SOLID STATE PHYSICS

PART II

Optical Properties of Solids

M. S. Dresselhaus

Contents

Chapter 1

Review of Fundamental Relations for Optical Phenomena

References:

- G. Bekefi and A.H. Barrett, *Electromagnetic Vibrations Waves and Radiation*, MIT Press, Cambridge, MA

- J.D. Jackson, *Classical Electrodynamics*, Wiley, New York, 1975

- Bassani and Pastori–Parravicini, *Electronic States and Optical Transitions in Solids*, Pergamon Press, NY (1975).

- Yu and Cardona, Fundamentals of Semiconductors, Springer Verlag (1996)

1.1   Introductory Remarks on Optical Probes

The optical properties of solids provide an important tool for studying energy band structure, impurity levels, excitons, localized defects, lattice vibrations, and certain magnetic excitations. In such experiments, we measure some observable, such as reflectivity, transmission, absorption, ellipsometry or light scattering; from these measurements we deduce the dielectric function $\varepsilon(\omega)$, the optical conductivity $\sigma(\omega)$, or the fundamental excitation frequencies. It is the frequency-dependent complex dielectric function $\varepsilon(\omega)$ or the complex conductivity $\sigma(\omega)$, which is directly related to the energy band structure of solids.

The central question is the relationship between experimental observations and the electronic energy levels (energy bands) of the solid. In the infrared photon energy region, information on the phonon branches is obtained. These issues are the major concern of Part II of this course.

1.2 The Complex dielectric function and the complex optical conductivity

The complex dielectric function and complex optical conductivity are introduced through Maxwell's equations (c.g.s. units)

$$\nabla \times \vec{H} - \frac{1}{c}\frac{\partial \vec{D}}{\partial t} = \frac{4\pi}{c}\vec{j} \tag{1.1}$$

$$\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} = 0 \tag{1.2}$$

$$\nabla \cdot \vec{D} = 0 \tag{1.3}$$

$$\nabla \cdot \vec{B} = 0 \tag{1.4}$$

where we have assumed that the charge density is zero.

The constitutive equations are written as:

$$\vec{D} = \varepsilon \vec{E} \tag{1.5}$$

$$\vec{B} = \mu \vec{H} \tag{1.6}$$

$$\vec{j} = \sigma \vec{E} \tag{1.7}$$

Equation 1.5 defines the quantity ε from which the concept of the complex dielectric function will be developed. When we discuss non–linear optics (see Chapter 11), these linear constitutive equations (Eqs. 1.5–1.7) must be generalized to include higher order terms in $\vec{E}\vec{E}$ and $\vec{E}\vec{E}\vec{E}$. From Maxwell's equations and the constitutive equations, we obtain a wave equation for the field variables $\vec{E}$ and $\vec{H}$:

$$\nabla^2 \vec{E} = \frac{\varepsilon\mu}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2} + \frac{4\pi\sigma\mu}{c^2}\frac{\partial \vec{E}}{\partial t} \tag{1.8}$$

and

$$\nabla^2 \vec{H} = \frac{\varepsilon\mu}{c^2}\frac{\partial^2 \vec{H}}{\partial t^2} + \frac{4\pi\sigma\mu}{c^2}\frac{\partial \vec{H}}{\partial t}. \tag{1.9}$$

For optical fields, we must look for a sinusoidal solution to Eqs. 1.8 and 1.9

$$\vec{E} = \vec{E}_0 e^{i(\vec{K}\cdot\vec{r}-\omega t)} \tag{1.10}$$

where $\vec{K}$ is a complex propagation constant and ω is the frequency of the light. A solution similar to Eq. 1.10 is obtained for the $\vec{H}$ field. The real part of $\vec{K}$ can be identified as a wave vector, while the imaginary part of $\vec{K}$ accounts for attenuation of the wave inside the solid. Substitution of the plane wave solution Eq. 1.10 into the wave equation Eq. 1.8 yields the following relation for K:

$$-K^2 = -\frac{\varepsilon\mu\omega^2}{c^2} - \frac{4\pi i\sigma\mu\omega}{c^2}. \tag{1.11}$$

If there were no losses (or attenuation), K would be equal to

$$K_0 = \frac{\omega}{c}\sqrt{\varepsilon\mu} \tag{1.12}$$

and would be real, but since there are losses we write

$$K = \frac{\omega}{c} \sqrt{\varepsilon_{\text{complex}} \mu} \tag{1.13}$$

where we have defined the complex dielectric function as

$$\varepsilon_{\text{complex}} = \varepsilon + \frac{4\pi i \sigma}{\omega} = \varepsilon_1 + i\varepsilon_2. \tag{1.14}$$

As shown in Eq. 1.14 it is customary to write ε_1 and ε_2 for the real and imaginary parts of $\varepsilon_{\text{complex}}$. From the definition in Eq. 1.14 it also follows that

$$\varepsilon_{\text{complex}} = \frac{4\pi i}{\omega} \left[\sigma + \frac{\varepsilon \omega}{4\pi i} \right] = \frac{4\pi i}{\omega} \sigma_{\text{complex}}, \tag{1.15}$$

where we define the complex conductivity σ_{complex} as:

$$\sigma_{\text{complex}} = \sigma + \frac{\varepsilon \omega}{4\pi i} \tag{1.16}$$

Now that we have defined the complex dielectric function $\varepsilon_{\text{complex}}$ and the complex conductivity σ_{complex}, we will relate these quantities in two ways:

1. to observables such as the reflectivity which we measure in the laboratory,

2. to properties of the solid such as the carrier density, relaxation time, effective masses, energy band gaps, etc.

After substitution for K in Eq. 1.10, the solution Eq. 1.11 to the wave equation (Eq. 1.8) yields a plane wave

$$\vec{E}(z,t) = \vec{E}_0 e^{-i\omega t} \exp\left(i\frac{\omega z}{c} \sqrt{\varepsilon \mu} \sqrt{1 + \frac{4\pi i \sigma}{\varepsilon \omega}} \right). \tag{1.17}$$

For the wave propagating in vacuum ($\varepsilon = 1$, $\mu = 1$, $\sigma = 0$), Eq. 1.17 reduces to a simple plane wave solution, while if the wave is propagating in a medium of finite electrical conductivity, the amplitude of the wave exponentially decays over a characteristic distance δ given by

$$\delta = \frac{c}{\omega \tilde{N}_2(\omega)} = \frac{c}{\omega \tilde{k}(\omega)} \tag{1.18}$$

where δ is called the optical skin depth, and $\tilde{k}$ is the imaginary part of the complex index of refraction (also called the extinction coefficient)

$$\tilde{N}(\omega) = \sqrt{\mu \varepsilon_{\text{complex}}} = \sqrt{\varepsilon \mu \left(1 + \frac{4\pi i \sigma}{\varepsilon \omega} \right)} = \tilde{n}(\omega) + i\tilde{k}(\omega). \tag{1.19}$$

This means that the *intensity* of the electric field, $|E|^2$, falls off to $1/e$ of its value at the surface in a distance

$$\frac{1}{\alpha_{\text{abs}}} = \frac{c}{2\omega \tilde{k}(\omega)} \tag{1.20}$$

3

where $\alpha_{\mathrm{abs}}(\omega)$ is the absorption coefficient for the solid at frequency ω.

Since light is described by a transverse wave, there are two possible orthogonal directions for the $\vec{E}$ vector in a plane normal to the propagation direction and these directions determine the *polarization* of the light. For cubic materials, the index of refraction is the same along the two transverse directions. However, for anisotropic media, the indices of refraction may be different for the two polarization directions, as is further discussed in §2.1.

1.3 Relation of Complex Dielectric Function to Observables

In relating $\varepsilon_{\mathrm{complex}}$ and $\sigma_{\mathrm{complex}}$ to the observables, it is convenient to introduce a complex index of refraction $\tilde{N}_{\mathrm{complex}}$

$$\tilde{N}_{\mathrm{complex}} = \sqrt{\mu\varepsilon_{\mathrm{complex}}} \tag{1.21}$$

where

$$K = \frac{\omega}{c}\tilde{N}_{\mathrm{complex}} \tag{1.22}$$

and where $\tilde{N}_{\mathrm{complex}}$ is usually written in terms of its real and imaginary parts (see Eq. 1.19)

$$\tilde{N}_{\mathrm{complex}} = \tilde{n} + i\tilde{k} = \tilde{N}_1 + i\tilde{N}_2. \tag{1.23}$$

The quantities $\tilde{n}$ and $\tilde{k}$ are collectively called **the optical constants** of the solid, where $\tilde{n}$ is the index of refraction and $\tilde{k}$ is the extinction coefficient. (We use the tilde over the optical constants $\tilde{n}$ and $\tilde{k}$ to distinguish them from the carrier density and wave vector which are denoted by n and k). The extinction coefficient $\tilde{k}$ vanishes for lossless materials. For non-magnetic materials, we can take $\mu = 1$, and this will be done in writing the equations below.

With this definition for $\tilde{N}_{\mathrm{complex}}$, we can relate

$$\varepsilon_{\mathrm{complex}} = \varepsilon_1 + i\varepsilon_2 = (\tilde{n} + i\tilde{k})^2 \tag{1.24}$$

yielding the important relations

$$\varepsilon_1 = \tilde{n}^2 - \tilde{k}^2 \tag{1.25}$$

$$\varepsilon_2 = 2\tilde{n}\tilde{k} \tag{1.26}$$

where we note that $\varepsilon_1, \varepsilon_2, \tilde{n}$ and $\tilde{k}$ are all frequency dependent.

Many measurements of the optical properties of solids involve the normal incidence reflectivity which is illustrated in Fig. 1.1. Inside the solid, the wave will be attenuated. We assume for the present discussion that the solid is thick enough so that reflections from the back surface can be neglected. We can then write the wave inside the solid for this one-dimensional propagation problem as

$$E_x = E_0 e^{i(Kz-\omega t)} \tag{1.27}$$

where the complex propagation constant for the light is given by $K = (\omega/c)\,\tilde{N}_{\mathrm{complex}}$.

On the other hand, in free space we have both an incident and a reflected wave:

$$E_x = E_1 e^{i(\frac{\omega z}{c}-\omega t)} + E_2 e^{i(\frac{-\omega z}{c}-\omega t)}. \tag{1.28}$$

4
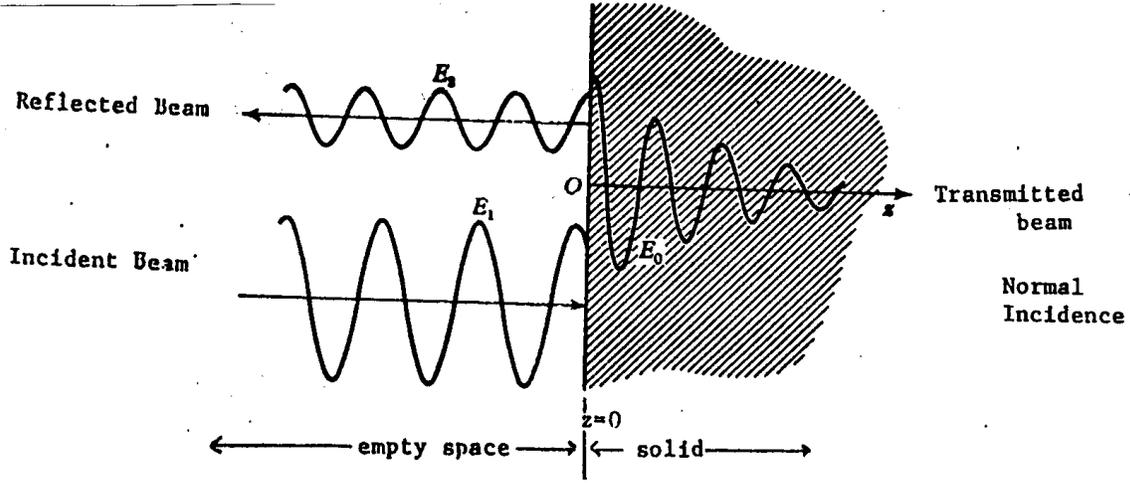
Figure 1.1: Schematic diagram for normal incidence reflectivity.

From Eqs. 1.27 and 1.28, the continuity of E_x across the surface of the solid requires that

$$E_0 = E_1 + E_2. \tag{1.29}$$

With $\vec{E}$ in the x direction, the second relation between E_0, E_1, and E_2 follows from the continuity condition for tangential H_y across the boundary of the solid. From Maxwell's equation (Eq. 1.2) we have

$$\nabla \times \vec{E} = -\frac{\mu}{c}\frac{\partial \vec{H}}{\partial t} = \frac{i\mu\omega}{c}\vec{H} \tag{1.30}$$

which results in

$$\frac{\partial E_x}{\partial z} = \frac{i\mu\omega}{c}H_y. \tag{1.31}$$

The continuity condition on H_y thus yields a continuity relation for $\partial E_x/\partial z$ so that from Eq. 1.31

$$E_0 K = E_1\frac{\omega}{c} - E_2\frac{\omega}{c} = E_0\frac{\omega}{c}\tilde{N}_{\text{complex}} \tag{1.32}$$

or

$$E_1 - E_2 = E_0\tilde{N}_{\text{complex}}. \tag{1.33}$$

The normal incidence reflectivity $\mathcal{R}$ is then written as

$$\mathcal{R} = \left|\frac{E_2}{E_1}\right|^2 \tag{1.34}$$

which is most conveniently related to the reflection coefficient r given by

$$r = \frac{E_2}{E_1}. \tag{1.35}$$

5

From Eqs. 1.29 and 1.33, we have the results

$$E_2 = \frac{1}{2}E_0(1 - \tilde{N}_{\text{complex}}) \tag{1.36}$$

$$E_1 = \frac{1}{2}E_0(1 + \tilde{N}_{\text{complex}}) \tag{1.37}$$

so that the normal incidence reflectivity becomes

$$\mathcal{R} = \left| \frac{1 - \tilde{N}_{\text{complex}}}{1 + \tilde{N}_{\text{complex}}} \right|^2 = \frac{(1 - \tilde{n})^2 + \tilde{k}^2}{(1 + \tilde{n})^2 + \tilde{k}^2} \tag{1.38}$$

where the reflectivity $\mathcal{R}$ is a number less than unity. We have now related one of the physical observables to the optical constants. To relate these results to the power absorbed and transmitted at normal incidence, we utilize the following relation which expresses the idea that all the incident power is either reflected, absorbed, or transmitted

$$1 = \mathcal{R} + \mathcal{A} + \mathcal{T} \tag{1.39}$$

where $\mathcal{R}$, $\mathcal{A}$, and $\mathcal{T}$ are, respectively, the fraction of the power that is reflected, absorbed, and transmitted as illustrated in Fig. 1.1. At high temperatures, the most common observable is the emissivity, which is equal to the absorbed power for a black body or is equal to $1 - \mathcal{R}$ assuming $\mathcal{T}=0$. As a homework exercise, it is instructive to derive expressions for $\mathcal{R}$ and $\mathcal{T}$ when we have relaxed the restriction of no reflection from the back surface. Multiple reflections are encountered in thin films.

The discussion thus far has been directed toward relating the complex dielectric function or the complex conductivity to physical observables. If we know the optical constants, then we can find the reflectivity. We now want to ask the opposite question. Suppose we know the reflectivity, can we find the optical constants? Since there are two optical constants, $\tilde{n}$ and $\tilde{k}$, we need to make two independent measurements, such as the reflectivity at two different angles of incidence.

Nevertheless, even if we limit ourselves to normal incidence reflectivity measurements, we can still obtain both $\tilde{n}$ and $\tilde{k}$ provided that we make these reflectivity measurements for all frequencies. This is possible because the real and imaginary parts of a complex physical function are not independent. Because of causality, $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$ are related through the Kramers–Kronig relation, which we will discuss in Chapter 6. Since normal incidence measurements are easier to carry out in practice, it is quite possible to study the optical properties of solids with just normal incidence measurements, and then do a Kramers–Kronig analysis of the reflectivity data to obtain the frequency–dependent dielectric functions $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ or the frequency–dependent optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$.

In treating a solid, we will need to consider contributions to the optical properties from various electronic energy band processes. To begin with, there are **intraband processes** which correspond to the electronic conduction by free carriers, and hence are more important in conducting materials such as metals, semimetals and degenerate semiconductors. These intraband processes can be understood in their simplest terms by the classical Drude theory, or in more detail by the classical Boltzmann equation or the quantum mechanical density matrix technique. In addition to the intraband (free carrier) processes, there are **interband**

6

processes which correspond to the absorption of electromagnetic radiation by an electron in an occupied state below the Fermi level, thereby inducing a transition to an unoccupied state in a higher band. This interband process is intrinsically a quantum mechanical process and must be discussed in terms of quantum mechanical concepts. In practice, we consider in detail the contribution of only a few energy bands to optical properties; in many cases we also restrict ourselves to detailed consideration of only a portion of the Brillouin zone where strong interband transitions occur. The intraband and interband contributions that are neglected are treated in an approximate way by introducing a core dielectric constant which is often taken to be independent of frequency and external parameters.

1.4   Units for Frequency Measurements

The frequency of light is measured in several different units in the literature. The relation between the various units are: $1 \text{ eV} = 8065.5 \text{ cm}^{-1} = 2.418 \times 10^{14} \text{ Hz} = 11{,}600 \text{ K}$. Also 1 eV corresponds to a wavelength of 1.2398 μm, and $1 \text{ cm}^{-1} = 0.12398 \text{ meV} = 3 \times 10^{10} \text{ Hz}$.

Chapter 2

Drude Theory–Free Carrier Contribution to the Optical Properties

2.1 The Free Carrier Contribution

In this chapter we relate the optical constants to the electronic properties of the solid. One major contribution to the dielectric function is through the "free carriers". Such free carrier contributions are very important in semiconductors and metals, and can be understood in terms of a simple classical conductivity model, called the Drude model. This model is based on the classical equations of motion of an electron in an optical electric field, and gives the simplest theory of the optical constants. The classical equation for the drift velocity of the carrier $\vec{v}$ is given by

$$m\frac{d\vec{v}}{dt} + \frac{m\vec{v}}{\tau} = e\vec{E}_0 e^{-i\omega t} \tag{2.1}$$

where the relaxation time τ is introduced to provide a damping term, $(m\vec{v}/\tau)$, and a sinusoidally time-dependent electric field provides the driving force. To respond to a sinusoidal applied field, the electrons undergo a sinusoidal motion which can be described as

$$\vec{v} = \vec{v}_0 e^{-i\omega t} \tag{2.2}$$

so that Eq. 2.1 becomes

$$(-mi\omega + \frac{m}{\tau})\vec{v}_0 = e\vec{E}_0 \tag{2.3}$$

and the amplitudes $\vec{v}_0$ and $\vec{E}_0$ are thereby related. The current density $\vec{j}$ is related to the drift velocity $\vec{v}_0$ and to the carrier density n by

$$\vec{j} = ne\vec{v}_0 = \sigma\vec{E}_0 \tag{2.4}$$

thereby introducing the electrical conductivity σ. Substitution for the drift velocity v_0 yields

$$\vec{v}_0 = \frac{e\vec{E}_0}{(m/\tau) - im\omega} \tag{2.5}$$

8

into Eq. 2.4 yields the complex conductivity

$$\sigma = \frac{ne^2\tau}{m(1 - i\omega\tau)}. \tag{2.6}$$

In writing σ in the Drude expression (Eq. 2.6) for the free carrier conduction, we have suppressed the subscript in σ_{complex}, as is conventionally done in the literature. In what follows we will always write σ and ε to denote the complex conductivity and complex dielectric constant and suppress subscripts "complex" in order to simplify the notation. A more elegant derivation of the Drude expression can be made from the Boltzmann formulation, as is done in Part I of the notes. In a real solid, the same result as given above follows when the effective mass approximation can be used. Following the results for the dc conductivity obtained in Part I, an electric field applied in one direction can produce a force in another direction because of the anisotropy of the constant energy surfaces in solids. Because of the anisotropy of the effective mass in solids, $\vec{j}$ and $\vec{E}$ are related by the tensorial relation,

$$j_\alpha = \sigma_{\alpha\beta}E_\beta \tag{2.7}$$

thereby defining the conductivity tensor $\sigma_{\alpha\beta}$ as a second rank tensor. For perfectly free electrons in an isotropic (or cubic) medium, the conductivity tensor is written as:

$$\overleftrightarrow{\sigma} = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{pmatrix} \tag{2.8}$$

and we have our usual scalar expression $\vec{j} = \sigma\vec{E}$. However, in a solid, $\sigma_{\alpha\beta}$ can have off-diagonal terms, because the effective mass tensors are related to the curvature of the energy bands $E(\vec{k})$ by

$$\left(\frac{1}{m}\right)_{\alpha\beta} = \frac{1}{\hbar^2}\frac{\partial^2 E(\vec{k})}{\partial k_\alpha \partial k_\beta}. \tag{2.9}$$

The tensorial properties of the conductivity follow directly from the dependence of the conductivity on the reciprocal effective mass tensor.

As an example, semiconductors such as CdS and ZnO exhibit the wurtzite structure, which is a non-cubic structure. These semiconductors are *uniaxial* and contain an *optic axis* (which for the wurtzite structure is along the c-axis), along which the velocity of propagation of light is independent of the polarization direction. Along other directions, the velocity of light is different for the two polarization directions, giving rise to a phenomenon called *birefringence*. Crystals with tetragonal or hexagonal symmetry are uniaxial. Crystals with lower symmetry have two axes along which light propagates at the same velocity for the two polarizations of light, and are therefore called *biaxial*.

Even though the constant energy surfaces for a large number of the common semiconductors are described by ellipsoids and the effective masses of the carriers are given by an effective mass tensor, it is a general result that for cubic materials (in the absence of externally applied stresses and magnetic fields), the conductivity for all electrons and all the holes is described by a single scalar quantity σ. To describe conduction processes in hexagonal materials we need to introduce two constants: $\sigma_\parallel$ for conduction along the high symmetry axis and $\sigma_\perp$ for conduction in the basal plane. These results can be directly demonstrated by summing the contributions to the conductivity from all carrier pockets.

9

In narrow gap semiconductors, $m_{\alpha\beta}$ is itself a function of energy. If this is the case, the Drude formula is valid when $m_{\alpha\beta}$ is evaluated at the Fermi level and n is the total carrier density. Suppose now that the only conduction mechanism that we are treating in detail is the free carrier mechanism. Then we would consider all other contributions in terms of the core dielectric constant $\varepsilon_{\text{core}}$ to obtain for the total complex dielectric function

$$\varepsilon(\omega) = \varepsilon_{\text{core}}(\omega) + 4\pi i\sigma/\omega \tag{2.10}$$

so that

$$\sigma(\omega) = \left(ne^2\tau/m^*\right)(1 - i\omega\tau)^{-1} \tag{2.11}$$

in which $4\pi\sigma/\omega$ denotes the imaginary part of the free carrier contribution. If there were no free carrier absorption, $\sigma = 0$ and $\varepsilon = \varepsilon_{\text{core}}$, and in empty space $\varepsilon = \varepsilon_{\text{core}} = 1$. From the Drude theory,

$$\varepsilon = \varepsilon_{\text{core}} + \frac{4\pi i}{\omega}\frac{ne^2\tau}{m(1 - i\omega\tau)} = (\varepsilon_1 + i\varepsilon_2) = (n_1 + ik_2)^2. \tag{2.12}$$

It is of interest to consider the expression in Eq. 2.12 in two limiting cases: low and high frequencies.

2.2 Low Frequency Response: $\omega\tau \ll 1$

In the low frequency regime ($\omega\tau \ll 1$) we obtain from Eq. 2.12

$$\varepsilon \simeq \varepsilon_{\text{core}} + \frac{4\pi ine^2\tau}{m\omega}. \tag{2.13}$$

Since the free carrier term in Eq. 2.13 shows a $1/\omega$ dependence as $\omega \to 0$, this term dominates in the low frequency limit. The core dielectric constant is typically 16 for geranium, 12 for silicon and perhaps 100 or more, for narrow gap semiconductors like PbTe. It is also of interest to note that the core contribution and free carrier contribution are out of phase.

To find the optical constants $\tilde{n}$ and $\tilde{k}$ we need to take the square root of ε. Since we will see below that $\tilde{n}$ and $\tilde{k}$ are large, we can for the moment ignore the core contribution to obtain:

$$\sqrt{\varepsilon} \simeq \sqrt{\frac{4\pi ne^2\tau}{m\omega}}\sqrt{i} = \tilde{n} + i\tilde{k} \tag{2.14}$$

and using the identity

$$\sqrt{i} = e^{\frac{\pi i}{4}} = \frac{1 + i}{\sqrt{2}} \tag{2.15}$$

we see that in the low frequency limit $\tilde{n} \approx \tilde{k}$, and that $\tilde{n}$ and $\tilde{k}$ are both large. Therefore the normal incidence reflectivity can be written as

$$\mathcal{R} = \frac{(\tilde{n} - 1^2) + \tilde{k}^2}{(\tilde{n} + 1^2) + \tilde{k}^2} \simeq \frac{\tilde{n}^2 + \tilde{k}^2 - 2\tilde{n}}{\tilde{n}^2 + \tilde{k}^2 + 2\tilde{n}} = 1 - \frac{4\tilde{n}}{\tilde{n}^2 + \tilde{k}^2} \simeq 1 - \frac{2}{\tilde{n}}. \tag{2.16}$$

Thus, the Drude theory shows that at low frequencies a material with a large concentration of free carriers (e.g., a metal) is a perfect reflector.

10

2.3 High Frequency Response; $\omega\tau \gg 1$

In this limit, Eq. 2.12 can be approximated by:

$$\varepsilon \simeq \varepsilon_{\text{core}} - \frac{4\pi n e^2}{m\omega^2}. \tag{2.17}$$

As the frequency becomes large, the $1/\omega^2$ dependence of the free carrier contribution guarantees that free carrier effects will become less important, and other processes will dominate. In practice, these other processes are the interband processes which in Eq. 2.17 are dealt with in a very simplified form through the core dielectric constant $\varepsilon_{\text{core}}$. Using this approximation in the high frequency limit, we can neglect the free carrier contribution in Eq. 2.17 to obtain

$$\sqrt{\varepsilon} \cong \sqrt{\varepsilon_{\text{core}}} = \text{real}. \tag{2.18}$$

Equation 2.18 implies that $\tilde{n} > 0$ and $\tilde{k} = 0$ in the limit of $\omega\tau \gg 1$, with

$$\mathcal{R} \rightarrow \frac{(\tilde{n}-1)^2}{(\tilde{n}+1)^2} \tag{2.19}$$

where $\tilde{n} = \sqrt{\varepsilon_{\text{core}}}$. Thus, in the limit of very high frequencies, the Drude contribution is unimportant and the behavior of all materials is like that for a dielectric.

2.4 The Plasma Frequency

Thus, at very low frequencies the optical properties of semiconductors exhibit a metal-like behavior, while at very high frequencies their optical properties are like those of insulators. A characteristic frequency at which the material changes from a metallic to a dielectric response is called the plasma frequency $\hat{\omega}_p$, which is defined as that frequency at which the real part of the dielectric function vanishes $\varepsilon_1(\hat{\omega}_p) = 0$. According to the Drude theory (Eq. 2.12), we have

$$\varepsilon = \varepsilon_1 + i\varepsilon_2 = \varepsilon_{\text{core}} + \frac{4\pi i}{\omega}\frac{ne^2\tau}{m(1-i\omega\tau)}\cdot\left(\frac{1+i\omega\tau}{1+i\omega\tau}\right) \tag{2.20}$$

where we have written ε in a form which exhibits its real and imaginary parts explicitly. We can then write the real and imaginary parts $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ as:

$$\varepsilon_1(\omega) = \varepsilon_{\text{core}} - \frac{4\pi n e^2\tau^2}{m(1+\omega^2\tau^2)} \qquad \varepsilon_2(\omega) = \frac{4\pi}{\omega}\frac{ne^2\tau}{m(1+\omega^2\tau^2)}. \tag{2.21}$$

The free carrier term makes a negative contribution to ε_1 which tends to cancel the core contribution shown schematically in Fig. 2.1.

We see in Fig. 2.1 that $\varepsilon_1(\omega)$ vanishes at some frequency $(\hat{\omega}_p)$ so that we can write

$$\varepsilon_1(\hat{\omega}_p) = 0 = \varepsilon_{\text{core}} - \frac{4\pi n e^2\tau^2}{m(1+\hat{\omega}_p^2\tau^2)} \tag{2.22}$$

which yields

$$\hat{\omega}_p^2 = \frac{4\pi n e^2}{m\varepsilon_{\text{core}}} - \frac{1}{\tau^2} = \omega_p^2 - \frac{1}{\tau^2}. \tag{2.23}$$

11
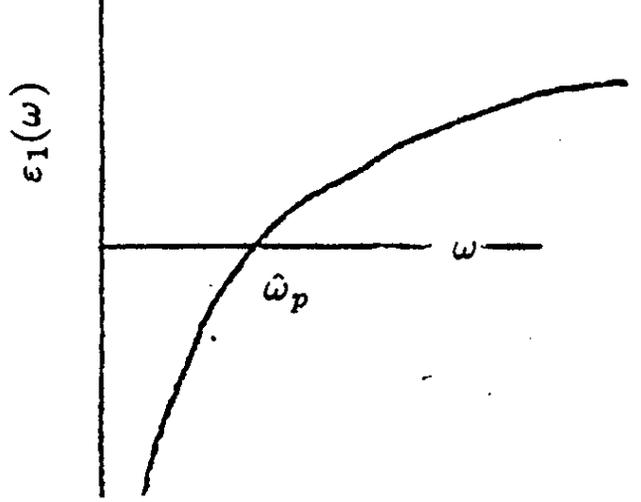
Figure 2.1: The frequency dependence $\varepsilon_1(\omega)$, showing the definition of the plasma frequency $\hat{\omega}_p$ by the relation $\varepsilon_1(\hat{\omega}_p) = 0$.

Since the term $(-1/\tau^2)$ in Eq. 2.23 is usually small compared with ω_p^2, it is customary to neglect this term and to identify the plasma frequency with ω_p defined by

$$\omega_p^2 = \frac{4\pi n e^2}{m \varepsilon_{\text{core}}} \tag{2.24}$$

in which screening of free carriers occurs through the core dielectric constant $\varepsilon_{\text{core}}$ of the medium. If $\varepsilon_{\text{core}}$ is too small, then $\varepsilon_1(\omega)$ never goes positive and there is no plasma frequency. The condition for the existence of a plasma frequency is

$$\varepsilon_{\text{core}} > \frac{4\pi n e^2 \tau^2}{m}. \tag{2.25}$$

The quantity ω_p in Eq. 2.24 is called the screened plasma frequency in the literature. Another quantity called the unscreened plasma frequency obtained from Eq. 2.24 by setting $\varepsilon_{\text{core}} = 1$ is also used in the literature.

The general appearance of the reflectivity as a function of photon energy for a degenerate semiconductor or a metal is shown in Fig. 2.2. At low frequencies, free carrier conduction dominates, and the reflectivity is $\simeq 100\%$. In the high frequency limit, we have

$$\mathcal{R} \sim \frac{(\tilde{n} - 1)^2}{(\tilde{n} + 1)^2}, \tag{2.26}$$

which also is large, if $\tilde{n} \gg 1$. In the vicinity of the plasma frequency, $\varepsilon_1(\omega_1)$ is small by definition; furthermore, $\varepsilon_2(\omega_p)$ is also small, since from Eq. 2.21

$$\varepsilon_2(\omega_p) = \left(\frac{4\pi}{m\omega_p}\right) \frac{n e^2 \tau}{1 + (\omega_p \tau)^2} \tag{2.27}$$

12

Figure 2.2: Reflectivity vs ω for a metal or a degenerate semiconductor in a frequency range where interband transitions are not important and the plasma frequency ω_p occurs near the minimum in reflectivity $\mathcal{R}$.
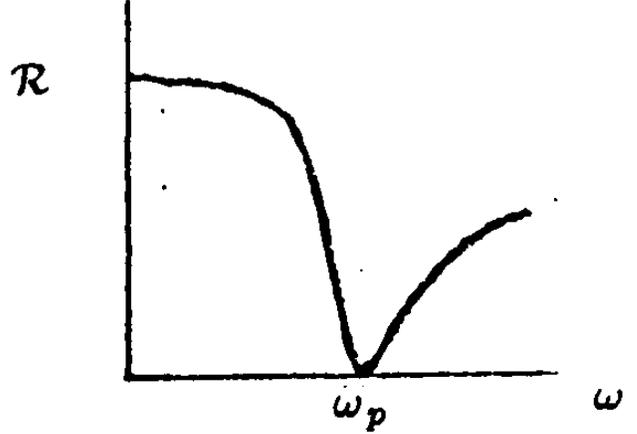
and if $\omega_p\tau \gg 1$

$$\varepsilon_2(\omega_p) \cong \frac{\varepsilon_{\text{core}}}{\omega_p\tau} \tag{2.28}$$

so that $\varepsilon_2(\omega_p)$ is often small. With $\varepsilon_1(\omega_p) = 0$, we have from Eq. 1.25 $\tilde{n} \cong \tilde{k}$, and $\varepsilon_2(\omega_p) = 2\tilde{n}\tilde{k} \simeq 2\tilde{n}^2$. We thus see that $\tilde{n}$ tends to be small near ω_p and consequently $\mathcal{R}$ is also small (see Fig. 2.2). The steepness of the dip at the plasma frequency is governed by the relaxation time τ; the longer the relaxation time τ, the sharper the plasma structure.

In metals, free carrier effects are almost always studied by reflectivity techniques because of the high optical absorption of metals at low frequency. For metals, the free carrier conductivity appears to be quite well described by the simple Drude theory. In studying free carrier effects in semiconductors, it is usually more accurate to use absorption techniques, which are discussed in Chapter 11. Because of the connection between the optical and the electrical properties of a solid through the conductivity tensor, transparent materials are expected to be poor electrical conductors while highly reflecting materials are expected to be reasonably good electrical conductors. It is, however, possible for a material to have its plasma frequency just below visible frequencies, so that the material will be a good electrical conductor, yet be transparent at visible frequencies. Because of the close connection between the optical and electrical properties, free carrier effects are sometimes exploited in the determination of the carrier density in instances where Hall effect measurements are difficult to make.

The contribution of holes to the optical conduction is of the same sign as for the electrons, since the conductivity depends on an even power of the charge ($\sigma \propto e^2$). In terms of the complex dielectric constant, we can write the contribution from electrons and holes as

$$\varepsilon = \varepsilon_{\text{core}} + \frac{4\pi i}{\omega}\left[\frac{n_e e^2 \tau_e}{m_e(1 - i\omega\tau_e)} + \frac{n_h e^2 \tau_h}{m_h(1 - i\omega\tau_h)}\right] \tag{2.29}$$

where the parameters n_e, τ_e, and m_e pertain to the electron carriers and n_h, τ_h, and m_h are for the holes. The plasma frequency is again found by setting $\varepsilon_1(\omega) = 0$. If there are

multiple electron or hole carrier pockets, as is common for semiconductors, the contributions from each carrier type is additive, using a formula similar to Eq. 2.29.

We will now treat another conduction process in Chapter 3 which is due to interband transitions. In the above discussion, interband transitions were included in an extremely approximate way. That is, interband transitions were treated through a frequency independent core dielectric constant $\varepsilon_{\mathrm{core}}$ (see Eq. 2.12). In Chapter 3 we consider the frequency dependence of this important contribution.

Chapter 3

Interband Transitions

3.1 The Interband Transition Process

In a semiconductor at low frequencies, the principal electronic conduction mechanism is associated with free carriers. As the photon energy increases and becomes comparable to the energy gap, a new conduction process can occur. A photon can excite an electron from an occupied state in the valence band to an unoccupied state in the conduction band. This is called an **interband transition** and is represented schematically by the picture in Fig. 3.1. In this process the photon is absorbed, an excited electronic state is formed and a hole is left behind. This process is quantum mechanical in nature. We now discuss the factors that are important in these transitions.

1. We expect interband transitions to have a **threshold energy** at the energy gap. That is, we expect the frequency dependence of the real part of the conductivity $\sigma_1(\omega)$ due to an interband transition to exhibit a threshold as shown in Fig. 3.2 for an allowed electronic transition.

2. The transitions are either **direct** (conserve crystal momentum $\vec{k}$: $E_v(\vec{k}) \rightarrow E_c(\vec{k})$) or **indirect** (a phonon is involved because the $\vec{k}$ vectors for the valence and conduction bands differ by the phonon wave vector $\vec{q}$). Conservation of crystal momentum yields $\vec{k}_{\text{valence}} = \vec{k}_{\text{conduction}} \pm \vec{q}_{\text{phonon}}$. In discussing the direct transitions, one might wonder about conservation of crystal momentum with regard to the photon. The reason we need not be concerned with the momentum of the photon is that it is very small in comparison to Brillouin zone dimensions. For a typical optical wavelength of 6000 Å, the wave vector for the photon $K = 2\pi/\lambda \sim 10^5 \text{cm}^{-1}$, while a typical dimension across the Brillouin zone is 10^8cm^{-1}. Thus, typical direct optical interband processes excite an electron from a valence to a conduction band without a significant change in the wave vector.

3. The transitions depend on the coupling between the valence and conduction bands and this is measured by the magnitude of the momentum matrix elements coupling the valence band state v and the conduction band state c: $|\langle v|\vec{p}|c\rangle|^2$. This dependence results from Fermi's "Golden Rule" (see Chapter A) and from the discussion on the perturbation interaction $\mathcal{H}'$ for the electromagnetic field with electrons in the solid (which is discussed in §3.2).
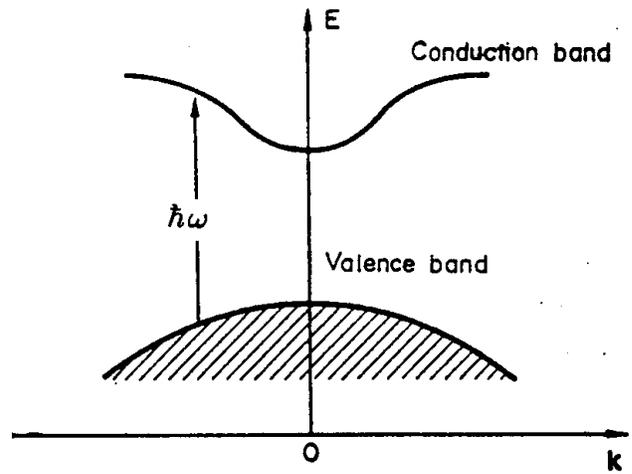
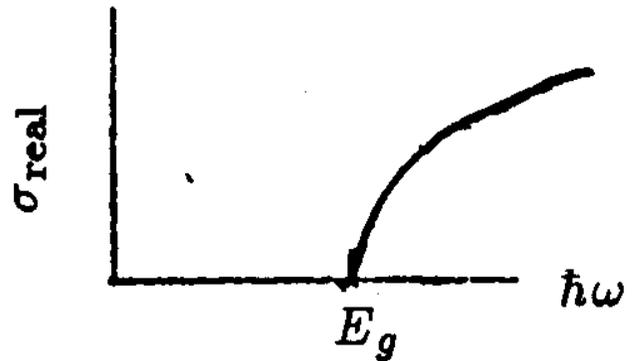Figure 3.1: Schematic diagram of an allowed interband transition.

Figure 3.2: Real part of the conductivity for an allowed optical transition. We note that $\sigma_1(\omega) = (\omega/4\pi)\varepsilon_2(\omega)$.

16

4. Because of the Pauli Exclusion Principle, an interband transition occurs from an occupied state below the Fermi level to an unoccupied state above the Fermi level.

5. Photons of a particular energy are more effective in producing an interband transition if the energy separation between the 2 bands is nearly constant over many $\vec{k}$ values. In that case, there are many initial and final states which can be coupled by the same photon energy. This is perhaps easier to see if we allow a photon to have a small band width. That band width will be effective over many $\vec{k}$ values if $E_c(\vec{k}) - E_v(\vec{k})$ doesn't vary rapidly with $\vec{k}$. Thus, we expect the interband transitions to be most important for $\vec{k}$ values near band extrema. That is, in Fig. 3.1 we see that states around $\vec{k} = 0$ make the largest contribution per unit bandwidth of the optical source. It is also for this reason that optical measurements are so important in studying energy band structure; the optical structure emphasizes band extrema and therefore provides information about the energy bands at **specific points** in the Brillouin zone.

Although we will not derive the expression for the interband contribution to the conductivity, we will write it down here to show how all the physical ideas that were discussed above enter into the conductivity equation. We now write the conductivity tensor relating the interband current density j_α in the direction α which flows upon application of an electric field E_β in direction β

$$j_\alpha = \sigma_{\alpha\beta} E_\beta \tag{3.1}$$

as

$$\sigma_{\alpha\beta} = -\frac{e^2}{m^2} \sum_{i,j} \frac{[f(E_i) - f(E_j)]}{E_i - E_j} \frac{\langle i|p_\alpha|j\rangle\langle j|p_\beta|i\rangle}{[-i\omega + 1/\tau + (i/\hbar)(E_i - E_j)]} \tag{3.2}$$

in which the sum in Eq. 3.2 is over all valence and conduction band states labelled by i and j. Structure in the optical conductivity arises through a singularity in the resonant denominator of Eq. 3.2 $[-i\omega + 1/\tau + (i/\hbar)(E_i - E_j)]$ discussed above under properties (1) and (5).

The appearance of the Fermi functions $f(E_i) - f(E_j)$ follows from the Pauli principle in property (4). The dependence of the conductivity on the momentum matrix elements accounts for the tensorial properties of $\sigma_{\alpha\beta}$ (interband) and relates to properties (2) and (3).

In semiconductors, interband transitions usually occur at frequencies above which free carrier contributions are important. If we now want to consider the total complex dielectric constant, we would write

$$\varepsilon = \varepsilon_{\text{core}} + \frac{4\pi i}{\omega} [\sigma_{\text{Drude}} + \sigma_{\text{interband}}]. \tag{3.3}$$

The term $\varepsilon_{\text{core}}$ contains the contributions from all processes that are not considered explicitly in Eq. 3.3; this would include both intraband and interband transitions **that are not treated explicitly**. We have now dealt with the two most important processes (intraband and interband) involved in studies of electronic properties of solids.

If we think of the optical properties for various classes of materials, it is clear from Fig. 3.3 that major differences will be found from one class of materials to another.
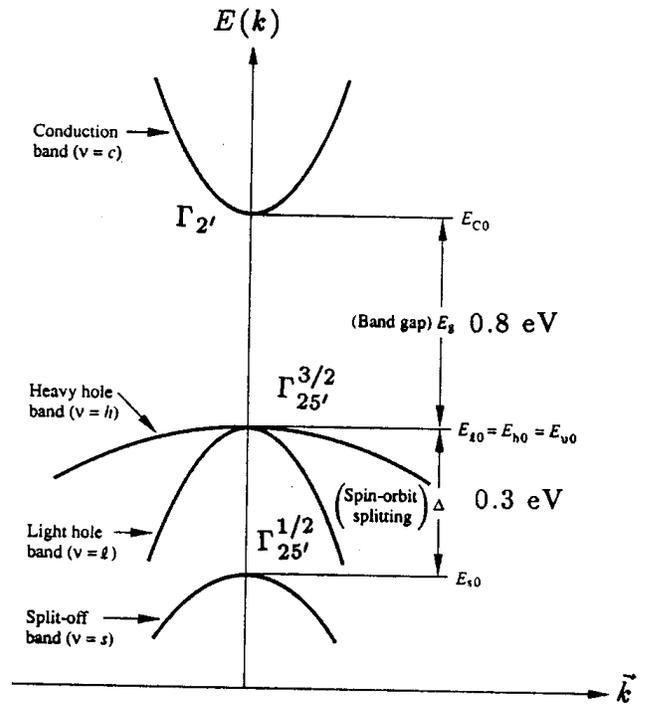
17

Figure 3.3: Structure of the valence band states and the lowest conduction band state at the Γ–point in germanium.

Figure 3.4: Absorption coefficient of germanium at the absorption edge corresponding to the transitions $\Gamma_{25'}^{3/2} \to \Gamma_{2'}(D_1)$ and $\Gamma_{25'}^{1/2} \to \Gamma_{2'}(D_2)$. The energy separation between the $\Gamma_{25'}^{1/2}$ and $\Gamma_{25'}^{3/2}$ bands is determined by the energy differences between the D_1 and D_2 structures.
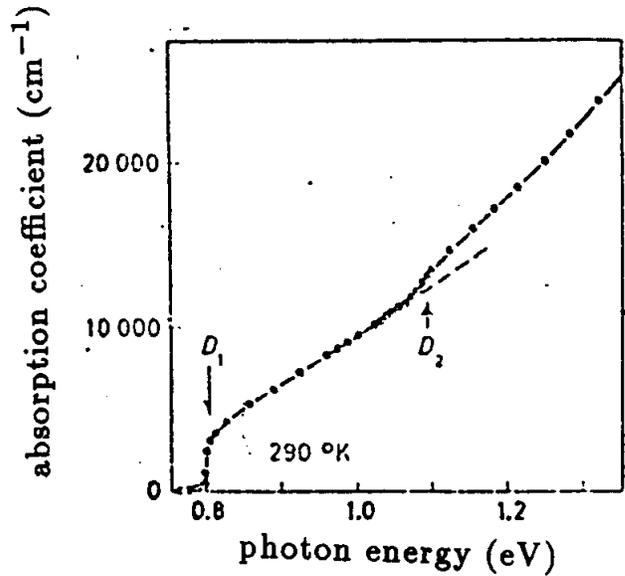
3.1.1 Insulators

Here the band gap is sufficiently large so that at room temperature, essentially no carriers are thermally excited across the band gap. This means that there is no free carrier absorption and that interband transitions only become important at relatively high photon energies (above the visible). Thus, insulators frequently are optically transparent.

3.1.2 Semiconductors

Here the band gap is small enough so that appreciable thermal excitation of carriers occurs at room temperature. Thus there is often appreciable free carrier absorption at room temperature either through thermal excitation or doping. In addition, interband transitions occur in the infrared and visible. As an example, consider the direct interband transition in germanium and its relation to the optical absorption. In the curve in Fig. 3.4, we see that the optical absorption due to optical excitation across the indirect bandgap at 0.7 eV is very small compared with the absorption due to the direct interband transition shown in Fig. 3.4. (For a brief discussion of the spin–orbit interaction as it affects interband transitions see §3.4.)

3.1.3 Metals

Here free carrier absorption is extremely important. Typical plasma frequencies are $\hbar\omega_p \cong$ 10 eV which occur far out in the ultraviolet. In the case of metals, interband transitions typically occur at frequencies where free carrier effects are still important. Semimetals, like metals, exhibit only a weak temperature dependence with carrier densities almost inde-

19

pendent of temperature. Although the carrier densities are low, the high carrier mobilities nevertheless guarantee a large contribution of the free carriers to the optical conductivity.

3.2 Form of the Hamiltonian in an Electromagnetic Field

A proof that the optical field is inserted into the Hamiltonian in the form $\vec{p} \to \vec{p} - e\vec{A}/c$ follows. Consider the classical equation of motion:

$$\frac{d}{dt}(m\vec{v}) = e\left[\vec{E} + \frac{1}{c}(\vec{v} \times \vec{H})\right] = e\left[-\vec{\nabla}\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} + \frac{1}{c}\vec{v} \times (\vec{\nabla} \times \vec{A})\right] \tag{3.4}$$

where ϕ and $\vec{A}$ are, respectively, the scalar and vector potentials, and $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields given by

$$\vec{E} = -\vec{\nabla}\phi - (1/c)\partial\vec{A}/\partial t$$

$$\vec{B} = \vec{\nabla} \times \vec{A}. \tag{3.5}$$

Using standard vector identities, the equation of motion Eq. 3.4 becomes

$$\frac{d}{dt}(m\vec{v} + \frac{e}{c}\vec{A}) = \vec{\nabla}(-e\phi) + \frac{e}{c}\vec{\nabla}(\vec{A} \cdot \vec{v}) \tag{3.6}$$

where $[\vec{\nabla}(\vec{A} \cdot \vec{v})]_j$ denotes $v_i \partial A_i/\partial x_j$ in which we have used the Einstein summation convention that repeated indices are summed and where we have used the vector relations

$$\frac{dA}{dt} = \frac{\partial \vec{A}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{A} \tag{3.7}$$

and

$$[\vec{v} \times (\vec{\nabla} \times \vec{A})]_i = v_j\frac{\partial A_j}{\partial x_i} - v_j\frac{\partial A_i}{\partial x_j}. \tag{3.8}$$

If we write the Hamiltonian as

$$\mathcal{H} = \frac{1}{2m}(\vec{p} - \frac{e}{c}\vec{A})^2 + e\phi \tag{3.9}$$

and then use Hamilton's equations

$$\vec{v} = \frac{\partial \mathcal{H}}{\partial \vec{p}} = \frac{1}{m}(\vec{p} - \frac{e}{c}\vec{A}) \tag{3.10}$$

$$\dot{\vec{p}} = -\vec{\nabla}\mathcal{H} = -e\vec{\nabla}\phi + \frac{e}{c}\vec{\nabla}(\vec{A} \cdot \vec{v}) \tag{3.11}$$

we can show that Eqs. 3.4 and 3.6 are satisfied, thereby verifying that Eq. 3.9 is the proper form of the Hamiltonian in the presence of an electromagnetic field, which has the same form as the Hamiltonian without an optical field except that $\vec{p} \to \vec{p} - (e/c)\vec{A}$. The same transcription is used when light is applied to a solid and is then called the Luttinger transcription. The Luttinger transcription is used in the effective mass approximation where the periodic potential is replaced by the introduction of $\vec{k} \to -(1/i)\vec{\nabla}$ and $m \to m^*$.

The reason why interband transitions depend on the momentum matrix element can be understood from perturbation theory. At any instance of time, the Hamiltonian for an electron in a solid in the presence of an optical field is

$$\mathcal{H} = \frac{(\vec{p} - e/c\vec{A})^2}{2m} + V(\vec{r}) = \frac{p^2}{2m} + V(\vec{r}) - \frac{e}{mc}\vec{A} \cdot \vec{p} + \frac{e^2 A^2}{2mc^2} \tag{3.12}$$

in which $\vec{A}$ is the vector potential due to the optical fields, $V(\vec{r})$ is the periodic potential. Thus, the one-electron Hamiltonian without optical fields is

$$\mathcal{H}_0 = \frac{p^2}{2m} + V(\vec{r}) \tag{3.13}$$

and the optical perturbation terms are

$$\mathcal{H}' = -\frac{e}{mc}\vec{A} \cdot \vec{p} + \frac{e^2 A^2}{2mc^2}. \tag{3.14}$$

Optical fields are generally very weak (unless generated by powerful lasers) and we usually consider only the term linear in $\vec{A}$, the linear response regime. The form of the Hamiltonian in the presence of an electromagnetic field is derived in this section, while the momentum matrix elements $\langle v|\vec{p}|c\rangle$ which determine the strength of optical transitions also govern the magnitudes of the effective mass components (see §3.3). This is another reason why optical studies are very important.

To return to the Hamiltonian for an electromagnetic field (Eq. 3.9), the coupling of the valence and conduction bands through the optical fields depends on the matrix element for the coupling to the electromagnetic field perturbation

$$\mathcal{H}' \cong -\frac{e}{mc}\vec{p} \cdot \vec{A}. \tag{3.15}$$

With regard to the spatial dependence of the vector potential we can write

$$\vec{A} = \vec{A}_0 \exp[i(\vec{K} \cdot \vec{r} - \omega t)] \tag{3.16}$$

where for a loss-less medium $K = \tilde{n}\omega/c = 2\pi\tilde{n}/\lambda$ is a slowly varying function of $\vec{r}$ since $2\pi\tilde{n}/\lambda$ is much smaller than typical wave vectors in solids. Here $\tilde{n}$, ω, and λ are, respectively, the real part of the index of refraction, the optical frequency, and the wavelength of light.

3.3 Relation between Momentum Matrix Elements and the Effective Mass

Because of the relation between the momentum matrix element $\langle v|\vec{p}|c\rangle$, which governs the electromagnetic interaction with electrons and solids, and the band curvature $(\partial^2 E/\partial k_\alpha \partial k_\beta)$, the energy band diagrams provide important information on the strength of optical transitions. Correspondingly, knowledge of the optical properties can be used to infer experimental information about $E(\vec{k})$.

We now derive the relation between the momentum matrix element coupling the valence and conduction bands $\langle v|\vec{p}|c\rangle$ and the band curvature $(\partial^2 E/\partial k_\alpha \partial k_\beta)$. We start with Schrödinger's equation in a periodic potential $V(\vec{r})$ having the Bloch solutions

$$\psi_{n\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}} u_{n\vec{k}}(\vec{r}), \tag{3.17}$$

$$\mathcal{H}\psi_{n\vec{k}}(\vec{r}) = E_n(\vec{k})\psi_{n\vec{k}}(\vec{r}) = \left[\frac{p^2}{2m} + V(\vec{r})\right]e^{i\vec{k}\cdot\vec{r}}u_{n\vec{k}}(\vec{r}) = E_n(\vec{k})e^{i\vec{k}\cdot\vec{r}}u_{n\vec{k}}(\vec{r}). \tag{3.18}$$

Since $\vec{p}$ is an operator $(\hbar/i)\vec{\nabla}$, we can write

$$\vec{p}\,e^{i\vec{k}\cdot\vec{r}}u_{n\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}(\vec{p} + \hbar\vec{k})u_{n\vec{k}}(\vec{r}). \tag{3.19}$$

Therefore the differential equation for $u_{n\vec{k}}(\vec{r})$ becomes

$$\left[\frac{p^2}{2m} + V(\vec{r}) + \frac{\hbar\vec{k}\cdot\vec{p}}{m} + \frac{\hbar^2 k^2}{2m}\right]u_{n\vec{k}}(\vec{r}) = E_n(\vec{k})u_{n\vec{k}}(\vec{r}) \tag{3.20}$$

giving the following differential equation for the periodic function $u_{nk}(\vec{r}) = u_{nk}(\vec{r} + \vec{R}_m)$

$$\left[\frac{p^2}{2m} + V(\vec{r}) + \frac{\hbar\vec{k}\cdot\vec{p}}{m}\right]u_{n\vec{k}}(\vec{r}) = \left[E_n(\vec{k}) - \frac{\hbar^2 k^2}{2m}\right]u_{n\vec{k}}(\vec{r}) \tag{3.21}$$

which we write as follows to put Eq. 3.21 in the canonical form for application of the perturbation theory formulae

$$\mathcal{H}_0 = \frac{p^2}{2m} + V(\vec{r}) \tag{3.22}$$

$$\mathcal{H}' = \frac{\hbar\vec{k}\cdot\vec{p}}{m} \tag{3.23}$$

$$\mathcal{E}_n(\vec{k}) = E_n(\vec{k}) - \frac{\hbar^2 k^2}{2m} \tag{3.24}$$

to yield

$$[\mathcal{H}_0 + \mathcal{H}']u_{n\vec{k}}(\vec{r}) = \mathcal{E}_n(\vec{k})u_{n\vec{k}}(\vec{r}). \tag{3.25}$$

Assume that we know the solution to Eq. 3.25 about a special point $\vec{k}_0$ in the Brillouin zone which could be a band extremum, such as $\vec{k}_0 = 0$. Then the perturbation formulae Eqs. 3.22–3.25 allow us to find the energy and wave function for states near $\vec{k}_0$. For simplicity, we carry out the expansion about the center of the Brillouin zone $\vec{k} = 0$, which is the most important case in practice; the extension of this argument to an energy extremum at arbitrary $\vec{k}_0$ is immediate. Perturbation theory then gives:

$$\mathcal{E}_n(\vec{k}) = E_n(0) + (u_{n,0}|\mathcal{H}'|u_{n,0}) + \sum_{n'\neq n}\frac{(u_{n,0}|\mathcal{H}'|u_{n',0})(u_{n',0}|\mathcal{H}'|u_{n,0})}{E_n(0) - E_{n'}(0)}. \tag{3.26}$$

The first order term $(u_{n,0}|\mathcal{H}'|u_{n,0})$ in Eq. 3.26 normally vanishes about an extremum because of inversion symmetry, with $\mathcal{H}'$ being odd under inversion and the two wavefunctions $u_{nk}(\vec{r})$ both being even or both being odd. Since

$$\mathcal{H}' = \frac{\hbar\vec{k}\cdot\vec{p}}{m} \tag{3.27}$$

the matrix element is then written as

$$(u_{n,0}|\mathcal{H}'|u_{n',0}) = \frac{\hbar}{m}\vec{k}\cdot(u_{n,0}|\vec{p}|u_{n',0}). \tag{3.28}$$

We now apply Eq. 3.26 to optical transitions, for the simplest case of a two band model. Here we assume that:

22

1. bands n and n' (valence (v) and conduction (c) bands) are close to each other and far from other bands

2. interband transitions occur between these two bands separated by an energy gap E_g.

We note that the perturbation theory is written in terms of the energy $\mathcal{E}_n(k)$

$$\mathcal{E}_n(k) = E_n(\vec{k}) - \frac{\hbar^2 k^2}{2m}. \tag{3.29}$$

Assuming that the first order term in perturbation theory (Eq. 3.26) can be neglected by parity (even and oddness) arguments, we obtain for $\mathcal{E}_n(k)$ about $\vec{k} = 0$

$$\mathcal{E}_n(\vec{k}) = E_n(0) + \frac{\hbar^2}{m^2} k_\alpha k_\beta \frac{|(v|p_\alpha|c)(c|p_\beta|v)|}{E_g} \tag{3.30}$$

or in terms of the energy eigenvalues of Schrödinger's equation (Eq. 3.18)

$$E_n(\vec{k}) = E_n(0) + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^2} k_\alpha k_\beta \frac{|(v|p_\alpha|c)(c|p_\beta|v)|}{E_g}. \tag{3.31}$$

We define the effective mass tensor by the relation

$$E_n(\vec{k}) = E_n(0) + \frac{\hbar^2}{2} \sum_{\alpha,\beta} k_\alpha k_\beta \left(\frac{1}{m^*}\right)_{\alpha\beta} \tag{3.32}$$

so that

$$\left(\frac{1}{m^*}\right)_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{m} + \frac{2}{m^2} \frac{|(v|p_\alpha|c)(c|p_\beta|v)|}{E_g} \tag{3.33}$$

where $\delta_{\alpha\beta}$ is the unit matrix. This discussion shows that the non-vanishing momentum matrix element is responsible for the inequality between the free electron m and the effective mass m^* in the solid. With regard to the optical properties of solids we note that the same momentum matrix element that governs the effective mass formula (Eq. 3.33) also governs the electromagnetic interaction given by Eq. 3.15. Thus small effective masses tend to give rise to strong coupling between valence and conduction bands and large values for $|(v|p|c)|^2$. On the other hand, small effective masses lead to a small density of states because of the $m^{*3/2}$ dependence of the density of states.

3.4 Spin-Orbit Interaction in Solids

Reference:

- Jones and March, pp. 85-87, 89-94.

- Eisberg and Resnick, Quantum Physics pp. 278-281.

A spin angular momentum $S_z = \hbar/2$ and a magnetic moment $\mu_B = |e|\hbar/2mc = 0.927 \times 10^{-20}$ erg/gauss is associated with each electron. The magnetic moment and spin angular momentum for the free electron are related by

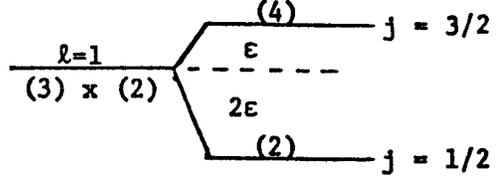$$\vec{\mu} = \frac{-|e|}{mc}\vec{S} = \frac{-|e|}{mc} \cdot \frac{\hbar}{2}\hat{S} \tag{3.34}$$

23

Figure 3.5: Schematic diagram showing the splitting of the $\ell = 1$ level by the spin–orbit interaction.

($\hat{S}$ is a unit vector along $\vec{S}$), and $\vec{\mu}$ and $\vec{S}$ are oppositely directed because the electron is negatively charged.

An electron in an atom sees a magnetic field because of its own orbital motion and consequently there is an interaction called the *spin-orbit interaction* whereby the magnetic field due to the orbital motion of the electron tends to line up its magnetic moment along the magnetic field:

$$\mathcal{H}'_{\text{S.O.}} = -\vec{\mu} \cdot \vec{H}. \tag{3.35}$$

$$\mathcal{H}'_{\text{S.O.}} = \frac{1}{2m^2c^2}(\nabla V \times \vec{p}) \cdot \vec{S} \tag{3.36}$$

since $e\vec{E} \sim -\vec{\nabla}V$. For an atom Eq. 3.36 results in

$$\mathcal{H}_{\text{S.O.atom}} = \xi(r)\vec{L} \cdot \vec{S}. \tag{3.37}$$

A detailed discussion of this topic is found in any standard quantum mechanics text.

This spin-orbit interaction gives rise to a spin-orbit splitting of the atomic levels corresponding to different values of the total angular momentum J

$$\vec{J} = \vec{L} + \vec{S} \tag{3.38}$$

where $\vec{L}$ and $\vec{S}$, respectively, denote the orbital and spin angular momentum. Thus

$$\vec{J} \cdot \vec{J} = (\vec{L} + \vec{S}) \cdot (\vec{L} + \vec{S}) = \vec{L} \cdot \vec{L} + \vec{S} \cdot \vec{S} + (\vec{L} \cdot \vec{S} + \vec{S} \cdot \vec{L}) \tag{3.39}$$

in which the operators $\vec{L}$ and $\vec{S}$ commute.

We take matrix elements in the $|j, \ell, s, m_j\rangle$ representation, because m_ℓ, m_s are not good quantum numbers, to obtain, with $j = |\ell - s|, (|\ell - s| + 1), \ldots, \ell + s$,

$$j(j+1) = \ell(\ell+1) + s(s+1) + 2\langle \vec{L} \cdot \vec{S} \rangle \tag{3.40}$$

so that the expectation value of $\vec{L} \cdot \vec{S}$ in the $|j, \ell, s, m_j\rangle$ representation becomes:

$$\langle \vec{L} \cdot \vec{S} \rangle = \frac{1}{2}[j(j+1) - \ell(\ell+1) - s(s+1)] \tag{3.41}$$

For p states, $\ell = 1, s = 1/2$ and $j = 3/2$ or $1/2$ as shown in Fig. 3.5. From Eq. 3.41 we can find the expectation value of $\langle \vec{L} \cdot \vec{S} \rangle$. In particular, we note that the degeneracy of an s-state is unaffected by the spin-orbit interaction. On the other hand, a d-state is split

up into a doublet $D_{5/2}$ (6-fold degenerate) and $D_{3/2}$ (4-fold degenerate). Thus, the spin-orbit interaction does not lift all the degeneracy for atomic states. To lift this additional degeneracy it is necessary to apply a magnetic field.

The magnitude of the spin-orbit interaction depends also on the expectation value of $\xi(r)$ defined by the following relation,

$$\langle n, j, \ell, s, m_j | \mathcal{H}'_{\text{S.O.}} | n, j, \ell, s, m_j \rangle = \langle j, \ell, s, m_j | \vec{L} \cdot \vec{S} | j, \ell, s, m_j \rangle \int_0^\infty R_{n\ell}\, \xi(r)\, R_{n\ell}\, dr \quad (3.42)$$

where the atomic wave function is written

$$\Phi = Y_{\ell m}(\theta, \phi) R_{n\ell}(r) \quad (3.43)$$

and $R_{n\ell}(r)$ denotes the radial part of the atomic wave function. We note that the integral over r in Eq. 3.42 increases rapidly with atomic number ($\sim Z^3$ or Z^4). The physical reason behind this sensitive dependence on Z is that heavier atoms have more electrons generating larger H fields, and therefore a greater spin-orbit splitting results.

References for tabulated spin-orbit splittings are:

- C.E. Moore – Atomic Energy Levels (National Bureau of Standards, Circular #467), vol. 1 (1949), vol. 2 (1952) and vol. 3 (1958). These references give the measured spectroscopic levels for any atom in a large number of excited configurations. The lowest Z values are in vol. 1, the highest in vol. 3.

- F. Herman and S. Skillman – Atomic Structure Calculation (Prentice-Hall, Inc. 1963). Most complete listing of calculated atomic levels.

- Landolt and Bornstein – Physical and Chemical Tables (many volumes in Reference section in the Science Library).

For most atomic species that are important in semiconductor physics, the spin-orbit interaction is important. Some typical values are:

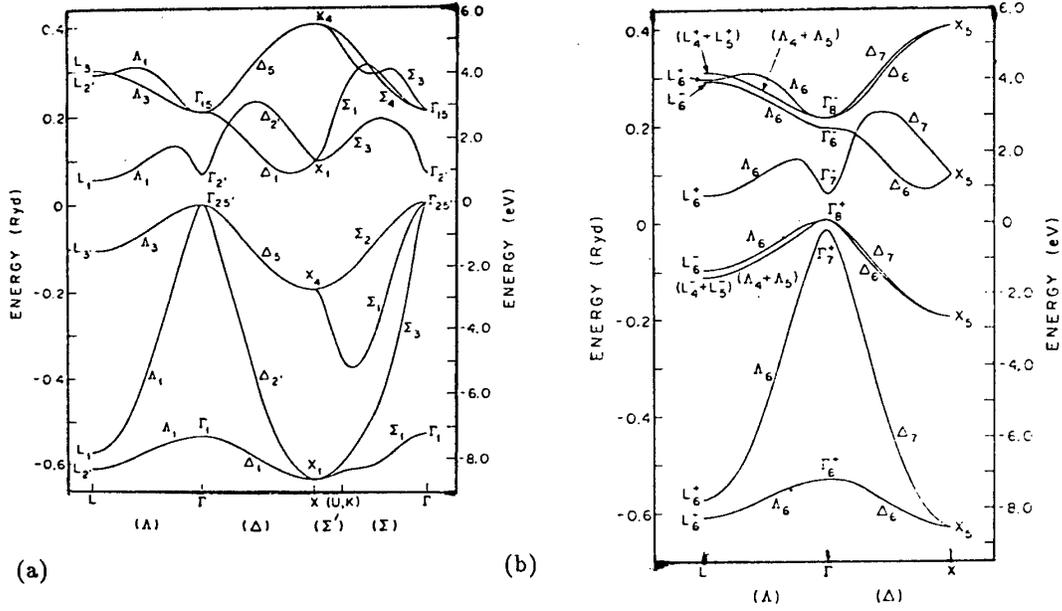| semiconductor	atomic number	Γ-point splitting
diamond	$Z = 6$	$\Delta = 0.006\text{eV}$
silicon	$Z = 14$	$\Delta = 0.044\text{eV}$
germanium	$Z = 32$	$\Delta = 0.290\text{eV}$
tin	$Z = 50$	$\Delta = 0.527\text{eV}$
InSb		
In	$Z = 49$	$\Delta = 0.274\text{eV}$
Sb	$Z = 51$	$\Delta = 0.815\text{eV}$
GaAs		
Ga	$Z = 31$	$\Delta = 0.103\text{eV}$
As	$Z = 33$	$\Delta = 0.364\text{eV}$
PbTe, HgTe		
Pb	$Z = 82$	$\Delta = 1.746\text{eV}$
Hg	$Z = 80$	$\Delta = 1.131\text{eV}$
Te	$Z = 52$	$\Delta = 1.143\text{eV}$

Figure 3.6: Energy bands of Ge: (a) without and (b) with spin–orbit interaction.

The listing above gives the Γ point splittings. The spin-orbit splittings are k-dependent and at the L-point are typically about 2/3 of the Γ point value.

The one-electron Hamiltonian for a solid including spin-orbit interaction is from Eq. 3.36

$$\mathcal{H} = \frac{p^2}{2m} + V(r) - \frac{1}{2m^2c^2}(\nabla V \times \vec{p}) \cdot \vec{S}. \tag{3.44}$$

When the electron spin is considered, the wave functions consist of a spatial and a spin part. The effect of the spin-orbit interaction is to introduce a partial lifting of the degeneracy of band states at high symmetry points in the Brillouin zone. Also, it is a convention in the literature to use a different labeling scheme for the energy bands when the spin-orbit interaction is included. To show the effect of the spin-orbit interaction on the energy bands of a semiconductor, consider the energy bands for germanium. We show in Fig. 3.6 the $E(\vec{k})$ vs. $\vec{k}$ along the $\Delta(100)$ axis, $\Lambda(111)$ axis and $\Sigma(110)$ axes for no spin-orbit interaction and with spin-orbit interaction.

As an example of the effect of the spin-orbit interaction, consider the valence band at the Γ-point ($\vec{k} = 0$) which is labeled by $\Gamma_{25'}$ when there is no spin-orbit interaction. The $\Gamma_{25'}$ band is triply degenerate at $\vec{k} = 0$, each of the three orbital levels containing a spin up and a spin down electron. With spin-orbit interaction, this band splits into the Γ_8^+ (doubly degenerate) band and the Γ_7^+ (non-degenerate) band. In the literature, the Γ_7^+ band is called the *split-off band.* In germanium the band gap is 0.8eV and the splitting between the Γ_8^+ and Γ_7^+ bands is 0.3eV. However, in InSb, the spin-orbit interaction is large and the separation between the upper valence band and the split-off band is 0.9eV, which is much larger than the band gap of 0.2eV between the valence and conduction bands.

26

Chapter 4

The Joint Density of States and Critical Points

References:

- Jones and March, *Theoretical Solid State Physics*: pp. 806-814

- Bassani and Pastori–Parravicini, *Electronic States and Optical Transitions in Solids*: chapter 5

- Yu and Cardona, Fundamentals of Semiconductors, pp. 251-258

- Madelung, *Introduction to Solid State Theory*: pp. 262-271

4.1   The Joint Density of States

The detailed calculation of the contribution to $\varepsilon(\omega)$ due to interband transitions is rather difficult. It is therefore instructive to obtain an approximate answer by use of the Fermi Golden Rule (Eq. A.32). The Golden Rule gives us the probability per unit time $W_{\vec{k}}$ that a photon of energy $\hbar\omega$ makes a transition at a given $\vec{k}$ point:

$$W_{\vec{k}} \cong \frac{2\pi}{\hbar} |\langle v|\mathcal{H}'|c\rangle|^2 \delta[E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega] \tag{4.1}$$

where the matrix element for the electromagnetic perturbation $\mathcal{H}'$ is taken between the valence and conduction band Bloch states at wave vector $\vec{k}$ and the δ-function $\delta[E_c - E_v - \hbar\omega]$ which expresses energy conservation is also evaluated at $\vec{k}$. In writing Eq. 4.1, we exploit the fact that the wave vector for the light is small compared to the Brillouin zone dimensions. Because the electronic states in the Brillouin zone are quasi–continuous functions of $\vec{k}$, to obtain the lineshape for an interband transition, we must integrate over $\vec{k}$. Recognizing that both the perturbation matrix elements and the joint density of states are $\vec{k}$-dependent, we obtain upon integration of Eq. 4.1 over $\vec{k}$ space

$$W = \frac{2\pi}{\hbar} \int |\langle v|\mathcal{H}'|c\rangle|^2 \frac{2}{8\pi^3} \delta(E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega) \, d^3k \tag{4.2}$$

27

for a 3D system. For 2D and 1D systems, we replace $[d^3k/(2\pi)^3]$ by $[d^2k/(2\pi)^2]$ and $[dk/(2\pi)]$, respectively. The perturbation Hamiltonian for the electromagnetic interaction is simply

$$\mathcal{H}' = -\frac{e\vec{A} \cdot \vec{p}}{mc} \tag{4.3}$$

where the time dependence of the vector potential $\vec{A}$ has already been taken into account, so that $\vec{A}$ is a vector with only spatial dependence. In taking matrix elements of the perturbation Hamiltonian, we need then only consider matrix elements of the momentum operator connecting the valence and conduction bands. In practical cases it is often not necessary to evaluate these matrix elements explicitly because it is precisely these momentum matrix elements that determine the experimentally measured effective masses (see §3.3). If we assume for simplicity that $|\langle v|\mathcal{H}'|c\rangle|^2$ is independent of $\vec{k}$, then the remaining integral is the joint density of states between the valence and conduction bands $\rho_{cv}(\hbar\omega)$. For a 3D system, we thus define $\rho_{cv}(\hbar\omega)$ as

$$\rho_{cv}(\hbar\omega) \equiv \frac{2}{8\pi^3} \int \delta[E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega]\, d^3k \tag{4.4}$$

and $\rho_{cv}(\hbar\omega)$ is the number of states per unit volume per unit energy range which occur with an energy difference between the conduction and valence bands equal to the photon energy. As explained above, $\rho_{cv}(\hbar\omega)$ can be evaluated in a similar manner for 2D and 1D systems.

We would now like to look at this joint density of states (Eq. 4.4) in more detail to see why the optical properties of solids give unique information about the energy band structure. The main point is that optical measurements provide information about the bands at *particular $\vec{k}$ points* in the Brillouin zone, usually points of high symmetry and near energy band extrema. This can be understood by casting $\rho_{cv}(\hbar\omega)$ in a more transparent form. We start with the definition of the joint density of states given in Eq. 4.4. It is convenient to convert this integral over $\vec{k}$-space to an integral over energy. This is done by introducing a constant energy surface S in k-space such that the energy difference $E_c - E_v = \hbar\omega$ is the photon energy. Then we can introduce the constant energy surfaces S and $S + dS$ in reciprocal space as corresponding to a constant energy difference between the conduction and valence bands at each $\vec{k}$ point and:

$$d^3k = dS\, dk_n \tag{4.5}$$

where dk_n is an element of a wave vector normal to S, as shown in Fig. 4.1.

By definition of the gradient, we have $|\nabla_k E|dk_n = dE$ so that for surfaces with energy difference $E_c - E_v$ we write:

$$|\nabla_k(E_c - E_v)|dk_n = d(E_c - E_v). \tag{4.6}$$

Therefore

$$d^3k = dk_n dS = dS\left[\frac{d(E_c - E_v)}{|\nabla_k(E_c - E_v)|}\right] \tag{4.7}$$

so that

$$\rho_{cv}(\hbar\omega) = \frac{2}{8\pi^3} \int \int \int \frac{dS\, d(E_c - E_v)\delta(E_c - E_v - \hbar\omega)}{|\nabla_k(E_c - E_v)|}. \tag{4.8}$$
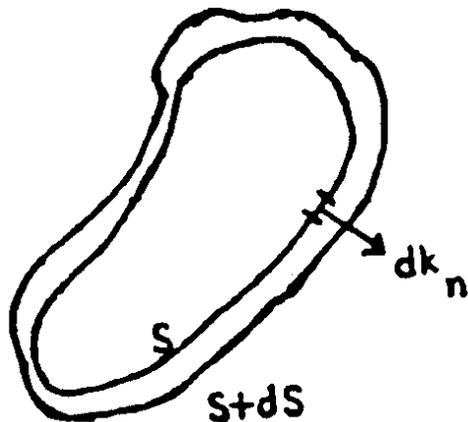
28

Figure 4.1: Adjacent constant energy difference surfaces in reciprocal space, S and $S + dS$, where the energy difference is between valence and conduction bands, and dk_n is the normal to these constant energy difference surfaces.

We now carry out the integral over $d(E_c - E_v)$ to obtain

$$\rho_{cv}(\hbar\omega) = \frac{2}{8\pi^3} \int \int \frac{dS}{|\nabla_k(E_c - E_v)|_{E_c - E_v = \hbar\omega}}. \tag{4.9}$$

Of special interest are those points in the Brillouin zone where $(E_c - E_v)$ is stationary and $\nabla_k(E_c - E_v)$ vanishes. At such points, called *joint critical points*, the denominator of the integrand in Eq. 4.9 vanishes and especially large contributions can be made to $\rho_{cv}(\hbar\omega)$. This can be understood on the basis of physical considerations. Around critical points, the photon energy $\hbar\omega = (E_c - E_v)$ is effective in inducing electronic transitions over a relatively larger region of the Brillouin zone than would be the case for transitions about non-critical points. The relatively large contributions to the transition probability for critical points gives rise to "structure" observed in the frequency dependence of the optical properties of solids. Critical points generally occur at high symmetry points in the Brillouin zone, though this is not necessarily the case.

As an illustration, let us consider the energy bands of the semiconductor germanium (see Fig. 4.2). Here we see that both the valence and conduction bands have extrema at the Γ point, $\vec{k} = 0$, although the lowest conduction band minimum is located at the L point. For the band extrema at $\vec{k} = 0$, the condition $[E_c(k = 0) - E_v(k = 0)] = \hbar\omega$ gives rise to critical points in the joint density of states. Notice also that around the L points, extrema occur in both valence and conduction bands, and a critical point therefore results. Since the energy difference $[E_c - E_v]$ has a relatively small gradient as we move away from the L point, this critical point participates more fully in the interband transitions. In fact, for germanium, Fig. 4.2 shows that there are large regions along the (100) and (111) axes where the energy separation between valence and conduction bands $(E_c - E_v)$ is roughly constant. These large regions in k-space make very large contributions to the dielectric function. We can see these features directly by looking at the frequency dependence of the real and imaginary parts of the dielectric function for germanium (see Fig. 4.3). Here we see that at low photon energies (below ~ 2 eV), where the interband transitions from the
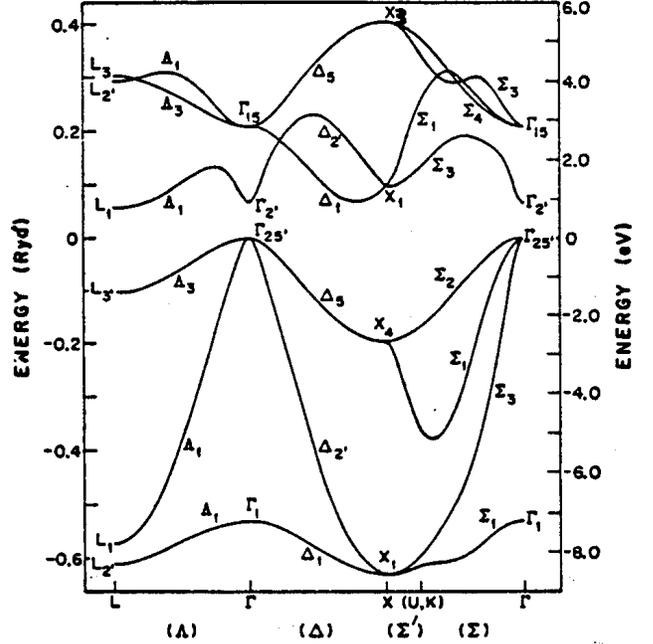
29

Figure 4.2: $E(\vec{k})$ for a few high symmetry directions in germanium, neglecting the spin-orbit interaction.

$\Gamma_{25'}$ valence band to the $\Gamma_{2'}$ conduction band dominate, the contributions to the real and imaginary parts of the dielectric function are small. On the other hand, the contributions from the large regions of the Brillouin zone along the (100) and (111) axes between 2 and 5 eV are very much more important, as is seen in Fig. 4.3 for both $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$.

In describing this contribution to the dielectric function of germanium we say that the valence and conduction bands track each other and in this way produce a large joint density of states over large regions of the Brillouin zone. A similar situation occurs in silicon and in common III-V semiconductors. The diagram in Fig. 4.2 shows that beyond ~ 5 eV there is no longer any significant tracking of the valence and conduction bands. Consequently, the magnitudes of $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ fall sharply beyond ~ 5 eV. The absolute magnitudes of ε_1 and ε_2 for germanium and other semiconductors crystallizing in the diamond or zincblende structure are relatively large. We will see shortly when we discuss the Kramers-Kronig relations in §6.1 that these large magnitudes of ε_1 and ε_2 are responsible for the large value of $\varepsilon_1(\omega \to 0)$ in these materials. For germanium $\varepsilon_1(0)$ is 16 from Fig. 4.3.

4.2 Critical Points

For a 3D system, critical points (often called Van Hove singularities) are classified into four categories depending on whether the band separations are increasing or decreasing as we move away from the critical point. This information is found by expanding $[E_c(\vec{k}) - E_v(\vec{k})]$ in a Taylor series around the critical point $\vec{k_0}$ which is at an energy difference extremum.
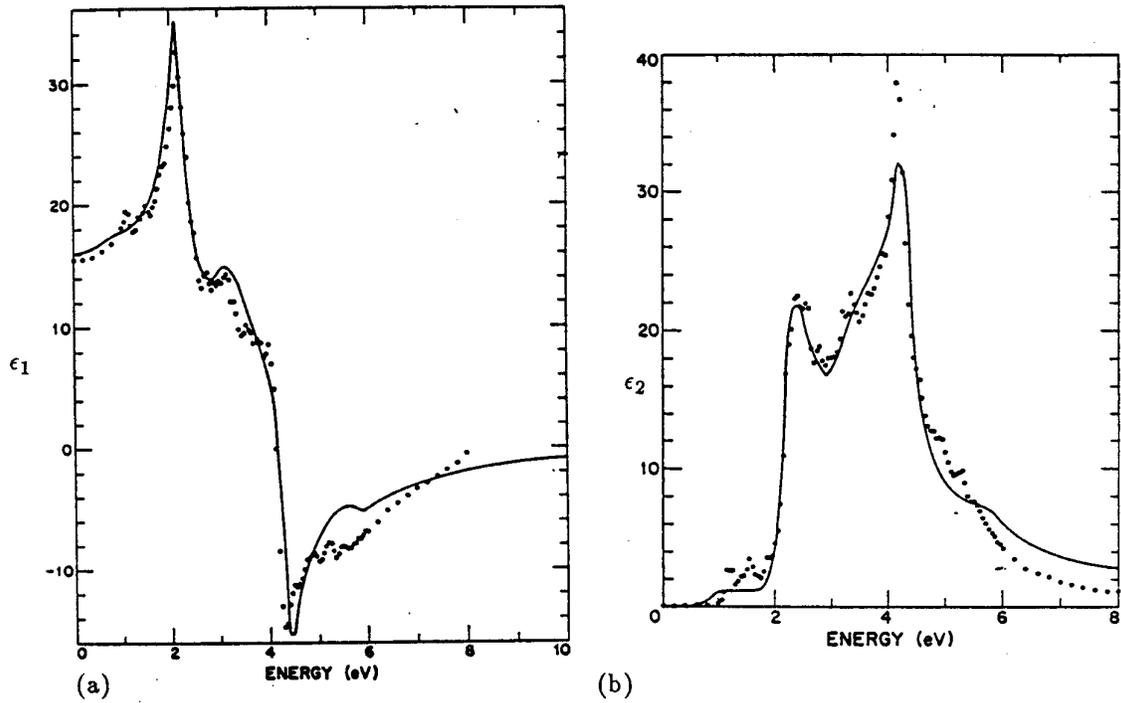
Figure 4.3: Frequency dependence of the real (ε_1) and imaginary (ε_2) parts of the dielectric function for germanium. The solid curves are obtained from an analysis of experimental normal-incidence reflectivity data while the dots are calculated from an energy band model.
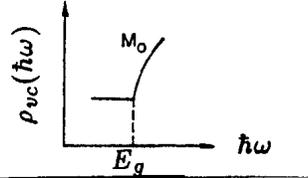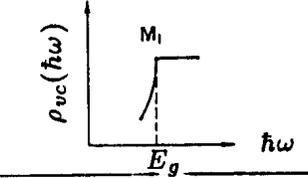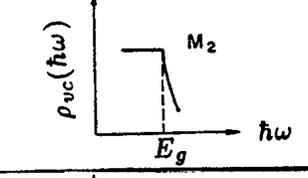
Type of singularity	Number of negative a_i's	Joint density of states
M_0 minimum	0	
M_1 saddlepoint	1	
M_2 saddlepoint	2	
M_3 maximum	3	
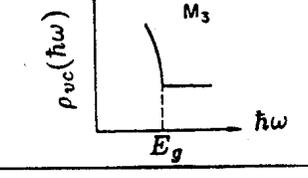
Figure 4.4: Summary of the joint density of states for a 3D system near each of the distinct type of critical point.

Thus,

$$E_c(\vec{k}) - E_v(\vec{k}) = E_g(\vec{k_0}) + \sum_{i=1}^{3} a_i(k_i - k_{0i})^2 \qquad (4.10)$$

where the energy gap at the expansion point is written as $E_g(\vec{k_0})$ and the sum is over the three directions x, y, and z. The coefficients a_i represent the second derivative of the energy difference $\frac{\partial^2}{\partial k_i^2}[E_c(\vec{k}) - E_v(\vec{k})]$. The classification of the critical points in a 3D system shown in Fig. 4.4 is made according to *how many a_i coefficients* in Eq. 4.10 are *negative*. The shapes given for the joint density of states curves of Fig. 4.4 are obtained as is here illustrated for the case of an M_0 singularity for a 3D system. In the case of 2D and 1D systems, there are 3 and 2 types of critical points, respectively, using the same definition of the coefficients a_i to define the type of critical point.

As an example we will calculate $\rho_{cv}(\hbar\omega)$ for an M_0 singularity in a 3D system assuming simple parabolic bands (see Fig. 4.5). Here,

$$E_c(\vec{k}) = \frac{E_g}{2} + \frac{\hbar^2 k^2}{2m_c} \qquad (4.11)$$

32
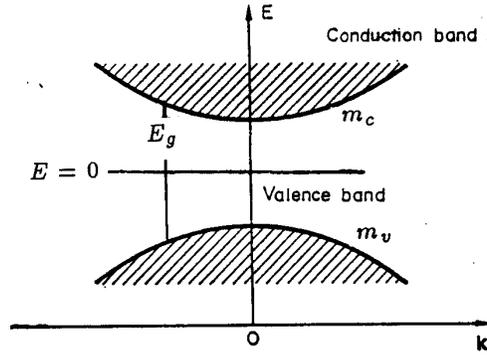
Figure 4.5: Bands associated with a M_0 critical point for a 3D system.

and

$$E_v(\vec{k}) = \frac{-E_g}{2} - \frac{\hbar^2 k^2}{2m_v} \tag{4.12}$$

where E_g is the energy gap, and m_c and m_v are effective masses for the conduction and valence bands, respectively, and m_v is taken as a positive number. We thus obtain

$$E_c(\vec{k}) - E_v(\vec{k}) = E_g + \frac{\hbar^2 k^2}{2}\left(\frac{1}{m_c} + \frac{1}{m_v}\right) = E_g + \frac{\hbar^2 k^2}{2m_r} \tag{4.13}$$

where we define the reduced mass m_r through the relation

$$\frac{1}{m_r} = \frac{1}{m_c} + \frac{1}{m_v}. \tag{4.14}$$

Taking the gradient of $E_c - E_v$ yields

$$\nabla_k(E_c - E_v) = \frac{\hbar^2 \vec{k}}{m_r} \tag{4.15}$$

so that the joint density of states becomes

$$\rho_{cv}(\hbar\omega) = \frac{2}{8\pi^3} \int \frac{dS}{|\nabla_k(E_c - E_v)|}\Big|_{E_c - E_v = \hbar\omega} \tag{4.16}$$

or

$$\rho_{cv}(\hbar\omega) = \frac{2}{8\pi^3}\left[\frac{4\pi}{\hbar^2}\left(\frac{k^2 m_r}{k}\right)\right]_{E_c - E_v = \hbar\omega} = \left[\frac{m_r}{\pi^2 \hbar^2}k\right]_{E_c - E_v = \hbar\omega}. \tag{4.17}$$

We evaluate k in Eq. 4.17 from the condition

$$E_c - E_v = \hbar\omega = E_g + \frac{\hbar^2 k^2}{2m_r} \tag{4.18}$$

or

$$k = \left[\frac{2m_r}{\hbar^2}(\hbar\omega - E_g)\right]^{1/2} \tag{4.19}$$

33

Figure 4.6: Two cases of band extrema which are associated with M_0 critical points. (a) Conduction band minimum and a valence band maximum and (b) Both bands showing minima.

so that

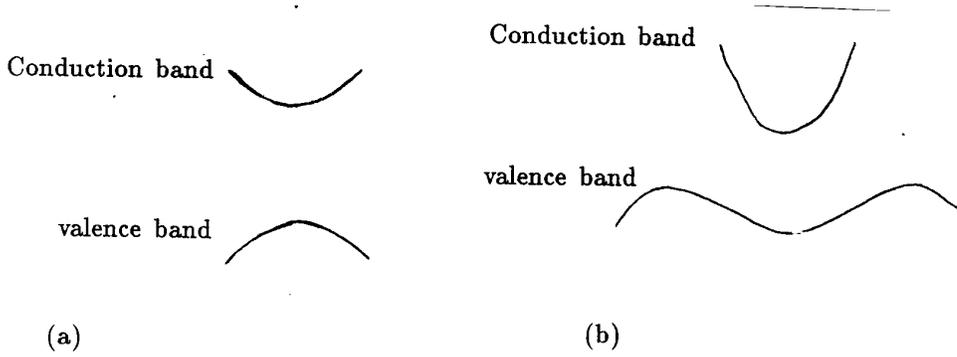$$\rho_{cv}(\hbar\omega) = \frac{1}{2\pi^2}\left[\frac{2m_r}{\hbar^2}\right]^{3/2}\sqrt{\hbar\omega - E_g} \tag{4.20}$$

as shown in Fig. 4.4 for an M_0 critical point. The expression for $\rho_{cv}(\hbar\omega)$ in Eq. 4.20 is not singular but represents a discontinuity in slope at $\hbar\omega = E_g$. This discontinuity in slope corresponds to a *threshold* for the absorption process, as discussed in Chapter 5.

On the other hand, the situation is quite different for the joint density of states corresponding to an M_0 critical point for a 3D system in a magnetic field, as we will see in Part III of the class notes. At a critical point, the joint density of states in a magnetic field *does* show singularities where the density of states in a magnetic field becomes infinite. These singularities in a magnetic field make it possible to carry out resonance experiments in solids, despite the quasi–continuum of the energy levels in the energy bands $E(\vec{k})$.

We note that we can have M_0-type critical points for bands that look like Fig. 4.6a or like Fig. 4.6b. It is clear that the difference $E_c - E_v$ in Fig. 4.6b varies more slowly around the critical point than it does in Fig. 4.6a. Thus, bands that tend to "track" each other have an exceptionally high joint density of states and contribute strongly to the optical properties. Examples of bands that track each other are found in common semiconductors like germanium along the Λ (111) direction (see Figs. 4.2 and 4.3).

In addition to the M_0 critical points, we have M_1, M_2, and M_3 critical points in 3D systems. The functional forms for the joint density of states for $\hbar\omega < E_g$ and $\hbar\omega > E_g$ are given in Table 4.1. From the table we see that in 2D, the M_0 and M_2 critical points correspond to discontinuities in the joint density of states at E_g, while the M_1 singularity corresponds to a saddle point logarithmic divergence. In the case of the 1D system, both the M_0 and M_1 critical points are singular.

Table 4.1: Functional form for the joint density of states for various types of singularities below and above the energy gap E_g for 3D, 2D, and 1D systems $\rho_{vc}(\hbar\omega)$.

	Type	$\hbar\omega < E_g$	$\hbar\omega > E_g$
3D	M_0	0	$(\hbar\omega - E_g)^{1/2}$
	M_1	$C - (E_g - \hbar\omega)^{1/2}$	C
	M_2	C	$C - (\hbar\omega - E_g)^{1/2}$
	M_3	$(E_g - \hbar\omega)^{1/2}$	0
2D	M_0	0	C
	M_1	$-\ln(E_g - \hbar\omega)$	$-\ln(\hbar\omega - E_g)$
	M_2	C	0
1D	M_0	0	$(\hbar\omega - E_g)^{-1/2}$
	M_0	$(E_g - \hbar\omega)^{-1/2}$	0

Chapter 5

Absorption of Light in Solids

References:

- Ziman, *Principles of the Theory of Solids*: Chapter 8

- Bassani and Pastori–Parravicini, *Electronic States and Optical Transitions in Solids*: chapter 5

- Yu and Cardona, Fundamentals of Semiconductors, Chapter 6

- Wolfe, Holonyak and Stillman, Physical Properties of Semiconductors, Chapter 7

5.1 The Absorption Coefficient

Measurement of the absorption of light is one of the most important techniques for optical measurements in solids. In the absorption measurements, we are concerned with the light intensity $I(z)$ after traversal of a thickness z of material as compared with the incident intensity I_0, thereby defining the absorption coefficient $\alpha_{\text{abs}}(\omega)$:

$$I(z) = I_0 e^{-\alpha_{\text{abs}}(\omega)z} \tag{5.1}$$

where the absorption constant is shown schematically in Fig. 5.1. Since the intensity $I(z)$ depends on the square of the field variables, it immediately follows that

$$\alpha_{\text{abs}}(\omega) = 2\frac{\omega \tilde{k}(\omega)}{c} \tag{5.2}$$

where the factor of 2 results from the definition of $\alpha_{\text{abs}}(\omega)$ in terms of the light intensity, which is proportional to the *square* of the fields. This expression tells us that the absorption coefficient is proportional to $\tilde{k}(\omega)$, the imaginary part of the complex index of refraction (extinction coefficient), so that $\tilde{k}$ is usually associated with power loss. We note that Eq. 5.2 applies to free carrier absorption in semiconductors in the limit $\omega\tau \gg 1$, and $\omega \gg \omega_p$.

We will now show that the frequency dependence of the absorption coefficient is quite different for the various physical processes which occur in the optical properties of solids. We will consider here the frequency dependence of the absorption coefficient for:
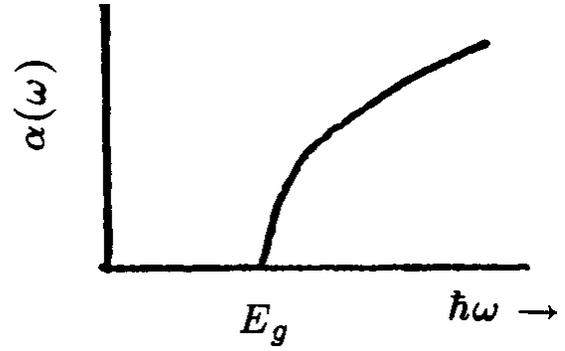
1. Free carrier absorption

Figure 5.1: Frequency dependence of the absorption coefficient near a threshold for interband transitions.

    (a) typical semiconductor $\alpha_{\mathrm{abs}}(\omega) \sim \omega^{-2}$

    (b) metals at low frequencies $\alpha_{\mathrm{abs}}(\omega) \sim \omega^{\frac{1}{2}}$

2. Direct interband transitions

    (a) form of absorption coefficient $\alpha_{\mathrm{abs}}(\omega) \sim \frac{(\hbar\omega - E_g)^{\frac{1}{2}}}{\hbar\omega}$

    (b) conservation of crystal momentum

    (c) relation between m^* and momentum matrix element

    (d) form of $\alpha_{\mathrm{abs}}(\omega)$ for direct forbidden transition $\sim \frac{(\hbar\omega - E_g)^{\frac{3}{2}}}{\hbar\omega}$

3. Indirect interband transitions

    (a) form of absorption coefficient $\alpha_{\mathrm{abs}}(\omega) \sim (\hbar\omega - E_g \pm \hbar\omega_q)^2$

    (b) phonon absorption and emission processes

The summary given above is for 3D systems. In the case of 2D and 1D systems, the functional dependence is sensitive to the dimensionality of the system for each process.

5.2 Free Carrier Absorption in Semiconductors

For free carrier absorption we use the relation for the complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ given by

$$\varepsilon(\omega) = \varepsilon_0 + \frac{4\pi i \sigma}{\omega} \tag{5.3}$$

where ε_0 is the core dielectric constant in the optical frequency range above the lattice mode frequencies and ε_0 is here assumed to be independent of ω. The electronic polarizability is related to the frequency dependent electrical conductivity by the frequency dependent Drude term

$$\sigma = \frac{ne^2\tau}{m^*(1 - i\omega\tau)}. \tag{5.4}$$

37

The plasma frequency ω_p is then given by the vanishing of $\varepsilon_1(\omega)$, that is $\varepsilon_1(\omega_p) = 0$ or

$$\omega_p^2 = \frac{4\pi n e^2}{m^* \varepsilon_0}. \tag{5.5}$$

For semiconductors, the core dielectric constant ε_0 is typically a large number and the contribution due to the free carriers is small at infrared and visible frequencies. For metals, the free carrier absorption is dominant over the entire optical frequency range.

For semiconductors, the typical frequency range of interest is that above the optical phonon frequencies, and for these frequencies it is generally true that $\omega\tau \gg 1$. We can then write:

$$\varepsilon(\omega) = \varepsilon_0 + \frac{4\pi i n e^2 \tau (1 + i\omega\tau)}{m^* \omega [1 + (\omega\tau)^2]} = \varepsilon_0 + \frac{i\varepsilon_0 \omega_p^2 \tau (1 + i\omega\tau)}{\omega[1 + (\omega\tau)^2]} \tag{5.6}$$

or for $\omega\tau \gg 1$

$$\varepsilon(\omega) \simeq \varepsilon_0 + \frac{i\varepsilon_0 \omega_p^2 \tau^2}{\omega^3 \tau^3} - \frac{\varepsilon_0 \omega_p^2}{\omega^2}. \tag{5.7}$$

In the range of interest for optical measurements in a semiconductor, the relation $\omega \gg \omega_p$ is satisfied. It is then convenient to express the complex dielectric function $\varepsilon(\omega)$ in terms of the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$ according to the definition $\varepsilon(\omega) = [\tilde{n}(\omega) + i\tilde{k}(\omega)]^2$ where $\tilde{n}(\omega)$ is the index of refraction and $\tilde{k}(\omega)$ is the extinction coefficient. We can then write for the real part of the dielectric function:

$$\varepsilon_1(\omega) \equiv \tilde{n}^2(\omega) - \tilde{k}^2(\omega) \approx \varepsilon_0 \tag{5.8}$$

where the index of refraction $\tilde{n}(\omega)$ is large and the extinction coefficient $\tilde{k}(\omega)$ is small. For the imaginary part of the dielectric function, we have

$$\varepsilon_2(\omega) \equiv 2\tilde{n}(\omega)\tilde{k}(\omega) \approx 2\sqrt{\varepsilon_0}\, \tilde{k}(\omega) = \frac{\varepsilon_0 \omega_p^2 \tau^2}{\omega^3 \tau^3} \tag{5.9}$$

which is small, since $\omega_p \ll \omega$. Thus the absorption coefficient can be written as:

$$\alpha_{\text{abs}}(\omega) = \frac{2\omega\tilde{k}(\omega)}{c} \simeq \frac{2\omega}{c} \frac{\varepsilon_0 \omega_p^2}{2\sqrt{\varepsilon_0}\omega^3 \tau} = \frac{\sqrt{\varepsilon_0}\omega_p^2}{c\omega^2 \tau} \tag{5.10}$$

and thus $\alpha_{\text{abs}}(\omega)$ is proportional to $1/\omega^2$ or to λ^2 for free carrier absorption in semiconductors for the case where $\omega\tau \gg 1$ and $\omega \gg \omega_p$. Figure 5.2 shows a plot of the optical absorption coefficient for InAs vs wavelength on a log-log plot for various carrier densities, showing that $\alpha_{\text{abs}}(\omega) \sim \lambda^p$ where p is between 2 and 3 for a wide range of donor concentrations. The dependence of the reflectivity spectra (vs wavelength) for various donor concentrations for heavily doped n-type InSb is shown in Fig. 5.3. The dependence of the plasma frequency on the carrier concentration is readily visible from these data.

5.3   Free Carrier Absorption in Metals

The typical limits for metals are somewhat different than for semiconductors. In particular we consider here the case where $\omega\tau \ll 1$, $\omega \ll \omega_p$, $|\varepsilon_0| \ll 4\pi\sigma/\omega$, so that $\tilde{n} \simeq \tilde{k}$. Thus we obtain

$$\varepsilon(\omega) \simeq \frac{4\pi i \sigma}{\omega} \simeq \frac{4\pi i n e^2 \tau}{\omega m^*} \simeq i\varepsilon_2(\omega) \equiv 2i\tilde{n}\tilde{k} \simeq 2i\tilde{k}^2 \tag{5.11}$$

38
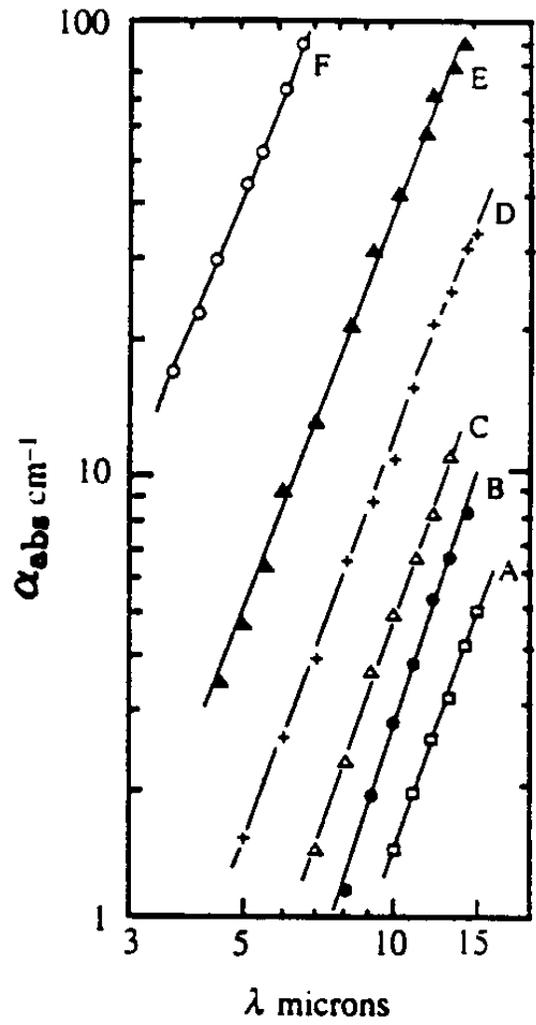
Figure 5.2: Free carrier absorption in n-type InAs at room temperature for six different carrier concentrations (in units of 10^{17} cm^{-3}) A: 0.28; B: 0.85; C: 1.4; D: 2.5; E: 7.8; and F: 39.0.

Figure 5.3: Plasma edges observed in the room temperature reflectivity spectra of n-type InSb with carrier concentration n varying between 3.5×10^{17} cm^{-3} and 4.0×10^{18} cm^{-3}. The solid curves are theoretical fits to the experimental points, including consideration of the energy dependence of m^* due to the strong interband coupling (called non-parabolic effects).
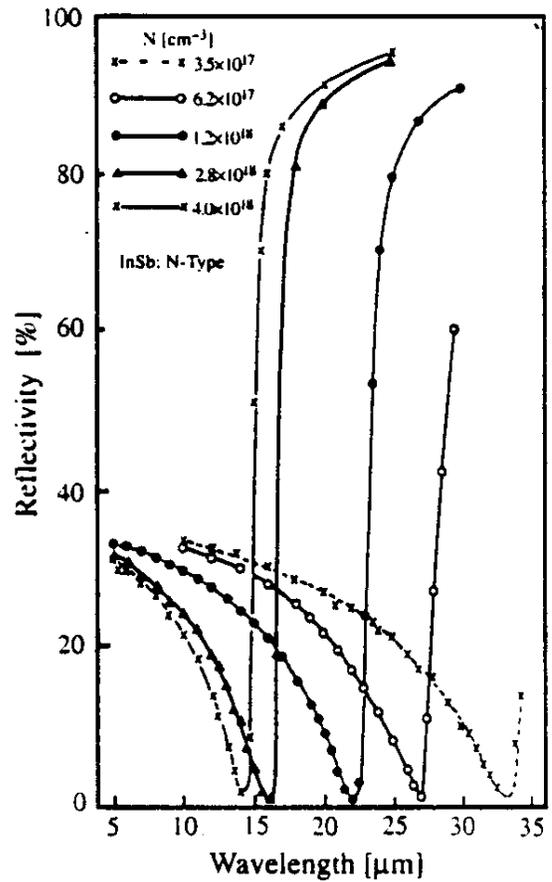
This gives us for the extinction coefficient $\tilde{k}(\omega)$

$$\tilde{k}(\omega) = \sqrt{\frac{2\pi ne^2\tau}{m^*\omega}} \tag{5.12}$$

and the absorption coefficient becomes:

$$\alpha_{\text{abs}}(\omega) = \frac{2\omega\tilde{k}(\omega)}{c} = \sqrt{\frac{8\pi\omega ne^2\tau}{m^*c^2}} \tag{5.13}$$

For this limit $\alpha_{\text{abs}}(\omega)$ *is proportional to* $\sqrt{\omega}$. Usually, the convenient observable for metals is the reflectivity. In the limit appropriate for metals, $\tilde{n} = \tilde{k}$, and both $\tilde{n}$ and $\tilde{k}$ are large. We thus have

$$\mathcal{R} = \frac{(\tilde{n}-1)^2 + \tilde{k}^2}{(\tilde{n}+1)^2 + \tilde{k}^2} = \frac{\tilde{n}^2 - 2\tilde{n} + 1 + \tilde{k}^2}{\tilde{n}^2 + 2\tilde{n} + 1 + \tilde{k}^2} = 1 - \frac{4\tilde{n}}{\tilde{n}^2 + \tilde{k}^2 + 2\tilde{n} + 1} \tag{5.14}$$

$$\mathcal{R} \approx 1 - \frac{4\tilde{n}}{\tilde{n}^2 + \tilde{k}^2} \approx 1 - \frac{2}{\tilde{n}}. \tag{5.15}$$

But from Eq. 5.12 and the condition $\tilde{n} \approx \tilde{k} \gg 1$, we obtain

$$\tilde{n}(\omega) \simeq \sqrt{\frac{2\pi me^2\tau}{m^*\omega}} \tag{5.16}$$

so that the reflectivity goes as

$$\mathcal{R}(\omega) \simeq 1 - 2\sqrt{\frac{m^*\omega}{2\pi ne^2\tau}}. \tag{5.17}$$

Equation 5.17 is known as the Hagen-Rubens relation which holds well for most metals in the *infrared region* of the spectrum. This formula also applies to *degenerate semiconductors* below the plasma frequency.

5.4 Direct Interband Transitions

To calculate the absorption due to direct interband transitions we go back to the definition for the absorption coefficient $\alpha_{\text{abs}}(\omega)$ which is defined as the power removed from the incident beam per unit volume per unit incident flux of electromagnetic energy:

$$\alpha_{\text{abs}}(\omega) = \frac{(\hbar\omega) \times \text{number of transitions/unit volume/unit time}}{\text{incident electromagnetic flux}}. \tag{5.18}$$

The incident electromagnetic flux is calculated from the Poynting vector

$$\vec{S} = \frac{c}{8\pi} Re(\vec{E}^* \times \vec{H}). \tag{5.19}$$

It is convenient to relate the field variables to the vector potential:

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} = \frac{i\omega}{c}\vec{A} \tag{5.20}$$

41

$$\mu \vec{H} = \vec{B} = \vec{\nabla} \times \vec{A}. \tag{5.21}$$

In non–magnetic materials we can take the permeability μ to be unity. In taking the curl of $\vec{A}$, we assume a plane wave form

$$\vec{A} = \vec{A}_0 e^{i(\vec{K}\cdot\vec{r}-\omega t)} \tag{5.22}$$

where the propagation constant for the light is denoted by the wave vector $\vec{K}$. We thus obtain for the Poynting vector

$$\vec{S} = \frac{c}{8\pi} Re\left[-\frac{i\omega}{c} \vec{A}^* \times (i\vec{K} \times \vec{A}) \right] \tag{5.23}$$

or

$$\vec{S} = \frac{\omega}{8\pi} Re\left[(\vec{A}^* \cdot \vec{A})\vec{K} - (\vec{A}^* \cdot \vec{K})\vec{A} \right]. \tag{5.24}$$

Utilizing the fact that for a transverse plane wave $\vec{A}^* \cdot \vec{K} = 0$, we obtain

$$\vec{S} = \frac{\omega}{8\pi} \frac{\tilde{n}\omega}{c} |A|^2 \hat{K} \tag{5.25}$$

where $\tilde{n}$ denotes the real part of the complex index of refraction and $\hat{K}$ is a unit vector along the Poynting vector. This quantity $|\vec{S}|$ in Eq. 5.25 becomes the denominator in Eq. 5.18 which is the expression defining the absorption coefficient. The transition probability/unit time/unit volume is calculated from the "Fermi Golden Rule"

$$W = \frac{2\pi}{\hbar} \mid \mathcal{H}'_{vc} \mid^2 \rho_{cv}(\hbar\omega). \tag{5.26}$$

If we wish to consider the absorption process at finite temperature, we also need to include the Fermi functions to represent the occupation of the states at finite temperature

$$f(E_v)[1 - f(E_c)] - f(E_c)[1 - f(E_v)] \tag{5.27}$$

in which the first group of terms represents the absorption process which depends on the valence band (v) being nearly full and the conduction band (c) being nearly empty. The second group of terms represents the emission process which proceeds if there are occupied conduction states and unoccupied valence states. Clearly, the Fermi functions in Eq. 5.27 simply reduce to $[f(E_v) - f(E_c)]$. The matrix elements $|\mathcal{H}'_{vc}|^2$ in Eq. 5.26 can be written in terms of the electromagnetic interaction Hamiltonian

$$\mathcal{H}'_{vc} = \langle v|\mathcal{H}'_{em}|c\rangle = -\left(\frac{e}{mc}\right)\langle v|\vec{A}(\vec{r},t)\cdot\vec{p}|c\rangle. \tag{5.28}$$

We show in §5.5 that the matrix element $\langle v|\vec{A}(\vec{r},t)\cdot\vec{p}|c\rangle$ coupling the valence and conduction bands for the electromagnetic interaction is diagonal in wave vector $\vec{k}$ since the wave vector for light $\vec{K}$ is small relative to Brillouin zone dimensions. As a result also the spatial dependence of the vector potential can be ignored. Thus the square of the matrix elements coupling the valence and conduction bands becomes

$$|\mathcal{H}'_{vc}|^2 = \left(\frac{e}{mc}\right)^2 |A|^2 |\langle v|p|c\rangle|^2, \tag{5.29}$$

42

where $|\langle v|p|c\rangle|^2$ couples states with the same electron wave vector in the valence and conduction bands. Since $|\langle v|p|c\rangle|^2$ is slowly varying with k in comparison to $\rho_{cv}(\hbar\omega)$ it is convenient to neglect the k dependence of $|\langle v|p|c\rangle|^2$ and evaluate this quantity at the M_i critical point. Thus for direct interband transitions, we obtain the following expression for the absorption coefficient

$$\alpha_{\text{abs}}(\omega) = \frac{(\hbar\omega)[\frac{2\pi}{\hbar}(\frac{e}{mc})^2|A|^2|\langle v|p|c\rangle|^2\rho_{cv}(\hbar\omega)][f(E_v) - f(E_c)]}{\frac{\omega}{8\pi}\frac{\tilde{n}\omega}{c}|A|^2} \tag{5.30}$$

or

$$\alpha_{\text{abs}}(\omega) = \frac{16\pi^2 e^2}{m^2 c\tilde{n}\omega}|\langle v|p|c\rangle|^2\rho_{cv}(\hbar\omega)[f(E_v) - f(E_c)] \tag{5.31}$$

where $\tilde{n}$ in Eqs. 5.30 and 5.31 denotes the index of refraction.

To get an idea of the functional forms of the quantities in Eq. 5.31, we will consider a rather simplified picture of two simple parabolic bands with an allowed optical transition, i.e., a non-vanishing momentum matrix element coupling them. Writing the joint density of states from Eq. 4.4 for the case of an M_0 critical point (as occurs near $k = 0$ for many semiconductors)

$$\rho_{cv}(\hbar\omega) = \frac{1}{2\pi^2}\left(\frac{2m_r}{\hbar^2}\right)^{3/2}\sqrt{\hbar\omega - E_g} \tag{5.32}$$

where m_r is the reduced mass for the valence and conduction bands, we can estimate the absorption coefficient $\alpha_{\text{abs}}(\omega)$. At very low temperature, a semiconductor has an essentially filled valence band and an empty conduction band; that is $f(E_v) = 1$ and $f(E_c) = 0$. We can estimate $|\langle v|p|c\rangle|^2$ from the effective mass sum–rule (Eq. 3.33)

$$|\langle v|p|c\rangle|^2 \simeq \frac{m_0 E_g}{2}\frac{m_0}{m^*} \tag{5.33}$$

where m_0 is the free electron mass. After substitution of Eqs. 5.32 and 5.33 into Eq. 5.31 we obtain the following frequency dependence for the absorption coefficient for direct allowed transitions:

$$\alpha_{\text{abs}}(\omega) \propto \frac{1}{\omega}\sqrt{\hbar\omega - E_g} \tag{5.34}$$

so that the direct optically–allowed interband transitions are characterized by a threshold at the energy gap E_g as shown in Fig. 5.1. We thus see a very different frequency dependence of $\alpha_{\text{abs}}(\omega)$ for the various physical processes.

It is sometimes convenient to express the absorption in terms of the imaginary part of the dielectric function

$$\varepsilon_2(\omega) = \frac{\tilde{n}c}{\omega}\alpha_{\text{abs}}(\omega) \tag{5.35}$$

which from Eq. 5.31 becomes

$$\varepsilon_2(\omega) = \left(\frac{4\pi e}{m\omega}\right)^2|\langle v|p|c\rangle|^2\rho_{cv}(\hbar\omega)[f(E_v) - f(E_c)]. \tag{5.36}$$

If we introduce the dimensionless quantity f_{vc}, which is usually called the oscillator strength and is defined by

$$f_{vc} = \frac{2|\langle v|p|c\rangle|^2}{m[E_c(k) - E_v(k)]} = \frac{2|\langle v|p|c\rangle|^2}{m\hbar\omega}, \tag{5.37}$$
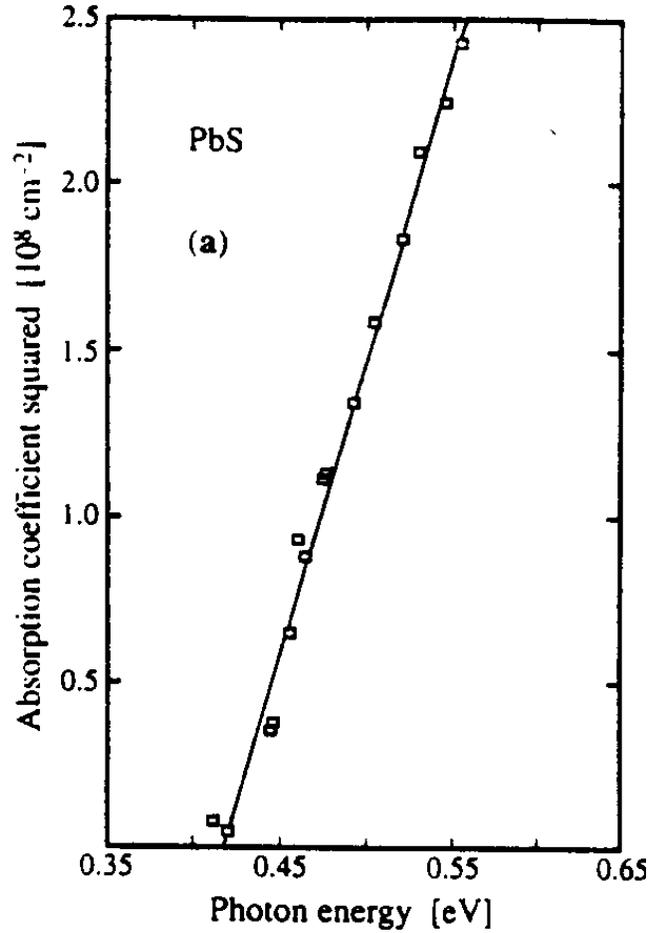
43

Figure 5.4: Plot of the square of the absorption coefficient of PbS as a function of photon energy showing the linear dependence of $[\alpha_{\mathrm{abs}}(\omega)]^2$ on $\hbar\omega$. The intercept with the x-axis defines the direct energy gap.

we obtain the following result for $\varepsilon_2(\omega)$ at $T = 0$

$$\varepsilon_2(\omega) = \left(\frac{8\pi^2 e^2 \hbar}{m\omega}\right) f_{vc} \rho_{cv}(\hbar\omega). \tag{5.38}$$

We further discuss how $\varepsilon_1(\omega)$ for interband transitions is obtained from $\varepsilon_2(\omega)$ in §6.2 using the Kramers–Kronig relation.

To illustrate the fit between these simple models and the behavior of the absorption coefficient near the fundamental absorption edge, we show in Fig. 5.4 a plot of $[\alpha_{\mathrm{abs}}]^2$ vs $\hbar\omega$ for PbS, with the intercept of $[\alpha_{\mathrm{abs}}]^2$ on the photon energy axis giving the direct energy band gap. By plotting $\alpha_{\mathrm{abs}}(\omega)$ on a log scale vs $\hbar\omega$, a value for the energy gap can also be obtained as shown in Fig. 5.5 for InSb.

The derivation of the functional form for the absorption coefficient for direct forbidden transitions proceeds as in the derivation of Eq. 5.31, except that $|\langle v|p|c\rangle|^2$ is now dependent

44

Figure 5.5: Semilogarithmic plot of the absorption coefficient of InSb at 5 K as a function of photon energy. The filled circles represent experimental results. The curves have been calculated using various models. Best results are obtained when the dependence of the matrix elements on k are included. The intercept with the x-axis gives the direct bandgap of InSb.
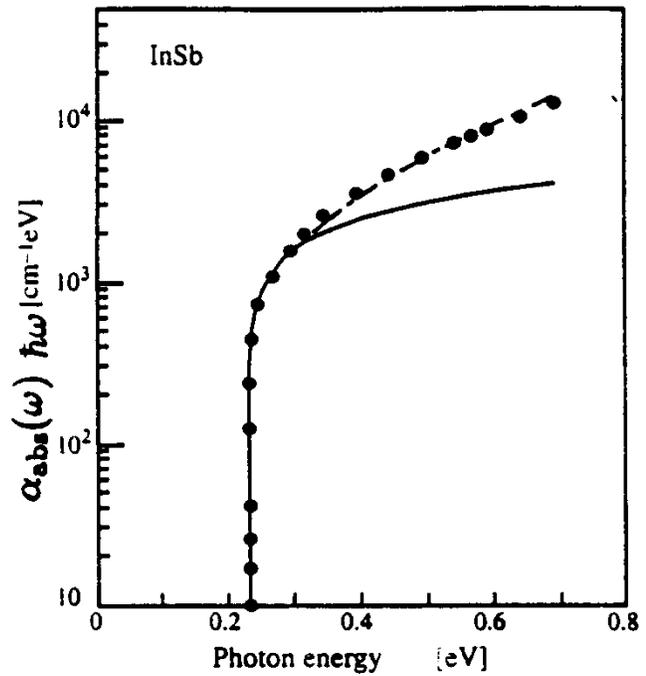
on k^2 so that $\alpha_{\mathrm{abs}}(\omega)$ follows a $(\hbar\omega - E_g)^{3/2}$ dependence for direct forbidden interband transitions.

5.4.1 Temperature Dependence of E_g

Because of the expansion and contraction of the lattice with temperature, the various band parameters, particularly the energy gap is temperature dependent. Although calculations are available to predict and account for the T dependence of the band gap at the fundamental absorption edge (threshold), $E_g(T)$ is best found by empirical fits. We give below expressions for such fits which are useful for research purposes

$$E_g(T) = 1.165 - 2.84 \times 10^{-4}T \quad \text{(eV)} \quad \text{Si}$$

$$E_g(T) = 0.742 - 3.90 \times 10^{-4}T \quad \text{(eV)} \quad \text{Ge}$$

$$E_g(T) = 1.522 - \frac{5.8 \times 10^{-4}T^2}{T+300} \quad \text{(eV)} \quad \text{GaAs}$$

$$E_g(T) = 2.338 - \frac{6.2 \times 10^{-4}T^2}{T+460} \quad \text{(eV)} \quad \text{GaP}$$

For Group IV and III–V compound semiconductors, $E_g(T)$ decreases with increasing T, as shown above, but for IV–VI compounds, $E_g(T)$ increases with increasing T.

5.4.2 Dependence of Absorption Edge on Fermi Energy

For lightly doped semiconductors, E_F lies in the bandgap and the absorption edge occurs at E_g, neglecting excitonic effects which are discussed in Chapter 7. However, for heavily doped semiconductors, E_F lies in the valence or conduction bands and the threshold for optical absorption is shifted. This shift in the absorption edge is often referred to as the Burstein shift, and is illustrated in Fig. 5.6 where it is shown that the threshold for absorption occurs when

$$\hbar\omega = E_g + \frac{\hbar^2 k_{\mathrm{BS}}^2}{2}\left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right) = E_g + \frac{\hbar^2 k_{\mathrm{BS}}^2}{2m_r^*} = E_F - 4k_B T \tag{5.39}$$

in which m_r^* is the reduced mass, $(1/m_r^*) = (1/m_e^*) + (1/m_h^*)$, k_{BS} is the wave vector corresponding to the Burstein shift defined in Eq. 5.39.

Referring to Eq. 5.27 where we introduce the probability that the initial state is occupied and the final state is unoccupied, we find that since doping affects the position of the Fermi level, the Fermi functions will depend on carrier concentration for heavily doped semiconductors. In particular the quantity $(1 - f_0)$ denoting the availability of final states will be affected by the Burstein shift. If we write

$$\frac{\hbar^2 k^2}{2m_e^*} = E - E_c \tag{5.40}$$

where E_c is the energy at the bottom of the conduction band, then the probability that the final state is empty is

$$1 - f_0 = \frac{1}{1 + \exp[(E_F - E)/k_B T]} = \frac{1}{1 + \exp\left[\frac{E_F - E}{k_B T} - \frac{(\hbar\omega - E_g)m_h^*}{(m_e^* + m_h^*)k_B T}\right]} \tag{5.41}$$

and Eq. 5.41 should be used for the probability of final states in evaluating $f(E_c)$ in Eq. 5.31.
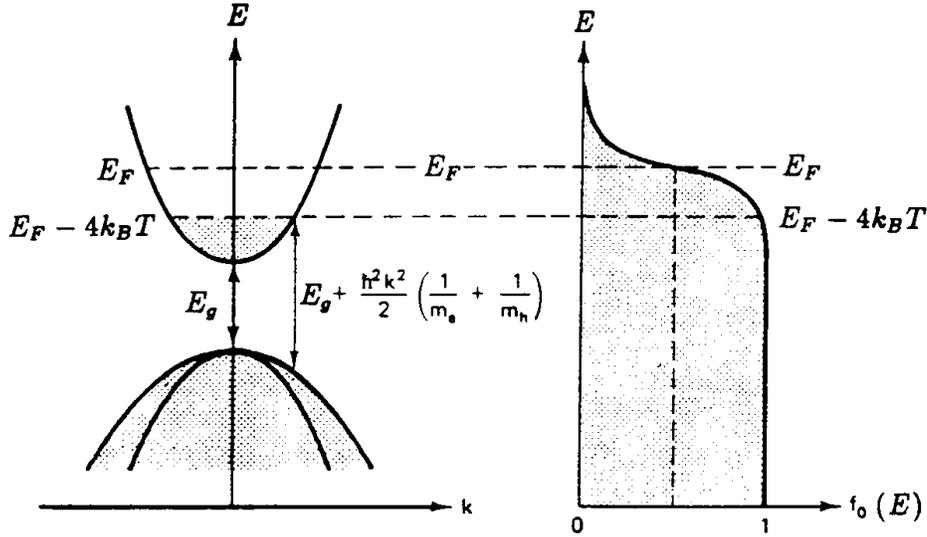
Figure 5.6: Diagram showing how the fundamental absorption edge of an n-type semiconductor is shifted to higher energy by doping. The wave vector for the Burstein shift k_{BS} is defined in Eq. 5.39.

5.4.3 Dependence of Absorption Edge on Applied Electric Field

The electron wave functions in the valence and conduction bands have an exponentially decaying amplitude in the energy gap. In the presence of an electric field $\vec{E}$, a valence band electron must tunnel through a triangular barrier to reach the conduction band. In the absence of photon absorption, the height of the barrier is E_g and its thickness is $E_g/e|\vec{E}|$ where $|\vec{E}|$ is the magnitude of the electric field, as shown in Fig. 5.7(a). The effect of the photon, as shown in Fig. 5.7(b), is to lower the barrier thickness to

$$t(\hbar\omega) = \frac{E_g - \hbar\omega}{e|\vec{E}|} \tag{5.42}$$

so that the tunneling probability is enhanced by photon absorption. Figure 5.8 shows that the absorption edge is effectively lowered by the presence of the electric field, and the effect of the electric field on α_{abs} is particularly pronounced below the zero field band gap. The effect of an electric field on the fundamental absorption edge is called the Franz–Keldysh effect.

5.5 Conservation of Crystal Momentum in Direct Optical Transitions

For clarity we now show why the momentum matrix elements coupling two Bloch states for a perfect crystal are diagonal in $\vec{k}$ and conserve crystal momentum. It is this property of the momentum matrix elements that is responsible for direct interband transitions. We
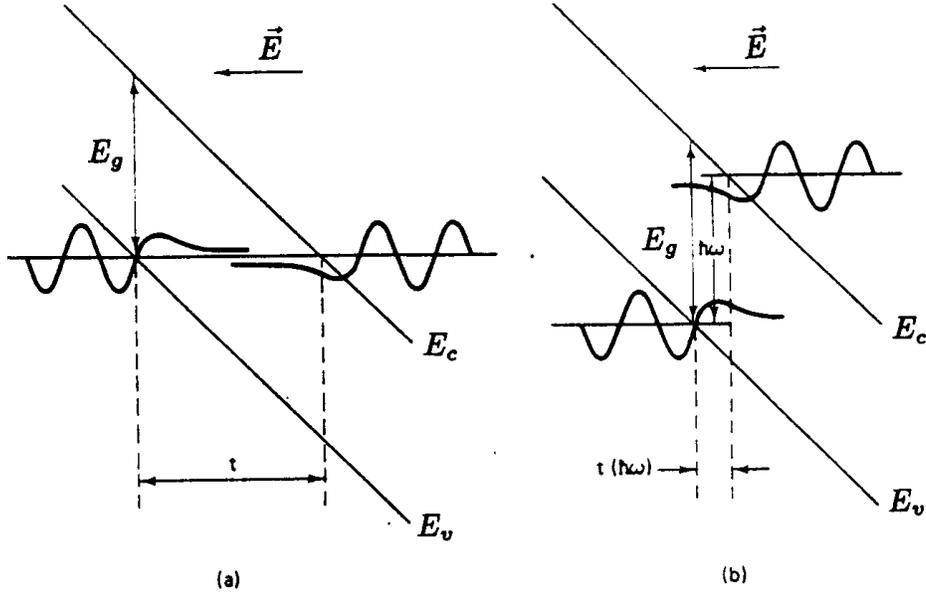
Figure 5.7: Energy band diagram in an electric field showing the wavefunction overlap (a) without and (b) with absorption of a photon of energy $\hbar\omega$.

write the momentum matrix elements coupling two bands (for example, the valence and conduction bands) as

$$\langle n'\vec{k}'|\vec{p}|n,\vec{k}\rangle = \int d^3r e^{-i\vec{k}'\cdot\vec{r}} u^*_{n'k'}(\vec{r})\left(\frac{\hbar}{i}\vec{\nabla}\right)e^{i\vec{k}\cdot\vec{r}} u_{nk}(\vec{r}). \tag{5.43}$$

Operating with $\vec{\nabla}$ on the product function of the Bloch state yields

$$\langle n'\vec{k}'|\vec{p}|n,\vec{k}\rangle = \int d^3r e^{-i\vec{k}'\cdot\vec{r}} u^*_{n'k'}(\vec{r})e^{i\vec{k}\cdot\vec{r}}(\hbar\vec{k} + \frac{\hbar}{i}\vec{\nabla})u_{nk}(\vec{r}). \tag{5.44}$$

Now the term in $\hbar\vec{k}$ can be integrated immediately to give $\hbar\vec{k}\delta_{nn'}\delta(\vec{k} - \vec{k}')$ and is thus diagonal in both band index and crystal momentum. This term therefore does not give rise to interband transitions. The remaining term in Eq. 5.44 is

$$\int d^3r e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} \frac{\hbar}{i} u^*_{n'k'}(\vec{r})\vec{\nabla}u_{nk}(\vec{r}). \tag{5.45}$$

The function $u^*_{n'k'}(\vec{r})\vec{\nabla}u_{nk}(\vec{r})$ in Eq. 5.45 is periodic under the translation $\vec{r} \to \vec{r} + \vec{R}_n$ where $\vec{R}_n$ is any lattice vector. But any spatially periodic function can be Fourier expanded

$$\sum_m F_m e^{i\vec{G}_m\cdot\vec{r}} = \frac{\hbar}{i}u^*_{n'k'}(\vec{r})\nabla u_{nk}(\vec{r}) \tag{5.46}$$

in terms of the reciprocal lattice vectors $\vec{G}_m$. We thus obtain for the integral in Eq. 5.45

$$\int d^3r e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} F_m e^{i\vec{G}_m\cdot\vec{r}} \tag{5.47}$$
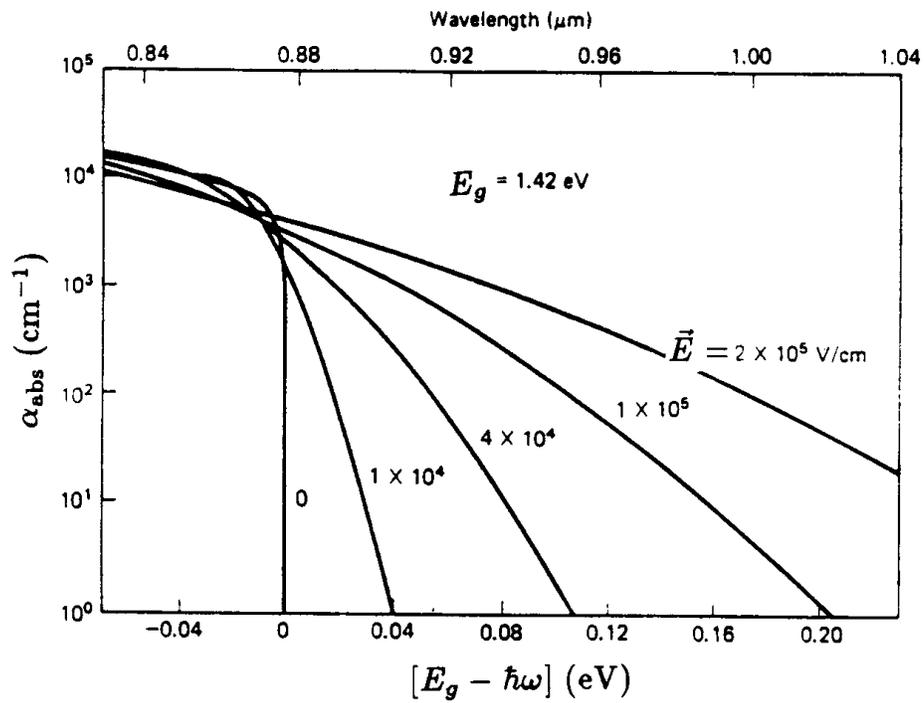
48

Figure 5.8: Electric field and photon energy dependence of the band-to-band absorption for GaAs.
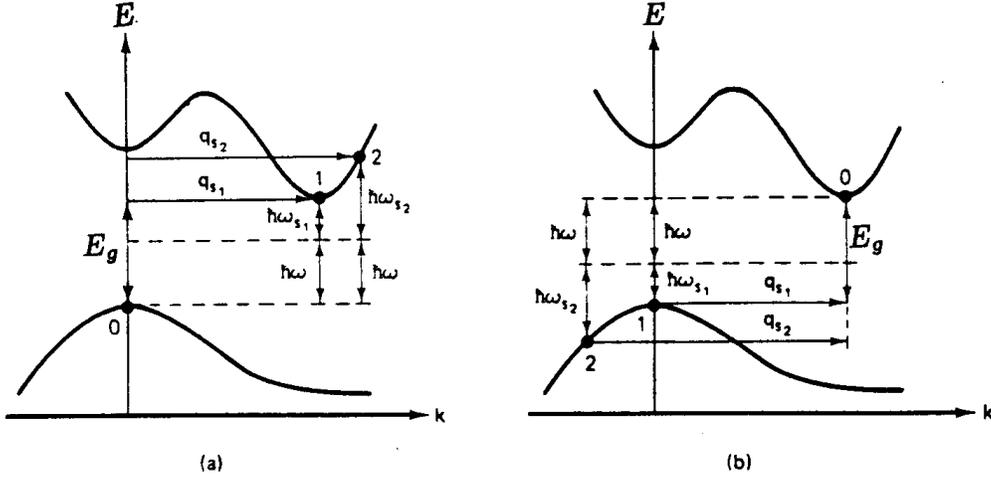
Figure 5.9: Indirect optically induced transitions of electrons (a) from the initial state 0 in the valence band to final states 1 and 2 in the conduction band and (b) from initial states 1 and 2 in the valence band to final state 0 in the conduction band. In both (a) and (b) a phonon labeled by $(\hbar\omega_s, q_s)$ is absorbed in the indirect transition process.

which vanishes unless

$$\vec{k} - \vec{k}' + \vec{G_m} = 0. \tag{5.48}$$

Since $\vec{k} - \vec{k}'$ must be within the first Brillouin zone, $\vec{k}$ and $\vec{k}'$ can only differ by the reciprocal lattice vector $\vec{G_m} \equiv 0$. Thus Eq. 5.47 vanishes unless $\vec{k} = \vec{k}'$ and we have demonstrated that because of the periodicity of the crystal lattice, the momentum matrix elements coupling two bands can only do so at the same value of crystal momentum $\vec{k}$. Since the probability for optical transitions involves the same momentum matrix elements as occur in the determination of the effective mass in the transport properties, study of the optical properties of a solid also bears an important relation to the transport properties of that material.

5.6   Indirect Interband Transitions

In making indirect transitions, the semiconductor can either emit or absorb a phonon of energy $\hbar\omega_q$

$$\hbar\omega = E_f - E_i \pm \hbar\omega_q \tag{5.49}$$

in which E_f and E_i are, respectively, the energies of the final and initial electron states and the $\pm$ signs refer to phonon emission ($+$ sign) or absorption ($-$ sign).

To review indirect interband transitions in a semiconductor, we derive below an expression for the absorption coefficient for the situation where a phonon is absorbed in the indirect process, as shown schematically in Fig. 5.9. Similar arguments can then be applied to the case where a phonon is emitted.

50

Figure 5.10: Schematic diagram of an indirect transition showing the notation used in the text. E_c is the energy of the L-point conduction band at wave vector k_c, while E_c is the thermal energy gap. The valence band maximum E_v is taken at the zero of energy. E_n and $(\vec{k}_n - \vec{k}_c)$, respectively, denote the energy and momentum of an excited electron, while E_p and $\vec{k}_p$, respectively, denote the corresponding parameters for the holes near $\vec{k} = 0$. It is customary to place the zero of energy at the valence band maximum.
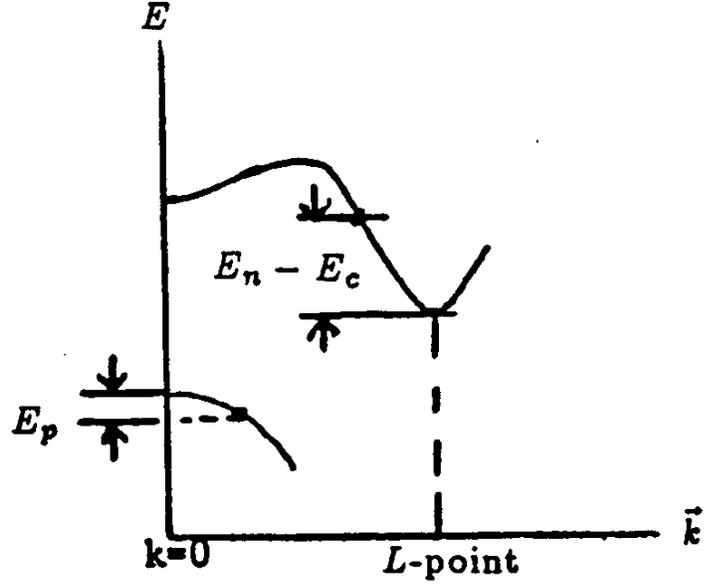
The conservation of energy principle is applied to the total process, consisting of the direct optical transition and the absorption of a phonon $\hbar\omega_q$, yielding

$$\hbar\omega = E_g - \hbar\omega_q + \frac{\hbar^2(\vec{k}_n - \vec{k}_c)^2}{2m_n} + \frac{\hbar^2 k_p^2}{2m_p} \tag{5.50}$$

in which the notation in Eq. 5.50 is defined in Fig. 5.10, and E_g is the thermal gap or energy difference between the conduction band minimum (e.g., at L) and the valence band maximum at Γ. The negative sign in front of the phonon energy $\hbar\omega_q$ in Eq. 5.50 corresponds to the phonon absorption process. In Eq. 5.50, the term $\hbar(\vec{k}_n - \vec{k}_c)$ denotes the difference between the crystal momentum $\hbar\vec{k}_n$ of an excited electron in the L–point conduction band and the crystal momentum $\hbar\vec{k}_c$ at the L–point conduction band minimum. Thus the kinetic energy of the excited electron with crystal momentum $\hbar\vec{k}_n$ is

$$E_n - E_c = \frac{\hbar^2(\vec{k}_n - \vec{k}_c)^2}{2m_n} \tag{5.51}$$

where E_n is the energy above the conduction band minimum E_c, and m_n in Eq. 5.51 is the effective mass of an electron near the conduction band minimum.

Since the valence band extremum is at $\vec{k} = 0$, then $\hbar\vec{k}_p$ is the crystal momentum for the hole that is created when the electron is excited, corresponding to the kinetic energy of the hole

$$E_p = \frac{\hbar^2 k_p^2}{2m_p}. \tag{5.52}$$

51

The sign convention that is used in this discussion is to take E_p as a positive number and the zero of energy is taken at the valence band maximum (see Fig. 5.10). In terms of these sign conventions, conservation of energy yields

$$\hbar\omega = E_g - \hbar\omega_q + (E_n - E_c) + E_p \tag{5.53}$$

and conservation of momentum requires

$$\vec{q} = \vec{k_n} - \vec{k_p} \tag{5.54}$$

where $\vec{q}$ is the wave vector for the absorbed phonon.

We now find the frequency dependence of the absorption edge for indirect transitions in order to make a distinction between direct and indirect transitions just from looking at the frequency dependence of the optical absorption data. Let us then consider the transition from some specific initial state E_p to a specific final state E_n. The density of states $\rho_c(E_n)$ (number of states/unit volume/unit energy range) for the final state conduction band has an energy dependence given by

$$\rho_c(E_n) \propto (E_n - E_c)^{1/2}. \tag{5.55}$$

Using the conservation of energy relation in Eq. 5.53, $\rho_c(E_n)$ can be expressed in terms of E_p as

$$\rho_c(E_n) \propto (\hbar\omega - E_g - E_p + \hbar\omega_q)^{1/2}. \tag{5.56}$$

Thus we see that transitions to a state E_n take place from a range of initial states, since E_p can vary between $E_p = 0$ where all of the kinetic energy is given to the electron and the opposite limit where $E_n - E_c = 0$ and all of the kinetic energy is given to the hole. Let the energy δ denote the range of possible valence band energies between these limits

$$\delta = \hbar\omega - E_g + \hbar\omega_q. \tag{5.57}$$

The density of initial states for the valence band has an energy dependence given by

$$\rho_v(E_p) \propto E_p^{1/2} \tag{5.58}$$

where we are using the convention $E_v \equiv 0$ for defining the zero of energy, so that E_p vanishes at the top of the valence band. Thus the effective density of states for the phonon absorption process is found by summing over all E_p values which conserve energy,

$$\rho(\hbar\omega) \propto \int_0^\delta \rho_c(E_n)\rho_v(E_p)dE_p \propto \int_0^\delta \sqrt{\delta - E_p}\,\sqrt{E_p}\,dE_p. \tag{5.59}$$

The integral in Eq. 5.59 can be carried out through integration by parts, utilizing the notation $u = E_p$, and $v = \delta - E_p$, and writing the limits of the integration in terms of the variable E_p

$$\int_0^\delta \sqrt{uv}\,du = \frac{\delta - 2v}{4}\sqrt{uv}\,\Big|_0^\delta + \frac{\delta^2}{4}\tan^{-1}\sqrt{\frac{u}{\delta - u}}\,\Big|_0^\delta = \frac{\delta^2\pi}{8}. \tag{5.60}$$

Substitution in Eq. 5.57 for δ in Eqs. 5.59 and 5.60 results in

$$\rho(\hbar\omega) \propto \frac{\pi}{8}(\hbar\omega - E_g + \hbar\omega_q)^2 \tag{5.61}$$

52

which gives the frequency dependence for the indirect interband transitions involving *phonon absorption*. Also, the probability for the absorption of a phonon is proportional to the Bose-Einstein factor

$$n(\hbar\omega_q) = \frac{1}{\exp(\hbar\omega_q/k_BT) - 1} \tag{5.62}$$

so that the absorption coefficient for indirect transitions in which a phonon is absorbed becomes

$$\alpha_{\text{abs}}(\omega) = \mathcal{C}_a \frac{(\hbar\omega - E_g + \hbar\omega_q)^2}{\exp(\hbar\omega_q/k_BT) - 1} \tag{5.63}$$

where $\mathcal{C}_a$ is a constant for the phonon absorption process.

To find the absorption coefficient for the indirect absorption process that involves the emission of a phonon, we must find the effective density of states for the emission process. The derivation in this case is very similar to that given above for phonon absorption, except that the energy conservation condition now involves the phonon energy with the opposite sign. Furthermore, the probability of emission of a phonon is proportional to $[n(\hbar\omega_q) + 1]$ which is given by

$$[n(\hbar\omega_q) + 1] = 1 + [e^{\hbar\omega_q/k_BT} - 1]^{-1} = \frac{1}{1 - e^{-\hbar\omega_q/k_BT}} \tag{5.64}$$

so that the absorption constant for phonon emission becomes

$$\alpha_{\text{ems}}(\omega) = \mathcal{C}_e \frac{(\hbar\omega - E_g - \hbar\omega_q)^2}{1 - \exp(-\hbar\omega_q/k_BT)} \tag{5.65}$$

where $\mathcal{C}_e$ is a constant for the phonon emission process.

At low temperatures, the phonon emission process dominates because there are so few phonons available for the absorption process. Furthermore, as a function of photon energy, different thresholds are obtained for the absorption and emission processes. In the absorption process, absorption starts when $\hbar\omega = E_g - \hbar\omega_q$ (see Fig. 5.11), while for the emission process, the optical absorption starts when $\hbar\omega = E_g + \hbar\omega_q$. So if we plot $\sqrt{\alpha_{\text{abs}}(\omega)}$ vs $\hbar\omega$, as is shown in Fig. 5.11, then $\sqrt{\alpha_{\text{abs}}(\omega)}$ in the low photon energy range will go as $\sqrt{\alpha_{\text{abs}}(\omega)} \propto (\hbar\omega - E_g + \hbar\omega_q)$ while $\sqrt{\alpha_{\text{ems}}(\omega)}$ will be proportional to $(\hbar\omega - E_g - \hbar\omega_q)$. Experimentally, a superposition of the absorption and emission processes will be observed.

Some experimental data illustrating indirect interband transitions are given in Fig. 5.12. The shift of the curves in Fig. 5.12 as a function of photon energy is due to the temperature dependence of the indirect gap in silicon. In Fig. 5.12 it is easy to separate out the lower energy absorption contribution which is associated with the phonon absorption process (compare Figs. 5.11 and 5.12). At higher energies it is also easy to separate out the phonon emission contribution. By carrying out measurements at several different temperatures it is possible to obtain a more accurate value for $\hbar\omega_q$. Figure 5.12 shows that the phonon absorption process becomes more favorable as the temperature is raised, while the emission process is less sensitive to temperature. The physical reason behind this is that for the absorption process to occur in the first place, phonons of the appropriate wave vector must be available. In Ge the phonon assisted process requires phonons of wave vector $\vec{q}$ extending from Γ to L, while for Si we need a phonon $\vec{q}$–vector from Γ to Δ_{min} (where Δ_{min} corresponds to the Δ point conduction band minimum). Since lattice vibrations are thermally excited, there are few available phonons at low temperatures, but more are

Figure 5.11: Schematic diagram showing the frequency dependence of the square root of the absorption coefficient for indirect interband transitions near the thresholds for the phonon emission and absorption processes. The curves are for four different temperatures. At the lowest temperature (T_4) the phonon emission process dominates, while at the highest temperature (T_1) the phonon absorption process is most important at low photon energies. The magnitude of twice the phonon energy is indicated.
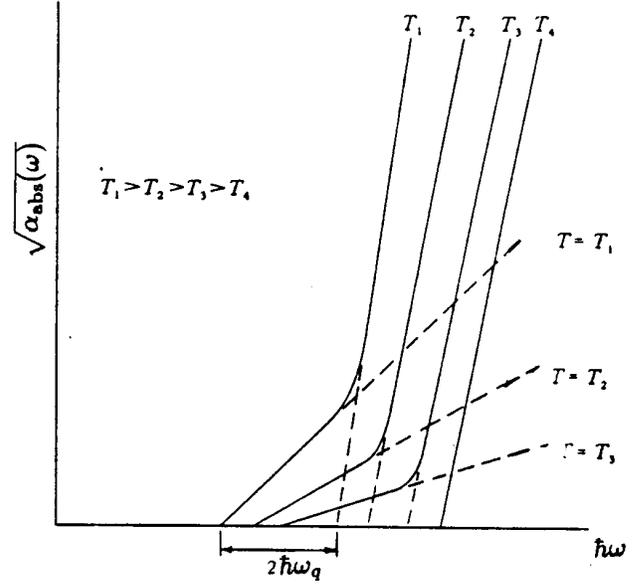
available at high temperatures. On the other hand, phonon emission does not depend upon the availability of phonons since the emission process itself generates phonons; for this reason the phonon emission process is relatively insensitive to temperature.

Since silicon is a relatively hard material (with a Debye temperature of $\theta_D = 658$ K) there will only be a few large wavevector phonons excited at room temperature. Therefore the phonon emission process will dominate in the optical absorption for photon energies where such emission is energetically possible. These arguments account for the different slopes observed for the phonon absorption and emission contributions to the absorption coefficient of Fig. 5.12.

Another complication that arises in real materials is that there are several types of phonons present for a given $\vec{q}$-vector, i.e., there are acoustic and optical branches and for each branch there are longitudinal and transverse modes. An example of the analysis of optical absorption data to obtain the frequencies of the various phonons at $q = 0$ is given in Fig. 5.13 where $\alpha_{\text{abs}}^{1/2}$ vs $\hbar\omega$ is plotted for the indirect gap semiconductor GaP, from which it is possible to measure $\hbar\omega_q$ for various LO, LA, TO and TA phonons. Today such optical data are seldom taken, because it is now customary to use inelastic neutron diffraction data to plot out the entire dispersion curve for each of the phonon branches. When the phonon frequencies are high, electron energy loss spectroscopy can be helpful in obtaining $\omega_q(q)$ for the various phonon branches as is discussed in Chapter 12.

54
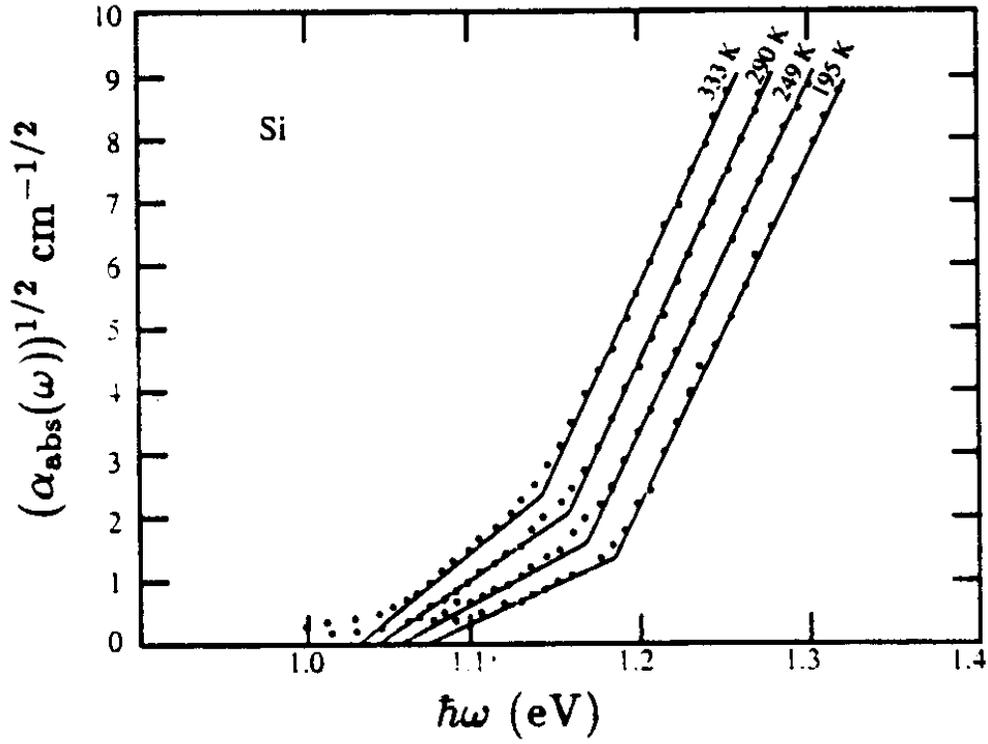
Figure 5.12: Plots of the square root of the absorption coefficients of Si versus photon energy at several temperatures. The two segments of a straight line drawn through the experimental points represent the two contributions associated with phonon absorption and emission. (From Macfarlane, et al., Phys. Rev. **111**, 1249 (1958)).

Figure 5.13: Plots of the square root of the absorption coefficients of GaP vs photon energy at two different temperatures. The labels denote the various absorption thresholds associated with the emission of various phonon modes. The observation of these phonon modes is made possible by the enhanced absorption associated with excitons at the absorption threshold (see Chapter 7). The apparent shift in the phonon frequencies is in most part due to the variation of the bandgap energy with temperature (see Figs. 5.11 and 5.12).
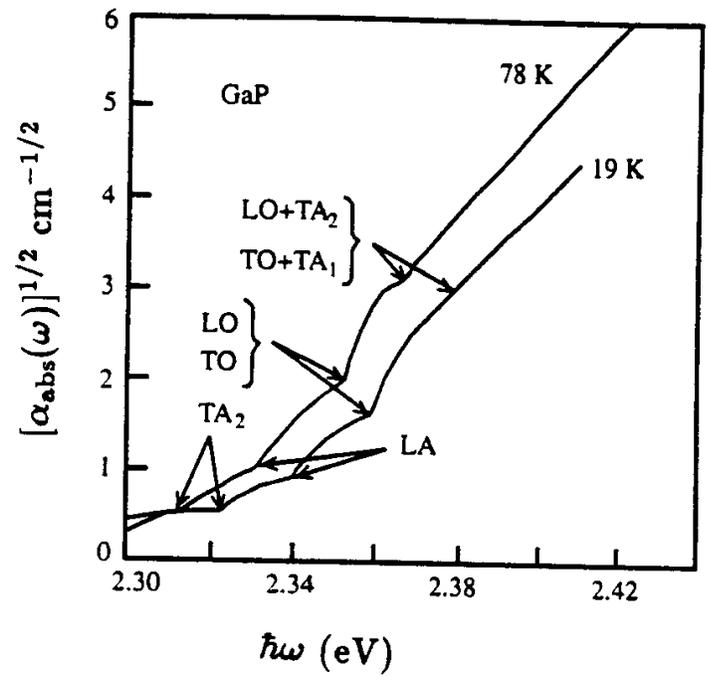
Chapter 6

Optical Properties of Solids Over a Wide Frequency Range

6.1 Kramers–Kronig Relations

References

- Jones and March, *Theoretical Solid State Physics*: pp. 787-793

- Jackson, *Classical Electrodynamics*: pp. 306-312

Measurement of the absorption coefficient gives the imaginary part of the complex index of refraction while the reflectivity is sensitive to a complicated combination of $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$. Thus from measurements such as $\alpha_{\mathrm{abs}}(\omega)$ we often have insufficient information to determine $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ independently. However, if we know either $\varepsilon_1(\omega)$ or $\varepsilon_2(\omega)$ over a wide frequency range, then $\varepsilon_2(\omega)$ or $\varepsilon_1(\omega)$ can be determined from the Kramers–Kronig relation given by

$$\varepsilon_1(\omega) - 1 = \frac{2}{\pi} \mathcal{P} \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{\omega'^2 - \omega^2} d\omega' \tag{6.1}$$

and

$$\varepsilon_2(\omega) = -\frac{2}{\pi} \mathcal{P} \int_0^\infty \frac{\omega' \varepsilon_1(\omega')}{\omega'^2 - \omega^2} d\omega' \tag{6.2}$$

in which $\mathcal{P}$ denotes the principal value. These relations are based on causality, linear response theory and the boundedness of physical observables.

The Kramers–Kronig relations relate $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ so that if either of these functions is known as a function of ω the other is completely determined. Because of the form of these relations, it is clear that the main contribution to $\varepsilon_1(\omega)$ comes from the behavior of $\varepsilon_2(\omega')$ near $\omega' \approx \omega$ due to the resonant denominator in Eqs. 6.1 and 6.2. What this means is that to obtain $\varepsilon_1(\omega)$ we really should know $\varepsilon_2(\omega')$ for all ω', but it is more important to know $\varepsilon_2(\omega')$ in the frequency range about ω than elsewhere. This property is greatly exploited in the analysis of reflectivity data, where measurements are available over a finite range of ω' values. Some kind of extrapolation procedure must be used for those frequencies ω' that are experimentally unavailable. We now give a derivation of the Kramers–Kronig relations after some introductory material.

This theorem is generally familiar to electrical engineers in another context. If a system is linear and obeys causality (i.e., there is no output before the input is applied), then the real and imaginary parts of the system function are related by a Hilbert transform. Let us now apply this causality concept to the polarization in a solid. We have the constitutive equation which defines the polarization of the solid:

$$\varepsilon \vec{E} = \vec{D} = \vec{E} + 4\pi \vec{P} \tag{6.3}$$

so that

$$\vec{P} = \frac{\varepsilon - 1}{4\pi} \vec{E} \equiv \alpha(\omega)\vec{E} \tag{6.4}$$

where $\alpha(\omega)$ defines the polarizability, and $\vec{P}$ is the polarization/unit volume or the response of the solid to an applied field $\vec{E}$. The polarizability $\alpha(\omega)$ in electrical engineering language is the system function

$$\alpha(\omega) = \alpha_r(\omega) + i\alpha_i(\omega) \tag{6.5}$$

in which we have explicitly written the real and imaginary parts $\alpha_r(\omega)$ and $\alpha_i(\omega)$, respectively. Let $E(t) = E_0\delta(t)$ be an impulse field at $t = 0$. Then from the definition of a δ-function, we have:

$$E(t) = E_0\delta(t) = \frac{E_0}{\pi} \int_{0^-}^{\infty} \cos\omega t\, d\omega. \tag{6.6}$$

The response to this impulse field yields an in-phase term proportional to $\alpha_r(\omega)$ and an out-of-phase term proportional to $\alpha_i(\omega)$, where the polarization vector is given by

$$\vec{P}(t) = \frac{E_0}{\pi} \int_{0^-}^{\infty} \left[\alpha_r(\omega) \cos\omega t + \alpha_i(\omega) \sin\omega t \right] d\omega, \tag{6.7}$$

in which $\alpha(\omega)$ is written for the complex polarizability (see Eq. 6.5). Since $\vec{P}(t)$ obeys causality and is bounded, we find that the integral of $\alpha(\omega)e^{-i\omega t}$ is well behaved along the contour C' as $R \to \infty$ and no contribution to the integral is made along the contour C' in the upper half plane (see Fig. 6.1). Furthermore, the causality condition that $\vec{P}(t)$ vanishes for $t < 0$ requires that $\alpha(\omega)$ have no poles in the upper half plane shown in Fig. 6.1.

To find an explicit expression for $\alpha(\omega)$ we must generate a pole on the real axis. Then we can isolate the behavior of $\alpha(\omega)$ at some point ω_0 by taking the principal value of the integral. We do this with the help of Cauchy's theorem. Since $\alpha(\omega)$ has no poles in the upper half-plane, the function $[\alpha(\omega)/(\omega - \omega_0)]$ will have a single pole at $\omega = \omega_0$ (see Fig. 6.2). If we run our contour just above the real axis, there are no poles in the upper-half plane and the integral around the closed contour vanishes:

$$\oint \frac{\alpha(\omega)d\omega}{\omega - \omega_0} = 0. \tag{6.8}$$

Let us now consider the integral taken over the various portions of this closed contour:

$$\int_{C'} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega + \int_{-R}^{\omega_0 - \epsilon} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega + \int_{C} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega + \int_{\omega_0 + \epsilon}^{R} \frac{\alpha(\omega)d\omega}{\omega - \omega_0} = 0. \tag{6.9}$$

The contribution over the contour C' vanishes since $\alpha(\omega)$ remains bounded, while $\frac{1}{\omega - \omega_0} \to 0$ as $R \to \infty$ (see Fig. 6.2). Along the contour C, we use Cauchy's theorem to obtain

$$\lim_{\epsilon \to 0} \int_{C} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega = -\pi i \alpha(\omega_0) \tag{6.10}$$
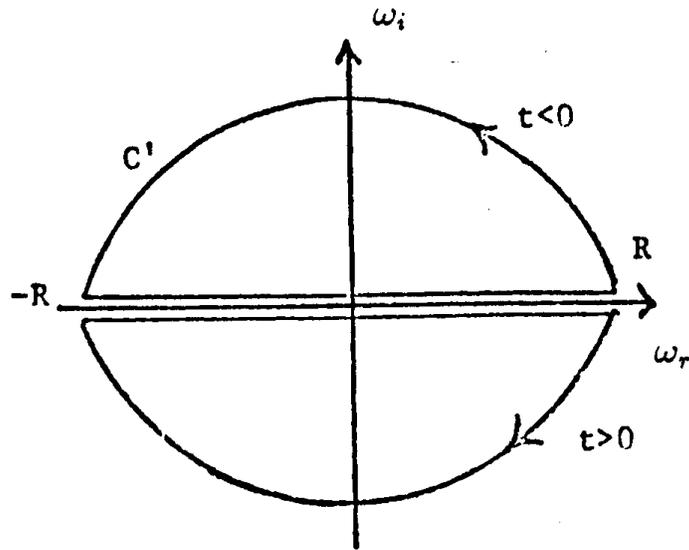
58

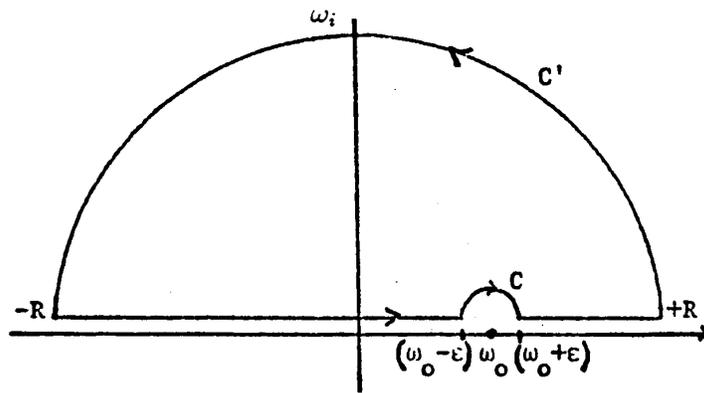Figure 6.1: Contours used in evaluating the complex polarizability integral of Eq. 6.7.

Figure 6.2: Contour used to evaluate Eq. 6.9.

59

in which $\alpha(\omega_0)$ is the residue of $\alpha(\omega)$ at $\omega = \omega_0$ and the minus sign is written because the contour C is taken clockwise. We further define the principal part $\mathcal{P}$ of the integral in the limit $R \to \infty$ and $\epsilon \to 0$ as

$$\lim_{\substack{R \to \infty \\ \epsilon \to 0}} \int_{-R}^{\omega_0 - \varepsilon} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega + \int_{\omega_0 + \epsilon}^{R} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega \to \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha(\omega)}{\omega - \omega_0} d\omega. \tag{6.11}$$

The vanishing of the integral in Eq. 6.8 thus results in the relation

$$\alpha_r(\omega_0) + i\alpha_i(\omega_0) = \frac{1}{\pi i} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha_r(\omega) + i\alpha_i(\omega)}{\omega - \omega_0} d\omega. \tag{6.12}$$

Equating real and imaginary parts of Eq. 6.12, we get the following relations which hold for $-\infty < \omega < \infty$;

$$\alpha_r(\omega_0) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha_i(\omega)}{\omega - \omega_0} d\omega \tag{6.13}$$

where $\alpha_r(\omega)$ is even and

$$\alpha_i(\omega_0) = \frac{-1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha_r(\omega)}{\omega - \omega_0} d\omega \tag{6.14}$$

where $\alpha_i(\omega)$ is odd.

We would like to write these relations in terms of integrals over positive frequencies. We can do this by utilizing the even- and oddness of $\alpha_r(\omega)$ and $\alpha_i(\omega)$. If we now multiply the integrand by $(\omega + \omega_0)/(\omega + \omega_0)$ and make use of the even- and oddness of the integrands, we get:

$$\alpha_r(\omega_0) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha_i(\omega)(\omega + \omega_0)}{\omega^2 - \omega_0^2} d\omega = \frac{2}{\pi} \mathcal{P} \int_{0}^{\infty} \frac{\omega \alpha_i(\omega) d\omega}{\omega^2 - \omega_0^2} \tag{6.15}$$

$$\alpha_i(\omega_0) = \frac{-1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\alpha_r(\omega)(\omega + \omega_0)}{\omega^2 - \omega_0^2} d\omega = -\frac{2}{\pi} \mathcal{P} \int_{0}^{\infty} \frac{\omega_0 \alpha_r(\omega) d\omega}{\omega^2 - \omega_0^2}. \tag{6.16}$$

We have now obtained the Kramers–Kronig relations. To avoid explicit use of the principal value of a function, we can subtract out the singularity at ω_0, by writing

$$\alpha_r(\omega_0) + i\alpha_i(\omega_0) = \frac{1}{\pi i} \int_{-\infty}^{\infty} \left(\frac{\alpha(\omega) - \alpha(\omega_0)}{\omega - \omega_0} \right) \left(\frac{\omega + \omega_0}{\omega + \omega_0} \right) d\omega. \tag{6.17}$$

Using the evenness and oddness of $\alpha_r(\omega)$ and $\alpha_i(\omega)$ we then obtain

$$\alpha_r(\omega_0) = \frac{2}{\pi} \int_{0}^{\infty} \frac{\omega \alpha_i(\omega) - \omega_0 \alpha_i(\omega_0)}{\omega^2 - \omega_0^2} d\omega \tag{6.18}$$

and

$$\alpha_i(\omega_0) = -\frac{2}{\pi} \int_{0}^{\infty} \frac{\omega_0 \alpha_r(\omega) - \omega_0 \alpha_r(\omega_0)}{\omega^2 - \omega_0^2} d\omega. \tag{6.19}$$

To obtain the Kramers–Kronig relations for the dielectric constant itself just substitute

$$\varepsilon(\omega) = 1 + 4\pi\alpha(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega) \tag{6.20}$$

to obtain

$$\varepsilon_1(\omega_0) - 1 = \frac{2}{\pi} \int_{0}^{\infty} \frac{\omega' \varepsilon_2(\omega') - \omega_0 \varepsilon_2(\omega_0)}{\omega'^2 - \omega_0^2} d\omega' \tag{6.21}$$

60

$$\varepsilon_2(\omega_0) = \frac{-2}{\pi} \int_0^\infty \frac{\omega_0 \varepsilon_1(\omega') - \omega_0 \varepsilon_1(\omega_0)}{\omega'^2 - \omega_0^2} d\omega'. \tag{6.22}$$

The Kramers–Kronig relations are very general and depend, as we have seen, on the assumptions of causality, linearity and boundedness. From this point of view, the real and imaginary parts of a "physical" quantity Q can be related by making the identification

$$Q_{\text{real}} \to \alpha_r \tag{6.23}$$

$$Q_{\text{imaginary}} \to \alpha_i. \tag{6.24}$$

Thus, we can identify $\varepsilon_1(\omega) - 1$ with $\alpha_r(\omega)$, and $\varepsilon_2(\omega)$ with $\alpha_i(\omega)$. The reason, of course, why the identification $\alpha_r(\omega)$ is made with $[\varepsilon_1(\omega) - 1]$ rather than with $\varepsilon_1(\omega)$ is that if $\varepsilon_2(\omega) \equiv 0$ for all ω, we want $\varepsilon_1(\omega) \equiv 1$ for all ω (the dielectric constant for free space).

Thus, if we are interested in constructing a Kramers–Kronig relation for the optical constants, then we again want to make the identification for $(\tilde{n} + i\tilde{k})$

$$[\tilde{n}(\omega) - 1] \to \alpha_r(\omega) \tag{6.25}$$

$$\tilde{k}(\omega) \to \alpha_i(\omega). \tag{6.26}$$

From Eqs. 6.21 and 6.22, we obtain the Kramers–Kronig relations for the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$

$$\tilde{n}(\omega) - 1 = \frac{2}{\pi} \int_0^\infty \frac{\omega' \tilde{k}(\omega') - \omega \tilde{k}(\omega)}{\omega'^2 - \omega^2} d\omega' \tag{6.27}$$

and

$$\tilde{k}(\omega) = -\frac{2}{\pi} \int_0^\infty \frac{\omega \tilde{n}(\omega') - \omega \tilde{n}(\omega)}{\omega'^2 - \omega^2} d\omega' \tag{6.28}$$

where we utilize the definition relating the complex dielectric function $\varepsilon(\omega)$ to the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$ where $\varepsilon(\omega) = [\tilde{n}(\omega) + i\tilde{k}(\omega)]^2$.

It is useful to relate the optical constants to the reflection coefficient $r(\omega) \exp[i\theta(\omega)]$ defined by

$$r(\omega) \exp[i\theta(\omega)] = \frac{\tilde{n}(\omega) - 1 + i\tilde{k}(\omega)}{\tilde{n}(\omega) + 1 + i\tilde{k}(\omega)} \tag{6.29}$$

and the reflectivity is given as $\mathcal{R}(\omega) = r^2(\omega)$. From Eq. 6.29, we can then write

$$\tilde{n}(\omega) = \frac{1 - r^2(\omega)}{1 + r^2(\omega) - 2r(\omega) \cos \theta(\omega)} \tag{6.30}$$

$$\tilde{k}(\omega) = \frac{2r(\omega) \sin \theta(\omega)}{1 + r^2(\omega) - 2r(\omega) \cos \theta(\omega)} \tag{6.31}$$

so that once $r(\omega)$ and $\theta(\omega)$ are found, the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$ are determined. In practice $r(\omega)$ and $\theta(\omega)$ are found from the reflectivity $\mathcal{R}$ which is measured over a wide frequency range and is modeled outside the measured range. A Kramers–Kronig relation can be written for the conjugate variables $\ln r(\omega)$ and $\theta(\omega)$, from which $\theta(\omega)$ is found:

$$\ln r(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\omega' \theta(\omega') - \omega \theta(\omega)}{\omega'^2 - \omega^2} d\omega' \tag{6.32}$$

$$\theta(\omega) = -\frac{2\omega}{\pi} \int_0^\infty \frac{\ln r(\omega') - \ln r(\omega)}{\omega'^2 - \omega^2} d\omega'. \tag{6.33}$$

where $\ln \mathcal{R}(\omega) = 2 \ln r(\omega)$.

From a knowledge of the frequency dependent reflectivity $\mathcal{R}(\omega)$, the reflection coefficient $r(\omega)$ and the phase of the reflectivity coefficient $\theta(\omega)$ can be found. We can then find the frequency dependence of the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$ which in turn yields the frequency dependent dielectric functions $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$. Starting with the experimental data for the reflectivity $\mathcal{R}(\omega)$ for germanium in Fig. 6.3(a), the Kramers–Kronig relations are used to obtain results for $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ for germanium as shown in Fig. 6.3(b).

The Kramers–Kronig relations for the conjugate variables $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$; $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$; and $\ln r(\omega)$ and $\theta(\omega)$ are widely used in quantitative studies of the optical properties of specific materials, as for example germanium in Fig. 6.3.

6.2   Optical Properties and Band Structure

If we are interested in studying the optical properties near the band edge such as the onset of indirect transitions or of the lowest direct interband transitions, then we should carry out absorption measurements (Chapter 5) to determine the absorption coefficient $\alpha_{\text{abs}}(\omega)$ and thus identify the type of process that is dominant (indirect, direct, allowed, forbidden, etc.) at the band edge. However, if we are interested in the optical properties of a semiconductor over a wide energy range, then we want to treat all bands and transitions within a few eV from the Fermi level on an equal footing. Away from the band edge, the absorption coefficients become too high for the absorption technique to be useful, and reflectivity measurements are made instead. Experimentally, it is most convenient to carry out reflectivity measurements at normal incidence. From these measurements, the Kramers–Kronig analysis (see §6.1) is used to get the phase angle $\theta(\omega)$ for some frequency ω_0, if the reflection coefficient $r(\omega)$ is known throughout the entire range of photon energies

$$\theta(\omega_0) = -\frac{2\omega_0}{\pi} \int_0^\infty \frac{\ln r(\omega) - \ln r(\omega_0)}{\omega^2 - \omega_0^2} d\omega. \tag{6.34}$$

From a knowledge of $r(\omega)$ and $\theta(\omega)$, we can then find the frequency dependence of the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$ using Eqs. 6.30 and 6.31 and the frequency dependent dielectric function

$$\varepsilon_1(\omega) = \tilde{n}^2 - \tilde{k}^2 \tag{6.35}$$

$$\varepsilon_2(\omega) = 2\tilde{n}\tilde{k}. \tag{6.36}$$

As an example of such an analysis, let us consider the case of the semiconductor germanium. The normal incidence reflectivity is given in Fig. 6.3(a) and the results of the Kramers–Kronig analysis described above are given for $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ in Fig. 6.3(b).

Corresponding to the structure in the reflectivity, there will be structure observed in the real and imaginary parts of the dielectric function. These structures in the reflectivity data are then identified with special features in the energy band structure. It is interesting to note that the indirect transition (0.66 eV) from the $\Gamma_{25'}$ valence band to the L_1 conduction band (see Part I of the notes) has almost no impact on the reflectivity data. Nor does the direct band gap, which is responsible for the fundamental absorption edge in germanium,

Figure 6.3: (a) Frequency dependence of the reflectivity of Ge over a wide frequency range. (b) Plot of the real $[\varepsilon_1(\omega)]$ and imaginary $[\varepsilon_2(\omega)]$ parts of the dielectric functions for Ge obtained by a Kramers–Kronig analysis of (a).
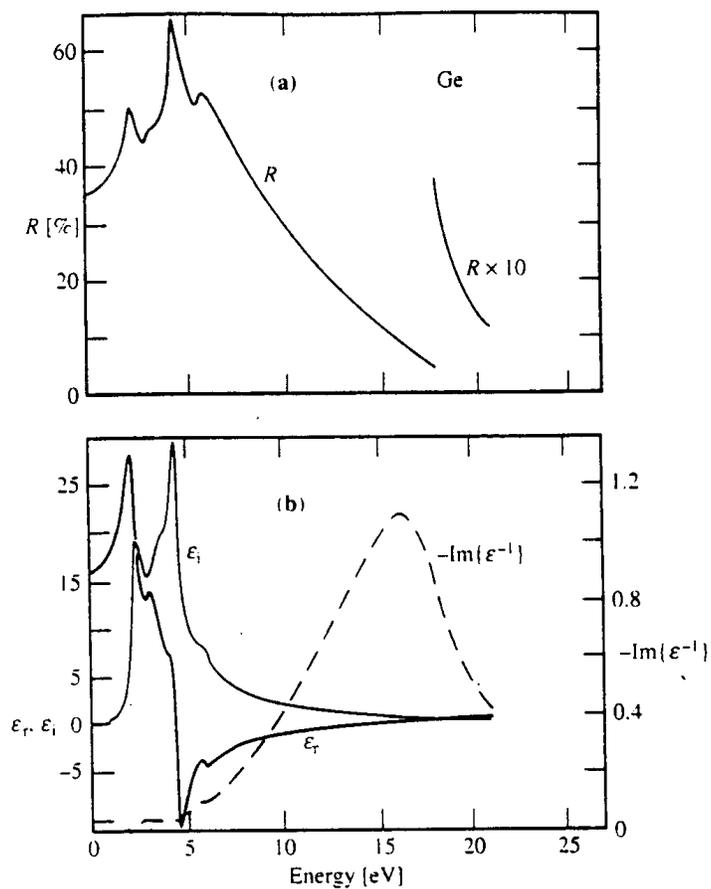
63

Table 6.1: The measured energies (eV) of the prominent structures in the optical spectra of some semiconductors with the diamond and zinc-blende structures. All energies are low temperature values except that of the E_0 transition in Si, which was measured at room temperature.

Transition	Si	Ge	GaAs	InP	GaP
E_0	4.185	0.898	1.5192	1.4236	2.869
$E_0 + \Delta_0$	4.229	1.184	1.859	1.532	2.949
E_1	3.45	2.222	3.017	3.15	3.785
$E_1 + \Delta_1$	–	2.41	3.245	3.835	–
E_0'	3.378	3.206	4.488	4.54	4.77
$E_0' + \Delta_0'$	–	3.39	4.659	–	–
E_2	4.33	4.49	5.11	5.05	5.21
E_1'	5.5	5.65	6.63	–	6.8

have a significant effect on the reflectivity data. These effects are small on the scale of the reflectivity structures shown in Fig. 6.3(a) and must be looked for with great care in a narrow frequency range where structure in the absorption data is found. The big contribution to the dielectric constant comes from interband transitions $L_{3'} \rightarrow L_1$ for which the joint density of states is large over large volumes of the Brillouin zone. The sharp rise in $\varepsilon_2(\omega)$ at 2.1 eV is associated with the $L_{3'} \rightarrow L_1$ transition. For higher photon energies, large volumes of the Brillouin zone contribute until a photon energy of about 5 eV is reached. Above this photon energy, we cannot find bands that track each other closely enough to give interband transitions with intensities of large magnitude.

6.3 Modulated Reflectivity Experiments

If we wish to study the *critical point* contribution to the optical reflectivity in more detail, it is useful to carry out modulated reflectivity measurements. If, for example, a small periodic perturbation is applied to a sample then there will be a change in reflectivity at the frequency of that perturbation. The frequency dependence of this change in reflectivity is small (parts in 10^3 or 10^4) but it is measurable. As an example, we show in Fig. 6.4, results for the reflectivity $R(\omega)$ and for the wavelength modulated reflectivity $(1/R)(dR/dE)$ of GaAs. Structure at E_0 would be identified with the direct band gap, while the structure at $E_0 + \Delta_0$ corresponds to a transition from the split-off valence band at $\vec{k} = 0$ which arises through the spin-orbit interaction. The transitions at E_1 and $E_1 + \Delta$ correspond to Λ point and L point transitions, also showing spin-orbit splitting. Also identified in Fig. 6.5 are the $E_{0'}$ transition from the Δ_7 valence band to the Δ_6 conduction band, and the E_2 transition from $X_5 \rightarrow X_5$ at the X point. Although the band structure and notation given in Fig. 6.5 applies to Ge in detail, the results for other group IV and III–V semiconductors is qualitatively similar, with values for the pertinent interband transitions given in Table 6.1 for Si, Ge, GaAs, InP and GaP.

In the vicinity of a critical point, the denominator in the joint density of states is small, so that a small change in photon energy can produce a significant change in the joint density

64

Figure 6.4: Reflectance and frequency modulated reflectance spectra for GaAs. (a) Room temperature reflectance spectrum and (b) the wavelength modulated spectrum $(1/R)(dR/dE)$ (the solid curve is experimental and the broken curve is calculated using a pseudopotential band structure model. Adapted from Yu and Cardona).
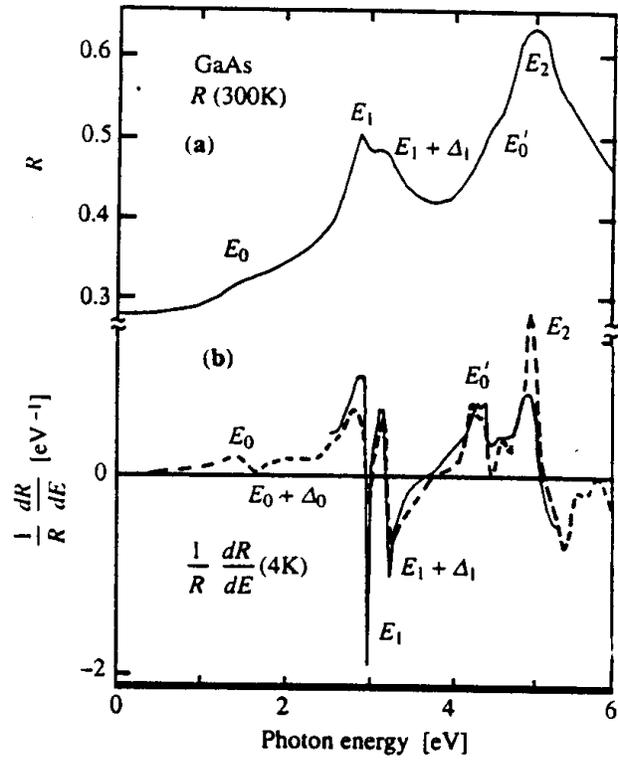
65

Figure 6.5: The band structure
of Ge including spin-orbit inter-
action and showing the various
direct transitions responsible for
the structures that are observed
in the imaginary part of the di-
electric function $\varepsilon_2(\omega)$ and in the
modulated reflectivity. Although
the band structure in this figure
is for Ge, a similar notation is
used to identify the various inter-
band transitions in other group
IV or III–V compound semicon-
ductors (see Table 6.1).
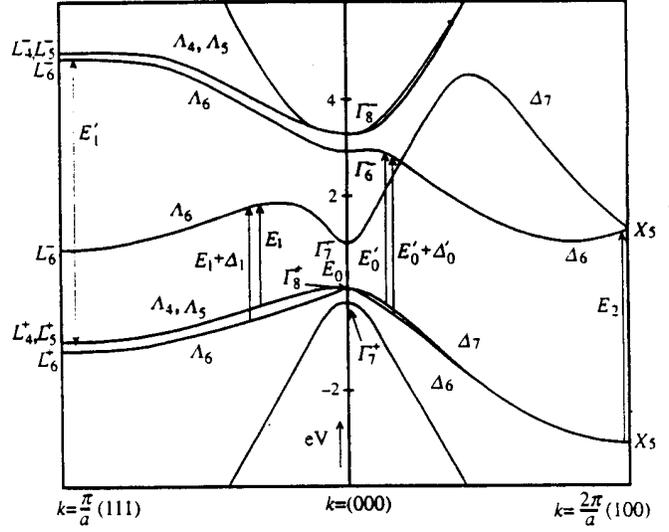
of states. Hence, modulation spectroscopy techniques emphasize critical points. There are
a number of parameters that can be varied in these modulation spectroscopy experiments:

electric field	–	electro-reflectance
wavelength	–	wavelength modulation
stress	–	piezoreflectance
light intensity	–	photo-reflectance
temperature	–	thermo-reflectance.

The various modulated reflectivity experiments are complementary rather than yielding
identical information. For example, certain structures in the reflectance respond more
sensitively to one type of modulation than to another. If we wish to look at structure
associated with the L point (111 direction) transitions, then a stress along the (100) direction
will not produce as important a symmetry change as application of stress along a (111)
direction; with a stress along a (111) direction, the ellipsoid having its longitudinal axis
along (111) will be affected one way while the other three ellipsoids will be affected in
another way. However, stress along the (100) direction treats all ellipsoids in the same way.

The reason why modulation spectroscopy emphasizes critical points can be seen by the
following argument. For a direct interband transition, the optical absorption coefficient has
a frequency dependence

$$\alpha_{abs}(\omega) = C \frac{\sqrt{(\hbar\omega - E_g)}}{\hbar\omega}. \tag{6.37}$$

Therefore, a plot of $\alpha_{abs}(\omega)$ vs. $\hbar\omega$ exhibits a threshold [Fig. 6.6(a)], but no singularity in
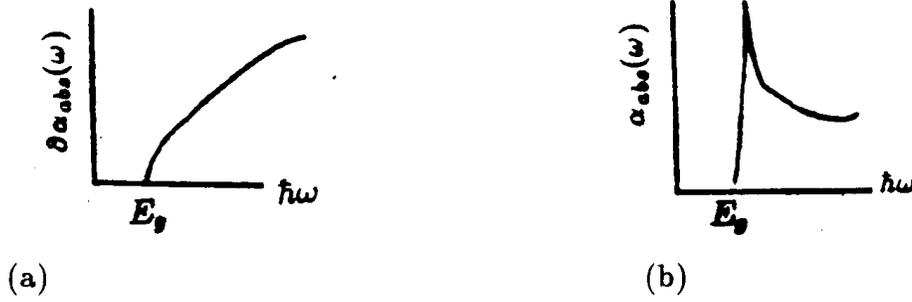
66

Figure 6.6: (a) The frequency dependence of the optical absorption coefficient showing a threshold for interband transitions at the band gap. (b) The derivative of (a) which is measured in the modulated reflectivity shows a sharp singularity associated with the threshold energy.

the frequency plot. However, when we take the derivative of Eq. 6.37

$$\frac{\partial \alpha_{abs}(\omega)}{\partial \omega} = \frac{C}{2\omega}(\hbar\omega - E_g)^{-1/2} - \frac{C}{\hbar\omega^2}(\hbar\omega - E_g)^{1/2} \tag{6.38}$$

a sharp structure is obtained in the modulated reflectivity due to the singularity in the first term of Eq. 6.37 at $\hbar\omega = E_g$ [see Fig. 6.6(b)]. If we modulate the light with any arbitrary parameter x, then

$$\frac{\partial \alpha_{abs}}{\partial x} = \frac{\partial \alpha_{abs}}{\partial \omega}\frac{\partial \omega}{\partial x}, \tag{6.39}$$

and structure in the reflectivity is expected as x is varied.

Thus all modulation parameters can be expected to produce singularities in the optical absorption. For some variables such as stress, the modulated signal is sensitive to both the magnitude and the *direction* of the stress relative to the crystal axes. For thermomodulation, the spectrum is sensitive to the magnitude of the thermal pulses, but the response is independent of crystalline direction. Thermomodulation is, however, especially sensitive to transitions from and to the Fermi level.

Thus, the various modulation techniques can be used in optical studies to obtain additional information about symmetry, which can then be used for more reliable identification of structure in the optical properties. The modulation technique specifically emphasizes interband transitions associated with particular points in the Brillouin zone. The identification of where in the Brillouin zone a particular transition is occurring is one of the most important and difficult problems in optical studies of solids. It is often not the case that we have reliable band models available to us when we start to do optical studies. For this reason, symmetry is a very powerful tool for the study of optical properties.

The high sensitivity of modulation spectroscopy provides valuable information about the band structure that would be difficult to obtain otherwise, and some examples are cited below. One example of the use of modulation spectroscopy is to determine the temperature dependence of the bandgap of a semiconductor, as shown in Fig. 6.7 for the direct Γ point gap in Ge. This measurement takes advantage of the high resolution of modulation

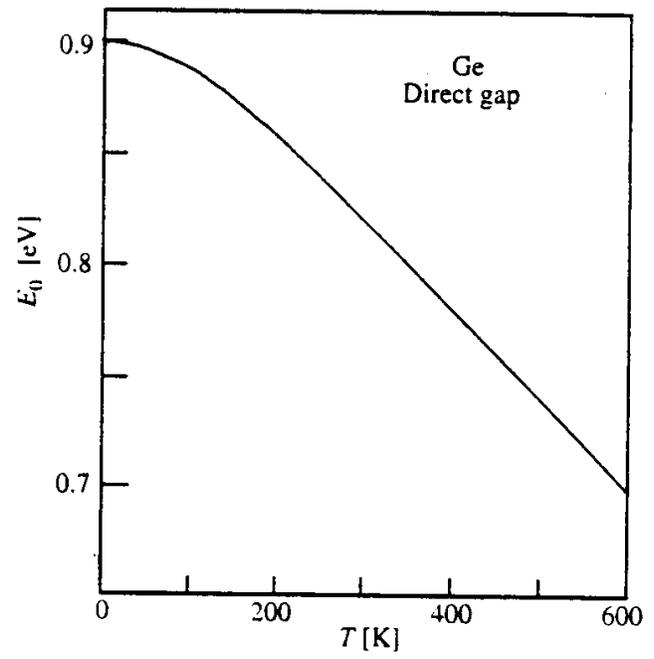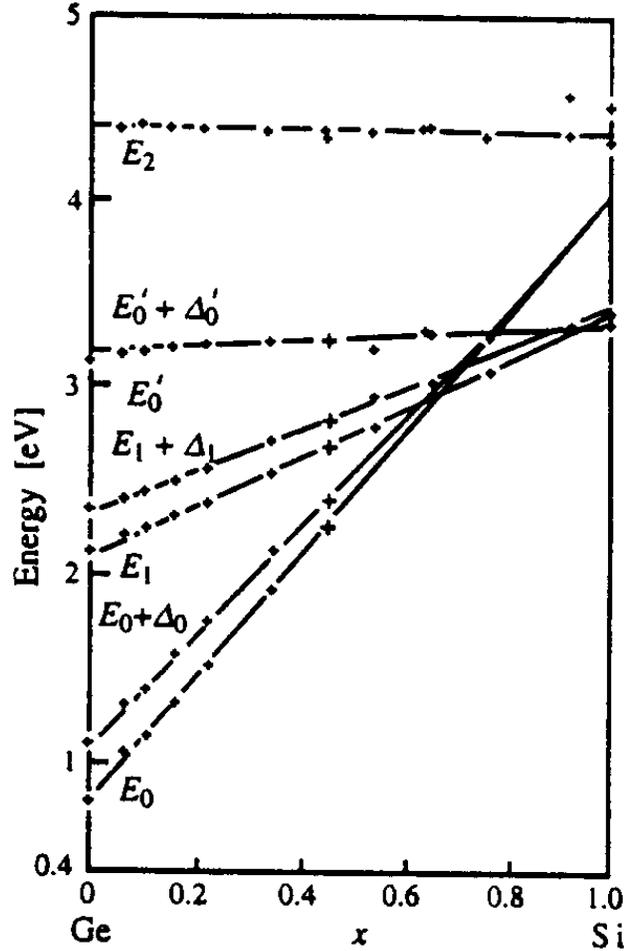Figure 6.7: Temperature dependence of the direct gap (E_0) of Ge.

Figure 6.8: Dependence of the energies of the E_0, $E_0 + \Delta_0$, E_1, $E_1 + \Delta_1$, E_0', and E_2 electro-reflectance peaks on x in the $Ge_{1-x}Si_x$ alloy system at room temperature.

spectroscopy and is especially useful for measurements at elevated temperatures. Another example is the dependence of the various band separations identified in Fig. 6.5 as a function of alloy concentration x in $Ge_{1-x}Si_x$ alloys (Fig. 6.8). Here again the high resolution of the modulation spectroscopy is utilized. A third example is the isotope dependence of the direct absorption edge of Ge as shown in Fig. 6.9. Modulation spectroscopy has also been applied to studying interband transitions in metals. For example, Fig. 6.10 shows modulated spectroscopy results from a gold surface taken with both the thermal modulation and piezoreflectance techniques. The results show that transitions involving states at the Fermi level (either initial or final states) are more sensitively seen using thermal modulation because small temperature variations affect the Fermi tail of the distribution function strongly. Thus, thermo-reflectance measurements on the noble metals give a great deal of well–resolved structure, compared with electro-reflectance and piezoreflectance measurements as illustrated in Fig. 6.10. In this figure, we see that in gold the piezoreflectance

Figure 6.9: Photo-modulated reflectivity of Ge showing the E_0 direct gap at $k = 0$ of single crystals of nearly isotopically pure ^{70}Ge, ^{74}Ge, and ^{76}Ge, at $T =$ 6 K. Note the remarkable dependence of E_0 on isotopic composition.
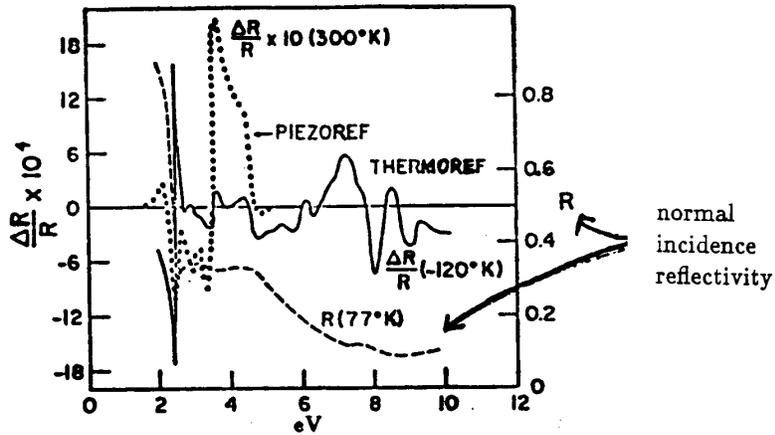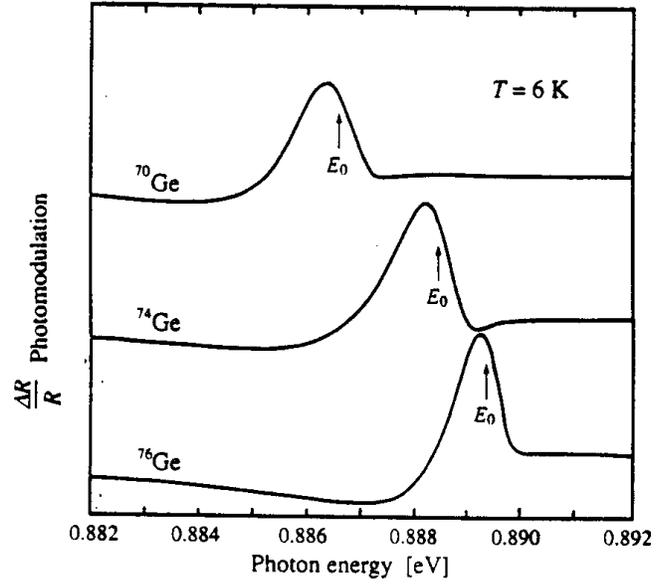

Figure 6.10: Thermo-reflectance and normal incidence reflectivity spectra of gold near liquid nitrogen temperature (from W.J. Scouler, Phys. Rev. Letters 18, 445 (1967)) together with the room temperature piezoreflectance spectrum (M. Garfunkel, J.J. Tiemann, and W.E. Engeler, Phys. Rev. 148, 698 (1966)).

is much more sensitive than ordinary reflectivity measurements near 4 eV, but the thermoreflectance technique is most powerful for transitions made to states near the Fermi level.

6.4   Ellipsometry and Measurement of Optical Constants

Ellipsometry is a standard method for measuring the complex dielectric function or the complex optical constants $\tilde{N} = \tilde{n} + i\tilde{k}$ of a material. Since two quantities are measured in an ellipsometry measurement, $\tilde{n}$ and $\tilde{k}$ can both be determined at a single frequency. The ellipsometry measurements are usually made over a range of frequencies, especially for frequencies well above the fundamental absorption edge where semiconductors become highly absorbing. At these higher frequencies very thin samples would be needed if the method of interference fringes were used to determine $\tilde{n}$, which is a very simple method for measuring the wavelength in a non-absorbing medium. One drawback of the ellipsometry technique is the high sensitivity of the technique to the quality and cleanliness of the surface. Ellipsometry is limited by precision considerations to measurements on samples with absorption coefficients $\alpha_{abs} > 1 - 10$ cm^{-1}. Ellipsometers can be made to operate in the near infrared, visible and near ultraviolet frequency regimes, and data acquisition can be made fast enough to do real time monitoring of $\varepsilon(\omega)$.

In the ellipsometry method the reflected light with polarizations "p" (parallel) and "s" (perpendicular) to the plane of incidence [see Fig. 6.11(a)] is measured as a function of the angle of incidence ϕ and the frequency of the light. The corresponding reflectances $R_s = |r_s|^2$ and $R_p = |r_p|^2$ are related to the complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega) = (\tilde{n} + i\tilde{k})^2$ by the Fresnel equations which can be derived from the boundary conditions on the fields at the interface between two surfaces with complex dielectric functions ε_a and ε_s as shown in Fig. 6.11(a). From the figure the complex reflection coefficients for polarizations s and p are

$$r_s = \frac{E_{sr}}{E_{si}} = \frac{\tilde{N}_a \cos\phi - \tilde{N}_s \cos\phi_t}{\tilde{N}_a \cos\phi + \tilde{N}_s \cos\phi_t} \tag{6.40}$$

and

$$r_p = \frac{E_{pr}}{E_{pi}} = \frac{\varepsilon_s \tilde{N}_a \cos\phi - \varepsilon_a \tilde{N}_s \cos\phi_t}{\varepsilon_s \tilde{N}_a \cos\phi + \varepsilon_a \tilde{N}_s \cos\phi_t} \tag{6.41}$$

in which

$$\tilde{N}_s \cos\phi_t = (\varepsilon_s - \varepsilon_a \sin^2\phi)^{1/2} \tag{6.42}$$

and r_s and r_p are the respective reflection coefficients, ε_s and $\tilde{N}_s$ denote the complex dielectric function and complex index of refraction within the medium, while ε_a and $\tilde{N}_a$ are the corresponding quantities outside the medium (which is usually vacuum or air). When linearly polarized light, that is neither s- nor p-polarized, is incident on a medium at an oblique angle of incidence, the reflected light will be elliptically polarized. The ratio (σ_r) of the complex reflectivity coefficients $r_p/r_s \equiv \sigma_r$ is then a complex variable which is measured experimentally in terms of its phase (or the phase shift relative to the linearly polarized incident light) and its magnitude, which is the ratio of the axes of the polarization ellipse of the reflected light [see Fig. 6.11(a)]. These are the two measurements that are made in ellipsometry. The complex dielectric function of the medium $\varepsilon_s(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ can
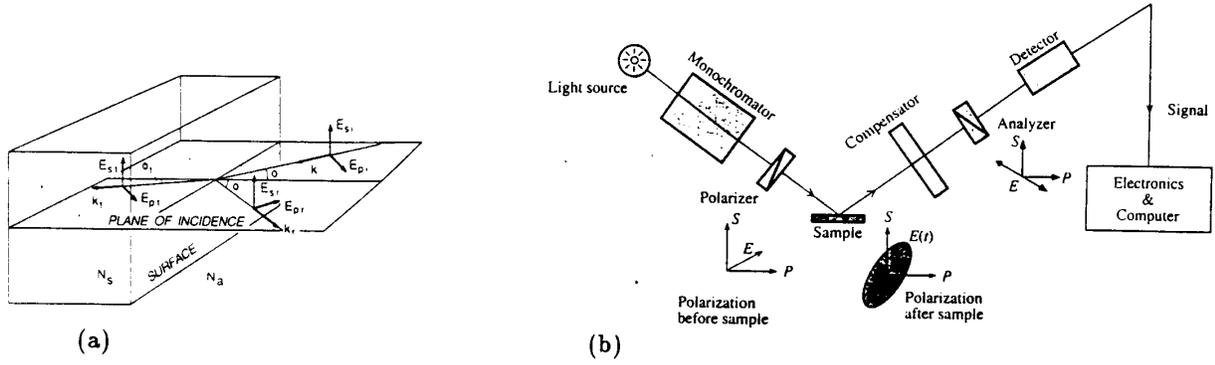
Figure 6.11: (a) Electric field vectors resolved into p and s components, for light incident (i), reflected (r), and transmitted (t) at an interface between media of complex indices of refraction $\tilde{N}_a$ and $\tilde{N}_s$. The propagation vectors are labeled by $\vec{k}_i$, $\vec{k}_r$, and $\vec{k}_t$. (b) Schematic diagram of an ellipsometer, where P and S denote polarizations parallel and perpendicular to the plane of incidence, respectively.

then be determined from the angle ϕ, the complex σ_r, and the dielectric function ε_a of the ambient environment using the relation

$$\varepsilon_s = \varepsilon_a \sin^2 \phi + \varepsilon_a \sin^2 \phi \tan^2 \phi \left(\frac{1 - \sigma_r}{1 + \sigma_r} \right)^2, \tag{6.43}$$

and in a vacuum environment $\varepsilon_a = 1$.

The experimental set-up for ellipsometry measurements is shown in Fig. 6.11(b). Light from a tunable light source is passed through a monochromator to select a frequency ω and the light is then polarized linearly along direction $\vec{E}$ to yield the I_s and I_p incident light intensities. After reflection, the light is elliptically polarized along $\vec{E}(t)$ as a result of the phase shift E_{pr} and E_{sr} have each experienced. The compensator introduces a phase shift $-\theta$ which cancels the $+\theta$ phase shift induced by the reflection at the sample surface, so that the light becomes linearly polarized again as it enters the analyzer. If the light is polarized at an angle of $\pi/2$ with respect to the analyzer setting, then no light reaches the detector. Thus at every angle of incidence and every frequency, $\varepsilon(\omega, \phi)$ is determined by Eq. 6.43 from measurement of the magnitude and phase of σ_r.

Another common method to determine the optical constants is by measurement of the normal incidence reflectivity over a wide frequency range and using the Kramers–Kronig analysis as discussed in §6.2 to determine the optical constants $\tilde{n}(\omega)$ and $\tilde{k}(\omega)$.

72

Chapter 7

Impurities and Excitons

7.1 Impurity Level Spectroscopy

Selected impurities are frequently introduced into semiconductors to make them n–type or p–type. The introduction of impurities into a crystal lattice not only shifts the Fermi level, but also results in a perturbation to the periodic potential, giving rise to bound impurity levels which often occur in the band gap of the semiconductor.

Impurities and defects in semiconductors can be classified according to whether they result in a minor or major perturbation to the periodic potential. Any disturbance to the periodic potential results in energy levels differing from the energy levels of the perfect crystal. However, when these levels occur within the energy band gap of a semiconductor or of an insulator, they are most readily identified, and these are the levels which give rise to well–defined optical spectra. Impurity levels are classified into two categories:

1. shallow levels,

2. deep levels,

corresponding, respectively, to a minor or a major perturbation of the periodic potential. Impurities are also classified according to whether they give rise to electron carriers (donors) or hole carriers (acceptors). We will now discuss the optical spectra for impurities.

7.2 Shallow Impurity Levels

An example of a shallow impurity level in a semiconductor is a hydrogenic donor level in a semiconductor like Si, Ge or the III-V compounds. Let us briefly review the origin of shallow donor levels in n–type semiconductors, where conduction is predominantly by electron carriers.

Suppose we add donor impurities such as arsenic, which has 5 valence electrons, to germanium which has 4 valence electrons (see Part I, Fig. 4.1). Each germanium atom in the perfect crystal makes 4 bonds to its tetrahedrally placed neighbors. For the arsenic impurity in the germanium lattice, four of the valence electrons will participate in the tetrahedral bonding to the germanium neighbors, but the fifth electron will be attracted back to the arsenic impurity site because the arsenic ion on the site has a positive charge.
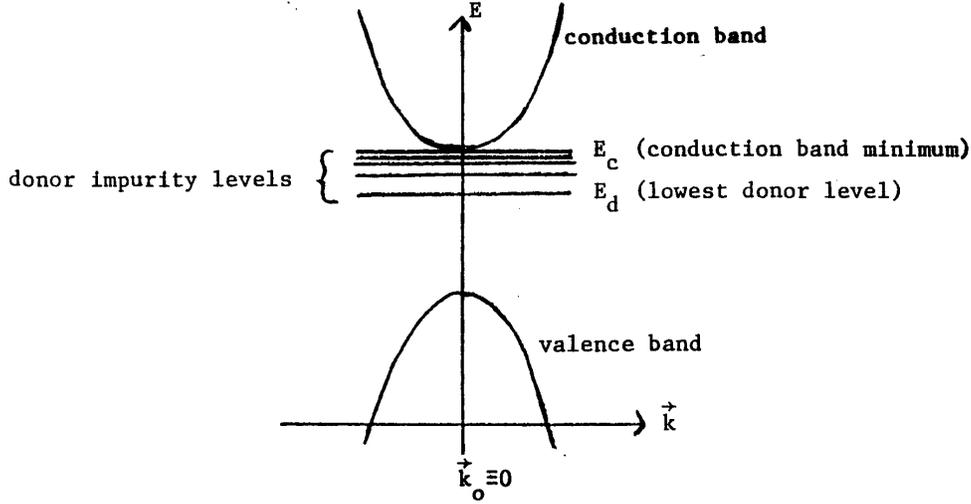
Figure 7.1: Hydrogenic impurity levels in a semiconductor.

This interaction is described by the Coulomb perturbation Hamiltonian,

$$\mathcal{H}'(r) = -\frac{e^2}{\varepsilon_0 r} \tag{7.1}$$

where ε_0 is the static dielectric constant which is 16 for germanium and 12 for silicon. This Coulomb interaction is a screened Coulomb potential, screened by the static dielectric constant. The approximation of taking ε_0 to be independent of distance is however not valid for values of r comparable to lattice dimensions, as discussed below.

In simple terms, $\mathcal{H}'$ given by Eq. 7.1 is the same as in the hydrogen atom except that the charge is now $e/\sqrt{\varepsilon_0}$ and the mass which enters the kinetic energy is the effective mass m^*. Since the levels in the hydrogen atom are given by the Bohr energy levels E_n^{hydrogen}

$$E_n^{\text{hydrogen}} = -\frac{m_0 e^4}{2\hbar^2 n^2} \tag{7.2}$$

then the energy levels in the hydrogenic impurity problem are to a first approximation given by hydrogenic levels E_n^{impurity}

$$E_n^{\text{impurity}} = -\frac{m^* e^4}{2\hbar^2 \varepsilon_0^2 n^2}. \tag{7.3}$$

The impurity levels are shown in Fig. 7.1 where the donor levels are seen to lie in the gap below the conduction band minimum.

For the hydrogen atom $E_1^{\text{hydrogen}} = -13.6\text{eV}$, but for germanium $E_1^{\text{impurity}} \sim 6 \times 10^{-3}$ eV where we have used a value of $m^* = 0.12 m_0$ representing an average of the effective mass over the entire conduction band pocket. From measurements such as the optical absorption we find that the thermal energy gap (which is the energy difference between the

L point lowest conduction band and the Γ point highest valence band) is 0.66 eV at room temperature. But the donor level manifold is only 6×10^{-3} eV wide (ranging from the E_1 level to the ionization limit) so that these impurity levels are very close to the bottom of the conduction band.

Another quantity of interest in this connection is the "orbital radius" of the impurity. Unless the orbital radius is greater than a few lattice dimensions, it is not meaningful to use a dielectric constant independent of $\vec{r}$ in constructing the perturbation Hamiltonian, since the dielectric constant used there is conceptually meaningful only for a continuum. Therefore, it is of interest to calculate the hydrogen Bohr radius using the usual recipe for the hydrogen atom

$$r_n^{\text{hydrogen}} = \frac{n^2 \hbar^2}{m_0 e^2} \tag{7.4}$$

where $\hbar = 1.054 \times 10^{-27}$ erg/sec, the mass of the free electron is $m_0 = 9.1 \times 10^{-28}$ g, and the charge on the electron is $e = 4.8 \times 10^{-10}$ esu. The value for the Bohr radius in the hydrogen atom is $r_1^{\text{hydrogen}} = 0.5$Å and for the screened hydrogenic states in the impurity problem, we have

$$r_n^{\text{impurity}} = \frac{n^2 \hbar^2 \varepsilon_0}{m^* e^2} \tag{7.5}$$

which is larger than the hydrogen Bohr radius by a factor $\varepsilon_0 m_0 / m^*$. Using typical numbers for germanium we get the ground state radius $r_1^{\text{impurity}} \sim 70$Å. Thus, the electron travels over many lattice sites in germanium and the dielectric constant approximation used in Eq. 7.1 is valid.

From this discussion we see that only a very small energy is needed to ionize a bound donor electron into the conduction band, and because this binding energy is small, these hydrogenic donor levels are called shallow impurity levels. Since $r_n^{\text{impurity}} \gg a$ where a is the lattice constant, these electrons are well localized in momentum space according to the uncertainty principle. Shallow donor levels are associated with the k–point where the conduction band minima occur.

Thus the simple hydrogenic view of impurity levels in a semiconductor predicts that the impurity spectrum should only depend on the host material and on the charge difference between the host and impurity. To see how well this model works, let us look at the experimental results summarized in Fig. 7.2. This picture is for silicon where the Bohr radius is ≈ 20Å. Here the agreement with the hydrogenic type model is good except for the ground state, where the dielectric constant approximation is not as valid as for germanium. The actual calculation referred to in this picture is solved for the case where the effective mass tensor components are included in the calculation. Such a calculation cannot be done exactly for an ellipsoidal constant energy surface

$$\left[-\frac{\hbar^2}{2} \left(\frac{\partial^2/\partial x^2}{m_l} + \frac{(\partial^2/\partial y^2 + \partial^2/\partial z^2)}{m_t} \right) - \frac{e^2}{\varepsilon r} \right] \psi(\vec{r}) = E\psi(\vec{r}) \tag{7.6}$$

for which $m_l \neq m_t$ but can be done exactly for the hydrogen atom for which $m_l = m_t$. In practice Eq. 7.6 is solved using a variational principle. To take into account that for small r, we have $\varepsilon(r) \to 1$ and for large r, we have $\varepsilon(r) \to \varepsilon_0$ where ε_0 is the static dielectric constant, a spatial dependence for $\varepsilon(r)$ thus needs to be assumed and this spatial dependence can be incorporated into the variational calculation. The inclusion of screening effects by the

75

Figure 7.2: Energy levels of donor states in silicon, experiment and theory. In recent work very good agreement is achieved between theory and experiment in transitions between shallow impurity states at low temperature (4 K) and low carrier concentrations ($\sim 10^{14}/\text{cm}^3$).
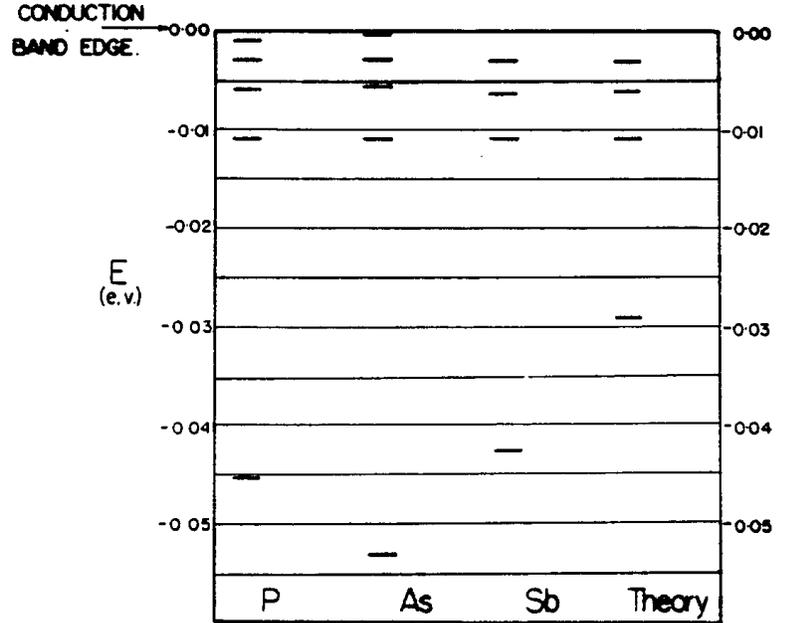
introduction of a spatial dependence to the dielectric function $\varepsilon(r)$, is called the "central cell correction" (see Part I §4.3).

The impurity spectra are studied most directly by infrared absorption and transmission measurements. As an example of such spectra we see in Fig. 7.3 the absorption spectrum from phosphorus impurities in Si. Note that the photon energies used in these measurements are small so that far infrared frequencies must be employed. The ground state donor level is a $1s$ state and allowed transitions are made to a variety of p–states. Since the constant energy surface is ellipsoidal, the $2p$ levels break up into a $2p(m_l = 0)$ level and a $2p(m_l = \pm 1)$ level which is doubly degenerate (see Fig. 7.3). Transitions from the $1s$ to both kinds of p levels occur, and account for the sharp features in the spectrum shown. The sensitivity of the spectra is somewhat improved using modulated spectroscopy techniques as shown in Fig. 7.4, where transitions to higher quantum states ($n = 6$) can be resolved and to higher angular momentum states (f levels where $\ell = 3$), noting that electric dipole transitions always occur between states of opposite parity. For both Figs. 7.3 and 7.4, the initial state is the $1s$ impurity ground state. Analysis of such spectra gives the location of the donor impurity levels including the location of the ground state donor level, which is more difficult to calculate because of the central cell correction.

In absorption measurements the impurity level transitions are observed as peaks. On the other hand, impurity spectra can also be taken using transmission techniques, where the impurity level transitions appear as minima in the transmission spectra.
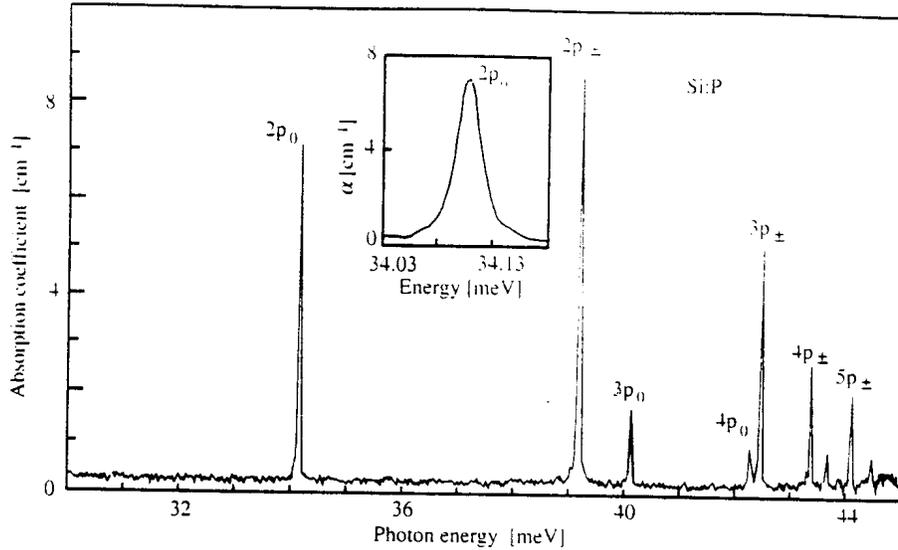
76

Figure 7.3: Absorption spectrum of phosphorus donors in Si for a sample at liquid helium temperature containing $\sim 1.2 \times 10^{14}$ cm^{-3} phosphorus. The inset shows the $2p_0$ line on an expanded horizontal scale.

7.3 Departures from the Hydrogenic Model

While the simple hydrogenic model works well for the donor states in silicon and germanium, it does not work so well for the degenerate valence bands. In this case the spectra are sensitive to the impurity species from column III in the periodic table, because the valence band masses are heavier and the effective Bohr radius is therefore more comparable to the lattice constant. Calculations for the acceptor impurity levels are now sufficiently accurate so that good agreement between theory and experiment is obtained in recent work.

It would be naive to assume that the simple hydrogenic model works for all kinds of impurity centers. If the effective Bohr radius is comparable with atomic separations, then clearly the Coulomb potential of the impurity center is not a small perturbation to the periodic potential seen by an electron. Specific cases where the impurity radius becomes small are materials with either (1) a large m^* or (2) a small ε_0 which imply a small interband coupling. When these conditions are put into Eqs. 7.3 and 7.5, we see that a small Bohr radius corresponds to a large E_n value. Thus "deep" impurity levels are not well described by simple effective mass theory. In order to make any progress at all with deep impurity level problems, we must consider the energy band structure throughout the Brillouin zone. When an electron is localized in real space, a suitable description in momentum space must include a large range of $\vec{k}$ values.

When the impurity concentration becomes so large that the Bohr orbits for neighboring impurity sites start to overlap, the impurity levels start to broaden, and eventually impurity bands are formed. These impurity bands tend to be only half filled because of the Coulomb repulsion which inhibits placement of both a spin up and a spin down carrier in the same
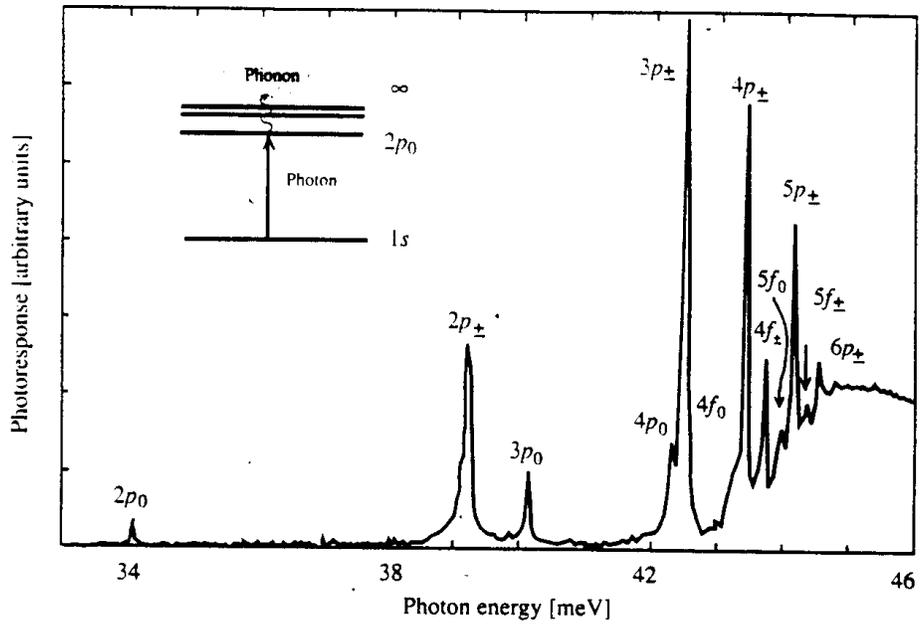
Figure 7.4: Photo-thermal ionization spectrum of phosphorus-doped Si measured by modulation spectroscopy. The inset shows schematically the photo-thermal ionization process for a donor atom.

impurity level. When these impurity bands lie close to a conduction or valence band extremum, the coalescence of these impurity levels with band states produces band tailing. This band tailing results in a smearing out of the threshold of the fundamental absorption edge as observed in absorption measurements. When the impurity band broadening becomes sufficiently large that the electron wavefunction extends to adjacent sites, metallic conduction can occur. The onset of metallic conduction is called the Mott metal–insulator transition.

7.4   Vacancies, Color Centers and Interstitials

Closely related to the impurity problem is the vacancy problem. When a compound semiconductor crystallizes, the melt usually is slightly off stoichiometry with respect to the concentration of anions and cations. As an example suppose we prepare PbTe with Pb and Te concentrations in the melt that are stoichiometric to 0.01%. This means that there will be a slight excess of one of the atoms or slight deficiency of the other. This deficiency shows up in the crystal lattice as a vacancy or the absence of an atom. Such a vacancy represents a strong local perturbation of the crystal potential which again cannot be modeled in terms of hydrogenic impurity models. Such vacancy centers further tend to attract impurity atoms to form vacancy-impurity complexes. Such defects are difficult to model theoretically because their spatial localization requires participation of energy states throughout the Brillouin zone. Such centers generally give rise to energy states within the band gap of semiconductors and insulators. Such defect centers are often studied by optical techniques.

One important defect in ionic insulating crystals is the F-center ("Farbe" or color center). We see in Fig. 7.5 that the negative ion vacancy acts like a $+ve$ charge (absence of a $-ve$ charge). This effective $+ve$ charge tends to bind an electron. The binding of an electron to a $-ve$ ion vacancy is called an *F-center*. These F-centers give rise to absorption bands in the visible. Without F-centers, these crystals are usually clear and transparent. The F-center absorption band causes crystals with defects to appear colored, having the color of the transmitted light. When the crystals are heated to high enough temperatures, these defects can be made to anneal and the absorption bands disappear. This is called *bleaching*.

Many other color centers are found in ionic crystals. For example, we can have a hole bound to a $+ve$ ion vacancy. We can also have a defect formed by a vacancy that is bound to any impurity atom, forming a vacancy-impurity complex, which can bind a charged carrier. Or we can have two adjacent vacancies (one $+ve$ and the other $-ve$) binding an electron and a hole. Further generalizations are also found. These defect centers are collectively called *color centers* and each color center has its characteristic absorption band. In Fig. 7.6 we see an example of absorption bands due to F-centers in several alkali halides. In all cases the absorption bands are very broad, in contrast with the sharp impurity lines which are observed in the far infrared for shallow impurity level transitions (see Figs. 7.3 and 7.4). In the case of the vacancy there is a considerable lattice distortion around each vacancy site as the neighboring atoms rearrange their electronic bonding arrangements.

A few comments are in order about the classification of point defects. In Fig. 7.7 various types of point defects are shown. Figure 7.7(a) illustrates a perfect ionic crystal. Figure 7.7(b) shows an ionic crystal with vacancies. This particular collection of vacancies is of the Schottky type (equal numbers of positive and negative ion vacancies). Schottky point

Figure 7.5: Diagram of a negative ion vacancy or F–center in an ionic crystal.
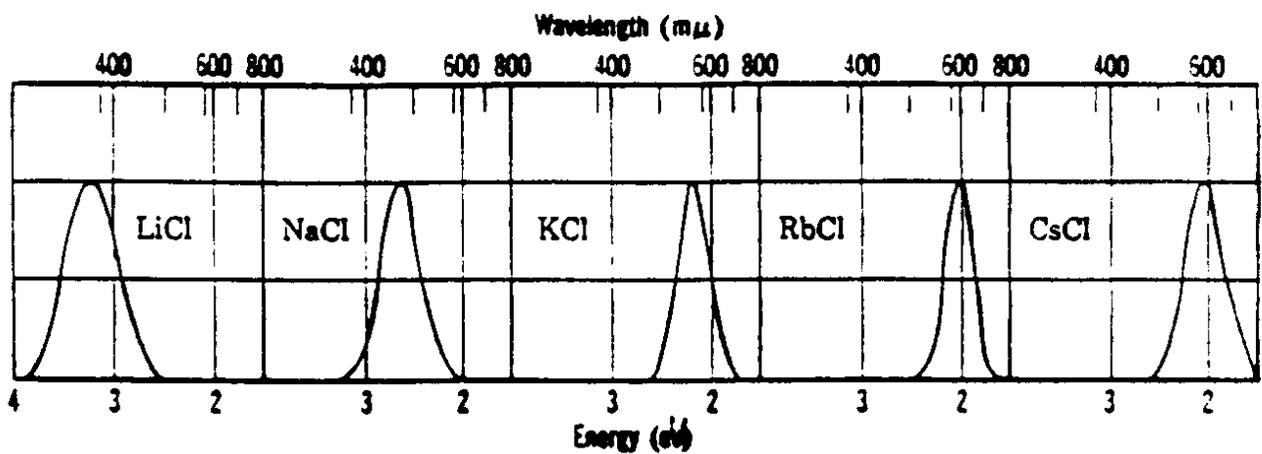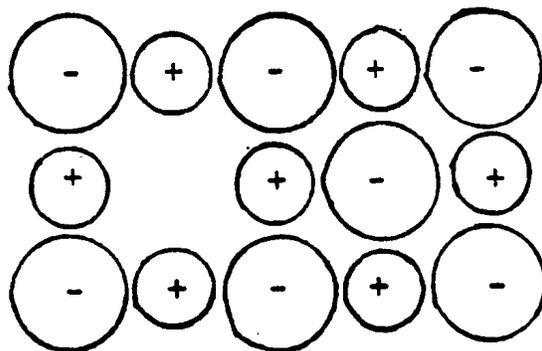

Figure 7.6: Examples of F–center absorption lines in various alkali halide ionic crystals.

defects also include neutral vacancies. Finally Fig. 7.7(c) shows both vacancies and interstitials. When a + (−) ion vacancy is near a + (−) ion interstitials, this defect configuration is called a Frenkel-type point defect.

We will now use simple statistical mechanical arguments to estimate the concentration of Schottky defects. Let E_s be the energy required to take an atom from a lattice site inside the crystal to the surface. If n is the number of vacancies, the change in internal energy resulting from vacancy generation is $U = nE_s$. Now the number of ways that n vacancy sites can be picked from N lattice sites is $N!/[(N-n)!n!]$, so that the formation of vacancies results in an increase in entropy of

$$S = k_B \ln \frac{N!}{(N-n)!n!} \tag{7.7}$$

and a change in free energy

$$F = U - TS = nE_s - k_B T \ln \frac{N!}{(N-n)!n!}. \tag{7.8}$$

Using Stirling's approximation for $\ln x!$ when x is large, we write

$$\ln x! \cong x \ln x - x. \tag{7.9}$$

Equilibrium is achieved when $(\partial F/\partial n) = 0$, so that at equilibrium we have

$$E_s = k_B T \left[\frac{\partial}{\partial n} \ln \frac{N!}{(N-n)!n!} \right] = k_B T \ln \frac{N-n}{n} \tag{7.10}$$

from which we write

$$\frac{n}{N-n} = e^{-\frac{E_s}{k_B T}} \tag{7.11}$$

or

$$\frac{n}{N} \sim \exp\left[-\frac{E_s}{k_B T}\right] \tag{7.12}$$

since $n \ll N$. The vacancy density is small because for $E_s \sim 1eV$, $T \sim 300K$, $(n/N) \sim e^{-40} \sim 10^{-17}$.

In the case of vacancy pair formation in an ionic crystal (Schottky defect), the number of ways to make n separated pairs is $[N!/(N-n)!n!]^2$ so that for Schottky vacancy pair formation

$$\frac{n_p}{N} \sim \exp\left[-\frac{E_p}{2k_B T}\right] \tag{7.13}$$

where n_p is the pair vacancy density and E_p is the energy required for pair formation.

These arguments can readily be extended to the formation of Frenkel defects and it can be shown that if N' is the density of possible interstitial sites, then the density of occupied interstitial sites is

$$n_i \simeq (NN')^{\frac{1}{2}} \exp\left[-\frac{E_i}{2k_B T}\right] \tag{7.14}$$

where E_i is the energy to remove an atom from a lattice site to form an interstitial defect site.

Figure 7.7: Schematic of various possible arrangements of both vacancies and interstitials. (a) a perfect ionic crystal, (b) an ionic crystal with positive and negative ion vacancies and (c) an ionic crystal with positive and negative ion vacancies and interstitials.
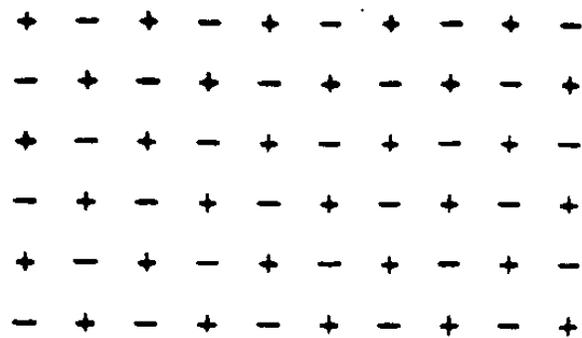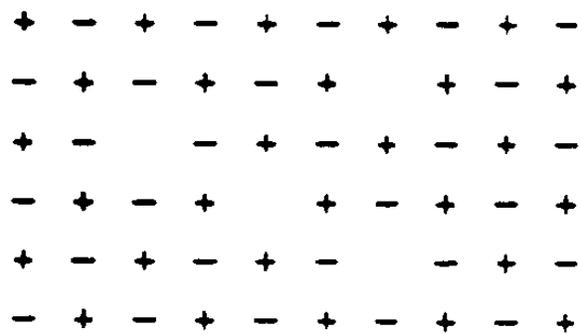
7.5   Spectroscopy of Excitons

An exciton denotes a system of an electron and a hole bound together by their Coulomb interaction. When a photon excites an electron into the con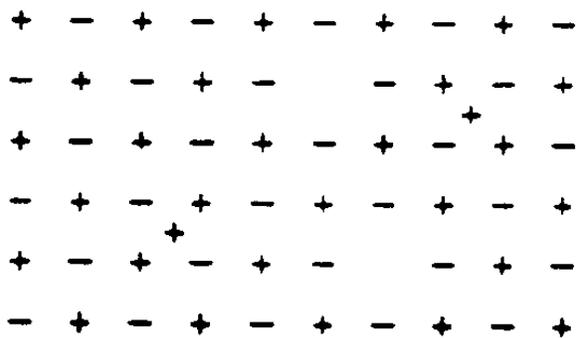duction band, a hole is left behind in the valence band; the electron, having a negative charge will be attracted to this hole and may (provided the energy is not too large) bind to the positively charged hole forming an exciton. Thus, the exciton binding energy is attractive and represents a lower energy state than the band states. Excitons are important in the optical spectra of bulk semiconductors especially at low temperature. Exciton levels are important for device applications since light emitting diodes and semiconductor lasers often involve excitons. However, because of the confinement of carriers in quantum wells, exciton effects become much more important in the case of quantum wells, superlattices and devices based on these deliberately structured materials (see Part I §8.3.1 of class notes). The topic of excitons in low dimensional semiconductor systems is discussed in §7.7.

   We will now use the effective mass approximation to find the exciton spectrum near an interband threshold and we assume that the exciton was created by a photon with energy slightly less than the direct energy gap E_g. The Schrödinger equation for the two-body exciton packet wave function Φ is written in the effective mass approximation as:

$$\left[\frac{p_e^2}{2m_e^*} + \frac{p_h^2}{2m_h^*} - \frac{e^2}{\varepsilon_0 |\vec{r}_e - \vec{r}_h|} \right] \Phi = E\Phi \tag{7.15}$$

thereby including the Coulomb binding energy of the electron–hole pair. For simplicity, we assume that the dielectric constant ε_0 is independent of $\vec{r}_e$ and $\vec{r}_h$ corresponding to a large spatial extension of the exciton in a semiconductor. We introduce new coordinates for the spatial separation $\vec{r}$ between the electron and hole

$$\vec{r} = \vec{r}_e - \vec{r}_h \tag{7.16}$$

and for the center of mass coordinate $\vec{\rho}$ given by

$$\vec{\rho} = \frac{m_e^* \vec{r}_e + m_h^* \vec{r}_h}{m_e^* + m_h^*}. \tag{7.17}$$

   We now separate the Schrödinger equation (Eq. 7.15) into an equation for the relative motion of the electron and hole in the exciton wave packet $F(\vec{r})$ and an equation of motion for the center of mass $G(\vec{\rho})$

$$\Phi(\vec{r}_e, \vec{r}_h) = F(\vec{r})G(\vec{\rho}). \tag{7.18}$$

Thus Eq. 7.15 becomes

$$\left[\frac{p_\rho^2}{2(m_e^* + m_h^*)} + \frac{p_r^2}{2m_r^*} - \frac{e^2}{\varepsilon_0 r} \right] F(\vec{r})G(\vec{\rho}) = E F(\vec{r})G(\vec{\rho}) \tag{7.19}$$

where the reduced effective mass μ^* is given by

$$\frac{1}{m_r^*} = \frac{1}{m_e^*} + \frac{1}{m_h^*} \tag{7.20}$$

83

to obtain an eigenvalue equation for $G(\vec{\rho})$

$$\left[\frac{p_\rho^2}{2(m_e^* + m_h^*)}\right]G(\vec{\rho}) = \Lambda G(\vec{\rho}) \tag{7.21}$$

which is of the free particle form and has eigenvalues

$$\Lambda(K) = \frac{\hbar^2 K^2}{2(m_e^* + m_h^*)} \tag{7.22}$$

where K is the wave vector of the exciton. The free particle solutions for the center of mass problem of Eq. 7.22 show that the exciton can move freely as a unit through the crystal. The momentum of the center of mass for a direct band gap exciton is small because of the small amount of momentum imparted to the excitation by the light.

We thus obtain the Schrödinger equation in the coordinate system of relative motion:

$$\left[\frac{p_r^2}{2m_r^*} - \frac{e^2}{\varepsilon_0 r}\right]F(\vec{r}) = E_n F(\vec{r}) \tag{7.23}$$

where Eq. 7.23 has the functional form of the Schrödinger equation for a hydrogen atom with eigenvalues E_n for quantum numbers n (where $n = 1, 2, \ldots$) given by

$$E_n = -\frac{m_r^* e^4}{2\hbar^2 \varepsilon_0^2 n^2}, \tag{7.24}$$

and the total energy for the exciton is then

$$E = \Lambda(K) + E_n. \tag{7.25}$$

The energy levels of Eq. 7.24 look like the donor impurity spectrum but instead of the effective mass of the conduction band m_e^* we now have the reduced effective mass m_r^* given by Eq. 7.20. Since m_r^* has a smaller magnitude than m_e^* as seen in Eq. 7.20, we conclude that the exciton binding energy is less than the impurity ionization energy for a particular solid. An example of a spectrum showing exciton effects is presented in Fig. 7.8. The points are experimental and the solid curves are a fit of the data points to $\varepsilon_2(\omega)$ for excitons given by

$$\varepsilon_2(\omega) = \frac{8\pi |(v|p|c)|^2 m_r^{*3}}{3\omega^2 \varepsilon_0^3} \sum_{n=1}^{\infty} \frac{1}{n^3} \delta(\omega - \omega_n), \tag{7.26}$$

where the sum is over all the exciton bound states. From Table 7.1 we see that the binding energy for excitons for GaAs is 4.9 meV and the effective Bohr radius is 112Å, which is many lattice spacings. The various exciton lines contributing to the exciton absorption profiles in Fig. 7.8 are unresolved even for the data shown for the lowest temperature of 21 K. A material for which the higher exciton energy levels ($n = 2, 3, \ldots$) of the Rydberg series are resolved is Cu_2O as can be seen in Fig. 7.9. The observation of these higher states is attributed to the forbidden nature of the coupling of the valence and conduction bands, giving rise to a strict selection rule that only allows coupling to exciton states with p symmetry. Since $n \geq 2$ is required to have p exciton states the $n = 1$ exciton is forbidden, and the exciton lines start at $n = 2$. The transitions are sharp, and well resolved exciton lines up to $n = 5$ can be identified in Fig. 7.9.

Table 7.1: Exciton binding energy (E_1) and Bohr radius (r_1) in some direct bandgap semiconductors with the zinc-blende structure (from Yu and Cardona).

Semiconductor	E_1 (meV)	E_1 (theory) (meV)	r_1 (Å)
GaAs	4.9	4.4	112
InP	5.1	5.14	113
CdTe	11	10.71	12.2
ZnTe	13	11.21	11.5
ZnSe	19.9	22.87	10.7
ZnS	29	38.02	10.22

Figure 7.8: Excitonic absorption spectra in GaAs near its bandgap for several sample temperatures. The lines drawn through the 21, 90 and 294 K data points represent fits with theory.
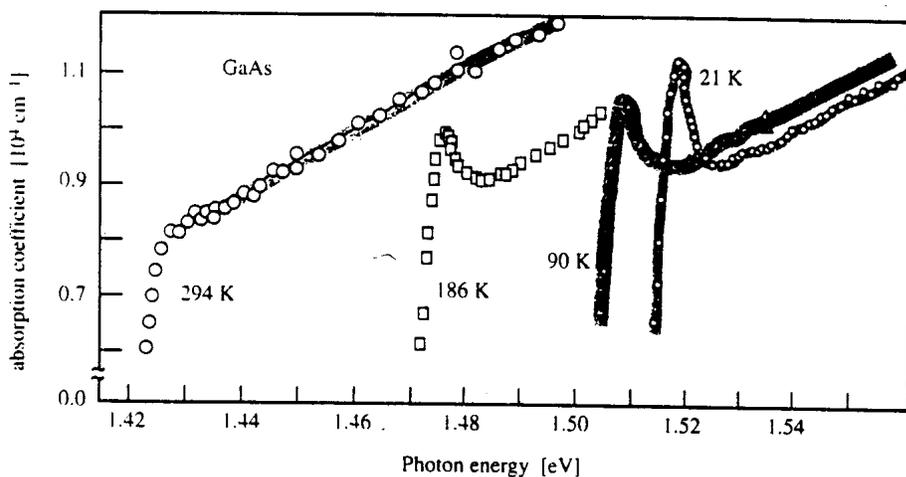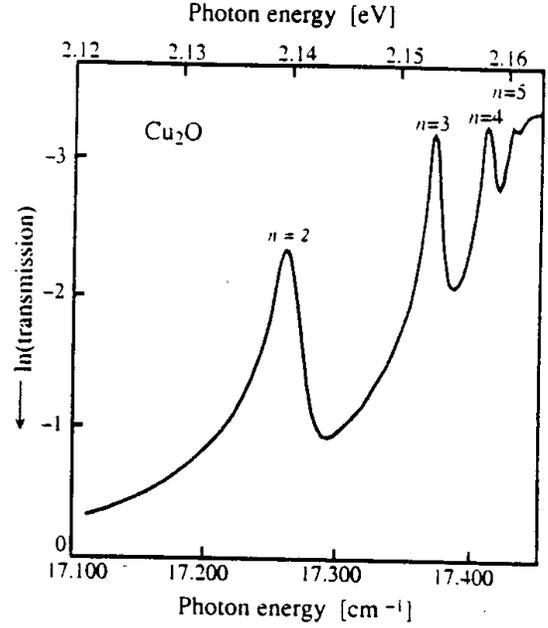
Figure 7.9: Low temperature absorption spectrum of Cu_2O (plotted as the log of the transmission) showing the excitonic p series associated with its "dipole-forbidden" band edge in Cu_2O.

The exciton spectrum appears quite similar to the impurity spectrum of shallow impurity states. These two types of spectra are distinguished through their respective dependences on impurity concentration. Suppose that we start with a very pure sample (10^{14} impurities/cm^3) and then dope the sample lightly (to 10^{16} impurities/cm^3). If the spectrum is due to donor impurity levels, the intensity of the lines would tend to increase and perhaps broaden somewhat. If, on the other hand, the spectrum is associated with an exciton, the spectrum would be attenuated because of screening effects associated with the charged impurities. Exciton states in 3D semiconductors are generally observed in very pure samples and at very low temperatures. The criterion is that the average Bohr orbit of the exciton is less than the distance between impurities. For the sake of this argument, consider an excitonic radius of ~ 100Å. If an impurity ion is located within this effective Bohr radius, then the electron–hole Coulomb interaction is screened by the impurity ion and the sharp spectrum associated with the excitons will disappear. A carrier concentration of 10^{16}/cm^3 corresponds to finding an impurity ion within every 100Å from some lattice point. Thus the electron–hole coupling can be screened out by a charged impurity concentration as low as 10^{16}/cm^3. Low temperatures are needed to yield an energy separation of the exciton levels that is larger than $k_B T$. Increasing the temperature shifts the absorption edge and broadens the exciton line in GaAs. At a temperature of 20 K we have $k_B T \simeq 1.7$ meV which is nearly as large as the exciton binding energy of 4.9 meV found in Table 7.1, explaining why no well resolved exciton spectrum for higher quantum states is observed. For the case of Cu_2O, the exciton binding energy of the ground state ($1s$), were it to exist, would be 97 meV, neglecting central cell corrections. The large exciton binding energy in Cu_2O also

helps with the resolution of the higher quantum exciton states.

7.6   Classification of Excitons

The exciton model discussed above is appropriate for a *free* exciton and a *direct* exciton. For the *direct* exciton, the initial excitation is accomplished in a $\vec{k}$-conserving process without the intervention of phonons. In materials like silicon and germanium, the thermal band gap corresponds to an indirect energy gap. For these materials, the exciton is formed by an indirect phonon-assisted process and the exciton is consequently called an *indirect* exciton.

Indirect excitons can be formed either with the emission or absorption of a phonon. Since excitons are more important at low temperatures, the emission process is much more likely than the absorption process. Because of the large difference in crystal momentum $\hbar \vec{k}$ between the valence band extremum and the lowest conduction band minimum in these indirect gap semiconductors, the exciton may acquire a large center of mass momentum corresponding to the momentum of the absorbed or emitted phonon $\hbar \vec{q}$. For the indirect exciton, a large range of crystal momentum $\hbar \vec{k}$ values are possible and hence the exciton levels spread out into bands as shown in the lower dashed rectangle of Fig. 7.10. This portion of the figure also appears in more detail in the upper left-hand corner. In Fig. 7.10 we also show in the upper right-hand corner the direct exciton associated with the Γ point conduction band for various temperatures. The shift in the absorption edge is associated with the decrease in band gap with increasing temperature. In Fig. 7.10, the individual exciton lines are not resolved – a lower temperature would be needed for that.

Addition of impurities to suppress the exciton formation does not help with the identification of bandgaps in semiconductors since the presence of impurities broadens the band edges. It is for this reason that energy gaps are best found from optical data in the presence of a magnetic field, to be discussed in connection with magnetism (Part III of this course).

For small distances from the impurity site or for small electron-hole separations, the effective mass approximation must be modified to consider central cell corrections explicitly. For example, central cell corrections are very important in Cu_2O so that the binding energy attributed to the $1s$ state is 133 meV, whereas the binding energy deduced from the Rydberg series shown in Fig. 7.9 indicates a binding energy of 97 meV.

The kinds of excitons we have been considering above are called *free* excitons. In contrast to these, are excitations called *bound* excitons. It is often the case that an electron and hole may achieve a lower energy state by locating themselves near some impurity site, in which case the exciton is called a *bound* exciton and has a larger binding energy. Bound excitons are observed in typical semi-conducting materials, along with free excitons.

Another category of excitons that occurs in semiconductors is the molecular excitons. Just as the energy of two hydrogen atoms decreases in forming molecular hydrogen H_2, the energy of two free (or bound) excitons may decrease on binding to form a molecular state. More complicated exciton complexes can be contemplated and some of these have been observed experimentally.

As the exciton density increases, further interaction occurs and eventually a quantum fluid called the electron-hole drop is formed. Unlike other fluids, both the negatively and positively charged particles in the electron-hole fluid have light masses. A high electron-hole density can be achieved in indirect band gap semiconductors such as silicon and germanium because of the long lifetimes of the electron-hole excitations in these materials. In treating
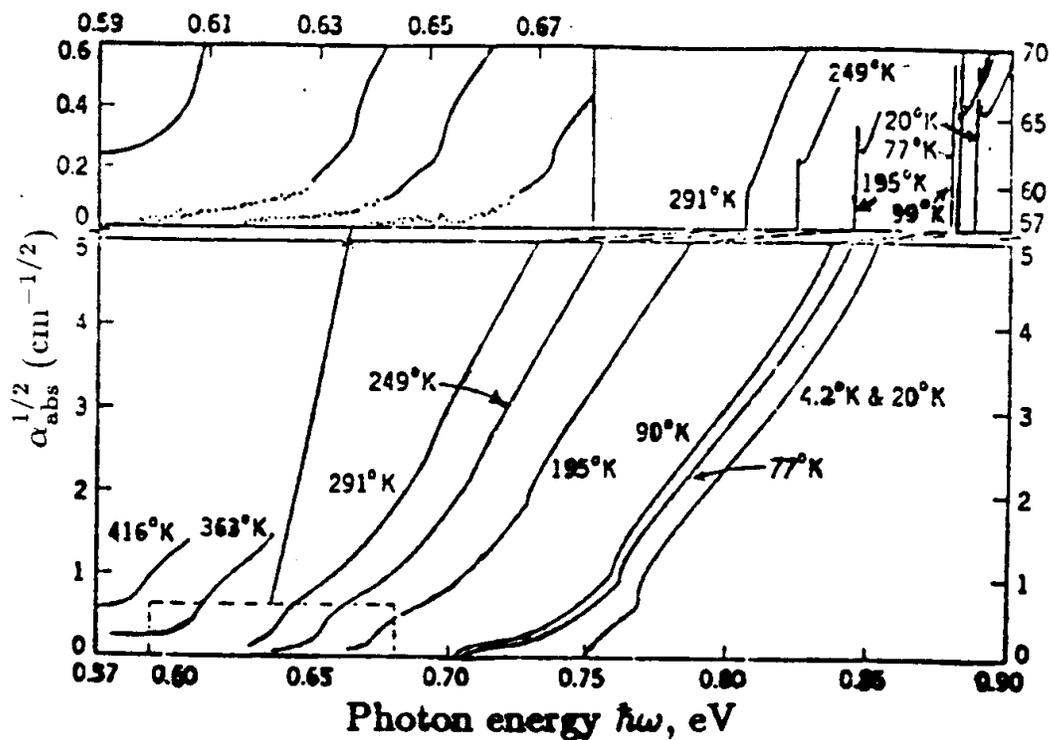
Figure 7.10: Plot of the square root of the absorption coefficient vs. $\hbar\omega$ for Ge for various temperatures showing the effect of the excitons. Features associated with both indirect and direct excitons are found.
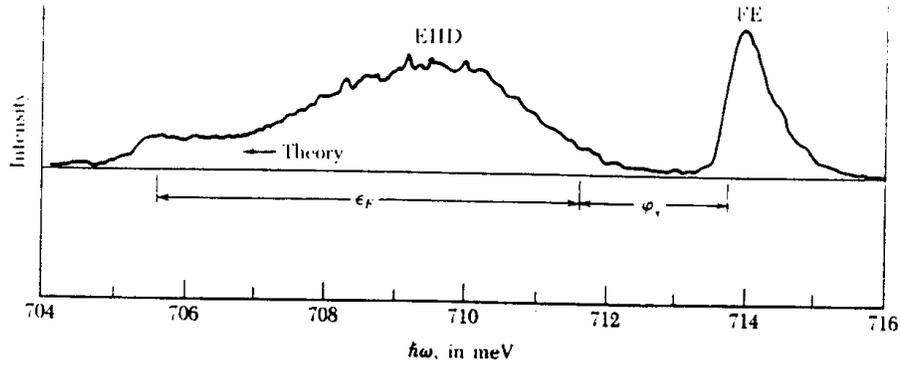
88

Figure 7.11: Recombination radiation (or photoluminescence spectrum) of free electrons (FE) and of electron-hole drops (EHD) in Ge at 3.04 K. The Fermi energy in the electron-hole drop is ϵ_F and the cohesive energy of the electron-hole drop with respect to a free exciton is $\phi_s = 1.8$ meV. The critical concentration and temperature for forming an electron-hole drop in Ge are respectively $2.6 \times 10^{17}/cm^3$ and 6.7 K.

the electron-hole drops theoretically, the electrons and holes are regarded as free particles moving in an effective potential due to the other electrons and holes. Because of the Pauli exclusion principle, no two electrons (or holes) can have the same set of quantum numbers. For this reason like particles tend to repel each other spatially, but unlike particles do not experience this repulsion. Thus electron-hole pairs are formed and these pairs can be bound to each other to form an electron-hole drop. These electron-hole drops have been studied in the emission or luminescence spectra (see Chapter 8). Results for the luminescence spectra of Ge and Si at very low temperatures ($T \leq 2K$) are shown in Fig. 7.11. Luminescence spectra for germanium provide experimental evidence for electron-hole drops for electron-hole concentrations exceeding $10^{17}/cm^3$.

In insulators (as for example alkali halides), excitons are particularly important, but here they tend to be well localized in space because the effective masses of any carriers that are well localized tend to be large. These localized excitons, called *Frenkel* excitons, are much more strongly bound and must be considered on the basis of a much more complicated theory. It is only for the excitons which extend over many lattice sites, the *Wannier* excitons, that effective mass theory can be used. And even here many-body effects must be considered to solve the problem with any degree of accuracy – already an electron bound to a hole is a two–body problem so that one–electron effective mass theory is generally not completely valid.

In studying the optical absorption of the direct gap, the presence of excitons complicates the determination of the direct energy gap, particularly in alkali halides where the exciton binding energy is large. Referring to Fig. 7.12(a), both Γ-point and L-point excitons are identified in the alkali halide ionic crystal KBr. The correspondence of the optical structure with the $E(\vec{k})$ diagram is shown by comparison of Figs. 7.12(a) and (b). Here it is seen that the Frenkel exciton lines dominate the spectrum at the absorption edge and we also see huge shifts in energy between the exciton lines and the direct absorption edge. These
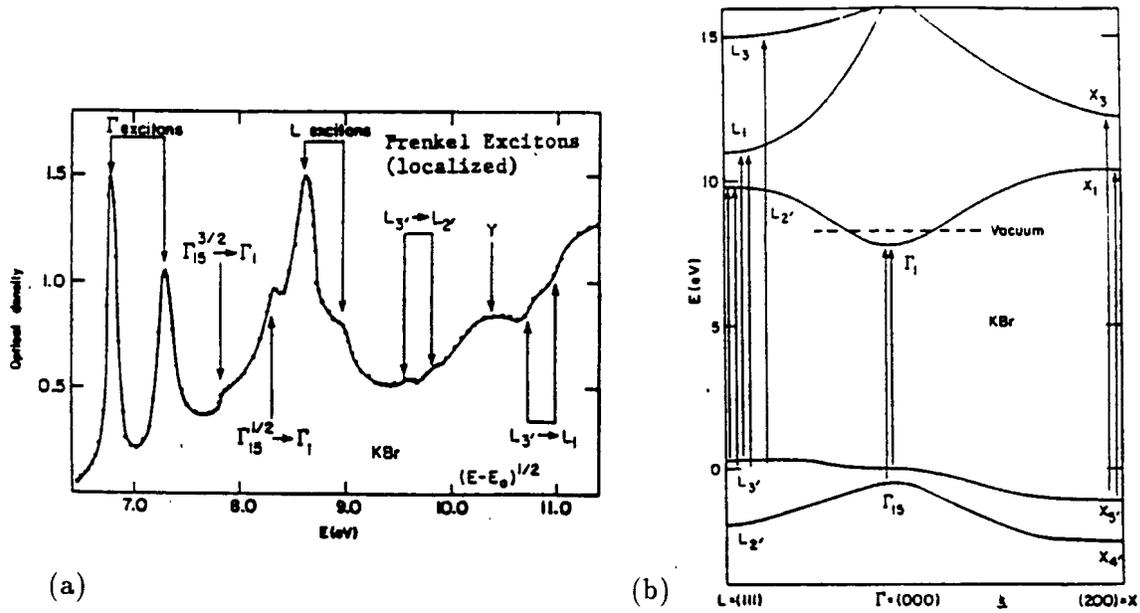
Figure 7.12: (a) A spectrum of the optical density of KBr showing Frenkel excitons. The optical density is defined as $\log(1/\mathcal{T})$ where $\mathcal{T}$ is the optical transmission. (b) The energy bands of KBr, as inferred from tight-binding calculations of the valence bands and the assignments of interband edges in optical experiments. We note that the spectrum is dominated by exciton effects and that direct band edge contributions are much less important and the binding energy is on the order of an electron volt.

figures show the dominance of strongly bound, localized Frenkel excitons in the spectra of alkali halides.

Excitons involve the presence of an electron-hole pair. If instead, an electron is introduced into the conduction band of an ionic crystal, a charge rearrangement occurs. This charge rearrangement partially screens the electron, thereby reducing its effective charge. When an electric field is now applied and the charge starts to move through the crystal, it moves together with this lattice polarization. The electron together with its lattice polarization is called a *polaron*. While excitons are important in describing the optical properties of ionic or partly ionic materials, polarons are important in describing the transport properties of such materials. The presence of polarons leads to thermally activated mobilities, which says that a potential barrier must be overcome to move an electron together with its lattice polarization through the crystal. The presence of polaron effects also results in an enhancement in the effective mass of the electron. Just as one categorizes excitons as weakly bound (Wannier) or strongly bound (Frenkel), a polaron may behave as a free particle with a relatively weak enhancement of the effective mass (a *large* polaron) or may be in a bound state with a finite excitation energy (a *small* polaron), depending on the strength of the electron-phonon coupling. Large polarons are typically seen in weakly ionic semiconductors, and small polarons in strongly ionic, large-gap materials. Direct evidence for large polarons in semiconductors has come from optical experiments in a magnetic field in the region where the cyclotron frequency ω_c is close to the optical phonon frequency ω_{LO}.

7.7 Optical Transitions in Quantum Well Structures

Optical studies are extremely important in the study of quantum wells and superlattices. For example, the most direct evidence for bound states in quantum wells comes from optical absorption measurements. To illustrate such optical experiments consider a GaAs quantum well bounded on either side by the wider gap semiconductor $Al_xGa_{1-x}As$. Because of the excellent lattice matching between GaAs and $Al_xGa_{1-x}As$, these materials have provided the prototype semiconductor superlattice for study of the 2D electron gas. The threshold for absorption is now no longer the band gap of bulk GaAs but rather the energy separation between the highest lying bound state of the valence band and the lowest bound state of the conduction band. Since the valence band of GaAs is degenerate at $\vec{k} = 0$ and consists of light and heavy holes, there will be two $n = 1$ levels in the valence band. Since $E_n \propto 1/m^*$ for the quantum well $n = 1$ bound state level, the heavy hole subband extremum will be closer in energy to the band edge than that of the light hole as shown on the left side of Fig. 7.13. Also the density of states for the heavy hole subband will be greater than that of the light hole subband by a factor of $2m_{hh}^*/m_{lh}^*$. The optical absorption will thus show two peaks near the optical threshold as illustrated in the diagram in Fig. 7.13. These data are for a sample with a quantum well width of 50Å, which is small enough to contain a single bound state ($n = 1$). Since the optical absorption from a single quantum well is very weak, the experiment is usually performed in superlattice structures containing a periodic array of many equivalent quantum wells. In forming the superlattice structure, it is important that the barrier between the quantum wells is not too small in extent, because for small spatial separations between quantum wells and low band offsets at the interfaces, the eigenfunctions in adjacent wells become coupled and we no longer have a 2D electron gas in the quantum well.

Figure 7.13: Optical excitations in a quantum well where the valence band has light and heavy holes (as in GaAs). The optical density is defined as $\log(1/\mathcal{T})$ where $\mathcal{T}$ is the transmission.
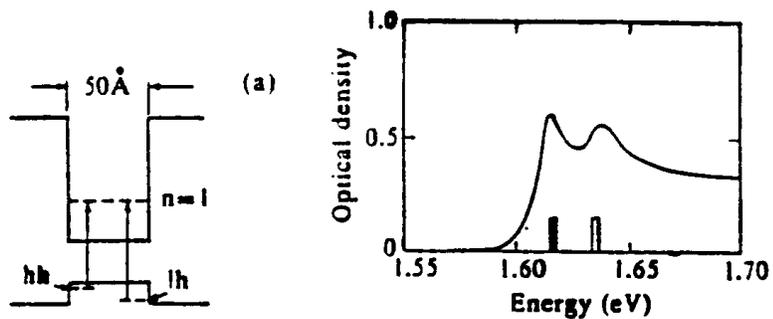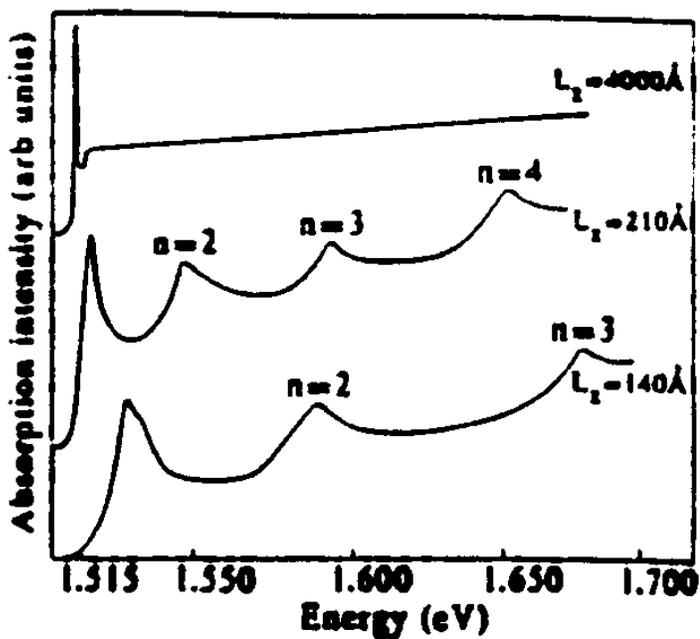

Figure 7.14: Frequency dependence of the absorption for GaAs/Al$_{0.2}$Ga$_{0.8}$As heterostructure superlattices of different thicknesses at optical frequencies.

92

For wider quantum wells containing several bound states (see Figs. 7.14 and 7.15), a series of absorption peaks are found for the various bound states, and the interband transitions follow the selection rule $\Delta n = 0$. This selection rule follows because of the orthogonality of wave functions for different states n and n'. Thus to get a large n matrix element for coupling valence and conduction band states, n' and n must be equal. As the width of the quantum well increases, the spectral features associated with transitions to the bound states become smaller in intensity and more closely spaced and eventually cannot be resolved. For the thickest films, the quantum levels are too close to each other to be resolved and only the bulk exciton peak is seen. For the 210Å quantum well (see Fig. 7.14), transitions for all 4 bound states within the quantum well are observed. In addition, excitonic behavior is observed on the $n = 1$ peak. For the 140 Å well, the transitions are broader, and effects due to the light and heavy hole levels can be seen through the distorted lineshape (see Fig. 7.14. To observe transitions to higher bound states, the spectra in Fig. 7.15 are taken for a quantum well width of 316 Å, for which transitions up to (6,6) are resolved. For such wide quantum wells, the contributions from the light holes are only seen clearly when a transition for a light hole state is not close to a heavy hole transition because of the lower density of states for the light holes (see Fig. 7.15).

Exciton effects are significantly more pronounced in quantum well structures than in bulk semiconductors, as can be understood from the following considerations. When the width d_1 of the quantum well is less than the diameter of the exciton Bohr orbit, the electron–hole separation will be limited by the quantum well width rather than by the larger Bohr radius, thereby significantly increasing the Coulomb binding energy and the intensity of the exciton peaks. Thus small quantum well widths enhance exciton effects. Normally sharp exciton peaks in bulk GaAs are observed only at low temperature ($T \ll 77$K); but in quantum well structures, excitons can be observed at room temperature, as shown in Fig. 7.14, which should be compared with Fig. 7.8 for 3D bulk GaAs.

The reason why the exciton line intensities are so much stronger in the quantum well structures is due to the reduction in the radius of the effective real space Bohr orbits, thereby allowing more k band–states to contribute to the optical transition. This argument is analogous to arguments made to explain why the exciton intensities for the alkali halides are huge [see Fig. 7.12(a)]. In the alkali halides the excitons have very small real space Bohr orbits so that large regions of k space can contribute to the exciton excitation.

In the case of the quantum well structures, two exciton peaks are observed because the bound states for heavy and light holes have different energies, in contrast to the case of bulk GaAs where the $j = 3/2$ valence band states are degenerate at $k = 0$. This property was already noted in connection with Fig. 7.13 for the bound state energies. Because of the large phonon density available at room temperature, the ionization time for excitons is only 3×10^{-13} sec. Also the presence of the electron–hole plasma strongly modifies the optical constants, so that the optical constants are strongly dependent on the light intensity, thereby giving rise to non–linear effects that are not easily observed in 3D semiconductors.

Because of the small binding energy of these exciton states, modest electric fields have a relatively large effect on the photon energy of the exciton peaks and on the optical constants. Application of an electric field perpendicular to the layers of the superlattice confines the electron and hole wave functions at opposite ends of the quantum well, as shown in Fig. 7.16.

Because of this spatial separation, the excitons become relatively long lived and now recombine on a time scale of 10^{-9} sec. Also because of the quantum confinement, it is

93

Figure 7.15: Transmission spectrum of a GaAs/AlGaAs multi-quantum well (well width = 316 Å) measured as a function of photon energy at low temperature (right panel). The peaks labeled (n, n) have been identified with optical transitions from the nth heavy hole (hh) and light hole (lh) subbands to the nth conduction subband as shown by arrows in the band diagram in the left panel. The values of the band offsets used in the analysis are given in the diagram, but these are not the most recent values.

94

Figure 7.16: Excitonic wave functions in a GaAs quantum well without (left) and with (right) an applied electric field. Because of the triangular potentials that are created by the electric field in the z–direction, the quantum well retains the electron and hole in a bound state at electric fields much higher than would be possible in the bulk classical ionization field.
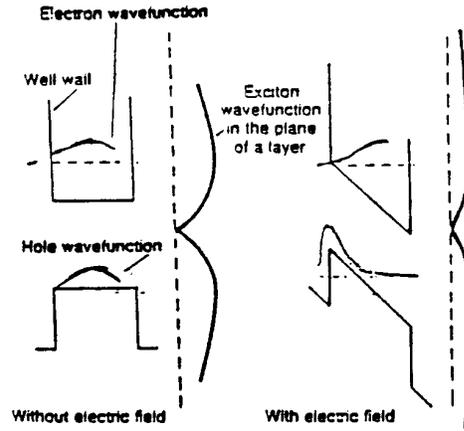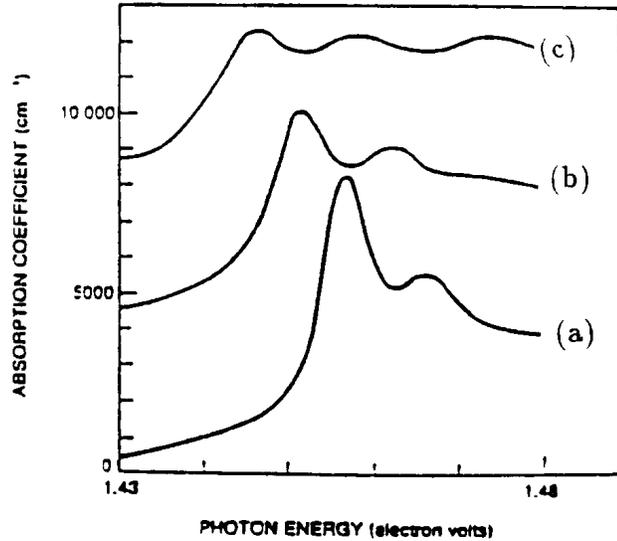

Figure 7.17: The absorption spectra in GaAs/Ga$_{1-x}$Al$_x$As heterostructures for various values of applied electric field illustrating the large changes in optical properties produced by the quantum confined Stark shift. The electric fields normal to the layer planes are: (a) 10^4 V/cm, (b) 5×10^4 V/cm, and (c) 7.5×10^4 V/cm.

95

possible to apply much higher (50 times) electric fields than is possible for an ionization field in a bulk semiconductor, thereby producing very large Stark red shifts of the exciton peaks, as shown in Fig. 7.17. This electric field effect on the exciton levels is called the *quantum confined Stark effect*. This effect is not observed in bulk semiconductors. The large electric field–induced change in the optical absorption that is seen in Fig. 7.17 has been exploited for device applications.

The following mechanism is proposed to explain the quantum confined Stark effect when the electric field is applied perpendicular to the layers. This electric field pulls the electrons and holes towards opposite sides of the layers as shown in Fig. 7.16 resulting in an overall net reduction in the attractive energy of the electron–hole pair and a corresponding Stark (electric field induced) shift in the exciton absorption. Two separate reasons explain the strong exciton peaks in quantum well structures. Firstly the walls of the quantum wells impede the electron and hole from tunneling out of the wells. Secondly, because the wells are narrow (e.g., ~ 100Å) compared to the three–dimensional (3D) exciton size (e.g., ~ 200Å), the electron–hole interaction, although slightly weakened by the separation of electron and hole, is still strong, and well defined excitonic states can still exist. Thus exciton resonances can remain to much higher fields than would be possible in the absence of this confinement, and large absorption shifts can be seen without excessive broadening.

Chapter 8

Luminescence and Photoconductivity

8.1 Classification of Luminescence Processes

Luminescence denotes the emission of radiation by a solid in excess of the amount emitted in thermal equilibrium. Since luminescence is basically a non-equilibrium phenomena, it requires excitation by light, electron beams, current injection, etc., which generally act to create excess electrons, holes, or both. The effects of electron–hole recombination give rise to **recombination radiation** or luminescence.

One classification of luminescent processes is based on the source of the excitation energy. The most important excitation sources are

1. **photoluminescence** by optical radiation,

2. **electroluminescence** by electric fields or currents,

3. **cathodoluminescence** by electron beams (or cathode rays),

4. **radioluminescence** by other energetic particles or high energy radiation.

A second classification of luminescent processes pertains to the time that the light is emitted relative to the initial excitation. If the emission is fast ($\overset{<}{\sim} 10^{-8}$ sec is a typical lifetime for an atomic excited state), then the process is **fluorescent**. The emission from most photoconductors is of the fluorescent variety. For some materials, the emission process is slow and can last for minutes or hours. These materials are **phosphorescent** and are called **phosphors**.

Let us first consider luminescent processes of the fluorescent type with fast emission times. The electronic transitions which follow the excitation and which result in luminescent emission are generally the same for the various types of excitations. Figure 8.1 shows a schematic diagram of the basic transitions in a semiconductor. These may be classified as follows:

1. Transitions involving chemical impurities or physical defects (such as lattice vacancies):

Figure 8.1: Basic transitions in a semi-conductor for the luminescent process. After H.F. Ivey, IEEE J.Q.E. *2*, 713 (1966). ● = electrons: ○ = holes.
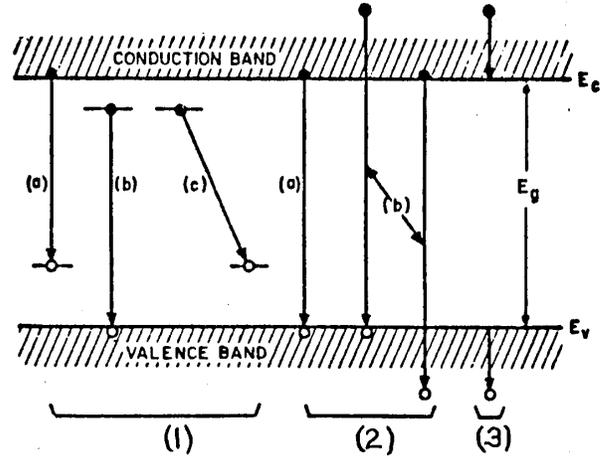
   (a) conduction band to acceptor.

   (b) donor to valence band.

   (c) donor to acceptor (pair emission).

2. Interband transitions:

   (a) intrinsic or edge emission corresponding very closely in energy to the band gap, though phonons and/or excitons may also be involved

   (b) higher energy emission involving energetic or "hot" carriers, sometimes related to **avalanche** emission, where "hot" carriers refers to highly energetic carriers well above thermal equilibrium levels.

3. Intraband transitions involving "hot" carriers, sometimes called deceleration emission.

It should be pointed out that the various transitions mentioned above do not occur in the same material or under the same conditions. Nor are all electronic transitions radiative. Phonon emission provides a non-radiative mechanism for the relaxation of an excited state in a solid to the lowest equilibrium ground state. An efficient luminescent material is one in which radiative transitions predominate over nonradiative ones.

When electron–hole pairs are generated by external excitations, radiative transitions resulting from the hole–electron recombination may occur. The radiative transitions in which the sum of electron and photon wavevectors is conserved are called direct transitions as opposed to indirect transitions which involve scattering agents such as phonons.

8.2   Emission and Absorption

For a given material the emission progagility will depend on the photon energy and on the temperature. The emission rate $R_{vc}(\omega)$ for the transition from the conduction band (c) to

Figure 8.2: Luminescence emission spectrum in an n-type InSb crystal with an electron concentration of $5 \times 10^{13} \text{cm}^{-3}$. The peak at 0.234 eV is due to inter-band recombinative emission. The peak at 0.212 eV (multiplied by 200) is due to phonon-assisted band-to-band transitions. (A. Mooradian and H.Y. Fan, Phys. Rev. *148*, 873 (1966).)
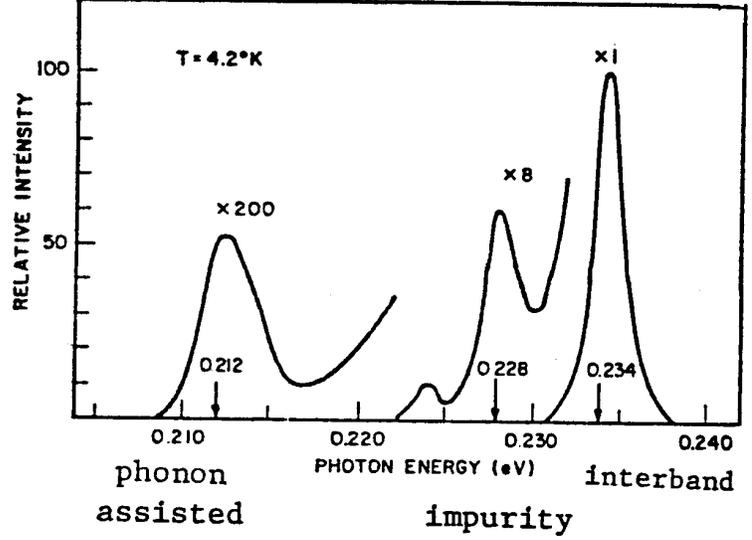
the valence band (v) is related to the absorption rate $P_{vc}(\omega)$ by the relation

$$R_{cv}(\omega) = P_{vc}(\omega)\rho(\omega) \tag{8.1}$$

where $\rho(\omega)$ is the Planck distribution at temperature T

$$\rho(\omega) = \frac{2}{\pi} \frac{\omega^2 n_r^3}{c^3[\exp(\hbar\omega/k_B T) - 1]} \tag{8.2}$$

and the absorption rate is given by

$$P_{vc}(\omega) = \frac{\alpha(\omega)c}{n_r} \tag{8.3}$$

so that the frequency and temperature dependence of the emission rate is given by

$$R_{cv}(\omega) = \frac{2}{\pi} \frac{\omega^2 n_r^2 \alpha\omega}{c^2[\exp(\hbar\omega/k_B T) - 1]}. \tag{8.4}$$

Basically, $R_{cv}(\omega)$ shows high emission at frequencies where the absorption is large, so that emission spectroscopy can be used as a technique to study various aspects of the band structure. For example, Fig. 8.2 shows luminescence from the conduction band of InSb to the valence band (band-to-band process) at 0.234 eV from the conduction band to an acceptor impurity level at 0.228 eV and luminescence that is phonon assisted at 0.212 eV involving phonon absorption.

For intrinsic or band-to-band transitions, the peak intensity occurs near the energy gap and the width of the spectral line (at half value of peak intensity) is proportional to

Figure 8.3: Direct and indirect intrinsic radiation recombination in Ge. The 70K spectrum is experimental and is in the energy range appropriate for indirect transitions assisted by longitudinal acoustic (LA) phonons. The circles are calculated from absorption data for both types of transitions. The free carrier densities at the direct and indirect conduction band minima and at the valence band maximum are denoted by n_0, n, and p, respectively.
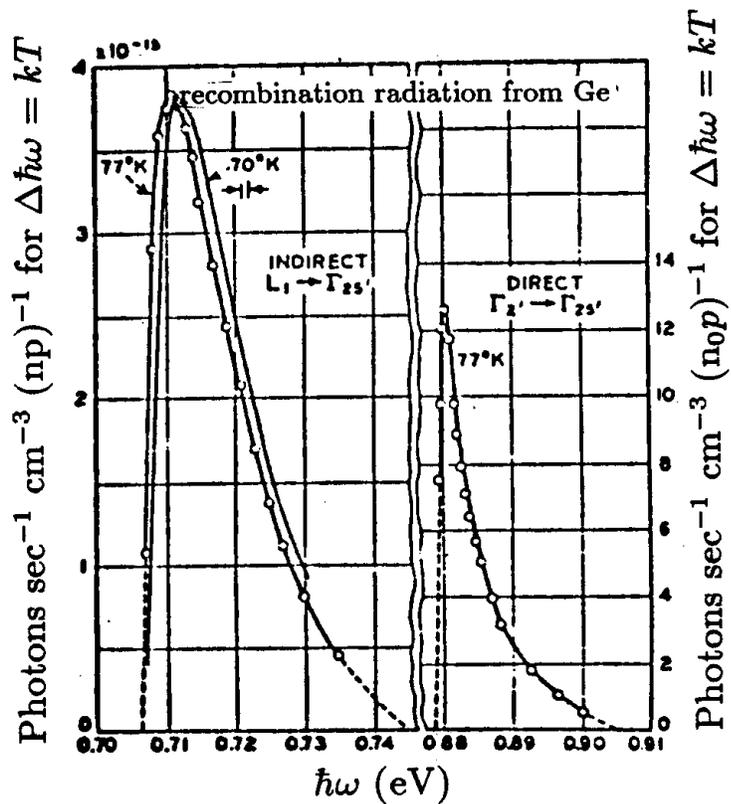
the thermal broadening energy $k_B T$. For extrinsic transitions, the peak emission intensity occurs near the transition energy, but the broadening is greater than for the intrinsic band-to-band emission shown in Fig. 8.3.

The general problem of luminescence is not only to determine the luminescent mechanisms and the emission spectra, as discussed above, but also to determine the **luminescent efficiency**. For a given input excitation energy, the radiative recombination process is in direct competition with the non-radiative processes. Luminescent efficiency is defined as the ratio of the energy associated with the radiative process to the total input energy.

Among the fastest emission luminescent processes, electroluminescence, or excitation by an electric field or current, has been one of the most widely utilized for device applications. Electroluminescence is excited in a variety of ways including intrinsic, injection, avalanche, and tunneling processes.

1. **Intrinsic process**. When a powder of a semiconductor, (e.g., ZnS) is embedded in a dielectric (plastic or glass), and exposed to an alternating electric field, usually at audio frequencies, electroluminescence may occur. Generally the efficiency is low ($\sim 1\%$) and such materials are used primarily in display devices. The mechanism is mainly due to impact ionization by accelerated electrons and/or field emission of electrons from trapping centers.

2. **Injection**. Under forward-bias conditions, the injection of minority carriers in a $p-n$ junction can give rise to radiative recombination. The energy level band diagram for a Cd-doped GaP $p-n$ junction is shown in the Fig. 8.4. Several different transitions for electron-hole recombination are indicated. The relative intensity of the red and the green bands can be varied by varying the impurity concentrations. The brightness of the red-light emission from the GaP $p-n$ junction at room temperature is sufficiently high to merit electro-luminescent applications, as an example, GaP light-emitting-diodes (LEDs) for numeric displays in pocket electronic calculators.

   At the present time, the highest electro-luminescent efficiency has been obtained experimentally in forward-biased GaAs diodes. This is expected because (1) the forward-bias injection is a very efficient method since electric energy can be converted directly into photons; (2) GaAs is a direct-gap semiconductor. Thus, the radiative recombination process is a first-order transitions process (no phonon involved); and (3) GaAs has the most advanced materials technology of all the direct-gap semiconductors.

3. **Avalanche**. When a $p-n$ junction or a metal semiconductor contact is reverse-biased into avalanche breakdown, the electron-hole pairs generated by impact ionization may result in emission of either interband (avalanche emission) or intraband (deceleration emission) transitions, respectively.

4. **Tunneling**. Electroluminescence can result from tunneling into forward-biased and reverse-biased junctions. In addition, light emission can occur in reverse-biased metal-semi-conductor contacts.

Fast emission luminescence also is of importance to semiconducting lasers. Luminescence is an **incoherence** emission process in contrast with laser action which involves the **coherent** emission of radiation in executing a radiative transition. The coherence is usually
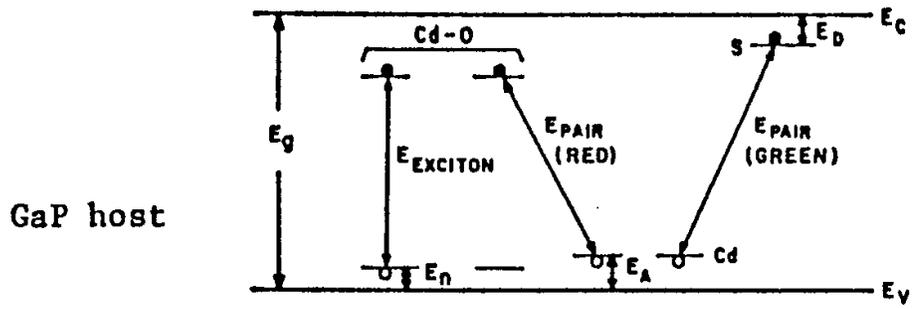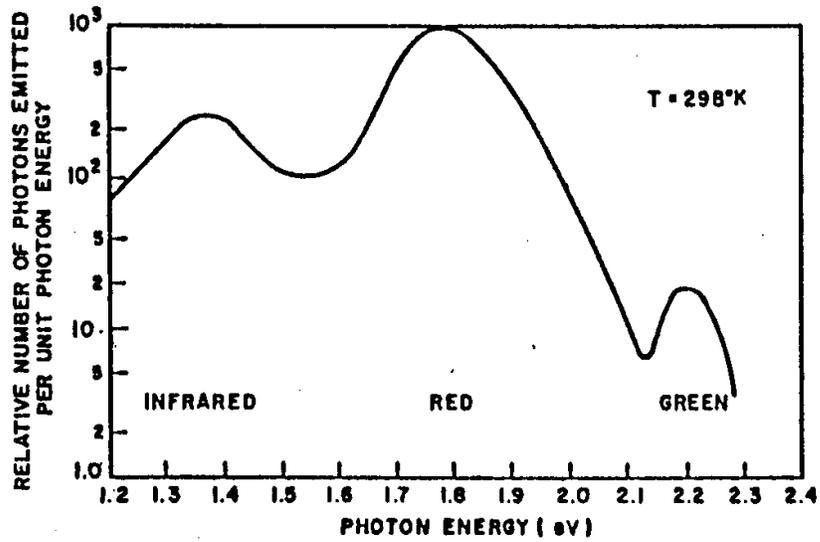
Figure 8.4: Energy level diagram for a Cd-doped GaP $p - n$ junction where Cd-O denotes an cadmium-oxygen complex. Transitions between the exciton level of the Cd-O complex to the acceptor level of Cd give rise to red light emission. Transitions between the donor level (S) and acceptor level (Cd) give rise to the green light emission. (b) Measured emission spectrum from a GaP diode in which the color associated with the various luminescent peaks are shown. (After M. Gershenzon, Bell Sys. Tech. J. *45*, 1599 (1966).)

Figure 8.5: Schematic of the phosphorescence process of the thallium$^+$ activator in KCl. The emission is downshifted from the absorption. This is an illustration of the Franck–Condon Principle.
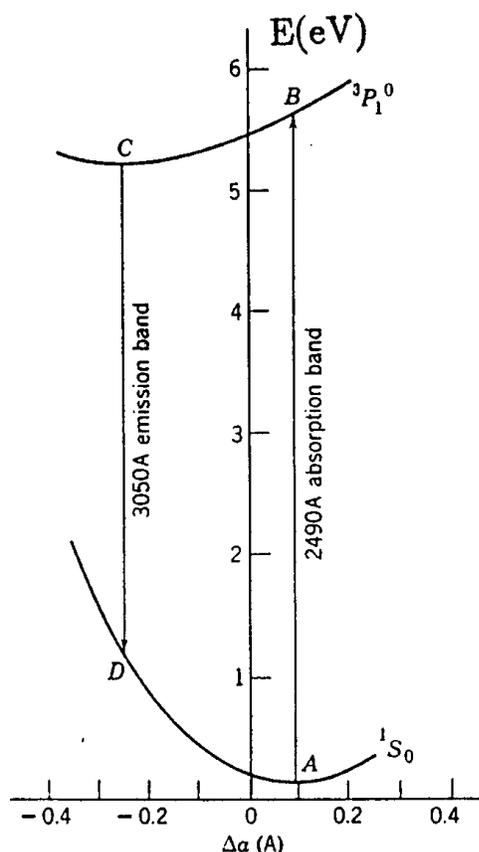
enhanced by polishing the sample faces to form an optical cavity. Examples of solid state lasers are the ruby laser and the direct gap semiconductor lasers. Optical and electrical pumping are the most common methods of exciting laser action in solid state lasers.

Finally, we conclude the discussion of electroluminescence in semiconductors with a short discussion of slow emission luminescence, i.e. phosphorescence. Phosphorescent materials exhibit afterglow effects and are consequently important in various optical display devices. These phosphors often do not exhibit large photoconductivities. That is to say, although the electrons that were produced survive for a long time, they are bound to particular defect centers and do not readily carry charge through the crystal.

In Fig. 8.5 we show an example of how a phosphor works in an alkali halide such as KCl with a small amount of Tl impurities. The thallium defects act as recombination centers. If these recombination centers are very efficient at producing recombination radiation they are called **activators**; Tl in KCl acts as an activator. In this system, the excitation occurs at higher energy than the emission. The **Franck-Condon principle** states that the atoms in the solid do not change their internuclear separations during an electronic transition. We now explain how emitted light is downshifted in frequency from the exciting light. The Tl$^+$ ion in the ground or unexcited state occupies some configuration close to the symmetric center of a K$^+$ ion which the Tl$^+$ ion might be replacing. When excited, the Tl$^+$ ion finds

103

Figure 8.6: Schematic diagram of the experimental arrangement for measuring the photoconductivity.
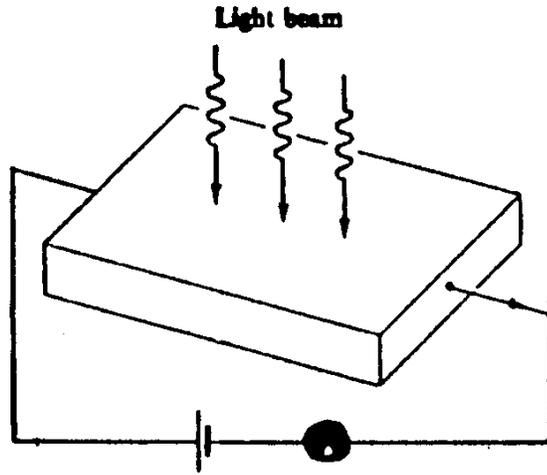
a lower energy state in a lower symmetry position near one of the Cl^- ions as shown on the top of Fig. 8.5. The absorption is made from the ground state energy (point A in Fig. 8.5) to an excited state with the same configuration. Phonon interactions then will bring the electron to the equilibrium position C. Achievement of equilibrium ($B \to C$) will take a longer time than the electronic transitions ($A \to B$). Emission from $C \to D$ again occurs in accordance with the Franck-Condon principle and the readjustment to the equilibrium configuration A proceeds by phonon processes.

8.3 Photoconductivity

Photoconductivity is observed when light is incident on a poorly conducting material, (e.g., an insulator or semiconductor), and the photon energy is sufficiently high to excite an electron from an occupied valence state to an unoccupied conduction state. In such interband transitions both the electron and hole will contribute to the electrical conductivity if a voltage is applied across the sample as shown in the schematic experimental arrangement in Fig. 8.6. Since the threshold for photoconduction occurs at $\hbar\omega = E_g$, measurement of the photoconductivity can be used to determine the band gap for non-conducting materials. Photoconductivity is often the concept used for the design of practical optical detectors.

The photoconduction process increases the electrical conductivity $\Delta\sigma$ due to the increase in the density of electrons (Δn) and (Δp) resulting from photo-excitation:

$$\frac{\Delta\sigma}{\sigma} = \frac{\Delta n\mu_n + \Delta p\mu_p}{n\mu_n + p\mu_p} \qquad (8.5)$$

in which $\mu_n + \mu_p$ are respectively, the electron and hole mobilities. Since the carriers are generated in pairs in the photo-excitation process $\Delta n = \Delta p$. In preparing materials for application as photoconductors, it is desirable to have a high mobility material with a low intrinsic carrier concentration, and long electron–hole recombination times to maximize the photo-excited carrier density concentration. CdS is an example of a good photoconductive

104

material. In CdS, it is possible to change the conductivity by $\sim$ **10 orders of magnitude** through carrier generation by light. These large changes in electrical conductivity can be utilized in a variety of device applications such as: light meters, photo-detectors, "electric eye" control applications, optically activated switches, and information storage.

To measure photo-currents, photo-excited carriers are collected at the external electrodes. In the steady state, free carriers are continually created by the incident light. At the same time, they annihilate each other through electron-hole recombination. To produce a large photocurrent, it is desirable to have a *long* free carrier *lifetime* τ' or a slow recombination time. If G is the rate of generation of electrons per unit volume due to photo-excitation, then the photo-excited electron density in the steady state will be given by

$$\Delta n = G\tau'. \tag{8.6}$$

The generation rate G will in turn be proportional to the photon flux incident on the photoconductor. Whereas slow recombination rates are essential to the operation of photoconductors, rapid recombination rates are necessary for luminescent materials.

In the recombination process, an electron and hole annihilate each other, emitting a photon in a radiative process. In real materials, the recombination process tends to be accelerated by certain defect sites. When such defects tend to be present in relatively greater concentrations at the surface, the process is called **surface recombination**. In bulk, the density of recombination centers can be made low for a very pure and "good" crystal. A typical concentration in a high quality Si crystal would be $\sim 10^{12}$ cm^{-3}.

Photo-excited carriers can also be eliminated from the conduction process by **electron** and **hole traps**. These traps differ from recombination centers insofar as traps preferentially eliminate a single type of carrier. In practice, hole traps seem to be more common than electron traps. For example, in the silver halides which are important in the photographic process, the hole is trapped almost as soon it is produced and photoconduction occurs through the electrons.

Electron and neutron irradiation produce both recombination centers and traps in photoconducting materials. Thus, special precautions must be exercised in using photo-detectors in a high radiation environment or on satellites which are expected to pass through regions of high radiation fluences.

Trapped electrons can be released by thermal or optical excitation. For example, consider a p-type sample of Ge which has been doped with Mn, Ni, Co, Fe. At low temperatures E_F will be near the top of the valence band and the acceptor impurity states will have very few electrons in them. Photons energetic enough to take an electron from the valence band to these impurity levels will result in hole carriers in the valence bands. The deep acceptor levels for these impurities are above the top of the valence band by 0.16 eV for Mn, 0.22eV for Ni, 0.25 eV for Co and 0.35 eV for Fe. The thresholds observed for photoconductivity in these p-type Ge samples are shown in Fig. 8.7 and the experimental results are in good agreement with this interpretation. The large increase in photoconductivity at 0.7eV corresponds to an interband transition and the threshold for this process is independent of the impurity species.

The excess carrier lifetime can be measured by using light pulses and observing the decay in the photocurrent through measurement of the voltage across a calibrated load resistor R in the external circuit as shown in Fig. 8.8. Using a light chopper, light pulses can be generated as indicated in Fig. 8.9. For each light pulse, the carrier density will build

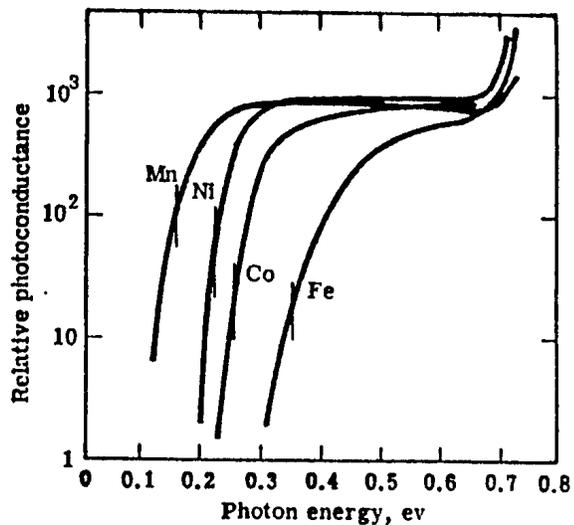Figure 8.7: Photoconductance spectrum in bulk Ge with various dopants.

Figure 8.8: Schematic of a circuit used to measure the excess carrier lifetime through decay in the photocurrent.
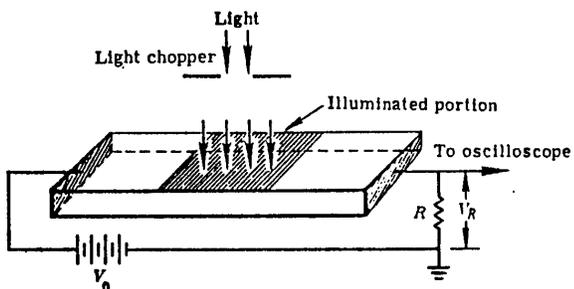

Figure 8.9: Schematic experimental time dependence of light pulses and of the corresponding photoconductivity signal.
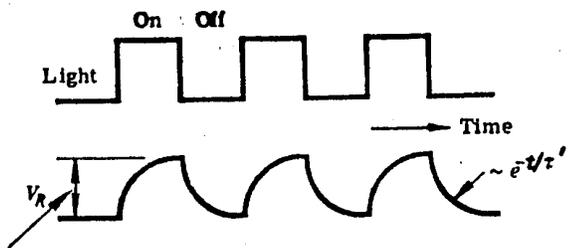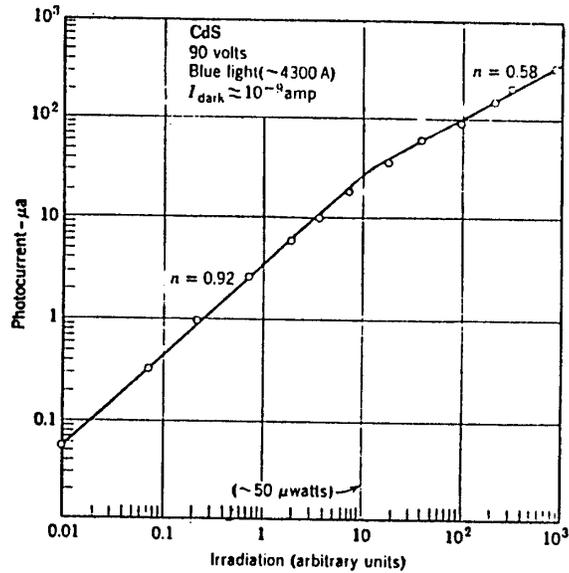
Figure 8.10: Experimental dependence of the photocurrent on light irradiation for CdS. A linear response is observed for low light levels.

up and then decay exponentially with a characteristic time equal to the lifetime τ' of the excess carriers.

In the interpretation of these experiments corrections must be made for surface recombination. To study a given material, the pulse repetition rate is adjusted to match approximately the excess carriers decay lifetime. For long lifetimes ($\sim 10^{-3}$ sec), a mechanical chopper arrangement is appropriate. On the other hand, for short lifetimes a spark source can be used to give light pulse of $\sim 10^{-8}$ sec duration. For extremely short lifetimes, lasers with pulses well below $\sim 10^{-10}$ seconds are available.

To get an idea of the magnitude involved in the photoconduction process, we show in Fig. 8.10 some data for CdS, a common photoconductor. This plot of photoconductive response versus illumination level shows that the photocurrent is almost a linear function of the illumination intensity for low intensities but is non-linear at high illumination levels. The **dark current** refers to the background current that flows in the absence of incident light. Thus, the Fig. 8.10 shows that an incident power as small as 5×10^{-8} watts results in a photocurrent 50 times greater than the dark current.

107

Chapter 10

Optical Study of Lattice Vibrations

References:

- Kittel, ISSP 6th edition, Ch 10.

- Ashcroft and Mermin, Ch 27.

- Yu and Cardona, Fundamentals of Semiconductors, pp. 251-258

10.1 Lattice Vibrations in Semiconductors

10.1.1 General Considerations

The lattice vibrations in semiconductors are described in terms of $3N$ branches for the phonon dispersion relations where N is the number of atoms per primitive unit cell. Three of these branches are the acoustic branches, and the remaining $3N-3$ are the optical branches. The optical lattice modes at $\vec{q} = 0$ are sensitively studied by infrared spectroscopy (optical reflectivity or transmission) for odd parity modes, including those for which the normal mode vibrations involve a dipole moment. Raman spectroscopy provides a complementary tool to infrared spectroscopy, insofar as Raman spectroscopy is sensitive to even parity modes. Since the group IV semiconductors have inversion symmetry, the optical phonon branch is Raman active but is not seen in infrared spectroscopy. The III–V compound semiconductors, however, do not have inversion symmetry, so that the optical modes for semiconductors such as GaAs are both infrared-active and Raman-active. A schematic optical absorption curve to a semiconductor is shown in Fig. 10.1.

Since the wavevector for light is very much smaller than the Brillouin zone dimensions, conservation of momentum requires the wave vector for the phonon $\vec{q}_{\mathrm{phonon}}$ that is created or absorbed to be much smaller than Brillouin zone dimensions, so that the wave vectors for phonons that are observed in first order infrared or Raman processes are close to $\vec{q} = 0$. Since thermal neutrons can have a wide range of momentum values, neutron spectroscopy using thermal neutrons as a probe allows exploration of the phonon branches over a wide range of $\vec{q}_{\mathrm{phonon}}$. Since heat in a semiconductor is dominantly carried by the acoustic phonons, information about the acoustic phonons is also provided by thermal conductivity studies.
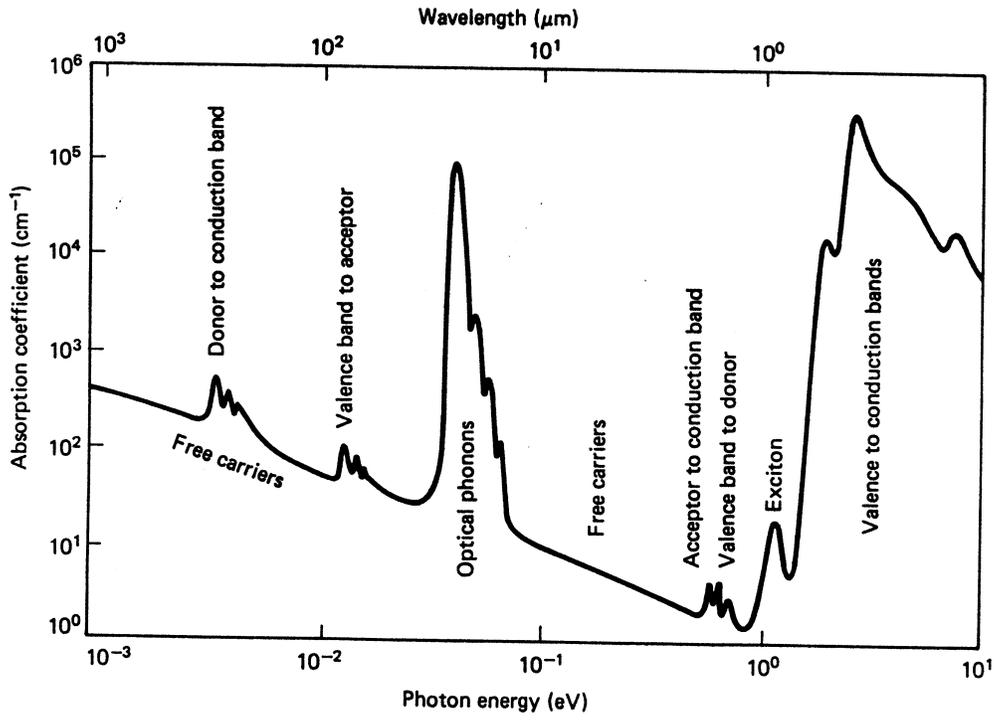
Figure 10.1: Hypothetical absorption spectrum for a typical III-V semiconductor as a function of phonon energy.

We now review the interaction of the electromagnetic field with an oscillating dipole due to a lattice vibration. Crystals composed of two different atomic species (like NaCl) can have vibrating ions at finite temperatures. When these ions are vibrating in an optic mode $\leftarrow \oplus \rightarrow\leftarrow \ominus \rightarrow$, a vibrating dipole is created and this dipole can interact with the electromagnetic field. In discussing this interaction, we wish to focus attention on the following points which are discussed more fully in the text below:

1. The existence of two characteristic frequencies for the vibrations in a solid in the presence of light:

   - ω_t = transverse optical phonon frequency (TO)

   - ω_ℓ = longitudinal optical phonon frequency (LO)

   The description of the LO and TO phonons is provided by the polariton model which accounts for the interaction between light and phonon excitations. Because of the very small wavevector of the incident photons, the phonons which are optically excited will also have very small wavevectors. Therefore, ω_t and ω_ℓ are taken as the phonon frequencies at $\vec{q} = 0$ for the TO and LO phonon dispersion curves.

2. These two frequencies are observable experimentally either through an infrared absorption, transmission, or reflection experiment (infrared activity) or through a scattering experiment (Raman activity). A transparent dielectric becomes lossy as ω

109

increases above ω_t. The transverse optical phonon frequency ω_t corresponds to a resonance in the dielectric function

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\text{const}}{\omega_t^2 - \omega^2} \tag{10.1}$$

where ε_∞ is the high frequency dielectric function (appropriate to electronic excitation processes) and a resonance in $\varepsilon(\omega)$ occurs at the TO phonon frequency $\omega = \omega_t$. The strong frequency dependence of the dielectric function (large dispersion) near ω_t is exploited in designing prisms for monochromators. The frequency ω_t is also called the *reststrahl* frequency.

3. The frequency ω_ℓ is the frequency at which the real part of the dielectric function vanishes $\varepsilon_1(\omega_\ell) = 0$. It will be shown below that ω_ℓ is the longitudinal optical phonon frequency corresponding to $\vec{q} = 0$ (zero wave vector). By group theory, it can be shown that the lattice modes at $\vec{q} = 0$ for a cubic crystal are three-fold degenerate. This degeneracy is lifted by the electromagnetic interaction in polar materials to give a splitting between the LO and TO modes. An example of the reflectivity of a normally transparent material in the region where phonon excitation processes dominate is shown in Fig. 10.2. From the diagram, we see that for $\omega_t < \omega < \omega_\ell$, the dielectric is both highly reflective and lossy. This range between ω_t and ω_ℓ is also observed as an absorption band in infrared absorption studies.

4. The dielectric function $\varepsilon(\omega)$ approaches the static dielectric constant ε_0 as $\omega \to 0$. Also, $\varepsilon(\omega)$ approaches the high frequency dielectric function ε_∞ as ω approaches frequencies that are large compared with ω_t and ω_ℓ. Even when we consider ω to be large, we are still thinking of ω as being very much smaller than typical interband electronic frequencies. Lattice modes typically are important in the wavelength range $10 \leq \lambda \leq 100\mu$m or $0.01 \leq \hbar\omega \leq 0.1$eV or $50 \leq \omega \leq 1000$ cm^{-1}.

5. The quantities ε_0, ε_∞, ω_t and ω_ℓ are not independent, but are related by a very general relation called the Lyddane–Sachs–Teller relation:

$$\frac{\omega_\ell^2}{\omega_t^2} = \frac{\varepsilon_0}{\varepsilon_\infty} \tag{10.2}$$

which is written here for a crystal with two atoms/unit cell.

10.2 Dielectric Constant and Polarizability

The polarizability α of an atom is defined in terms of the local electric field at the atom,

$$p = \alpha E_{\text{local}}. \tag{10.3}$$

The polarizability is an atomic property, the the dielectric constant will depend on the manner in which the atoms are assembled to form a crystal. For a non-spherical atom α will be a tensor. The polarization of a crystal may be approximated as the product of the polarizabilities of the atoms times the local electric fields,

$$P = \sum_j N_j p_j = \sum_j N_j \alpha_j E_{\text{local}}(j), \tag{10.4}$$
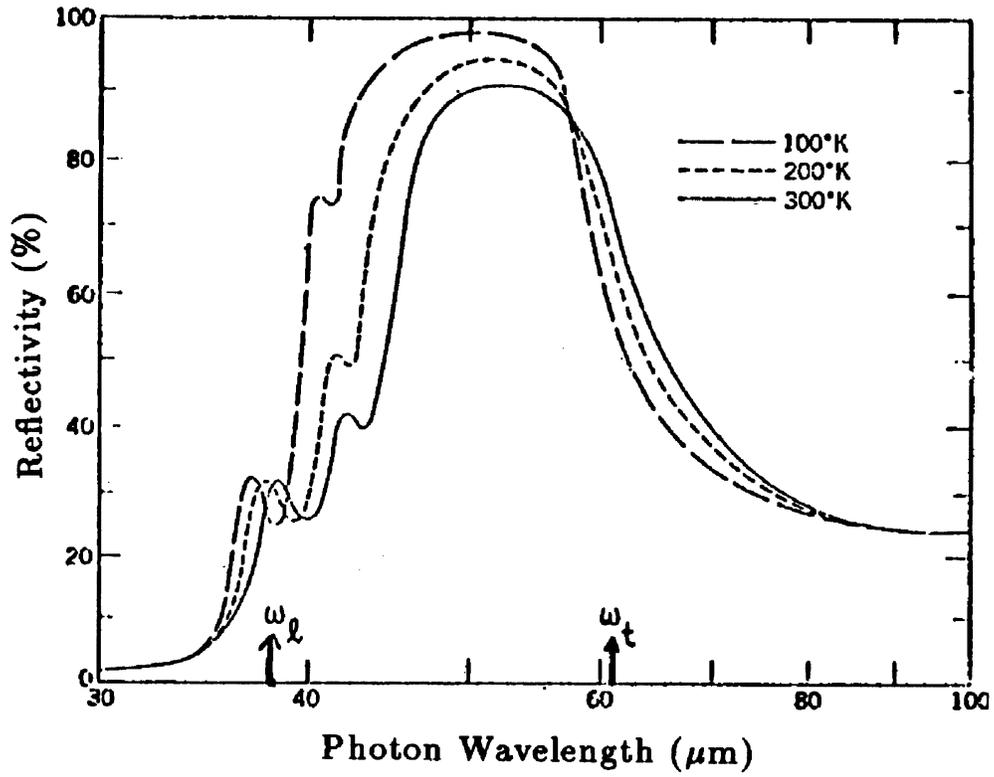
110

Figure 10.2: Reflectivity of a thick crystal of NaCl vs. wave length at several temperatures. The nominal values of ω_ℓ and ω_t at room temperature correspond to wavelengths of 38 and 61 microns, respectively. The additional structure seen in the reflectivity spectrum near ω_ℓ is associated with defects.

where N_j is the concentration and α_j the polarizability of atoms or ions j, and $E_{\text{local}}(j)$ is the local field at atomic sites j. If the local field is given by the Lorentz relation, then

$$P = \left(\sum_j N_j \alpha_j\right)\left(E + \frac{4\pi}{3}P\right). \tag{10.5}$$

Solving for the susceptibility

$$\chi = \frac{P}{E} = \frac{\sum_j N_j \alpha_j}{1 - \frac{4\pi}{3}\sum_j N_j \alpha_j}. \tag{10.6}$$

Using the definition $\epsilon = 1 + 4\pi\chi$ one obtains the Clausius–Mossotti relation

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi}{3}\sum_j N_j \alpha_j. \tag{10.7}$$

This relates the dielectric constant to the electronic polarizability, but only for crystal structures for which the Lorentz local field relation applies.

10.3 Polariton Dispersion Relations

The statements 1–5 in §10.1 provide an overview on optical studies of lattice modes. In this section we discuss the polariton dispersion relations which describe the interaction of light with the electric dipole moment associated with infrared absorption, and the LO–TO splitting of the normal mode vibration of the atoms in the solid arising from these dispersion relations.

Consider the equation of motion of an ion in a solid using the normal mode coordinate $\vec{r}$, so that harmonic motion yields

$$m\ddot{\vec{r}} = -\kappa\vec{r} + e\vec{E} = -m\omega^2\vec{r} \tag{10.8}$$

where

$$\vec{E} = \vec{E}_0 e^{-i\omega t} \tag{10.9}$$

and $-\kappa\vec{r}$ represents a lattice restoring force while $e\vec{E}$ is the force due to the actual electric field $\vec{E}$ at an ion site. Maxwell's equations give us

$$\text{curl } \vec{H} = \frac{1}{c}\dot{\vec{D}} = \frac{1}{c}(\dot{\vec{E}} + 4\pi\dot{\vec{P}}) = -\frac{i\omega}{c}(\vec{E} + 4\pi\vec{P}) \tag{10.10}$$

$$\text{curl } \vec{E} = -\frac{1}{c}\dot{\vec{H}} = \frac{i\omega}{c}\vec{H} \tag{10.11}$$

We also have a constitutive equation which tells us that the total polarization arises from an ionic contribution $N'e\vec{r}$ where N' is the number of optical modes per unit volume and from an electronic contribution $n\alpha\vec{E}$, where n is the electron concentration and α is the electronic polarizability:

$$\vec{P} = N'e\vec{r} + n\alpha\vec{E}. \tag{10.12}$$

Equations 10.8, 10.10, 10.11 and 10.12 represent 4 equations in the 4 variables $\vec{E}, \vec{H}, \vec{r}$, and $\vec{P}$.

112

In writing Eq. 10.12 for the polarization vector $\vec{P}$, we have considered two degrees of freedom: the ion system and the electron system. We further assume that these polarizations are accomplished independently. In formulating this calculation, the electric field in all equations is the applied electric field, since it is assumed that the lattice polarization effects are weak. In more sophisticated treatments, we must consider the effect of local field corrections when the dielectric function is large, as occurs for example in the case of ferroelectrics.

We now seek plane wave solutions for transverse wave propagation: $(\vec{E}, \vec{H})$ in the xy plane and perpendicular to the Poynting vector, $\vec{S} = [c/(8\pi)]\mathrm{Re}(\vec{E}^* \times \vec{H})$, and the Poynting vector is taken along the z direction

$$E_x = E_x^0 e^{-i(\omega t - Kz)} \tag{10.13}$$

$$H_y = H_y^0 e^{-i(\omega t - Kz)} \tag{10.14}$$

$$P_x = P_x^0 e^{-i(\omega t - Kz)} \tag{10.15}$$

$$r_x = r_x^0 e^{-i(\omega t - Kz)}. \tag{10.16}$$

Here K is the wave vector for the light, $K = 2\pi/\lambda$. Using values for λ typical for lattice modes in NaCl, we have $\lambda \sim 60\mu\mathrm{m}$ and $K \sim 10^3 \mathrm{cm}^{-1}$. Substitution of the harmonic solutions in Eqs. 10.13–10.16 into the 4 equations (Eqs. 10.8, 10.10, 10.11 and 10.12) for the four variables $\vec{E}$, $\vec{H}$, $\vec{r}$, and $\vec{P}$ yields:

$$iKH_y - \frac{i\omega}{c}E_x - \frac{4\pi i\omega}{c}P_x = 0 \tag{10.17}$$

$$-iKE_x + \frac{i\omega}{c}H_y = 0 \tag{10.18}$$

$$-\omega^2 r_x + \frac{\kappa}{m}r_x - \frac{e}{m}E_x = 0 \tag{10.19}$$

$$P_x - N'e r_x - n\alpha E_x = 0. \tag{10.20}$$

Equations 10.17–10.20 form 4 equations in 4 unknowns. To have a non-trivial solution to Eqs. 10.17–10.20, the coefficient determinant must vanish. We arrange the coefficient determinant following the order of the variables in Eqs. 10.13–10.16: $(E_x \quad H_y \quad P_x \quad r_x)$:

$$\begin{vmatrix} \omega/c & -K & 4\pi\omega/c & 0 \\ K & -\omega/c & 0 & 0 \\ e/m & 0 & 0 & \omega^2 - \kappa/m \\ -n\alpha & 0 & 1 & -N'e \end{vmatrix} = 0. \tag{10.21}$$

Multiplying out the determinant in Eq. 10.21, we get a quadratic equation in ω^2

$$\omega^4[1 + 4\pi n\alpha] - \omega^2\left[c^2 K^2 + \frac{\kappa}{m} + \frac{4\pi N'e^2}{m} + \frac{4\pi n\alpha\kappa}{m}\right] + K^2 c^2 \frac{\kappa}{m} = 0. \tag{10.22}$$

Equation 10.22 is more conveniently written in terms of the parameters ε_∞, ε_0, and ω_T where these parameters are defined in Eqs. 10.23, 10.25 and 10.27 given below:

1. The high frequency dielectric constant ε_∞ is written as $\varepsilon_\infty = 1 + 4\pi P_\infty / E$, and is the parameter normally used to express the optical core dielectric constant when discussing electronic processes studied by optical techniques. From the equation of motion (Eq. 10.8), we conclude that at high frequencies ($\omega \gg \omega_T$ and we show below that ω_T is the transverse optical frequency), the ionic displacement is small, for otherwise the acceleration would tend to ∞. Thus as the frequency increases, the ions contribute less and less to the polarization vector. We thus have the result $P_\infty = n\alpha E$, so that the electronic contribution dominates and

$$\varepsilon_\infty = 1 + 4\pi n\alpha. \tag{10.23}$$

2. The low frequency ($\omega \ll \omega_T$) dielectric constant is written as ε_0. At $\omega = 0$ the equation of motion Eq. 10.8 yields $\vec{r} = e\vec{E}/\kappa$ so that the polarization vector at zero frequency is

$$\vec{P_0} = [\frac{N'e^2}{\kappa} + n\alpha]\vec{E}; \tag{10.24}$$

and

$$\varepsilon_0 = 1 + 4\pi[\frac{N'e^2}{\kappa} + n\alpha]. \tag{10.25}$$

At a general frequency ω, we must from Eqs. 10.8 and 10.12 write

$$\varepsilon(\omega) = 1 + 4\pi[\frac{N'e^2}{\kappa - m\omega^2} + n\alpha]. \tag{10.26}$$

3. Finally, we introduce a frequency ω_T defined as

$$\omega_T^2 \equiv \frac{\kappa}{m} \tag{10.27}$$

which depends only on the restoring forces and not on the externally applied field. Of course, these restoring forces will depend on internal fields, since electromagnetic interactions are responsible for producing these lattice vibrations in the first place. We will later identify ω_T with ω_t, the transverse optical phonon frequency. Substitution of $\varepsilon_\infty, \varepsilon_0$, and ω_T into Eq. 10.22 yields the polarization dispersion relation

$$\omega^4 \varepsilon_\infty - \omega^2[c^2 K^2 + \omega_T^2 \varepsilon_0] + \omega_T^2 c^2 K^2 = 0. \tag{10.28}$$

Equation 10.28 has two solutions

$$\omega^2 = \frac{1}{2\varepsilon_\infty}(\omega_T^2 \varepsilon_0 + c^2 K^2) \pm \left(\frac{1}{4\varepsilon_\infty^2}(\omega_T^2 \varepsilon_0 + c^2 K^2)^2 - \omega_T^2 K^2 \frac{c^2}{\varepsilon_\infty}\right)^{1/2} \tag{10.29}$$

which are shown graphically in Fig. 10.3. Each solution in Eq. 10.29 is twofold degenerate, since $\vec{E}$ can be chosen in any arbitrary direction perpendicular to the propagation vector. The coupled excitation of the transverse optical phonon to the electromagnetic radiation is called the **polariton** and the picture in Fig. 10.3 is called the **polariton** dispersion relation. There is also a longitudinal direction for both the light and the lattice vibrations; for this case there is no coupling between the light and the phonons
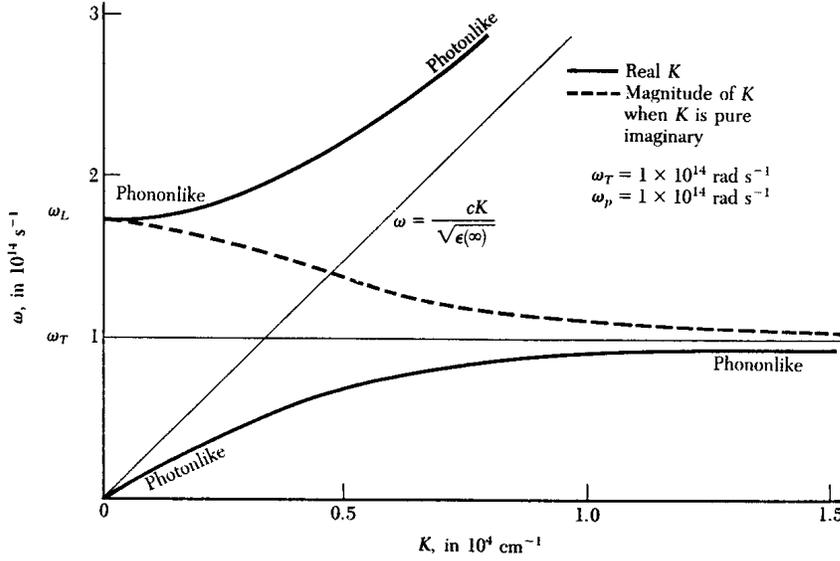
Figure 10.3: Polariton dispersion relations showing the coupling between the transverse lattice vibrations and the electromagnetic radiation. In this figure, we clearly see the splitting of the LO and TO modes ($\omega_L - \omega_T$) induced by the ionicity of the solid.

and the frequency is the same as in the absence of light. We therefore obtain a total of 6 modes for the 3 coupled optical lattice modes and the three electromagnetic modes (two transverse modes representing photons and one longitudinal mode). It is of interest to examine the solutions of Eq. 10.29 for small and large K vectors where we must remember that the scale of the K–vectors for light is a scale of $10^3 - 10^4$ cm^{-1} rather than 10^8cm^{-1} which describes the Brillouin zone dimension, corresponding to Brillouin zone dimensions. Thus the whole picture shown in Fig. 10.3 occurs essentially at $\vec{q} = 0$ when plotting phonon dispersion relations $\omega_q(\vec{q})$ for wave vectors $\vec{q}$ in the Brillouin zone.

At small K vectors ($|K| \ll 10^4$ cm^{-1}), we have two solutions to Eq. 10.29. The <u>positive solution</u> is given by

$$\omega^2 = \frac{1}{\varepsilon_\infty}(\omega_T^2 \varepsilon_0 + c^2 K^2) \tag{10.30}$$

which gives

$$\omega_T^2 \varepsilon_0 / \varepsilon_\infty \equiv \omega_L^2, \tag{10.31}$$

thus defining the frequency ω_L. In writing this solution we neglected the term $c^2 K^2$ as $K \to 0$. This solution corresponds to the phonon branch with finite frequency at $K = 0$ and hence is an optical phonon mode. We will call this frequency ω_L and later we will identify ω_L with the longitudinal optical phonon mode frequency, ω_ℓ. We shall see that the above definition is equivalent to taking frequency ω_ℓ as the frequency where the real part of the dielectric function vanishes $\varepsilon_1(\omega_\ell) \equiv 0$. We also remember, that the longitudinal optical (LO) phonon does not interact with the electromagnetic field. For a phonon–electromagnetic interaction, we require that the electric field be transverse to the direction of propagation.

115

With regard to the *negative solution* of Eq. 10.29, we expand the square root term in Eq. 10.29 to obtain:

$$\omega^2 = \frac{\omega_T^2 K^2 c^2}{\omega_T^2 \varepsilon_0 + c^2 K^2} \tag{10.32}$$

or

$$\omega^2 \simeq \frac{c^2 K^2}{\varepsilon_0} \tag{10.33}$$

yielding the photon–like mode with a linear K dependence

$$\omega = \frac{cK}{\sqrt{\varepsilon_0}} \quad \text{for } \omega \ll \omega_T. \tag{10.34}$$

At large K values ($|K| \sim 10^5 \mathrm{cm}^{-1}$), we solve the quadratic equation given by Eq. 10.29 in the large K limit and obtain positive and negative solutions. Using a binomial expansion for Eq. 10.29, we obtain the following positive and negative solutions. For the positive solution, i.e., K large, we obtain

$$\omega^2 \simeq \frac{1}{\varepsilon_\infty}(\omega_T^2 \varepsilon_0 + c^2 K^2) = \frac{c^2 K^2}{\varepsilon_\infty}. \tag{10.35}$$

This is clearly the photon-like mode, since

$$\omega = \frac{cK}{\sqrt{\varepsilon_\infty}} \quad \text{for } \omega \gg \omega_T. \tag{10.36}$$

This result is almost identical to Eq. 10.34 obtained in the low K limit, except that now we have ε_∞ instead of ε_0. Correspondingly, the phonon-like mode for large K arises from the negative solution:

$$\omega^2 \simeq \frac{\omega_T^2 K^2 c^2}{\omega_T^2 \varepsilon_0 + c^2 K^2} \simeq \omega_T^2. \tag{10.37}$$

We have thus introduced two frequencies: ω_T and ω_L and from the definition of ω_L we obtain the Lyddane–Sachs–Teller relation

$$\frac{\omega_\ell^2}{\omega_T^2} = \frac{\varepsilon_0}{\varepsilon_\infty}. \tag{10.38}$$

Now, ω_T and ω_L have well-defined meanings with regard to the dielectric function as can be seen in Fig. 10.3. From Eq. 10.12, we have for the polarization due to ions and electrons:

$$\vec{P} = N'e\vec{r} + n\alpha\vec{E} \tag{10.39}$$

while the equation of motion, Eq. 10.8, ($F = ma$) gives

$$-m\omega^2 \vec{r} = -\kappa\vec{r} + e\vec{E} \tag{10.40}$$

yielding

$$\vec{r} = \frac{e\vec{E}}{\kappa - m\omega^2} = \frac{e\vec{E}/m}{\omega_T^2 - \omega^2} \tag{10.41}$$

so that

$$\frac{P}{E} = \frac{\varepsilon(\omega) - 1}{4\pi} = \frac{N'e^2/m}{\omega_T^2 - \omega^2} + \frac{\varepsilon_\infty - 1}{4\pi}, \tag{10.42}$$

116

since the electronic polarizability term is $n\alpha = (\varepsilon_\infty - 1)/4\pi$. We therefore obtain:

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{4\pi N'e^2/m}{\omega_T^2 - \omega^2} \tag{10.43}$$

where ε_∞ represents the contribution from the electronic polarizability and the resonant term represents the lattice contribution. Neglecting damping, we have the result $|\varepsilon(\omega)| \to \infty$ as $\omega \to \omega_T$, where the transverse optical phonon frequency $\omega = \omega_T$ is interpreted as the frequency at which the dielectric function $\varepsilon(\omega)$ is resonant. The name *reststrahl* frequency denotes that frequency ω_T where light is maximally absorbed by the medium.

We would now like to get a more physical idea about ω_ℓ. So far ω_ℓ has been introduced as the phonon mode of the polariton curve in Fig. 10.3 near $k = 0$. From Eq. 10.43 we have the relation

$$\varepsilon_0 = \varepsilon_\infty + \frac{4\pi N'e^2}{m\omega_T^2} \tag{10.44}$$

where ε_0 is defined by $\varepsilon_0 \equiv \varepsilon(\omega = 0)$, so that

$$\frac{4\pi N'e^2}{m} = \omega_T^2(\varepsilon_0 - \varepsilon_\infty) \tag{10.45}$$

and $\omega_t = \omega_T$ is the frequency where $\varepsilon(\omega)$ is resonant. Thus from Eqs. 10.1 and 10.43, we can write

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{(\varepsilon_0 - \varepsilon_\infty)}{(1 - \omega^2/\omega_t^2)} = \varepsilon_\infty + \frac{(\varepsilon_0 - \varepsilon_\infty)}{(1 - \omega^2/\omega_T^2)} \tag{10.46}$$

so that $\omega_T = \omega_t$. We define ω_ℓ as the frequency at which the dielectric function vanishes $\varepsilon(\omega_\ell) \equiv 0$ so that setting $\varepsilon(\omega) = 0$ in Eq. 10.46 yields

$$\varepsilon_\infty = \frac{(\varepsilon_\infty - \varepsilon_0)}{(1 - \omega_\ell^2/\omega_t^2)} \tag{10.47}$$

or

$$\frac{\omega_\ell^2}{\omega_t^2} = \frac{\varepsilon_0}{\varepsilon_\infty}. \tag{10.48}$$

Thus, the frequency ω_ℓ, which yields a zero in the dielectric function, also satisfies the Lyddane-Sachs-Teller relation (Eq. 10.48).

We illustrate the properties of ω_ℓ and ω_t in Fig. 10.4 where we see that the frequency dependence of the dielectric function $\varepsilon(\omega)$ has two special features:

- a zero of $\varepsilon(\omega)$ occurring at ω_ℓ

- an infinity or pole of $\varepsilon(\omega)$ occurring at ω_t.

For $\omega_t < \omega < \omega_\ell$, the dielectric function $\varepsilon(\omega)$ is negative, so that losses must occur and transmission is consequently poor. The frequency difference between the two characteristic frequencies ω_ℓ and ω_t depends on the ionicity of the crystal. Thus, predominantly covalent materials like InSb which have weak ionicity have a smaller $\omega_\ell - \omega_t$ splitting than alkali halide crystals which are highly ionic. For weakly polar materials like InSb, the treatment of the electric field given here is adequate. For highly polar materials, one must also consider the local fields, as distinct from the applied field. These local fields tend to increase the
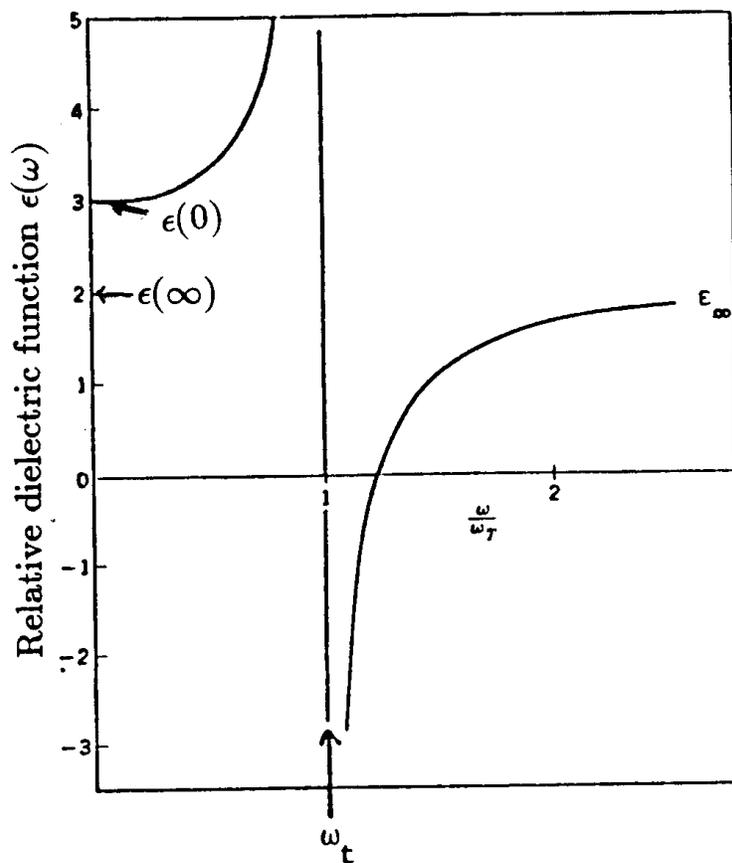
Figure 10.4: The dielectric function $\varepsilon(\omega)$ plotted as a function of normalized frequency ω/ω_T. When damping is included, the real part of the dielectric function remains finite at ω_T.

118

separation between ω_t and ω_ℓ, pulling ω_t to low frequencies. Since mechanically hard materials tend to have high Debye temperatures and high phonon frequencies, the passage of ω_t toward zero for ferroelectric materials (extremely high dielectric function and capable of spontaneous polarization) is referred to as the appearance of a "soft mode".

The Lyddane-Sachs-Teller relation is more general than the derivation given here would imply. This relation can be extended to cover anisotropic materials with any number of optical modes. In this context we can write the frequency dependence of the symmetrized dielectric tensor function associated with symmetry μ as

$$\varepsilon_\mu(\omega) = \varepsilon_\mu(\infty) + \sum_{j=1}^{p} \frac{f_{\mu,j}\omega_{T,j}^2}{\omega_{T,j}^2 - \omega^2 - i\gamma_j\omega} \tag{10.49}$$

where $f_{\mu,j}$ is the oscillator strength, γ_j is the damping of mode j, and p is the number of modes with symmetry μ. An example where this would apply is the case of tetragonal symmetry where μ could refer to the in-plane modes (E_u symmetry) or to the out-of-plane modes (A_{2u} symmetry). Figure 10.5 shows the measured reflectivity for the lattice modes of TeO$_2$ which has 4 formula units per unit cell (12 atoms/unit cell) can be described by a model based on Eq. 10.49 for polarization of the electromagnetic field parallel and perpendicular to the tetragonal axis.

Setting the damping terms in Eq. 10.49 to zero, $\gamma_j = 0$, we obtain the result

$$\frac{\varepsilon(\omega)}{\varepsilon(\infty)} = \prod_{j=1}^{p} \left(\frac{\omega_{l,j}^2 - \omega^2}{\omega_{t,j}^2 - \omega^2} \right) \tag{10.50}$$

which leads to the generalized Lyddane–Sachs–Teller relation

$$\frac{\varepsilon_0}{\varepsilon_\infty} = \frac{\varepsilon(0)}{\varepsilon(\infty)} = \prod_{j=1}^{p} \left(\frac{\omega_{l,j}^2}{\omega_{t,j}^2} \right). \tag{10.51}$$

Equation 10.51 can be generalized for anisotropic crystals by writing Eq. 10.50 for each component, keeping in mind that the optical selection rules differ for each component. The dependence of the reflectivity on polarization and on temperature is illustrated for the tetragonal crystal TeO$_2$ in Fig. 10.5.

To find the LO and TO modes associated with Eq. 10.49, we would look for zeros and poles of the dielectric function for a general direction of light propagation. For example, in a tetragonal crystal we can write

$$\varepsilon(\theta, \omega) = \frac{\varepsilon_\|(\omega)\varepsilon_\perp(\omega)}{\varepsilon_\|(\omega)\cos^2\theta + \varepsilon_\perp(\omega)\sin^2\theta}. \tag{10.52}$$

The observation of LO and TO phonon frequencies by optical measurements is made using two basically different techniques. In one approach, we make absorption, reflection or transmission measurements, while in the other approach, light scattering measurements are made. These are often complementary methods for the following reason. Many important crystals have inversion symmetry (e.g., the NaCl structure). In this case, the phonon modes are purely odd or purely even. If the odd parity modes have dipole moments and couple directly to the electromagnetic fields, then they are infrared active. On the other hand,

119

Figure 10.5: Reflectivity in paratel-
luride, TeO$_2$, for (a) $\vec{E}$ parallel and
(b,c) perpendicular to the tetragonal
axis at 295 K (b). The polarization
$\vec{E} \parallel$ the tetragonal axis has only the
A_{2u} modes allowed whereas for $\vec{E} \perp$ the
tetragonal axis has only the E_{2u} modes
allowed. The points are experimental
and the solid line is a model based on
Eq. 10.49. (After Korn, et al., Phys.
Rev. $\underline{B8}$, 768 (1973).)
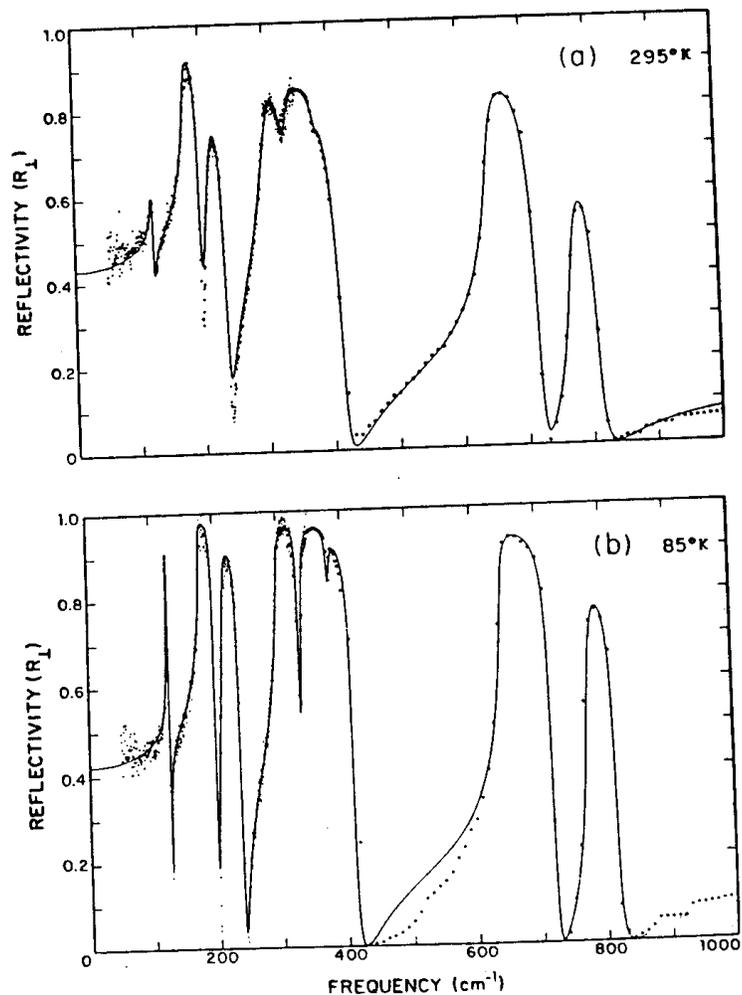
the even parity modes are not infrared active but instead may be Raman active and can be observed in a light scattering experiment. Thus, by doing both infrared absorption and Raman scattering measurements we can find both even and odd parity optical phonon modes, except for the silent modes which because of other symmetry requirements are neither infrared nor Raman active. These concepts are discussed in detail in the group theory course.

In modeling the phonon and free carrier contributions to the dielectric function it can happen that these phenomena occur over a common frequency range. In this case, we write the complex dielectric function for an isotropic semiconductor as follows in analyzing optical data

$$\frac{\varepsilon_\mu(\omega)}{\varepsilon_\mu(\infty)} = \left(1 - \frac{\omega_p^2}{\omega(\omega + i\gamma_p)}\right) + \sum_{j=1}^{p} \frac{\omega_{L,j}^2 - \omega_{T,j}^2}{\omega_{T,j}^2 - \omega^2 - i\omega\gamma_j} \tag{10.53}$$

where the first and second terms are, respectively, the free carrier and the infrared-active phonon contributions to the dielectric function. In Eq. 10.53, ω_p is the screened electronic plasma frequency ($\omega_p^2 = 4\pi n e^2/m^*\varepsilon(\infty)$, and $\varepsilon(\infty)$ is the core dielectric constant used to approximate the higher frequency electronic polarizability). The phonon contribution to Eq. 10.53 depends on $\omega_{L,j}$ and $\omega_{T,j}$ which are the j–th longitudinal and transverse optic mode frequencies, while γ_j and γ_p are the phonon and plasma damping factors, respectively.

The model given by Eq. 10.53 can, for example, be used to model the optical properties of the anisotropic compound La_2CuO_4 which becomes a high T_c superconductor, upon addition of a small concentration of Sr. In this case it is important to obtain polarized reflectivity measurements on oriented single crystals, and to carry out the Kramers–Kronig analysis of reflectivity data for each of the polarization components separately.

10.4 Light Scattering

Light scattering techniques provide an exceedingly useful tool to study fundamental excitations in solids, such as phonons, because light can be scattered from solids inelastically, whereby the incident and scattered photons have different frequencies. Inelastic light scattering became an important tool for the study of excitations in solids in the mid-1960's with the advent of laser light sources, because the inelastically scattered light is typically only $\sim 10^{-7}$ of the intensity of the incident light.

In the light scattering experiments shown schematically in Fig. 10.6, conservation of energy gives:

$$\omega = \omega_0 \pm \omega_q \tag{10.54}$$

and conservation of momentum gives:

$$\vec{K} = \vec{K}_0 \pm \vec{q} \tag{10.55}$$

where the "0" subscript refers to the incident light, $\vec{K}$ refers to the wave vector of the light and "$\vec{q}$" refers to the wave vector for the excitation in the solid. Since $K_0 = 2\pi/\lambda$ is very small compared with the Brillouin zone dimensions, measurement of the angular dependence of $\omega_q(\vec{q})$ can then be used to provide dispersion relations for the excitations near $\vec{q} = 0$. If $\omega_q \ll \omega_0$, then $|\vec{K}| \simeq |\vec{K}_0|$, and we have $|q| \simeq 2|\vec{K}_0|\sin(\theta/2)$ so that $|q_{max}| = 2K_0$.
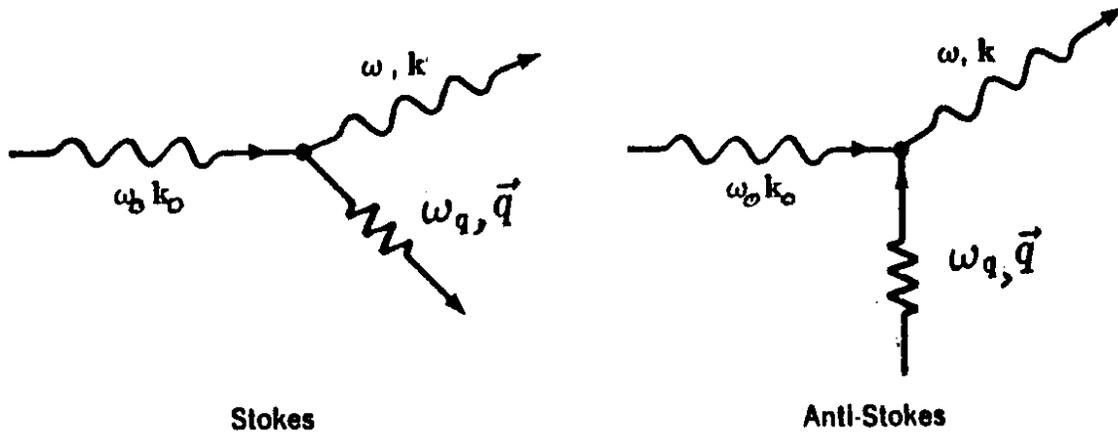
Stokes　　　　　　　　　　　　　　**Anti-Stokes**

Figure 10.6: Raman scattering of a photon showing both phonon emission (Stokes) and absorption (anti–Stokes) processes. The scattering process is called Brillouin scattering when an acoustic phonon is involved and polariton (Raman) scattering when an optical phonon is involved. Similar processes occur with magnons, plasmons or any other excitation of the solid with the correct symmetry.

If the excitation is an **acoustic** phonon, the inelastic light scattering process is called **Brillouin scattering**, while light scattering by **optical** phonons is called **Raman scattering**. Raman and Brillouin scattering also denote light scattering processes due to other elementary excitations in solids.

The light scattering can be understood on the basis of classical electromagnetic theory. When an electric field $\vec{E}$ is applied to a solid, a polarization $\vec{P}$ results

$$\vec{P} = \overset{\leftrightarrow}{\alpha} \cdot \vec{E} \tag{10.56}$$

where $\overset{\leftrightarrow}{\alpha}$ is the polarizability tensor of the atom in the solid, indicating that the positive charge moves in one direction and the negative charge in the opposite direction under the influence of the applied field. In the light scattering experiments, the electric field is oscillating at an optical frequency ω_0

$$\vec{E} = \vec{E}_0 \sin \omega_0 t. \tag{10.57}$$

The lattice vibrations in the solid modulate the polarizability of the atoms themselves

$$\alpha = \alpha_0 + \alpha_1 \sin \omega_q t. \tag{10.58}$$

so that the polarization which is induced by the applied electric field is:

$$\vec{P} = \vec{E}_0 (\alpha_0 + \alpha_1 \sin \omega_q t) \sin \omega_0 t$$

$$= \vec{E}_0 \left[\alpha_0 \sin(\omega_0 t) + \tfrac{1}{2}\alpha_1 \cos(\omega_0 - \omega_q)t - \tfrac{1}{2}\alpha_1 \cos(\omega_0 + \omega_q)t \right]. \tag{10.59}$$

Thus we see in Fig. 10.7 that light will be scattered elastically at frequency ω_0 (Raleigh scattering) and also inelastically, being modulated downward by the natural vibration frequency ω_q of the atom (Stokes process) or upward by the same frequency ω_q (anti-Stokes

Figure 10.7: Schematic diagram of light scattering spectrum showing the central unshifted Rayleigh line, the upshifted anti-Stokes line (emission process), and the downshifted Stokes line (absorption process). The ratio of the Stokes to anti-Stokes can be used to estimate the temperature of the phonon system.
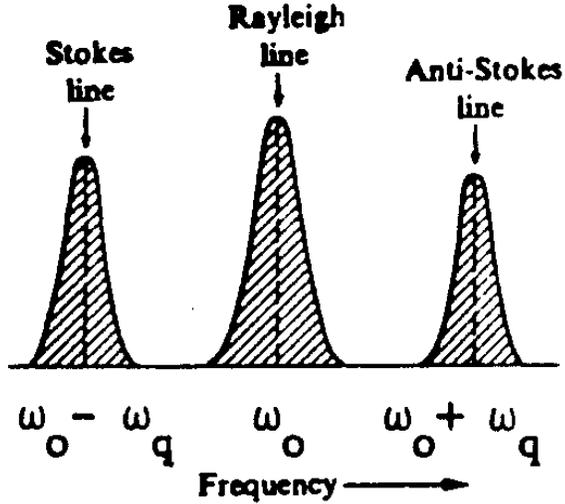
process). The light scattering process can also be viewed from a quantum mechanical perspective. If the "system" is initially in a state E'', then light scattering can excite the "system" to a higher energy state E' shown in Fig. 10.8a by absorption of an excitation energy $(E' - E'')$. Similarly, the "system" can initially be in a state E' and light scattering can serve to bring the system to a final state of lower energy E'' by emission of an excitation of energy $(E' - E'')$ as shown in Fig. 10.8b. The matrix element of the polarization vector between initial and final states is written (when expressed in terms of quantum mechanics) as

$$\vec{P}_{nm} = \int \Psi_n^* \vec{P} \Psi_m d^3r = \vec{E} \cdot \int \Psi_n^* \overset{\leftrightarrow}{\alpha} \Psi_m d^3r \tag{10.60}$$

where the polarizability $\overset{\leftrightarrow}{\alpha}$ is a second rank symmetrical tensor. The Stokes and anti-Stokes processes arise from consideration of the phase factors in this matrix element: Ψ_m has a phase factor $e^{-iE_m t/\hbar}$ while Ψ_n^* has a phase factor $e^{+iE_n t/\hbar}$. The polarizability tensor has a phase factor $e^{\pm i\omega_q t}$ so that the integration implied by Eq. 10.60 yields

$$E_m - E_n \pm \hbar\omega_q = 0. \tag{10.61}$$

We should remember that the optical absorption process is governed by the **momentum** matrix element which is a radial vector. Of particular significance is the case of a crystal with inversion symmetry whereby the momentum operator is an odd function, but the polarizability tensor is an even function. This characteristic feature has an important consequence; namely electronic absorption processes are sensitive to transitions between states of opposite parity (parity meaning even or odd), while light scattering is sensitive to transitions between states of similar parity. For this reason, light scattering and optical absorption are considered to be complementary spectroscopies, and together form basic tools for the study of the optical properties of elementary excitations in solids.

It is important to draw a clear distinction between Raman scattering and fluorescence. In Raman scattering, the intermediate states shown in Fig. 10.8a,b are "virtual" states and

123

$\hbar\omega_0$

$\hbar\omega_0 - \hbar\omega_q$

E'

$\hbar\omega_q$

E''

Stokes Process

$\hbar\omega_0$

$\hbar\omega_0 + \hbar\omega_q$

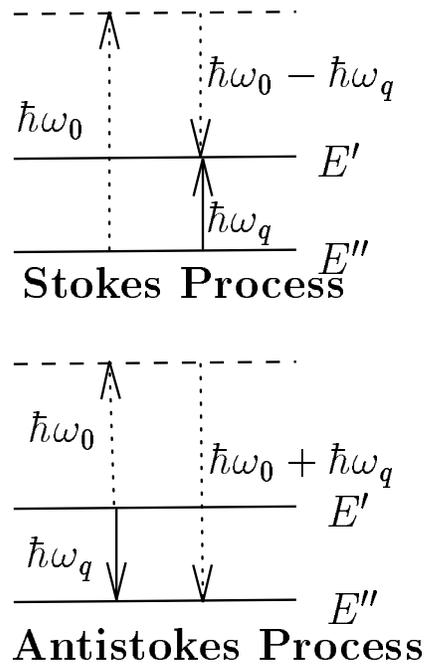E'

$\hbar\omega_q$

E''

Antistokes Process

Figure 10.8: Schematic energy level diagram for the (a) Stokes and (b) anti-Stokes processes. In this figure the solid lines denote real processes and the dashed lines virtual processes.

don't have to correspond to eigenstates of the physical "system"— any optical excitation frequency will in principle suffice. In fluorescence, on the other hand, the optical excitation state must be a real state of the system and in this case a real absorption of light occurs, followed by a real emission at a different frequency.

The major reason why these two processes are sometimes confused is that Raman scattering in solids often has a much higher intensity when $\hbar\omega_0$ is equal to an energy band gap and this effect is called resonant Raman scattering. In such cases, the fluorescent emission differs from the Raman process because fluorescent phenomena take a finite time to occur.

Typical Raman traces are shown in Fig. 10.9 for several III-V compound semiconductors. The laser wavelength is 1.06 μm (Nd:YAG laser) which is a photon energy below the band gap for each material. The scattered light is collected at 90° with respect to the incident light and both the LO and TO phonon modes at $\vec{q} = 0$ are observed. For the case of the group IV semiconductors there is no LO–TO splitting and only a single optical Raman-allowed mode is observed (at 519 cm^{-1} for Si). What is measured in Fig. 10.9 is the frequency shift between the incident and scattered light beams. For the range of phonon wave vectors where Raman scattering can be carried out, this technique is the most accurate method available for the measurement of the dispersion relations near the Brillouin zone center.

By doing the Raman scattering experiment with polarized light, it is possible to get information on the symmetry of the lattice vibrations by monitoring the polarization of both the incident and scattered radiation. This approach is important in the identification of phonon frequencies with specific lattice normal modes.

The inelastic neutron scattering technique, though less accurate than Raman scattering, has the advantage of providing information about phonons throughout the Brillouin zone. By using neutrons of low energy (thermal neutrons), it is possible to make the neutron wavelengths comparable to the lattice dimensions, in which case the inelastic scattering by
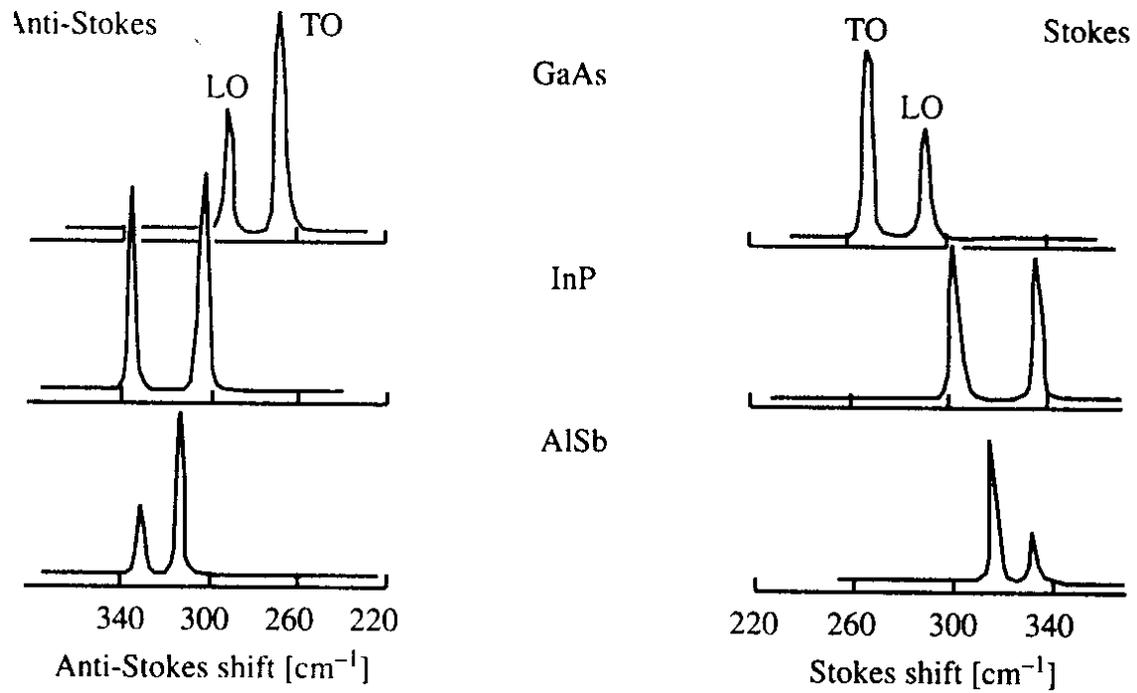
Figure 10.9: Raman spectra of three zinc-blende-type semiconductors showing the TO and LO phonons in both Stokes and anti-Stokes scattering.

a lattice vibration can cause a large momentum transfer to the neutron.

10.5   Feynman Diagrams for Light Scattering

Feynman diagrams are useful for keeping track of various processes that may occur in an inelastic scattering process that absorbs or creates an excitation. The basic notation used in drawing Feynman diagrams consists of propagators such as electrons, phonons or photons and vertices where interactions occur, as shown in Fig. 10.10(g).

The rules in drawing Feynman diagrams are:

- Excitations such as photon, phonons and electron-hole pairs in Raman scattering are represented by lines (or propagators) as shown in Fig. 10.10(g). These propagators can be labeled with properties of the excitations, such as their wavevectors, frequencies and polarizations.

- The interaction between two excitations is represented by an intersection of their propagators. This intersection is known as a *vertex* and is sometimes highlighted by a symbol such as a filled circle or empty rectangle.

- Propagators are drawn with an arrow to indicate whether they are created or annihilated in an interaction. Arrows pointing towards a vertex represent excitations which are annihilated. Those pointing away from the vertex are created.

- When several interactions are involved they are always assumed to proceed sequentially from the left to the right as a function of time.

- Once a diagram has been drawn for a certain process, other possible processes are derived by permuting the time order in which the vertices occur in this diagram.

The basic diagram for the Raman process is given in Fig. 10.10(a) taken from the Yu and Cardona book on "Fundamentals of Semiconductors." The other permutations of (a) obtained by different orders of the vertices are given in Figs. 10.10(b–f). We then use the Fermi Golden rule for each diagram, multiplying the contributions from each vertex. For example, the first vertex in Fig. 10.10(a) contributes a term to the scattering probability per unit time of the form
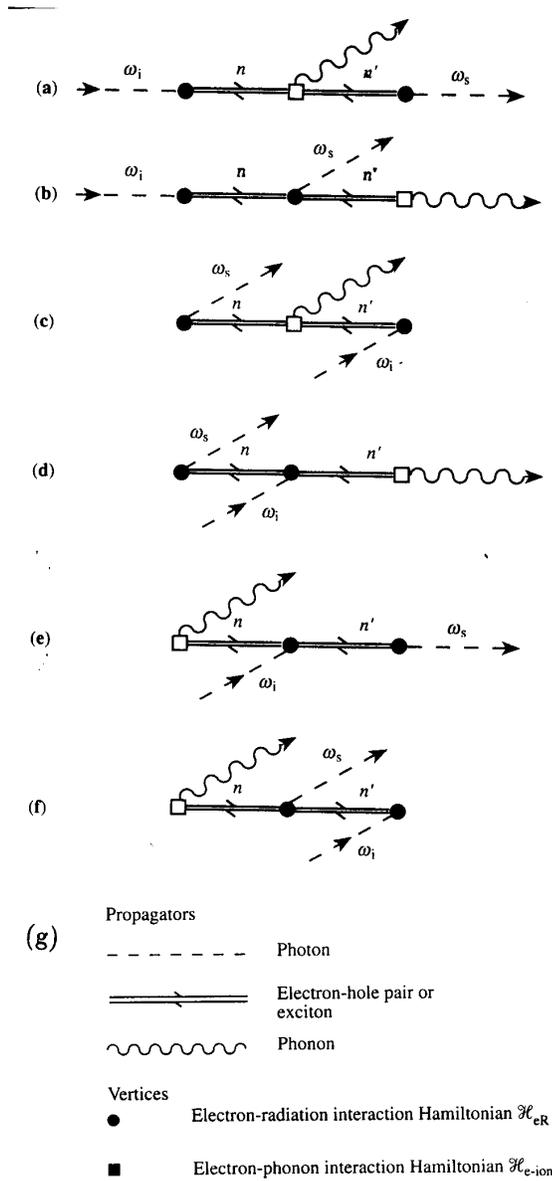
$$\frac{\langle n|\mathcal{H}_{eR}(\omega_i)|i\rangle}{[\hbar\omega_i - (E_n - E_i)]}$$

where the sign $(+)$ corresponds to absorption and $(-)$ to emission and $\mathcal{H}_{eR}(\omega_i)$ denotes the interaction between the electron and the electromagnetic radiation field. The interaction for the second vertex $\mathcal{H}_{e-ion}(\omega_i)$ between the electron and the lattice vibrations of the ion (or the electron-phonon interaction) and the corresponding energy denominator is

$$\hbar\omega_i - (E_n - E_i) - \hbar\omega_q - (E_{n'} - E_n) = [\hbar\omega_i - \hbar\omega_q - (E_{n'} - E_i)]$$

and for the third vertex the denominator becomes $[\hbar\omega_i - \hbar\omega_q - \hbar\omega_s - (E_{n'} - E_i)]$ but since the initial and final electron energies are the same, energy conservation requires $\delta(\hbar\omega_i -$

Figure 10.10: Feynman diagrams for the six scattering processes that contribute to one-phonon (Stokes) Raman scattering. (Taken from Yu and Cardona.) (g) Symbols used in drawing Feynman diagrams to represent Raman scattering.

127

$\hbar\omega_q - \hbar\omega_s$) to yield the probability per unit time for Raman scattering for diagram (a):

$$P_{\mathrm{ph}}(\omega_s) = \left(\frac{2\pi}{\hbar}\right)\left|\sum_{n,n'}\frac{\langle i|\mathcal{H}_{eR}(\omega_s)|n'\rangle\langle n'|\mathcal{H}_{e-\mathrm{ion}}|n\rangle\langle n|\mathcal{H}_{eR}(\omega_i)|i\rangle}{[\hbar\omega_i-(E_n-E_i)][\hbar\omega_i-\hbar\omega_q-(E_{n'}-E_i)]}\right|^2 \\ \times\delta(\hbar\omega_i - \hbar\omega_q - \hbar\omega_s). \tag{10.62}$$

Then summing over the other 5 diagrams yields the result

$$P_{\mathrm{ph}}(\omega_s) = \left(\frac{2\pi}{\hbar}\right)\left|\sum_{n,n'}\frac{\langle i|\mathcal{H}_{eR}(\omega_i)|n\rangle\langle n|\mathcal{H}_{e-\mathrm{ion}}|n'\rangle\langle n'|\mathcal{H}_{eR}(\omega_s)|i\rangle}{[\hbar\omega_i-(E_n-E_i)][\hbar\omega_i-\hbar\omega_q-(E_{n'}-E_i)]}\right. \\ +\frac{\langle i|\mathcal{H}_{eR}(\omega_i)|n\rangle\langle n|\mathcal{H}_{eR}(\omega_s)|n'\rangle\langle n'|\mathcal{H}_{e-\mathrm{ion}}|i\rangle}{[\hbar\omega_i-(E_n-E_i)][\hbar\omega_i-\hbar\omega_s-(E_{n'}-E_i)]} \\ +\frac{\langle i|\mathcal{H}_{eR}(\omega_s)|n\rangle\langle n|\mathcal{H}_{e-\mathrm{ion}}|n'\rangle\langle n'|\mathcal{H}_{eR}(\omega_i)|i\rangle}{[-\hbar\omega_s-(E_n-E_i)][-\hbar\omega_s-\hbar\omega_q-(E_{n'}-E_i)]} \\ +\frac{\langle i|\mathcal{H}_{eR}(\omega_s)|n\rangle\langle n|\mathcal{H}_{eR}(\omega_i)|n'\rangle\langle n|\mathcal{H}_{e-\mathrm{ion}}|n'\rangle}{[-\hbar\omega_s-(E_n-E_i)][-\hbar\omega_s+\hbar\omega_i-(E_{n'}-E_i)]} \tag{10.63} \\ +\frac{\langle i|\mathcal{H}_{e-\mathrm{ion}}|n\rangle\langle n|\mathcal{H}_{eR}(\omega_i)|n'\rangle\langle n'|\mathcal{H}_{eR}(\omega_s)|i\rangle}{[-\hbar\omega_q-(E_n-E_i)][-\hbar\omega_q+\hbar\omega_i-(E_{n'}-E_i)]} \\ \left.+\frac{\langle i|\mathcal{H}_{e-\mathrm{ion}}|n\rangle\langle n|\mathcal{H}_{eR}(\omega_s)|n'\rangle\langle n'|\mathcal{H}_{eR}(\omega_i)|i\rangle}{[-\hbar\omega_q-(E_n-E_i)][-\hbar\omega_q-\hbar\omega_s-(E_{n'}-E_i)]}\right|^2 \\ \times\delta(\hbar\omega_i - \hbar\omega_s - \hbar\omega_q).$$

Although Eq. 10.63 is not generally used to calculate scattering intensities directly, Feynman diagrams similar to those in Fig. 10.10 are widely used in physics.

10.6   Raman Spectra in Quantum Wells and Superlattices

Raman spectroscopy has also been used to study quantum well and superlattice phenomena. One important example is the use of Raman spectroscopy to elucidate zone folding phenomena in the phonon branches of a superlattice of quantum wells. Since the Raman effect is highly sensitive to phonon frequencies, this technique can be used to characterize quantum wells and superlattices with regard to the composition of an alloy constituent (e.g., the composition x of an alloy such as $\mathrm{Si}_x\mathrm{Ge}_{1-x}$). The Raman effect can then be used to determine the amount of strain in each constituent from measurement of the phonon frequencies.

Zone folding effects in the phonon dispersion relations are demonstrated in a superlattice of [GaAs (13.6Å)/AlAs (11.4Å)] $\times 1720$ periods. The observed Raman spectra are shown in Figs. 10.11(a) and (b), demonstrating the zone folding of the LA branch. The difference in the force constants between the GaAs and AlAs constituents causes splittings of the zone–folded phonon branch, as shown in Fig. 10.11(c). The peaks in the Raman spectrum at ~ 64 cm^{-1} and ~ 66 cm^{-1} are identified and labeled with the zone folded modes of the LA branch with symmetries $A_1^{(1)}$ and $B_2^{(1)}$, consistent with the polarization of the incident and scattered photons. At higher frequencies the Raman spectrum of Fig. 10.11(a) shows additional structure related to the zone folded LO phonon branch. Here we note that the normally three–fold levels of T symmetry of the cubic crystal are split into E and B_2 symmetries in the superlattice because of its lower tetragonal symmetry. The two–fold level of E symmetry can be further split by the LO–TO splitting which occurs in ionic solids.
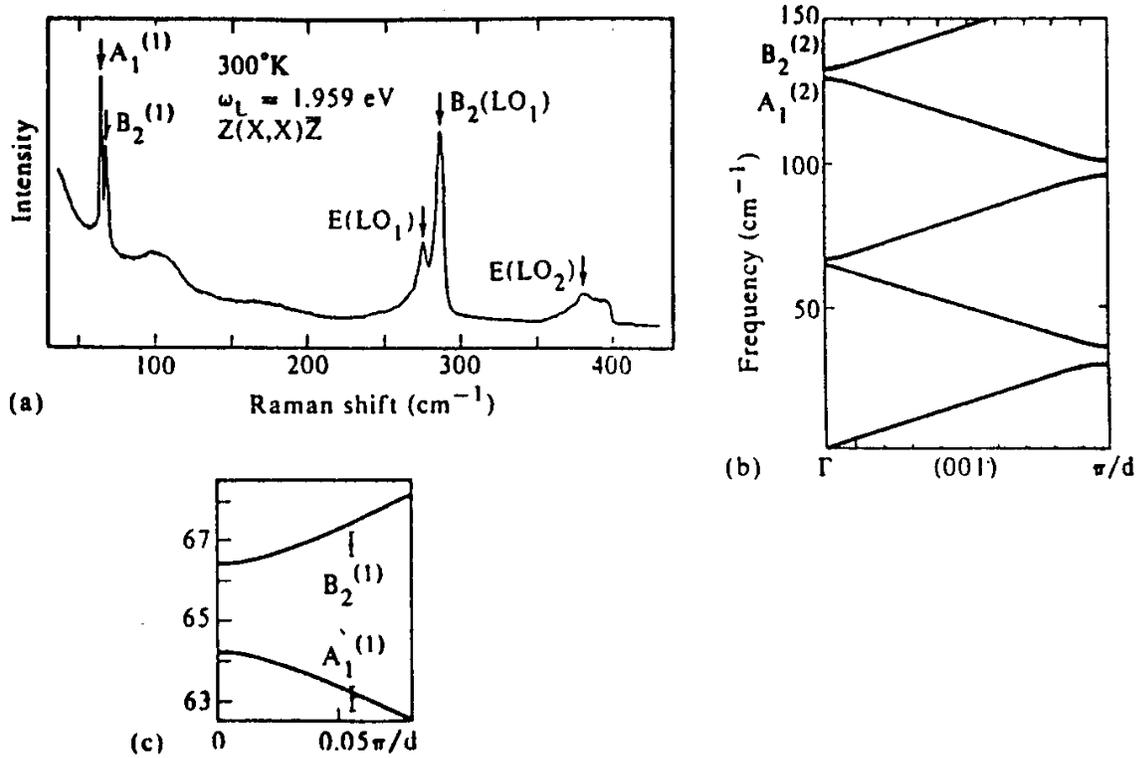
Figure 10.11: (a) Raman spectra of a superlattice consisting of 1720 periods of a 13.6 Å GaAs quantum well and a 11.4 Å AlAs barrier. The polarizations for the incident and scattered light are arranged so that only longitudinal phonons are observed. (b) Dispersion of the LA phonons in the superlattice. (c) An expanded view of the 65 cm^{-1} region of the zone folded LA branch near $\vec{k} \approx 0$. (C. Colvard, T.A. Grant, M.V. Klein, R. Merlin, R. Fischer, H. Morkoc and A.C. Gossard, *Phys. Rev.* B$\underline{31}$, 2080 (1985).)

As another example, Raman spectroscopy can be used as a compositional characterization technique to confirm the chemical composition of a semiconductor alloy. This characterization is based on the identification of the Raman-active modes and the measurement of their frequency shifts and their relative intensities. The strain induced by the lattice mismatch at the interface between $Si_{0.5}Ge_{0.5}$ and a GaAs (110) surface is responsible for the dependence of the frequency shifts of the Ge-Ge, Si-Si and Si-Ge phonon lines on the thickness of the quantum wells in the spectra shown in Fig. 10.12 for $Si_{0.5}Ge_{0.5}$ layers of various thicknesses on a GaAs (110) surface. Since phonon frequencies depend on $(K/M)^{1/2}$ (where K is the force constant and M is the ion mass) the mode frequencies of the Ge–Ge, Ge–Si and Si–Si optical mode vibrations are very different as seen in Fig. 10.12. Therefore the amount of interface strain can be sensitively monitored by Raman scattering. Note the disappearance of the GaAs Raman lines (associated with the substrate) as the thickness of the $Si_{0.5}Ge_{0.5}$ overlayer increases.

Figure 10.12: Raman spectra for various thicknesses of $Si_{0.5}Ge_{0.5}$ on an GaAs (110) substrate. Here the dependence of the Si–Si, Ge–Ge, and Si–Ge bond lengths on the thickness of the $Si_{0.5}Ge_{0.5}$ layer can readily be seen. The samples were grown at 720 K and the measurements were made at 80 K using a laser with a wavelength of 457.9 nm. (G. Abstreiter, H. Brugger, T. Wolf, H. Jorke and H.J. Herzog, *Phys. Rev. Lett.* <u>54</u>, 2441 (1985).

131
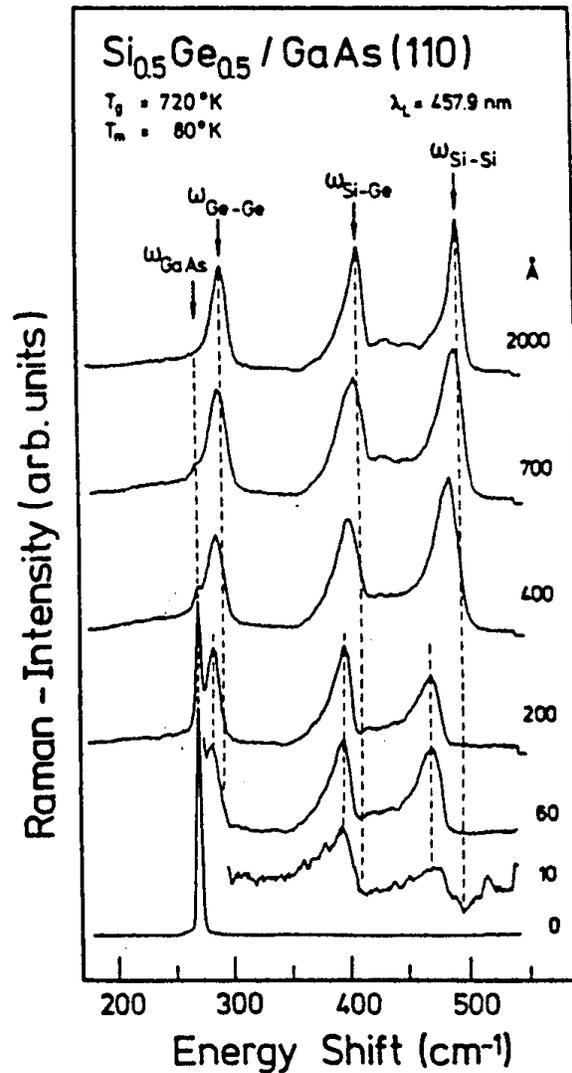
Chapter 11

Non-Linear Optics

References:

- Bloembergen, <u>Non-Linear Optics</u>, W.A. Benjamin, 1965.

- Loudon, <u>Quantum Theory of Light</u>, 2nd Ed., Oxford, 1983.

- Yariv, <u>Quantum Electronics</u>, 3rd Ed., Wiley, 1989.

- Yariv, <u>Optical Electronics</u>, 3rd Ed., Holt, Rinehart and Winston, 1985.

- Yariv and Pearson, <u>Progress in Quantum Electronics</u>, Vol. I. Pt. I.

11.1 Introductory Comments

Non-linear optics became an important field of activity in the 1960's with the advent of high power laser sources such as:

CO_2	at	10.6 μm
YAG:Nd^{3+}	at	1.06 μm
ruby	at	6943 Å
argon	at	5145 Å

The topics we will consider under heading of non-linear optics are:

1. harmonic generation

2. paramagnetic oscillation

3. frequency conversion.

The significant point in non-linear optics is that, when the electromagnetic fields become strong enough, the dielectric function becomes dependent on the electric field $\vec{E}$. We write the total dielectric function as

$$\varepsilon_T = \varepsilon + \overleftrightarrow{\varepsilon}_{\text{NL}} \cdot \vec{E} \tag{11.1}$$

where ε is the linear response and $\overleftrightarrow{\varepsilon}_{\mathrm{NL}} \cdot \vec{E}$ represents the non-linear term. Thus, non-linear effects become more important as the magnitude of the electric field is increased. To observe non-linear effects, we require fields of magnitude $|\vec{E}| \sim 10^6$ volt/cm. Fields of this magnitude are readily available with high power laser sources. Non–linear effects are of great importance in quantum well structures.

It is sometimes more convenient to express these non-linear effects in terms of the polarization per unit volume $\vec{P}$

$$\vec{P} = \overleftrightarrow{\chi}_0 \, \vec{E} + \overleftrightarrow{\chi}_1 \, \vec{E}\vec{E} + \overleftrightarrow{\chi}_2 \, \vec{E}\vec{E}\vec{E} + \cdots \tag{11.2}$$

where $\overleftrightarrow{\chi}_0$ is the linear susceptibility tensor, $\overleftrightarrow{\chi}_1$ is the lowest order non-linear susceptibility important in non-linear materials with no center of inversion; and $\overleftrightarrow{\chi}_2$ is the second order non-linear susceptibility which comes into play for non-linear effects in cubic crystals with a center of inversion for which the first order term vanishes by symmetry.

For the present discussion we will only consider the lowest order non-linear term:

$$\vec{P}^{NL} = \overleftrightarrow{\chi}_1 \cdot \vec{E}\vec{E} \tag{11.3}$$

where NL denotes non–linear and the non-linear susceptibility is in reality a 3rd rank tensor in the same sense that χ_0 for the linear response is a second rank tensor in an actual solid state medium. Likewise $\overleftrightarrow{\chi}_2$ in Eq. 11.2 is a 4th rank tensor. We will not make any use of the tensorial properties of χ_1 here, because we want to keep things simple.

Maxwell's equations in this case become

$$\vec{\nabla} \times \vec{E} + \frac{\mu_0}{c}\left(\frac{\partial \vec{H}}{\partial t}\right) = 0 \qquad \vec{\nabla} \times \vec{H} - \frac{\varepsilon_0}{c}\left(\frac{\partial \vec{E}}{\partial t}\right) = \frac{4\pi}{c}\left(\frac{\partial}{\partial t}\right)\overleftrightarrow{\chi}_1 \cdot \vec{E}\vec{E} \tag{11.4}$$

where the term in $\chi_1 \cdot \vec{E}\vec{E}$ is the non-linear term. We thus obtain the non-linear wave equation

$$\nabla^2 \vec{E} - \frac{\varepsilon_0 \mu_0}{c^2}\left(\frac{\partial^2 \vec{E}}{\partial t^2}\right) = \frac{4\pi \mu_0}{c^2}\left(\frac{\partial^2}{\partial t^2}\right)\overleftrightarrow{\chi}_1 \cdot \vec{E}\vec{E}. \tag{11.5}$$

The wave equation <u>without the non-linear term</u> has plane wave eigen functions. That is, if we have an incident field with more than one frequency

$$E = \left[\vec{E}_1 e^{i(K_1 z - \omega_1 t)} + \vec{E}_2 e^{i(K_2 z - \omega_2 t)}\right], \tag{11.6}$$

the wave equation

$$\nabla^2 \vec{E} - \frac{\varepsilon_0 \mu_0}{c^2}\left(\frac{\partial^2 \vec{E}}{\partial t^2}\right) = 0 \tag{11.7}$$

is valid for each of the waves in Eq. 11.6. This insures that no mixing occurs and each frequency propagates independently through the linear medium.

Now what does the non-linear term do? Here we have to take a product $\vec{E}\vec{E}$. Clearly we will get terms at $2\omega_1$ and $2\omega_2$ (frequency doubling or second harmonic generation) and also at $[\omega_1 + \omega_2]$ (frequency mixing). This means that we now no longer have plane wave solutions

$$\vec{E}_i = \vec{\mathcal{E}}_i e^{i(K_i z - \omega_i t)} \tag{11.8}$$

133

for $\vec{E}_i$ polarized along $\hat{x}$, where $\vec{\mathcal{E}}_i$ is a constant amplitude. Instead we must look for a more generalized form. For example, we could seek a solution of the form of a modified plane wave

$$\vec{E}_i = \vec{\mathcal{E}}_i(z)e^{i(K_i z - \omega_i t)} \tag{11.9}$$

where $\vec{\mathcal{E}}_i(z)$ now has a weak z dependence. This approach is in the spirit of perturbation theory. Now in taking spatial derivatives of E_i we will get two terms

$$\frac{\partial E_i}{\partial z} = \left[\frac{\partial \mathcal{E}_i}{\partial z} + iK_i \mathcal{E}_i\right]e^{i(K_i z - \omega_i t)} \tag{11.10}$$

where the term in $iK_i\mathcal{E}_i$ is the large term and the term in $\partial\mathcal{E}_i/\partial z$ represents the small perturbation. Thus in taking two derivatives we will get

$$\frac{\partial^2 E_i}{\partial z^2} = iK_i\left[2\frac{\partial \mathcal{E}_i}{\partial z} + iK_i\mathcal{E}_i\right]e^{i(K_i z - \omega_i t)}, \tag{11.11}$$

retaining only the lowest order term in the perturbation.

We will now show that coupling to frequency ω_3 is possible for waves at frequencies ω_1 and ω_2 provided that $\omega_3 = \omega_1 + \omega_2$. From the wave equation Eq. 11.5, we see that if the incident field has two frequencies ω_1 and ω_2, we will get a perturbation driving term on the right hand side of the non-linear wave equation and also a perturbation term on the left hand side of this equation due to the z dependence of $\mathcal{E}_i(z)$. Assuming a solution at some frequency ω_3 (to be determined from the wave equation, Eq. 11.7) we can write an equation for the right and left hand perturbation terms

$$2iK_3\left(\frac{\partial \mathcal{E}_3}{\partial z}\right)e^{i(K_3 z - \omega_3 t)} = -(\omega_1 + \omega_2)^2\left(\frac{4\pi\mu_0}{c^2}\right)\chi_1 \cdot \mathcal{E}_1\mathcal{E}_2 e^{i(K_1 z - \omega_1 t)}e^{i(K_2 z - \omega_2 t)} \tag{11.12}$$

To satisfy the right hand side of Eq. 11.5 we need to match the time phase terms on the left hand side, yielding

$$\omega_3 = \omega_1 + \omega_2, \tag{11.13}$$

which indicates that mixing has occurred.

Because of the dispersion properties of crystals, the wave vector for light K will be a function of ω and we cannot in general cancel the phases for all frequencies. Thus, some mismatch $\Delta K = K_3 - (K_1 + K_2)$ will generally occur. Phase matching is achieved when $\Delta K \equiv 0$ and in this case the three waves will be coherent. In free space, phase matching is automatically satisfied since there is no nonlinear response in free space. In a solid, $K = \tilde{n}\omega/c$ where $\tilde{n}$ is the index of refraction. Solids have the property that the optical constants are frequency-dependent so that in general $\tilde{n}_{\omega_1} + \tilde{n}_{\omega_2} \neq \tilde{n}_{\omega_3}$. If, however, ΔK is small, phase matching is approximately satisfied. In fact, provided that the phase changes by less than π, some coherence will be achieved. We thus introduce a phase coherence length ℓ_c defined as $\ell_c = \pi/\Delta K$ over which some degree of coherence is achieved.

11.2 Second Harmonic Generation

For the non–linear process corresponding to second harmonic generation we have $\omega_1 = \omega_2 = \omega$ and $\omega_3 = 2\omega$. The non-linear contribution to the polarization will be proportional

to $(E_\omega)^2$ and the power generated at 2ω will be proportional to $(E_\omega)^4$. If phase matching is achieved, the power produced at the second harmonic will be maximized. For phase matching we require

$$\omega_3 \tilde{n}_3 = \omega_1 \tilde{n}_1 + \omega_2 \tilde{n}_2 \qquad (11.14)$$

or

$$2\omega \tilde{n}_3 = 2\omega \tilde{n}_1, \qquad (11.15)$$

which can be written more conveniently as $\tilde{n}_{2\omega} = \tilde{n}_\omega$, which says that we require the index of refraction at 2ω to be equal to the index at frequency ω for phase matching.

For solids, the index of refraction $\tilde{n}$ will generally be frequency dependent, so phase matching would seem difficult to achieve. By using anisotropic materials and selecting particular directions of propagation, it is sometimes possible to arrange matters so that $\tilde{n}_\omega$ for one polarization direction is equal to $\tilde{n}_{2\omega}$ for another polarization direction. Efficient harmonic generation has been achieved using the semiconductor Te, which crystallizes in a hexagonal structure and has a rather different index of refraction for the polarization $\vec{E} \parallel \vec{c}$ than for $\vec{E} \perp \vec{c}$. To achieve frequency doubling in Te, it is convenient to use a CO_2 laser source, since Te which has a band gap of 0.344 eV is quite transparent at both ω and 2ω for the CO_2 laser line at 10.6μm. With frequency doubling, it is possible to convert infrared light to visible radiation and thus to utilize the highly developed technology for the detection of visible light signals. The process of second harmonic generation need not be considered a small or weak effect. High conversion efficiencies ($> 50\%$) can be achieved.

11.2.1 Parametric Oscillation

Here laser power is applied to a non-linear crystal at a pump frequency ω_3. Oscillations are induced in the crystal at frequencies ω_1 and ω_2, the signal and idler frequencies respectively. The signal and idler frequencies are determined by the frequency condition

$$\omega_3 = \omega_1 + \omega_2 \qquad (11.16)$$

and the phase matching condition

$$\Delta K = 0, \qquad K_3 = (K_1 + K_2). \qquad (11.17)$$

It is only when phase matching occurs that the two waves will interact sufficiently to produce any measurable non–linear effects. The phase matching condition can also be written as

$$\omega_3 \tilde{n}_3 = \omega_1 \tilde{n}_1 + \omega_2 \tilde{n}_2 \qquad (11.18)$$

where $\tilde{n}_1$ and $\tilde{n}_2$ are refractive indices at frequencies ω_1 and ω_2 and are determined by the propagation direction and polarization of the modes at ω_1 and ω_2. In non-cubic materials, the index of refraction depends on the direction of the $\vec{E}$ field relative to the crystallographic directions. Thus, by changing the propagation direction of the pump frequency relative to the crystal optic axis (e.g., the "c" axis in a hexagonal crystal like Te), it is possible to "tune" the signal and idler frequencies ω_1 and ω_2. Furthermore, since the indices are temperature dependent, "tuning" can also be accomplished by varying the temperature of the non-linear crystal; tuning with application of uniaxial stress can also be accomplished. Parametric oscillation need not be a small effect. Using a Q-switched ruby laser as a pump at 6943

Å on a non-linear LiNbO$_3$ crystal, a signal at 1.04μm and an idler at 2.08μm have been achieved with 60 kW of signal power generated for 270 kW of input power or a conversion efficiency of $\sim 20\%$. Tuning with a parametric oscillator between 0.54μm and 3.7μm has also been achieved.

11.2.2 Frequency Conversion

For frequency conversion, two frequencies are applied to a non-linear crystal and the sum frequency (up converter) or difference frequency (down converter) is generated. In this experiment, we might impose a high power signal at a pump frequency ω_3 and a lower power signal at ω_2. These signals mix in the non-linear medium to produce a signal at

$$\omega_1 = \omega_3 - \omega_2 \qquad \text{(down converter)} \tag{11.19}$$

or at

$$\omega_1 = \omega_3 + \omega_2 \qquad (\text{up} - \text{converter}). \tag{11.20}$$

Phase matching determines whether up-conversion or down-conversion actually occurs. The power from the signal at ω_3 drives the system at frequencies ω_1 and ω_2, and as a function of length of the non-linear crystal, the amounts of power at ω_1 and ω_2 can be varied. Thus, by choosing the length properly the conversion of power to frequency ω_1 can be maximized.

Frequency conversion is attractive for practical applications because up-conversion can be exploited to convert an infrared signal into the visible region where detectors are fast and sensitive. Down-converters can be exploited to create a different frequency in the far infrared where high power sources have been unavailable until about 1965, when some far infrared lasers were first built. For example, the two strong CO$_2$ laser lines at 10.6μm and 9.6μm can be mixed to get a far infrared signal.

Chapter 12

Electron Spectroscopy and Surface Science

References:

- Demuth, Physics in a Technological World, French, Editor, AIP, NY 1988.

- Burns, Solid State Physics, Academic Press 1985, Chapt. 7.

12.1 Photoemission Electron Spectroscopy

12.1.1 Introduction

Photoemission is one of the most important of the electron spectroscopy techniques. In photoemission, the excitation particle is the photon and the excited particle is the electron. What we measure is the dielectric response function which gives us information about elementary excitations and the electronic structures of the solid. Depending on the excitation energies, photoemission measures the density of states, energy distribution of the joint density of states, and probes the valence bands and the core levels. With the advent of the use of synchrotron radiation and angle-resolved spectroscopy, the mapping of the electron energy bands has become possible. Other electron spectroscopies or related surface techniques include:

1. ESCA (Electron Spectroscopy for Chemical Analysis) which measures the chemical shift and thereby probes the local environment and oxidation state of the compound – also called x-ray photoelectron spectroscopy (XPS)

2. AES (Auger Electron Spectroscopy) is a two electron process mainly used for elemental analysis of surface constituents

3. X-ray Fluorescence again is mainly used for elemental (chemical) analysis

4. ELS (Electron Loss Spectroscopy) which like optical spectroscopy gives the dielectric function of the material – also called electron energy loss spectroscopy (EELS)

5. LEED (Low Energy Electron Diffraction) mainly used for structural analysis.

6. RHEED (Reflection High Energy Electron Diffraction) mainly used for in situ characterization of superlattices during layer-by-layer growth.

7. STM (Scanning Tunneling Microscopy) which is used to obtain atomic resolution of atoms and molecules on surfaces.

All of the above, except LEED, involve an inelastic scattering mechanism.

In photoemission the photoelectric current (I) can be written in the form

$$I = I(E, \theta_e, \phi_e, \vec{\sigma}; \hbar\omega, \vec{\rho_p}, \theta_p, \phi_p) \tag{12.1}$$

where $E, \theta_e, \psi_e, \vec{\sigma}$ are respectively the kinetic energy, polar angle, azimuthal angle, and spin of the electron, and $\omega, \vec{\rho_p}, \theta_p, \phi_p$ are respectively the frequency, polarization, polar angle, and azimuthal angle of the photon. For the various experimental measurements, different variables are held constant. The most common quantities that are measured in photoemission experiments include

1. EDC (Energy Distribution Curves) where the photoelectron current is observed as a function of electron energy $I = I(E)$ with all other parameters held constant

2. CIS (Constant Initial State Spectroscopy) where $\hbar\omega - E$ is kept constant, and the photoelectron current is measured as a function of electron energy and photon energy $I = I(E, \hbar\omega)|_{E-\hbar\omega=\text{const}}$

3. CFS (Constant Final State Spectroscopy) where $I = I(\hbar\omega)$ is measured

4. ARPS (Angular Resolved Photoemission Spectroscopy) where the angles are allowed to vary.

Historically, photoemission was first used to study the work function of solids, and to study core levels in molecules. Both UPS (Ultraviolet Photoelectron Spectroscopy) and XPS (X-ray Photoelectron Spectroscopy) were used to probe core levels.

In the photoemission process light incident on a sample is absorbed in a length characterized by the optical skin depth. In this optical skin depth, electrons can be excited to ionization states and are eventually emitted. Because of the much stronger interaction of electrons with matter (in contrast to the case of photons), the characteristic absorption length for the photo-excited electrons is much smaller than that for the exciting photons and as a consequence, only electrons that are excited close to the surface will be emitted. By applying an electric field to the sample, some of the electrons generated in the optical excitation process are collected and their kinetic energy is measured. The following two quantities are observed in standard photoemission studies:

1. the photoelectric yield (defined as the number of electrons that are produced per unit of incident photon flux) as a function of photon energy,

2. the energy distribution of the emitted electrons, for various values of the incident photon energy.

These measurements provide information on interband transitions through analysis of structure in the photoelectric yield curves, and on the density of valence states through the shape
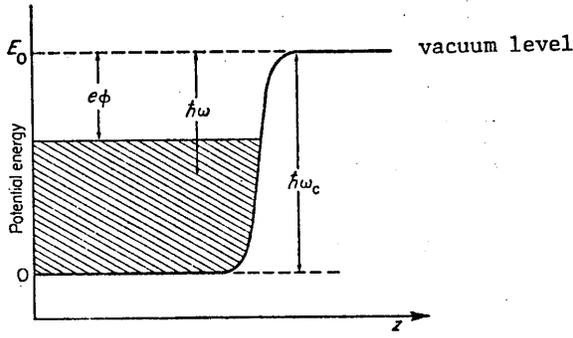
Figure 12.1: The potential barrier at a surface showing the work function.

of the electron distribution curves. Because this technique provides one of the few methods for studying the density of states (particularly for low lying valence states), this has become an important measurement technique. Furthermore because of its surface sensitivity, ultraviolet photoemission spectroscopy provides a useful tool for contrasting electronic states characteristic of the surface relative to states characterizing the bulk.

In describing the photoemission process in a metal, we make use of the model for the potential barrier at a surface shown in Fig. 12.1. In so–doing, we illustrate one of the classical applications of the photoemission process in measuring the work function of a solid.

In this model we assume that our surface is in the $x-y$ plane, and that p_z is the electron momentum (in the direction normal to this surface) which the electron acquires through the photo-excitation process. From the diagram, we see that it will be possible for this electron to escape from the surface (by photoemission) provided that

$$\frac{p_z^2}{2m} + \hbar\omega \geq E_0 \tag{12.2}$$

where E_0 is called the <u>vacuum level</u>, located at an energy $\hbar\omega_0$ above the band extremum. The work function $e\phi$ represents the minimum kinetic energy that an electron must be given by the light in order to escape from the surface (see Fig. 12.1). For the electron to retain any information about its initial state the mean free path (see Fig. 12.2) must be greater than the penetration depth of the exciting radiation.

The photocurrent I will then be proportional to the number of electrons escaping from the surface of the metal

$$I = e \int_0^\infty p_z n(p_z) D(p_z) dp_z \tag{12.3}$$

where $D(p_z)$ is the escape probability for an electron of momentum p_z and $n(p_z)$ is the corresponding electron density, which can be expressed through the Fermi distribution function

$$n(p_z) = \frac{2}{\hbar^3} \int_{-\infty}^\infty dp_x dp_y \frac{1}{e^{(\frac{p^2}{2m}-E_F)/k_B T} + 1} \tag{12.4}$$

The discussion given above is appropriate to photoemission from a metal. In the case of semiconductors there is an energy gap and the Fermi level lies in this energy gap, in
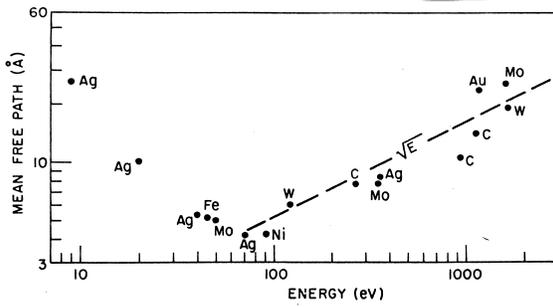
139

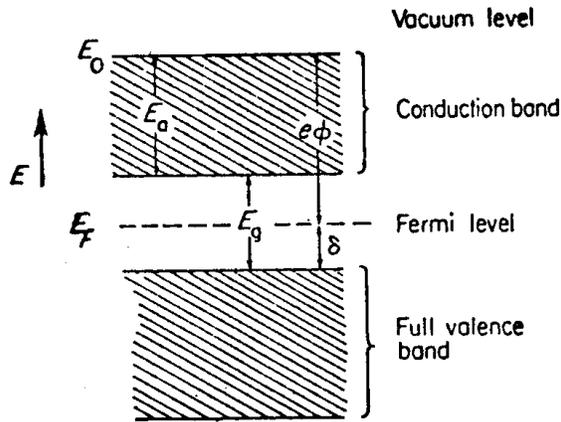Figure 12.2: The mean free path for an electron in various materials as a function of energy.

Figure 12.3: Schematic diagram of the states near the Fermi energy of a semiconductor.

contrast to the situation in metals where the Fermi level lies at the top of the occupied electron states within an energy band. For semiconductors, the work function $e\phi$ is still defined relative to the Fermi level but the threshold energy is now increased to $e\phi + \delta$ where δ is the energy difference between the highest lying valence band maximum and the Fermi level as seen in Fig. 12.3.

For semiconductors it is customary also to refer to the electron affinity, denoted by E_a on the diagram and representing the energy difference between the vacuum level E_0 and the bottom of the conduction band. We can see how interband transitions are detected in the photoemission process by the following argument. Suppose that $E_a < E_g$. The threshold for photoemission requires the incident photons to have an energy of at least:

$$\hbar\omega > e\phi + \delta = E_a + E_g. \tag{12.5}$$

The photoemission process will dominate until the photo-emitted electrons are themselves energetic enough to make electron-hole pairs through collisions with other electrons. The threshold for this secondary interband transition process (whereby the photo-excited electron has enough energy to produce a second photoelectron) is

$$\hbar\omega > (E_a + E_g) + E_g = E_a + 2E_g. \tag{12.6}$$

140
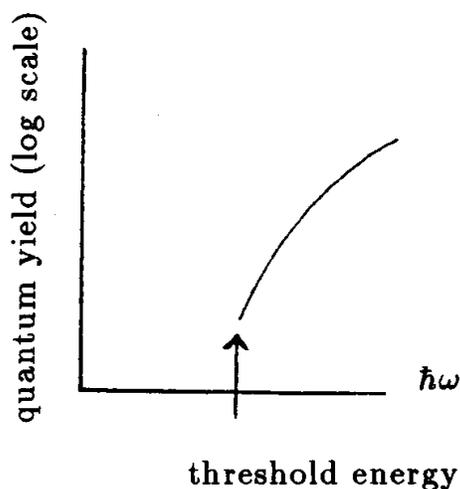
Figure 12.4: Quantum yield vs. photon energy above the threshold energy.

The threshold for this interband process gives rise to structure in the photoemission distribution curves identified with interband transitions.

We call this general field of study photoemission spectroscopy. The advent of the theoretical development of the electronic structure of solids in the 50's made the microscopic understanding of photoemission possible. Furthermore, instrumental advances in high vacuum technology in the late 60's extended the photon range of ultraviolet photoelectron spectroscopy (UPS) to greater than 6 eV. Simultaneously the use of synchrotron radiation as a light source gives tunability of the excitation frequency and high power densities from the visible to hard x–ray frequencies. Angle resolved techniques are now commonly used to gain an understanding of the electronic structure of solids.

In the photoemission process, three basic things must happen: (3-step model)

1. optical excitation of an electron from an occupied state

2. transport of the photo-excited electron to the surface

3. the electron must escape from the surface and into the vacuum region

In order for the electron to escape from the surface into the vacuum region where the electron is collected, it must have sufficient kinetic energy. Measurements of the photoelectric yield exhibit a threshold energy and thereby provide a measure of the work function. Photoelectric yield data are plotted in terms of the quantum yield (defined as the number of electrons emitted per incident photon) vs. photon energy as shown in Fig. 12.4.

12.1.2 Energy Distribution Curves

Of greater interest however is the energy distribution of the photo-emitted electrons $N(E)$ which is defined as the number of electrons emitted with energy E in the range ΔE relative to the total number of electrons produced per photon. We show that the probability that an electron of energy E is produced is proportional to the initial density of states at energy $(E-$

$\hbar\omega$) written as $g(E - \hbar\omega)$. The intensity profile of the electrons emitted in a photoemission experiment will contain both the primary electrons which suffer no inelastic collisions, and the secondary electrons that suffer at least one inelastic collision. The photoelectric current can then be written as

$$I(E,\omega) = I_p(E,\omega) + I_s(E,\omega) \tag{12.7}$$

where I_p, I_s are respectively identified with primary and secondary electrons. I_p depends on three factors according to the three-step model:

$$I_p(E,\omega) = P(E,\omega) \cdot T(E) \cdot D(E) \tag{12.8}$$

where $P(E,\omega)$ is the probability that a photoelectron of energy E is excited by a photon of energy $\hbar\omega$, $T(E)$ is the transmission function of the excited electrons and $D(E)$ is the escape function of the excited electron. We can write $T(E)$ as

$$T(E) = \frac{\lambda_e(E)/\lambda_{ph}(\omega)}{1 + \lambda_e(E)/\lambda_{ph}(\omega)} \tag{12.9}$$

where λ_e is the mean free path of the electrons and λ_{ph} is the attenuation length of photon. Likewise, we can write $D(E)$ as

$$D(E) = \begin{cases} \frac{1}{2}[1 - (\frac{E_F + \phi}{E})^{\frac{1}{2}}] & \text{for } E > E_F + e\phi \\ \\ 0 & \text{otherwise} \end{cases} \tag{12.10}$$

where E_F denotes the Fermi level and ϕ is the work function. If we consider only bulk states, and direct transitions, the energy distribution takes the form

$$P(E,\omega) \sim \Sigma_{n,n'} \int d^3k \delta(E_{n'}(\vec{k}) - E_n(\vec{k}) - \hbar\omega)\delta(E_{n'}(\vec{k}) - E) \tag{12.11}$$

The first δ function represents the joint density of states for optical absorption and the second δ function selects out the energy that is set by the energy analyzer. Thus the structures of the EDCs (energy distribution curves) mimic those of the joint density of states and thus give information concerning the joint density of states. The experimental data for the photo-emitted electron energy distribution are taken for a variety of photon energies as shown in the curves in Fig. 12.5 and $N(E)$ is plotted as a function of $(E - \hbar\omega)$ in order to relate the electron energy distributions to the density of states at the same initial energy. Peaks in the density of initial states give rise to peaks in $N(E)$ at the same value of $E - \hbar\omega$, independent of the energy of the photons involved in the excitation process. Each curve in Fig. 12.5 for aluminum is labeled by the incident photon energy. The dashed curve is a density of states curve for the occupied electron states in aluminum obtained from the interpretation of these data. Note the threshold appearing at the Fermi level. Since the onset of interband transitions corresponds to discontinuities in the density of states spectrum, the EDC curves can also be used to identify interband transitions. To interpret valence band states we make use of the fact that d-bands have low dispersion and therefore a high density of states over a narrow energy region, while s- and p-bands have a low density of states over a wide energy region (see Fig. 12.5).

Since laboratory ultraviolet sources are weak and difficult to work with, it is common to use a monolayer of cesium on the surface to lower the work function and the threshold
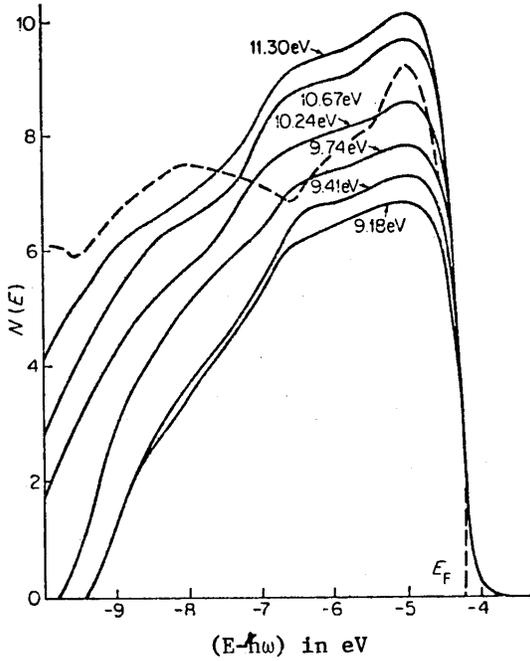
Figure 12.5: Energy distribution curves for photoelectrons in aluminum for various photon energies. The dashed curve shows the density of states over a wide energy region.

energy for the photoemission process. This allows the photoemission experiments to be carried out at somewhat lower photon energies where laboratory sources are more intense. More recently, intense synchrotron radiation ultraviolet sources have become available at a few of the national accelerator facilities and this has accelerated the development of photoemission spectroscopy research.

If we allow non–direct transitions to occur, then the energy distribution function $P(E, \omega)$ takes the form

$$P(E, \omega) \sim \Sigma_{nn'} \int d^3k \, d^3k' |\langle n'|\vec{p}|n\rangle|^2 \delta(E_{n'}(\vec{k'}) - E_n(\vec{k}) - \hbar\omega)\delta(E_{n'}(\vec{k'}) - E). \qquad (12.12)$$

If we rewrite $P(E, \omega)$ as

$$P(E, \omega) \sim \Sigma_n \int d^3k \, \delta[E - \hbar\omega - E_n(\vec{k})] \sum_{n'} \int d^3k' \, \delta[E_{n'}(\vec{k'}) - E] |\langle n'|\vec{p}|n\rangle|^2 \qquad (12.13)$$

then $P(E, \omega)$ is expressed as a weighted average of the initial and final density of states. Thus indirect transitions can be invoked to explain stationary structures in the EDCs as we scan the photon frequency.

Modifications to the three step model have been made to include the possibility of an energy dependent electron mean free path

$$\lambda_e(E) = v_g T_e(E) = \frac{1}{\hbar} |\nabla_k E(\vec{k})| T_e(E) \qquad (12.14)$$

and the possibility that the transmission function $T_e(E)$ is described by more than one core state that couples to Bloch states. In Eq. 12.14 v_g denotes the group velocity.

143

12.1.3 Angle Resolved Photoelectron Spectroscopy

Advances in angle-resolved photoemission have made photoemission an even more powerful experimental technique especially for the study of electronic band structure. If one measures the kinetic energy and the propagation direction of the electron by the conservation of wave vector parallel to the surface, we obtain

$$\vec{K}_{||} = \vec{k}_{||} + \vec{G}_{||} \tag{12.15}$$

where $\vec{K}$ and $\vec{k}$ are respectively the wave vectors in vacuum and in the solid and $\vec{G}$ is a reciprocal lattice vector of the solid. Wave vector conservation together with the energy conservation

$$E = E_f(\vec{k}) \tag{12.16}$$

where $E_f(\vec{k})$ denotes the energy of the electron and the zero of energy is taken at the vacuum level gives

$$E = \frac{\hbar^2}{2m}(K_\perp^2 + \vec{K}_{||}^2) \tag{12.17}$$

$$\frac{\hbar^2 K_\perp^2}{2m} = E_f(\vec{k}) - \frac{\hbar^2(\vec{k}_{||} + \vec{G}_{||})^2}{2m} \tag{12.18}$$

allowing the determination of $E_f(\vec{k})$ as a function of $\vec{k}_{||}$ from photoemission data. Thus with a good band structure calculation the energy for all the directions of the bulk photoemission are determined. The functional form of Eq. 12.18 is especially suitable for layered materials due to the fact that the $k_\perp$ dispersion is very small. Thus for layered materials the band structure is approximately 2-dimensional. Each peak in the EDC will give rise to a point on the E vs. $\vec{k}_{||}$ plot and thus E vs. k can then be mapped uniquely. For a three-dimensional system, a knowledge of the energy band structure is needed since $k_\perp$ is not determined if $E_f(\vec{k})$ is not known.

12.1.4 Synchrotron Radiation Sources

Before the availability of synchrotron radiation, photoemission was carried out using a few strong discrete line sources. Synchrotron radiation has provided us with a strong tunable source from the infrared to the x-ray region of the electromagnetic spectrum. Synchrotron radiation is emitted by electrons in circular accelerators, such as synchrotrons and storage rings. This radiation is a consequence of the centripetal acceleration of the particle moving in a circular path at relativistic velocities (close to the velocity of light). The energy to produce this radiation is supplied by particle accelerators.

Synchrotron radiation has a number of properties which make it extremely useful. First, this photon source can be extremely intense, several orders of magnitude more intense than other broad-band sources. Second, it can have a very broad frequency spectrum, including the ultraviolet and x-ray regions where there are no other intense, tunable sources. The center of the spectrum is near the energy

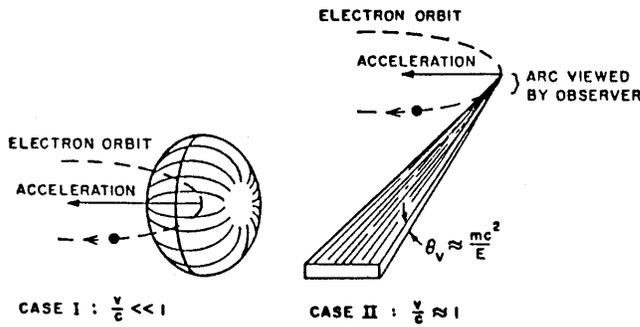$$\hbar\omega = \frac{\gamma^3 \hbar c}{R} \tag{12.19}$$

144

Figure 12.6: Radiation emission pattern of electrons in circular motion: Case I, non-relativistic electrons. Case II, relativistic electrons. Synchrotron radiation sources operate under case II.

where

$$\gamma = \frac{E}{mc^2} \tag{12.20}$$

is the ratio of the accelerator energy to the particle's rest energy, and

$$R = \frac{\gamma mc^2}{eB} \tag{12.21}$$

is the radius of the circular path. For example, a 1 GeV accelerator with a 1 tesla magnetic field and a radius of 3 meters gives $\hbar\omega$ of about 1 keV. The third important property of synchrotron radiation is that it is highly collimated, being confined to a narrow "searchlight" beam in the direction tangent to the orbit with angular spread $\sim \gamma^{-1}$ radians as shown in Fig. 12.6.

A fourth useful property is the high degree of polarization of the radiation in the plane of the orbit. Fifth, devices called "wigglers" and "undulators" have recently been developed which enhance the intensity of the radiation in a particular part of the spectrum, adding one or two orders of magnitude to the already high intensity in the desired region.

The first experiments using synchrotron radiation were carried out in the so-called parasitic mode, at synchrotrons being used for particle physics research where the accelerators are optimized for the particular set of particle-physics experiments being carried out. Furthermore these synchrotrons operate in a pulsed mode, where the electrons are accelerated in bunches up to the maximum energy desired. The desired maximum energy may vary from experiment to experiment. An example of the spectral distribution available from the Stanford SLAC facility is shown in Fig. 12.7.

Dedicated sources of synchrotron radiation are becoming increasingly available, such as the National Light Source at Brookhaven National Laboratory, for both the UV and x-ray regions. These are operated as storage rings, where the electron beam is maintained at a constant energy for long periods of time. Some synchrotron radiation work is still being done in the parasitic mode at storage rings being used for colliding-beam high-energy physics experiments, but here the beam currents are lower and hence the radiation is weaker.

The new synchrotron radiation sources have made possible many new experiments, not only advances in photoemission such as angle-resolved experiments, but also advances in crystal structure determination, microlithography, x-ray fluorescence, and the determination of local environments on surfaces using x-ray absorption fine structure (EXAFS).
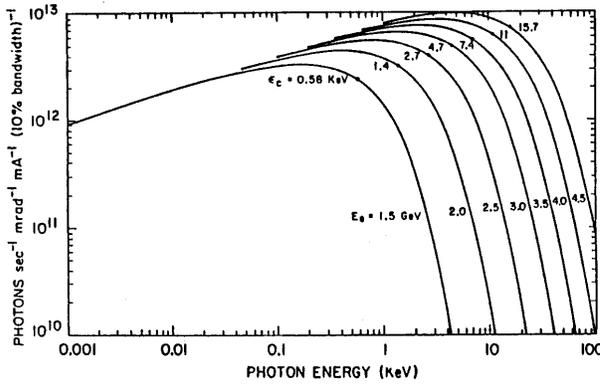
Figure 12.7: Photon intensity vs. photon energy for various maximum accelerator energies E_0 showing the spectral distribution of a synchrotron radiation source. The photon energy at which the maximum photon intensity occurs is denoted by ε_c on the figure.

12.2 Surface Science

12.2.1 Introduction

Many electron devices depend on the electronic properties of surfaces. Because of the geometrical effect of a two-dimensional surface, atoms at a surface have fewer neighbors than similar atoms in the bulk. Therefore, the electronic energy levels at surfaces are different from what they are in the bulk. For example, a silicon atom in bulk silicon is surrounded by four tetrahedral bonds. On the surface, a silicon atom will have fewer bonds, and the surface valence electrons that do not participate in bonding are described as dangling bonds. These surface valence electrons give rise to new electronic states called surface states. When the surface states are located in the band gap of a semiconductor or insulator, they are more readily detected. A probe of electronic energy states with a skin depth δ that is large $\delta \gg a$ compared with a lattice constant a is sensitive to the bulk electronic states because the surface atoms comprise a small fraction of the total number of atoms that are probed. On the other hand, a probe with a short skin depth (such as electrons in the ten and hundred eV range) is especially sensitive to the surface atoms.

In addition, impurity atoms are preferentially adsorbed on the surface. These impurity atoms also give rise to surface states. Since adsorbed impurity atoms are important in catalyzed chemical reactions, there is considerable interest in studying these surface states. At the present time there is a great deal of work being done on the study of surfaces and on their electronic surface states. One reason is the availability of new experimental probes, such as the photoemission experiments already discussed and the STM probes discussed in this chapter, using ultra-high vacuum equipment. A second reason is the recent improvement in the calculation of surface states and of the total energy of different surface structures. A third reason is related to the smaller dimensions of semiconductor electronic devices and the increasing importance of surfaces in these devices.

A number of different experimental techniques are used to study surfaces. The photoemission process discussed in §12.1 is sensitive to surface states especially for incident photon energies close to the photoemission threshold. In this limit, only electrons near the surface will have enough kinetic energy to escape because of the strong electron-electron interaction. As discussed above, photoemission provides information on the occupied valence states.
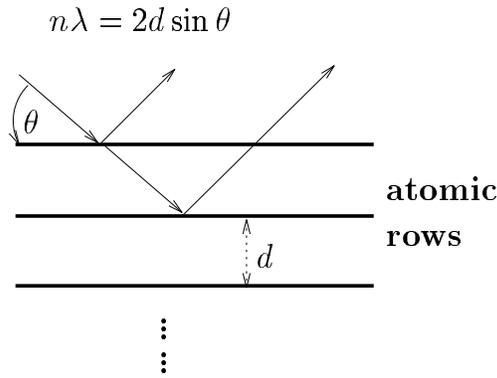
$$n\lambda = 2d\sin\theta$$

atomic rows

Figure 12.8: Bragg condition for x-ray diffraction from rows of atoms.

12.2.2 Electron Diffraction

Another common technique is low energy electron diffraction (LEED). This technique is especially sensitive to atomic arrangements on the surface and is analogous to the x-ray diffraction techniques that are used to establish the crystal structure in the bulk solid. LEED experiments can be carried out to study the structure of clean surfaces or of adsorbed species on surfaces. The positions of the LEED spots on a photograph establish the periodicity of the intrinsic surface structure.

Because of the small penetration depth for low energy electrons ($E < 100$eV), the LEED technique emphasizes the surface structure. The LEED spot pattern that is formed is due to the constructive interference of reflections of the electron beam through scattering of <u>rows</u> of atoms rather than planes of atoms as occur in three-dimensional x-ray diffraction illustrated in Fig. 12.8. Since the Bragg law in this case corresponds to rows of atoms, the surface structure that is probed is indexed by a two-dimensional lattice. In many cases the surface structure forms a superlattice relative to the substrate. This rearrangement of the surface atoms takes place because of the dangling bonds, which would otherwise occur at the surface. The rearrangement partially satisfies the bonding requirements. Within the surface, rows of atoms may move closer together or farther apart, and the surface atoms may move in or out relative to the inner layers of atoms. Such a rearrangement of the surface atoms is called <u>reconstruction</u>. Surface reconstruction is illustrated in Fig. 12.9 for the case of a (110) GaAs surface. The corresponding change in the surface density of states is shown in Fig. 12.10. The surface is highly sensitive to the presence of adsorbed atoms. Figure 12.11 shows the modification to the surface density of states of a (110) surface of GaAs upon exposure to oxygen, as measured in this case by photoemission (see §12.1).

The notation used to describe the surface structure is in terms of (1) the length of the lattice vectors in the superlattice relative to those of the substrate and (2) the angle of rotation of the superlattice coordinate system relative to that of the substrate. Illustrated in Fig. 12.12 are examples of $(2 \times 2)R0°$ and $(\sqrt{3} \times \sqrt{3})R30°$ superlattices. Note in the case of the $(\sqrt{3} \times \sqrt{3})R30°$ superlattice that the coordinate system of the superlattice makes an angle of $30°$ with respect to that of the substrate.

The LEED technique is used to study the structure of pristine surfaces. In many cases, the surface structure may be different from the bulk structure because the surface atoms
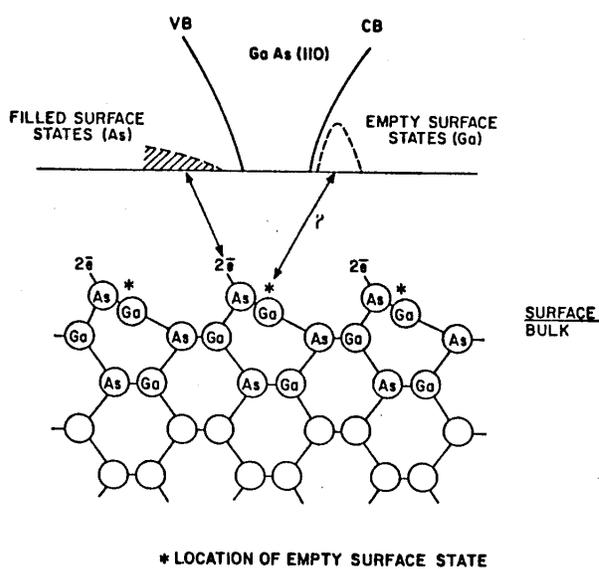
147

Figure 12.9: A schematic diagram of the electronic and spatial configurations of the GaAs (110) surface. The As atoms have moved outward and the Ga atoms inward compared to the positions in the bulk of the crystal.
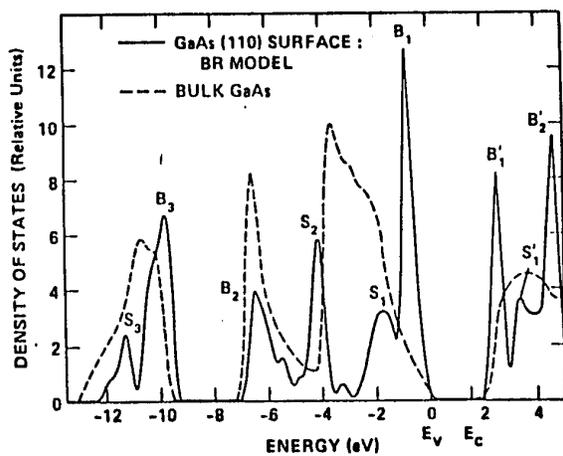

Figure 12.10: The local density of surface states (solid line) for the bond relaxation model of the GaAs (110) surface. Electronic states located on the first two layers of Ga and As atoms are shown.
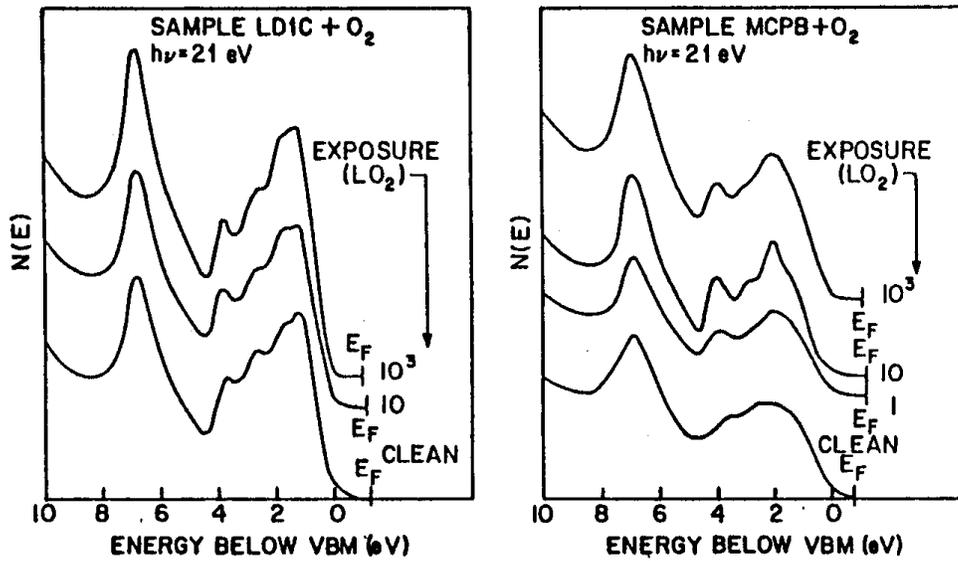
148

Figure 12.11: Energy distribution curves from the upper part of the valence band of two different GaAs (110) crystals. The effects of a small oxygen exposure on the Fermi level pinning and valence band structure are also shown.
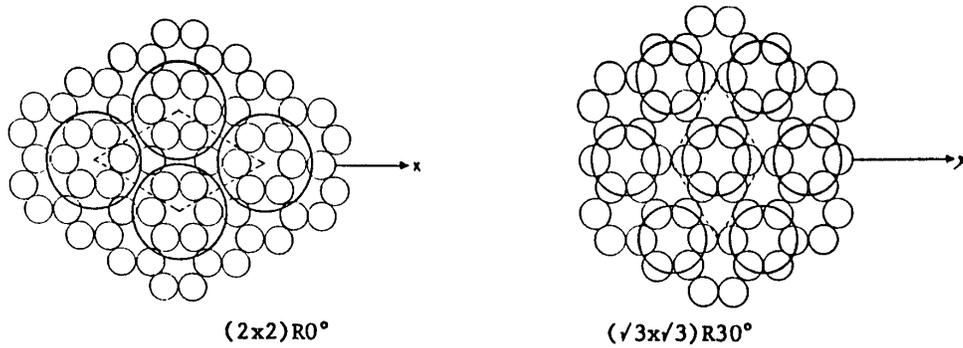
Figure 12.12: This figure illustrates two possible commensurate structures for adsorbed atoms on a honeycomb triangular lattice substrate. (a) The $(2 \times 2)R0°$ structure and (b) the $(\sqrt{3} \times \sqrt{3})R30°$ structure.

149
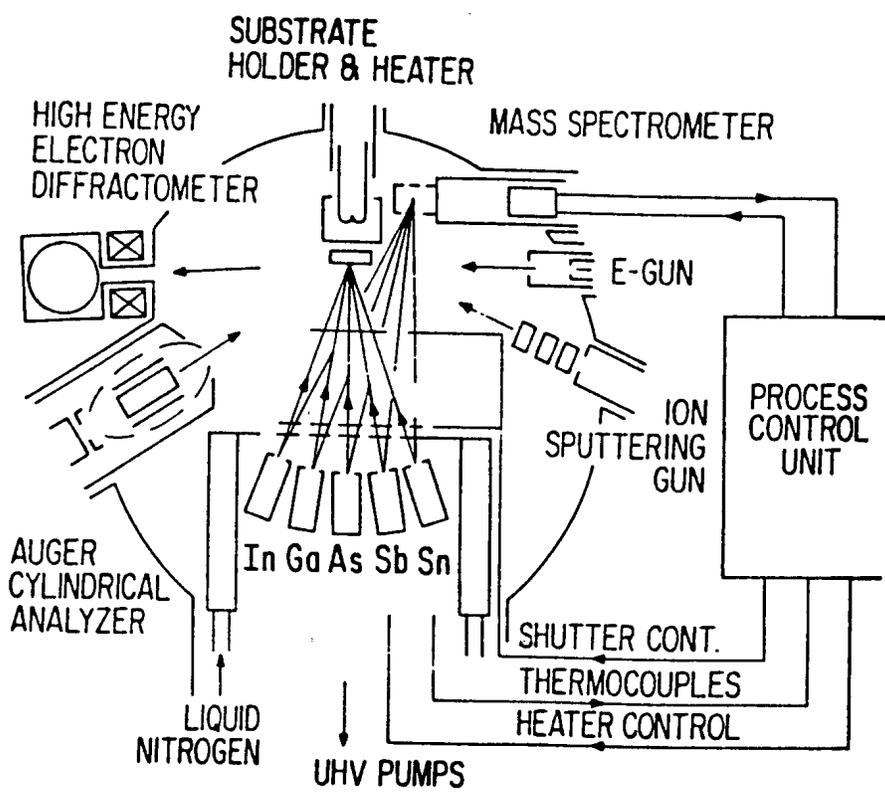
Figure 12.13: Schematic diagram of an MBE system.

have fewer nearest neighbors. In general, one would thus expect the bond lengths (normal to the surface) for the surface atoms to be slightly shorter than in the bulk.

In addition, the LEED techniques can be used to determine the structure of adsorbed species in the coverage range from below one monolayer to perhaps two monolayers. The analysis of LEED patterns in general is more complex than for x-ray diffraction because of multiple electron scattering phenomena. For many surface structures, other evidence in addition to LEED data is needed to determine the surface structure unambiguously.

The standard in situ characterization technique for MBE growth of semiconductor superlattices (see Fig. 12.13) is RHEED (reflection high energy electron diffractometry). The RHEED measurements are carried out at almost glancing angles of incidence to accentuate the surface sensitivity. The RHEED measurements provide a diffraction pattern on a fluorescent screen which is used to monitor the growth, providing information on the:

- structure of the growing surface and smoothness of surface

- surface reconstruction

- growth dynamics through observation of intensity oscillations

- evolution of surface impurities.

Some examples of RHEED patterns from the growth of InSb on a CdTe (001) substrate, looking down a [110] direction (glancing angle) are seen in Fig. 12.14. After 5 sec of InSb
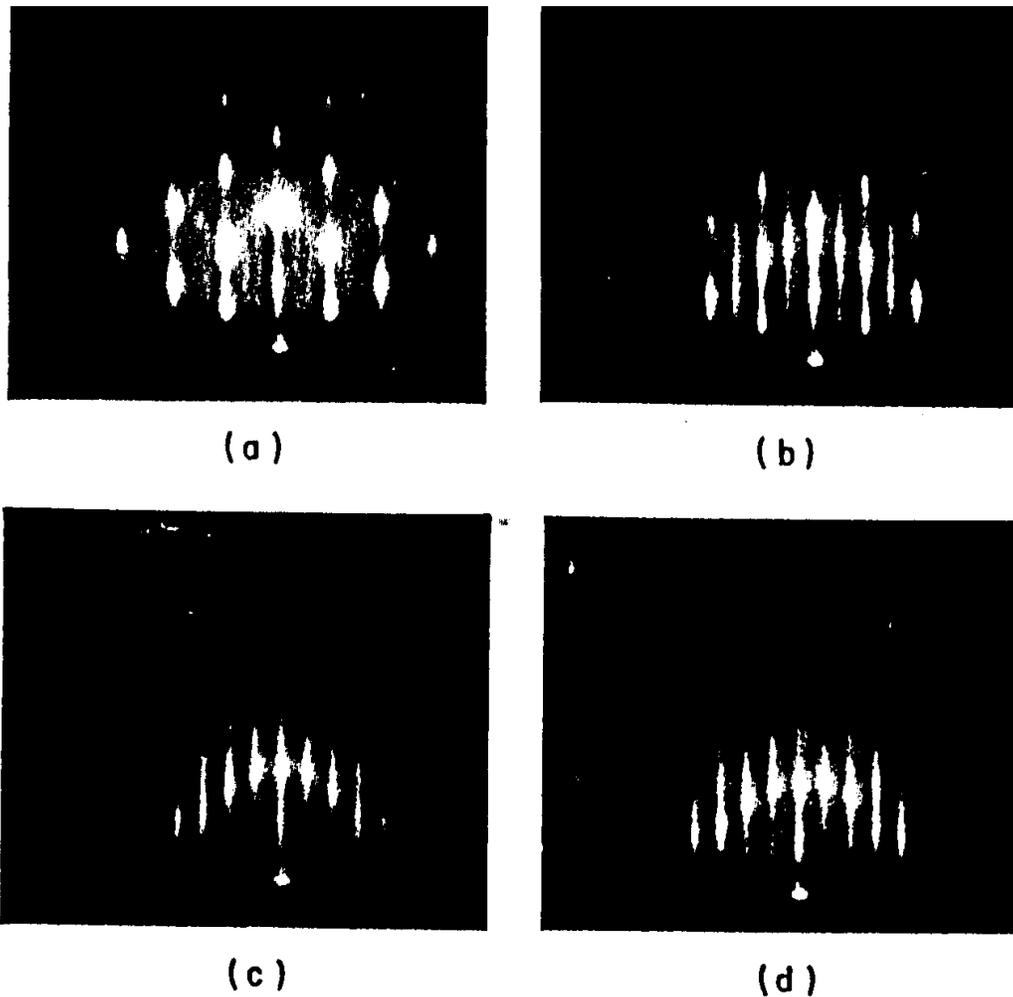
150

Figure 12.14: Reflection high–energy electron diffraction patterns at 20KeV and at a glancing [110] angle characterizing the growth of InSb on CdTe (001): (a) after 5 sec; (b) after 75 sec; (c) after 115 sec; (d) after 160 sec, to yield an InSb quantum well of 160Å. (After L. A. Kolodziejski, Gunshor, Otsuka, Datta, Becker, and Nurmikko, IEEE J. Quantum Electronics, vol. QE22, 1666 (1986)).

growth ($\sim 1\text{Å/sec}$ growth rate), a spotty diffraction pattern is seen (Fig. 12.14a), indicative of the formation of islands of InSb on the CdTe substrate. As the growth proceeds over a ~ 75 sec time period, the islands get large enough to join up and form a smooth surface, yielding the more uniform streak pattern. This streak pattern develops further with increasing deposition time. The diffraction patterns for later times (75 sec and beyond) show a central diffraction line characteristic of the bulk material, and some weak sidebands characteristic of the surface reconstruction. The surface reconstruction sidebands indicate a surface structure that is also periodic but with a different structure than the bulk, due to the relaxation of the solid to tie up the dangling bonds at the surface.

12.2.3 Electron Energy Loss Spectroscopy, EELS

Electron energy loss spectroscopy (EELS or ELS) is another commonly used technique in surface science. In electron energy loss spectroscopy, a primary electron with an energy of perhaps 100 eV will excite an electron in a filled initial state to an empty excited state. The electronic structure of the valence band states is determined by examination of the energy spectrum of the emitted secondary electrons. In the interpretation of these energy loss studies, no correction need be made for the work function for the electron, since the same potential energy drop is experienced for both the primary and secondary electrons at the surface potential barrier.

The EELS technique is conceptually the same as Raman scattering (§10.4 of notes Part II) or inelastic neutron scattering. An incident electron of energy E_i is scattered by an electron in the solid, imparting (or absorbing) an energy $\hbar\omega$ to (from) the electron in the solid to achieve an energy E_f for the scattered electron using conservation of energy:

$$E_i - E_f = \hbar\omega. \tag{12.22}$$

Likewise momentum is conserved to yield the relation

$$\vec{k}_i - \vec{k}_f = \vec{q}. \tag{12.23}$$

Unlike the case of Raman scattering, the incident electrons in the EELS experiment can have a large range of wave vectors $\vec{k}_i$ so that the change in momentum for the electron in the solid can be comparable to Brillouin zone dimensions. Since the incident electrons typically have energies up to $\sim 100\text{eV}$ with wave vectors up to $\vec{k}_i \approx 5\text{Å}^{-1}$, the EELS technique can probe a wider wave vector range in the $E(\vec{k})$ diagram than is commonly probed in an optical reflectivity measurement.

EELS is different from optical absorption and Raman scattering in that it is sensitive to different aspects of the electronic structure of solids because the probe is a charged particle rather than a photon. An incident light wave is characterized by its electric field $\vec{E}$. The rate of absorption of the light or the power loss is proportional to

$$Im(\vec{E} \cdot \vec{D}) \propto Im(\varepsilon)|\vec{E}|^2 \propto \varepsilon_2(\omega)|\vec{E}|^2 \tag{12.24}$$

where the imaginary $Im(\varepsilon) = \varepsilon_2(\omega)$ and $\vec{D} = \varepsilon\vec{E}$ is the displacement vector while $\varepsilon = \varepsilon_1 + i\varepsilon_2$ is the complex dielectric constant. Optical absorption occurs preferentially at peaks in $\varepsilon_2(\omega)$. On the other hand, an incident electron sets up a free charge density $\rho(r)$ which determines the displacement vector $\vec{D}$ through the Maxwell equation

$$\vec{\nabla} \cdot \vec{D} = 4\pi\rho \tag{12.25}$$

152

Figure 12.15: Schematic diagram of the Auger process. $E = E_{\ell'} - E_{\ell} - E_{\ell''}$
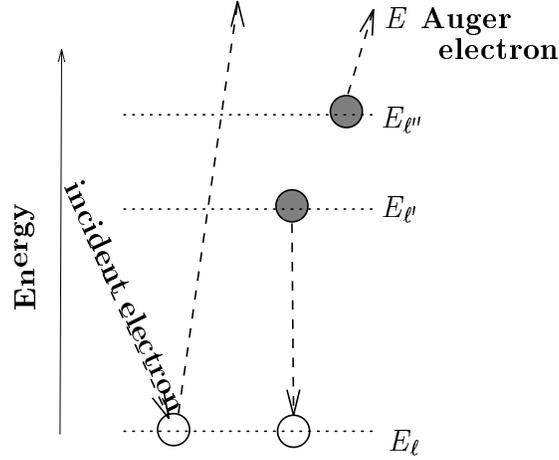
For an incident electron beam, the electron energy loss rate is proportional to

$$Im(\vec{E} \cdot \vec{D}) = Im(\frac{1}{\varepsilon})|\vec{D}|^2 = \frac{\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2}|\vec{D}|^2 \tag{12.26}$$

Thus in the EELS experiment peaks in the energy loss rate occur both near peaks in $\varepsilon_2(\omega)$ (if $|\varepsilon_1(\omega)| \gg |\varepsilon_2(\omega)|$ in this frequency region), and also near zeros in $\varepsilon_1(\omega)$ (where $\varepsilon_2(\omega)$ often remains small), corresponding to longitudinal modes. The most important longitudinal mode is the plasma mode, which is usually a prominent feature in EELS data.

The standard EELS technique is limited by the small penetration depth of the electrons relative to that for photons. EELS is thus primarily a surface technique. By using electron beams of higher energy and near normal incidence, EELS can be applied to study the electronic structure of the bulk. To emphasize the electronic structure near the surface, low energy electrons are used at grazing angles of incidence.

12.2.4 Auger Electron Spectroscopy (AES)

To determine the chemical species present on a surface, Auger Electron Spectroscopy (AES) is commonly used. In this technique an electron beam of several keV is incident on the surface. An electron in this primary incident beam will excite an inner core electron, creating a hole in this inner core electron state, which we will label as state ℓ. An electron in some higher-lying state ℓ' in the same atom will quickly fall into the hole state ℓ. At the same time, a second electron in a state ℓ'' also in the same atom will be ionized. This second electron will acquire kinetic energy E such that energy is conserved in the total Auger process. Thus in the Auger process (see Fig. 12.15) there are three quantum states involved: ℓ, ℓ', ℓ''. What is measured is the energy spectrum of the emitted secondary electrons. In this spectrum, peaks in the intensity profile are identified with specific core states of the atom participating in the Auger process. Since each atomic species has its own characteristic core state spectrum, the Auger spectrum provides an excellent tool for the identification of atomic species. Furthermore, the intensity of the Auger lines provides a measure of the concentration of each atomic species. Because of the strong interaction

153

Figure 12.16: Schematic diagram for the x-ray fluorescence process.
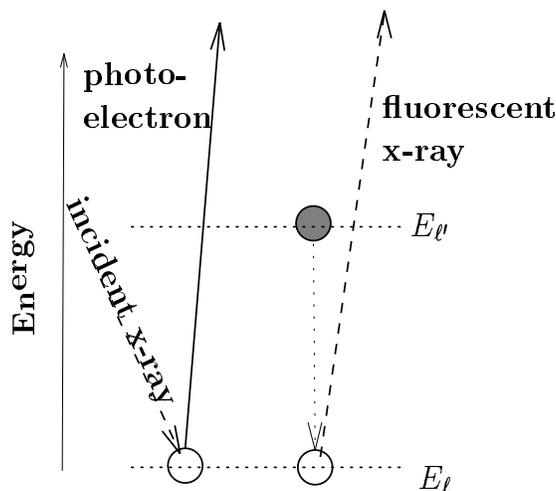
of electrons with matter, electrons will be emitted only from atoms near the surface and for this reason Auger electron spectroscopy preferentially studies the chemical species at or near (within $\sim 40\mathring{A}$) the free surface. Typical Auger electron spectroscopy equipment contains an Argon ion sputtering gun, permitting the removal of surface atoms so that the Auger experiments can be carried out as a function of depth into the surface by a method called depth profiling. Use of the depth profiling technique is destructive to the sample, leaving a tiny hole behind.

X-ray fluorescence measurements are also used to identify the chemical species present in a given species. In the x-ray fluorescence technique, x-rays (energies ~ 50keV) are incident on a sample and eject an electron from a core state ℓ in the atom of the solid. An electron in a level ℓ' (see Fig. 12.16) will fall into the state ℓ, releasing a photon with energy $E_{\ell'} - E_\ell$ which is measured.

The emitted x-rays thus have characteristic energies corresponding to the core level for each atom, thereby allowing identification of the chemical species. The intensity of the characteristic emission lines are related to the concentration of each of the chemical species. X-ray fluorescence (see Fig. 12.16) is a non-destructive method for chemical analysis and has a penetration depth of $\sim 1\mu$m. This technique is usually available when doing scanning electron microscopy (SEM) studies and is called EDX.

12.2.5 EXAFS

Another powerful experimental method for studying solids is Extended X-ray Absorption Fine Structure (EXAFS), which has been applied to the study of surfaces. Figure 12.17 shows a typical x-ray absorption spectrum, in the region of an inner-shell ionization energy. The photon energy at the absorption edge is equal to the minimum of threshold energy required to excite an electron from an inner tightly-bound atomic level to an unbound or continuum state. The region of higher photon energies, which is about 1-2 keV above the absorption edge, is called the EXAFS region; small oscillations in the absorption strength are produced by interference between the wave function for the outgoing electron state and the wave-functions reflected from neighboring atoms in the solid or on the surface. (See

154

Figure 12.17: Schematic of x-ray absorption spectrum showing the threshold region (including pre-edge and edge region) and the EXAFS spectrum.

155
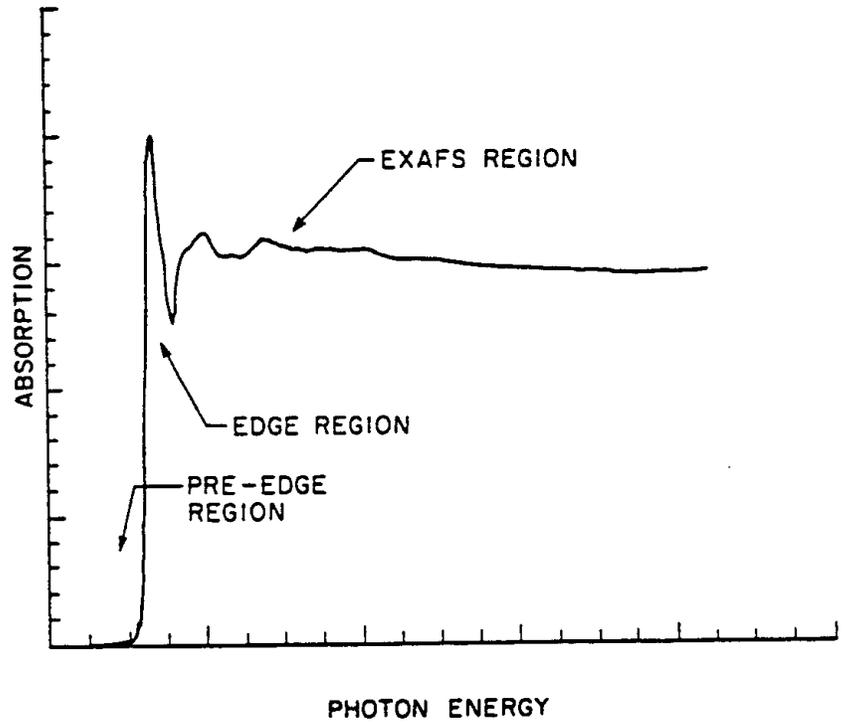
EXAFS structure is shown in Fig. 12.17).

To analyze the EXAFS region, the final state of the electron $\phi_f(r)$ is written as

$$\phi_f(r) = \phi_{f,\ell}(r) + \Sigma_{k_j}\phi_{SC}(r - R_j) \tag{12.27}$$

where $\phi_{f,\ell}(r)$ is the final state of the absorption process, centered at the particular atom with angular momentum quantum number ℓ and which corresponds to a particular energy E above the threshold, and ϕ_{SC} is a reflected wave corresponding to the scattering of the electron in state $\phi_{f,\ell}(r)$ from a neighboring atom at position R_j. If one knows, from atomic physics experiments or theory, the phase shifts δ_j corresponding to scattering from each of the neighboring chemical species, then the position R_j of these neighbors can in principle be determined, since the EXAFS amplitude is proportional to: (see a chapter on scattering theory in a quantum mechanics text such as Schiff, Chapter 5 or Sakurai, Chapter 7.)

$$\alpha(E) \propto |\langle\phi_i|\mathcal{H}'|\phi_f\rangle|^2 \propto \sum_j F(|\vec{R}_j|)Im(f(\pi)e^{2ik|\vec{R}_j|}e^{2i\delta_j}) \tag{12.28}$$

where ϕ_i is the initial atomic state, $\mathcal{H}'$ is the optical perturbation Hamiltonian, F is a smooth function of $|\vec{R}_j|$, and $k \propto (2mE/\hbar^2)^{\frac{1}{2}}$ is the wave vector of the ejected electrons, while $f(\pi)$ is the amplitude for scattering the ejected electron at $180°$, back towards the emitting atom. Because of the factor $\exp(2ik|\vec{R}_j|)$, the EXAFS amplitude is essentially the Fourier transform of the probability distribution of nearest-neighbor separations $|\vec{R}_j|$. The Fourier transform of the probability distribution, is then compared to the function calculated for a given model of the structure to test the validity of that model. In Eq. 12.28, ϕ_f is the wavefunction for the final state and δ_j is phase shift for an atom at position j. Figure 12.18 gives an example of the power of the EXAFS technique. Here EXAFS spectra are shown for a clean ruthenium surface and for surfaces exposed to two different oxidation conditions. The analysis of each trace is shown on the right to extract the pertinent nearest neighbor distances.

Equipment to carry out the various surface science experiments mentioned above is very expensive and the techniques are generally useful in many areas of solid state research. Therefore we have a number of experimental systems available through the Central Facilities of the Center for Materials Science and Engineering (Building 13). The LEED, ESCA (Electron Spectroscopy for Chemical Analysis) and AES equipment is in the Surface Analytical Laboratory (4th floor), EELS measurements can be done with the transmission electron microscope (basement 13-1027) and x-ray fluorescence measurements can be made with the scanning electron microscope using the KEVEX attachments (2nd floor) or with an electron microprobe (basement level). A range of scanning electron microscopes (SEM), transmission electron microscopes (TEM), and scanning transmission microscopes (STEM) are also available in the Building 13 Central Facilities.

12.2.6 Scanning Tunneling Microscopy

The scanning tunneling microscope (STM) provides a unique and powerful new tool for the direct determination of real space surface structure at the atomic level, including nonperiodic structures. In this microscope, a small metal tip is brought close enough to the surface to permit electron tunneling between the tip and the surface. The tip scans the surface in two–dimensions (hence the name scanning tunneling microscope). By adjusting the height

Figure 12.18: Extended absorption edge fine structure (EXAFS) is shown in (a) for thinly dispersed pure ruthenium metal, (b) for ruthenium partially covered with O_2 at $25°$, and (c) for ruthenium mostly converted to RuO_2 at $400°C$. Curves d, e and f are the Fourier transforms of a, b and c, from which the number and bond distances of nearest-neighbor atoms may be derived.
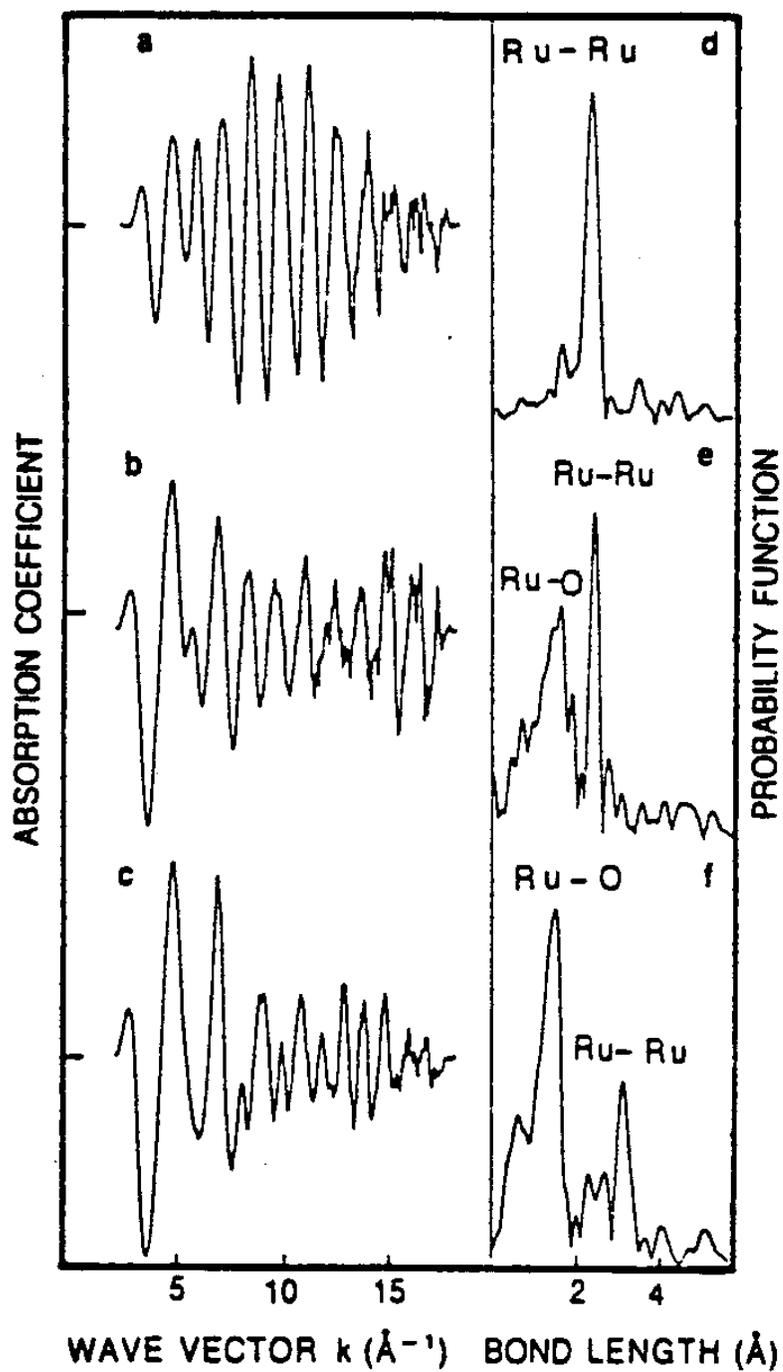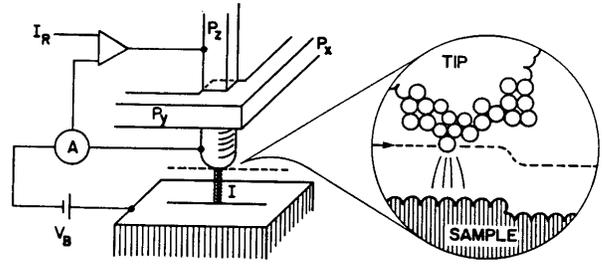
Figure 12.19: Schematic of a scanning tunneling microscope and of its operation.

of the tip above the surface to maintain a constant tunneling current, it is possible to obtain a contour map of the surface. The announcement of the successful observation of surface structure on an atomic scale with the STM in 1982 created tremendous excitement in the solid state community.

Some useful references and reviews are listed below.

- *IBM Journal of Research and Development* **30**, #4 & #5 (1986)

- G. Binnig and H. Rohrer, *IBM Journal of Research and Development* **30**, 355 (1986); *Scientific American*, August 1985, p. 50.

- J.E. Demuth, R.J. Hamers, R.M. Tromp and M.E. Welland, *IBM Journal of Research and Development* **30**, 396 (1986)

- J.E. Demuth, Physics in a Technological World, French, Editor, AIP, NY 1988.

- *Inelastic Electron Tunneling Spectroscopy*, T. Wolfram, editor, Springer Series in Solid–State Sciences 4, Springer, Berlin (1977).

- *Tunneling in Solids*, C.B. Duke, ed. by H. Ehrenreich, F. Seitz, and D. Turnbull, *Solid State Physics, Supplement #10* Academic Press, New York (1969).

- M. Amrein, A. Stasiak, H. Gross, E. Stoll, and G. Travaglini, *Science* **240**, 514, (1988).

Referring to Fig. 12.19, the piezo–drives P_x and P_y scan the metal tip over the surface. The control unit (CU) applies the appropriate voltage to the piezo–drive P_z to maintain a constant tunneling current J_T at constant tunnel voltage V_T. For a constant work function, the voltages applied to the piezo–drives P_x, P_y, and P_z yield the topography of the surface directly, whereas modulation of the tunnel distance z by Δz gives a measure of the work function. The dotted line in Fig. 12.19 indicates the z displacement in a y scan at a surface step.

The very high resolution of the STM depends on the exponential dependence of the tunneling current on the distance z between the tip and the scanned surface. If Φ is the average barrier height for tunneling (the average work function $\Phi = (\Phi_1 + \Phi_2)/2$ between the tip and the surface), then the tunneling current is given by

$$J_T = J_0 \, exp \, (-2\kappa z) \tag{12.29}$$

where

$$\hbar^2 \kappa^2 = 2m\Phi \qquad (12.30)$$

and m is the free electron mass, so that

$$J_T = J_0 \; exp \left(-A\Phi^{1/2}z\right) \qquad (12.31)$$

where

$$A = 2(2m/\hbar^2)^{1/2} = 1.025 \quad \text{Å}^{-1}\text{eV}^{-1/2} \qquad (12.32)$$

and

$$J_0 = (e^2/\hbar)(\kappa/4\pi^2 z). \qquad (12.33)$$

The tunneling probe tip in a typical instrument may be prepared to have a radius between a few thousand angstroms to $1\mu m$, but containing some sharp mini-tips close to the atomic limit as shown schematically in Fig. 12.19. The extreme sensitivity of the tunneling current on the gap width selects the extremal mini-tip protuberance ($\simeq 10$Å) for operation of the STM.

The very fine tips of the STM can resolve monatomic steps within 10 Å lateral resolution, as indicated in Fig. 12.20 for the case of steps on a clean Au (100) surface (G. Binnig, H. Rohrer, Ch. Gerber and E. Stoll, *Surface Science* **114**, 321 (1984)). To give the order of magnitude of the sensitivity of the STM due to the exponential dependence of the current on distance z, an increase in distance of 1 Å results in an order of magnitude decrease in tunneling current.

From Eq. 12.31, we see that scanning the tunneling tip at constant tunneling current implies that $\Phi^{1/2}z = $ constant. Thus for constant work function of the substrate, the displacement in the z direction is adjusted to yield $z = $ constant during the scan. The voltage applied to the piezo–drive P_z to achieve $z = $ constant thus provides a record of the surface topography along the scan. Figure 12.20 shows scanning tunneling micrographs of a gold (100) surface, exhibiting atomically flat terraces with monolayer steps. The tunneling current is sensitive not only to the topographical features but also to the local electronic structure. A good deal of effort has been devoted to disentangling the contributions to the tunneling current from each of these two effects.

Electronic and chemical surface properties manifest themselves primarily in the voltage dependence of the tunneling current. They appear as specific features in the local $I - V$, $V - z$ or $I - z$ characteristics, where z is the distance between the tip and the surface. In practice, electronic or chemical images are obtained by recording dI/dV or dI/dz while scanning and controlling the gap width, and while keeping the average current constant. Depicted in Fig. 12.21 is an example of such scanning tunneling spectroscopic (STS) imaging on a Ni (100) surface (R. Garcia, J.J. Saenz and N. Garcia, *Phys. Rev.* **B33**, 4439 (1986)). In trace (a) showing a plot of dI/dV vs V, the strong peak at 0.8 V is attributed to surface nickel oxide. In (b), the surface is imaged with respect to that spectroscopic feature (top) by taking scans of dI/dV at 0.8 V in the y direction. On the bottom of (b), topographic $I = $ constant scans are taken at the indicated voltages. Whereas the STS images differ dramatically, the STM images remain essentially unchanged. Because of the usually close relation of the tip–to–sample spacing at $I = $ constant with the topography, of dI/dV with the local density of states, and of dI/dz with the local barrier height (or work function), it is possible to separate the various effects observed with STM. Associated images are
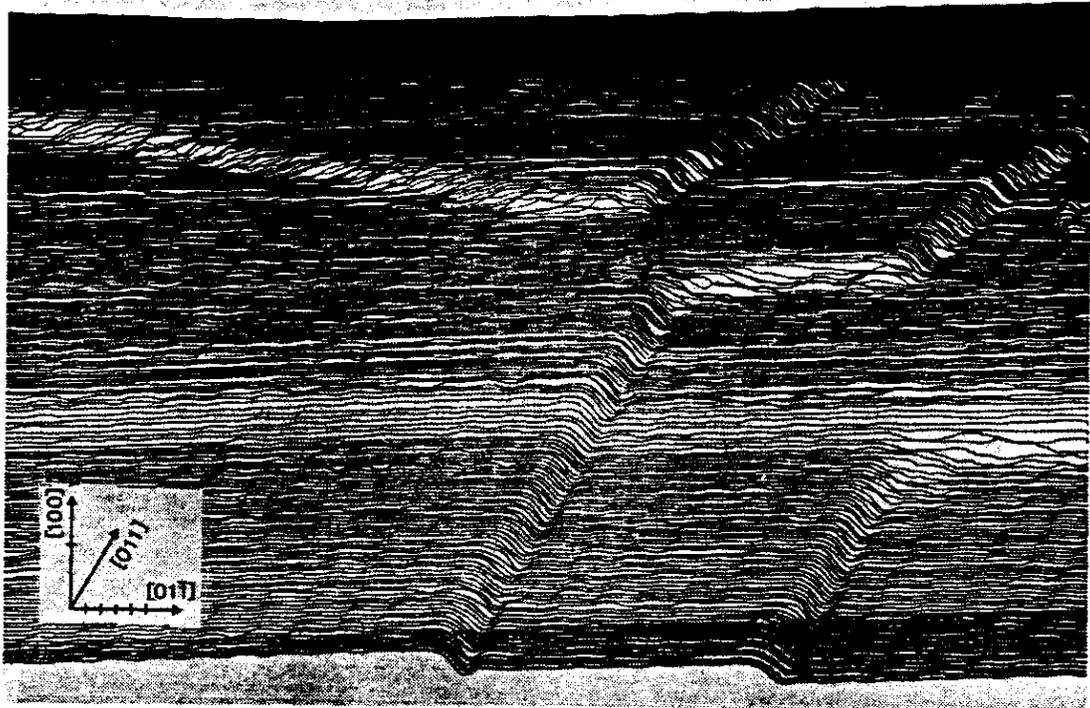
Figure 12.20: STM image of a clean Au (100) surface obtained at a constant tunneling current of 1 nA, showing the terraces and monolayer step lines. The wavy structure can be resolved into individual atomic rows. The divisions on the axes correspond to spacings of 5 Å(G. Binnig, H. Rohrer, Ch. Gerber and E. Stoll, *Surface Science* **114**, 321 (1984)).

Figure 12.21: Spectroscopic and structural imaging of NiO on a Ni (100) surface. Shown in (a) is dI/dV vs. V from an oxide–covered region. The strong peak at 0.8 V is characteristic of NiO. The STS and STM images shown in (b) were obtained at the indicated bias voltages of 0.8 and 1.3 V. An oxide island to the left is evident in the STS image obtained at 0.8 V. Spatial separations in units of the NiO lattice spacing are indicated at the left (bottom). The oxide island is hardly noticeable in the STM images. The divisions on the y axis correspond to spacings of 5 Å (R. Garcia, J.J. Saenz and N. Garcia, *Phys. Rev.* **B33**, 4439 (1986)).
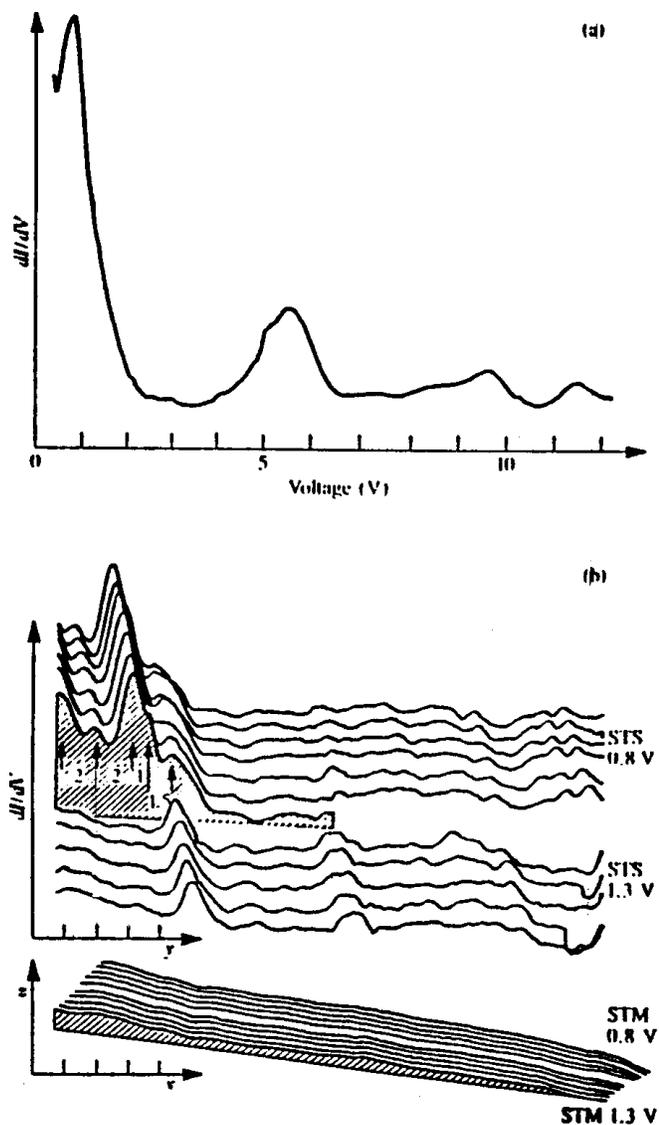
Figure 12.22: STM relief of a (111) Si surface showing the (7×7) unit cells with a superposed model for the (7×7) structure. The diagram covers a scan area of 60Å×60Å. Original work by G. Binnig, H. Rohrer, Ch. Gerber and E. Weibel, *Phys. Rev. Lett.* **50**, 120 (1983).
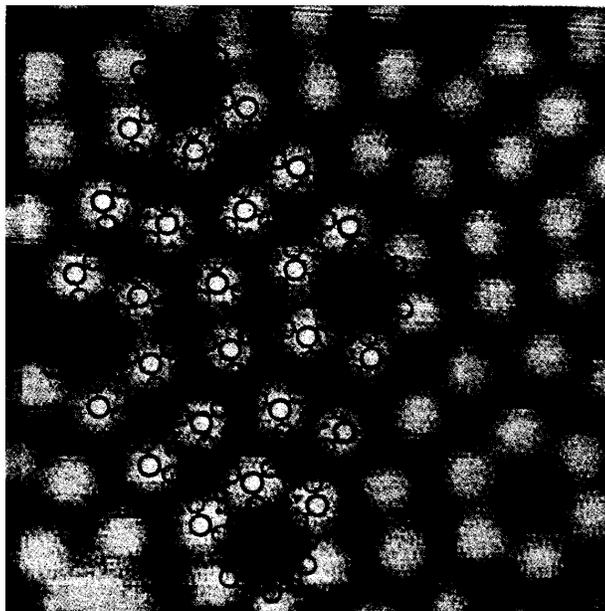
often referred to as "topographical or STM" images, "spectroscopic or STS" images, and "work–function profiles", respectively.

An important aspect of the STM is its apparent nondestructive nature. In normal operation, no perceptible irreversible damage to the sample surface occurs as a result of its use. On the other hand, the STM can be used to intentionally induce permanent local structural or chemical modifications.

One of the most noteworthy achievements of the STM has been the direct observation of surface reconstruction in silicon (G. Binnig, H. Rohrer, Ch. Gerber and E. Weibel, *Phys. Rev. Lett.* **50**, 120 (1983); G. Binnig and H. Rohrer, *IBM Journal of Research and Development* **30**, 355 (1986); J.E. Demuth, R.J. Hamers, R.M. Tromp and M.E. Welland, *IBM Journal of Research and Development* **30**, 396 (1986)). Surface reconstruction pertains to a different surface structure relative to that of the bulk due to the broken bonds at the surface discontinuity. It has been known for some time that the surface structure of semiconductors differs from that of the bulk because of the different number of nearest neighbors available for bonding. It had been conjectured that the (111) surface of Si when heated above 900°C exhibits a (7×7) surface reconstruction, though many uncertainties remained about whether the (7×7) reconstruction was correct and where the atoms were located within the unit cell. With the STM, the (7×7) surface reconstruction has been vividly demonstrated, as shown in Fig. 12.22. This picture of the (7×7) reconstruction was obtained on a sample previously heated at 900°C in high vacuum to remove any surface oxide layer. The STM micrograph was taken at a 2.9 *V* positive tip potential. The (7×7) rhombohedral unit cell is clearly seen in the scan in Fig. 12.22, bounded by lines on minima with deep corners. Each unit cell contains 12 maxima and the diagonals are determined to be 46 ± 1 Å and 29 ± 4 Å, in good agreement with the crystallographically determined values of 46.56 Å and 26.88 Å, respectively.

Figure 12.23: Schematic diagram of the Atomic Force Microscope (AFM). An STM is used to measure the displacements of the probe tip as it scans the surface of an insulating sample.
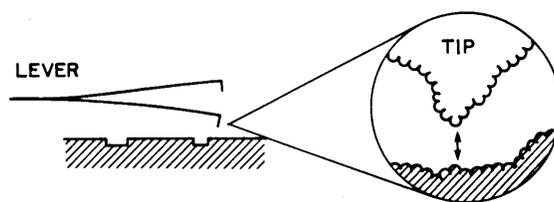

Figure 12.24: AFM images of grooves etched in Si for three increased magnifications. The highest magnification scans show the details of the groove substructure which is only 50Å high. (Y. Martin, C.C. Williams, and H.K. Wickramasinghe, *J. Appl. Phys.* **61**, 4723 (1987)).

Following the successful development of the scanning tunneling microscope, several related instruments have emerged. One of the more important of these instruments is the atomic force microscope (AFM). The forces measured by this instrument are the interatomic forces between the surface atoms under investigation and the apex atoms of a very sharp diamond tip fixed on a conducting cantilever (see Fig. 12.23). Bending of the cantilever by the interatomic forces is monitored by the tunneling current between the cantilever and an STM tip. Scanning the diamond tip across the conducting or insulating surface under investigation at constant interatomic force yields a topographical image of the surface. The AFM image is composed of contours of constant force between the imaged surface and a probe tip, and permits measurement of surface contours of both conductors and insulators; the present versions of the STM are not able to make topographical maps of insulators.

Although atomic resolution has been achieved on graphite surfaces, this technique will be more widely exploited in the study of larger-scale features on real surfaces under normal laboratory conditions. Recent work at IBM (see Fig. 12.24) shows an application of the AFM to examine the groves in a reactive-ion-etched silicon wafer after oxidation in air. The different microstructures on top of and in the grooves are associated with the etching method and are clearly resolved by the AFM over a wide range of scales, even though the surface is insulating. At the highest resolution, features as small as 50Å high are easily

163

seen.

Chapter 13

Amorphous Semiconductors

References

- Kittel, Introduction to Solid State Physics

- Ziman, Models of Disorder, Cambridge, 1979.

13.1 Introduction

There are many materials which are of scientific and technological interest which are not single crystals, or even microcrystalline. The general category of amorphous materials is defined as including materials which have no crystalline order; that is, their x–ray diffraction patterns consist of thick diffuse rings or halos instead of sharp spots. Usually excluded from this definition are polycrystalline materials which consist of small crystallites in random orientations. Since, as we shall discuss below, amorphous materials often do possess considerable short–range order, the distinction between polycrystalline and amorphous materials blurs as the crystallite size is reduced. There are a number of substances, such as Ge and Si, which can be prepared in all three forms: single–crystal, polycrystalline and amorphous.

Amorphous materials can be prepared in several ways, depending on the material. Materials called glasses can be formed by cooling from the liquid state. These materials are not in thermal equilibrium and can be classified as super-cooled liquids with an extremely high viscosity. Some materials cannot be cooled fast enough to avoid crystallization. Splat cooling (or rapid solidification) is often used to prepare metallic glasses. In this process a liquid stream is shot onto a thermally conducting substrate. Some of the splat–cooled metals exhibit microcrystalline ordering. There are also a number of techniques for deposition from the vapor state, such as sputtering, or for condensation from a chemically reactive vapor such as silane (SiH_4) to prepare (hydrogenated) amorphous Si.

Many amorphous materials can be called semiconductors in the sense that they are neither good conductors nor good insulators, but instead are poor conductors. Many are also similar to their crystalline counterparts in that they possess an optical gap. The reason for this general behavior seems to be that, even though the amorphous structure is quite random over long distances, there still seems to be considerable short–range order with local bonding requirements generally satisfied. Although these materials are full of defects,
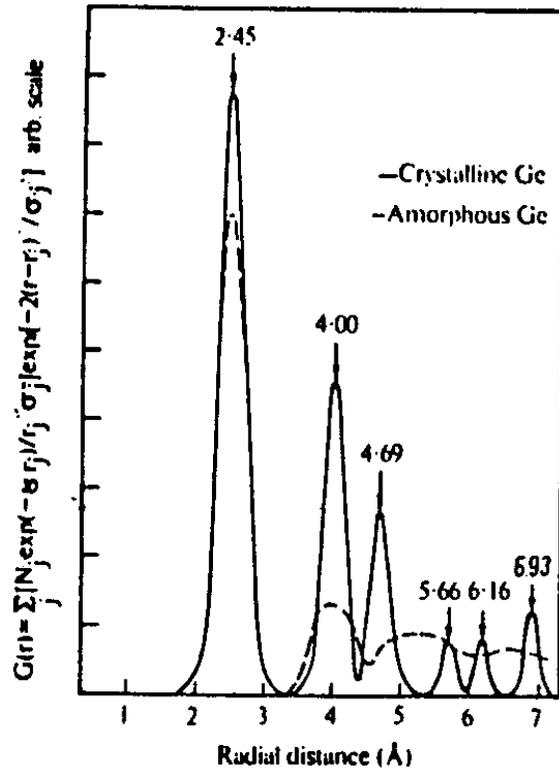
Figure 13.1: Distribution of neighbor distances $G(r)$ for Ge, from EXAFS data, shown for both crystalline and amorphous germanium.

there are not many electrons which are involved in the process to carry electric current. In addition, because of the spatial disorder resulting in strong carrier scattering, the carriers mobilities are low. A further consequence of the tendency for bonding to be satisfied is that the electrical properties of many amorphous semiconductors tend to be insensitive to the presence of large concentrations of impurities.

There are two major categories of amorphous semiconductors. The first consists of the tetrahedrally bonded materials, primarily Ge and Si but also including amorphous III–V semiconductors. The second major category consists of the chalcogenide or lone–pair semiconductors, which means the elements Se, S or Te (column VI) and compounds and alloys containing these elements.

13.1.1 Structure of Amorphous Semiconductors

The major attribute of the structure of amorphous materials is the lack of long–range order or the absence of a periodic lattice. In spite of this, there is considerable similarity in the <u>local</u> environments of amorphous and crystalline materials. For example, the EXAFS measurements for Ge shown in Fig. 13.1 indicate that the first and second–nearest–neighbor distances are the same and that differences only appear at the third–and higher–neighbor distances.

Not only are these short–range bond lengths usually preserved, but also bond <u>angles</u> tend to be the same. Some amorphous materials have been modeled successfully as continuously

166

Figure 13.2: The diamond structure showing the atoms 1 and 2 are in a staggered configuration and are crystallographically distinct in the perfect crystal.
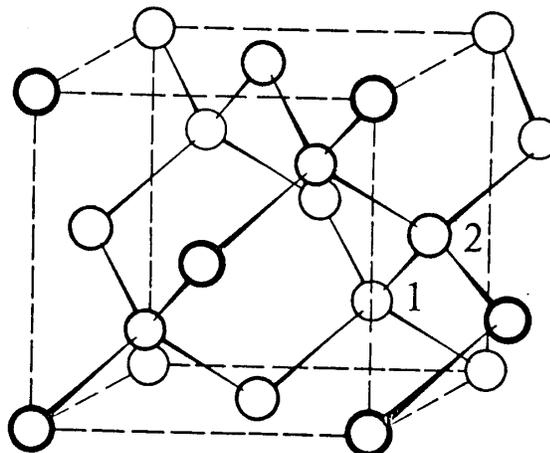
perturbed from their crystalline form. However there is also some evidence from x–ray diffraction for more drastic changes. For example, one can view crystalline germanium or silicon as containing distorted 6–fold rings (see Fig. 13.2 for the diamond structure). Amorphous germanium and silicon also seem to contain some 5–fold rings.

13.1.2   Electronic States

It is evident that Bloch's theorem no longer holds in amorphous materials; hence electronic states cannot be characterized by a $\vec{k}$ vector confined to a single Brillouin zone. In other words, $\vec{k}$ is no longer a good quantum number. Thus one can no longer use the powerful energy–band theory which predicts bands of extended electronic states with forbidden gaps, and which we use to differentiate conductors from insulators.

In practice the electronic properties of an amorphous material do not differ as drastically as one might expect from those of the crystalline material. Consider the data for the resistivity of a number of materials at temperatures near their melting points, as shown in Fig. 13.3. Where small jumps do occur at the melting point, these are correlated with small discontinuous volume changes. When the volume does not change, as for HgTe or CdTe, neither does the resistivity. Evidently the conduction properties are remarkably similar even though the structure has changed drastically. We note here that for the column IV semiconductors (Si, Ge, Sn), the molten material is metallic and octahedrally coordinated.

Without the powerful simplification of Bloch's theorem, it is extremely difficult to calculate the electronic states in amorphous materials. Without a $\vec{k}$–vector one cannot calculate $E(\vec{k})$. Instead one attempts to calculate directly a density of states $\rho(E)$, and also an average energy–dependent mobility $\mu(E)$. One starts by noting the similarity of amorphous and crystalline electronic properties and especially their similarity with respect to short–range order. Recently much progress has been made using various types of cluster models. The general result, as shown in Fig. 13.4, is that the crystalline density of states, which has sharp features called Van Hove singularities at critical points where $\partial E/\partial\vec{k} = 0$, is smoothed out and broadened at the critical points (see §4.2). This is shown experimentally

Figure 13.3: Resistivity as a function of temperature for several materials in both the crystalline and liquid states. The melting temperature in each case is indicated by an arrow.
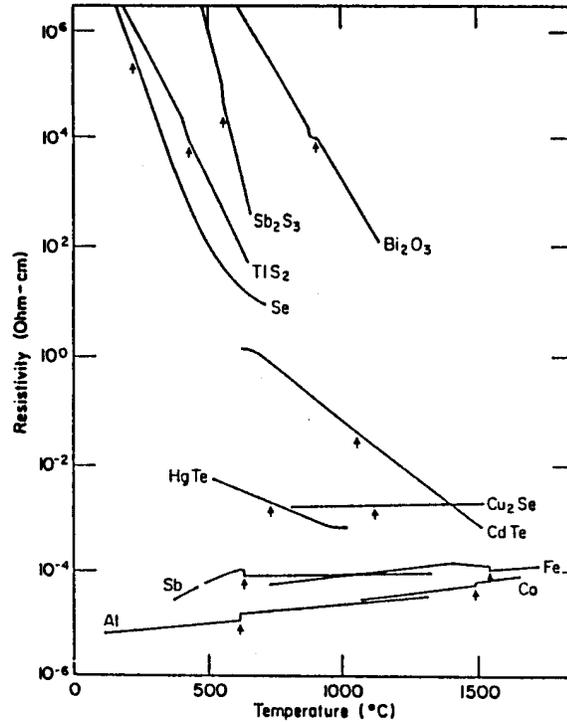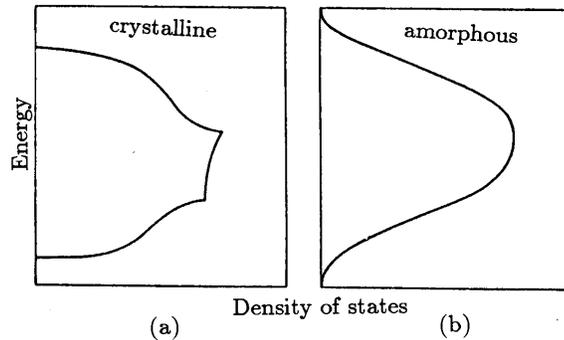

Figure 13.4: (a) Density of electronic states as a function of energy for a single band of a crystalline solid. The sharp behavior at the band edges and in the interior represents the effects of Van Hove singularities. (b) Density of electronic states as a function of energy for a single band of an amorphous solid. All Van Hove singularities have disappeared.
(The Van Hove singularities are the M_0, M_1, M_2, M_3 singularities in the joint density of states discussed in §4.2).)

168

Figure 13.5: Ultraviolet photoemission results (top) for the density of states for trigonal (solid line) and amorphous (dashed line) Se. Photoemission results (bottom) on trigonal (solid line) and amorphous (dashed line) Te.
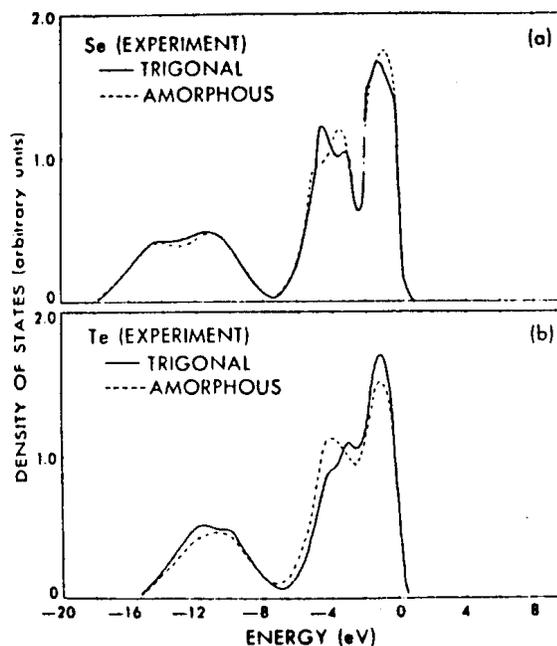
in the photoemission results in Fig. 13.5 for the valence bands of trigonal (crystalline) and amorphous Se and Te.

In order to find the transport properties of an amorphous material, one needs to know not only the density of states $\rho(E)$ at each energy E, but also the mobility $\mu(E)$ which may be a function of $\rho(E)$. Broadly, one distinguishes extended states, similar to states in periodic crystals which have finite amplitude throughout the material, and finite mobility, from localized states which have a significant amplitude only in a small region of the material and have extremely small mobility. An example of such a localized state, in a nearly perfect crystal, is an impurity state in which the electron is localized in a hydrogen–like orbit around a donor ion. Such states have sharp features in a density–of–states diagram near the conduction–band minimum or, for acceptors near, the valence–band maximum, as shown in Fig. 13.6(b). If there is a large enough impurity concentration, these states can broaden into impurity bands which can merge into the conduction or valence bands. However, if the material itself is disordered or amorphous, the states near the band edges are themselves localized.

This subject of localized *vs.* extended states has been treated extensively, most notably by Mott and by Anderson who shared the Nobel Prize in Physics in 1977 for this work. Mott developed the concept of the mobility edge, postulating a relatively sharp demarcation between localized and extended states, giving rise to a mobility gap which is considerably larger than the forbidden gap in the density of states, as illustrated in Fig. 13.7. The mechanism for localization in states near the energy band edges is illustrated in Fig. 13.8, and is due to Fritzsche. According to this model disorder produces a spatial variation in the conduction and valence band edges, giving rise to local conduction–band minima or valence–band maxima which trap electrons or holes.
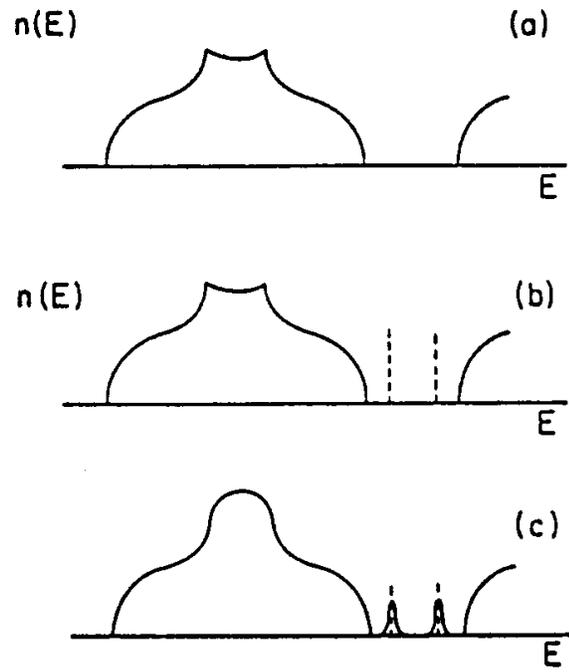
169

Figure 13.6: Density of states $n(E)$ for (a) a perfect crystal, (b) a crystal with a few donors and acceptors, and (c) a crystal with a larger number of imperfections where the impurity levels have broadened into impurity bands.
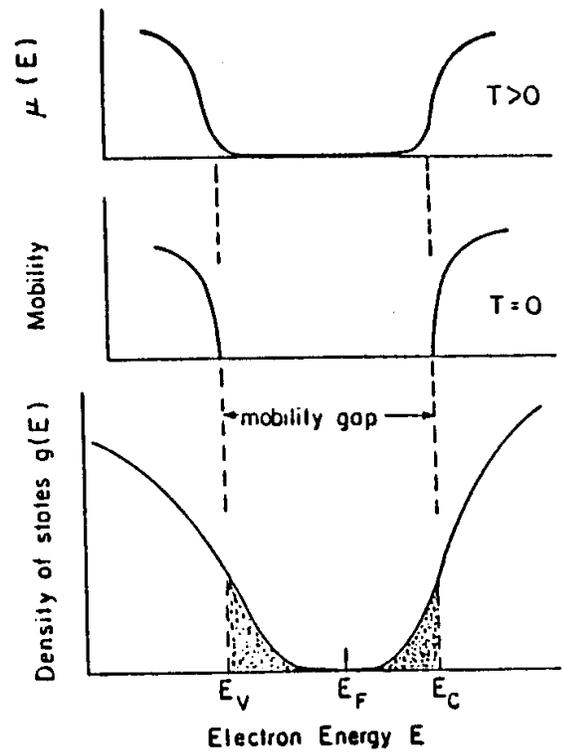

Figure 13.7: Sketch of the Mott–CFO (Cohen–Fritzsche–Ovshinsky) model for covalent disordered semiconductors having a three–dimensional cross–linked network structure. The critical energies E_c and E_v define the mobility gap. For $T > 0$, the mobility $\mu(E)$ may be finite in the gap because of thermally assisted tunneling. Here E_F denotes the Fermi energy. The distribution of localized gap states may be non–monotonic when defect states of a certain energy are prevalent.
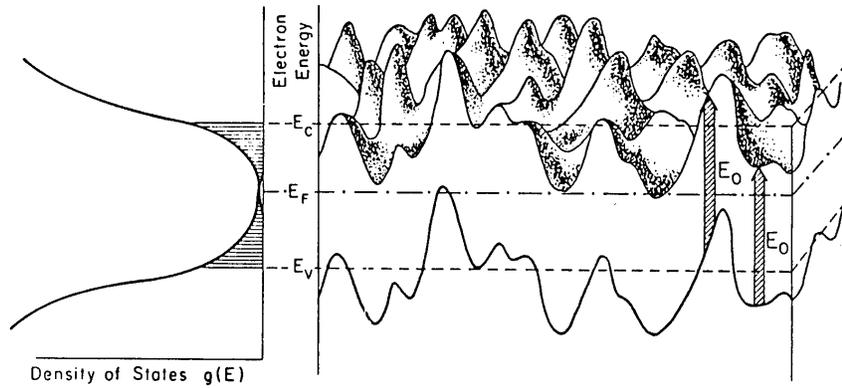
170

Figure 13.8: Potential fluctuations of the initial and final electron states for the optical transitions corresponding to the optical gap E_0. The left hand side shows the density of states. The region of localized states lies between E_c and E_v. Note that the short range potential wells which give rise to many of the localized states are not shown here. This figure shows only that part of the long wavelength potential fluctuations which cause a parallel shift of the valence and conduction band states. The part which causes a spatial variation of E_0 is omitted for clarity (after Fritzsche).

An important difference between the tetrahedrally–bonded amorphous semiconductors and the chalcogenide materials is that the former have large numbers of unpaired spins, as observed in electron spin resonance experiments, and the chalcogenides have no measurable density of spins. A large number of unpaired spins is expected in a material containing a large number of broken or "dangling" bonds, each bond being occupied only by one electron instead of two electrons of opposite spin. The lack of unpaired spins has been explained by the valence–alternation–pair model of Kastner, Adler and and Fritzsche. The chalcogen atoms have 4 electrons in an outer (unfilled) p shell. In the lowest–energy bonding configuration, two of these electrons form bonds with neighboring atoms, and two are in non–bonding or "lone–pair" states. Thus both crystalline and amorphous Se, for example, contain chains of atoms, each bonded to two neighbors. In the amorphous material a Se atom can also be triply–bonded in a trigonal configuration, leaving the fourth p electron in a higher–energy, non–bonding state (with unpaired spin). However, the total energy can be reduced if this extra electron migrates to another triply–bonded Se atom nearby. First, two of the bonds on this second atom break, leaving only one electron in a bonding orbital. The two electrons from the broken bonds join the single electron in the lone–pair orbitals. Then the new electron can enter this atom as a fourth lone–pair electron. Kastner, Adler and Fritzsche argued that this is an energetically favorable configuration. The result, as shown in Fig. 13.9, is a large density of equal number of positive and negative ions but with all electron spins paired. Structurally, this picture implies that chalcogenide glasses contain a large number of linked chains (by triply–bonded atoms) as well as nearby broken chains (ending in singly–bonded atoms), which provides an explanation for the fact that these materials are more resistant to crystallization ("better glasses") than the tetrahedrally–bonded materials.

At the left of Fig. 13.9 two selenium atoms, each of which is triply bonded, serve to
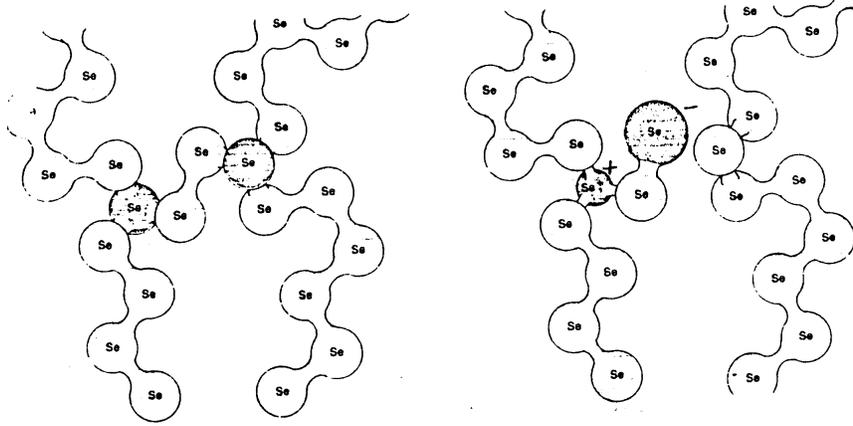
171

Figure 13.9: Valence–alternation pairs can form in a neutral chalcogenide–glass matrix without any major displacement of its atoms, leading to a sharp reduction in total energy.

cross–link two molecular chains of doubly bonded atoms. A valence–alternation pair can be produced (right) by a spontaneous break of the cross–linkage, combined with the simultaneous transfer of an electron from one of the triply bonded selenium atoms to an atom near the one where the cross–link was broken. Since such electronic transfers reduce the total energy of the solid, nearly all the trigonally bonded selenium atoms become members of a valence–alternation pair. Important physical consequences follow, including the almost complete disappearance of electrons with unpaired spins and the appearance of large but equal concentrations of positively and negatively charged traps in chalcogenide glasses. The consequences of this high disorder (in the chalcogenides) for the electronic density of states is shown in Fig. 13.10.

On the left is shown the broadened conduction band and valence bands, with the mobility edges, for an amorphous tetrahedrally bonded semiconductor. Cohen, Ovshinsky and Fritzsche postulated that (whatever the details of the structure), the high disorder in the chalcogenides produces overlapping densities of states, as shown in the right side of the figure, so that electrons will lower their energy by migrating to new localized states, creating large charge separation as in the valence–alternation model.

The distinguishing feature of the bands in amorphous solids is the replacement of the sharp band edges present in crystals by what are called "band tails" or localized states, that extend into the energy gap. The localized states are separated from the extended states in the main part of the bands by "mobility edges". The region that lies between the mobility edges of the valence and conduction bands is the "mobility gap" (see Fig. 13.10). It plays the same role in amorphous semiconductors that the energy gap plays in crystalline semiconductors. Chemical impurities or defects in the configuration of local bands can lead to sharp structural changes (not shown) in the mobility gap. The result of the large density of localized states in the mobility gap is a high density of positively and negatively charged traps, which decrease the mobility of the carriers and make the material less sensitive to efforts to control its conductivity by doping.

This density of states model explains the fact that the chalcogenide glasses are much less sensitive to doping than the tetrahedrally coordinated materials. Impurity states, in-
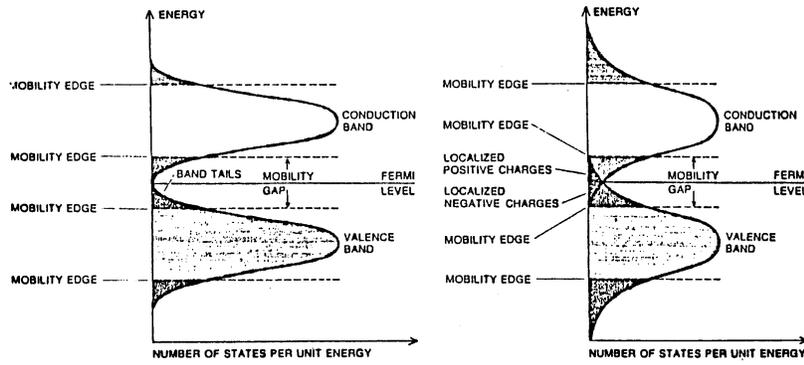
172

Figure 13.10: Amorphous semiconductors that are not strongly disordered (left) have valence and conduction bands similar to those in the corresponding crystalline semiconductor. If the disorder is large, as is expected in multicomponent glasses (right), the band tails of the valence and conduction bands can overlap in the mobility gap. This leads to a redistribution of electric charge as electrons move from one localized state to another in order to lower their energy.

troduced in or just outside the band tails, make only a negligible change in the already–appreciable density of states in this region. If, on the other hand, the bands do not overlap, impurity states in or just outside the band tails can make a large change in the density of states. Thus they can, at finite temperatures, become a source of carriers in the conducting or extended states.

13.1.3 Optical Properties

Amorphous semiconductors have optical spectra similar to their crystalline counterparts in that they possess an optical gap or absorption edge. However, all sharp features, including the band edge absorption, are considerably broadened, as shown for the case of the chalcogenide semiconductor As_2S_3 in Fig. 13.11. The reflectance spectra for crystalline, amorphous and liquid Ge are given in Fig. 13.12. These data show that the amorphous material more closely resembles the crystalline material than the liquid which shows metallic behavior at low frequencies.

As was the case for the photoemission measurements, the broadened optical spectra result from the broadened density of states for the amorphous materials. In fact, since $\vec{k}$ is no longer a good quantum number, one can expect that transitions would be allowed between any pair of valence and conduction band states. The absorption coefficient in this picture is proportional to

$$\alpha(\omega) = (\text{const}/\omega) \int dE \, \rho(E)\rho(E + \hbar\omega)| \, M(E) \, |^2 \tag{13.1}$$

where $\rho(E)$ is the density of states, ω is the optical frequency, and $M(E)$ is a generalized momentum matrix element. Mott has argued that $M = 1$ for transitions involving two extended states <u>and</u> for one extended and one localized state, but that M is negligible for transitions involving two localized states (which will have negligible spatial overlap). Thus

173

Figure 13.11: Absorption edge of crystalline c–As$_2$S$_3$ for 2 directions of light polarization relative to the c–axis compared with the absorption edge of amorphous a–As$_2$S$_3$.
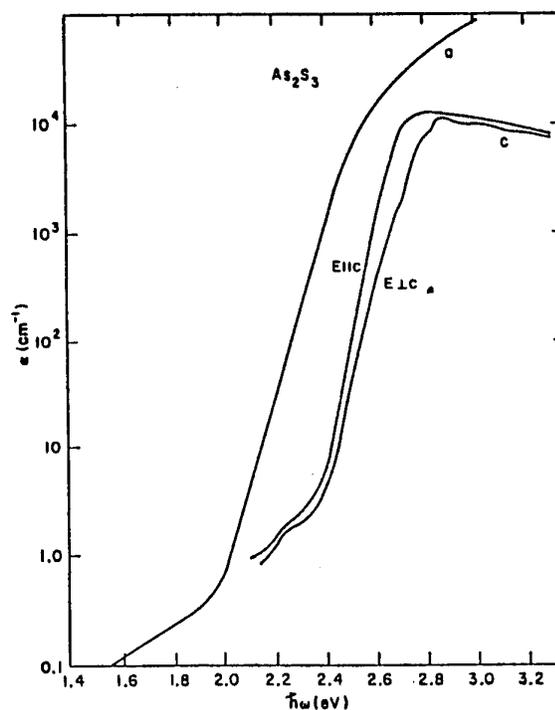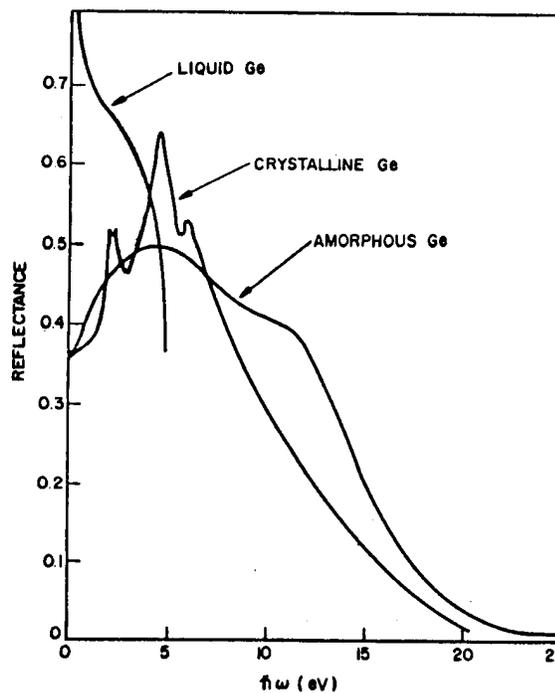

Figure 13.12: Fundamental reflection spectra due to electronic transitions in crystalline, amorphous and liquid Ge. The results are consistent with the metallic transport properties of liquid Ge.

174

the optical gap should be at a different energy $\hbar\omega = E_{opt}$ than the mobility edge. Mott argued that the density of states $\rho(E) \simeq (E - E_0)$ is a linear function of E near the band edges E_0, giving

$$\alpha(\omega) = \text{const} \; \frac{(\hbar\omega - E_{opt})^2}{\omega}. \tag{13.2}$$

Unfortunately this argument is not of general validity, but is nevertheless used to estimate E_{opt} by plotting $(\alpha\hbar\omega)^{\frac{1}{2}}$ vs. ω and extrapolating the straight–line behavior to zero frequency.

The amorphous chalcogenides exhibit remarkable luminescence behavior, with a large shift to lower energy of the luminescence peak relative to E_{opt}, by as much as $1/2$ E_{opt}. This is attributed to a large electron–lattice interaction: the excited state of the optical transition produces an atomic or bond rearrangement sufficient to cause a large shift in the energies of both the excited and ground states.

13.1.4 Transport Properties

Transport measurements on amorphous semiconductors have proved difficult to interpret, partly because of differences in the measured transport results arising from differences in methods of sample preparation. Because of the low mobilities, Hall data have been difficult to obtain, and thermopower data have been difficult to interpret. Attempts to measure mobilities using transition–time methods resulted in the discovery of non–dispersive transport: a pocket of charge injected at one side of the sample does not propagate to the other side with fixed velocity but instead spreads out in time because of the large number of traps which have a large distribution of release times. These measurements have been exploited by Professor Kastner's group at MIT to give new data on the electronic density of states in amorphous As_2Se_3.

13.1.5 Applications of Amorphous Semiconductors

The most successful application of amorphous semiconductors has been the use of amorphous Se films for Xerography. In this process, one surface of the film is charged. When light reflected from the white area of the original page strikes the Se film, electron–hole pairs are formed, which then migrate to the surface and neutralize the charge. These areas on the Se film do not attract the small charged black "toner" particles, resulting in areas which do not print (white on the copy). The unaffected (black) areas retain their charge, do attract toner, and do print on the copy.

Another important application is for solar cells and thin film transistors, where the substantial reduction in cost of producing large areas of amorphous rather than crystalline films has the potential to offset their lower efficiency.

An effect which caused some excitement several years ago was the observation by Ovshinsky and others of reversible switching behavior in chalcogenide semiconductors. This has been shown to be due to an electronic mechanism, the filling of traps with carriers above a threshold current, producing a sharp drop in the resistance. A second type of switching is associated with the formation of small crystalline regions. These effects have been exploited in computer memory devices.

13.2 Amorphous Semiconductor Superlattices.

The extraction of quantitative information from the study of amorphous semiconductor superlattices offers considerable challenge, because the number of variables is large (band offsets, masses, band gaps, mobility edges, chemistry, etc.) On the other hand, superlattices introduce one element of order (z–axis periodicity) in an otherwise disordered system; the superlattice periodicity may perhaps be exploited to learn new physics about this class of materials.

Early achievements in the field of amorphous semiconductor superlattices (B. Abeles and T. Tiedje, *Phys. Rev. Lett.* <u>51</u>, 2003 (1983)) indicated that superlattices can be synthesized with alternate layers of amorphous semiconductors such as a–Si:H, a–Ge:H, a–SiN$_x$:H and a–Si$_{1-x}$C$_x$:H where the <u>a</u> denotes amorphous and the :H denotes the addition of hydrogen to tie up the dangling bonds in the amorphous semiconductor. Unlike the case in crystalline materials, lattice matching is not an issue in the synthesis of amorphous semiconductor superlattices.

The amorphous superlattices are prepared by a plasma–assisted chemical vapor deposition (CVD) method in which the composition of the reactive gases is changed periodically in the reaction chamber. This process has some similarities to the MOCVD technique discussed in connection with heterojunction crystalline semiconductor superlattices. The plasma assisted technique allows deposition to occur at lower substrate temperatures, thereby achieving sharper interfaces. The amorphous films can be deposited on quartz substrates. The residence time of the gases in the reactor (SiH$_4$ for preparing a–Si:H, and SiH$_4$ + NH$_3$ for preparing a–Si$_{1-x}$N$_x$:H) can be as short as 1 sec while the time to grow a monolayer is $\sim$ 3 sec. (see Fig. 13.13). Thus the gases in the reactor can be exchanged rapidly enough to achieve sharp interfaces. The plasma discharge is maintained continuously while the gases are changed.

The superlattice periodicity is monitored during the growth process, and the periodicity is confirmed after the film is deposited by x–ray diffraction, as shown in Fig. 13.13. Because of the random atomic arrangements in the layer planes of the two constituents a–Si:H and a–Si$_{1-x}$N$_x$:H, there is no periodicity within the layers d_1 and d_2, so that the only periodicity found with the x–ray characterization experiment is that due to the periodicity $d = d_1 + d_2$. From the width of the x–ray peaks, the authors deduce an rms fluctuation in the layer thickness of $\Delta d \sim 5$ Å where $d = 41 + 27 = 68$ Å.

Although no direct observation has been made of bound states in the quantum wells of amorphous semiconductors, the optical absorption measurements of Fig. 13.14 show an increase in the optical bandgap E_g as the quantum well width decreases, where the optical gap was determined from the energy dependence of the absorption coefficient using the relation $\alpha \sim (\hbar\omega - E_g)^2/\omega$, which normally is valid for bulk amorphous semiconductors.

The temperature dependence of the photoluminescence of the amorphous semiconductor superlattice is similar to that of the bulk, showing an $\exp(-T/T_0)$ dependence, except that for the superlattice the characteristic temperature, T_0 increases as the width of the quantum well decreases, as shown in Fig. 13.15. Also, the energy width E_0 of the localized state distribution in the Urbach tail for amorphous semiconductors shows a similar increase as the width of the quantum well decreases (see Fig. 13.15), indicating that the distribution of localized states broadens as the layer thickness decreases. It is expected that the superlattices will have a small effect on localized states that are deep in the band tail, but

Figure 13.13: X–ray (1.54 Å) diffraction pattern vs. scattering angle (lower scale) and d spacing (upper scale) of a a–Si:H (41 Å)/a–SiN$_x$:H (27 Å) superlattice with 41 periods on a quartz substrate. The inset shows the energy–band diagram assumed for the superlattice. The conduction band offset $U = 1.05$ eV is indicated on the figure.

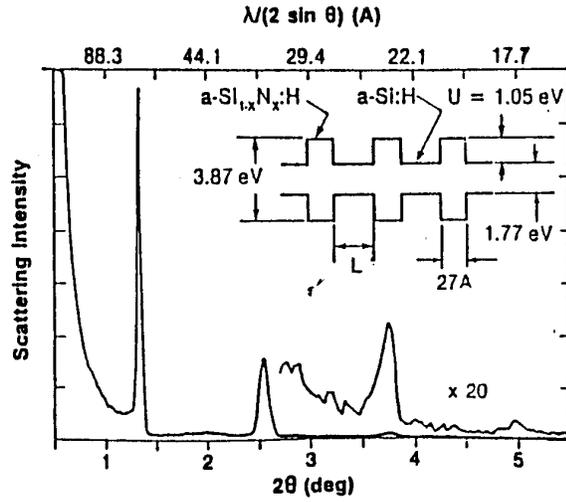Figure 13.14: Optical–absorption coefficient α vs. photon energy E for a–Si:H/a–SiN$_x$:H superlattices with varying a–Si:H layer thickness L and a constant a–SiN$_x$:H layer thickness of 27 Å. Also given in the figure is the absorption coefficient α for a–SiN$_x$:H films prepared under the same conditions.
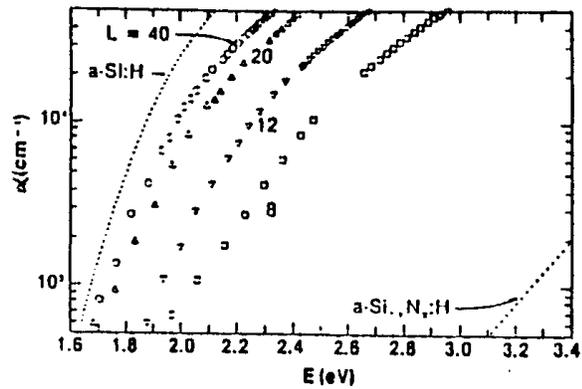
177

Figure 13.15: Dependence of the optical gap on the a–Si:H sublayer thickness L with the a–SiN$_x$:H thickness held fixed at 27Å. The solid line is a calculated curve based on the band diagram in Fig. 13.13, assuming effective masses of unity. Dependence of the Urbach slope parameter E_0 (left scale) and photoluminescence quenching parameter T_0 (right side) on the a–Si:H layer thickness.
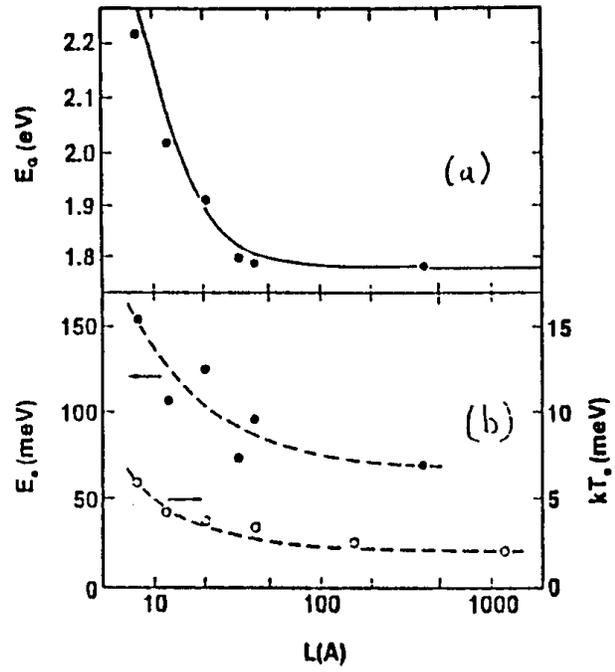
a large effect on the energy of the weakly localized shallow states.

Appendix A

Time Dependent Perturbation Theory

References

- Eisberg, *Fundamentals of Modern Physics*, Ch. 9

- Schiff, *Quantum Mechanics*, Ch. 8

A.1   General Formulation

To proceed further with the formal development of the optical properties of solids, we need to consider how to handle the effect of time-dependent electromagnetic fields quantum mechanically. The most important case of interest is the one where the external field is a sinusoidal function of time. For most practical applications, the external fields are sufficiently weak, so that their effect can be handled within the framework of perturbation theory. If the perturbation has an explicit time dependence, it must be handled by time-dependent perturbation theory. Practical problems which are handled by time-dependent perturbation theory include such subjects as magnetic resonance (nuclear and electronic spin), cyclotron resonance and optical properties of solids. We give here a brief review of the subject.

In doing time-dependent perturbation theory we write the total Hamiltonian $\mathcal{H}$ as:

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}'(t) \tag{A.1}$$

where $\mathcal{H}_0$ is the unperturbed Hamiltonian and $\mathcal{H}'(t)$ is the time-dependent perturbation. We assume here that we know how to solve the unperturbed time independent problem for its eigenvalues E_n and corresponding eigenfunctions u_n.

$$\mathcal{H}_0 u_n = E_n u_n. \tag{A.2}$$

Since $\mathcal{H}'(t)$ has an explicit time dependence, then "energy" is no longer a "constant of the motion". Since we no longer have stationary, time-independent solutions, we must use the time-dependent form of Schrödinger's equation, which is:

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H}\psi = (\mathcal{H}_0 + \mathcal{H}')\psi. \tag{A.3}$$

Now, if we didn't have the perturbation term $\mathcal{H}'(t)$ to contend with, we would set

$$\psi(\vec{r}, t) = u_n(\vec{r})e^{-iE_n t/\hbar} \tag{A.4}$$

where $u_n(\vec{r})$ is independent of time and satisfies Eq. A.2. Thus all the time dependence of $\psi(\vec{r}, t)$ is contained in the phase factor $e^{-iE_n t/\hbar}$. For $\mathcal{H}'(t) = 0$, it immediately follows that

$$i\hbar \frac{\partial \psi}{\partial t} = E_n \psi \tag{A.5}$$

which yields the time-independent Schrödinger equation. With the perturbation present, we expand the time dependent functions $\psi(\vec{r}, t)$ in terms of the complete set $u_n(\vec{r})e^{-iE_n t/\hbar}$

$$\psi(\vec{r}, t) = \sum_n a_n(t)u_n(\vec{r})\, e^{-iE_n t/\hbar} \tag{A.6}$$

where the $a_n(t)$ are the time-dependent expansion coefficients. Substituting Eq. A.6 in the time-dependent Schrödinger equation (Eq. A.3) we obtain:

$$i\hbar \sum_n \dot{a}_n(t)u_n e^{-iE_n t/\hbar} + \sum_n a_n(t)u_n E_n e^{-iE_n t/\hbar} = \sum_n a_n(t)[\mathcal{H}_0 + \mathcal{H}'(t)]u_n e^{-iE_n t/\hbar}$$
$$= \sum_n a_n(t)[E_n + \mathcal{H}'(t)]u_n e^{-iE_n t/\hbar} \tag{A.7}$$

where $\dot{a}_n(t)$ denotes the time derivative $da_n(t)/dt$. We note that because of Eq. A.2 the second term on the left hand side of Eq. A.7 is canceled by the first term on the right hand side.

We now multiply on the left hand side of Eq. A.7 by $u_k^*(\vec{r})$ and integrate over all space. If we make use of the orthogonality of the eigenfunctions

$$\int u_k^*(\vec{r})u_n(\vec{r})d^3r = \delta_{n,k} \tag{A.8}$$

we obtain from Eq. A.7

$$i\hbar \sum_n \dot{a}_n(t)u_n e^{-iE_n t/\hbar} = \sum_n a_n(t)\mathcal{H}'(t)u_n e^{-iE_n t/\hbar} \tag{A.9}$$

the result:

$$i\hbar\, \dot{a}_k e^{-iE_k t/\hbar} = \sum_n a_n\langle k|\mathcal{H}'(t)|n\rangle e^{-iE_n t/\hbar} \tag{A.10}$$

where we have written the matrix element

$$\langle k|\mathcal{H}'(t)|n\rangle = \int u_k^*(\vec{r})\mathcal{H}'(t)u_n(\vec{r})d^3r. \tag{A.11}$$

Since $\mathcal{H}'(t)$ is time-dependent, so is the matrix element time-dependent, even though, the matrix element is taken between stationary states. We thus obtain the result

$$i\hbar\dot{a}_k(t) = \sum_n a_n(t)\langle k|\mathcal{H}'(t)|n\rangle e^{i(E_k - E_n)t/\hbar}. \tag{A.12}$$

If we set

$$\hbar\omega_{kn} = E_k - E_n \tag{A.13}$$

180

where ω_{kn} is the Bohr frequency between states k and n, we have

$$\dot{a}_k(t) = \frac{1}{i\hbar} \sum_n a_n(t) e^{i\omega_{kn}t} \langle k|\mathcal{H}'(t)|n\rangle \qquad (A.14)$$

in which the indicated matrix element is taken between eigenstates of the unperturbed Hamiltonian $\mathcal{H}_0$. So far, no perturbation theory has been used and the result given in Eq. A.14 is exact. We notice that the unperturbed Hamiltonian is completely absent from Eq. A.14. Nevertheless, its energy eigenvalues appear in ω_{kn} and its eigenfunctions in the matrix element $\langle k|\mathcal{H}'(t)|n\rangle$.

In applying perturbation theory, we consider the matrix element $\langle k|\mathcal{H}(t)|n\rangle$ to be small, and we write each time-dependent amplitude as an expansion in perturbation theory

$$a_n = a_n^{(0)} + a_n^{(1)} + a_n^{(2)} + \cdots = \sum_{i=0}^{\infty} a_n^{(i)} \qquad (A.15)$$

where the superscript gives the order of the term. Thus $a_n^{(0)}$ is the $zero^{th}$ order term and $a_n^{(i)}$ is the ith order correction to a_n. From Eq. A.14, we see that $a_k(t)$ changes its value with time only because of the time dependent perturbation. Thus, the unperturbed situation (0th order perturbation theory) must give no time dependence in $zero^{th}$ order

$$\dot{a}_m^{(0)} = 0 \qquad (A.16)$$

and the first order correction yields:

$$\dot{a}_m^{(1)} = 1/i\hbar \sum_n a_n^{(0)} \langle m|\mathcal{H}'(t)|n\rangle e^{i\omega_{mn}t}. \qquad (A.17)$$

In the application of perturbation theory we assume, for example, that if we start in an eigenstate $n = \ell$, only the coefficient $a_\ell^{(0)}$ will be appreciably large. Then all other terms in the sum can be neglected. This gives us in 1st order perturbation theory:

$$\dot{a}_m^{(1)} = \frac{1}{i\hbar} a_\ell^{(0)} \langle m|\mathcal{H}'|\ell\rangle e^{i\omega_{m\ell}t} \qquad (A.18)$$

where $a_\ell^{(0)}$ is approximately unity.

For many cases of interest, this integration over the time variable can be performed and $a_m^{(1)}$ rather than its time derivative is obtained. The two simple cases that can be integrated easily are:

1. The perturbation $\mathcal{H}'$ is constant but is turned on at some time ($t = 0$) and we look at the amplitudes of the wave function in the various states after the perturbation has been acting for some time $t > 0$.

2. The perturbation $\mathcal{H}'$ has a sinusoidal time dependence with frequency ω. This is the situation for all resonant phenomena.

Let us first consider case (1). Then

$$a_m^{(1)}(t) = \frac{1}{i\hbar} \int_0^t \langle m|\mathcal{H}'|\ell\rangle e^{i\omega_{m\ell}t'} dt' = \frac{\langle m|\mathcal{H}'|\ell\rangle}{i\hbar} \frac{[e^{i\omega_{m\ell}t} - 1]}{i\omega_{m\ell}}. \qquad (A.19)$$
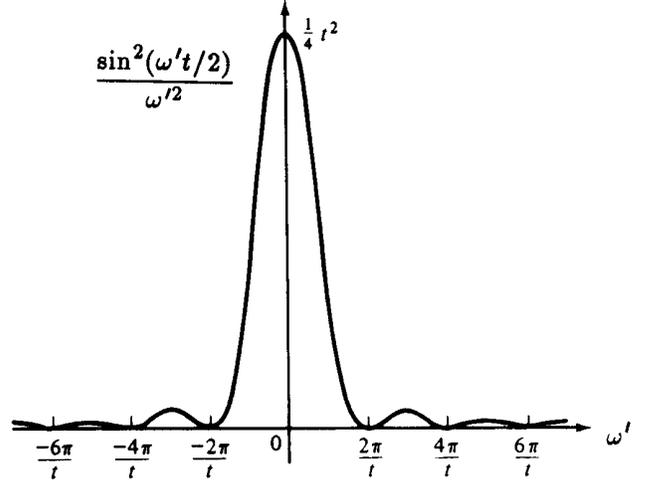
181

Figure A.1: Plot of $\sin^2(\omega' t/2)/\omega'^2$ vs. ω', a function which enters the calculation of time–dependent perturbation problems.

Similarly, for case (2), we can write

$$\mathcal{H}'(t) = \mathcal{H}'(0)e^{\pm i\omega t} \tag{A.20}$$

to show the explicit time dependence, so that upon integration we obtain for the amplitudes $a_m^{(1)}(t)$

$$a_m^{(1)}(t) = \frac{1}{i\hbar}\langle m|\mathcal{H}'(0)|\ell\rangle \int_0^t e^{i(\omega_{m\ell}\pm\omega)t'}dt' = \frac{1}{i\hbar}\,\langle m|\mathcal{H}'(0)|\ell\rangle \frac{e^{i(\omega_{m\ell}\pm\omega)t}-1}{i(\omega_{m\ell}\pm\omega)}. \tag{A.21}$$

We interpret the time dependent amplitudes $|a_m^{(1)}(t)|^2$ as the probability of finding the system in a state m after a time t has elapsed since the perturbation was applied; the system was initially in a state $\ell \neq m$.

We thus obtain for case (1) given by Eq. A.19

$$|a_m^{(1)}(t)|^2 = \left(\frac{|\langle m|\mathcal{H}'|\ell\rangle|^2}{\hbar^2}\right)\left(\frac{|e^{i\omega_{m\ell}t}-1|^2}{\omega_{m\ell}^2}\right) \tag{A.22}$$

$$|a_m^{(1)}(t)|^2 = \left(\frac{|\langle m|\mathcal{H}'|\ell\rangle|^2}{\hbar^2}\right)\left(\frac{4\sin^2(\omega_{m\ell}t/2)}{\omega_{m\ell}^2}\right) \tag{A.23}$$

Clearly for case (2), the same result follows except that $\omega_{m\ell}$ is replaced by $(\omega_{m\ell}\pm\omega)$ where ω is the applied frequency and a resonant denominator results for the transition probability amplitude. It is clear from the above arguments that for both cases (1) and (2), the explicit time dependence is contained in an oscillatory term of the form $[\sin^2(\omega' t/2)/\omega'^2]$ where $\omega' = \omega_{m\ell}$ for the case (1) and $\omega' = \omega_{m\ell} \pm \omega$ for case (2). This function was previously encountered in diffraction theory and looks like that shown in Fig. A.1. Of special interest here is the fact that the main contribution to this function comes for $\omega' \cong 0$, with the height

182

of the main peak proportional to $t^2/4$ and the width proportional to $1/t$. This means that the area under the central peak goes as t. If we think of $|a_m(t)|^2$ as the probability of finding the system in a state m, then for case (2), where we have a perturbation with frequency ω, the system attempts to make a transition from a state ℓ to a state m with a transition probability proportional to the time the perturbation acts. If we then wait long enough, a system in an energy state ℓ will make a transition to a state m, if photons of the resonant frequency $\omega_{\ell m}$ are present.

A.2   Fermi Golden Rule

Since the transition probability is proportional to the time the perturbation acts, it is therefore useful to deal with a quantity called the *transition probability per unit time* and the relation giving this quantity is called the Golden Rule (named by Fermi and often called Fermi's Golden Rule).

In deriving the Golden Rule from Eq. A.19, we must consider the system exposed to the perturbation for a time sufficiently long so that we can make a meaningful measurement within the framework of the Heisenberg uncertainty principle:

$$\Delta E \Delta t \sim h \tag{A.24}$$

so that the uncertainty in energy (or frequency) during the time that the perturbation acts is

$$\Delta E \sim h/t \tag{A.25}$$

or

$$\Delta \omega_{\ell m} \sim 2\pi/t. \tag{A.26}$$

But this is precisely the period of the oscillatory function shown in Fig. A.1. In this context, we must think of the concept of transition probability/unit time as encompassing a range of energies and times consistent with the uncertainty principle. In the case of solids, it is quite natural to do this anyhow, because the wave vector $\vec{k}$ is a quasi-continuous variable. That is, there are a large number of k states which have energies close to a given energy. The quantum states labeled by wave vector $\vec{k}$ are close together in a solid having about 10^{22} atoms/cm^3. Since the photon source itself has a bandwidth, we would automatically want to consider a range of energy differences $\delta \hbar \omega'$. From this point of view, we introduce the transition probability/unit time W_m for making a transition to a state m

$$W_m = \frac{1}{t} \sum_{m' \approx m} |a_{m'}^{(1)}(t)|^2 \tag{A.27}$$

where the summation is carried out over a range of energy states consistent with the uncertainty principle; $\Delta \omega_{mm'} \sim 2\pi/t$.

Substituting for $|a_{m'}^{(1)}(t)|^2$ from Eq. A.23, we have

$$|a_m^{(1)}(t)|^2 = \left(\frac{4|\langle m|\mathcal{H}'|\ell\rangle|^2}{\hbar^2} \right) \left(\frac{\sin^2(\omega' t/2)}{\omega'^2} \right) \tag{A.28}$$

183

and the summation is replaced by an integration over a narrow energy range weighted by the density of states $\rho(E_m)$ which gives the number of states per unit energy range. We thus obtain

$$W_m = \frac{1}{\hbar^2 t} \int |4\mathcal{H}'_{m'\ell}|^2 \left(\frac{\sin^2(\omega_{m'\ell} t/2)}{\omega^2_{m'\ell}} \right) \rho(E_{m'}) \, dE_{m'} \tag{A.29}$$

where we have written $\mathcal{H}'_{m'\ell}$ for the matrix element $\langle m'|\mathcal{H}'|\ell \rangle$. But, by hypothesis, we are only considering energies within a small energy range $E_{m'}$ around E_m and over this range the matrix elements and density of final states will not be varying. However, the function $[\sin^2(\omega' t/2)/\omega'^2]$ will be varying rapidly, as can be seen from Fig. A.1. Therefore, it is adequate to integrate Eq. A.29 only over the rapidly varying function $[\sin^2(\omega t/2)]/\omega^2$. Writing $dE = \hbar d\omega'$, we obtain;

$$W_m \simeq \left(\frac{4|\mathcal{H}'_{m\ell}|^2 \rho(E_m)}{t\hbar^2} \right) \int \left(\frac{\sin^2 \frac{\omega' t}{2}}{\omega'^2} \right) d\omega'. \tag{A.30}$$

The most important contribution to the integral in Eq. A.30 comes from values of ω close to ω'. On the other hand, we know how to do this integral between $-\infty$ and $+\infty$, since

$$\int_{-\infty}^{\infty} (\sin^2 x/x^2) dx = \pi. \tag{A.31}$$

Therefore we can write an approximate relation from Eq. A.30 by setting $x = \omega' t/2$

$$W_m \cong (2\pi/\hbar)|\mathcal{H}'_{m\ell}|^2 \rho(E_m) \tag{A.32}$$

which is often called Fermi's Golden Rule. In the subsequent sections, we will apply the Fermi Golden Rule to calculate the optical properties of solids.

If the initial state is a discrete level (such as donor impurity level) and the final state is a continuum (such as conduction band), then the Fermi Golden Rule (Eq. A.32) as written yields the transition probability per unit time and $\rho(E_m)$ is interpreted as the density of *final* states. Likewise if the final state is discrete and the initial state is a continuum, W_m also gives the transition probability per unit time, only in this case $\rho(E_m)$ is interpreted as the density of initial states.

For many important applications in solid state physics, the transitions of interest are between a continuum of initial states and a continuum of final states. In this case the Fermi Golden Rule must be interpreted in terms of a joint density of states, whereby the initial and final states are separated by the photon energy $\hbar\omega$ inducing the transition. These issues are discussed in Chapter 4.

A.3 Time Dependent 2nd Order Perturbation Theory

This second order treatment is needed for indirect optical transitions, where

$$\mathcal{H} = \mathcal{H}_0 + \lambda \mathcal{H}' \tag{A.33}$$

and $\lambda \ll 1$. Here $\mathcal{H}_0 \psi_0 = i\hbar \partial \psi_0 / \partial t$. Expand ψ, the solution to Eq. A.33 in terms of the complete set of functions denoted by $\psi_0 \equiv |n, \vec{k}\rangle$

$$\psi = \sum_{n,\vec{k}} a_n(\vec{k}, t) e^{-\frac{i}{\hbar} E_n(\vec{k}) t} |n, \vec{k}\rangle \tag{A.34}$$

184

where $|n,\vec{k}\rangle$ is a Bloch function describing the eigenstates of the unperturbed problem

$$\mathcal{H}\psi = i\hbar\dot{\psi} \tag{A.35}$$

$$\sum_{n\vec{k}} a_n(\vec{k},t)E_n(\vec{k})e^{-\frac{i}{\hbar}E_n(\vec{k})t}|n,\vec{k}\rangle + \sum_{n,\vec{k}} a_n(\vec{k},t)e^{-\frac{i}{\hbar}E_n(\vec{k})t}\lambda\mathcal{H}'|n,\vec{k}\rangle$$

$$= i\hbar\sum_{n,\vec{k}}\dot{a}_n(\vec{k},t)e^{-\frac{i}{\hbar}E_n(\vec{k})t}|n,\vec{k}\rangle + \sum_{n,\vec{k}} a_n(\vec{k},t)e^{-\frac{i}{\hbar}E_n(\vec{k})t}|n,\vec{k}\rangle \tag{A.36}$$

which gives

$$\dot{a}_m(\vec{k}',t) = \frac{1}{i\hbar}\sum_{n,\vec{k}} a_n(\vec{k},t)e^{\frac{i}{\hbar}(E_m(\vec{k}')t - E_n(\vec{k})t)}\langle m,\vec{k}'|\lambda\mathcal{H}'|n,\vec{k}\rangle \tag{A.37}$$

We expand

$$a_m(\vec{k}',t) = a_m^{(0)} + \lambda a_m^{(1)} + \lambda^2 a_m^{(2)} + \dots \tag{A.38}$$

and let $a_j(\vec{k},0) = 1$, and all others $a_n(\vec{k},0) = 0$ where $n \neq j$.

To first order, as before,

$$\lambda^2\dot{a}_m^{(1)}(\vec{k}',t) = \frac{1}{i\hbar}\lambda a_n^{(0)}(\vec{k},t)\exp\left[\frac{i}{\hbar}[E_m(\vec{k}') - E_n(\vec{k})]t\right]\langle m,\vec{k}'|\lambda\mathcal{H}'|n,\vec{k}\rangle \tag{A.39}$$

or

$$a_m^{(1)}(\vec{k}',t) = \frac{1}{i\hbar}\int_0^t dt'\ \exp\left[\frac{i}{\hbar}[E_m(\vec{k}') - E_n(\vec{k})]t\right]\langle m,\vec{k}'|\lambda\mathcal{H}'|n,\vec{k}\rangle \tag{A.40}$$

To second order

$$\lambda^2\dot{a}_m^{(2)}(\vec{k}',t) = \frac{1}{i\hbar}\sum_{n,\vec{k}}\lambda a_n^{(1)}(\vec{k},t)\exp\left[\frac{i}{\hbar}[E_m(\vec{k}') - E_n(\vec{k})]t\right]\langle m,\vec{k}'|\lambda\mathcal{H}'|n,\vec{k}\rangle \tag{A.41}$$

or

$$\begin{aligned}
\dot{a}_m^{(2)}(\vec{k}',t) &= -\frac{1}{\hbar^2}\sum_{n,\vec{k}} a_n^{(1)}(\vec{k},t)\exp\left\{\frac{i}{\hbar}[E_m(\vec{k}') - E_n(\vec{k})]t\right\}\langle m,\vec{k}'|\lambda\mathcal{H}'|n,\vec{k}\rangle \\
&\times \int_0^t dt'\ \exp\left\{\frac{i}{\hbar}[E_n(\vec{k}') - E_i(\vec{k})]t'\right\}\langle n,\vec{k}'|\lambda\mathcal{H}'|i,\vec{k}\rangle
\end{aligned} \tag{A.42}$$

We write the time dependence of the perturbation Hamiltonian explicitly as

$$\mathcal{H}' = \sum_\alpha \mathcal{H}e^{-i\omega_\alpha t} \tag{A.43}$$

and then Eq. A.42 can be written, after integrating twice

$$|a_f^{(2)}(\vec{k}_f,t)|^2 = 2\pi\hbar t\sum_{m,\vec{k},\alpha,\alpha'}\frac{|\langle f|\mathcal{H}'_{\alpha'}|m,\vec{k}\rangle|^2|\langle m,\vec{k}|\mathcal{H}'_\alpha|i\rangle|^2}{(E_m(\vec{k}) - E_i - \hbar\omega_\alpha)^2}\delta(E_f - E_i - \hbar\omega_\alpha - \hbar\omega_{\alpha'}) \tag{A.44}$$

This second-order time-dependent perturbation theory expression is used to derive the probability of an indirect interband transition.

Appendix B

Harmonic Oscillators, Phonons, and the Electron-Phonon Interaction

B.1 Harmonic Oscillators

In this section we review the solution of the harmonic oscillator problem in quantum mechanics using raising and lowering operators. The Hamiltonian for this problem is written as:

$$\mathcal{H} = \frac{p^2}{2m} + \frac{1}{2}\kappa x^2. \tag{B.1}$$

Classically, we know that the frequency of oscillation is given by $\omega = \sqrt{\kappa/m}$ so that

$$\mathcal{H} = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2. \tag{B.2}$$

We define the lowering and raising operators a and $a^\dagger$, respectively, by

$$a = \frac{p - i\omega m x}{\sqrt{2\hbar\omega m}} \tag{B.3}$$

and

$$a^\dagger = \frac{p + i\omega m x}{\sqrt{2\hbar\omega m}}. \tag{B.4}$$

Since $[p, x] = \hbar/i$, then it follows that

$$[a, a^\dagger] = 1 \tag{B.5}$$

so that

$$\mathcal{H} = \frac{1}{2m}\left[(p + i\omega m x)(p - i\omega m x) + m\hbar\omega\right] \tag{B.6}$$

$$= \hbar\omega[a^\dagger a + 1/2]. \tag{B.7}$$

Let

$$N = a^\dagger a \tag{B.8}$$

denote the number operator and we denote its eigenstates by $|n\rangle$, so that

$$N|n\rangle = n|n\rangle \tag{B.9}$$

where n is any real integer. However

$$\langle n|N|n\rangle = \langle n|a^\dagger a|n\rangle = \langle y|y\rangle = n \geq 0 \tag{B.10}$$

where $|y\rangle = a|n\rangle$ implies that n is a non–negative integer. We note with regard to Eq. B.10 that the absolute value square of any wavefunction cannot be negative, because quantum mechanically, this quantity signifies a probability. Hence n is positive number or zero.

The action of the lowering operator is found from consideration of

$$Na|n\rangle = a^\dagger aa|n\rangle = (aa^\dagger - 1)a|n\rangle = (n-1)a|n\rangle. \tag{B.11}$$

Hence we find that

$$a|n\rangle = c|n-1\rangle. \tag{B.12}$$

However from Eq. B.10, we have

$$\langle n|a^\dagger a|n\rangle = |c|^2, \tag{B.13}$$

and also from Eq. B.10 we have

$$\langle n|a^\dagger a|n\rangle = n, \tag{B.14}$$

so that

$$c = \sqrt{n} \tag{B.15}$$

and

$$a|n\rangle = \sqrt{n}|n-1\rangle. \tag{B.16}$$

Since the operator a lowers the quantum number of the state, a is called the annihilation or lowering operator. From this argument you can also see that n has to be an integer. The null state is obtained for $n = 0$.

To obtain the raising operator consider,

$$Na^\dagger|n\rangle = a^\dagger aa^\dagger|n\rangle = a^\dagger(1 + a^\dagger a)|n\rangle = (n+1)a^\dagger|n\rangle. \tag{B.17}$$

Hence we obtain

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \tag{B.18}$$

so that $a^\dagger$ is called a creation operator or a raising operator. Finally, for the Hamiltonian in Eq. B.7 we write

$$\mathcal{H}|n\rangle = \hbar\omega[N + 1/2]|n\rangle = \hbar\omega(n + 1/2)|n\rangle \tag{B.19}$$

so that the eigenvalues for the harmonic oscillator are written as:

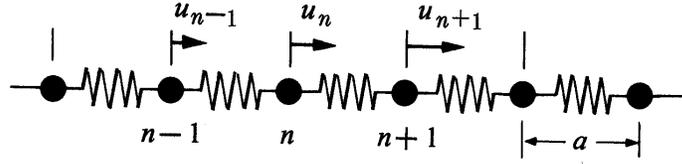$$E = \hbar\omega(n + 1/2) \qquad n = 0, 1, 2, \ldots. \tag{B.20}$$

187

Figure B.1: 1D spring model

B.2 Phonons

In this section we relate the lattice vibrations to harmonic oscillators and identify the quanta of the lattice vibrations with phonons. Consider the 1-D model with springs shown in Fig. B.1 The Hamiltonian for this case is written as

$$\mathcal{H} = \sum_{s=1}^{N} \left(\frac{p_s^2}{2m_s} + \frac{1}{2}\kappa(q_{s+1} - q_s)^2 \right) \tag{B.21}$$

This equation doesn't look like a set of independent harmonic oscillators since q_s and q_{s+1} are coupled. To obtain normal mode solutions we write

$$q_s = \left(1/\sqrt{N}\right) \sum_k Q_k e^{iksa}$$

$$p_s = \left(1/\sqrt{N}\right) \sum_k P_k e^{iksa}. \tag{B.22}$$

These Q_k's and P_k's are called phonon coordinates. It can be verified that

$$[p_s, q_{s'}] = (\hbar/i)\delta_{ss'} \tag{B.23}$$

implies that

$$[P_k, Q_{k'}] = (\hbar/i)\delta_{kk'}. \tag{B.24}$$

The Hamiltonian for 1D lattice vibrations in phonon coordinates is

$$\mathcal{H} = \sum_k \left(\frac{1}{2}P_k^\dagger P_k + \frac{1}{2}\omega_k^2 Q_k^\dagger Q_k \right) \tag{B.25}$$

and gives rise to the 1–D phonon dispersion relation (see Fig. B.2)

$$\omega_k \equiv \sqrt{2\kappa(1 - \cos ka)} = \left(4\kappa/m\right)^{1/2} |\sin(ka/2)|. \tag{B.26}$$

This is all in Kittel ISSP, see pp 611-615 (Sixth edition). Again let

$$a_k = \frac{iP_k^\dagger + \omega_k Q_k}{\sqrt{2\hbar\omega_k}}, \tag{B.27}$$

$$a_k^\dagger = \frac{-iP_k + \omega_k Q_k^\dagger}{\sqrt{2\hbar\omega_k}} \tag{B.28}$$

188

$$\frac{\omega}{(4C/M)^{1/2}} = \left| \sin \frac{1}{2} Ka \right|$$
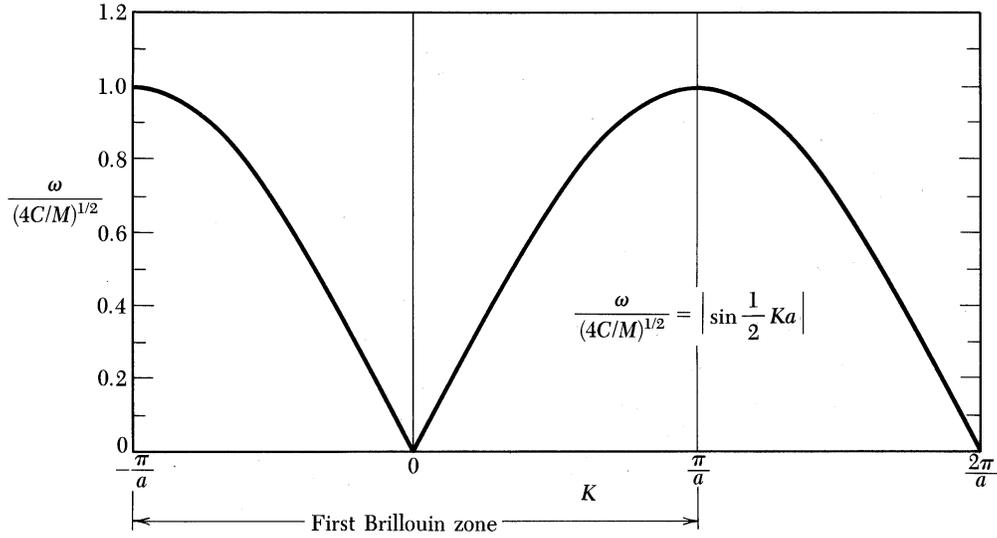
Figure B.2: Phonon dispersion relation.

represent the annihilation and creation operators. The Hamiltonian written in terms of the creation and annihilation operators becomes

$$\mathcal{H} = \sum_k \hbar \omega_k (a_k^\dagger a_k + 1/2) \tag{B.29}$$

yielding energy eigenvalues

$$E = \sum_k (n_k + 1/2) \hbar \omega_k \tag{B.30}$$

The quantum excitation in this case is called a phonon, and the state vector of a system of phonons is written as $|n_1, n_2, \ldots, n_k, \ldots\rangle$. To annihilate or create a phonon in mode k we then write

$$a_k |n_1, n_2, \ldots, n_k, \ldots\rangle = \sqrt{n_k} \, |n_1, n_2, \ldots, n_k - 1, \ldots\rangle \tag{B.31}$$

$$a_k^\dagger |n_1, n_2, \ldots, n_k, \ldots\rangle = \sqrt{n_k + 1} \, |n_1, n_2, \ldots, n_k + 1, \ldots\rangle \tag{B.32}$$

from which the probabilities n_k and $(n_k + 1)$ are obtained for the annihilation and creation processes.

B.3 Phonons in 3D Crystals

We give some examples of the phonon is 3D crystals. The first example is the zone center atomic displacements in graphite shown in Fig. B.3. Graphite has 4 carbon atoms per unit cell, thus 12 zone center modes. There are 3 acoustic modes and 9 optic modes.

The next example is the phonon dispersion curves for diamond shown in Fig. B.4. Diamond has 2 carbon atoms per fcc unit cell, thus 6 branches. The zone center optic modes are Raman active. There are 3 acoustic branches and 3 are optic modes.

The next example is the phonon dispersion curves for silicon shown in Fig. B.5. Silicon
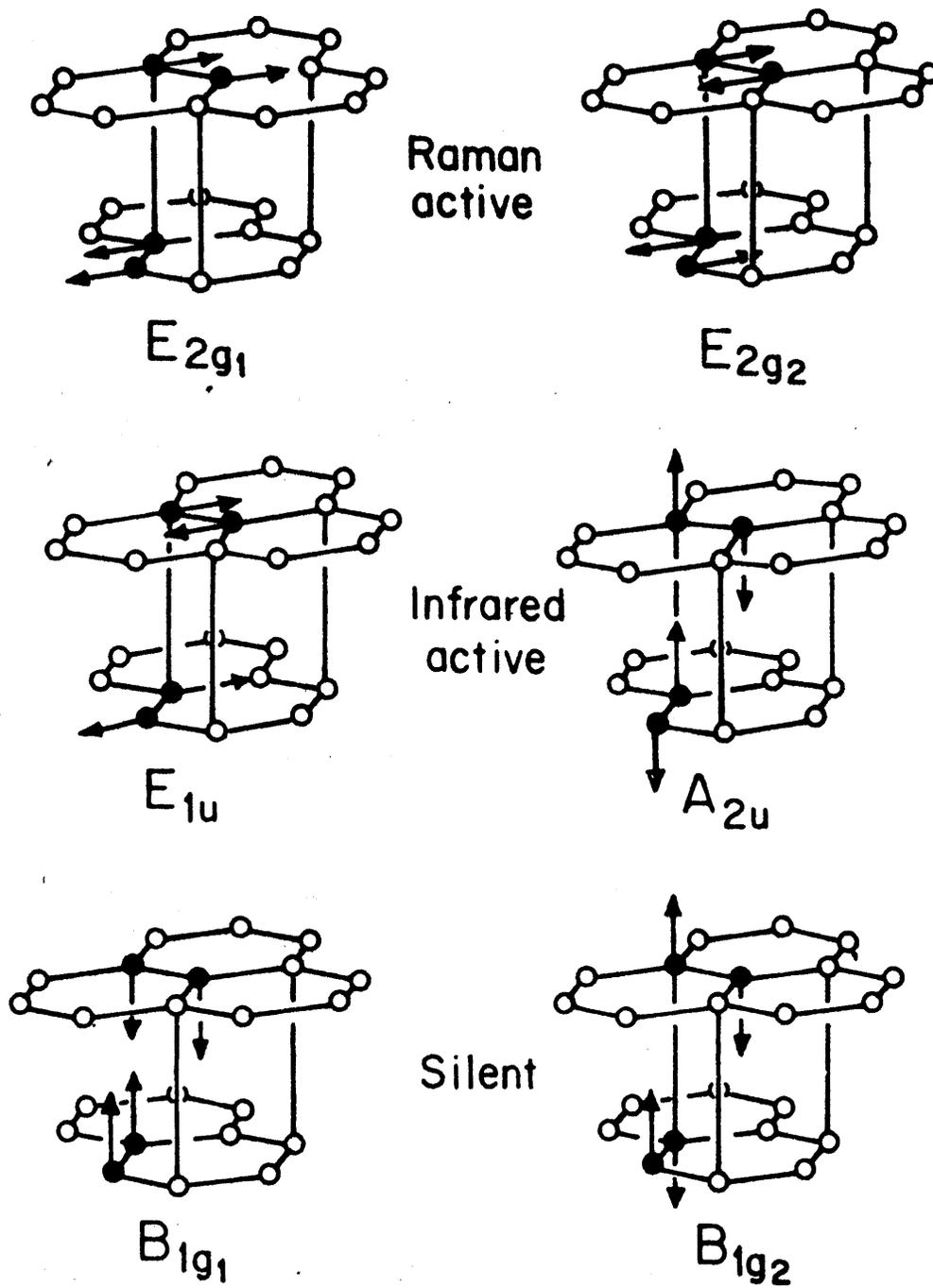
189

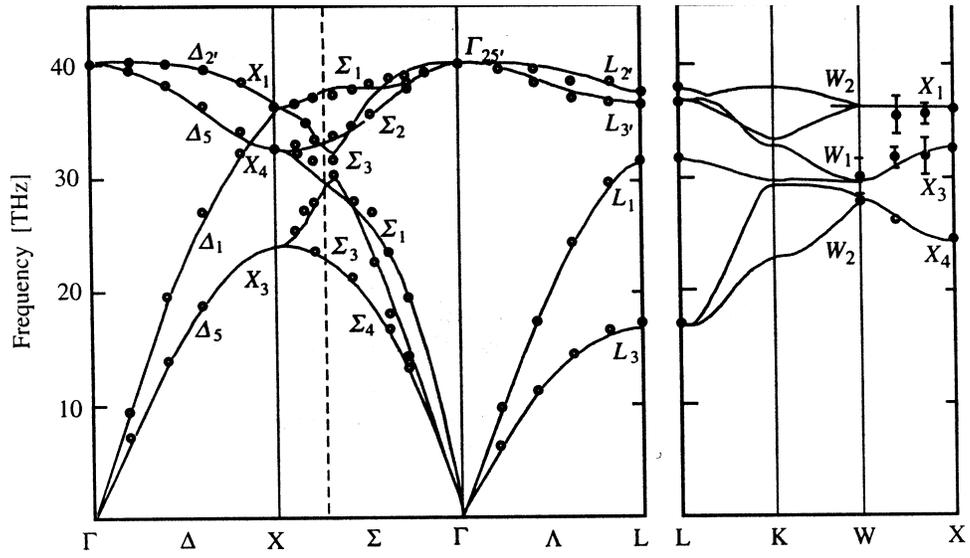Figure B.3: Zone center optical phonon modes in graphite.

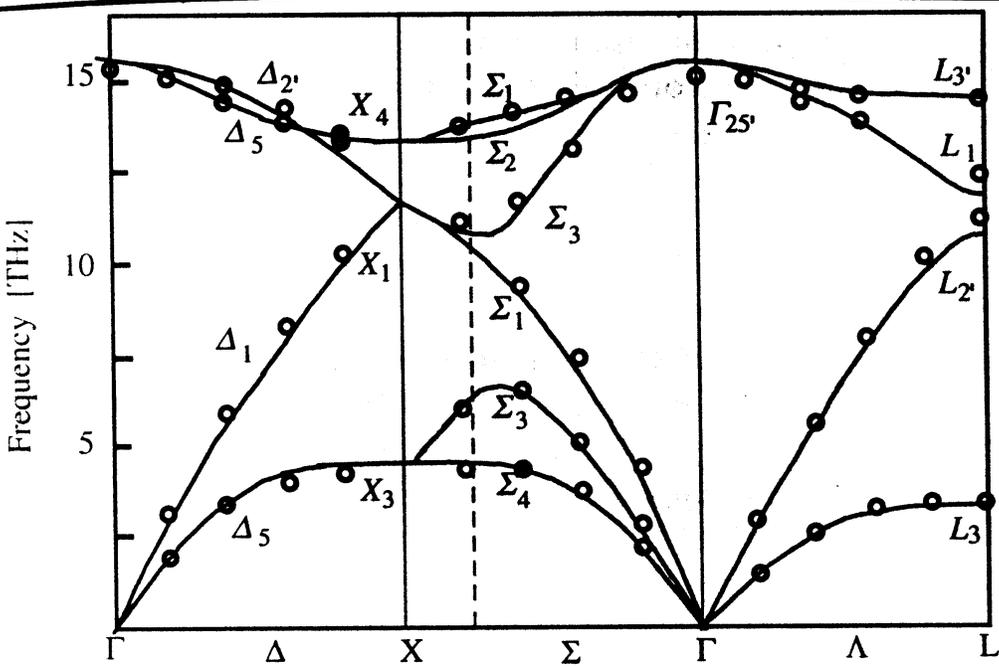Figure B.4: Phonon dispersion curves in diamond.

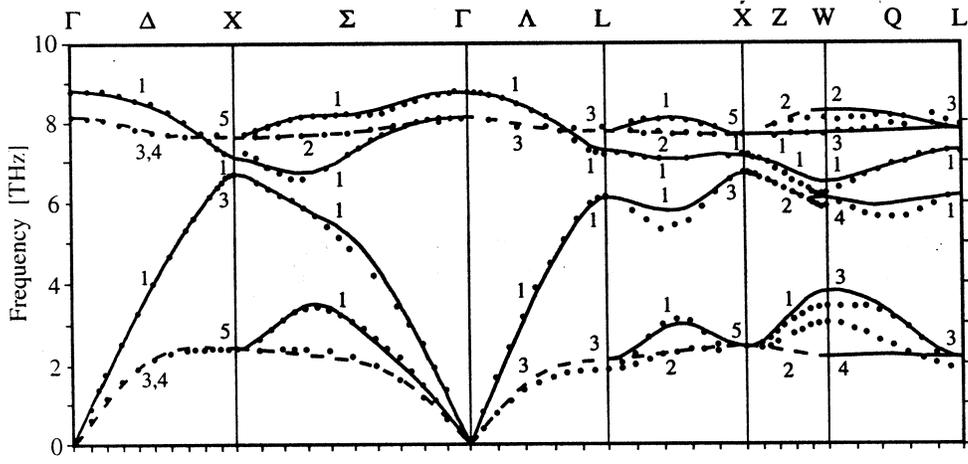Figure B.5: Phonon dispersion curves in silicon.

Figure B.6: Phonon dispersion curves in silicon.

like diamond has 2 atoms per fcc unit cell, thus 6 branches. The zone center optic modes are Raman active. There are 3 acoustic branches and 3 are optic modes.

The next example is the phonon dispersion curves for GaAs shown in Fig. B.6. GaAs like diamond has 2 atoms per fcc unit cell, thus 6 branches. However the two atoms are different and GaAs lacks inversion symmetry. The zone center optic modes are both infrared and Raman active. There are 3 acoustic branches and 3 are optic modes.

B.4 Electron-Phonon Interaction

The basic Hamiltonian for the electron-lattice system is

$$\mathcal{H} = \sum_k \frac{p_k^2}{2m} + \frac{1}{2}\sum_{kk'}{}' \frac{e^2}{|\vec{r}_k - \vec{r}_{k'}|} + \sum_i \frac{P_i^2}{2M} + \frac{1}{2}\sum_{ii'}{}' V_{\text{ion}}(\vec{R}_i - \vec{R}_{i'}) + \sum_{k,i} V_{\text{el-ion}}(\vec{r}_k - \vec{R}_i) \quad \text{(B.33)}$$

where

$$\mathcal{H} = \mathcal{H}_{\text{electron}} + \mathcal{H}_{\text{ion}} + \mathcal{H}_{\text{electron-ion}}. \quad \text{(B.34)}$$

The electron-ion interaction term can be separated into two parts: the interaction of electrons with ions in their equilibrium positions, and an additional term due to lattice vibrations:

$$\mathcal{H}_{\text{el-ion}} = \mathcal{H}^0_{\text{el-ion}} + \mathcal{H}_{\text{el-phonon}} \quad \text{(B.35)}$$

$$\sum_{k,i} V_{\text{el-ion}}(\vec{r}_k - \vec{R}_i) = \sum_{k,i} V_{\text{el-ion}}(\vec{r}_k - (\vec{R}_i^0 + \vec{s}_i)) \quad \text{(B.36)}$$

$$= \sum_{k,i} V_{\text{el-ion}}(\vec{r}_k - \vec{R}_i^0)$$

$$- \sum_{k,i} \vec{s}_i \cdot \nabla V_{\text{el-ion}}(\vec{r}_k - \vec{R}_i^0) \quad \text{(B.37)}$$

$$= \mathcal{H}^0_{\text{el-ion}} + \mathcal{H}_{\text{el-phonon}}. \quad \text{(B.38)}$$

192

In solving the Hamiltonian $\mathcal{H}$ of Eq. B.33 we seek a solution of the total problem in the form

$$\Psi = \psi(\vec{r}_1, \vec{r}_2, \cdots \vec{R}_1, \vec{R}_2, \cdots)\varphi(\vec{R}_1, \vec{R}_2, \cdots) \tag{B.39}$$

such that

$$\mathcal{H}\Psi = E\Psi. \tag{B.40}$$

We then use an adiabatic approximation, which solves the electron part of the Hamiltonian by

$$(\mathcal{H}_{\text{electron}} + \mathcal{H}_{\text{el}-\text{ion}}^0)\psi = E_{\text{el}}\psi. \tag{B.41}$$

Neglecting the $\mathcal{H}_{\text{el}-\text{phonon}}$ term, which we consider as a perturbation, we write:

$$\mathcal{H}_{\text{ion}}\varphi = (E - E_{\text{el}})\varphi = E_{\text{ion}}\varphi \tag{B.42}$$

and we have thus decoupled the electron-lattice system.

Equation B.42 gives us the phonon spectra and harmonic oscillator like wave functions, as discussed in the previous section (§B.2). The term that was left out in the above discussion is the electron–phonon interaction

$$\mathcal{H}_{\text{el}-\text{phonon}} = -\sum_{k,i} \vec{s}_i \cdot \nabla V_{\text{el}-\text{ion}}(\vec{r}_k - \vec{R}_i^0) \tag{B.43}$$

which we now treat as a perturbation. We rewrite Eq. B.42 by introducing the normal coordinates

$$\vec{s}_i = \frac{1}{\sqrt{NM}} \sum_{\vec{q},j} Q_{\vec{q},j} e^{i\vec{q}\cdot\vec{R}_i^0} \hat{e}_j \tag{B.44}$$

where j is polarization index and $\hat{e}_j$ is a unit displacement vector for mode j. Hence we obtain

$$\mathcal{H}_{\text{el}-\text{phonon}} = -\sum_{k,i} \frac{1}{\sqrt{NM}} \sum_{\vec{q},j} Q_{\vec{q},j} e^{i\vec{q}\cdot\vec{R}_i^0} \hat{e}_j \cdot \nabla V_{\text{el}-\text{ion}}(\vec{r}_k - \vec{R}_i^0) \tag{B.45}$$

where

$$Q_{\vec{q},j} = \left(\frac{\hbar}{2\omega_{\vec{q},j}}\right)^{\frac{1}{2}} (a_{\vec{q},j} + a_{-\vec{q},j}^\dagger). \tag{B.46}$$

Writing the time dependence explicitly for the raising and lowering operators

$$a_{\vec{q},j}(t) = a_{\vec{q},j} e^{-i\omega_{\vec{q},j}t} \tag{B.47}$$

$$a_{\vec{q},j}^\dagger(t) = a_{\vec{q},j}^\dagger e^{i\omega_{\vec{q},j}t} \tag{B.48}$$

we obtain

$$\begin{aligned}
\mathcal{H}_{\text{el}-\text{phonon}} &= -\sum_{\vec{q},j} \left(\frac{\hbar}{2MN\omega_{\vec{q},j}}\right)^{\frac{1}{2}} (a_{\vec{q},j} e^{-i\omega_{\vec{q},j}t} + a_{\vec{q},j}^\dagger e^{i\omega_{\vec{q},j}t}) \\
&\quad \times \sum_{k,i} (e^{i\vec{q}\vec{R}_i^0} + e^{i\vec{q}\vec{R}_i^0})\hat{e}_j \cdot \nabla V_{\text{el}-\text{ion}}(\vec{r}_k - \vec{R}_i^0)
\end{aligned} \tag{B.49}$$

which can be written as

$$\mathcal{H}_{\text{el}-\text{phonon}} = -\sum_{\vec{q},j} \left(\frac{\hbar}{2NM\omega_{\vec{q},j}}\right)^{\frac{1}{2}} \left(a_{\vec{q},j} \sum_{k,i} e^{i(\vec{q}\vec{R}_i^0 - \omega_{\vec{q},j}t)}\hat{e}_j \cdot \nabla(\vec{r}_k - \vec{R}_i^0) + c.c.\right) \tag{B.50}$$

If we are only interested in the interaction of one electron and a phonon on a particular branch, say the longitudinal acoustic branch, then we drop the summation over j and k and write

$$\mathcal{H}_{\text{el}-\text{phonon}} = -\sum_{\vec{q}} \left(\frac{\hbar}{2NM\omega_{\vec{q}}}\right)^{\frac{1}{2}} \left(a_{\vec{q}} \sum_i e^{i(\vec{q}\cdot\vec{R}_i^0 - \omega_{\vec{q}}t)}\hat{e} \cdot \vec{\nabla} V_{\text{el}-\text{ion}}(\vec{r} - \vec{R}_i^0) + c.c.\right) \quad \text{(B.51)}$$

where the 1st term in the bracket corresponds to phonon absorption and the c.c. term corresponds to phonon emission.

With $\mathcal{H}_{\text{el}-\text{phonon}}$ in hand, we can solve transport problems (e.g., τ due to phonon scattering) and optical problems (e.g., indirect transitions) directly, since all these problems involve matrix elements $\langle f|\mathcal{H}_{\text{el}-\text{phonon}}|i\rangle$ coupling initial and final states i and f, respectively.