OPTICAL PROPERTIES OF SOLIDS

OPTICAL PROPERTIES OF SOLIDS

Frederick Wooten
Department of Applied Science
University of California
Davis, California

1972

ACADEMIC PRESS New York and London

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 72-187257

PRINTED IN THE UNITED STATES OF AMERICA

To My Father

CONTENTS

PREFACE

The present book attempts to fill a need for a fundamental textbook which explains the optical properties of solids. It is based on two short courses I gave in the Department of Applied Science and a series of fifteen lectures at Chalmers Tekniska Högskola, Göteborg, Sweden presented at the invitation of Professors Stig Hagström, Gösta Brogren, and H. P. Myers.

This book is meant to explain a number of important concepts rather than present a complete survey of experimental data. Its emphasis is almost entirely on intrinsic optical properties and photoelectric emission. Little is said concerning imperfections, color centers, etc. However, the principles are general, so the book serves as a stepping stone to the more advanced review articles and papers on a wide variety of topics.

The book assumes a background in quantum mechanics, solid state physics, and electromagnetic theory at about the level of a senior-year undergraduate or first-year graduate student in physics. Problems and exercises have been included to elaborate more fully on some aspects of the physics, to gain familiarity with typical characteristics of optical properties, and to develop some skills in mathematical techniques.

The central theme of the book is the dielectric function (a macroscopic quantity) and its relationship to the fundamental microscopic electronic properties of solids. The emphasis is on basic principles, often illustrated by simple models. The necessary mathematics needed to understand the models is generally carried through to completion, with no steps missing, and no "it can be shown" statements. Thus, the text is intended to be suitable for self-study, as well as for use in a one-semester first-year graduate course.

ACKNOWLEDGMENTS

There are many people who have contributed in some way to the making of this book. I thank especially those who have permitted the use of their figures from published work. In a more direct personal way, Dr. Louis F. Wouters first aroused in me a latent interest in light and the photoelectric effect and provided the opportunity for my initial research on photoelectric emission. I am grateful to Dr. James E. Carothers and Dr. Jack N. Shearer for their encouragement and support over the years. Dr. Tony Huen, my friend and colleague, aided substantially with many helpful discussions in our day-to-day collaboration. Others whom I thank for helpful discussions are Professor William E. Spicer, Dr. Erkki Pajanne, Dr. Birger Bergersen, Professor Lars Hedin, Professor Stig Lundquist, Dr. Per-Olov Nilsson, Major L. P. Mosteller, Dr. Geoffrey B. Irani, Captain Harry V. Winsor, Lt. George Fuller, Professor Ching-Yao Fong, and Dr. David Brust. Thanks also to Mrs. Peggy Riley, Mrs. Kathryn Smith, and Mrs. Donna Marshall who typed parts of the manuscript.

Chapter 1

INTRODUCTION

This book presents an introduction to the fundamental optical spectra of solids. The aim is to develop an understanding of the relationship of measurable optical properties to the dielectric function and the microscopic electronic structure of solids.

The usual way to determine the optical properties of a solid is to shine monochromatic light onto an appropriate sample and then to measure the reflectance or transmittance of the sample as a function of photon energy. Other methods, such as ellipsometry, are sometimes used. However, these methods are of no concern here. The choice of experimental technique is largely one of convenience, not of the basic information obtained. We shall concentrate on reflectivity. Details of experimental technique and methods of data analysis are left to other monographs and papers, some of which are included in the references throughout the book. One exception is the inclusion of a discussion of the analysis of normal incidence reflectance data with the use of the Kramers–Kronig equations; but, the importance here lies in the physics and great generality contained in the Kramers–Kronig equations, not in the experimental techniques for measuring the reflectance at normal incidence.

In recent years, photoelectric emission and characteristic energy loss experiments have proven useful as methods of studying electronic band structure. These experiments are closely related to optical experiments in terms of the kind of information they provide. They are discussed at various points throughout the book.

There are experimental techniques other than the optical types which

provide information on band structure. These include cyclotron resonance, de Haas–van Alphen effect, galvanomagnetic effects, and magnetoacoustic resonance. However, even though often of high accuracy, these experiments yield information pertaining to energy levels only within a few kT of the Fermi surface. Ion neutralization spectroscopy and soft x-ray emission provide information over a wide energy range, but they have not proven as useful as optical methods.

This chapter discusses briefly the kinds of experiments that are most typical and indicates the kind of information that can be obtained. To provide a framework for discussions of optical measurements, photoelectric emission, and characteristic energy loss spectra, we begin in Section 1.1 with a reminder of some of the ideas of band theory, but no more than that. The reader is assumed to have an adequate understanding of the basic ideas of band theory. Next is a brief introduction to optical reflectivity. This is followed by a discussion of photoelectric emission. Since even an elementary discussion of the physics of photoelectric emission is not included in most textbooks on solid-state physics, it is included here. The final section consists of a brief introduction to characteristic energy loss spectra.

For those unfamiliar with the optical spectrum, it is suggested that Table 1.1 be memorized. Optical data are often presented in terms of frequency or wavelength, but band structure is discussed in terms of energy (eV). It is useful to know how to convert units easily.

TABLE 1.1 *Relationship of Wavelength to Energy, Frequency, and Color*

12,400 Å $\leftrightarrow$ 1 eV $\leftrightarrow$ $\omega = 1.5 \times 10^{15}$ sec^{-1}
6,200 Å $\leftrightarrow$ 2 eV $\leftrightarrow$ red
5,800 Å $\longleftrightarrow$ yellow
5,200 Å $\longleftrightarrow$ green
4,700 Å $\longleftrightarrow$ blue

1.1 Band Theory of Solids

The band theory of solids is based on a one-electron approximation. That is, an electron is assumed to be acted on by the field of the fixed atomic cores plus an average field arising from the charge distribution of all the other outer-shell electrons. The atomic cores consist of the nuclei and all inner-shell electrons not appreciably perturbed by neighboring atoms. If the solid is also a perfect crystal, the total crystal potential energy $V(\mathbf{r})$ must have the periodicity of the crystal lattice. On the basis of this model, the solutions of the Schrödinger equation

$$(\hbar^2/2m)\nabla^2\psi + [\mathscr{E} - V(\mathbf{r})]\psi = 0 \tag{1.1}$$

are Bloch functions

$$\psi(\mathbf{k}, \mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp i\mathbf{k} \cdot \mathbf{r} \tag{1.2}$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a function having the periodicity of the lattice.

The simplest solution to Eq. (1.1) is for the case in which $V(\mathbf{r})$ is constant and can be taken as zero. It leads to free electrons and plane waves for wave functions. The energy of an electron is then given by

$$\mathscr{E} = \hbar^2 k^2 / 2m \tag{1.3}$$

If we include the periodicity of the lattice, but say that the perturbing potential is arbitrarily weak, the energy of an electron can be expressed as

$$\mathscr{E} = (\hbar^2/2m) \left| \mathbf{k} + \mathbf{G} \right|^2 \tag{1.4}$$

where $\mathbf{G}$ is a reciprocal lattice vector. The energy bands are then best represented in the reduced zone scheme. Figure 1.1 shows the free-electron

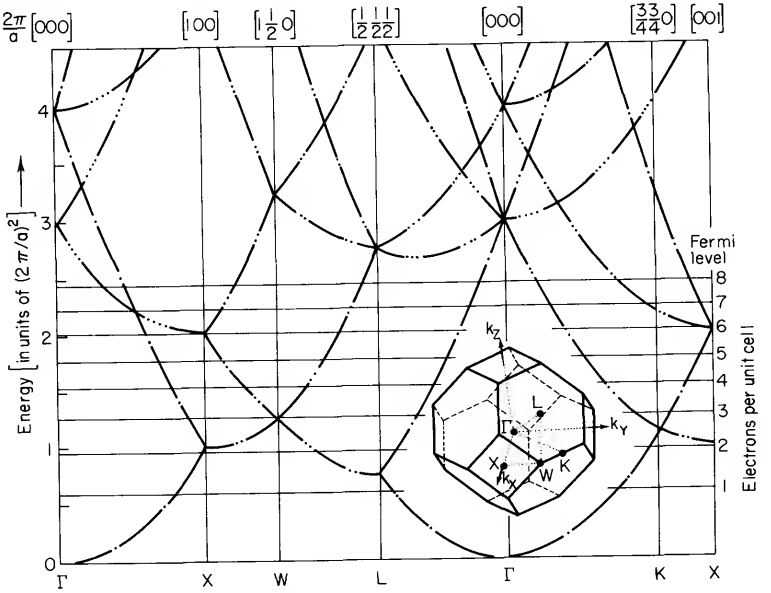

Fig. 1.1 Free-electron energy band structure in the reduced zone scheme for face-centered-cubic lattices. The Fermi level is shown for different numbers of outershell electrons per unit cell. The degeneracy (other than the twofold spin degeneracy) of each energy band segment is indicated by the number of dots on the corresponding line. Symmetry points in the reduced zone (insert) are identified by Greek or Roman letters. The lattice constant (unit cube edge) is denoted by a. This diagram applies to such crystals as Al, Cu, Ag, Ge, and GaAs: it includes most of the solids discussed in detail in this book. [From F. Herman, Atomic Structure, *in* "An Atomistic Approach to the Nature and Properties of Materials" (J. A. Pask, ed.). Wiley, New York, 1967.]

energy band structure for a face-centered cubic crystal. The energy bands are shown for a number of important directions in **k**-space.

In a real crystal, the finite periodic perturbation of the lattice lifts many of the degeneracies of the free-electron model. An example is shown in Fig. 1.2.

If spin–orbit coupling is included in the crystal Hamiltonian, degeneracies

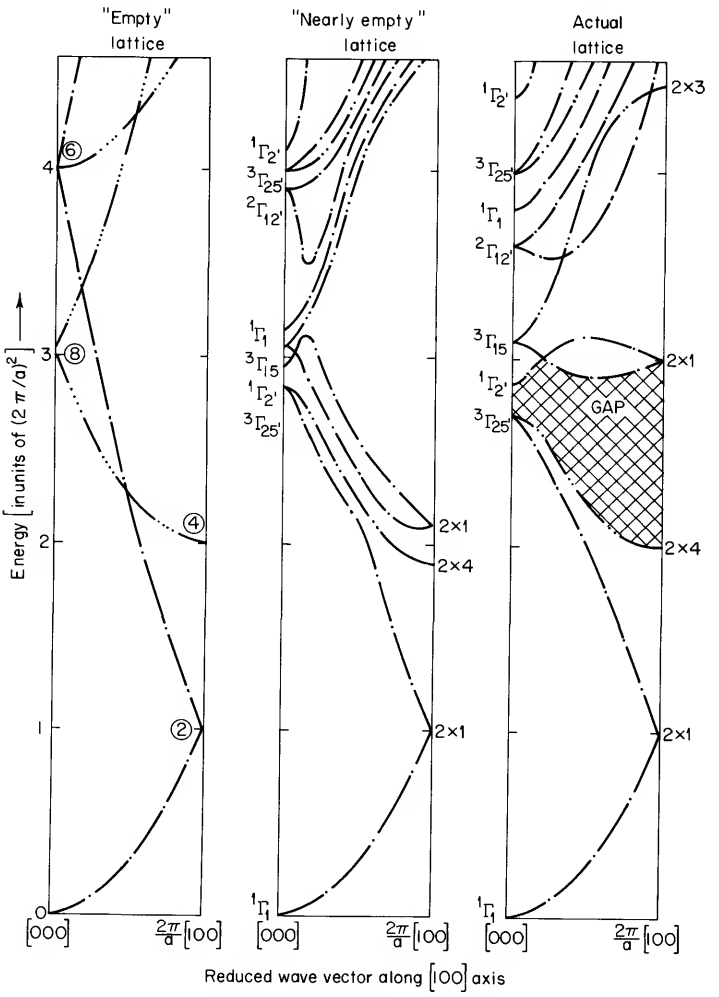

Fig. 1.2 Comparison of free-electron, nearly free-electron, and actual energy band models for the germanium crystal, for the [100] direction in the reduced zone. The spin–orbit splitting has been omitted. [From F. Herman, Atomic Structure, *in* "An Atomistic Approach to the Nature and Properties of Materials" (J. A. Pask, ed.). Wiley, New York, 1967.]

will be further lifted at some points in **k**-space. In general, throughout this book, however, we will speak of bands for which electrons have the same energy independent of whether they are spin up or spin down.

The agreement between experiment and theory for a wide variety of types of materials ranging from insulators to metals suggests that the one-electron model for solids is generally quite adequate. Many experiments point to the existence of a sharp Fermi surface in metals as expected for a one-electron model. There are, of course, some inadequacies in the treatment of the one-electron band model. These include the poor representation of electron–electron correlation effects and the variation in potential for electrons in different states. However, the general features of the one-electron band picture are real. In fact, more sophisticated treatments often give as much insight into the success of one-electron methods as they do in actually improving the results.

We shall assume that the one-electron band model is correct. Besides, it is not a model to be discarded lightly. It has the highly desirable feature that it is possible to use the Fermi–Dirac distribution function for a statistical description of the total electron population. It also means that when an electron changes its energy, there is no resultant change in any of the other electrons in the system. Thus, we can treat a change in energy of a single electron as a change in energy of the system.

1.2 Optical Reflectivity

When light of sufficient energy shines onto a material, it induces transitions of electrons from occupied states below the Fermi energy to unoccupied states above the Fermi energy. Clearly, a quantitative study of these transitions must provide some understanding of the initial and final states for the transitions and hence some knowledge of the band structure. But what sort of experiments are to be carried out, and how are they to be interpreted?

The most common experiments consist of shining a beam of monochromatic light onto a sample and measuring the fraction of the incident beam that is transmitted or reflected. In the spectral regions of greatest interest, optical absorption is generally quite high, so that often a negligibly small fraction of the incident light is transmitted. For example, in the visible and ultraviolet regions, many materials have a mean absorption depth of the order of only 100 Å. It is generally not feasible to make films that thin which are of sufficient quality to get meaningful data. Thus, most experiments are measurements of the reflectivity.

Figure 1.3 shows a scheme for measuring normal incidence reflectance. The measurements must usually extend out to photon energies of at least

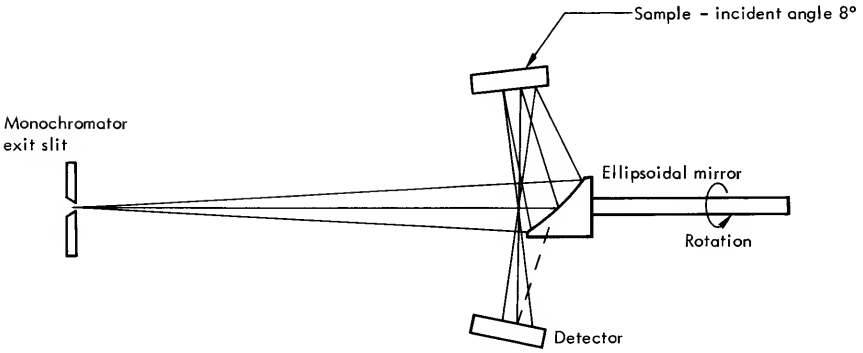

Fig. 1.3 Schematic of optical system for near-normal-incidence reflectivity measurements. The arrangement reduces the number of reflections to one. This is an important and valuable feature in the vacuum-ultraviolet, where the reflectance of solids is low and the use of a multiple-reflection scheme drastically reduces light intensity. The ellipsoidal mirror diverts the light beam consecutively onto the sample and the detector. Continuous rotation thus produces a sequence of "incident" and "reflected" pulses which are electronically measured and recorded. [From T. Huen, G. B. Irani, and F. Wooten, *Appl. Opt.* **10**, 552 (1971).]

10 eV if the major optical transitions are to be covered. Sometimes, of course, it is necessary to go to considerably higher energies to get sufficient data. However, most of the interesting structure usually lies at photon energies below 25 eV. This is a spectral region that is reasonably accessible with commercially available monochromators and light sources.

Figure 1.4 shows $(h\nu)^2\varepsilon_2$ for GaAs as determined from an analysis of the optical reflectance of GaAs over the spectral region 0–25 eV. This curve is a measure of the optical absorption in GaAs. The peaks and valleys in the "absorption" curve are of course related in some way to possible transitions between states in the energy bands. However, the interpretation of this structure is nontrivial. It is impossibly complicated without at least some knowledge of the band structure as a starting point. Thus, advancement in knowledge of the electronic band structure of solids has been a joint effort of theory and experiment. That this is so should not be surprising. Observe the band diagram for GaAs as shown in Fig. 1.5. It is hopeless to derive such a diagram solely from optical reflectance measurements. On the other hand, it is possible to determine band gaps quite accurately and provide checks on theoretical calculations.

The reason that optical data are often so difficult to interpret is that they depend on a summation over all possible transitions. For example, the photon energy required for the transition from L_3 to L_3 indicated in Fig. 1.5 is $\hbar\omega = 6$ eV. However, if one shines monochromatic light of that energy onto GaAs, the absorption depends on all possible transitions for which the occupied initial state and the unoccupied final state are separated by 6 eV. Furthermore, Fig. 1.5 shows the band structure only along certain

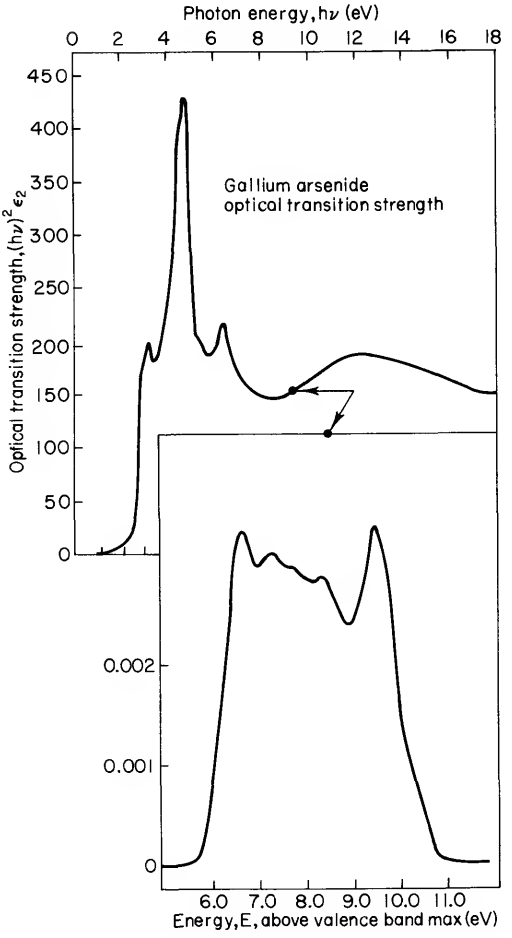

Fig. 1.4 Illustration that much more information can be obtained from a combination of photoemission and optical measurements than from either one alone. The upper curve, $(h\nu)^2\varepsilon_2$ versus $h\nu$, is a plot of a function related to the optical absorption strength as determined from a Kramers–Kronig analysis of the optical reflectivity. The lower curve gives a photoemission energy distribution curve for $h\nu = 10.4$ eV. Note that there is strong structure in the energy distribution curve, whereas the optical curve is quite smooth in that energy region. [From W. E. Spicer and R. C. Eden, *Proc. 9th Int. Conf. Phys. Semicond. (Moscow, USSR)* **2**, 65 (1968).]

directions in **k**-space. It is necessary to consider transitions at all points in the Brillouin zone. These considerations make it clear that disentangling everything is not easy. Fortunately, it is also not completely hopeless. Certain transitions, especially near symmetry points, are sometimes suf-

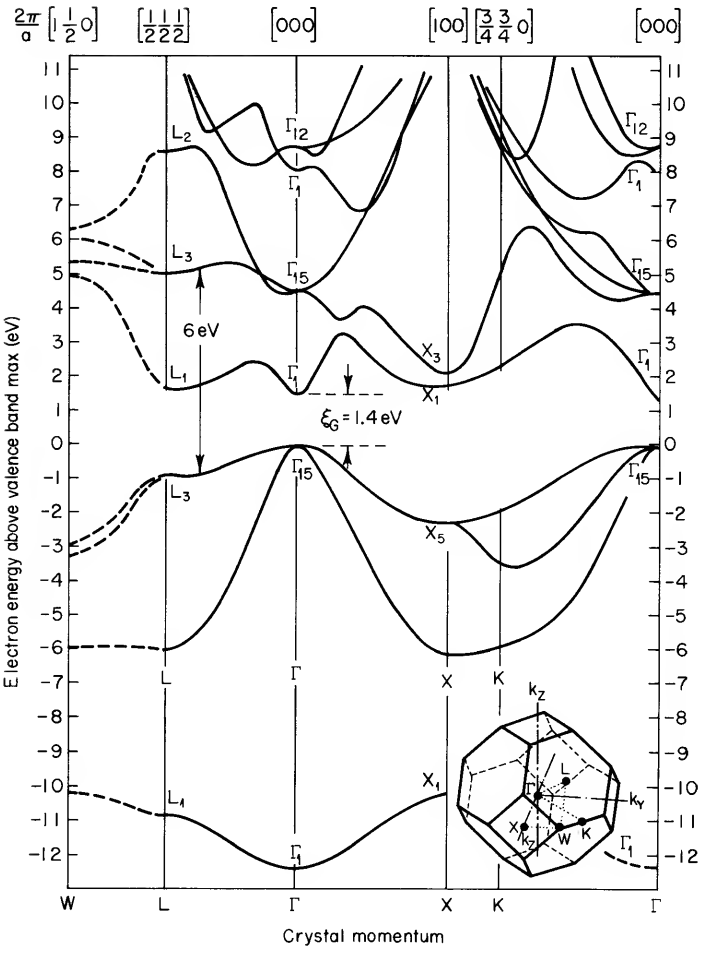

Fig. 1.5 Energy band diagram for GaAs. Compare this diagram with Fig. 1.1 to see the effect of the finite periodic crystal potential in GaAs. [Based on M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* **141**, 789 (1966).]

ficiently strong that they show up quite clearly and can be readily interpreted. These often provide a kind of framework for the rest of the structure.

1.3 Photoemission

When light shines onto a material, it may excite some of the electrons to energies sufficiently high that they can escape from the material. This is

the photoelectric effect. The optical absorption that occurs in photoemission is exactly the same process that we have just discussed in relation to reflectance measurements. Thus, there is a close relationship between the information that can be extracted from photoemission experiments and from purely optical measurements.

In photoemission, interest is focused on the escaping electrons. These electrons constitute current which contains information on the band structure. Indeed, photoemission has emerged during the last decade as a powerful tool for investigating the electronic band structure of solids.

Let us consider briefly on the basis of a simple three-step model what happens in the photoelectric effect. First, a photon is absorbed and an electron makes a transition from an occupied state to one of the higher-energy empty states. The electron then moves through the crystal, possibly suffering numerous collisions. Finally, the excited electron may reach the surface and have sufficient momentum to escape over the surface barrier. These steps are illustrated in Figs. 1.6 and 1.7.

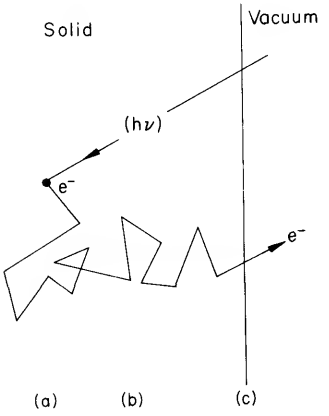

Fig. 1.6 (a) Photoexcitation of an electron, (b) random motion through the crystal, (c) and electron escape. [From W. E. Spicer and F. Wooten, *Proc. IEEE* **51**, 1119 (1963).]

From the simple three-step mechanistic model of photoemission, it is clear that a detailed study of photoemission from a solid should provide information on the band structure as well as electron transport at energies well above the Fermi energy. The difficulty lies in extracting the information contained in the data. We shall consider some of these difficulties later.

The basic experiment in photoemission consists of shining monochromatic light onto a sample (cathode) placed near the center of a spherical collector (anode) and measuring the photoelectric current as a function of voltage. This is illustrated in Fig. 1.8. By varying the photon energy, a series of $I-V$ curves can be obtained. The form of these curves and their relationship to the energy band profile of solids are illustrated in Figs. 1.9–1.12. However, it is not just a series of $I-V$ curves that is desired. The point

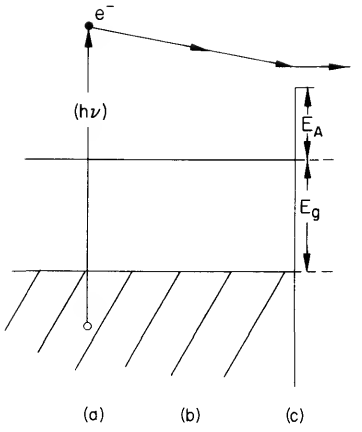

Fig. 1.7 (a) Photoexcitation of an electron, (b) random motion through the crystal, and (c) electron escape in terms of a simple energy band diagram for a semiconductor. The continuous loss in energy during diffusion to the surface as indicated in the figure represents energy loss from a number of nearly elastic collisions. Loss of energy between excitation and escape contributes to the difficulty of interpreting photoemission experiments. [From W. E. Spicer and F. Wooten, *Proc. IEEE* **51**, 1119 (1963).]

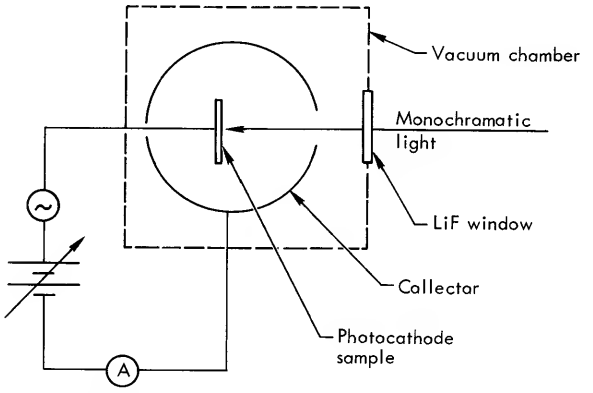

Fig. 1.8 Schematic diagram of basic arrangement for photoemission measurements. The photocathode–collector voltage is slowly varied during measurements of photocurrent versus retarding voltage. For quantum yield measurements, the collector is biased sufficiently positive to collect all photoelectrons. The LiF window transmits to about 11 eV; at higher photon energies, a windowless experiment must be performed. The pressure in the vacuum chamber must be of the order of 10^{-9} Torr or less to maintain a clean sample surface during measurements.

is that it is the structure in the energy distribution of photoelectrons that is related to the electronic band structure. This information can, in principle, be obtained by numerically differentiating the current–voltage curves just mentioned. However, a much more satisfactory method is to electronically differentiate the $I-V$ curve. One way this can be done is by superimposing a small ac voltage on a variable dc retarding voltage between the photo-

cathode and the spherical collector. Then, as the retarding voltage is slowly swept over the range from saturation current to zero current, the ac component of photocurrent can be detected (see Fig. 1.13.). The ac current component is directly proportional to the photoelectron energy distribution.

Figure 1.4 shows the photoelectron energy distribution curve for GaAs with $h\nu = 10.4$ eV. Note that near this photon energy, there is no structure in the optical absorption, yet there are five peaks in the photoemission curve. This is a striking example of the value of photoemission measurements. It illustrates the way in which photoemission helps to unravel band structure by resolving many of the transitions which contribute to optical absorption at a particular photon energy. Note also that since the final-state energy of the electrons is measured, photoemission determines in an absolute sense the energy of the initial and final states. Purely optical measurements determine only differences in energy between states.

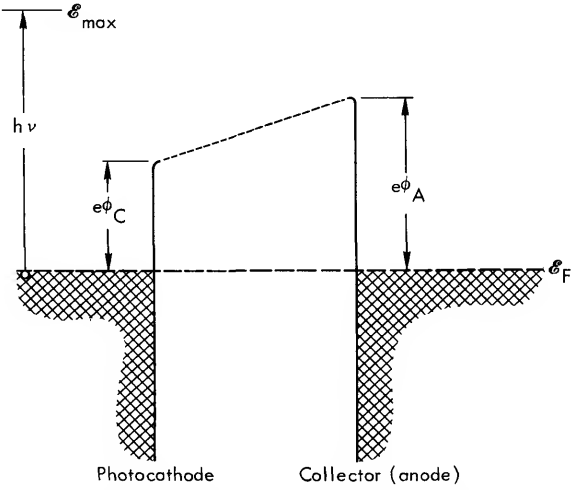

Fig. 1.9 Energy profile for photocathode and collector made of different metals. The maximum energy for a photoelectron just outside the photocathode is $h\nu - e\phi_C$, the energy of an electron excited from the Fermi energy but losing energy $e\phi_C$ in escaping from the photocathode. Once outside the photocathode, the electron experiences a built-in field $\phi_C - \phi_A$ arising from the contact potential difference between the two metals. On heading toward the collector, the electron must overcome the built-in field and thus loses more kinetic energy. It arrives at the collector (if it has sufficient energy) with a maximum kinetic energy $h\nu - e\phi_A$. Thus a retarding voltage of $(h\nu/e) - \phi_A$ is sufficient to completely cut off the photocurrent. Most electrons will be emitted with lower energies. Those emitted with zero energy require an accelerating voltage $\phi_A - \phi_C$ in order to be collected. The resultant $I-V$ curve based on this energy profile is illustrated in Fig. 1.10.

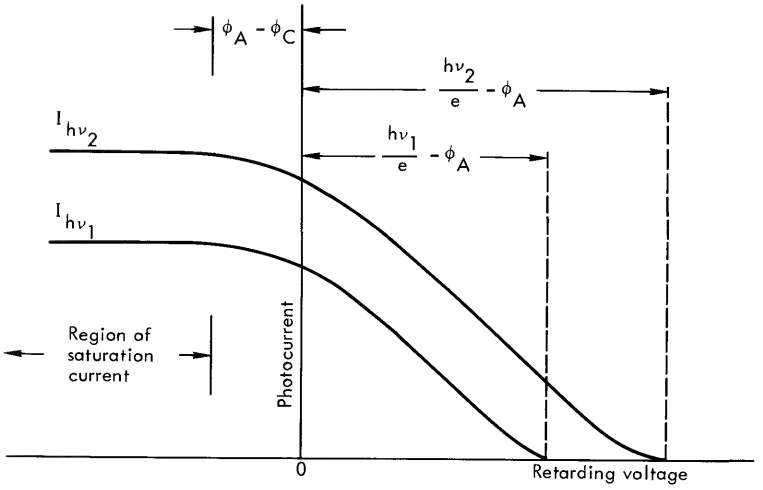

Fig. 1.10 Photocurrent versus retarding voltage for the photocathode and collector of Fig. 1.9. Note that the end of the region of saturation current and the beginning of photocurrent cutoff depend only on the contact potential difference. The cutoff voltage for zero photocurrent depends on photon energy and the work function of the collector.

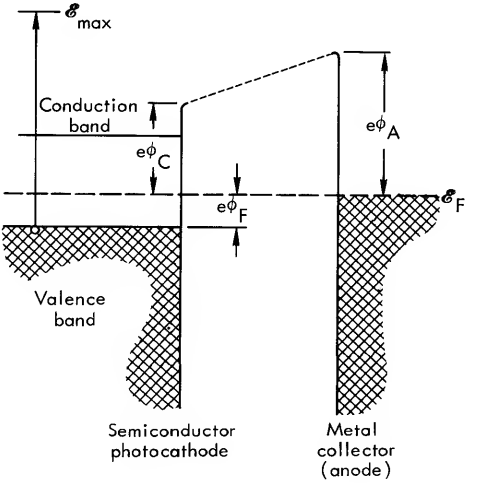

Fig. 1.11 Energy profile for a semiconductor photocathode and metal anode. The threshold for photoemission from a semiconductor is from the valence band maximum to the vacuum level, not from the Fermi level to the vacuum level. The effect is to change the cutoff voltage for zero current. Compare Figs. 1.10 and 1.12.

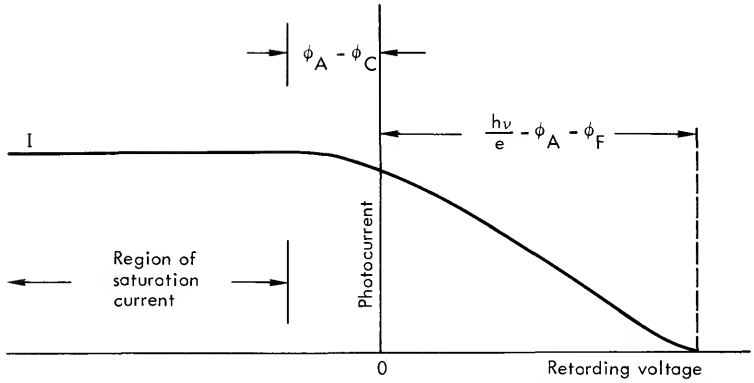

Fig. 1.12 Photocurrent versus retarding voltage for the semiconductor photocathode and metal anode of Fig. 1.11. The cutoff voltage differs from that for a metal photocathode by ϕ_F, the Fermi potential measured with respect to the valence band maximum.

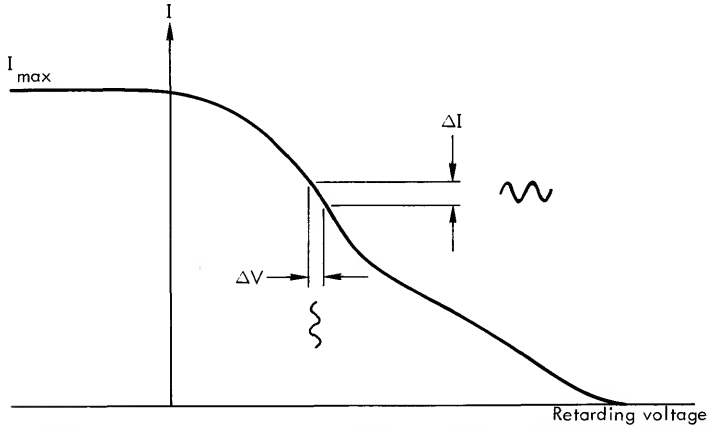

Fig. 1.13 Modulation of I–V curve with small ac voltage permits direct measurement of $\Delta I/\Delta V$. This is the method used to obtain the energy distribution curve for GaAs shown in Fig. 1.4.

Figure 1.4 shows a photoelectron energy distribution curve for only one photon energy. It is necessary to repeat the measurement for many photon energies. Then, by studying the movement of structure in energy and its change in magnitude, one can deduce considerable information about the band structure and selection rules. However, to compare the magnitude of different curves requires another measurement, namely the quantum yield. This consists in measuring the total photocurrent response to a monochromatic light beam of known intensity. The quantum yield, which is the number of electrons emitted per photon absorbed, is just the

ratio of photoemission current to "photon current." Then, the energy distribution curves are normalized such that the area under a curve equals the quantum yield.

One of the major difficulties in interpreting photoemission data is separating out the effects of electron scattering. In an ideal crystalline solid, the dominant scattering mechanisms are electron–phonon scattering and electron–electron scattering. The former is nearly elastic and mostly just randomizes the motion of electrons; the latter can lead to large energy losses. Both can have important effects on photoemission. Some of these effects are discussed at appropriate points in later chapters.

1.4 Characteristic Energy Loss Spectra

There are numerous ways of exciting electrons in a solid other than by photon absorption. One way that is related to optical spectra is the excitation of electrons by other electrons. This is done by shining a beam of monoenergetic electrons at a sample and analyzing the energy of the transmitted or reflected beam. It is found that the incident electrons lose energy in discrete amounts. The loss spectrum arises both from the excitation of single electrons in the solid, just as happens with photon absorption, and also from the excitation of collective oscillations called plasmons. From these measurements, it is also possible to deduce optical constants. It is also feasible to make measurements corresponding to a photon energy range not always accessible by traditional optical means.

Chapter 2

MAXWELL'S EQUATIONS AND THE DIELECTRIC FUNCTION

The optical property usually available directly from experiment is the frequency-dependent reflectance or transmittance; the property most directly related to the electronic structure of a solid is the dielectric function. To interpret experimental measurements in terms of the fundamental electronic properties of the solid requires an understanding of Maxwell's equations, the nature of the interaction between electromagnetic fields and matter, and an understanding of the dielectric function from a fundamental microscopic viewpoint.

We begin this chapter with a discussion of Maxwell's equations using a notation standard in most texts on electromagnetic theory. After describing the properties of the medium, a simple case of the interaction of light and a solid is considered. This introduces some of the basic ideas. Then, we take a closer look at the description of fields and sources (light or electron beams) and make some changes in notation. The new notation differs only slightly from the old. It is made to conform with that used by many active workers in the field and to emphasize the physics involved in the interactions. Finally, the results are generalized and the dielectric tensor is introduced in such a way as to make clear the relationships and distinctions between longitudinal and transverse dielectric functions.

2.1 Maxwell's Microscopic Equations

The relationship of the microscopic properties of matter to the macroscopic optical properties is best understood by beginning with the microscopic form of Maxwell's equations. They are,

$$\nabla \cdot \mathbf{e} = 4\pi\rho_{\text{micro}} \tag{2.1}$$

$$\nabla \times \mathbf{e} = -\frac{1}{c}\frac{\partial \mathbf{b}}{\partial t} \tag{2.2}$$

$$\nabla \cdot \mathbf{b} = 0 \tag{2.3}$$

$$\nabla \times \mathbf{b} = \frac{1}{c}\frac{\partial \mathbf{e}}{\partial t} + \frac{4\pi}{c}\mathbf{j}_{\text{micro}} \tag{2.4}$$

The equations are written in Gaussian units, the system most commonly used in quantum mechanics and modern literature on optical properties of solids. The vectors $\mathbf{e} = \mathbf{e}(\mathbf{r}, t)$ and $\mathbf{b} = \mathbf{b}(\mathbf{r}, t)$ are the microscopic electric and magnetic fields, respectively; ρ_{micro} and $\mathbf{j}_{\text{micro}}$ are the microscopic charge and current densities, respectively. These equations are completely general, applying to dielectrics as well as metals.

With the microscopic form of Maxwell's equations, one considers the contribution of each charged particle. For example, in a classical model in which the electrons and atomic nuclei within a system are treated as point charges, the charge density is given by

$$\rho_{\text{micro}}(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{2.5}$$

where q_i is the charge of the ith particle and $\delta(\mathbf{r} - \mathbf{r}_i)$ is the Dirac delta function.

In the treatment of optical problems from a quantum mechanical viewpoint, it is not possible to treat electrons as distinguishable charged particles. Then, it is necessary to specify the charge density for electrons in terms of the spatial probability distribution and electronic charge $-e$ as

$$\rho_{\text{microelectronic}}(\mathbf{r}) = -e\Psi^*(\mathbf{r})\,\Psi(\mathbf{r}) \tag{2.6}$$

where $\Psi(\mathbf{r})$ is the electronic wave function of the system. The atomic nuclei can still be treated as point charges, but it is frequently more convenient to consider a smoothed-out, positive charge distribution.

2.2 Maxwell's Macroscopic Equations

It is clearly not feasible to work with the microscopic form of Maxwell's equations for real systems containing of the order of 10^{22} electrons per

cm^3. It is possible, though, to rewrite Maxwell's equations in terms of macroscopic quantities and still retain the form of the microscopic equations. This is an essential step if Maxwell's equations are to be of any use in the study of solids. That such a step is possible is because optical measurements provide a probe with a spatial resolution of the order of a wavelength of light. Since a small volume element of a solid with characteristic dimensions of the order of a wavelength of light contains many millions of atoms, solids can be treated as continuous from the viewpoint of interpreting optical measurements.

Macroscopic quantities can be defined in terms of their microscopic counterparts as follows. The electric field strength $\mathbf{E}(\mathbf{r})$ and magnetic induction $\mathbf{B}(\mathbf{r})$ are

$$\mathbf{E}(\mathbf{r}) = \langle \mathbf{e}(\mathbf{r}) \rangle = (1/\Delta V) \int_{\Delta V} \mathbf{e}(\mathbf{r} + \boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{2.7}$$

$$\mathbf{B}(\mathbf{r}) = \langle \mathbf{b}(\mathbf{r}) \rangle = (1/\Delta V) \int_{\Delta V} \mathbf{b}(\mathbf{r} + \boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{2.8}$$

where $\int d\boldsymbol{\xi}$ symbolizes $\int dx \int dy \int dz$. The charge and current densities are given by

$$\rho^{\text{total}}(\mathbf{r}) = (1/\Delta V) \int_{\Delta V} \rho_{\text{micro}}(\mathbf{r} + \boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{2.9}$$

$$\mathbf{J}^{\text{total}}(\mathbf{r}) = (1/\Delta V) \int_{\Delta V} \mathbf{j}_{\text{micro}}(\mathbf{r} + \boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{2.10}$$

In Eqs. (2.7)–(2.10), the volume element ΔV is taken to be of linear dimensions small compared with the wavelength of light but large enough to contain many atoms. The same procedure is also valid when time dependence is included. The charge density and current density are designated as total densities in anticipation of separating these densities into "free" and "bound" parts or induced and external parts.

By starting with Maxwell's microscopic equations and following the averaging procedure just described, the following macroscopic equations are obtained:

$$\nabla \cdot \mathbf{E} = 4\pi\rho^{\text{total}} \tag{2.11}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{2.12}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.13}$$

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{J}^{\text{total}} \tag{2.14}$$

It is implicit that the fields are total macroscopic fields.

No properties of the medium have yet been introduced. Before doing so, we shall first discuss formal solutions of Maxwell's equations.

2.3 Formal Solutions of Maxwell's Equations

Imagine that an electron is suddenly moved. The moving electron constitutes a current and produces a magnetic field satisfying Eq. (2.14). The presence of a magnetic field where none existed before means there has been a changing magnetic field. This in turn implies an electric field produced by the time-varying magnetic field and described by Eq. (2.12). Thus, once a field has been started propagating through free space, it is self-sustaining. If the **E** field begins to collapse, it produces a **B** field. A collapsing **B** field in turn produces an **E** field. We thus expect the field vectors **E** and **B** to satisfy a wave equation. We know that such is the case, and we shall later carry through an analysis for propagation of electromagnetic waves in a medium which also includes an energy-absorption mechanism. For now, the treatment will be somewhat more formal. We shall show that the fields **E** and **B** are derivable from potential functions and that these potentials satisfy a wave equation. We shall also consider certain useful properties of these potential functions.

Equation (2.13) implies the existence of a vector potential **A** such that

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{2.15}$$

Equation (2.12) then becomes

$$\nabla \times \left(\mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}\right) = 0 \tag{2.16}$$

Now, a vector whose curl is zero must be the gradient of some scalar potential function, thus

$$\mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} = -\nabla\phi \tag{2.17}$$

Note that if $\mathbf{B} = \nabla \times \mathbf{A}$, it is possible to choose a new vector potential

$$\mathbf{A}' = \mathbf{A} + \nabla\psi \tag{2.18}$$

such that

$$\mathbf{B} = \nabla \times \mathbf{A}' \equiv \nabla \times \mathbf{A} \tag{2.19}$$

There is no unique vector potential; it can be shifted arbitrarily by any gradient of a scalar field. Since $\nabla \times \nabla\psi \equiv 0$, the new vector potential $\mathbf{A}'$ still gives the correct magnetic induction $\mathbf{B}$. The transformation described by Eq. (2.18) is called a gauge transformation.

Since the vector potential $\mathbf{A}$ is also included in the expression for $\mathbf{E}$, Eq. (2.17), it is necessary if $\mathbf{A}$ is changed to $\mathbf{A}' = \mathbf{A} + \nabla\psi$ to also change ϕ to $\phi' = \phi - (1/c)(\partial\psi/\partial t)$. Then, neither $\mathbf{B}$ nor $\mathbf{E}$ is changed.

One way to restrict $\mathbf{A}$ is to specify the divergence of $\mathbf{A}$. A common choice is to let $\nabla \cdot \mathbf{A} = 0$. This is called the Coulomb gauge for reasons soon to be made clear. $\mathbf{A}$ is restricted in the Coulomb gauge, but is not uniquely defined. It is still possible to make transformations. For example, taking the divergence of Eq. (2.18) in the Coulomb gauge yields $\nabla \cdot \mathbf{A}' = \nabla^2\psi$ and it is only necessary to require that $\nabla^2\psi$ be zero everywhere.

The most convenient choice for $\mathbf{A}$ depends on the type of problem to be analyzed. For applications to optical properties of solids, the Coulomb gauge is usually preferable.

Substituting Eq. (2.17) into (2.11), we obtain

$$\nabla^2\phi + \frac{1}{c}\frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -4\pi\rho^{\text{total}} \tag{2.20}$$

In the Coulomb gauge, $\nabla \cdot \mathbf{A} = 0$, and the scalar potential satisfies Poisson's equation

$$\nabla^2\phi = 4\pi\rho^{\text{total}} \tag{2.21}$$

Now, we see that the scalar potential is just the instantaneous (i.e., unretarded) potential familiar from electrostatics. That is why the gauge with $\nabla \cdot \mathbf{A} = 0$ is called the Coulomb gauge. Of course, the electric field is not determined just by the unretarded scalar potential; it is specified by Eq. (2.17), and the effects of time retardation must be included in the vector potential $\mathbf{A}$.

Now, return to the vector potential $\mathbf{A}$. Substituting Eq. (2.17) into (2.14) gives

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial}{dt}\left(-\frac{1}{c}\frac{\partial \mathbf{A}}{dt} - \nabla\phi\right) + \frac{4\pi}{c}\mathbf{J}^{\text{total}} \tag{2.22}$$

Substituting for $\mathbf{B}$ from Eq. (2.15) and using the identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2\mathbf{A} \tag{2.23}$$

we obtain

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{dt^2} - \nabla\left(\nabla \cdot \mathbf{A} + \frac{1}{c}\frac{\partial\phi}{dt}\right) = -\frac{4\pi}{c}\mathbf{J}^{\text{total}} \tag{2.24}$$

Choose the Coulomb gauge once again and **A** satisfies the equation

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{dt^2} = -\frac{4\pi}{c}\mathbf{J}^{\text{total}} + \frac{1}{c}\nabla\frac{\partial \phi}{dt} \tag{2.25}$$

We want to discuss Eq. (2.25) in greater depth, to gain some further insight into the significance of the Coulomb gauge and the wave equation which **A** must satisfy. First, though, it is necessary to digress briefly and discuss the resolution of fields and currents into longitudinal and transverse parts.

Recall that for electromagnetic waves in homogeneous media, the fields **E** and **B** are prependicular to the direction of propagation, that is, perpendicular to the direction of maximum spatial variation in **E** and **B**; the fields **E** and **B** are transverse! Static fields are longitudinal; **E** and **B** are then decoupled and each is parallel to the direction of its maximum spatial variation. The decomposition of vector fields into longitudinal and transverse parts is discussed further in Appendix A.

For illustration, we take the case of a single plane wave

$$\mathbf{E} = \mathbf{E}_0 \exp - i(\omega t - \mathbf{q} \cdot \mathbf{r}) \tag{2.26}$$

Taking the curl and divergence of Eq. (2.26), we get

$$\nabla \times \mathbf{E} = i\mathbf{q} \times \mathbf{E}$$

$$\nabla \cdot \mathbf{E} = i\mathbf{q} \cdot \mathbf{E} \tag{2.28}$$

If we now write the electric field in terms of its transverse and longitudinal parts, Eqs. (2.27) and (2.28) become

$$\nabla \times (\mathbf{E}^{\mathrm{T}} + \mathbf{E}^{\mathrm{L}}) = i\mathbf{q} \times (\mathbf{E}^{\mathrm{T}} + \mathbf{E}^{\mathrm{L}}) \tag{2.29}$$

$$\nabla \cdot (\mathbf{E}^{\mathrm{T}} + \mathbf{E}^{\mathrm{L}}) = i\mathbf{q} \cdot (\mathbf{E}^{\mathrm{T}} + \mathbf{E}^{\mathrm{L}}) \tag{2.30}$$

Now, the wave vector **q** is parallel to the direction of propagation, as is $\mathbf{E}^{\mathrm{L}}$; but $\mathbf{E}^{\mathrm{T}}$ is perpendicular to the direction of propagation. Thus,

$$i\mathbf{q} \times \mathbf{E}^{\mathrm{L}} = 0 \tag{2.31}$$

$$i\mathbf{q} \cdot \mathbf{E}^{\mathrm{T}} = 0 \tag{2.32}$$

These in turn require that

$$\nabla \times \mathbf{E}^{\mathrm{L}} = 0 \tag{2.33}$$

$$\nabla \cdot \mathbf{E}^{\mathrm{T}} = 0 \tag{2.34}$$

and hence

$$\nabla \times \mathbf{E} = \nabla \times \mathbf{E}^{\mathrm{T}} \tag{2.35}$$

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{E}^{\mathrm{L}} \tag{2.36}$$

Similar relationships hold for any vector field (see Appendix A). It is not necessary, for example, to use a plane wave such as Eq. (2.26) which depends upon a third vector $\mathbf{q}$.

We now return to Eq. (2.25). The term $(1/c)\,\nabla(\partial\phi/\partial t)$ corresponds to a current density. We expect that it is related only to the longitudinal part of the total current density. Why should that be so? It is because only the longitudinal part of the electric field is derivable from a potential function ϕ. This is apparent from Eq. (2.33). Because the curl of $\mathbf{E}^{\mathrm{L}}$ is zero, it must be derivable from a potential function. We now show explicitly that $(1/c)\,\nabla(\partial\phi/\partial t)$ is related only to the longitudinal current density.

The solution to Eq. (2.21) is obtained from electrostatics. Thus,

$$\phi(\mathbf{r}, t) = \int \frac{\rho^{\text{total}}(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \tag{2.37}$$

$$\frac{\partial\phi(\mathbf{r}, t)}{\partial t} = \int \frac{(\partial/\partial t)\,\rho^{\text{total}}(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \tag{2.38}$$

Using the continuity equation

$$\nabla' \cdot \mathbf{J}_{\text{total}} = -\partial\rho^{\text{total}}/\partial t \tag{2.39}$$

where ∇' indicates the operation is with respect to $\mathbf{r}'$, Eq. (2.38) becomes

$$\frac{\partial\phi}{\partial t} = -\int \frac{\nabla' \cdot \mathbf{J}^{\text{total}}}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \tag{2.40}$$

Next, decompose the current density into transverse and longitudinal parts. Then, since the divergence of a transverse vector field is zero,

$$\nabla' \cdot \mathbf{J}^{\mathrm{T}}_{\text{total}} = 0 \tag{2.41}$$

and Eq. (2.40) can be expressed in terms of the longitudinal current density alone. Then, operating on Eq. (2.40) with the gradient operator $\nabla^2 = \nabla^2(\mathbf{r})$, the result is

$$\nabla^2 \frac{\partial\phi}{\partial t} = -\nabla^2 \int \frac{\nabla' \cdot \mathbf{J}^{\mathrm{L}}_{\text{total}}}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \tag{2.42}$$

Finally, using the property of a delta function that

$$\nabla^2 \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} \right) = -4\pi\delta(\mathbf{r} - \mathbf{r}') \tag{2.43}$$

Eq. (2.42) simplifies as follows

$$\nabla^2 \frac{\partial\phi}{\partial t} = 4\pi \int \nabla' \cdot \mathbf{J}^{\mathrm{L}}_{\text{total}}(\mathbf{r}', t)\, \delta(\mathbf{r} - \mathbf{r}')\, d\mathbf{r}' \tag{2.44}$$

$$\nabla \cdot \left(\nabla \frac{\partial \phi}{\partial t} \right) = 4\pi \nabla \cdot \mathbf{J}^{\mathrm{L}}_{\mathrm{total}}(\mathbf{r}, t) \tag{2.45}$$

$$\nabla \frac{\partial \phi}{\partial t} = 4\pi \mathbf{J}^{\mathrm{L}}_{\mathrm{total}} \tag{2.46}$$

We see that $\nabla(\partial\phi/\partial t)$ determines only the longitudinal part of the current density. If we decompose the current density into its transverse and longitudinal parts, and use Eq. (2.46), we can rewrite Eq. (2.25) as a wave equation for $\mathbf{A}$ expressed solely in terms of the transverse current density,

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi}{c} \mathbf{J}^{\mathrm{T}}_{\mathrm{total}} \tag{2.47}$$

The Coulomb gauge is sometimes referred to as the transverse gauge. The reason is now clear. In the Coulomb gauge, or transverse gauge, the vector potential $\mathbf{A}$ is also transverse, so we could label the vector potential as $\mathbf{A}^{\mathrm{T}}$ in Eq. (2.47).

2.4 Analysis of Charge and Current Densities

It is sometimes convenient to split the macroscopic charge and current densities into two parts: a bound and a free contribution. For example, the current density in a solid may be considered to consist of two contributions, one arising from the motion of electrons bound to nuclei, and thus restricted to localized motion, and the other arising from electrons free to move through the solid. Such a viewpoint is common and often helpful, but splitting current densities in this way is not unique. Electrons which are bound at one frequency of electromagnetic radiation might better be described as free at a higher frequency. Nonetheless, for now, we shall write

$$\mathbf{J}^{\mathrm{total}} = \mathbf{J}^{\mathrm{free}} + \mathbf{J}^{\mathrm{bound}} \tag{2.48}$$

We could also describe the charge density as the sum of two contributions: a free and a bound part. This is sometimes done. We shall not do it, because it is confusing. What is meant by free current density does not correspond to the movement of free charge density. This is made clear by a simple example: Current flow in an ideal metal is the result of electrons moving freely through the metal, but there is no charge density, free or bound, because the electronic charge is exactly balanced by the positive background of atomic nuclei. Charge density ρ refers to a net charge density.

There are two ways of obtaining a net charge density in some region of a solid. One is the displacement of charge within the medium, building

up net charge density at one place and depleting it in another; the second is the introduction of a net extra charge density from an external source. Rather than splitting the charge density into bound and free parts, we will describe the partioning as polarization and external charge densities. Thus,

$$\rho^{\text{total}} = \rho^{\text{pol}} + \rho^{\text{ext}} \tag{2.49}$$

What is the nature of the polarization charge density? In the presence of an electric field, the atoms in the solid are polarized. The electronic charge distribution of each atom is displaced with respect to the nucleus. This is illustrated in Fig. 2.1, which is a greatly simplified drawing of a region in a solid which contains many atoms but is less than a wavelength in dimensions. If the polarization is uniform, there is no net charge moved into or out of the region. For nonuniform polarization, though, there is a net change in the charge within the region. This is illustrated in Fig. 2.1(b) and is sometimes described by

$$\rho^{\text{bound}} = -\nabla \cdot \mathbf{P} \tag{2.50}$$

where the polarization $\mathbf{P}$ is the dipole moment per unit volume. Since ρ^{bound} is the net charge density arising from polarization, we shall refer to it as the polarization charge density,

$$\rho^{\text{pol}} = -\nabla \cdot \mathbf{P} \tag{2.51}$$

as a reminder of the origin of this part of the total macroscopic charge density. In the presence of a time-dependent electric field, the resulting time-dependent polarization gives a current density

$$\mathbf{J}^{\text{pol}} = \partial \mathbf{P}/\partial t \tag{2.52}$$

which contributes to the total macroscopic current density.

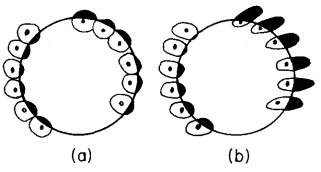

Fig. 2.1 Idealized model of a solid. (a) No electric field. (b) Polarization in presence of a spatially varying electric field. The charge removed from the region considered does not equal the charge that enters.

There may also be a contribution to the current density arising from electron spin. Such a contribution can be included in terms of the macroscopic magnetization $\mathbf{M}$, defined as the magnetic dipole moment per

unit volume and corresponding to a current density contribution

$$\mathbf{J}^{\text{mag}} = c\nabla \times \mathbf{M} \tag{2.53}$$

Magnetic effects are usually small. They will be of no direct concern to us, but we shall nonetheless formally include magnetization in a modified version of Maxwell's equations. However, the very small magnetic effects that arise from the motion of conduction electrons, such as Landau diamagnetism, will be completely ignored.

Including the effects of magnetization and polarization, we have

$$\mathbf{J}^{\text{bound}} = (\partial \mathbf{P}/\partial t) + c\nabla \times \mathbf{M} \tag{2.54}$$

The current density $\mathbf{J}^{\text{free}}$ consists of two parts. One is that contribution $\mathbf{J}^{\text{cond}}$ arising from the motion of conduction electrons in the presence of an electric field; the other consists of a current density $\mathbf{J}^{\text{ext}}$ introduced into the system from an external source. Of course, we are not so much interested in the usual electrical conductivity of solids; we include the absorption of photons in $\mathbf{J}^{\text{cond}}$ and think in terms of an optical conductivity which in the limit of zero frequency equals the dc electrical conductivity. The latter is strictly true for cubic materials; for noncubic systems, some complications arise because the optical conductivity describes the response of the system to a transverse electric field, whereas the dc electrical conductivity refers to a longitudinal field. These points are discussed in Section 2.9.

We now use Eqs. (2.48)–(2.54) to rewrite Maxwell's equations (2.11) and (2.14) as

$$\nabla \cdot \mathbf{E} = -4\pi\nabla \cdot \mathbf{P} + 4\pi\rho^{\text{ext}} \tag{2.55}$$

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\frac{\partial \mathbf{P}}{\partial t} + 4\pi\nabla \times \mathbf{M} + \frac{4\pi}{c}\mathbf{J}^{\text{cond}} + \frac{4\pi}{c}\mathbf{J}^{\text{ext}} \tag{2.56}$$

Defining two new vectors, the displacement

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \tag{2.57}$$

and the magnetic field strength

$$\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M} \tag{2.58}$$

we finally write Maxwell's equations in their common macroscopic form:

$$\nabla \cdot \mathbf{D} = 4\pi\rho^{\text{ext}} \tag{2.59}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{2.60}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.61}$$

$$\nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c}\mathbf{J}^{\text{cond}} + \frac{4\pi}{c}\mathbf{J}^{\text{ext}} \tag{2.62}$$

It should be noted that these equations are exact.

2.5 Properties of the Medium

So far, we have little more than hinted at the properties of the medium. It is clear that the polarization and free current density are related to the electronic field strength and that the magnetization is related to the magnetic induction and hence to the magnetic field strength.

Consider the polarization. The relationship of the ith spatial component of the polarization can, for example, be expressed in terms of the electronic field components by a power series of the form

$$P_i = \sum_j \chi_{ij} E_j + \sum_{j,k} \gamma_{ijk} E_j E_k + \cdots \tag{2.63}$$

With the advent of lasers, it is now quite common to observe nonlinear optical effects. However, the concern here is only with linear optics, and only linear terms will be retained in expressions such as Eq. (2.63).

We now simplify the mathematics by considering only isotropic media. Then, within the linear approximation, we make the usual assumptions

$$\mathbf{P} = \chi_e \mathbf{E} \tag{2.64}$$

$$\mathbf{M} = \chi_m \mathbf{H} \tag{2.65}$$

$$\mathbf{J}^{\text{cond}} = \sigma \mathbf{E} \tag{2.66}$$

where χ_e is the electric susceptibility, χ_m is the magnetic susceptibility, and σ is the conductivity.

Two alternative parameters which are used to characterize the medium are defined by

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{2.67}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{2.68}$$

The parameter ε is usually called the dielectric constant or dielectric function. The parameter μ is the magnetic permeability; it equals unity for nonmagnetic materials.

The dielectric function introduced in Eq. (2.67) is real. It is a function of the space and time variables because we have been working with the space–time representation of Maxwell's equations. More precisely, it is a response function, or linear integral operator, that connects the field

$\mathbf{D}(\mathbf{r}, t)$ with the field $\mathbf{E}(\mathbf{r}', t')$ existing at all other positions and all earlier times. Thus, in general,

$$\mathbf{D}(\mathbf{r}, t) = \int d\mathbf{r}' \int_{-\infty}^{t} dt' \, \varepsilon(\mathbf{r}, \mathbf{r}', t, t') \, \mathbf{E}(\mathbf{r}', t') \tag{2.69}$$

The dielectric function as a response function is treated in more detail in Chapter 6. There, and throughout most of this book, we shall be concerned mostly with $\varepsilon(\mathbf{q}, \omega)$, which is the Fourier transform of the general response function $\varepsilon(\mathbf{r}, \mathbf{r}', t, t')$.

The preceding assumptions represent an attempt to describe the properties of matter and it should be clearly understood that we have made approximations. It should also be noted that χ_e, χ_m, σ, and ε are macroscopic quantities because they are defined in terms of macroscopic quantities. Indeed, Eqs. (2.64)–(2.68) are of little use as they stand; at this point, they represent little more than a conviction that such relationships are reasonable. Since our main interest is to understand microscopic properties, we must relate these parameters to microscopic quantities that can be expressed explicitly in a way that gives some insight into the fundamental physics involved. This is a point to which we shall frequently return.

The properties of the medium can now be included in Maxwell's equations. We then get the following set of equations, which is no longer exact:

$$\nabla \cdot (\varepsilon \mathbf{E}) = 4\pi \rho^{\text{ext}} \tag{2.70}$$

$$\nabla \times \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t} \tag{2.71}$$

$$\nabla \cdot (\mu \mathbf{H}) = 0 \tag{2.72}$$

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial}{\partial t} (\varepsilon \mathbf{E}) + \frac{4\pi\sigma}{c} \mathbf{E} + \frac{4\pi}{c} \mathbf{J}^{\text{ext}} \tag{2.73}$$

Note that the charge and current densities are not known *a priori*; they must be codetermined with the solutions of Maxwell's equations.

2.6 Interaction of Light with the Medium

When we consider the interaction of light with a medium, the equations (2.70)–(2.73) can be simplified. Since there are no external sources, $\rho^{\text{ext}} = 0$, and, for isotropic media, there is no spatial variation in ε. Thus,

$$\nabla \cdot \mathbf{E} = 0 \tag{2.74}$$

$$\nabla \times \mathbf{E} = -\frac{\mu}{c}\frac{\partial \mathbf{H}}{\partial t} \tag{2.75}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{2.76}$$

$$\nabla \times \mathbf{H} = \frac{\varepsilon}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi\sigma}{c}\mathbf{E} \tag{2.77}$$

Using the vector identity

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} \tag{2.78}$$

and Eqs. (2.75) and (2.77), we get the equation

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = -\frac{\varepsilon\mu}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} - \frac{4\pi\sigma\mu}{c^2}\frac{\partial \mathbf{E}}{\partial t} \tag{2.79}$$

Using Eq. (2.74), we then obtain the wave equation for a plane wave propagating in an energy-absorbing medium:

$$\nabla^2\mathbf{E} = \frac{\varepsilon\mu}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} + \frac{4\pi\sigma\mu}{c^2}\frac{\partial \mathbf{E}}{\partial t} \tag{2.80}$$

The solutions are necessarily restricted to transverse plane waves because $\nabla \cdot \mathbf{E} = 0$ in the absence of a net charge density. The conductivity which appears in Eq. (2.80) should be called the optical conductivity. This is because the energy absorption with which we are concerned is that arising from electronic transitions accompanying photon absorption. These transitions correspond to a transverse current density $\mathbf{J}^{\mathrm{T}}$ that does not include the conventional current such as is obtained with a battery connected across the sample. The latter is a longitudinal current; it originates from a longitudinal electric field which is derivable from a scalar potential function. Nonetheless, we shall see later that at sufficiently long wavelengths, in the infrared, the transverse optical conductivity approaches the ordinary dc electrical conductivity for isotropic materials. For anisotropic materials, the optical conductivity and dielectric functions must be treated as tensors. This is done in Section 2.9.

Since experiments on optical properties of solids are usually conducted with monochromatic light, we shall consider first the propagation of a single plane wave within an isotropic medium. Anticipating that the wave vector must be complex to describe energy dissipation of the wave, we write

$$\mathbf{E} = \mathbf{E}_0 \exp i(\hat{\mathbf{q}} \cdot \mathbf{r} - \omega t) \tag{2.81}$$

where $\mathbf{E}_0$ is perpendicular to the wave vector $\hat{\mathbf{q}}$. Note that we may assume a sinusoidal (plane wave) variation of $\mathbf{E}$ only in a region large compared with the lattice constant. This is possible and quite satisfactory for a theory

of the optical properties of solids, but it is not suitable for the x-ray region.

Substituting Eq. (2.81) into (2.80), we find

$$\hat{q}^2 = \mu \frac{\omega^2}{c^2} \left(\varepsilon + i \frac{4\pi\sigma}{\omega} \right) \tag{2.82}$$

We now define a complex refractive index $\hat{n}$ such that

$$\hat{q} = (\omega/c)\,\hat{n} = (\omega/c)\,(n + ik) \tag{2.83}$$

where n is the refractive index and k is the extinction coefficient. Now we can rewrite Eq. (2.81) as

$$\mathbf{E} = \mathbf{E}_0 \left[\exp - \left(\frac{\omega}{c} \mathbf{k} \cdot \mathbf{r} \right) \right] \exp i \left(\frac{\omega}{c} \mathbf{n} \cdot \mathbf{r} - \omega t \right) \tag{2.84}$$

The first exponential factor in Eq. (2.84) describes the attenuation of wave amplitude with distance. The absorption coefficient, which describes the fractional decrease in intensity with distance, is defined as

$$\alpha = -\frac{1}{I} \frac{dI}{dr} \tag{2.85}$$

where I is the intensity. Since the intensity is proportional to the square of the wave amplitude, we find from Eqs. (2.84) and (2.85) that

$$\alpha = 2\omega k/c = 4\pi k/\lambda \tag{2.86}$$

where λ is the wavelength of the light in vacuum.

The second exponential factor is Eq. (2.84) describes a wave traveling with phase velocity c/n, hence the earlier identification of n as the refractive index.

Equations (2.82) and (2.83) can be used to obtain expressions for ε and σ in terms of n and k. Thus,

$$\varepsilon = (n^2 - k^2)/\mu \tag{2.87}$$

$$4\pi\sigma/\omega = 2nk/\mu \tag{2.88}$$

We now define a complex dielectric function ε as

$$\hat{\varepsilon} = \varepsilon_1 + i\varepsilon_2 = \hat{n}^2/\mu \tag{2.89}$$

where ε_1 is the old ε of Eq. (2.87), the ε that appears in the usual versions of Maxwell's equations when the properties of the medium are included. Thus,

$$\varepsilon_1 = (n^2 - k^2)/\mu \tag{2.90}$$

$$\varepsilon_2 = 2nk/\mu = 4\pi\sigma/\omega \tag{2.91}$$

Equations (2.90) and (2.91) show that ε_1 and ε_2 are not independent quantities. We shall see later that ε_1 and ε_2, as well as n and k, are related in a quite fundamental way by means of the Kramers–Kronig dispersion relations.

2.7 External Sources and Induced Responses

We have written Maxwell's equations in a form and notation corresponding to common usage in many applications of electromagnetic theory. However, Maxwell's equations can be written using a notation that more clearly emphasizes the physics involved in studying the optical properties of solids. For example, a light wave incident on a sample can be described by its electric field. This electric field is an external field that acts as a probe. It induces an electric field in the sample. The total electric field, which is the sum of the external field and the induced field, is the electric field $\mathbf{E}$ that appears in the usual version of Maxwell's equations; however, the field that is controlled by the experimenter is the external field. Furthermore, it is the vector potential related to the external electric field that enters in the Hamiltonian describing the interaction of light with the medium; that is, it is the external field that acts as a perturbation on the system. With the use of perturbation theory, we can then calculate the induced current and charge densities and the induced fields.

We now want to see how external and total fields as well as external and total charge densities and currents are related. We will then rewrite Maxwell's equations so as to emphasize this viewpoint.

The presence of an external field induces a field in the medium, that is, it polarizes the medium. Thus

$$\mathbf{E}^{\text{total}} = \mathbf{E}^{\text{ext}} + \mathbf{E}^{\text{pol}} \tag{2.92}$$

where $\mathbf{E}^{\text{total}}$ is the electric field $\mathbf{E}$ that appears in the usual version of Maxwell's equations and $\mathbf{E}^{\text{ext}}$ is the displacement that appears in Eq. (2.59). That is,

$$\mathbf{E} = \mathbf{E}^{\text{total}}, \qquad \mathbf{D} = \mathbf{E}^{\text{ext}} \tag{2.93}$$

The presence of an induced field in the medium results in an induced charge density if the divergence of the induced field is nonzero. This induced charge density is the same as the polarization or bound charge density that we talked of earlier in discussing the polarization of matter.

If there is an induced charge density in the medium, it means that charge has been depleted from some regions and added to others; there is no net extra charge induced. Any net extra charge density must be supplied externally, by an electron beam incident on the sample, for example. Such a beam also supplies an external current.

The various relationships among fields, among currents, and among charge densities are illustrated in Table 2.1. We can use these relationships to rewrite Maxwell's equations (2.70)–(2.73) as

$$\nabla \cdot \mathbf{E}^{\text{ext}} = 4\pi\rho^{\text{ext}} \tag{2.94}$$

$$\nabla \times \mathbf{E}_{\text{total}} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{2.95}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.96}$$

$$\nabla \times \mathbf{B} = \frac{\mu\varepsilon}{c}\frac{\partial \mathbf{E}^{\text{total}}}{\partial t} + \frac{4\pi\sigma\mu}{c}\mathbf{E}^{\text{total}} + \frac{4\pi\mu}{c}\mathbf{J}^{\text{ext}} \tag{2.97}$$

Similar equations can be written to express relationships connecting external, total, and induced magnetic field. This has not been done here because we are considering only nonmagnetic materials. We have carried the magnetic permeability μ along so that the extension to magnetic effects can be made more easily, and for conceptual clarity. The latter is made more explicit in Section 2.8, where the relationships between μ and the longitudinal and transverse dielectric functions are discussed.

TABLE 2.1 *Relationships among Fields, among Currents, and among Charge Densities*

$$\mathbf{E} = \mathbf{E}^{\text{total}} = \begin{cases} \mathbf{D} & \rightarrow \mathbf{E}^{\text{ext}} \\ + & \\ -4\pi\mathbf{P} & \rightarrow \mathbf{E}^{\text{ind}} \end{cases}$$

$$\mathbf{J} = \mathbf{J}^{\text{total}} = \begin{cases} \mathbf{J}^{\text{bound}} = \begin{cases} \mathbf{J}^{\text{pol}} \\ + \\ \mathbf{J}^{\text{mag}} \end{cases} = \mathbf{J}^{\text{ind}} \\ + \\ \mathbf{J}^{\text{free}} = \begin{cases} \mathbf{J}^{\text{cond}} \\ + \\ \mathbf{J}^{\text{ext}} \rightarrow \mathbf{J}^{\text{ext}} \end{cases} \end{cases}$$

$$\rho = \rho^{\text{total}} = \begin{cases} \rho^{\text{bound}} \rightarrow \rho^{\text{pol}} \rightarrow \rho^{\text{ind}} \\ + \\ \cdots \quad \rightarrow \rho^{\text{ext}} \end{cases}$$

2.8 Fourier Analysis of Maxwell's Equations

We have been working mostly in terms of space and time variables, $\mathbf{r}$ and t. However, in Section 2.6, we described the electric field in terms of a single Fourier component, i.e., a single monochromatic wave. The result was that the dielectric function, as it appears in Eq. (2.82), for example, is dependent on a single ω and $\mathbf{q}$; the dielectric function $\varepsilon(\mathbf{q}, \omega)$ in Eq. (2.82) is just the Fourier transform of the dielectric response function introduced in Eq. (2.69). Usually, we shall be concerned only with the spectral decomposition of ε that results from temporal dispersion; occasionally, we shall be concerned with the dependence of ε on $\mathbf{q}$ that results from spatial dispersion. In general, it is simplest to arrive at these results by first making a Fourier analysis of Maxwell's equations.

We assume that all fields and sources can be decomposed into a complete (continuous) set of plane waves varying as $\exp(i\mathbf{q} \cdot \mathbf{r} - i\omega t)$ for all $\mathbf{q}$ and ω. Thus, for example,

$$\mathbf{E}(\mathbf{r}, t) = \int d\mathbf{q} \int_{-\infty}^{\infty} d\omega\, \mathbf{E}(\mathbf{q}, \omega) \exp(i\mathbf{q} \cdot \mathbf{r} - i\omega t) \tag{2.98}$$

where $\int d\mathbf{q}$ symbolizes $\int dq_x \int dq_y \int dq_z$. The Fourier transform of $\mathbf{E}(\mathbf{r}, t)$ is

$$\mathbf{E}(\mathbf{q}, \omega) = \frac{1}{(2\pi)^4} \int d\mathbf{r} \int_{-\infty}^{\infty} dt\, \mathbf{E}(\mathbf{r}, t) \exp(-i\mathbf{q} \cdot \mathbf{r} + i\omega t) \tag{2.99}$$

where $\int d\mathbf{r}$ symbolizes $\int dx \int dy \int dz$. Equations (2.98) and (2.99) can be shown to be compatible with the use of the relations

$$\int d\mathbf{r} \exp[-i(\mathbf{q} - \mathbf{q}') \cdot \mathbf{r}] = (2\pi)^3 \delta(\mathbf{q} - \mathbf{q}') \tag{2.100}$$

$$\int dt \exp[i(\omega - \omega')t] = 2\pi \delta(\omega - \omega') \tag{2.101}$$

where δ is the Dirac delta function.

Taking the Fourier transform of Maxwell's equations, including the properties of the medium as given by Eqs. (2.94) and (2.97), we obtain

$$i\mathbf{q} \cdot \mathbf{E}^{\text{ext}}(\mathbf{q}, \omega) = 4\pi\rho^{\text{ext}}(q, \omega) \tag{2.102}$$

$$\mathbf{q} \times \mathbf{E}^{\text{total}}(\mathbf{q}, \omega) = (\omega/c)\, \mathbf{B}(\mathbf{q}, \omega) \tag{2.103}$$

$$\mathbf{q} \cdot \mathbf{B}(\mathbf{q}, \omega) = 0 \tag{2.104}$$

$$i\mathbf{q} \times \mathbf{B}(\mathbf{q}, \omega) = -i\frac{\omega\mu\varepsilon}{c} \mathbf{E}^{\text{total}}(\mathbf{q}, \omega) + \frac{4\pi\sigma\mu}{c} \mathbf{E}^{\text{total}}(\mathbf{q}, \omega) + \frac{4\pi\mu}{c} \mathbf{J}^{\text{ext}}(\mathbf{q}, \omega) \tag{2.105}$$

Note that now $\varepsilon = \varepsilon(\mathbf{q}, \omega)$, $\sigma = \sigma(\mathbf{q}, \omega)$, and $\mu = \mu(\mathbf{q}, \omega)$. We have used the shorter notation for simplicity.

The solutions for **B** and **E**, corresponding to Eqs. (2.15) and (2.17), are

$$\mathbf{B}(\mathbf{q}, \omega) = i\mathbf{q} \times \mathbf{A}(\mathbf{q}, \omega) \tag{2.106}$$

$$\mathbf{E}(\mathbf{q}, \omega) = i(\omega/c)\,\mathbf{A}(\mathbf{q}, \omega) - i\mathbf{q}\phi(\mathbf{q}, \omega) \tag{2.107}$$

where $\mathbf{E}(\mathbf{q}, \omega) = \mathbf{E}^{\text{total}}(\mathbf{q}, \omega)$ and $\mathbf{A}(\mathbf{q}, \omega) = \mathbf{A}^{\text{total}}(\mathbf{q}, \omega)$, and the potential is given by the Fourier transform of Eq. (2.21) as

$$q^2\phi(\mathbf{q}, \omega) = 4\pi\rho^{\text{total}}(\mathbf{q}, \omega) \tag{2.108}$$

Note that whenever fields or sources are not identified as external, induced, or total, they are meant to be total fields or sources.

Equation (2.105) can be separated into two equations corresponding to the transverse and longitudinal parts of the fields and current densities.

It follows from Eq. (2.104) that $\mathbf{B}^{\text{L}}(\mathbf{q}, \omega) = 0$. In the Coulomb gauge, $\mathbf{A}^{\text{L}}(\mathbf{q}, \omega) = 0$. We thus need only the potential function $\phi(\mathbf{q}, \omega)$ to determine the longitudinal fields and current densities. Equation (2.107) thus yields

$$\mathbf{E}^{\text{L}}(\mathbf{q}, \omega) = -i\mathbf{q}\phi(\mathbf{q}, \omega) \tag{2.109}$$

and the longitudinal part of Eq. (2.105) yields

$$0 = -i\omega\left(\varepsilon + i\frac{4\pi\sigma}{\omega}\right)\mathbf{E}^{\text{L}}(\mathbf{q}, \omega) + 4\pi\mathbf{J}^{\text{L}}_{\text{ext}}(\mathbf{q}, \omega) \tag{2.110}$$

Realize, though, that Eq. (2.110) represents a special case! We are still considering isotropic media so that **E** and **J** are parallel. In general, there is a coupling between longitudinal and transverse fields and currents. These aspects are considered later in this chapter.

Introducing the complex dielectric function from Eqs. (2.89)–(2.91), and expressing the external current density as the difference between the total and induced current densities, yields

$$4\pi\mathbf{J}^{\text{L}}_{\text{ind}}(\mathbf{q}, \omega) = 4\pi\mathbf{J}^{\text{L}}_{\text{total}}(\mathbf{q}, \omega) - i\omega\varepsilon\mathbf{E}^{\text{L}}_{\text{total}}(\mathbf{q}, \omega) \tag{2.111}$$

Using Eq. (2.109) and the Fourier transform of Eq. (2.46),

$$\omega\mathbf{q}\phi(\mathbf{q}, \omega) = 4\pi\mathbf{J}^{\text{L}}_{\text{total}}(\mathbf{q}, \omega) \tag{2.112}$$

Eq. (2.111) becomes

$$\mathbf{J}^{\text{L}}_{\text{ind}}(\mathbf{q}, \omega) = (i\omega/4\pi)\,(1 - \hat{\varepsilon})\,\mathbf{E}^{\text{L}}_{\text{total}}(\mathbf{q}, \omega) \tag{2.113}$$

An equation similar to Eq. (2.113) can be obtained for the transverse parts of the field and current densities. Because $\mathbf{B}^{\text{T}}(\mathbf{q}, \omega) \neq 0$, it is best to first

solve Eq. (2.106) in terms of the vector potential $\mathbf{A}(\mathbf{q}, \omega)$. This gives

$$-\mathbf{q} \times [\mathbf{q} \times \mathbf{A}(\mathbf{q}, \omega)] = -i\frac{\omega\mu}{c}\left(\varepsilon + i\frac{4\pi\sigma}{\omega}\right)\left[i\frac{\omega}{c}\mathbf{A}(\mathbf{q}, \omega)\right] + \frac{4\pi\mu}{c}\mathbf{J}^{\text{ext}}(\mathbf{q}, \omega) \quad (2.114)$$

which, on introducing the complex dielectric function, and taking only the transverse current density, simplifies to

$$\left(\frac{q^2}{\mu} - \frac{\omega^2}{c^2}\hat{\varepsilon}\right)\mathbf{A}(\mathbf{q}, \omega) = \frac{4\pi}{c}\mathbf{J}^{\text{T}}_{\text{ext}}(\mathbf{q}, \omega) \quad (2.115)$$

Once again substituting the difference between the total and induced current densities for the external current density gives

$$\left(\frac{q^2}{\mu} - \frac{\omega^2}{c^2}\hat{\varepsilon}\right)\mathbf{A}(\mathbf{q}, \omega) = \frac{4\pi}{c}[\mathbf{J}^{\text{T}}_{\text{total}}(\mathbf{q}, \omega) - \mathbf{J}^{\text{T}}_{\text{ind}}(\mathbf{q}, \omega)] \quad (2.116)$$

The total current density $\mathbf{J}^{\text{T}}_{\text{total}}(\mathbf{q}, \omega)$ is found from the Fourier transform of Eq. (2.47) to be

$$\frac{4\pi}{c}\mathbf{J}^{\text{T}}_{\text{total}}(\mathbf{q}, \omega) = \left(q^2 - \frac{\omega^2}{c^2}\right)\mathbf{A}(\mathbf{q}, \omega) \quad (2.117)$$

This, together with

$$\mathbf{E}^{\text{T}}(\mathbf{q}, \omega) = i(\omega/c)\,\mathbf{A}(\mathbf{q}, \omega) \quad (2.118)$$

from Eq. (2.107), can be used to rewrite Eq. (2.116) as

$$\left[q^2\left(1 - \frac{1}{\mu}\right) - \frac{\omega^2}{c^2}(1 - \hat{\varepsilon})\right]\mathbf{E}^{\text{T}}(\mathbf{q}, \omega) = i\frac{4\pi\omega}{c^2}\mathbf{J}^{\text{T}}_{\text{ind}}(\mathbf{q}, \omega) \quad (2.119)$$

This is hardly an improvement in terms of simplicity. The only apparent similarity between Eqs. (2.113) and (2.119) is that each provides a relationship between a total electric field and an induced current density.

Before proceeding with a simplification of Eq. (2.119), it is worth noting that the real dielectric function ε that appears in Maxwell's equations is a longitudinal dielectric function. It is only the longitudinal electric field that is determined by external charge densities and for which the relationship between external and total electric fields is given by

$$\mathbf{E}^{\text{L}}_{\text{ext}}(\mathbf{q}, \omega) = \varepsilon\mathbf{E}^{\text{L}}_{\text{total}}(\mathbf{q}, \omega) \quad (2.120)$$

This is the reason for the relative simplicity of Eq. (2.113). It is clear from Eq. (2.120) that the dielectric function appearing in Maxwell's equations is a longitudinal dielectric function. Thus ε and the corresponding complex dielectric function $\hat{\varepsilon}$ will now be designated as ε^{L} and $\hat{\varepsilon}^{\text{L}}$. Then, if a new

dielectric function $\hat{\varepsilon}^{\mathrm{T}}$ is defined by

$$q^2\left(1-\frac{1}{\mu}\right) \equiv \frac{\omega^2}{c^2}(\hat{\varepsilon}^{\mathrm{T}} - \hat{\varepsilon}^{\mathrm{L}}) \tag{2.121}$$

Equation (2.119) can be written as

$$\mathbf{J}^{\mathrm{T}}_{\mathrm{ind}}(\mathbf{q}, \omega) = (i\omega/4\pi)(1 - \hat{\varepsilon}^{\mathrm{T}})\,\mathbf{E}^{\mathrm{T}}(\mathbf{q}, \omega) \tag{2.122}$$

Problem 2.8 shows how one is led to the definition given by Eq. (2.121). It is clear from Eq. (2.121) that unless $1/\mu(\mathbf{q}, \omega)$ has poles of strength ω^2/c^2q^2 or higher, then in the long-wavelength limit ($q \to 0$), the transverse and longitudinal dielectric functions are equal. Physically, this just means that in the long-wavelength limit (for isotropic materials), the medium cannot distinguish between electric fields parallel or perpendicular to $\mathbf{q}$.

The properties of the medium have now been accounted for by the two functions $\hat{\varepsilon}^{\mathrm{L}}(\mathbf{q}, \omega)$ and $\hat{\varepsilon}^{\mathrm{T}}(\mathbf{q}, \omega)$ rather than the functions $\hat{\varepsilon}(\mathbf{q}, \omega) = \hat{\varepsilon}^{\mathrm{L}}(\mathbf{q}, \omega)$ and $\mu(\mathbf{q}, \omega)$. Equations (2.113) and (2.122) are now of the same form and the Fourier components of the induced current densities and total electric fields can be written

$$\mathbf{J}^{\mathrm{T,L}}_{\mathrm{ind}}(\mathbf{q}, \omega) = (i\omega/4\pi)\left[1 - \hat{\varepsilon}^{\mathrm{T,L}}(\mathbf{q}, \omega)\right]\mathbf{E}^{\mathrm{T,L}}(\mathbf{q}, \omega) \tag{2.123}$$

for isotropic media.

The description of a medium in terms of transverse and longitudinal dielectric functions is equivalent to a description in terms of the usual (longitudinal) dielectric function and the magnetic permeability. Such a description is possible because the magnetization enters Maxwell's equations only through the term $\nabla \times \mathbf{M}$. Since it is only another current term, it can be included with the polarization and conduction currents by means of a dielectric function. This is the basis of Problem 2.8.

2.9 The Dielectric Tensor

In the preceding section, we derived two equations (2.123), relating an induced current to a total electric field. One relates the longitudinal induced current to the longitudinal electric field; the other relates the transverse induced current to the transverse electric field. For nonmagnetic isotropic or cubic materials, these are the only possibilities; it is not possible to induce transverse currents with longitudinal fields or longitudinal currents with transverse fields.

In an anisotropic medium, the polarization and induced currents generally lie in a direction different from that of the electric field. It is then possible, e.g., to induce a longitudinal current with a purely transverse electric field. This situation can be handled by representing the dielectric function as a tensor.

The real part of the complex dielectric tensor is symmetric even for an anisotropic medium. Thus, it is always possible to find a set of axes, the principal dielectric axes, such that the real dielectric tensor can be put into diagonal form. The conductivity tensor, which is proportional to the imaginary dielectric tensor, is also symmetric and can be diagonalized. However, the directions of the principal axes of the real dielectric and conductivity tensors are not generally the same; but the two sets of principal axes do coincide for crystals with symmetry at least as high as orthorhombic. Only systems with at least orthorhombic symmetry will be considered here. Thus, we can combine the real dielectric and conductivity tensors in a complex dielectric tensor such that

$$\hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega) = \begin{vmatrix} \varepsilon_{xx}(\mathbf{q}, \omega) & 0 & 0 \\ 0 & \varepsilon_{yy}(\mathbf{q}, \omega) & 0 \\ 0 & 0 & \varepsilon_{zz}(\mathbf{q}, \omega) \end{vmatrix} \tag{2.124}$$

The choice of principal axes is clearly related to the crystal symmetry and a convenient set of axes can be easily chosen in specific cases.

How is the dielectric tensor (2.124) related to the earlier dielectric functions we have used? Starting with Eq. (2.105) and following the general procedures of the preceding section, it is possible to derive an equation of the same form as Eqs. (2.123), namely

$$\mathbf{J}^{\text{ind}}(\mathbf{q}, \omega) = (i\omega/4\pi)\,[\mathbf{1} - \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega]\,\mathbf{E}(\mathbf{q}, \omega) \tag{2.125}$$

It is important to note though, that Eq. (2.125) is not obtained simply by adding Eqs. (2.123). To see the relationship of $\hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega)$ to the earlier dielectric functions, by all means at least read Problem 2.7. Later, we shall derive quantum mechanical expressions for the dielectric tensor. Here, we want to see how to find $\hat{\boldsymbol{\varepsilon}}^{\text{L}}(\mathbf{q}, \omega)$ or $\hat{\boldsymbol{\varepsilon}}^{-1}(\mathbf{q}, \omega)$, given $\hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega)$.

Suppose we are concerned with the longitudinal response of the system to a longitudinal perturbation. We want to know $\hat{\boldsymbol{\varepsilon}}^{\text{L}}(\mathbf{q}, \omega)$.

A longitudinal perturbation causes a longitudinal induced current. We can pick out the longitudinal component of the induced current density by just taking the scalar product of $\mathbf{J}^{\text{ind}}(\mathbf{q}, \omega)$ with the unit vector $\mathbf{q}/q$ in the direction of propagation $\mathbf{q}$. Thus

$$J^{\text{L}}_{\text{ind}}(\mathbf{q}, \omega) = \mathbf{J}^{\text{ind}}(\mathbf{q}, \omega) \cdot \mathbf{q}/q \tag{2.126}$$

where $J^{\text{L}}_{\text{ind}}(\mathbf{q}, \omega)$ is a scalar quantity. We can now make a vector out of $J^{\text{L}}_{\text{ind}}(\mathbf{q}, \omega)$ by just multiplying it by the unit vector. Thus,

$$\mathbf{J}^{\text{L}}_{\text{ind}}(\mathbf{q}, \omega) = \mathbf{J}^{\text{ind}}(\mathbf{q}, \omega) \cdot \mathbf{q}\mathbf{q}/q^2 \tag{2.127}$$

The product of two vectors just sitting side by side, with no scalar product or vector product indicated, is known as a dyad. It is a useful concept

for working in a linear vector space. Later, when working with state vectors labeling a quantum mechanical system, we shall write dyads in Dirac notation as e.g., $|\mathbf{k}\rangle\langle\mathbf{k}|$. Thus, we see that a dyad is simply a projection operator.

In three dimensions, the product of two vectors can be written as

$$\begin{aligned}\mathbf{AB} &= (\mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z)(\mathbf{i}B_x + \mathbf{j}B_y + \mathbf{k}B_z) \\ &= \begin{cases} \mathbf{ii}A_xB_x + \mathbf{ij}A_xB_y + \mathbf{ik}A_xB_z \\ \mathbf{ji}A_yB_x + \mathbf{jj}A_yB_y + \mathbf{jk}A_yB_z \\ \mathbf{ki}A_zB_x + \mathbf{kj}A_zB_y + \mathbf{kk}A_zB_z \end{cases}\end{aligned} \tag{2.128}$$

This product of vectors gives a sum of nine dyads. The combination is known as a dyadic. It is clear from Eq. (2.128) that a dyadic is just a tensor written in a form that emphasizes the vector nature of the tensor rather than its transformation properties.

The dyad which projects out the longitudinal component of the induced current or electric field is now defined to be

$$\mathbf{1}^{\mathrm{L}} = \mathbf{qq}/q^2 \tag{2.129}$$

We can also define a dyad which projects out the transverse components as

$$\mathbf{1}^{\mathrm{T}} = \mathbf{1} - \mathbf{1}^{\mathrm{L}} \tag{2.130}$$

where $\mathbf{1}$ is the unit dyadic.

We can now find $\mathbf{J}^{\mathrm{L}}_{\mathrm{ind}}(\mathbf{q}, \omega)$ by operating on Eq. (2.125) with the dyad $\mathbf{1}^{\mathrm{L}}$. If the electric field is expressed as the sum of its longitudinal and transverse parts, the result is

$$\begin{aligned}\mathbf{J}^{\mathrm{L}}_{\mathrm{ind}}(\mathbf{q}, \omega) &= (i\omega/4\pi)\left[\mathbf{1}^{\mathrm{L}} \cdot \mathbf{1} - \mathbf{1}^{\mathrm{L}} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega)\right] \cdot \mathbf{E}(\mathbf{q}, \omega) \\ &= (i\omega/4\pi)\left[\mathbf{1}^{\mathrm{L}} - \mathbf{1}^{\mathrm{L}} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega]\right] \cdot \left[\mathbf{1}^{\mathrm{L}} \cdot \mathbf{E}(\mathbf{q}, \omega) + \mathbf{1}^{\mathrm{T}} \cdot \mathbf{E}(\mathbf{q}, \omega)\right] \\ &= (i\omega/4\pi)\left[\mathbf{1}^{\mathrm{L}} - \mathbf{1}^{\mathrm{L}} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega) \cdot \mathbf{1}^{\mathrm{L}}\right] \cdot \mathbf{E}(\mathbf{q}, \omega) \\ &\quad - (i\omega/4\pi)\,\mathbf{1}^{\mathrm{L}} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega) \cdot \mathbf{1}^{\mathrm{T}} \cdot \mathbf{E}(\mathbf{q}, \omega)\end{aligned} \tag{2.131}$$

It is clear from Eq. (2.131) that the dielectric tensor describing the longitudinal current induced by a longitudinal field is

$$\hat{\boldsymbol{\varepsilon}}^{\mathrm{L}}(\mathbf{q}, \omega) = \mathbf{1}^{\mathrm{L}} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega) \cdot \mathbf{1}^{\mathrm{L}} \tag{2.132}$$

and that $\mathbf{1}^{\mathrm{L}} \cdot \hat{\boldsymbol{\varepsilon}}(\mathbf{q}, \omega) \cdot \mathbf{1}^{\mathrm{T}}$ describes the longitudinal current induced by a transverse field.

It is helpful to consider several simple specific cases. First, consider a cubic crystal with light impinging at normal incidence and directed along the x axis of the crystal. Light waves are transverse, so we want to find the

transverse dielectric tensor. The transverse dyad is

$$\begin{aligned}\mathbf{1}^{\mathrm{T}} &= \mathbf{1} - \mathbf{1}^{\mathrm{L}} \\ &= \begin{pmatrix} \mathbf{ii} & 0 & 0 \\ 0 & \mathbf{jj} & 0 \\ 0 & 0 & \mathbf{kk} \end{pmatrix} - \begin{pmatrix} \mathbf{ii} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\end{aligned} \tag{2.133}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors along the x, y, and z axes. In matrix notation,

$$\mathbf{1}^{\mathrm{T}} = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \tag{2.134}$$

Thus, the transverse dielectric tensor is found to be

$$\hat{\varepsilon}^{\mathrm{T}}(q_x, \omega) = \begin{vmatrix} 0 & 0 & 0 \\ 0 & \varepsilon^{\mathrm{T}}_{yy}(q_x, \omega) & 0 \\ 0 & 0 & \varepsilon^{\mathrm{T}}_{zz}(q_x, \omega) \end{vmatrix} \tag{2.135}$$

Now, suppose an electron beam is incident upon the crystal and directed along the x axis. The probe is now a longitudinal one, and the appropriate longitudinal dielectric tensor is

$$\begin{aligned}\hat{\varepsilon}^{\mathrm{L}}(q_x, \omega) &= \mathbf{1}^{\mathrm{L}} \cdot \hat{\varepsilon}(q_x, \omega) \cdot \mathbf{1}^{\mathrm{L}} \\ &= \begin{vmatrix} \hat{\varepsilon}^{\mathrm{L}}_{xx}(q_x, \omega) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}\end{aligned} \tag{2.136}$$

For a cubic crystal,

$$\hat{\varepsilon}^{\mathrm{T}}_{yy}(q_x, \omega) = \hat{\varepsilon}^{\mathrm{T}}_{zz}(q_x, \omega) \tag{2.137}$$

The equality holds because of the cubic symmetry of a cubic crystal.

The dielectric function for a longitudinal wave traveling in the x direction is $\hat{\varepsilon}^{\mathrm{L}}_{xx}(q_x, \omega)$. In general,

$$\hat{\varepsilon}^{\mathrm{L}}_{xx}(q_x, \omega) \neq \begin{cases} \hat{\varepsilon}^{\mathrm{T}}_{yy}(q_x, \omega) \\ \hat{\varepsilon}^{\mathrm{T}}_{zz}(q_x, \omega) \end{cases} \tag{2.138}$$

However, in accordance with the discussion following Eq. (2.122), we assert that in the limit of long wavelengths, for cubic materials,

$$\text{cubic:} \qquad \hat{\varepsilon}^{\mathrm{L}}_{xx}(0, \omega) = \begin{cases} \hat{\varepsilon}^{\mathrm{T}}_{yy}(0, \omega) \\ \hat{\varepsilon}^{\mathrm{T}}_{zz}(0, \omega) \end{cases} \tag{2.139}$$

For light waves and (usually) for plasmons, $q \approx 0$. Thus, the same information concerning the dielectric properties of cubic solids can be obtained

from characteristic electron energy loss experiments as from the more usual optical measurements. At frequencies up to the ultraviolet regions of the spectrum, conventional optical measurements are usually simplest; but at higher energies, say 15 eV or more, characteristic energy loss experiments may be much simpler to reliably carry out. For example, the plasma resonance at about 15 eV in aluminum has been studied only by characteristic electron energy loss experiments.

Now consider a uniaxial crystal with optical axis along the x axis. For a transverse wave propagating in the x direction, the appropriate dielectric function is $\hat{\varepsilon}^{\mathrm{T}}_{yy}(q_x, \omega) = \hat{\varepsilon}^{\mathrm{T}}_{zz}(q_x, \omega)$, just as for a cubic material. The dielectric function for a longitudinal wave propagating in the x direction is $\hat{\varepsilon}^{\mathrm{L}}_{xx}(q_x, \omega)$. Now, however, in the limit of long wavelengths, the transverse and longitudinal dielectric functions are not equal, i.e., for noncubic materials,

$$\text{noncubic:} \qquad \hat{\varepsilon}^{\mathrm{L}}_{xx}(0, \omega) \neq \begin{cases} \hat{\varepsilon}^{\mathrm{T}}_{yy}(0, \omega) \\ \hat{\varepsilon}^{\mathrm{T}}_{zz}(0, \omega) \end{cases} \tag{2.140}$$

The reason for the nonequality expressed in Eq. (2.140) is simply that the properties of the medium are in general significantly different along the optical axis compared with a direction perpendicular to the optical axis. Thus, the response of the medium to an electric field along the optical axis is not the same as the response to an electric field perpendicular to the optical axis, even at long wavelengths. Does this mean that it is not possible to use characteristic energy loss experiments to determine the optical properties of anisotropic materials? No! It simply requires, in principle, a different sample orientation to get the same information as would be obtained in an optical experiment. All that need be done is to, say, have the electron beam incident on the sample in a direction parallel to the y axis. Then the electric field is normal to the x axis, just as for a light wave traveling parallel to the x axis. The characteristic energy loss experiment for an electron beam parallel to the y axis determines the dielectric function $\hat{\varepsilon}^{\mathrm{L}}_{yy}(q_y, \omega)$. The optical reflectance for light parallel to the x axis determines $\hat{\varepsilon}^{\mathrm{T}}_{yy}(q_x, \omega) = \hat{\varepsilon}^{\mathrm{T}}_{zz}(q_x, \omega)$. For long wavelengths,

$$\hat{\varepsilon}^{\mathrm{L}}_{yy}(0, \omega) = \begin{cases} \hat{\varepsilon}^{\mathrm{T}}_{yy}(0, \omega) \\ \hat{\varepsilon}^{\mathrm{T}}_{zz}(0, \omega) \end{cases} \tag{2.141}$$

The point is that, to compare longitudinal and transverse dielectric functions in the long-wavelength limit, it is necessary that the polarization of the medium be in the same direction, or an equivalent one, for both modes of excitation. This condition is always satisfied in cubic materials; it can sometimes be satisfied in other materials.

PROBLEMS

2.1 Estimate the maximum velocity of electrons in a free-electron metal having a density typical of real metals. What is a typical maximum velocity for an excited electron when light of wavelength 600 Å is used in optical experiments? How important are relativistic corrections to Maxwell's equations?

2.2 Show that the polarization current does not lead to energy absorption.

2.3 Derive a wave equation for the vector potential **A** in terms of the properties of the complex dielectric function.

2.4 Derive expressions for n and k as functions of ε_1 and ε_2.

2.5 Define a complex conductivity $\hat{\sigma} = \sigma_1 + i\sigma_2$, where σ_1 is the usual real conductivity σ that appears in Maxwell's equations. Maxwell's equations, including the properties of the medium, can now be written in terms of a complex conductivity rather than a complex dielectric function. What is the relationship of $\hat{\sigma}$ to $\hat{\varepsilon}$? What is the relationship of $\mathbf{J}^{\text{ind}}(\mathbf{q}, \omega)$ to $\mathbf{E}(\mathbf{q}, \omega)$ in terms of $\hat{\sigma}(\mathbf{q}, \omega)$?

Define a transverse "optical" conductivity σ^{T}. What is the relationship of σ^{T} to σ?

2.6 What is the ratio of **H** to **E**? Show that whereas the electric and magnetic field energies are equal in vacuum, where $\varepsilon_1 = \mu = 1$, the magnetic field carries most of the energy in a metal.

2.7 With the use of Eqs. (2.112) and (2.117), show that Eq. (2.105) can be written as

$$q^2\mathbf{A} = -\frac{i\omega}{c}\mu\hat{\varepsilon}^{\text{L}}\mathbf{E} + \mu\left(q^2 - \frac{\omega^2}{c^2}\right)\mathbf{A} + \mu\frac{\omega}{c}\mathbf{q}\phi - \frac{4\pi\mu}{c}\mathbf{J}^{\text{ind}}$$

With the use of the defining equation (2.121), show that this can be transformed to

$$\mathbf{J}^{\text{ind}} = \frac{i\omega}{4\pi}(1 - \hat{\varepsilon}^{\text{L}})\,\mathbf{E} + \frac{i\omega}{4\pi}(\hat{\varepsilon}^{\text{L}} - \hat{\varepsilon}^{\text{T}})\,\mathbf{E}^{\text{T}}$$

By analogy with this last result, a more general dielectric tensor [the one appearing in Eq. (2.125)] can be defined as

$$\hat{\varepsilon}(\mathbf{q}, \omega) = \hat{\varepsilon}^{\text{L}}(\mathbf{q}, \omega) + [\hat{\varepsilon}^{\text{T}}(\mathbf{q}, \omega) - \hat{\varepsilon}^{\text{L}}(\mathbf{q}, \omega)] \cdot \mathbf{1}^{\text{T}}$$

so that

$$\mathbf{J}^{\text{ind}}(\mathbf{q}, \omega) = (i\omega/4\pi)\,[\mathbf{1} - \hat{\varepsilon}(\mathbf{q}, \omega)] \cdot \mathbf{E}(\mathbf{q}, \omega)$$

Show that

$$\hat{\varepsilon}^{L}(\mathbf{q}, \omega) = \mathbf{1}^{L} \cdot \hat{\varepsilon}(\mathbf{q}, \omega) \cdot \mathbf{1}^{L}$$

$$\hat{\varepsilon}^{T}(\mathbf{q}, \omega) = \mathbf{1}^{T} \cdot \hat{\varepsilon}(\mathbf{q}, \omega) \cdot \mathbf{1}^{T}$$

Show that the dielectric tensor describing the longitudinal current induced by a transverse field is

$$\mathbf{1}^{L} \cdot \hat{\varepsilon} \cdot \mathbf{1}^{T} = \mathbf{1}^{L} \cdot \hat{\varepsilon}^{T} \cdot \mathbf{1}^{T}$$

and that the dielectric tensor describing the transverse current induced by a longitudinal field is

$$\mathbf{1}^{T} \cdot \hat{\varepsilon} \cdot \mathbf{1}^{L} = \mathbf{1}^{T} \cdot \hat{\varepsilon}^{L} \cdot \mathbf{1}^{L}$$

2.8 From the definitions $\mathbf{B} = \mu\mathbf{H}$ and $\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M}$, show that

$$\mathbf{q} \times 4\pi\mathbf{M} = \mathbf{q} \times \left(1 - \frac{1}{\mu}\right)\mathbf{B}$$

Use this result to show that Eq. (2.105) can be written as

$$i\mathbf{q} \times \mathbf{B} = -i\frac{\omega}{c}\left[\frac{c^2q^2}{\omega^2}\left(1 - \frac{1}{\mu}\right) + \hat{\varepsilon}^{L}\right]\mathbf{E} + i\frac{cq^2}{\omega}\left(1 - \frac{1}{\mu}\right)\mathbf{E}^{L} + \mathbf{J}^{ext}$$

With the aid of the scalar and vector potentials, show that this can be transformed to

$$\mathbf{J}^{ind} = \frac{i\omega}{4\pi}(1 - \hat{\varepsilon}^{L})\,\mathbf{E}^{L} + \frac{i\omega}{4\pi}\left[1 - \hat{\varepsilon}^{L} - \frac{c^2q^2}{\omega^2}\left(1 - \frac{1}{\mu}\right)\right]\mathbf{E}^{T}$$

This result suggests the definition of a transverse dielectric function as given in Eq. (2.121).

2.9 Consider a uniaxial crystal with the optical axis taken as the z direction. Determine $\mathbf{1}^{L} \cdot \hat{\varepsilon}(\mathbf{q}, \omega) \cdot \mathbf{1}^{T}$ for a light wave propagating in a direction 45° from the optical axis. Is the coupling between longitudinal modes and transverse driving forces (and vice versa) likely to be strong? When is the coupling zero? Does it depend on the direction of propagation?

FURTHER READING

R. P. Feynman, R. B. Leighton, and M. Sands, "The Feynman Lectures on Physics," Addison Wesley, Reading, Massachusetts, 1964.

J. D. Jackson, "Classical Electrodynamics," Wiley, New York, 1962.

J. Lindhard, On the Properties of a Gas of Charged Particles, *Dan. Mat. Fys. Medd.* **28**, No. 8, 1–57 (1954).

H. Ehrenreich, *in* "The Optical Properties of Solids" (J. Tauc, ed.), Academic Press, New York, 1966.

F. Stern, Elementary Optical Properties of Solids, *Solid State Phys.* **15**, (1963).

L.D. Landau and E. M. Lifshitz, "Electrodynamics of Continuous Media," Pergamon Press, Oxford, 1960.

Chapter 3

ABSORPTION AND DISPERSION

This chapter consists mostly of a rather elementary treatment of absorption and dispersion. It includes some simple examples of applications to optical properties and photoemission.

The classical theory of absorption and dispersion is due mainly to Lorentz and Drude. The Lorentz model is applicable to insulators; its quantum mechanical analog includes all direct interband transitions; i.e., all transitions for which the final state of an electron lies in a different band but with no change in **k**-vector in the reduced zone scheme. The Drude model is applicable to free-electron metals; its quantum mechanical analog includes intraband transitions, where intraband transitions are taken to mean all transitions not involving a reciprocal lattice vector.

Both the Lorentz and Drude models are largely *ad hoc*, but still useful as starting points and for developing a feeling for optical properties. We shall see that many features of these classical models have quantum mechanical counterparts which are easily understood as generalizations of their classical analogs.

3.1 The Lorentz Oscillator

Consider an atom with electrons bound to the nucleus in much the same way as a small mass can be bound to a large mass by a spring. This is the Lorentz model. The motion of an electron bound to the nucleus is then

described by

$$m\frac{d^2\mathbf{r}}{dt^2} + m\Gamma\frac{d\mathbf{r}}{dt} + m\omega_0{}^2\mathbf{r} = -e\mathbf{E}_{\text{loc}} \tag{3.1}$$

where m is the electronic mass and e is the magnitude of electronic charge. The field $\mathbf{E}_{\text{loc}}$ is the local electric field acting on the electron as a driving force. It is a microscopic field but is written as $\mathbf{E}_{\text{loc}}$ to eliminate confusion with the electronic charge $-e$ and to conform with common usage. The term $m\Gamma(d\mathbf{r}/dt)$ represents viscous damping and provides for an energy loss mechanism. The actual loss mechanism is radiation damping for a free atom, but it arises from various scattering mechanisms in a solid. The damping term in Eq. (3.1) is written in the form in which it often appears in describing the electrical conductivity metals. The term $m\omega_0{}^2\mathbf{r}$ is a Hooke's law restoring force.

In the context of a classical model, there are two approximations in Eq. (3.1). The nucleus has been assumed to have infinite mass, otherwise the reduced mass should have been used. We could have simply included the reduced mass, but our goal is to understand solids and there we can quite accurately take the mass of the lattice as infinite. We have also neglected the small force $-e\mathbf{v} \times \mathbf{b}/c$ arising from the interaction of the electron with the magnetic field of the light wave. It is negligible because the velocity of the electron is small compared with c.

The local field can be taken to vary in time as $e^{-i\omega t}$; thus the solution to Eq. (3.1) is

$$\hat{\mathbf{r}} = \frac{-e\mathbf{E}_{\text{loc}}/m}{(\omega_0{}^2 - \omega^2) - i\Gamma\omega} \tag{3.2}$$

and the induced dipole moment is

$$\hat{\mathbf{p}} = \frac{e^2\mathbf{E}_{\text{loc}}}{m}\frac{1}{(\omega_0{}^2 - \omega^2) - i\Gamma\omega} \tag{3.3}$$

Note that it is important to be consistent in the form of the time variation used to describe time-dependent fields. The use of a time variation $e^{i\omega t}$ leads to a complex refractive index $\hat{n} = n - ik$ rather than the convention $\hat{n} = n + ik$ chosen earlier.

We now assume that the displacement r is sufficiently small that a linear relationship exist between $\hat{\mathbf{p}}$ and $\mathbf{E}_{\text{loc}}$, namely

$$\hat{\mathbf{p}} = \hat{\alpha}(\omega)\mathbf{E}_{\text{loc}} \tag{3.4}$$

where $\hat{\alpha}(\omega)$ is the frequency-dependent atomic polarizability. From Eqs.

(3.3) and (3.4), the polarizability for a one-electron atom is seen to be

$$\hat{\alpha}(\omega) = \frac{e^2}{m} \frac{1}{(\omega_0{}^2 - \omega^2) - i\Gamma\omega} \tag{3.5}$$

The polarizability is complex because of the inclusion of a damping term. As a result, the polarization differs in phase from the local field at all frequencies.

If there are N atoms per unit volume, the macroscopic polarization is

$$\mathbf{P} = N\langle \mathbf{p} \rangle = N\hat{\alpha}\langle \mathbf{E}_{\text{loc}} \rangle = \chi_e \mathbf{E} \tag{3.6}$$

To relate the microscopic atomic polarizability to the macroscopic electric susceptibility, it is necessary to know the relationship between the microscopic field $\mathbf{E}_{\text{loc}}$ and the macroscopic field $\mathbf{E}$. Except for some limiting ideal cases, this is a problem of considerable complexity. It is discussed briefly in Appendix B. In general, $\langle \mathbf{E}_{\text{loc}} \rangle \neq \mathbf{E}$ since $\langle \mathbf{E}_{\text{loc}} \rangle$ is usually an average over atomic sites, not over regions between sites. For free-electron metals, though, we can argue that since the conduction electrons are not bound, the field felt by the conduction electrons is on the average just the macroscopic field $\mathbf{E}$. Then, of course, we should let $\omega_0 = 0$ in Eq. (3.1) because the conduction electrons are not bound. The result is just the Drude model for metals. However, what we shall do is something in between. We will keep the restoring force term, but still assume for simplicity that $\langle \mathbf{E}_{\text{loc}} \rangle = \mathbf{E}$. Such a model contains all the essential features to describe the optical properties; but it must be remembered that in the detailed analysis of specific real solids, it is necessary to consider carefully what is the correct field to use. Proceeding with our assumptions, then, we have

$$\hat{\mathbf{P}} = N\hat{\alpha}\mathbf{E} = \chi_e \mathbf{E} \tag{3.7}$$

We are now ready to get an expression for the dielectric function in terms of the atomic polarizability. But we now have an energy loss mechanism explicitly included with the result that the atomic polarizability is now complex. This means also that the fields $\mathbf{E}$, $\mathbf{P}$, and $\mathbf{D}$ are not in phase. The most convenient way to handle the situation is to generalize some earlier results. In analogy with Eq. (2.67), we define a complex displacement $\hat{\mathbf{D}}$ such that

$$\hat{\mathbf{D}} = \hat{\varepsilon}\mathbf{E} = \mathbf{E} + 4\pi\hat{\mathbf{P}} = \hat{\mathbf{E}}^{\text{ext}} \tag{3.8}$$

This is equivalent to defining $\hat{\mathbf{D}}$ as

$$\hat{\mathbf{D}} = \mathbf{D} + i(4\pi/\omega)\,\mathbf{J} \tag{3.9}$$

The physical quantities $\mathbf{E}$, $\mathbf{D}$, $\mathbf{J}$, etc. are generally written in complex nota-

tion as, e.g.,

$$\mathbf{D} = \mathbf{D}_0 \exp i(\mathbf{q} \cdot \mathbf{r} - \omega t) \tag{3.10}$$

because this notation explicitly shows the phase, in addition to greatly simplifying the mathematical manipulations. Values for these physical quantities are obtained by taking the real part of the complex expressions used for these quantities. Although $\hat{\mathbf{D}}$ can also be written in complex notation, the values for the physical quantities that $\hat{\mathbf{D}}$ represents are not obtained by taking the real part of $\hat{\mathbf{D}}$. The quantity $\hat{\mathbf{D}}$ is truly a complex quantity and represents the two real quantities $\mathbf{D}$ and $\mathbf{J}$. The true values for $\hat{\mathbf{D}}$ must be obtained from the right-hand side of Eq. (3.9) by taking the real parts of $\mathbf{D}$ and $\mathbf{J}$, i.e.,

$$\hat{\mathbf{D}}(\text{true}) = \text{Re}(\hat{\mathbf{D}}) + i(4\pi/\omega)\,\text{Re}(\mathbf{J}) \tag{3.11}$$

Having recognized that there is a truly complex $\hat{\mathbf{D}}$, we shall from here on generally follow convention and write simply $\mathbf{D}$. We shall explicitly designate complex quantities only for properties of the medium, e.g., the complex dielectric function $\hat{\varepsilon}$ and the complex polarizability $\hat{\alpha}$.

Now, from Eqs. (3.7) and (3.8), we get

$$\hat{\varepsilon} = 1 + 4\pi N\hat{\alpha} \tag{3.12}$$

Using Eq. (3.5), this becomes

$$\hat{\varepsilon} = 1 + \frac{4\pi Ne^2}{m} \frac{1}{(\omega_0{}^2 - \omega^2) - i\Gamma\omega} \tag{3.13}$$

From Eq. (3.13) and the definitions is Eqs. (2.89)–(2.91), we get, for non-magnetic materials,

$$\varepsilon_1 = n^2 - k^2 = 1 + \frac{4\pi Ne^2}{m} \frac{(\omega_0{}^2 - \omega^2)}{(\omega_0{}^2 - \omega^2)^2 + \Gamma^2\omega^2} \tag{3.14}$$

$$\varepsilon_2 = 2nk = \frac{4\pi Ne^2}{m} \frac{\Gamma\omega}{(\omega_0{}^2 - \omega^2)^2 + \Gamma^2\omega^2} \tag{3.15}$$

If we consider classical atoms with more than one electron per atom, we can extend the previous results. Let N_j be the density of electrons bound with resonance frequency ω_j. Then,

$$\hat{\varepsilon} = 1 + \frac{4\pi e^2}{m} \sum_j \frac{N_j}{(\omega_j{}^2 - \omega^2) - i\Gamma_j\omega} \tag{3.16}$$

$$\sum_j N_j = N \tag{3.17}$$

We shall shortly derive a corresponding quantum mechanical equation which can be written

$$\hat{\varepsilon} = 1 + \frac{4\pi e^2}{m} \sum_j \frac{Nf_j}{(\omega_j^2 - \omega^2) - i\Gamma_j \omega} \tag{3.18}$$

There is a formal similarity between Eqs. (3.16) and (3.18), but the meanings of some corresponding terms are quite different. In Eq. (3.16), ω_j is the resonance frequency of a bound electron, whereas in Eq. (3.18), it is the transition frequency of an electron between two atomic states separated in energy by $\hbar\omega_j$. The parameter f_j, called the oscillator strength, is a measure of the relative probability of a quantum mechanical transition. We shall show that for free atoms, it satisfies a sum rule

$$\sum_j f_j = 1 \tag{3.19}$$

which is the quantum mechanical analogy to Eq. (3.17).

Now, return to Eqs. (3.14) and (3.15) and consider the frequency dependence of ε_1 and ε_2 for a solid made of a collection of single-electron classical atoms. The frequency dependence is illustrated graphically in Fig. 3.1.

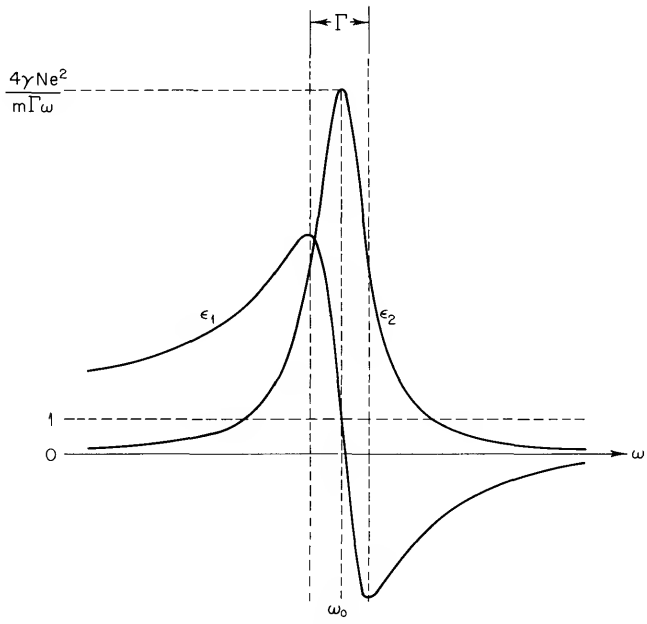

Fig. 3.1 Frequency dependence of ε_1 and ε_2.

Figure 3.1 shows that except for a narrow region near ω_0, ε_1 increases with increasing frequency. This is called normal dispersion. However, there is a region near ω_0 where ε_1 decreases with increasing frequency. This is called anomalous dispersion. We can find the width of the region of anomalous dispersion as follows. Equating the derivative of Eq. (3.14) to zero, we find

$$(\omega_0^2 - \omega_m^2)^2 = \pm \omega_0^2 \Gamma^2 \tag{3.20}$$

where ω_m is the frequency at which ε_1 is a maximum or a minimum. If the region of anomalous dispersion is reasonably narrow, $\omega_m \approx \omega_0$,

$$(\omega_0 - \omega_m) = \pm \Gamma/2 \tag{3.21}$$

and the full width of the region of anomalous dispersion is Γ. In the absence of an energy loss mechanism, there is a singularity at ω_0.

If

$$\Gamma \approx 0 \tag{3.22}$$

ε_2 versus ω is a bell-shaped curve which is symmetric about ω_0. Small values of Γ compared with ω_0 cause little distortion. From Eq. (3.15), we find that the maximum value of ε_2 is

$$\varepsilon_2(\max) = \frac{4\pi N e^2/m}{\Gamma \omega_0} \tag{3.23}$$

assuming the maximum occurs exactly at ω_0. Also, the full width of the ε_2 curve at half maximum is Γ.

Figure 3.1 shows the contribution of the electronic polarizability to the dielectric constant. There are also other contributions. For example, in ionic crystals, in the infrared region, there is an absorption spectrum and polarization associated with the direct stimulation of vibrational modes of the ions by means of electromagnetic radiation. The Lorentz model also describes that situation.

Figure 3.2 shows the general form of the polarizability to be expected

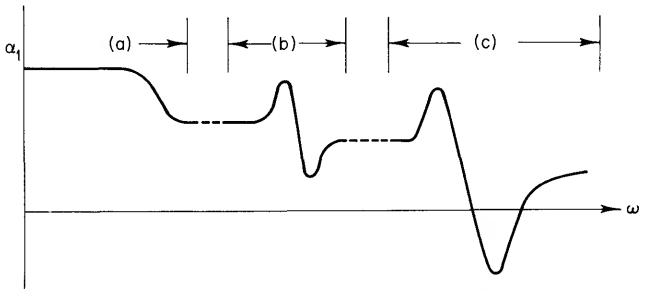

Fig. 3.2 Frequency dependence of contributions to the polarizability arising from orientation of (a) permanent dipoles (microwave), (b) ionic lattice vibrations (infrared), and (c) displacement of electrons (visible and ultraviolet).

in a material consisting of three discrete modes of oscillation. Although all the modes of oscillation contribute to the polarizability and to the dielectric constant, the contributions of ionic motions are small at optical frequencies because of the large inertia of ions compared with electrons. We shall consider only electronic contributions to the dielectric constant. In that context, references to the low-frequency dielectric constant of a material will mean the dielectric constant at the low-frequency end of the visible region but at a frequency high compared with lattice vibrations or molecular oscillations in the crystal.

We now want to consider the implications of the frequency dependence of ε_1 and ε_2 for the optical properties of solids. The reflectivity of solids at normal incidence is shown in Appendix C to be given by

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \tag{3.24}$$

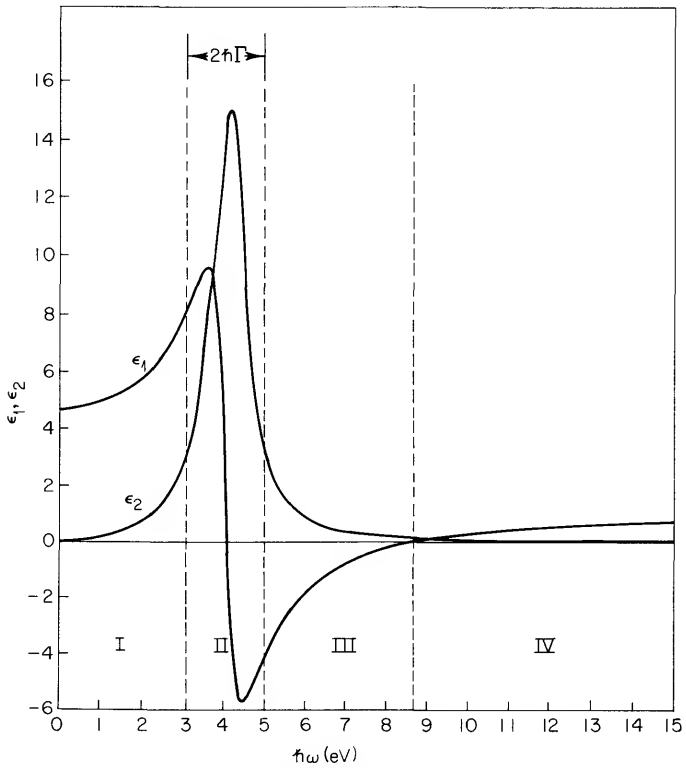

Fig. 3.3 Spectral dependence of ε_1 and ε_1. The curves are calculated for the case in which $\hbar\omega_0 = 4$ eV, $\hbar\Gamma = 1$ eV, and $4\pi Ne^2/m = 60$. The onset of region IV is defined by $\varepsilon_1 = 0$.

Using Eqs. (2.90) and (2.91), we find that for nonmagnetic materials,

$$n = \{\tfrac{1}{2}[(\varepsilon_1{}^2 + \varepsilon_2{}^2)^{1/2} + \varepsilon_1]\}^{1/2} \tag{3.25}$$

$$k = \{\tfrac{1}{2}[(\varepsilon_1{}^2 + \varepsilon_2{}^2)^{1/2} - \varepsilon_1]\}^{1/2} \tag{3.26}$$

Now, from Eqs. (3.14), (3.15), and (3.24)–(3.26), we can analyze the frequency-dependent behavior of a solid in terms of whether it is primarily reflecting, absorbing, or transparent. The results are summarized in Figs. 3.3–3.5.

In region I, $\omega \ll \omega_0$, $\varepsilon_2 = 2nk = 0$, and $\varepsilon_1 = n^2 - k^2 > 1$. We may thus conclude that $k = 0$, $n > 1$, and $\varepsilon_1 = n^2$.

Insulators, such as KCl, typically have a refractive index of about 1.5 in region I. Thus, region I is characterized by high transparency, no absorption, and a small reflectivity for insulators. This is illustrated in Fig.

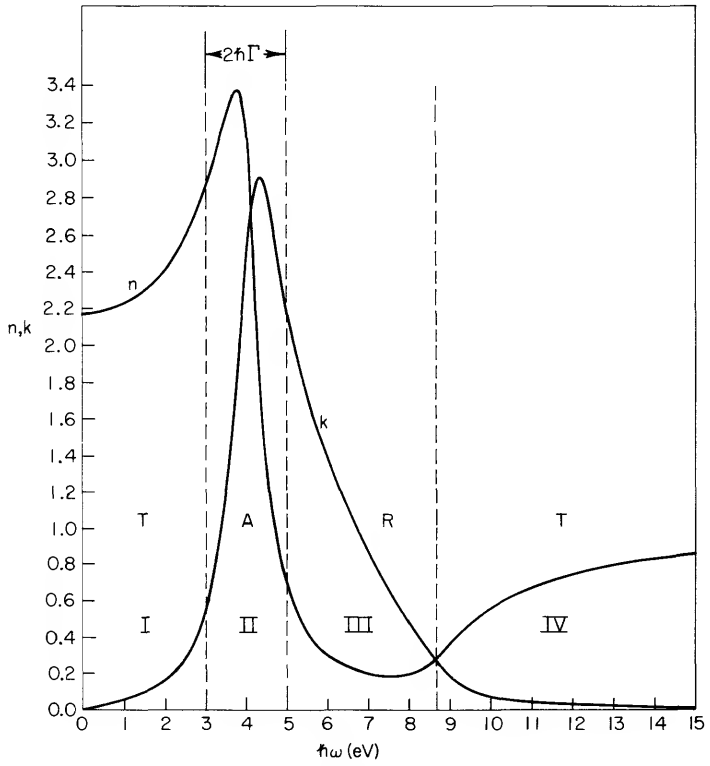

Fig. 3.4 Spectral dependence of n and k. The curves are calculated from the values of ε_1 and ε_2 given in Fig. 3.3. The regions I, II, III, and IV can be seen to be primarily transmitting (T), absorbing (A), reflecting (R), and transmitting (T), respectively. These results follow from consideration of Eq. (3.24) and the realization that strong absorption takes place only in the neighborhood of a transition frequency.

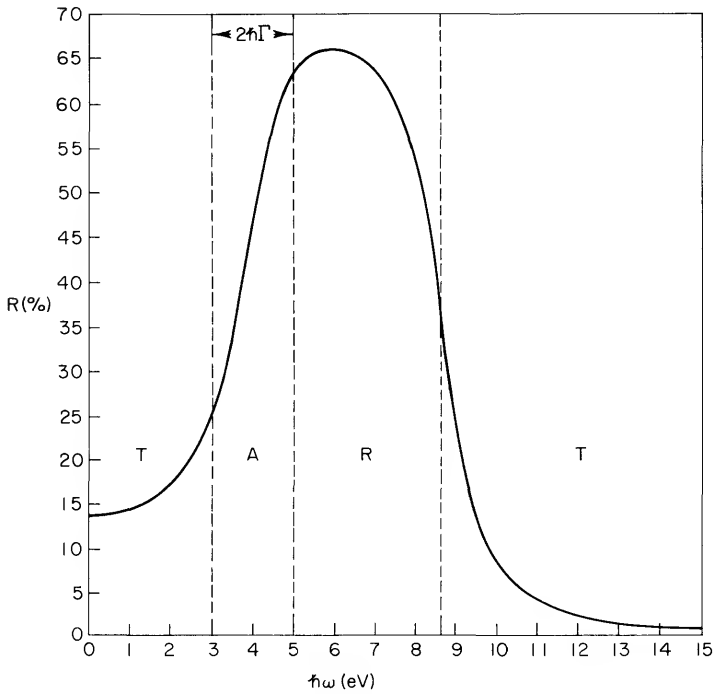

Fig. 3.5 Spectral dependence of reflectivity. The curve is calculated from the n and k values given in Fig. 3.4.

3.6 for the reflectivity of KCl. Of course, since the treatment developed here does not include local field corrections, it is not quantitatively applicable to highly ionic materials. However, for highly polarizable materials such as Si and Ge, there is probably no need to include local field corrections.

The difficulty with applying the present formulas to real materials, even in the absence of local field corrections, is that real materials correspond to a collection of Lorentz oscillators with different frequencies spread out over bands. Nonetheless, if we think of the frequency of a Lorentz oscillator as corresponding to the transition frequency across the band gap of an insulator or semiconductor, we can make some estimates of the optical properties. We can even include approximate band structure effects by using an effective mass rather than the free-electron mass. For example, Fig. 3.7 shows that the reflectivity of Si rises sharply at about 3 eV. This corresponds to a frequency $\omega_0 = 4.5 \times 10^{15}$ sec^{-1}. If we take this as an approximate value for the average spring frequency to be used in Eq. (3.14), and assume four valence electrons per Si atom, each with the mass of a free electron, then $\varepsilon_1(\omega \to 0) = 15$. That is in fairly good agreement with the experimental low-frequency value $\varepsilon_1 = 12$.

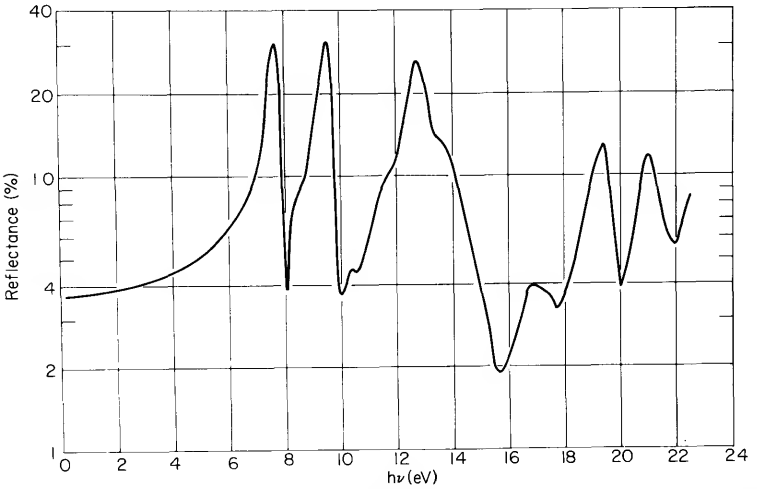

Fig. 3.6 The spectral dependence of the reflectance of KCl. The region of transparency extends to about 7 eV. Above 7 eV, there are a number of sharp peaks related to narrow energy bands and excitons. [From H. R. Philipp and H. Ehrenreich, *Phys. Rev.* **131**, 2016 (1963).]

The refractive index for more highly polarizable materials such as Si and Ge is higher than for ionic insulators. For Si, $n = 3.5$, and for Ge at low frequencies, $n = 4$. As a result, the reflectivity can be appreciable in region I even though there is no absorption. The reflectivity arises from the induced polarization current corresponding to the valence electrons oscillating out of phase with the incident radiation. There is no absorption for this process, but the interference of the incident beam with the waves reradiated by the valence electrons does lead to appreciable reflectivity.

That the Lorentz model is qualitatively correct for semiconductors and insulators is also indicated by the dependence of ε_1 on band gap. Thus, if we identify $\hbar\omega_0$ as corresponding approximately to the band gap, then ε_1 should decrease with increasing band gap. That indeed is the case. The band gaps of Ge, Si and KCl are, respectively, 0.8, 1.1, and 7.5 eV, whereas the low-frequency optical dielectric constants are, respectively, 4, 3.5, and 1.5.

Region II of Figs. 3.3–3.5 is characterized by strong absorption. There may also be appreciable reflectivity in this region. That simply means that although the values of n and k may be high, leading to appreciable reflectivity, the light that is not reflected is strongly absorbed in the material.

In region III, $\omega \gg \omega_0$, and the electrons of the insulator respond as if they were free electrons. This is because the photon energy is much greater than the binding energy of the electron. The insulator thus has a metallic reflectance. Of course, for good insulators, this region lies well into the vacuum-ultraviolet and cannot be observed visually. However, for semi-

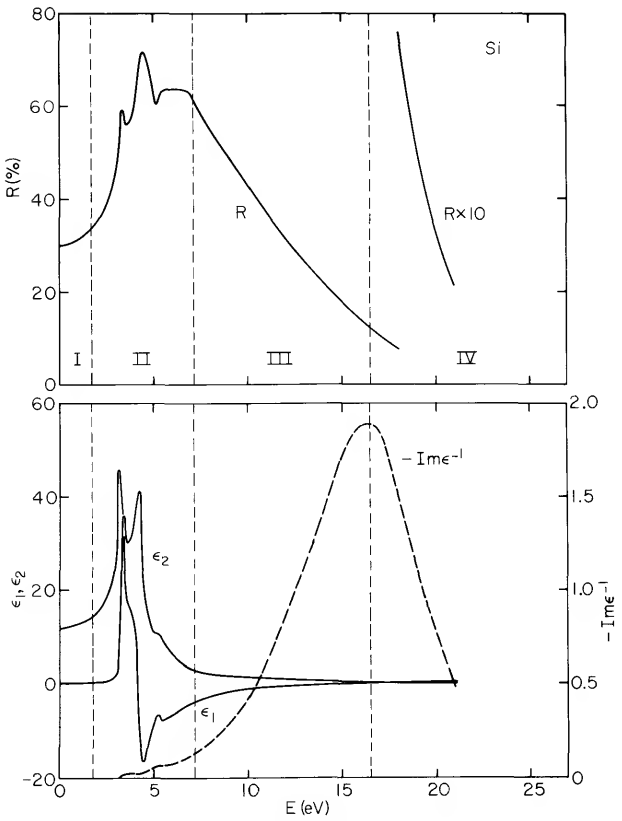

Fig. 3.7 The spectral dependence of the reflectance and dielectric functions of Si. Regions I, II, III, and IV correspond to the regions with the same designation shown in Figs. 3.1, 3.3, and 3.4. [H. R. Philipp and H. Ehrenreich, *Phys. Rev.* **129**. 1550 (1963).]

conductors like Ge and Si, the band gap lies in the infrared and the region of metallic reflectance is in the visible. Thus, KCl is transparent to the eye, but Ge and Si have a metallic appearance.

The onset of region IV is defined by $\varepsilon_1 = 0$. This happens at a frequency ω_p called the plasma frequency. From Eq. (3.14), assuming $\omega \gg \omega_0 \gg \Gamma$, we find

$$\omega_p^2 = 4\pi Ne^2/m \tag{3.27}$$

3.2 The Drude Model for Metals

The Drude model for metals is obtained directly from the Lorentz model for insulators simply by equating the restoring force to zero. The conduction

electrons of a metal are not bound. Furthermore, because the wave function for a free electron is distributed fairly uniformly throughout the metal, the field acting on the electron is just the average field. Thus, there is no need to make corrections for the local field.

From Eqs. (3.14) and (3.15), taking $\omega_0 = 0$, we have

$$\varepsilon_1 = 1 - \frac{4\pi Ne^2}{m} \frac{1}{(\omega^2 + \Gamma^2)} \tag{3.28}$$

$$\varepsilon_2 = \frac{4\pi Ne^2}{m} \frac{\Gamma}{\omega(\omega^2 + \Gamma^2)} \tag{3.29}$$

The origin of the viscous damping term for a free-electron metal is the ordinary scattering of electrons associated with electrical resistivity. In the next chapter, when we derive the properties of a free-electron metal in terms of a complex conductivity rather than a complex dielectric function, we shall see that $\Gamma = \tau^{-1}$, where τ is the mean free time between collisions. If we now make the substitution $\Gamma = \tau^{-1}$, and use Eq. (3.27), we get from Eqs. (3.28) and (3.29)

$$\varepsilon_1 = 1 - \frac{\omega^2\tau^2}{(1 + \omega^2\tau^2)} \tag{3.30}$$

$$\varepsilon_2 = \frac{\omega_p{}^2\tau}{\omega(1 + \omega^2\tau^2)} \tag{3.31}$$

Since the Drude model is obtained directly from the Lorentz model simply by setting ω_0 equal to zero, the optical properties of a free-electron metal should resemble those for an insulator at frequencies greater than ω_0. As we saw in the preceding section, the frequency range $\omega > \omega_0$ in an insulator corresponds to the region in which the electrons are effectively free, so that it might be more accurate to say that an insulator responds like a metal to photons of energy $\hbar\omega > \hbar\omega_0$.

A plot of the dielectric functions and the optical constants for a Drude metal is shown in Figs. 3.8–3.10. The corresponding reflectivity is shown in Fig. 3.11. It is clear from Fig. 3.11 that for an ideal free-electron metal, the reflectivity approaches unity below the plasma frequency. Above the plasma frequency, the metal is transparent and the reflectivity decreases rapidly with increasing frequency. That this describes the behavior of real free-electron metals is shown in Figs. 3.12 and 3.13.

The plasma frequency typically lies in the visible or ultraviolet spectral region. That corresponds to $\omega > 10^{15}$ sec^{-1}. The mean free collision time for electrons in metals is typically $\tau \approx 10^{-14}$ sec. Thus, in the region of the

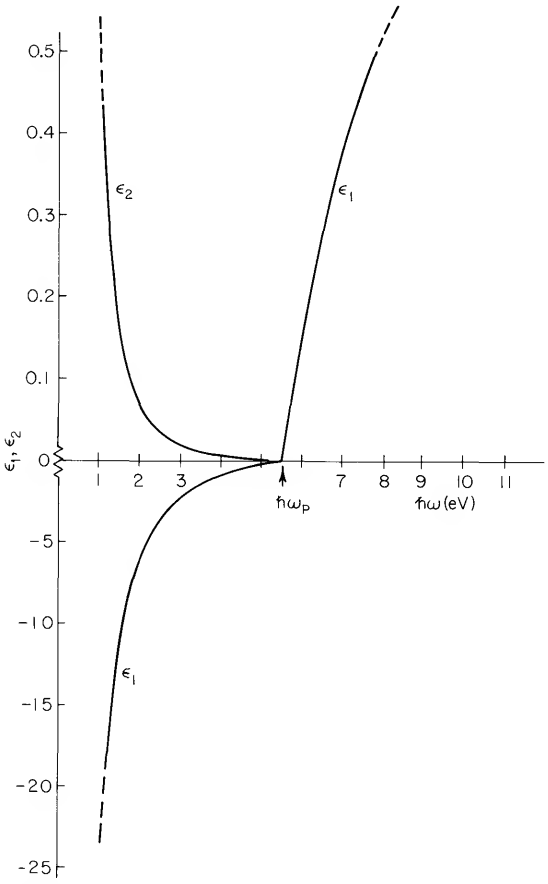

Fig. 3.8 Spectral dependence of ε_1 and ε_2 for a free-electron metal. The calculations are for the case in which $4\pi Ne^2/m = \omega_p{}^2 = 30$ eV^2 and $\hbar\Gamma = 0.02$ eV. Note the difference in scale of the ordinate along the positive and negative axes. The magnitude of ε_1 is much greater than that of ε_2 for the frequency range shown. For $\hbar\omega < \hbar\Gamma$, $|\varepsilon_2/\varepsilon_1| \to \Gamma/\omega$ and ε_2 dominates.

plasma frequency, $\omega\tau \gg 1$, and from Eq. (3.30), we get

$$\varepsilon_1 = n^2 - k^2 = 1 - (\omega_p{}^2/\omega^2) \tag{3.32}$$

From Fig. 3.9, it is clear that $n \gg k$ just above the plasma frequency and so Eq. (3.32) simplifies to

$$n^2 \approx 1 - (\omega_p{}^2/\omega^2) \tag{3.33}$$

for $\hbar\omega > \hbar\omega_p$. Just at the plasma frequency, $n \approx 0$. But what does it mean that the refractive index is zero? The index of refraction is defined in terms of the phase velocity as $v_p = c/n$. Thus, a zero value for n means an infinite

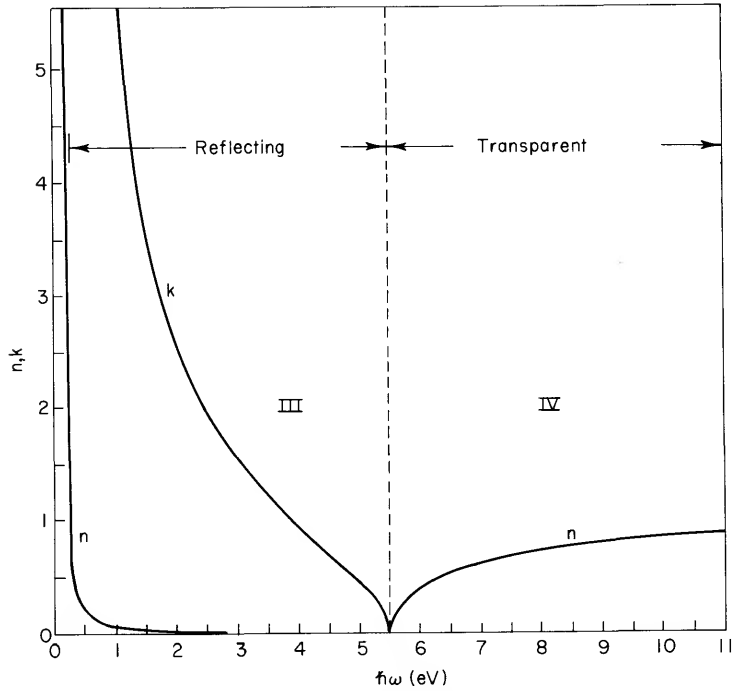

Fig. 3.9 Spectral dependence of n and k for a free-electron metal. The curves are calculated from the values of ε_1 given in Fig. 3.8. Regions III and IV correspond to the same regions as shown in Fig. 3.4. Region II, the region of strong absorption, is the range $0 \leqq \hbar\omega \leqq 0.02$ eV in this case. It is, in general, the range $0 \leqq \hbar\omega \leqq \hbar\Gamma$. Region I does not exist for metals.

phase velocity and an infinite wavelength. That the wavelength becomes infinite means the electrons are all oscillating in phase; however, there is no polarization charge density as with a true plasma oscillation. The distinction is made clear in the next section as well as in Chapter 9.

3.3 A Qualitative Look at Real Metals

We will now use the Lorentz and Drude models in a discussion of the optical behavior of some real metals. Real metals exhibit aspects of both models. To see the role of both models in describing real metals, consider the schematic band diagram for a metal as shown in Fig. 3.14.

Two typical transitions are illustrated in Fig. 3.14. The first of these, called an intraband transition, corresponds to the optical excitation of an electron from below the Fermi energy to another state above the Fermi energy but within the same band. These transitions are described by the

Drude model. There is no threshold energy for such transitions; however, they can occur only in metals. Insulators do not have partially filled bands that would allow excitation of an electron from a filled state below the Fermi energy to an empty state within the same band. That is, of course, what makes an insulator nonconducting.

The second transition illustrated in Fig. 3.14 is a direct interband transition. It is the optical excitation of an electron to another band. It is called a direct or vertical transition because it involves only the excitation of an electron by a photon. Since the momentum of a photon is very small compared with the range of values of crystal momentum in the Brillouin zone, conservation of total crystal momentum of the electron plus photon means that the value of wave vector **k** for the electron is essentially unchanged in the reduced zone scheme. There are nonvertical or indirect transitions between bands, and we shall consider them later; but for present purposes, a discussion of direct transitions is sufficient to illustrate the characteristic optical properties of real metals.

Direct interband transitions have a threshold energy. For the band

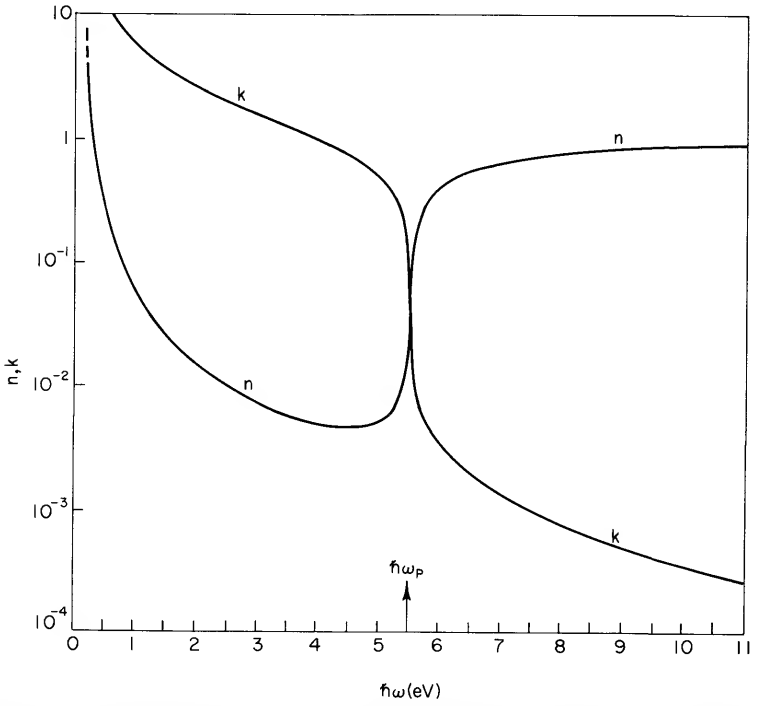

Fig. 3.10 A semilogarithmic plot of the n and k values for a free-electron metal as taken from Fig. 3.9.

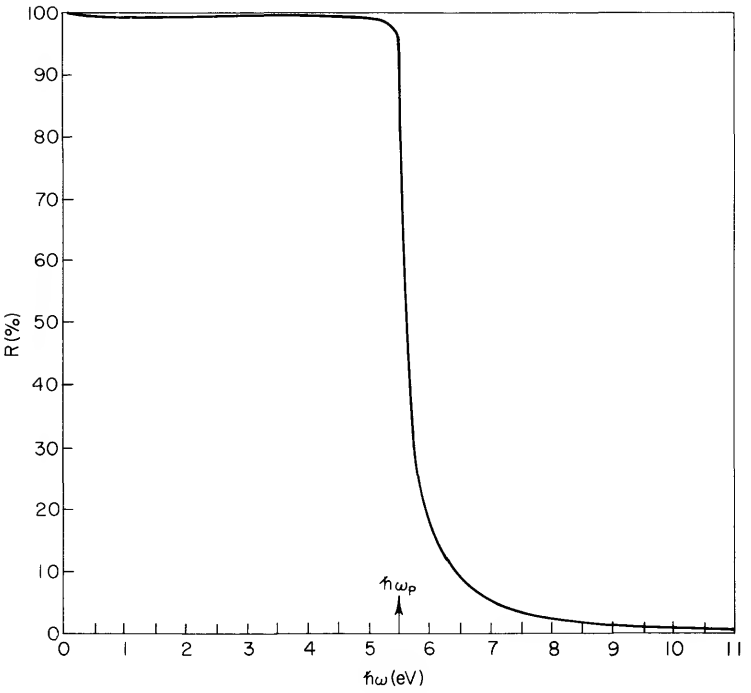

Fig. 3.11 Spectral dependence of reflectivity for a free-electron metal. The curve is calculated from the n and k values given in Fig. 3.9.

diagram shown in Fig. 3.14, the threshold energy is the energy $\hbar\omega_0$ for the transition from the Fermi energy at $\mathbf{k}_0$ to the same state $\mathbf{k}_0$ in the next higher band. This threshold energy is analogous to that for the excitation of an electron across the band gap in an insulator.

Now, how do we use these models to understand real metals, and where do we begin? The beginning is an experimental determination of the reflectance over a wide frequency range. From the reflectance, the dielectric function can be obtained using, e.g., a Kramers–Kronig analysis as discussed in Chapter 6. The dielectric function can then sometimes be split into bound (Lorentz) and free (Drude) contributions and interpreted in terms of the fundamental electronic band structure of the solid. We shall see how this can be done in some cases by following through the steps used by Ehrenreich and Philipp in their classic paper on the optical properties of silver and copper.

The reflectance of silver is shown in Fig. 3.15. A band diagram for the noble metals is shown schematically in Fig. 3.16. The d bands lie several volts below the Fermi energy. Thus, only interband transitions of conduction electrons are possible at low energies, and an onset of interband

transitions from the d bands to higher-lying conduction band states takes place in the visible or ultraviolet. Under these conditions, it is feasible to separate the dielectric function into free and bound parts,

$$\varepsilon_1 = \varepsilon_1^{\,f} + \varepsilon_1^{\,b} \tag{3.34}$$

$$\varepsilon_2 = \varepsilon_2^{\,f} + \varepsilon_2^{\,b} \tag{3.35}$$

where $\varepsilon_1^{\,f}$ and $\varepsilon_2^{\,f}$ are described by Eqs. (3.30) and (3.31) and $\varepsilon_1^{\,b}$ and $\varepsilon_2^{\,b}$ are described by Eqs. (3.14) and (3.15). Band structure effects are included by using an effective mass in these equations rather than using the free-electron mass. The separation depends upon the fact that $\varepsilon_2^{\,f}$ and $\varepsilon_2^{\,b}$ lie in separate spectral regions. Thus, $\varepsilon_2^{\,f}$ can be determined by fitting the data at low frequencies to the Drude model using the relaxation time and effective mass as adjustable parameters. Having determined $\varepsilon_2^{\,f}(\omega)$ in this manner, $\varepsilon_2^{\,b}(\omega)$ is found by subtracting $\varepsilon_2^{\,f}(\omega)$ from the experimentally determined $\varepsilon_2(\omega)$. Then, as shown in Chapter 6, $\varepsilon_1^{\,f}(\omega)$ and $\varepsilon_1^{\,b}(\omega)$ can be determined

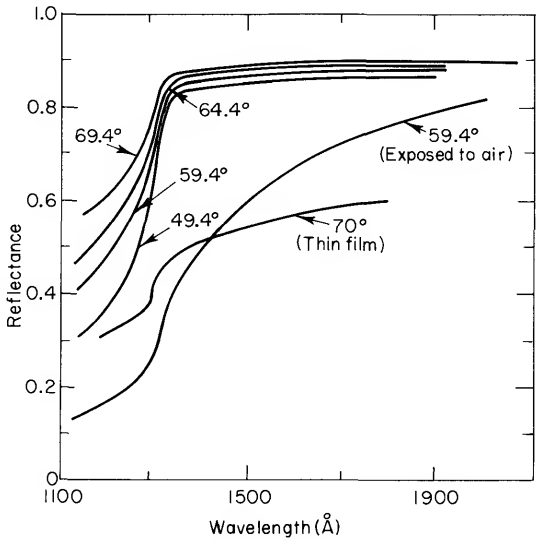

Fig. 3.12 Basal plane reflectance of zinc for nonpolarized light. The angles of incidence are measured with respect to the surface normal. The sample was a single crystal of high-purity zinc cleaved and maintained in a vacuum chamber at pressure less than 10^{-9} Torr. Note the drastic change in reflectance after exposure to air. This illustrates the importance of careful sample preparation if the measurements are to be useful. A comparison with the best earlier work on reflectance from evaporated thin films [T. M. Jelinek, R. N. Hamm, and E. T. Arakawa, ORNL TM-1164 (1965).] is also shown. [From L. P. Mosteller, Jr. and F. Wooten, *Phys. Rev.* **171**, 743 (1968).]

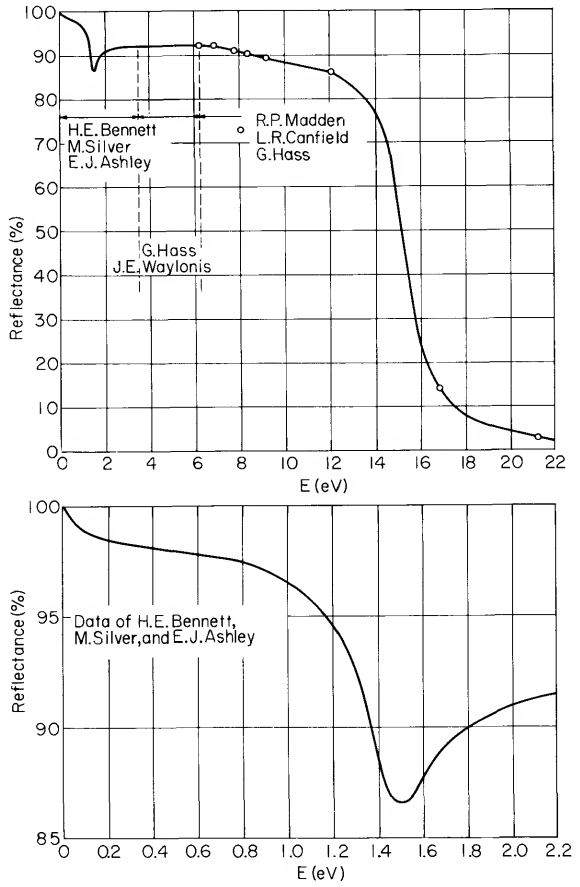

Fig. 3.13 Reflectance of aluminum. The decrease in reflectance at $\hbar\omega = 1.4$ eV arises from a weak interband transition. The large decrease in reflectance at $\hbar\omega = 14.7$ eV identifies the plasma resonance. [From H. Ehrenreich, H. R. Philipp, and B. Segall, *Phys. Rev.* **132**, 1918 (1963).]

from a Kramers–Kronig analysis of $\varepsilon_2^{\,f}(\omega)$ and $\varepsilon_2^{\,b}(\omega)$, respectively. The results of such an analysis for silver are shown in Figs. 3.17 and 3.18.

The reflectance of silver is high at photon energies below 3.9 eV. At 3.9 eV, the reflectance drops sharply, then rises almost immediately. The reflectance well into the vacuum-ultraviolet region is characteristic of interband transitions spread over a wide spectral region.

The reflectance of silver near 3.9 eV is striking, but we can interpret this behavior quite simply. The sharp decrease in reflectance at 3.9 eV identifies the plasma frequency. We previously discussed such behavior for

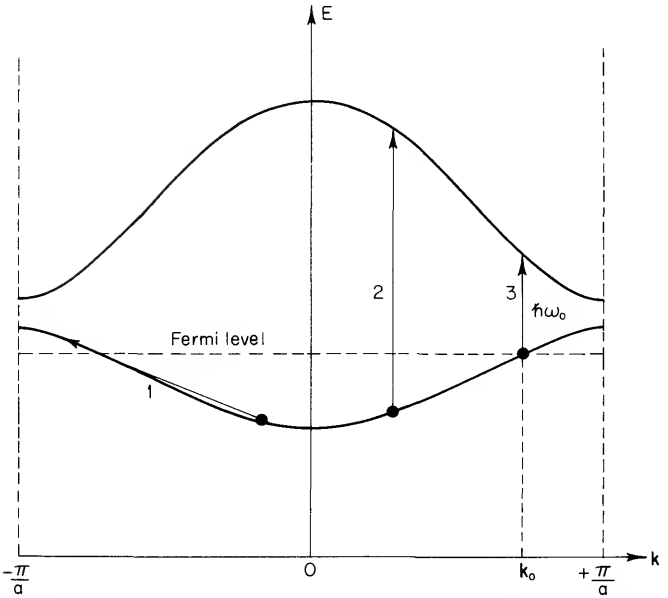

Fig. 3.14 A schematic band diagram for a metal. The threshold for direct transitions is from the Fermi energy at $\mathbf{k}_0$ to the same state $\mathbf{k}_0$ in the next higher band.

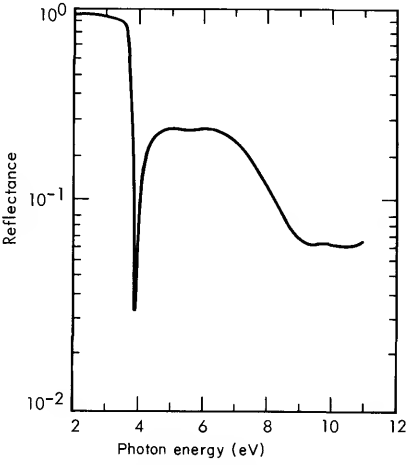

Fig. 3.15 Reflectance of Ag. [From G. B. Irani, T. Huen, and F. Wooten, *Phys. Rev.* **3B**, 2385 (1971).]

an ideal free-electron metal and we have seen such characteristic behavior for the real metals aluminum and zinc. In all these cases, the reflectance remained high at photon energies up to the plasma frequency. Then, it dropped sharply where $\varepsilon_1(\omega)$ passed through zero. The new feature seen

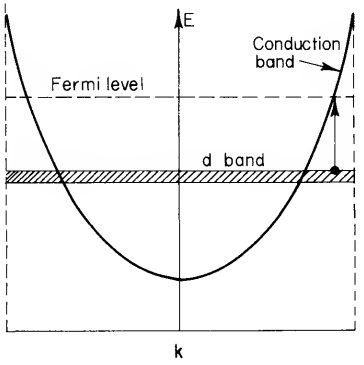

Fig. 3.16 Schematic band diagram for the noble metals.

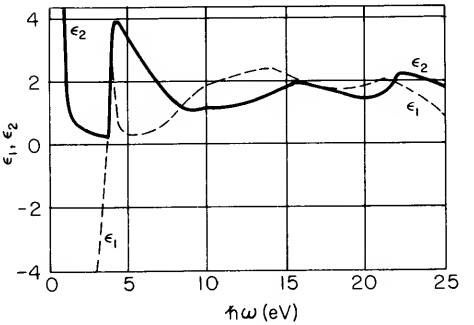

Fig. 3.17 Spectral dependence of the real and imaginary dielectric functions for Ag. [From H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962).]

for silver is a rapid rise in reflectance just above the plasma frequency. To understand that, we must briefly reconsider plasma oscillations.

We shall consider plasma oscillations in detail in Chapter 9. For now, it is sufficient to know that a plasma oscillation is a collective oscillation of the electrons. (See Fig. 3.19.) These oscillations are approximately normal modes of the system if they are only weakly damped. Now, an oscillation is a normal mode if it can maintain itself in the absence of an external field. Thus, we must have

$$\mathbf{E}^{\text{ext}} = \mathbf{D} = \hat{\varepsilon}\mathbf{E} = 0 \tag{3.36}$$

The net charge separation in the electron gas produces the field $\mathbf{E}$ which acts as a restoring force for the oscillations. Since $\mathbf{E} \neq 0$, the condition for plasma oscillations is

$$\hat{\varepsilon} = \varepsilon_1 + i\varepsilon_2 = 0 \tag{3.37}$$

Thus, the plasma frequency is defined by $\varepsilon_1 = 0$ with $\varepsilon_2 \ll 1$.

The response of an electron gas to a transverse electromagnetic wave at the plasma frequency is a collective oscillation in the sense that it involves

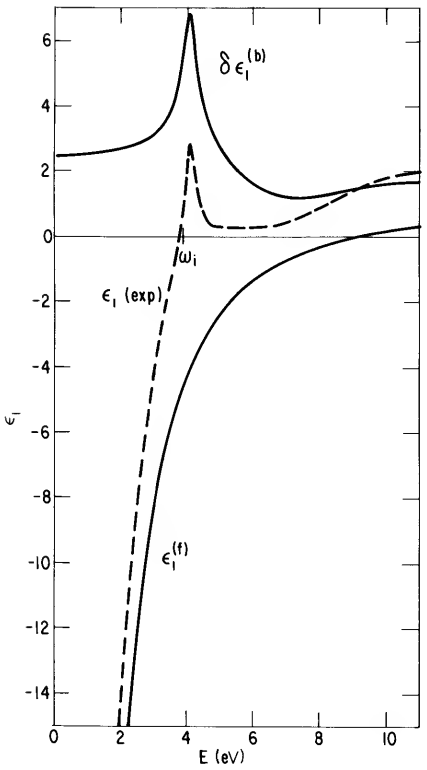

Fig. 3.18 Decomposition of the experimental values of $\varepsilon_1(\omega)$ for Ag into free and bound contributions $\varepsilon_1^{(f)}$ and $\delta\varepsilon_1^{(b)}$. The threshold energy for interband transitions is indicated by ω_i. [From H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962).]

the motion of many electrons moving in phase over an extended region because $\lambda \to \infty$. However, it is a forced oscillation of the system. It exists only in the presence of an external electromagnetic driving force. Thus when one speaks of the plasma resonance in the context of optical studies, it refers to this externally forced collective oscillation, but it is not a true plasma oscillation. The true plasma oscillation is a cooperative collective oscillation that persists even after the external field is removed. This is because, as illustrated in Fig. 3.19, there is a mechanism for sustaining the oscillation.

Now the unusual behavior of Ag is understandable. The sharp decrease in reflectance at 3.9 eV is indeed the characteristic decrease at the plasma frequency. However, the reason $\varepsilon_1(\omega)$ passes through zero at $\hbar\omega = 3.9$ eV, and not at $\hbar\omega = 9.2$ eV as calculated for silver using a free-electron model, is because of the influence of the d states which lie 4 eV below the Fermi energy in silver. As shown in Fig. 3.18, the effect of the d states is to make a positive contribution to $\varepsilon_1(\omega)$ and so shift to lower energies the point at which $\varepsilon_1(\omega)$ passes through zero. Almost immediately, though, there is a

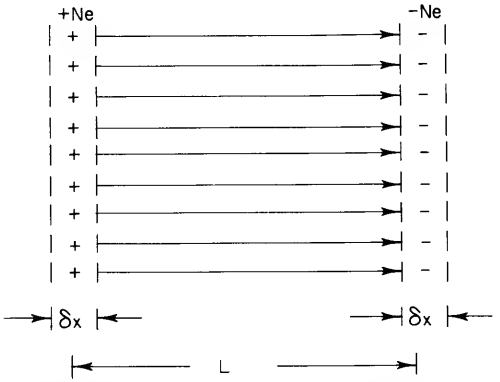

Fig. 3.19 Schematic diagram of the origin of plasma oscillations in one dimension. The equilibrium concentration of electrons is N. The net charge is zero because of a uniform background density of positive ions. Now, imagine that a region of electron charge has been uniformly translated a distance δx without disturbing the rest of the system. This leaves a region of thickness δx having a net positive charge density and creates another region of net negative charge density as illustrated in the figure. Because of the force between the two regions of unbalanced charge, they will be attracted toward each other. In the absence of a damping mechanism, there will be an overshoot, leading to the situation depicted in the figure, but with the charges reversed. The system will continue to oscillate in this manner at a characteristic frequency known as the plasma frequency. The oscillations are called plasma oscillations. The plasma frequency can be easily derived for this simple case. Recognizing that the polarization is the dipole moment per unit volume, we have

$$\mathbf{P} = -(NeA\,\delta x)L/AL = -Ne\,\delta x$$

In the absence of an external field, this gives rise to the total electric field

$$\mathbf{E} = -4\pi\mathbf{P} = 4\pi Ne\,\delta x$$

The force on an electron in the polarized region is thus

$$m\ddot{x} = -e\mathbf{E} = -4\pi Ne^2\,\delta x$$

This is just the equation for a simple harmonic oscillator with characteristic frequency

$$\omega_p^2 = 4\pi Ne^2/m$$

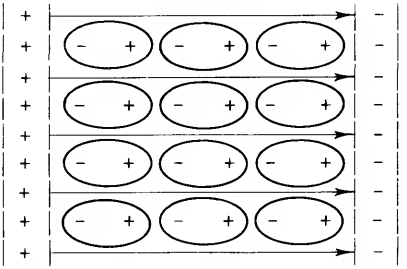

Fig. 3.20 Screening of plasma oscillations by polarization of localized electrons. Compare with Fig. 3.19. The d electrons are highly localized on atomic sites. However, their polarizability may be high enough to cause appreciable screening of the forces between fluctuations in the charge densities of conduction electrons. The result can be a significant shift in the plasma frequency.

rise in $\varepsilon_2(\omega)$ (Fig. 3.17) because of real transitions from d states. The plasma oscillations are quenched and the reflectance rises.

The plasma oscillation in silver consists of a collective oscillation, or hybrid resonance, involving both s and d electrons. It can also be thought of as a collective oscillation of the conduction electrons screened by the bound d electrons. The latter viewpoint is illustrated schematically in Fig. 3.20.

The reflectance of copper is shown in Fig. 3.21, and the decomposition of the experimental values of $\varepsilon_1(\omega)$ into free and bound contributions is shown in Fig. 3.22. Although the band structures of copper and silver are qualitatively the same, copper does not exhibit a plasma resonance. The d states in copper lie only 2 eV below the Fermi energy, so that the positive contribution of $\varepsilon_1{}^b$ comes in a frequency region where the negative contribution from $\varepsilon_1{}^f$ is of too large a magnitude for ε_1 to reach zero.

The difference in color between silver and copper arises from the positions

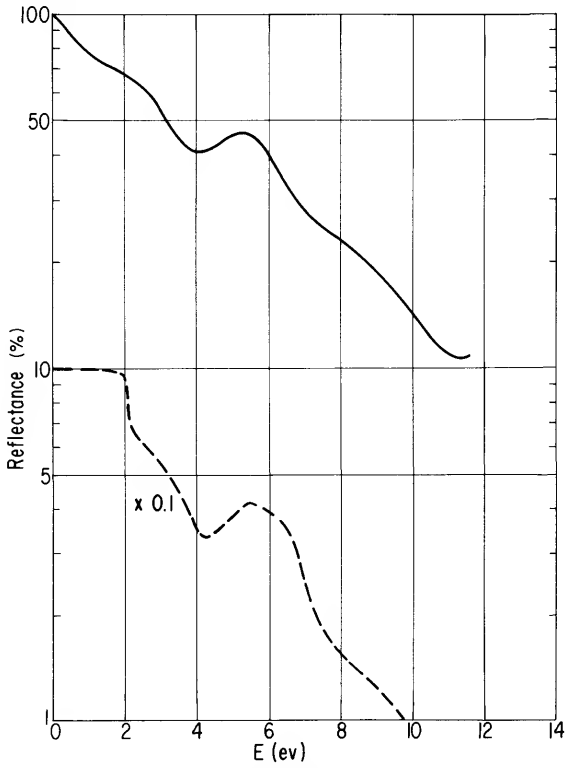

Fig. 3.21 Spectral dependence of the reflectance of Cu (lower curve) and Ni (upper curve). [Cu data from H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962); Ni data from H. Ehrenreich, H. R. Philipp, and D. J. Olechna, *Phys. Rev.* **131**, 2469 (1963).]

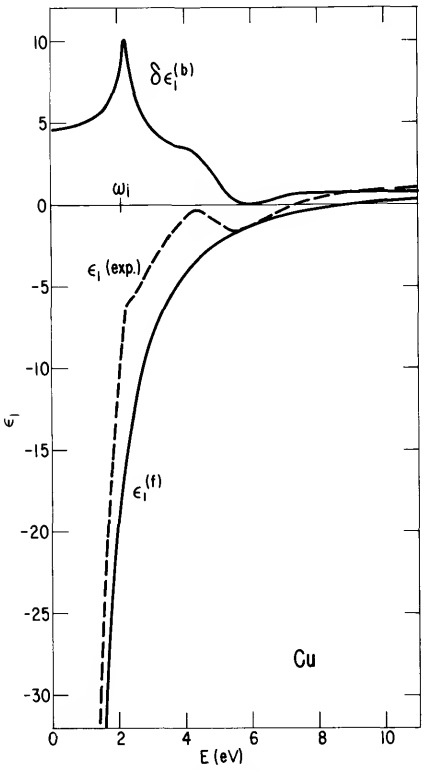

Fig. 3.22 Decomposition of the experimental values of $\varepsilon_1(\omega)$ for Cu into free and bound contributions $\varepsilon_1^{(f)}$ and $\delta\varepsilon_1^{(b)}$. The threshold energy for interband transitions is indicated by ω_i. [H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962).]

of the d bands. In silver, the d-band absorption does not begin until $\hbar\omega =$ 4 eV, or at wavelengths less than 3000Å. This is in the ultraviolet, so the reflectance is high throughout the visible. In copper, however, the d bands lie 2 eV below the Fermi energy, so that strong absorption begins at $\lambda =$ $=$ 6000Å. This wavelength is in the red. Thus, the reflectance of copper decreases in the red spectral region and continues to decrease as the wavelength decreases. Since the reflectance of copper is highest towards the red end of the visible spectral region, copper has a reddish or "copper" color.

The reflectance of nickel is also shown in Fig. 3.21. In nickel and in the other transition metals, the Fermi level intersects the d bands. Interband transitions set in at about 0.3 eV, so it is not feasible to fit the low-frequency region to the Drude model. Interband transitions cause the reflectance to drop almost immediately from the limiting reflectance of unity as ω approaches zero.

3.4 Photoemission from Copper

Having taken a first look at the optical properties of solids based on some fairly simple models, it is now appropriate to consider some photoemission results which can be readily understood in a qualitative way. Copper presents such a case.

Figure 3.23 shows a schematic energy diagram for Cu with a monolayer of Cs to lower the work function from 4.5 eV to 1.55 eV. The application of a monolayer of Cs, other alkali metals, or a compound such as Cs_2O is frequently done in order to lower the work function. There are several reasons for wanting to lower the threshold for photoemission. With practical photocathodes, it is simply the desire to get as high a quantum efficiency as possible at long wavelengths. In band structure studies, a higher quantum efficiency makes it easier to measure photocurrents. Of greater importance, it allows one to probe deeper-lying levels with conventional light sources and permits a study of conduction band states between the old vacuum level and the new, lower vacuum level. Of course, it is necessary that the only effect of the applied monolayer be to lower the work function. It must not destroy the fundamental shape of the energy distribution curve (EDC) for the underlying solid. That is something that can be determined only by experiment.

Figure 3.24 shows the EDC for Cu at four photon energies. There is little difference among the four curves.

Now look at Fig. 3.25 and note the dramatic change in the EDC on going from $h\nu = 3.7$ to 3.9 eV. The large spike in the EDC near zero energy results from excitation of d electrons. These states were inaccessible for photoemission with $h\nu = 3.7$ eV, but are accessible with $h\nu = 3.9$ eV.

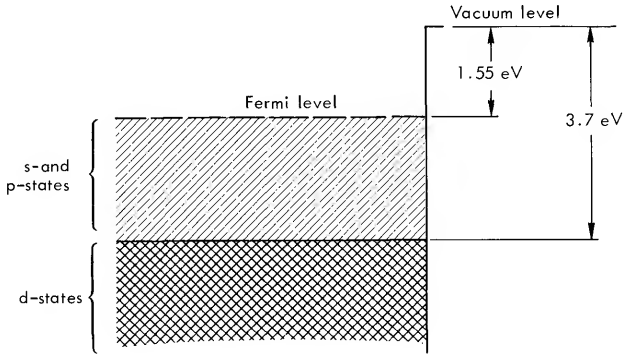

Fig. 3.23 Simplified energy band diagram for Cu with a surface monolayer of Cs. The threshold for photoemission from s- and p-like states is 1.55 eV. The threshold for emission from d states is 3.7 eV.

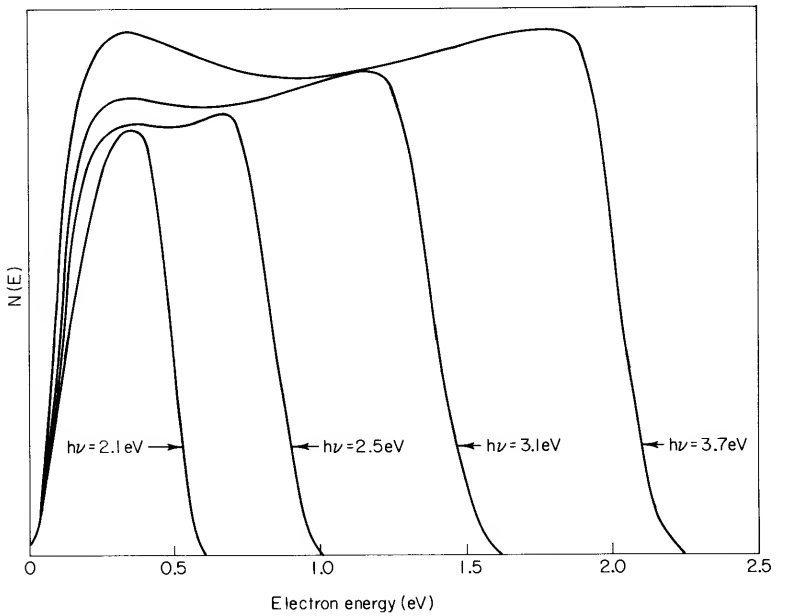

Fig. 3.24 Photoemission from Cu with a surface monolayer of Cs. At photon energies less than 3.7 eV, electrons excited from d states cannot escape. Thus these curves represent electrons excited from s- and p-like states as illustrated in Fig. 3.23. [From C. N. Berglund and W. E. Spicer, *Phys. Rev.* **136**, A1044 (1964).]

The large peak results because the density of states is much higher for d electrons than for s and p electrons; there are more d electrons in the outer shells of Cu and the d bands are flatter. Thus, in agreement with optical studies, photoemission shows that the d states in Cu lie 2.1 eV below the Fermi energy ($h\nu = 3.7$ eV minus $e\phi_F = 1.55$ eV). Note, though, that by drawing an energy level diagram in the manner of Fig. 3.23, we have avoided any discussion of selection rules. We have been concerned here only with energy conservation.

3.5 Quantum Theory of Absorption and Dispersion

The polarizability and the dielectric function obtained by a quantum mechanical analysis are of the same form as those obtained with the classical Lorentz model. To show this, we again consider first the interaction of a free atom with an electromagnetic field. We are concerned only with systems initially in their ground state, so that we need consider only the effects of polarization and absorption, not emission.

At times $t < 0$, the eigenstates of the unperturbed system are given by

$$H_0\phi_n = \mathscr{E}_n\phi_n \tag{3.38}$$

where H_0 is a time-independent Hamiltonian. At time $t = 0$, the interaction with the radiation is turned on. The Hamiltonian

$$H = H_0 + H'(t) \tag{3.39}$$

is then time-dependent, and the wave function for the atom must be found from the time-dependent Schrödinger equation

$$i\hbar\,\partial\psi/\partial t = (H_0 + H')\,\psi \tag{3.40}$$

The solutions may be written as

$$\begin{aligned}\psi(\mathbf{r}, t) &= \sum_n a_n(t)\phi_n(\mathbf{r})e^{-i\mathscr{E}_n t/\hbar} \\ &= \sum_n a_n\phi_n e^{-i\mathscr{E}_n t/\hbar}\end{aligned} \tag{3.41}$$

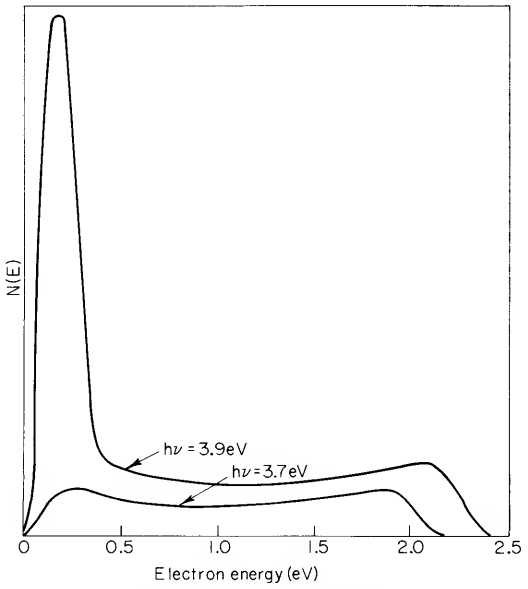

Fig. 3.25 Photoemission from Cu with a surface monolayer of Cs. At $h\nu = 3.7$ eV, electrons excited from d states are unable to escape. At $h\nu = 3.9$ eV, a large spike in $N(E)$ results from electrons escaping after excitation from the high density of d states. [From C. N. Berglund and W. E. Spicer, *Phys. Rev.* **136**, A1044 (1964).]

Substituting Eq. (3.41) into (3.40) and using (3.38) yields

$$i\hbar \sum_i (da_n/dt)\, \phi_n e^{-i\delta_n t/\hbar} = \sum_n a_n H' \phi_n e^{-i\delta_n t/\hbar} \tag{3.42}$$

Multiplying Eq. (3.42) from the left by

$$\phi_m^* e^{i\delta_m t/\hbar} \tag{3.43}$$

and integrating over all space, we get

$$i\hbar\, da_m/dt = \sum_n a_n H'_{mn} e^{i(\delta_m - \delta_n)t/\hbar} \tag{3.44}$$

where H'_{mn} is the matrix element

$$H'_{mn} = \int \phi_m^* H' \phi_n \, d\mathbf{r} \tag{3.45}$$

We take the atom to be initially in the ground state so that

$$a_n = \delta_{n0}, \qquad t = 0 \tag{3.46}$$

Then, within the approximation of first-order perturbation theory, Eq. (3.44) gives

$$\begin{aligned} i\hbar\, da_m/dt &= H'_{m0} e^{i(\delta_m - \delta_0)t/\hbar} \\ &= H'_{m0} e^{i\omega_{m0} t} \end{aligned} \tag{3.47}$$

We now need an expression for the perturbation term in the Hamiltonian. We shall keep in mind the classical oscillator by writing

$$H' = e\mathbf{E}_{\text{loc}} \cdot \mathbf{r} \tag{3.48}$$

The electric field to use is that which exists at the momentary location of the electron. Again, because we are concerned with radiation of long wavelength compared with atomic dimensions, it is sufficient for present purposes to use the electric field existing at the nucleus.

Take the electromagnetic field to be polarized in the x direction so that

$$\mathbf{E}_{\text{loc}} = E_x \cos \omega t = \tfrac{1}{2} E_x (e^{i\omega t} + e^{-i\omega t}) \tag{3.49}$$

Then, Eq. (3.48) becomes

$$H' = \tfrac{1}{2} eE_x x (e^{i\omega t} + e^{-i\omega t}) \tag{3.50}$$

and Eq. (3.45) becomes

$$H'_{mn} = \tfrac{1}{2} eE_x x_{m0} (e^{i\omega t} + e^{-i\omega t}) \tag{3.51}$$

where

$$x_{m0} = \int \phi_m^* x \phi_n \, d\mathbf{r} \tag{3.52}$$

We now use Eq. (3.51) in Eq. (3.47) and integrate from $t = 0$, when $a_m = 0$, to time t. Thus,

$$a_m = \frac{eE_x x_{m0}}{2\hbar}\left[\frac{1 - e^{i(\omega_{m0}+\omega)t}}{\omega_{m0}+\omega} + \frac{1 - e^{i(\omega_{m0}-\omega)t}}{\omega_{m0}-\omega}\right] \tag{3.53}$$

The induced dipole moment is

$$p = \langle -ex \rangle = \int \psi^*(-ex)\psi \, d\mathbf{r} \tag{3.54}$$

Substituting for ψ^* and ψ from Eq. (3.41), we get

$$p = -\sum_{k,l}\int a_k^* \phi_k^* e^{i\mathscr{E}_k t/\hbar}(ex) a_l \phi_l e^{-i\mathscr{E}_l t/\hbar}\, d\mathbf{r} \tag{3.55}$$

At times $t > 0$, the initial condition (3.46) is replaced by

$$a_n \approx \delta_{n0} \tag{3.56}$$

Thus, there will be significant contributions to the dipole moment only from those terms in Eq. (3.55) for which one of the coefficients is $a_0 \approx 1$. The term for which both a_k and a_l are a_0 contributes nothing since the integrand is then an odd function. Making the appropriate substitutions for the coefficients, it becomes clear that a single dummy index is sufficient and that

$$p = -\sum_m \int [a_m^* \phi_m^* e^{i\mathscr{E}_m t/\hbar}(ex)\, a_0 \phi_0 e^{-i\mathscr{E}_0 t/\hbar} + a_0^* \phi_0^* e^{i\mathscr{E}_0 t/\hbar}(ex)\, a_m \phi_m e^{-i\mathscr{E}_m t/\hbar}]\, d\mathbf{r} \tag{3.57}$$

Taking $a_0 = a_0^* = 1$, and $(\mathscr{E}_m - \mathscr{E}_0) = \hbar\omega_{m0}$, this can be rewritten as

$$p = -e\sum_m (x_{m0} a_m^* e^{i\omega_{m0}t} + x_{0m} a_m e^{-i\omega_{m0}t}) \tag{3.58}$$

Now, substituting for a_m and a_m^* from Eq. (3.53) and its complex conjugate, Eq. (3.58) becomes

$$p = \frac{e^2 E_x}{2\hbar}\sum_m |x_{m0}|^2 \left(\frac{e^{-i\omega t} + e^{i\omega_{m0}t}}{\omega_{m0}+\omega} + \frac{e^{i\omega t} - e^{-\omega_{m0}t}}{\omega_{m0}-\omega} + \frac{e^{i\omega t} - e^{-i\omega_{m0}t}}{\omega_{m0}+\omega} + \frac{e^{-i\omega t} - e^{-i\omega_{m0}t}}{\omega_{m0}-\omega}\right) \tag{3.59}$$

Equation (3.59) is a spatial average for the induced dipole moment. We implicitly took a spatial average when we said that it is sufficient to take the electric field at the nucleus for the field at the momentary location of

the electron. If we now take a time average, we need consider only those terms having the same frequency as the driving force, that is, the frequency ω of the electromagnetic radiation. Then,

$$p = \frac{e^2 E_x}{2\hbar} \sum_m |x_{m0}|^2 \left(\frac{1}{\omega_{m0} + \omega} + \frac{1}{\omega_{m0} - \omega} \right) (e^{i\omega t} + e^{-i\omega t}) \quad (3.60)$$

Comparing Eq. (3.60) with Eqs. (3.4) and (3.49), we see that the atomic polarizability is

$$\alpha = \sum_m \frac{2e^2 |x_{m0}|^2}{\hbar} \frac{\omega_{m0}}{\omega_{m0}^2 - \omega^2} \quad (3.61)$$

The polarizability in Eq. (3.61) is real because no allowance has been made for an energy absorption process. An energy absorption mechanism can be included phenomenologically by recognizing that the excited states are really quasistationary states, not exact stationary states. They are stationary states for the unperturbed crystal, but not for the total system of crystal plus radiation.

We now assume that because of absorption of light, the probability of finding an electron in the ground state decreases according to $e^{-\Gamma_0 t}$. The amplitude of the state thus decreases as $e^{-\Gamma_0 t/2}$. At the same time, the probability of finding the electron in the mth excited state grows as $e^{\Gamma_m t}$. We can handle this situation by making the substitution

$$\mathscr{E}_m \to \mathscr{E}_m + \tfrac{1}{2} i\hbar\Gamma_m \quad (3.62)$$

in Eq. (3.41) and carrying the analysis through as before. For simplicity, we shall ignore broadening of the initial state and assume all other states have the same broadening, so that $\Gamma_m = \Gamma$. Then, if we let $\Gamma t \to 0$, we get, instead of Eq. (3.60),

$$\begin{aligned} p &= \frac{e^2 E_x}{2\hbar} \sum_m |x_{m0}|^2 \left[\left(\frac{1}{\omega_{m0} + \omega + \frac{1}{2} i\Gamma} + \frac{1}{\omega_{m0} - \omega - \frac{1}{2} i\Gamma} \right) e^{i\omega t} \right. \\ &\quad \left. + \left(\frac{1}{\omega_{m0} - \omega + \frac{1}{2} i\Gamma} + \frac{1}{\omega_{m0} + \omega - \frac{1}{2} i\Gamma} \right) e^{-i\omega t} \right] \\ &= \frac{e^2 E_x}{2\hbar} \sum_m |x_{m0}|^2 \left(\frac{2\omega_{m0} e^{i\omega t}}{\omega_{m0}^2 - \omega^2 + \frac{1}{4}\Gamma^2 - i\Gamma\omega} \right. \\ &\quad \left. + \frac{2\omega_{m0} e^{-i\omega t}}{\omega_{m0}^2 - \omega^2 + \frac{1}{4}\Gamma^2 + i\Gamma\omega} \right) \end{aligned} \quad (3.63)$$

If we now ignore $\frac{1}{4}\Gamma^2$ relative to ω_{m0}^2, the polarization can be rewritten as

$$p = \frac{2e^2E_x}{\hbar} \sum_m |x_{m0}|^2 \left[\frac{\omega_{m0}(\omega_{m0}^2 - \omega^2)\cos \omega t}{(\omega_{m0}^2 - \omega^2)^2 + \Gamma^2\omega^2} + \frac{\omega_{m0}\Gamma\omega \sin \omega t}{(\omega_{m0}^2 - \omega^2)^2 + \Gamma^2\omega^2} \right] \tag{3.64}$$

The first term in Eq. (3.64) is in phase with the driving field and leads to the real part of the dielectric function

$$\varepsilon_1 = 1 + \frac{8\pi Ne^2}{\hbar^2} \sum_m \frac{|x_{m0}|^2 \hbar\omega_{m0}(\omega_{m0}^2 - \omega^2)}{(\omega_{m0}^2 - \omega^2)^2 + \Gamma^2\omega^2} \tag{3.65}$$

for an assemblage of N noninteracting atoms per unit volume. The second term in Eq. (3.64) is 90° out of phase with the driving field. It is the energy absorption term and leads to the imaginary part of the dielectric function,

$$\varepsilon_2 = \frac{8\pi Ne^2}{\hbar^2} \sum_m \frac{|x_{m0}|^2 \hbar\omega_{m0}\Gamma\omega}{(\omega_{m0}^2 - \omega^2)^2 + \Gamma^2\omega^2} \tag{3.66}$$

If we make the substitution

$$f_{m0} = (2m\hbar\omega_{m0}/\hbar^2)|x_{m0}|^2 \tag{3.67}$$

where f_{m0} is defined as the oscillator strength, then we can combine Eqs. (3.65) and (3.66) into a complex dielectric function given by

$$\hat{\varepsilon} = 1 + \frac{4\pi Ne^2}{m} \sum_m \frac{f_{m0}}{\omega_{m0}^2 - \omega^2 - i\Gamma\omega} \tag{3.68}$$

It should be noted once again that the question of the proper local field to use may be a difficult one. For free atoms, it is possible to use the field existing at the nucleus, but atoms in solids interact, and the variation in microscopic field within a unit cell must be considered.

If the electrons are localized, and that is what has been implicitly assumed so far, the proper local field to use is just the microscopic field existing at the position of the electron due to all sources except from that electron itself. If it is more appropriate to consider the electron as a spatial charge distribution, it is necessary to generalize the meaning of the local field as we have used it and as it was originally introduced by Lorentz. This point is discussed in Appendix B.

3.6 Oscillator Strengths and Sum Rules

We have introduced an oscillator strength defined by

$$f_{m0} = (2m\hbar\omega_{m0}/\hbar^2)|x_{m0}|^2 \tag{3.69}$$

and stated that it satisfies the sum rule

$$\sum_m f_{m0} = 1 \tag{3.70}$$

We have been concerned only with upward transitions from the ground state; but oscillator strengths also describe downward transitions and the sum rule is more general. That is, the fixed state from which and to which transitions take place need not be the ground state. We will now prove the more general sum rule

$$\sum_m f_{mn} = 1 \tag{3.71}$$

and put some substance into the previous analogies we have made with classical oscillators.

From the definition (3.69) we have, rewriting $\hbar\omega_{mn}$ as an energy difference,

$$\begin{aligned}\sum_m f_{mn} &= \sum_m (2m/\hbar^2)(\mathscr{E}_m - \mathscr{E}_n) x_{nm} x_{mn} \\ &= (2m/\hbar^2) \sum_m \tfrac{1}{2} [x_{nm}(\mathscr{E}_m - \mathscr{E}_n) x_{mn} + x_{nm}(\mathscr{E}_m - \mathscr{E}_n) x_{mn}]\end{aligned} \tag{3.72}$$

Using

$$(\mathscr{E}_m - \mathscr{E}_n) x_{mn} = [H, x]_{mn} \tag{3.73}$$

and, from a change of dummy indices,

$$(\mathscr{E}_m - \mathscr{E}_n) x_{nm} = -[H, x]_{nm} \tag{3.74}$$

which hold for $V = V(\mathbf{r})$, we get

(3.75)

Using the closure property

$$\sum_m \langle n \mid A \mid m \rangle \langle m \mid B \mid n \rangle = \langle n \mid AB \mid n \rangle \tag{3.76}$$

we can express the summation on the RHS of Eq. (3.75) as

$$\begin{aligned}\sum_m f_{mn} &= -(m/\hbar^2) \{[H, x] x - x[H, x]\}_{nn} \\ &= -(m/\hbar^2) [[H, x], x]_{nn}\end{aligned} \tag{3.77}$$

We now have the summation of f_{mn} over all final states expressed solely in terms of the initial state. Breaking up the Hamiltonian into kinetic and potential energy parts, Eq. (3.77) becomes

$$\sum_m f_{mn} = -\frac{m}{\hbar^2} \left[\left[\frac{p^2}{2m}, x \right], x \right] - \frac{m}{\hbar^2} \left[[V, x], x \right] \tag{3.78}$$

If $V = V(\mathbf{r})$, the last term is zero. Then, using the commutation relation

$$[p, x] = -i\hbar \tag{3.79}$$

Eq. (3.78) reduces to the TRK sum rule for single-electron atoms,

$$\sum_m f_{mn} = 1 \tag{3.80}$$

For an atom with Z electrons, the sum rule becomes

$$\sum_m f_{mn} = Z \tag{3.81}$$

Note that whereas our original treatment of the classical oscillator is based on a Hooke's law spring constant, the purely quantum mechanical development is more general, depending only on $V = V(\mathbf{r})$.

We have defined the oscillator strength in terms of matrix elements of the electric dipole moment. It is often more convenient to work in terms of momentum matrix elements, in which case the oscillator strength becomes

$$f_{mn} = 2 \mid p_{mn} \mid^2 / m\hbar\omega_{mn} \tag{3.82}$$

where

$$p_{mn} = \int \phi_m^*(-i\hbar\nabla)\phi_n \, d\mathbf{r} \tag{3.83}$$

Equation (3.82) is readily obtained from Eq. (3.79) by means of Eqs. (3.73) and (3.74), and the commutation relation

$$[p^2, x] = -2i\hbar p \tag{3.84}$$

which is a direct consequence of the commutation relation (3.79).

As an example of the relationship between oscillator strengths and classical oscillators, consider the relationship between power absorption for transitions between quantum states with that for the corresponding classical oscillator. The quantum mechanical result for power absorption can be found with the aid of Eq. (3.53). The result is

$$\begin{aligned} P_{\mathrm{QM}} &= (\text{transitions per second}) \times (\text{energy absorbed per transition}) \\ &= (d/dt)\,(a_m^* a_m)\,(\hbar\omega_{mn}) \\ &= (\pi e^2 \mid \mathbf{E} \mid^2 \mid x_{mn} \mid^2 / 2\hbar^2)\,\hbar\omega_{mn} \end{aligned} \tag{3.85}$$

Power absorption for the classical oscillator can be found with the aid of Eq. (3.1) to be

$$P_{\text{classical}} = \pi e^2 \mid \mathbf{E} \mid^2 / 4m \tag{3.86}$$

From Eqs. (3.85), (3.86), and (3.69), we have

$$P_{\mathrm{QM}} / P_{\text{classical}} = (2m\hbar\omega_{mn}/\hbar^2) \mid x_{mn} \mid^2 = f_{mn} \tag{3.87}$$

Thus, a quantum mechanical transition from state n to state m behaves like a classical oscillator with strength f_{mn}. An equivalent statement is that f_{mn} is the number of classical oscillators which is equivalent in terms of power absorption to a quantum mechanical system making transitions from state n to state m.

Another analogy is apparent when a comparison is made between Eqs. (3.16) and (3.68). We see that the quantum mechanical and classical results agree if the oscillator strength is interpreted as the fraction of an electron bound by a classical oscillator with spring constant $\omega_{mn}^2 m$.

3.7 Applications of Sum Rules

In Appendix D, it is shown that the sum rule for a solid is modified by the periodic potential of the lattice to become

$$\sum_m f_{mn} = 1 - \frac{m}{\hbar^2}\frac{\partial^2 \mathscr{E}_n}{\partial k^2} = 1 - \frac{m}{m_n^*} \tag{3.88}$$

where m_n^* is the effective mass for the electron in the initial state n. If the strength of the periodic potential becomes zero, corresponding to a free-electron gas, we see that $\sum_m f_{mn} = 0$. This means that there can be no photon absorption by a free-electron gas. This is readily apparent from an examination of a diagram of $\mathscr{E}$ versus $\mathbf{k}$ for a free-electron gas as illustrated in Fig. 3.26. It is not possible in such a case to conserve both energy and momentum without a third body. This is the basis of the arguments, mostly in the 1930's, that the photoelectric effect in metals must be a surface effect. It is now generally thought to be primarily a volume effect. But, let us return to some simple applications of sum rules!

We first note that the oscillator strength is proportional to the square of a dipole matrix element. Since the magnitude of the matrix element depends on the overlap of wave functions, it will be strongly energy-dependent. Transitions from deep-lying atomic levels to outer discrete levels have very small matrix elements because of the small overlap. The effect of the matrix element is much more important in this respect than the tendency for an increase in the oscillator strength, because of the factor $\hbar\omega_{mn}$. Because the oscillator strength for higher discrete levels drops so rapidly, transitions to the continuum make an important contribution to the sum rule for transitions from deep-lying levels. This is illustrated in Table 3.1, listing the oscillator strengths for the Lyman series.

It is, of course, a difficult problem to calculate such accurate oscillator strengths for transitions in solids. What is of perhaps greater interest at this point is to illustrate the effect of the exclusion principle on the sum rule. We will illustrate this for an idealized system having only four states.

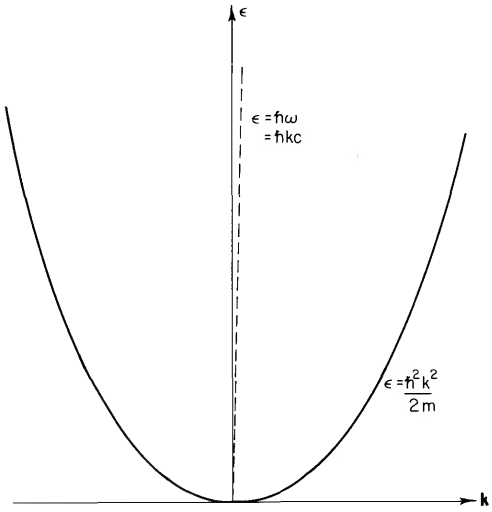

Fig. 3.26 Energy versus **k**-vector for photons and free electrons. The slope of the dispersion curve for photons is nearly vertical in a diagram for which the energy covers the range $0 \sim 100$ eV for both photons and electrons. The two curves are not parallel until the energy approaches 10^6 ev, an energy well above the range of interest for optical properties and where relativistic effects are important. It is clear that there is no way to induce a transition of a free electron from one energy to another by the absorption of a photon.

TABLE 3.1 *Oscillator Strengths for the Lyman Series in Hydrogen*

Wavelength (Å)	Transition $n = 1 \rightarrow m$	Oscillator strength (f_m)
1216	2	0.4162
1026	3	0.0791
973	4	0.0290
950	6–8	0.0158
	9–continuum	0.0101
≤912	Continuum	0.4359
		1.0000

The states are shown as a set of four energy levels in Fig. 3.27(a). The oscillator strengths, which have been arbitrarily assigned, have been chosen to conform with the sum rule $\Sigma_m f_{mn} = 1$. Now, suppose the system has only one electron in the ground state. Then, the possible transitions are as shown in Fig. 3.27(b). If there are two electrons and the system is in the ground state, the sum rule becomes $\Sigma_m f_{m0} = 2$; either electron can make the transitions.

Now, suppose the system consists of three electrons as shown in Fig.

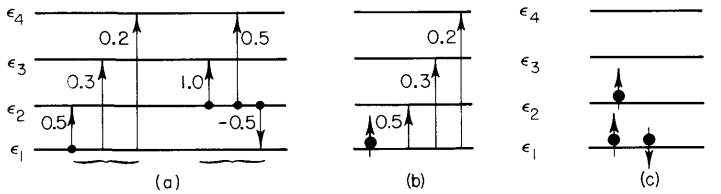

Fig. 3.27 An energy level scheme for a system having only four one-electron energy levels. Possible transitions and oscillator strengths are shown. The oscillator strengths are chosen arbitrarily except for the sum-rule requirement.

3.27(c). The electron in level $\mathscr{E}_2$ can make transitions to $\mathscr{E}_3$ or $\mathscr{E}_4$, and the total oscillator strength for allowed transitions is $1.0 + 0.5 = 1.5$. That is, the total oscillator strength for allowed transitions of the electron in level $\mathscr{E}_2$ is enhanced because level $\mathscr{E}_1$ is occupied. Similarly, the total oscillator strength for transitions of the spin-up electron in level $\mathscr{E}_1$ is depressed, because transitions to level $\mathscr{E}_2$ (with conservation of spin) are not allowed. The total oscillator strength for transitions of the spin-up electron in level 1 is decreased to 0.5. The spin-down electron in level $\mathscr{E}_1$ is not affected. Note, though, that whereas the sum of oscillator strengths for all allowed transitions of a particular electron may differ from unity, the more general sum rule $\Sigma_m f_{mn} = Z$ is satisfied. The point is that the sum rule applies to the entire system. It cannot be applied rigorously to particular groups of electrons, nor does it apply to imperfections. We shall see in the next section, though, that if care is used, the sum rule is often useful even for nonrigorous applications.

There is a sum rule for solids,

$$\int_0^\infty \omega\varepsilon_2(\omega)\, d\omega = \tfrac{1}{2}\pi\omega_p{}^2 \tag{3.89}$$

which is analogous to the sum rule $\Sigma_m f_{mn} = Z$ for atoms and is in a more useful form than the sum rule expressed in Eq. (3.88) for solids. We can see the origin of this sum rule as follows. The rate of energy absorption per unit volume from an electric field is

$$\frac{d\mathscr{E}}{dt} = \mathrm{Re}\left(\mathbf{E}\cdot\frac{\partial}{\partial t}\hat{\mathbf{D}}\right) = \mathrm{Re}\left[\mathbf{E}\cdot(\varepsilon_1 + i\varepsilon_2)\frac{\partial}{\partial t}\mathbf{E}\right] \tag{3.90}$$

With $\mathbf{E}$ having a time dependence $e^{-i\omega t}$, this becomes

$$d\mathscr{E}/dt = \omega\varepsilon_2\,|\,\mathbf{E}\,|^2 \tag{3.91}$$

Thus, the integral

$$\int_0^\infty \omega\varepsilon_2(\omega)\, d\omega \tag{3.92}$$

is a measure of the energy absorption for all frequencies.

We will now evaluate the integral (3.92) using our classical expression (3.15) for ε_2:

$$\int_0^\infty \omega\varepsilon_2(\omega)\,d\omega = \frac{4\pi Ne^2\Gamma}{m}\int_0^\infty \frac{\omega^2\,d\omega}{(\omega_{m0}^2-\omega^2)^2+\Gamma^2\omega^2}$$

$$= {\omega_p}^2\Gamma\int_0^\infty \frac{\omega^2\,d\omega}{(\omega_{m0}-\omega)^2(\omega_{m0}+\omega)^2+\Gamma^2\omega^2} \tag{3.93}$$

For Γ small, significant contributions come only for $\omega_{m0}\approx\omega$. Thus,

$$\int_0^\infty \omega\varepsilon_2(\omega)\,d\omega = \frac{{\omega_p}^2\Gamma\omega_{m0}^2}{4\omega_{m0}^2}\int \frac{d\omega}{(\omega_{m0}-\omega)^2+\Gamma^2}$$

$$= \frac{{\omega_p}^2\Gamma}{4}\,\frac{2}{\Gamma}\tan^{-1}\left(\frac{2\omega-2\omega_{m0}}{\Gamma}\right)\Bigg|_0^\infty$$

$$= \tfrac{1}{2}\pi{\omega_p}^2 \tag{3.94}$$

We have derived the sum rule under rather restricted conditions. It happens, though, that the result is exact even for interacting many-electron systems. This is shown rigorously in Chapter 6.

Sum rules are frequently defined in terms of an effective number of electrons per atom, n_{eff}, contributing to the optical properties over a finite frequency range. Thus,

$$\int_0^{\omega_c} \omega\varepsilon_2(\omega)\,d\omega = \tfrac{1}{2}\pi(4\pi N_a e^2/m)\,n_{\text{eff}}\ (\omega_c) \tag{3.95}$$

where now N_a is the density of atoms.

We will now consider a simple example to illustrate why one must be careful in the interpretation of curves of n_{eff} versus ω. Consider a metal having two electrons per atom in the valence-conduction band and ten electrons per atom in lower-lying d states. Assume the d states lie sufficiently low in energy that transitions of valence electrons are essentially exhausted before transitions from d states begin and that all other core states lie so deep they can be neglected. Then a plot of the absorption versus frequency might look as shown in Fig. 3.28.

First inclination might be to identify the area V as the number of valence electrons per atom and the area D as the number of electrons per atom in the highest d states. Then, if $V = 2.2$, for example, we might say that the effective number of valence electrons is 2.2. Alternatively, we might say that there are, of course, only two valence electrons per atom, but they have an effective mass of 0.9. But remember the example (Fig. 3.27) illustrating

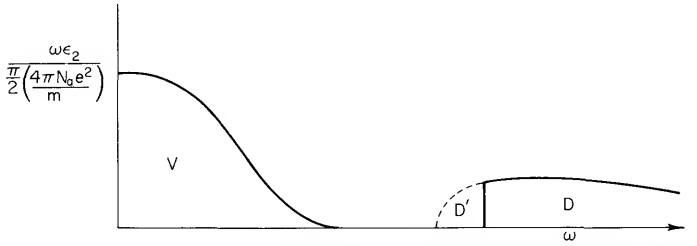

Fig. 3.28 Absorption versus frequency for a metal like Cd. [After N. V. Smith, *Advan. Phys.* **16**, 629 (1967).]

the effect of the exclusion principle on sum rules. The point is that in Fig. 3.28, an area D' is missing from the contribution due to excitation of d electrons; no excitations are possible from the d states to valence band states below the Fermi energy. The sum rule says simply that

$$V + D = 2 + 10 = 12$$

We also "know" that $D + D' = 10$, so that $V = 2 + D'$. Thus, the strength of transitions from the valence band is enhanced by the effect of the exclusion principle in prohibiting excitation of d electrons to valence states below the Fermi energy.

In some cases, it may be possible to estimate the area D'. If, for example, the valence band density of states is reasonably well known, it may be feasible to extrapolate to lower energies and determine D'. This was done by N. V. Smith for mercury, but opportunities to analyze such a simple case are rare.

We see, then, that sum rules provide a useful guide to interpreting the results of experiments and also serve as a check on the validity of the experimental data. The areas under particular bumps in a plot of $\omega\varepsilon_2$ versus ω must, however, be viewed with care. For example, if an experiment is conducted only over the region of Drude absorption, exact agreement with the Drude model cannot be expected; if it is obtained, it should not be taken seriously.

There is another sum rule which is related to plasma excitations. If an electron enters a solid, it produces the field **D**. Then, the rate of energy absorption is given by

$$\frac{d\mathscr{E}}{dt} = \frac{\omega\varepsilon_2}{|\hat{\varepsilon}_2|^2} |\mathbf{D}|^2 \tag{3.96}$$

The relevant factor is

$$\frac{\varepsilon_2}{|\hat{\varepsilon}|^2} = \frac{\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2} = -\operatorname{Im}\left(\frac{1}{\hat{\varepsilon}}\right) \tag{3.97}$$

for which there is a sum rule analogous to Eq. (3.89), namely

$$-\int_0^\infty \omega \,\mathrm{Im}(1/\hat{\varepsilon})\, d\omega = \tfrac{1}{2}\pi\omega_p{}^2 \tag{3.98}$$

Just as in Eq. (3.95), it is possible to define an effective number of electrons per atom contributing to the optical properties over a finite frequency range. This is illustrated in Fig. 3.29 for aluminum. The plot corresponding to the sum rule (3.95) rises rapidly at low frequencies because of the free-electron behavior of aluminum. The plot corresponding to the sum rule (3.98) begins to increase only near the region of plasmon excitation. Both plots saturate at a value near 3, the number of valence-conduction electrons per atom. Such plots for semiconductors and metals with d electrons are not so simple. Interband transitions play a dominant role and the sum rules do not saturate in the spectral region of most optical measurements (<25 eV). An example of the latter is shown in Fig. 3.30.

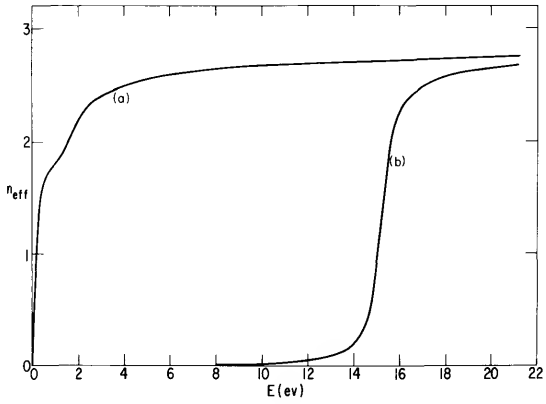

Fig. 3.29 The effective number of electrons per Al atom versus $\hbar\omega$ obtained from a numerical integration of the experimental values of (a) ε_2 and (b) $-\mathrm{Im}\,\hat{\varepsilon}^{-1}$. [H. Ehrenreich, H. R. Philipp, and B. Segall, *Phys. Rev.* **132**, 1918 (1963).]

3.8 The Absorption Coefficient, Optical Conductivity, and Dielectric Function

We shall now derive some useful relationships between α, σ, ε_2, and the transition rate W_{ji}. The absorption coefficient, as defined in Eq. (2.85), can

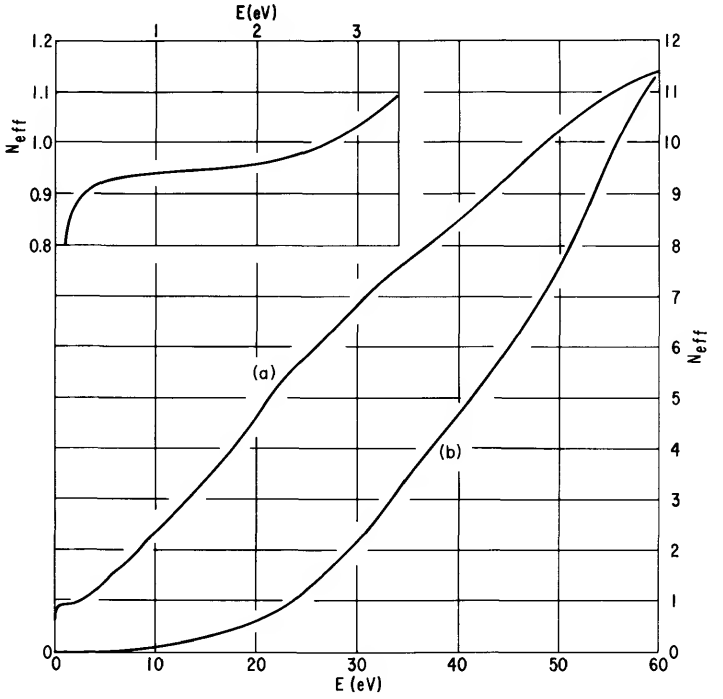

Fig. 3.30 The effective number of electrons per Au atom versus $\hbar\omega$ obtained from a numerical integration of the experimental values for (a) ε_1 and (b) $-\operatorname{Im}\hat{\varepsilon}^{-1}$. [From B. R. Cooper, H. Ehrenreich, and H. R. Philipp, *Phys. Rev.* **138**, A494 (1965).]

be written as

$$\alpha = \frac{\text{power dissipation per unit volume}}{(\text{energy density})\,(\text{energy velocity})}$$

$$= \frac{\sigma\langle E^2\rangle}{(\varepsilon_1/4\pi)\langle E^2\rangle(v)} = \frac{4\pi\sigma}{\varepsilon_1 v} \tag{3.99}$$

Now, in the case of weak absorption, $\varepsilon_1 = n^2 - k^2 \approx n^2$, and $v \approx c/n$, so that

$$\varepsilon_1 v = nc \tag{3.100}$$

Although $\varepsilon_1 \neq n^2$ for strong absorption, and the treatment of energy velocity requires some care, it happens that Eq. (3.100) is true even for strong absorption. Thus

$$\alpha = 4\pi\sigma/nc \tag{3.101}$$

Another expression for α is obtained by expressing the power absorption per unit volume as the quantum mechanical transition rate per unit volume, W_{ji}, times the energy $\hbar\omega_{ji}$ absorbed per transition. Then,

$$\alpha_{ji} = 4\pi\hbar\omega_{ji}W_{ji}/nc\langle E^2\rangle \tag{3.102}$$

If we now use the result

$$\varepsilon_1\langle E^2\rangle/4\pi = \hbar\omega \tag{3.103}$$

from Appendix E, we get

$$\alpha_{ji} = \varepsilon_1 W_{ji}/nc \tag{3.104}$$

for the absorption coefficient corresponding to transitions from state ϕ_i to state ϕ_j.

Other useful relationships that follow from Eqs. (3.101)–(3.104) and (2.91) are

$$\sigma_{ji} = \varepsilon_1 W_{ji}/4\pi \tag{3.105}$$

$$\varepsilon_{2ji} = 4\pi\sigma_{ji}/\omega_{ji} = \varepsilon_1 W_{ji}/\omega_{ji} \tag{3.106}$$

PROBLEMS

3.1 Assume that the electrons surrounding the nucleus of an atom form a spherical cloud of uniform density and radius r_0. Show that a small displacement of this electronic cloud with respect to the nucleus leads to a linear restoring force. What is the polarizability? Assume r_0 is the Bohr radius and calculate the polarizability of atomic hydrogen based on this model. How does it compare with the true polarizability? What is the magnitude of the electric field strength experienced by an electron at the distance r_0 from the nucleus? Compare this with the magnitude of field strengths available as light sources. Is the linear relationship between dipole moment and field strength justified?

3.2 Show that $n = k$ in the region of high absorption.

3.3 Derive an expression for σ based on the Lorentz model.

3.4 Show that if the local field is given by the Lorentz–Lorenz model (Appendix B), Eq. (3.12) is replaced by

$$\frac{\hat{\varepsilon} - 1}{\hat{\varepsilon} + 2} = \frac{4\pi}{3} N\hat{\alpha}$$

3.5 Assume a free-electron metal in which the maximum velocity of electrons corresponds to an energy of 10 eV. How far does the electron

travel during one period of a light wave? Is this sufficient to say that, on the average, the conduction electrons in a metal experience the macroscopic field **E**?

3.6 Show that if the Lorentz–Lorenz local field condition is incorporated in Eq. (3.1), the equation can be written in a form that includes the local field correction as a change in the resonant frequency. Show that ω_0^2 is replaced by $\omega^2_{\text{eff}} = \omega_0^2 - (4\pi Ne^2/m)$. If N is sufficiently large, the effective resonant frequency approaches zero. Is this possible with reasonable values of N?

3.7 Show that the group velocity, $v_g = \partial\omega/\partial(2\pi/\lambda)$ is given by $v_g = c[1 - (\omega_p^2/\omega^2)]^{1/2} = nc$ at the plasma frequency. Thus the group velocity is zero and there is no propagation of energy at the plasma frequency. This is why we refer to plasma oscillations rather than waves. To obtain a finite group velocity, the q dependence of the dielectric function must be taken into account.

3.8 Show that in the presence of interband transitions (bound electrons), the condition for plasma resonance can be written as

$$\hat{\varepsilon}(\omega) = [1 + \delta\hat{\varepsilon}^b(\omega)]\{1 - [(\omega_p^*)^2/\omega^2]\} = 0$$

where $(\omega_p^*)^2 = 4\pi Ne^2/m(1 + \delta\hat{\varepsilon}^b)$ is the square of the screened plasma frequency and τ has been neglected.

3.9 Show that the area under the Drude curve of σ/c versus $\hbar\omega$ equals the number of electrons per atom and thus satisfies a sum rule.

3.10 Assume that the local field correction for an insulator is given by the Lorentz–Lorenz local field. Show that in such a case the imaginary part of the polarizability is given by

$$\alpha_2 = 9nk/2\pi N[(n^2 + k^2)^2 + 4(n^2 - k^2 + 1)]$$

Derive a sum rule for α_2 analogous to that of Eq. (3.99).

3.11 There are several ways to calculate power absorption for a quantum mechanical system. One is to include the lifetime of states in the presence of the perturbation. Another method, to be used in later sections, is to consider monochromatic light inducing transitions over a range of neighboring states. Still another method is to consider a discrete transition induced by "white light." Calculate Eq. (3.96) using the latter method. Take the second term of Eq. (3.53) to give the transition amplitude and interpret E_x as the spectral density of the driving electric field. Then, $|E|^2$ is proportional to the spectral power density. Thus, the probability of finding an electron in state m, after starting in state n, is $\int a_m^* a_n \, d\omega$. It will be seen that the time

factor has a sharp maximum at $\omega = \omega_{mn}$, so that the other factors involving ω can be taken outside the integral and evaluated at ω_{mn}, and the limits of integration can be extended to $\pm\infty$. It is implicit in this that there are no phase relations between the different frequency components, so that each frequency range $\Delta\omega$ makes an additive contribution to the absorption.

Calculate the power absorbed by a classical oscillator driven by a band of frequencies with constant spectral power density in the neighborhood of the resonant frequency.

FURTHER READING

J. C. Slater, "Insulators, Semiconductors and Metals," McGraw-Hill, New York, 1967.
M. Garbuny, "Optical Physics," Academic Press, New York, 1965.
F. Seitz, "The Modern Theory of Solids," McGraw-Hill, New York, 1940.
H. A. Lorentz, "The Theory of Electrons," Dover Publications, Inc., New York, 1952.
P. K. L. Drude, "Theory of Optics," Dover, New York, 1959.
J. N. Hodgson, "Optical Absorption and Dispersion in Solids," Chapman and Hall, London, 1970.

Chapter 4

FREE-ELECTRON METALS

The first realistic model for a metal is due to Drude. He simply treated a metal as a free-electron gas and neglected the positive-ion background. Sommerfeld modified the model later by substituting Fermi–Dirac statistics for Boltzmann statistics. The model is remarkably simple, yet the simple ideas are still useful in the interpretation of the optical properties of metals.

In this chapter, we shall examine the properties of nearly free-electron metals in the spirit of the Drude model but in the light of modern understanding of the electronic structure of metals. This means we will be interested in seeing what can be learned about the Fermi surface as well as interpreting the traditional optical properties such as reflectivity and absorption coefficient.

The discussion is limited largely to low optical frequencies. In the infrared, most transitions are intraband transitions. These are the transitions characteristic of a free-electron metal. They involve the excitation of an electron to another state within the same band and therefore can take place only in metals. Intraband transitions persist to higher frequencies, but then, in real metals, interband transitions become important and the distinction between insulators and metals begins to break down. The effects of interband transitions are discussed in other chapters.

4.1 Classical Theory of Free-Electron Metals

We derived in Chapter 3 an expression for the dielectric function of a free-electron metal based on a rather simple phenomenological model. The model includes a viscous damping term, which leads to energy absorption by the system. We can now be more explicit about the mechanism for energy absorption; namely, energy absorption must result from electron scattering by the same mechanism that results in Joule heating when an electric current passes through a metal. This is the essence of the treatment of electrical and optical properties set forth by Drude in 1900. We shall now consider some aspects of Drude's treatment, but with an emphasis more appropriate to our present purposes.

An electron in a metal makes random collisions with the defects present. These defects may consist of various impurities and crystal inperfections; but we shall generally concern ourselves only with scattering by the crystal lattice, that is, electron–phonon scattering. The defects are the phonons present at finite temperature. Making reference to Fig. 4.1, the mean free time between collisions is

$$\bar{t} = (t_1 + t_2 + \cdots + t_n)/n \tag{4.1}$$

and the mean free path is

$$\bar{l} = (l_1 + l_2 + \cdots + l_n)/n \tag{4.2}$$

Both $\bar{t}$ and $\bar{l}$ are functions of temperature.

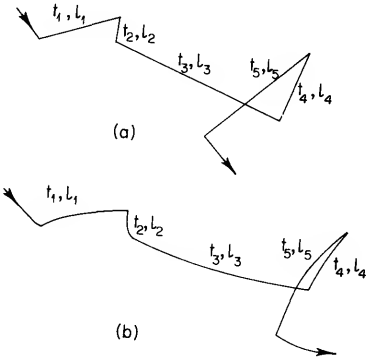

Fig. 4.1 Trajectories of an electron in a metal. (a) No field present; (b) in a constant electric field.

Assume that the probability that an electron makes a collision in a time interval dt is dt/τ, where τ is a constant. Then, if we consider $n(0)$ electrons at time $t = 0$, the number of electrons that will not have suffered any collisions

between time $t = 0$ and $t = t$ is

$$n(t) = n(0)e^{-t/\tau} \tag{4.3}$$

The constant τ is the relaxation time; it is also the mean free time between collisions, as we now show:

$$\bar{t} = \frac{1}{n(0)} \int_0^\infty tn(t) \frac{dt}{\tau} = \int_0^\infty te^{-t/\tau} \frac{dt}{\tau} = \tau \tag{4.4}$$

In the presence of a constant electric field, the electrons experience an acceleration $-e\mathbf{E}/m$. If, after each collision, an electron starts off again in a random direction, the average net distance traveled after n collisions will be

$$\begin{aligned} \mathbf{L} &= -\frac{e\mathbf{E}}{2m}(t_1^2 + t_2^2 + \cdots + t_n^2) \\ &= -\frac{e\mathbf{E}}{2m} n\bar{t^2} = -\frac{ne\mathbf{E}}{2m} \int_0^\infty t^2 e^{-t/\tau} \frac{dt}{\tau} \\ &= -\frac{ne\mathbf{E}\tau^2}{m} \end{aligned} \tag{4.5}$$

These n collisions will take on the average a time $n\tau$. Thus the average velocity in the presence of the electric field is

$$\bar{\mathbf{v}} = \mathbf{L}/n\tau = -e\mathbf{E}\tau/m \tag{4.6}$$

We can look at the interaction between electron and electric field in a slightly different way. In the presence of a constant electric field, with no energy loss mechanism, an electron will aquire in a time interval dt' an increment in velocity of $-(e\mathbf{E}/m)\,dt'$. If the constant field were applied only for the time interval dt', the electrons would maintain the velocity increment $-(e\mathbf{E}/m)\,dt'$ at all later times t. Suppose, now, that the energy loss mechanism we have been discussing is included. Then, the velocity increment aquired at time t' in a time interval dt' will have been damped, at time t, as a result of scattering, by the factor $\exp[-(t - t')/\tau]$. To determine the average velocity at time t, we need only sum up, by means of an integration, all the contributions to the velocity at time t resulting from velocity increments acquired at all earlier times t'. Thus, for a constant electric field, the average velocity is given by

$$\begin{aligned} \bar{\mathbf{v}} &= \int_{-\infty}^{t} -(e\mathbf{E}/m) \exp[-(t - t')/\tau]\, dt' \\ &= -e\mathbf{E}\tau/m \end{aligned} \tag{4.7}$$

in agreement with Eq. (4.6).

The increment in velocity, $d\mathbf{v}(t)$, can be considered as the response to an electric field acting for a time interval dt' at time t' with the response function defined as

$$G(t - t') = -(e/m)\exp[-(t - t')/\tau] \tag{4.8}$$

If the field varies sinusoidally in time,

$$\mathbf{E} = \mathbf{E}(0)e^{-i\omega t} \tag{4.9}$$

then,

$$\bar{\mathbf{v}} = \int_{-\infty}^{t} -\frac{e\mathbf{E}(0)}{m}(\exp - i\omega t')(\exp -(t - t')/\tau)\,dt'$$

$$= -\frac{e\mathbf{E}}{m}\frac{\tau}{1 - i\omega\tau} \tag{4.10}$$

Since the conductivity is related to the average velocity according to

$$\sigma = -Ne\bar{\mathbf{v}}/\mathbf{E} = Ne^2\tau/m \tag{4.11}$$

where N is the density of conduction electrons, the conductivity for a time-dependent field is given by

$$\hat{\sigma} = \frac{Ne^2\tau}{m}\frac{1}{1 - i\omega\tau} \tag{4.12}$$

That is, the conductivity is complex, and is related to the ordinary dc conductivity σ_0 according to

$$\hat{\sigma} = \sigma_0/(1 - i\omega\tau) \tag{4.13}$$

It has real and imaginary parts given by

$$\sigma_1 = \sigma_0/(1 + \omega^2\tau^2), \qquad \sigma_2 = \sigma_0\omega\tau/(1 + \omega^2\tau^2) \tag{4.14}$$

The usual way of deriving expressions for transport properties, including the complex conductivity, is via the Boltzmann equation. It is also necessary to apply Fermi–Dirac statistics to the behavior of the electrons. This subject is well treated in numerous texts. We shall just summarize some of the essential features here.

The Boltzmann equation is

$$\frac{\partial f}{\partial t} + \dot{\mathbf{k}}\cdot\nabla_{\mathbf{k}}f + \mathbf{v}\cdot\nabla_{\mathbf{r}}f = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}} \tag{4.15}$$

where $f = f(\mathbf{k}, \mathbf{r}, t)$ is a distribution function describing the distribution of electrons in $\mathbf{k}$-space and real space at time t. The right-hand side of Eq. (4.15) describes the rate of change of f resulting from collisions of electrons

with the lattice. The Boltzmann equation is, in general, extremely difficult to solve. It is easily solved, however, under certain restrictions. If we consider a steady-state system, $\partial f/\partial t = 0$. Then, if the disturbance from an electric field is small, we can write

$$f = f_0 + f_1 \tag{4.16}$$

where f_0 is the distribution function in equilibrium, i.e., the Fermi–Dirac distribution function, and f_1 is a small displacement from equilibrium. Finally, if we assume that the term describing lattice collisions can be expressed in terms of a relaxation time $\tau(\mathbf{k})$ such that

$$(\partial f/\partial t)_{\text{coll}} = -f_1/\tau(\mathbf{k}) \tag{4.17}$$

the Boltzmann equation can be solved. If it is further assumed that τ is a constant, the results are those we have already derived from a different approach.

There are several aspects of the relaxation time worth noting. The concept of a relaxation time is introduced for its simplicity, but it is useful only if, in fact, it is possible to describe many processes in terms of a lifetime which is in itself relatively simple. Ideally, that means that τ should be well approximated by a constant or a simple function of energy. It is also desirable that the same τ be useful in explaining more than one property. For example, the ratio of thermal to electrical conductivity is proportional to temperature and is the same for all metals (Wiedemann–Franz law). This requires that the lifetime appropriate for electrical conductivity be the same as that for thermal conductivity.

We have shown earlier that $\bar{t} = \tau$ on the basis of a simple explicit model in which scattering is isotropic. It should be noted that the assumption of isotropic scattering is important. The relaxation time that appears in the Boltzmann equation is the same as the mean free time between collisions only for isotropic scattering. Isotropic scattering results from scattering from acoustic phonons, but scattering from impurities is preferentially in the forward direction. Forward scattering requires more scattering events to return to equilibrium and the relaxation time is longer than for isotropic scattering. As a result, the relaxation time for impurity scattering is greater than the mean free time between collisions. Also, in general, the lifetime is a tensor for anisotropic solids. We shall consider only isotropic scattering, in which case $\bar{t} = \tau$, and we identify both characteristic times as simply the lifetime.

There are two final remarks to be made. The mass which appears in these equations for the transport properties must be some effective mass. We shall see later in this chapter what is the appropriate effective mass for optical properties. Also, we have implicitly assumed that the electric field does not vary appreciably over a mean free path. We should expect that for

strong absorption of light near the surface of a metal, the field might well be attenuated so rapidly in space that this assumption may not be valid. We shall return to this point and some of its interesting features in Section 4.3, where we discuss the anomalous skin effect.

4.2 The Classical Skin Effect

The dielectric function for the Drude model of a free-electron metal is given by

$$\hat{\varepsilon} = 1 - \frac{4\pi Ne^2}{m} \frac{1}{\omega(\omega + i/\tau)} \tag{4.18}$$

We expect that at sufficiently low frequencies and high purity, the lifetime is determined by electron–phonon scattering, and we can use the ordinary dc electrical conductivity. By low frequencies, it is meant that $\omega \ll 1/\tau$. From Eqs. (4.18) and (4.11), we thus obtain

$$\begin{aligned} \hat{\varepsilon} &= 1 - \frac{4\pi\sigma}{\omega\tau} \frac{1}{(\omega + i/\tau)} \\ &\approx 1 - \frac{4\pi\sigma\tau}{\omega} \left(\omega - \frac{i}{\tau}\right) \end{aligned} \tag{4.19}$$

The real and imaginary parts of ε are

$$\varepsilon_1 = 1 - 4\pi\sigma\tau \tag{4.20}$$

$$\varepsilon_2 = 4\pi\sigma/\omega \tag{4.21}$$

Thus, with $\omega\tau \ll 1$, $|\varepsilon_1| \ll |\varepsilon_2|$. These restrictions, together with Eqs. (3.25) and (3.26), give

$$n = k, \qquad \omega\tau \ll 1 \tag{4.22}$$

Using Eqs. (2.84), (2.88), and (4.22), we find that the decay in magnitude of the electric vector within the metal is given by

$$\mathbf{E} = \mathbf{E}_0 \exp[-(2\pi\sigma\mu\omega/c^2)^{1/2} r] = \mathbf{E}_0 \exp(-r/\delta_{\mathrm{Cl}}) \tag{4.23}$$

The term

$$\delta_{\mathrm{Cl}} = c/(2\pi\sigma\mu\omega)^{1/2} \tag{4.24}$$

is known as the classical skin depth. It is the same result obtained in texts on electromagnetic theory. Table 4.1 gives values of δ_{Cl} calculated for copper using the dc electrical conductivity at 300°K and at 4°K.

TABLE 4.1 *Classical Skin Depth and Relaxation Parameters in Copper*

ω, sec^{-1}		10^6	10^8	10^{10}	10^{12}	10^{14}	10^{16}
λ, cm		1.9×10^5	1.9×10^3	1.9×10	1.9×10^{-1}	1.9×10^{-3}	1.9×10^{-5}
v_F/ω, cm		1.4×10^2	1.4	1.4×10^{-2}	1.4×10^{-4}	1.4×10^{-6}	1.4×10^{-8}
4°K:	δ_{Cl}	1.2×10^{-4}	1.2×10^{-5}	1.2×10^{-6}	1.2×10^{-7}	1.2×10^{-8}	1.2×10^{-9}
	$\omega\tau$	5×10^{-4}	5×10^{-2}	5	5×10^2	5×10^4	5×10^6
300°K	δ_{Cl}	1.6×10^{-2}	1.6×10^{-3}	1.6×10^{-4}	1.6×10^{-5}	1.6×10^{-6}	1.6×10^{-7}
	$\omega\tau$	4×10^{-8}	4×10^{-6}	4×10^{-4}	4×10^{-2}	4	4×10^2

[a] The classical skin depth δ_{Cl} in copper is given for two temperatures as calculated using the dc electrical conductivity of copper. The parameter v_F/ω is the distance traveled by an electron on the Fermi surface in the time $1/2\pi$ of the period of the light wave.

Clearly, the last column in Table 4.1 is not valid; it is simply not correct to use the dc electrical conductivity for $\omega\tau \gg 1$. The use of the dc electrical conductivity implicitly assumes that the electric field acting on an electron is constant during a time of the order of several lifetimes. That is, Ohm's law requires equilibrium. Equilibrium is established by electron–phonon collisions, and it requires several such events during the period of one oscillation of the electric field.

Deviations from Ohm's law that occur when the electric field varies appreciably during a time τ, that is when $\omega\tau > 1$, are easily handled. One just requires that the equations for the current be such that if $\mathbf{J}^{ind}$ differs from its equilibrium value $\sigma\mathbf{E}$, then the rate of change of $\mathbf{J}^{ind}$ as a result of collisions must be such as to make $\mathbf{J}^{ind}$ tend exponentially toward its equilibrium value with time constant τ. Thus,

$$(d/dt)\,\mathbf{J}^{ind} = (1/\tau)(\sigma_0\mathbf{E} - \mathbf{J}^{ind}) \tag{4.25}$$

For electric fields varying as $e^{-i\omega t}$, this becomes

$$\mathbf{J}^{ind} = \sigma_0\mathbf{E}/(1 - i\omega\tau) \tag{4.26}$$

But this is just what we derived in the preceding section [Eq. (4.13)] from a different viewpoint!

What we want to emphasize now is that Ohm's law, which gives

$$\mathbf{J}^{ind} = \sigma_0\mathbf{E} \tag{4.27}$$

describes a local, instantaneous relationship between the total electric field and the induced current. This is the model assumed in the classical theory of the skin effect. At sufficiently high frequencies, the model must be modified. The current at a particular position and a particular time depends upon the electric field at that same position; but it depends upon the electric field at earlier times. That is, the inertia of the electrons leads to relaxation effects! The Drude model of metals incorporates these relax-

ation effects into the model by means of a viscous damping term. In Section 4.1, we have derived the same result for the conductivity, but in a manner such as to emphasize that the contributions to the current at time t depend upon the electric field at previous times of the order of τ.

Figure 4.2 illustrates how the nature of the interaction between photons and electrons in metals changes with increasing frequency (and electron mean free path). Typical values for the parameters in each region are given for copper in Table 4.1. Figures 4.3 and 4.4 show where the regions characterized in Fig. 4.2 appear in an $\omega-\tau$ diagram.

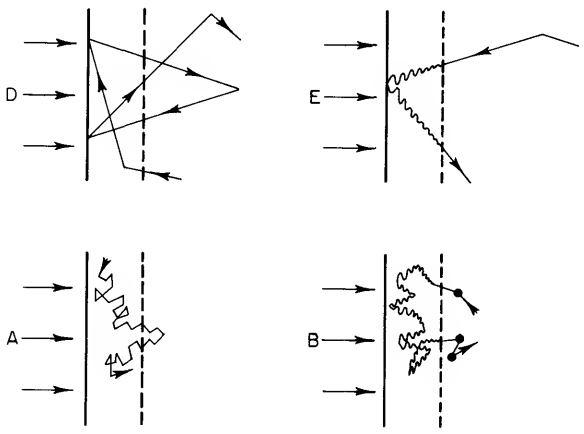

Fig. 4.2. Electron trajectories in the skin layer. The dashed line marks the skin layer, the zigzag line represents a colliding electron, the wavy lines represent the oscillatory motion of the electron due to the alternating field, and the arrows indicate the incident light waves. The optical characteristics in the four extreme regions illustrated depend upon the relative amplitudes of the mean free path l, the skin depth δ, and the mean distance v/ω traveled by an electron in a time $1/2\pi$ of the period of a light wave. (A) Classical skin effect, $l \ll \delta$ and $l \ll v/\omega$. (B) Relaxation region, $v/\omega \ll l \ll \delta$. (D) Anomalous skin effect, $\delta \ll l$ and $\delta \ll v/\omega$. (E) Extreme anomalous skin effect, $v/\omega \ll \delta \ll l$. [From H. B. G. Casimir and J. Ubbink, *Philips Tech. Rev.* **28**, 300 (1967).]

Region A, the region of the classical skin effect, is of concern only at infrared and longer wavelengths. In this region, the mean free path of an electron is sufficiently small and the frequency of the light wave low enough that an electron suffers many collisions during the time it spends in the skin layer and during one period of the light wave. Thus, this region is well described by a local, instantaneous relationship between induced current and total electric field.

At high frequencies, as we have seen, inertial effects of the electron become important. This is region B, the relaxation region. Here, the electron suffers many collisions within the skin layer, but the field of the light wave

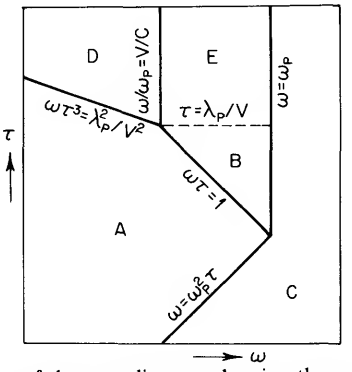

Fig. 4.3. A logarithmic plot of the ω–τ diagram showing the regions illustrated in Fig. 4.2. The new region, region C, is characterized by transmission. [From H. B. G. Casimir and J. Ubbink, *Philips Tech. Rev.* **28**, 300 (1967).

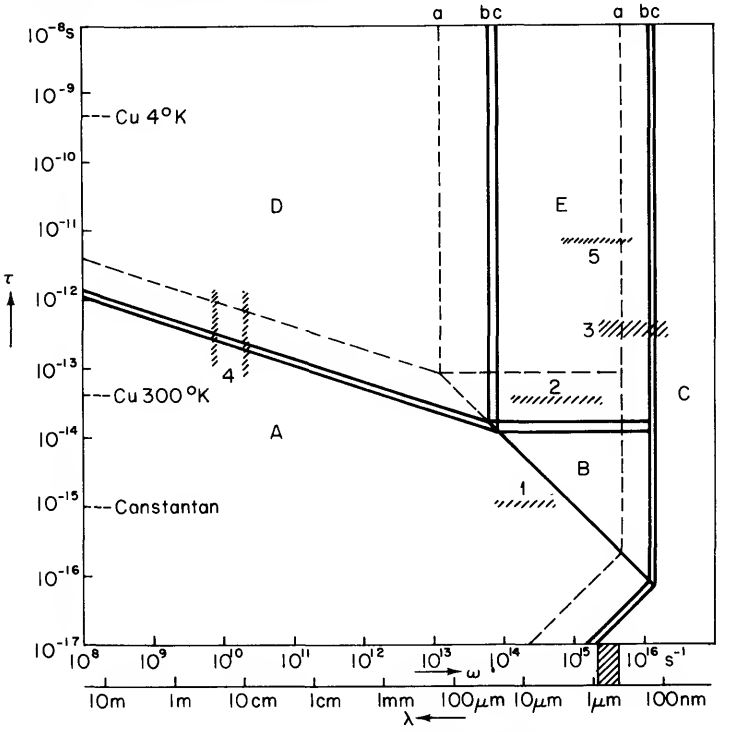

Fig. 4.4 The ω–τ diagram for three monovalent metals: (a) cesium, (b) constantan, (c) copper. The hatched area along the horizontal axis corresponds to the visible spectrum; (A) normal skin effect, (B) relaxation, (C) transmission, (D) anomalous skin effect, (E) anomalous reflection. The finely shaded strips indicate the ω–τ region where measurements have been carried out by: (1) E. Hagen and H. Rubens, *Ann. Phys.* **11**, 873 (1903). (2) K. Försterling and V. Freedericksz, *Ann. Phys.* **40**, 200 (1913). (3) R. W. Wood, *Phys. Rev.* **44**, 353 (1933). (4) R. G. Chambers, *Proc. Roy. Soc.* **A215**, 481 (1952); M.A. Biondi, *Phys. Rev.* **102**, 964 (1956). [From H. B. G. Casimir and J. Ubbink, *Philips Tech. Rev.* **28**, 300 (1967).]

also oscillates many times between collisions. Because of the inertia of the electrons, the induced current lags behind the field by an increasing amount as the light frequency increases. The phase lag approaches 90° for extreme relaxation. In this region, electron collisions are of little importance. The electrons respond to the rapidly oscillating electric field as free electrons, only occasionally undergoing a collision. Thus, the electrons act to screen the external field, so the reflection is very high, and because of the phase lag, absorption is negligible. This is characteristic of many metals in the optical region before the onset of interband transitions. Note, though, that here, absorption refers to the fraction of the incident light absorbed, not to the absorption coefficient.

At very high frequencies, the reflectance drops and transmission is dominant. We have already seen that transmission takes place above the plasma frequency. Then, there is no longer a skin effect. Transmission also occurs below the plasma frequency if $\omega \gg \omega_p^2\tau$. This follows from Eqs. (3.30) and (3.31) if $\omega\tau \gg 1$, for then, $\varepsilon_1 \to 1$ and $\varepsilon_2 \to 0$, so that the reflectance and absorption approach zero. Thus regions A and B are separated from region C, the region of transparency, by the two lines $\omega = \omega_p$ and $\omega = \omega_p^2\tau$ as shown in Fig. 4.3.

4.3 The Anomalous Skin Effect

There is another way in which the local, instantaneous relationship implicit in Eq. (4.27) may break down: The electric field may vary appreciably over a distance comparable to the mean free path. Then, the velocity of an electron at a particular position depends upon the electric field at other positions within distances of the order of the mean free path. The electron "remembers" other fields at other positions! This nonlocal spatial relationship is analogous to the noninstantaneous relationship between the electric field and induced current at frequencies $\omega > 1/\tau$. The most common example of a nonlocal spatial relationship between current and field is the anomalous skin effect. This effect was first studied carefully by Pippard [1]. It is illustrated by region D of Figs. 4.2–4.4. Here, the mean free path is greater than the skin depth, where now, the skin depth is the anomalous skin depth, which differs from the classical skin depth. We shall shortly derive an approximate equation for the anomalous skin depth.

Region D can be reached from region A if the mean free path can be substantially increased from the usual values, which are of the order of tens or hundreds of angstroms at room temperature. This can be done if a sample of high purity, say pure copper, is cooled to very low temperatures. However, alloys such as constantan cannot exhibit the anomalous skin effect. The mean free path in alloys is limited by scattering from the lattice because of a lack of perfect periodicity.

To analyze the nonlocal spatial relationship between current and field, it is convenient to work with the Fourier transforms of the actual fields. Thus,

$$\mathbf{E}(\mathbf{q}) = [1/(2\pi)^3] \int \mathbf{E}(\mathbf{r}) \exp(-i\mathbf{q} \cdot \mathbf{r})\, d\mathbf{r} \tag{4.28}$$

Since the fields are also generally harmonic in time as well as in space, a complete description in terms of Fourier components leads to

$$\mathbf{J}^{\text{ind}}(\mathbf{q}, \omega) = \hat{\sigma}(\mathbf{q}, \omega)\, \mathbf{E}(\mathbf{q}, \omega) \tag{4.29}$$

The description in terms of a (generally complex) conductivity tensor which depends on both frequency and wave vector is equivalent to the description in terms of a frequency- and wavevector-dependent dielectric tensor as emphasized in Chapter 2.

Electric fields are strongly damped in metals, but magnetic fields may be strong (see Problem 2.6). The magnetic fields induce currents and these currents in turn give rise to weak electric fields. Thus, it is reasonable to define a surface impedance as

$$Z = \mathbf{E}_0 / \int \mathbf{J}^{\text{ind}} \cdot d\mathbf{S} \tag{4.30}$$

where $\mathbf{E}_0$ is the electric field component along the surface of the metal, $\mathbf{J}^{\text{ind}}$ is the current density within the metal, and $d\mathbf{S}$ is a differential element of area. At low frequencies, the displacement current is negligible, and Eq. (2.77) can be simplified to

$$\nabla \times \mathbf{H} = 4\pi\sigma\mathbf{E}/c = (4\pi/c)\, \mathbf{J}^{\text{ind}} \tag{4.31}$$

for the interaction of light with a metal. Since the induced current decreases rapidly with increasing depth into the metal, we can evaluate the integral $\int \mathbf{J}^{\text{ind}} \cdot d\mathbf{S}$ by using Stokes theorem and an integration contour that is parallel to $\mathbf{H}$ at the surface and extends into the metal to a depth at which the fields can be taken as zero. In this case, Eq. (4.30) becomes

$$\begin{aligned} Z &= 4\pi\mathbf{E}_0 / \left[c \int_S (\nabla \times \mathbf{H}) \cdot d\mathbf{S}\right] = (4\pi/c)\, \mathbf{E}_0 / \int_l \mathbf{H} \cdot d\mathbf{l} \\ &= (4\pi/c)\, \mathbf{E}_0 / \mathbf{H}_0 \end{aligned} \tag{4.32}$$

where $\mathbf{E}_0$ and $\mathbf{H}_0$ designate fields at the surface of the metal. It is clear from Eq. (4.32) that the impedance of free space must be $4\pi/c$.

At low frequencies, such as we have assumed, the displacement current is negligible and the surface impedance is a pure resistance. The surface resistance or surface impedance is directly related to the reflectivity according to

$$R = \left| \frac{Z - (4\pi/c)}{Z + (4\pi/c)} \right|^2 \tag{4.33}$$

[The derivation of Eq. (4.33) is given in Appendix C.] Thus, from a direct measurement of the reflectivity, the surface resistance can be easily determined.

Pippard measured the surface resistance of metals at low temperature and found that the surface resistance does not agree with the bulk resistance. The reason is simply that the measured surface resistance depends on the penetration depth, that is, the skin depth, and when the skin depth is less than the mean free path, Ohm's law is not valid in the surface region where energy is absorbed. In fact, we should not even expect a simple exponential form for the way in which the field decays inside the metal. If the decay were exponential, electrons leaving the surface of the metal would acquire momentum from the field within the skin depth and carry it inward a distance of the order of l. Currents would exist at depths greater than δ (see Fig. 4.2D). But this is inconsistent with the field and current being confined to a depth of the order of δ. The field must distribute itself in such a way as to reduce these effects. This is a very difficult problem. The Boltzmann equation must be solved taking into account the spatial variation of $\mathbf{E}$ as well as the nature of electron scattering at the surface.

An exact treatment is difficult, but some understanding of the physics involved can be gotten from a simple model. Electrons traveling nearly perpendicular to the surface spend only a small fraction of a mean free path within the skin depth (see Fig. 4.5). They cannot acquire nearly so much momentum from the field as electrons traveling nearly parallel to the surface. We can thus say that the effective concentration of electrons contributing to absorption is

$$n_{\text{eff}} \approx (\delta/l)n \tag{4.34}$$

or that the effective conductivity is

$$\sigma_{\text{eff}} \approx (\delta/l)\sigma \tag{4.35}$$

We can now find a self-consistent solution to the problem. We say that

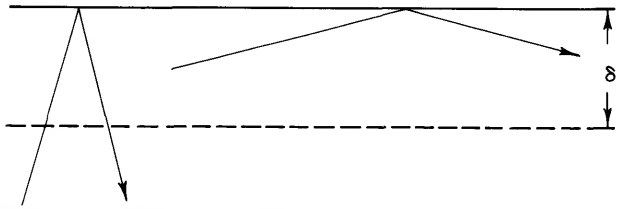

Fig. 4.5 Electron trajectories near the surface. Only electrons traveling nearly parallel to the surface are effective in absorbing and screening electromagnetic radiation when $l \gg \delta$.

a metal with this conductivity would have a skin depth

$$\delta = c/(2\pi\sigma_{\text{eff}}\mu\omega)^{1/2} \tag{4.36}$$

which is just a plausible modification of Eq. (4.24). From the last two equations and

$$\sigma = Ne^2\tau/m = Ne^2 l/mv_F \tag{4.37}$$

where v_F is the velocity of electrons on the Fermi surface, we get

$$\begin{aligned}\delta &= c/(2\pi Ne^2\delta\mu\omega)^{1/2} \\ &= (mc^2 v_F/2\pi Ne^2\mu\omega)^{1/3}\end{aligned} \tag{4.38}$$

The important point is that a measurement of the skin depth is a measure of the electron velocity at the Fermi surface. The principal contribution is from those electrons traveling parallel to the surface. If the Fermi surface is anisotropic, the surface impedance will be different for different orientations of a metal crystal. In principle, and to some degree in practice, the geometry of Fermi surfaces can be determined in this manner.

We shall return to a discussion of the anomalous skin effect and its relationship to the Fermi surface in the next section. For now, we shall examine in a little more detail some aspects of the relationship of electron transport properties to the anomalous skin effect.

When the electron mean free path is greater than the distance over which electric fields vary appreciably, an electron "remembers" the fields existing at other positions. Because of collisions, its memory decays exponentially as $e^{-t/\tau}$. The problem, then, is to compute the current density at a point $\mathbf{r}$ in such a way as to account for spatial variations of the electric field. One way is to consider the arrival of electrons at $\mathbf{r}$ coming from all other points $\mathbf{r}'$ where they suffered their last collision. (See Fig. 4.6.) The number of electrons in a volume element $dV(\mathbf{r}')$ reaching the point $\mathbf{r}$ without colliding is thus

$$dN(\mathbf{r}) \propto \frac{\exp(-|\mathbf{r}-\mathbf{r}'|/l)}{|\mathbf{r}-\mathbf{r}'|^2} dV(\mathbf{r}') \tag{4.39}$$

where the exponential decay factor gives the arrival probability at $\mathbf{r}$ from $\mathbf{r}'$ without collision, and the inverse square of $|\mathbf{r}-\mathbf{r}'|$ accounts for the isotropic distribution of initial directions of electrons at $\mathbf{r}'$. This is equivalent to all the electrons acting as if they had experienced an electric field $\mathbf{E}(\mathbf{r}', t')\exp[-(t-t')/\tau]$ and had suffered no collisions. We can then treat the problem as if the electrons suffered no collisions if we replace the actual field $\mathbf{E}(\mathbf{r}', t')$ by the effective field $\mathbf{E}(\mathbf{r}', t')\exp[-(t-t')/\tau]$. If we further

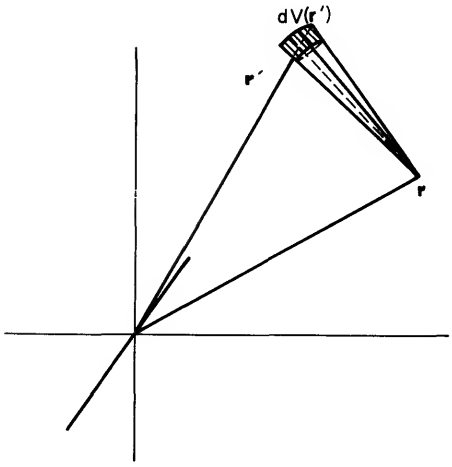

Fig. 4.6 The volume element $dV(\mathbf{r}')$ represents a general region in which some electrons are scattered and then travel to $\mathbf{r}$ without suffering further collisions.

assume that the field does not change rapidly with time, $\omega\tau < 1$, the contribution $d\mathbf{J}(\mathbf{r})$ to the current at $\mathbf{r}$ arising from electrons at $\mathbf{r}'$ is

$$d\mathbf{J}(\mathbf{r}) \propto \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\,\frac{\mathbf{E}\cdot(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\frac{\exp(-|\mathbf{r}-\mathbf{r}'|/l)}{|\mathbf{r}-\mathbf{r}'|^2}\,dV(\mathbf{r}') \tag{4.40}$$

where the first factor is just the unit vector from $\mathbf{r}'$ to $\mathbf{r}$, the second factor is the component of electric field in the direction of $(\mathbf{r}-\mathbf{r}')$, and the third factor is from Eq. (4.39).

Equation (4.40) can be easily integrated to get the current at $\mathbf{r}$ for the special case in which $\mathbf{E}$ is constant everywhere. Then, we must get $\mathbf{J}^{\text{ind}} = \sigma\mathbf{E}$. In this manner, it is found that the constant needed in Eq. (4.40) is $3\sigma/4\pi l$. Thus,

$$\mathbf{J}^{\text{ind}}(\mathbf{r}) = \frac{3\sigma}{4\pi l}\int \frac{(\mathbf{r}-\mathbf{r}')\,[\mathbf{E}\cdot(\mathbf{r}-\mathbf{r}')]\exp(-|\mathbf{r}-\mathbf{r}'|/l)\,dV(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^4} \tag{4.41}$$

which gives an explicit (but not completely rigorous) nonlocal relationship between induced current and electric field.

To solve a particular problem requires knowledge of the surface. For example, for specular reflection of electrons from surfaces, the electrons "remember" the fields from before striking the surface; diffuse surface reflection wipes out all previous field effects. These cases have been treated rigorously by Reuter and Sondheimer [2], and the results are generally in excellent agreement with the rather simple treatment given here.

Now let us return to a consideration of Figs. 4.2–4.4. We are now in a

position to see how the remaining boundaries between the various regions of Figs. 4.3 and 4.4 are determined.

The boundary between regions A and D is specified by the condition $l = \delta$, where the electron mean free path is given by $l = v\tau$. Using the classical limit for δ, we obtain from Eqs. (4.11), (4.20), and (4.24)

$$\omega\tau^3 \approx \lambda_p^2/v^2 \tag{4.42}$$

where

$$\lambda_p^2 \equiv c^2/\mu\omega_p^2 \tag{4.43}$$

and a factor of 2 has been ignored.

In the region of the anomalous skin effect, collisions are of little importance since the mean free path is greater than the skin depth. Because collisions are unimportant, the relaxation time is unimportant. Thus, if the frequency is increased until $\omega\tau > 1$, there is no change in properties when passing through $\omega\tau = 1$. The $\omega\tau = 1$ boundary between regions A and B does not extend into region D.

The significant change that occurs on increasing the frequency, starting in region D, is that during the time an electron spends in the skin layer, it experiences an increasing number of oscillations of the alternating electric field. The boundary between regions D and E is thus determined by $\delta/v = 1/\omega$. Using this condition together with Eq. (4.38) yields $\omega/\omega_p \approx v/c$.

Region E is called the region of anomalous reflection. Here, the electrons in the skin layer respond to the electric field as essentially free electrons. Region E differs only slightly from region B. Both regions are characterized by free-electron behavior. In region B, there is some absorption arising from occasional collisions with the lattice. In region E, collisions occur to a lesser extent and almost exclusively at the surface. Thus, the boundary between regions B and E corresponds to the mean free path being approximately equal to the penetration depth, the condition for which lattice collisions are of equal importance with surface collisions. In these regions, the penetration depth is approximately λ_p, which leads to $\tau = \lambda_p/v$.

4.4 Optical Properties and the Fermi Surface

We have already touched on one aspect of the relationship of optical properties to the Fermi surface. In Section 4.3, it was shown that in the region of the anomalous skin effect, only electrons with trajectories nearly parallel to the surface are effective in absorbing electromagnetic radiation. It is clear from these arguments and Eq. (4.38) that the surface resistance of a single crystal is different for different crystal planes. Thus, a measurement of the surface resistance for different orientations of a single crystal in the

anomalous region allows a determination of the shape of the Fermi surface.

Figure 4.7 shows the variation in surface impedance of copper as measured by Pippard, and Fig. 4.8 shows the Fermi surface of copper as inferred from the measurements. The distortion from a spherical Fermi surface such as to contact the Brillouin-zone boundaries in the ⟨111⟩ directions was one of the first experimental demonstrations of multiple-connectivity of a Fermi surface.

The difficulties in practice are several. It is necessary to prepare samples

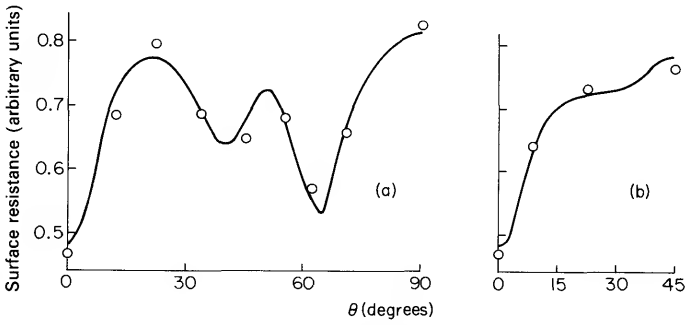

Fig. 4.7 A plot of surface resistance (arbitrary units) versus orientation for copper in the extreme anomalous region. The angle θ subtends from the (001) to the (110) directions. [From A. B. Pippard, *Roy. Soc. (London) Phil. Trans.* **A250**, 325 (1957).]

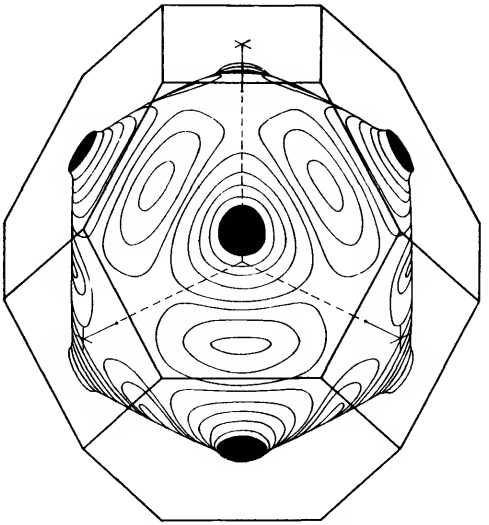

Fig. 4.8 The Fermi surface of copper. [From A. B. Pippard, *Roy. Soc. (London) Phil. Trans.* **A250**, 325 (1957).]

of sufficient purity that imperfections other than phonons do not limit the mean free path; otherwise, cooling the sample to liquid helium temperatures to achieve $l \gg \delta$ will be to no avail. The surfaces of the sample must also be microscopically smooth and strain-free so that the surface region is characteristic of the bulk material. On the theoretical side, the interpretation of measurements is especially difficult for Fermi surfaces composed of several distinct sheets. In most cases, it is necessary to have a reasonable model of the Fermi surface in order to interpret the experiments.

To determine the shape of the Fermi surface requires a single crystal. However, the surface impedance in the anomalous region is also useful even for a polycrystalline sample. It is directly related to the total area S_F of the Fermi surface and is not affected by anisotropies in scattering mechanisms. To see this it is only necessary to show (Problem 4.4) that the quantity σ/l can be obtained from measurements of the anomalous skin effect. Then, for a polycrystalline sample, a measurement of the anomalous skin effect gives $\bar{\sigma}/\bar{l}$. This, in turn, is directly related to the area of the Fermi surface according to

$$\bar{\sigma}/\bar{l} = e^2 S_F / 12\pi^3 \hbar \tag{4.44}$$

The latter follows directly from the nearly-free-electron model of a metal. It must be noted, though, that the area of the Fermi surface is the surface enclosing the electrons in **k**-space *excluding* regions defined by a zone boundary. The latter is discussed in more detail later in this section.

There is another source of information on the Fermi surface. The dielectric function determined from experiments (usually) in the near-infrared is related to the area of the Fermi surface [3]. We shall discuss in later chapters how the dielectric function is determined. For now, we write the dielectric function for a nearly-free-electron metal as

$$\hat{\varepsilon} = A - \frac{4\pi N e^2}{m_0} \frac{1}{\omega(\omega + i/\tau)} \tag{4.45}$$

The constant A may differ from unity because of local field effects and core polarization. It is also implicit that the relaxation time is isotropic.

It is sometimes possible to write the dielectric function in the form of Eq. (4.45) even in the presence of interband transitions if the interband transitions are spread over a sufficiently wide range. However, we shall neglect interband transitions.

By measuring the reflectance of metals at long wavelengths, before the onset of interband transitions, it is possible to determine $\hat{\varepsilon}$ and, hence, the optical effective mass m_0. We shall now show that m_0 is directly related to the area of the Fermi surface. Consider the interaction of electrons in a cubic metal with light polarized in the x direction. Each electron then responds to the electric vector of the light according to its effective mass.

Thus, $1/m_0$, which is an average inverse effective mass, is given by

$$\frac{1}{m_0} = \frac{2}{N} \sum_{\mathbf{k}_{\text{occ}}} \frac{1}{m^*_{xx}(\mathbf{k})} \tag{4.46}$$

where the summation is over all occupied states, the factor 2 accounts for the two spins per state $|\mathbf{k}\rangle$, and $m^*_{xx}(\mathbf{k})$ is a diagonal element of the effective mass tensor related to band structure by

$$1/m^*_{xx}(\mathbf{k}) = (1/\hbar^2)\, \nabla^2_{k_x} \mathscr{E}(\mathbf{k}) \tag{4.47}$$

Replacing the summation in (4.46) by an integration over all occupied states gives

$$1/m_0 = [2/(2\pi)^3 \hbar^2 N] \int \nabla^2_{k_x} \mathscr{E}(\mathbf{k})\, d\mathbf{k} \tag{4.48}$$

For a cubic material, the diagonal elements of the inverse effective mass tensor are equal. Since

$$\nabla_{\mathbf{k}}{}^2 \mathscr{E}(\mathbf{k}) = (\nabla^2_{k_x} + \nabla^2_{k_y} + \nabla^2_{k_z}) \mathscr{E}(\mathbf{k}) \tag{4.49}$$

Eq. (4.48) can be written as

$$1/m_0 = (1/4\pi^3 \hbar^2 N) \int \tfrac{1}{3} \nabla_{\mathbf{k}}{}^2 \mathscr{E}(\mathbf{k})\, d\mathbf{k} \tag{4.50}$$

Now, by using Green's theorem, Eq. (4.50) can be rewritten in terms of a surface integral as

$$1/m_0 = (1/12\pi^3 \hbar^2 N) \int_S \nabla_{\mathbf{k}} \mathscr{E}(\mathbf{k}) \cdot d\mathbf{S} \tag{4.51}$$

Substituting

$$\nabla_{\mathbf{k}} \mathscr{E}(\mathbf{k}) = \hbar \mathbf{v}_{\mathbf{k}} \tag{4.52}$$

in (4.51), we get

$$1/m_0 = (1/12\pi^3 \hbar N) \int_S \mathbf{v}_{\mathbf{k}} \cdot d\mathbf{S} \tag{4.53}$$

Note that a full band makes no contribution to $1/m_0$ because $\mathbf{v}_{\mathbf{k}} = 0$ at the boundaries of a Brillouin zone.

The electron concentration N can be written in terms of the area S_{F0} the Fermi surface would have if it were spherical:

$$N = \frac{2}{(2\pi)^3} \frac{4\pi k^3_{F0}}{3} = \frac{k_{F0} S_{F0}}{12\pi^3} \tag{4.54}$$

Thus, for metals in which all the conduction electrons are contained within the first Brillouin zone, the surface over which the integral (4.53) ranges is just the Fermi surface, and

$$1/m_0 = (1/\hbar k_{F0} S_{F0}) \int_{S_F} \mathbf{v}_F \cdot d\mathbf{S}_F \tag{4.55}$$

The Fermi velocity averaged over the Fermi surface is

$$\langle v_F \rangle = (1/S_F) \int_{S_F} \mathbf{v}_F \cdot d\mathbf{S}_F \tag{4.56}$$

hence

$$1/m_0 = (1/\hbar k_{F0})(S_F/S_{F0}) \langle v_F \rangle \tag{4.57}$$

This last equation provides a rather direct connection between optical properties and the Fermi surface. Of course, only a quite simple case has been treated. Optical properties are generally quite complicated and we cannot expect to be able simply to use optical reflectance measurements to study Fermi surfaces. What is important is the ideas involved, to see that, at least in principle, there is a relationship between optical properties and Fermi surfaces, and that the relationship is conceptually quite simple in limiting cases.

We have yet to generalize the analysis of optical properties and Fermi surfaces and see the range and limitations of its applicability to more complex metals. Already, though, it is clear that, at least in principle, the results should be of considerable interest. Optical measurements can be made on large classes of metals, alloys, and liquid metals which are not amenable to the usual methods of studying Fermi surfaces (de Haas–van Alphen effect, cyclotron resonance, etc.). Because of rather severe restrictions on the electron mean free path, the latter are usually limited to samples of high purity and low crystal structure defect concentrations.

We shall now resume the analysis of Fermi surfaces in terms of optical properties, blithely ignoring any complicating factors.

For metals in which the Fermi surface extends through more than one Brillouin zone, the surface over which the integral (4.53) ranges is just that surface which encloses all the electrons in an extended zone scheme. This is illustrated in Fig. 4.9 for a two-dimensional case.

Those parts of the surface which lie on the Brillouin zone faces are not part of the Fermi surface; electrons in those states have energies less than $\mathscr{E}_F$. However, because $\mathbf{v}_k = 0$ on the Brillouin-zone faces, these regions make no contribution to the integral in (4.53). Thus, even for disconnected Fermi surfaces, the optical mass is given by Eqs. (4.55) and (4.57).

There is one further step to be made in order to simplify the results and

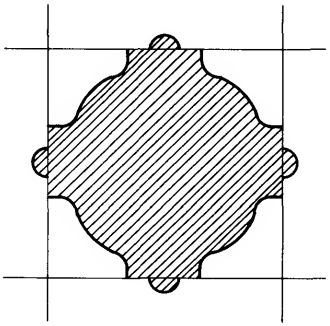

Fig. 4.9 Two-dimensional representation of the intersection of the Fermi surface with the Brillouin-zone boundaries.

make them more useful. We must be able to separate the contributions from S_F and $\langle v_F \rangle$ in Eq. (4.57). This requires more than optical measurements. One way is by means of electronic specific heat measurements. The electronic specific heat is proportional to the density of states at the Fermi surface, and the density of states is proportional to an effective mass. Following through in much the same way as we derived an expression for the optical mass, it is possible to derive an expression for a thermal mass m_t. The crucial difference between the optical mass and the thermal mass is that the averages are taken in different ways. One is an average of the reciprocal effective mass of band theory; the other is a direct average of the effective mass of band theory. Expressed in terms of velocities averaged over the Fermi surface,

$$1/m_0 = (1/\hbar k_{F0})(S_F/S_{F0})\langle v_F \rangle \tag{4.58}$$

$$m_t = \hbar k_{F0}(S_F/S_{F0})\langle 1/v_F \rangle \tag{4.59}$$

The ratio of these masses is

$$m_t/m_0 = (S_F/S_{F0})^2 \langle v_F \rangle \langle 1/v_F \rangle \tag{4.60}$$

The Schwartz inequality requires

$$\langle v_F \rangle \langle 1/v_F \rangle \geq 1 \tag{4.61}$$

Hence,

$$m_t/m_0 \geq (S_F/S_{F0})^2 \tag{4.62}$$

We now consider what sort of information about the Fermi surface can be obtained from Eq. (4.62) if the optical and thermal masses have been determined from experiment. For simple metals, in which all the conduction electrons are contained within the first Brillouin zone, there are two possibilities:

1. The Fermi surface is spherical. Then, v_F is constant, the equality holds in (4.62), and $m_t/m_0 = 1$.

2. The Fermi surface is not spherical. Then, since a sphere has the minimum surface area per unit volume, $S_F > S_{F0}$, and $m_t/m_0 > 1$.

If the Fermi surface intersects the Brillouin-zone faces, it may be that $S_F < S_{F0}$. As was pointed out when discussing Fig. 4.9, the surface area on the Brillouin-zone faces is not part of the Fermi surface. If the area of contact is large enough to offset the restriction (4.61) imposed by the Schwartz inequality, then it is possible that $m_t/m_0 < 1$.

It is clear that m_0 is related to the Fermi surface. However, there are difficulties in attempting quantitative studies of Fermi surfaces. It is more likely that the ideas mentioned here will be more easily applied to the study of *changes* in the Fermi surface upon alloying. This was first done by Cohen and Heine in an attempt to reconcile the Hume-Rothery rules for alloy phases with modern knowledge of Fermi surfaces. Even in these sorts of applications, there are difficulties. For example, it is not always clear what is the proper choice for the electron concentration N. To apply the Hume-Rothery rules to alloys containing Fe, Co, and Ni, it is necessary to assume that these atoms contribute no electrons to the concentration N. Thus, Fe–Al acts as if it has an electron/atom ratio of $\frac{3}{2}$. Ambiguity in N is also likely to be a problem with multivalent solutes added to monovalent metals.

There are other complications. These include the following:

1. The lattice spacing changes on alloying, producing changes in band structure.
2. If the alloys are ordered, the Brillouin-zone structure differs from the parent metal.
3. The integral in Eq. (4.55) is not very sensitive to the shape of the Fermi surface if, as in the noble metals, the Fermi surface only slightly contacts the Brillouin-zone faces. That is, the integral does not differ much from what it would be if the Fermi surface were a sphere.

Finally, there are corrections for many-body effects. In an ideal crystal, many-body effects arise from electron–phonon and electron–electron interactions. These interactions lead to a quasiparticle description of the electron.

A quasielectron is visualized as follows. Imagine an electron moving through a solid. As it moves, it attracts neighboring ions and repels other electrons. Thus, the electron surrounds itself with a positive screening cloud. The polarization cloud tends to move with the electron. Of course, the same ions cannot move along with the electron; the cloud is made up of just neighboring ions and these neighbors change as the electron moves. The system of real electron plus surrounding cloud is called a quasielectron. In an ionic crystal, where the electron–ion interaction is particularly strong, the composite particle is called a polaron.

The interaction between two quasielectrons is a screened interaction

and is much weaker than the bare Coulomb interaction. Thus, the quasielectrons behave much like free electrons but with a renormalized mass. That is, the quasielectron acts as a free electron whose mass is greater than the true electron mass because it carries along a cloud of other particles.

The thermal mass of an electron as determined from specific heat measurements includes effects from electron–phonon and electron–electron interactions. Thus, the thermal mass is increased over the value calculated from one-electron band theory. At optical frequencies, the lattice can no longer follow the motion of the electrons, so the electron–phonon (i.e., electron–ion) interaction is not effective. However, the electron–electron interaction may still be of some importance and modify the optical mass. Since many-body effects act in different ways to modify the thermal and optical masses, comparisons are made more difficult and deductions concerning the nature of the Fermi surface may be almost hopeless.

PROBLEMS

4.1 Show that if $\omega\tau \ll 1$, the reflectivity of a metal is given by the Hagen–Rubens relation:

$$R \approx 1 - (2\omega/\pi\sigma)^{1/2}$$

4.2 Show that the reflectivity of a metal in the relaxation region is given by

$$R \approx 1 - (2/\omega_p\tau)$$

4.3 Carry through an analysis of the skin effect in terms of a complex wave vector $\hat{\mathbf{q}}$. How is the imaginary part of $\hat{\mathbf{q}}$ related to the classical skin depth?

4.4 Read R. G. Chambers, *Proc. Roy. Soc.* **A215**, 481 (1956). Discuss the determination of σ/l from anomalous skin effect measurements.

4.5 Show that when $\omega\tau \ll 1$, the complex wave vector is

$$\hat{\mathbf{q}} = (1 + i)/\delta_{\text{Cl}}$$

so that in the region of the classical skin effect, electromagnetic waves are damped as "quickly" as they propagate.

4.6 Show that the surface resistance is independent of the scattering time in the region of the anomalous skin effect.

4.7 Derive a relationship between the reflectivity and the skin depth in the region of the anomalous skin effect.

4.8 Show how the surface resistance in the anomalous region is related to the curvature of the Fermi surface.

4.9 Show that the surface conductivity in the anomalous region is proportional to $(\sigma/l)^{1/3}$ so long as $(l/\delta)^2 \gg (1 + \omega^2\tau^2)^{3/2}$. Thus $\omega\tau$ can be greater than one, if l/δ is sufficiently large, and relaxation effects are still negligible.

4.10 Read and discuss the following articles: (a) H. E. Bennett *et al.*, *Phys. Rev.* **165**, 755 (1968). (b) N. V. Smith, *Phys. Rev.* **2**, 2840 (1970). (c) M. Thèye, *Phys. Rev.* **2**, 3060 (1970).

4.11 Read W. P. Dumke, *Phys. Rev.* **124**, 1813 (1961). Discuss the differences and similarities in the classical and quantum theories of free-carrier absorption.

4.12 Derive Eq. (4.59) for the thermal mass.

4.13 Is it necessary to impose any restrictions on the magnitude of $\omega\tau$ in order to determine m_0 from reflectance measurements? Does it matter if τ is frequency-dependent?

4.14 What is the relationship between S_{F0} and the area for the Fermi surface of a perfect free electron gas of the same density? Is S_{F0} related to the effective mass m^* of band structure?

REFERENCES AND FURTHER READING

1. A. B. Pippard, *Proc. Roy. Soc.* **A191**, 385 (1947).
2. G. E. H. Reuter and E. H. Sondheimer, *Proc. Roy. Soc.* **A195**, 336 (1948).
3. M. H. Cohen and V. Heine, *Advan. Phys.* **7**, 395 (1958).
4. H. B. G. Casimir and J. Ubbink, *Philips Tech. Rev.* **28**, 300 (1967).
5. A. B. Pippard, "Dynamics of Conduction Electrons," Gordon and Breach, New York (1965).
6. L. G. Schulz, *Advan. Phys.* **6**, 102–144 (1957).
7. W. A. Harrison and M. B. Webb, eds., "The Fermi Surface," Wiley, New York, 1960.
8. M. H. Cohen, Optical Constants, Heat Capacity and the Fermi Surface, *Phil. Mag.* **3**, 762–775 (1958).
9. M. P. Givens, Optical Properties of Metals, *Solid State Phys.* **6** (1958).
10. A. V. Sokolov, "Optical Properties of Metals," Blackie, London, 1965.
11. T. S. Moss, "Optical Properties of Semiconductors," Butterworths, London and Washington, D.C., 1959.
12. J. N. Hodgson, "Optical Absorption and Dispersion in Solids," Chapman and Hall, London, 1970.

Chapter 5

INTERBAND TRANSITIONS

In this chapter, we develop the quantum mechanical theory of direct and indirect transitions between bands. We then examine the optical properties of some representative solids in terms of their electronic band structure. Next, we consider exciton absorption. Excitons, which are quasiparticles consisting of an electron–hole pair, lie outside the one-electron band picture, but they are too important to dismiss and they can be conveniently discussed with reference to a one-electron band diagram. Finally a discussion of photoemission illustrates the power as well as the difficulties of this relatively new method of studying band structure. It also introduces the new and still somewhat controversial concept of nondirect transitions.

5.1 Periodic Perturbation

We want to consider the effect of an electromagnetic field in exciting an electron from an initial state ϕ_i to some other state ϕ_j.

Consider a field producing a perturbation of the form

$$H' = \tfrac{1}{2} V(\mathbf{r}) (e^{i\omega t} + e^{-i\omega t}) \tag{5.1}$$

From first-order perturbation theory (see Section 3.5), the rate of change of

amplitude for finding the electron in state ϕ_j is

$$i\hbar\, da_j/dt = V_{ji}(e^{i(\omega_{ji}+\omega)t} + e^{i(\omega_{ji}-\omega)t}) \tag{5.2}$$

where

$$V_{ji} = \int \phi_j^* [V(\mathbf{r})/2]\, \phi_i\, d\mathbf{r} \tag{5.3}$$

Integrating Eq. (5.2) from time $t = 0$, when the perturbation is first turned on, we obtain

$$a_j = \frac{V_{ji}}{\hbar}\left(\frac{1 - e^{i(\omega_{ji}+\omega)t}}{\omega_{ji} + \omega} + \frac{1 - e^{i(\omega_{ji}-\omega)t}}{\omega_{ji} - \omega}\right) \tag{5.4}$$

The amplitude a_j is appreciable only if $\omega \approx \omega_{ji}$. If we consider only the absorption of energy, so that $\omega_{ji} = (\mathscr{E}_j - \mathscr{E}_i)/\hbar$ is positive, then only the second term on the RHS of Eq. (5.4) need be considered. Thus the probability of finding an electron excited to the state ϕ_j is

$$|a_j|^2 = \frac{4|V_{ji}|^2}{\hbar^2}\,\frac{\sin^2[\frac{1}{2}(\omega_{ji} - \omega)t]}{(\omega_{ji} - \omega)^2} \tag{5.5}$$

In real systems, we must consider a range of final states having energies near $\mathscr{E}_j$. Thus if $\rho(\mathscr{E}_j)\, d\mathscr{E}_j = \rho(\mathscr{E}_j)\hbar\, d\omega_{ji}$ is the number of states having energies between $\mathscr{E}_j$ and $\mathscr{E}_j + d\mathscr{E}_j$, then the probability of finding the electron excited into one of these states is

$$\int |a_j|^2 \rho(\mathscr{E}_j)\, d\mathscr{E}_j = \frac{4}{\hbar}\int |V_{ji}|^2 \frac{\rho(\mathscr{E}_j)\, \sin^2[\frac{1}{2}(\omega_{ji} - \omega)t]\, d\omega_{ji}}{(\omega_{ji} - \omega)^2} \tag{5.6}$$

Because of the factor $1/(\omega_{ji} - \omega)^2$, the only significant contribution to the integral comes from frequencies $\omega_{ji} \approx \omega$. In the narrow band of frequencies near $\omega_{ji} = \omega$, we can assume that $\rho(\mathscr{E}_j)$ and $|V_{ji}|^2$ are slowly varying functions of $\mathscr{E}_j$. They can be taken outside the integral and Eq. (5.6) rewritten as

$$\int |a_j|^2 \rho(\mathscr{E}_j)\, d\mathscr{E}_j = \frac{4|V_{ji}|^2 \rho(\mathscr{E}_j)}{\hbar}\int_{-\infty}^{\infty} \frac{\sin^2[\frac{1}{2}(\omega_{ji} - \omega)t]\, d\omega_{ji}}{(\omega_{ji} - \omega)^2} \tag{5.7}$$

The region of integration has been extended to include all frequencies, $-\infty < \omega < \infty$. This is possible because only the region near $\omega_{ji} = \omega$ contributes. Carrying out the integration, we obtain

$$\int |a_j|^2 \rho(\mathscr{E}_j)\, d\mathscr{E}_j = (2\pi t/\hbar)\, |V_{ji}|^2 \rho(\mathscr{E}_j) \tag{5.8}$$

Thus, the probability of finding the electron in an excited state increases linearly with time.

That significant contributions to the integral in Eq. (5.7) come only

for frequencies $\omega_{ji} \approx \omega$ is an expression of the conservation of energy for the unperturbed states. It is not an exact restriction except in the limit $t \to \infty$. All this can be seen most simply from an inspection of a plot of $[\sin^2(\frac{1}{2}\omega t)]/\omega^2$ versus ω as shown in Fig. 5.1. The central peak, of height $t^2/4$, becomes sharper with increasing time. In the limit $t \to \infty$, only transitions for which $\omega = 0$ contribute. The area under the curve is proportional to the probability of finding an electron in the excited state. Since the peak height increases as t^2, whereas the width decreases as t^{-1}, the area under the curve increases linearly with time as we expect.

Recognizing that only transitions for which $\omega_{ji} \approx \omega$ are possible, we can rewrite Eq. (5.8) as

$$|a_j|^2 = (2\pi t/\hbar)\,|V_{ji}|^2 \;\; \delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.9}$$

The use of a δ function specifies the conservation of energy requirement, but it only has meaning as part of an integrand. Thus, Eq. (5.9) for $|a_j|^2$ can be used only when multiplied by the appropriate density of states and an integration performed.

The transition rate from state ϕ_i to state ϕ_j is given by

$$W_{ji} = (d/dt)\,|a_{ji}|^2 = (2\pi/\hbar)\,|V_{ji}|^2\,\delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.10}$$

5.2 Direct Interband Transitions

We are now ready to consider interband transitions. We consider first a solid having simple parabolic bands with maxima and minima at $\mathbf{k} = 0$,

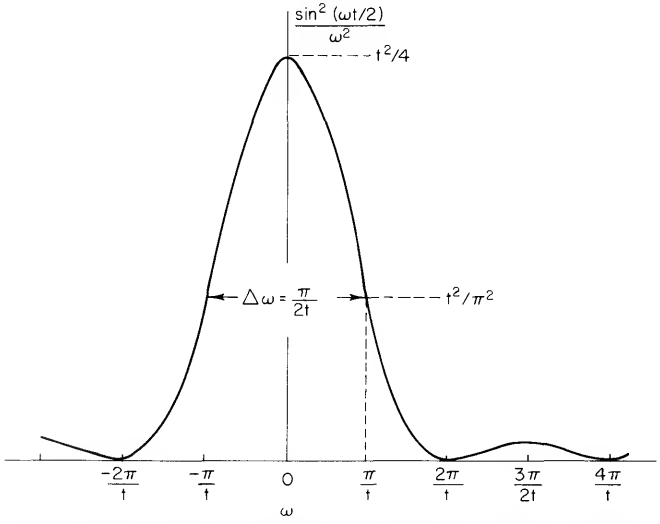

Fig. 5.1 Relative probability of finding an electron excited to a state having an energy $\hbar\omega$ greater than the initial state as determined from first-order perturbation theory.

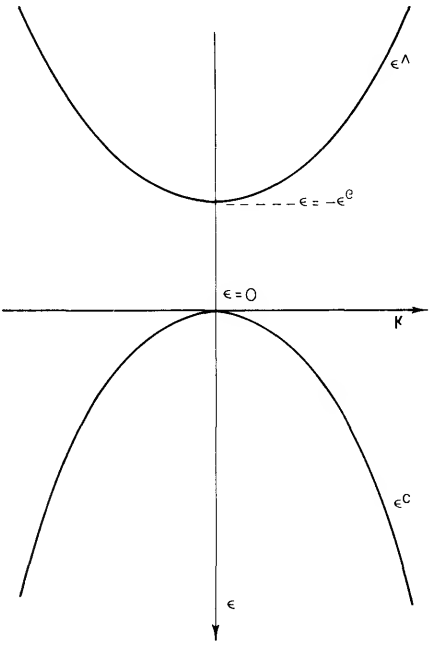

Fig. 5.2 Schematic energy band diagram.

as shown in Fig. 5.2. The zero of energy is taken at the bottom of the conduction band. Then, the energy of an electron in the conduction band is

$$\mathscr{E}_c(\mathbf{k}'') = \hbar^2(k'')^2/2m_e \tag{5.11}$$

and in the valence band is

$$\mathscr{E}_v(\mathbf{k}') = -\mathscr{E}_G - [\hbar^2(k')^2/2m_h] \tag{5.12}$$

where m_e and m_h are, respectively, effective masses for electrons in the conduction band and holes in the valence band.

The light wave can be described by the vector potential

$$\mathbf{A} = \tfrac{1}{2}A\mathbf{a}_0\{[\exp i(\mathbf{q}\cdot\mathbf{r} - \omega t)] + \exp -i(\mathbf{q}\cdot\mathbf{r} - \omega t)\} \tag{5.13}$$

where $\mathbf{a}_0$ is the unit polarization vector and $\mathbf{q}$ is the photon wave vector.

For transverse fields, we can work in the Coulomb gauge, $\nabla\cdot\mathbf{A} = 0$, so that, as shown in Appendix E, the interaction between photons and electrons can be described by the perturbation Hamiltonian

$$H' = -(ie\hbar/mc)\,\mathbf{A}\cdot\nabla \tag{5.14}$$

Substituting from Eq. (5.13) yields

$$H' = -(ie\hbar A/2mc)\{[\exp i(\mathbf{q}\cdot\mathbf{r} - \omega t)] + \exp -i(\mathbf{q}\cdot\mathbf{r} - \omega t)\}(\mathbf{a}_0\cdot\nabla) \tag{5.15}$$

Now, consider a transition from a state i in the valence band to state j in the conduction band. We describe these states by the Bloch functions

$$\psi_i = \Omega^{-1/2}(\exp i\mathbf{k}' \cdot \mathbf{r})\, u_v(\mathbf{r}, \mathbf{k}')$$
$$\psi_j = \Omega^{-1/2}(\exp i\mathbf{k}'' \cdot \mathbf{r})\, u_c(\mathbf{r}, \mathbf{k}'') \tag{5.16}$$

We want to determine the transition rate as given by Eq. (5.10). For this, we need to know the perturbation matrix element. If we consider only absorption processes, then we need retain only the first term in Eq. (5.15). The spatial part of the matrix element is then

$$\begin{aligned} H'_{ji}(\mathbf{r}) &= \int \psi_j^*[H'(\mathbf{r})]\, \psi_i \, d\mathbf{r} \\ &= \int \psi_j^* \left[-\frac{ie\hbar A}{2mc} (\exp i\mathbf{q} \cdot \mathbf{r})\, \mathbf{a}_0 \cdot \nabla \right] \psi_i \, d\mathbf{r} \\ &= -\frac{ie\hbar A}{2mc\Omega} \int u_c^*(\mathbf{r}, \mathbf{k}'')(\exp -i\mathbf{k}'' \cdot \mathbf{r}) \exp i\mathbf{q} \cdot \mathbf{r})\, \mathbf{a}_0 \cdot \nabla (\exp i\mathbf{k}' \cdot \mathbf{r})\, u_v(\mathbf{r}, \mathbf{k}')\, d\mathbf{r} \\ &= -\frac{ie\hbar A}{2mc\Omega} \int u_c^*(\mathbf{a}_0 \cdot \nabla u_v + i u_v \mathbf{a}_0 \cdot \mathbf{k}') \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot \mathbf{r}]\, d\mathbf{r} \end{aligned} \tag{5.17}$$

Because of the periodicity of the functions u_c and u_v, the integral can be expressed as a sum of integrals over unit cells. Referring to Fig. 5.3, we see that

$$\exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot \mathbf{r}] = \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot (\mathbf{R}_n + \mathbf{r}')] \tag{5.18}$$

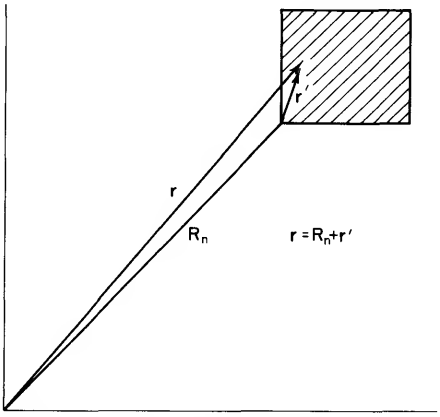

Fig. 5.3 The position $\mathbf{r}$ can be expressed as the vector sum of the distance $\mathbf{R}_n$ to the nth unit cell and the distance $\mathbf{r}'$ from the cell origin to the point $\mathbf{r}$.

Thus, $H'_{ji}(\mathbf{r})$ can be expressed as

$$H'_{ji}(\mathbf{r}) = \sum_n \int_{\text{cell}} f(\mathbf{r}', \mathbf{k}', \mathbf{k}'', \mathbf{q}) \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot \mathbf{R}_n] \, d\mathbf{r}'$$

$$= \sum_n \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot \mathbf{R}_n] \int_{\text{cell}} f(\mathbf{r}', \mathbf{k}', \mathbf{k}'', \mathbf{q}) \, d\mathbf{r} \qquad (5.19)$$

The summation $\Sigma_n \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot \mathbf{R}_n]$ is approximately equal to zero unless $(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') = \mathbf{K}$, where $\mathbf{K}$ is a reciprocal lattice vector. (See Problem 5.1.) Working in the reduced zone scheme, we take $\mathbf{K} = 0$, so that

$$\mathbf{k}'' = \mathbf{k}' + \mathbf{q} \qquad (5.20)$$

Since $\mathbf{q}$ for a light is a small fraction of a reciprocal lattice vector, we may take

$$\mathbf{k}'' = \mathbf{k}', \qquad \mathbf{q} = 0 \qquad (5.21)$$

Transitions of this kind are called direct, or vertical, interband transitions. With these assumptions,

$$\sum_n \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'') \cdot \mathbf{R}_n] = N_c \qquad (5.22)$$

where N_c is the number of unit cells in the crystal. Taking Δ as the volume of a unit cell and Ω as the volume of the crystal, so that $N_c = \Omega/\Delta$, the perturbation matrix element is

$$H'_{ji} = -\frac{ie\hbar A}{2mc\Delta} \int_{\text{cell}} u_c^*(\mathbf{k}') [\mathbf{a}_0 \cdot \nabla u_v(\mathbf{k}') + i\mathbf{a}_0 \cdot \mathbf{k}' u_v(\mathbf{k}')] \, d\mathbf{r} \qquad (5.23)$$

The second term in the integral is zero since Bloch functions, and the periodic parts of Bloch functions, are orthonormal. Should it happen that the first term is also zero because of the particular form of symmetry of u_c and u_v, it will be necessary to reconsider the restrictions on $\mathbf{k}'$ and $\mathbf{k}''$. That is, we will have to consider what are called forbidden direct transitions. These are transitions for which it is necessary to recognize that $\mathbf{q} \neq 0$, and hence Eq. (5.20) rather than (5.21) must be used. However, for now, we shall consider the case in which the first term in H'_{ji} in Eq. (5.23) is nonzero.

Introducing the momentum operator

$$\mathbf{p} = -i\hbar\nabla \qquad (5.24)$$

and the matrix element of the momentum operator

$$\mathbf{p}_{ji} = -(i\hbar/\Delta) \int_{\text{cell}} u_c^* \nabla u_v \, d\mathbf{r} \qquad (5.25)$$

into Eq. (5.23), we get for allowed direct transitions

$$H'_{ji} = -(ie\hbar A/2mc\Delta) \int_{\text{cell}} u_c^* \mathbf{a}_0 \cdot \nabla u_v \, d\mathbf{r}$$

$$= (eA/2mc)\, \mathbf{a}_0 \cdot \mathbf{p}_{ji} \tag{5.26}$$

Substituting this into Eq. (5.10), we find that the transition rate of an electron from state ψ_i to state ψ_j is

$$W_{ji} = (\pi e^2 A^2/2m^2\hbar c^2)\, |\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2\, \delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.27}$$

Of course, in a real solid, electrons do not make transitions simply between two Bloch states. We must consider a range of states between which transitions can take place. The density of states in $\mathbf{k}$-space (including spin) is $(4\pi^3)^{-1}$ per unit volume of crystal. Multiplying this by W_{ji} and integrating over all allowed values of $\mathbf{k}$ gives the total transition rate for electrons excited from the valence band to the conduction band. Note that in the integration we need consider only the density of states in the valence band. The density of states in the conduction band need not be considered because for direct transitions, $\mathbf{k}' = \mathbf{k}''$, and there is a one-to-one correspondence between states in the valence band and states in the conduction band. Thus, the total transition rate from valence band to conduction band is

$$W_{cv} = \frac{\pi e^2 A^2}{2m^2\hbar c^2} \int \frac{d\mathbf{k}}{4\pi^3} |\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2\, \delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.28}$$

Let us now eliminate the factor A^2 in Eq. (5.28). Comparing Eq. (5.13) with Eq. (E.18) from Appendix E, we see that

$$A^2 = 8\pi\hbar c^2/\varepsilon_1\omega \tag{5.29}$$

Thus,

$$W_{cv} = (e^2/\varepsilon_1\pi m^2\omega) \int d\mathbf{k}\, |\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2\, \delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.30}$$

Using Eq. (3.106), the imaginary part of the dielectric function is

$$\varepsilon_2(\omega) = (e^2/\pi m^2\omega^2) \int d\mathbf{k}\, |\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2\, \delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.31}$$

We shall discuss the dielectric function and its relationship to the electronic band structure in considerable detail in the next section. For the moment, let us calculate a numerical estimate of the absorption coefficient to compare with experiment.

The factor $|\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2$ appearing in Eq. (5.30) is dependent on $\mathbf{k}$. However, if we simply use the conditions of spherical symmetry implicit in Eqs. (5.11) and (5.12), we can define a mean square momentum operator $\langle p_{ji}^2 \rangle$ by

$$\langle |\mathbf{a}_0 \cdot \mathbf{p}_{ji}(\mathbf{k})|^2 \rangle = \langle p_{ji}^2(k_x) \rangle + \langle p_{ji}^2(k_y) \rangle + \langle p_{ji}^2(k_z) \rangle = 3\langle p_{ji}^2 \rangle \tag{5.32}$$

where $\langle p_{ji}^2 \rangle$ is assumed constant over the range of allowed $\mathbf{k}$ values. Then, Eq. (5.30) becomes

$$W_{cv} \approx \frac{3e^2\langle p_{ji}^2 \rangle}{\varepsilon_1 \pi m^2 \omega} \int d\mathbf{k}\, \delta(\mathscr{E}_{ji} - \hbar\omega) \tag{5.33}$$

Assuming parabolic bands as given by Eqs. (5.11) and (5.12),

$$\mathscr{E}_{ji} = \mathscr{E}_c - \mathscr{E}_v = \mathscr{E}_G + (\hbar^2 k^2 / 2m_r) \tag{5.34}$$

where m_r is the reduced mass defined by

$$1/m_r = (1/m_e) + (1/m_h) \tag{5.35}$$

Now, the integration over $\mathbf{k}$-space in Eq. (5.33) can be converted to an integration over $\mathscr{E}_{ji}$. We have, from

$$d\mathbf{k} = 4\pi k^2\, dk \tag{5.36a}$$

in spherical coordinates, and from Eq. (5.34), that

$$d\mathbf{k} = (4\pi/\hbar^3) m_r [2m_r(\mathscr{E}_{ji} - \mathscr{E}_G)]^{1/2}\, d\mathscr{E}_{ji} \tag{5.36b}$$

Thus, Eq. (5.31) becomes

$$W_{cv} = \frac{6e^2(2m_r)^{3/2}\langle p_{ji}^2 \rangle}{m^2\hbar^3\varepsilon_1\omega} \int (\mathscr{E}_{ji} - \mathscr{E}_G)^{1/2}\, \delta(\mathscr{E}_{ji} - \hbar\omega)\, d\mathscr{E}_{ji}$$

$$= \frac{6e^2(2m_r)^{3/2}\langle p_{ji}^2 \rangle (\hbar\omega - \mathscr{E}_G)^{1/2}}{m^2\hbar^3\varepsilon_1\omega} \tag{5.37}$$

Defining an average oscillator strength

$$f = 2\langle p_{ji}^2 \rangle / m\hbar\omega \tag{5.37}$$

analogous to the rigorous definition of Eq. (3.82), we can rewrite Eq. (5.37) as

$$W_{cv} = 3e^2(2m_r)^{3/2} f (\hbar\omega - \mathscr{E}_G)^{1/2} / m\hbar^2\varepsilon_1 \tag{5.39}$$

We can now make an estimate of the absorption coefficient for a material in which direct transitions are allowed. Assuming $m_e = m_h = m$, so that $2m_r = m$, and using Eqs. (3.104) and (5.39), we get

$$\alpha_{cv} = 3e^2 f [m(\hbar\omega - \mathscr{E}_G)]^{1/2} / \hbar^2 nc \tag{5.40}$$

If we now assume an oscillator strength of unity, take $n = 4$, which is typical of many semiconductors, and calculate the absorption coefficient for $\hbar\omega - \mathscr{E}_G = 0.01$ eV, we find $\alpha_{cv} \approx 2 \times 10^4\ \text{cm}^{-1}$. This is, in fact, in good agreement with the values of α for direct transitions in Ge near threshold. However, the agreement is not obtained simply by a direct measurement of α near threshold. This is because the formation of excitons causes α to tend to a constant value near threshold. Excitons arise from the Coulomb interaction between the excited electron and the hole left behind, something we have neglected, and thus our present results are not rigorous at energies just above threshold. However, corrections can be made for exciton formation, and when that is done, good agreement is obtained between theory and experiment. We shall discuss excitons in greater detail in later sections.

5.3 Joint Density of States and Critical Points

Equation (5.31) for $\varepsilon_2(\omega)$ arising from direct interband transitions can be rewritten using the property of the δ function:

$$\delta[g(x)] = \sum_n |g'(x_n)|^{-1}\, \delta(x - x_n); \qquad [g(x_n) = 0, \quad g'(x_n) \neq 0] \tag{5.41}$$

In the case of Eq. (5.31), we have

$$\delta[\mathscr{E}_{ji}(\mathbf{k}) - \hbar\omega] = \sum_n [\delta(\mathbf{k} - \mathbf{k}_n)/|\nabla_{\mathbf{k}}\mathscr{E}_{ji}(\mathbf{k})|] \tag{5.42}$$

so that

$$\varepsilon_2(\omega) = \frac{e^2}{\pi m^2 \omega^2} \int d\mathbf{k} \sum_n \frac{|\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2\, \delta(\mathbf{k} - \mathbf{k}_n)}{|\nabla_{\mathbf{k}}\mathscr{E}_{ji}(\mathbf{k})|} \tag{5.43}$$

where $\mathbf{k}_n$ represents all the values of $\mathbf{k}$ such that $\mathscr{E}_{ji}(\mathbf{k}_n) = \hbar\omega$. The points $\mathbf{k}_n$ thus define a surface in $\mathbf{k}$-space. We can now describe $\varepsilon_2(\omega)$ in terms of an integration over that surface in $\mathbf{k}$-space. Namely,

$$\varepsilon_2(\omega) = \frac{e^2}{\pi m^2 \omega^2} \int |\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2 \frac{dS}{|\nabla_{\mathbf{k}}\mathscr{E}_{ji}(\mathbf{k})|} \tag{5.44}$$

where dS is a surface element in $\mathbf{k}$-space such that

$$\mathscr{E}_{ji}(\mathbf{k}) = \hbar\omega \tag{5.45}$$

To see more clearly how one makes the transformation from Eq. (5.43) to Eq. (5.44), we shall consider in more detail the surface defined by Eq. (5.45). We start by considering an electron energy band.

An expression for the density of states $\rho(\mathscr{E})$ for an electron energy band can be derived as follows. The number of states per unit volume with energy

between $\mathscr{E}$ and $\mathscr{E} + d\mathscr{E}$ is $\rho(\mathscr{E})\, d\mathscr{E}$. Now, consider two surfaces of constant energy in **k**-space, one having energy $\mathscr{E}$ and the other having energy $\mathscr{E} + d\mathscr{E}$. Two such surfaces are illustrated in Fig. 5.4. Take a cylindrical volume element of area dS on the constant-energy surfaces and of height $dk_\perp$, where $dk_\perp$ is the perpendicular distance between the two-constant energy surfaces in **k**-space. Then,

$$\rho(\mathscr{E})\, d\mathscr{E} = \int_{\mathbf{k}(E)}^{\mathbf{k}(\mathscr{E}+d\mathscr{E})} \rho(\mathbf{k})\, d\mathbf{k}$$

$$= \int \rho(\mathbf{k})\, dS_\varepsilon\, dk_\perp \tag{5.46}$$

where $\rho(\mathbf{k})$ is the density of states in **k**-space. Now, $\rho(\mathbf{k}) = 1/4\pi^3$ including spin, and

$$d\mathscr{E} = \nabla_{\mathbf{k}}\mathscr{E} \cdot d\mathbf{k}$$

$$= |\nabla_{\mathbf{k}}\mathscr{E}|\, dk_\perp \tag{5.47}$$

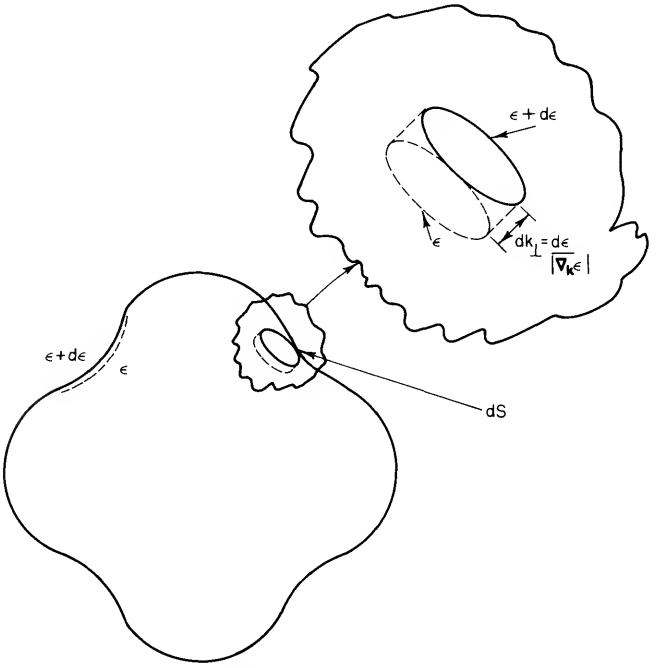

Fig. 5.4 Element of area dS on a constant-energy surface in **k**-space. The element of volume between two surfaces of constant energy $\mathscr{E}$ and $\mathscr{E} + d\mathscr{E}$ is seen from the insert to be $dV = dS\, d\mathscr{E}/|\nabla_{\mathbf{k}}\mathscr{E}|$.

Substituting these into Eq. (5.46), we find

$$\rho(\mathscr{E}) = (1/4\pi^3) \int \frac{dS_{\mathscr{E}}}{|\nabla_{\mathbf{k}} \mathscr{E}|} \tag{5.48}$$

As we would expect, the density of states is highest for flat bands, that is, as $|\nabla_{\mathbf{k}} \mathscr{E}| \to 0$.

Assume that the electron energy $\mathscr{E}_i(\mathbf{k})$ in band i can be expanded to second order about $\mathbf{k}_0$:

$$\mathscr{E}_i(\mathbf{k}) = \mathscr{E}_i(\mathbf{k}_0) + \boldsymbol{\alpha}^i \cdot (\mathbf{k} - \mathbf{k}_0) + (\mathbf{k} - \mathbf{k}_0) \cdot \boldsymbol{\beta}^i \cdot (\mathbf{k} - \mathbf{k}_0) \tag{5.49}$$

Equation (5.49) defines an electron energy band in the neighborhood of $\mathbf{k}_0$ We now define an optical energy band $\mathscr{E}_{ji}(\mathbf{k})$ as the difference between two electron energy bands:

$$\mathscr{E}_{ji}(\mathbf{k}) = \mathscr{E}_{ji}(\mathbf{k}_0) + \boldsymbol{\alpha}^{ji} \cdot (\mathbf{k} - \mathbf{k}_0) + (\mathbf{k} - \mathbf{k}_0) \cdot \boldsymbol{\beta}^{ji} \cdot (\mathbf{k} - \mathbf{k}_0) \tag{5.50}$$

where

$$\mathscr{E}_{ji}(\mathbf{k}) = \mathscr{E}_j(\mathbf{k}) - \mathscr{E}_i(\mathbf{k})$$
$$\boldsymbol{\alpha}^{ji} = \boldsymbol{\alpha}^j - \boldsymbol{\alpha}^i, \qquad \boldsymbol{\beta}^{ji} = \boldsymbol{\beta}^j - \boldsymbol{\beta}^i$$

Clearly, by direct comparison of Eqs. (5.48)–(5.50), the optical energy band has a density of states

$$\rho(\mathscr{E}_{ji}) = \frac{1}{4\pi^3} \int \frac{dS}{|\nabla_{\mathbf{k}} \mathscr{E}_{ji}|_{\mathscr{E}_{ji} = \hbar\omega}}$$

$$= 2J_{ji} \tag{5.51}$$

The density of states J_{ji} as given by Eq. (5.51) is called the joint density of states and is usually denoted by J_{cv} for transitions from the valence band to the conduction band. It does not include the factor of two for spin. Of course, optical transitions are weighted by matrix elements, so Eq. (5.51) cannot be directly substituted into Eq. (5.44). However, $|\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2$ is a slowly varying function in many cases. Thus, in the neighborhood of regions where J_{ji} is a strongly varying function, $|\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2$ can be taken outside the integral. Then $\varepsilon_2(\omega)$ is directly proportional to the joint density of states:

$$\varepsilon_2(\omega) = 8(\pi e/m\omega)^2 |\mathbf{a}_0 \cdot \mathbf{p}_{ji}|^2 J_{cv} \tag{5.52}$$

Points in $\mathbf{k}$-space for which

$$\nabla_{\mathbf{k}} \mathscr{E}_{ji} = 0 \tag{5.53}$$

are called critical points or van Hove singularities [1]. They were discussed first by van Hove in connection with neutron scattering and later applied by

Phillips to a discussion of electronic energy bands and optical properties [2].

Critical points are sources of prominent structure in the joint density of states and in the optical constants. They can occur for both electronic and optical energy bands. In the case of an electronic energy band, a critical point is defined by

$$\nabla_{\mathbf{k}} \mathscr{E}_j = 0 \tag{5.54}$$

Such critical points occur only at highly symmetric points in the Brillouin zone. If also

$$\nabla_{\mathbf{k}} \mathscr{E}_i = 0 \tag{5.55}$$

then the two electronic critical points necessarily produce an optical critical point

$$\nabla_{\mathbf{k}} \mathscr{E}_{ji} = \nabla_{\mathbf{k}} \mathscr{E}_j - \nabla_{\mathbf{k}} \mathscr{E}_i = 0 \tag{5.56}$$

However, optical critical points can also occur for

$$\nabla_{\mathbf{k}} \mathscr{E}_j = \nabla_{\mathbf{k}} \mathscr{E}_i \neq 0 \tag{5.57}$$

The latter usually occur on symmetry planes or lines, but may be at a general point in the Brillouin zone. A critical point defined by Eq. (5.56) is called a symmetry critical point.

There are four types of critical points corresponding to four types of singularities which may appear in the joint density of states. This can be seen by examining a quadratic expansion for the energy in the vicinity of a critical point.

In the neighborhood of a critical point $\mathbf{k}_c$, and with a suitable choice of coordinate axes, a constant-energy surface can be expressed as

$$\mathscr{E}(\mathbf{k}) = \mathscr{E}(\mathbf{k}_c) + \beta_1 k_1^{\,2} + \beta_2 k_2^{\,2} + \beta_3 k_3^{\,2}$$

$$\mathbf{k}_1 = (\mathbf{k} - \mathbf{k}_c)_1, \qquad \mathbf{k}_2 = (\mathbf{k} - \mathbf{k}_c)_2, \qquad \mathbf{k}_3 = (\mathbf{k} - \mathbf{k}_c)_3 \tag{5.58}$$

The four types of analytical critical points correspond to the four possible choices of signs (plus or minus) for the coefficients β_1, β_2, and β_3.

Suppose that the coefficients β_1, β_2, and β_3 are all positive. Then the surfaces of constant energy are ellipsoids in $\mathbf{k}$-space. The volume of an ellipsoid with surface energy $\mathscr{E}(\mathbf{k})$ is

$$\text{volume in } \mathbf{k}\text{-space} = \frac{4\pi}{3} \frac{[\mathscr{E}(\mathbf{k}) - \mathscr{E}(\mathbf{k}_c)]^{3/2}}{(\beta_1 \beta_2 \beta_3)^{1/2}} \tag{5.59}$$

Since

$$\int \rho(\mathscr{E}_{ji})\, d\mathscr{E}_{ji} = \frac{\text{volume in } \mathbf{k}\text{-space}}{4\pi^3}$$

$$= \frac{1}{3\pi^2} \frac{[\mathscr{E}_{ji}(\mathbf{k}) - \mathscr{E}_{ji}(\mathbf{k}_c)]^{3/2}}{(\beta_1\beta_2\beta_3)^{1/2}} \tag{5.60}$$

we obtain

$$J_{ji}(\mathscr{E}) = \frac{1}{2}\rho_{ji}(\mathscr{E}) = \frac{1}{2}\frac{d}{d\mathscr{E}}\left\{\frac{1}{3\pi^2}\frac{[\mathscr{E}(\mathbf{k}) - \mathscr{E}(\mathbf{k}_c)]^{3/2}}{(\beta_1\beta_2\beta_3)^{1/2}}\right\}$$

$$= \frac{[\mathscr{E}(\mathbf{k}) - \mathscr{E}(\mathbf{k}_c)]^{1/2}}{4\pi^2(\beta_1\beta_2\beta_3)^{1/2}} \tag{5.61}$$

This is just the well-known square root dependence of the density of states above a minimum. It is designated as an M_0-type critical point. The subscript in the notation designates the number of negative coefficients β_1, β_2, β_3 (or the number of negative masses). The joint density of states in the neighborhood of an M_0-type critical point is shown in Fig. 5.5 along with the other three types of critical points.

The derivation of the joint density of states for an M_1 critical point is

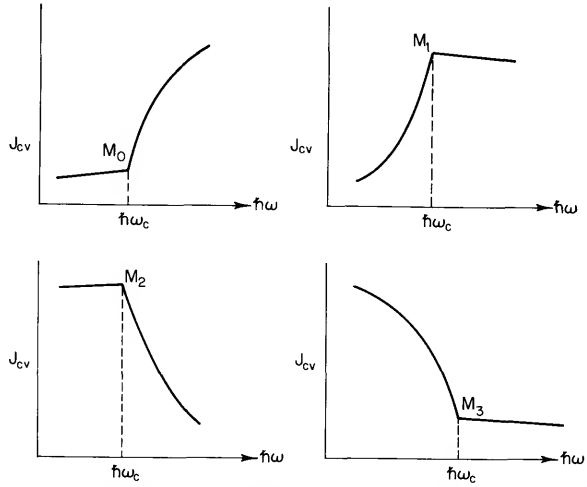

Fig. 5.5 Van Hove singularities in the joint density of states J_{cv}. These singularities appear superimposed on a (sometimes) rather smooth background in J_{cv} arising from the totality of transitions throughout the Brillouin zone. Thus, e.g., the line sloping slightly down and to the left from M_0 represents the smooth background in the neighborhood of M_0. The figures are drawn such as to emphasize the critical point structure; the true zero in J_{cv} would generally be well off scale in these figures.

given in Appendix F. Critical points of type M_1 and M_2 are known as saddle points because the constant-energy surfaces have the general shape of a saddle. They are sometimes designated S_1- and S_2-type critical points.

The properties of the four types of critical points are summarized in Table 5.1.

TABLE 5.1 *Joint Density of States J_{cv} for Critical Points*

Critical point	Type	β_1	β_2	β_3	J_{cv}	
					$\mathscr{E} < \mathscr{E}_0$	$\mathscr{E} > \mathscr{E}_0$
M_0	Minimum	+	+	+	0	$C_0(\mathscr{E} - \mathscr{E}_0)^{1/2}$
$M_1(=S_1)$	Saddle point	+	+	−	$C_1 - C_1'(\mathscr{E}_0 - \mathscr{E})^{1/2}$	C_1
$M_2(=S_2)$	Saddle point	+	−	−	C_2	$C_2 - C_2'(\mathscr{E} - \mathscr{E}_0)^{1/2}$
M_3	Maximum	−	−	−	$C_3(\mathscr{E}_0 - \mathscr{E})^{1/2}$	0

We have just considered three-dimensional critical points. Should it happen that one or two of the coefficients β in Eq. (5.58) are negligible because of principal effective masses of large magnitude, then it is of interest to consider one- or two-dimensional critical points.

Clearly, critical points provide important information concerning the electronic band structure. Knowledge of symmetry critical points is especially helpful since they can be used as end points for interpolation schemes. However, although critical points are in principle of special significance, they may make only a small contribution to $\varepsilon_2(\omega)$. Their identification in an absorption or reflection spectrum may be quite difficult. They are generally superimposed on a broad background and the structure due to the critical points themselves is further weakened by broadening. It is not usually sufficient just to use high-quality single crystals and make measurements at low temperature in order to resolve critical-point structure.

The structure in the neighborhood of critical points is greatly enhanced if a derivative technique is used. There are a number of ways to modulate the optical properties in order to be able to measure a derivative of some sort. These include electroreflectance, piezoreflectance, and magnetoreflectance; all are discussed in detail by Cardona [3]. Here, we shall discuss only wavelength-modulation spectroscopy.

Wavelength-modulation spectroscopy requires a wavelength-modulated monochromatic light beam. If the wavelength (or frequency) of the beam is then slowly changed so as to sweep through the spectral region of interest, and the ac component of the reflectance is detected by synchronous phase-sensitive detection techniques, the derivative $dR/d\lambda$ or $dR/d\omega$ can be measured directly. This has the advantage of discriminating against the

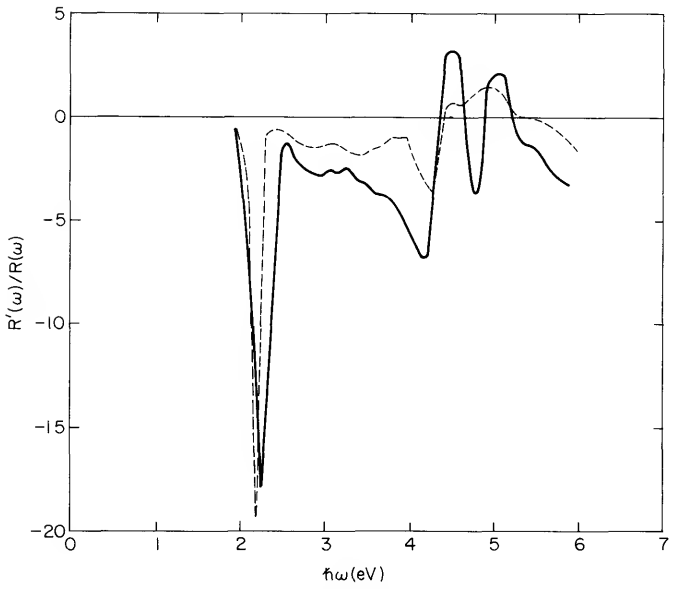

Fig. 5.6 Frequency modulation spectrum of copper. The measured values of $R'(\omega)/R(\omega)$ are at $T = 7°K$. The agreement between theory (——) and experiment (— — —) confirms the assignments of transitions between critical points as identified by a number of workers. However, it is found that volume-effect contributions from all over the Brillouin zone are more important than critical points. [From C. Y. Fong, M. L. Cohen, R. R. L. Zucca, J. Stokes, and Y. R. Shen, *Phys. Rev. Lett.* **25**, 1486 (1970).]

large but rather structureless background. Compare, for example, the rich structure in the wavelength modulation spectrum of Cu (Fig. 5.6) with the ordinary reflectance spectrum of Cu (Fig. 3.21). Of course, critical points are not the only sources of structure in the derivative spectrum, but when structure can be identified as arising from critical points, it is especially helpful in interpreting the electronic band structure. We shall discuss some examples of derivative reflectance spectroscopy in the next section.

5.4 Direct Transitions in Germanium

Figure 5.7 shows the room-temperature reflectance spectrum of Ge. The band structure is shown in Fig. 5.8 and a comparison of calculated and experimental values of ε_2 is shown in Fig. 5.9. We now want to discuss some of the transitions shown in these figures and see how they are related to the critical points discussed in the preceding section.

The reflectance spectrum is dominated by a doublet peak at 2.3 eV and a large maximum at 4.5 eV. The latter peak arises from transitions $\Sigma_2 \rightarrow \Sigma_3$

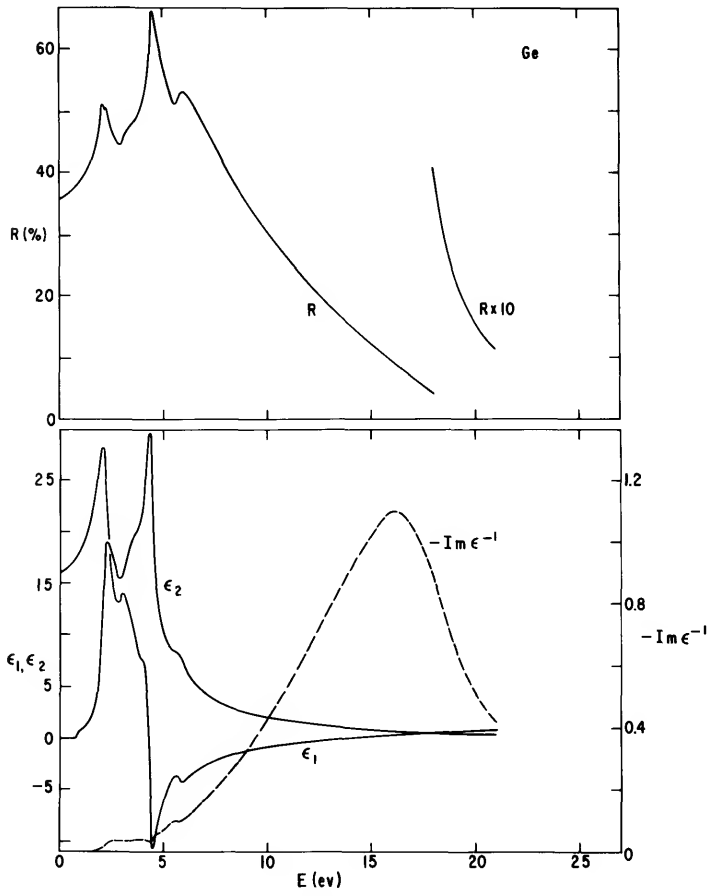

Fig. 5.7 Room-temperature reflectivity and optical constants of Ge. [From H. R. Phillip and H. Ehrenreich, *Phys. Rev.* **129**, 1550 (1963).]

and points around $X_4 \rightarrow X_1$, but probably also contains contributions from extended regions of the Brillouin zone. Nonetheless, one can gain an understanding of the 4.5-eV peak simply in terms of the $X_4 \rightarrow X_1$ and $\Sigma_2 \rightarrow \Sigma_3$ transitions as shown in the band diagram of Fig. 5.8. Note that these transitions are at M_1- and M_2-type saddle points. In fact, it is necessary to have two saddle points to produce a peak; a single saddle point produces only an edge. The formation of a peak from two saddle points is illustrated in Fig. 5.10. The 4.5-eV reflectance peak in Ge can be understood, in a somewhat oversimplified picture, as arising from the accidental near-coincidence of the two saddlepoint transitions $X_4 \rightarrow X_1$ and $\Sigma_2 \rightarrow \Sigma_3$.

The doublet peak at 2.3 eV arises mainly from M_1 saddlepoint transitions

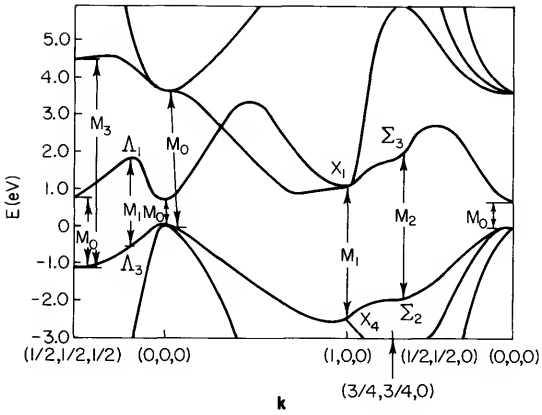

Fig. 5.8 The band structure of Ge. [From J. C. Phillips, D. Brust, and G. F. Bassani, *Proc. Int. Conf. Phys. Semicond. Exeter London* p. 564 (1962).]

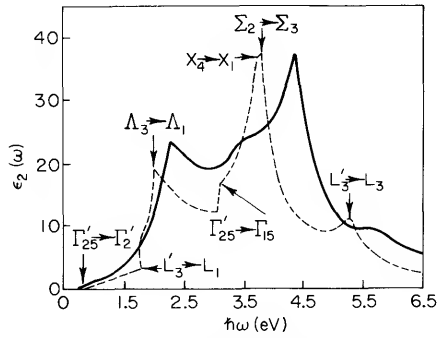

Fig. 5.9 Imaginary part of the dielectric constant $\varepsilon_2(\omega)$ for Ge. (——) Experiment; (— — —) theory. [From D. Brust, J. C. Phillips, and G. F. Bassani, *Phys. Rev. Lett.* **9**, 94 (1962).]

$\Lambda_3 \rightarrow \Lambda_1$. Saravia and Brust [4] have shown that the spin–orbit splitting of Λ_3 near L is about two-thirds of the splitting at $\Gamma_{25'}$. The Λ_1 level is not spin–orbit split [5]. Since the spin–orbit splitting at $\Gamma_{25'}$ is known from theory and confirmed by experiment to be 0.29 eV [6], the observed splitting of 0.20 eV for the 2.3-eV peak confirms the two-thirds rule.

The weaker peak at 6 eV arises from $L_{3'} \rightarrow L_3$ transitions. It is also a doublet, although this is not apparent from the room-temperature reflectance measurements of Fig. 5.7. The spin–orbit splitting of the $L_{3'} \rightarrow L_3$ transition can be seen even at room temperature, however, if sufficient care is taken in sample preparation. The splitting is found to be 0.2 eV, in agreement with theoretical predictions for the splitting of the $L_{3'}$ state [7].

One can learn more about these and other transitions from a study of the wavelength-modulated reflectance spectrum. Some results for Ge

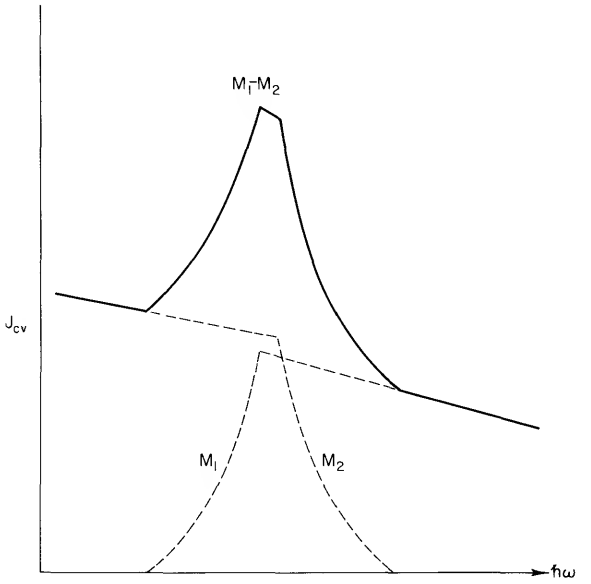

Fig. 5.10 The production of a peak in the joint density of states by the superposition of two saddle points.

are shown in Fig. 5.11. From a decomposition of this reflectance spectrum into its components, Zucca and Shen [8] have verified the assignments just discussed as well as a number of others. The decomposition was made according to these general rules: (1) The low-temperature spectrum of a composite line was decomposed into a minimum number of individual lines with simple line shapes. (2) Recomposition of these individual lines with broadened linewidths had to yield the high-temperature spectrum of the composite line. (3) Similarity in the spectra of different semiconductors was used as a guideline in the decomposition. The latter rule is, of course, always useful. It is always of help in making assignments of transitions to study the systematics of a related group of materials.

Figure 5.11 suggests that the temperature dependence of band structure can also be studied quite nicely by an analysis of the temperature dependence of the wavelength-modulated reflectance spectra. Zucca and Shen have in fact made some qualitative deductions of this type from their studies.

The emphasis in this section has been on an analysis of three peaks in the reflectance spectrum of Ge. These peaks can all be understood in terms of transitions near saddle points. However, the peaks observed in the reflectance spectrum of many semiconductors and most insulators are caused by excitons associated with critical points. Excitons are especially

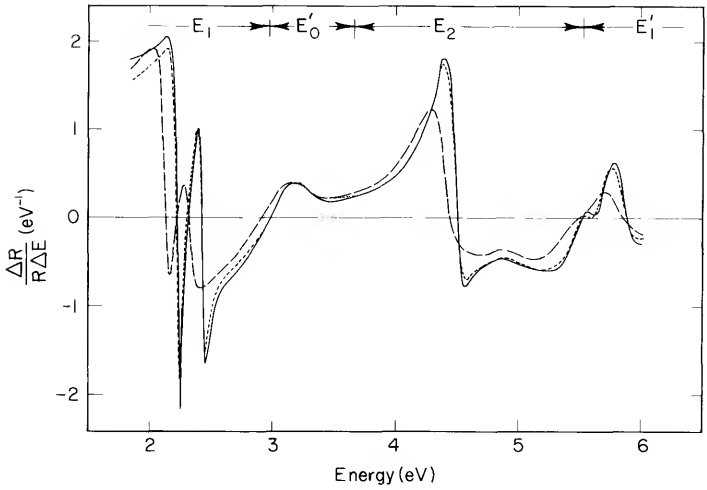

Fig. 5.11 Logarithmic derivative of the reflectivity spectrum of Ge at (——) 5, (– – –) 80, and (— —) 300°K. [From R. R. L. Zucca and Y. R. Shen, *Phys. Rev.* **B1**, 2668 (1971).]

important in insulators because of the importance of the electron–hole interaction in the absence of strong screening. This is why the reflectance spectrum of insulators such as KCl (see Fig. 3.6) is usually dominated by peaks rather than edges. Even the peaks for Ge at 2.3 eV and 4.5 eV probably have associated excitons. This is suggested by a much greater line narrowing at low temperatures than is predicted by simple interband effects.

5.5 Direct Transitions in Silver: Effects of Temperature and Alloying

Direct interband transitions in metals are spread rather widely in energy. This wide distribution is oscillator strengths is apparent from the rather smooth and gradual rise in n_{eff}, the effective number of electrons per atom contributing to the optical properties over a given frequency range, as shown in Fig. 5.12. Note that n_{eff} rises rapidly to a value near unity, the free-electron value, and remains nearly constant until the threshold for direct interband transitions is reached. The deviation from unity in the free-electron value of n_{eff} arises from departures of the effective mass from the free-electron mass and perhaps also from many-body effects as discussed in Chapter 4. Assuming that core states can be neglected, n_{eff} should saturate at the value 11 (1s plus 10d electrons). It is clear from Fig. 5.12 that oscillator strengths for interband transitions in Ag and Cu are spread over a very wide energy range.

We shall now consider in more detail the interband transitions near thresh-

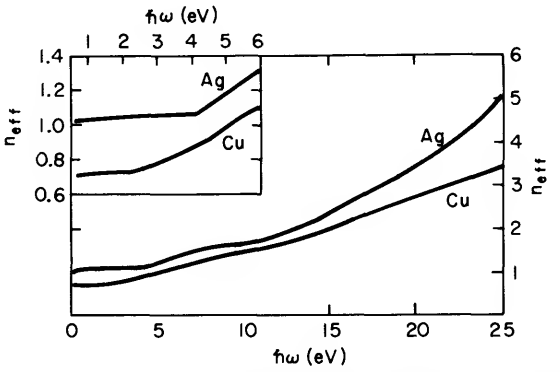

Fig. 5.12 Effective numbers of electrons per atom contributing to the optical properties over a given energy range versus energy. [From H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962).]

old in Ag, particularly what can be learned about these transitions from studies of alloying and temperature variation. We shall be concerned with the two transitions indicated in the schematic band diagram of Fig. 5.13 for Ag near L.

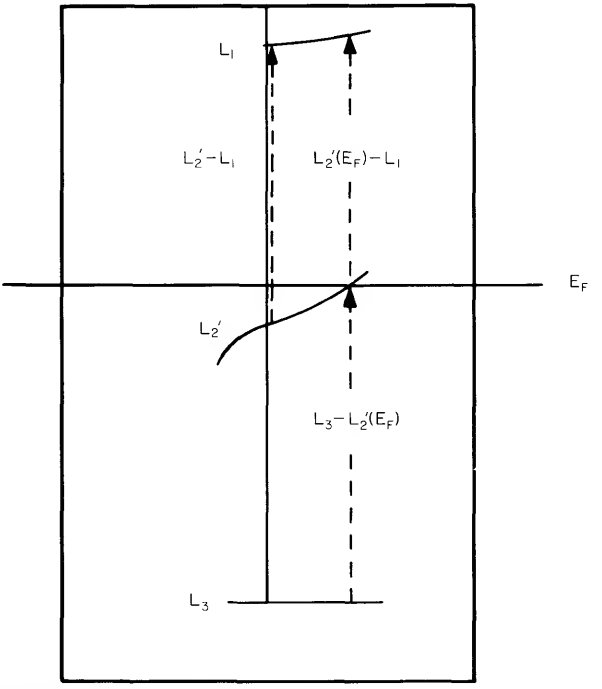

Fig. 5.13 Schematic band diagram for Ag near L. [From H. G. Liljenvall and A. G. Mathewson, *J. Phys. C Metal Phys. Suppl.* No. 3, S341 (1970).]

The transition L_3–$L_{2'}$ is the transition from the d band to the Fermi surface that is responsible for the sharp rise in ε_2 at 3.9 eV as shown in Fig. 3.17. This transition has been well established by a number of workers [9–13]. The $L_{2'}$–L_1 transition from the Fermi level to a higher-lying conduction band is hidden under the stronger L_3–$L_{2'}$ transition. This transition was identified from studies of dilute alloys of Al, Cd, Zn, and In in Ag [11–13] as well as from studies of the temperature dependence of the optical properties of pure Ag [14]. That it overlapped the L_3–$L_{2'}$ transition had also been suggested earlier by Mueller and Phillips [15] based on band structure calculations, and by Berglund and Spicer [16] from their photoemission studies.

Figure 5.14 show the reflectance of a series of Ag–Al alloys. The dielectric functions and the loss functions obtained from an analysis of these

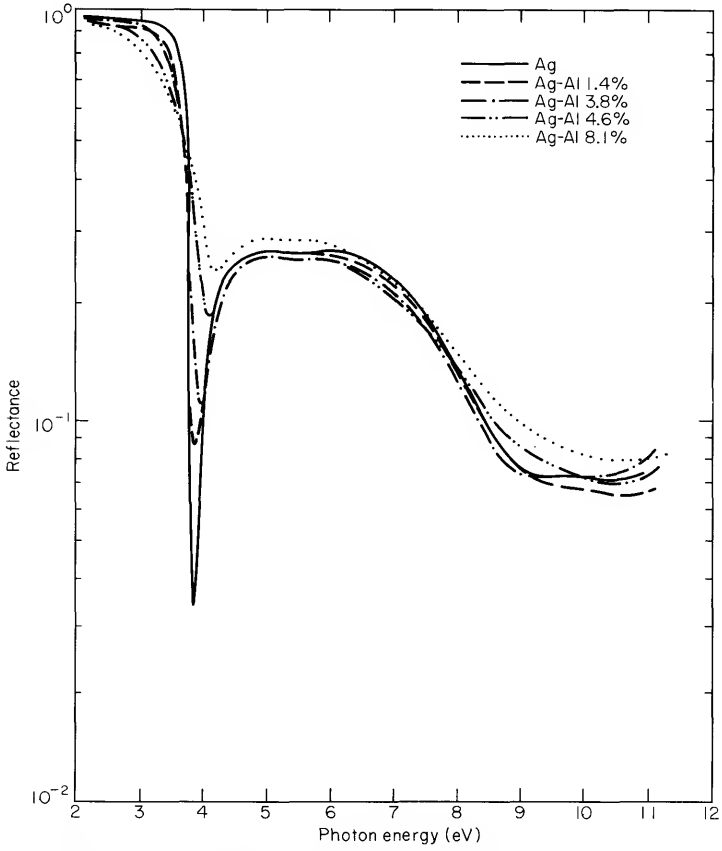

Fig. 5.14 Reflectance of Ag and Ag–Al alloys at 300°K. [From G. B. Irani, T. Huen, and F. Wooten, *Phys. Rev.* **3B**, 2385 (1971).]

data are shown in Figs. 5.15–5.17. Figure 5.15 shows a peak in the ε_1 curves for Ag and the Ag–Al alloys near 4 eV. This peak is the contribution to ε_1 of $\varepsilon_1{}^b$, which is shown separately in the insert. Increasing solute concentration damps and broadens the peak in $\varepsilon_1{}^b$; it also moves to slightly higher energy the peak in ε_1 and the energy at which ε_1 crosses zero. The latter is an indication of a shift in the plasma frequency. The plasma frequency and

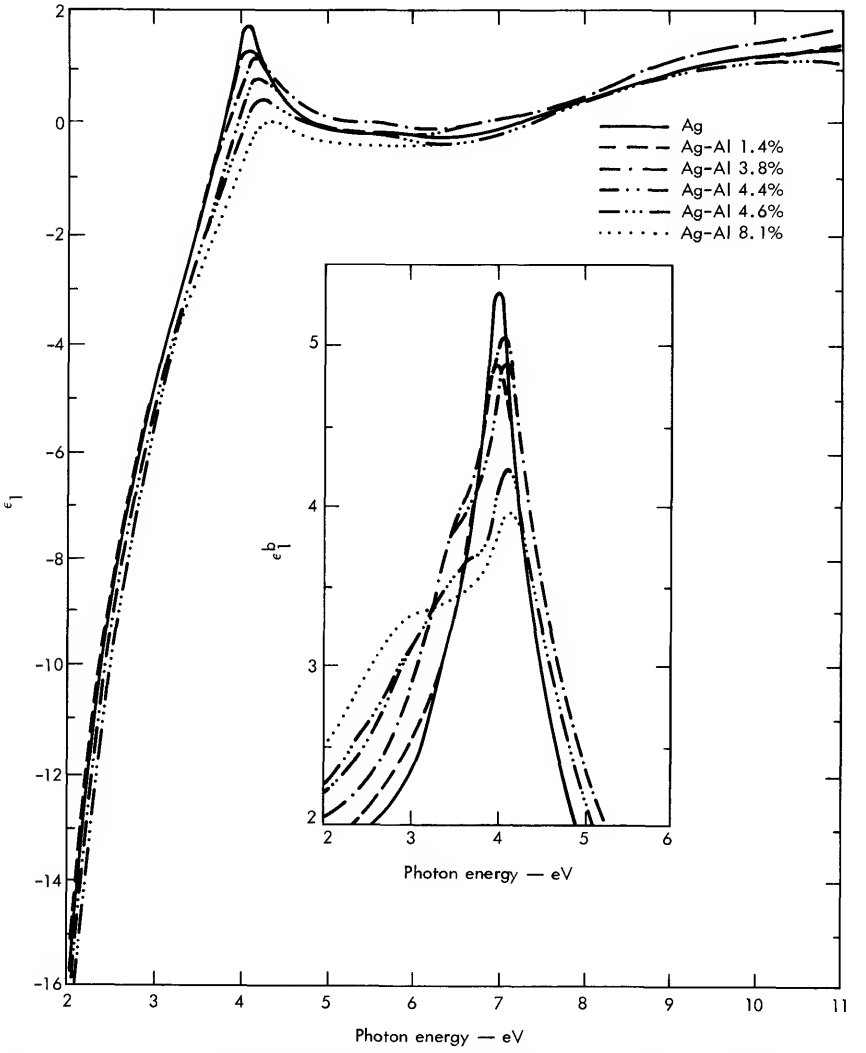

Fig. 5.15 The real dielectric functions ε_1 and $\varepsilon_1{}^b$ versus photon energy for Ag and Ag–Al alloys at 300°K. [From G. B. Irani, T. Huen, and F. Wooten, *Phys. Rev.* **3B**, 2385 (1971).]

its shift to higher energy with increasing solute concentration are defined more clearly by the peak in the electron loss functions

$$-\mathrm{Im}(1/\hat{\varepsilon}) = \varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2) \tag{5.62}$$

shown in Fig. 5.17. The shift in plasma frequency to higher energies is expected because of the increased electron concentration obtained on substituting Al for Ag.

The main peak in $\varepsilon_1{}^b$ near 4 eV shifts slowly to higher energy as Al is added to Ag. This peak is associated with the sharp rise in ε_2 at 3.9 eV and is identified mainly with transitions from d states to the Fermi surface, the

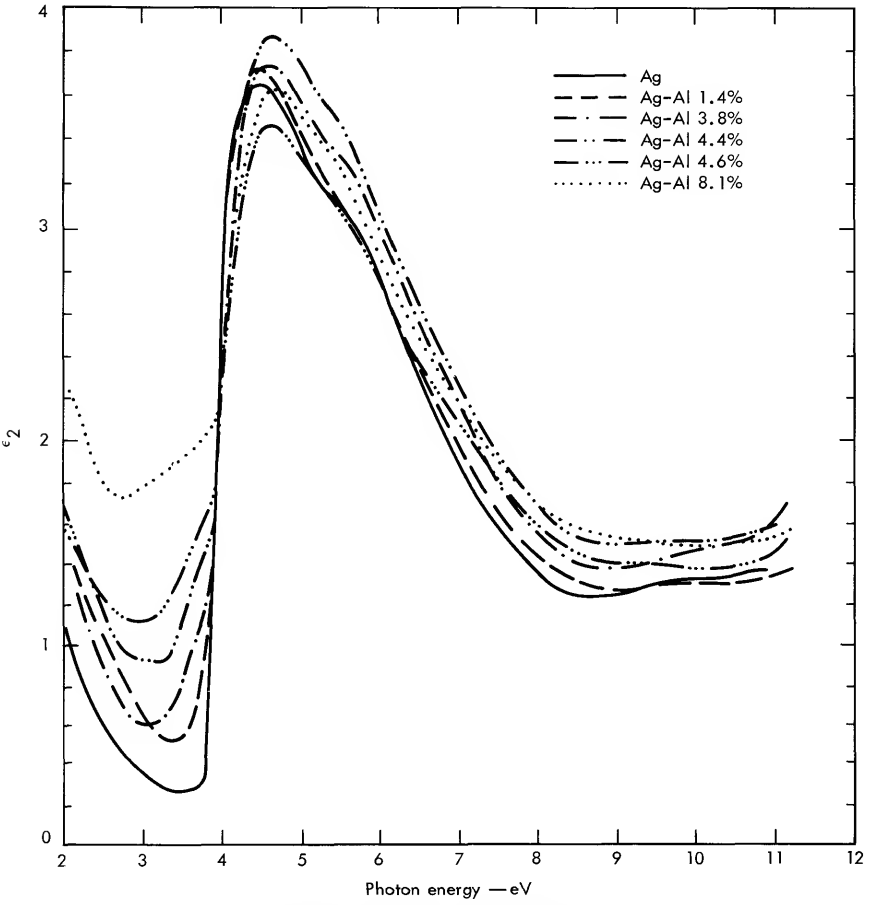

Fig. 5.16 The imaginary dielectric function ε_2 versus photon energy for Ag and Ag–Al alloys at 300°K. [From G. B. Irani, T. Huen, and F. Wooten, *Phys. Rev.* **3B**, 2385 (1971).]

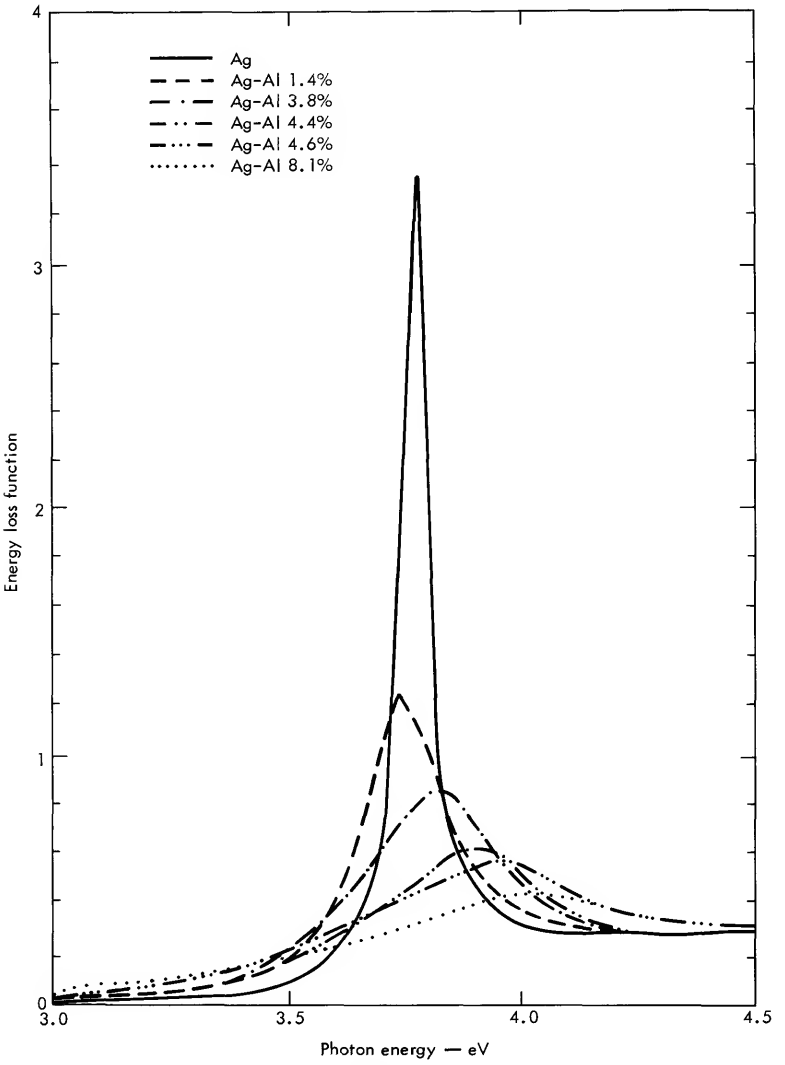

Fig. 5.17 The energy loss functions versus photon energy for Ag and Ag–Al alloys at 300°K. [From G. B. Irani, T. Huen, and F. Wooten, *Phys. Rev.* **3B**, 2385 (1971).]

L_3–$L_{2'}$ transition. A second structure emerges from $\varepsilon_1{}^b$ and moves more quickly to lower energy with increasing Al concentration. The corresponding increase in ε_2 can be seen in Fig. 5.16. This arises from the $L_{2'}$–L_1 transition. The rapid shift of the $L_{2'}$–L_1 transition to lower energy is a result of the great sensitivity of the L_1 level to crystal potential [17–19]. This same

shift also increases the magnitude of ε_2 in the region where $\varepsilon_1 \simeq 0$, and so increases the damping of the plasma resonance, as can be seen in Fig. 5.17.

A calculation of the band structure of Ag at two different lattice constants by Jacobs [20] shows that the $L_{2'}$–L_1 gap moves to lower energies relative to the L_3–$L_{2'}$ gap with increasing lattice constant. Thus, even in pure Ag, it should be possible to separate these transitions if the lattice constant is changed or the potential is changed. One way to do this is to vary the temperature. This has been done by Liljenvall and Mathewson [14] and some of their results are shown in Figs. 5.18 and 5.19. However, increasing the temperature changes the lattice constant only slightly and the effects of volume expansion can usually be neglected. Thus, the effects of temperature illustrated in Figs. 5.18 and 5.19 can be attributed mostly to a change in the average crystal potential because of increased motion of the atoms about their only slightly changed mean position.

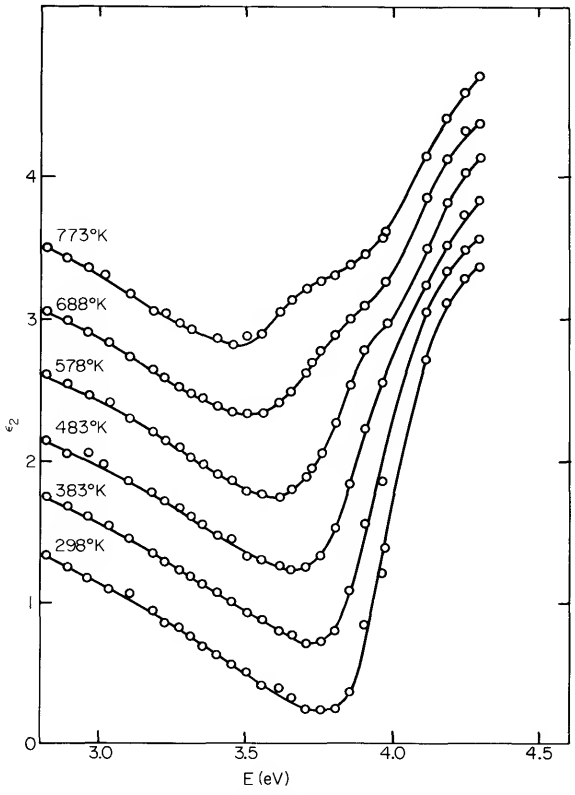

Fig. 5.18 The imaginary dielectric function ε_2 versus photon energy for Ag at different temperatures. [From H. G. Liljenvall and A. G. Mathewson, *J. Phys. C Metal Phys. Suppl.* No. 3, 5341 (1970).]

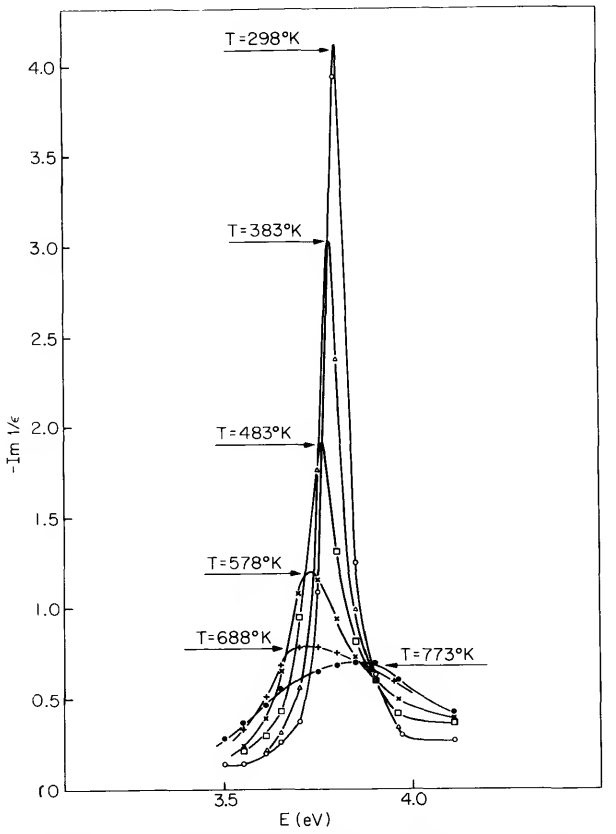

Fig. 5.19 The energy loss function versus photon energy for Ag at different temperatures. [From H. G. Liljenvall and A. G. Mathewson, *J. Phys. C Metal. Phys. Suppl.* No. 3, 5341 (1970).]

The effect of increasing temperature is the same as alloying except that increasing the temperature moves the electron loss peak to lower energies. The latter shift arises from the contribution to $\varepsilon_1{}^b$ of the transition $L_{2'}$–L_1. As this contribution to ε_1 moves to lower energy, it also shifts to lower energy the point at which $\varepsilon_1 \simeq 0$. This shift also makes a contribution in the alloys, but there it is offset by the greater effect of an increased electron concentration. The mechanism for increased damping of the loss function, however, is the same for both cases.

There has so far been no success in determining unambiguously the line shapes of these transitions, and hence the nature of the critical points, as has been done for Ge and many semiconductors.

5.6 Indirect Transitions

We have so far considered only the interaction of electrons with an electromagnetic field. The interaction with the perfect lattice is included implicitly via an effective mass. If we now also include interactions with imperfections in the crystal, then indirect transitions are possible. One such transition is illustrated in Fig. 5.20.

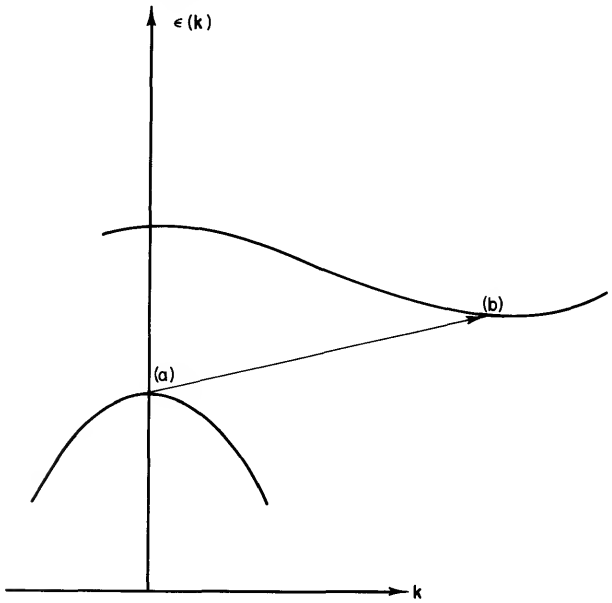

Fig. 5.20 An indirect transition from (a) the valence band maximum to (b) a conduction band minima. There are various mechanisms via intermediate states that make this transition possible. Some are illustrated in later figures.

The only imperfections we will consider explicitly are lattice vibrations. That is, even in an otherwise perfect crystal, the existence of a finite temperature means the crystal lattice is vibrating and the perfect periodicity of the lattice is destroyed. The quantized lattice vibrations (phonons) constitute another means of interaction with electrons. More than that, even at 0°K, an energetic electron can excite a lattice vibration, i.e., create a phonon. Thus we can expect indirect interband transitions at all temperatures, but a higher transition rate at higher temperatures, where the equilibrium phonon density is higher.

We shall now consider indirect transitions of the type illustrated in Fig. 5.20. Let us assume that direct transitions are allowed but that we are

working in a photon energy range corresponding to the indirect band gap. Thus direct transitions are energetically impossible.

We assume that solutions are available for the unperturbed crystal. That is, $H_0\psi = \mathscr{E}\psi$ leads to eigenfunctions ψ_n such that $H_0\psi_n = \mathscr{E}_n\psi_n$. When the perturbation is present, we have to solve

$$i\hbar\,\partial\Psi/\partial t = (H_0 + H')\Psi \tag{5.63}$$

where

$$\Psi = \sum_n a_n(t)\psi_n e^{-i\mathscr{E}_n t/\hbar} \tag{5.64}$$

Following the procedures of Sections 3.4 and 5.1, this leads to

$$i\hbar\, da_m/dt = \sum_n a_n V_{mn}(e^{i(\omega_{mn}+\omega)t} + e^{i(\omega_{mn}-\omega)t}) \tag{5.65}$$

where

$$V_{mn} = \int \psi_m^* H'(\mathbf{r})\psi_n\, d\mathbf{r} \tag{5.66}$$

In first-order perturbation theory, we then have

$$i\hbar\, da_m/dt = V_{m0}(e^{i(\omega_{m0}+\omega)t} + e^{i(\omega_{m0}-\omega)t}) \tag{5.67}$$

for the system originally in the ground state. In the present case, even though direct transitions are allowed according to symmetry, it is necessary to have a two-step process to excite the indirect transition, and so we must go to second-order perturbation theory.

The general procedure for second-order perturbation theory is as follows. We ignore the coefficient a_0 in Eq. (5.65) since it did not lead to a constant transition rate for the transition under consideration. (If it had, we would not need to go to second-order perturbation theory.) We now consider all other a_n in Eq. (5.65). However, we use the results of first-order perturbation theory for the time dependence of a_n. Thus, in second order, Eq. (5.65) becomes

$$i\hbar\, da_m^{(2)}/dt = \sum_{n\neq 0} a_n^{(1)} V_{mn}(e^{i(\omega_{mn}+\omega)t} + e^{i(\omega_{mn}-\omega)t}) \tag{5.68}$$

The superscripts (1) or (2) on a_n indicate whether the coefficient is the result of a first-order or second-order perturbation calculation. The summation in Eq. (5.68) can be over all n. The indication $n \neq 0$ simply serves as a reminder that there was no contribution to da_m/dt in first-order and so there is no need to include it.

We are now ready to derive the transition rate for indirect transitions. These transitions can be thought of as a two-step process: (1) the absorption of energy by the absorption of a photon and (2) the conservation of

crystal momentum $\hbar\mathbf{k}$ by emission or absorption of a phonon. This process is illustrated in Fig. 5.21. We will consider only photon absorption, but since phonons have energies comparable to kT, we will have to consider both absorption and emission of phonons. In fact, in many cases, it is even necessary to consider the emission and/or absorption of several phonons for a single transition. However, we shall consider explicitly only single-phonon emission or absorption.

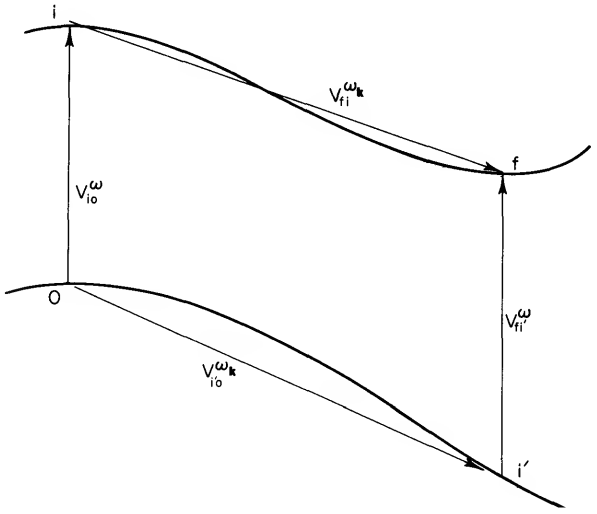

Fig. 5.21 Two mechanisms for an indirect transition from the valence band maximum at 0 to the conduction band minimum at f. One mechanism is a virtual direct transition from the initial state 0 to intermediate state i followed by a phonon-assisted transition to final state f. A second mechanism is a phonon-assisted transition to intermediate state i' followed by a direct transition to f. The latter can also be described (more accurately) as a vertical transition of an electron from i' to f with the positive hole left behind being scattered to 0. However, it is simpler in following through the mathematics to adopt the former viewpoint.

We assume that the Hamiltonian for the interaction can be written as a sum of terms of the form given in Eq. (5.1). Thus

$$H' = \tfrac{1}{2}A^{\text{photon}}[(\exp i\omega t) + \exp -i\omega t]$$

$$+ \tfrac{1}{2}V^{\text{phonon}}[(\exp i\omega_{\mathbf{k}} t) + \exp - i\omega_{\mathbf{k}} t] \qquad (5.69)$$

where A^{photon} and V^{phonon} are the appropriate operators for electron–photon and electron–phonon interactions and $\hbar\omega_{\mathbf{k}}$ is the energy of a phonon of wave vector $\mathbf{k}$. The operators for the interactions are chosen to symbolize the fact that the origin of the interaction of an electron with the light wave is through a vector potential, whereas the electron–phonon interaction arises

from the Coulomb forces between electron and atomic cores. Generally, the wave vector of the light wave is unimportant, but the wave vector for phonons is very important. The wave vector for photons and phonons is contained implicitly in the operators A^{photon} and V^{phonon}.

We can now define matrix elements

$$V_{mn}^{\omega} \equiv \int \psi_m^*(\tfrac{1}{2}A^{\text{photon}})\psi_n \, d\mathbf{r}$$

$$V_{mn}^{\omega_{\mathbf{k}}} \equiv \int \psi_m^*(\tfrac{1}{2}V^{\text{phonon}})\psi_n \, d\mathbf{r} \tag{5.70}$$

First-order perturbation theory then yields

$$i\hbar \, da_n^{(1)}/dt = V_{n0}^{\omega} \exp[i(\omega_{n0} - \omega)t] + V_{n0}^{\omega_{\mathbf{k}}}(\exp[i(\omega_{n0} + \omega_{\mathbf{k}})t] + \exp[i(\omega_{n0} - \omega_{\mathbf{k}})t]) \tag{5.71}$$

A term $V_{n0}^{\omega} e^{i(\omega_{n0}+\omega)t}$ is not included in Eq. (5.70) because it corresponds to photon emission, a process of no concern here.

Integration of Eq. (5.71) gives

$$a_n^{(1)} = \frac{V_{n0}^{\omega}}{\hbar}\left\{\frac{1 - \exp[i(\omega_{n0} - \omega)t]}{\omega_{n0} - \omega}\right\} + \frac{V_{n0}^{\omega_{\mathbf{k}}}}{\hbar}\left\{\frac{1 - \exp[i(\omega_{n0} \pm \omega_{\mathbf{k}})t]}{\omega_{n0} \pm \omega_{\mathbf{k}}}\right\} \tag{5.72}$$

These are the coefficients we need to substitute into Eq. (5.68). In actual cases, though, we do not need the complete infinite set of coefficients $a_n^{(1)}$. Indeed, usually only a few are necessary. As a specific example, and one that is applicable to a number of semiconductors, we will consider transitions from a valence band maximum at $\mathbf{k} = 0$ to a conduction band minimum at a $\mathbf{k}$ value well removed from the center of the Brillouin zone (see Fig. 5.21). We shall assume all other valence bands and conduction bands to be sufficiently far removed that they need not be considered. That is, they will correspond to states such that the denominators in terms such as appear in Eq. (5.71) are so large that there is an insignificant contribution to absorption. Thus, in the example of Fig. 5.21, only two intermediate states, i and i', need be considered. The appropriate matrix elements for transitions to or from these intermediate states are indicated in the figure. They are chosen according to whether the transition corresponds to photon absorption or electron–phonon interactions.

The only coefficients $a_n^{(1)}$ we need from first-order perturbation theory are $a_i^{(1)}$ and $a_{i'}^{(1)}$. With these, Eq. (5.68) becomes

$$i\hbar\, da_f/dt = a_i^{(1)} V_{fi}^{\omega_{\mathbf{k}}} \left(\exp[i(\omega_{fi} + \omega_{\mathbf{k}})t] + \exp[i(\omega_{fi} - \omega_{\mathbf{k}})t]\right)$$

$$+ a_{i'}^{(1)} V_{fi'}^{\omega} \exp[i(\omega_{fi} - \omega)t] \tag{5.73}$$

where a term $a_{i'}^{(1)} V_{fi'}^{\omega} e^{i(\omega_{fi} + \omega)t}$ is not included because it corresponds to photon emission. Note that in using Eq. (5.68), it is necessary to remember that the ω appearing there is the angular frequency of the time-dependent perturbation. Thus, in rewriting it as in Eq. (5.73), it was necessary to make the transformation $\omega \to \omega_{\mathbf{k}}$ for the contribution of electron–phonon interactions to the transition.

We can now substitute for $a_i^{(1)}$ and $a_{i'}^{(1)}$ in Eq. (5.73). Once again, though, rather than blind substitution, we will pick out from Eq. (5.72) only those terms that are important. For example, when we consider the factor $a_i^{(1)} V_{fi}^{\omega_{\mathbf{k}}}$ appearing in Eq. (5.73), we realize that the matrix element $V_{fi}^{\omega_{\mathbf{k}}}$ accounts for the transition via a phonon from state i to final state f. Thus, what we need in $a_i^{(1)}$ is only those terms corresponding to the transition via a photon from state 0 to state i. Following this procedure, Eq. (5.73) becomes

$$i\hbar \frac{da_f^{(2)}}{dt} = \frac{V_{fi}^{\omega_{\mathbf{k}}} V_{i0}^{\omega}}{\hbar} \left\{ \frac{\exp[i(\omega_{fi} \pm \omega_{\mathbf{k}})t] - \exp[i(\omega_{fi} \pm \omega_{\mathbf{k}} + \omega_{i0} - \omega)t]}{\omega_{i0} - \omega} \right\}$$

$$+ \frac{V_{fi'}^{\omega} V_{i'0}^{\omega_{\mathbf{k}}}}{\hbar} \left\{ \frac{\exp[i(\omega_{fi'} - \omega)t] - \exp[i(\omega_{fi'} - \omega + \omega_{i'0} \pm \omega_{\mathbf{k}})t]}{\omega_{i'0} \pm \omega_{\mathbf{k}}} \right\} \tag{5.74}$$

The next step is to integrate Eq. (5.74) to determine $a_f^{(2)}$ and hence $|a_f^{(2)}|^2$. We can anticipate some of the results and thus simplify what has to be done. The term proportional to $\exp[i(\omega_{fi} \pm \omega_{\mathbf{k}})t]$ will lead to an integrated term containing $[\sin^2 \frac{1}{2}(\omega_{fi} \pm \omega_{\mathbf{k}})\, t]/(\omega_{fi} \pm \omega_{\mathbf{k}})^2$. Just as in Section 5.1, this leads to appreciable absorption only if $\omega_{\mathbf{k}} = \mp\, \omega_{fi}$. That is, energy must be conserved for the indirect transition via a phonon and, hence, energy must then also be conserved for the direct transition $0 \to i$ via photon absorption. There is nothing wrong with that except there would be no need then to consider indirect transitions via second-order perturbation theory. We have excluded this case by specifying that we are concerned only with the situation for which the (allowed) direct transitions are energetically impossible. The same argument holds for the term $\sim\exp[i(\omega_{fi'} - \omega)t]$. Eliminating these two terms, and making the substitution

$$\omega_{fi} + \omega_{i0} = \omega_{fi'} + \omega_{i'0} = \omega_{f0} \tag{5.75}$$

Eq. (5.74) becomes

$$i\hbar \frac{da_f^{(2)}}{dt} = -\frac{V_{fi}^{\omega_k} V_{io}^{\omega}}{\hbar(\omega_{i0} - \omega)} \exp[i(\omega_{f0} \pm \omega_k - \omega)t]$$

$$-\frac{V_{fi'}^{\omega} V_{i'0}^{\omega_k}}{\hbar(\omega_{i'0} \pm \omega_k)} \exp[i(\omega_{f0} \pm \omega_k - \omega)t] \tag{5.76}$$

Conservation of energy for the overall transition requires that

$$\omega = \omega_{f0} \pm \omega_k = \omega_{fi'} + \omega_{i'0} \pm \omega_k \tag{5.77}$$

Hence,

$$\omega_{i'0} \pm \omega_k = \omega - \omega_{fi'} \tag{5.78}$$

Substitution of Eq. (5.78) into (5.76) and integration from $t = 0$ to t gives

$$a_f^{(2)} = \left[\frac{V_{fi}^{\omega_k} V_{i0}^{\omega}}{\hbar^2(\omega_{i0} - \omega)} + \frac{V_{fi'}^{\omega} V_{i'0}^{\omega_k}}{\hbar^2(\omega - \omega_{fi'})}\right] \times \left\{\frac{\exp[i(\omega_{f0} \pm \omega_k - \omega)t] - 1}{\omega_{f0} \pm \omega_k - \omega}\right\} \tag{5.79}$$

The probability of finding an electron in state f at time t is given by $|a_f^{(2)}|^2$. As before, we will also have to integrate over a range of possible transitions or final states. We saw earlier that integrals of the form $\int d\omega\, [\sin^2(\frac{1}{2}\omega t)]/\omega^2$ are necessary to get a transition probability that increases linearly with time. Thus all cross terms can be neglected in finding $|a_f^{(2)}|^2$. A typical term that does contribute to $|a_f^{(2)}|^2$ is

$$|a_f^{(2)}|_{ia}^2 = \left[\frac{V_{fi}^{\omega_k} V_{i0}^{\omega}}{\hbar^2(\omega_{i0} - \omega)} \frac{\exp[i(\omega_{f0} - \omega_k - \omega)t] - 1}{\omega_{f0} - \omega_k - \omega}\right] \times \left[\frac{(V_{fi}^{\omega_k} V_{i0}^{\omega})^*}{\hbar^2(\omega_{i0} - \omega)} \frac{\exp[-i(\omega_{f0} - \omega_k - \omega)t] - 1}{\omega_{f0} - \omega_k - \omega}\right]$$

$$= \frac{4|V_{fi}^{\omega_k}|^2 |V_{i0}^{\omega}|^2 \sin^2 \frac{1}{2}(\omega_{f0} - \omega_k - \omega)t}{\hbar^4(\omega_{i0} - \omega)^2 (\omega_{f0} - \omega_k - \omega)^2} \tag{5.80}$$

where the subscripts ia on $|a_f^{(2)}|_{ia}^2$ indicate a vertical transition to intermediate state i followed by phonon absorption to reach the final state f. We can simplify this result by rewriting it in terms of a δ function for energy conservation. The result, obtained by following the procedure used in Section 5.1, is

$$|a_f^{(2)}|_{ia}^2 = \frac{2\pi t |V_{fi}^{\omega_k}|^2 |V_{i0}^{\omega}|^2}{\hbar^4(\omega_{i0} - \omega)^2} \delta(\mathscr{E}_{f0} - \hbar\omega_k - \hbar\omega) \tag{5.81}$$

Following the same procedure for the other three terms in Eq. (5.79), and differentiating with respect to time to get the transition rate, we obtain

$$W_{f0} = \frac{2\pi|V_{fi}^{\omega_{\mathbf{k}}}|^2\,|V_{i0}^{\omega}|^2}{\hbar^4(\omega_{i0}-\omega)^2}\left[\delta(\mathscr{E}_{f0}-\hbar\omega_{\mathbf{k}}-\hbar\omega)+\delta(\mathscr{E}_{f0}+\hbar\omega_{\mathbf{k}}-\hbar\omega)\right]$$

$$+\frac{2\pi|V_{fi'}^{\omega}|^2\,|V_{i'0}^{\omega_{\mathbf{k}}}|^2}{\hbar^4(\omega-\omega_{fi'})^2}\left[\delta(\mathscr{E}_{f0}-\hbar\omega_{\mathbf{k}}-\hbar\omega)+\delta(\mathscr{E}_{f0}+\hbar\omega_{\mathbf{k}}-\hbar\omega)\right] \quad (5.82)$$

The first term in Eq. (5.82) gives the transition rate from state 0 to state f via a direct transition to intermediate state i followed by a transition to final state f via phonon absorption. The matrix element $V_{fi}^{\omega_{\mathbf{k}}}$ is zero unless $\mathbf{k}$(phonon) = $\mathbf{k}$(electron in state f). Thus, only those states f satisfying conservation of crystal momentum as well as conservation of energy for the overall process can be reached from a given initial state for a fixed value of $\hbar\omega$. The two possibilities along one direction in $\mathbf{k}$-space are illustrated in Fig. 5.22. Considering all directions in $\mathbf{k}$-space, many final states with essentially the same energy can be reached from the single initial state 0. We shall ignore the small differences in energy arising from the difference in energy of phonons having slightly different $\mathbf{k}$ values. Thus, referring to Fig. 5.23,

$$W_{ia} = \frac{2\pi|V_{fi}^{\omega_{\mathbf{k}}}|^2\,|\overline{V_{i0}^{\omega}}|^2}{\hbar^4(\omega_{i0}-\omega)^2}\,\rho_c(\hbar\omega+\hbar\omega_{\mathbf{k}}-\mathscr{E}_G) \quad (5.83)$$

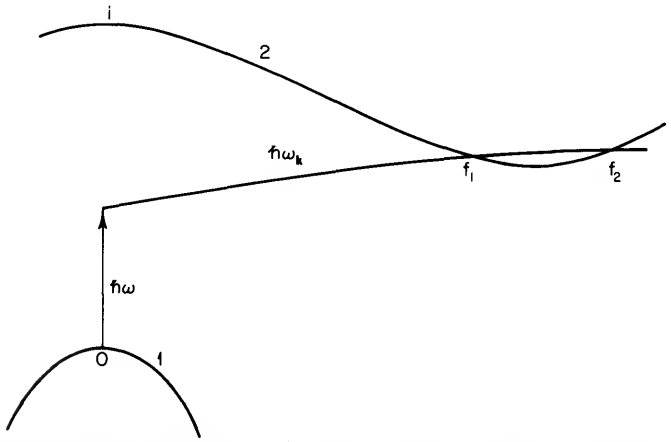

Fig. 5.22 The energy conditions for an indirect transition. (1) Valence band, (2) conduction band. The phonon dispersion curve $\hbar\omega_{\mathbf{k}}$ is superposed on the electronic band diagram so as to show the final states reached by a combination of photon and phonon absorption. The energy scale for the phonon dispersion curve is exaggerated. Note that the end of the arrow indicating photon absorption does not correspond to the intermediate state for the overall indirect transition. It merely indicates the amount of energy contributed by the photon. The true intermediate state is indicated by i.

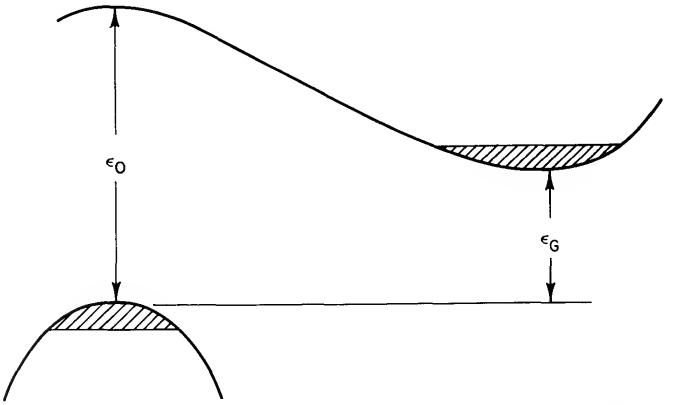

Fig. 5.23 The energy conditions for indirect transitions over a range of initial and final states. The zero of energy is taken at the conduction band minimum. $\mathscr{E}_{max} = \hbar\omega \pm \hbar\omega_{\mathbf{k}} - \mathscr{E}_G$.

is the transition rate from state 0, via intermediate state i, followed by phonon absorption to a group of final states with the same energy and having wave vector near $\mathbf{k}_f$. Because the electron–photon interaction depends on the photon polarization (which is not spherically symmetric), we assume an appropriate average matrix element $|\overline{V_{i0}^{\omega}}|$ in Eq. (5.83). If we now consider transitions from a group of initial states near $\mathbf{k} = 0$, we obtain in the region near threshold for indirect transitions from the valence band to the conduction band, via phonon absorption,

$$W_{ca} = \frac{2\pi|V_{fi}^{\omega_{\mathbf{k}}}|^2\,|\overline{V_{i0}^{\omega}}|^2}{\hbar^4(\omega_{i0} - \omega)^2} \int_0^{\hbar\omega + \hbar\omega_{\mathbf{k}} - \mathscr{E}_G} \rho_v(\mathscr{E} - \hbar\omega - \hbar\omega_{\mathbf{k}})\rho_c(\mathscr{E})\,d\mathscr{E} \qquad (5.84)$$

The limits of integration are determined from inspection of Fig. 5.23. Equation (5.84) assumes $|V_{fi}^{\omega_{\mathbf{k}}}|$ to be constant, which is reasonable for small variations in $\mathbf{k}$ when $\mathbf{k}$ is large. Also, $|\overline{V_{i0}^{\omega}}|$ has been assumed constant, i.e., independent of $\mathbf{k}_0$, where we have taken $\mathbf{k}_0 = 0$. The latter is equivalent to our initial assumption that direct transitions are allowed (although, in the present case, lack of energy conservation for the vertical transition prohibits such transitions in reality). See, in this regard, the discussion at the end of the section.

We have neglected several things in Eq. (5.84). We have neglected the various phonon branches and the modes in each branch. Diamond, for example, has three modes in the acoustic branch and three modes in the optical branch. Also, there may be a number of equivalent conduction-band minima. This should all be realized, but for purposes of illustration, we shall proceed with Eq. (5.84). The important feature is that

$$W_{ca} \propto |V_{fi}^{\omega_{\mathbf{k}}}|^2\,|\overline{V_{i0}^{\omega}}|^2 \int_0^{\hbar\omega + \hbar\omega_{\mathbf{k}} - \mathscr{E}_G} \rho_v(\mathscr{E} - \hbar\omega - \hbar\omega_{\mathbf{k}})\,\rho_c(\mathscr{E})\,d\mathscr{E} \qquad (5.85)$$

For parabolic energy bands, this becomes

$$W_{ca} \propto |V_{fi}^{\omega_k}|^2 \, |\overline{V_{i0}^{\omega}}|^2 \int_0^{\hbar\omega + \hbar\omega_k - \mathscr{E}_G} (\hbar\omega + \hbar\omega_k - \mathscr{E}_G)^{1/2} \mathscr{E}^{1/2} \, d\mathscr{E}$$

$$\propto |V_{fi}^{\omega_k}|^2 \, |\overline{V_{i0}^{\omega}}|^2 (\hbar\omega + \hbar\omega_k - \mathscr{E}_G)^2 \tag{5.86}$$

Now, the probability of phonon absorption is proportional to the phonon density N_p. That is,

$$|V_{fi}^{\omega_k}|^2 \propto N_p = \frac{1}{\exp(\hbar\omega_k/kT) - 1} \tag{5.87}$$

Thus, when direct transitions are allowed, so that $|V_{i0}^{\omega}|$ is nearly constant, the absorption coefficient for the mechanism under discussion varies as

$$\alpha_{ca} \propto \frac{(\hbar\omega + \hbar\omega_k - \mathscr{E}_G)^2}{\exp(\hbar\omega_k/kT) - 1} \tag{5.88}$$

The probability of phonon emission is proportional to $N_p + 1$. Using this, the absorption coefficient for phonon emission following a virtual direct transition is easily found to be

$$\alpha_{ce} \propto \frac{\exp(\hbar\omega_k/kT)}{\exp(\hbar\omega_k/kT) - 1} (\hbar\omega - \hbar\omega_k - \mathscr{E}_G)^2 \tag{5.89}$$

There remains one case to consider, namely, when direct transitions are forbidden. This results when the first term in the last line of Eq. (5.17) is zero because of the symmetry of the conduction-band and valence-band wave functions. Then, it is necessary to consider the previously neglected second term. The result is that

$$|\overline{V_{i0}^{\omega}}|^2 \propto (k')^2 \neq \text{const} \tag{5.90}$$

so that for transitions from near $\mathbf{k} = 0$, $|\overline{V_{i0}^{\omega}}|^2$ is proportional to the energy of the initial state measured with respect to the valence-band maximum and must be included in the integration. Thus, e.g.,

$$\alpha \propto \int_0^{\hbar\omega \pm \hbar\omega_k - \mathscr{E}_G} (\hbar\omega \pm \hbar\omega_k - \mathscr{E}_G)^{3/2} \mathscr{E}^{1/2} \, d\mathscr{E}$$

$$\propto (\hbar\omega \pm \hbar\omega_k - \mathscr{E}_G)^3 \tag{5.91}$$

when direct transitions are forbidden.

In summary, the essential features for indirect transitions are:

$$\alpha = \frac{A(\hbar\omega + \hbar\omega_{\mathbf{k}} - \mathscr{E}_G)^2}{\exp(\hbar\omega_{\mathbf{k}}/kT) - 1} + \frac{B[\exp(\hbar\omega_{\mathbf{k}}/kT)]\ (\hbar\omega - \hbar\omega_{\mathbf{k}} - \mathscr{E}_G)^2}{\exp(\hbar\omega_{\mathbf{k}}/kT) - 1} \tag{5.92}$$

when vertical transitions are allowed, and,

$$\alpha = \frac{C(\hbar\omega + \hbar\omega_{\mathbf{k}} - \mathscr{E}_G)^3}{\exp(\hbar\omega_{\mathbf{k}}/kT) - 1} + \frac{D[\exp(\hbar\omega_{\mathbf{k}}/kT)]\,(\hbar\omega - \hbar\omega_{\mathbf{k}} - \mathscr{E}_G)^3}{\exp(\hbar\omega_{\mathbf{k}}/kT) - 1} \tag{5.93}$$

when vertical transitions are forbidden.

These results are modified when exciton effects are included. A thorough analysis has been given by Elliott [21], who finds that for allowed indirect transitions to unbound electron–hole states, the absorption coefficient varies as the $\frac{3}{2}$ power of the energy in excess of threshold. Allowed indirect transitions into exciton states give a $\frac{1}{2}$-power variation. For forbidden indirect transitions, the absorption coefficient for transitions to unbound electron–hole states varies as the $\frac{5}{2}$ power of the energy above threshold, and for transitions to exciton states, it rises as the $\frac{3}{2}$ power. We will see these effects when we study the absorption edge of Ge in the next section.

Now, recall that conservation of momentum is required in each step of the indirect transition. However, conservation of energy is not required for these virtual transitions to or from intermediate states. It is required only for the overall indirect process. What is required is that the time spent in the intermediate state be compatible with the uncertainty principle. The uncertainty in energy of the intermediate state is $\Delta\mathscr{E} \approx |\mathscr{E}_0 - \hbar\omega|$, where $\mathscr{E}_0$ is the direct band gap. The lifetime of the intermediate state must then be

$$\tau_i \approx \hbar/|\mathscr{E}_0 - \hbar\omega| \tag{5.94}$$

The uncertainty in energy enters as a square term in the denominator of expressions such as Eq. (5.84) for the transition rate. The point of interest here is that the appearance of this energy difference term in the denominator usually makes one of the processes illustrated in Fig. 5.21 very unlikely.

In all the preceding discussions, we have been talking about an idealized system in which the electrons, phonons, and photons are all decoupled. Transitions have been between states of the unperturbed system. A strictly rigorous treatment would not have the system decoupled. The true eigen-

states of the system must be for the solid plus radiation field. This is of course impractical and not necessary. A more realistic way of viewing the system is not to worry so much about including the photons, but at least to recognize that one should not decouple the electrons and phonons. That is, each state should be recognized as having both electron and phonon character. We could depict this situation schematically by making fuzzy $\mathscr{E}(\mathbf{k})$ lines in the traditional one-electron band structure diagrams. Then, an indirect transition can be considered to go from an initial state 0 to a final state f without an intermediate state. In the actual mathematics, though, the final result of an attempt along these lines will be mathematically equivalent to the original prescription in which electrons and phonons are decoupled.

5.7 The Absorption Edge in Ge, AgBr, and AgBr(Cl)

We have derived expressions for the absorption coefficient for indirect transitions for the case in which the threshold is from $\mathbf{k} = 0$ in the valence band to $\mathbf{k}$ near the Brillouin zone edge in the conduction band. This is just the situation which is applicable to Ge.

The band diagram for Ge is shown in Fig. 5.8. The indirect band gap threshold is from Γ_{25} to L_1. Measured absorption curves near the absorption edge are shown in Fig. 5.24, taken from the classic work of MacFarlane *et al.* [22].

Below 20°K, the absorption consists of two components, one beginning at 0.75 eV and the other at 0.77 eV. At higher temperatures, four components are apparent, beginning at energies of 0.705, 0.725, 0.745, and 0.760 eV. At still higher temperatures, the onset of absorption is at even lower energies. These various components of the absorption curve are a result of phonon emission or absorption during the indirect transition.

We now define a quantity $A \equiv \alpha\hbar\omega$ and call it, simply, the absorption. A careful study of the absorption curve shows that the components come in pairs. We denote the two components of the ith pair by A_{ia} and A_{ie}, corresponding to absorption and emission of a phonon. The corresponding threshold energies are denoted by $\mathscr{E}_{ia}$ and $\mathscr{E}_{ie}$, and their mean value is denoted by $\mathscr{E}_i$. Clearly $(\mathscr{E}_{ie} - \mathscr{E}_{ia})$ is just twice the energy $k\theta$ of the phonons emitted or absorbed during the indirect transition. Some of the quantities, as determined from a detailed analysis of curves such as those in Fig. 5.24, are listed in Table 5.2.

At 4.2°K, we expect phonons to be emitted but not absorbed; there just are no phonons to absorb. If we make the assignment at 4.2°K as resulting from emission of two types of phonons, one of effective temperature θ_1 and the other of effective temperature θ_2, the overall results are self-consistent.

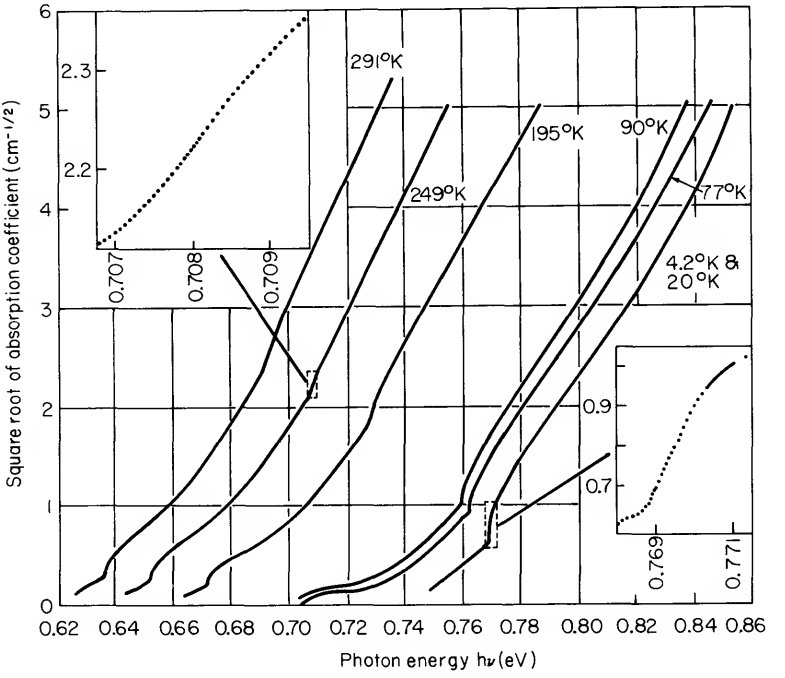

Fig. 5.24 The absorption-edge spectrum of Ge at various temperatures. [From G. G. MacFarlane, T. P. McLean, J. E. Quarrington, and V. Roberts, *Phys. Rev.* **108**, 1377 (1957).]

TABLE 5.2 *Values of Threshold Energies and Effective Phonon Temperatures Obtained from an Analysis of the Absorption Curves in Fig. 5.24*

T (°K)	$\mathscr{E}_{2a}$	$\mathscr{E}_{1a}$	$\mathscr{E}_{1e}$	$\mathscr{E}_{2e}$	$\bar{\mathscr{E}}_1$	$\bar{\mathscr{E}}_2$	θ_1 (°K)	θ_2 (°K)
4.2	—	—	0.7485	0.7686	—	—	—	—
77	0.7063	0.7264	0.7419	0.7614	0.7342	0.7339	90	320
291	0.6367	0.6570	0.6724	0.6913	0.6647	0.6640	88	317

At 77°K, the phonon density is sufficiently high that the two types of phonons emitted at 4.2°K can now also be absorbed during the transition.

At much higher temperatures, many more processes involving phonons are possible, but as the temperature is raised, the free-carrier concentration also increases and the effects of phonon absorption or emission are soon masked by free-carrier (Drude) absorption.

The variation in $\bar{\mathscr{E}}_1$ and $\bar{\mathscr{E}}_2$ is a result of the temperature dependence of the band gap. We shall not discuss that aspect of the problem. However, the close agreement of $\bar{\mathscr{E}}_1$ and $\bar{\mathscr{E}}_2$ at each temperature and the nearly

constant values of θ_1 and θ_2 are support for the correctness of the analysis given above.

The basic shape of the components A_{1a} and A_{1e} is given by

$$A_1 = a_1(\hbar\omega - \mathscr{E}_1)^{3/2} \tag{5.95}$$

where a_1 is temperature-dependent. The basic shape of the components A_{2a} and A_{2e} is given by

$$A_2 = a_2[(\hbar\omega - \mathscr{E}_2)^{1/2} + b(\hbar\omega - \mathscr{E}_2 - 0.0010)^{1/2} + c(\hbar\omega - \mathscr{E}_2 - 0.0027)^{3/2}] \tag{5.96}$$

where a_2 is temperature-dependent but b and c are constants.

The interpretation of Eqs. (5.95) and (5.96) is as follows. The initial square root variation in A_2 is from allowed transitions into the ground state of the exciton. The contribution from the second term is probably a result of transitions to the first excited state of the exciton. The third term arises from allowed indirect transitions producing unbound electron–hole pairs. The threshold for these transitions is 0.0027 eV above the exciton ground state. This energy is thus the binding energy for indirect excitons in Ge.

The $\frac{3}{2}$-power dependence for A_1 would seem to indicate no exciton production for single-phonon emission processes. However, if this were so, the values of $\bar{\mathscr{E}}_1$ should be larger than $\bar{\mathscr{E}}_2$ by just the exciton binding energy, 0.0027 eV. We see from Table 5.2 that this is not so. Presumably, excitons are involved, but it is not possible to see their contribution to the experimental data.

Indirect phonon-assisted transitions also are evident at the absorption threshold for a number of other materials. As a second example, we consider AgBr. The band structure is shown in Fig. 5.25 and the absorption coefficient is given in Fig. 5.26.

At 4.7°K, only phonon emission contributes to the transition $L_{3'} \rightarrow \Gamma_1$. At higher temperatures, the threshold shifts to lower energy because phonon absorption can also contribute. A single phonon of energy 0.00805 eV appears to be important at threshold. A more detailed analysis such as was outlined for germanium shows good agreement between experiment and theory.

Figure 5.27 shows the low-temperature absorption coefficient for AgBr + 1% AgCl. The important difference from Fig. 5.26 is that a new component appears midway between the two thresholds for the pure AgBr crystal. This is a temperature-independent component. It arises because the selection rules are not strictly valid for a disordered crystal. It is possible to have zero-phonon indirect transitions. The impurities act to conserve momentum for the transition.

As a final point, it should be noted that the absorption coefficient in

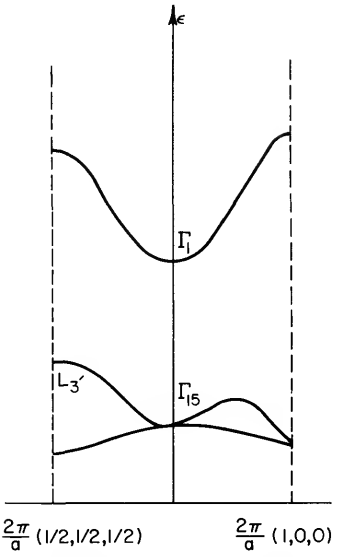

Fig. 5.25 Schematic band structure diagram for AgBr.

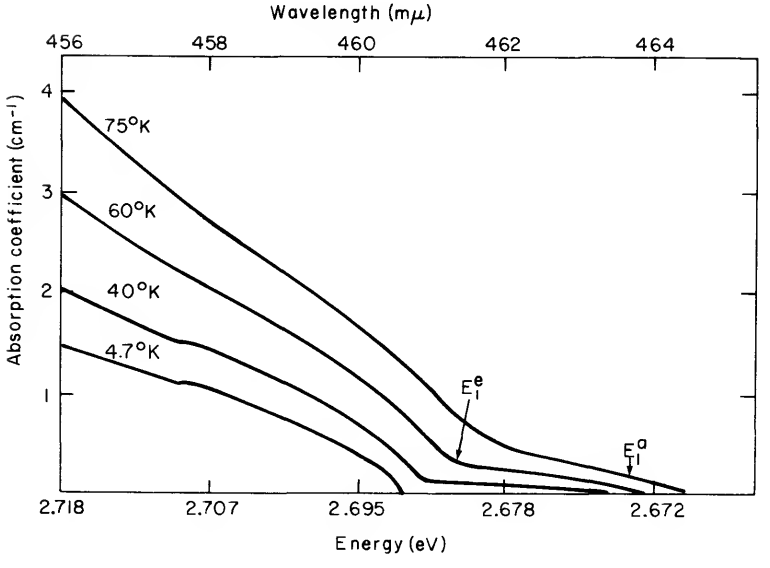

Fig. 5.26 The absorption edge for pure AgBr at several different temperatures. E_1^e and E_1^a denote, respectively, the thresholds corresponding to emission and absorption of a 0.009-eV phonon. [From B. L. Joesten and F. C. Brown, *Phys. Rev.* **148**, 919 (1966).]

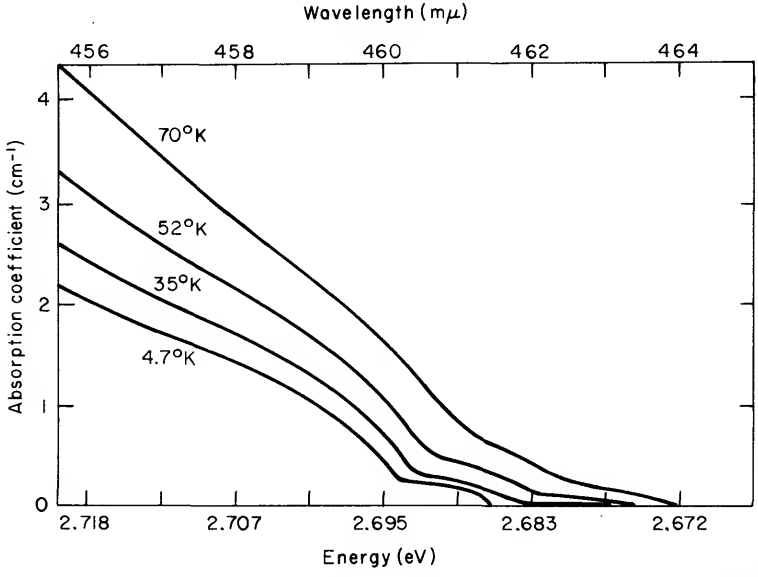

Fig. 5.27 The absorption edge at different temperatures for AgBr containing 1 mole% AgCl. Comparison of this figure with Fig. 5.26 shows that a disorder-induced component sets in exactly halfway between the edges associated with phonon emission and phonon absorption. [From B. L. Joesten and F. C. Brown, *Phys. Rev.* **148**, 919 (1966).]

Figs. 5.24, 5.26, and 5.27 is quite small. It is several orders of magnitude smaller than for typical allowed vertical transitions in semiconductors and about four orders of magnitude smaller than for metals in the visible spectrum. As a result, fairly thick samples ($\gtrsim 1$ mm) are required to determine optical properties from transmission measurements.

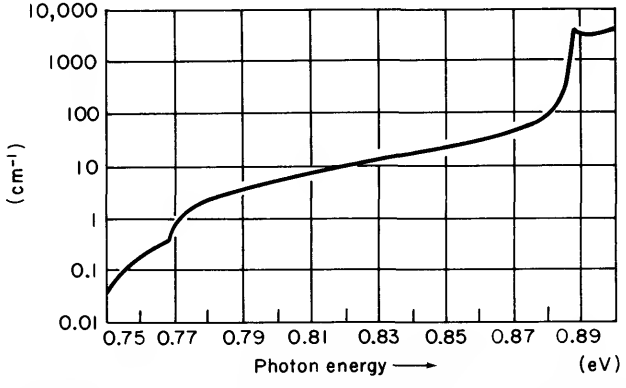

Fig. 5.28 The absorption edge in germanium at 20°K [From T. P. McLean, *in* "Progress in Semiconductors" (A. F. Gibson, ed.), Vol. 5, p. 55. Heywood, London, 1960.]

We return now to the absorption in germanium. The absorption edge for germanium is shown in Fig. 5.28. Here, the range has been extended to include the absorption edge arising from direct transitions at $\mathbf{k} = 0$. This occurs just below 0.89 eV. The small peak is a broadened exciton line, a "direct" exciton. Note the fairly good agreement with our previous estimate (Section 5.2) for the absorption coefficient near the direct gap threshold. There, we estimated $\alpha \approx 5 \times 10^3$ cm^{-1} for $\hbar\omega = 0.01$ eV above the direct gap threshold.

5.8 Excitons

We have already mentioned excitons and discussed their effect on the absorption coefficient near threshold. We shall now consider excitons themselves. However, the treatment will be only semiquantitative. The study of excitons constitutes a field in itself. A rigorous treatment is left to monographs and review articles.

In the tight-binding approximation, we can imagine an exciton of a single atom. We can visualize it in its simplest form as the interaction of an excited electron with the hole it left in the shell it came from. Such a localized excitation is called a Frenkel exciton. But it cannot really be restricted to a particular atom. Atoms in a solid interact and the quantum of energy can jump to a neighboring atom and still conserve the energy of the entire system. In this process, the exciton carries no net charge, only energy. Because the exciton cannot be restricted to a single atom, it is an excited state of the whole system.

A Frenkel exciton is characterized by a small radius for the orbit of the excited electron. However, if the orbit of the excited electron is such as to include a large number of atoms, the Coulomb interaction between electron and hole is screened by the dielectric function. Then, the effective-mass approximation of band theory is more appropriate and the excitations are called Wannier excitons. Stated in a different way, the approaches are as follows. For Frenkel excitons, the electron and hole are close to one another and interact via Coulomb and exchange interactions. For Wannier excitons, the electron and hole are widely separated and it is simplest to consider the electron and hole as essentially free and put in the Coulomb interaction as a perturbation. The latter is more appropriate for the excitons formed at the absorption edge in germanium. For these indirect transitions, the exciton wave packet which describes the motion of the electron–hole pair can be constructed of hole wave functions centered at $\mathbf{k} = 0$ and linear combinations of electron wave functions at $\mathbf{k}_c$ and $-\mathbf{k}_c$, where $\mathbf{k}_c$ is at the conduction band minimum. We shall emphasize these Wannier excitons.

In material like Ge and Cu_2O with small band gaps and large dielectric

constants, electrons and holes are quite accurately treated as independent particles. The weak, screened Coulomb interaction then can produce bound pairs which still retain much of their free character.

We consider first an idealized case with valence-band maximum and conduction-band minimum at $\mathbf{k} = 0$ and

$$\mathscr{E}_c(\mathbf{k}) = \mathscr{E}_c(0) + (\hbar^2 k^2/2m_e^*) \tag{5.97}$$

$$\mathscr{E}_v(\mathbf{k}) = \mathscr{E}_v(0) + (\hbar^2 k^2/2m_h^*) \tag{5.98}$$

The Hamiltonian for the system of the electron and hole is

$$H = -\frac{\hbar^2 \nabla_e^2}{2m_e^*} - \frac{\hbar^2 \nabla_h^2}{2m_h^*} - \frac{e^2}{\varepsilon|\mathbf{r}_e - \mathbf{r}_h|} \tag{5.99}$$

This is just the same as for the hydrogen atom and we can immediately write down the solution

$$\mathscr{E}(\text{exciton}) = \mathscr{E}_G - \frac{\mu(e/\varepsilon)^2}{2\hbar^2 n^2} + \frac{\hbar^2 K^2}{2M^*} \tag{5.100}$$

where

$$1/\mu = (1/m_e^*) + (1/m_h^*) \tag{5.101}$$

$$M^* = m_e^* + m_h^* \tag{5.102}$$

and n is the principal quantum number. $\mathscr{E}_G$ is the band gap energy, μ is the reduced mass, and M^* is the effective mass of the exciton. Thus, there exists a series of hydrogenlike energy levels. These excitonic levels cannot rigorously be included in a one-electron energy band diagram. They are, nonetheless, shown in band diagrams and it is useful to do so. By convention, these levels are shown as existing below the conduction band as in Figure 5.29. The levels become closer together for higher energies and finally merge into a continuum. The continuum is taken as the bottom of the conduction band, where the electron is free.

The absorption spectrum of a thin film of Cu_2O is shown in Fig. 5.30. Because of the dielectric screening between electron and hole, the radius of the $n = 3$ exciton state for Cu_2O is 140Å. This large distance is sufficient to justify the use of dielectric screening in the model. The symmetry of the valence and conduction bands in cuprous oxide is such as to result in a very weak oscillator strength for transitions to the ground state ($n = 1$) of the exciton. The uppermost valence band is d-like and the lowest conduction band is s-like. The probability of creating s-like hydrogenic excitons under such conductions is very small. The oscillator strength for the $n = 1$

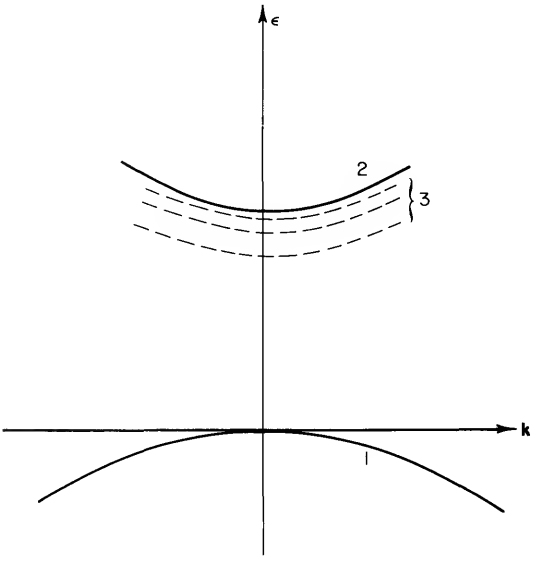

Fig. 5.29 Exciton levels for direct transitions. (1) Valence band, (2) conduction band, (3) exciton levels. A similar diagram can be drawn for excitons associated with indirect transitions.

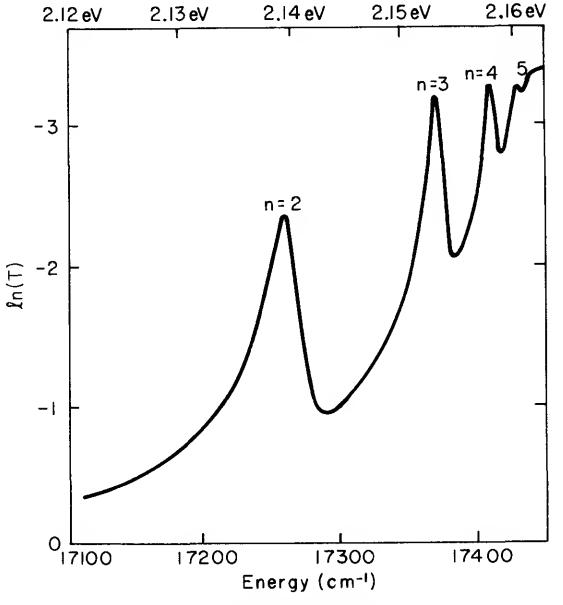

Fig. 5.30 The logarithm of the transmission as a function of photon energy of a Cu_2O sample at 77°K showing the details of the yellow series of exciton lines [From P. W. Baumeister, *Phys. Rev.* **121**, 359 (1961).]

transition in Cu_2O is observed to be about 10^{-9}. However, for higher excited states of the exciton, the hole can be d-like and the electron p-like, and transitions can at least be of the dipole type rather than of the quadrupole type. Thus, dipole transitions can be expected to begin at $n = 2$. Even so, the oscillator strengths are small, $\sim 10^{-6}$, because d holes are quite localized and the p electron has no amplitude at the origin, so the overlap is small.

We have mentioned that for a Wannier exciton, the electron and hole interact via a screened Coulomb interaction. But what dielectric constant should be used? This can be determined from an estimate of the angular frequency of the exciton as found by equating angular momentum to Planck's constant. Thus,

$$\omega = \hbar/\mu r^2 \tag{5.103}$$

If $\omega < \omega_0$, where ω_0 is an optical mode vibrational frequency, the lattice can respond and the low-frequency lattice dielectric constant should be used. If $\omega > \omega_0$ the high-frequency lattice dielectric constant should be used, that is, the high-frequency dielectric constant for lattice vibrations, not electronic excitations.

As a final example of excitons, consider Fig. 5.31, the absorption spectrum

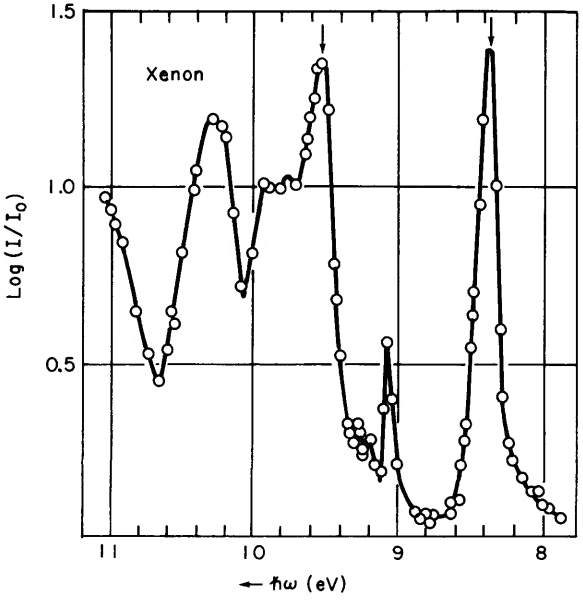

Fig. 5.31 Absorption spectrum of annealed solid xenon at 21°K. [After G. Baldini, *Phys. Rev.* **128**, 1562 (1962).]

of solid xenon. Xenon has excited atomic states at wavelengths of 1470Å and 1295Å. These two wavelengths are indicated by the vertical arrows in Fig. 29. Clearly, xenon as a solid has some of the absorption characteristic of the free atoms. This is not unexpected, but what is the origin of the remaining structure?

The weak lines at 9.1 and 9.2 eV correspond to Wannier excitons with $n = 2$ and 3. The radii (>8Å) are sufficiently large to call these Wannier excitons. For $n = 1$, the radius is so small that the exciton is more properly regarded as a Frenkel exciton; it appears at 8.4 eV, corresponding to the lowest atomic absorption line. These cases are very difficult to treat quantitatively.

We have seen evidence for exciton formation during indirect transitions near the absorption edge in Ge. These are Wannier excitons, but the two levels observed for the indirect transition are not hydrogenlike levels. They are the result of the splitting of a single hydrogenic level. The splitting arises because the valence band edge is degenerate and the conduction band edge consists of four minima in **k**-space. The degenerate valence band leads to degenerate exciton levels. For the direct exciton, the degeneracy is retained. However, for the indirect exciton, the electron can be associated with any of the four minima. The effect is to remove some of the degeneracy for the exciton.

5.9 Direct and Indirect Transitions in Photoemission

The measurement of a photoelectron energy distribution curve (EDC) typically reveals various kinds of structure. This structure is ultimately related to the electronic band structure, but the way it shows up in an EDC depends on the nature of the optical transition. We want to consider how to distinguish between direct and indirect transitions, and how the band structure is revealed in each of these cases.

Let us first consider indirect transitions. We have seen in Section 5.6 that the absorption coefficient for indirect transitions to energy $\mathscr{E}$ from an initial energy $\mathscr{E} - \hbar\omega$ is proportional to the product of the initial density of states and the final density of states. Thus, the probability that an electron will be excited to an energy $\mathscr{E}$ by a photon of energy $\hbar\omega$ is

$$P(\mathscr{E}, \hbar\omega) \propto \eta(\mathscr{E} - \hbar\omega)\eta(\mathscr{E}) \tag{5.104}$$

where $\eta(\mathscr{E})$ is the optical density of states. The optical density of states is defined such that it includes matrix element effects; it is the density of states as determined from photoemission experiments rather than the true band structure density of states. Experiments indicate that there is a close relation-

ship between the two densities of states. However, interest is usually focused on the positions of structure rather than absolute magnitudes. Having realized that there is a difference between the optical density of states and the true density of states, we shall now refer simply to the density of states for either one.

Now, consider a hypothetical metal with the density of states shown in Fig. 5.32. Absorption of photons with energy $\hbar\omega$ excites electrons from filled states below the Fermi level to empty states in the conduction band. If the transitions are indirect, the coupling between densities of states is described by Eq. (5.104). If we assume constant matrix elements, we can give a graphic illustration of this coupling. This is done in Fig. 5.33. Curve (a) in Figure 5.33 shows the density of empty conduction band states. Absorption of photons with energy $\hbar\omega$ excites electrons into these empty states. The valence band states which couple into the empty conduction band states are found by shifting the valence band density of states by $\hbar\omega$ as shown in curve (b). Then, the internal density of photoexcited electrons is found by multiplying curves (a) and (b). The result is curve (c), which is the graphical analog to Eq. (5.104). We see that in this illustration, there are three peaks in the energy distribution of internal photoexcited electrons.

Some of the excited electrons shown in Fig. 5.33(c) may reach the vacuum surface. If they have sufficient energy they may escape. The escape function and electron scattering effects are discussed in more detail later. Here, we note only that electrons reaching the surface have a probability of escaping given by an energy-dependent escape function. This is shown in curve (d) of Figure 5.33. Multiplying curves (c) and (d) thus gives the external distribution of photoelectrons if internal scattering is unimportant. We note that now there are only two peaks in the energy distribution. The low-energy peak in the internal distribution of photoexcited electrons lies below the vacuum level. Those electrons cannot escape.

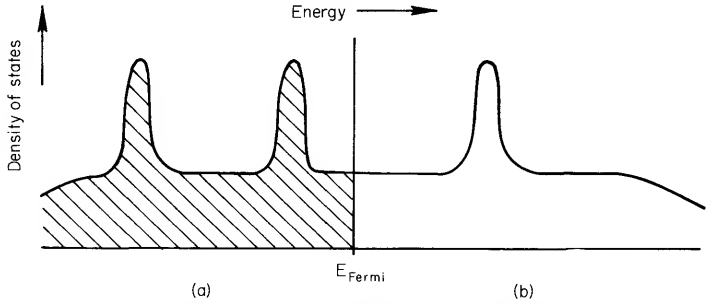

Fig. 5.32 Density of states for a hypothetical metal. (a) Valence band (filled states), (b) conduction band (empty states). [From A. J. Blodgett, Jr. and W. E. Spicer, *Phys. Rev.* **146**, 390 (1966).]

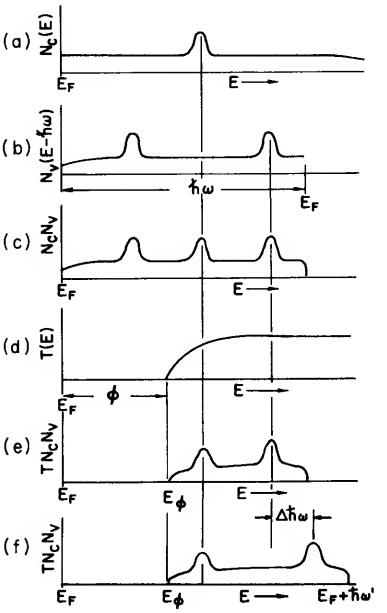

Fig. 5.33 Relationship between density of states (taken from Fig. 5.32) and photoelectron energy distribution for indirect transitions with constant matrix elements. (a) Conduction band (empty), (b) valence band (filled), (c) photoexcited electrons, (d) escape function, (e) photoemitted electrons ($\hbar\omega$), (f) photoemitted electrons ($\hbar\omega'$). [From A. J. Blodgett, Jr. and W. E. Spicer, *Phys. Rev.* **146**, 390 (1966).]

We could measure the EDC for curve (e) of Fig. 5.33, but what could we say about the two peaks? We could say very little, only that there is clearly some structure in the density of states. However, if we repeated the experiment with a higher photon energy, $\hbar\omega'$, we could already learn quite a lot. We could measure curve (f). Comparison of curves (e) and (f) reveals one peak in the EDC that is fixed in energy and another peak that has moved in energy by an amount $\Delta\hbar\omega = \hbar\omega' - \hbar\omega$. We can then easily deduce that the peak that is fixed in energy corresponds to a high density of states in the empty conduction band. Increasing the photon energy simply means the electrons excited into the fixed peak originate from lower-lying levels. The peak that moves in energy by an amount exactly equal to the change in photon energy corresponds to a peak in the density of filled valence band states. Structure in the EDC that obeys this equal-increment rule can usually be attributed to indirect transitions from the valence band.

Direct transitions show up in the EDC as structure which abruptly appears and disappears at different photon energies and which does not

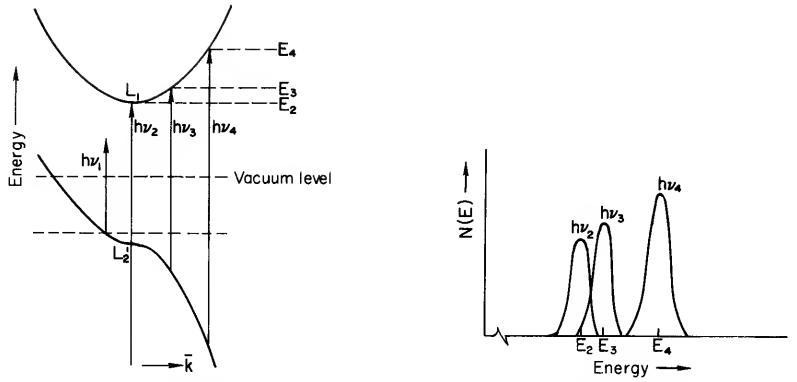

Fig. 5.34 Direct transitions in copper and their contributions to photoemission. The important feature of the peak is that it moves to higher energy at a somewhat slower rate than the increase in photon energy. Note that $E_3 - E_2 \neq h\nu_3 - h\nu_2$ and $E_4 - E_3 \neq h\nu_4 - h\nu_3$. [From C. N. Berglund and W. E. Spicer, *Phys. Rev.* **136**, A1030 (1964).]

obey the equal increment rule. That an increase in photon energy by $\Delta\hbar\omega$ does not result in a peak which moves to higher energy by $\Delta\hbar\omega$ is illustrated in Fig. 5.34. The restriction to vertical transitions means that a change in photon energy requires both a new initial state and a new final state. This causes shifts such as illustrated in Fig. 5.34 as well as the abrupt appearance and disappearance of transitions.

Even structure that obeys the equal-increment rule may result from direct transitions. This is so for flat bands, a common situation for solids with d electrons. Thus, one cannot always tell from photoemission measurements the difference between direct transitions and indirect transitions to or from flat bands. One way one might think of distinguishing between direct and indirect transitions is to measure the temperature dependence and absolute magnitude of the quantum yield. This has been done for some materials and it has led to the suggestion that there is a class of transitions that behave like indirect transitions insofar as they are described by Eq. (5.104), but they are temperature-independent and of a strength characteristic of allowed direct transitions. These transitions have been called nondirect to distinguish them from the usual indirect transitions. They do not seem to fit so readily into the conventional one-electron model.

5.10 Nondirect Transitions: Photoemission from Cs_3Bi

Cs_3Bi is representative of a class of semiconductors made from alkali metals and antimony or bismuth. These materials, although even now relatively unknown as semiconductors, have been of great significance.

They have been used for commerical photocathodes since the 1930's. An interesting historical point is that the development of photoemission as a tool for fundamental band structure studies followed from the study of these photocathodes, all done in industrial laboratories. The same sensitivity to visible light and high quantum yield (0.1 electron/photon) that made them ideal practical photocathode materials also permitted basic experiments on photoemission to be performed with relative ease. Studies on these materials also introduced a feature which is not yet fully resolved, namely, nondirect transitions.

On the basis of energy-distribution measurements for Cs_3Sb, Apker, Taft, and Dickey recognized that there is structure in the energy distribution and that this structure is related to the band gap [23]. Others, especially Spicer, Sommer, and Philipp, continued with studies of these materials [24–28]. Spicer, in particular, recognized the potential usefulness of photoemission as a tool for the study of electronic band structure. However, to study most materials which have photoemission thresholds in the vacuum ultraviolet (VUV), it was necessary to develop techniques for working conveniently in this spectral region. The first VUV studies on photoelectron energy distributions and their relationship to electronic band structure were reported in 1963, for copper [29] and the semiconductors Cs_3Bi, Cs_3Sb, Na_3Sb, and K_3Sb [30]. This was quickly followed by the comprehensive studies on copper and silver by Berglund and Spicer [31], and hundreds of other studies since then.

A photoelectron energy distribution curve for Cs_3Bi is shown in Fig. 5.35. The low-energy peak is due to scattered electrons. It remains fixed in position, but increases in magnitude as the photon energy is increased. The two high-energy peaks move to higher energy with increasing photon energy, thus indicating that this structure in $N(\mathscr{E})$ arises from structure in the valence band density of states. These two peaks are due to spin–orbit splitting of the 6p bismuth atomic orbitals [32]. The magnitude of the spin–orbit splitting in the solid is approximately equal to the atomic spin–orbit splitting because the bismuth atoms are well separated from each other in the solid. Similar results are found for the entire class of semiconductors made from alkali metals and antimony or bismuth. In all cases, the magnitude of the spin–orbit splitting as well as the multiplicity is in agreement with theory. In the cubic materials, two peaks are observed; in the hexagonal materials, the highest-energy peak is further split into a doublet.

That the two high-energy peaks in $N(\mathscr{E})$ for Cs_3Bi are indeed due to some kind of indirect transitions from the valence band can be shown graphically by plotting the energy distributions versus the energy in the valence band from which the electrons are excited. This is done in Fig. 5.36 for a number of different photon energies. Spicer called these transitions nondirect [32,

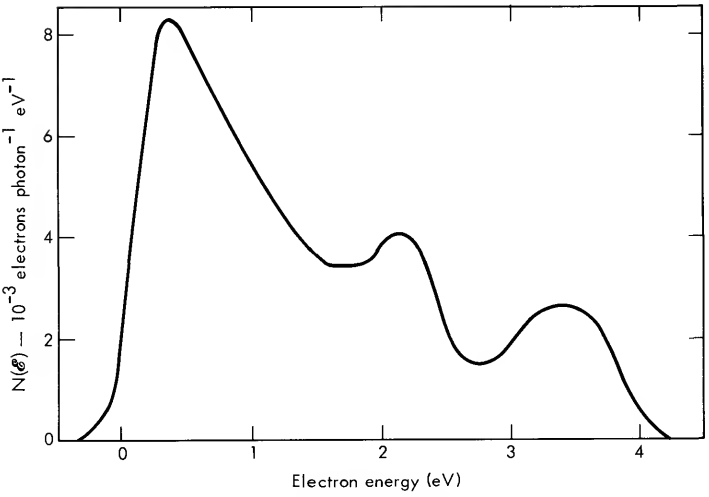

Fig. 5.35 Photoelectron energy distribution for Cs_3Bi for $h\nu = 5.9$ eV [30].

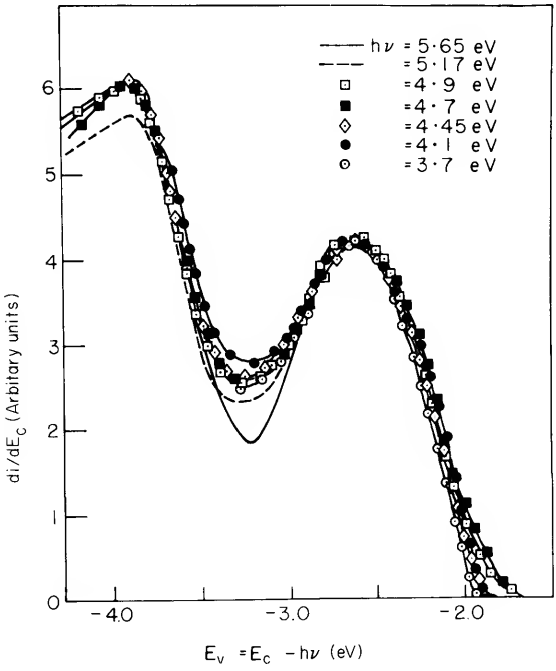

Fig. 5.36 Energy distributions for Cs_3Bi plotted versus the energy in the valence band $\mathscr{E}_v$ from which the electrons are excited. This is done by subtracting the photon energy $h\nu$ from the external kinetic energy of the photoelectrons. Energy is measured from the vacuum level. The values of $h\nu$ indicate the energy of the incident radiation. [From W. E. Spicer, *Phys. Rev. Lett.* **11**, 243 (1963).]

33]. He introduced this term to distinguish them from indirect transitions. They cannot be the usual indirect transitions involving phonons because the optical absorption is large ($\alpha > 10^5$ cm^{-1}) and independent of temperature. On the other hand the bands are not flat, as can be seen from the widths of the structure in the energy-distribution curves. Thus, it seems that there is no dependence on **k** in the optical absorption process.

It seems clear that **k** conservation will be unimportant for materials with low hole mobility. Then the hole must be fairly localized and the wave function must contain all **k** if Bloch functions are used. Highly ionic materials should be expected to exhibit such behavior. However, it was the appearance of nondirect transitions in nearly all the early photoelectron energy distribution measurements that seemed to strike at the heart of band theory and continues to be the source of some controversy.

More recently, measurements and experimental techniques have been improved and direct transitions have been observed in most materials [34, 35]. Nonetheless, there is still in many cases a large background in the energy distribution that is most easily explained on the basis of nondirect transitions. However, we shall leave this to the interested reader to follow in the current literature. After discussing transport effects in the next section, we shall return to a discussion of photoemission results that can be explained in terms of conventional direct and indirect transitions.

5.11 Transport and Escape Cone Effects on Photoemission

For purposes of determining electronic band structure, one would like the external photoelectron energy distribution to be a representative sample of the internal distribution of excited electrons. However, between excitation and escape many things may happen which lead to distortion of the idealized EDC. We shall consider several of the more important possibilities.

There are two dominant scattering mechanisms in an ideal crystal. They are electron–phonon scattering and electron–electron scattering. Electron-phonon scattering is nearly elastic and is usually assumed to be isotropic. It randomizes the internal angular distribution of photoelectrons and increases the average total path length traveled by an electron for a particular net displacement. Electron–phonon scattering thus increases the probability that an electron will suffer an electron–electron collision before it reaches the vacuum surface. In many cases, the mean free path for electron–phonon scattering, l_p, is large compared with that for electron–electron scattering, l_e. Then, it can be neglected. When it must be included in the analysis, it can usually be assumed constant. That is because it is usually important only over the range of electron energies for which $l_p < l_e$. This is usually a small energy range and an average value of l_p can be assumed.

Electron–electron scattering is assumed to be such that all possible transitions are equally probable (see Fig. 5.37). The transition matrix elements are taken to be independent of initial and final states for all energy-conserving transitions. There is evidence that such an assumption is valid to first order. The experimental evidence is that such an assumption seems to have worked quite well for nearly all photoemission experiments that have been analyzed on this basis [36–38]. Ambiguity as to the validity of the assumption arises because of the inability to accurately separate out the effects of varying matrix elements for the initial optical excitation. Theoretical calculations of electron–electron scattering in silicon using momentum- and energy-dependent matrix elements gave excellent agreement with the simpler model described here [39]. The agreement probably results from the averaging over so many possible transitions.

Assuming the simple model for electron–electron scattering just described, the probability that an electron with energy $\mathscr{E}$ will be scattered to an

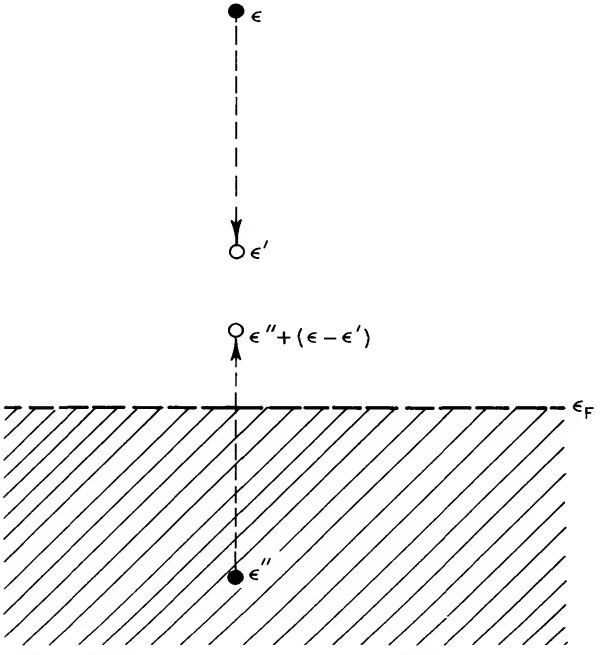

Fig. 5.37 An electron of energy $\mathscr{E}$ is assumed to be able to scatter to any state of lower energy with equal probability. Thus the probability of being scattered to energy between $\mathscr{E}'$ and $\mathscr{E}' + d\mathscr{E}'$ is proportional to $\rho(\mathscr{E}')\, d\mathscr{E}'$. In losing the energy $\mathscr{E} - \mathscr{E}'$, the primary electron excites a secondary electron from an energy $\mathscr{E}''$ below the Fermi energy to an unoccupied state above the Fermi energy. Since the secondary electron absorbs the energy $\mathscr{E} - \mathscr{E}'$ lost by the primary, the final state energy of the secondary electron is $\mathscr{E}'' + (\mathscr{E} - \mathscr{E}')$. All possible primary losses and secondary excitations must be considered.

energy between $\mathscr{E}'$ and $\mathscr{E}' + d\mathscr{E}'$ is then

$$P_s(\mathscr{E}', \mathscr{E})\, d\mathscr{E}' = C\left[\int_{\mathscr{E}_F-(\mathscr{E}-\mathscr{E}')}^{\mathscr{E}_F} \rho(\mathscr{E}'')\rho(\mathscr{E}'' + [\mathscr{E} - \mathscr{E}'])\, d\mathscr{E}''\right]\rho(\mathscr{E}')\, d\mathscr{E}' \tag{5.105}$$

where C is a constant. The integration over $\mathscr{E}''$ is such as to include all possible energy-conserving secondary excitations. The probability that an electron with energy $\mathscr{E}$ will be scattered to some other energy is then just

$$P_S(\mathscr{E}) = \int_{\mathscr{E}_F}^{\mathscr{E}} P_S(\mathscr{E}', \mathscr{E})\, d\mathscr{E}' \tag{5.106}$$

The constant C is chosen to give results consistent with quantum yield and shape of the EDC for all photon energies.

It is clear from Eqs. (5.105) and (5.106) that electron–electron scattering must be strongly energy-dependent. In metals, l_e can vary over several orders of magnitude, decreasing to 10–20 Å for electrons with energy about 10 eV above the Fermi energy. For semiconductors, the energy dependence is even more striking. An excited electron must have an energy in excess of the conduction band minimum by at least the band gap in order to be able to excite a secondary electron. This is because the primary electron must lose energy equal to or greater than the band gap energy in order to excite a secondary electron from the valence band to the conduction band. Thus, l_e is infinite at energies less than $\mathscr{E}_G + \mathscr{E}_c$, where $\mathscr{E}_G$ is the band gap energy and $\mathscr{E}_c$ is the conduction band minimum. It falls to tens of angstroms several band gap energies above $\mathscr{E}_c$. Requirements of momentum conservation may take the threshold for pair production several times the band gap energy.

The effect of electron–electron scattering in producing a low-energy peak in the EDC is illustrated in Fig. 5.38 for a semiconductor. For all materials, those electrons that have suffered electron–electron scattering usually undergo large energy losses. Thus, the EDC is usually easy to separate into two regions, a high-energy region which retains the structure characteristic of the initial optical excitation and a low-energy peak dominated by scattered electrons.

During the course of its motion through the crystal, an excited electron may reach the vacuum surface. If its normal component of momentum corresponds to an energy greater than the surface barrier height, it may escape. Most analyses of photoemission are based on a phenomenological model that utilizes the concept of an escape cone. The concept of an escape cone can be understood as follows. Consider an electron with momentum $\hbar\mathbf{k}$ traveling toward the vacuum surface at an angle θ with respect to the

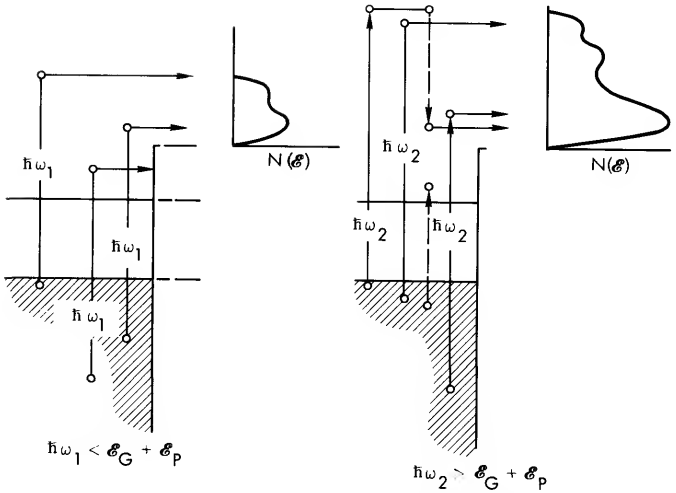

Fig. 5.38 Left: A hypothetical EDC for a semiconductor with $\hbar\omega_1 < \mathscr{E}_G + \mathscr{E}_p$, where $\mathscr{E}_p$ is the threshold energy for pair production. Here, the only electrons escaping are those which have suffered no scattering or only nearly elastic scattering. The structure in the EDC is characteristic of that in the initial optical excitation. Right: An EDC for photon energy $\hbar\omega_2 > \hbar\omega_1$ such that $\hbar\omega_2 > \mathscr{E}_G + \mathscr{E}_p$. Some electrons can escape even after having suffered large energy losses. These electrons contribute to a low-energy peak which often consists mostly of scattered electrons. The peaks at higher energy contain few scattered electrons and thus are representative of the structure in the initial optical excitation.

surface normal. On the basis of a free-electron model, the condition for escape is

$$\hbar^2 k_x^2/2m \geqslant \mathscr{E}_F + e\phi \tag{5.107}$$

where $\hbar k_x$ is the normal component of momentum and

$$\cos\theta = k_x/k \tag{5.108}$$

The escape cone for electrons with energy $\mathscr{E} = \hbar^2 k_x^2/2m$ is defined by the critical angle θ_c, which is the maximum value of θ such that an electron can still escape. This critical angle is found with the help of Eqs. (5.107) and (5.108) to be

$$\cos\theta_c = [(\mathscr{E}_F + e\phi)/\mathscr{E}]^{1/2} \tag{5.109}$$

One way in which the escape cone concept is used is to assume that the photoelectrons are excited such that their initial directions are isotropically distributed. Then, the probability that an electron is within the escape cone is given by

$$T_0(\mathscr{E}) = \tfrac{1}{2}\{1 - [(\mathscr{E}_F - e\phi)/\mathscr{E}]^{1/2}\} \tag{5.110}$$

This result is found simply from the ratio of the solid angle included by the escape cone to the total solid angle 4π. The effects of electron scattering are then sometimes included by multiplying this semiclassical escape function by an attenuation factor $e^{-x/L}$, where L is the mean attenuation length and x is the depth at which the electron is initially excited. Of course, real solids are not free-electron metals. For real metals, it has often been found in practice that it is sufficient to take the bottom of the bands, i.e., the bottom of the potential well, at the Fermi level for purposes of calculating a threshold function. Then, Eq. (5.110) is transformed to

$$T(\mathscr{E}) \approx \tfrac{1}{2}\{1 - [e\phi/(\mathscr{E} - \mathscr{E}_F)]^{1/2}\} \tag{5.111}$$

This probably works because the bands begin to be free-electronlike in the vicinity of the Fermi energy for many metals. For semiconductors, a similar modification is to put the bottom of the potential well at $\mathscr{E}_c$, the conduction band minimum. Then,

$$T(\mathscr{E}) \approx \tfrac{1}{2}\{1 - [\mathscr{E}_A/(\mathscr{E} - \mathscr{E}_c)]^{1/2}\} \tag{5.112}$$

where $\mathscr{E}_A$ is the electron affinity.

The escape function $T(\mathscr{E})$ is sometimes determined entirely from a self-consistent phenomenological analysis of experimental data [36, 37]. It then implicitly includes effects of electron–phonon scattering since these effects are too difficult to include in approximate analytic expressions. Quantum mechanical reflection effects at the barrier seem to be unimportant in practice.

One can view transport effects as nuisances which make it difficult to obtain the electronic density of states from photoemission data. Alternatively, one can view these effects as providing an opportunity to learn about electron scattering in an energy range not generally accessible. The former viewpoint is more common. However, the hindrance to band structure studies is more one of determining absolute magnitudes rather than the hiding of structure. For indirect transitions, or ones that behave approximately like indirect transitions, even the effects of transport and the escape cone can be eliminated. One can understand this in the following way. Focus attention on a particular energy $\mathscr{E}_f$ in the external energy distribution. If this energy is sufficiently high, the only electrons escaping are those that have, at most, undergone nearly elastic scattering. These electrons originated at some initial energy $\mathscr{E}_i$ within the solid. Now, if the photon energy is increased by $\Delta\hbar\omega$, the electrons emerging at energy $\mathscr{E}_f$ originated at energy $\mathscr{E}_i - \Delta\hbar\omega$. Thus, we can visualize that as the photon energy is increased, we can watch structure pass by. Since we have focused on a particular final energy, the escape cone and transport effects remain constant. Thus, we can determine the valence-band optical density of

states. Most methods of analysis of photoemission are just some analytic variant of this procedure.

5.12 Photoemission and Electron Transport in Al and GaAs

Aluminum is quite well understood. Its band structure is sufficiently well known that the photoemission data can be utilized to determine electron transport properties in an energy range not generally accessible.

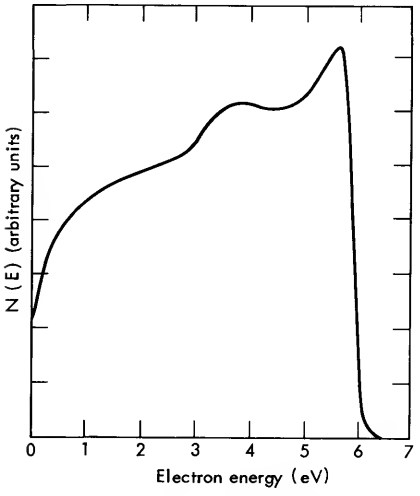

Fig. 5.39 Photoelectron energy distribution curve for Al for $h\nu = 10.2$ eV. The energy distribution was measured within 1 hr after deposition of the Al film. [From T. Huen and F. Wooten, *Solid State Commun.* **9**, 871 (1971).]

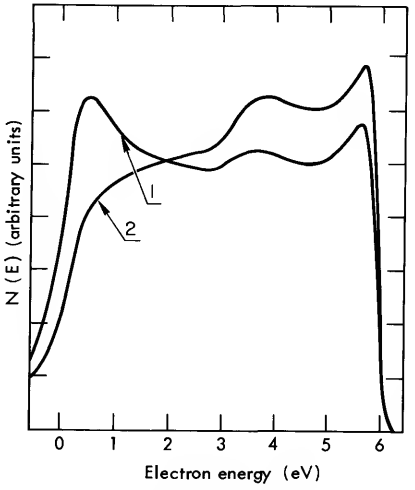

Fig. 5.40 Photoelectron energy distribution curves for Al for $h\nu = 10.2$ eV. (1) after 2–3 hr; (2) after 24 hr. The curves show the effect of aging in an environment at 10^{-10} Torr. [From T. Huen and F. Wooten, *Solid State Commun.* **9**, 871 (1971).]

Figure 5.39 shows a photoelectron energy distribution curve for clean Al. The sample was prepared and maintained in a vacuum chamber with an ambient pressure of 2×10^{-10} Torr. Figure 5.40 shows the effects (in this rather extreme case) of even a small degree of contamination. Fortunately, most materials are not nearly this sensitive to surface contamination. Also, optical reflectance is usually less sensitive to surface contamination than is photoemission.

Photoemission from Al has been analyzed by a Monte Carlo analysis [38, 40]. The procedure was the following. The photoelectrons were as-

sumed to be excited within the solid with an exponential spatial distribution corresponding to the optical absorption coefficient. They were distributed in energy in proportion to the product of initial and final density of states and then started off in random directions. The mean free path for electron–phonon scattering, l_p, was taken as an adjustable parameter. Electron–electron scattering was taken as described by Eq. (5.106). The external energy distribution was then calculated and the parameters l_p and C from Eq. (5.105) were adjusted to give agreement with the EDC and quantum yield. In this manner, the cross section for electron–electron scattering was obtained. The results are shown in Fig. 5.41.

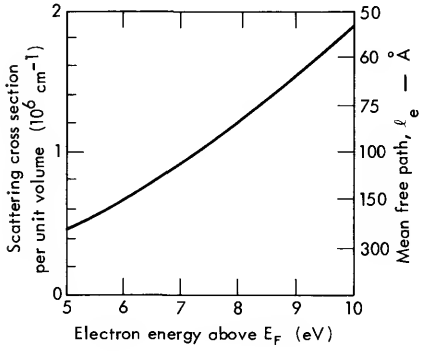

Fig. 5.41 Calculated cross section and mean free path for electron–electron scattering in Al. The mean free path for electron–phonon scattering was taken as 200 Å. [From T. Huen and F. Wooten, *Solid State Commun.* **9**, 871 (1971).]

The analysis outlined here has been based on indirect transitions. The experimental photoemission data are in excellent agreement with such a model. It is also consistent with the free-electronlike optical reflectance of Al. Furthermore, although the temperature dependence of photoemission for Al has not been measured, it has been measured for Zn [41]. Since Zn also exhibits free-electronlike optical reflectance, and since the temperature dependence of photoemission from Zn is significant and consistent with a model involving phonon-assisted indirect transitions [42], it seems probable that the transitions in both Zn and Al are largely indirect even in the vacuum-ultraviolet. Nonetheless, a direct transition analysis is also in fair agreement with experiment and the issue is not finally resolved [43].

Many workers analyze their data in terms of the kind of approach used here for Al. However, the usual method is an analytic approximation to the model. Whereas Monte Carlo methods can handle multiple electron–electron scattering events, analytic approximations are usually restricted to a single electron–electron scattering event per primary photoelectron. Fortunately, this is usually quite adequate since electrons which have suffered more than one inelastic scattering event are usually at too low

an energy to escape. Electron–phonon scattering and second-order electron–electron scattering effects are included implicitly in the empirical threshold escape function $T(\mathscr{E})$. When the density of states is unknown, the analysis consists of an iteration process to reach a selfconsistent result.

Optical transitions in GaAs are direct. Thus, as $\hbar\omega$ increases, peaks in the EDC such as those shown in Fig. 1.4 appear and disappear according to whether the corresponding direct transitions are allowed or not. The energy distribution curves and spectral dependence of quantum yield can be explained without the need to include electron–phonon scattering. This is because the excited electrons with sufficient energy to escape also have sufficient energy to excite secondary electrons. Thus, electron–electron scattering dominates.

If GaAs is coated with a surface layer of Cs_2O, the work function (the difference between vacuum level and Fermi level) can be lowered to 1.2 eV. Thus, heavily doped p-type GaAs:Cs_2O, for which the Fermi level coincides with the valence band maximum, has the conduction band minima above

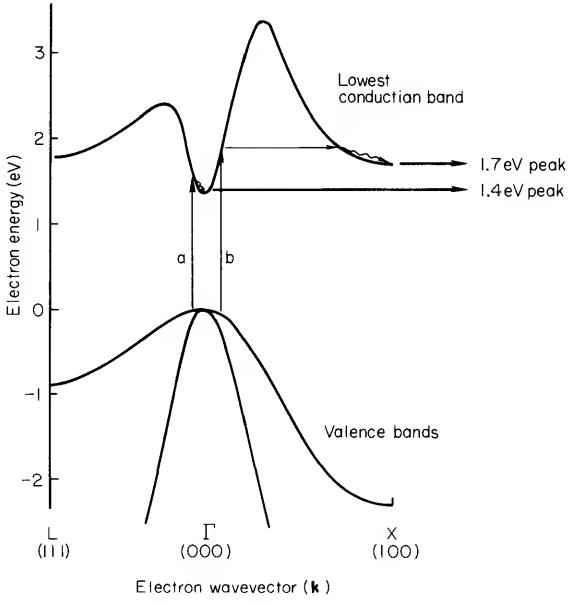

Fig. 5.42 Schematic of the GaAs band structure near the energy gap showing the relevant excitation-escape processes. For photon energies below 1.7 eV, the electrons thermalize in the Γ minimum before escaping into vacuum. For greater photon energies, an increasing number of electrons thermalize in the X minima. [From R. C. Eden, J. L. Moll, and W. E. Spicer, *Phys. Rev. Lett.* **18**, 597 (1967).]

the vacuum level. Then, photoemission from these states can be measured. For these low-energy electrons, the mean free path for electron–electron scattering is infinite and electron–phonon scattering plays an important role. This has important applications for infrared photocathodes [44]. It also has interesting consequences for band structure studies.

Figure 5.42 illustrates what happens in GaAs : Cs_2O with electrons in the energy region just above the conduction band minima. For $1.4 \leq \hbar\omega \leq 1.7$ eV, electrons are excited to states near the conduction band minimum.

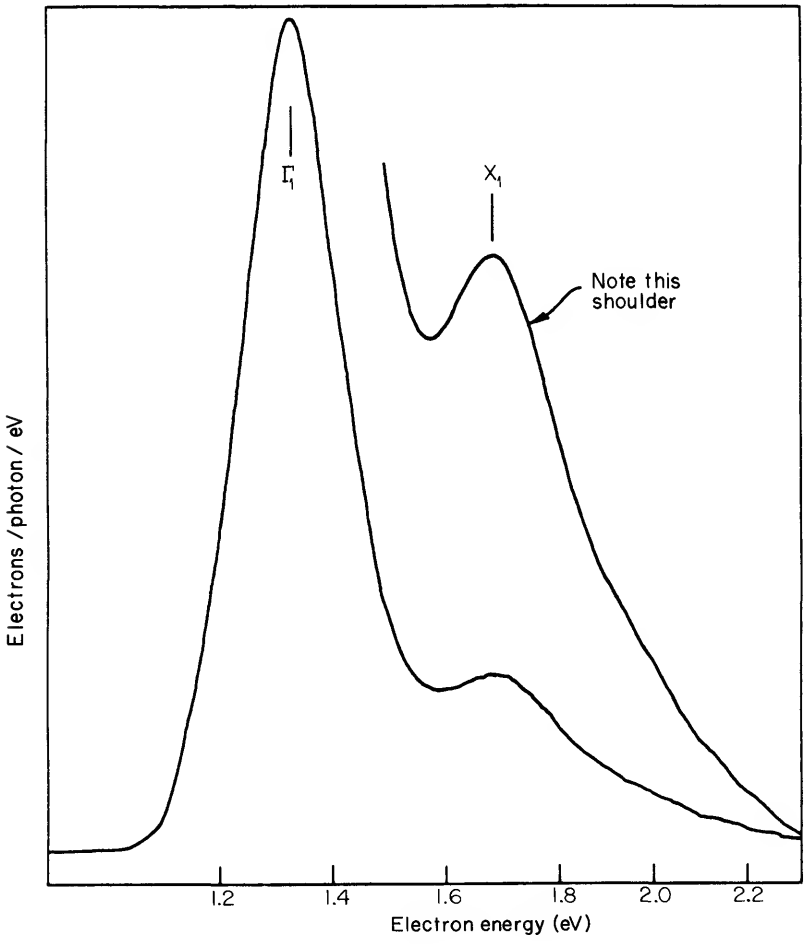

Fig. 5.43 Energy distribution curve for p^+ GaAs:Cs_2O for a photon energy of 2.2 eV. [From L. W. James, R. C. Eden, J. L. Moll, and W. E. Spicer, *Phys. Rev.* **174**, 909 (1968).]

These electrons can escape into the vacuum because the vacuum level lies below the conduction band minima. However, because the optical absorption depth is quite long ($1/\alpha > 5000$ Å) these electrons diffuse to the surface, becoming thermalized in the Γ_1 minimum in the process. For photon energies greater than 1.7 eV, some of the excited electrons become thermalized in the X_1 minima. Thus, even though the optical transitions are direct, the X_1 conduction band minima can also be seen in the photoelectron energy distribution curve. This is shown in Fig. 5.43 for $h\nu = 2.2$ eV.

One might also expect to see a peak arising from the L_1 conduction band minima. It happens, though, that the L_1 and X_1 minima are at nearly the same energy. Thus, the stronger peak, arising from the X_1 minima, dominates. The contribution from the L_1 minima is barely discernible as a shoulder on the X_1 peak. That the shoulder is real has been verified by measuring the second derivative of the I–V curve, i.e., the first derivative of the energy distribution curve. EDC derivative spectroscopy has all the

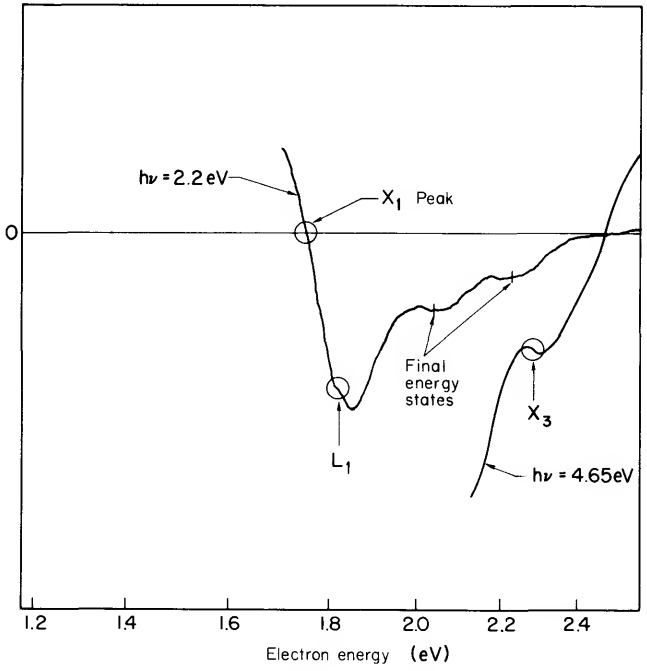

Fig. 5.44 Magnified derivative of EDC for p^+ GaAs:Cs_2O for a photon energy of 2.2 eV. Also included are data for $h\nu = 4.65$ eV showing only that portion corresponding to the X_3 conduction band minima. See Fig. 1.5 for the X_3 minima. [From L. W. James, R. C. Eden, J. L. Moll, and W. E. Spicer, *Phys. Rev.* **174**, 909 (1968).]

advantages already mentioned in Sections 5.3 and 5.4 for optical reflectance derivative spectroscopy. The results for GaAs:Cs_2O are shown in Fig. 5.44.

The X_1 peak is identified in Fig. 5.44 by the point at which the EDC derivative passes through zero. The kink in the derivative curve corresponds to the L_1 minima. In contrast with Fig. 5.43, its location can be quite accurately determined with respect to X_1. That the kink is due to the L_1 minima is suggested by its consistency with band structure calculations and the observation that the kink appears at the same energy over the photon energy range 2.0–4.25 eV. The structure identified as final-energy states corresponds to final states for optical transitions. The electrons contributing to this structure have emerged unscattered. Since the original optical transitions are direct, this structure changes with changing photon energy.

PROBLEMS

5.1 Show that $\Sigma_n \exp[i(\mathbf{k}' + \mathbf{q} - \mathbf{k}'')\cdot \mathbf{R}_n] = 0$ unless $\mathbf{k}' + \mathbf{q} - \mathbf{k}'' = \mathbf{K}$, where $\mathbf{K}$ is a reciprocal lattice vector.

5.2 Under what conditions is it reasonable to assume an oscillator strength of unity in making the estimate of absorption coefficient following Eq. (5.38)?

5.3 Derive the joint density of states J_{cv} for one-dimensional and two-dimensional critical points.

5.4 Derive Eqs. (6)–(8) of N. V. Smith, *Phys. Rev.* **2**, 2841 (1970). Refer to N. F. Mott and H. Jones, "*The Theory of Metals and Alloys*," Chapter 3, Sect. 5. Dover, New York, 1958.

5.5 Assume an isotropic solid with band maxima and minima at $k = 0$. Show that the absorption coefficient for forbidden direct transitions is proportional to $(h\omega - \mathscr{E}_0)^{3/2}/h\omega$, where $\mathscr{E}_0$ is the direct band gap energy. Show that for indirect transitions from the valence band maximum to conduction band minima near the Brillouin zone boundary, the absorption coefficient is given by Eq. (5.91).

5.6 It follows from Eq. (5.5) that the transition probability between two discrete states increases quadratically with time as long as $|(\omega_{ji} - \omega)\tau| \ll 1$. We arrived at a transition probability that increases linearly with time, Eq. (5.9), by integrating over a range of final states. The integration includes transitions which strictly conserve energy between unperturbed states, $\omega_{ji} = \omega$, and transitions for which energy is not conserved, $\omega_{ji} \neq \omega$. The former increase quadratically with time; the latter oscillate periodically. The total result is a transition probability that increases linearly with time.

In light of this understanding of transition rates, read the paragraph preceding Eq. (5.94) and then discuss the validity of the following argument: For indirect transitions, the transition from the intermediate state to the final state must take place in a time $t < \tau_i$. From Eq. (5.94), it is clear that the greater the difference between the direct band gap energy and the photon energy, the shorter will be the lifetime of the intermediate state. We might expect the probability amplitude (or just the probability itself?) for electron–phonon interactions to be proportional to the time available for such an interaction. We would then expect an expression for the transition rate to be proportional to τ^2 and hence contain a term $(\mathscr{E}_0 - \hbar\omega)^2$ in the denominator as in Eq. (5.84).

5.7 Read S. J. Nettel, *Phys. Rev.* **150**, 421 (1966) and H. Ehrenreich, *in* "The Optical Properties of Solids" (J. Tauc, ed.), pp. 106–110, Academic Press, New York, 1966. Discuss the differences in intraband, interband, and Drude absorption as generalized by these authors. Where their viewpoint differs, whose do you favor and why?

5.8 Read G. P. Pells and H. Montgomery, *J. Phys. C Metal Phys. Suppl.* No. 3, p. S330 (1970). Discuss their arguments concerning the importance of nondirect transitions.

5.9 Read N. V. Smith, *Phys. Rev.* **3B**, 1862 (1971). This paper shows that direct transitions occur to a much greater degree than was apparent in the original work of Berglund and Spicer. However, derivative techniques would remove a large, but smooth background due to nondirect transitions. Based on this paper, can you argue that nondirect transitions are unimportant in Cu?

FURTHER READING

D. L. Greenaway and G. Harbeke, "Optical Properties and Band Structure of Semiconductors," Pergamon Press, Oxford, 1968.

J. Tauc, ed,. "The Optical Properties of Solids," Academic Press, New York, 1966.

F. Stern, Elementary Theory of the Optical Properties of Solids, *Solid State Phys.* **15** (1963).

J. C. Phillips, The Fundamental Optical Spectra of Solids, *Solid State Phys.* **18** (1966).

M. Cardona, "Modulation Spectroscopy," Academic Press, New York, 1969.

R. S. Knox, "Theory of Excitrons," Academic Press, New York, 1963.

N. F. Mott and H. Jones, "The Theory of the Properties of Metals and Alloys," Dover, New York, 1958.

S. Nudelman and S. S. Mitra, eds,. "Optical Properties of Solids," Plenum Press, New York, 1969.

R. W. Willardson and A. C. Beer, eds., Optical Properties of III–V Compounds, Vol. 3, "Semiconductors and Semimetals," Academic Press, New York, 1969.

F. Abeles, ed., "Optical Properties and Electronic Structure of Metals and Alloys," North-Holland Publ., Amsterdam, 1966.

R. A. Smith, "Semiconductors," Academic Press, New York, 1963.
R. A. Smith, "Wave Mechanics of Crystalline Solids," Chapman and Hall, London, 1969.
T. S. Moss, "Optical Properties of Semiconductors," Butterworths, London and Washington, D.C., 1961.
D. L. Dexter, "Excitons," Wiley (Interscience), New York, 1965.
W. E. Spicer, Optical Density of States Ultraviolet Photoelectric Spectroscopy, *J. Res. Nat. Bur. Std.* **74A**, No. 3, 397–415 (1970).
N. V. Smith, Photoemission Properties of Metals, *Critical Rev. Solid State Sci.* **2**, 45–83 (1971).
A. V. Sokolov, "Optical Properties of Metals," Blackie, London, 1965.
T. S. Moss, "Optical Properties of Semiconductors," Butterworths, London and Washington, D.C., 1959.
J. N. Hodgson, "Optical Absorption and Dispersion in Solids," Chapman and Hall, London, 1970.

REFERENCES

1. L. van Hove, *Phys. Rev.* **89**, 1189 (1963).
2. J. C. Phillips, *Phys. Rev.* **104**, 1263 (1956).
3. M. Cardona, "Modulation Spectroscopy," Academic Press, New York, 1969.
4. L. R. Saravia and D. Brust, *Phys. Rev.* **176**, 915 (1968).
5. R. H. Parmenter, *Phys. Rev.* **100**, 573 (1955).
6. F. Herman *et al.*, *in* "Methods on Computational Physics" (B. Alder, S. Fernbach, and M. Rotenberg, eds.), p. 193. Academic Press, New York, 1968.
7. L. M. Roth and B. Lax, *Phys. Rev. Lett.* **3**, 217 (1959).
8. R. R. L. Zucca and Y. R. Shen, *Phys. Rev.* **1B**, 2668 (1970).
9. H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962).
10. B. R. Cooper, H. Ehrenreich, and H. R. Philipp, *Phys. Rev.* **138**, A494 (1965).
11. G. B. Irani, T. Huen, and F. Wooten, *Phys. Rev.* **3B**, 2385 (1971).
12. E. L. Green and L. Muldauer, *Phys. Rev.* **2**, 330 (1970).
13. R. M. Morgan and D. W. Lynch, *Phys. Rev.* **172**, 628 (1968).
14. H. G. Liljenvall and A. G. Mathewson, *J. Phys. C: Metal Phys. Suppl.* No. 3, S341 (1970).
15. F. M. Mueller and J. C. Phillips, *Phys. Rev.* **157**, 600 (1967).
16. C. N. Berglund and W. E. Spicer, *Phys. Rev.* **136**, A1030, A1044 (1964).
17. C. E. Morris and D. W. Lynch, *Phys. Rev.* **182**, 719 (1969).
18. F. Herman and S. Skillman, *Proc. Int. Conf. Semicond. Prague, 1960* (J. Bardeen, ed.), p. 20. Czechoslovakian Acad. of Sci., Prague, 1961.
19. J. C. Phillips, *Phys. Rev.* **125**, 1931 (1962).
20. R. L. Jacobs, *J. Phys. C Solid State Phys.* **1**, 1296 (1968).
21. R. J. Elliott, *Phys. Rev.* **108**, 1384 (1957).
22. G. G. MacFarlane, T. P. MacLean, J. E. Quarrington, and V. Roberts, *Phys. Rev.* **108**, 1377 (1957).
23. L. Apker, E. Taft, and J. Dickey, *J. Opt. Soc. Amer.* **43**, 78 (1953).
24. W. E. Spicer, *Phys. Rev.* **112**, 114 (1958).
25. W. E. Spicer, *J. Appl. Phys.* **31**, 2077 (1960).
26. A. H. Sommer, "Photoemissive Materials," Wiley, New York, 1968.
27. E. Taft and H. Philipp, *Phys. Rev.* **115**, 1583 (1959).
28. H. Philipp, E. A. Taft, and L. Apker, *Phys. Rev.* **120**, 49 (1960).
29. C. N. Berglund and W. E. Spicer, *Bull. Amer. Phys. Soc.* **8**, 613 (1963).
30. W. E. Spicer, J. P. Hernàndez, and F. Wooten, *Bull. Amer. Phys. Soc.* **8**, 614 (1963).
31. C. N. Berglund and W. E. Spicer, *Phys. Rev.* **136**, A1030, A1044 (1964).
32. W. E. Spicer, *Phys. Rev. Lett.* **11**, 243 (1963).

33. W. E. Spicer, *Phys. Lett.* **20**, 325 (1966).
34. D. E. Eastman and J. K. Cashion, *Phys. Rev. Lett.* **24**, 310 (1970).
35. N. V. Smith, *Phys. Rev.* **3B**, 1862 (1971).
36. W. F. Krolikowski and W. E. Spicer, *Phys. Rev.* **185**, 882 (1969).
37. A. J. Blodgett, Jr. and W. E. Spicer, *Phys. Rev.* **146**, 390 (1966).
38. F. Wooten, T. Huen, and R. N. Stuart, *in* "Optical Properties and Electronic Structure of Metals and Alloys" (F. Abeles, ed.) North-Holland Publ., Amsterdam, 1966.
39. E. O. Kane, *J. Phys. Soc. Japan Suppl.* **21**, 37 (1966); *Phys. Rev.* **159**, 624 (1967).
40. T. Huen and F. Wooten, *Solid State Commun.* **9**, 871 (1971).
41. L. P. Mosteller, T. Huen, and F. Wooten, *Phys. Rev.* **184**, 364 (1969).
42. L. P. Mosteller, Jr. and F. Wooten, *Phys. Rev.* **171**, 743 (1968).
43. R. Y. Koyama and N. V. Smith, *Phys. Rev.* **2B**, 3049 (1970).
44. R. L. Bell and W. E. Spicer, *Proc. IEEE* **58**, 1788 (1970).

Chapter 6

DISPERSION RELATIONS AND SUM RULES

A dispersion relation is an integral formula relating a dispersive process to an absorption process. An example of dispersion is the dependence of refractive index n on energy $\hbar\omega$. It leads, e.g., to the angular dispersion of white light by a prism and to the variation in velocity with wavelength. A dispersion relation relates n to the extinction coefficient k.

In this chapter, we shall show how dispersion relations follow rigorously from the requirement of causality. Causality means there can be no effect before the cause. Thus light cannot be reflected or absorbed by a system before the arrival of the primary light wave. The dispersion relations that can be derived for linear systems subject to causality are quite general and very useful. They can be used to derive sum rules and to analyze experimental reflectivity data to obtain the optical constants. We shall consider examples of both of these applications of dispersion relations.

6.1 Linear Response Functions and Kramers-Kronig Relations

Under the action of an external stimulus, a system responds in its own characteristic way. The relationship of the response to the stimulus is given by a response function. We have already encountered several examples of response functions. A particular example is the Green's function given by Eq. (4.8). It describes the increment in velocity of an electron as the response to an electric field. In general, the induced response to an external

stimulus can be written

$$X(\mathbf{r}, t) = \int_{-\infty}^{\infty} G(\mathbf{r}, \mathbf{r}', t, t') f(\mathbf{r}', t')\, d\mathbf{r}'\, dt' \tag{6.1}$$

for a linear system. Equation (6.1) describes the response $X(\mathbf{r}, t)$ of the system at location $\mathbf{r}$ and time t to stimuli $f(\mathbf{r}', t')$ acting at all times t' and places $\mathbf{r}'$. The function $G(\mathbf{r}, \mathbf{r}', t, t')$ is called the response function.

The stimulus of interest here will generally be an electromagnetic wave. We shall assume that the wavelengths of the variable fields are sufficiently long that spatial dispersion can be neglected. This is not true for the anomalous skin effect, as we have already seen. It is also not true for plasma oscillations and the effects of impurities. However, spatial dispersion is unimportant for interpreting most of the fundamental optical spectra in the visible and ultraviolet. Neglecting spatial dispersion is the same as making a local approximation. That is, it assumes that what happens at a particular place depends only on the fields existing at that place. If we make this assumption, and realize also that the flow of time is uniform (or assume it, in order to avoid philosphical problems), then

$$G(\mathbf{r}, \mathbf{r}', t, t') = \delta(\mathbf{r} - \mathbf{r}')\, G(t - t') \tag{6.2}$$

and Eq. (6.1) becomes

$$X(t) = \int_{-\infty}^{\infty} G(t - t') f(t')\, dt' \tag{6.3}$$

If we include the additional requirement that the system be causal, that is, that there can be no response before there is a stimulus, then

$$G(t - t') \equiv 0, \qquad t < t' \tag{6.4}$$

Equation (6.4) follows directly from Eq. (6.1) if a δ-function stimulus is assumed. Taking $f(t') = \delta(t' - t_0)$,

$$X(t) = \int G(t - t')\, \delta(t' - t_0)\, dt'$$

$$= G(t - t_0) \tag{6.5}$$

Equation (6.4) then follows directly from the requirement of causality. We thus see that the response function is a Green's function which describes the response of the system at time t to a stimulus at some earlier time t'.

As is so often the case with folded integrals such as in Eq. (6.3), it is convenient to make a Fourier transformation to convert the folded products

into simple products. The Fourier transforms can be written

$$f(\omega) = \int f(t) \exp(i\omega t)\, dt \tag{6.6}$$

$$X(\omega) = \int X(t) \exp(i\omega t)\, dt \tag{6.7}$$

$$G(\omega) = \int G(t - t') \exp[i\omega(t - t')]\, dt \tag{6.8}$$

From Eqs. (6.7) and (6.3), we get

$$\begin{aligned} X(\omega) &= \int dt\, (\exp i\omega t) \left[\int G(t - t') f(t')\, dt' \right] \\ &= \int dt'\, f(t') \left[\int G(t - t')\, (\exp i\omega t)\, dt \right] \\ &= \int dt'\, f(t')\, (\exp i\omega t') \left\{ \int G(t - t')\, \exp i\omega(t - t')\, dt \right\} \end{aligned} \tag{6.9}$$

Substituting on the RHS from Eqs. (6.6) and (6.8), we get

$$X(\omega) = G(\omega) f(\omega) \tag{6.10}$$

Thus, in terms of Fourier transforms, a monochromatic stimulus $f(\omega)$ is just multiplied by some number $G(\omega)$ to give the response $X(\omega)$.

If $G(\omega)$ has a pole at $\omega = \omega_0$, i.e., if the denominator of $G(\omega)$ vanishes at $\omega = \omega_0$, then there can be a finite response in the absence of a stimulus. What this means is that ω_0 corresponds to the angular frequency of a normal mode of the system. We have already seen an example of this in our earlier discussions of plasma oscillations. There, we found the condition [Eq. (3.36)]

$$\mathbf{D} = \mathbf{E}^{\text{ext}} = \hat{\varepsilon}\mathbf{E} = 0 \tag{6.11}$$

if a plasma oscillation is to sustain itself in the absence of an external field. Note, though, that Eq. (6.11) is not written in the same form as Eq. (6.10). If we rewrite Eq. (6.11) as

$$\mathbf{E} = \hat{\varepsilon}^{-1}\mathbf{E}^{\text{ext}} \neq 0 \tag{6.12}$$

then we see that $\hat{\varepsilon}^{-1}$ is the response function. Since the poles of the response function give the normal modes or elementary excitations of the system, it is the zeros of the dielectric function that give these natural frequencies.

There are some fundamental and very important properties of the response function. They include relationships between the real and imaginary parts of response functions that have direct applicability to the analysis of experimental data and the derivation of useful sum rules. We shall now develop some of these relationships.

If we let ω be complex, $\hat{\omega} = \omega_1 + i\omega_2$, then Eq. (6.8) becomes

$$G(\hat{\omega}) = \int G(t - t') \exp[i\omega_1(t - t')] \exp[-\omega_2(t - t')]\, dt \tag{6.13}$$

The factor $\exp[i\omega_1(t - t')]$ is bounded at all frequencies; $\exp[-\omega_2(t - t')]$ is bounded only in the upper half-plane for $t - t' > 0$ and only in the lower half-plane for $t - t' < 0$. Thus the requirement of causality, $G(t - t') = 0$ for $t - t' < 0$, requires that the integral [Eq. (6.13)] be evaluated in the upper half-plane.

Now, let ω be on the real axis. Then, from Cauchy's theorem,

$$G(\omega) = \frac{1}{i\pi} \mathscr{P} \int_{-\infty}^{\infty} \frac{G(\omega')\, d\omega'}{\omega' - \omega} \tag{6.14}$$

where $\mathscr{P}$ stands for principal value. This result is obtained by integrating over the contour shown in Fig. 6.1 and assuming that $G(\hat{\omega})$ decreases such that as the radius of the semicircle approaches infinity, the contribution to the integral over the semicircle approaches zero. If we now split $G(\omega)$ into its real and imaginary parts, we get

$$\text{Re } G(\omega) = \frac{1}{\pi} \mathscr{P} \int_{-\infty}^{\infty} \frac{\text{Im } G(\omega')}{\omega' - \omega} d\omega' \tag{6.15}$$

$$\text{Im } G(\omega) = -\frac{1}{\pi} \mathscr{P} \int \frac{\text{Re } G(\omega')}{\omega' - \omega} d\omega' \tag{6.16}$$

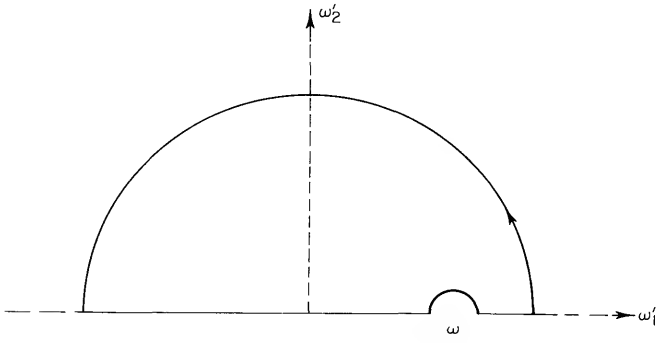

Fig. 6.1 Integration contour for obtaining Eq. (6.14).

We see that the real and imaginary parts of $G(\omega)$ are not independent; they are connected by means of formulas called dispersion relations.

Let us now consider a specific example. We consider a system under the influence of an electric field. The system is polarized by the field and from Eqs. (3.5) and (3.7), we can describe the polarization as

$$\mathbf{P}(\omega) = \frac{Ne^2}{m} \frac{1}{({\omega_0}^2 - \omega^2) - i\Gamma\omega} \mathbf{E}(\omega) \tag{6.17}$$

Comparison with Eq. (6.10) shows that

$$G(\omega) = \frac{Ne^2}{m} \frac{1}{{\omega_0}^2 - \omega^2 - i\Gamma\omega} \tag{6.18}$$

Now, we ask if $G(\omega)$ is causal. We know, of course, there can be no polarization induced before the arrival of the electric field. The question is whether our mathematical description of the system is consistent with causality. If $G(\omega)$ is causal, then, clearly, there are dispersion relations connecting the real and imaginary parts of the polarizability. We now check $G(\omega)$ for causal behavior. Note, though, that here we are using the total electric field as the stimulus, not an external field.

The singularities in $G(\omega)$, as given by Eq. (6.18), occur for

$$\omega \approx -\tfrac{1}{2}i\Gamma \pm \omega_0, \qquad \Gamma^2 \ll {\omega_0}^2 \tag{6.19}$$

Thus Eq. (6.18) becomes

$$G(\omega) = -\frac{Ne^2}{m} \frac{1}{(\omega + \omega_0 + \frac{1}{2}i\Gamma)(\omega - \omega_0 + \frac{1}{2}i\Gamma)} \tag{6.20}$$

The inverse transform of Eq. (6.20) is

$$\begin{aligned} G(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{-i\omega t}\, d\omega \\ &= -\frac{1}{2\pi} \frac{Ne^2}{m} \int_{-\infty}^{\infty} \frac{e^{-i\omega t}}{(\omega + \omega_0 + \frac{1}{2}i\Gamma)(\omega - \omega_0 + \frac{1}{2}i\Gamma)}\, d\omega \end{aligned} \tag{6.21}$$

Now, for $t < 0$, the integral must be evaluated in the upper half-plane where $e^{-i\omega t}$ is bounded. Since there are no singularities in the upper half-plane,

$$G(t) = 0, \qquad t < 0 \tag{6.22}$$

and $G(t)$ and $G(\omega)$ are causal. For $t > 0$, we must evaluate (6.21) in the

lower half-plane. Then,

$$G(t) = -\frac{Ne^2}{2\pi m} 2\pi i \sum (\text{Residues})$$

$$= -\frac{Ne^2}{m} i \left\{ \frac{\exp[-i(-\omega_0 - \frac{1}{2}i\Gamma)t]}{-2\omega_0} + \frac{\exp[-i(\omega_0 - \frac{1}{2}i\Gamma)t]}{2\omega_0} \right\}$$

$$= -\frac{Ne^2}{m\omega_0} (\sin \omega_0 t) \exp\left(-\frac{\Gamma}{2} t\right), \qquad t > 0 \tag{6.23}$$

The polarization as a function of time is

$$\mathbf{P}(t) = \int_{-\infty}^{\infty} G(t - t')\mathbf{E}(t')\, dt'$$

$$= -\frac{Ne^2}{m\omega_0} \int_{-\infty}^{t} \left\{ \exp\left[-\frac{\Gamma}{2}(t - t') \right] \right\} [\sin \omega_0(t - t')]\, \mathbf{E}(t')\, dt' \tag{6.24}$$

Suppose we look at the response to an electric field pulse described by a δ function. Then, if

$$\mathbf{E}(t') = \mathbf{E}_0\, \delta(t') \tag{6.25}$$

Eq. (6.25) yields

$$\mathbf{P}(t) = -\frac{Ne^2}{m\omega_0} \mathbf{E}_0 \left[\exp\left(-\frac{\Gamma}{2} t \right) \right] \sin \omega_0 t, \qquad t > 0 \tag{6.26}$$

The system is polarized at time $t = 0$ by a δ-function impulse and oscillates with frequency ω_0. The amplitude of the polarization decays with a damping constant $\Gamma/2$.

The causality conditions we imposed result in a mathematical description in which no output can occur before an input, but the output may be delayed with respect to the input.

We have demonstrated explicitly that our model for the polarizability is a causal one. Since, from Eq. (3.12), $\hat{\varepsilon} - 1 = 4\pi N\hat{\alpha}$, $\hat{\varepsilon} - 1$ must also be causal. Thus, from quite general principles, we can write down dispersion relations for the dielectric function. From inspection of Eqs. (6.15) and (6.16), we see immediately that

$$\varepsilon_1(\omega) - 1 = \frac{1}{\pi} \mathscr{P} \int_{-\infty}^{\infty} \frac{\varepsilon_2(\omega')}{\omega' - \omega}\, d\omega' \tag{6.27}$$

$$\varepsilon_2(\omega) = -\frac{1}{\pi} \mathscr{P} \int_{-\infty}^{\infty} \frac{[\varepsilon_1(\omega') - 1]}{\omega' - \omega}\, d\omega' \tag{6.28}$$

We are interested in physical systems with real inputs and real outputs. This means we must have

$$G(-\omega) = G^*(\omega) \tag{6.29}$$

In terms of the dielectric function, it means that

$$\hat{\varepsilon}(-\omega) = \hat{\varepsilon}^*(\omega) \tag{6.30}$$

Thus

$$\varepsilon_1(-\omega) = \varepsilon_1(\omega) \tag{6.31}$$

and

$$\varepsilon_2(-\omega) = -\varepsilon_2(\omega) \tag{6.32}$$

We can use these results to rewrite the dispersion relations in the more usual form in terms of integrals over positive frequencies

$$\varepsilon_1(\omega) - 1 = \frac{2}{\pi} \mathscr{P} \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{(\omega')^2 - \omega^2} d\omega' \tag{6.33}$$

$$\varepsilon_2(\omega) = -\frac{2\omega}{\pi} \mathscr{P} \int_0^\infty \frac{[\varepsilon_1(\omega') - 1]}{(\omega')^2 - \omega^2} d\omega' \tag{6.34}$$

Equations (6.33) and (6.34) are known as the Kramers–Kronig dispersion relations.

We have seen how the condition of causality imposes some general relationships between the real and imaginary parts of any causal function. We have used a purely mathematical treatment. We can also see that some kind of a relationship must exist between real and imaginary parts in terms of a simple picture. Figure 6.2(a) shows a single pulse of radiation incident on a system at time $t = 0$. Suppose the system can absorb only a narrow band of frequencies. Then, showing only a single component for convenience, the absorbed component is as shown in Fig. 6.2(b). What remains is shown in Fig. 6.2(c), which is clearly nonsense. No real physical system can have an output before the arrival of the input signal. What this means is that the response function of the system cannot simply describe the absorption, it must also describe the way in which all other frequency components are shifted in phase so that they cancel the absorbed component for times $t < 0$. Thus the response function must have real and imaginary parts to describe both absorption and phase shifts, and the real part at a single frequency must be related to the imaginary part at all other frequencies and vice versa. There must be dispersion relations to satisfy causality.

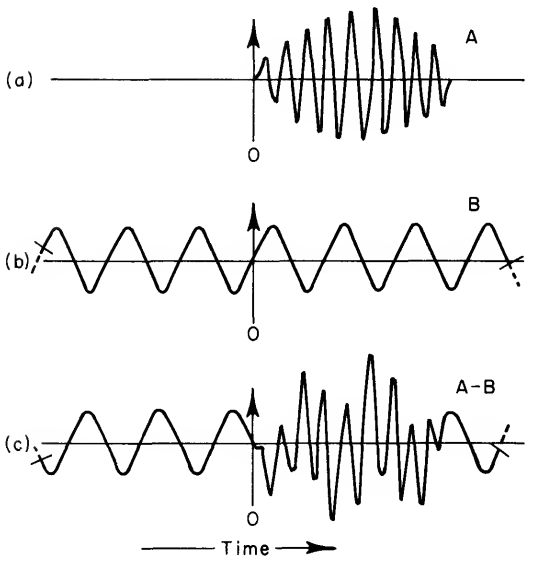

Fig. 6.2 This figure illustrates schematically the basic reason for the logical connection of causality and dispersion. (a) An input A which is zero for times t less than zero is formed as a superposition of many Fourier components (b) such as B, each of which extends from $t = -\infty$ to $t = \infty$. These components produce the zero-input signal by destructive interference for $t < 0$. It is impossible to design a system which absorbs just the component B without affecting other components, for in this case, the output (c) would contain the complement of B during times before the onset of the input wave, in contradication with causality. Thus causality implies that absorption of one frequency must be accompanied by a compensating shift of phase of other frequencies; the required phase shifts are prescribed by the dispersion relation. [From J. S. Toll, *Phys. Rev.* 104, 1760 (1965).]

There are several points to be made with regard to Eq. (6.34). Note that the integral can be broken up into two terms, one arising from $\varepsilon_1(\omega')$ in the integrand and the other from -1. The latter integral is

$$\frac{2\omega}{\pi} \mathscr{P} \int_0^\infty \frac{d\omega'}{(\omega')^2 - \omega^2} = 0 \tag{6.35}$$

Thus Eq. (6.34) could be written

$$\varepsilon_2(\omega) = -\frac{2\omega}{\pi} \mathscr{P} \int_0^\infty \frac{\varepsilon_1(\omega')}{(\omega')^2 - \omega^2} d\omega' \tag{6.36}$$

as it sometimes is. This destroys some of the symmetry between Eqs. (6.33) and (6.34). More important, it is necessary to include $\varepsilon_1(\omega') - 1$ in the integrand rather than just $\varepsilon_1(\omega')$. This is because $\varepsilon_1(\omega') \to 1$ as $\omega' \to \infty$ and thus $[\varepsilon_1(\omega') - 1]/(\omega' - \omega)$ approaches zero more rapidly than $1/\omega'$. This is necessary for the integral over the infinite semicircle to approach

zero and give no net contribution. Even so, Eq. (6.34) is not strictly true for all materials. For metals, there is a contribution $4\pi\sigma/\omega$ to ε_2. This is not bounded as $\omega \to 0$. If we include the residue at $\omega = 0$ arising from this term and follow through with the same steps as before, we get

$$\varepsilon_2(\omega) = \frac{4\pi\sigma_0}{\omega} - \frac{2\omega}{\pi}\mathscr{P}\int_0^\infty \frac{[\varepsilon_1(\omega') - 1]}{(\omega')^2 - \omega^2} \tag{6.37}$$

where σ_0 is the dc conductivity.

6.2 Reflectivity and Phase Shift Dispersion Relations

There are a number of ways to determine optical constants. A common way is to measure the reflectivity at normal incidence and use dispersion relations to determine the optical properties. We will now see how this works.

The reflectivity for normal incidence is given by

$$R(\omega) = \hat{r}(\omega)\hat{r}^*(\omega) \tag{6.38}$$

where

$$\hat{r}(\omega) = (n - 1 + ik)/(n + 1 + ik) \tag{6.39}$$

We can write the complex reflectivity amplitude as

$$\hat{r}(\omega) = \rho(\omega)e^{i\theta(\omega)} \tag{6.40}$$

where now

$$R(\omega) = \rho^2(\omega) \tag{6.41}$$

We shall now assume that $\hat{r}(\omega)$ can be analytically continued into the complex plane and that it has all the necessary mathematical properties for the integral over the appropriate semicircle to approach zero as the radius approaches infinity. This is discussed in Appendix G. Then, if we write Eq. (6.40) as

$$\ln r(\omega) = \ln \rho(\omega) + i\theta(\omega) \tag{6.42}$$

and argue that the reflectance must obey causality, we can look at Eqs. (6.15) and (6.16) and immediately write the dispersion relations as

$$\ln \rho(\omega) = \frac{1}{\pi}\mathscr{P}\int_{-\infty}^{\infty} \frac{\theta(\omega')}{\omega' - \omega}\, d\omega' \tag{6.43}$$

$$\theta(\omega) = -\frac{1}{\pi}\mathscr{P}\int_{-\infty}^{\infty} \frac{\ln \rho(\omega')}{\omega' - \omega}\, d\omega' \tag{6.44}$$

Again, since the input and output functions must be real, we require

$$\hat{r}(-\omega) = \hat{r}^*(\omega) \tag{6.45}$$

This makes it possible to rewrite Eq. (6.44) as an integral over positive frequencies.

$$\theta(\omega) = -\frac{2\omega}{\pi}\,\mathscr{P}\int_0^\infty \frac{\ln \rho(\omega')}{(\omega')^2 - \omega^2}\,d\omega' \tag{6.46}$$

We now need an expression for the phase shift θ; the magnitude of the reflectivity amplitude $\rho(\omega)$ is determined directly from experiment. The integration of Eq. (6.46) must be done numerically. It is also necessary to have some means of extrapolating experimental results over regions for which data are not available. After $\theta(\omega)$ is determined, the optical constants n and k can be determined from

$$\rho(\omega)e^{i\theta} = (n - 1 + ik)/(n + 1 + ik) \tag{6.47}$$

A brief discussion of the numerical integration of Eq. (6.46) and methods of extrapolation are given in Appendix G.

6.3 Sum Rules

We have already encountered sum rules in Section 3.5 and seen their usefulness in several applications in Section 3.6. We will now show how these sum rules follow from the Kramers–Kronig equations.

Equation (6.33) can be written as

$$\varepsilon_1(\omega) - 1 = \frac{2}{\pi}\,\mathscr{P}\int_0^{\omega_c} \frac{\omega'\varepsilon_2(\omega')}{(\omega')^2 - \omega^2}\,d\omega' + \frac{2}{\pi}\,\mathscr{P}\int_{\omega_c}^\infty \frac{\omega'\varepsilon_2(\omega')}{(\omega')^2 - \omega^2}\,d\omega' \tag{6.48}$$

Let ω_c be a cutoff frequency such that there is no absorption at higher frequencies. Then, $\varepsilon_2(\omega) = 0$ for $\omega > \omega_c$. If we determine $\varepsilon_1(\omega)$ for $\omega \gg \omega_c$, then ω' can be neglected in the denominator of the first integral in Eq. (6.48). The second integral is zero because $\varepsilon_2(\omega') = 0$ for $\omega' > \omega_c$. Under these conditions, Eq. (6.48) becomes

$$\varepsilon_1(\omega) = 1 - (2/\pi\omega^2)\int_0^{\omega_c} \omega'\varepsilon_2(\omega')\,d\omega', \qquad \omega \gg \omega_c \tag{6.49}$$

At sufficiently high frequencies, all the electrons can be considered free and the real dielectric function is given by the Drude result, Eq. (3.32),

$$\varepsilon_1(\omega) = 1 - (\omega_p^2/\omega^2) = 1 - (4\pi Ne^2/m) \tag{6.50}$$

where m is the free-electron mass. Comparison of Eqs. (6.49) and (6.50) shows that

$$\int_0^\infty \omega\varepsilon_2(\omega)\,d\omega = \tfrac{1}{2}\pi\omega_p{}^2 \tag{6.51}$$

where the upper limit of integration has been extended to ∞ because $\varepsilon_2(\omega) = 0$ for $\omega > \omega_c$. Equation (6.51) is the sum rule related to the rate of energy absorption by transverse fields (photons).

The sum rule related to the rate of energy absorption by longitudinal fields (electrons) can be obtained in the same way as Eq. (6.51). The real and imaginary parts of the reciprocal of the dielectric function are also related by dispersion relations. By direct analogy with Eq. (6.33), we can write

$$\mathrm{Re}\left(\frac{1}{\hat{\varepsilon}(\omega)}\right) - 1 = \frac{2}{\pi}\mathscr{P}\int_0^\infty \frac{\omega'\,\mathrm{Im}[1/\hat{\varepsilon}(\omega')]}{(\omega')^2 - \omega^2}\,d\omega' \tag{6.52}$$

The number one has been subtracted from $\mathrm{Re}(1/\hat{\varepsilon})$ so as to ensure that the integral along the semicircle vanishes when its radius tends to infinity. That is,

$$\mathrm{Re}\left(\frac{1}{\hat{\varepsilon}}\right) - 1 = \frac{\varepsilon_1}{\varepsilon_1{}^2 + \varepsilon_2{}^2} - 1 \tag{6.53}$$

For frequencies sufficiently high that absorption no longer takes place, $\varepsilon_2(\omega) = 0$, and

$$\begin{aligned}\mathrm{Re}\left(\frac{1}{\hat{\varepsilon}}\right) - 1 &= \frac{1}{\varepsilon_1} - 1 \\ &= \frac{\omega_p{}^2}{\omega^2 - \omega_p{}^2} \to \frac{\omega_p{}^2}{\omega^2}, \qquad \omega \gg \omega_p\end{aligned} \tag{6.54}$$

Following through just as in the derivation of Eq. (6.51), we obtain

$$\int_0^\infty \omega'\,\mathrm{Im}(-1/\hat{\varepsilon}(\omega'))\,d\omega' = \tfrac{1}{2}\pi\omega_p{}^2 \tag{6.55}$$

A sum rule for the real, static dielectric constant is obtained directly from Eq. (6.33) simply by setting $\omega = 0$. Thus

$$\varepsilon_1(0) = 1 + \frac{2}{\pi}\int_0^\infty \frac{\varepsilon_2(\omega')\,d\omega'}{\omega'} \tag{6.56}$$

In the case of a metal, both $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ approach infinity as ω tends

toward zero. For semiconductors and insulators, we see from Eq. (6.56) that the static dielectric constant $\varepsilon_1(0)$ is strongly influenced by interband transitions at low frequencies. We thus expect relatively high static dielectric constants for narrow gap materials. This is what we indeed found to be the case in Section 3.1.

PROBLEMS

6.1 Derive Kramers–Kronig relations between the derivatives $d\varepsilon_1/d\omega$ and $d\varepsilon_2/d\omega$.

6.2 For studies of lattice vibrations by means of infrared reflectance, one takes $\varepsilon_1(\omega) \to \varepsilon_\infty \neq 1$ as $\omega \to \infty$. Here, ε_∞ is the infrared dielectric constant; it is the low-frequency optical dielectric constant $\varepsilon_1(0)$ used for analysis of reflectance in the visible and ultraviolet. How are Eqs. (6.33) and (6.34) modified for the analysis of infrared reflectance?

6.3 What are the Kramers–Kronig relations for the refractive index n and the attenuation coefficient k?

6.4 Show by carrying out an integration by parts, that the dispersion relations (6.33) and (6.34) can be put in the form

$$\varepsilon_1(\omega) - 1 = (1/\pi) \int_0^\infty [d\varepsilon_2(\omega')/d\omega'] \ln[|(\omega')^2 - \omega^2|^{-1}]\, d\omega'$$

$$\varepsilon_2(\omega) = -(1/\pi) \int_0^\infty [d\varepsilon_1(\omega')/d\omega'] \ln([\omega' + \omega]/|\omega' - \omega|)\, d\omega'$$

Discuss the behavior of the reflectivity in the neighborhood of $\omega' = \omega$. What is the effect of a sharp edge in $\varepsilon_2(\omega)$ resulting from the sudden onset of interband transitions?

6.5 Assume a frequency ω sufficiently high that $\omega\tau \gg 1$, so collisions can be neglected. Then, the conductivity $\sigma(\omega)$ is $Ne^2/m\omega$. It is all right to use this formula for numerical values of $\sigma(\omega)$ in many cases. However, a literal interpretation of this procedure says there is no damping mechanism. Show that in this case, the response of the system as indicated by current flow $J(t)$ is such that a δ-function pulse

$$\delta(t) = (1/2\pi) \int_{-\infty}^{\infty} e^{-i\omega t}\, d\omega = \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t}\, d\omega$$

applied at $t = 0$ produces a current signal $J(t)$ which exists for $t < 0$. This result violates causality. Show that formally, even in the absence of collisions, one must write

$$\hat{\sigma} = i(Ne^2/m)\,[(1/\omega) - i\pi\,\delta(\omega)]$$

This result can be obtained by switching off the collisions gradually by letting

$$\omega = \lim_{\Gamma \to 0} (\omega + i\Gamma)$$

in the formula for the response function. Show that the new term in $\delta(\omega)$ gives a current which just cancels the previous contribution to the current for $t < 0$.

FURTHER READING

J. S. Toll, Causality and the Dispersion Relation: Logical Foundations, *Phys. Rev.* **104**, 1760 (1956).

P. C. Martin, Sum Rules, Kramers–Kronig Relations, and Transport Coefficients in Charged Systems, *Phys. Rev.* **161**, 143 (1967).

H. C. Bolton, Some Practical Properties of Kronig–Kramers Transforms, *Phil. Mag.* **19**, 487 (1969).

L. D. Landau and E. M. Lifshitz, "Electrodynamics of Continuous Media," Pergamon Press, Oxford, 1960.

F. Stern, Elementary Theory of the Optical Properties of Solids, *Solid State Phys.* **15** (1963).

J. des Cloizeaux, Linear response, generalized susceptibility, and dispersion theory, *in* "Theory of Condensed Matter," Int. At. Energy Agency, Vienna, 1968.

Chapter 7

SELF-CONSISTENT FIELD APPROXIMATION

In this chapter, we shall derive an expression for the dielectric function in the self-consistent field approximation. A number of authors have used this approach to study the response of electrons to an external perturbation [1–5]. However, we shall follow quite closely the steps outlined by Ehrenreich and Cohen in their classic paper [1].

Ehrenreich and Cohen considered the response of an electron gas to an external charge density. Under these conditions, the system is acted on by a scalar potential. Thus an analysis of the response of the system leads to a derivation of the longitudinal dielectric function. The longitudinal dielectric function is much easier to derive than the transverse dielectric function. Since the two are equal in the long-wavelength limit, it is quite satisfactory to use the longitudinal dielectric function for optical properties. Then, in Chapter 9, where plasmas are considered, the same longitudinal dielectric function can be used even when $q \not\approx 0$.

7.1 Self-Consistent Field Approximation

From Eq. (3.8), we can write the dielectric function as

$$\hat{\varepsilon} = (\mathbf{E} + 4\pi\mathbf{P})/\mathbf{E} \tag{7.1}$$

Using $\mathbf{E}^{\text{ext}} = \mathbf{E} + 4\pi\mathbf{E}$, $\hat{\varepsilon}$ can be written as

$$\frac{1}{\hat{\varepsilon}} = \frac{\mathbf{E}^{\text{ext}} - 4\pi\mathbf{P}}{\mathbf{E}^{\text{ext}}} = \frac{\mathbf{E}^{\text{ext}} - 4\pi\hat{\alpha}\mathbf{E}_{\text{loc}}}{\mathbf{E}^{\text{ext}}} \tag{7.2}$$

Now, assuming that wave functions are available so that $\hat{\alpha}$ can be calculated, what do we use for $\mathbf{E}_{\text{loc}}$? One possibility is to say that $\mathbf{E}_{\text{loc}} \approx \mathbf{E}^{\text{ext}}$. This approximation yields the Hartree–Fock dielectric function

$$1/\hat{\varepsilon}_{\text{HF}} = 1 - 4\pi\hat{\alpha} \tag{7.3}$$

Such an approximation is valid only when the interaction of an electron with the induced fields in the rest of the medium is very weak. The dielectric function obtained in this manner is called the Hartree–Fock dielectric function simply because it is equivalent to the usual Hartree–Fock approximations in the calculation of the ground-state energy of the electron gas. The Hartree–Fock approximation actually includes Coulomb interactions; the labeling $\hat{\varepsilon}_{\text{HF}}$ has to do only with the fact that $\hat{\varepsilon}_{\text{HF}}$ leads to the Hartree–Fock ground-state energy (a topic beyond the scope of this book).

The next degree of approximation is to at least include the effect of the induced fields in the medium. Thus, if we take $\mathbf{E}_{\text{loc}} \approx \mathbf{E}$, the dielectric function obtained from Eq. (7.1) is

$$\hat{\varepsilon}_{\text{RPA}} = 1 + 4\pi\hat{\alpha} \tag{7.4}$$

and is known as the dielectric function in the random-phase approximation (RPA). It is also known as the self-consistent field approximation (SCF) as well as by a variety of other names corresponding to the method used to reach the same result as Eq. (7.4). The approach here will be to study the time-dependent interaction of a single electron with the self-consistent field arising from the external perturbation and the induced fields. We shall designate the dielectric function $\hat{\varepsilon}_{\text{RPA}} \equiv \hat{\varepsilon}_{\text{SCF}}$ as $\hat{\varepsilon}$ because this is really the approximation that has been used throughout this book and is the common approximation used in the literature. It is usually a good approximation except for dielectric materials such as NaCl. For insulators, local field effects such as discussed in Appendix B must be included.

Before proceeding with the derivation of $\hat{\varepsilon}$ in the SCF approximation, it is worth noting several aspects of the approximation. One is that $\hat{\varepsilon}$ can be computed with the use of Feynman diagrams and time-dependent perturbation theory. If Eq. (7.4) is expanded as

$$\hat{\varepsilon}_{\text{RPA}}^{-1} = 1 - (4\pi\hat{\alpha}) + (4\pi\alpha)^2 - (4\pi\alpha)^3 + \ldots \tag{7.5}$$

each term in the expansion corresponds to one of the terms in the series

of Feynman diagrams. It is clear from inspection of Eqs. (7.3) and (7.5) that $\hat{\varepsilon}_{\mathrm{HF}}^{-1}$ is the lowest-order term in an expansion of $\hat{\varepsilon}_{\mathrm{RPA}}^{-1}$. Further discussion of this approach (which is beyond the scope of this book) is found in the literature [6–8].

The second point, the important one, is that it should be clear that $\hat{\varepsilon}_{\mathrm{RPA}}$ is equivalent to neglecting local field corrections. By taking the local field to be the average macroscopic field, we include all long-range self-consistent field effects, but we are neglecting the short-range screened interactions between electrons. (See Problem 7.1.)

We are now ready to derive the longitudinal dielectric function following the treatment of Ehrenreich and Cohen [1]. The basic idea is that some external perturbation causes a redistribution of charge within the system. The redistributed charge produces an induced potential which acts to screen the original external potential. The total self-consistent potential ϕ consists of the original external potential ϕ^{ext} plus the induced screening potential ϕ^{ind}. The final distribution of charge must be consistent with the total potential.

We shall first consider a free-electron gas. We begin with the Liouville equation of motion for the density operator

$$i\hbar\, \partial\rho_{\mathrm{op}}/\partial t = [H, \rho_{\mathrm{op}}] \tag{7.6}$$

and the single-particle Hamiltonian

$$H = H_0 + V(\mathbf{r}, t) \tag{7.7}$$

The Hamiltonian for the unperturbed system satisfies the Schrödinger equation

$$H_0|\mathbf{k}\rangle = \mathscr{E}_{\mathbf{k}}|\mathbf{k}\rangle \tag{7.8}$$

for the free-electron gas. The eigenstates for the unperturbed system are

$$|\mathbf{k}\rangle = \Omega^{-1/2} \exp i\mathbf{k}\cdot\mathbf{r} \tag{7.9}$$

where Ω is the volume of the system.

We now write

$$\rho_{\mathrm{op}} = \rho_{\mathrm{op}}^{\circ} + \rho_{\mathrm{op}}' \tag{7.10}$$

where $\rho_{\mathrm{op}}^{\circ}$ is the density operator for the unperturbed system, and we make the Fourier expansion of the total perturbing potential energy

$$V(\mathbf{r}, t) = \sum_{\mathbf{q}'} V(\mathbf{q}', t) \exp i\mathbf{q}'\cdot\mathbf{r} \tag{7.11}$$

With these expansions for ρ_{op} and V, Eq. (7.6) becomes

$$i\hbar\,(\partial/\partial t)\,(\rho_{\mathrm{op}}^{\circ} + \rho_{\mathrm{op}}') = [H_0 + \sum_{\mathbf{q}'} V(\mathbf{q}', t) \exp i\mathbf{q}'\cdot\mathbf{r}, \rho_{\mathrm{op}}^{\circ} + \rho_{\mathrm{op}}'] \tag{7.12}$$

Using the Liouville equation for the unperturbed system

$$i\hbar\,\partial\rho^{\circ}_{\mathrm{op}}/\partial t = [H_0, \rho^{\circ}_{\mathrm{op}}] \tag{7.13}$$

Eq. (7.12) becomes

$$i\hbar\,\partial\rho'_{\mathrm{op}}/\partial t = [H_0, \rho'_{\mathrm{op}}] + \left[\sum_{\mathbf{q}'} V(\mathbf{q}', t)\exp i\mathbf{q}'\cdot\mathbf{r}, \rho^{\circ}_{\mathrm{op}} + \rho'_{\mathrm{op}}\right] \tag{7.14}$$

Taking matrix elements between states $|\mathbf{k}\rangle$ and $|\mathbf{k}+\mathbf{q}\rangle$, and keeping only linear terms, the equation of motion for the matrix elements of the perturbation term in the density operator is

$$\begin{aligned} i\hbar\,(\partial/\partial t)\,\langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle &= \langle \mathbf{k}+\mathbf{q}|[H_0, \rho'_{\mathrm{op}}]|\mathbf{k}\rangle \\ &\quad + \langle \mathbf{k}+\mathbf{q}|\left[\sum_{\mathbf{q}'} V(\mathbf{q}', t)\exp i\mathbf{q}'\cdot\mathbf{r}, \rho^{\circ}_{\mathrm{op}}\right]|\mathbf{k}\rangle \end{aligned} \tag{7.15}$$

Now let us look at the terms on the RHS of Eq. (7.15) and see how they can be simplified. The first term is

$$\begin{aligned} \langle \mathbf{k}+\mathbf{q}|[H_0, \rho'_{\mathrm{op}}]|\mathbf{k}\rangle &= \langle \mathbf{k}+\mathbf{q}|H_0\rho'_{\mathrm{op}}|\mathbf{k}\rangle - \langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}H_0|\mathbf{k}\rangle \\ &= \mathscr{E}_{\mathbf{k}+\mathbf{q}}\langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle - \mathscr{E}_{\mathbf{k}}\langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle \\ &= (\mathscr{E}_{\mathbf{k}+\mathbf{q}} - \mathscr{E}_{\mathbf{k}})\,\langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle \end{aligned} \tag{7.16}$$

The second term on the RHS of Eq. (7.15) is

$$\begin{aligned} &\langle \mathbf{k}+\mathbf{q}|\left[\sum_{\mathbf{q}'} V(\mathbf{q}', t)\exp i\mathbf{q}'\cdot\mathbf{r}, \rho^{\circ}_{\mathrm{op}}\right]|\mathbf{k}\rangle \\ &= \langle \mathbf{k}+\mathbf{q}\left|\sum_{\mathbf{q}'} V(\mathbf{q}', t)\,(\exp i\mathbf{q}'\cdot\mathbf{r})\,\rho^{\circ}_{\mathrm{op}}\right|\mathbf{k}\rangle \\ &\quad - \langle \mathbf{k}+\mathbf{q}\left|\rho^{\circ}_{\mathrm{op}}\sum_{\mathbf{q}'} V(\mathbf{q}', t)\,(\exp i\mathbf{q}'\cdot\mathbf{r})\right|\mathbf{k}\rangle \end{aligned} \tag{7.17}$$

The density operator operating on a state $|\mathbf{k}\rangle$ gives the occupation probability of that state. Thus,

$$\rho^{\circ}_{\mathrm{op}}|\mathbf{k}\rangle = f(\mathscr{E}_{\mathbf{k}})|\mathbf{k}\rangle \tag{7.18}$$

where $f(\mathscr{E}_{\mathbf{k}})$ is the Fermi–Dirac distribution function. Using this result,

Eq. (7.17) becomes

$$\langle \mathbf{k}+\mathbf{q}\left|\left[\sum_{\mathbf{q}'} V(\mathbf{q}',t)\exp i\mathbf{q}'\cdot\mathbf{r},\rho^{\circ}_{\mathrm{op}}\right]\right|\mathbf{k}\rangle$$

$$=\left[f(\mathscr{E}_{\mathbf{k}})-f(\mathscr{E}_{\mathbf{k}+\mathbf{q}})\right]\langle \mathbf{k}+\mathbf{q}\left|\sum_{\mathbf{q}'} V(\mathbf{q}',t)(\exp i\mathbf{q}'\cdot\mathbf{r})\right|\mathbf{k}\rangle \tag{7.19}$$

Substituting explicit expressions for the plane wave states $|\mathbf{k}\rangle$ and $|\mathbf{k}+\mathbf{q}\rangle$ from Eq. (7.9), we find that

$$\langle \mathbf{k}+\mathbf{q}\left|\sum_{\mathbf{q}'} V(\mathbf{q}',t)(\exp i\mathbf{q}'\cdot\mathbf{r})\right|\mathbf{k}\rangle$$

$$=\Omega^{-1}\int\{\exp[-i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}]\}\left[\sum_{\mathbf{q}'} V(\mathbf{q}',t)(\exp i\mathbf{q}'\cdot\mathbf{r})\right](\exp i\mathbf{k}\cdot\mathbf{r})\,d\mathbf{r}$$

$$=\sum_{\mathbf{q}'} V(\mathbf{q}',t)\int\Omega^{-1}\exp[i(\mathbf{q}'-\mathbf{q})\cdot\mathbf{r}]\,d\mathbf{r}$$

$$=\sum_{\mathbf{q}'} V(\mathbf{q}',t)\delta_{\mathbf{q}\mathbf{q}'}=V(\mathbf{q},t) \tag{7.20}$$

so that Eq. (7.19) reduces to

$$\langle \mathbf{k}+\mathbf{q}\left|\left[\sum_{\mathbf{q}'} V(\mathbf{q}',t)\exp i\mathbf{q}'\cdot\mathbf{r},\rho^{\circ}_{\mathrm{op}}\right]\right|\mathbf{k}\rangle=\left[f(\mathscr{E}_{\mathbf{k}})-f(\mathscr{E}_{\mathbf{k}+\mathbf{q}})\right]V(\mathbf{q},t) \tag{7.21}$$

Substituting Eqs. (7.16) and (7.21) into Eq. (7.15) yields

$$i\hbar\,(\partial/\partial t)\langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle=(\mathscr{E}_{\mathbf{k}+\mathbf{q}}-\mathscr{E}_{\mathbf{k}})\langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle$$

$$+\left[f(\mathscr{E}_{\mathbf{k}})-f(\mathscr{E}_{\mathbf{k}+\mathbf{q}})\right]V(\mathbf{q},t) \tag{7.22}$$

We now have a time-dependent equation for the contribution to the density operator arising from the induced charge density. It is rather straightforward to carry through from here to a formal expression for the dielectric function. Before following through the steps, however, let us digress briefly to discuss the dielectric function we are working toward and how we might expect to get there. Should we consider turning on the perturbation at time $t=0$ and doing some sort of integration from time $t=0$ to time t? That is the procedure we followed in Chapter 3–5. However, that leads to a real dielectric function unless a phenomenological damping term is included. We included such a damping term in our earlier treatments in order to arrive at a dielectric function having real and imag-

inary parts. We did it because we wanted to account for energy absorption in the medium as well as shielding of external fields. In fact, we are more interested in the former than in the latter.

The necessity of a damping term is clear if we consider the response of the system to an external impulse. In the absence of an energy loss mechanism, the response of the system persists forever. We do not want such an unphysical result. One way to avoid this dilemma is to turn on the perturbation gradually, with a time dependence proportional to $e^{\eta t}$. This is a mathematical device known as the adiabatic approximation. It ensures that in the infinite past, there was no perturbation, but that a finite perturbation exists for positive values of time. One can let $\eta \to 0$ at the end of the calculation. The adiabatic approximation leads directly to a complex dielectric function. This is because it satisfies causality (Problem 7.1).

We now assume that the external potential $\phi^{\text{ext}}(\mathbf{r}, t)$ is turned on with a time dependence

$$\phi^{\text{ext}}(\mathbf{r}, t) = \phi^{\text{ext}}(\mathbf{r}, \omega) e^{-i\omega t} e^{\eta t} \tag{7.23}$$

We need consider only a single Fourier component of the potential since in the linear approximation all the Fourier components are independent.

The induced screening potential, the total potential, and the corresponding density fluctuations all have the same time dependence $e^{-i\omega t} e^{\eta t}$ as the external potential acting on the system. Thus, Eq. (7.22) yields

$$\langle \mathbf{k} + \mathbf{q} | \rho'_{\text{op}} | \mathbf{k} \rangle = \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q}}) - f(\mathscr{E}_{\mathbf{k}})}{(\mathscr{E}_{\mathbf{k}+\mathbf{q}} - \mathscr{E}_{\mathbf{k}}) - \hbar\omega - i\hbar\eta} V(\mathbf{q}, t) \tag{7.24}$$

We now have an equation for the induced density matrix in terms of the total perturbing potential energy; i.e., the self-consistent potential energy. What we "know," what we can directly control, is the external potential. What we can most easily calculate, though, is the induced screening potential. Our course, then, is to relate the total potential to the screening potential by means of the dielectric functions and then to express the induced screening potential in an appropriate explicit form.

Working in terms of the potential energies corresponding to the total, external, and induced potentials, we have

$$V(\mathbf{q}, t) = V^{\text{ext}}(\mathbf{q}, t) + V^{\text{ind}}(\mathbf{q}, t) \tag{7.25}$$

and

$$V(\mathbf{q}, t) = V^{\text{ext}}(\mathbf{q}, t)/\hat{\varepsilon}(\mathbf{q}, \omega) \tag{7.26}$$

From these last two equations, we obtain

$$V(\mathbf{q}, t) = V^{\text{ind}}(\mathbf{q}, t)[1 - \hat{\varepsilon}(\mathbf{q}, \omega)] \tag{7.27}$$

The induced screening potential energy is related to the induced change in electron density by Poisson's equation:

$$\nabla^2 V^{\text{ind}}(\mathbf{r}, t) = -4\pi e^2 \, \delta n \tag{7.28}$$

The induced fluctuation in electron density within the system can be obtained from the density matrix as follows.

The density operator is defined by

$$\rho_{\text{op}} = \sum_{\mathbf{m}} |\mathbf{m}\rangle P_{\mathbf{m}} \langle \mathbf{m}| \tag{7.29}$$

where $P_{\mathbf{m}}$ is the probability of finding the system in one of the states $|\mathbf{m}\rangle$. Since we are concerned with an independent-particle model, we take $P_{\mathbf{m}}$ as the probability of finding an electron in the one-electron state $|\mathbf{m}\rangle$. The charge density within the medium is then given by

$$n = \text{Tr}\,[\rho_{\text{op}} \, \delta(\mathbf{r}_e - \mathbf{r})] \tag{7.30}$$

where $\delta(\mathbf{r}_e - \mathbf{r})$ is the electron position operator. We will now evaluate the trace in the representation of plane-wave states appropriate to a free-electron gas. Thus,

$$\begin{aligned} n &= \sum_{\mathbf{l}} \langle \mathbf{l}|\rho_{\text{op}} \, \delta(\mathbf{r}_e - \mathbf{r})|\mathbf{l}\rangle \\ &= \sum_{\mathbf{l},\mathbf{m}} \langle \mathbf{l}|\rho_{\text{op}}|\mathbf{m}\rangle \langle \mathbf{m}|\mathbf{l}\rangle \langle \mathbf{l}|\delta(\mathbf{r}_e - \mathbf{r})|\mathbf{l}\rangle \\ &= \sum_{\mathbf{l},\mathbf{m}} \langle \mathbf{l}|\rho_{\text{op}}|\mathbf{m}\rangle \langle \mathbf{m}|\mathbf{l}\rangle \end{aligned} \tag{7.31}$$

Substituting from Eq. (7.9) for the eigenstates of the unperturbed system, Eq. (7.31) becomes

$$n = \sum_{\mathbf{l},\mathbf{m}} \Omega^{-1} \{\exp[i(\mathbf{l} - \mathbf{m}) \cdot \mathbf{r}]\} \langle \mathbf{l}|\rho_{\text{op}}|\mathbf{m}\rangle \tag{7.32}$$

Now, to get density matrix elements of the same form as in Eq. (7.22), we let $\mathbf{m} = \mathbf{k}$, $\mathbf{l} = \mathbf{k} + \mathbf{q}$. Thus

$$n = \Omega^{-1} \sum_{\mathbf{k},\mathbf{q}} (\exp i\mathbf{q} \cdot \mathbf{r}) \langle \mathbf{k} + \mathbf{q}|\rho_{\text{op}}|\mathbf{k}\rangle \tag{7.33}$$

$$n_0 + \delta n = \Omega^{-1} \sum_{\mathbf{k},\mathbf{q}} (\exp i\mathbf{q} \cdot \mathbf{r}) \langle \mathbf{k} + \mathbf{q}|\rho^{\circ}_{\text{op}} + \rho'_{\text{op}}|\mathbf{k}\rangle \tag{7.34}$$

where Eq. (7.34) follows directly from Eq. (7.33). From inspection of Eq.

(7.34), it is clear that

$$\delta n = \Omega^{-1} \sum_{\mathbf{k},\mathbf{q}} (\exp i\mathbf{q}\cdot\mathbf{r}) \langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle \tag{7.35}$$

Note that the induced screening charge is directly related to the perturbation density operator ρ'_{op}. That ρ'_{op} does not include the external test charge follows from the definition of the density operator; it describes how the charge fluctuation density of the system itself is distributed among the various states.

From Eqs. (7.28) and (7.35), we have

$$\nabla^2 V^{\mathrm{ind}}(\mathbf{r}, t) = -(4\pi e^2/\Omega) \sum_{\mathbf{k},\mathbf{q}} (\exp i\mathbf{q}\cdot\mathbf{r}) \langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle \tag{7.36}$$

Expressing $V^{\mathrm{ind}}(\mathbf{r}, t)$ as

$$V^{\mathrm{ind}}(\mathbf{r}, t) = \sum_{\mathbf{q}} V^{\mathrm{ind}}(\mathbf{q}, t) \exp i\mathbf{q}\cdot\mathbf{r} \tag{7.37}$$

Eq. (7.36) yields

$$V^{\mathrm{ind}}(\mathbf{q}, t) = (4\pi e^2/q^2\Omega) \sum_{\mathbf{k}} \langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle \tag{7.38}$$

where the factor $4\pi e^2/q^2$ is the Fourier transform of the Coulomb potential.

Using Eqs. (7.24) and (7.27), we can now find an expression for the dielectric function. The first step is to sum Eq. (7.24) over all values of $\mathbf{k}$,

$$\sum_{\mathbf{k}} \langle \mathbf{k}+\mathbf{q}|\rho'_{\mathrm{op}}|\mathbf{k}\rangle = V(\mathbf{q}, t) \sum_{\mathbf{k}} \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q}}) - f(\mathscr{E}_{\mathbf{k}})}{(\mathscr{E}_{\mathbf{k}+\mathbf{q}} - \mathscr{E}_{\mathbf{k}}) - \hbar\omega - i\hbar\eta} \tag{7.39}$$

and notice that the LHS of Eq. (7.39) is simply related to $V^{\mathrm{ind}}(\mathbf{q}, t)$ by means of Eq. (7.38). Thus

$$V^{\mathrm{ind}}(\mathbf{q}, t) = \frac{4\pi e^2}{q^2\Omega} V(\mathbf{q}, t) \sum_{\mathbf{k}} \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q}}) - f(\mathscr{E}_{\mathbf{k}})}{(\mathscr{E}_{\mathbf{k}+\mathbf{q}} - \mathscr{E}_{\mathbf{k}}) - \hbar\omega - i\hbar\eta} \tag{7.40}$$

Substituting for $V(\mathbf{q}, t)$ from Eq. (7.27) yields the Lindhard dielectric function

$$\varepsilon(\mathbf{q}, \omega) = 1 - \lim_{\eta\to 0} \frac{4\pi e^2}{q^2\Omega} \sum_{\mathbf{k}} \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q}}) - f(\mathscr{E}_{\mathbf{k}})}{(\mathscr{E}_{\mathbf{k}+\mathbf{q}} - \mathscr{E}_{\mathbf{k}}) - \hbar\omega - i\hbar\eta} \tag{7.41}$$

Equation (7.41) has been derived for a free-electron gas. However, to apply the dielectric function to real solids, we should at least include the periodicity of the lattice. It is an easy step to generalize the previous results using Bloch functions rather than simple plane waves. We shall assume that the core states of the atoms composing the solid are tightly bound and can be neglected, and that the valence bands and conduction bands are sufficiently broad that local field corrections are unimportant. Then, using

the Bloch functions,

$$|\mathbf{k}l\rangle = \Omega^{-1/2}(\exp i\mathbf{k}\cdot\mathbf{r})\,u_{\mathbf{k}l}(\mathbf{r}) \tag{7.42}$$

where l is a band index and $u_{\mathbf{k}l}(\mathbf{r})$ is the spatially periodic part of the wave function, we can carry the treatment through just as for plane waves (see Problem 7.3). The result is

$$\hat{\varepsilon}(\mathbf{q},\omega) = 1 - \lim_{\eta\to 0}\frac{4\pi e^2}{q^2\Omega}\sum_{\mathbf{k},l,l'}|(\mathbf{k}+\mathbf{q},l'|\mathbf{k}l)|^2 \times \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}) - f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega - i\hbar\eta} \tag{7.43}$$

where

$$(\mathbf{k}+\mathbf{q},l'|\mathbf{k}l) = (1/\Delta)\int_{\text{cell}} u^*_{\mathbf{k}+\mathbf{q},l'}(\mathbf{r})u_{\mathbf{k}l}(\mathbf{r})\,d\mathbf{r} \tag{7.44}$$

and Δ is the volume of a unit cell.

We arrived at Eq. (7.44) by neglecting local field effects. What would we have to do to include local field effects? Bloch functions would still provide satisfactory wave functions. However, we could not simply work within the reduced zone scheme. We must include Umklapp processes, i.e., work with a repeated zone scheme. This can perhaps best be seen by looking at the expression for the potential energy. Equation (7.11) is not an expansion in a complete set of basis functions when working in the reduced zone. The shortest wavelengths in the expansion, corresponding to $\mathbf{q}$ at the Brillouin zone boundary, are just equal to two lattice spacings. These are clearly not short enough to describe rapid variations of potential within a unit cell. To describe such variations in potential requires including all reciprocal lattice vectors $\mathbf{G}$ in the expansion. We must write

$$V(\mathbf{r},t) = \sum_{\mathbf{q},\mathbf{G}} V(\mathbf{q},\mathbf{G},t)\exp[i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}] \tag{7.45}$$

The treatment goes through more or less as before, but is more tedious to carry out. It is discussed in papers by Adler and Wiser [2, 3]. The way to define polarizability including local field effects is discussed briefly in Appendix B.

7.2 Special Cases and Applications

In order to develop a better feeling for the dielectric function, we shall consider several limiting forms of the dielectric functions we have just

derived. We also want to see the relationship of the present form of the dielectric function to the earlier forms. Presumably, they must be equivalent and we shall show that they are indeed, but it does take a little manipulating to show it. Other limiting forms of $\hat{\varepsilon}(\mathbf{q}, \omega)$, more appropriate to a discussion of plasmons, will be taken up in Chapter 9.

Let us consider first the zero-frequency, small-wavevector limit for a free-electron gas. Equation (7.41) can be simplified in this limiting case by using the approximations

$$\mathscr{E}_{\mathbf{k}+\mathbf{q}} - \mathscr{E}_{\mathbf{k}} \approx \mathbf{q} \cdot \nabla_{\mathbf{k}} \mathscr{E}(\mathbf{k}) \tag{7.46}$$

$$f(\mathscr{E}_{\mathbf{k}+\mathbf{q}}) - f(\mathscr{E}_{\mathbf{k}}) \approx \mathbf{q} \cdot (\partial f/\partial \mathscr{E}) \nabla_{\mathbf{k}} \mathscr{E}(\mathbf{k}) \tag{7.47}$$

and approximating the summation over $\mathbf{k}$ by an integral over all the corresponding energy states. Thus, with the transformation

$$\sum_{\mathbf{k}} \to \Omega \int N(\mathscr{E})\, d\mathscr{E} \tag{7.48}$$

where $N(\mathscr{E})$ is the density of states, the real dielectric function is

$$\begin{aligned} \varepsilon_1(\mathbf{q}, 0) &= 1 - \frac{4\pi e^2}{q^2} \int \frac{\mathbf{q} \cdot (\partial f/\partial \mathscr{E}) \nabla_{\mathbf{k}}(\mathscr{E}_{\mathbf{k}})}{\mathbf{q} \cdot \nabla_{\mathbf{k}}(\mathscr{E}_{\mathbf{k}})} N(\mathscr{E})\, d\mathscr{E} \\ &= 1 - \frac{4\pi e^2}{q^2} \int \frac{\partial f}{\partial \mathscr{E}} N(\mathscr{E})\, d\mathscr{E} \end{aligned} \tag{7.49}$$

In the zero-temperature limit, $f(\mathscr{E})$ is unity for $\mathscr{E} < \mathscr{E}_F$ and zero for $\mathscr{E} > \mathscr{E}_F$. Thus $\partial f/\partial \mathscr{E}$ is essentially a δ function, i.e.,

$$\lim_{T \to 0} (\partial f/\partial \mathscr{E}) = -\delta(\mathscr{E} - \mathscr{E}_F) \tag{7.50}$$

so that Eq. (7.49) becomes

$$\begin{aligned} \varepsilon_1(\mathbf{q}, 0) &= 1 + (4\pi e^2/q^2) N(\mathscr{E}_F) \\ &= 1 + (\lambda^2/q^2) \end{aligned} \tag{7.51}$$

where

$$\lambda^2 = 4\pi e^2 N(\mathscr{E}_F) \tag{7.52}$$

Thus the effective potential (energy) in the medium is

$$V(q, 0) = \frac{V^{\text{ext}}(q, 0)}{\varepsilon_1(q, 0)} = \frac{V^{\text{ext}}(q, 0)}{1 + (\lambda/q)^2} \tag{7.53}$$

In the limit $q \to 0$, $\varepsilon_1(q, 0) \to \infty$, and an external field of long wavelength is perfectly screened out. For q small, but nonzero, Eq. (7.53) expresses the fact that because of polarization of the medium, the Coulomb force due to an external potential is effective only over distances of the order of λ^{-1}. The characteristic distance λ^{-1} is exactly the same as that obtained for the screening distance of an electron gas by elementary Thomas–Fermi theory.

What we have just derived is the static dielectric constant of a free-electron gas. However, in the study of the optical properties of solids, we are more interested in the finite-frequency, long-wavelength limit. Thus, for real solids, we want to determine $\hat{\varepsilon}(q, \omega)$ from Eq. (7.43).

In Appendix H, it is shown that in the limit $q \to 0$,

$$|(\mathbf{k} + \mathbf{q}, l'|\mathbf{k}l)|^2 = \delta_{ll'} + (1 - \delta_{ll'})\,(q/m\omega_{l'l})^2\,|P^{\mu}_{l'l}|^2 \tag{7.54}$$

where

$$P^{\mu}_{l'l} = (1/\Delta_{\text{cell}}) \int u_{\mathbf{k}l'} p^{\mu} u_{\mathbf{k}l}\, d\mathbf{r}$$

$$\hbar\omega_{l'l} = \mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l} \tag{7.56}$$

and p^{μ} is the momentum operator associated with the direction of propagation $\mathbf{q}$. Other quantities in Eq. (7.43) can also be found to lowest order in q. What follows is the separation of complex quantities in Eq. (7.43) into real and imaginary parts, and the expansion of these parts to lowest order so as to determine $\varepsilon_1(q, \omega)$ and $\varepsilon_2(q, \omega)$.

From the general mathematical theorem

$$\lim_{\eta \to +0} \frac{1}{X - i\eta} = \mathscr{P}\frac{1}{X} + i\pi\,\delta(X) \tag{7.57}$$

where $\mathscr{P}$ denotes the principal value, we have

$$\frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}) - f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega - i\hbar\eta} = \mathscr{P}\left\{\frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}) - f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega}\right\} + i\pi[f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}) - f(\mathscr{E}_{\mathbf{k}l})]\,\delta(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega) \tag{7.58}$$

Working with only the real part of the RHS of Eq. (7.58), we have

$$\sum_{\mathbf{k}} \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}) - f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega} = \sum_{\mathbf{k}} \frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega} - \sum_{\mathbf{k}} \frac{f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega} \tag{7.59}$$

The point in breaking up the expression into two terms is to allow the

easy introduction of a dummy index so that the Fermi–Dirac function $f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'})$ can be replaced by $f(\mathscr{E}_{\mathbf{k}l'})$. Thus, making the substitution $\mathbf{k}+\mathbf{q}\to\mathbf{k}$ in the first term on the RHS of Eq. (7.59) yields

$$\sum_{\mathbf{k}}\frac{f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'})-f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}-\mathscr{E}_{\mathbf{k}l}-\hbar\omega}=\sum_{\mathbf{k}}\frac{f(\mathscr{E}_{\mathbf{k}l'})}{\mathscr{E}_{\mathbf{k}l'}-\mathscr{E}_{\mathbf{k}-\mathbf{q},l}-\hbar\omega}-\sum\frac{f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}-\mathscr{E}_{\mathbf{k}l}-\hbar\omega} \tag{7.60}$$

That the substitution $\mathbf{k}+\mathbf{q}\to\mathbf{k}$ is allowed is because the summation over the set of states $\mathbf{k}+\mathbf{q}$ includes all the same states as the summation over $\mathbf{k}$.

Now, in order to further simplify our task, let us first calculate the intraband contribution to $\varepsilon_1(q,\omega)$. For that case, we take $l=l'$, so that the initial and final states are within the same band. Then, the appropriate contributions from Eqs. (7.43), (7.54), and (7.60) yield

$$\varepsilon_1^{\text{intraband}}(q,\omega)=1-\frac{4\pi e^2}{q^2\Omega}\sum_{\mathbf{k}l}f(\mathscr{E}_{\mathbf{k}l})\left(\frac{1}{\mathscr{E}_{\mathbf{k}l}-\mathscr{E}_{\mathbf{k}-\mathbf{q},l}-\hbar\omega}\right.$$

$$\left.-\frac{1}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l}-\mathscr{E}_{\mathbf{k}l}-\hbar\omega}\right)=1-\frac{4\pi e^2}{q^2\Omega}\sum_{\mathbf{k}l}f(\mathscr{E}_{\mathbf{k}l})\left[\frac{\mathscr{E}_{\mathbf{k}l}-\mathscr{E}_{\mathbf{k}-\mathbf{q},l}+\hbar\omega}{(\mathscr{E}_{\mathbf{k}l}-\mathscr{E}_{\mathbf{k}-\mathbf{q},l})^2-(\hbar\omega)^2}\right.$$

$$\left.-\frac{\mathscr{E}_{\mathbf{k}+\mathbf{q},l}-\mathscr{E}_{\mathbf{k}l}+\hbar\omega}{(\mathscr{E}_{\mathbf{k}+\mathbf{q},l}-\mathscr{E}_{\mathbf{k}l})^2-(\hbar\omega)^2}\right] \tag{7.61}$$

Recognizing that for small q, $(\mathscr{E}_{\mathbf{k}+\mathbf{q},l}-\mathscr{E}_{\mathbf{k}l})^2\approx(\mathscr{E}_{\mathbf{k}l}-\mathscr{E}_{\mathbf{k}-\mathbf{q},l})^2\ll(\hbar\omega)^2$, we get

$$\varepsilon_1^{\text{intraband}}(q,\omega)=1+\frac{4\pi e^2}{q^2\Omega(\hbar\omega)^2}\sum_{\mathbf{k}l}f(\mathscr{E}_{\mathbf{k}l})\left[(\mathscr{E}_{\mathbf{k}l}-\mathscr{E}_{\mathbf{k}-\mathbf{q},l})-(\mathscr{E}_{\mathbf{k}+\mathbf{q},l}-\mathscr{E}_{\mathbf{k}l})\right] \tag{7.62}$$

Now, for small q,

$$\mathscr{E}_{\mathbf{k}+\mathbf{q},l}=\mathscr{E}_{\mathbf{k}l}+q\frac{\partial\mathscr{E}_{\mathbf{k}l}}{\partial k}+\frac{1}{2}q^2\frac{\partial^2\mathscr{E}_{\mathbf{k}l}}{\partial k^2} \tag{7.63}$$

$$\mathscr{E}_{\mathbf{k}-\mathbf{q},l}=\mathscr{E}_{\mathbf{k}l}-q\frac{\partial\mathscr{E}_{\mathbf{k}l}}{\partial k}+\frac{1}{2}q^2\frac{\partial^2\mathscr{E}_{\mathbf{k}l}}{\partial k^2} \tag{7.64}$$

so that Eq. (7.62) becomes

$$\varepsilon_1^{\text{intraband}}(q,\omega)=1-\frac{4\pi e^2}{\Omega(\hbar\omega)^2}\sum_{\mathbf{k}l}f(\mathscr{E}_{\mathbf{k}l})\frac{\partial^2\mathscr{E}_{\mathbf{k}l}}{\partial k_\mu{}^2} \tag{7.65}$$

Now making the transformation

$$\sum_{\mathbf{k}} \rightarrow (\Omega/4\pi^3) \int d\mathbf{k} \tag{7.66}$$

Eq. (7.65) yields

$$\varepsilon_1^{\text{intraband}}(q, \omega) = 1 - \left(\frac{e}{\pi\hbar\omega}\right)^2 \sum_l \int d\mathbf{k}\, f(\mathscr{E}_{\mathbf{k}l}) \frac{\partial^2 \mathscr{E}_{\mathbf{k}l}}{\partial k_\mu{}^2} \tag{7.67}$$

To determine the interband contribution to $\varepsilon_1(q, \omega)$, we consider all terms l, l' such that $l \neq l'$. Then, the appropriate contributions from Eqs. (7.43), (7.54), and (7.60) yield

$$\delta\varepsilon_1^{\text{interband}}(q, \omega) = -\frac{4\pi e^2}{q^2\Omega} \sum_{ll}{}' \left(\frac{q}{m\omega_{l'l}}\right)^2 |P_{l'l}^{\mu}|^2 \times \sum_{\mathbf{k}} \left[\frac{f(\mathscr{E}_{\mathbf{k}l'})}{\mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}-\mathbf{q},l} - \hbar\omega} - \frac{f(\mathscr{E}_{\mathbf{k}l})}{\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega}\right] \tag{7.68}$$

where the prime on the summation indicates terms with $l = l'$ are to be excluded. Now, since $l \neq l'$, we can ignore the q dependence of $\mathscr{E}$ in the denominators of Eq. (7.68) since the only significant contribution to the energy differences arises from the difference in energy of the bands l, l'. Thus,

$$\delta\varepsilon_1^{\text{interband}}(q, \omega) = -\frac{4\pi e^2}{\hbar q^2\Omega} \sum_{ll}{}' \left(\frac{q}{m\omega_{l'l}}\right)^2 |P_{l'l}^{\mu}|^2 \times \sum_{\mathbf{k}} \left[\frac{f(\mathscr{E}_{\mathbf{k}l'})}{\omega_{l'l} - \omega} - \frac{f(\mathscr{E}_{\mathbf{k}l})}{\omega_{l'l} - \omega}\right] \tag{7.69}$$

The last factor in Eq. (7.69) can be simplified as follows. Looking only at the term $f(\mathscr{E}_{\mathbf{k}l'})/(\omega_{l'l} - \omega)$, and making the transformation $l' \rightarrow l$,

$$\frac{f(\mathscr{E}_{\mathbf{k}l'})}{\omega_{l'l} - \omega} \rightarrow \frac{f(\mathscr{E}_{\mathbf{k}l})}{\omega_{ll'} - \omega} \tag{7.70}$$

Now, since $\omega_{ll'} = -\omega_{l'l}$, we can rewrite this as

$$\frac{f(\mathscr{E}_{\mathbf{k}l'})}{\omega_{l'l} - \omega} \rightarrow -\frac{f(\mathscr{E}_{\mathbf{k}l})}{\omega_{l'l} + \omega} \tag{7.71}$$

Making this substitution into Eq. (7.69), we obtain

$$\delta\varepsilon_1^{\text{interband}}(q,\omega) = \frac{4\pi e^2}{\hbar m^2 \Omega} \sum_{ll'}{}' \left(\frac{1}{\omega_{l'l}}\right)^2 |P^{\mu}_{l'l}|^2 \times \sum_{\mathbf{k}} f(\mathscr{E}_{\mathbf{k}l}) \left[\frac{1}{\omega_{l'l} + \omega} + \frac{1}{\omega_{l'l} - \omega}\right] \tag{7.72}$$

Introducing the oscillator strength

$$f^{\mu}_{l'l} = (2/\hbar\omega_{l'l}m)\,|P^{\mu}_{l'l}|^2 \tag{7.73}$$

and making the transformation (7.66) from a sum to an integral, Eq. (7.72) becomes

$$\delta\varepsilon_1^{\text{interband}}(q,\omega) = \frac{e^2}{m\pi^2} \sum_{ll'}{}' \int d\mathbf{k}\, \frac{f(\mathscr{E}_{\mathbf{k}l}) f^{\mu}_{l'l}}{\omega^2_{l'l} - \omega^2} \tag{7.74}$$

The total real dielectric function is thus given by

$$\varepsilon_1(q,\omega) = \varepsilon_1^{\text{intraband}}(q,\omega) + \delta\varepsilon_1^{\text{interband}}(q,\omega) \tag{7.75}$$

and Eqs. (7.67) and (7.74) if core states are neglected.

We must now go through a similar series of steps to determine $\varepsilon_2(q,\omega)$. From Eq. (7.43) and the imaginary part of Eq. (7.58), we obtain

$$\varepsilon_2(q,\omega) = -(4\pi^2 e^2/q^2\Omega) \sum_{\mathbf{k}ll'} |(\mathbf{k}+\mathbf{q}, l'|\mathbf{k}l)|^2 \times [f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'}) - f(\mathscr{E}_{\mathbf{k}l})]\, \delta(\mathscr{E}_{\mathbf{k}+\mathbf{q},l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega) \tag{7.76}$$

Once again, the intraband contribution is obtained from the terms with $l = l'$. Thus, using Eq. (7.54),

$$\varepsilon_2^{\text{intraband}}(q,\omega) = - (4\pi^2 e^2/q^2\Omega) \sum_{l} [f(\mathscr{E}_{\mathbf{k}+\mathbf{q},l}) - f(\mathscr{E}_{\mathbf{k}l})]\, \delta(\mathscr{E}_{\mathbf{k}+\mathbf{q},l} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega) \tag{7.77}$$

Breaking this expression up into two terms, making the substitution $\mathbf{k} + \mathbf{q} \to \mathbf{k}$ as was done in Eq. (7.60), and making the transformation (7.66), we obtain

$$\varepsilon_2^{\text{intraband}}(q,\omega) = (e^2/q^2\pi) \sum_{l} \int d\mathbf{k}\, f(\mathscr{E}_{\mathbf{k}l}) \times [\delta(\mathscr{E}_{\mathbf{k}+\mathbf{q},l} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega) - \delta(\mathscr{E}_{\mathbf{k}-\mathbf{q},l} - \mathscr{E}_{\mathbf{k}l} + \hbar\omega)] \tag{7.78}$$

Following the same general procedures as before, and using Eqs. (7.54), (7.66) and (7.76) to determine the interband contribution yields

$$\delta\varepsilon_2^{\text{interband}}(q,\omega) = (e^2/\pi m^2\omega^2)\sum_{ll'}{}' \int d\mathbf{k}\,[f(\mathscr{E}_{\mathbf{k}l}) - f(\mathscr{E}_{\mathbf{k}l'})]\,\delta(\mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l} - \hbar\omega) \tag{7.79}$$

The total imaginary dielectric function is given by

$$\varepsilon_2(q,\omega) = \varepsilon_2^{\text{intraband}}(q,\omega) + \delta\varepsilon_2^{\text{interband}}(q,\omega) \tag{7.80}$$

and Eqs. (7.78) and (7.79).

Now let us see what we can learn from some simple limiting cases. For example, if we consider a monovalent, nearly-free-electron metal, and neglect interband transitions, the real dielectric function is given by Eq. (7.67). If we make the substitution

$$1/m^* = (1/\hbar^2)\,\partial^2\mathscr{E}/\partial k^2 \tag{7.81}$$

we get

$$\varepsilon_1(\omega) = 1 - (e^2/m^*\pi^2\omega^2)\sum_l \int d\mathbf{k}\, f(\mathscr{E}_{\mathbf{k}l}) \tag{7.82}$$

For a simple monovalent metal, we need consider only one band, so we can drop the summation over l. Also, the Fermi–Dirac distribution function is unity over half the volume of the Brillouin zone and zero elsewhere. Thus,

$$\sum_l \int_{\text{BZ}} d\mathbf{k} f(\mathscr{E}_{\mathbf{k}l}) = \tfrac{1}{2}\int_{\text{BZ}} d\mathbf{k} = \tfrac{1}{2}(8\pi^3/\Delta) \tag{7.83}$$

where Δ is the volume of a unit cell. The real dielectric function then reduces to

$$\begin{aligned}\varepsilon_1(\omega) &= 1 - (4\pi e^2/m^*\omega^2\Delta) \\ &= 1 - (4\pi N e^2/m^*\omega^2)\end{aligned} \tag{7.84}$$

in agreement with our earlier results.

For insulators, the second term of Eq. (7.65) vanishes. Thus, from Eqs. (7.65), (7.74), and (7.75),

$$\varepsilon_1(\omega) = 1 + \frac{e^2}{m\pi^2}\sum_{ll'}{}' \int d\mathbf{k}\,\frac{f(\mathscr{E}_{\mathbf{k}l})f_{l'l}^{\mu}}{\omega_{l'l}^2 - \omega^2} \tag{7.85}$$

Now let us consider an idealized insulator for which only two bands are important, namely the valence band l and a higher-lying conduction band l'. Then, $f(\mathscr{E}_{\mathbf{k}l}) = 1$ since all valence band states can be assumed to be

occupied. Also, $f_{l'l}^{\mu} = 1$ because of the sum rule for oscillator strengths. Thus,

$$\varepsilon_1(\omega) = 1 + [e^2/m\pi^2(\omega_{l'l}^2 - \omega^2)] \int d\mathbf{k} \tag{7.86}$$

The integral $\int d\mathbf{k}$ has already been obtained in Eq. (7.83). One must simply recognize that for an idealized insulator, there are two electrons per atom in each band. Then, following through as before, we obtain

$$\varepsilon_1(\omega) = 1 + [4\pi N e^2/m(\omega_{l'l}^2 - \omega^2)] \tag{7.87}$$

This is of the same form as Eq. (3.14) except that here we have not included lifetime broadening of the states. The way to include broadening in all that we have done is outlined in Problem 7.5.

There are several other things that can be learned from a study of Eq. (7.85). One is that the plasma frequency must be greater than the band gap. To see this, we write the double summation as $\Sigma_{l \le L, l' > L}$, where L is the highest-lying valence band. That this is permissible is easy to see. All states $l > L$ make no contribution, since for these states, $f(\mathscr{E}_{\mathbf{k}l}) = 0$. All states $l' \le L$ can be excluded since for each term $f_{l'l}^{\mu}$ appearing in the summation, there will be a corresponding term $f_{ll'}^{\mu}$ also appearing. Because $f_{ll'}^{\mu} = -f_{l'l}^{\mu}$ and the denominator of Eq. (7.85) is symmetric in l and l', there is no net contribution from terms $l' \le L$. Thus,

$$\varepsilon_1(\omega) = 1 + \frac{e^2}{m\pi^2} \sum_{l \le L, l' > L} \int d\mathbf{k} \frac{f_{l'l}^{\mu}}{\omega_{l'l}^2 - \omega^2} \tag{7.88}$$

Now, as we have seen earlier, the condition for the existence of plasma oscillations is that $\varepsilon_1(\omega) = 0$. The only way in which $\varepsilon_1(\omega)$ can be zero is for the integral in Eq. (7.88) to be negative. Since all the individual terms in the integral are positive, the integral can be negative only if the denominator is negative. That is, only if $\omega_{l'l} < \omega$ can plasma oscillations exist. Thus, the plasma frequency must lie above the band gap.

PROBLEMS

7.1 Prove that using the factor $e^{\eta t}$ in Eq. (7.23) is equivalent to satisfying causality.

7.2 Show that $4\pi e^2/q^2$ is the Fourier transform of the Coulomb potential.

7.3 Derive Eq. (7.43) with the help of the summation technique used in Section 5.2.

7.4 Read H. Ehrenreich and M. H. Cohen, *Phys. Rev.* **115**, 786 (1959). Modify Eq. (28) of this paper by introducing a collision term as described

in footnote 19 of H. Ehrenreich and H. R. Philipp, *Phys. Rev.* **128**, 1622 (1962). Use this to derive Eq. (1) of the paper by Ehrenreich and Philipp.

7.5 Show that, by writing the summation in Eq. (7.85) in the form $\Sigma_{l \leq L, l'}$, where L is the highest-lying valence band, the plasma frequency can be obtained rather simply if core states can be neglected and $\omega \gg \omega_{l'L}$ for all bands l' that contribute appreciably to absorption.

7.6 Read S. L. Adler, *Phys. Rev.* **126**, 413 (1962). Derive Eq. (1.23) of Adler's paper starting with Eq. (1.19). Note that it is the assumption of Eq. (1.9) that permits the derivation of a general dielectric function from which transverse and longitudinal dielectric functions can be projected.

7.7 Read N. Wiser, *Phys. Rev.* **129**, 62 (1963). Find the trivial error in Eq. (10) of Wiser's paper. Summarize the important results of this paper.

REFERENCES AND FURTHER READING

1. H. Ehrenreich and M. H. Cohen, Self-Consistent Field Approach to the Many-Electron Problem, *Phys. Rev.* **115**, 786 (1959).
2. S. L. Adler, Quantum Theory of the Dielectric Constant in Real Solids, *Phys. Rev.* **126**, 413 (1962).
3. N. Wiser, Dielectric Constant with Local Field Effects Included, *Phys. Rev.* **129**, 62 (1963).
4. H. Ehrenreich, Electromagnetic Transport in Solids: Optical Properties and Plasma Effects, *in* "Optical Properties of Solids" (J. Tauc, ed.). Academic Press, New York, 1965.
5. P. Nozieres and D. Pines, A Dielectric Formulation of the Many Body Problem: Application to the Free Electron Gas, II, *Nuovo Cimento* **9**, 470 (1958).
6. P. Nozieres, "Theory of Interacting Fermi Systems," Benjamin, New York, 1963.
7. D. Pines, "Elementary Excitations in Solids," Benjamin, New York, 1963.
8. R. D. Mattuck, "Guide to Feynman Diagrams in the Many Body Problem," McGraw-Hill, New York, 1967.

Chapter 8

CURRENT-CURRENT CORRELATIONS AND THE FLUCTUATION-DISSIPATION THEOREM

An equation relating voltage fluctuations (noise) in electrical systems to the electrical resistance was derived in 1928 by Nyquist [1]. This was an accomplishment of great fundamental importance, for Nyquist was able to describe an irreversible process (the dissipation of energy in an electrical resistor) in terms of a thermal equilibrium property (voltage fluctuations) of the system. The Nyquist relation has since been extended and generalized by Callen and Welton [2]. Further advances, with applications to optical properties, have been made by Kubo [3].

The Kubo formalism is abstract, and has not been used by experimentalists studying the optical properties of solids. However, it is completely rigorous for linear systems, and enables the complete conductivity tensor to be expressed in terms of current fluctuations occurring in a system in thermal equilibrium in the absence of perturbing radiation. It is thus of great theoretical interest and is useful for casting new insights into the optical properties of solids.

In this chapter, it is first shown that the absorption of a system can be expressed as a product of two terms, one representing the probe and the other representing the thermal equilibrium current fluctuations of the system. Next, a simple relationship concerning current fluctuations is derived. Finally, these results are used in a simple application of the fluctuation-dissipation theorem to derive the conductivity of a Drude-like system.

8.1 Transition Rate and Current-Current Correlations

Consider a system in state $|s\rangle$ that absorbs a photon in state $|\mathbf{q}\rangle$. After absorption of a photon, the system is in state $|s'\rangle$. Using occupation number formalism to describe the photons in state $|\mathbf{q}\rangle$, the transition rate is

$$W_{s's} = (2\pi/\hbar^2)\,|\langle s', n_{\mathbf{q}} - 1\,|\,H^{\text{ext}}|\,s, n_{\mathbf{q}}\rangle|^2\,\delta(\omega - \omega_{s's}) \tag{8.1}$$

where $n_{\mathbf{q}}$ specifies the photon density when the system is in state $|s\rangle$ and $n_{\mathbf{q}} - 1$ specifies the photon density when the system has been excited to state $|s'\rangle$.

The Hamiltonian

$$H^{\text{ext}} = -(1/c)\int d\mathbf{r}\,\mathbf{J}(\mathbf{r})\cdot\mathbf{A}^{\text{ext}}(\mathbf{r}, t) \tag{8.2}$$

describes the coupling of the external perturbation (the light wave) to the system. The operator $\mathbf{A}^{\text{ext}}(\mathbf{r}, t)$ is given (see Appendix E) in the Schrödinger picture by

$$\mathbf{A}^{\text{ext}}(\mathbf{r}, t) = (2\pi\hbar c^2/\omega)^{1/2}\sum_{\mathbf{q},\eta}\{b_{\mathbf{q},\eta}\boldsymbol{\eta}[\exp i(\mathbf{q}\cdot\mathbf{r} - \omega t)]$$

$$+ b^{\dagger}_{\mathbf{q},\eta *}\boldsymbol{\eta}^{*}\exp - i(\mathbf{q}\cdot\mathbf{r} - \omega t)\} \tag{8.3}$$

where $\boldsymbol{\eta}$ is a polarization index, and where $b_{\mathbf{q},\eta}$ and $b^{\dagger}_{\mathbf{q},\eta *}$ are photon annihilation and creation operators, respectively.

If we consider monochromatic polarized light incident on an isotropic system, and if we consider only the absorption of light, we can take

$$\mathbf{A}^{\text{ext}}(\mathbf{r}, t) = (2\pi\hbar c^2/\omega)^{1/2}\,b_{\mathbf{q}}\exp i(\mathbf{q}\cdot\mathbf{r} - \omega t) \tag{8.4}$$

Then the matrix element for the interaction is, from Eqs. (8.2) and (8.4), and following the conventions of Chapter 3,

$$\langle|H^{\text{ext}}|\rangle = -(2\pi\hbar/\omega)^{1/2}\,\langle s', n_{\mathbf{q}} - 1|\int d\mathbf{r}\,\mathbf{J}(\mathbf{r})b_{\mathbf{q}}\exp i\mathbf{q}\cdot\mathbf{r}|s, n_{\mathbf{q}}\rangle \tag{8.5}$$

The action of the photon annihilation operator changes this to

$$\langle|H^{\text{ext}}|\rangle = -(2\pi\hbar n_{\mathbf{q}}/\omega)^{1/2}\,\langle s', n_{\mathbf{q}} - 1|\int d\mathbf{r}\,\mathbf{J}(\mathbf{r})\exp i\mathbf{q}\cdot\mathbf{r}|s, n_{\mathbf{q}} - 1\rangle \tag{8.6}$$

Equation (8.6) is further simplified by introducing the Fourier transform

$$\mathbf{J}^{*}(\mathbf{q}) = [1/(2\pi)^3]\int d\mathbf{r}\,\mathbf{J}(\mathbf{r})\exp i\mathbf{q}\cdot\mathbf{r} \tag{8.7}$$

so that the transition rate can be written as

$$W_{s's} = (2\pi)^6 (4\pi^2 n_{\mathbf{q}}/\hbar\omega) \langle s|\mathbf{J}(\mathbf{q})|s'\rangle \langle s'|\mathbf{J}^*(\mathbf{q})|s\rangle \delta(\omega - \omega_{s's}) \tag{8.8}$$

Equation (8.8) is not yet very useful. The δ function really has meaning only as part of an integrand, and the system is never in a pure state. If we describe the system by a mixture of states, with each initial state weighted by a density operator to give the proper statistical distribution of initial states, we can replace Eq. (8.8) by the more general formula

$$\sum_{s's} W_{s's} = (2\pi)^6 (4\pi^2 n_{\mathbf{q}}/\hbar\omega) \sum_{s's} \rho_s \langle s|\mathbf{J}(\mathbf{q})|s'\rangle \langle s'|\mathbf{J}^*(\mathbf{q})|s\rangle \delta(\omega - \omega_{s's}) \tag{8.9}$$

A summation over final states has also been included in Eq. (8.9). It is now convenient to replace the δ function by

$$\delta(\omega - \omega_{s's}) = (1/2\pi) \int_{-\infty}^{\infty} \exp[i(\omega - \omega_{s's})t]\, dt \tag{8.10}$$

so that the transition rate becomes

$$\sum_{s's} W_{s's} = (2\pi)^6 (2\pi n_{\mathbf{q}}/\hbar\omega) \int_{-\infty}^{\infty} dt\, (\exp i\omega t)$$

$$\times \sum_{s's} \rho_s \langle s|\mathbf{J}(\mathbf{q})|s'\rangle \langle s'|\mathbf{J}^*(\mathbf{q})|s\rangle \exp[-i(\omega_{s'} - \omega_s)t] \tag{8.11}$$

Equation (8.11) can be further modified by expressing the time-dependent terms $\exp[-i\omega_{s'}t]$ and $\exp(i\omega_s t)$ as exponential operators acting on the states $|s'\rangle$ and $|s\rangle$. Thus,

$$\sum_{s's} W_{s's} = (2\pi)^6 (2\pi n_{\mathbf{q}}/\hbar\omega) \int_{-\infty}^{\infty} dt\, e^{i\omega t} \sum_{s's} \rho_s$$

$$\times \langle s|\mathbf{J}(\mathbf{q})|s'\rangle \langle s'|e^{-iH_0 t/\hbar}\mathbf{J}^*(\mathbf{q})e^{iH_0 t/\hbar}|s\rangle \tag{8.12}$$

The operators $\mathbf{J}(\mathbf{q})$ and $\mathbf{J}^*(\mathbf{q})$ are time-independent in the Schrodinger picture. The corresponding time-dependent operators in the interaction picture are

$$\mathbf{J}_{\mathbf{q}}(t) = e^{iH_0 t/\hbar}\mathbf{J}(\mathbf{q})e^{-iH_0 t/\hbar} \tag{8.13}$$

$$\mathbf{J}_{\mathbf{q}}{}^*(t) = e^{-iH_0 t/\hbar}\mathbf{J}^*(\mathbf{q})e^{iH_0 t/\hbar} \tag{8.14}$$

Using these operators, and the condition of closure for a complete set of states,

$$\sum_{s'} |s'\rangle \langle s'| = 1 \tag{8.15}$$

Eq. (8.12) becomes

$$\sum_{s's} W_{s's} = (2\pi)^6 (2\pi n_{\mathbf{q}}/\hbar\omega) \int dt\, e^{i\omega t} \sum_{s} \rho_s \langle s|\mathbf{J}_{\mathbf{q}}(0)\mathbf{J}_{\mathbf{q}}^*(t)|s\rangle \tag{8.16}$$

The transition rate has now been expressed as a product of two factors. One depends on the probe (the light beam with intensity proportional to the photon population density $n_{\mathbf{q}}$); the other depends on the correlation between current fluctuations in the ground state of the system at different times. The correlation function contains all the information about the system that can be obtained in an optical experiment. The power absorption of the system is just given by the product of transition rate and energy $\hbar\omega$ absorbed per transition. Thus,

$$\text{Power} = (2\pi)^6 2\pi n_{\mathbf{q}} \int dt\, e^{i\omega t} \sum_{s} \rho_s \langle s|\mathbf{J}_{\mathbf{q}}(0)\mathbf{J}_{\mathbf{q}}^*(t)|s\rangle \tag{8.17}$$

8.2 Current Fluctuations

We have just found the power absorption of a system to depend upon the correlation of current fluctuations in the system in thermal equilibrium. The expectation value of $\mathbf{J}_{\mathbf{q}}(t)$ is of course zero, but the presence of spontaneous fluctuations in $\mathbf{J}_{\mathbf{q}}(t)$ leads to a finite expectation value for $\mathbf{J}_{\mathbf{q}}{}^2(t) = \mathbf{J}_{\mathbf{q}}{}^2$.

A convenient way to determine $\langle s|\mathbf{J}_{\mathbf{q}}{}^2|s\rangle$ is through the Heisenberg equation of motion for $\mathbf{J}_{\mathbf{q}}(t)$ and the relationship

$$(d/dt)\mathbf{J}_{\mathbf{q}}(t) = \dot{\mathbf{J}}_{\mathbf{q}}(t) = -i\omega\mathbf{J}_{\mathbf{q}}(t) \tag{8.18}$$

From Eq. (8.18),

$$\langle \mathbf{J}_{\mathbf{q}}{}^2 \rangle = (1/\omega^2) \langle \dot{\mathbf{J}}_{\mathbf{q}}{}^2 \rangle \tag{8.19}$$

and thus the expectation value of $\mathbf{J}_{\mathbf{q}}{}^2$ in the state $|s\rangle$ can be written

$$\langle s|\mathbf{J}_{\mathbf{q}}{}^2|s\rangle = (1/\omega^2) \sum_{s'} \langle s|\dot{\mathbf{J}}_{\mathbf{q}}|s'\rangle \langle s'|\dot{\mathbf{J}}_{\mathbf{q}}{}^*|s\rangle \tag{8.20}$$

Now, introducing the equation of motion for $\mathbf{J}_{\mathbf{q}}$ in operator form,

$$\dot{\mathbf{J}}_{\mathbf{q}} = (i/\hbar)(H\mathbf{J}_{\mathbf{q}} - \mathbf{J}_{\mathbf{q}}H) \tag{8.21}$$

Eq. (8.20) becomes

$$\begin{aligned} \langle s|\mathbf{J}_{\mathbf{q}}{}^2|s\rangle &= (1/\hbar^2\omega^2) \langle s|H\mathbf{J}_{\mathbf{q}} - \mathbf{J}_{\mathbf{q}}H|s'\rangle \langle s'|H\mathbf{J}_{\mathbf{q}}{}^* - \mathbf{J}_{\mathbf{q}}{}^*H|s\rangle \\ &= [(\mathscr{E}_s - \mathscr{E}_{s'})^2/\hbar^2\omega^2]\, |\langle s'|\mathbf{J}_{\mathbf{q}}|s\rangle|^2 \end{aligned} \tag{8.22}$$

For a Fourier component having wave vector $\mathbf{q}$ and frequency ω,

$$|\mathscr{E}_s - \mathscr{E}_{s'}| = \hbar\omega \tag{8.23}$$

Thus

$$\langle s|\mathbf{J_q}^2|s\rangle = \sum_{s'} |\langle s'|\mathbf{J_q}|s\rangle|^2 \tag{8.24}$$

This simple identity is used in the next section in the calculation of the conductivity for a simple system.

8.3 The Fluctuation-Dissipation Theorem and the Conductivity

In Chapter 4, we discussed the classical theory of nearly-free-electron metals. It is helpful to reconsider some aspects of that model as a simple example of the relationship between fluctuations and dissipation within a system.

Imagine that we can watch the motion of a single electron in a metal. The motion is random as a result of the thermal motion of the lattice and all other electrons in the metal. It is much like the Brownian motion of colloidal particles which can be observed with the aid of a microscope. These fluctuations exist even in thermal equilibrium.

If an electric field is applied to the metal, the electrons undergo a forced motion superposed on the existing random motion. Along with the applied driving force, there is a damping force which arises from impacts with the lattice (phonons). These impacts are exactly the same as those which result in random motion of the electrons in thermal equilibrium, but now they act to produce a constant average viscous damping force proportional to the velocity.

It is clear that the frictional force and the random force are of the same origin and so they must be related. The relationship between these microscopic forces is stated in quite general terms by the fluctuation–dissipation theorem.

We have already calculated the dissipation (the power absorption) in the system in Section 8.1. In the more general treatment of Callen and Welton [2], the dissipation calculation also includes a term for emission of energy. At normal temperatures and for typical optical frequencies, that is a negligible contribution. Now, rather than derive the generalized fluctuation–dissipation theorem, we will relate the dissipation to current fluctuations and obtain the optical conductivity of a simple system.

We can define the conductivity of the system in terms of power absorption in the usual way as

$$\text{Power} = \sigma E^2 \tag{8.25}$$

The question is, "What do we use for E^2?" We have just seen that in the simple model of a metal, the frictional forces leading to energy absorption are of the same origin as the random forces causing fluctuations. Based on this simple physical model, we assert that the appropriate E^2 to use is just that value of $\mathbf{E} \cdot \mathbf{E}^*$ produced by the fluctuations in current. The fluctuating electric fields and the dissipation mechanism are all part of the same thing! That is, we imagine the fluctuations in electric field to be produced by the same fluctuations in current we wish to calculate. In this way, using Eqs. (8.4), (8.17), and (8.25), we obtain

$$\sigma(\omega) = [(2\pi)^6/\hbar\omega] \int dt\, e^{i\omega t} \sum_s \rho_s \langle s|\mathbf{J}_\mathbf{q}(0)\mathbf{J}_\mathbf{q}^*(t)|s\rangle \tag{8.26}$$

for $\hbar\omega \gg kT$. The generalization of this result to obtain the elements of the conductivity tensor at all frequencies yields (following Kubo [3])

$$\sigma_{xx}(\omega) = \frac{(2\pi)^6}{\hbar\omega \coth(\hbar\omega/2kT)} \int dt\, e^{i\omega t} \sum_s \rho_s \langle s|J_{q_x}(0) J_{q_x}^*(t)|s\rangle \tag{8.27}$$

An explicit calculation of Eq. (8.26) is, in general, impossibly complicated. There is one simple and instructive case, however, for which an explicit calculation can be easily made. If we assume a constant correlation time τ for all states of the system, such that $\mathbf{J}_\mathbf{q}(t) = \mathbf{J}_\mathbf{q}(0)e^{-t/\tau}$ then Eq. (8.26) becomes

$$\sigma(\omega) = [(2\pi)^6/\hbar\omega] \int_{-\infty}^{\infty} dt\, e^{i\omega t} e^{-|t|/\tau} \sum_s \rho_s \langle s|\mathbf{J}_\mathbf{q}{}^2|s\rangle \tag{8.28}$$

Note that τ is a measure of the average time a current persists in a particular state; it is the same as the relaxation time as we used it in Chapter 4. Using Eq. (8.24), we now get

$$\sigma(\omega) = [(2\pi)^6/\hbar\omega] \int dt\, e^{i\omega t} e^{-|t|/\tau} \sum_{s's} \rho_s |\langle s'|\mathbf{J}_\mathbf{q}|s\rangle|^2 \tag{8.29}$$

Now, if the current operator $\mathbf{J}_\mathbf{q}$ is replaced by its Fourier integral expansion

$$\mathbf{J}_\mathbf{q} = [1/(2\pi)^3] \int d\mathbf{r}\, \mathbf{J}(\mathbf{r}) \exp -i\mathbf{q} \cdot \mathbf{r} \tag{8.30}$$

where

$$\mathbf{J}(\mathbf{r}) = -(e/2m) \sum_i [\mathbf{p}_i\, \delta(\mathbf{r} - \mathbf{r}_i) + \delta(\mathbf{r} - \mathbf{r}_i)\mathbf{p}_i] \tag{8.31}$$

and if the dipole approximation $\exp i\mathbf{q} \cdot \mathbf{r} \approx 1$ is used, Eq. (8.29) becomes

$$\sigma(\omega) = \frac{1}{\hbar\omega} \frac{e^2}{m^2} \int_{-\infty}^{\infty} dt\, e^{i\omega t} e^{-|t|/\tau} \sum_{s's,i} \rho_s |\langle s'|\mathbf{p}_i|s\rangle|^2 \tag{8.32}$$

Using the definition of oscillator strength

$$f_{s's} = 2|\mathbf{p}_{s's}|^2/m\hbar\omega_{s's} \tag{8.33}$$

and carrying out the integration and summations in Eq. (8.32) yields

$$\sigma(\omega) = Ne^2\tau/m(1 + \omega^2\tau^2) \tag{8.34}$$

We have thus arrived at the same simple expression for the conductivity as is obtained with the Drude model of a metal.

PROBLEM

8.1 We arrived at the Drude expression for the conductivity, Eq. (8.34), by making some quite special assumptions. Show that the following procedure leads to an expression for the conductivity which is essentially equivalent to that of the Lorentz model for interband transitions. Start with Eq. (8.26) and insert a projection operator $|s'\rangle\langle s'|$ such that the correlation function becomes

$$\sum_{ss'} \rho_s \langle s|\mathbf{J}_{\mathbf{q}}(0)|s'\rangle \langle s'|\mathbf{J}^*_{-\mathbf{q}}(t)|s\rangle$$

Rewrite the operator $\mathbf{J}^*_{-\mathbf{q}}(t)$ in terms of the Schrödinger picture using Eq. (8.14), but recognize that $\mathbf{J}^*(\mathbf{q})$ in the Schrödinger picture is equivalent to $\mathbf{J}_{\mathbf{q}}{}^*(0)$ in the interaction picture. Show that the operators $\exp(-iH_0t/\hbar)$ and $\exp(iH_0t/\hbar)$ operating on the states $\langle s'|$ and $|s\rangle$ lead to a term $\exp(-i\omega_{s's}t)$ and hence, following the procedure of Section 8.3, lead to a Lorentzian-like expression for the conductivity.

REFERENCES AND FURTHER READING

1. H. Nyquist, *Phys. Rev.* **32**, 110 (1928).
2. H. B. Callen and T. A. Welton, *Phys. Rev.* **83**, 34 (1951).
3. R. Kubo, *J. Phys. Soc. Japan* **12**, 570 (1957).
4. V. L. Bonch-Bruevich, Fundamentals of Solid State Optics, *in* "Optical Properties of Solids" (J. Tauc, ed.), Academic Press, New York, 1965.
5. P. C. Martin, Measurements and Correlation Functions, *in* "Many-Body Physics" (C. DeWitt and R. Balian, eds.), Gordon and Breach, New York, 1968.

Chapter 9

PLASMONS AND CHARACTERISTIC ENERGY LOSSES

We have discussed plasmons in a number of places in previous chapters. We have seen the sharp drop in reflectance for free-electron metals at the plasma frequency and the dramatic shift in the plasma frequency of silver caused by the d bands. All of this earlier discussion is related to bulk plasmons or volume plasmons. Surface plasmons also exist, and we shall discuss them briefly at the end of this chapter, but what is meant here by the term plasmon is a quantized collective oscillation of electrons in the bulk of a material.

Confusion sometimes results from the common implication that volume plasmons are actually excited in the usual normal-incidence optical reflectance experiment, but the implication is wrong. Volume plasmons are not excited by light at normal incidence. The excitation of plasmons by light at nonnormal incidence is discussed in some of the references listed at the end of this chapter. This is a field of considerable recent interest and activity, but it is one which is not so directly related to the central theme of this book. Thus the discussion within this chapter is limited to interactions at normal incidence.

Plasmons are oscillations of fluctuations in charge density. These oscillations cannot be produced by divergenceless transverse electromagnetic waves. The appearance in the optical reflectance of what are called plasma oscillations results simply from the fact that (for isotropic material) the

transverse and longitudinal dielectric functions are numerically equal in the long-wavelength limit. Even anisotropic materials can be included if the prescription in Section 2.9 is followed. Thus, when an optical reflectance spectrum is measured to determine the dielectric function and plasma frequency, it does not mean that a plasmon is actually excited during the experiment. It simply means that one has determined the frequency at which a plasma oscillation would occur if one had a means of exciting it.

To excite a plasmon, it is necessary to produce a fluctuation in charge density. The simplest way to do this is with a beam of electrons or other charged particles. Such a beam can excite collective modes of oscillation by means of the long-range Coulomb forces. It also can excite all the usual single-electron and exciton transitions.

Because an electron beam can excite single-electron and many-electron modes, it provides a probe for determining the dielectric function. The essence of the experiment consists in shining a monoenergetic beam of electrons onto a sample and measuring the distribution in energy and angle of the electrons transmitted or reflected by the sample. A Kramers–Kronig analysis of the loss function obtained in this way allows a determination of ε_1 and ε_2. The utility of the experiment is that it makes possible the determination of optical properties in an energy range which is inaccessible or for which surface contamination or some other sample preparation problem makes optical reflectance measurements not feasible. Of course, this utility depends upon the equivalence of the longitudinal and transverse dielectric functions.

In this chapter, we shall first consider some aspects of single-electron excitations that will clarify the role of Coulomb forces in introducing a new feature into the excitation spectrum, namely plasmons. After discussing plasmons in free-electron metals, we shall see how well the optical properties obtained from characteristic energy loss measurements agree with those obtained from optical measurements. Finally, we shall briefly discuss surface plasmons (which sometimes can be excited by light) and their effect on optical properties and photoemission.

9.1 Single-Electron Excitations in Metals

We consider first a free-electron metal of unit volume. The ground state is that in which two electrons are assigned to each of the lowest $N/2$ energy levels. All the states in $\mathbf{k}$-space are occupied out to a radius k_F, the radius of the Fermi sphere. The excited states of the system are states in which an electron has been removed from the Fermi sphere, leaving a hole behind. If the perturbation causing the transition imparts momentum $\hbar\mathbf{q}$ to an

electron in state $\mathbf{k}$, then the energy imparted is

$$\hbar\omega = \frac{\hbar^2(\mathbf{k}+\mathbf{q})^2}{2m} - \frac{\hbar^2 k^2}{2m}$$

$$= \frac{\hbar^2}{m}\left(\frac{q^2}{2} + \mathbf{k}\cdot\mathbf{q}\right) \tag{9.1}$$

The region in $\mathbf{k}$-space from which electrons can be excited is illustrated in Fig. 9.1. The solid circle of radius k_F defines the surface of the Fermi sphere. A change in momentum $\hbar\mathbf{q}$ must take an electron from a state in the Fermi sphere to an unoccupied state outside the Fermi sphere. Thus, the possible initial states are found to lie within the shaded crescent found by translating the Fermi sphere by $-\mathbf{q}$.

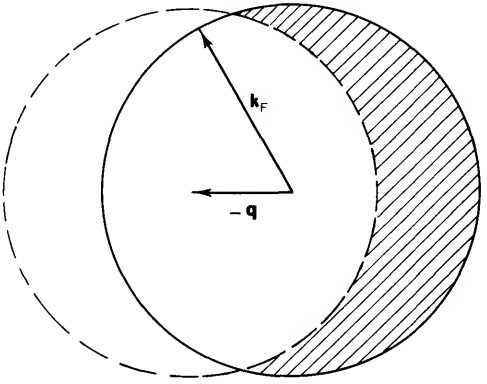

Fig. 9.1 Region in $\mathbf{k}$-space from which electrons can be excited with momentum change $\hbar\mathbf{q}$. The possible initial states lie within the shaded crescent found by translating the Fermi sphere (solid circle of radius k_F) by $-\mathbf{q}$.

Now let us find the possible values of $\hbar\omega$ consistent with Eq. (9.1). If $q \leq 2k_F$, transitions are possible from states on the Fermi surface to states infinitesimally outside the Fermi surface. For these transitions, even though of possibly large q, the energy change is zero. The most common example of this is ordinary electron–phonon scattering. If $q > 2k_F$, transitions must take place from a state in the Fermi sphere to a state lying outside the Fermi sphere. The minimum-energy change is for a transition of the type illustrated in Fig. 9.2. Thus,

$$\begin{aligned} \hbar\omega_{\min} &= 0, & q &\leq 2k_F \\ &= \frac{\hbar^2}{m}\left(\frac{q^2}{2} - qk_F\right), & q &> 2k_F \end{aligned} \tag{9.2}$$

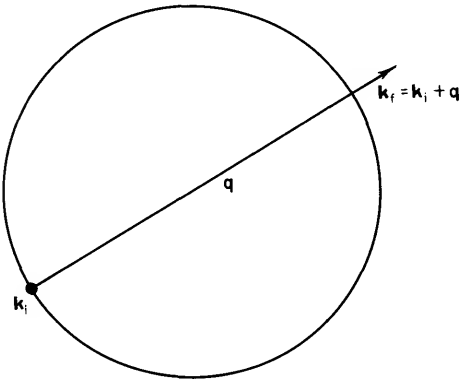

Fig. 9.2 A transition of minimum energy for $q > 2k_F$. The initial state $\mathbf{k}_i$ lies on the Fermi surface. The final state $\mathbf{k}_f$ is antiparallel to $\mathbf{k}_i$ and lies outside the opposite surface.

The maximum change in energy is for a transition from a state $\mathbf{k}_i$ on the Fermi surface to a final state $\mathbf{k}$ such that $\mathbf{k}_i \| \mathbf{q}$. Thus,

$$\hbar\omega_{\max} = \frac{\hbar^2}{m}\left(\frac{q^2}{2} + qk_F\right), \qquad q \geqslant 0 \tag{9.3}$$

The spectrum of possible single-electron (electron–hole pair) excitations is shown in Fig. 9.3. For each value of q, the range of possible excitation energies lies along a vertical line between the curves for $\hbar\omega_{\min}$ and $\hbar\omega_{\max}$.

If electron–electron interactions are included, single-electron excitations

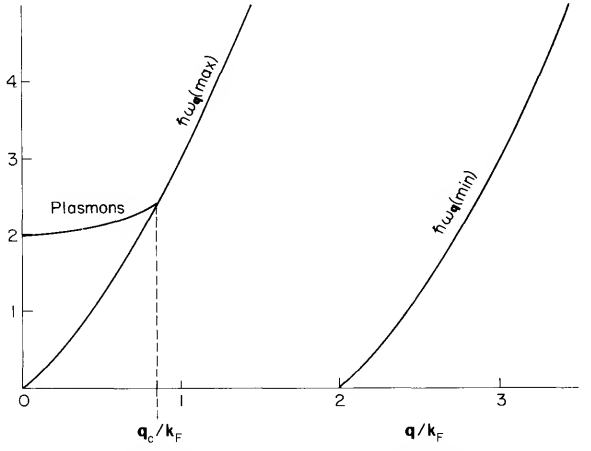

Fig. 9.3 Spectrum of excitations for a free-electron gas. A collective plasma mode splits off from the continuum of single-electron excitations below a critical wave vector q_c. In this illustration, $\hbar\omega_p$ has been taken equal to twice the Fermi energy, hence $\hbar\omega_p = \hbar^2 k_F{}^2/m$.

still exist, although the energies are slightly shifted. This is because the Coulomb interactions are strongly screened, as was discussed in Section 7.2. However, Coulomb interactions introduce a new feature, the possibility of collective oscillations, or plasmons.

9.2 Plasmons in Simple Metals

Plasma oscillations can exist because for small q, the dielectric function does not always act as a screening factor. The long-range Coulomb interactions then make possible collective oscillations of large numbers of electrons. We have already seen that the condition for plasma oscillations is $\varepsilon_1 = 0$, and that in the limit $q \to 0$.

$$\varepsilon_1 = 1 - (\omega_p{}^2/\omega^2) \tag{9.4}$$

The derivation of the plasma frequency as given in Fig. 3.19 is for $q = 0$; it is implicit in the model used there. However, we are now in a position to determine the q dependence of the plasma frequency. We simply require that $\hat{\varepsilon}(q, \omega)$ as given by Eq. (7.41) equal zero. If we assume damping to be small, we need be concerned only with $\varepsilon_1(q, \omega)$. Then, since for a free-electron metal only the equivalent of intraband transitions arise, we can follow through more or less the same steps as those leading to Eq. (7.84). The only difference is that in order to include the q dependence of the dielectric function, we cannot neglect the terms $(\mathscr{E}_{\mathbf{k}+\mathbf{q},l} - \mathscr{E}_{\mathbf{k}l})^2$ and $(\mathscr{E}_{\mathbf{k}l} - \mathscr{E}_{\mathbf{k}-\mathbf{q},l})^2$ in the denominator of Eq. (7.61). If these terms are included, we get, instead of Eq. (7.65),

$$\varepsilon_1(q,\omega) = 1 - \frac{4\pi e^2}{\Omega} \sum_{\mathbf{k}l} \frac{f(\mathscr{E}_{\mathbf{k}l})\, \partial^2 \mathscr{E}_{\mathbf{k}l}/\partial k_\mu{}^2}{(\hbar\omega_q)^2 - (q\, \partial \mathscr{E}_{\mathbf{k}l}/\partial k_\mu)^2} \tag{9.5}$$

Then, following through as in Eqs. (7.81)–(7.84), we get, instead of Eq. (7.84),

$$\varepsilon_1(q, \omega) = 1 - \frac{4\pi N e^2}{m\omega_q{}^2} \frac{1}{1 - (\hbar^2 q^2 \langle k^2 \rangle / m^2 \omega_q{}^2)} \tag{9.6}$$

For a free-electron gas,

$$\langle k^2 \rangle = \tfrac{3}{5} k_F{}^2 \tag{9.7}$$

and

$$\hbar k_F / m = v_F \tag{9.8}$$

where v_F is the velocity of an electron on the Fermi surface. Using these results and assuming

$$\hbar^2 q^2 \langle k^2 \rangle / m^2 \omega_q{}^2 \ll 1 \tag{9.9}$$

Eq. (9.6) becomes

$$\varepsilon_1(q,\omega) \approx 1 - \frac{4\pi N e^2}{m\omega_q{}^2}\left(1 + \tfrac{3}{5}\frac{q^2 v_F{}^2}{\omega_q{}^2}\right)$$

$$= 1 - \frac{\omega_p{}^2}{\omega_q{}^2}\left(1 + \tfrac{3}{5}\frac{q^2 v_F{}^2}{\omega_q{}^2}\right) \tag{9.10}$$

Since the second term in parentheses in Eq. (9.10) is small, we can approximate ω_q there by ω_p, the infinite-wavelength plasma frequency. Then, setting $\varepsilon_1(q,\omega) = 0$, we get

$$\omega_q \approx \omega_p\left(1 + \tfrac{3}{10}\frac{q^2 v_F{}^2}{\omega_p{}^2}\right) \tag{9.11}$$

The dispersion curve for plasma oscillations is also shown in Fig. 9.3. The plasma dispersion curve merges into the continuum of single-electron excitations at a cutoff wave vector q_c.

9.3 The Plasmon Cutoff Wave Vector

An estimate of the cutoff wave vector q_c can be obtained from a classical analysis of plasma oscillations. We shall now carry through such an analysis.

We assume the usual jellium model of a free-electron metal. The positive ions are smoothed out to form a uniform background of positive charge. The electrons move about at random, but on the average, the electronic charge density just cancels out the uniform positive background.

The potential energy associated with the Coulomb interaction between the jth and kth electrons is

$$e^2/r_{jk} = \sum_{\mathbf{q}} V_{\mathbf{q}} \exp i\mathbf{q}\cdot\mathbf{r}_{jk} \tag{9.12}$$

where (see Problem 7.3)

$$V_{\mathbf{q}} = 4\pi e^2/q^2, \qquad q \neq 0 \tag{9.13}$$

is the Fourier transform of the Coulomb potential. For $q = 0$,

$$V_0 = \int (e^2/r_{jk})\, d\mathbf{r}_k \tag{9.14}$$

which is just the potential energy of the jth electron due to a uniform distribution of one electronic charge. Thus, if we want the Fourier expansion of the potential energy of the jth electron due to all other electrons as well as a uniform positive background, we can simply exclude the terms

with $q = 0$. Then, the positive contributions to the potential energy arising from electron–electron interactions are exactly cancelled by the negative contributions arising from the interaction with the uniform positive background. The result is

$$V(\mathbf{r}_j) = \sum_{\substack{k \\ k \neq j}} \sum_{\mathbf{q}}{}' (4\pi e^2/q^2) \exp i\mathbf{q} \cdot \mathbf{r}_{jk} \tag{9.15}$$

where the prime on the summation denotes that the term with $q = 0$ is to be excluded. The force on the jth electron is

$$\mathbf{F}_j = m\ddot{\mathbf{r}}_j = -\nabla V(\mathbf{r}_j) \tag{9.16}$$

so that the acceleration of the jth electron is

$$\dot{\mathbf{v}}_j = \ddot{\mathbf{r}}_j = -\frac{4\pi e^2}{m} i \sum_k \sum_{\mathbf{q}}{}' \frac{\mathbf{q}}{q^2} \exp i\mathbf{q} \cdot \mathbf{r}_{jk} \tag{9.17}$$

The term with $k = j$ is not now explicitly excluded in the summation of Eq. (9.17); it is zero because of the symmetric distribution of the $\mathbf{q}$ vectors.

We can relate the acceleration of the jth electron to density fluctuations (the origin of plasma oscillations) as follows. The electron particle density is

$$\rho(\mathbf{r}) = \sum \delta(\mathbf{r} - \mathbf{r}_k) = \sum_{\mathbf{q}} \rho_{\mathbf{q}} \exp i\mathbf{q} \cdot \mathbf{r} \tag{9.18}$$

where

$$\rho_{\mathbf{q}} = \int \rho(\mathbf{r}) \exp(-i\mathbf{q} \cdot \mathbf{r})\, d\mathbf{r} = \sum_k \exp -i\mathbf{q} \cdot \mathbf{r}_k$$

Thus, Eq. (9.17) becomes

$$\dot{\mathbf{v}}_j = -\frac{4\pi e^2}{m} i \sum_{\mathbf{q}}{}' \frac{\mathbf{q}}{q^2} \rho_{\mathbf{q}} \exp i\mathbf{q} \cdot \mathbf{r}_j \tag{9.20}$$

We already know that plasma oscillations are oscillations of density fluctuations. Thus, to determine the dispersion equation for plasma oscillations, we need to know the equation describing the oscillations of the various Fourier components of the density fluctuations. From Eq. (9.19), we have

$$\begin{aligned} \ddot{\rho}_{\mathbf{q}} &= (d/dt)\left(\sum_k -i\mathbf{q} \cdot \mathbf{v}_k \exp -i\mathbf{q} \cdot \mathbf{r}_k\right) \\ &= \sum_k \left[-(\mathbf{q} \cdot \mathbf{v}_k)^2 - i\mathbf{q} \cdot \dot{\mathbf{v}}_k\right] \exp -i\mathbf{q} \cdot \mathbf{r}_k \end{aligned} \tag{9.21}$$

Substituting for $\dot{\mathbf{v}}_k$ from Eq. (9.20) yields

$$\ddot{\rho}_{\mathbf{q}} = -\sum_k (\mathbf{q}\cdot\mathbf{v}_k)^2 \exp(-i\mathbf{q}\cdot\mathbf{r}_k)$$

$$-\frac{4\pi e^2}{m}\sum_k{}' \sum_{\mathbf{q}'} \frac{\mathbf{q}\cdot\mathbf{q}'}{(q')^2}\rho_{\mathbf{q}'}\exp[i(\mathbf{q}'-\mathbf{q})\cdot\mathbf{r}_k] \tag{9.22}$$

Now, we argue on physical grounds that only those terms with $\mathbf{q}' = \mathbf{q}$ contribute to the double summation in Eq. (9.22). To see that this must be approximately true, we can use Eq. (9.19) to rewrite the summation over $\mathbf{q}'$ as

$$\sum_{\mathbf{q}'}{}' \frac{\mathbf{q}\cdot\mathbf{q}'}{(q')^2}\rho_{\mathbf{q}'}\exp[i(\mathbf{q}'-\mathbf{q})\cdot\mathbf{r}_k] = \sum_{\mathbf{q}'}{}' \frac{\mathbf{q}\cdot\mathbf{q}'}{(q')^2}\rho_{\mathbf{q}'}\rho_{\mathbf{q}'-\mathbf{q}} \tag{9.23}$$

It is clear that the density fluctuations $\rho_{\mathbf{q}}$ must be small if $\mathbf{q} \neq 0$, for the concentration of electrons is large and fairly uniformly distributed. Furthermore, the $\mathbf{q}$ vectors that appear in the phase factors $\exp i\mathbf{q}\cdot\mathbf{r}_k$ are fairly uniformly distributed; for each $\mathbf{q}$, there is, at least approximately, a $-\mathbf{q}$. Also, the time average of the density fluctuations must be zero. The point of all this is to say, again, that unless $\mathbf{q} = \mathbf{q}'$, the double summation in Eq. (9.22) is approximately zero, at least for reasonably small values of q. This is known as the random-phase approximation.

Using the random-phase approximation, Eq. (9.22) becomes

$$\ddot{\rho}_{\mathbf{q}} = -\sum_k (\mathbf{q}\cdot\mathbf{v}_k)^2 \exp(-i\mathbf{q}\cdot\mathbf{r}_k) - \frac{4\pi e^2 N}{m}\rho_{\mathbf{q}}$$

$$= -\omega_p{}^2\rho_{\mathbf{q}} - \sum_k (\mathbf{q}\cdot\mathbf{v}_k)^2 \exp -i\mathbf{q}\cdot\mathbf{r}_k \tag{9.24}$$

In the limit $q \to 0$, we see that

$$\ddot{\rho}_{\mathbf{q}} = -\omega_p{}^2\rho_{\mathbf{q}} \tag{9.25}$$

and the density fluctuations oscillate with the infinite-wavelength plasma frequency ω_p. These oscillations exist because of the long-range electron–electron interactions. For small but finite q, Eq. (9.24) can be rewritten, using Eq. (9.19), as

$$\ddot{\rho}_{\mathbf{q}} = -\sum_k [\omega_p{}^2 + (\mathbf{q}\cdot\mathbf{v}_k)^2] \exp -i\mathbf{q}\cdot\mathbf{r}_k \tag{9.26}$$

The term $(\mathbf{q}\cdot\mathbf{v}_k)^2$ arises from the thermal motions of the electrons. It is the origin of the dispersion of plasma oscillations.

From Eq. (9.26), we may expect that plasma oscillations exist if

$$\omega_p \gg qv_F \quad \text{or} \quad q^2 \ll q_c^2 \tag{9.27}$$

where $q_c \approx \omega_p/v_F$ and v_F is the velocity of an electron at the Fermi energy, the maximum energy of an electron at $T = 0°K$. However, for $q \gg q_c$, the plasma behaves like a system of free particles.

An estimate of the cutoff wave vector q_c can be obtained by putting in typical values for ω_p and v_F. For typical metals,

$$q_c \approx \omega_p/v_F \approx 10^8 \quad \text{cm}^{-1} \tag{9.28}$$

corresponding to a minimum wavelength of ~ 6Å. This is not surprising. Plasma oscillations are collective modes involving many electrons. The concept of organized oscillations has little meaning when the wavelength of the oscillations is comparable to the mean distance between electrons.

9.4 Characteristic Energy Loss Spectra

When a beam of monoenergetic electrons impinges on a solid, the excitations in the solid can be recognized by the energy losses of the electrons [1]. These energy losses, whether arising from plasmon excitations, interband transitions, exciton creation, or any other excitation, are characteristic of the material. The initial electron energy has no effect on the values of the energy losses.

Figure 9.4 shows the energy loss spectra for a beam of 20-keV electrons

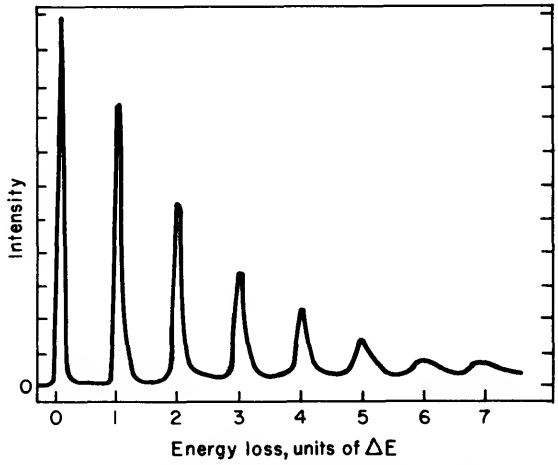

Fig. 9.4 Energy loss spectra for a beam of 20-keV primary electrons passing through an Al foil of thickness 2580Å. The unit of energy loss, ΔE, is about 15 eV, the plasmon excitation energy for Al. [From L. Marton, J. A. Simpson, H. A. Fowler, and N. Swanson, *Phys. Rev.* **126**, 182 (1962).]

incident on an Al foil. In this case, the characteristic energy loss peaks are all due to the excitation of one or more plasmons by electrons passing through the foil.

Other experiments, utilizing different electron beam energies, have shown the same characteristic energy losses as in Fig. 9.4. However, the intensities

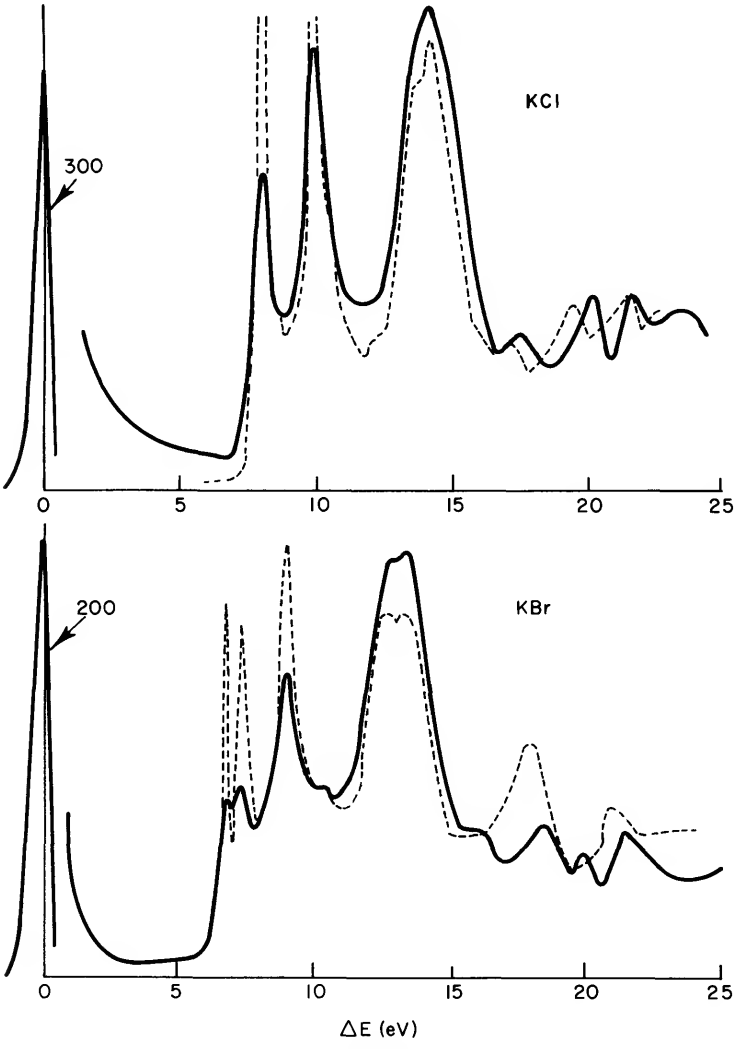

Fig. 9.5 Characteristic energy loss spectra of 50-keV electrons for KCl and KBr. The broken lines represent optical measurements of $-\mathrm{Im}(1/\hat{\varepsilon})$. [From M. Creuzburg and H. Raether, *Solid State Commun.* **2**, 345 (1964).]

of peaks differ, for even though the initial electron energy has no effect on the values of the energy losses, it does strongly affect the cross section and the angular distribution of scattered electrons. These topics are discussed at length in the references and we shall not go into them here. We want to show only that if one does a careful characteristic energy loss experiment, the optical properties can be obtained.

Figure 9.5 shows the energy loss functions for KCl and KBr as obtained from characteristic energy loss experiments. Also shown are the loss functions obtained from the optical data. The peaks at 13.9 and 13.4 eV for KCl and KBr, respectively, are the plasma resonances associated with the collective oscillation of the electrons in the valence band. The other peaks are from excitons and interband transitions.

The excellent agreement between the curves shown in Fig. 9.5, as well as for many other materials, shows that characteristic energy loss measurements provide a useful tool for the study of optical properties.

9.5 Surface Plasmons

The bulk plasmons we have considered so far are purely longitudinal. They cannot couple to transverse electromagnetic waves. However, at the surface of a solid, an oscillation of surface charge density fluctuations is possible. These surface plasmons exist in a number of modes. Some of these modes can radiate and hence can also be excited by light.

For a bounded slab we have, referring to Fig. 9.6,

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\hat{\varepsilon}\mathbf{E}) = \lim_{V \to 0} \left\{ \left[\int \mathbf{n} \cdot (\hat{\varepsilon}\mathbf{E})\, dS \right] \Big/ V \right\}$$

$$= \hat{\varepsilon}_{\mathrm{d}} E_{z=0+} - \hat{\varepsilon}_{\mathrm{m}} E_{z=0-} \tag{9.29}$$

where $\mathbf{n}$ is the unit surface normal for the Gaussian pillbox. In the absence of any external charge, the electric field arises only from polarization charges on the boundary. Thus, by symmetry,

$$E_{z=0-} = -E_{z=0+} \tag{9.30}$$

From Eqs. (9.29) and (9.30), we then get

$$\hat{\varepsilon}_{\mathrm{m}} = -\hat{\varepsilon}_{\mathrm{d}} \tag{9.31}$$

as the condition for a nontrivial solution, that is, one which allows a finite electric field at the boundary.

If we neglect damping and assume a free-electron metal with

$$\varepsilon = 1 - (\omega_{\mathrm{p}}^2/\omega^2) \tag{9.32}$$

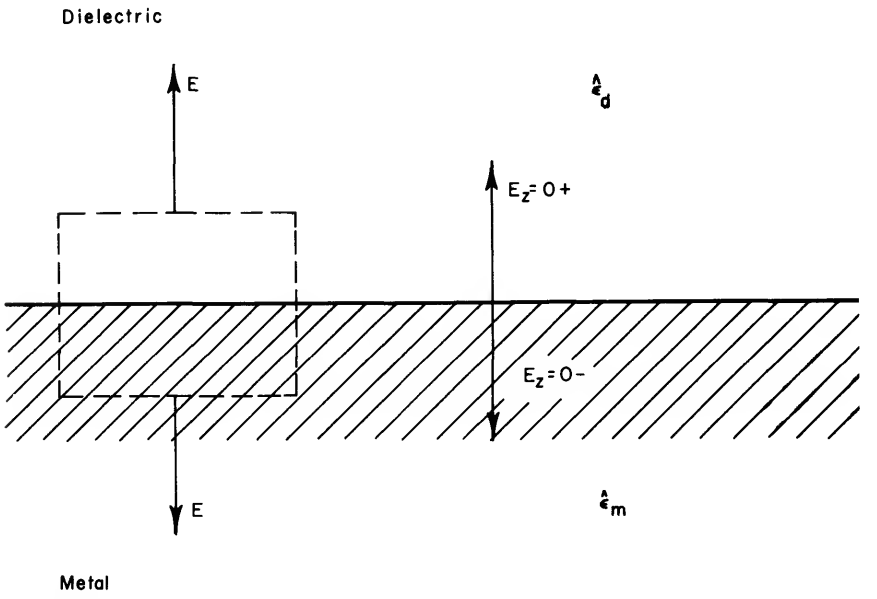

Fig. 9.6 Boundary condition at a metal–dielectric interface in the absence of external charge. The electric field arises only from surface polarization charge and, by symmetry, is of the same magnitude but opposite sign on opposite sides of the interface.

then Eq. (9.31) yields for the surface plasmon frequency

$$\omega_S = \omega_p/(\varepsilon_d + 1)^{1/2} \tag{9.33}$$

For the case in which the dielectric is simply the vacuum, we get

$$\omega_S = \omega_p/\sqrt{2} \tag{9.34}$$

The model presented here is much too simple to allow a discussion of radiative and nonradiative modes or of the effects of retardation and hydrodynamic dispersion. All these are essential for a full understanding of the effects of surface plasmons. However, the intent here is merely to point out the existence of surface plasmons and to give a few examples of their effect. A full quantitative treatment of the many interesting aspects of surface plasmons is to be found in the references.

The oscillation of surface charge fluctuations at the frequency $\omega_p/\sqrt{2}$ cannot ordinarily be excited by light. One mechanism that leads to surface plasmon absorption is the coupling of a surface plasmon to light by means of surface roughness. An example of the effect is shown in Fig. 9.7. The dip in reflectance near $\lambda = 3400$Å arises from surface plasmon absorption.

The reflectance is for an unannealed film of pure silver. It is the same film for which the reflectance after annealing is shown in Figs 3.15 and 5.14. Stanford has measured decreases in reflectance as much as 30% for samples with an rms surface roughness of 24Å [2].

If a sample is sufficiently thick, each surface can sustain independent surface oscillations. However, if $qd \ll 1$, where q is the plasmon wave vector and d is the sample thickness, the oscillations become coupled. A splitting of frequencies results. One mode corresponds to electrons flowing back and forth across the sample foil at nearly the full bulk plasmon frequency ω_p. This mode can be excited by light polarized parallel to the plane of incidence. The optical excitation of this mode is sometimes referred to as the optical excitation of volume plasmons, but it is not the same as the volume plasmons which we have discussed in this book.

Surface plasmons can of course be excited by an electron beam. An interesting example of this is shown in Fig. 9.8, which shows the energy distribution of photoelectrons emitted from cesium. The high-energy peak results from electrons excited from the valence band followed by escape with only inelastic scattering. The second peak is separated from the first by 2.1 eV, the surface plasmon energy for cesium. Similar results have been obtained for the other alkali metals. The implication is that although surface plasmons are important, the mean escape depth for photoelectrons is sufficiently short because of electron–electron scattering

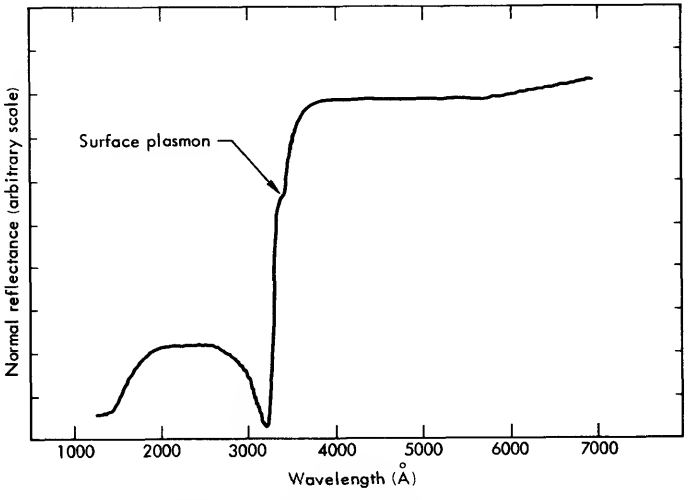

Fig. 9.7 Normal-incidence reflectance for an unannealed film of silver. The dip in reflectance at 3400Å arises from surface plasmon excitation made possible by surface roughness. This dip in reflectance is not present in the annealed film (same sample) shown in Figs. 3.15 and 5.14. [From T. Huen, G. B. Irani, and F. Wooten, *Appl. Opt.* **10**, 552 (1971).]

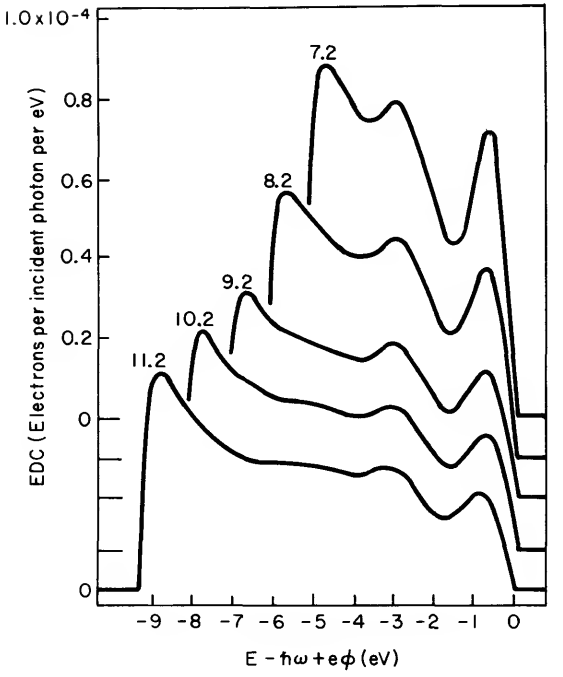

Fig. 9.8 Photoelectron distribution curves for Cs plotted versus the energy of the initial state for unscattered electrons. The zero of energy is thus the Fermi energy. The peak at -3 eV arises from excitation of a surface plasmon by an escaping photoelectron. Values given on the various curves are $\hbar\omega$ in electron-volts. [From N. V. Smith and G. B. Fisher, *Phys. Rev.* **3**, 3662 (1971).]

that the contribution of volume plasmon excitation to the mean free path is unimportant at these low energies.

PROBLEMS

9.1 Show that the average distance between electrons in a free-electron metal is $d = (3/4\pi N)^{1/3}$. Assuming an average interelectronic force e^2/d^2, show that this is equivalent to a characteristic frequency e^2/md^3. Show that this is comparable to the plasma frequency, and thus electrons are not really free except at frequencies well above the plasma frequency.

9.2 Show that the ratio of the energy of a plasmon to the energy of an electron at the Fermi energy in a free-electron metal is $\hbar\omega_p/0.94\mathscr{E}_F r_s^{1/2}$, where r_s is the interelectronic spacing measured in Bohr radii. At metallic

densities, $1.8 < r_S < 6$, and thus thermal excitation of plasmons does not take place at room temperature.

9.3 Show, starting from Eq. (7.41), that

$$\lim_{\substack{q \to 0 \\ \omega \to \infty}} \hat{\varepsilon}(q, \omega) = 1 - (\omega_p{}^2/\omega^2)$$

9.4 Verify the curves plotted in Fig. 9.3.

9.5 Use the results of Problem 9.2 to estimate the cutoff wave vector for plasmons. Is the estimate reasonable?

REFERENCES AND FURTHER READING

1. L. Marton, L. B. Leder, and H. Mendlowitz, Characteristic Energy Losses of Electrons in Solids, *in* "Advances in Electronics and Electron Physics," Vol. 7. Academic Press, New York, 1955.
2. J. L. Stanford, *J. Opt. Soc. Amer.* **60**, 49 (1970).
3. H. Raether, Solid State Excitations by Electrons, *in* "Springer Tracts in Modern Physics," Vol. 38. Springer-Verlag, Berlin, 1965.
4. P. Resibois, Individual and Collective Excitations in an Electron Gas, *in* "Elementary Excitations in Solids" (A. A. Maradudin and G. F. Nardelli, eds.). Plenum Press, New York, 1969.
5. D. Pines, "Elementary Excitations in Solids," Benjamin, New York, 1963.
6. D. Pines and P. Nozieres, "The Theory of Quantum Liquids," Vol. I. Benjamin, New York, 1966.
7. J. Daniels, et al, Optical Constants of Solids by Electron Spectroscopy, *in* "Springer Tracts in Modern Physics," Vol. 54, Springer Verlag, Berlin, 1970.
8. W. Steinmann, Optical Plasma Resonances in Solids, *Phys. Status Solidi* **28**, 437 (1968).
9. E. Tosatti, Anisotropy of Optical Constants and Electron Energy Loss, II, *Nuovo Cimento* **LXIIIB**, 54 (1969).
10. E. A. Stern and R. A. Ferrell, Surface Plasma Oscillations of a Degenerate Electron Gas, *Phys. Rev.* **120**, 130 (1960).
11. H. Frohlich and H. Pelzer, Plasma Oscillations and Energy Loss of Charged Particles in Solids, *Proc. Phys. Soc. (London)* **A68**, 525 (1955).
12. K. L. Kliewer and R. Fuchs, Optical Properties of an Electron Gas: Further Studies of a Nonlocal Description, *Phys. Rev.* **185**, 905 (1969).
13. A. R. Melnyk and M. J. Harrison, Theory of Optical Excitation of Plasmons in Metals, *Phys. Rev.* **2B**, 835 (1970).
14. I. Lindau and P. O. Nilsson, Experimental Verification of Optically Excited Longitudinal Plasmons, *Phys. Scripta* **3**, 87 (1971).
15. S. Raimes, "The Wave Mechanics of Electrons in Metals," North-Holland, Amsterdam, 1961.

Appendix A

DECOMPOSITION OF A VECTOR FIELD INTO LONGITUDINAL AND TRANSVERSE PARTS

Any vector field that describes a real physical phenomenon can be uniquely decomposed into two vector fields, one of which is irrotational and the other divergenceless. We shall discuss this decomposition in somewhat more detail than was done in Chapter 2, but with less mathematical rigor than can be found in Morse and Feshbach [1].

Consider a vector field that satisfies the equation

$$\nabla^2 \mathbf{W} = -\mathbf{F} \tag{A.1}$$

The solution is, in analogy with Poisson's equation for the scalar potential,

$$\mathbf{W}(\mathbf{r}) = (1/4\pi) \int \frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{A.2}$$

Using the vector identity

$$\nabla^2 \mathbf{W} = \nabla(\nabla \cdot \mathbf{W}) - \nabla \times (\nabla \times \mathbf{W}) \tag{A.3}$$

and defining

$$\nabla \cdot \mathbf{W} = -U, \qquad \nabla \times \mathbf{W} = \mathbf{A} \tag{A.4}$$

we get

$$\mathbf{F} = \nabla U + \nabla \times \mathbf{A} \tag{A.5}$$

That a vector field can be uniquely expressed as the sum of a gradient and

a curl is known as Helmholtz's theorem. Given a vector field $\mathbf{F}$, the decomposition into a gradient and a curl can be carried out using Eqs. (A.2), (A.4), and (A.5).

We now define two new vectors

$$\mathbf{F}^{\mathrm{L}} = \nabla U, \qquad \mathbf{F}^{\mathrm{T}} = \nabla \times \mathbf{A} \tag{A.6}$$

so that

$$\mathbf{F} = \mathbf{F}^{\mathrm{L}} + \mathbf{F}^{\mathrm{T}} \tag{A.7}$$

and

$$\begin{aligned} \nabla \times \mathbf{F}^{\mathrm{L}} &= 0, \qquad \nabla \cdot \mathbf{F}^{\mathrm{L}} = \nabla \cdot \mathbf{F} \\ \nabla \cdot \mathbf{F}^{\mathrm{T}} &= 0, \qquad \nabla \times \mathbf{F}^{\mathrm{T}} = \nabla \times \mathbf{F} \end{aligned} \tag{A.8}$$

We call $\mathbf{F}^{\mathrm{L}}$ and $\mathbf{F}^{\mathrm{T}}$ the longitudinal and transverse parts of the vector field $\mathbf{F}$, respectively. The reason for the names longitudinal and transverse can be seen from some simple cases. For example, if we expand a transverse field in plane waves, the vector $\mathbf{F}^{\mathrm{T}}$ is perpendicular to the direction of wave propagation. This follows directly from Eqs. (A.8).

REFERENCE

1. P. M. Morse, and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill, New York, 1953.

Appendix B

THE LOCAL FIELD

The concept of a local field as first introduced by Lorentz was used to describe the force acting on an electron. It was defined to be the microscopic electric field acting on a particular electron due to all other charged particles in the system plus all applied fields. Thus, for a localized electron, the local field is the total microscopic field at the position of the electron minus the singularity caused by the electron itself. The term local field when applied to nonlocalized electrons is a misnomer. We shall use it anyway. What is meant is that the original concept of a local field can be generalized and defined in such a way as to give the correct polarization for an electron cloud and which, in the limit of point charges, reduces to the classical Lorentz local field.

We shall discuss the local field for two cases. The first is for an insulator consisting of noninteracting atoms and for which it is sufficiently accurate to assume that the electrons are all localized point charges. The second is for the case in which the electrons are spread throughout the crystal in an electron cloud.

B.1 Insulators

For gases at low pressures, where the interactions between molecules can be neglected, it is to be expected that the locally acting field will be identical to the externally applied field. At high pressures, however, and

particularly in the condensed phase of the solid state, the field acting on a molecule may be considerably modified by the polarization of the surroundings. To take this effect into account, we consider the following simple model of an insulator. We assume an insulator to consist of polarizable molecules or atoms arranged in a cubic array and idealized for mathematical purposes as point dipoles. Figure B.1 illustrates the effect of an external field on the dipoles within the insulator. The external field is produced in this example by placing the insulator between the plates of a charged capacitor. The average macroscopic field within the insulator is **E**. It is produced by those charges on the capacitor plates which are uncompensated by opposite charges at the ends of the dipole chains in the insulator.

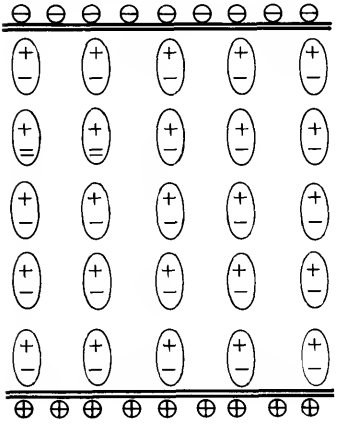

Fig. B.1 The atoms in an insulator placed in an applied field are polarized. Here, the applied field is produced by a charged capacitor. The macroscopic field **E** within the insulator is less than the total external field $\mathbf{E}^{\text{ext}}$. The field $\mathbf{E}^{\text{ext}}$ is determined by the total surface charge density on the capacitor plates; **E** is less than $\mathbf{E}^{\text{ext}}$ because the free ends of the dipole chains in the insulator cancel a fraction of the total surface charge on the capacitor plates. The reduced field **E** is given by $\mathbf{E} = \mathbf{E}^{\text{ext}}/\varepsilon_1$.

Figure B.2 shows a reference atom surrounded by an imaginary sphere of such extent that beyond it, the insulator can be treated as a continuum insofar as its interactions with the reference atom are concerned. If the atoms within this sphere were removed with no change in the rest of the system, the field acting on the reference atom would stem from two sources: The macroscopic field **E** and the field $\mathbf{E}_2$ produced by the free ends of the dipole chains that line the walls of the cavity. Of course, there are atoms within the sphere, and these cannot be neglected, but they are so near the reference atom that they must be considered individually on a microscopic basis. The individual atoms within the cavity contribute a field $\mathbf{E}_3$ to the local field acting on the reference atom. Thus

$$\mathbf{E}_{\text{loc}} = \mathbf{E} + \mathbf{E}_2 + \mathbf{E}_3 \tag{B.1}$$

where $\mathbf{E}_{\text{loc}}$ is the field produced at an atomic site by all other atoms acting on the reference atom.

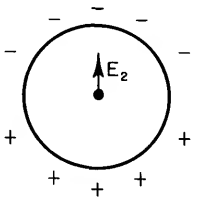

Fig. B.2 The Lorentz field $\mathbf{E}_2$ arises from the free ends of the dipole chains that terminate on the walls of the fictitious spherical cavity.

The field $\mathbf{E}_2$ produced by the polarization charges on the surface of the fictitious cavity was first calculated by Lorentz. The Lorentz field is calculated by assuming all dipoles to be aligned parallel to the applied field. Then, since the polarization is the dipole moment per unit volume and is readily shown to be equal to the surface density of induced charge, the field $\mathbf{E}_2$ can be calculated using Coulomb's law. That the polarization can be expressed as surface charge density can be seen most simply by imagining the insulator in Fig. B.1 as one large dipole. Let Q be the induced charge at each surface of the polarized insulating slab of area A and thickness d. The dipole moment for the slab is then Qd and the dipole moment per unit volume is $Qd/Ad = Q/A$. Thus the polarization is the surface charge density, i.e., $P = Q/A$.

We are now ready to calculate the Lorentz field $\mathbf{E}_2$ in the fictitious cavity. Referring to Fig. B.3, which illustrates a cross section of the Lorentz cavity, we see that since P is the magnitude of the charge density per unit area normal to the external field, the charge density per unit area of cavity surface is $-P \cos \theta$. Thus the charge dQ on an annular ring as shown in Fig. B.3 is

$$dQ = -2\pi r^2 P \cos \theta \sin \theta \, d\theta \tag{B.2}$$

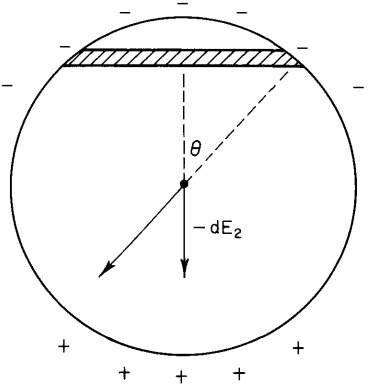

Fig. B.3 Lorentz field arising from an annular ring of charge on the Lorentz spherical cavity.

From Coulomb's law, this gives rise to a force per unit charge at the center of the cavity given by

$$dE_2 = -dQ\,(\cos\theta)/r^2$$

$$= 2\pi P \cos^2\theta \sin\theta\, d\theta \tag{B.3}$$

Thus the total Lorentz field is

$$\mathbf{E}_2 = \int d\mathbf{E}_2 = \int_0^\pi 2\pi \mathbf{P} \cos^2\theta \sin\theta\, d\theta$$

$$= \tfrac{4}{3}\pi \mathbf{P} \tag{B.4}$$

The field $\mathbf{E}_3$ produced by the dipoles within the cavity is the only term in Eq. (B.1) which depends on crystal structure. However, as may readily be seen from a simple sketch, if the crystal structure is cubic and all atoms can be idealized as point dipoles aligned parallel to the applied field, then $\mathbf{E}_3 = 0$. Thus, for simple isotropic solids,

$$\mathbf{E}_{\text{loc}} = \mathbf{E} + \tfrac{4}{3}\pi \mathbf{P} \tag{B.5}$$

Ideally, $\mathbf{E}_{\text{loc}}$ as given by Eq. (B.5) is the local field which should be used to describe the force on a bound electron in an insulator. In fact, it is not really correct. It is wrong to assume that the atoms within the Lorentz cavity are point dipoles. There may be considerable overlap of adjacent atoms. As a result, the field which is effective in polarizing an atom is between $\mathbf{E}$ and $\mathbf{E} + \frac{4}{3}\pi\mathbf{P}$, tending to the simple $\mathbf{E}$.

B.2 Nonlocalized Electrons

We saw in the discussion of the Lorentz model in Section 3.1 that the polarization is related to the microscopic local electric field $\mathbf{E}_{\text{loc}}$ through the equation

$$\mathbf{P} = N\alpha\langle \mathbf{E}_{\text{loc}} \rangle \tag{B.6}$$

In the Lorentz model, $\mathbf{E}_{\text{loc}}$ is averaged over discrete local sites. To generalize this model for nonlocalized electrons, the polarization can be written

$$\mathbf{P}(\mathbf{r}) = N \int d\mathbf{r}'\, \alpha(\mathbf{r}, \mathbf{r}')\mathbf{e}(\mathbf{r}') \tag{B.7}$$

where $\alpha(\mathbf{r}, \mathbf{r}')$ is a microscopic polarizability which provides a weighting

factor to account for variations in polarizability throughout the unit cell and $\mathbf{e}(\mathbf{r}')$ is the total microscopic electric field. [Recall that in Eq. (B.6), $\mathbf{E}_{loc}$ is the total microscopic field acting on an electron minus the field due to the electron itself. Here, the total microscopic field $\mathbf{e}(\mathbf{r}')$ itself can be used since, for electrons spread throughout the crystal, one electron makes a negligible contribution to the total field at a point.] The generalized local field $\bar{\mathbf{E}}_{loc}(\mathbf{R})$ can then be defined in terms of a weighted average over the unit cell such that the correct polarization is obtained. Here, $\mathbf{R}$ is the macroscopic position variable indicating the position of the unit cell. The bar in $\bar{\mathbf{E}}_{loc}(\mathbf{R})$ indicates that the local field is now a macroscopic field, but it is not generally equal to the average macroscopic field $\mathbf{E}$. The latter is an unweighted average of the microscopic field $\mathbf{e}(\mathbf{r})$. For a free-electron gas, the spatial charge distribution is uniform and all regions are equally polarizable. Then, $\bar{\mathbf{E}}_{loc}$ and $\mathbf{E}$ are equal. However, for electrons described by Bloch functions, there is some nonuniformity of the electron cloud and $\bar{\mathbf{E}}_{loc} \neq \mathbf{E}$.

FURTHER READING

C. Kittel, "Introduction to Solid State Physics," Wiley, New York, 1971.
N. Wiser, *Phys. Rev.* **129**, 62 (1963).
S. L. Adler, *Phys. Rev.* **126**, 413 (1962).
H. Ehrenreich, Electromagnetic Transport in Solids, *in* "The Optical Properties of Solids" (J. Tauc, ed.). Academic Press, New York, 1966.

Appendix C

REFLECTION AT NORMAL INCIDENCE

The reflectance for light impinging at normal incidence onto an ideal solid surface can be derived rather simply from a consideration of the boundary conditions for **E** and **H** at the interface. From a consideration of Fig. C.1, we can write

$$E_i + E_r = E_t \tag{C.1}$$

where the subscripts i, r, and t represent, respectively, the incident, reflected, and transmitted waves at the interface. Similar equations hold for H, but with a change in sign for H_r. That is, having adopted the convention implicit in Eq. (C.1), we must remember that **H** is perpendicular to **E** in the sense that $\mathbf{E} \times \mathbf{H}$ is in the direction of the wave propagation. Thus if E_i and E_r are in the same direction at the interface, H_i and H_r must be in opposite directions, so that

$$H_i - H_r = H_t \tag{C.2}$$

In the vacuum, $E = H$, whereas in the medium,

$$H = (\hat{n}/\sqrt{\mu})E \tag{C.3}$$

as can be shown by substituting plane-wave expressions of the form $\exp i(\hat{\mathbf{q}} \cdot \mathbf{r} - \omega t) = \exp i[(\omega/c)\hat{\mathbf{n}} \cdot \mathbf{r} - \omega t]$ into Maxwell's equations (2.75) and (2.77) for **E** and **H**. Thus, Eqs. (C.1) and (C.2) become

$$E_i + E_r = E_t \tag{C.4}$$

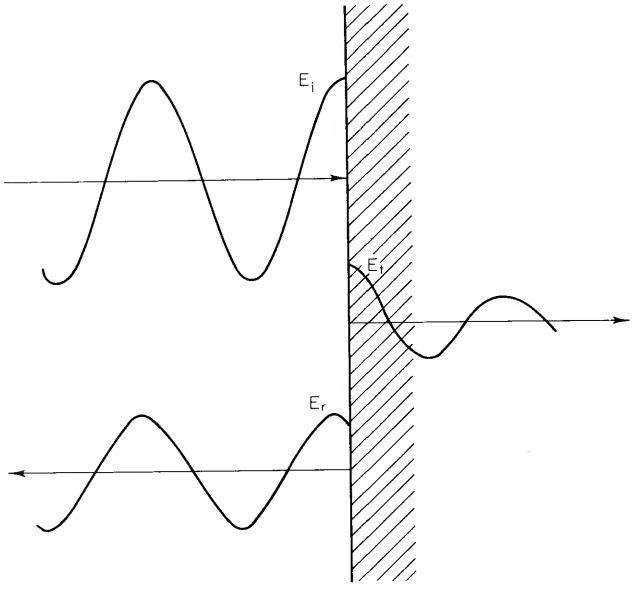

Fig. C.1 Illustration of incident, reflected, and transmitted waves at a vacuum–crystal interface.

$$E_i - E_r = \hat{n}E_t \tag{C.5}$$

where the permeability μ has been taken as unity. Equations (C.4) and (C.5) are easily solved to yield a reflectance wave amplitude

$$\hat{r} = E_r/E_i = (1 - \hat{n})/(1 + \hat{n}) \tag{C.6}$$

The reflectivity is then given by

$$R = \hat{r}^*\hat{r} = |(1 - \hat{n})/(1 + \hat{n})|^2$$

$$= [(1 - n)^2 + k^2]/[(1 + n)^2 + k^2]$$

The reflectivity can also be put in a form commonly used by radio engineers and when dealing with the anomalous skin effect. Recalling from Chapter 4 that the surface impedance is

$$Z = (4\pi/c)\mathbf{E}/\mathbf{H} \tag{C.8}$$

we obtain for nonmagnetic materials, with the help of Eqs. (C.3) and (C.7),

$$R = \left| \frac{Z - (4\pi/c)}{Z + (4\pi/c)} \right|^2 \tag{C.9}$$

Appendix D

THE *f*-SUM RULE FOR A CRYSTAL

The *f*-sum rule for a periodic crystal can be derived by direct comparison of a Taylor's series expansion for the energy of an electron with the corresponding terms in a second order perturbation analysis.

A Taylor's series expansion of the energy in the neighborhood of $\mathbf{k}$ gives

$$\mathscr{E}_n(\mathbf{k}+\mathbf{q}) = \mathscr{E}_n(\mathbf{k}) + \frac{\partial \mathscr{E}_n(\mathbf{k})}{\partial k_\alpha} q_\alpha + \frac{1}{2}\frac{\partial^2 \mathscr{E}_n(\mathbf{k})}{\partial k_\alpha \, \partial k_\beta} q_\alpha q_\beta + \cdots \tag{D.1}$$

where $[\partial \mathscr{E}_n(\mathbf{k})/\partial k_\alpha]\, q_\alpha$ means the partial derivative evaluated at $\mathbf{k}$ and the subscripts α and β are directional components.

Second-order perturbation theory as described in most textbooks on quantum mechanics yields

$$\mathscr{E}_n(\mathbf{k}+\mathbf{q}) = \mathscr{E}_n(\mathbf{k}) + (n\mathbf{k}'|H'|n\mathbf{k}) + \sum_{n'}{}' \frac{|(n'\mathbf{k}|H'|n\mathbf{k})|^2}{\mathscr{E}_n(\mathbf{k}) - \mathscr{E}_{n'}(\mathbf{k})} \tag{D.2}$$

where $|n\mathbf{k})$ denotes $u_{n\mathbf{k}}(\mathbf{r})$, the periodic part of the Bloch function which is a solution to the Schrödinger equation in the absence of a perturbation, and H' is the perturbation. The *f*-sum rule follows from direct comparison of the third term on the right-hand side of Eq. (D.1) with the corresponding term in Eq. (D.2).

Equating these terms yields

$$\frac{1}{2}\frac{\partial^2 \mathscr{E}_n(\mathbf{k})}{\partial k_\alpha \, \partial k_\beta} q_\alpha q_\beta = \sum_{n'} \frac{|(n'\mathbf{k}|H'|n\mathbf{k})|^2}{\mathscr{E}_n(\mathbf{k}) - \mathscr{E}_{n'}(\mathbf{k})} \tag{D.3}$$

We now need to evaluate the right-hand side of Eq. (D.3).

The perturbation term in Eq. (D.3) is given by Eq. (H.7) of Appendix H as

$$H' = H_{\mathbf{k}+\mathbf{q}} - H_{\mathbf{k}} = \frac{\hbar \mathbf{q} \cdot \mathbf{p}}{m} + \frac{\hbar^2 \mathbf{q} \cdot \mathbf{k}}{m} + \frac{\hbar^2 q^2}{2m} \tag{D.4}$$

Let us evaluate some of the matrix elements needed for Eq. (D.3). Taking the first two perturbation terms of Eq. (D.4), we obtain

$$\left(n\mathbf{k}' \left| \frac{\hbar \mathbf{q} \cdot \mathbf{p}}{m} + \frac{\hbar^2 \mathbf{q} \cdot \mathbf{k}}{m} \right| n\mathbf{k} \right) = \frac{\hbar \mathbf{q}}{m} \cdot \langle n'\mathbf{k} | \mathbf{p} | n\mathbf{k} \rangle \tag{D.5}$$

where $|n\mathbf{k}\rangle$ represents the full Bloch function $\psi_{n\mathbf{k}}(\mathbf{r}) = (\exp i\mathbf{k} \cdot \mathbf{r}) u_{n\mathbf{k}}(\mathbf{r})$. That the expression on the RHS of (D.5) is correct is most easily seen by substituting the full expression for a Bloch function on the RHS of (D.5) and showing that one obtains the LHS of (D.5), which is in terms of the periodic part of the Bloch function.

The last term in Eq. (D.4) is not an operator and does not require the use of perturbation theory, but it must be included in the second-order energy shift. It is second order because of the second-order dependence on the shift $\mathbf{q}$ in going from $\mathbf{k}$ to $\mathbf{k} + \mathbf{q}$. Thus the second-order shift in energy, corresponding to the third term on the RHS of Eqs. (D.1) and (D.2), is

$$\begin{aligned} \Delta\mathscr{E}\,(\text{second order}) &= \frac{1}{2} \frac{\partial^2 \mathscr{E}_n(\mathbf{k})}{\partial k_\alpha \, \partial k_\beta} q_\alpha q_\beta \\ &= \frac{\hbar^2 q^2}{2m} + \frac{\hbar^2}{m^2} \sum_{n'}{}' \frac{\langle n'\mathbf{k} | p_\alpha | n\mathbf{k} \rangle \langle n\mathbf{k} | p_\beta | n'\mathbf{k} \rangle}{\mathscr{E}_n(\mathbf{k}) - \mathscr{E}_{n'}(\mathbf{k})} q_\alpha q_\beta \end{aligned} \tag{D.6}$$

Recalling that $(1/\hbar^2)\, \partial^2 \mathscr{E} / \partial k^2 = 1/m^*$, we see that we can obtain from Eq. (D.6) an equation for the elements of the effective mass tensor. Thus,

$$\frac{m}{m^*_{\alpha\beta}} = \delta_{\alpha\beta} + \frac{2}{m} \sum_{n'}{}' \frac{\langle n'\mathbf{k} | p_\alpha | n\mathbf{k} \rangle \langle n\mathbf{k} | p_\beta | n'\mathbf{k} \rangle}{\mathscr{E}_n(\mathbf{k}) - \mathscr{E}_{n'}(\mathbf{k})} \tag{D.7}$$

Comparison of the summation in Eq. (D.7) with the definition of oscillator strength given in Eq. (3.82) shows that Eq. (D.7) is equivalent to the f-sum given in Eq. (3.88) for an isotropic crystal. We can now write it in the more general form

$$m/m^*_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{n'}{}' f^{\alpha\beta}_{n'n} \tag{D.8}$$

where $f^{\alpha\beta}_{n'n}$ is the oscillator strength.

Appendix E

INTERACTION OF RADIATION WITH MATTER

The Hamiltonian for a system in the presence of an external vector field $\mathbf{A}^{\text{ext}}$ and scalar field ϕ^{ext} can be written

$$H = \frac{1}{2m}\sum_{i=1}^{N}\left[\mathbf{p}_i - \frac{e}{c}\mathbf{A}^{\text{ext}}(\mathbf{r}_i, t)\right]^2 + \sum_{i=1}^{N} V_i(\mathbf{r}_i)$$

$$+ \tfrac{1}{2}\sum_{i,j}{}' \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + I + \sum_{i=1}^{N} e\phi^{\text{ext}}(\mathbf{r}_i, t) \tag{E.1}$$

where the indices i and j are summed over all N electrons (per unit volume), $V_i(\mathbf{r}_i)$ is the potential energy of the ith electron in the field of all ion cores, and the Coulomb interaction of an electron with itself is excluded as indicated by a prime on the third summation. Here, I represents the interaction of the ion cores with each other.

The total Hamiltonian can be rewritten as

$$H = \tilde{H}_0 + H^{\text{ext}} \tag{E.2}$$

where

$$\tilde{H}_0 = \frac{1}{2m}\sum_{i=1}^{N} \mathbf{p}_i{}^2 + \sum_{i=1}^{N} V_i(\mathbf{r}_i) + \tfrac{1}{2}\sum_{i,j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + I \tag{E.3}$$

and

$$H^{\text{ext}} = -\frac{e}{2mc}\sum_{i=1}^{N}\left[\mathbf{p}_i \cdot \mathbf{A}^{\text{ext}}(\mathbf{r}_i, t) + \mathbf{A}^{\text{ext}}(\mathbf{r}_i, t)\cdot \mathbf{p}_i\right] + \sum_{i=1}^{N} e\phi^{\text{ext}}(\mathbf{r}_i, t) \tag{E.4}$$

The term $e^2A^2_{\text{ext}}/(2mc)^2$ has been neglected in Eq. (E.4) because we are interested in the linear response of the system to weak fields.

Equation (E.1) includes the full Coulomb interaction between electrons and thus includes the presence of polarization fields induced by the external fields. One then computes the response of the system treating $\mathbf{A}^{\text{ext}}$ and ϕ^{ext} as perturbations inducing transitions in the system. In this case, $\tilde{H}_0$ does not correspond exactly to the Hamiltonian for the unperturbed system since it includes the polarization effects induced by the external fields. In the solution of the Schrödinger equation

$$-i\hbar\,\partial\Psi/\partial t = H\Psi \tag{E.5}$$

the appropriate functions in which to expand Ψ are thus not the eigenfunctions of the unperturbed system but the eigenfunctions for an electron acted on by the self-consistent potential arising from the presence of an external potential.

An alternative method is to treat the polarization field (the induced field) as an additional "external" field. One then computes the response of the system to the effective fields, i.e., the total fields

$$\mathbf{A} = \mathbf{A}^{\text{ext}} + \mathbf{A}^{\text{ind}} \tag{E.6}$$

$$\phi = \phi^{\text{ext}} + \phi^{\text{ind}} \tag{E.7}$$

This is the procedure that is usually more convenient for nonrelativistic systems. It also is perhaps more satisfying in that it recognizes explicitly that the electrons in a solid respond to the total field, including the induced field due to other electrons, and not just to the external field. This is the approach used throughout this book, except for Chapter 8.

If we treat the electrons as responding to the total electric field, we can write the total Hamiltonian of the system, the radiation field, and the interaction energy as

$$H = (1/2m)\sum_i\left[\mathbf{p}_i - \frac{e}{c}\mathbf{A}(\mathbf{r}_i, t)\right]^2 + \sum_i V_i(\mathbf{r}_i) + H^0_{\text{Coul}} + I + \sum_i e\phi_i(\mathbf{r}_i, t) \tag{E.8}$$

We can write this as

$$H = H_0 + H' \tag{E.9}$$

where

$$H_0 = (1/2m)\sum_i \mathbf{p}_i^2 + \sum_i V_i(\mathbf{r}_i) + H^0{}_{\text{Coul}} \tag{E.10}$$

and $\Sigma_i V_i(\mathbf{r}_i) + H^0{}_{\text{Coul}}$ is the self-consistent field in the pure crystal, and

$$H' = (-e/2mc)\sum_i [\mathbf{p}_i \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}_i] + \sum_i e\phi_i \tag{E.11}$$

is treated as the perturbation term. Again, we have dropped the term in $\mathbf{A}^2$.

In the formulation of Eqs. (E.8)–(E.11), the Coulomb interaction $H^0{}_{\text{Coul}}$ contains only the residual interactions between electrons in the absence of an external field. Thus the wave functions which are solutions to

$$H_0\psi = \mathscr{E}\psi \tag{E.12}$$

are exact solutions to the unperturbed system. These are the eigenfunctions calculated by band theorists who use a self-consistent potential. Now, of course, the properties of the medium are incorporated into the perturbation term itself. Thus we will be calculating the response of the system to a perturbation which already includes the properties of the system. Formally, we do this by relating the total fields to the energy density via the dielectric function. We will return to this point when we normalize the vector potential $\mathbf{A}$.

If we consider the interaction of electrons with an electromagnetic field, we can take the scalar field as zero if we work in the Coulomb gauge. Then, Eq. (E.11) describing the interaction of the system with the radiation becomes

$$H' = -(e/2mc)\sum_i (\mathbf{p}_i \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}_i) \tag{E.13}$$

In general,

$$\mathbf{p} \cdot \mathbf{A} - \mathbf{A} \cdot \mathbf{p} = -i\hbar\, \nabla \cdot \mathbf{A} \tag{E.14}$$

so that $\mathbf{p} \cdot \mathbf{A} \neq \mathbf{A} \cdot \mathbf{p}$. However, since we have chosen to work in the Coulomb gauge, $\nabla \cdot \mathbf{A} = 0$, and hence $\mathbf{p} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{p}$. Thus, introducing the current operator

$$\mathbf{J}(\mathbf{r}) = (e/2m)\sum_i [\mathbf{p}_i\, \delta(\mathbf{r} - \mathbf{r}_i) + \delta(\mathbf{r} - \mathbf{r}_i)\mathbf{p}_i] \tag{E.15}$$

the interaction term becomes

$$H' = -(1/c)\int d\mathbf{r}\, \mathbf{J}(\mathbf{r}) \cdot \mathbf{A}(\mathbf{r}, t) \tag{E.16}$$

It is clear that Eq. (E.15) can be interpreted as the operator for the current. It is a sum of terms of the form of the velocity of the ith particle times the density operator $\delta(\mathbf{r} - \mathbf{r}_i)$ of the ith particle. The current operator is expressed in a symmetric form so as to fulfill the requirement that it be a Hermitian operator. However, Eq. (E.15) does not describe the true current in the presence of an electromagnetic field. The velocity of an electron in the presence of an electromagnetic field is given by

$$\mathbf{v} = (\mathbf{p}/m) - (e/mc)\,\mathbf{A}(\mathbf{r}, t) \tag{E.17}$$

not by $\mathbf{p}/m$. Thus the true current operator is given by

$$\mathbf{J}^{\text{true}}(\mathbf{r}) = \mathbf{J}(\mathbf{r}) - (e/mc)\mathbf{A}(\mathbf{r}, t)\rho(\mathbf{r}) \tag{E.18}$$

The term $\mathbf{J}(\mathbf{r})$ is known as the paramagnetic current. The term proportional to $\mathbf{A}(\mathbf{r}, t)$ is called the diamagnetic current. We shall ignore the diamagnetic current as we are interested only in linear response and we already have the vector potential $\mathbf{A}(\mathbf{r}, t)$ included in the interaction Hamiltonian of Eq. (E.16).

The vector potential $\mathbf{A}(\mathbf{r}, t)$ can be represented by the sum of plane waves

$$\mathbf{A}(\mathbf{r}, t) = \left(\frac{2\pi\hbar c^2}{\varepsilon_1 \omega}\right)^{1/2} \sum_{\mathbf{q},\eta} \{b_{\mathbf{q}\eta}\boldsymbol{\eta}[\exp i(\mathbf{q}\cdot\mathbf{r} - \omega t)] + b^{\dagger}_{\mathbf{q}\eta}\boldsymbol{\eta}^* \exp -i(\mathbf{q}\cdot\mathbf{r} - \omega t)\} \tag{E.19}$$

where $\boldsymbol{\eta}$ is a polarization index, and $b_{\mathbf{q}\eta}$ and $b^{\dagger}_{\mathbf{q}\eta}$ are photon annihilation and creation operators, respectively. (The reader unfamiliar with annihilation and creation operators can simply let these operators be unity.) The normalization factor is chosen to give an energy density in the medium corresponding to one photon per unit volume for each mode specified by $\mathbf{q}, \boldsymbol{\eta}$. That is, we take

$$\mathscr{U}_{\mathbf{q}\eta} = \varepsilon_1 \langle E^2 \rangle_{\mathbf{q}\eta}/4\pi = (\varepsilon_1/4\pi)(\omega^2/c^2)\langle A^2 \rangle_{\mathbf{q}\eta} = \hbar\omega \tag{E.20}$$

where the zero-point energy has been ignored.

When working with $\mathbf{A}^{\text{ext}}(\mathbf{r}, t)$, as in Chapter 8, rather than $A(\mathbf{r}, t)$, the appropriate results are obtained simply by setting $\varepsilon_1 = 1$ in Eqs. (E.19) and (E.20).

Appendix F

M_1 CRITICAL POINTS

We shall derive an expression for the density of states in the neighborhood of an M_1-type critical point. We shall consider an electron energy band, but the joint density of states for an optical energy band is of the same form.

If the energy in the neighborhood of a point $\mathbf{k}_c$ in $\mathbf{k}$-space can be expressed as a quadratic in $(\mathbf{k} - \mathbf{k}_c)$, then $\mathbf{k}_c$ is a critical point. Now, consider the constant-energy surface which, with a suitable choice of coordinate axes, can be expressed as

$$\mathscr{E}(\mathbf{k}) = \mathscr{E}(\mathbf{k}_c) + \frac{\hbar^2 k_1^2}{2m_1} + \frac{\hbar^2 k_2^2}{2m_2} + \frac{\hbar^2 k_3^2}{2m_3} \tag{F.1}$$

Take $\mathscr{E}(\mathbf{k}_c)$ as zero for convenience and let

$$\hbar^2/2m_1 = \beta_1, \qquad \hbar^2/2m_2 = \beta_2, \qquad \hbar^2/2m_3 = \beta_3 \tag{F.2}$$

Then, for an M_1-type critical point

$$\mathscr{E} = \beta_1 k_1^2 + \beta_2 k_2^2 - \beta_3 k_3^2 \tag{F.3}$$

For $\mathscr{E} > 0$, the constant-energy surfaces are hyperboloids of one sheet as illustrated in Fig. F.1. Such a hyperboloid resembles a saddle, and for this reason the critical point is known as a saddle point. It is usually designated by M_1 or S_1, where the subscript indicates the number of negative signs in equations such as Eq. (F.1).

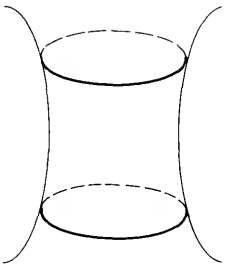

Fig. F.1 A hyperboloid of one sheet.

We cannot enclose all states within a constant-energy surface for a hyperboloid as we did for the ellipsoid in Section 5.3. Here, the surface is unbounded. We must introduce some kind of arbitrary cutoff in the integration. This is allowable since we are interested in only a small neighborhood of the critical point $\mathbf{k}_c$ where the quadratic expansion is valid. The main requirement is that the results not be sensitive to the choice of cutoff. We shall use the volume bounded by the surfaces of constant energy and the planes perpendicular to k_3 and passing through $k_3 = \pm k_0$.

These boundaries are illustrated in Fig. F.2, which represents a plane passing through the hyperboloid of Fig. F.1. The volume enclosed within these boundaries is given by

$$\Omega = 2 \int_0^{k_0} A \, dk_3 \tag{F.4}$$

where A is the area of the ellipse defined by

$$\frac{{k_1}^2}{(\mathscr{E}/\beta_1)[1 + (\beta_3/\mathscr{E}){k_3}^2]} + \frac{{k_2}^2}{(\mathscr{E}/\beta_2)[1 + (\beta_3/\mathscr{E}){k_3}^2]} = 1 \tag{F.5}$$

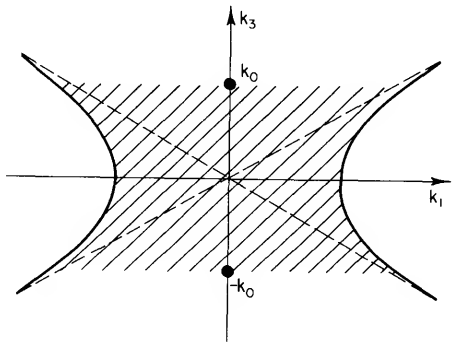

Fig. F.2 A plane through the hyperboloid of Fig. F.1 showing the region of integration.

The area is

$$A = \frac{\pi\mathscr{E}}{(\beta_1\beta_2)^{1/2}}\left(1 + \frac{\beta_3}{\mathscr{E}}k_3^{\,2}\right) \tag{F.6}$$

Carrying out the integration yields

$$\Omega = \frac{2\pi\mathscr{E}}{(\beta_1\beta_2)^{1/2}}\left(k_0 + \frac{\beta_3 k_0^{\,3}}{3\mathscr{E}}\right) \tag{F.7}$$

For $\mathscr{E} < 0$, the constant-energy surfaces are paraboloids. Referring to Fig. F.3, we see that

$$\Omega = 2\int_{(-\mathscr{E}/\beta_3)^{1/2}}^{k_0} \frac{\pi\mathscr{E}}{(\beta_1\beta_2)^{1/2}}\left(1 + \frac{\beta_3}{\mathscr{E}}k_3^{\,2}\right)dk_3$$

$$= \frac{2\pi\mathscr{E}}{(\beta_1\beta_2)^{1/2}}\left(k_0 + \frac{\beta_3 k_0^{\,3}}{3\mathscr{E}}\right) + \frac{4\pi}{3(\beta_1\beta_2\beta_3)^{1/2}}(-\mathscr{E})^{3/2} \tag{F.8}$$

We now need only multiply by the density of states in **k**-space per unit volume of crystal, $1/4\pi^3$ (including spin), and differentiate with respect to

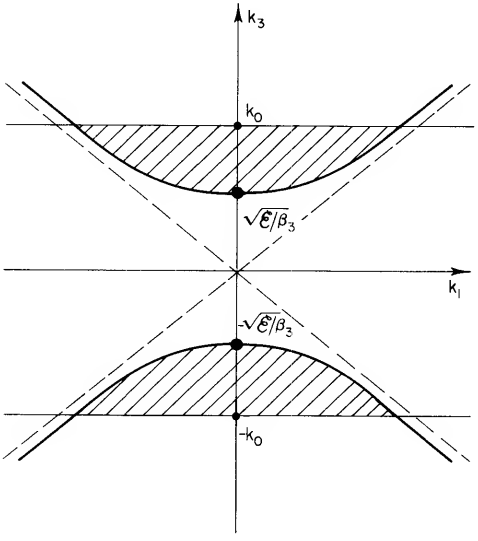

Fig. F.3 Section of paraboloids showing region of integration.

$\mathscr{E}$ to obtain the density of states near the saddle point. Using

$$\frac{d}{d\mathscr{E}} = \frac{\partial}{\partial \mathscr{E}} + \frac{\partial k_0}{\partial \mathscr{E}} \frac{\partial}{\partial k_0} \tag{F.9}$$

$$= \frac{\partial}{\partial \mathscr{E}} - \frac{1}{2\beta_3 k_0} \frac{\partial}{\partial k_0}$$

the density of states is found to be

$$\rho(\mathscr{E}) = \frac{1}{4\pi^3} \left[\frac{\pi k_0}{(\beta_1\beta_2)^{1/2}} - \frac{2\pi}{(\beta_1\beta_2\beta_3)^{1/2}} (-\mathscr{E})^{1/2} - \frac{\pi\mathscr{E}}{\beta_3(\beta_1\beta_2)^{1/2} k_0} \right], \quad \mathscr{E} < 0$$

$$= \frac{1}{4\pi^3} \left[\frac{\pi k_0}{(\beta_1\beta_2)^{1/2}} - \frac{\pi\mathscr{E}}{\beta_3(\beta_1\beta_2)^{1/2} k_0} \right], \quad \mathscr{E} > 0 \tag{F.10}$$

If k_0 is sufficiently large, it is seen that

$$\rho(\mathscr{E}) = c_1 - c_2(-\mathscr{E})^{1/2} - O(\mathscr{E}), \quad \mathscr{E} < 0$$

$$= c_1 - O(\mathscr{E}), \quad \mathscr{E} > 0 \tag{F.11}$$

This is illustrated in Fig. F.4.

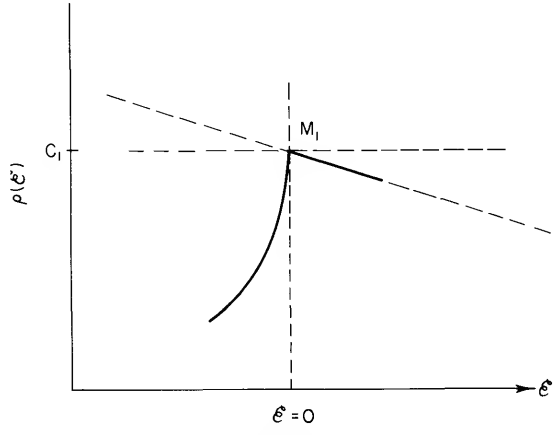

Fig. F.4 Contribution to the density of states from an M_1 critical point.

From Eq. (F.11) and Fig. F.4, we see that the density of states near a saddle point can be fairly well described as a sum of two terms, a smooth function $\rho(\varepsilon)$ and a term $\delta\rho(\varepsilon)$ which is different from zero on that side of the critical point where the number of sheets of constant energy has increased by one.

Appendix G

REFLECTANCE AND PHASE-SHIFT DISPERSION RELATIONS

In section G.1, it is shown that the complex reflectivity amplitude $\hat{r}(\omega)$ can be analytically continued into the complex frequency plane and that Eq. (6.46) for the phase shift can be derived from an appropriate contour integration.

Section G.2 consists of a brief discussion of the numerical integration of the phase-shift equation.

G.1 The Phase-Shift Dispersion Relation

We have seen in Section 6.2 that the derivation of the phase-shift dispersion relation requires that the complex reflectivity amplitude $\hat{r}(\omega)$ be analytically continued into the complex frequency plane. We now show that $\hat{r}(\omega)$, and hence $\ln \hat{r}(\omega)$, can be analytically continued into the upper half of the complex frequency plane, which we denote as domain D. A contour integration as indicated in Fig. G.1 then yields the desired dispersion relation.

The linear causal relation between $\mathbf{E}_r(\omega)$ and $\mathbf{E}_i(\omega)$ requires analyticity of $r(\omega)$ in D as shown by the following. We can write

$$\mathbf{E}_r(t) = \int_{-\infty}^{+\infty} \mathbf{E}_r(\omega) e^{-i\omega t} d\omega \tag{G.1}$$

$$\mathbf{E}_r(\omega) = \hat{r}(\omega)\mathbf{E}_i(\omega) \tag{G.2}$$

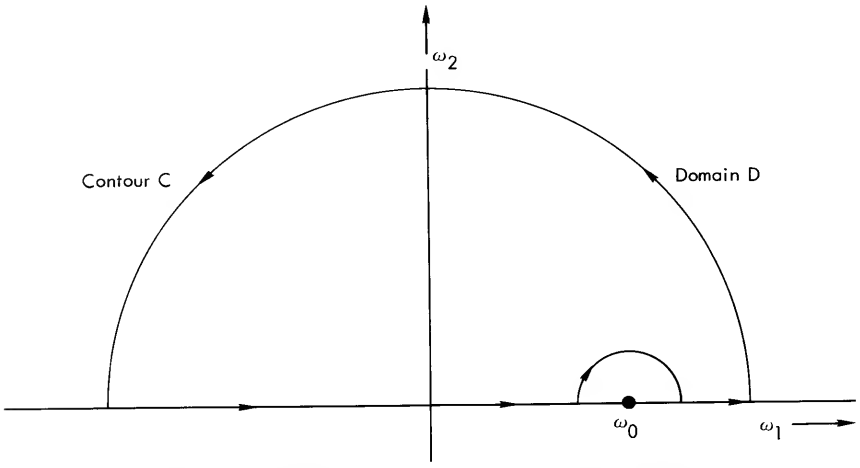

Fig. G.1 Integration contour in complex frequency plane.

and

$$\mathbf{E}_i(\omega) = (1/2\pi) \int_{-\infty}^{+\infty} \mathbf{E}_i(t') \exp(i\omega t')\, dt' \tag{G.3}$$

Combining these relations, we have

$$\mathbf{E}_r(t) = \int_{-\infty}^{+\infty} \hat{r}(\omega)\, (\exp\, -i\omega t) \left[(1/2\pi) \int_{-\infty}^{+\infty} \mathbf{E}_i(t')\, (\exp\, i\omega t')\, dt' \right] d\omega$$

$$= (1/2\pi) \int_{-\infty}^{+\infty} \mathbf{E}_i(t')\, dt' \int_{-\infty}^{+\infty} \hat{r}(\omega) \exp\, -\, i\omega(t\, -\, t')\, d\omega$$

$$= \int_{-\infty}^{+\infty} G(t\, -\, t')\mathbf{E}_i(t')\, dt' \tag{G.4}$$

where

$$G(t\, -\, t') \equiv (1/2\pi) \int_{-\infty}^{+\infty} \hat{r}(\omega) \exp\, -i\omega(t\, -\, t')\, d\omega \tag{G.5}$$

is the transform of $r(\omega)$. If this transform can be inverted,

$$\hat{r}(\omega) = \int_{-\infty}^{+\infty} G(T) e^{i\omega T}\, dT, \qquad T \equiv t\, -\, t' \tag{G.6}$$

Causality requires $G(T < 0) \equiv 0$, and integration over negative time

difference T can be omitted. From Eq. (G.6), it is seen that $\hat{r}(\hat{\omega})$ is analytic in D by virtue of the factor $\exp[-T\,\mathrm{Im}(\hat{\omega})]$, where T and $\mathrm{Im}(\hat{\omega})$ are positive. $G(T)$ is the ratio of reflected and incident fields and must be defined, so $r(\hat{\omega})$ and $d\hat{r}(\hat{\omega})/d\hat{\omega}$ are defined in D.

We further need to show $\hat{r}(\hat{\omega})$ has no zeros in D since these are branch points of $\ln[\hat{r}(\hat{\omega})]$. From Eq. (6.39), a zero requires $\hat{n}(\hat{\omega}) = 1$, $\hat{\varepsilon}(\hat{\omega}) = 1$, and $\varepsilon_2(\hat{\omega}) = 0$. Since absorption processes are physically inescapable, $\varepsilon_2(\hat{\omega}) \neq 0$, and $\hat{r}(\hat{\omega}) \neq 0$. Thus $\ln[\hat{r}(\hat{\omega})]$ has no branch points in D and the same is true of $\ln \rho(\hat{\omega})$.

We want to derive a dispersion relation for $\ln \hat{r}(\omega)$, i.e., we want an integral relationship between $\ln \rho(\omega)$ and $\theta(\omega)$ so we can eliminate one of the unknowns in Eq. (6.42). However, $|\ln \hat{r}(\hat{\omega})| \to \infty$ as $|\hat{\omega}| \to \infty$ and so a function other than $\ln \hat{r}(\hat{\omega})$ must be used for the integration over the contour shown in Fig. G.1. The function

$$f(\hat{\omega}) = \frac{(1 + \omega_0\hat{\omega}) \ln[\hat{r}(\hat{\omega})]}{(1 + \hat{\omega}^2)(\omega_0 - \hat{\omega})} \tag{G.7}$$

suffices. It can be integrated over the contour C. It is physically plausible that $\hat{r}(\hat{\omega}) < b(\hat{\omega})^{-s}$, $|\hat{\omega}| \to \infty$, where b and s are positive constants. For $|\hat{\omega}| \to \infty$, the absolute value of Eq. (G.7) is less than

$$\left| \frac{(1 + \omega_0\hat{\omega})(-s \ln \hat{\omega} + \ln b)}{(1 + \hat{\omega}^2)(\omega_0 - \hat{\omega})} \right| \to 0, \qquad |\hat{\omega}| \to \infty$$

Thus the contribution to the integral from the large semicircle may be neglected as we let its radius become infinite.

Integrating Eq. (G.7), we have

$$\int_c \frac{(1 + \omega_0\hat{\omega}) \ln[\hat{r}(\hat{\omega})]}{(1 + \hat{\omega}^2)(\omega_0 - \hat{\omega})}\, d\hat{\omega} = 2\pi i \sum_j \mathrm{Res}(\hat{\omega}_j). \tag{G.8}$$

This becomes

$$\mathscr{P} \int_{-\infty}^{+\infty} \frac{(1 + \omega_0\omega) \ln[\hat{r}(\hat{\omega})]}{(1 + \omega^2)(\omega_0 - \omega)}\, d\omega + i\pi \ln[\hat{r}(\omega_0)] = 2\pi i \left[\frac{1 + i\omega_0}{2i(\omega_0 - i)} \right] \ln[r(i)]$$

$$= i\pi \ln[r(i)] \tag{G.9}$$

since the only pole enclosed by the contour is $\hat{\omega} = i$, and the only pole on the contour is $\hat{\omega} = \omega_0$. Principal value is indicated by $\mathscr{P}$. From Eq. (G.6), we see $\hat{r}(i)$ is real, so $\ln[\hat{r}(i)]$ is real. Equating real parts of Eq. (G.9), we find

$$\theta(\omega_0) = \frac{1}{\pi} \mathscr{P} \int_{-\infty}^{+\infty} \frac{1 + \omega_0\omega}{1 + \omega^2} \ln[\rho(\omega)] \frac{d\omega}{\omega_0 - \omega} \tag{G.10}$$

We may rewrite this as an integral over positive frequencies by remembering that $\hat{r}(\omega)$ is the transform of the ratio of physical, thus real, quantities. Namely, the real fields are given by

$$\mathbf{E}_r(t) = \int_{-\infty}^{+\infty} \mathbf{E}_r(\omega) e^{-i\omega t}\, d\omega \tag{G.11}$$

"Real" implies

$$\mathbf{E}_r(t) = \mathbf{E}_r{}^*(t) \tag{G.12}$$

Therefore,

$$\mathbf{E}_r{}^*(\omega) = \mathbf{E}_r(-\omega) \tag{G.13}$$

and similarly

$$\mathbf{E}_i{}^*(\omega) = \mathbf{E}_i(-\omega) \tag{G.14}$$

Using Eq. (G.2), we see that

$$\hat{r}^*(\omega) = \mathbf{E}_r{}^*(\omega)/\mathbf{E}_i{}^*(\omega) = \mathbf{E}_r(-\omega)/\mathbf{E}_i(-\omega) \tag{G.15}$$

so

$$\hat{r}^*(\omega) = \hat{r}(-\omega) \tag{G.16}$$

and

$$\ln[\rho(\omega)] = \ln[\rho(-\omega)] \tag{G.17}$$

Considering integration of Eq. (G.10) over negative frequencies and using Eq. (G.17), we have

$$\frac{1}{\pi} \mathscr{P} \int_{-\infty}^{0} \frac{(1 + \omega_0\omega) \ln \rho(\omega)}{(1 + \omega^2)(\omega_0 - \omega)}\, d\omega = \frac{1}{\pi} \mathscr{P} \int_{\infty}^{0} \frac{(1 - \omega_0\omega) \ln \rho(-\omega)}{(1 + \omega^2)(\omega_0 + \omega)}\, d(-\omega)$$

$$= \frac{1}{\pi} \mathscr{P} \int_{0}^{\infty} \frac{(1 - \omega_0\omega) \ln \rho(\omega)}{(1 + \omega^2)(\omega_0 + \omega)}\, d\omega \tag{G.18}$$

Thus Eq. (G.10) becomes

$$\theta(\omega_0) = \frac{2\omega_0}{\pi} \mathscr{P} \int_{0}^{+\infty} \frac{\ln \rho(\omega)}{\omega_0{}^2 - \omega^2}\, d\omega \tag{G.19}$$

The singularity at $\omega = \omega_0$ is removed by subtracting

$$\frac{2\omega_0}{\pi} \mathscr{P} \int_{0}^{+\infty} \frac{\ln \rho(\omega_0)}{\omega_0{}^2 - \omega^2}\, d\omega = 0 \tag{G.20}$$

from Eq. (G.19) and then, using Eq. (6.42) to obtain the desired dispersion relation,

$$\theta(\omega_0) = \frac{\omega_0}{\pi} \int_0^\infty \frac{\ln[R(\omega)/R(\omega_0)]}{\omega_0^2 - \omega^2} d\omega \tag{G.21}$$

Equation (G.20) is proved through a similar contour integration. The lack of a singularity at $\omega = \omega_0$ is seen by applying L'Hospital's rule.

There are several notable features of Eq. (G.21). A fractional error in reflectance, that is, a measurement which yields the same percentage error in reflectance at all frequencies, does not affect the phase. Such errors still yield the correct reflectance ratio $R(\omega)/R(\omega_0)$. Also, no contribution to the phase results from regions with reflectance equal to $R(\omega_0)$.

G.2 Numerical Integration of the Phase-Shift Equation

The integrand of the phase-shift equation (G.21) is a smooth, well-behaved function and can be approximated between two neighboring frequencies by a linear interpolation. The integral can then be evaluated by Simpson's rule, summing trapezoidal areas defined by pairs of points, where each point equals the value of the integrand for a particular frequency. Although it is not obvious from Eq. (G.21), a significant contribution to the integral comes from the reflectance at frequencies well removed from ω when $dR/d\omega$ is large. Thus the extrapolation of reflectance to regions outside the range of experimental measurement must be done with great care.

Evaluation of the phase-shift equation requires separation of the integration into spectral regions. This is depicted in Fig. G.2 for an idealized reflectance spectrum, a spectrum which, for purposes of discussion, we imagine we have obtained partly with the use of three different monochromators. Thus, for example, the regions of "our" experimental data (regions b, c, d) have been taken respectively with the use of an infrared, a visible, and a vacuum-ultraviolet monochromator. The range of experimental data is extended to higher energy by using the measurements (region e) of other investigators. Extrapolations are used for the remaining regions, (a) (f), and (g).

For metals (as assumed implicitly here) the reflectance approaches unity at zero frequency. For nonmetals, the reflectance can usually be extrapolated with sufficient accuracy from the available experimental data. The phase contribution from regions (a)–(e) can then be calculated numerically by applying Simpson's rule.

The high-energy regions (e), (f), and (g) are the most difficult to treat accurately and are the largest source of error in the phase calculation. Most experiments do not include the full range of interband transitions

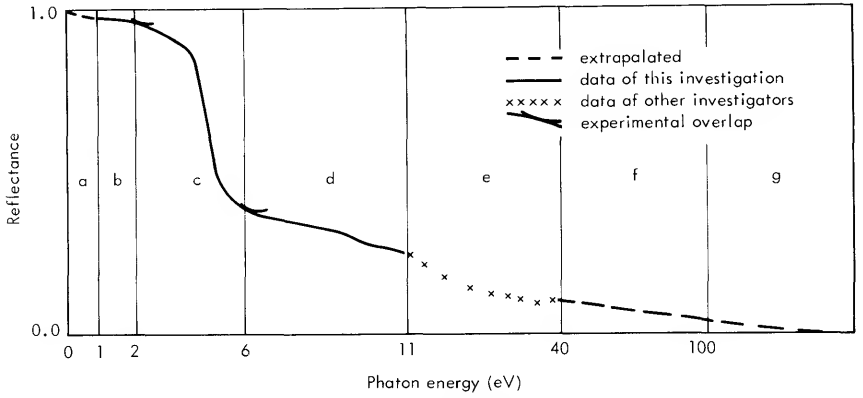

Fig. G.2 Idealized composite reflectance curve. Spectral regions: (a) low-energy, linear extrapolation of data from region (b); (b) near-infrared; (c) visible and ultraviolet; (d) vacuum-ultraviolet; (e) data from other investigators; (f) empirical power-law extrapolation; (g) free-electron-gas asymptotic reflectance extrapolation.

and thus the reflectance at even fairly high frequency cannot be assumed to vary as $(\omega/\omega_p)^{-4}$, the asymptotic form, but is characterized by structure. In our hypothetical case, we assume that some data (region e) are available from other investigators.

In region (f), we assume that there is still some contribution from interband transitions and approximate the reflectance by

$$R(\omega) = R_{ef}(\omega_{ef}/\omega)^s, \qquad \omega > \omega_{ef}; \quad s > 0, \tag{G.22}$$

where ω_{ef} is the frequency of the last measured reflectance value R_{ef} at the boundary between regions (e) and (f), and s is a parameter. In region (g), we assume that the frequency is sufficiently high that the reflectance is accurately described by the free-electron asymptotic limit

$$R(\omega) = R_{fg}(\omega_{fg}/\omega)^4, \qquad \omega > \omega_{fg} \tag{G.23}$$

Substituting Eqs. (G.22) and (G.23) into Eq. (G.21), and using series expansions, the integration can be performed, yielding

$$\theta_h(\omega_0) = \frac{1}{2\pi} \ln\left[\frac{R_{fg}}{R(\omega_0)}\right] \ln\left[\frac{1 - (\omega_0/\omega_{ef})}{1 + (\omega_0/\omega_{ef})}\right]$$

$$+ \frac{1}{\pi} \sum_{n=0}^{\infty} \left[s \frac{(\omega_0/\omega_{ef})^{2n+1}}{(2n+1)^2} + (4 - s) \frac{(\omega_0/\omega_{fg})^{2n+1}}{(2n+1)^2} \right] \tag{G.24}$$

where $\theta_h(\omega_0)$ is the contribution to the phase at ω_0 arising from the two high-energy regions (f) and (g). The parameter s and the frequency ω_{fg} are

chosen to give a best fit to the calculation of the optical constants. What determines a best fit depends on the particular situation. One possibility is that the parameters are chosen such that the calculated optical constants are in agreement with those obtained at a few particular frequencies by an independent experiment such as reflectance as a function of angle. Another, useful for intrinsic semiconductors and insulators, is to require that ε_2 be zero at frequencies less than the band gap. There are, of course, numerous variations of extrapolation schemes and methods to obtain a best fit. The experimentalist must work out the details of the most suitable approach for his particular problem.

FURTHER READING

M. Cardona, Optical Constants of Insulators: Dispersion Relations, *in* "Optical Properties of Solids" (S. Nudelman and S. S. Mitra, eds.). Plenum Press, New York, 1969.

F. Stern, Elementary Theory of the Optical Properties of Solids, *Solid State Phys.* **15** (1963).

D. L. Greenaway and G. Harbeke, "Optical Properties and Band Structure of Semiconductors," Pergamon Press, London, 1968.

D. M. Roessler, *Brit. J. Appl. Phys.* **16**, 1359 (1965).

Appendix H

$\mathbf{k} \cdot \mathbf{p}$ PERTURBATION THEORY

We can derive Eq. (7.54) with the help of a form of $\mathbf{k} \cdot \mathbf{p}$ perturbation theory. We begin with the Schrödinger equation

$$[(\mathbf{p}^2/2m) + V]\psi_{\mathbf{k}l} = \mathscr{E}_{\mathbf{k}l}\psi_{\mathbf{k}l} \tag{H.1}$$

Substituting the Bloch function

$$\psi_{\mathbf{k}l} = (\exp i\mathbf{k} \cdot \mathbf{r})u_{\mathbf{k}l}(\mathbf{r}) \tag{H.2}$$

and operating with $\mathbf{p}^2$ gives

$$\left[\frac{\mathbf{p}^2}{2m} + \frac{\hbar}{m}\mathbf{k} \cdot \mathbf{p} + \frac{\hbar^2 k^2}{2m} + V(\mathbf{r})\right]u_{\mathbf{k}l} = \mathscr{E}_{\mathbf{k}l}u_{\mathbf{k}l} \tag{H.3}$$

Defining an operator

$$H_{\mathbf{k}} \equiv \left[\frac{\mathbf{p}^2}{2m} + \frac{\hbar}{m}\mathbf{k} \cdot \mathbf{p} + \frac{\hbar^2 k^2}{2m} + V(\mathbf{r})\right] \tag{H.4}$$

Eq. (H.3) becomes

$$H_{\mathbf{k}}u_{\mathbf{k}l} = \mathscr{E}_{\mathbf{k}l}u_{\mathbf{k}l} \tag{H.5}$$

What we need to know is the function $u_{\mathbf{k}+\mathbf{q},l'}$. Then, from Eq. (7.44), we can determine Eq. (7.54). We can find $u_{\mathbf{k}+\mathbf{q},l'}$ in terms of $u_{\mathbf{k}l'}$ by perturbation theory if $\mathbf{q}$ is small. By starting with the function $\psi_{\mathbf{k}+\mathbf{q},l'}$, and following the

same steps as those leading to Eq. (H.5), we find

$$H_{\mathbf{k}+\mathbf{q}}u_{\mathbf{k}+\mathbf{q},l'} = \mathscr{E}_{\mathbf{k}+\mathbf{q},l'}u_{\mathbf{k}+\mathbf{q},l'} \tag{H.6}$$

$$H_{\mathbf{k}+\mathbf{q}} = H_{\mathbf{k}} + \frac{\hbar\mathbf{q}\cdot(\mathbf{p}+\hbar\mathbf{k})}{m} + \frac{\hbar^2 q^2}{2m} \tag{H.7}$$

The difference $H_{\mathbf{k}+\mathbf{q}} - H_{\mathbf{k}}$ as obtained from Eq. (H.7) can be regarded as a perturbation. The state $u_{\mathbf{k}+\mathbf{q},l'}$ can then be found by doing perturbation theory around the state $u_{\mathbf{k}l'}$, belonging to $H_{\mathbf{k}}$. The periodic part of the Bloch function to first order in $\mathbf{q}$ is then found by time-independent perturbation theory to be

$$\begin{aligned} u_{\mathbf{k}+\mathbf{q},l'} &\equiv |\mathbf{k}+\mathbf{q}, l') \\ &= |\mathbf{k}l') + \sum_{l''}{}' |\mathbf{k}l'') \frac{(\mathbf{k}l''|H_{\mathbf{k}+\mathbf{q}} - H_{\mathbf{k}}|\mathbf{k}l')}{\mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l''}} \\ &= |\mathbf{k}l') + \frac{\hbar}{m}\sum_{l''}{}' |\mathbf{k}l'') \frac{(\mathbf{k}l''|\mathbf{q}\cdot\mathbf{p}|\mathbf{k}l')}{\mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l''}} \end{aligned} \tag{H.8}$$

if the state $|\mathbf{k}l')$ is nondegenerate. Note that the terms in $\mathbf{q}\cdot\mathbf{k}$ and q^2 appearing in Eq. (H.7) do not show up in the last line of Eq. (H.8). These terms are not operators and so can be taken outside the integral. What remains is then an integral over two orthogonal functions. Thus the integral for these terms is zero.

The name $\mathbf{k}\cdot\mathbf{p}$ perturbation theory comes from the (usual) appearance of $\mathbf{k}\cdot\mathbf{p}$ in the matrix element of Eq. (H.8). Here, we are using $\mathbf{q}$ as the perturbation in wave vector and thus have $\mathbf{q}\cdot\mathbf{p}$ appearing in Eq. (H.8).

Using Eq. (H.8), we obtain

$$(\mathbf{k}l|\mathbf{k}+\mathbf{q}, l') = (\mathbf{k}l|\mathbf{k}l') + \frac{\hbar}{m}\sum_{l''}{}' \frac{(\mathbf{k}l|\mathbf{k}l'')(\mathbf{k}l''|\mathbf{q}\cdot\mathbf{p}|\mathbf{k}l')}{\mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l''}} \tag{H.9}$$

The term $(\mathbf{k}l|\mathbf{k}l'')$ appearing in Eq. (H.9) is zero unless $l = l''$ because the periodic parts of Bloch functions are orthogonal. Thus Eq. (H.9) simplifies to

$$(\mathbf{k}l|\mathbf{k}+\mathbf{q}, l') = \delta_{ll'} + \frac{\hbar}{m}\sum_{l}{}' \frac{(\mathbf{k}l|\mathbf{q}\cdot\mathbf{p}|\mathbf{k}l')}{\mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l}} \tag{H.10}$$

Defining

$$\begin{aligned} P^{\mu}_{ll'} &= (\mathbf{k}l|p^{\mu}|\mathbf{k}l') \\ &= (1/\Delta_{\text{cell}})\int u_{\mathbf{k}l}p^{\mu}u_{\mathbf{k}l'}\,d\mathbf{r} \end{aligned} \tag{H.11}$$

and

$$\hbar\omega_{l'l} = \mathscr{E}_{\mathbf{k}l'} - \mathscr{E}_{\mathbf{k}l} \tag{H.12}$$

where p^μ is the momentum operator associated with the direction of propagation $\mathbf{q}$, we obtain

$$(\mathbf{k}l|\mathbf{k} + \mathbf{q}, l') = \delta_{ll'} + \frac{\hbar q}{m} \sum_{l}{}' \frac{P^\mu_{ll'}}{\hbar\omega_{l'l}} \tag{H.13}$$

Of course, there is really no summation left. Both l and l' are given. We merely require that the summation give no contribution for $l = l'$ by writing

$$(\mathbf{k}l|\mathbf{k} + \mathbf{q}, l') = \delta_{ll'} + (1 - \delta_{ll'})\,(q/m\omega_{l'l})\,P^\mu_{l'l} \tag{H.14}$$

Equation (7.54) follows directly from Eq. (H.14).

INDEX