mathbf{D}$, the electric field, $\mathbf{E}$, and the polarization, $\mathbf{P}$, are

$$\mathbf{D}(\omega) = \epsilon(\omega)\mathbf{E} = \epsilon_0 \mathbf{E} + \mathbf{P}(\omega) \tag{6.11}$$

where ϵ_0 is the free-space permittivity. In the low-frequency limit (6.11) become

$$\mathbf{D}(0) = \epsilon(0)\mathbf{E} = \epsilon_0 \mathbf{E} + \mathbf{P}(0) \tag{6.12}$$

where $\epsilon(0) = \epsilon_r(0)\epsilon_0$ is the static permittivity. Physically, the source of $\mathbf{D}(0)$ is the "true" charge (i.e., space and surface charges), while the source of $\mathbf{P}(0)$ is the "polarization" charge, usually atomic core and ionic dipoles. Since measurements of the so-called "static" dielectric constant do not include piezoelectric polarization, it is necessary to add an additional term to (6.12) to include this effect. [In principle, $\mathbf{P}(0)$ should include piezoelectric polarization.]

From Fig. 6.2 we see that this polarization must be proportional to the strain induced by the phonons. Ignoring the tensor nature of this interaction, we have

$$\mathbf{D}(0) = \epsilon(0)\mathbf{E}(\mathbf{r}, t) + e_{pz} \nabla u(\mathbf{r}, t) \tag{6.13}$$

where the piezoelectric constant, e_{pz}, has units of coulomb per square meter. With no true charge, the sources of the electric field are piezoelectric, ionic,

and atomic polarization,

$$\mathbf{E}(\mathbf{r}, t) = -\frac{e_{pz}}{\epsilon(0)} \nabla u(\mathbf{r}, t) \tag{6.14}$$

Using (6.9), (6.10), and (6.14), the scattering potential in terms of the displacement is

$$\Delta U(\mathbf{r}, t) = \frac{-qe_{pz}}{\epsilon(0)} u(\mathbf{r}, t) \tag{6.15}$$

With (6.5) and (6.6), equation (6.15) can be expressed in terms of the strain as

$$\Delta U(\mathbf{r}, t) = \frac{iqe_{pz}}{\epsilon(0)q_s} \nabla \cdot \mathbf{u}(\mathbf{r}, t) \tag{6.16}$$

Comparing (6.8) and (6.16), we see that the scattering potentials for the deformation potential and piezoelectric interactions are separated in phase by 90°. The two acoustic phonon scattering mechanisms therefore operate independently.

6.1.3 Optical Phonons

The optical phonons also scatter carriers by two independent processes. These are referred to as *deformation potential scattering* (the same as for acoustic phonons) and *polar mode scattering*. The deformation potential scattering by optical phonons is similar to that for acoustic phonons and the polar mode scattering is due to the polarization of atoms within the unit cell. The displacements, $u(\mathbf{r})$, of a chain of atoms from their Bravais lattice sites for longitudinal optical (LO) and transverse optical (TO) phonons are shown in Fig. 6.3.

In a manner similar to acoustic phonons it can be seen that the expansion and contraction of the unit cell is dominated by the longitudinal optical phonons. The main difference in Fig. 6.3 at the zone center is that the atoms in the unit cell vibrate against one another. Because of this, for optical phonon deformation potential scattering, it is necessary to consider the *relative* displacement between atoms in the unit cell,

$$\delta\mathbf{u}(\mathbf{r}, t) \equiv \mathbf{u}_1(\mathbf{r}, t) - \mathbf{u}_2(\mathbf{r}, t) \tag{6.17}$$

where $u_1(\mathbf{r}, t)$ and $\mathbf{u}_2(\mathbf{r}, t)$ have the form given by (6.4) and (6.5). The scattering potential due to modulation of the conduction and valence edges must then be proportional to this relative displacement and

$$\Delta U(\mathbf{r}, t) = D\, \delta u(\mathbf{r}, t) \tag{6.18}$$

where

$$\delta\mathbf{u}(\mathbf{r}, t) = \mathbf{a}\, \delta u(\mathbf{r}, t) \tag{6.19}$$

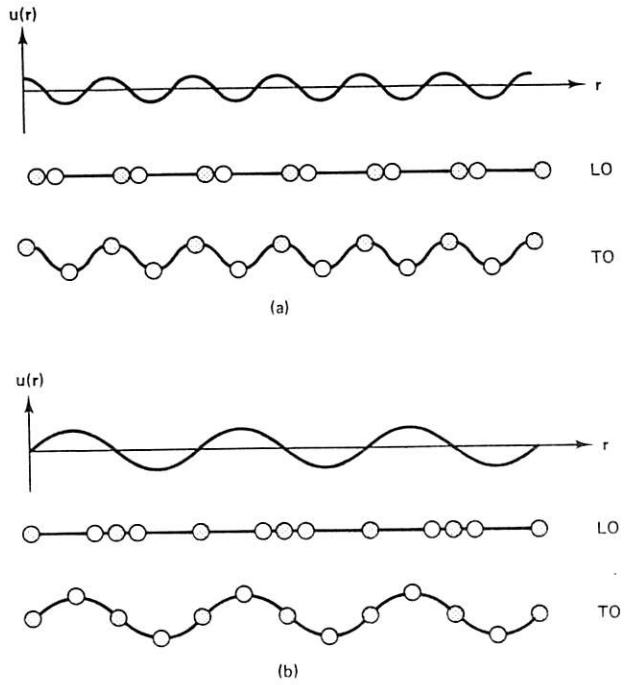


Figure 6.3 Displacements of a diatomic chain for LO and TO phonons at (a) the center and (b) the edge of the Brillouin zone. The lighter mass atoms are indicated by open circles. For zone edge optical phonons only the lighter atoms are displaced.

for optical phonon deformation potential scattering. In (6.18) the deformation potential constant, D, has units of energy per unit length. A similar treatment can be used for intervalley phonon scattering.

The optical phonon polar mode scattering is due to the electric field caused by the polarization of the ions in the unit cell. This polarization is caused mainly by the longitudinal component and is equivalent to the ionic polarization, $\mathbf{P}_i$, which is discussed in Chapter 7. The scattering potential is obtained from (6.9) and (6.10), where the internal electric field is deduced from the low- and high-frequency limits of (6.11),

$$\mathbf{D}(0) = \epsilon(0)\mathbf{E} = \epsilon_0\mathbf{E} + \mathbf{P}(0) \tag{6.20}$$

and

$$\mathbf{D}(\infty) = \epsilon(\infty)\mathbf{E} = \epsilon_0\mathbf{E} + \mathbf{P}(\infty) \tag{6.21}$$

Note that in (6.20) the total low-frequency polarization is due to atomic and

ionic polarization,

$$\mathbf{P}(0) = \mathbf{P}(\infty) + \mathbf{P}_i \tag{6.22}$$

Using (6.22) in (6.20) and subtracting (6.21), we obtain

$$\epsilon(0)\mathbf{E} = \epsilon(\infty)\mathbf{E} + \mathbf{P}_i \tag{6.23}$$

or

$$\mathbf{D}(0) = \epsilon(\infty)\mathbf{E} + \mathbf{P}_i \tag{6.24}$$

From (6.24) we can determine the internal fields induced by the optical phonon polarization of the unit cell.

The polarization of a unit cell, $\mathbf{P}_i(\mathbf{r}, t)$, is determined by the relative displacement of the ions in a unit cell, $\delta\mathbf{u}(\mathbf{r}, t)$, and the effective ionic charge, e^*, such that

$$\mathbf{P}_i(\mathbf{r}, t) = \frac{e^*}{\Omega}\,\delta\mathbf{u}(\mathbf{r}, t) \tag{6.25}$$

In this equation $\Omega = V/N$ is the volume of the N primitive or Wigner–Seitz unit cells and e^* is the Born effective charge given by

$$e^* = \Omega\omega_{LO}\epsilon(\infty)\rho^{1/2}\left[\frac{1}{\epsilon(\infty)} - \frac{1}{\epsilon(0)}\right]^{1/2} \tag{7.174}$$

where ρ is the mass density. This equation is derived in Chapter 7. Assuming no space or surface charges, (6.24) and (6.25) give an internal field,

$$\mathbf{E}(\mathbf{r}, t) = -\frac{e^*}{\Omega\epsilon(\infty)}\,\delta\mathbf{u}(\mathbf{r}, t) \tag{6.26}$$

Using (6.9), (6.10), and (6.26), the scattering potential for polar mode scattering is

$$\Delta U(\mathbf{r}, t) = \frac{-qe^*}{\Omega\epsilon(\infty)}\int \delta\mathbf{u}(\mathbf{r}, t)\cdot d\mathbf{r} \tag{6.27}$$

or with (6.5) and (6.19),

$$\Delta U(\mathbf{r}, t) = \frac{iqe^*}{\Omega\epsilon(\infty)q_s}\,\delta u(\mathbf{r}, t) \tag{6.28}$$

A comparison of (6.18) and (6.28) shows that the scattering potentials for deformation potential and polar mode scattering by optical phonons are out of phase by 90° and are thus independent.

Equation (6.28) is sometimes written in the form

$$\Delta U(\mathbf{r}, t) = \frac{iqe_c^*}{\Omega\epsilon_0 q_s}\,\delta u(\mathbf{r}, t) \tag{6.29}$$

where a Callen effective charge, e_c^* is used. The relationship to the Born effective charge is given by

$$e^* = \epsilon_r(\infty)e_c^* \tag{6.30}$$

6.2 SCREENING

In the derivations of the scattering potentials we assumed, in all cases, that there were no space charges (also, no surface charges). That is, we assumed that the charge carriers were uniformly distributed in the material such that

$$-\rho = q(n - p + N_a^- - N_d^+) = 0 \tag{6.31}$$

In the vicinity of a crystal potential perturbation caused by an impurity or phonon, however, charge carriers can be accumulated or depleted by the scattering potential. This space charge produces an additional potential given by

$$\nabla^2\psi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon(0)} \tag{4.99}$$

where

$$-\rho(\mathbf{r}) = q[n(\mathbf{r}) - p(\mathbf{r}) + N_a^-(\mathbf{r}) - N_d^+(\mathbf{r})] \tag{6.32}$$

which screens the effects of the scattering potential.

In (6.32), $n(\mathbf{r})$, $p(\mathbf{r})$, $N_a^-(\mathbf{r})$, and $N_d^+(\mathbf{r})$ are the *total* electron, hole, ionized acceptor, and ionized donor concentrations as a function of distance, $\mathbf{r}$, from the center of the perturbing potential. These total concentrations can be split into two components,

$$\begin{aligned} n(\mathbf{r}) &= n + \delta n(\mathbf{r}) \\ N_d^+(\mathbf{r}) &= N_d^+ + \delta N_d^+(\mathbf{r}), \quad \text{etc.} \end{aligned} \tag{6.33}$$

a uniform concentration and an excess (or deficit) concentration which varies with $\mathbf{r}$ according to the variation of $\psi(\mathbf{r})$. Excess carriers due to built-in potentials in nonuniform materials are discussed in some detail in Chapter 8. For our purposes here, we assume that $q\psi(\mathbf{r}) \ll kT$, so that the excess concentrations are, approximately,

$$\begin{aligned} \delta n(\mathbf{r}) &\simeq \frac{qn}{kT}\psi(\mathbf{r}) \\ \delta N_d^+(\mathbf{r}) &\simeq \frac{-qN_d^+}{kT}\psi(\mathbf{r}), \quad \text{etc.} \end{aligned} \tag{6.34}$$

With this approximation we can define an *effective total* electron

concentration,

$$n^*(\mathbf{r}) \equiv n(\mathbf{r}) - p(\mathbf{r}) + N_a^-(\mathbf{r}) - N_d^+(\mathbf{r}) \tag{6.35}$$

which, using (6.33), (6.34), and (6.31), takes the form

$$n^*(\mathbf{r}) = \frac{qn^*}{kT}\psi(\mathbf{r}) \tag{6.36}$$

From (4.99), (6.32), (6.35), and (6.36), Poisson's equation for the potential is

$$\nabla^2\psi(\mathbf{r}) = \frac{q^2n^*}{\epsilon kT}\psi(\mathbf{r}) \tag{6.37}$$

where the *effective*, uniform electron concentration, n^*, is to be determined later. If we define an effective Debye length,

$$\lambda^2 \equiv \frac{\epsilon kT}{q^2n^*} \tag{6.38}$$

the differential equation for the potential is

$$\nabla^2\psi(\mathbf{r}) = \frac{1}{\lambda^2}\psi(\mathbf{r}) \tag{6.39}$$

For a spherically symmetric potential, (6.39) is

$$\frac{d^2}{dr^2}[r\psi(r)] = \frac{r\psi(r)}{\lambda^2} \tag{6.40}$$

The physically significant solution to this equation is

$$\psi(r) = \frac{C}{r}\exp\left(\frac{-r}{\lambda}\right) \tag{6.41}$$

where the constant of integration,

$$C = \frac{Zq^2}{4\pi\epsilon} \tag{6.42}$$

for ionized impurity scattering. From (6.41) we see that the accumulation or depletion of charge carriers produces an exponential decay of the scattering potential with a characteristic length λ. This characteristic length is controlled by n^*, the effective electron concentration.

6.2.1 Degenerate Statistics

To determine $n^*(\mathbf{r})$ and, subsequently, n^*, it is necessary to introduce an energy band formalism for nonuniform materials. This formalism is summarized in Fig. 6.4, where we show the terminology for (a) a semiconductor

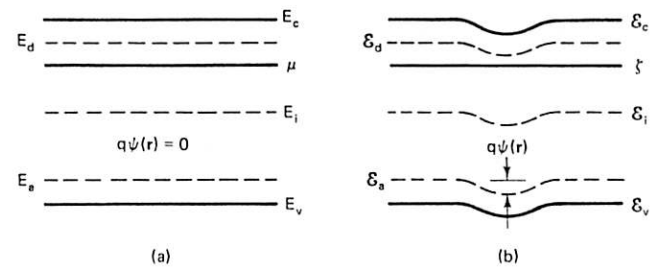


Figure 6.4 Energy band diagrams and terminology for (a) uniform material and (b) the same material with potential perturbations.

with uniform charge carrier and ionized impurity distributions, and (b) the same semiconductor with nonuniform distributions caused by perturbing potentials.

For a material with uniform doping and band structure as in Fig. 6.4(a), the Fermi energy is flat in thermal equilibrium and equal to the chemical potential energy, μ. For the same material with perturbing potentials due to impurities and phonons as in Fig. 6.4(b), the Fermi energy is flat, also in thermal equilibrium, and equal to the electrochemical potential energy,

$$\zeta = \mu(\mathbf{r}) - q\psi(\mathbf{r}) \tag{6.43}$$

All the other energies have the form

$$\mathcal{E}_c(\mathbf{r}) = E_c - q\psi(\mathbf{r}), \quad \text{etc.} \tag{6.44}$$

Since the perturbed material is in thermal equilibrium, we can simply modify the equilibrium distributions for n, p, N_a^-, and N_d^+ derived in Chapter 4 with this formalism to obtain $n(\mathbf{r})$, $p(\mathbf{r})$, $N_a^-(\mathbf{r})$, and $N_d^+(\mathbf{r})$.

From (4.68) and (4.65) the total electron concentration is

$$n(\mathbf{r}) = N_cF_{1/2}[\eta_c(\mathbf{r})] \tag{6.45}$$

where

$$\eta_c(\mathbf{r}) = \frac{\mu - E_c + q\psi(\mathbf{r})}{kT} \tag{6.46}$$

Equations (4.78) and (4.76) give the total hole concentration,

$$p(\mathbf{r}) = N_vF_{1/2}[\eta_v(\mathbf{r})] \tag{6.47}$$

where

$$\eta_v(\mathbf{r}) = \frac{E_v - \mu - q\psi(\mathbf{r})}{kT} \tag{6.48}$$

The total ionized acceptor concentration from (4.98) is

$$N_a^-(\mathbf{r}) = \frac{N_a}{1 + g_a \exp\{[E_a - \mu - q\psi(\mathbf{r})]/kT\}} \tag{6.49}$$

while the ionized donor concentration from (4.97) is

$$N_d^+(r) = \frac{N_d}{1 + g_d \exp\{[\mu - E_d + q\psi(\mathbf{r})]/kT\}} \tag{6.50}$$

Equations (6.45) and (6.47) for the electron and hole concentrations can be simplified in the following manner. From (4.67) the Fermi–Dirac integral is

$$F_{1/2}[\eta(\mathbf{r})] = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2}\, dx}{1 + \exp[x - \eta(\mathbf{r})]} \tag{6.51}$$

where $\eta(\mathbf{r})$ is given by either (6.46) or (6.48). The exponential in the denominator in (6.51) involving $q\psi(\mathbf{r})/kT$ is then expanded in a Taylor series and the denominator is divided into the numerator. Neglecting terms in $\psi^2(\mathbf{r})$ and higher, which is equivalent to the assumption $q\psi(\mathbf{r}) \ll kT$, (6.45) and (6.47) become

$$n(\mathbf{r}) = n + \frac{q\psi(\mathbf{r})}{kT} \frac{dn}{d\eta_c} \tag{6.52}$$

and

$$p(\mathbf{r}) = p - \frac{q\psi(\mathbf{r})}{kT} \frac{dp}{d\eta_v} \tag{6.53}$$

From the recurrence relationship (5.83),

$$\frac{dn}{d\eta_c} = N_c F_{-1/2}(\eta_c) = n \frac{F_{-1/2}(\eta_c)}{F_{1/2}(\eta_c)} \tag{6.54}$$

and

$$\frac{dp}{d\eta_v} = N_v F_{-1/2}(\eta_v) = p \frac{F_{-1/2}(\eta_v)}{F_{1/2}(\eta_v)} \tag{6.55}$$

Equations (6.49) and (6.50) for the ionized acceptor and donor concentrations can be simplified in a similar manner to obtain

$$N_a^-(\mathbf{r}) = N_a^- + \frac{N_a^0 N_a^-}{N_-} \frac{q\psi(\mathbf{r})}{kT} \tag{6.56}$$

and

$$N_d^+(r) = N_d^+ - \frac{N_d^0 N_d^+}{N_d} \frac{q\psi(\mathbf{r})}{kT} \tag{6.57}$$

The effective electron concentration, n^*, which controls the screening is obtained as follows: (6.52), (6.53), (6.56), and (6.57) are substituted into (6.35). Equation (6.31) for charge neutrality is applied to (6.35) and n^* is then obtained from (6.36) as

$$n^* = \frac{dn}{d\eta_c} + \frac{dp}{d\eta_v} + \frac{N_a^0 N_a^-}{N_a} + \frac{N_d^0 N_d^+}{N_d} \tag{6.58}$$

Equation (6.58) is valid for screening in degenerately doped semiconductors under the assumption that the perturbation energy is substantially less than the thermal energy.

6.2.2 Nondegenerate Statistics

When the doping of the material is such that the Fermi energy is greater than $\mathcal{E}_v + 4kT$ and less than $\mathcal{E}_c - 4kT$, η is negative, and $F_{1/2}(\eta) = F_{-1/2}(\eta)$. From (6.54), (6.55), and (6.58) the effective screening concentration is

$$n^* = n + p + \frac{N_a^0 N_a^-}{N_a} + \frac{N_d^0 N_d^+}{N_d} \tag{6.59}$$

where $N_a^0 = N_a - N_a^-$ and $N_d^0 = N_d - N_d^+$. Eliminating the neutral concentrations,

$$n^* = n + p + N_a^- \left(1 - \frac{N_a^-}{N_a}\right) + N_d^+ \left(1 - \frac{N_d^+}{N_d}\right) \tag{6.60}$$

For nondegenerate n- or p-type material, (6.60) can be further simplified. In n-type material, for example, all the acceptors will usually be ionized, so $N_a^- = N_a$. The acceptors are ionized by electrons from the donors, with the remaining electrons from the donors contributing to conduction. Therefore, $N_d^+ = n + N_a$. Using these arguments and space charge neutrality, in (6.60) the effective electron screening concentration is

$$n^* = n + \frac{(n + N_a)(N_d - N_a - n)}{N_d} \tag{6.61}$$

For p-type material, the effective hole screening concentration is

$$p^* = p + \frac{(p + N_d)(N_a - N_d - p)}{N_a} \tag{6.62}$$

Since N_d and N_a are usually constant in a material, (6.61) and (6.62) are useful in examining screening under conditions where n or p vary.

6.3 COLLISION INTEGRAL

Boltzmann's equation for the time rate of change of the electron distribution function under the influence of internal and applied forces is

$$\frac{\partial f}{\partial t} = -\frac{1}{\hbar}\mathbf{F}\cdot\nabla_k f - \mathbf{v}\cdot\nabla_r f + \left.\frac{\partial f}{\partial t}\right|_c \tag{5.4}$$

In Chapter 5 the first and second terms on the right-hand side of this equation were evaluated under the assumption that the third term could be put in the form

$$\left.\frac{\partial f}{\partial t}\right|_c = \frac{-(f - f_0)}{\tau_m} \tag{5.6}$$

That is, we assumed that the time rate of change of the distribution function due to collisions (the collision term) could be described by a momentum relaxation time, τ_m. In this section we examine the conditions under which this assumption is valid and show how τ_m can be obtained from the scattering potentials derived in Section 6.1. An equation relating the collision term or momentum relaxation time to the basic scattering process is called a *collision integral*. The scattering process itself can be described, quantum mechanically, by a matrix element or, classically, by a differential scattering cross section. We examine these two treatments and develop the relationship between them.

6.3.1 Quantum Treatment

The Hamiltonian for an electron undergoing a scattering process is

$$\mathbf{H} = \mathbf{H}_0 + \Delta\mathbf{U} \tag{6.63}$$

where $\mathbf{H}_0$ is the unperturbed energy operator and $\Delta\mathbf{U}$ is one (or more) of the scattering potentials, in operator form, derived in Section 6.1. Since the process evolves in time, Schrödinger's equation is

$$(\mathbf{H}_0 + \Delta\mathbf{U})\psi(t) = i\hbar\frac{\partial\psi(t)}{\partial t} \tag{6.64}$$

Solutions to (6.64) are obtained by constructing time-dependent wavefunctions from a set of time-independent Bloch wavefunctions,

$$\psi(t) = \sum_k A_k(t)\psi_k \exp\left[\frac{(-i\mathscr{E}_k t)}{\hbar}\right] \tag{7.13}$$

where ψ_k are given by (2.10).

This scattering problem is formally equivalent to the optical transition problem described in Chapter 7. For an electron that is scattered from a

state with wavevector $\mathbf{k}$ to one with wavevector $\mathbf{k}'$, the scattering rate is

$$S_{kk'} \equiv \frac{|A_k(t)|^2}{t} \tag{6.65}$$

Using (7.36) this can be written as

$$S_{kk'} = \frac{2\pi}{\hbar}|H_{kk'}|^2\delta(\mathscr{E}_k - \mathscr{E}_{k'}) \tag{6.66}$$

where from (7.19) the matrix element is

$$H_{kk'} = \frac{1}{N}\int_V \psi_k^*\,\Delta\mathbf{U}\psi_{k'}\,d\mathbf{r} \tag{6.67}$$

In (6.67), N is the number of primitive or Wigner–Seitz unit cells and V is the crystal volume.

For an electron to be scattered from an initial state $\mathbf{k}$ to one of the $(N - 1)$ states $\mathbf{k}'$, the initial state $\mathbf{k}$ must be occupied and the final state $\mathbf{k}'$ must be unoccupied. Conversely, an electron in one of the occupied $(N - 1)$ states $\mathbf{k}'$ can be scattered into the unoccupied state $\mathbf{k}$. Considering these two competing processes and summing over all $(N - 1)$ values of $\mathbf{k}'$, the time rate of *increase* of the distribution function due to collisions is

$$\left.\frac{\partial f}{\partial t}\right|_c = N_s\sum_{k'}^{N-1}[S_{k'k}f_{k'}(1 - f_k) - S_{kk'}f_k(1 - f_{k'})] \tag{6.68}$$

where N_s is the number of scattering centers and f_k is the nonequilibrium distribution function at energy $\mathscr{E}(\mathbf{k})$. Since the number of unit cells in the crystal, N, is very large, the summation over the $(N - 1)$ values of $\mathbf{k}'$ can be approximated by an integration over the $(N - 1) \simeq N$ values of $\mathbf{k}'$ in the Brillouin zone. From (2.30), (1.20), and (1.12), each value of $\mathbf{k}'$ occupies a reciprocal volume,

$$\Omega_{k'} = \frac{(2\pi)^3}{V} \tag{6.69}$$

The integral approximation of (6.68) is, therefore,

$$\left.\frac{\partial f}{\partial t}\right|_c = \frac{N_s V}{(2\pi)^3}\int_{\Omega_K}[S_{k'k}f_{k'}(1 - f_k) - S_{kk'}f_k(1 - f_{k'})\,d\mathbf{k}' \tag{6.70}$$

When (6.70) is used in (5.4), an integrodifferential form of Boltzmann's equation is obtained which is quite general and valid for arbitrary degeneracy.

It is instructive to examine (6.70) in thermal equilibrium. Under this condition there is no change in the distribution function and the left-hand side of (6.70) must equal zero. For this to be true for any value of $\mathbf{k}'$,

$$S_{k'k} = S_{kk'}\frac{f_{0k}(1 - f_{0k'})}{f_{0k'}(1 - f_{0k})} \tag{6.71}$$

where the subscript zero denotes the equilibrium Fermi–Dirac distribution function,

$$f_{0k} = \left[1 + \exp\left(\frac{\mathscr{E}_k - \mathscr{E}_f}{kT}\right)\right]^{-1} \tag{6.72}$$

If the material is nondegenerate, $\mathscr{E}_k - \mathscr{E}_f \gg kT$, and (6.71) reduces to

$$S_{k'k} = S_{kk'} \exp\left(\frac{\mathscr{E}_{k'} - \mathscr{E}_k}{kT}\right) \tag{6.73}$$

From (6.73) it can be seen that $S_{k'k} = S_{kk'}$ only when $\mathscr{E}_{k'} = \mathscr{E}_k$ or $|\mathbf{k}'| = |\mathbf{k}|$. That is, the scattering rate from a state $\mathbf{k}$ to $\mathbf{k}'$ is equal to its inverse only for elastic collisions. It is only under this condition that a universal momentum relaxation time can be defined.

Assuming elastic collisions we will now evaluate the nonequilibrium distribution coefficients for an arbitrary force field. With this we can then obtain the relationship between the momentum relaxation time and the matrix element, or using (6.67) the scattering potential, for a scattering process that conserves energy.

Consider an electron with initial wavevector $\mathbf{k}$ scattering into a final state $\mathbf{k}'$ under the influence of an arbitrary force, $\mathbf{G}$. This force can include electric, magnetic, and thermal fields. From (5.18) the nonequilibrium distribution function in the relaxation time approximation is

$$f_k = f_{0k} + \frac{\partial f_{0k}}{\partial \mathscr{E}_k} \frac{\hbar}{m^*} \mathbf{k}\cdot\mathbf{G} \tag{6.74}$$

For a collision at the origin of the reciprocal-space coordinate system in Fig. 6.5,

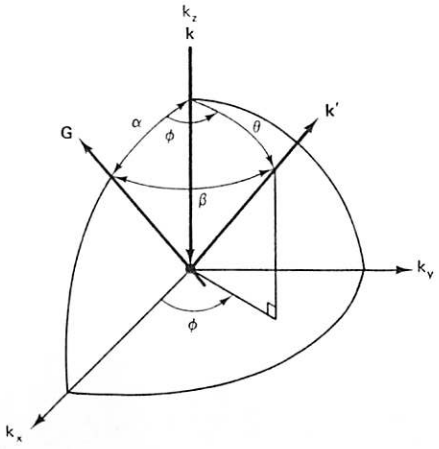


Figure 6.5 Spherical coordinate system in reciprocal space for an electron with wavevector $\mathbf{k}$ (along the k_z axis) scattering into a state with wavevector $\mathbf{k}'$ in an arbitrary force field $\mathbf{G}$. The scattering center is at the origin. For simplicity the event is rotated so that $\mathbf{G}$ has no k_y component.

$$\mathbf{k}\cdot\mathbf{G} = kG\cos\alpha \tag{6.75}$$

and

$$\mathbf{k}'\cdot\mathbf{G} = kG\cos\beta \tag{6.76}$$

The magnitudes of $\mathbf{k}$ and $\mathbf{k}'$ are equal when energy is conserved.

Using (6.74), (6.75), and (6.76), the expression involving distribution functions in the integrand of (6.70) is

$$f_{k'}(1 - f_k) - f_k(1 - f_{k'}) = \frac{\partial f_0}{\partial \mathscr{E}} \frac{\hbar}{m^*} kG[\cos\beta(1 - \cos\alpha) - \cos\alpha(1 - \cos\beta)] \tag{6.77}$$

Eliminating $\cos\beta$ with the equation for a spherical triangle (see Fig. 6.5) gives

$$\cos\beta = \cos\alpha\cos\theta + \sin\alpha\sin\theta\cos\phi \tag{6.78}$$

Equation (6.77) becomes

$$f_{k'}(1 - f_k) - f_k(1 - f_{k'}) = \frac{\partial f_0}{\partial \mathscr{E}} \frac{\hbar}{m^*} kG[\sin\alpha\sin\theta\cos\phi - \cos\alpha(1 - \cos\theta)] \tag{6.79}$$

Inserting (6.79) into (6.70), the collision term is

$$\left.\frac{\partial f}{\partial t}\right|_c = -\frac{\partial f_0}{\partial \mathscr{E}} \frac{\hbar}{m^*} G \frac{N_s V}{(2\pi)^3} \int_{\Omega_K} S_{kk'}[\cos\alpha(1 - \cos\theta) - \sin\alpha\sin\theta\cos\phi]k\, d\mathbf{k}' \tag{6.80}$$

Equation (6.80) can be simplified to give an expression for τ_m in the following manner: First, note that the differential volume in $\mathbf{k}$-space is

$$d\mathbf{k} = k^2 \sin\theta\, d\theta\, d\phi\, dk \tag{6.81}$$

Then, integrate ϕ from 0 to 2π, which eliminates the ϕ-dependent term. Finally, from (6.74) and (6.75),

$$-\frac{\partial f_0}{\partial \mathscr{E}} \frac{\hbar}{m^*} kG\cos\alpha = -(f - f_0) \tag{6.82}$$

Following these steps, (6.80) becomes

$$\frac{1}{\tau_m} = \frac{N_s V}{(2\pi)^2 k} \int_k \int_0^\pi S_{kk'} \sin\theta(1 - \cos\theta)\, d\theta\, k^3\, dk \tag{6.83}$$

Equation (6.66) for the scattering rate, $S_{kk'}$, shows that the integration over

k is zero except at the point where energy is conserved. Using (6.66) and assuming a parabolic band, the relationship between the momentum relaxation time and the matrix element for the scattering process is

$$\frac{1}{\tau_m} = \frac{N_s V m^{*2} v}{2\pi\hbar^4} \int_0^\pi |H_{kk'}|^2 \sin\theta(1 - \cos\theta)\, d\theta \tag{6.84}$$

6.3.2 Classical Treatment

The classical derivation of the momentum relaxation time proceeds rather simply. For N_s/V scattering centers per unit volume with scattering cross section, σ_m, the mean free time between collisions for an electron with velocity, v, is

$$\frac{1}{\tau_m} = \frac{N_s v}{V} \sigma_m \tag{6.85}$$

The scattering cross section is determined by setting a scattering center with differential cross section, $\sigma(\theta)$, at the origin in Fig. 6.5. The θ-dependence allows for different scattering mechanisms. An electron scattered by the center into the solid angle (θ, ϕ) *loses* $(1 - \cos\theta)$ of its initial momentum in the incident direction. Taking into account all possible scattering angles yields

$$\sigma_m = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \sigma(\theta) \sin\theta(1 - \cos\theta)\, d\theta\, d\phi \tag{6.86}$$

Using (6.85) and (6.86), the momentum relaxation time is

$$\frac{1}{\tau_m} = \frac{2\pi N_s v}{V} \int_0^\pi \sigma(\theta) \sin\theta(1 - \cos\theta)\, d\theta \tag{6.87}$$

This is the classical collision integral.

Since the quantum and classical integrals have the same angular dependence, a relationship can be obtained between the differential scattering cross section, $\sigma(\theta)$, and the matrix element, $H_{kk'}$, for a given scattering process. Equating (6.84) and (6.87), we obtain

$$\sigma(\theta) = \left(\frac{Vm^*}{2\pi\hbar^2} |H_{kk'}|\right)^2 \tag{6.88}$$

6.4 MATRIX ELEMENTS

In principle the calculation of a matrix element for electron scattering from a given scattering potential using

$$H_{kk'} = \frac{1}{N} \int_V \psi_k^* \, \Delta\mathbf{V} \, \psi_{k'} \, d\mathbf{r} \tag{6.67}$$

is relatively straightforward. In detail, however, the procedure is often quite laborious, involving a number of approximations and assumptions. Here, we simply indicate the general procedure and refer the reader to the literature for the detailed treatment.

6.4.1 General Procedure

The usual procedure for evaluating a matrix element is first to expand the scattering potential in a Fourier series,

$$\Delta U(\mathbf{r}) = \sum_g A_g \exp(i\mathbf{g}\cdot\mathbf{r}) \tag{6.89}$$

where the Fourier coefficients are

$$A_g = \frac{1}{V} \int_V \Delta U(\mathbf{r}) \exp(-i\mathbf{g}\cdot\mathbf{r})\, d\mathbf{r} \tag{6.90}$$

For Bloch wavefunctions,

$$\psi_k(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) u_k(\mathbf{r}) \tag{2.10}$$

$$H_{kk'} = \frac{1}{N} \sum_g \int_V \exp(-i\mathbf{k}\cdot\mathbf{r}) u_k^*(\mathbf{r}) A_g \exp(i\mathbf{g}\cdot\mathbf{r}) \times \exp(i\mathbf{k}'\cdot\mathbf{r}) u_{k'}(\mathbf{r})\, d\mathbf{r} \tag{6.91}$$

$$H_{kk'} = \frac{1}{N} \sum_g A_g \int_V u_k^*(\mathbf{r}) u_{k'}(\mathbf{r}) \exp[i(\mathbf{g} + \mathbf{k}' - \mathbf{k})\cdot\mathbf{r}]\, d\mathbf{r} \tag{6.92}$$

Since the integral is zero except when

$$\mathbf{g} = \mathbf{k} - \mathbf{k}' \tag{6.93}$$

(6.92) is

$$H_{kk'} = \frac{A_g}{N} \int_V u_k^*(\mathbf{r}) u_{k'}(\mathbf{r})\, d\mathbf{r} \tag{6.94}$$

For parabolic bands $u_k(\mathbf{r}) = u_{k'}(\mathbf{r})$, and the matrix element is simply the Fourier coefficient that satisfies (6.93)

$$H_{kk'} = A_{k-k'} \tag{6.95}$$

where $A_{k-k'}$ is given by (6.90).

6.4.2 Screening Factor

Since the matrix element for ionized impurity scattering is relatively easy to obtain, we will derive it as an example of the procedure. Also, by comparing the screened and unscreened matrix elements for ionized impurity

scattering, the screening factor for a general scattering process can be deduced.

Inserting the unscreened potential for ionized impurity scattering (6.1) into (6.90) yields

$$A_g = \frac{Zq^2}{4\pi\epsilon(0)V}\int_V \exp(-i\mathbf{g}\cdot\mathbf{r})\frac{d\mathbf{r}}{\mathbf{r}} \tag{6.96}$$

For the differential volume element

$$d\mathbf{r} = r^2 \sin\theta\, d\theta\, d\phi\, dr \tag{6.97}$$

$$A_g = \frac{Zq^2}{\epsilon(0)V}\int_0^\infty r\exp(-igr)\,dr \tag{6.98}$$

The integral is evaluated with (3.21) to obtain

$$A_g = \frac{Zq^2}{\epsilon(0)V\,|\,g\,|^2} \tag{6.99}$$

or from (6.95),

$$H_{kk'} = \frac{Zq^2}{\epsilon(0)V\,|\,k-k'\,|^2} \tag{6.100}$$

For the screened potential, (6.41),

$$A_g = \frac{Zq^2}{4\pi\epsilon(0)V}\int_V \exp\left(-\frac{\mathbf{r}}{\lambda}\right)\exp(-i\mathbf{g}\cdot\mathbf{r})\frac{d\mathbf{r}}{\mathbf{r}} \tag{6.101}$$

Following the procedures above gives us

$$A_g = \frac{Zq^2}{\epsilon(0)V(|\,g\,|^2 + 1/\lambda^2)} \tag{6.102}$$

or

$$H_{kk'} = \frac{Zq^2}{\epsilon(0)V(|\,k-k'\,|^2 + 1/\lambda^2)} \tag{6.103}$$

A comparison of the screened equation (6.103) and the unscreened equation (6.100) indicates that the screening factor for a scattering process in general is

$$\Lambda = \frac{|\,k-k'\,|^2}{|\,k-k'\,|^2 + 1/\lambda^2} \tag{6.104}$$

In Table 6.1 we have summarized the scattering potentials and matrix elements for various scattering mechanisms. Screening can be accounted for by multiplying the matrix elements by (6.104).

TABLE 6.1 Scattering Potentials and Matrix Elements for Various Scattering Mechanisms[a]

Scattering Mechanisms	Scattering Potential	Matrix Element
Impurities		
Ionized	$\frac{Zq^2}{4\pi\epsilon(0)r}$	$\frac{Zq^2}{\epsilon(0)V\|k-k'\|^2}$
Neutral	$\frac{\hbar^2}{m^*}\left(\frac{r_B}{r^5}\right)^{1/2}$	$\frac{2\pi\hbar^2}{m^*V}\left(\frac{20r_B}{k}\right)^{1/2}$
Acoustic phonons		
Deformation potential	$\mathscr{E}_A\nabla\cdot\mathbf{u}$	$\mathscr{E}_A\left(\frac{\hbar}{2V\rho\omega_s}\right)^{1/2}(\mathbf{a}\cdot\mathbf{q}_s)\left(n_q+\frac{1}{2}\pm\frac{1}{2}\right)^{1/2}$
Piezoelectric	$\frac{iqe_{pz}}{\epsilon(0)q_s}\nabla\cdot\mathbf{u}$	$\frac{qe_{pz}}{\epsilon(0)}\left(\frac{\hbar}{2V\rho\omega_s}\right)^{1/2}\left(n_q+\frac{1}{2}\pm\frac{1}{2}\right)^{1/2}$
Optical phonons		
Deformation potential	$D\delta u$	$D\left(\frac{\hbar}{2V\rho\omega_{LO}}\right)^{1/2}\left(n_q+\frac{1}{2}\pm\frac{1}{2}\right)^{1/2}$
Polar	$\frac{iqe^*}{\omega\epsilon(\infty)q_s}\delta u$	$\frac{qe^*}{\Omega\epsilon(\infty)q_s}\left(\frac{\hbar}{2V\rho\omega_{LO}}\right)^{1/2}\left(n_q+\frac{1}{2}\pm\frac{1}{2}\right)^{1/2}$

[a] r_B = Bohr radius; n_q = phonon occupation number; $e^* = \Omega\omega_{LO}\epsilon(\infty)\rho^{1/2}[1/\epsilon(\infty) - 1/\epsilon(0)]^{1/2}$.

6.5 RELAXATION TIMES

With the matrix elements listed in Table 6.1, momentum relaxation times can be calculated from (6.84) for the various scattering mechanisms. Assuming isotropic parabolic energy bands, ionized impurity scattering can be described by the Brooks–Herring equation [H. Brooks, *Adv. Electron. Electron Phys.* 7, 158 (1955)],

$$\frac{1}{\tau_{II}(x)} = \frac{2.41Z^2N_I}{\epsilon_r^2(0)T^{3/2}}g(n^*, T, x)\left(\frac{m}{m^*}\right)^{1/2}x^{-3/2}\ \text{second}^{-1} \tag{6.105}$$

where the screening term

$$g(n^*, T, x) = \ln(1+b) - \frac{b}{1+b} \tag{6.106}$$

and

$$b = 4.31\times10^{13}\frac{\epsilon_r(0)T^2}{n^*}\left(\frac{m^*}{m}\right)x \tag{6.107}$$

In these equations N_I is the total ionized impurity concentration in cm^{-3} and n^* is the effective screening concentration in cm^{-3} given by (6.58).

For neutral impurity scattering, we use Erginsoy's result [C. Erginsoy, *Phys. Rev. 79,* 1013 (1950)],

$$\frac{1}{\tau_{NI}} = 1.22 \times 10^{-7}\epsilon_r(0)N_N\left(\frac{m}{m^*}\right)^2 \text{ second}^{-1} \tag{6.108}$$

where N_N is the total neutral impurity concentration in cm^{-3}. Notice that this momentum relaxation time is independent of the carrier energy, x. Usually, neutral impurities have an appreciable effect on carrier scattering only for relatively uncompensated samples at low temperatures.

The momentum relaxation time for deformation potential scattering by acoustic phonons was first calculated by Bardeen and Shockley (J. Bardeen and W. Shockley, *Phys. Rev. 80,* 72 (1950)]. Their result is

$$\frac{1}{\tau_{DA}(x)} = \frac{4.17 \times 10^{19}\mathscr{E}_A^2 T^{3/2}}{C_l}\left(\frac{m^*}{m}\right)^{3/2} x^{1/2} \text{ second}^{-1} \tag{6.109}$$

for $\mathscr{E}_A$ in eV and C_l in dyn/cm^2. C_l is the spherically averaged longitudinal elastic constant indicated by (6.111) below.

For materials with no inversion symmetry, the acoustic phonons also scatter carriers by means of a piezoelectric interaction. A momentum relaxation time for this process was first formulated by Meijer and Polder [H. J. G. Meijer and D. Polder, *Physica 19,* 255 (1953)]. With spherical averaging of the elastic and piezoelectric constants over a cubic crystal structure [J. D. Zook, *Phys. Rev. 136,* A849 (1964)], this is given by

$$\frac{1}{\tau_{PA}(x)} = 1.05 \times 10^7 h_{14}^2\left[\left(\frac{3}{C_l} + \frac{4}{C_t}\right)\right] T^{1/2}\left(\frac{m^*}{m}\right)^{1/2} x^{-1/2} \text{ second}^{-1} \tag{6.110}$$

In (6.110) $h_{14} = e_{14}/\epsilon(0)$ is the piezoelectric constant in V/cm and the average longitudinal and transverse elastic constants are

$$C_l = \tfrac{1}{5}(3C_{11} + 2C_{12} + 4C_{44}) \tag{6.111}$$

and

$$C_t = \tfrac{1}{5}(C_{11} - C_{12} + 3C_{44}) \tag{6.112}$$

in dyn/cm^2. For a hexagonal crystal structure the momentum relaxation time is anisotropic.

For impurities and acoustic phonons the scattering processes are, to a good approximation, elastic. For optical phonons, however, the phonon energy is comparable to the thermal energy of the carriers and the scattering processes are inelastic. Despite this, a momentum relaxation time can still be defined for deformation potential scattering by optical phonons [W. A. Harrison, *Phys. Rev. 104,* 1281 (1956)]. This is given by

$$\frac{1}{\tau_{DO}(x)} = \frac{2.07 \times 10^{19}\mathscr{E}_A^2 T^{1/2}\theta}{C_l[\exp(\theta/T) - 1]}\left(\frac{m^*}{m}\right)^{3/2}\left[\left(x + \frac{\theta}{T}\right)^{1/2} + \exp\left(\frac{\theta}{T}\right)\left(x - \frac{\theta}{T}\right)^{1/2}\right] \text{ second}^{-1} \tag{6.113}$$

for $\mathscr{E}_A$ in eV and C_l in dyn/cm^2. θ in this equation is the longitudinal optical phonon temperature,

$$\theta \equiv \frac{\hbar\omega_{LO}}{k} \tag{6.114}$$

$\mathscr{E}_A$ is the acoustic phonon deformation potential constant, which is related to the optical phonon deformation potential constant, D, by

$$\mathscr{E}_A^2 = \frac{C_l D^2}{\rho\omega_{LO}^2} \tag{6.115}$$

For polar scattering of carriers by optical phonons a universal relaxation time can be defined only for temperatures much less than or much greater than the optical phonon temperature. It is thus necessary to use a variational method to solve the Boltzmann equation and determine the carrier scattering. However, Ehrenreich [H. Ehrenreich, *J. Appl. Phys. 32,* 2155 (1961)] has developed a relaxation time based on a variational calculation for polar scattering which gives the correct solutions to the Boltzmann equation at low and high temperatures. This is given by,

$$\frac{1}{\tau_{PO}(x)} = \frac{1.04 \times 10^{14}[\epsilon_r(0) - \epsilon_r(\infty)]\theta^{1/2}(\theta/T)^r}{\epsilon_r(0)\epsilon_r(\infty)[\exp(\theta/T) - 1]}\left(\frac{m^*}{m}\right)^{1/2} x^{-r} \text{ second}^{-1} \tag{6.116}$$

where r varies with (θ/T) as shown in Fig. 6.6.

The results of this section for the momentum relaxation times are summarized in Table 6.2 in the form

$$\tau_i(x) = \tau_i x^{r_i} \tag{6.117}$$

These expressions are valid only for scattering in isotropic parabolic energy bands. For scattering in more complex bands, see D. L. Rode, *Semiconductors and Semimetals,* Vol. 10, *Transport Phenomena,* ed. R. K. Willardson and A. C. Beer (New York: Academic Press, 1975) or J. D. Wiley (ibid.). It should also be noted that the momentum relaxation times were derived under the assumption that screening can be neglected for *phonon* scattering. This is usually a good assumption for samples with nondegenerate doping.

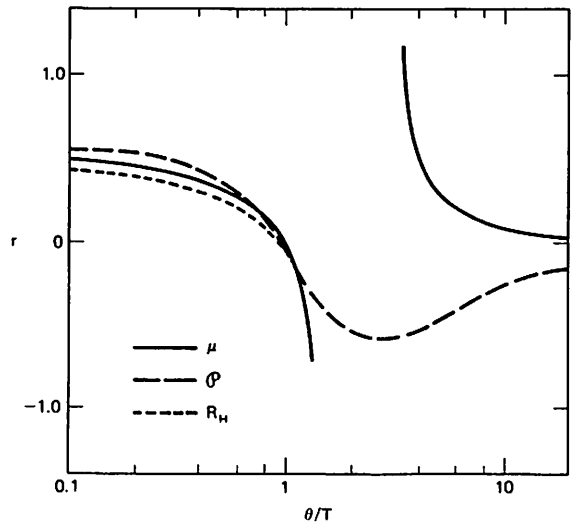


Figure 6.6 Variation of the parameter r in Equation (6.116) with (θ/T) obtained by equating variational solutions for the mobility, μ, thermoelectric power, $\mathcal{P}$, and Hall coefficient, R_H, with the corresponding expressions in the relaxation time approximation. [From H. Ehrenreich, General Electric Research Lab. Rep. No. 61-RL-(27626), June 1961.]

TABLE 6.2 Momentum Relaxation Times and Reduced Energy Dependence for Materials with Isotropic Parabolic Bands[a]

Scattering Mechanisms	τ_i (sec)	r_i
Impurities		
Ionized	$\dfrac{0.414\epsilon_r^2(0)T^{3/2}}{Z^2N_I(\text{cm}^{-3})g(n^*, T, x)}\left(\dfrac{m^*}{m}\right)^{1/2}$	$\frac{3}{2}$
Neutral	$\dfrac{8.16 \times 10^6}{\epsilon_r(0)N_N(\text{cm}^{-3})}\left(\dfrac{m^*}{m}\right)^2$	0
Acoustic phonons		
Deformation potential	$\dfrac{2.40 \times 10^{-20}C_l(\text{dyn/cm}^2)}{\mathscr{E}_A^2(\text{eV})T^{3/2}}\left(\dfrac{m}{m^*}\right)^{3/2}$	$-\frac{1}{2}$
Piezoelectric	$\dfrac{9.54 \times 10^{-8}}{h_{14}^2(\text{V/cm})(3/C_l + 4/C_t)T^{1/2}}\left(\dfrac{m}{m^*}\right)^{1/2}$	$\frac{1}{2}$
Optical phonons		
Deformation potential	$\dfrac{4.83 \times 10^{-20}C_l(\text{dyn/cm}^2)[\exp(\theta/T) - 1]}{\mathscr{E}_A^2(\text{eV})T^{1/2}\theta}\left(\dfrac{m}{m^*}\right)^{3/2}$	$\cong -\frac{1}{2}$
Polar	$\dfrac{9.61 \times 10^{-15}\epsilon_r(0)\epsilon_r(\infty)[\exp(\theta/T) - 1]}{[\epsilon_r(0) - \epsilon_r(\infty)]\theta^{1/2}(\theta/T)^r}\left(\dfrac{m}{m^*}\right)^{1/2}$	$r\left(\dfrac{\theta}{T}\right)$

[a] N_I = concentration of ionized impurities; $g(n^*, T, x) = \ln(1 + b) - b/(1 + b)$; $b = 4.31 \times 10^{13}[\epsilon_r(0)T^2/n^*(\text{cm}^{-3})](m^*/m)x$; N_N = concentration of neutral impurities; $C_l = \frac{1}{3}(3C_{11} + 2C_{12} + 4C_{44})$; $C_t = \frac{1}{3}(C_{11} - C_{12} + 3C_{44})$; $\theta = \hbar\omega_{LO}/k$.

6.6 COMBINED SCATTERING

In most calculations of transport properties it is necessary to consider several scattering processes at the same time. If these scattering mechanisms are independent of one another, the matrix elements or differential scattering cross sections for each process can be added to obtain the total scattering. In the relaxation time approximation we see from (6.84) or (6.87) that this is equivalent to adding the reciprocal times for each process,

$$\frac{1}{\tau_m(x)} = \sum_i \frac{1}{\tau_i(x)} \tag{6.118}$$

where the $\tau_i(x)$ are given by (6.105), (6.108), (6.109), (6.110), (6.113), and/or (6.116). The desired transport property is then obtained by averaging the appropriate expression involving $\tau_m(x)$ over the electron distribution. From (5.54) this procedure is

$$\langle \tau_m^s(x)x^t \rangle = \frac{2}{3}\frac{\int_0^\infty \tau_m^s(x)(-\partial f_0/\partial x)x^{t-3/2}\,dx}{\int_0^\infty f_0 x^{1/2}\,dx} \tag{6.119}$$

For some combinations of scattering mechanisms it is possible to evaluate (6.119) analytically. The usual procedure is to integrate the numerator by parts and obtain a solution in terms of Fermi–Dirac integrals of order j, which are tabulated in Appendix B. For problems involving ionized impurity scattering this procedure is complicated by the energy or x dependence of the screening term, $g(n^*, T, x)$, given by (6.106) and (6.107). Since it is a slowly varying function of x, however, reasonable approximations can be made. The usual procedure is to evaluate $g(n^*, T, x)$ at a constant energy, $x = x_m$, and remove it from the integral. The value of x_m is determined by the condition that the integrand remaining after the removal of $g(n^*, T, x_m)$ be a maximum [E. M. Conwell and V. F. Weisskopf, *Phys. Rev.* *77*, 388 (1950)]. Typically, x_m has a value of about 3 or so.

For most combinations of scattering mechanisms it is necessary to evaluate (6.119) numerically. As an example of this, the temperature dependence of the mobility,

$$\mu = \frac{q\langle \tau_m \rangle}{m^*} \tag{5.50}$$

for high-purity n-type GaAs is shown in Fig. 6.7. Here the mobility for each relevant scattering mechanism was calculated separately from (5.50), (6.105), (6.108), (6.109), (6.110), (6.116), and (6.119) and then combined to compare with experimental data.

As can be seen, the mobility of this GaAs sample is dominated by ionized impurity scattering at low temperatures and by polar optical phonon

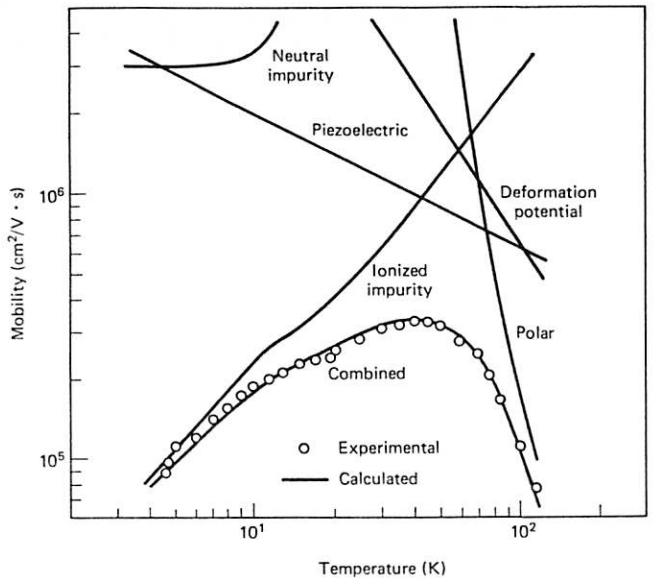


Figure 6.7 Temperature dependence of the mobility for n-type GaAs showing the separate and combined scattering processes. [From C. M. Wolfe, G. E. Stillman, and W. T. Lindley, *J. Appl. Phys. 41*, 3088 (1970).]

scattering at high temperatures. Deformation potential optical phonon scattering is not important for Γ conduction bands [H. Ehrenreich and A. W. Overhauser, *Phys. Rev. 104*, 331 (1956)]. This mobility behavior is typical for polar semiconductors. In these calculations the singly ionized N_I in (6.105) is given by

$$N_I = n + 2N_a \tag{6.120}$$

n^* in (6.107) by (6.61), and N_N in (6.108) by

$$N_N = N_d - N_a - n \tag{6.121}$$

N_d and N_a were obtained by analyzing the experimental temperature dependence of n with (4.90). For n-type material this equation is

$$\frac{n(n + N_a)}{N_d - N_a - n} = \frac{N_c}{g_d} \exp\left(\frac{-\Delta\mathscr{E}_d}{kt}\right) \tag{6.122}$$

The other parameters required in the analysis are typically obtained from other, independent measurements. These are listed for GaAs in Table 6.3 together with the appropriate parameters for other materials for which the analysis above is valid.

TABLE 6.3 Parameters for Calculating the Transport Properties of n-Type Semiconductors with Isotropic Parabolic Bands

Material	$\frac{m^*}{m}$	$\epsilon_r(0)$	$\epsilon_r(\infty)$	θ (K)	$\mathscr{E}_A$ (eV)	C_l (10^{12} dyn/cm²)	$h\bar{\tau}_4\left(\frac{3}{C_l}+\frac{4}{C_t}\right)$ (10^3 V²/dyn)
→ GaN	0.218	9.87	5.80	1044	8.4	2.65	18.32
GaP	0.13	11.10	9.11	580	13.0	1.66	1.15
→ GaAs	0.067	12.53	10.90	423	6.3	1.44	2.04
GaSb	0.042	15.69	14.44	346	8.3	1.04	
InP	0.082	12.38	9.55	497	6.8	1.21	0.137
InAs	0.025	14.54	11.74	337	5.8	1.0	0.192
InSb	0.0125	17.64	15.75	274	7.2	0.79	0.409
ZnS	0.312	8.32	5.13	506	4.9	1.28	6.87
ZnSe	0.183	9.20	6.20	360	4.2	1.03	0.620
ZnTe	0.159	9.67	7.28	297	3.5	0.84	0.218
CdS	0.208	8.58	5.26	428	3.3	0.85	32.5
CdSe	0.130	9.40	6.10	303	3.7	0.74	16.7
CdTe	0.096	10.76	7.21	246	4.0	0.70	0.445
HgSe	0.0265	25.6	12.0	268	4	0.80	0.445
HgTe	0.0244	20.0	14.0	199	4	0.61	0.445
PbS		175	17	300	20		
PbSe		250	24	190	24	0.71	
PbTe		400	33	160	25		

PROBLEMS

6.1. In a collision with an acoustic phonon, show that an electron with initial velocity v_i will gain or lose at most only

$$\frac{4u_s}{v_i} - 4\left(\frac{u_s}{V_i}\right)^2$$

of its initial energy, where u_s is the sound velocity.

6.2. An acoustic wave of the form $A \exp[i(\mathbf{q}\cdot\mathbf{r} - \omega t)]$ propagates through an n-type semiconductor with a parabolic band where it produces a variation in the energy of the electrons

$$\mathscr{E} = \mathscr{E}_1 A \exp[i(\mathbf{q}\cdot\mathbf{r} - \omega t)]$$

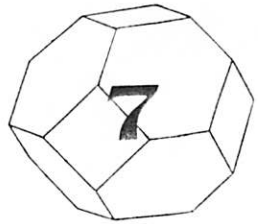
Since the force exerted on an electron is $\mathbf{F} = -\nabla_r\mathscr{E}$, show that in the relaxation time approximation, a good approximation to the electron distribution is

$$f = f_0 + \frac{\partial f_0}{\partial\mathscr{E}} \frac{i\tau_m \mathbf{v}\cdot\mathbf{q}\mathscr{E}}{1 + i\tau_m \mathbf{v}\cdot\mathbf{q}}$$

Does this distribution provide conduction?

6.3. (a) Evaluate r_H for ionized impurity scattering using the momentum relaxation time determined in the Brooks–Herring approximation.
(b) Plot the temperature variation of r_H using parameters appropriate to GaAs.

6.4. Use the Rutherford scattering cross section to derive the mobility for ionized impurity scattering in the Conwell–Weisskopf approximation.
(a) Discuss the validity of the Born approximation for ionized impurity scattering.
(b) Discuss the differences between C-W and B-H approximation, particularly in the temperature range where there is carrier freeze-out.



Optical Properties

When light is incident on a semiconductor, the optical phenomena of absorption, reflection, and transmission are observed. From these optical effects, we obtain much of the information we have concerning the energy band structure and electronic processes in semiconductors. Figure 7.1 shows a hypothetical absorption spectrum as a function of photon energy for a typical semiconductor. As can be seen, a number of processes can contribute to absorption. At high energies photons are absorbed by the transitions of electrons from filled valence band states to empty conduction band states. For energies just below the lowest forbidden energy gap, radiation is absorbed due to the formation of excitons and electron transitions between band and impurity states. The transitions of free carriers within energy bands produce an absorption continuum which increases with decreasing photon energy. Also, the crystalline lattice itself can absorb radiation, with the energy being given off in optical phonons. Finally, at low energies, or long wavelengths, electronic transitions can be observed between impurities and their associated bands.

Many of these processes have important technological applications. For example, intrinsic photodetectors utilize band-to-band absorption, while semiconductor lasers generally operate by means of transitions between impurity and band states. In this chapter we examine these optical processes in detail.

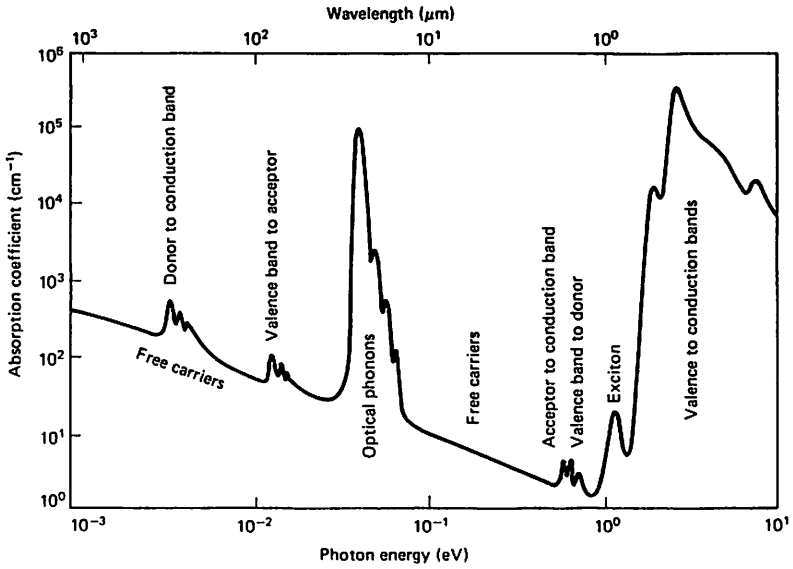


Figure 7.1 Hypothetical absorption spectrum for a typical semiconductor as a function of photon energy.

7.1 ELECTRON–PHOTON INTERACTION

To examine the interaction between an electron and a photon, let us represent the photon by a vector potential, **A**, defined by

$$\mathbf{E} = -\frac{\partial}{\partial t}\mathbf{A} \tag{7.1}$$

$$\mu\mathbf{H} = \nabla_r \times \mathbf{A} \tag{7.2}$$

and

$$\nabla_r \cdot \mathbf{A} = 0 \tag{7.3}$$

We will take the vector potential to have the form of a plane wave,

$$\mathbf{A} = \tfrac{1}{2}A\mathbf{a}\exp\left[i(\mathbf{q}\cdot\mathbf{r} - \omega t)\right] + \tfrac{1}{2}A\mathbf{a}\exp\left[-i(\mathbf{q}\cdot\mathbf{r} - \omega t)\right] \tag{7.4}$$

where **a** is the unit polarization vector in the direction of **E** and **q** is the wavevector. For a photon in a semiconductor, the wavevector is related to the frequency by

$$|\mathbf{q}| = \frac{\omega\eta}{c} \tag{7.5}$$

where c is the velocity of light and η is the refractive index of the material. The energy of the photon is simply

$$\mathscr{E} = \hbar\omega \tag{7.6}$$

Also, since (7.4) is a transverse electromagnetic wave

$$\mathbf{a}\cdot\mathbf{q} = 0 \tag{7.7}$$

The classical Hamiltonian of an electron with wavevector **k** interacting with a light wave of vector potential **A** is

$$\mathbf{H} = \frac{1}{2m}(\hbar\mathbf{k} - q\mathbf{A})^2 \tag{7.8}$$

Expanding (7.8), we have

$$\mathbf{H} = \frac{1}{2m}(\hbar^2\mathbf{k}^2 - \hbar q\mathbf{k}\cdot\mathbf{A} - \hbar q\mathbf{A}\cdot\mathbf{k} + q^2\mathbf{A}^2) \tag{7.9}$$

Using the operator form of **k** and (7.3), we obtain the quantum mechanical Hamiltonian

$$\mathbf{H} = \frac{1}{2m}(-\hbar^2\nabla_r^2 + i2q\hbar\mathbf{A}\cdot\nabla_r + q^2\mathbf{A}^2) \tag{7.10}$$

For low light levels the term with the vector potential squared can be neglected, to obtain

$$\mathbf{H} = \frac{-\hbar^2}{2m}\nabla_r^2 + \frac{iq\hbar}{m}\mathbf{A}\cdot\nabla_r$$

$$= \mathbf{H}_0 + \mathbf{H}' \tag{7.11}$$

That is, the Hamiltonian consists of a term $\mathbf{H}_0$ which corresponds to the unperturbed electron energy and a term $\mathbf{H}'$ due to the electron–photon interaction.

Since this interaction can result in a change of state for the electron with time, it is necessary to solve the time-dependent Schrödinger equation,

$$(\mathbf{H}_0 + \mathbf{H}')\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{7.12}$$

Let us construct wavefunction solutions to (7.12) which are linear combinations of the wavefunctions, Ψ_n, for the unperturbed time-independent system

$$\Psi = \sum_n A_n(t)\psi_n \exp\left(\frac{-i\mathscr{E}_n t}{\hbar}\right) \tag{7.13}$$

where the ψ_n satisfy

$$\mathbf{H}_0\psi_n = \mathscr{E}_n\psi_n \tag{7.14}$$

Using (7.13) in (7.12), we obtain

$$\sum_n A_n(\mathbf{H}_0\psi_n + \mathbf{H}'\psi_n)\exp\left(\frac{-i\mathscr{E}_n t}{\hbar}\right) = \sum_n \left(A_n\mathscr{E}_n\psi_n + i\hbar\frac{dA_n}{dt}\psi_n\right)\exp\left(\frac{-i\mathscr{E}_n t}{\hbar}\right) \tag{7.15}$$

From (7.14) the first term on the left is equal to the first term on the right in (7.15), and

$$i\hbar\sum_n \frac{dA_n}{dt}\psi_n \exp\left(\frac{-i\mathscr{E}_n t}{\hbar}\right) = \sum_n A_n\mathbf{H}'\psi_n \exp\left(\frac{-i\mathscr{E}_n t}{\hbar}\right) \tag{7.16}$$

If we multiply (7.16) by $\psi_m^*\exp(i\mathscr{E}_m t/\hbar)$ and integrate over the volume of the crystal, we have

$$i\hbar\sum_n \frac{dA_n}{dt}\exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_n)t}{\hbar}\right]\int_V \psi_m^*\psi_n\, d\mathbf{r} = \sum_n A_n \exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_n)t}{\hbar}\right]\int_V \psi_m^*\mathbf{H}'\psi_n\, d\mathbf{r} \tag{7.17}$$

Because of the orthogonality of the unperturbed wavefunctions,

$$\int_V \psi_m^*\psi_n\, d\mathbf{r} = N\delta_{mn} \tag{3.14}$$

(7.17) reduces to

$$i\hbar\frac{dA_m}{dt} = \sum_n A_n H_{mn}(t)\exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_n)t}{\hbar}\right] \tag{7.18}$$

where

$$H_{mn}(t) \equiv \frac{1}{N}\int_V \psi_m^*\mathbf{H}'\psi_n\, d\mathbf{r} \tag{7.19}$$

is the *matrix element* for an electron transition from state n with energy $\mathscr{E}_n$ to state m with energy $\mathscr{E}_m$ and N is the number of primitive unit cells in the crystal. Equation (7.18) is an exact differential equation for the time-dependent coefficients of the wavefunction of (7.13).

To solve for the coefficients, A_m, we will use first-order perturbation theory. Let us assume that at $t = 0$, the system starts in a time-independent state with energy $\mathscr{E}_0$ and we take

$$A_0(0) = 1, \qquad A_n(0) = 0 \tag{7.20}$$

Equation (7.18) is now

$$i\hbar\frac{dA_m}{dt} = H_{mo}(t)\exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_0)t}{\hbar}\right] \tag{7.21}$$

Integrating (7.21) with respect to time, the coefficients are given by the expression

$$A_m(t) = \frac{1}{i\hbar}\int_0^t H_{m0}(t)\exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_0)t}{\hbar}\right]dt \tag{7.22}$$

Since the probability of a transition from state 0 to state m is given by $|A_m(t)|^2$, we must now evaluate the integral in (7.22).

From (7.19) and (7.11) we have

$$H_{m0}(t) = \frac{iq\hbar}{mN}\int_V \psi_m^*(\mathbf{A}\cdot\nabla_r)\psi_0\, d\mathbf{r} \tag{7.23}$$

Equation (7.4) for the vector potential, $\mathbf{A}$, has two terms: the first corresponds to stimulated absorption and the second corresponds to stimulated emission. For our current purpose, we ignore the stimulated emission term, so that the matrix element is

$$H_{m0}(t) = \frac{iq\hbar A}{2mN}\exp(-i\omega t)\int_V \psi_m^*\exp(i\mathbf{q}\cdot\mathbf{r})(\mathbf{a}\cdot\nabla_r)\psi_0\, d\mathbf{r} \tag{7.24}$$

$$H_{m0}(t) = H_{m0}\exp(-i\omega t) \tag{7.25}$$

In (7.25) we have separated the matrix element into a time-dependent term and the time-independent term

$$H_{m0} = \frac{iq\hbar A}{2mN}\int_V \psi_m^*\exp(i\mathbf{q}\cdot\mathbf{r})(\mathbf{a}\cdot\nabla_r)\psi_0\, d\mathbf{r} \tag{7.26}$$

Equation (7.22) now has the form

$$A_m(t) = \frac{H_{m0}}{i\hbar}\int_0^t \exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega)t}{\hbar}\right]dt \tag{7.27}$$

which can be integrated to obtain

$$A_m(t) = \frac{H_{m0}}{(\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega)}\left\{1 - \exp\left[\frac{i(\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega)t}{\hbar}\right]\right\} \tag{7.28}$$

The transition probability is therefore

$$|A_m(t)|^2 = \frac{4|H_{m0}|^2\sin^2\dfrac{(\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega)t}{2\hbar}}{(\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega)^2} \tag{7.29}$$

If we let

$$x = \frac{\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega}{2\hbar} \tag{7.30}$$

the transition probability takes the form

$$|A_m(t)|^2 = \frac{|H_{m0}|^2}{\hbar^2}\frac{\sin^2 xt}{x^2} \tag{7.31}$$

Figure 7.2 shows a plot of the term $(\sin^2 xt)/x^2$ in (7.31). As can be seen, the height of this term increases with t^2, while the width is inversely proportional to t. The area under the curve is

$$\int_{-\infty}^{\infty} \frac{\sin^2 xt}{x^2}\,dx = \pi t \tag{7.32}$$

Thus, for times sufficiently long that the transition is completed, we can make the approximation

$$\frac{\sin^2 xt}{x^2} \simeq \pi t\delta(x) \tag{7.33}$$

where $\delta(x)$ is the Dirac delta function [L. I. Schiff, *Quantum Mechanics* (New York: McGraw-Hill, 1955), p. 197]. The transition probability is then

$$|A_m(t)|^2 = \frac{\pi|H_{m0}|^2 t}{\hbar^2}\,\delta\left(\frac{\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega}{2\hbar}\right) \tag{7.34}$$

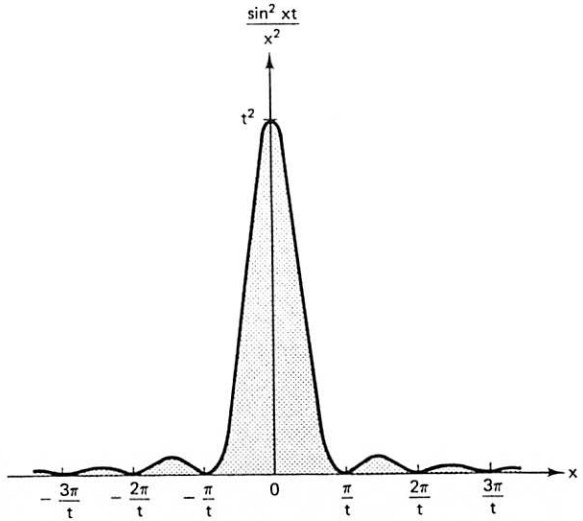


Figure 7.2 Plot of $(\sin^2 xt)/x^2$ for the transition probability in (7.31).

Since

$$\delta(ax) = \frac{1}{a}\,\delta(x) \tag{7.35}$$

(7.34) becomes

$$|A_m(t)|^2 = \frac{2\pi|H_{m0}|^2 t}{\hbar}\,\delta(\mathscr{E}_m - \mathscr{E}_0 - \hbar\omega) \tag{7.36}$$

Equation (7.36) tells us that the probability of an electron making a transition from state 0 with energy $\mathscr{E}_0$ to state m with energy $\mathscr{E}_m$ is zero unless the photon energy, $\hbar\omega$, is equal to the difference in energy between the two states. That is, energy must be conserved. We also see that the transition probability increases with time. With (7.36) the transition probability can be evaluated when the *time-dependent* matrix element, H_{m0}, for a given transition is known.

7.2 BAND-TO-BAND ABSORPTION

7.2.1 Direct Transitions

Let us first examine the matrix element for a direct electron transition from a valence band state with wavevector $\mathbf{k}$ to a conduction band state with wavevector $\mathbf{k}'$. As indicated in Fig. 7.3, the initial and final states for a direct transition are uniquely determined by the photon energy, $\hbar\omega$, and the energy band structure. For simplicity we assume parabolic bands. From (7.26) the matrix element is

$$H_{k'k} = \frac{iq\hbar A}{2mN}\int_V \psi_{k'}^*\exp(i\mathbf{q}\cdot\mathbf{r})(\mathbf{a}\cdot\nabla_r)\psi_k\,d\mathbf{r} \tag{7.37}$$

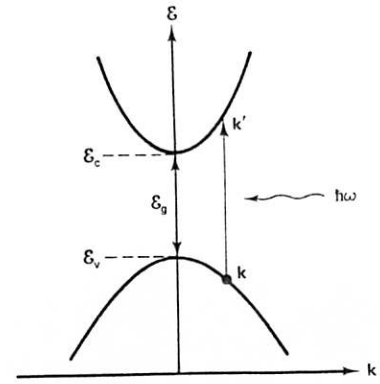


Figure 7.3 Direct optically induced transition of an electron from a valence band state with wavevector $\mathbf{k}$ to a conduction band state with wavevector $\mathbf{k}'$.

where ψ_k and $\psi_{k'}$ are the wavefunctions of the valence and conduction band states, respectively. According to (2.10), these wavefunctions have the form

$$\psi_k = \exp(i\mathbf{k}\cdot\mathbf{r})u_k(\mathbf{r}) \tag{7.38}$$

Using (7.38) in (7.37) and performing the operation, ∇_r, the matrix element is

$$H_{k'k} = \frac{iq\hbar A}{2mN}\int_V \exp[i(\mathbf{k} - \mathbf{k}' + \mathbf{q})\cdot\mathbf{r}]u_{k'}^*[\mathbf{a}\cdot\nabla_r u_k + i(\mathbf{a}\cdot\mathbf{k})u_k]\,d\mathbf{r} \tag{7.39}$$

Since the Bloch functions are periodic in $\mathbf{R}$, the integral over the crystal volume can be taken as a sum of integrals over the primitive unit cell,

$$H_{k'k} = \frac{iq\hbar A}{2mN}\sum_R \exp[i(\mathbf{k} - \mathbf{k}' + \mathbf{q})\cdot\mathbf{R}]\int_\Omega u_{k'}^*[\mathbf{a}\cdot\nabla_r u_k + i(\mathbf{a}\cdot\mathbf{k})u_k]\,d\mathbf{r} \tag{7.40}$$

For conservation of wavevector the sum over $\mathbf{R}$ in (7.40) is simply N and the matrix element is, finally,

$$H_{k'k} = \frac{iq\hbar A}{2m}\int_\Omega u_{k'}^*[\mathbf{a}\cdot\nabla_r u_k + i(\mathbf{a}\cdot\mathbf{k})u_k]\,d\mathbf{r} \tag{7.41}$$

Thus the matrix element consists of two terms: one involving $\nabla_r u_k$ and the other with u_k. Because of the orthogonality of the Bloch function between the bands, the term involving u_k would be zero for $\mathbf{k}$ equal to $\mathbf{k}'$. Therefore, taking into account the small wavevector of the photon, this term is much smaller than the term with $\nabla_r u_k$. For this reason the $\nabla_r u_k$ term in the matrix element of (7.41) is referred to as *allowed*, while the u_k term is referred to as *forbidden*.

Let us first consider the allowed term only. The matrix element for an allowed direct transition is

$$H_{k'k} = \frac{iq\hbar A}{2m}\int_\Omega u_{k'}^*(\mathbf{a}\cdot\nabla_r u_k)\,d\mathbf{r} \tag{7.42}$$

Since

$$\mathbf{p} = \hbar\mathbf{k} = -i\hbar\nabla_r \tag{7.43}$$

we will define a crystal momentum matrix element by

$$\mathbf{p}_{k'k} \equiv -i\hbar\int_\Omega u_{k'}^*\,\nabla_r u_k\,d\mathbf{r} \tag{7.44}$$

The matrix element in (7.42) then takes the form

$$H_{k'k} = \frac{-qA}{2m}(\mathbf{a}\cdot\mathbf{p}_{k'k}) \tag{7.45}$$

and from (7.36) the probability in the entire crystal that an electron will make a transition from a state with wavevector $\mathbf{k}$ to a state with wavevector $\mathbf{k}'$ is

$$|A_{k'}(t)|^2 = \frac{2\pi t}{\hbar}\left(\frac{qA}{2m}\right)^2(\mathbf{a}\cdot\mathbf{p}_{k'k})^2\delta(\mathscr{E}_{k'} - \mathscr{E}_k - \hbar\omega) \tag{7.46}$$

The transition *rate* over the whole crystal is just the transition probability (7.46) divided by t.

To determine the total probability for a band-to-band transition, we assume $\hbar\omega$ to be monochromatic and sum (7.46) over all N allowed values of $\mathbf{k}$. Taking into account the volume occupied by each value of $\mathbf{k}$ from (2.29), the probability that an initial valence band state will be occupied from (4.41), and the probability that a final conduction band state will be unoccupied from (4.42), the total probability for a band-to-band transition is

$$P = \frac{2V}{(2\pi)^3}\int_{\Omega_K}|A_{k'}(t)|^2 f_0(1 - f_0)\,d\mathbf{k} \tag{7.47}$$

The factor of 2 in (7.47) accounts for a possible change of spin during absorption. From (7.47) the transition probability per unit volume per unit time or the transition *rate* per unit volume is

$$r = \frac{2}{(2\pi)^3}\int_{\Omega_K}\frac{|A_{k'}(t)|^2}{t}f_0(1 - f_0)\,d\mathbf{k} \tag{7.48}$$

Using (7.46) in (7.48) we have

$$r = \frac{2}{4\pi^2\hbar}\left(\frac{qA}{2m}\right)^2\int_{\Omega_K}(\mathbf{a}\cdot\mathbf{p}_{k'k})^2\delta(\mathscr{E}_{k'} - \mathscr{E}_k - \hbar\omega)f_0(1 - f_0)\,d\mathbf{k} \tag{7.49}$$

Assuming parabolic bands, we can see from Fig. 7.3 that

$$\begin{aligned}\mathscr{E}_{k'} - \mathscr{E}_k &= \mathscr{E}_g + \frac{\hbar^2k'^2}{2m_e} + \frac{\hbar^2k^2}{2m_h}\\ &\simeq \mathscr{E}_g + \frac{\hbar^2k^2}{2m_r}\end{aligned} \tag{7.50}$$

where m_r is the reduced mass of the electron and hole,

$$m_r = \frac{m_e m_h}{m_e + m_h} \tag{7.51}$$

Compared with the delta function, the other terms under the integral in (7.49) are slowly varying functions of k. If we take them out of the integral, (7.49) becomes

$$r = \frac{q^2A^2\omega f}{16\pi^2 m}f_0(1 - f_0)\int_{\Omega_K}\delta\left(\mathscr{E}_g + \frac{\hbar^2k^2}{2m_r} - \hbar\omega\right)4\pi k^2\,dk \tag{7.52}$$

where we have defined a dimensionless *oscillator strength*,

$$f \equiv \frac{2(\mathbf{a}\cdot\mathbf{p}_{k'k})^2}{\hbar m\omega} \tag{7.53}$$

The oscillator strength, f, in (7.53) is given approximately by the f-sum rule [A. H. Wilson, *The Theory of Metals* (Cambridge: Cambridge University Press, 1954), p. 47] as

$$f \simeq 1 + \frac{m}{m_h} \tag{7.54}$$

The integral in (7.52) can be evaluated as

$$\begin{aligned} 4\pi \int_{\Omega_K} \delta\left(\mathscr{E}_g + \frac{\hbar k^2}{2m_r} - \hbar\omega\right) k^2\, dk \\ &= 4\pi \frac{d}{d(\hbar\omega)} \left| \int_{\Omega_K} k^2\, dk \right|_{\hbar\omega} = \mathscr{E}_g + \frac{\hbar^2 k^2}{2m_r} \\ &= \frac{4\pi}{3}\left(\frac{2m_r}{\hbar^2}\right)^{3/2} \frac{d}{d(\hbar\omega)} (\hbar\omega - \mathscr{E}_g)^{3/2} \\ &= \frac{2\pi}{\hbar^3}(2m_r)^{3/2}(\hbar\omega - \mathscr{E}_g)^{1/2} \end{aligned} \tag{7.55}$$

With (7.52) and (7.55) the total allowed transition rate per unit volume between the conduction and valence band is

$$r = \frac{q^2A^2\omega f(2m_r)^{3/2}}{8\pi\hbar^3 m} f_0(1 - f_0)(\hbar\omega - \mathscr{E}_g)^{1/2} \tag{7.56}$$

The absorption coefficient, α, can be obtained from (7.56) by means of

$$\alpha = \frac{r}{\Phi} \tag{7.57}$$

where Φ is the quantum flux or the number of photons crossing the unit area in unit time. From (7.57) we can see that α has units of reciprocal length. The quantum flux, Φ, can be determined from the average value of the Poynting vector $\mathbf{S}$ of the radiation, which is the energy crossing unit area in unit time, by

$$\Phi = \frac{\langle \mathbf{S} \rangle}{\hbar\omega} \tag{7.58}$$

where

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \tag{7.59}$$

From (7.1), (7.2), and (7.4),

$$\begin{aligned} \mathbf{E} &= A\omega\mathbf{a} \sin(\mathbf{q}\cdot\mathbf{r} - \omega t) \\ \mu\mathbf{H} &= -A(\mathbf{q} \times \mathbf{a}) \sin(\mathbf{q}\cdot\mathbf{r} - \omega t) \end{aligned} \tag{7.60}$$

The average value of $\mathbf{S}$ over a period is, therefore,

$$\langle \mathbf{S} \rangle = \frac{1}{2} |\, \mathbf{q} \,| \frac{\omega A^2}{\mu} \tag{7.61}$$

or, using (7.5),

$$\langle \mathbf{S} \rangle = \tfrac{1}{2}\eta\epsilon_0 c\omega^2 A^2 \tag{7.62}$$

Using (7.56), (7.58), and (7.62) in (7.57), the absorption coefficient for allowed direct band-to-band transitions is

$$\alpha = \frac{q^2(2m_r)^{3/2} f}{4\pi\epsilon_0\eta c m\hbar^2} f_0(1 - f_0)(\hbar\omega - \mathscr{E}_g)^{1/2} \tag{7.63}$$

If we assume that all valence band states are full and all conduction band states are empty, α is given by

$$\alpha = 2.7 \times 10^5 \left(\frac{2m_r}{m}\right)^{3/2} \frac{f}{\eta} (\hbar\omega - \mathscr{E}_g)^{1/2}\ \text{cm}^{-1} \tag{7.64}$$

for $(\hbar\omega - \mathscr{E}_g)$ in units of eV.

The analysis of absorption for *forbidden* direct transitions is somewhat similar to that given above for allowed transitions. From (7.41) the matrix element is

$$H_{k'k} = \frac{-q\hbar A}{2m} (\mathbf{a}\cdot\mathbf{k}) \int_\Omega u_{k'}^* u_k\, d\mathbf{r} \tag{7.65}$$

Using (7.65) in (7.36) the probability for a forbidden transition between states with wavevectors $\mathbf{k}$ and $\mathbf{k}'$ is

$$|\, A_{k'}(t) \,|^2 = \frac{2\pi t}{\hbar}\left(\frac{q\hbar A}{2m}\right)^2 |\, \mathbf{a}\cdot\mathbf{k} \,|^2 f' \delta(\mathscr{E}_{k'} - \mathscr{E}_k - \hbar\omega) \tag{7.66}$$

where

$$f' = \left| \int_\Omega u_{k'}^* u_k\, d\mathbf{r} \right|^2 \tag{7.67}$$

For $\mathbf{k}$ not equal to $\mathbf{k}'$, $0 < f' \ll 1$. To determine the transition rate per unit volume, we must integrate (7.66) over the first Brillouin zone as in (7.48). The term $|\, \mathbf{a}\cdot\mathbf{k} \,|^2$ in (7.66), however, introduces an additional $\mathbf{k}$ dependence

in the integral. We will use the average value of $|\mathbf{a}\cdot\mathbf{k}|$, which is $\frac{1}{3}k^2$, to obtain

$$r = \frac{q^2A^2\hbar f'}{24\pi^2m^2} f_0(1 - f_0) \int_{\Omega_K} \delta\left(\mathscr{E}_g + \frac{\hbar^2k^2}{2m_r} - \hbar\omega\right) 4\pi k^2\, dk \quad (7.68)$$

which can be compared to (7.52) for an allowed transition.

When the integral in (7.68) is evaluated in a manner similar to (7.55), the total forbidden transition rate per unit volume is

$$r = \frac{q^2A^2f'(2m_r)^{5/2}}{12\pi\hbar^4m^2} f_0(1 - f_0)(\hbar\omega - \mathscr{E}_g)^{3/2} \quad (7.69)$$

Using (7.69), (7.58), and (7.62) in (7.57), the forbidden absorption coefficient is

$$\alpha = \frac{q^2(2m_r)^{5/2}f'}{6\pi\epsilon_0\eta cm^2\hbar^2} f_0(1 - f_0) \frac{(\hbar\omega - \mathscr{E}_g)^{3/2}}{\hbar\omega} \quad (7.70)$$

or

$$\alpha = 1.8 \times 10^5 \left(\frac{2m_r}{m}\right)^{5/2} \frac{f'}{\eta} \frac{(\hbar\omega - \mathscr{E}_g)^{3/2}}{\hbar\omega} \text{ cm}^{-1} \quad (7.71)$$

for a completely full valence band, a completely empty conduction band, and $\hbar\omega$ and $\mathscr{E}_g$ in eV. By comparing (7.71) with (7.64), we can see that the absorption coefficient for a forbidden transition is much smaller than for an allowed transition. Also, notice that the two expressions have a different dependence on $(\hbar\omega - \mathscr{E}_g)$.

7.2.2 Indirect Transitions

For direct optically induced transitions, we found that the initial and final electron states were uniquely determined by the photon energy, $\hbar\omega$. For indirect optically induced transitions, however, this is not the case. That is, an indirect transition requires the absorption or emission of a phonon to conserve the wavevector. Since the simultaneous absorption of a photon and a phonon is a higher-order process, one would expect indirect transition probabilities to be much less than those for direct transitions. However, because of the additional degree of freedom introduced by the phonon energy, $\hbar\omega_s$, transitions to many more states are possible.

This is illustrated in Fig. 7.4(a), where a valence band electron in initial state 0 can make a transition to final state 1 in the conduction band with the absorption of a photon of energy, $\hbar\omega$, and a phonon of energy, $\hbar\omega_{s1}$, and wavevector, $\mathbf{q}_{s1}$; or to final state 2 with the absorption of a photon of the same energy, $\hbar\omega$, and a phonon of energy, $\hbar\omega_{s2}$, and wavevector, $\mathbf{q}_{s2}$. Figure 7.4(b) shows similar possible transitions from initial states 1 and 2 in the valence band to a final state 0 in the conduction band. We can also envision

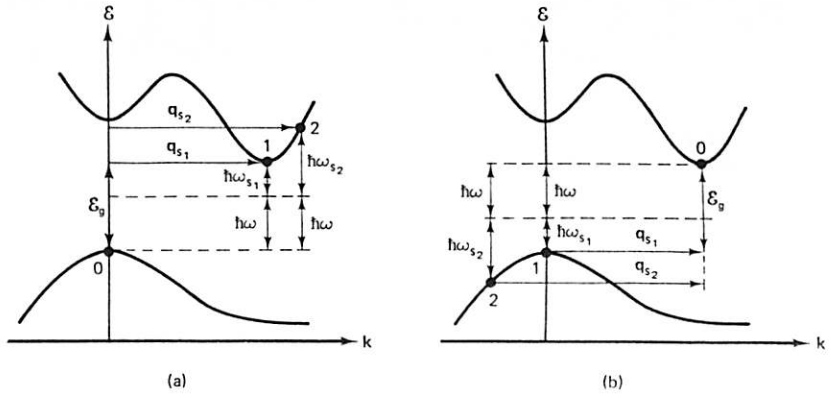


Figure 7.4 Indirect optically induced transitions of electrons (a) from initial state 0 in the valence band to final states 1 and 2 in the conduction band and (b) from initial states 1 and 2 in the valence band to final state 0 in the conduction band.

transitions involving the absorption of a photon and the emission of phonons. Thus, for a given photon energy a range of possible energies is available for indirect transitions. Notice also that indirect band-to-band absorption begins at photon energies below the bandgap such that

$$\hbar\omega \geq \mathscr{E}_g - \hbar\omega_s \quad (7.72)$$

where $\hbar\omega_s$ is the phonon energy.

As shown in Fig. 7.5, an indirect transition can be described by a direct transition from state 0 in the valence band to a short-lived virtual state I_c in the conduction band with simultaneous absorption or emission of a phonon to scatter the electron from I_c to conduction band state 1. In a similar manner, the same indirect transition can be described as a transition from 0 to a

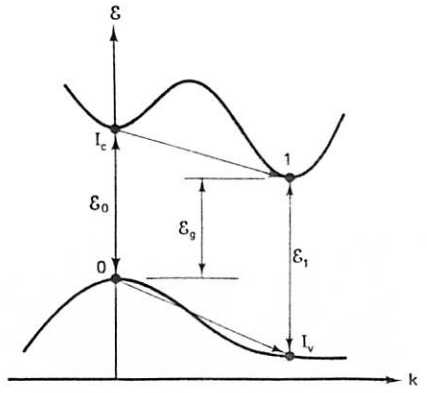


Figure 7.5 Analysis of indirect transitions by means of a direct transition from 0 to a virtual state I_c with simultaneous absorption or emission of a phonon to scatter from I_c to 1, or emission of a phonon to scatter from 0 to virtual state I_v with a simultaneous direct transition from I_v to 1.

virtual state I_v in the valence band with a simultaneous direct transition from I_v to 1. In such a treatment conservation of energy can be relaxed in the transitions to the virtual states because of the short times the electron remains in these states. Of course, energy must be conserved in the complete indirect process.

Indirect transitions described in terms of these virtual states can be analyzed by second-order perturbation theory [R. A. Smith, *Wave Mechanics of Crystalline Solids* (New York: Wiley, 1961)]. In such a treatment, the absorption coefficient for allowed indirect transitions involving conduction band virtual states is given by

$$\alpha_c(\pm\omega_s) = \frac{g_c q^2 m_h^{3/2} f_c \omega_s \mathscr{E}_0(\hbar\omega \pm \hbar\omega_s - \mathscr{E}_g)^2}{32\pi\epsilon_0 \eta c m m_e^{1/2} \hbar\omega l_c kT(\mathscr{E}_0 - \hbar\omega)^2} \frac{\pm 1}{\exp(\pm\hbar\omega_s/kT) - 1} \tag{7.73}$$

where the + and − signs are for the absorption and emission of phonons, respectively. The absorption coefficient for allowed indirect transitions involving valence band virtual states is

$$\alpha_v(\pm\omega_s) = \frac{g_c q^2 m_e^{3/2} f_v \omega_s \mathscr{E}_1(\hbar\omega \pm \hbar\omega_s - \mathscr{E}_g)^2}{32\pi\epsilon_0 \eta c m m_h^{1/2} \hbar\omega l_v kT(\mathscr{E}_1 - \hbar\omega)^2} \frac{\pm 1}{\exp(\pm\hbar\omega_s/kT) - 1} \tag{7.74}$$

In (7.73) and (7.74), g_c is the number of conduction band minima, f_c and f_v are oscillator strengths for the appropriate transitions, and l_c and l_v are the mean free paths for electron scattering in each band. The total allowed indirect absorption coefficient is

$$\alpha = \alpha_c(+\omega_s) + \alpha_c(-\omega_s) + \alpha_v(+\omega_s) + \alpha_v(-\omega_s) \tag{7.75}$$

Forbidden indirect transitions can also be analyzed by means of virtual states. In this case the absorption coefficient for transitions through conduction band virtual states is

$$\alpha_c(\pm\omega_s) = \frac{g_c q^2 m_h^{5/2} f_c' \omega_s(\hbar\omega \pm \hbar\omega_s - \mathscr{E}_g)^3}{48\pi\epsilon_0 \eta c m^2 m_e^{1/2} \hbar\omega l_c kT(\mathscr{E}_0 - \hbar\omega)^2} \frac{\pm 1}{\exp(\pm\hbar\omega_s/kT) - 1} \tag{7.76}$$

with a similar expression for transitions through valence band virtual states.

From (7.63), (7.70), (7.73), and (7.76) it can be seen that the dependence of the absorption coefficient on $(\hbar\omega - \mathscr{E}_g)$ is different for each of the band-to-band absorption processes. That is, direct allowed absorption varies as the $\frac{1}{2}$ power of $(\hbar\omega - \mathscr{E}_g)$, direct forbidden as the $\frac{3}{2}$ power, indirect allowed as the 2 power, and indirect forbidden as the 3 power. It is by this difference that the absorption processes can be distinguished experimentally.

7.2.3 Fermi Energy Dependence

Although band-to-band absorption is an intrinsic process, it can be strongly affected by the doping of the semiconductor. For lightly doped materials the fundamental absorption edge (energy at which appreciable band-to-band absorption begins) is determined by the energy gap, $\mathscr{E}_g$. For heavily doped semiconductors, however, when the Fermi energy lies in the conduction or valence bands, the position of the Fermi level must be taken into account. Figure 7.6 shows such a situation in an n-type, direct energy gap semiconductor. Since the Fermi level, ζ, is in the conduction band, essentially all states in the valence band and from the bottom of the conduction band to $\zeta - 4kT$ in the conduction band are occupied with electrons. If a photon of energy, $\hbar\omega = \mathscr{E}_g$, is incident on the material, there are no unoccupied states in the conduction band to which a transition can be made. Thus no band-to-band absorption occurs. As can be seen, no absorption is obtained until

$$\hbar\omega = \mathscr{E}_g + \frac{\hbar^2 k^2}{2}\left(\frac{1}{m_e} + \frac{1}{m_h}\right) \tag{7.77}$$

and the fundamental absorption edge has been shifted from $\hbar\omega = \mathscr{E}_g$ to this value by the doping. This is referred to as the Burstein–Moss shift [E. Burstein, *Phys. Rev. 93*, 632 (1954); T. S. Moss, *Proc. Phys. Soc. London B76*, 775 (1954)].

An expression for the fundamental absorption edge in terms of the

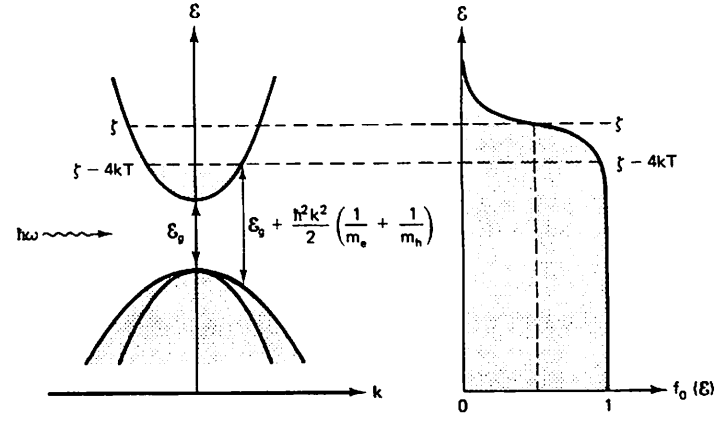


Figure 7.6 Diagram showing how the fundamental absorption edge of an n-type semiconductor is shifted to higher energy by doping.

Fermi energy can be obtained from (7.77) in the form

$$\hbar\omega = \mathscr{E}_g + \frac{\hbar^2 k^2}{2m_e}\left(1 + \frac{m_e}{m_h}\right) \tag{7.78}$$

Using (2.118), equation (7.78) becomes

$$\hbar\omega = \mathscr{E}_g + (\mathscr{E} - \mathscr{E}_c)\left(1 + \frac{m_e}{m_h}\right) \tag{7.79}$$

or

$$\hbar\omega = \mathscr{E}_g + (\zeta - \mathscr{E}_c - 4kT)\left(1 + \frac{m_e}{m_h}\right) \tag{7.80}$$

Equations (7.63) and (7.70) for direct band-to-band absorption can be modified to account for the Burstein–Moss shift by determining the probability of an unoccupied conduction band state. From (4.42) and (7.79),

$$1 - f_0 = \frac{1}{1 + \exp\left[(\zeta - \mathscr{E})/kT\right]} \tag{7.81}$$

$$1 - f_0 = \frac{1}{1 + \exp\left[\dfrac{\zeta - \mathscr{E}_c}{kT} - \dfrac{(\hbar\omega - \mathscr{E}_g)m_h}{(m_e + m_h)kT}\right]} \tag{7.82}$$

$$1 - f_0 = \frac{1}{1 + \exp\left[\dfrac{(\zeta - \mathscr{E}_c)m_e - (\hbar\omega - \mathscr{E}_g)m_r}{m_e kT}\right]} \tag{7.83}$$

where the reduced mass, m_r, is given by

$$m_r = \frac{m_e m_h}{m_e + m_h} \tag{7.84}$$

Thus the shift of absorption with doping for n-type material can be accounted for with (7.83).

7.2.4 Temperature Dependence

Another factor that controls the fundamental absorption edge is the temperature of the sample. This is reflected primarily in the expansion and contraction of the lattice with temperature and its effect on the energy gap. This dependence can be seen in Fig. 2.14. Since the temperature dependence of the energy gap varies considerably among semiconductors, it is best determined from experimental results. For Si an empirical fit to the experimental data is given by

$$\mathscr{E}_g(T) = 1.165 - 2.84 \times 10^{-4} T \text{ in eV} \tag{7.85}$$

where T is in K. For Ge the temperature dependence of the energy gap is

$$\mathscr{E}_g(T) = 0.742 - 3.90 \times 10^{-4} T \tag{7.86}$$

The form of the temperature dependence for the III–V compound semiconductors is somewhat different. For GaAs the experimental data are best fit by

$$\mathscr{E}_g(T) = 1.522 - \frac{5.8 \times 10^{-4} T^2}{T + 300} \tag{7.87}$$

while GaP is

$$\mathscr{E}_g(T) = 2.338 - \frac{6.2 \times 10^{-4} T^2}{T + 460} \tag{7.88}$$

Notice that in (7.85) to (7.88) the energy gap decreases with increasing temperature. Although this behavior holds for most semiconductors, for some materials such as the IV–VI compounds the energy gap increases with increasing temperature.

7.2.5 Electric Field Dependence

The electric field dependence of the fundamental absorption edge is referred to as the Franz–Keldysh effect [W. Franz, *Z. Naturforsch. 13a*, 484 (1958); L. V. Keldysh, *Sov. Phys.-JETP 7*, 788 (1958)]. As indicated in Fig. 7.7, this electroabsorption process can be thought of as photon-assisted tunneling through the energy gap. That is, the electron wavefunctions in the valence and conduction bands have an exponentially decaying amplitude in the energy gap. In the presence of an electric field a valence band electron must tunnel through a triangular barrier to reach the conduction band. When there is no photon absorption as in Fig. 7.7(a), the height of this barrier is $\mathscr{E}_g$ and its thickness, t, can be determined from

$$q\mathbf{E} = \nabla_r \mathscr{E} = \frac{\mathscr{E}_g}{t} \tag{7.89}$$

In one dimension the barrier thickness is

$$t = \frac{\mathscr{E}_g}{qE} \tag{7.90}$$

However, with photon absorption as in Fig. 7.7(b), the barrier height is reduced to $\mathscr{E}_g - \hbar\omega$ and the barrier thickness becomes

$$t(\hbar\omega) = \frac{\mathscr{E}_g - \hbar\omega}{qE} \tag{7.91}$$

Obviously, the tunneling probability is considerably enhanced with photon absorption and depends on the electric field as well as the photon energy.

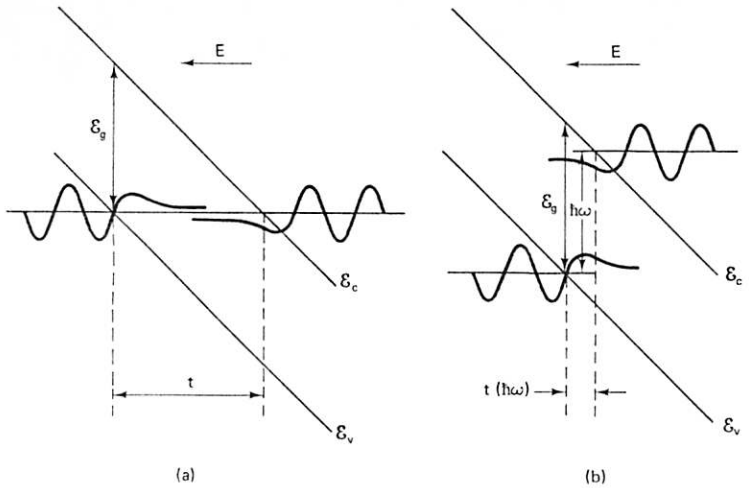


Figure 7.7 Energy band diagram in an electric field showing the wavefunction overlap (a) without and (b) with absorption of a photon of energy $\hbar\omega$.

An analysis [K. Tharmalingham, *Phys. Rev. 130*, 2204 (1963)] of this photon-assisted tunneling process indicates that the electroabsorption coefficient for a direct energy gap is given by

$$\alpha(\hbar\omega, E) = 1.0 \times 10^4 \frac{f}{\eta}\left(\frac{2m_r}{m}\right)^{4/3} E^{1/3} \int_\beta^\infty |\,\mathrm{Ai}(z)\,|^2\, dz \text{ in cm}^{-1} \quad (7.92)$$

where

$$\beta = 1.1 \times 10^5 \left(\frac{2m_r}{m}\right)^{1/3} \frac{\mathcal{E}_g - \hbar\omega}{E^{2/3}} \quad (7.93)$$

for E in volts/cm and $\mathcal{E}_g - \hbar\omega$ in eV. In (7.92) and (7.93) Ai(z) is the Airy function, f is the oscillator strength given by (7.54), and m_r is the reduced electron–hole mass given by (7.84). Figure 7.8 shows calculated values of the electroabsorption coefficient for GaAs, assuming a uniform electric field.

7.2.6 Magnetic Field Dependence

When a magnetic field is applied to a semiconductor, the components of electron motion perpendicular to the field describe circular orbits at the cyclotron frequency, ω_c. From (5.147) the cyclotron frequency for parabolic bands is

$$\omega_c = \frac{qB}{m^*} \quad (7.94)$$

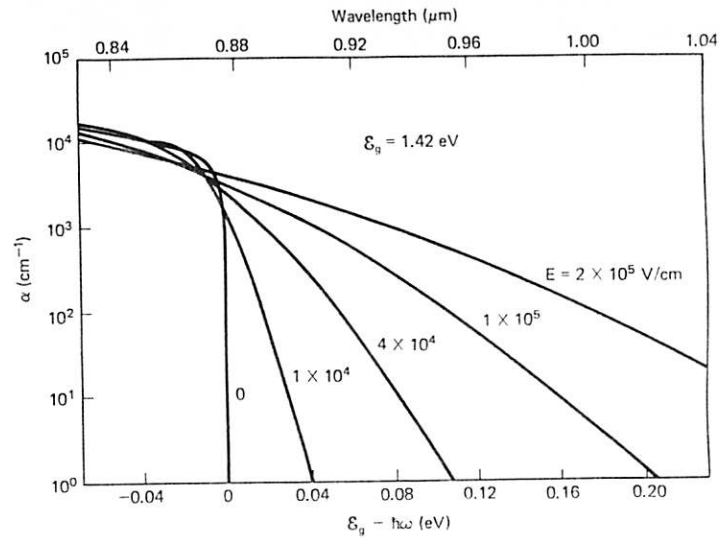


Figure 7.8 Electric field and photon energy dependence of the band-to-band absorption for GaAs. [After G. E. Stillman and C. M. Wolfe, "Avalanche Photodiodes" in *Semiconductors and Semimetals*, Vol. 12, *Infrared Detectors II*, ed. R. K. Willardson and A. C. Beer (New York: Academic Press, 1977), p. 291.]

An analysis of this effect in the effective mass approximation shows that the electron wavefunctions are described by Schrödinger's equation for a simple harmonic oscillator. As shown in Fig. 7.9, the allowed energy levels are quantized into Landau levels given by

$$\mathcal{E} = \mathcal{E}_c + \frac{\hbar^2 k^2}{2m_e} + \hbar\omega_{ce}\left(n + \frac{1}{2}\right) \quad (7.95)$$

for the conduction band, and by

$$\mathcal{E} = \mathcal{E}_v - \frac{\hbar^2 k^2}{2m_h} - \hbar\omega_{ch}\left(n + \frac{1}{2}\right) \quad (7.96)$$

for the valence bands. In (7.95) and (7.96) **k** is in a direction perpendicular to the field and n takes on all integer values, including zero. From (7.94), (7.95), and (7.96) the magnetic field dependence of the absorption edge or energy gap is simply

$$\mathcal{E}_g(B) = \mathcal{E}_g(0) + \frac{q\hbar B}{2m_r} \quad (7.97)$$

where m_r is the reduced electron–hole mass.

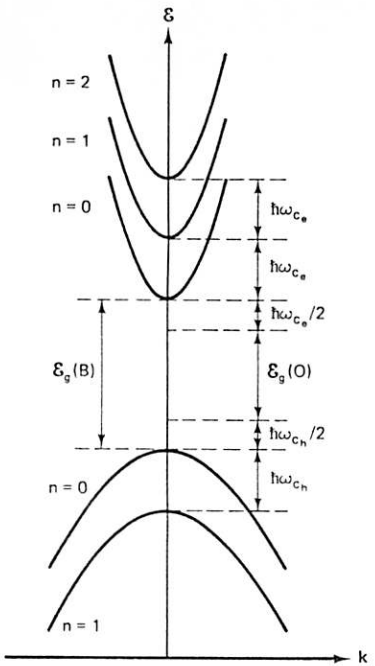


Figure 7.9 Splitting of the conduction and valence bands into Landau levels by a magnetic field.

7.3 EXCITON ABSORPTION

As indicated in Fig. 7.1, some structure in the absorption spectra of semiconductors is often observed just below the fundamental absorption edge. This structure is due to exciton absorption. This phenomenon can be understood in the following way. The energy band calculations in Chapter 2 were for one electron only, where the rest of the crystal was treated as a periodic potential. In this case there are no allowed energy states between the minima of the conduction bands and the maxima of the valence bands. In the creation of an electron–hole pair, as in band-to-band absorption, we looked at a perturbation of these one-electron energy bands. We ignored, however, the Coulombic interaction between the electron and the hole. For sufficiently low thermal energy this Coulombic attraction binds the electron and hole together to produce the quasi-particle known as an exciton. This is the first excited state of the one-electron energy bands.

In Chapter 3 we used the effective mass approximation to examine a similar problem: the interaction between an electron and a donor atom. The only difference between that problem and this one is that here the positively charged hole is free to move. Thus this problem is exactly analogous to the hydrogen atom with a reduced mass corresponding to the relative motion

of the electron and hole. From (3.33) the n binding energies of the exciton are given by

$$\mathscr{E}_{xn} = 13.6 \left(\frac{1}{n\epsilon_r}\right)^2 \left(\frac{m_r}{m}\right) \quad \text{in eV} \tag{7.98}$$

where the reduced mass is

$$\frac{1}{m_r} = \frac{1}{m_e} + \frac{1}{m_h} \tag{7.99}$$

and ϵ_r is the dielectric constant of the semiconductor. In a similar manner the orbital radii of the exciton are

$$r_{xn} = 0.53 n^2 \epsilon_r \left(\frac{m}{m_r}\right) \quad \text{in Å} \tag{7.100}$$

As indicated in Fig. 7.10, an exciton is free to move throughout the crystal with a dispersion relationship given by

$$\mathscr{E}_{xn}(\mathbf{k}) = \mathscr{E}_{xn} + \frac{\hbar^2 k^2}{2M} \tag{7.101}$$

Since this represents the collective motion of the electron and the hole, the appropriate effective mass for (7.101) is

$$M = m_e + m_h \tag{7.102}$$

To determine quantitatively the amount of absorption due to exciton for-

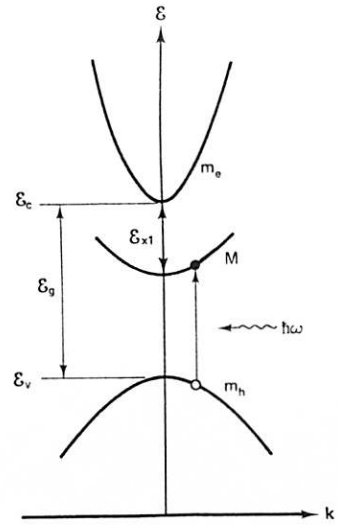


Figure 7.10 Creation of an exciton by photon absorption below the fundamental absorption edge.

mation, we can use the same procedure as for band-to-band absorption. The appropriate modification of (7.64) for exciton absorption is

$$\alpha_n = 2.7 \times 10^5 \left(\frac{2m_r}{m}\right)^{3/2} \frac{f}{\eta} (\hbar\omega - \mathscr{E}_g - \mathscr{E}_{xn})^{1/2} \text{ cm}^{-1} \tag{7.103}$$

In this case the reduced mass is

$$\frac{1}{m_r} = \frac{1}{M} + \frac{1}{m_h} \tag{7.104}$$

7.4 FREE CARRIER ABSORPTION

When the energy of the incident radiation is too small to create electron–hole pairs or excitons, other absorption processes can occur. The absorption of radiation by electrons in the conduction bands or by holes in the valence bands produces an absorption background below the fundamental absorption edge which increases with wavelength as indicated in Fig. 7.1. In this process the electric field of the incident radiation accelerates the free carriers, which, in turn, are decelerated by collisions with the lattice. Thus the energy of the radiation field is converted to heat.

This problem can be examined semiclassically by using Maxwell's equations to characterize the radiation in the semiconductor:

$$\begin{aligned} \nabla \times \mathbf{H} &= \sigma^* \mathbf{E} + \frac{\partial \mathbf{D}}{\partial t}, \qquad \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{E} &= -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \qquad \nabla \cdot \mathbf{D} = 0 \\ \mathbf{D} &= \epsilon \mathbf{E} \end{aligned} \tag{7.105}$$

In these equations, we have assumed there are no space-charge regions in the semiconductor, so the divergence of the electric field is zero. Also, a complex conduction current, $\sigma^*\mathbf{E}$, is included in the equation for the curl of the magnetic field. From (5.52) and (5.145) the high-frequency conductivity is

$$\sigma^* = \frac{q^2 n \langle \tau_m^* \rangle}{m^*} \tag{7.106}$$

where

$$\langle \tau_m^* \rangle = \left\langle \frac{\tau_m}{1 + \omega^2 \tau_m^2} \right\rangle + i \left\langle \frac{\omega \tau_m^2}{1 + \omega^2 \tau_m^2} \right\rangle \tag{7.107}$$

For $\omega\tau_m \gg 1$,

$$\langle \tau_m^* \rangle = \left\langle \frac{1}{\omega^2 \tau_m} \right\rangle + \frac{i}{\omega} \tag{7.108}$$

and

$$\begin{aligned} \sigma^* &= \frac{q^2 n}{m^* \omega^2} \left\langle \frac{1}{\tau_m} \right\rangle + i \frac{q^2 n}{m^* \omega} \\ &= \sigma_r + i\sigma_i \end{aligned} \tag{7.109}$$

The magnetic field can be eliminated from (7.105) to obtain an equation for a propagating radiation wave with losses,

$$\nabla^2 \mathbf{E} = \epsilon\mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma^* \mu_0 \frac{\partial \mathbf{E}}{\partial t} \tag{7.110}$$

Choosing a traveling-wave solution of the form

$$\mathbf{E} = \mathbf{a}E \exp [i(\mathbf{q}\cdot\mathbf{r} - \omega t)] \tag{7.111}$$

the wavevector $\mathbf{q}$ must satisfy

$$q = \frac{\omega}{c} \left(\epsilon_r + i \frac{\sigma^*}{\epsilon_0 \omega}\right)^{1/2} \tag{7.112}$$

where $c = (\epsilon_0 \mu_0)^{-1/2}$ is the free-space velocity of light. Since in free space,

$$q = \frac{\omega}{c} = \frac{2\pi}{\lambda} \tag{7.113}$$

(7.112) indicates that in the semiconductor the phase velocity and wavelength of the radiation are reduced by a complex refractive index

$$\eta^* \equiv \left(\epsilon_r + i \frac{\sigma^*}{\epsilon_0 \omega}\right)^{1/2} \tag{7.114}$$

Setting

$$\eta^* = \eta + ik \tag{7.115}$$

and using (7.109) in (7.114), we obtain

$$\eta^2 - k^2 = \epsilon_r - \frac{\sigma_i}{\epsilon_0 \omega} \tag{7.116}$$

$$2\eta k = \frac{\sigma_r}{\epsilon_0 \omega} \tag{7.117}$$

For wave propagation in the z direction, (7.111) is

$$\mathbf{E} = \hat{x}E \exp [i(qz - \omega t)] \tag{7.118}$$

where

$$q = \frac{\omega}{c}(\eta + ik) \tag{7.119}$$

Therefore, (7.118) has the form

$$\mathbf{E} = \hat{x}E \exp\left(\frac{-\omega k z}{c}\right) \exp\left[i\left(\frac{\omega \eta z}{c} - \omega t\right)\right] \tag{7.120}$$

In (7.120) it can be seen that the velocity and wavelength of the radiation in the semiconductor are decreased by η, while the electric field is attenuated by

$$\frac{\omega k}{c} = \frac{\sigma_r}{2\eta c \epsilon_0} \tag{7.121}$$

Since the intensity or power of the radiation is proportional to $|\mathbf{E}^2|$, the absorption coefficient for free carriers is

$$\alpha = \frac{\sigma_r}{\eta c \epsilon_0} \tag{7.122}$$

or using (7.109), we obtain

$$\alpha = \frac{q^2 n}{\eta c \epsilon_0 m^* \omega^2}\left\langle\frac{1}{\tau_m}\right\rangle \tag{7.123}$$

In terms of the free-space wavelength, (7.123) becomes

$$\alpha = \frac{q^2 n \lambda^2}{4\pi^2 \eta c^3 \epsilon_0 m^*}\left\langle\frac{1}{\tau_m}\right\rangle \tag{7.124}$$

Thus, free carrier absorption depends on the concentration of free carriers, n, the process by which they convert the light into heat through τ_m, and it increases with the square of the free-space wavelength, λ.

From (7.109), (7.116), and (7.117) we see that the free carriers also lower the refractive index,

$$\eta^2 = \epsilon_r - \frac{q^2 n}{\epsilon_0 m^* \omega^2} + k^2 \tag{7.125}$$

If we define a plasma frequency,

$$\omega_p^2 \equiv \frac{q^2 n}{\epsilon m^*} \tag{7.126}$$

(7.125) becomes

$$\eta^2 = \epsilon_r\left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right] + \frac{\epsilon_r^2}{4\eta^2}\left(\frac{\omega_p}{\omega}\right)^4\left\langle\frac{1}{\omega\tau_m}\right\rangle^2 \tag{7.127}$$

For $\omega\tau_m \gg 1$, the last term on the right-hand side is small and

$$\eta^2 = \epsilon_r\left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right] \tag{7.128}$$

In terms of the free-space wavelength,

$$\eta^2 = \epsilon_r - \frac{q^2 n \lambda^2}{4\pi^2 \epsilon_0 m^* c^2} \tag{7.129}$$

Although this effect is relatively small in semiconductors, heavily doped regions can be used to guide light into higher index, more lightly doped regions for some optoelectronic applications.

7.5 REFLECTIVITY

Equations (7.105) can also be used to determine the reflectivity of an ideal semiconductor surface. Let us consider a plane wave normally incident in the positive z direction on a semiconductor surface at $z = 0$. A semiconductor with complex refractive index given by (7.114) occupies the space with positive z values. At the surface, part of the incident radiation is reflected in the minus z direction and the rest penetrates the semiconductor, where its velocity and wavelength are reduced by η and its intensity attenuated by absorption.

For the region outside the semiconductor ($z < 0$), the electric field of the radiation is

$$E_x(z) = E_i \exp\left[i\omega\left(\frac{z}{c} - t\right)\right] + E_r \exp\left[-i\omega\left(\frac{z}{c} + t\right)\right] \tag{7.130}$$

where E_i is the magnitude of the incident electric field and E_r is the magnitude of the reflected electric field. From (7.105) the corresponding expression for the magnetic field for $z < 0$ is

$$\mu_0 c H_y(z) = E_i \exp\left[i\omega\left(\frac{z}{c} - t\right)\right] - E_r \exp\left[-i\omega\left(\frac{z}{c} + t\right)\right] \tag{7.131}$$

In the semiconductor ($z > 0$) the expression for the transmitted electric field is

$$E_x(z) = E_t \exp\left[i\omega\left(\frac{\eta^* z}{c} - t\right)\right] \tag{7.132}$$

and the magnetic field from (7.105) is

$$\mu_0 c H_y(z) = \eta^* E_t \exp\left[i\omega\left(\frac{\eta^* z}{c} - t\right)\right] \tag{7.133}$$

At the surface ($z = 0$) the electric and magnetic fields are continuous. Setting (7.130) equal to (7.132) and (7.131) equal to (7.133) at $z = 0$, we obtain

$$\begin{aligned} E_t &= E_i + E_r \\ \eta^* E_t &= E_i - E_r \end{aligned} \tag{7.134}$$

Eliminating E_t from these equations yields

$$\frac{E_r}{E_i} = \frac{1 - \eta^*}{1 + \eta^*} \tag{7.135}$$

Since the intensity of the light depends on the square of the electric field, the reflectivity is defined as

$$R \equiv \left(\frac{1 - \eta^*}{1 + \eta^*}\right)^2 \tag{7.136}$$

Using (7.115), we have

$$R = \frac{(\eta - 1)^2 + k^2}{(\eta + 1)^2 + k^2} \tag{7.137}$$

7.6 LATTICE ABSORPTION

Figure 7.1 shows an absorption band due to optical phonons which is almost as strong as the absorption due to electron transitions from the valence to conduction bands. This absorption is due to electric dipole coupling between photons and phonons which arises from the motion of charged atoms in the crystal. (The structure on the high-energy side is due to multiple phonon absorption.) Thus this absorption is strong for semiconductors with an ionic component of bonding and weak for those with purely covalent bonding. Since the photon is a transverse electromagnetic wave, it couples most strongly to the transverse optical phonons. For photon frequencies between the transverse and longitudinal optical phonon frequencies, the reflection is very high. If the radiation from a hot body is reflected several times from an ionic crystal, the dominant remaining frequencies will be those in this band of energy. For this reason this absorption is called the "restrahlen" or "residual ray" band.

Let us consider the interaction between a traveling light wave with electric field given by (7.111) and an ionic crystal with a basis of two atoms per primitive unit cell. Due to an electronegativity difference, the cations have an effective charge $+e^*$ and the anions $-e^*$. If we consider only interactions between ions in the unit cell, the equations of motion for the

displacement of the ions from their Bravais lattice sites are, from (3.72),

$$\begin{aligned} M_1 \frac{\partial^2 \mathbf{u}_1}{\partial t^2} &= -\frac{1}{2}\alpha(\mathbf{u}_1 - \mathbf{u}_2) + e^* \mathbf{E}_l \\ M_2 \frac{\partial^2 \mathbf{u}_2}{\partial t^2} &= -\frac{1}{2}\alpha(\mathbf{u}_2 - \mathbf{u}_1) - e^* \mathbf{E}_l \end{aligned} \tag{7.138}$$

In these equations α is the coupling coefficient between ions. The effect of the photon electric field $\mathbf{E}$ on the displacement is included by way of the local field $\mathbf{E}_l$ at the primitive or Wigner–Seitz unit cell.

Defining the relative displacement between ions as $\delta\mathbf{u} \equiv \mathbf{u}_1 - \mathbf{u}_2$, the coupled equations of motion (7.138) become

$$\frac{\partial^2 \delta\mathbf{u}}{\partial t^2} = -\frac{\alpha}{2M_r}\delta\mathbf{u} + \frac{e^*}{M_r}\mathbf{E}_l \tag{7.139}$$

where the reduced ion mass is

$$\frac{1}{M_r} = \frac{1}{M_1} + \frac{1}{M_2} \tag{7.140}$$

Considering a solution for (7.139) of the form

$$\delta\mathbf{u} = \mathbf{b}\,\delta u \exp\left[i(\mathbf{q}_s \cdot \mathbf{r} - \omega_s t)\right] \tag{7.141}$$

conservation of energy and momentum requires that the frequencies and wavevectors of the photon (7.111) and phonons (7.141) be equal. That is, $\omega_s = \omega$ and $q_s = q \simeq 0$. Thus the photon interaction is with the optical branch of the phonon dispersion curves shown in Fig. 3.23. Under these conditions (7.139) becomes

$$-\omega^2 \delta\mathbf{u} = -\frac{\alpha}{2M_r}\delta\mathbf{u} + \frac{e^*}{M_r}\mathbf{E}_l \tag{7.142}$$

or

$$\delta\mathbf{u} = \frac{e^*/M_r}{\omega_0^2 - \omega^2}\mathbf{E}_l \tag{7.143}$$

where

$$\omega_0^2 \equiv \frac{\alpha}{2M_r} \tag{7.144}$$

From (7.143) we see that the relative displacement between ions, $\delta\mathbf{u}$, resonates at the photon frequency $\omega = \omega_0$, and the lattice absorbs energy from the light wave. The interaction between a photon and the lattice in this frequency range is sometimes treated as a quasi-particle known as a "polariton."

The ability of the ions to respond to electromagnetic radiation constitutes an ionic contribution to the polarization of the crystal in addition to the contribution from the atomic cores. For N/V Wigner–Seitz cells per unit volume this ionic polarization is

$$\mathbf{P}_i = \frac{Ne^*}{V}\,\delta\mathbf{u} \tag{7.145}$$

Substituting (7.143) in this equation gives

$$\mathbf{P}_i(\omega) = \frac{Ne^{*2}/M_rV}{\omega_0^2 - \omega^2}\,\mathbf{E}_l \tag{7.146}$$

Notice that this frequency-dependent ionic polarization produces a frequency-dependent permittivity or dielectric constant given by

$$\mathbf{D}(\omega) = \epsilon(\omega)\mathbf{E} = \epsilon_0\mathbf{E} + \mathbf{P}(\omega) \tag{7.147}$$

where $\mathbf{P}(\omega)$ is the total polarization of the crystal. At low frequencies the permittivity

$$\epsilon(0) = \epsilon_r(0)\epsilon_0 \tag{7.148}$$

is due to both ionic and atomic polarization. At high frequencies the ions can no longer respond to the field and the permittivity,

$$\epsilon(\infty) = \epsilon_r(\infty)\epsilon_0 \tag{7.149}$$

is due to atomic polarization only. Thus the total polarization is

$$\mathbf{P}(\omega) = \mathbf{P}_i(\omega) + \mathbf{P}(\infty) \tag{7.150}$$

where from (7.147),

$$\mathbf{P}(\infty) = [\epsilon(\infty) - \epsilon_0]\mathbf{E} \tag{7.151}$$

The polarization of the lattice also screens the applied electric field from the unit cell, so that the local field and the applied field are different. In cubic crystals the relationship is

$$\mathbf{E}_l = \mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P} \tag{7.152}$$

For simplicity in the analysis to follow, we will assume that $\mathbf{E}_l = \mathbf{E}$. With this assumption we use (7.146) and (7.151) in (7.150) and insert (7.150) in (7.147), to obtain

$$\epsilon(\omega)\mathbf{E} = \frac{Ne^{*2}/M_rV}{\omega_0^2 - \omega^2}\,\mathbf{E} + \epsilon(\infty)\mathbf{E} \tag{7.153}$$

Defining a frequency,

$$\omega_1^2 \equiv \frac{Ne^{*2}}{M_rV\epsilon(\infty)} \tag{7.154}$$

(7.153) gives

$$\epsilon_r(\omega) = \epsilon_r(\infty)\left(1 + \frac{\omega_1^2}{\omega_0^2 - \omega^2}\right) \tag{7.155}$$

and

$$\epsilon_r(0) = \epsilon_r(\infty)\left(\frac{\omega_0^2 + \omega_1^2}{\omega_0^2}\right) \tag{7.156}$$

as the relationship between the low- and high-frequency dielectric constants. Although the frequencies ω_0 and ω_1 in (7.155) and (7.156), as defined by (7.144) and (7.154), appear to be somewhat arbitrary, they are intimately related to the transverse and longitudinal optical phonon frequencies.

To examine these relations, let us first consider a transverse lattice wave in the form of (7.141). Assuming a transverse displacement $\delta\mathbf{u}_{TO}$ with frequency ω_{TO} in the equation of motion (7.139), we obtain

$$\omega_{TO}^2\,\delta\mathbf{u}_{TO} = \omega_0^2\,\delta\mathbf{u}_{TO} - \frac{e^*}{M_r}\mathbf{E} \tag{7.157}$$

Taking the curl of this equation yields

$$\nabla \times \delta\mathbf{u}_{TO} = i\mathbf{q} \times \delta\mathbf{u}_{TO} \neq 0 \tag{7.158}$$

since the displacement is perpendicular to the direction of propagation. For the electric field associated with the displacement, however,

$$\nabla \times \mathbf{E} = 0 \tag{7.159}$$

Thus

$$\omega_{TO}^2 = \omega_0^2 \tag{7.160}$$

and the transverse phonon frequency is not affected by the electric field.

Next consider a longitudinal wave with displacement $\delta\mathbf{u}_{LO}$ and frequency ω_{LO}. From the equation of motion, (7.139),

$$\omega_{LO}^2\,\delta\mathbf{u}_{LO} = \omega_0^2\,\delta\mathbf{u}_{LO} - \frac{e^*}{M_r}\mathbf{E} \tag{7.161}$$

Taking the divergence of this equation gives

$$\nabla\cdot\delta\mathbf{u}_{LO} = i\mathbf{q}\cdot\delta\mathbf{u}_{LO} \neq 0 \tag{7.162}$$

since the displacement is in the direction of propagation. For the electric

flux density associated with the displacement,

$$\nabla \cdot \mathbf{D} = 0 \tag{7.163}$$

Using (7.143) in (7.153), we have

$$\mathbf{D} = \frac{Ne^*}{V}\,\delta\mathbf{u}_{LO} + \epsilon(\infty)\mathbf{E} \tag{7.164}$$

or

$$\nabla \cdot \mathbf{E} = \nabla \cdot \left[\frac{\mathbf{D}}{\epsilon(\infty)} - \frac{Ne^*}{V\epsilon(\infty)}\,\delta\mathbf{u}_{LO}\right]$$

$$= -\frac{Ne^*}{V\epsilon(\infty)}\,\nabla \cdot \delta\mathbf{u}_{LO} \tag{7.165}$$

and (7.161) is

$$\omega_{LO}^2 = \omega_0^2 + \omega_1^2 \tag{7.166}$$

Thus the longitudinal optical phonon frequency is higher than the transverse frequency because of the field interaction.

Using (7.166) and (7.160) in (7.156), we find that the dielectric constants are related to the optical phonon frequencies by

$$\frac{\epsilon_r(0)}{\epsilon_r(\infty)} = \left(\frac{\omega_{LO}}{\omega_{TO}}\right)^2 \tag{7.167}$$

This equation is known as the Lyddane–Sachs–Teller relation [R. H. Lyddane, R. G. Sachs, and E. Teller, *Phys. Rev. 59*, 613 (1941)]. From (7.155) the frequency dependence of the dielectric constant is

$$\epsilon_r(\omega) = \epsilon_r(\infty)\frac{\omega_{LO}^2 - \omega^2}{\omega_{TO}^2 - \omega^2} \tag{7.168}$$

Notice from this equation that the dielectric constant is infinite for $\omega = \omega_{TO}$, zero for $\omega = \omega_{LO}$, and negative for ω between ω_{TO} and ω_{LO}. This behavior is shown in Fig. 7.11(a).

This behavior of the dielectric constant strongly affects the reflectivity of the crystal. Neglecting the carrier concentration in (7.125) gives us

$$\epsilon_r = \eta^2 - k^2 \tag{7.169}$$

and the reflectivity is

$$R = \frac{(\eta - 1)^2 + k^2}{(\eta + 1)^2 + k^2} \tag{7.137}$$

In the region between ω_{TO} and ω_{LO} where ϵ_r is negative, k^2 must be much greater than η^2 according to (7.168). From (7.137) this produces a reflectivity

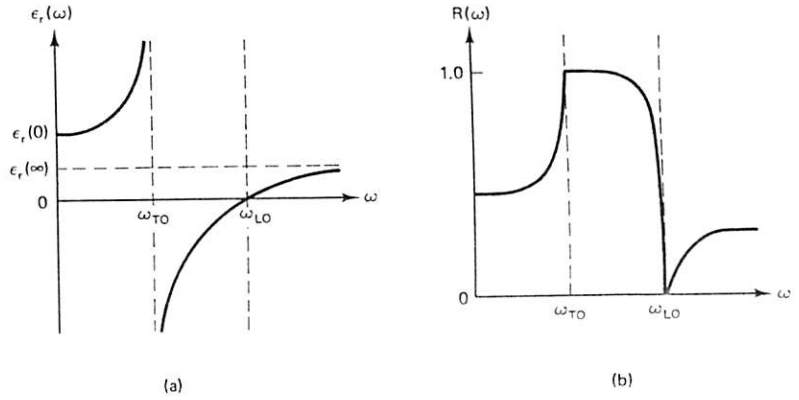


Figure 7.11 Frequency variation of the (a) dielectric constant and (b) reflectivity in the vicinity of the optical phonon frequencies.

near unity which accounts for the "restrahlen" band. The behavior of the reflectivity is illustrated schematically in Fig. 7.11(b).

Using the Lyddane–Sachs–Teller relation, an equation relating the effective charge of the ions to more easily measurable quantities can be obtained. Equation (7.154) gives

$$e^{*2} = \frac{M_r V\epsilon(\infty)}{N}\,\omega_1^2 \tag{7.170}$$

where from (7.160) and (7.166),

$$\omega_1^2 = \omega_{LO}^2 - \omega_{TO}^2 \tag{7.171}$$

Using (7.167) in (7.171), we have

$$e^{*2} = \frac{M_r V}{N}\,\omega_{LO}^2\epsilon^2(\infty)\left[\frac{1}{\epsilon(\infty)} - \frac{1}{\epsilon(0)}\right] \tag{7.172}$$

Since the mass density (g/cm^3)

$$\rho = \frac{M_r}{\Omega} = \frac{M_r N}{V} \tag{7.173}$$

(7.172) can be expressed as

$$e^* = \Omega\omega_{LO}\epsilon(\infty)\rho^{1/2}\left[\frac{1}{\epsilon(\infty)} - \frac{1}{\epsilon(0)}\right]^{1/2} \tag{7.174}$$

In terms of the transverse optical frequency, (7.174) has the somewhat simpler form

$$e^* = \Omega\omega_{TO}\rho^{1/2}[\epsilon(0) - \epsilon(\infty)]^{1/2} \tag{7.175}$$

Since as indicated in Chapter 6 the definition of this quantity is not unique, (7.175) is called the Born effective ionic charge.

7.7 IMPURITY ABSORPTION

The last absorption process we discuss is that due to direct band-to-impurity transitions. From the discussion in Section 7.2 the transition probability has the same form as for direct band-to-band transitions. From (7.63) and (7.53) the proportionality of the absorption coefficient for an electron transition from an initial to a final state is

$$\alpha_{if} \propto N_I f_i (1 - f_f) \frac{(\hbar\omega - \Delta\mathscr{E}_{if})^{1/2}}{\hbar\omega} \tag{7.176}$$

where N_I is the total concentration of impurities, f_i the probability that the initial state is full, $1 - f_f$ the probability that the final state is empty, and $\Delta\mathscr{E}_{if}$ the energy difference between the initial and final state.

As an example, consider the transition of an electron from the valence band to an empty donor in n-type material. In n-type material the valence band is full, so $f_i = 1$. The concentration of empty donors is

$$N_I(1 - f_f) = N_d^+ \tag{7.177}$$

where N_d^+ is given by (4.97). Substituting (4.97) into (7.170) and using the appropriate notation,

$$\alpha_{v,d} \propto \frac{N_d}{1 + g_d \exp[(\mathscr{E}_f - \mathscr{E}_d)/kT]} \frac{[\hbar\omega - (\mathscr{E}_d - \mathscr{E}_v)]^{1/2}}{\hbar\omega} \tag{7.178}$$

Similar expressions can be obtained for transitions involving full donors from (4.88) and full acceptors from (4.93).

PROBLEMS

7.1. Determine the energy at which appreciable band-to-band absorption begins at 300 K for a p-type GaAs sample with a hole concentration of 1×10^{20} cm^{-3}. Take into account both the light- and heavy-hole bands, which are degenerate at Γ.

7.2. Determine the free-carrier absorption at 300 K and 0.905 μm (free-space wavelength) for a GaAs sample with $n = N_d - N_a = 9.8 \times 10^{17}$ cm^{-3} and $N_a = 9.8 \times 10^{17}$ cm^{-3}. For this sample ionized impurity scattering is the dominant scattering mechanism.

7.3. In the density-of-states representation we have a hypothetical intrinsic semiconductor (direct) as shown in Fig. P7.3 and have no problems with k-selection rules relative to generation or recombination transitions.

(a) At low temperature we pump a thin sample of this material and measure its transmission in order to determine the absorption coefficient $\alpha = \alpha(\mathscr{E}_p) = \alpha(h\nu_p)$. The magnitude of the absorption depends on the joint density of upper and lower states separated by $h\nu_p$, appropriate filling factors (probabilities), and an assumed constant coupling between valence and conduction band states. How does α vary as a function of $\mathscr{E}_p$?

(b) Suppose that the pump intensity I_0 can be increased to very high levels. At larger and larger I_0 and fixed $\mathscr{E}_p = h\nu_p$, how does the absorption of the thin sample change (vary)? Justify your answer.

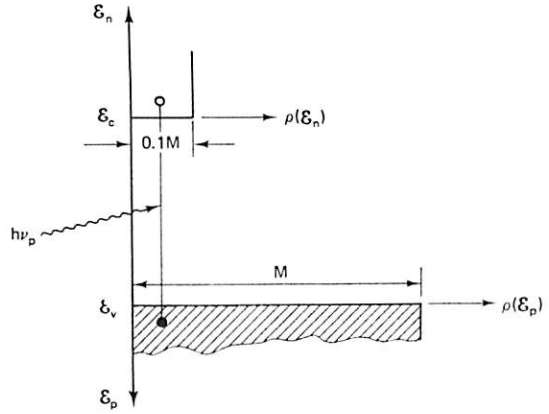


Figure P7.3

7.4. In a semiconductor laser Fabry–Perot resonator of length L the number M of half-wavelengths in the structure is $M(\lambda'_M/2) = L$, where $\lambda' \equiv \lambda/\eta$, λ is the free-space wavelength and η is the index of refraction of the semiconductor medium. (Index of refraction is the ratio of the speed of light in free space to that in the medium.) Find the spacing $d\lambda$ between adjacent wavelength modes in the resonator. Remember that the index of refraction, $\eta = \eta(\lambda)$ is a function of wavelength.

7.5. Given a semiconductor laser of length L and index of refraction η, where η is a function of λ, $\eta(\lambda)$.

(a) Derive the formula for the spectral separation of the laser modes, $d\lambda$.

(b) Suppose that one end of the original cavity is coated with an antireflective coating, and an external mirror is placed a distance l from this end. What is the formula for $d\lambda$ in this case?

(c) Define the quantity $\eta - \lambda(d\eta/d\lambda) = \Delta$ as the index dispersion expression. What is the expression for this quantity from part (a)? Part (b)? (In terms of L, l, λ, and $d\lambda$)

(d) If the typical mode spacing for no external mirror is 3 Å and a typical value of Δ is 6, what is a practical limit on l/L such that a monochrometer with a resolution of 0.3 Å can still "see" separate modes in the case of the arrangement in part (b)?

The quantity Δ, called the index dispersion, is useful in interpreting laser mode data and inferring behavior of density of states and absorption.

triangles. Reasonable agreement is thus obtained between the experimental form factor values and those computed assuming the unpaired electrons to be in a state of pure e_g symmetry.

It is important to observe that the scale factor, k, here determined experimentally, is such as to make the unpaired $3d$ charge distribution for Ni^{++} much more compact in the solid than it is for the free atom. This is in contrast to the case of Mn^{++} where experiments[13] show that the charge distribution is expanded in the solid.

These experimental results may be compared with the recent calculations of Watson and Freeman[5] for the Ni^{++} ion. These Hartree-Fock self-consistent field calculations allow the wave functions of electrons with opposite spins to have different radial dependencies (spin polarization) and lead to a contraction of the charge distribution (both for the free-atom case and the case where the Ni^{++} ion is placed in an octahedral array of point charges).[14] Unfortunately, the magnitude of the contraction is much too small to explain the observations reported here. Nevertheless, the important fact that the relation of the observed f_S to the free-atom f_S is just opposite for the cases of Ni^{++} and Mn^{++} would lead one to look for the origin of this effect in the outstanding difference between the two ions: namely their differing spin configurations. These experimental results also serve to suggest that whereas effects such as spin polarization and crystalline environment have a large influence on f_S, their effect on f_A is small.

The author wishes to thank Professor E. Uchida and Dr. Y. Nakazumi for the single crystal of NiO.

†Performed under the auspices of the U. S. Atomic Energy Commission.

[1]R. Nathans, C. G. Shull, G. Shirane, and A. Andresen, J. Phys. Chem. Solids 10, 138 (1959).

[2]R. Nathans and A. Paoletti, Phys. Rev. Letters 2, 254 (1959).

[3]S. J. Pickart and R. Nathans, Suppl. J. Appl. Phys. 31, 372S (1960).

[4]R. Nathans, S. J. Pickart, and H. A. Alperin, Bull. Am. Phys. Soc. 5, 455 (1960).

[5]R. E. Watson and A. J. Freeman, Phys. Rev. 120, 1125 (1960); R. E. Watson and A. J. Freeman, Phys. Rev. 120, 1134 (1960).

[6]R. Nathans, Suppl. J. Appl. Phys. 31, 350S (1960).

[7]H. A. Alperin, U. S. Naval Ordnance Laboratory Report NAVORD-6907, June, 1960 (unpublished).

[8]J. D. Dunitz and L. E. Orgel, J. Phys. Chem. Solids 3, 20 (1957).

[9]We are assuming here that the radial part of the wave functions is the same for all eight electrons so that the three electrons of t_{2g} symmetry with their spins in one direction "cancel" the three electrons of like symmetry with oppositely directed spins.

[10]R. J. Weiss and A. J. Freeman, J. Phys. Chem. Solids 10, 147 (1959).

[11]R. E. Watson and A. J. Freeman, Acta Cryst. (to be published).

[12]The form factor of Ni^{++} in $KNiF_3$ has been measured up to the value $\sin\theta/\lambda = 0.32$ by V. Scatturin, L. Corliss, N. Elliott, and J. Hastings, Acta Cryst. (to be published). Their result, which is for a powder measurement, is in approximate agreement with the curve presented here.

[13]J. M. Hastings, N. Elliott, and L. M. Corliss, Phys. Rev. 115, 13 (1959).

[14]It should be noted that Mn^{++} which has all its spins in the same direction has no spin polarization effect.

TUNNELLING FROM A MANY-PARTICLE POINT OF VIEW*

J. Bardeen
University of Illinois, Urbana, Illinois
(Received December 16, 1960)

Giaever[1] and more recently Nicol, Shapiro, and Smith[2] have observed the tunnelling current flowing between two metals separated by a thin oxide layer. The most interesting results are obtained when one or both of the metals are superconducting, in which case they find direct evidence for a gap in the quasi-particle spectrum of the superconductor. They were able to account for the data quantitatively on the assumption that the only relevant factor is the density of states in energy. This is to be expected if the transition probability for transfer of an electron from one side to the other is given by the familiar expression $(2\pi/\hbar)|M|^2\rho_f$, where M is the matrix element and ρ_f the energy density of final states, and if it is further assumed that M can be treated as a constant. It is implied that M is not only independent of energy for the small energy differences involved, but is also unchanged when the metal goes from normal to superconducting.

However, it is not immediately obvious that these assumptions are justified. We give here a discussion of tunnelling from a many-particle point of view and show that it is plausible to treat M as a constant in the interpretation of the experiments.

We suppose that the barrier extends from x_a to x_b, with metal a to the left of x_a and metal b to the right of x_b. Consider two many-particle states of the entire system, Ψ_0 and Ψ_{mn}, which differ in the transfer of an electron from a to b. We suppose that Ψ_0 and Ψ_{mn} may be defined in terms of quasi-particle occupation numbers of metals a and b, so that Ψ_{mn} differs from Ψ_0 in the transfer of an electron from state m in a to state n in b, all other occupation numbers remaining the same. Of course m may correspond to a normally occupied state, in which case there will be a hole in m in Ψ_{mn}.

The quasi-particles do not correspond to plane waves, but to waves which are reflected at the barrier and which drop exponentially with distance into the barrier region. For example, in the free-electron approximation for the normal state the wave function would be of the form (WKB approximation):

$$\psi_m = Cp_x^{-1/2} e^{i(p_y y + p_z z)} \sin(p_x x + \gamma), \quad x < x_a \tag{1a}$$

$$\psi_m = \tfrac{1}{2} C |p_x|^{-1/2} e^{i(p_y y + p_z z)} \times \exp\left(-\int_{x_a}^{x} |p_x| dx\right), \quad x_a < x < x_b \tag{1b}$$

where $C = (2p_x/L)^{1/2}$ is a normalization constant, and in the barrier region, $|p_x| = (2\mu U - p_y^2 - p_z^2)^{1/2}$, where $U(x)$ is the potential energy. We have taken units such that $\hbar = 1$. Beyond x_b, we assume that the wave function representing the electron in state m drops smoothly to zero, instead of oscillating, so that it is not a good solution for $x > x_b$.

Thus we assume that Ψ_0 is a solution of the Schrödinger equation with energy W_0 for $x < x_b$, but there is a region to the right of x_b where it is not a good solution. Similarly, we assume that Ψ_{mn} with energy W_{mn} is a solution for $x > x_a$, but not for the region to the left of x_a where the wave function for quasi-particle n drops to zero. Both Ψ_0 and Ψ_{mn} are good solutions in the barrier region $x_a < x < x_b$.

We form a time-dependent solution as a linear combination of Ψ_0 and various final states, Ψ_{mn}, by the usual method:

$$\Psi = a(t)\Psi_0 e^{-iW_0 t} + \sum_{mn} b_{mn}(t)\Psi_{mn} e^{-iW_{mn} t}, \tag{2}$$

and substitute into the Schrödinger equation. This gives for the matrix element for the transition

$$M_{mn} = \int \Psi_0^* (H - W_{mn}) \Psi_{mn} d\tau. \tag{3}$$

Since the integrand vanishes except over a region to the left of x_a, we need to integrate only over this region.

We may express the integral in a more symmetric form by subtracting $\Psi_{mn}(H - W_0)\Psi_0^*$, which vanishes to the left of x_b. Since we are interested only in final states such that $W_{mn} \simeq W_0$, the result may be written

$$M_{mn} = \int_a [\Psi_0^* H \Psi_{mn} - \Psi_{mn} H \Psi_0^*] d\tau, \tag{4}$$

where the subscript a indicates that the integration is to be taken over the region to the left of x_a.

The matrix element may be expressed in terms of that of the current density operator, J, in the barrier region as follows. We introduce into the integrand a step function $S(x)$ which is equal to unity between a point x_0 to the left of the important region of integration and a point x_1 in the barrier and which vanishes elsewhere. Then integrations with respect to y_i and z_i vanish and integration with respect to x_i gives

$$M_{mn} = -\frac{1}{2\mu} \sum_i \int \cdots \int (\Psi_0^* \nabla_i^2 \Psi_{mn} - \Psi_{mn} \nabla_i^2 \Psi_0^*) S(x_i) d\tau_1 \cdots d\tau_N$$

$$= -i[J_{mn}(x_1) - J_{mn}(x_0)], \tag{5}$$

where $J_{mn}(x)$ is the matrix element of the x component of the current density operator defined by

$$J_{mn}(x) = -\frac{i}{2\mu} \sum_i \int \cdots \int [\Psi_0^* \partial \Psi_{mn}/\partial x_i - \Psi_{mn} \partial \Psi_0^*/\partial x_i] \delta(x - x_i) d\tau_1 \cdots d\tau_N. \tag{6}$$

According to our assumptions, $J_{mn}(x_0) = 0$ and further, $J_{mn}(x_1)$ is independent of position as long as x_1 is in the barrier region. Thus

$$M_{mn} = -iJ_{mn}(x_1). \tag{7}$$

It is easily verified that this method leads to the usual results for barrier penetration problems.

The quasi-particle energy in a superconductor is $E = (\epsilon^2 + \Delta^2)^{1/2}$, where $\epsilon(k)$ is the normal state energy measured from the Fermi surface. In calculating the density of final states, it is simplest to take the conventions that $f = 1$ and E is negative for the normally occupied states in the Fermi sea, so that for $k < k_F$, the probability of a hole excitation is $1-f$ and the energy of a hole is $-E$. This is the procedure which has been used in the interpretation of the experimental data.[1,2] In each region of k space the density of states in energy in the superconductor differs from that in the normal metal by the factor

$$\rho_S = \rho_n[|E|/(E^2 - \Delta^2)^{1/2}], \tag{8}$$

with $\rho_S = 0$ in the gap, $|E| < \Delta$. Agreement with experiment is obtained if it is assumed that M_{mn} (or J_{mn}) is the same for the corresponding transitions in normal and superconducting states.

Usually coherence factors, which have a marked effect on transition probabilities, are introduced in calculations of matrix elements between quasi-particle states of a superconductor.[3] To see why such factors are not expected to occur in tunnelling, one must consider the pairing in the vicinity of the barrier. Coherence factors would be introduced if one simply paired complex conjugate wave functions of the type (1a), (1b), so that pairing extends into the barrier region. However, if one looks at the problem more closely, from the viewpoint of the more general Gor'kov equations[4] which allow for a variation of the energy gap parameter with position, one sees that Δ will drop to zero very rapidly in the barrier. In effect electrons in this region are not paired and the wave function is essentially the same as in the normal state.

For an effective interaction $v(\vec{r}_1, \vec{r}_2)$, the position-dependent energy gap function is defined by

$$\Delta(\vec{r}_1, \vec{r}_2) = \langle N-2|\psi(\vec{r}_1)\psi(\vec{r}_2)|N\rangle_{\mathrm{av}}\, v(\vec{r}_1, \vec{r}_2). \tag{9}$$

Simply from the fact that ψ is very small in the barrier region, one expects that $\Delta(\vec{r}_1, \vec{r}_2)$ is small when either $\vec{r}_1$ or $\vec{r}_2$ is in the barrier. Note that the tail in the barrier of the wave function of a typical electron at the Fermi surface is very much smaller than that of one of the few electrons moving normal to the interface which has an appreciable probability of penetrating. However, $\Delta(\vec{r}_1, \vec{r}_2)$ is expected to rise very rapidly to normal values in the superconductor, $\Delta(\vec{r}_1, \vec{r}_2) \simeq \delta(\vec{r}_1 - \vec{r}_2)\Delta$, with $\Delta = \text{const}$.

Since, according to (6), the matrix element depends only on the wave function in the barrier, it would be expected to be the same as for the corresponding transition in the normal state. Actually, the wave function in the barrier would be changed slightly because of the difference in quasi-particle energies in normal and superconducting states, but this would have a negligible effect on the matrix elements. Thus the only significant difference in the tunnelling current comes from the density of states factor.

The method described here can also be used for calculating the tunnelling current between the valence and conduction bands of a semiconductor, as observed in the Esaki diode. It can be generalized to take indirect transitions into account.

The author is indebted to Dr. W. A. Harrison and Dr. J. R. Schrieffer for discussions of the theory of tunnelling in a superconductor.

---

*This work was supported in part by the Office of Ordnance Research, U. S. Army.

[1]Ivar Giaever, Phys. Rev. Letters 5, 147, 464 (1960).

[2]J. Nicol, S. Shapiro, and P. H. Smith, Phys. Rev. Letters 5, 461 (1960).

[3]J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. 408, 1175 (1957).

[4]L. P. Gor'kov, J. Exptl. Theoret. Phys. U.S.S.R. 34, 735 (1958) [translation: Soviet Phys.-JETP 34(7), 505 (1958)].

Tunneling from an Independent-Particle Point of View

WALTER A. HARRISON
General Electric Research Laboratory, Schenectady, New York
(Received February 20, 1961)

A method is developed for calculating wave functions through regions of varying band structure. This method is applied to tunneling problems using the transition-probability approach of Bardeen. It is found that the experiments of Giaever involving tunneling into superconductors cannot be understood strictly in terms of an independent quasi-particle model of the superconductor. The observed proportionality of the tunneling probability to the density of states depends upon the matrix elements being constant which, in turn, depends upon a many-particle feature of the problem. This feature does not carry over to fluctuations in the density of states arising from band structure, and contributions to the current are not expected to be proportional to the density of states in that case. Instead, a projection in wave-number space of the appropriate constant-energy surface enters. Tunneling systems are discussed which involve semiconductors, semimetals, and transition metals as well as simple metals. Finally, alterations in the properties arising from alterations in the nature of the boundary regions are discussed.

I. INTRODUCTION

BARDEEN[1] has discussed tunneling from a many-particle point of view. He did not, however, discuss systems in which the band structure varies with position. In the following treatment we restrict ourselves to an independent-particle approximation, but extend the work to allow for variations in the local band structure.

We proceed by first developing a method for construction of one-particle wave functions, taking particular care that these conserve current locally. Matrix elements are then calculated in terms of these functions and the tunneling current determined.

Within this framework we consider the Giaever experiment[2] (tunneling into superconductors) using an independent-particle model of the superconductor. The failure of this approach sheds some further light on the treatment by Bardeen.

The same formulas are applied to systems involving tunneling into semiconductors, semimetals, and transition metals as well as into simple metals. Finally, we discuss the dependence of the tunneling properties on the nature of the boundary between different regions of the system.

II. DETERMINATION OF WAVE FUNCTIONS

We consider first a simple system which is divided by the plane $x=0$ into two regions of known band structure, and an electron energy which lies within the allowed bands in both regions. If suitable boundary conditions were applied to either one of these regions alone, we could construct a complete set of Bloch functions corresponding to the appropriate band structure. We will assume that wave functions for the composite system may be given, to a good approximation, by a linear combination of an incoming and an outgoing Bloch function on each side of the boundary, with each of these waves having the same component of wave number parallel to the boundary. This latter condition corresponds to an assumption of specular reflection and transmission by the boundary and will be relaxed in Sec. VI. This cannot be an exact eigenfunction for general band structure since we are required to match the wave functions on the entire plane $x=0$, but have only the coefficients of four waves at our disposal.

The problem of constructing wave functions is now reduced to the problem of obtaining two matching conditions upon the wave function; these, in conjunction with normalization and external boundary conditions, will uniquely determine the eigenstates. It will be a great mathematical simplification to assume reflection symmetry in the band structure at all points; this would correspond to time-reversal symmetry in a one-dimensional problem.

It will be convenient to represent the *Bloch* waves by *plane* waves of the same wave number, and to represent linear combinations of Bloch waves by linear combinations, ϕ, of the corresponding plane waves. Since it is the *Bloch* functions which are to be matched smoothly at the boundary, the function ϕ may not be smooth at the boundary. We write generalized boundary conditions:

$$\beta\phi \text{ continuous, } \alpha\partial\phi/\partial x \text{ continuous,} \tag{1}$$

across the boundary, where α and β may depend upon energy and transverse component of wave number and are different for different band structures. These parameters may be related to the current-density operator by noting that we may associate

$$J_x=\frac{\hbar}{2mi}\left[\phi^*\alpha\frac{\partial}{\partial x}\beta\phi-\phi\alpha\frac{\partial}{\partial x}\beta\phi^*\right], \tag{2}$$

with the x component of current density. This is permissible since with $\alpha=\beta=1$, this J_x is the current density for free electrons and J_x is conserved across all boundaries between differing band structures. The particular ordering of terms will be appropriate when the equations are generalized to continuously varying band structures and will guarantee the conservation of current in the volume of the material.

We may identify the velocity associated with a state of wave number k_x as $\hbar\alpha\beta k_x/m$ by noting the form of the current density, (2). This, in turn, must equal $\partial H/\partial p_x$

[1] J. Bardeen, Phys. Rev. Letters **6**, 57 (1961).
[2] I. Giaever, Phys. Rev. Letters **5**, 147, 464 (1960).

$=(1/\hbar)\partial E/\partial k_x$ in a one-particle approximation. Thus the product $\alpha\beta k_x$ may be evaluated in terms of the band structure as

$$\alpha\beta k_x=(m/\hbar^2)\partial E/\partial k_x. \tag{3}$$

Equation (3) will provide sufficient information about the matching parameters for most of our treatment, but further knowledge will be needed when we treat specialized boundary regions in Sec. VI.

In terms of these parameters the problem of obtaining wave functions for a system in which the band structure changes discontinuously at planes of constant x is completely determined. Within regions of constant band structure, the x variation of ϕ is obtained from the equation

$$\partial^2\phi/\partial x^2+k_x^2\phi=0$$

(noting the assumed reflection symmetry in the band structure). At planes of discontinuity we apply the conditions, (1).

It may be noted that the matching conditions introduce discontinuities in the slope and value of ϕ, and that these are required in order to conserve electron flux at the boundaries. The method strongly resembles the earlier treatment by the author[3] of wave functions in perturbed monovalent metals utilizing a cellular method. In that case variations of band structure in three dimensions were allowed.

The method is readily generalized to continuous variations of the band structure by taking the planes of discontinuity closer and closer together and, in the limit, obtaining a differential equation for the x variation of ϕ;

$$\frac{\partial}{\partial x}\frac{\alpha}{\beta}\frac{\partial}{\partial x}\phi+\beta\alpha k_x^2\phi=0. \tag{4}$$

Here ϕ, α, β, and k_x are functions of position which are assumed to be slowly varying over a single atomic cell. The latter three need not be slowly varying over a wavelength of ϕ; that is, we have not yet made a WKB approximation.

In order to treat problems involving tunneling, we must extend the above treatment to electron energies which lie in the forbidden band. We do this simply by analytically continuing the band structure ($E[\mathbf{k}]$) and the parameters α and β into the space of complex $\mathbf{k}$. This is the procedure used by James[4] in treating one-dimensional problems and by Peterson[5] in developing a crystalline WKB approximation.

III. CALCULATION OF THE TUNNELING CURRENT

We now have a well-defined procedure for constructing wave functions, and we may proceed to calculate eigenstates and the tunneling current. We will use the very convenient procedure developed by Bardeen.[1]

Bardeen has written the probability per unit time of the transition of an electron in a state a on one side of the tunneling region to a state b on the other side:

$$P_{ab}=(2\pi/\hbar)|M_{ab}|^2\rho_b f_a(1-f_b),$$

where M_{ab} is the matrix element for the transition, ρ_b is the density of states at b, and f_a and f_b are the probabilities of occupation of the states a and b, respectively. Further, he shows that we may write the matrix element in terms of the matrix element of the current operator between states a and b which are constructed to continue to drop exponentially beyond the tunneling region.

$$M_{ab}=-i\hbar J_{ab}, \tag{5}$$

where the x component of the current density operator J_{ab} is to be evaluated in the tunneling region between the states a and b which decay in opposite directions.

M_{ab} vanishes unless the transverse wave number k_t is the same for the initial and final states (specular transmission); thus ρ_b is a density of states for fixed k_t. We sum over all states a of fixed k_t, sum over k_t, multiply by 2 for spin and multiply by the electronic charge e to obtain the total current to the right. Subtracting the current to the left, we have finally

$$j=\frac{4\pi e}{\hbar}\sum_{k_t}\int_{-\infty}^{\infty}|M_{ab}|^2\rho_a\rho_b(f_a-f_b)dE. \tag{6}$$

The integral over energy is taken at fixed transverse wave number k_t.

In evaluating the matrix elements, we construct states which are sinusoidal in a positive-energy region and drop exponentially in an adjacent negative-energy region. We assume that, except near the transition region, the band structure is uniform. We assume for the time that the band structure is slowly varying in the transition region and make a WKB approximation. In Sec. VI we consider the consequences of making an abrupt approximation instead.

The WKB approximation is readily applied to Eq. (4) by assuming α, β, and k_x vary slowly in the x direction. The wave functions are found to be of the form

$$(\alpha\beta k_x)^{-\frac{1}{2}}\exp\left(i\int^x k_x dx\right),$$

except near turning points. The occurrence of the factor $(\alpha\beta k_x)^{-\frac{1}{2}}$ could have been forseen from the form of the current operator and the requirement of current conservation. Connection formulas across regions where $\alpha\beta k_x$ is small are the same as the usual ones. If x_a is the point on the left of the tunneling region at which $\alpha\beta k_x=0$, the x variation of a left-hand state is given by

[3] W. A. Harrison, Phys. Rev. **110**, 14 (1958).
[4] H. M. James, Phys. Rev. **76**, 1602 (1949).
[5] G. A. Peterson, Bull. Am. Phys. Soc. **5**, 161 (1960).

$$C_a(\alpha\beta k_x)^{-\frac{1}{2}}\cos\left(\int_x^{x_a} k_x dx+\gamma_a\right), \quad x<x_a,$$
$$(\tfrac{1}{2}C_a)(\alpha\beta k_x)^{-\frac{1}{2}}\exp\left(-\int_{x_a}^{x}|k_x|dx\right), \quad x>x_a, \tag{7}$$

where k_x is chosen to make the local energy independent of position. C_a is to be determined by normalization; if L_a is the extension of the crystal to the left of x_a, then

$$C_a{}^2=2(\alpha\beta k_x)_a/L_a.$$

Similarly, right-hand states may be constructed by replacing the subscript a by b.

We may now evaluate the matrix element for transition using its relation to the current-density matrix element [Eq. (5)], the form of the current density operator [Eq. (2)], and the wave function [Eq. (7)].

$$|M_{ab}|^2=\left(\frac{\hbar^2}{2m}\right)^2\frac{(\alpha\beta k_x)_a}{L_a}\times\frac{(\alpha\beta k_x)_b}{L_b}\exp\left(-2\int_{x_a}^{x_b}|k_x|dx\right). \tag{8}$$

This is to be substituted into Eq. (6) for the current density;

$$j=\frac{2e}{h}\sum_{k_t}\int_{-\infty}^{\infty}\exp\left(-2\int_{x_a}^{x_b}|k_x|dx\right)(f_a-f_b)dE. \tag{9}$$

In the determination of (9) the density of states was written as

$$\rho=(L/\pi)(\partial E/\partial k_x)^{-1}. \tag{10}$$

We notice the conspicuous absence of the density of states factor in Eq. (9). This is a direct consequence of our independent-particle model and the resultant reciprocal relation between the particle velocity [Eq. (3)] and the density of states [Eq. (10)]. We will see, in Sec. VI, that the exact form of the integrand in the expression for the current depends upon the nature of the barrier region, but only a quite artificial model would restore the simple proportionality to the density of states.

IV. TUNNELING INTO SUPERCONDUCTORS

The absence of the densities of states in the integrand in Eq. (9) is contrary to the experimental results of Giaever and to the result given by Bardeen. The inconsistency results from a failure of the independent-particle model. In order to obtain agreement with experiment, we may (following Bardeen) assert that the density of states is that for quasi-particles, but the current entering the matrix element is given by the value for normal metals. From the independent-particle point of view, this violates charge conservation at the boundary; from the many-particle point of view this is the assertion that back-flow around the quasi-particles should be neglected. In either framework, the essential problem is gauge invariance, and its resolution is beyond the realm of the independent-particle model. It seems remarkable that the simple experimental result depends so directly upon the subtleties of the many-particle system.

In any case, we do not expect this essential breakdown of the independent-particle model to occur in tunneling involving normal metals, semimetals, and semiconductors. Neither, then, do we expect to find the simple proportionality of the ac conductance to the density of states.

V. TUNNELING INTO NONSUPERCONDUCTING MATERIALS

We may obtain a more informative description of Eq. (9) by replacing the sum over k_t by an integral over the projection S of a constant energy surface onto the plane of the barrier; i.e., $d^2k_t=dS$. After rearranging factors, we obtain for the current per unit area,

$$j=\frac{e}{2\pi^2\hbar}\int_{-\infty}^{\infty}dE(f_a-f_b)\int dS\, e^{-\eta}, \tag{11}$$

with

$$\eta=2\int_{x_a}^{x_b}|k_x|dx.$$

η depends on energy and on the transverse wave number associated with dS. The second integral is over regions of k_t corresponding to positive-energy states on both sides. Thus if we define the "shadow" of a constant-energy surface to be its projection in wave-number space on a plane parallel to the barrier, the integral over S is over the overlap of the shadows from the two sides for the energy E.

The current between corresponding energy shells on the two sides of the barrier is proportional to the difference in the probability of occupation of the states in the two shells and to a surface integral over the overlap of their shadows. If one or both of the energy surfaces is small, as in a semiconductor or a semimetal, η becomes essentially constant and the integral over S is simply the exponential times the area of the shadow overlap. If, on the other hand, both of the surfaces are large, as when both sides are simple metals, the exponential becomes small for high angles of incidence. Only electrons moving approximately normal to the surface tunnel and the integral is equal to the value of the exponential at normal incidence times that area of the constant-energy surface for which appreciable numbers of electrons tunnel. This area depends only upon the properties of the tunneling region and may be readily evaluated for simple models by expanding $|k_x|$ for small k_t and integrating over S.

A. Simple Metals

The integral over S in Eq. (9) is a constant and the integral over energy is equal simply to the difference in

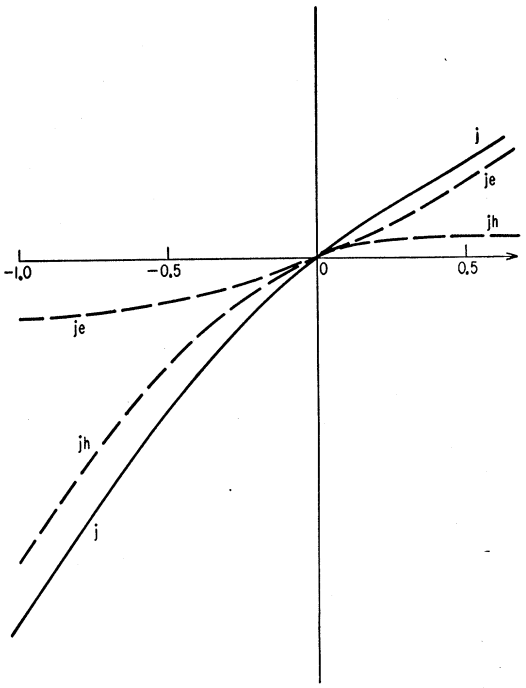


FIG. 1. The current-voltage characteristic for tunneling from a metal into an idealized semimetal. The hole mass was taken equal to three times the electron mass. Voltages are measured in units of the electron Fermi energy; the scale of the ordinate is arbitrary. The currents associated with the electron and the hole band are shown separately as well as the total current.

Fermi energy (or to the applied voltage). The behavior is Ohmic as observed by Giaever for normal metals.

B. Semimetals

We consider the tunneling between a simple metal and a semimetal. There is Fermi surface in both the electron band and the hole band and two contributions must be added. We calculate a sample characteristic neglecting the dispersal of the Fermi surface over the Brillouin zone and take η the same for both holes and electrons; we assume spherical surfaces and an effective-mass approximation for both signs of carriers with the result shown in Fig. 1.

We note that the irregularities are somewhat "washed out" and would be further reduced by the inclusion of other bands. The success of the nearly free-electron approximation for treating bismuth[6] suggests that there will be many other band edges near the Fermi energy and that the collection of all bands resembles a single free-electron band. This would suggest a generally linear characteristic with minor fluctuations.

C. Transition Metals

The very large fluctuations in the density of states in this case would have provided spectacular results if the tunneling were proportional to the density of states. However, the Fermi surfaces associated with both *d* and *s* bands are expected usually to be large. The current into each is determined by an area of surface which depends only on the properties of the insulating barrier, so the contributions from the two bands should be comparable. Only near band edges should there be fluctuations in the characteristic and these should be more pronounced near a *d*-band edge since the energy surface and its shadow area change more rapidly with energy in that case. Such irregularities have not been observed.

[6] W. A. Harrison, J. Phys. Chem. Solids **17**, 171 (1960).

D. Esaki Diode

The application to the tunnel diode is straightforward. The band structure is continued into the plane of complex $\mathbf{k}$ in the evaluation of η. In this case, with a reasonable model of the tunneling region, η will depend upon the applied voltage.

An interesting feature of the analysis is the failure of the one-dimensional density of states to enter the integral. The same feature occurs in the treatment of tunneling as a Zener current in a uniform field.[7] In that analysis, an artificial density of states enters which depends only on the applied field and the magnitude of a reciprocal lattice vector.

VI. NATURE OF THE TRANSITION REGION

It has been necessary in this treatment to make two specific assumptions about the boundary between different regions of the system. We have assumed that this region is sufficiently planar that the transmission of the barrier is specular. We have also assumed that the band structure varies sufficiently slowly that we may make a WKB approximation.

The experiments on superconductors shed no light on either of these questions. The one-dimensional density of states which enters Eq. (6) is proportional to the three-dimensional density of states which would enter if the transmission were diffuse. Furthermore, the assertion that the matrix element is the same for the superconducting and normal metals is sufficient to assure its relative constancy whether the boundary is gradual or sharp.

In the tunnel diode it is reasonable to expect that both assumptions are valid and the question remains open with respect to the boundary region with deposited or oxidized films. The observed diffuse reflection at metallic surfaces[8] would suggest diffuse transmission also. This is not conclusive, however, since macroscopically diffuse reflection is not inconsistent with microscopic specular reflection if the transition region is irregular.

We consider briefly some modifications expected if

[7] E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1959).

[8] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, New York, 1954), p. 248.

either of these assumptions as to the nature of the boundary is not valid.

A. Sharpness of the Boundary

If the boundary region were sharp, we should proceed with the calculation of the wave functions by matching rather than with the WKB approximation. The boundary conditions of Eq. (1) are directly applicable. We assume sharp boundaries and uniform band structures in each of the three regions, numbered 1, 2 and 3 (the tunneling region being numbered 2). We obtain

$$|M_{ab}|^2=\left(\frac{\hbar^2}{2m}\right)^2\frac{1}{L_1}\frac{1}{L_3}\frac{16\beta_1^2\nu_2^2\beta_3^2}{(\nu_1^2+\nu_2^2)(\nu_2^2+\nu_3^2)}e^{-\eta},$$

where

$$\nu_i=\beta_i/\alpha_i k_{xi}. \tag{12}$$

When multiplied by the density-of-states factors, this becomes

$$|M_{ab}|^2\rho_a\rho_b=\frac{4}{\pi^2}\frac{\nu_1\nu_2^2\nu_3}{(\nu_1^2+\nu_2^2)(\nu_2^2+\nu_3^2)}e^{-\eta}. \tag{13}$$

In contrast to the result of the WKB approximation, this is not independent of energy. Furthermore, the dependence upon the matching parameters is not simply through the product $\alpha\beta k$, which we were able to evaluate in terms of the band structure. We must consider these parameters in more detail.

The complete Bloch function may be written as $u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$. If $u_{\mathbf{k}}$ is normalized such that its average value is unity, the factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ may be associated with our function ϕ. When the band structure changes, we must match the *Bloch functions* at the cell boundaries, whereas our matching condition is that $\beta\phi$ be matched at the boundary. We may therefore associate β with the value of the normalized $u_{\mathbf{k}}$ at the cell boundary.

$$\beta^2=\frac{|u_k(\text{cell})|^2\Omega}{\int_{\text{cell}}|u_{\mathbf{k}}(\mathbf{r})|^2 d\tau}.$$

Here Ω is the cell volume, β is of order unity at the edge of an s band and zero at the edge of a p band, and β^2 is equal to the parameter γ introduced by Bardeen[9] in his extension of the Wigner-Seitz cellular method. Our matching conditions are equivalent to those used in the earlier extension of the cellular method to nonperiodic structures.[3]

From Eqs. (3), (10), and (12) we see that ν is proportional to β^2 times a density of states. Thus, Eq. (13) indicates that the density of states has reappeared in the integrand for the tunneling current. However when the density of states becomes sufficiently large, it dominates in the denominator and the result is *inversely* proportional to the density of states.

If we were to attempt to use the form of Eq. (13) to obtain agreement with the superconducting experiments within an independent-particle approximation, it would be necessary to take β in the superconductor (and in the same metal when it is normal) to be negligible compared to that in the tunneling region. This seems quite inappropriate and we maintain our conclusion that the independent-particle model fails in that case.

The dependence upon the density of states given in Eq. (13) offers some hope for interesting behavior with semimetals and transition metals, but the hope is weak. This dependence on the details of the wave functions seems unrealistic when the boundary is not ideal. Further, a sample comparison between an exact calculation and a WKB approximation for free-electron tunneling indicates that the WKB approximation becomes quite good as the thickness of the transition approaches and exceeds a single electron wavelength.

B. Diffusion of Transmission

There is no unique way to introduce a diffuse transmission. One might, however, expect three qualitative conclusions to hold in any case.

First, if the boundary is gradual we might still expect the cancellation of the density of states by the velocity in the matrix element even in the diffuse case since this arose from current conservation. In particular, we should not expect the simple proportionality to the density of states to reappear.

Second, we expect no appreciable modification of the results for simple metals (and superconductors) which have isotropic electronic properties.

Third, we expect significant differences in tunneling into semimetals and degenerate semiconductors depending upon the nature of the transmission since their Fermi surfaces lie in limited regions of wave-number space. However, the general smoothing of fluctuations because of the cancellation of the density of states and the contribution of many bands should remain.

[9] J. Bardeen, J. Chem. Phys. **6**, 367 (1938).