Optical effects in solids

David B. Tanner

Department of Physics, University of
Florida, Gainesville, FL 32611-8440, USA

TABLE OF CONTENTS

1. INTRODUCTION

The way in which light interacts with material objects is determined by the optical properties of the materials. Why might you want to think about these optical properties? There are at least two reasons. One is that you can make use of known optical materials to design and build devices to manipulate light: mirrors, lenses, filters, polarizers, and a host of other gadgets. The second is that you can measure the optical properties of some new material and obtain a wealth of information about the low energy excitations that govern the material's physics. Figure 1 is a chart that identifies some of these excitations and indicates the part of the spectrum where they might be expected to appear.

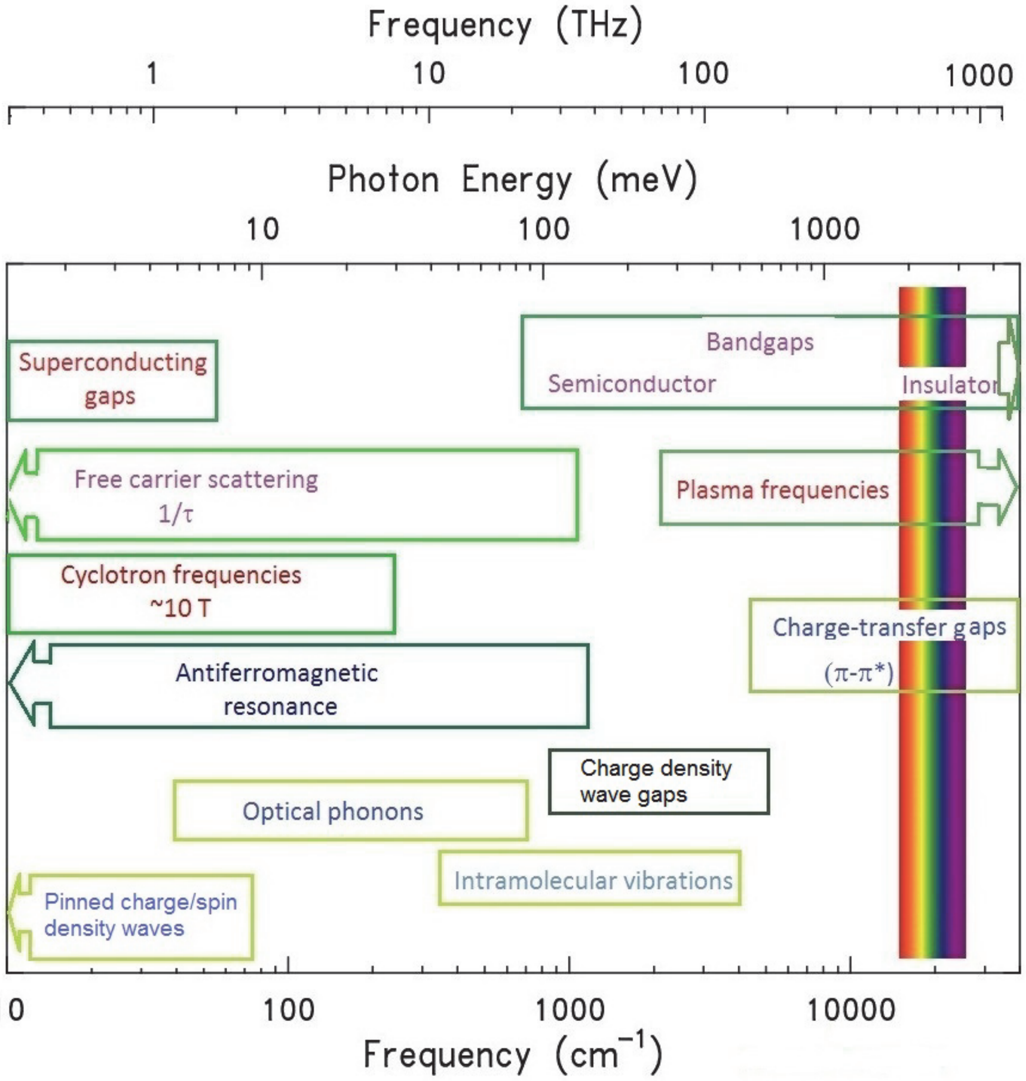

Fig. 1. Chart showing optical processes in solids, with an indication of the frequencies where these processes typically may be studied. Frequencies are given on three scales: the uppermost scale shows THz, 10^{12} cycles/sec; the second shows photon energies in meV; and the bottom shows wavenumbers, $\tilde{\nu}$ (in cm^{-1}), defined by $\tilde{\nu} = 1/\lambda$ with λ the wavelength measured in cm.

In common parlance, "optical" can be a synonym of "visual," and hence related to human eyesight. This interpretation would restrict discussion to the visible part of the electromagnetic spectrum, light with wavelengths of 390 to 780 nm, indicated by the little rainbow in Fig. 1. (I will need to introduce the variety of measures used for the wavelength or the frequency or the photon energy of electromagnetic waves. Three are shown in Fig. 1: frequency f in THz, photon energy E in meV, and another frequency unit, wavenumbers or cm^{-1}. The latter may be unfamiliar to persons who have not worked in the field of optical effects in solids; it is the inverse of the wavelength in cm. The visible spectrum spans 385–770 THz, 1.59–3.18 eV and 12,800–25,600 cm^{-1}. Units used in papers about the optical properties of solids are discussed in Appendix A.)

Of course, here I will *not* make the interpretation that optical means visible; instead, materials properties will be considered over a wide range of frequencies or wavelengths. Plausible ranges are discussed towards the end of this chapter.

Even if you were constrained to the use of your own eyes, you would see that solids have a wide range of optical properties. Silver is a lustrous metal used for centuries in coins and fine tableware, with a high reflectance over the whole visible range. Silicon is a crystalline semiconductor and the foundation of modern electronics. With its surface oxide freshly etched off, silicon is also rather reflective, although not as good a mirror as silver.[*] Salt (sodium chloride) is a transparent ionic insulator, is necessary for life, and makes up about 3.5% (by weight) of seawater. A crystal of salt is transparent over the entire visible spectrum; because the refractive index is about 1.5, the reflectance is everywhere about 4%.

If you had ultraviolet eyes, you would see these materials differently. Silver would be a poor reflector, with at most 20% reflectance and trailing off to zero at the shortest wavelengths. In contrast, the reflectance of silicon would be better than in the visible, reaching up to 75%. Sodium chloride would be opaque over much of the spectrum, with a reflectance a bit higher than in the visible. Those with infrared eyes would also see things differently from visible or uv-sensitive individuals. Silver would have a reflectance above 99%. Silicon would appear opaque at the shortest infrared wavelengths but would then become transparent, so that you could see through even meter-thick crystals.[†] Sodium chloride remains transparent over much of the infrared, but an opaque and highly reflecting "reststrahlen" (residual ray) region occurs at long wavelengths. In the reststrahlen band, NaCl has a reflectivity not much below that of a metal.

If you put on your solid-state-physics hat, you can understand the optical properties of these materials, at least qualitatively. Silver is a nearly free-electron metal, with one electron per atom in the metallic Fermi surface. These mobile electrons give the high electrical conductivity; they form a plasma that makes silver opaque and highly reflective below the plasma frequency.[‡] Silicon is a semiconductor with a gap between the filled valence band and

---

[*] Silver reflects about 98% of red light and about 80% of violet light; silicon reflects about 33% of red and 50% of violet.

[†] Here ultra-high-purity is assumed. Moreover, in the far-infrared region there is a band caused by lattice vibrational effects—multiphonons in this case—where silicon is opaque unless rather thin.

[‡] The connection between high absorption, high conductivity, and high reflectance is not intuitive. For the moment, I'll just assert that all three go together. So at wavelengths where a material is opaque, it also has increased reflectance. The more intense is the absorption, the higher the conductivity and also the higher

the empty conduction band. Photons with energies below the gap can propagate without loss in silicon. Photons with energies above the gap are absorbed, generating electron-hole pairs. This absorption renders silicon opaque and, as mentioned, increases the reflectance. Sodium chloride is an insulating crystal, with a band gap in the ultraviolet. Similar to silicon, photons with energy larger that the gap are absorbed. Sodium chloride has two atoms per unit cell; these occur as ions, Na$^+$ and Cl$^-$; an electric field displaces these ion, producing induced dipoles in the solid. With a two-atom basis, the lattice vibrations have an optical branch, and the reststrahlen band is a result of the light exciting this optical branch.

Now let me return to the question of the range of wavelengths (or the range of light frequencies or of photon energies) over which I can discuss the optical properties of solids. The electromagnetic spectrum extends over a huge range; one of many existing cartoons illustrating the "electromagnetic spectrum" is shown in Fig. 2. This chart shows wavelengths from km to pm along with corresponding frequencies and photon energies. So the question is what part of this spectrum might be used to study the optics of solids?

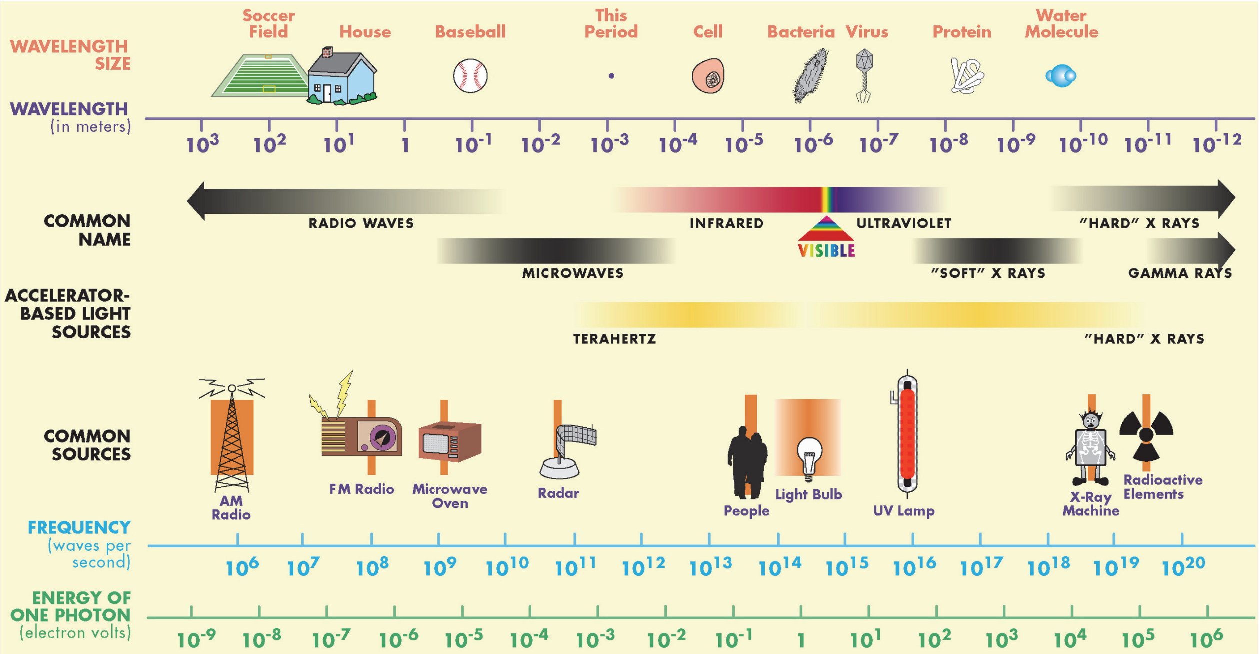

Fig. 2. Electromagnetic spectrum, after a diagram from SURA.

To start, I'll want to use continuum electrodynamics, so the short wavelength limit is set by a requirement that the wavelength be larger than the spacing between atoms. When the wavelength is less than the interatomic distances, diffraction effects dominate. X-ray diffraction is essential for determining crystal structure but beyond my scope. At somewhat longer wavelengths, continuum electrodynamics is fine, but the materials properties are essentially a superposition of atomic transitions. Solid-state effects contribute of course, but minimally for wavelengths shorter than something on the order of 50 nm.[*]

---

the reflectance. A hand-waving argument says that high conductivity means large currents in response to applied electric fields; the power loss or absorption goes as $\mathbf{j} \cdot \mathbf{E} = \sigma E^2$. See page 54 for further discussion.

[*] Using $\lambda f = c$, $\tilde{\nu} = 1/\lambda$, and $E = hf$ with λ the wavelength, f the frequency in Hz, $\tilde{\nu}$ the frequency in cm^{-1},

As wavelengths get longer and longer, there is of course no problem with continuum electrodynamics. However, practically speaking, the physics that governs the electromagnetic response at dc and audio frequencies is the same as the physics at ultra-high radio frequencies and even microwaves. So the lowest frequencies that I will consider are around a few GHz.

There is a second reason for setting a long-wavelength limit. One GHz (4 μeV photon energy) corresponds to 30 cm wavelength, and this large scale raises an experimental issue: to shine light on an solid, one sends a beam that one would like to consider to be composed of plane electromagnetic waves. Cartoons of a typical experimental setups are shown in Fig. 3. The left panel shows a reflectance ($\mathscr{R}$) experiment and the right a transmittance ($\mathscr{T}$) experiment. Light comes from a source (of known properties) that can emit a range of wavelengths, encounters the sample, and goes to a detector where it is converted to an electrical signal, is amplified, and recorded, yielding a spectrum of $\mathscr{R}$ or $\mathscr{T}$ *vs.* the wavelength or the frequency. These ideas are only a good approximation when the wavelength is small compared to the size of the sample or of the experimental apparatus. When this condition is no longer the case, diffraction effects (by the sample not by the atomic lattice) as well as waveguide effects in the surrounding apparatus become important.*

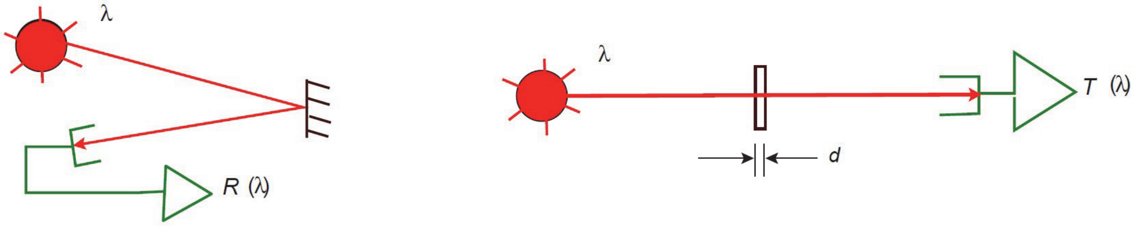

Fig. 3. Cartoon of experiment where one measures reflectance (left) or transmittance (right).

It makes little sense to be very precise in specifying the wavelength or frequency limits over which optical concepts are important for the physics of solids. So I'll say that I will consider the wavelength range to be several cm to several tens of nm, the frequency range to be a few GHz to a few PHz, and the photon energy range to be tens of μeV to tens of eV. This range covers the bands labeled microwaves, infrared, visible, and ultraviolet in Fig. 2. There is a factor of a million between one end and the other; that should be enough for everybody.†

In the following chapters, I'll remind you a little bit about electromagnetism, including Maxwell's equations and their plane-wave solutions. I'll then restrict myself for some time to local, nonmagnetic, isotropic, homogeneous, and linear solids.‡ The motivations for these

---

or wavenumber, E the photon energy, c the speed of light, and h Planck's constant, 50 nm corresponds to a frequency of 6 PHz (petaHertz), a wavenumber of 200,000 cm^{-1}, and photon energies of 25 eV.

[*] One may of course measure materials properties all the way to zero frequency (infinite wavelength) by electrical means: apply contacts and measure resistance, capacitance, etc. I will discuss the connection of optical measurements to dc electrical properties a number of times.

[†] In fact there are few materials that have been studied over the entire range. A much more typical range is in wavelength from, say, 0.3 mm to 300 nm, far infrared to near ultraviolet, a range of 10^3.

[‡] Much of this discussion can apply to liquids as well, and even to a dilute gas, but the physics discussion will rely on solid-state physics ideas: Fermi surfaces, band structure, etc.

restrictions are that the materials should be local, so that the current at point **r** is a function only of the fields at **r**; nonmagnetic, because most solids are nonmagnetic and because even magnetic materials only show themselves to be magnetic at rather low frequencies; isotropic, so that the properties do not depend on the direction or polarization of the light; homogeneous, so that the response functions do not depend on spatial position; and linear, so that I may make a Fourier decomposition of the fields and treat each component independently. With these approximations, I'll introduce the idea of a complex dielectric function, discuss classical theories of free carrier response in metals, interband absorption in semiconductors and insulators, and lattice vibrations (phonons). Next, I will show data from the literature to illustrate some of the concepts of these theories.

I return to electromagnetism to calculate the reflection and transmission by a thin film or slab, as a way to link experimental measurements to the optical properties of the material making up the film. The next step is to introduce simple quantum mechanics, leading to a discussion of free-electron metals followed by a presentation of the quantum-mechanical perturbation-theory of optical absorption, culminating in an important sum rule for the conductivity. The sum rule result motivates an interlude about causality where I obtain the Kramers-Kronig relations between the absorptive and dispersive parts of the response function, discuss the analysis of reflectivity by Kramers-Kronig methods, and derive other sum rules. I return to solids to address the band structure of simple solids and the interband absorption edge in semiconductors. Simple treatments of the optics of superconductors, distinctly quantum-mechanical solids, and of materials with strong correlations and interactions follow.

After this, it will be time to relax the initial conditions and discuss (one at a time!) nonlocal properties, mostly the anomalous skin effect in a pure metal, wave propagation in anisotropic materials, magneto-optics, and randomly inhomogeneous materials.[*] The last chapter is a survey of experimental techniques. Several appendices discuss units, some mathematics, and other things "optical."

---

[*] I'll leave the huge subjects of quantum optics and nonlinear optics to others; I think it is better to say nothing than to make short and probably superficial treatments.

2. MAXWELL'S EQUATIONS AND PLANE WAVES IN MATTER

2.1 Optical constants

The response of materials to light is described by a number of quantities, often called "optical constants." Among these are:

- ϵ, the dielectric constant.
- σ, the electrical conductivity.
- χ, the susceptibility.*
- n, the refractive index.
- κ, the extinction coefficient.
- δ, the electromagnetic skin depth.
- Z, the surface impedance.

and many others. See Appendix G for a longer but still incomplete list.

These quantities are neither constant nor independent. They are functions of the frequency, temperature, pressure, external magnetic field, and many other things. By knowing two of these, one that describes the absorption in the solid (such as the electrical conductivity or the extinction coefficient) and one that describes dispersion (such as the dielectric constant or the refractive index), all of the others may calculated.

2.2 Maxwell's equations

In my initial Electricity and Magnetism course, the professor said that the subject is governed by equations written in the 19th century by Maxwell,[1] and that some day a teacher would come in and write these equations on the blackboard on the first day of the class and then proceed to develop a theory based on these equations. Of course in a junior-level class, he did not do this, but I can do it here.

There are two versions: Maxwell's microscopic equations and Maxwell's macroscopic equations. The first are more fundamental, because they describe the microscopic fields arising from every charge in the Universe and from the motion of these charges as well. The second are more fun, because they average over the charges in macroscopic media† and allow great simplification of the subject.

After restricting myself to macroscopic charges and currents, I have ρ_{ext}, the external charge density, and $\mathbf{j}_{\text{free}}$, the free current density as sources in Maxwell's equations. The external charge density is essentially the charge imbalance in the medium; it is zero for electrically neutral objects.‡ The free current density is the result of the motion of free charges in the metal. These charges are typically the free electrons in a metal, but could include doped or thermally excited free carriers in a semiconductor or the diffusion of ions in

---

* Electric *and* magnetic.

† In the absence of external fields, the positive charges (the nuclei) in a uniform medium are perfectly screened by the negative charges (the electrons) so that both may be omitted completely from the charges and currents in Maxwell's equations. External fields may polarize these charges (pushing + in one direction and − in the opposite direction); the effects of such currents and dipole moments are the subject of our optical properties studies.

‡ And it does *not* include any charge imbalance caused by external fields. This polarization gets included in the dipole moment/unit volume $\mathbf{P}$.

an electrolyte. Note that the free carriers are compensated by bound ions in the electrically neutral material so these charges do not appear as external charge density.

So let me begin by writing Maxwell's equations for macroscopic media. I'll use cgs-Gaussian units;[2] Appendix B gives them in SI units.

$$\nabla \cdot \mathbf{D} = 4\pi\rho_{\text{ext}} \tag{1a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1b}$$

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

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{j}_{\text{free}} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \tag{1d}$$

I must add a connection to classical mechanics to these equations. The force $\mathbf{F}$ on a particle with electric charge q satisfies the Lorentz law,

$$\mathbf{F} = q(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}), \tag{2}$$

where $\mathbf{v}$ is the particle's velocity vector.

The definitions of the auxiliary fields are

$$\begin{aligned} \mathbf{D} &= \mathbf{E} + 4\pi\mathbf{P} \\ \mathbf{H} &= \mathbf{B} - 4\pi\mathbf{M} \end{aligned} \tag{3}$$

The quantities in Eqs. 1, 2, and 3 are all functions of space and time. The vector $\mathbf{E}$ is called the "electric field" and the vector $\mathbf{D}$ is called the "electric displacement field." In vacuum, $\mathbf{D}$ and $\mathbf{E}$ are proportional to each other, with the multiplicative constant ϵ_0 depending on the physical units.* Inside a material they are different on account of the polarization of the material. The vector $\mathbf{P}$ is the "polarization field" or the "electric dipole moment/unit volume." The vector $\mathbf{B}$ is called the "magnetic field," and it and $\mathbf{E}$ are defined as the vector fields necessary to make the Lorentz law correctly describe the forces on a moving charged particle. $\mathbf{B}$ is also called the "magnetic flux density," or the "magnetic induction." The vector $\mathbf{H}$ is also sometimes called the "magnetic field". Other names include the "magnetic field intensity," the "magnetic field strength," and the "magnetizing field." In vacuum, $\mathbf{B}$ and $\mathbf{H}$ are proportional to each other, with the multiplicative constant μ_0 depending on the physical units.* Inside a material they are different on account of the magnetization of the material. The scalar ρ_{ext} is the "free volume charge density" and the vector $\mathbf{j}_{\text{free}}$ is the "free electric current density." The vector $\mathbf{P}$ is the "polarization field," also known as the "electric polarization," "electric polarization density," or "electric dipole moment per unit volume." The vector $\mathbf{M}$ is the "magnetization field" or the "magnetic dipole moment/unit volume."

Three of the four Maxwell equations have names: Equation 1a is Gauss' Law; Eq. 1c is Faraday's law of induction; and Eq. 1d is Ampére's law with Maxwell's correction. I'll just call it Ampére's law. Eq. 1b is sometimes called Gauss' law for magnetism but is sometimes called the no magnetic monopole law. Others say that it has no name. I'll call it the "no-monopole law."

---

* Unity in cgs-Gaussian units.

2.3 Total, free, and bound charges and currents

If I view charges as being small point-like particles moving here and there in space, then a classical-physics-style definition of the microscopic charge density in a solid might be:

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

whereas the current would be written.

$$\mathbf{j}_{\text{micro}}(\mathbf{r},t) = \sum_i q_i \mathbf{v}_i(t) \delta(\mathbf{r} - \mathbf{r}_i(t)).$$

Here, the sum i runs over all electrons and nuclei, q_i is the charge of the ith particle, which is located at (time-dependent coordinate) $\mathbf{r}_i$ and moving with velocity $\mathbf{v}_i$.

There are quantum mechanical versions of the charge density and current density also. To use them requires a solution of Schrödinger's equation in the solid. Having that, I could for example find the contribution of the electrons* (with charge $-e$) to the total charge density to be $\rho_e = -e\Psi_e^*\Psi_e$ with Ψ_e some total wave function of the electrons, and, similarly $\mathbf{j}_e = -(\hbar e/2mi)\left[\Psi_e^*\nabla\Psi_e - (\nabla\Psi_e^*)\Psi_e\right]$.

Accounting for all the particles in a material with 10^{22} particles per cubic centimeter is of course too hard, so I'll average over some volume ΔV. The scale of V is taken to be large with respect to the interatomic spacing a and small with respect to the electromagnetic wavelength λ, or

$$a^3 \ll \Delta V \ll \lambda^3.$$

The inequalities are easy to satisfy because $a \sim 0.1$ nm and, according to the discussion on page 4, λ is bigger than 10 nm, so there is a factor of 10^6 between their cubes. The charge and current then are written as averaged quantities, $\rho(\mathbf{r},t) = qn(\mathbf{r},t)$ and $\mathbf{j}(\mathbf{r},t) = \rho(\mathbf{r},t)\mathbf{v}_d(\mathbf{r},t)$, where q is the average charge of the particles n is their number density, and $\mathbf{v}_d$ the average velocity in ΔV, known as the "drift velocity."

The electrons and nuclei do not know that I would like to average them out, so they still respond to applied fields, producing dipole moments in the material, $\mathbf{P}$ and $\mathbf{M}$. These bound or polarization charges, electric polarization currents, and magnetization currents are determined by the physics of the solid. Then, the macroscopic polarization charge density ρ_{pol}, the polarization current density $\mathbf{j}_{\text{pol}}$, and the magnetization current density $\mathbf{j}_{\text{mag}}$ are defined in terms of polarization $\mathbf{P}$ and magnetization $\mathbf{M}$ as $\rho_{\text{pol}} = -\nabla\cdot\mathbf{P}$, $\mathbf{j}_{\text{pol}} = c\frac{\partial\mathbf{P}}{\partial t}$, and $\mathbf{j}_{\text{mag}} = c\nabla\times\mathbf{M}$.

Maxwell's macroscopic equations reduce to the microscopic equations if one recognizes that the microscopic, free, and bound charge and current density are related by $\rho_{\text{micro}} = \rho_{\text{ext}} + \rho_{\text{pol}}$, and $\mathbf{j}_{\text{micro}} = \mathbf{j}_{\text{free}} + \mathbf{j}_{\text{pol}} + \mathbf{j}_{\text{mag}}$, and then uses Eq. 3 to eliminate $\mathbf{D}$ and $\mathbf{H}$.

---

* Because of a choice made by Franklin, the electron is negative. I'll take e to be a positive number and put the sign in explicitly as needed.

2.4 Maxwell's equations for solids

I am almost to a point where I can begin to address optical effects in solids. But there are two things still to do. First, the world is in general electrically neutral,[*] so I will take $\rho_{\text{ext}} = 0$. For notational convenience, I will drop the "free" subscript on the current, with the understanding that it is the free current that is meant.[†] Thus, $\mathbf{j}_{\text{free}} \to \mathbf{j}$.

Henceforth, I'll use the form of Maxwell's macroscopic equations written just below. I'll claim that these cover the most general cases that occur in the studies of optical effects in solids. I'll make further simplifications in the next several pages.

$$\nabla \cdot \mathbf{D} = 0 \tag{4a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{4b}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{4c}$$

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \tag{4d}$$

Eq. 4 has a certain pleasing symmetry. After I've defined and used the complex dielectric function, this symmetry will become more perfect.

2.5 Plane-wave solutions

I write the electric field as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{q}\cdot\mathbf{r} - \omega t)} \tag{5}$$

and the magnetic field as

$$\mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0 e^{i(\mathbf{q}\cdot\mathbf{r} - \omega t)}, \tag{6}$$

where $\mathbf{E}_0$ and $\mathbf{H}_0$ are constant vectors giving the amplitudes of the fields (quite possibly complex quantities) and direction,[‡] $\mathbf{q}$ is the wave vector of the field (measured in cm^{-1} in cgs), and ω is the angular frequency (in radians/sec or s^{-1}). I choose to use $\mathbf{H}$ rather than $\mathbf{B}$, as does Jackson.[2]

The quantity $\sqrt{-1}$ does not appear in Eqs. 1, 2, 3, or 4. Hence the only mechanism to have equations containing real and imaginary quantities is through writing the fields as complex quantities, as in Eqs. 5 and 6. It would be perfectly valid to write $\mathbf{E} = \mathbf{E}_0 \cos(\mathbf{q} \cdot \mathbf{r} - \omega t + \phi_e)$ and $\mathbf{H} = \mathbf{H}_0 \cos(\mathbf{q} \cdot \mathbf{r} - \omega t + \phi_m)$, where the phases ϕ_e and ϕ_m allow for differing phases in electric and magnetic fields.[§] In this case, all quantities in electromagnetism would be purely real quantities. I could do this; at a minimum it would give fine training in the use

---

[*] Anyone who has been shocked by a static discharge after crossing a rug on a dry day or who has seen a lightning strike from a thunderstorm knows (1) that this statement is not always true but (2) that the charge imbalance does not last.

[†] And see the discussion on page 16 where the distinction between free and bound currents will be blurred as well.

[‡] Hence, to be explicit I could write $\mathbf{E}_0 = \hat{\boldsymbol{e}}\mathcal{E}_0 e^{i\phi}$, where $\hat{\boldsymbol{e}}$ is a unit vector pointing in the field direction, $\mathcal{E}_0$ is the field magnitude, and ϕ is a phase which, when combined with $(\mathbf{q} \cdot \mathbf{r} - \omega t)$, specifies where the zeros, crests, and valleys of the wave occur.

[§] Alternatively, write $\mathbf{E} = \mathbf{E}_c \cos(\mathbf{q} \cdot \mathbf{r} - \omega t) + \mathbf{E}_s \sin(\mathbf{q} \cdot \mathbf{r} - \omega t)$ and a similar equation for $\mathbf{H}$.

of trigonometric identities, because as soon as I took a derivative, there will be both sines and cosines in the math.

It is conventional to say that one writes the fields as complex quantities but when one wants to evaluate the observable fields, the real part should be taken. This statement is basically true, though one has to be careful in cases where two complex fields are multiplied together.

I know from freshman physics that the crests and valleys of the wave repeat in space (at fixed time) by a translation of the wavelength λ. From this assertion I get $|\mathbf{q}|\lambda = 2\pi$ or $|\mathbf{q}| = 2\pi/\lambda$. At a point in space, the wave repeats every time the time advances by the period T; hence $\omega = 2\pi/T = 2\pi f$ with f the frequency. Moreover, $\lambda f = v$ and $\omega = qv$ with v the wave speed ($= c$ in vacuum). See Appendix A for a discussion of units used for length, time, frequency, energy, fields, and many other quantities used in optical studies.

The use of plane-wave fields may seem arbitrary or restrictive, but Fourier tells us that

$$\mathbf{E}(\mathbf{r},t) = \int d^3q \int d\omega\, \mathbf{E}_0(\mathbf{q},\omega) e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)},$$

with

$$\mathbf{E}_0(\mathbf{q},\omega) = \int d^3r \int dt\, \mathbf{E}(\mathbf{r},t) e^{-i(\mathbf{q}\cdot\mathbf{r}-\omega t)}.$$

I can write an arbitrary field in terms of the Fourier integral, do the usual trick of exchanging order of integration and differentiation, and find (for linear, local materials) that Maxwell's equations apply to each Fourier component.

2.6 Converting differential equations to algebraic ones

Now, I insert plane waves for each of the fields in Eq. 4. See Appendix C for the rules of how vector and partial differential operators act on plane wave functions. For example, Eq. 4a becomes

$$\nabla \cdot \mathbf{D} = i\mathbf{q} \cdot \mathbf{D} = 0$$

After doing this type of thing for all four of Maxwell's equations, cleaning up the signs, and canceling as many of the appearances of i as I can, I get

$$\mathbf{q} \cdot \mathbf{D} = 0 \tag{7a}$$

$$\mathbf{q} \cdot \mathbf{B} = 0 \tag{7b}$$

$$\mathbf{q} \times \mathbf{E} = \frac{\omega}{c}\mathbf{B} \tag{7c}$$

$$\mathbf{q} \times \mathbf{H} = -\frac{\omega}{c}\mathbf{D} - \frac{4\pi i}{c}\mathbf{j} \tag{7d}$$

with

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \qquad \mathbf{H} = \mathbf{B} - 4\pi\mathbf{M} \tag{8}$$

(I exchanged the order on the right hand side of Eq. 7d for convenience in the steps coming up.)

2.7 Vector directions

Vectors whose inner product is zero are orthogonal. Hence, $\mathbf{q}$ is perpendicular to both $\mathbf{D}$ and $\mathbf{B}$. The cross product produces a vector normal to the plane containing the other two. Hence, Eq. 7c tells me that $\mathbf{B}$ is perpendicular to $\mathbf{E}$ (and to $\mathbf{q}$, which I already knew). Equation 7d says that $\mathbf{H}$ is perpendicular to a particular linear combination of $\mathbf{D}$ and $\mathbf{j}$ and, moreover, that $\mathbf{j}$ is perpendicular to $\mathbf{q}$ because $\mathbf{D}$ is.

2.8 Electromagnetic waves in vacuum

If there is no medium, there are neither electric nor magnetic dipoles nor free currents. Hence $\mathbf{D} = \mathbf{E}$ and $\mathbf{H} = \mathbf{B}$. Equation 7 becomes

$$\mathbf{q} \cdot \mathbf{E} = 0 \tag{9a}$$

$$\mathbf{q} \cdot \mathbf{H} = 0 \tag{9b}$$

$$\mathbf{q} \times \mathbf{E} = \frac{\omega}{c}\mathbf{H} \tag{9c}$$

$$\mathbf{q} \times \mathbf{H} = -\frac{\omega}{c}\mathbf{E}. \tag{9d}$$

$\mathbf{q}$, $\mathbf{E}$, and $\mathbf{H}$ form a right handed orthogonal set, which I may orient respectively along the $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ Cartesian axes. Then* $\hat{\boldsymbol{q}} \times \hat{\boldsymbol{e}} = \hat{\boldsymbol{h}}$ and $\hat{\boldsymbol{q}} \times \hat{\boldsymbol{h}} = -\hat{\boldsymbol{e}}$, making the magnitudes of Eqs. 9c and 9d be $qE = \omega H/c$ and $qH = \omega E/c$. Solving the second of these for H and substituting in the first, I find $q = \omega/c$; either of these then give $H = E$. So in vacuum, electromagnetic waves travel at the speed of light and have equal (cgs-Gaussian) amplitudes for the electric and magnetic components. The energy density[2] $U = (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H})/8\pi$ becomes in vacuum $U_{\text{vac}} = (E^2 + H^2)/8\pi$. Half the energy is carried by the electric component and half by the magnetic component.

Figure 4 shows a cartoon of the wave in vacuum.† The wave is traveling to the right, with the $\mathbf{E}$ field (blue) in the vertical plane and the $\mathbf{H}$ field (red) in the horizontal plane. These waves are plane waves, so the fields have spatial variation in the propagation direction

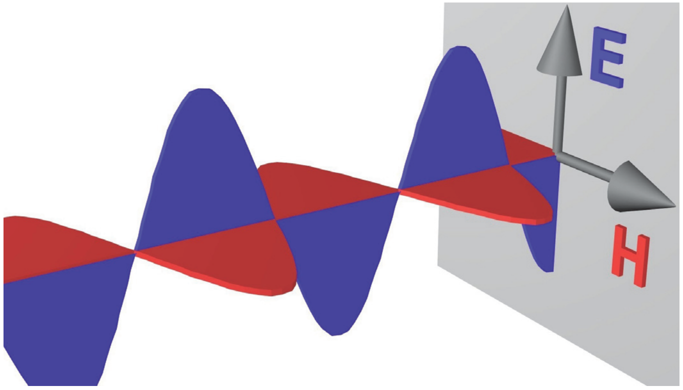

Fig. 4. Electromagnetic wave in vacuum.

---

* I'll define unit vectors parallel to any vector $\mathbf{A}$ as $\hat{\boldsymbol{a}}$ and the magnitude (length) of the vector as A.

† If you look on the web for such cartoons you may note that more than half of them show the fields incorectly, failing to satisfy $\hat{\boldsymbol{q}} \times \hat{\boldsymbol{e}} = \hat{\boldsymbol{h}}$.

while being constant in amplitude for perpendicular directions. Figure 5 is an attempt to illustrate this condition. It shows the **E** field at its crests (red arrow and red plane) and troughs (blue arrow and blue plane). The propagation vector (black) is normal to the planes of constant phase.

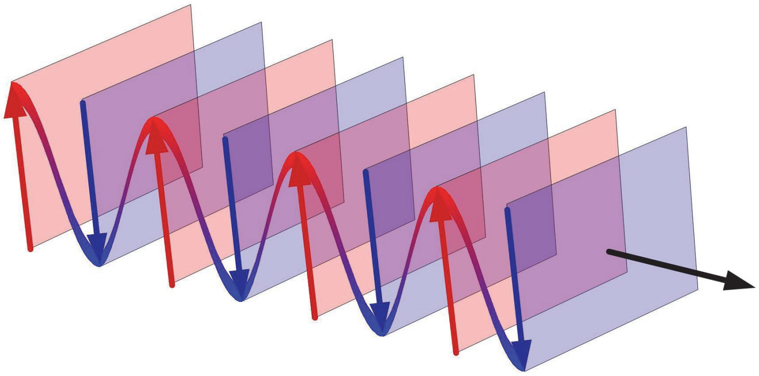

Fig. 5. Plane waves.

2.9 Five easy simplifications

Now, I return to Eqs. 7 and 8 and discuss how they change as simplifying assumptions about the material's properties are made.

2.9.1 Local response

If I write Ohm's law as $\mathbf{j} = \sigma\mathbf{E}$, I am making an assumption that the response is local, that the current $\mathbf{j}$ at point $\mathbf{r}$ depends only on the electric field $\mathbf{E}$ at that point and the conductivity σ. This statement is never completely true; the current possesses a certain momentum and, once set in motion, takes a certain time (e. g., the relaxation time τ) or distance (e. g., the mean free path ℓ) to relax back to zero. Even so, local electrodynamics is the usual case, because the mean free path is typically short compared to the lengths over which the field itself varies, and so the current is indeed related to the field by a local relation. However, in pure metals at low temperatures, the mean free path can become longer than the electromagnetic penetration depth (i. e., the skin depth). In that case, currents generated within the skin depth can travel deep into the metal, where the electric field is essentially zero. Nonlocal effects appear also in superconductors. I'll discuss nonlocal electrodynamics in chapter 12. Meantime, local behavior will be assumed.

2.9.2 Non-magnetic materials

If the material is non-magnetic, $\mathbf{M} = 0$ everywhere and $\mathbf{B} = \mathbf{H}$. Then Eq. 7 shows that $\mathbf{q}$, $\mathbf{D}$, and $\mathbf{H}$ (which is parallel to $\mathbf{B}$) form a right-handed set. $\mathbf{E}$ is perpendicular to $\mathbf{H}$, lying in the plane containing $\mathbf{D}$ and $\mathbf{q}$. $\mathbf{E}$ need not be perpendicular to $\mathbf{q}$, although it cannot be exactly parallel to it either. Analogous to the vacuum case, I'll set $\hat{\boldsymbol{q}} \| \hat{\boldsymbol{x}}$, $\hat{\boldsymbol{j}} \| \hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{h}} \| \hat{\boldsymbol{z}}$.

2.9.3 Linear materials

For many materials, the electric dipole moment/unit volume obeys

$$\mathbf{P} = \overleftrightarrow{\chi}_e \cdot \mathbf{E} \tag{10}$$

where $\overleftrightarrow{\chi}_e$ is a 3×3 tensor, the electric susceptibility. Then

$$\mathbf{D} = \mathbf{E} + 4\pi\overleftrightarrow{\chi}_e \cdot \mathbf{E} = (1 + 4\pi\overleftrightarrow{\chi}_e) \cdot \mathbf{E} \equiv \overleftrightarrow{\epsilon} \cdot \mathbf{E} \tag{11}$$

where $\overleftrightarrow{\epsilon} = 1 + 4\pi\overleftrightarrow{\chi}_e$ is the dielectric tensor.

Similarly

$$\mathbf{j} = \overleftrightarrow{\sigma} \cdot \mathbf{E}. \tag{12}$$

with $\overleftrightarrow{\sigma}$ the conductivity tensor.

And (if the material becomes magnetic again)

$$\mathbf{M} = \overleftrightarrow{\chi}_m \cdot \mathbf{H} \tag{13}$$

where $\overleftrightarrow{\chi}_m$ is a 3×3 tensor, the magnetic susceptibility. Then

$$\mathbf{H} = \mathbf{B} - 4\pi\overleftrightarrow{\chi}_m \cdot \mathbf{H}$$

and (after a bit of algebra)

$$\mathbf{B} = \overleftrightarrow{\mu} \cdot \mathbf{H} \tag{14}$$

where $\overleftrightarrow{\mu} = 1 + 4\pi\overleftrightarrow{\chi}_m$ is the permeability tensor.

Equations 10–14 indicate linear relations between applied fields and materials response. The relation is linear because if the **E** field is doubled, the polarization **P** and current **j** are doubled. Linearity is not a big restriction. Everything is linear if the fields are small enough. Everything is *non*-linear if the fields are big enough. Most optical fields are actually rather weak; the exception being high power and short pulsed lasers.

Substituting Eqs. 11 and 12 into Eq. 7d, I get

$$\begin{aligned} \mathbf{q} \times \mathbf{H} &= -\frac{\omega}{c}\overleftrightarrow{\epsilon} \cdot \mathbf{E} - \frac{4\pi i}{c}\overleftrightarrow{\sigma} \cdot \mathbf{E} \\ &= -\frac{\omega}{c}[\overleftrightarrow{\epsilon} + \frac{4\pi i}{\omega}\overleftrightarrow{\sigma}] \cdot \mathbf{E} \end{aligned} \tag{15}$$

I will spend quite a bit of time on the consequences of this equation in Chapter `<anisotropy>`, which is about anisotropic media (crystals). For now let me just note that $\mathbf{q} \times \mathbf{H}$ produces a vector that is perpendicular to **q** and **H** but **E** is not parallel to that vector, because the multiplication of a tensor with a vector produces a new vector that is in general oriented different from the initial vector.*

---

* Only if the tensor is in diagonal form *and* the field is oriented along one of the principal axes of the tensor is the result of the multiplication not rotated.

2.9.4 Isotropic systems

Amorphous solids, polycrystalline solids, and liquids are isotropic: their properties are uniform for all directions of fields and propagation. In the case of crystals, there is no requirement for isotropy. Indeed, the optical properties must reflect the symmetry of the underlying crystal structure. On the one hand, the majority of all possible crystals are anisotropic; of the seven crystal families only the cubic is by symmetry required to have isotropic optical properties. On the other hand, a large fraction of commonly encountered materials (silver, silicon, and sodium chloride for example) are cubic. On the remaining hand, many of the most interesting solid state systems, such as the high-temperature superconductors, graphene, organic solids, etc. are highly anisotropic. Many persons think their remarkable properties are a consequence of this anisotropy.

The optics of anisotropic crystals are worked out in Chapter `<anisotropy>`. For the time being, I will assume isotropy. In this case, the tensors of Eqs. 10 to 13 become scalars. Formally, one writes

$$\overset{\leftrightarrow}{\sigma} = \sigma \overset{\leftrightarrow}{\mathbf{1}}$$

with $\overset{\leftrightarrow}{\mathbf{1}}$ the identity tensor (diagonal, with all diagonal elements equal to 1). In isotropic media, **D**, **E**, and **P** are parallel, as are **H**, **B**, and **M**.

2.9.5 Homogeneous media

There are many materials whose properties vary with position. One class of such materials is exemplified by paint or ink, a collection of particles, small with respect to the wavelength, suspended in a host material. The optical properties of each type of grain and of the host are different but, because of the small grain size, the material has an apparent (or effective) uniform response. A second class of materials consist of a plane surface where the optical properties change from, say that of a metal to vacuum.

At present, I will consider the material to be spatially homogeneous, so that the response functions, such as the conductivity, satisfy

$$\sigma \neq \sigma(\mathbf{r}).$$

2.10 Maxwell's equations for local, non-magnetic, linear, isotropic, and homogeneous materials

With the assumptions of the previous section, I can write for local, nonmagnetic, linear, isotropic, and homogeneous materials:

$$\mathbf{j} = \sigma_1 \mathbf{E} \tag{16}$$

and

$$\mathbf{D} = \epsilon_1 \mathbf{E}. \tag{17}$$

The meaning of the subscripts will become clear in a moment.

Equation 17 replaces $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, Eq. 3. With it and Eq. 16, I can rewrite Eq. 7:

$$\mathbf{q} \cdot \mathbf{E} = 0 \tag{18a}$$

$$\mathbf{q} \cdot \mathbf{H} = 0 \tag{18b}$$

$$\mathbf{q} \times \mathbf{E} = \frac{\omega}{c}\mathbf{H} \tag{18c}$$

$$\mathbf{q} \times \mathbf{H} = -\frac{\omega}{c}[\epsilon_1 + \frac{4\pi i}{\omega}\sigma_1]\mathbf{E} \tag{18d}$$

Only the fourth one has changed.

I am almost to the starting point where I can discuss the optical properties of simple* materials. But first I have to say what the subscripts on ϵ_1 and σ_1 mean, given that I am using electric fields like

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}$$

so that the field amplitudes (and the materials related coefficient $[\epsilon_1 + \frac{4\pi i}{\omega}\sigma_1]$) contain real and imaginary parts. I will use the first part of the next chapter to explain and then eliminate these subscripts.

---

* Simple == local, nonmagnetic, linear, isotropic, and homogeneous solids.

3. THE COMPLEX DIELECTRIC FUNCTION ϵ AND REFRACTIVE INDEX N

3.1 Conductivity and dielectric constant

In this this chapter I'll define the complex dielectric function as I'll use it for the remainder of the text. The first topic is the relaxation between conductivity (σ) and dielectric function (ϵ). For a moment, let me return to Eq. 4d, substituting into it Eqs. 16 and 17:

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\sigma_1 \mathbf{E} + \frac{\epsilon_1}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{19}$$

As discussed on page 9, the Maxwell equations are only complex if complex fields are used with them. But how am I to think about σ_1 and ϵ_1? Are they real numbers or complex numbers? Well, if I use sines and cosines for the fields, they had better both be real numbers. If I use complex fields (as I plan to do) I can take one of two philosophical positions.

The first position omits the subscripts and identifies $\sigma\mathbf{E}$ as the free current density. Similarly $\epsilon\frac{\partial \mathbf{E}}{\partial t}$ contains the bound current density and Maxwell's displacement current. One can defend this distinction. For example, the conduction electrons in a metal are free carriers and their motion in response to an electric field is the free current. Because the carriers have inertia, their response will lag the field as frequency increases. This phase delay can be represented by having σ be complex, with the imaginary part being an inductive response that results from this inertia. Similarly, the remaining electrons in the solid are bound to their atoms, and their response is the polarization current. Inertial and dissipative effects will ensure that ϵ also is complex. I have 4 independent quantities (the real and the imaginary parts of these functions) but can identify them with the free and with the bound carriers.

The second position asserts that the conductivity and dielectric function in Eq. 19 are both real quantities, just as they would be if real-valued fields were used. It goes on to note that the distinction between free carrier and bound carrier response is clear in the case of well understood and well characterized systems such as simple metals. However, for some new material, perhaps with no good theoretical understanding, it would not be possible to make such a clear distinction. It is better to reduce the number of materials "constants" to two.

I'll take the second position. There are two independent quantities, the real conductivity σ_1 and the real dielectric function ϵ_1. The first leads to a current that is in phase with $\mathbf{E}$ (and hence is dissipative) and the second to a current that is 90° out of phase (and hence dispersive). I will combine them into a single complex dielectric function.

3.2 The complex dielectric function

Using this approach, I can substitute our plane-wave fields into Eq. 19 or just immediately copy Eq. 18d (our plane-wave-generated algebraic equation):

$$\mathbf{q} \times \mathbf{H} = -\frac{\omega}{c}[\epsilon_1 + \frac{4\pi i}{\omega}\sigma_1]\mathbf{E}. \tag{20}$$

The next step *defines* a complex dielectric function ϵ as

$$\epsilon \equiv \epsilon_1 + \frac{4\pi i}{\omega}\sigma_1. \tag{21}$$

With this definition, I arrive at the last form of Maxwell's equations, appropriate for local, nonmagnetic, linear, isotropic, and homogeneous media. I'll put them in a box to make them stand out.

$$\mathbf{q} \cdot \mathbf{E} = 0 \tag{22a}$$
$$\mathbf{q} \cdot \mathbf{H} = 0 \tag{22b}$$
$$\mathbf{q} \times \mathbf{E} = \frac{\omega}{c}\mathbf{H} \tag{22c}$$
$$\mathbf{q} \times \mathbf{H} = -\frac{\omega}{c}\epsilon\mathbf{E} \tag{22d}$$

These equations have a very agreeable symmetry. The materials properties enter only through the presence of ϵ in Eq. 22d.

Equations 22a and 22b tell me that $\mathbf{q}$ is perpendicular to $\mathbf{E}$ and $\mathbf{H}$, just as they were in vacuum. Equation 22c tells me that $\mathbf{H}$ is perpendicular to $\mathbf{q}$ and $\mathbf{E}$; the three vectors form a right handed set, with unit vectors $\{\hat{\boldsymbol{q}}, \hat{\boldsymbol{e}}, \hat{\boldsymbol{h}}\}$. Then, with $\hat{\boldsymbol{q}} \times \hat{\boldsymbol{e}} = \hat{\boldsymbol{h}}$, the vector part of Eq. 22d gives $\hat{\boldsymbol{q}} \times \hat{\boldsymbol{h}} = -\hat{\boldsymbol{e}}$ as it should, telling nothing new about vector directions.

3.3 The optical conductivity

I'll write complex quantities as $Z = Z_1 + iZ_2$, with Z_1 the real part and Z_2 the imaginary part.* The dielectric function is one such complex quantity, $\epsilon = \epsilon_1 + i\epsilon_2$. From the definition, Eq. 21, I get immediately

$$\epsilon_2 = \frac{4\pi}{\omega}\sigma_1. \tag{23}$$

There are cases where it convenient to work with a complex conductivity $\sigma = \sigma_1 + i\sigma_2$. In these cases, Eq. 22d becomes

$$\mathbf{q} \times \mathbf{H} = -[\frac{4\pi i}{c}\sigma + \frac{\omega}{c}]\mathbf{E}, \tag{24}$$

---

* Some authors use primes instead, $Z = Z' + iZ''$.

where the second term in the square brace preserves Maxwell's displacement current in the vacuum.[*] By comparison of Eqs. 20 and 24 I find

$$\sigma_2 = \frac{\omega}{4\pi}(1 - \epsilon_1). \tag{25}$$

σ_2 is positive when ϵ_1 is negative. I will show in Chapter 4 that metals and insulators both may have ϵ_1 negative over certain frequency ranges and positive over others. A negative σ_2 should not worry you. Recall that the reactance of inductors ($i\omega L$, positive) and capacitors ($1/i\omega C$, negative) have opposite signs, allowing resonant circuits to be constructed.[†]

Finally, combining Eqs. 23 and 25 with the definitions of the complex dielectric function and complex conductivity, I find that the relation between the two is

$$\epsilon \equiv 1 + \frac{4\pi i}{\omega}\sigma. \tag{26}$$

When considering a specific type of material (metal, insulator, ionic solid, superconductor) it is sometimes better to work out the conductivity and other times better to derive the dielectric function. Equation 26 allows one to translate back and forth between them.

3.4 The complex refractive index, N

With the directions removed, Eqs. 22c and 22d become $qE = \frac{\omega}{c}H$ and $qH = \frac{\omega}{c}\epsilon E$ with both E and H varying as $e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}$. I'll solve the first for H and substitute into the second to find

$$q = \frac{\omega}{c}\sqrt{\epsilon}. \tag{27}$$

This motivates the definition of the complex refractive index,

$$N = n + i\kappa \equiv \sqrt{\epsilon} \tag{28}$$

The real part of N, n is the refractive index and the imaginary part κ is the extinction coefficient.[‡]

The electric field for plane waves becomes

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} = \mathbf{E}_0 e^{i(N\frac{\omega}{c}\hat{\boldsymbol{q}}\cdot\mathbf{r}-\omega t)}. \tag{29}$$

For definiteness, let $\hat{\boldsymbol{q}} = \hat{\boldsymbol{x}}$, so $\hat{\boldsymbol{q}} \cdot \mathbf{r} = x$, and

$$\mathbf{E} = \mathbf{E}_0 e^{i(n\frac{\omega}{c}x-\omega t)} e^{-\kappa\frac{\omega}{c}x}. \tag{30}$$

Recall that in vacuum, $q = \omega/c$, so the wave vector (wavelength) in the material is larger (smaller) by a factor n. The wave decays with a decay length $\delta = c/\omega\kappa$. (In terms of the

---

[*] The vacuum has $\sigma = 0$! In fact, the vacuum has $\epsilon = 1 + i0$ and $\sigma = 0 + i0$.

[†] A negative σ_1 *should* worry you.

[‡] This is a counterexample to the $Z = Z_1 + iZ_2$ convention. Note also that some (older) texts write $N = n(1 + i\kappa)$. Furthermore, it is more common to use k than κ. I choose the latter because I want to use k as a wave vector.

vacuum* wavelength $\lambda_0 = 2\pi c/\omega = c/f$, the wavelength inside is $\lambda = \lambda_0/n$ and the decay length is $\delta = \lambda_0/2\pi\kappa$. The wave crests move at c in vacuum and at $v = c/n$ in the medium. Equation 30 can then be rewritten as

$$\mathbf{E} = \mathbf{E}_0 e^{2\pi i(\frac{x}{\lambda} - ft)} e^{-\frac{x}{\delta}}. \tag{31}$$

The observable (real part) of the electric field†, $\mathbf{E} = \mathbf{E}_0 \cos\left[2\pi(x/\lambda - ft)\right] e^{-x/\delta}$, is shown in Fig. 6. The wavelength (λ) and the damping length (δ) are both equal to 1 here.

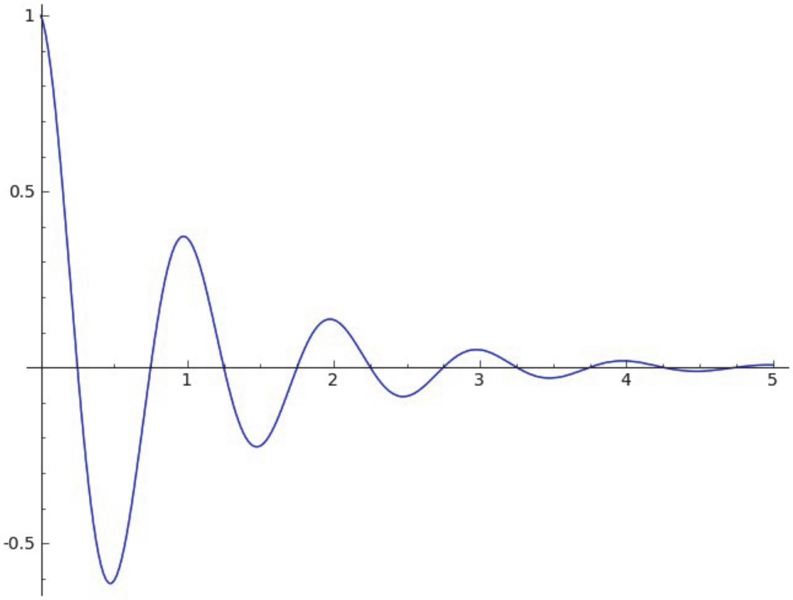

Fig. 6. Damped sinusoidal wave.

With directions removed, Eq. 22c is $H = (cq/\omega)E$, and, then, with $q = \omega N/c$, I get

$$H = NE = \sqrt{\epsilon}E. \tag{32}$$

N and ϵ are both generally complex, so the phase of $\mathbf{H}$ is different from the phase of $\mathbf{E}$.

Squaring Eq. 28 gives

$$N^2 = n^2 - \kappa^2 + 2in\kappa = \epsilon = \epsilon_1 + i\epsilon_2$$

so that it is easy to see that

$$\epsilon_1 = n^2 - \kappa^2 \tag{33}$$

and

$$\epsilon_2 = 2n\kappa. \tag{34}$$

It takes a bit more labor to calculate things the other way, but after this labor I get

$$n = \sqrt{\frac{|\epsilon| + \epsilon_1}{2}} \tag{35a}$$

---

[*] The word "wavelength" sometimes means the vacuum wavelength, c/f, and sometimes means the wavelength inside the medium, c/nf. If not specified or obvious from context, take it to mean the vacuum wavelength.

[†] I take $\mathbf{E}_0$ to be real.

$$\kappa = \sqrt{\frac{|\epsilon| - \epsilon_1}{2}} \tag{35b}$$

where $|\epsilon| = \sqrt{\epsilon_1^2 + \epsilon_2^2}$.

3.5 Poynting vector and intensity

The Poynting vector **S** describes energy flow in the medium, with the time-averaged Poynting vector for our harmonically-varying fields being[2]

$$\mathbf{S} = \frac{c}{8\pi}\mathbf{E} \times \mathbf{H}^* \tag{36}$$

Because both **E** and **H** contain a factor $e^{i(qx-\omega t)}$, the effect of the complex conjugation is to eliminate this term; the time average of the cosine is accounted for by the factor of 8π in the denominator; the time-dependent Poynting vector has 4π in the denominator.[2] Now, $\hat{\boldsymbol{e}} \times \hat{\boldsymbol{h}} = \hat{\boldsymbol{q}}$ so $\mathbf{S} \| \mathbf{q}$. Then, taking the real part of Eq. 36, I find

$$\mathbf{S} = \hat{\boldsymbol{q}}\frac{c}{8\pi}n|\mathbf{E}|^2 = \hat{\boldsymbol{q}}\frac{c}{8\pi}n|\mathbf{E}_0|^2 e^{-2\frac{\omega}{c}\kappa x}$$

The time average of **S** is related to the wave intensity I, the energy flux/unit area. Both are proportional to $|\mathbf{E}|^2$. I'll be sloppy and write

$$I = I_0 e^{-2\frac{\omega}{c}\kappa x} \equiv I_0 e^{-\alpha x} \tag{37}$$

where I_0 is the intensity in the y-z plane at $x = 0$ and

$$\alpha = 2\frac{\omega}{c}\kappa \tag{38}$$

is the absorption coefficient (units: cm^{-1}) of the material. Equation 37 is known as Lambert's Law or the Beer-Lambert Law.* Note that because $\epsilon_2 = 2n\kappa$, one can also write

$$\alpha = \frac{\omega}{c}\frac{\epsilon_2}{n} = \frac{4\pi\sigma_1}{nc}$$

which is useful when $\epsilon_2 \ll \epsilon_1$ so that $n \approx \sqrt{\epsilon_1}$.

If $\epsilon_2 = 0$ and $\epsilon_1 > 0$, then $n = \sqrt{\epsilon_1}$ and $\kappa = 0$. Because κ governs the damping of the wave, a material with positive ϵ_1 and $\epsilon_2 = 0$ is nonabsorbing; the electromagnetic wave propagates to infinity without loss of amplitude. If $\epsilon_1 < 0$ and $\epsilon_2 = 0$, then $\kappa = \sqrt{-\epsilon_1}$ and $n = 0$. The electric field is damped as $e^{-\kappa\frac{\omega}{c}x}$ with no oscillatory part at all. (The wavelength is infinite.)†

---

* Beer's Law is written in terms of the concentration c of a molecule in solution, and breaks the absorption coefficient into two parts: $\alpha = c\varepsilon$ where ε is a molecular constant called the molar extinction coefficient.

† It is not immediately obvious whether this damping is accompanied by energy absorption. I'll need to consider energy flow across an interface to determine that, in fact, there is no absorption. I can argue here that the current **j** is 90° out of phase with the electric field **E**, so the time average dissipation, $\langle \mathbf{j} \cdot \mathbf{E} \rangle$, is zero.

3.6 Normal-incidence reflectance

I'll discuss reflection and transmission of a sample (with a front and a back surface) in Chapter 6. For the moment I want to consider the simpler case of light incident normally on an interface between two semi-infinite media, one with complex dielectric function ϵ_a (refractive index N_a), occupied by the incoming and reflected fields, and a second with complex dielectric function ϵ_b (refractive index N_b), occupied by the transmitted field. See Fig. 7.

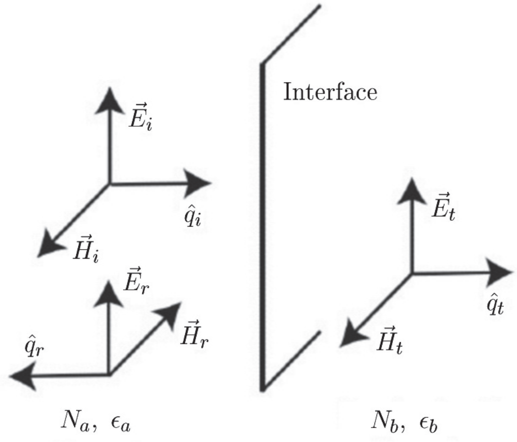

Fig. 7. Field vectors for incoming, reflected, and transmitted waves at an interface between two media.

Let me orient the $\hat{\boldsymbol{q}}$ vector of the incoming field along the $\hat{\boldsymbol{x}}$ axis and its $\mathbf{E}$ field along the $\hat{\boldsymbol{y}}$ axis, so that the $\mathbf{H}$ field is along $\hat{\boldsymbol{z}}$. The interface is in the y-z plane at $x = 0$. The incident fields are

$$\mathbf{E}_i = E_0 \hat{\boldsymbol{y}} e^{i(qx-\omega t)}$$

and

$$\mathbf{H}_i = N_a E_0 \hat{\boldsymbol{z}} e^{i(qx-\omega t)},$$

where I've used Eq. 32 to write the magnetic field amplitude.

The reflected field is traveling along the $-x$ axis, so $\mathbf{q}\cdot\mathbf{r} = -qx$ or $\hat{\boldsymbol{q}} = -\hat{\boldsymbol{x}}$. Its amplitude is proportional to E_0 and to the amplitude reflection coefficient r.

$$\mathbf{E}_r = rE_0 \hat{\boldsymbol{y}} e^{i(-qx-\omega t)}$$

and

$$\mathbf{H}_r = -N_a r E_0 \hat{\boldsymbol{z}} e^{i(-qx-\omega t)}.$$

Reversing $\hat{\boldsymbol{q}}$ means that one of $\hat{\boldsymbol{e}}$ and $\hat{\boldsymbol{h}}$ (but not both!) must reverse also. I chose $\hat{\boldsymbol{h}}$. Both fields have $q = \omega N_a / c$.

The transmitted field is in the second medium (b). Hence, it has $q = \omega N_b/c$. It also is linear in $\mathbf{E}_0$ and in the amplitude transmission coefficient t.

$$\mathbf{E}_t = tE_0\hat{\boldsymbol{y}}e^{i(qx-\omega t)}$$

and

$$\mathbf{H}_t = tN_bE_0\hat{\boldsymbol{z}}e^{i(qx-\omega t)}.$$

All fields are parallel to the interface, so the boundary conditions are that the tangential components of the total fields are continuous across the interface* so lets dot with the appropriate unit vector, cancel $\mathbf{E}_0$, which appears in every field, and evaluate the boundary conditions at the interface, $x = 0$. I can evaluate the boundary conditions at any convenient time, zero being the most convenient. The boundary conditions produce two equations:

E is continuous ($\mathbf{E}_i + \mathbf{E}_r = \mathbf{E}_t$):

$$1 + r = t. \tag{39}$$

H is continuous ($\mathbf{H}_i + \mathbf{H}_r = \mathbf{H}_t$):

$$N_a(1 - r) = N_bt. \tag{40}$$

Substitute t from Eq. 39 into Eq. 40 and solve for r to find:

$$r = \frac{N_a - N_b}{N_a + N_b}. \tag{41}$$

This clearly gives the intuitive answer when $N_a = N_b$.

Now that I know r, it is easy to find t using Eq. 39:

$$t = \frac{2N_a}{N_a + N_b}. \tag{42}$$

For the moment, I'll specialize to the case where the incident and reflected waves are traveling in vacuum, so $N_a = 1$. I can let $N_b \equiv N$ and write Eq. 41 as

$$r = \frac{1 - N}{1 + N}. \tag{43}$$

It seems that for $N > 1$ and mostly real (a common case) the amplitude reflection coefficient is negative. This fact is usually expressed by saying the reflection from less-dense to more-dense media has a 180° phase shift.

---

* The condition on **H** is actually that the fields are discontinuous by the surface current density $4\pi\mathbf{K}/c$ but for the moment let me assume no surface charges or currents.

In most experiments, one measures the intensity of incident and reflected light.[*] The normal-incidence reflectance is proportional to $|\mathbf{E}_r|^2$, i. e., $\mathscr{R} = rr^*$. With $N = n + i\kappa$, the reflectance is

$$\begin{aligned}\mathscr{R} &= \left|\frac{1-N}{1+N}\right|^2 \\ &= \frac{(n-1)^2+\kappa^2}{(n+1)^2+\kappa^2}.\end{aligned} \tag{44}$$

The range of $\mathscr{R}$ is $0 \leftrightarrow 1$.

To see how $\mathscr{R}$ depends on the refractive index, let me first look at the case where the material is weakly absorbing, so that κ can be neglected, and $\mathscr{R} \approx (n-1)^2/(n+1)^2$. Then, if n is not large (say in the range of 1.1 to 1.9, common values for glass and other materials in the visible), $\mathscr{R}$ is in the range 0.002 to 0.1. It requires $n = 6$ to raise $\mathscr{R}$ to 0.5 or so. $\mathscr{R} = 1$ only when n is very very large

If κ is large, n can be close to unity and the reflectance still will be large. In the case[†] when $n \ll 1$, $\mathscr{R}$ approaches 1 independent of κ.

3.7 What if my solid *is* magnetic?

If I keep the requirements of linear, isotropic, and homogeneous materials, but allow there to be an induced magnetization $\mathbf{M}$, I will have a non-zero magnetic susceptibility χ_m. The magnetization in a linear material is proportional to the magnetizing field $\mathbf{H}$: $\mathbf{M} = \chi_m\mathbf{H}$. Then, using $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$, I see that $\mathbf{B} = \mu\mathbf{H}$ (rather than $\mathbf{B} = \mathbf{H}$).

In a microscopic picture, the magnetic dipoles within the material are rearranged by the external field. Materials with $\mu > 1$ are "paramagnetic" (aluminum, platinum, MnO_2); those with with $\mu < 1$ are "diamagnetic" (bismuth, silver, water, benzene).[‡] Typical deviations from one are small.

I have to modify Eq. 22 when $\mathbf{M} \neq 0$. Using $\mathbf{B} = \mu\mathbf{H}$, Eq. 22c becomes

$$\mathbf{q} \times \mathbf{E} = \frac{\omega}{c}\mu\mathbf{H}. \tag{45}$$

Having the permeability in Eq. 22c increases the symmetry; now two equations have materials properties in them. Note that μ is complex, with $\mu = \mu_1 + i\mu_2$.

---

[*] Detectors typically can respond to amplitudes at low frequencies (up to GHz) but not at optical frequencies (THz to PHz and more). Instead they respond to the intensity or power in the optical beam.

[†] That case occurs in metals. In the extreme limit of materials with $n = 0$ and κ finite, $\mathscr{R} = 1$. Any value for κ suffices here. This result is the proof for the assertion in the footnote on page 20 that there is no absorption when $\epsilon_2 = 0$.

[‡] The conduction electrons of a metal exhibit Pauli paramagnetism, a definite quantum mechanical effect. The core electrons or filled shells exhibit core diamagnetism, another quantum effect. The sum of these terms determines whether the material is overall paramagnetic or diamagnetic.

Continuing with the algebra which gave Eqs. 27 and 28 I find

$$q = \frac{\omega}{c}\sqrt{\epsilon\mu}$$

and

$$N = \sqrt{\epsilon\mu}. \tag{46}$$

The algebra which gave Eq. 32 leads to

$$H = \sqrt{\frac{\epsilon}{\mu}}E.$$

Finally, the normal incidence amplitude reflectivity (Eq. 41) becomes

$$r = \frac{\dfrac{N_a}{\mu_a} - \dfrac{N_b}{\mu_b}}{\dfrac{N_a}{\mu_a} + \dfrac{N_b}{\mu_b}}.$$

and the reflectance (intensity) is $\mathscr{R} = rr*$.

3.8 Negative index materials

Problems

Slow glass: 1 meter thick; 1, 100 year storage time.

4. SEMICLASSICAL THEORIES FOR ϵ

4.1 A polarizable medium

In my junior E&M class the lecturer worked out the case of a capacitor with a dielectric inside and found that the capacitance is increased by a factor of ϵ_1, the dielectric constant. (Dissipation is usually not considered at this initial stage.) The enhancement is attributed to the polarization of the atoms or molecules in the dielectric medium, typically viewed a little spheres or ellipsoids, each with a net dipole moment $\mathbf{p}$. This dielectric constant is generally not taken to be a function of frequency.*

I can quantify this a bit more by considering a single atom between the capacitor plates, with the plates charged up to produce a field $\mathbf{E}$ at the site of the atom. The atom thereby acquires a dipole moment, whose magnitude is determined by the atomic polarizability α_e:

$$\mathbf{p} = \alpha_e \mathbf{E}. \tag{47}$$

As long as the atom has at least one excited state (of the right symmetry) the polarizability can be calculated using second-order perturbation theory.[3] In cgs-Gaussian units, the polarizability has dimensions of volume, cm^3. Hence, the polarizability is often considered as a measure of the size of the charge-distribution. (Recall[2] that for a conducting sphere, $\alpha_e = r^3$.)

Now consider placing N atoms, all the same, in the volume V between the capacitor plates. At *low densities* of atoms, the dipole moment per unit volume so produced is

$$\mathbf{P} = \frac{1}{V} \sum_{i=1}^{N} \alpha_{ei} \mathbf{E} = n\alpha_e \mathbf{E}. \tag{48}$$

Eq. 48 ignores the effects that the polarization of one atom has on its neighbors. I'll work this out in some detail on page 42. For now, because $\mathbf{P} = \chi_e \mathbf{E}$, I will write the dielectric susceptibility as $\chi_e = n\alpha_e$ and

$$\epsilon_c = 1 + 4\pi n \alpha_e. \tag{49}$$

The subscript (c) appearing in Eq. 49 stands for "core" and is meant to represent the response of tightly bound core electrons to the optical properties. As long as $\hbar\omega \ll E_0 - E_1$ where E_0 is the ground-state binding energy of these core electrons and E_1 is their first available excited state† ϵ_c can be taken as a real constant. It is often called ϵ_∞, the limiting high frequency value of the dielectric function because, as I shall show, other contributions to ϵ tend to zero as $\omega \to \infty$. As I work out simple models for the dielectric function of solids, I'll include ϵ_c where appropriate.

---

[*] Although it is. But if I neglect the frequency dependence I get to call it the dielectric "constant."

[†] The first excited state is at the Fermi energy (the chemical potential) or higher.

4.2 Drude absorption by free carriers in metals and semiconductors

The first serious model for optical properties that I will introduce is the one due to Drude.[4,5] Drude's model of metallic conduction predates the quantum theory of solids, and so omitted several important considerations. The most notable of these are (1) the fact that electrons in a metal obey Fermi-Dirac statistics and are a temperature low compared to the Fermi temperature and (2) Bloch's theorem, that electrons can effectively propagate without dissipation in a perfect periodic potential. I'll introduce the theory keeping these concepts in mind, but at the beginning—before worrying about values of things like the mean free path or the effective mass—they make little difference.

4.2.1 Components of the Drude model

The Drude model is based on the following ideas:

- The metal has a density n mobile carriers, where n is a number/cm^3 (number density, not mass density).
- These carriers are typically electrons,[*] with charge $q = -e$.
- The carriers are free. This statement means that the internal restoring force is zero, or the potential is constant.
- The carriers do not interact; the electron-electron potential is neglected.
- If set in motion, the carriers relax to equilibrium due to damping from collisions. This damping is taken as a frictional or viscous force, proportional to the drift velocity of the carriers.[†]
- Newton's second law for an electron whose coordinate is at $\mathbf{r}$ is

$$\sum \mathbf{F} = m\mathbf{a} = m\ddot{\mathbf{r}}$$

- The forces are the external electric field $-e\mathbf{E}$ and the damping $\Gamma\mathbf{v} = -\Gamma\dot{\mathbf{r}}$.
- Linear response, so coordinates, velocities, accelerations have the same time dependence as the field $\mathbf{E}$.
- The electric field is the applied field $\mathbf{E}_{\text{ext}}$ and not some local field because the electrons sample all parts of the crystal and average over the dipole fields of the atoms. See page 42 for more on this point.

4.2.2 Equation of motion and its solution

With the above, the equation of motion of one electron is

$$m\ddot{\mathbf{r}} = -e\mathbf{E}_{\text{ext}} - \Gamma\dot{\mathbf{r}} \tag{50}$$

The electric field is the plane wave field of Eqs. 5 or 29. Note that there are two sets of coordinates here. The electron has a location ($\mathbf{r}$) which varies in time as it moves. The electric field has a value everywhere in space according to Eq. 5. Of course it is the value of

---

[*] But not always. Holes are easy to accommodate; change the sign of the charge carrier to positive. The equation for the conductivity will be the same whether one has holes or electrons.

[†] This is where the Fermi-Dirac statistics first rears its head. The electrons at the Fermi energy have very high velocities; the drift velocity is the deviation of from equilibrium. Of course Drude took the equilibrium speed as zero or as a thermal speed for classical particles.

$\mathbf{E}$ at the location of the particle that matters. For simplicity I'll orient coordinates fo that $\hat{\boldsymbol{q}} = \hat{\boldsymbol{x}}$, making $\mathbf{E}$ be only a function of the x coordinate and is constant for all y and z:

$$\mathbf{E}_{\text{ext}} = \mathbf{E}_0 e^{i(qx-\omega t)}.$$

To maintain a nice right-handed set of fields, I'll take $\hat{\boldsymbol{e}} = \hat{\boldsymbol{y}}$, so the force on the electron is in the $\hat{\boldsymbol{y}}$ direction.[*]

Because the equation of motion contains only the velocity and acceleration of the electron, it makes sense to calculate the current, which, as was discussed on page 8, is

$$\mathbf{j} = -ne\mathbf{v}, \tag{51}$$

where n is the number density of electrons. So let me write

$$\mathbf{v} = \mathbf{v}_0 e^{i(qx-\omega t)} = \dot{\mathbf{r}}, \tag{52}$$

with $\mathbf{v}_0$ a complex constant vector. Taking a time derivative of $\mathbf{v}$ gives the acceleration as $\ddot{\mathbf{r}} = -i\omega\mathbf{v}_0 e^{i(qx-\omega t)}$. Note that all three terms in Eq. 50 have the same complex exponential factor $e^{i(qx-\omega t)}$, allowing that factor to be hidden. The equation of motion is now an algebraic one,

$$-i\omega m\mathbf{v} = -e\mathbf{E}_{\text{ext}} - \Gamma\mathbf{v}.$$

It is convenient (and conventional) to make the replacement $\Gamma = m/\tau$, where τ is the relaxation time of the charge carriers.[†] With this definition I can solve for the velocity

$$\mathbf{v} = \frac{-e}{m(1/\tau - i\omega)}\mathbf{E}. \tag{53}$$

The velocity is in phase with the applied field at low frequencies ($\omega \ll 1/\tau$); it also is independent of frequency. When $\omega \gg 1/\tau$ it is 90°out of phase with the electric field and the amplitude of the motion decreases as $1/\omega$.

4.2.3 Ohm's law and the Drude conductivity

Now, substitute this velocity into Eq. 51 to find

$$\mathbf{j} = \frac{ne^2}{m(1/\tau - i\omega)}\mathbf{E}.$$

This equation is in the form of Ohm's law, $\mathbf{j} = \sigma\mathbf{E}$. The Drude conductivity is

$$\sigma = \frac{ne^2\tau/m}{1 - i\omega\tau}. \tag{54}$$

At dc ($\omega = 0$) the conductivity is

$$\sigma_{dc} = ne^2\tau/m. \tag{55}$$

σ_{dc} is proportional to the carrier density n, the relaxation time τ, and inversely proportional to the mass of the carriers.

---

[*] $-\hat{\boldsymbol{y}}$ actually, since the electron has a negative charge.

[†] Suppose when $\mathbf{v} = \mathbf{v}_0$ the electric field is turned off. Equation 50 becomes $m\dot{\mathbf{v}} = -m\mathbf{v}/\tau$. Its solution is $\mathbf{v} = \mathbf{v}_0 e^{-t/\tau}$. τ is the time required to relax the velocity to $1/e$ of the initial value.

4.2.4 Some numbers

Silver or copper have a dc conductivity of $\sigma_{\mathrm{dc}} = 6.2 \times 10^5\ \Omega^{-1}\mathrm{cm}^{-1}$ (i. e., a resistivity of $1.6\ \mu\Omega$-cm). After reading Appendix A, I know to multiply σ_{dc} by $30c$ to get the conductivity into esu units; it is $5.6 \times 10^{17}\ \mathrm{s}^{-1}$. This is a fairly high (angular) frequency; it corresponds to about 370 eV photon energy. The approximation $\sigma_{\mathrm{dc}} \gg \omega$ is certainly valid at low frequencies ($\omega\tau < 1$).

I can estimate the mean free time from the dc conductivity. Solving Eq. 55 for τ, I obtain $\tau = m\sigma_{\mathrm{dc}}/ne^2$ (with σ_{dc} in s^{-1}!). Using $n = 5.9 \times 10^{22}\ \mathrm{cm}^{-3}$ (silver or gold), $e = 4.8 \times 10^{-10}$ esu, and $m = 9.1 \times 10^{-28}$ g, I get $\tau = 3.8 \times 10^{-14}$ s. If I convert this to a frequency (by inversion) $1/\tau = 2.6 \times 10^{13}\ \mathrm{s}^{-1}$. I need to be clear about the units here, because there has been occasional confusion about what this time and rate mean.[6] τ is the mean free time between collisions and $1/\tau$ is the rate at which collisions occur; it also represents the angular frequency at which $\omega\tau = 1$. If the static field driving the current were suddenly to be extinguished, the currents would decay to zero as $j \propto e^{-t/\tau}$. Hence, I divide by 2π to convert to ordinary frequency and then by c to convert to cm^{-1}. I multiply by $\hbar$ to convert to an energy equivalent. Hence, to two digit precision,* I have

$$1/\tau = 2.6{\times}10^{13}\ \mathrm{s}^{-1} == 4.2\ \mathrm{THz} == 140\ \mathrm{cm}^{-1} == 17\ \mathrm{meV}/\hbar.$$

4.2.5 The real and imaginary parts of σ

The conductivity may be split into its real and imaginary parts by multiplying by unity, written as $(1 + i\omega\tau)/(1 + i\omega\tau)$, to get

$$\sigma_1 = \frac{ne^2\tau/m}{1+\omega^2\tau^2} = \frac{\sigma_{dc}}{1+\omega^2\tau^2} \qquad (56a)$$

$$\sigma_2 = \omega\tau\frac{ne^2\tau/m}{1+\omega^2\tau^2} = \frac{\omega\tau\sigma_{dc}}{1+\omega^2\tau^2}. \qquad (56b)$$

Figures 8 and 9 show the real and imaginary parts of the conductivity of a Drude metal. The first show these functions on linear scales and the second on logarithmic scale. The parameters are $\sigma_{\mathrm{dc}} = 6.2 \times 10^5\ \Omega^{-1}\mathrm{cm}^{-1}$ and $1/\tau = 140\ \mathrm{cm}^{-1}$ (19 meV). These are meant to mimic the optical conductivity of silver.

The real part, σ_1, is constant at low frequencies but, as frequency increases, it rolls off, falling as $1/\omega^2$ at frequencies above $1/\tau$. The imaginary part starts off linear in ω and has a maximum at $\omega = 1/\tau$ (where $\sigma_1 = \sigma_2$) and falls as $1/\omega$at high frequencies.

The shape of the Drude conductivity curve is always the same: a peak in $\sigma_1(\omega)$ at zero frequency and a hump-like feature† in $\sigma_2(\omega)$. Two quantities control the details. One is the area under the $\sigma_1(\omega)$ curve, which (as I shall show) is proportional to ne^2/m and hence is often called the "spectral weight." The second is $1/\tau$, the full width at half maximum.

---

* Probably more precision than the calculation deserves.

† A derivative actually, or rather the negative of the derivative.

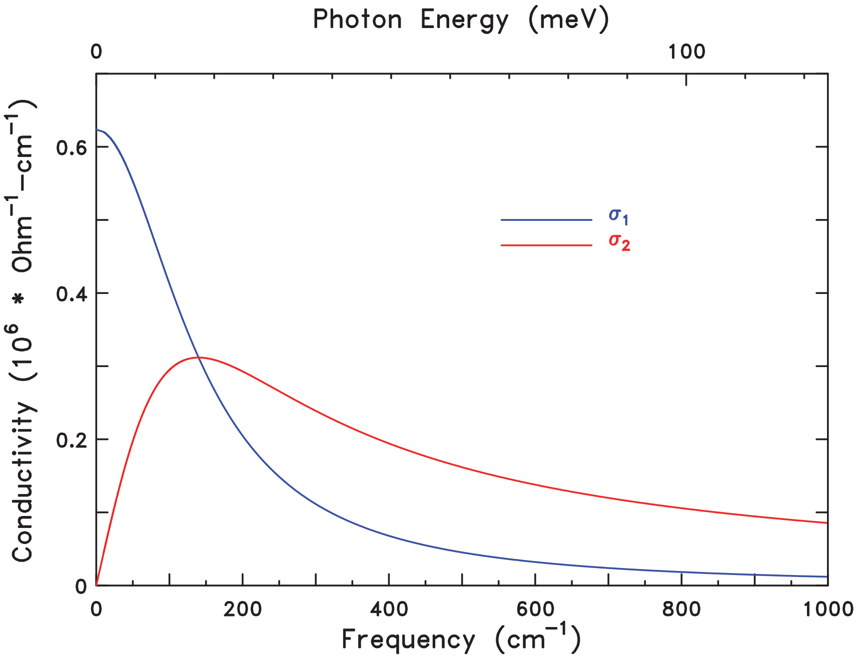

Fig. 8. Real and imaginary parts of the conductivity of a Drude metal like silver.

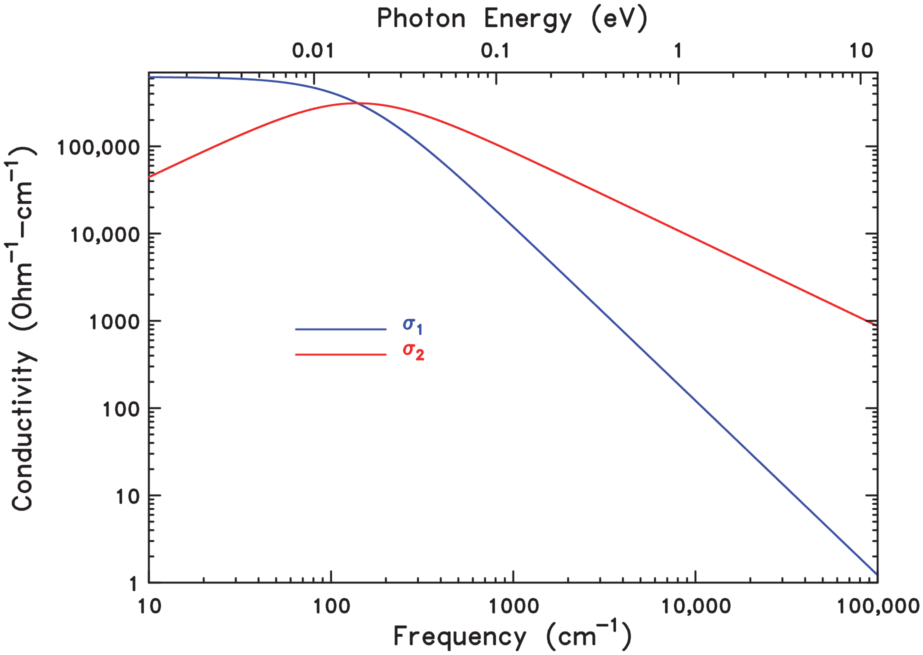

Fig. 9. Real and imaginary parts of the conductivity of a Drude metal like silver. The curves are plotted on a log-log scale.

4.2.6 Complex dielectric function, $\epsilon(\omega)$

I've already written Eq. 26: $\epsilon = 1 + 4\pi i\sigma/\omega$. But the Drude conductivity, Eq. 54, considers only the conduction electrons in the metal. As discussed on page 25, there is also the polarization of the ion cores. So, I'll replace the above equation with $\epsilon = \epsilon_c + 4\pi i\sigma/\omega$. Substituting Eq. 54 and doing a little algebra, I get

$$\epsilon(\omega) = \epsilon_c - \frac{4\pi ne^2/m}{\omega^2 + i\omega/\tau} \tag{57}$$

The Drude contribution comes in with a minus sign in the real part (but a positive imaginary part).

Look at the numerator in the second term: $4\pi ne^2/m$. This quantity has dimension of [esu^2/cm^3-g]. But Coulomb's law in cgs-Gaussian units, $F = q_1q_2/r^2$, tells me that [esu^2 = dyne-cm^2 = g-cm^3/s^2] with [dyne = g-cm/s^2], making the units of $4\pi ne^2/m$ [1/s^2], a frequency squared.* So I'll define the plasma frequency, ω_p to be

$$\omega_p = \sqrt{\frac{4\pi ne^2}{m}}, \tag{58}$$

and write

$$\epsilon(\omega) = \epsilon_c - \frac{\omega_p^2}{\omega^2 + i\omega/\tau}. \tag{59}$$

Moreover, Eq. 55 becomes $\sigma_{\text{dc}} = \omega_p^2\tau/4\pi$.

4.2.7 The real and imaginary parts of ϵ

The real and imaginary parts of ϵ are obtained by multiplying the top and bottom of the second term by $\omega - i/\tau$ and doing a small amount of algebra:

$$\epsilon_1 = \epsilon_c - \frac{\omega_p^2}{\omega^2 + 1/\tau^2} \tag{60a}$$

$$\epsilon_2 = \frac{\omega_p^2}{\omega\tau(\omega^2 + 1/\tau^2)}. \tag{60b}$$

These functions are plotted in Fig. 10. The parameters are $\omega_p = 73,000\,\text{cm}^{-1}$ (9.0 eV), $1/\tau = 140\,\text{cm}^{-1}$ (17 meV) and $\epsilon_c = 5.1$. These are meant to mimic the dielectric function of silver.The real part is negative and large in magnitude at low frequencies, The low frequency limit of Eq. 60a is $\epsilon_1 \approx -\omega_p^2\tau^2$ (neglecting ϵ_c, which is almost always appropriate). As the frequency becomes large with respect to $1/\tau$, the dielectric function becomes that of pure free carriers, $\epsilon_1 \approx \epsilon_c - \omega_p^2/\omega^2$

The zero of ϵ_1 occurs at $\omega_p/\sqrt{\epsilon_c}$. I'll call this frequency the screened plasma frequency.

The imaginary part goes as $1/\omega$ at low frequencies, diverging as $\omega \to 0$.† At high frequencies, $\epsilon_2 \sim 1/\omega^3$.

---

* I knew this of course, because ϵ is just a number and the denominator of Eq. 57 has units of frequency squared.

† This behavior makes ϵ_2 not a very useful function for conducting materials.

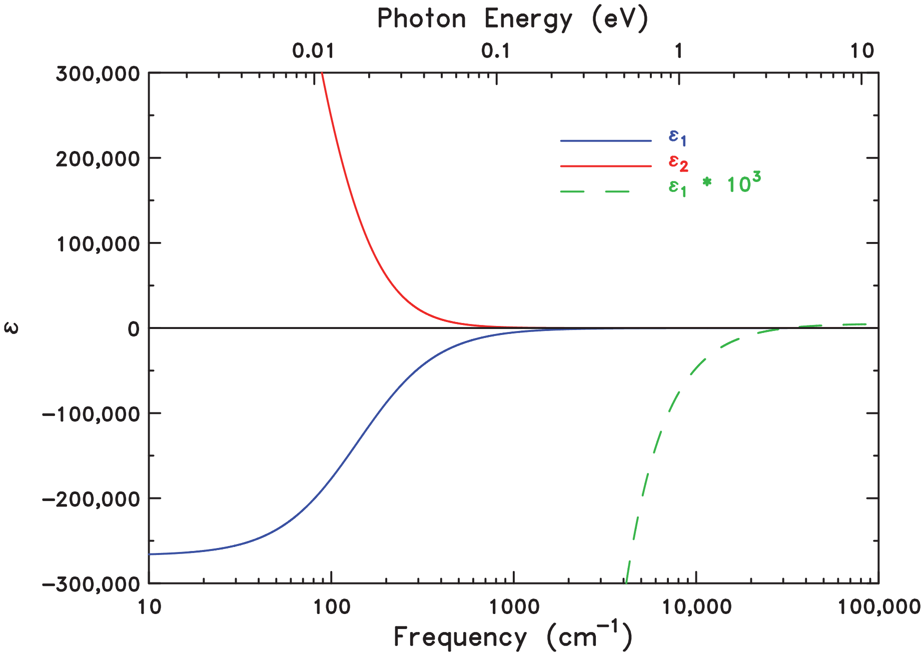

Fig. 10. Real and imaginary parts of the dielectric function of a Drude metal. The dashed line shows ϵ_1 multiplied by 1000, showing the zero crossing at 32,000 cm^{-1} (4 eV).

4.2.8 Some more numbers

Plasma frequencies are typically in the visible to ultraviolet for metals, but can be in the infrared or far infrared for low-density metals and for doped semiconductors. Using the same parameters as on page 28 ($n = 5.9 \times 10^{22}$ cm^{-3}, $e = 4.8 \times 10^{-10}$ esu, and $m = 9.1 \times 10^{-28}$ g), I calculate $\omega_p = 1.4 \times 10^{16}$ s^{-1}. If I convert this to a frequency, I get

$$\omega_p = 2.2{\times}10^{15} \text{ Hz} == 72{,}000 \,\text{cm}^{-1} == 9.0 \text{ eV}/\hbar.$$

Using the value for the mean free time estimated on page 28, $\tau = 2{\times}10^{-14}$, the low frequency value for ϵ_1 is $-\omega_p^2\tau^2 = -250,000$.

4.2.9 Factors that affect the optical conductivity

Three materials-dependent quantities come into the conductivity of a metal: the carrier density n, the dynamical or effective mass m, and the mean free time τ. I will show in a moment (page 36) that the first two combine to give the plasma frequency; the result is that only n/m can be extracted.*

The mean free time in the Drude model represents the time between electron collisions (or more correctly,[7] the time the electron has before it suffers its next collision) the electron moves under the influence of the field between the collisions and contributes to current flow. I'll say a lot more about this in Chapter 7. For now, I'll assert that the mean free time is affected by the purity and crystal quality of the metal and by the temperature. The mean

---

* If I measure the Hall effect or other magneto-optical parameters such as cyclotron frequency, I can determine n or m and then the optical parameters will provide the other.

free time is longer when purity is high and temperature is low. Empirically. these scattering rates are independent and add, following Matthiessen's rule:

$$\frac{1}{\tau} = \frac{1}{\tau(T)} + \frac{1}{\tau_i}$$

where $\tau(T)$ is the temperature-dependent scattering from thermal excitations in the metal and τ_i is the temperature-independent scattering from impurities or other defects. $\tau(T)$ has a complex temperature dependence, but at high temperatures many metals have $1/\tau(T) \sim T$; at low temperatures the power law is stronger: T^2, T^3, or T^5. At the lowest temperatures, the scattering rate (and the resistivity) saturates at the impurity-limited value. The ratio between the room-temperature resistivity and the low-temperature resistivity is called the residual resistivity ratio (RRR) and is a measure of purity. It is also a measure of relative scattering times, because the carrier density and mass in simple metals do not change much with temperature.*

As an example, Fig. 11 shows a simulation at three temperatures of conductivity (left panel) and dielectric constant (right panel) for a metal of modest purity, RRR = 10. (I choose this value rather than 100 or 1000, values which are possible in high-purity metals, to avoid entering the anomalous skin effect regime. See Chapter 12.) Relative to room temperature, the mean free time is assumed 3× longer for the 100 K calculation and 10× longer for the

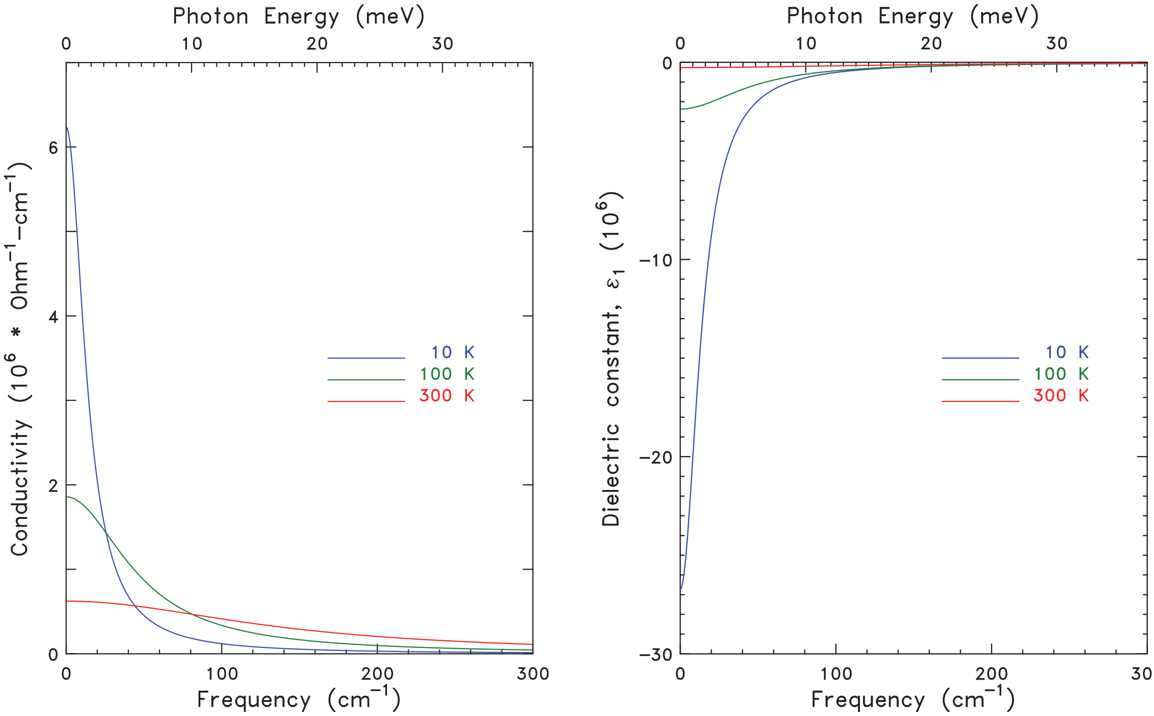

Fig. 11. Left: Optical conductivity of a hypothetical metal with modest purity. Curves are shown for room temperature, 100 K, and low temperatures. Right: Dielectric function of the same metal.

---

[*] There are other systems with other behaviors. Semiconducting materials have a much stronger temperature dependence to the carrier density than to the mean free time. Heavy-Fermion metals have a huge increase in carrier effective mass at low temperatures.

10 K calculation. The dc conductivity goes up and the width of the peak becomes narrower with decreasing temperature; the areas under the three conductivity curves stay the same. The dielectric constant follows $-\omega_p^2/\omega^2$ at high frequencies; with decreasing temperature the $1/\omega^2$ behavior continues to lower frequencies. (The curve marked 100 K occupies only a bit of the panel here; it is identical to the ϵ_1 shown in Fig. 10.) The limiting low frequency value of $-\omega_p^2\tau^2$ increases 100× as the mean free time becomes 10 times longer.

4.2.10 Fields and currents

This is a good time to take a look at the fields and currents in the metal. Recall that the physically observable field is the real part of the complex field. Hence, if I could put probes in and measure the field, I would see

$$\mathbf{E} = \mathrm{Re}[E_0\hat{\boldsymbol{y}}e^{-i\omega t}] = E_0\hat{\boldsymbol{y}}\cos(\omega t)$$

and if I could put a current meter in I would see

$$\mathbf{j} = \mathrm{Re}[\sigma\hat{\boldsymbol{y}}\mathbf{E}e^{-i\omega t}] = \sigma_1 E_0\hat{\boldsymbol{y}}\cos(\omega t) + \sigma_2 E_0\hat{\boldsymbol{y}}\sin(\omega t)$$

so if $\sigma_1 \gg \sigma_2$ (low frequencies), the current is in-phase with the field and their time average $\langle \mathbf{j}\cdot\mathbf{E}\rangle$ equals half of the product of peak values. This is the Ohmic loss in the metal. If, however, $\sigma_1 \ll \sigma_2$ (high frequencies), the field follows the cosine and the current mostly follows the sine; the time average of their product is nearly zero. The electrical analogy is a nearly ideal inductor.

4.2.11 The refractive index

It is the refractive index ($N \equiv \sqrt{\epsilon}$) which most directly governs wave propagation in the medium. Figures 12 and 13 show the real and imaginary parts of the refractive index of a Drude metal. These are calculated using the same parameters as for the dielectric function.

There are several regimes to discuss. At low frequencies, n and κ are both large and relatively close in magnitude. (See Fig. 12 for the low-frequency region.) At very low frequencies, where $\epsilon_1 \ll \epsilon_2$ (i. e., where $\omega \ll 1/\tau$), Eq. 35 gives $n \approx \kappa \approx \sqrt{\epsilon_2/2}$. See Figs. 10–13 for the behavior of ϵ and N. For the Drude metal, then,

$$n \approx \kappa \approx \sqrt{\frac{\omega_p^2\tau}{2\omega}} = \sqrt{\frac{2\pi\sigma_{\mathrm{dc}}}{\omega}} \quad \text{when } \omega \ll \frac{1}{\tau} \tag{61}$$

Both n and κ fall with increasing frequency, with n falling faster than κ.

At intermediate frequencies, the refractive index of a metal can be less than one, sometimes much less than one. This behavior is seen in the calculation shown in Figs. 12 and 13 once $\omega \gg 1/\tau$ up until the screened plasma frequency. In this regime, the phase velocity, $v = c/n$ is bigger than c. I leave it to the reader to show that the group velocity is smaller than c. The waves in the material are of course strongly damped; with the electric field penetration length, $\delta = \lambda_0/2\pi\kappa$, smaller than the vacuum wavelength.

At a certain frequency (32,000 cm^{-1} in this case) κ falls rapidly and n jumps up to a value bigger than 1. This jump occurs exactly where ϵ crosses zero, the screened plasma frequency. Above this frequency, $n \approx \sqrt{\epsilon_c}$ and κ is small.

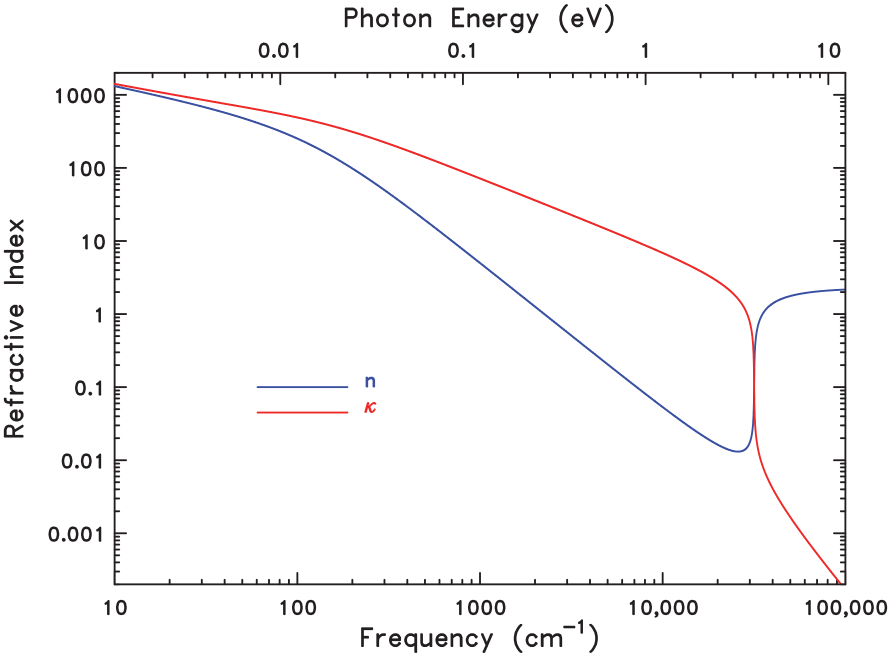

Fig. 12. Real and imaginary parts of the refractive index of a Drude metal. Data are shown on a log-log scale.

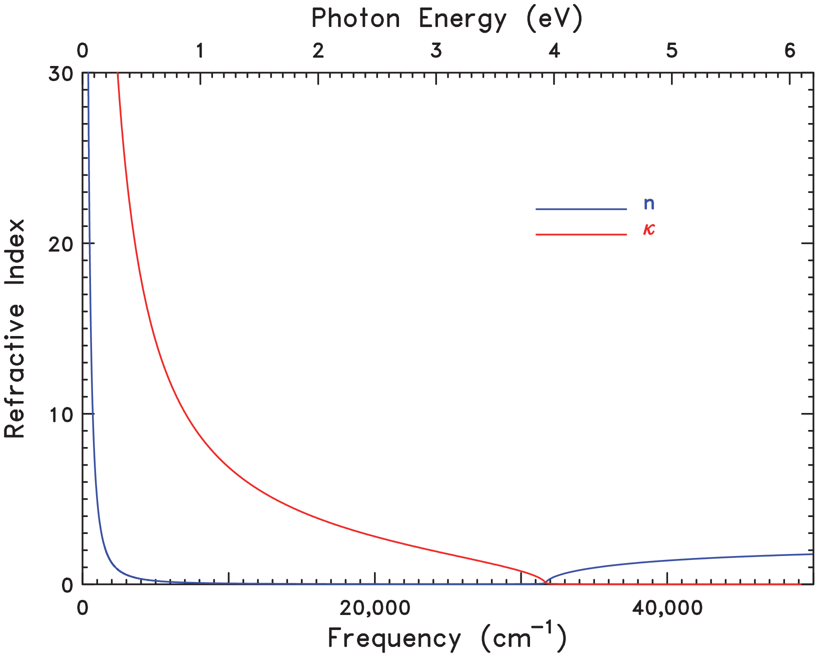

Fig. 13. Real and imaginary parts of the refractive index of a Drude metal. Data are shown on linear scale.

The behavior of N is easy to understand in the extreme case when I can neglect $1/\tau$. (Even if τ is finite, this is a good first approximation when $\omega \gg 1/\tau$.) Then

$$\epsilon = \epsilon_c - \frac{\omega_p^2}{\omega^2}$$

and

$$N = \begin{cases} 0 + i\sqrt{\dfrac{\omega_p^2}{\omega^2} - \epsilon_c}, & \text{for } \omega < \omega_p/\sqrt{\epsilon_c} \\ \sqrt{\epsilon_c - \dfrac{\omega_p^2}{\omega^2}} + i0, & \text{for } \omega > \omega_p/\sqrt{\epsilon_c} \end{cases}$$

Below the screened plasma frequency, this approximation gives a pure imaginary value $i\kappa$ for N. The real part is zero. Above the screened plasma frequency, $N = n$ is purely real and has the high-frequency limit of $n = \sqrt{\epsilon_c}$. Metals should be transparent above $\omega_p/\sqrt{\epsilon_c}$, and some are.

4.2.12 Reflectance

Once I have N, I can calculate the reflectance $\mathscr{R}$ from Eq. 44. At low frequencies, both parts of N are very large, so the difference between n and 1 is large. In this case, it is sometimes convenient to write Eq. 44 as

$$\mathscr{R} = 1 - \frac{4n}{(n+1)^2 + \kappa^2}. \tag{62}$$

Note that although Eq. 62 might look like an expansion, it is exact. The calculated reflectance, using parameters identical to those for Figs. 8–10, 12, and 13, is shown in Fig. 14 as the red curve, marked "300 K."

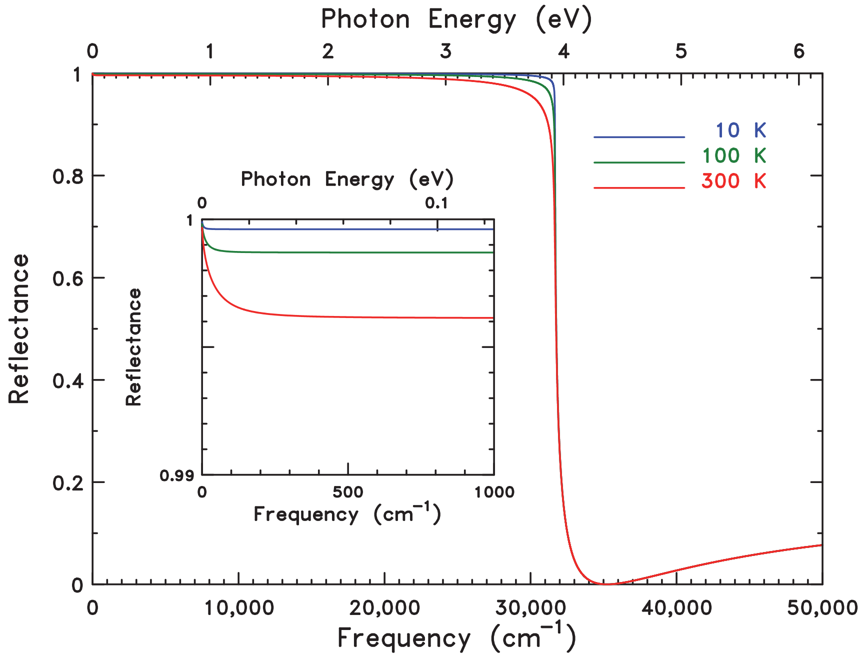

Fig. 14. Normal incidence reflectance of a Drude metal.

A good approximation in the far infrared, where $n \approx \kappa \gg 1$, is to write $n + 1 \approx n$ and then, using Eqs. 61 and 62, obtain the so-called Hagen-Rubens[8] equation

$$\mathscr{R} \approx 1 - \sqrt{\frac{2\omega}{\pi\sigma_{\rm dc}}} \tag{63}$$

The reflectance is 100% at $\omega = 0$ and falls in accord with Eq. 63 with increasing frequency.

Once ω is bigger than $1/\tau$ the reflectance flattens out. This behavior is shown in the inset of Fig. 14, where one can see the Hagen-Rubens regime at the left and the flattening out at higher frequencies. The reflectance is above 99.5% in each case.

At high frequencies, the reflectance falls rapidly from a value close to unity to a value close to zero. This drop occurs around 30,000 cm^{-1} (3.5 eV) for the parameters used for Fig. 14. This "plasma edge" occurs at the same frequencies where the rapid fall of κ and rise of n occur in Fig. 13. There is a minimum value of the reflectance, corresponding to the conditions* $n \approx 1$. (The approximation becomes an equality when $\kappa \ll 1$.)

The simulations for 100 K and 10 K (actually τ increased by 3× and 10×) are shown also in Fig. 14. As the mean free time becomes longer, the turnover between the flat and high reflectance an the steep plasma edge becomes sharper and sharper.

4.3 The plasma frequency

I called $\omega_p = \sqrt{4\pi ne^2/m}$ the plasma frequency. What does this mean? There are no peaks in σ_1 (or ϵ_2) at this frequency. The dielectric function changes sign at $\omega = \omega_p/\sqrt{\epsilon_c}$ but does so smoothly.

To see why the name is what it is, consider a slab of metal of thickness L. The metal contains positive ions bound to their lattice positions by the interatomic forces in the metal. Permeating the entire slab is the gas of conduction electrons. The metal is electrically neutral, with a number density of electrons of $n_e = N_{\text{electrons}}/V$ and a number density of ions of $n_i = N_{\text{ions}}/V$, where V is the volume of the slab. For conceptual simplicity, let me assume that the ions are singly charged,† so that the number of charges is $N_{\text{electrons}} = N_{\text{ions}} \equiv N$ and the number density is $n_e = n_i \equiv n$

Now, displace the electrons an amount z_0 in the upward ($\hat{\boldsymbol{z}}$) direction. Take $z_0 \ll L$. Figure 15 shows the slab with the displaced electrons. Because every electron is displaced by the same amount, there is a net dipole moment in the slab. The dipoles all point downward, from the negative electrons to the positive ions. The dipoles are $\mathbf{p} = -QL\hat{\boldsymbol{z}}$ where Q is the magnitude of the charge on each surface. This quantity is the product of the number density of charges n, their individual charge e and the volume of uncompensated charge Az_0 with A the area of the slab surface: $Q = neAz_0$. Hence the dipole moment/unit volume, $\mathbf{P} = \mathbf{p}/AL$, is

$$\mathbf{P} = -\frac{neAz_0L}{AL}\hat{\boldsymbol{z}} = -nez_0\hat{\boldsymbol{z}}.$$

---

* The exact location of this "plasma minimum" is a balance of the frequency where $n - 1 \sim 0$ and the value of κ near this frequency. To have a deep minimum, κ must have a relatively small value. It is *not* the same as the frequency where $\epsilon_1 = 0$ although these two frequencies are close to each other. In the case where $\omega_p \gg 1/\tau$, then I can write $\epsilon(\omega) = \epsilon_c - \omega_p^2/\omega^2$. With ϵ_c a real constant, the zero of $\epsilon(\omega)$ is at $\omega_p/\sqrt{\epsilon_c}$ and the reflectance minimum is when $n = \sqrt{\epsilon} = 1 = 1^2 = \epsilon$, making the minimum occur at $\omega_p/\sqrt{\epsilon_c - 1}$. If $\epsilon_c = 1$, the minimum is at infinity. If ϵ_c has dispersion (and an imaginary part), the minimum is still near where $\text{Re}(\epsilon_c(\omega)) = 1$ but the exact location can be pushed up or down depending on the functional form of $\text{Im}(\epsilon_c)$.

† This assumption is by no means essential; a solid of trivalent ions plus electrons would have $N_{\text{electrons}} = 3N_{\text{ions}}$.

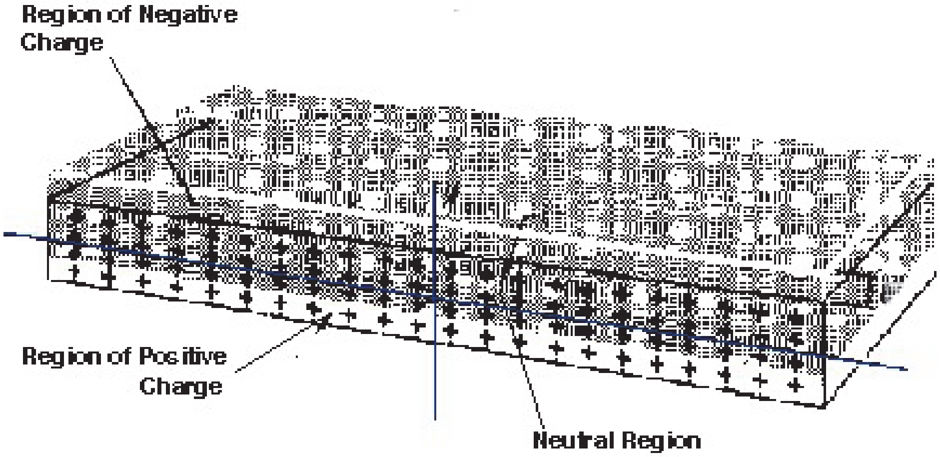

Fig. 15. A slab of free electrons and fixed positive ions. Each electron has been displaced upwards by a small amount z_0, creating a region of negative charge on the top of the slab and a region of positive charge on the bottom. The central region remains overall neutral.[9]

At the same time, the uncompensated surfaces generate an electric field **E**, just as in the case of a charged capacitor. **E** points upwards, from the positive charge on the bottom to the negative charge on top. Gauss' law tells me that $E = 4\pi Q/A\hat{\mathbf{z}} = 4\pi nez\hat{\mathbf{z}} = -4\pi\mathbf{P}$.

The field exerts a force on the electrons, $\mathbf{F} = -e\mathbf{E} = m\ddot{\mathbf{z}}$. The equation of motion of the electrons is thus

$$m\ddot{\mathbf{z}} = -4\pi ne^2\mathbf{z}.$$

The solution to this equation is that of a harmonic oscillator, $\mathbf{z} = \mathbf{z}_0 e^{-i\omega_0 t}$ with ω_0 the resonant frequency of the oscillator. Substitution gives

$$-m\omega_0^2\mathbf{z}_0 = -4\pi ne^2\mathbf{z}_0.$$

So, I get the frequency of the free oscillations to be

$$\omega_0 = \sqrt{\frac{4\pi ne^2}{m}} \equiv \omega_p.$$

This equation is identical to Eq. 58, the plasma frequency I defined when calculating the Drude dielectric function of the metal. This plasma oscillation is a normal mode of the electrons in the metal slab. The frequency of oscillation is the plasma frequency.

If the atomic core polarizability is included in the problem, it becomes analogous to a capacitor with a dielectric material between the plates. The Gauss' law problem finds **D**; the electric field **E**, dipole moment/unit volume **P**, and the restoring force on the electrons are all reduced by a factor of ϵ_c, and the resonant frequency becomes

$$\omega_0 = \frac{\omega_p}{\sqrt{\epsilon_c}}$$

A measurement of reflectance over the appropriate frequency range can provide me with the plasma frequency. There are four methods in common use. I can fit the reflectance to a model dielectric function that includes a Drude term; ω_p is a parameter from the fit. I can carry out Kramers-Kronig analysis (Chapter 9) and analyze the $1/\omega^2$ term in ϵ_1. Alternatively, a sum-rule analysis of $\sigma_1(\omega)$ (also Chapter 9) provides an estimate of ω_p. Finally, I can measure the reflection and transmission of a thin film (Chapter 6) and extract the optical constants and plasma frequency from those data. Each method then gives the ratio of carrier density to effective mass, n/m.

Table 1 gives a calculated plasma frequency for a number of elements. The assumptions are: (1) the the carrier density is the atomic density times the valence, and (2) the carriers have the free-electron mass.* These assumptions are excellent for the monovalent alkali and noble metals and poor for multivalent metals. The last two columns show the numbers for two carrier densities a factor of 10 apart, taking the mass as the free electron mass. The frequency for any other density can be calculated by multiplying by the square root of the carrier density ratio.

Table 2. Calculated plasma frequencies for a variety of metals.

Valence	Metal	n 10^{22} cm^{-3}	ω_p 10^{15} s $^{-1}$	ν_p cm^{-1}	E_p eV	c/ω_p nm
1	Li	4.7	12.23	64,800	8.03	24.5
	Na	2.65	9.18	48,700	6.04	32.7
	K	1.4	6.67	35,400	4.39	44.9
	Rb	1.15	6.05	32,000	3.97	49.6
	Cs	0.91	5.38	28,500	3.53	55.7
	Cu	8.45	16.40	86,900	10.77	18.3
	Ag	5.85	13.64	72,300	8.96	22.0
	Au	5.9	13.70	72,600	9.00	21.9
2	Be	24.2	27.75	147,200	18.25	10.8
	Mg	8.6	16.54	87,700	10.87	18.1
	Ca	4.6	12.10	64,100	7.95	24.8
	Sr	3.56	10.64	56,400	6.99	28.2
	Ba	3.2	10.09	53,500	6.63	29.7
	Zn	13.1	20.42	108,300	13.43	14.7
	Cd	9.28	17.19	91,100	11.29	17.5
3	Al	18.06	23.97	127,100	15.76	12.5
	Ga	15.3	22.07	117,000	14.51	13.6
	In	11.49	19.12	101,400	12.57	15.7
4	Pb	13.2	20.50	108,700	13.48	14.6
	Sn(w)	14.48	21.47	113,800	14.11	14.0
	$1{\times}10^{21}$	0.1	1.78	9460	1.17	168
	$1{\times}10^{22}$	1	5.64	29,900	3.71	53.2

---

* The charge is e, in esu!

Metal plasma frequencies are typically rather high, in the ultraviolet for ordinary metals. Table 3 lists a few experimental values.

Table 3. Experimental plasma frequencies

Material	ω_p s^{-1}	ν_p cm^{-1}	E_p eV
K	6×10^{15}	31,000	3.9
Na	8×10^{15}	48,000	6.0
Al	2.4×10^{16}	140,000	18.
"Si"	2.6×10^{16}	160,000	20.

4.4 Lorentz model: Interband absorption in semiconductors and insulators

The Lorentz model is designed for insulators, or, at least for electrons which are bound by some force to an atom or ion in the solid. I'll also use it for the contribution to the dielectric function from optically active vibrational modes (phonons). This latter role for the model begins on page 49.

I will view the potential for these bound electrons as harmonic, i. e., the electrons are attached to their equilibrium positions by a spring. The spring constant will be the same for all the electrons in a particular orbital. On the one hand, the harmonic potential (a spring) is quite different from the Coulomb potential that—as is well known—governs the electronic properties of a atom, molecule, or solid. On the other hand, whatever the potential is, I know that it will look harmonic for small displacements from equilibrium. My assumption of linear response suggests that the displacements are indeed small.

4.4.1 Dilute limit

To begin, I'll consider the dilute limit, just as I did on page 25. The reason I want to start with the dilute limit is that an electric field applied* to the molecule polarizes the molecule. Near the location of each polarized molecule, there is also a dipolar field[2]

$$\mathbf{E}_d(\mathbf{r}) = \frac{3(\mathbf{p}\cdot\hat{\boldsymbol{r}})\hat{\boldsymbol{r}} - \mathbf{p}}{r^3} \tag{64}$$

where $\mathbf{r}$ is the distance from the origin to the field point, $\hat{\boldsymbol{r}} = \mathbf{r}/r$, $\mathbf{p} = \alpha_e \mathbf{E}_{\rm loc}$ and $\mathbf{E}_{\rm loc}$ is the local field at the position of the molecule. I want the strength of these dipolar fields to

---

* I think it is easy to be confused about the meaning of the "applied field." In continuum electrodynamics it of course means the electric field applied to the medium from external sources (lamps, lasers, generators). From the point of view of a microscopic picture of the medium as a body consisting of polarizable entities, it is of course the field seen by the polarizable entities. These fields are not necessarily the same. The second one is the superposition of the external field and the fields emitted by all the other polarizable entities. So I need to be careful with the use of "applied field;" perhaps "external field" and "local field" are better terms. I also need to be aware that when I send light into the material from an external source, some of it will be reflected at the surfaces of the material and some will be absorbed on its way to the point where I want to evaluate the field. I can measure the total field at any point by inserting a (small) test charge at that point and seeing what force the test charge feels. Dilute media are those in which the external field *is* the applied field.

be negligible at the other nearby molecules so as to make the local field (the field responsible for the polarization) equal to the field at the molecule from the external electromagnetic wave. The dipolar field varies in space, looping from the + end of the dipole around to the − end. Its magnitude is of order $E_d \approx (\alpha_e/r^3)\mathbf{E}_{\text{applied}}$. I'll take for the polarizability α of the molecule the polarizability of a conducting sphere, $\alpha_e \sim a_0^3$ with a_0 the sphere's radius. To keep the field from the polarized molecule smaller than, say, 0.1% of the external field, the nearby molecules have to be separated by ~10× the molecular size, whereas in close packing the separation is of order the molecular size. The density is 1000 times smaller than the close-packed density; hence, the system is a gas.

The molecule has zero dipole moment without an external field. The electron is considered to be bound to the position where $\mathbf{p} = 0$ by a harmonic force.

$$\mathbf{F}_h = -K\mathbf{x} = -m\omega_0^2\mathbf{x}$$

where K is the spring constant, m is the electronic mass,* and $m\omega_0^2$ is a complicated way to write the spring constant.†

A second force acting on the electron is a damping force proportional to the velocity, just as in the case of the Drude model, $\mathbf{F}_v = -\Gamma\mathbf{v} \equiv -m\gamma\dot{\mathbf{x}}$, where $\gamma = 1/\tau$ is the damping rate. This is formally identical with the Drude model for a metal but the underlying physics is a bit different. In a metal the damping is associated with collisions of the Fermi surface electrons with impurities, phonons, etc. In semiconductors and insulators there is no Fermi surface; the chemical potential lies in a forbidden band gap between valence and conduction band. The damping may be considered as due to collisions of photoexcited carriers with impurities, phonons, other electrons, with excited state lifetimes before relaxation to the ground state, and with the emission of excitations like phonons, with a corresponding scattering of the electron. The remaining force is the force due to the external electric field, which I'll take as $\mathbf{F}_e = -e\mathbf{E} = -e\mathbf{E}_0 e^{-i\omega t}$. Newton's law is then written as

$$m\ddot{\mathbf{x}} = -m\omega_0^2\mathbf{x} - m\gamma\dot{\mathbf{x}} - e\mathbf{E}_0 e^{-i\omega t}. \tag{65}$$

In linear response, the location of the electron is $\mathbf{x} = \mathbf{x}_0 e^{-i\omega t}$. Taking the first derivative (bringing down $-i\omega$) and the second derivative (bringing down $-\omega^2$), Eq. 65 for the position of the electron becomes

$$m\omega_0^2\mathbf{x} - i\omega m\gamma\mathbf{x} - m\omega^2\mathbf{x} = -e\mathbf{E}.$$

or

$$\mathbf{x} = \frac{-e/m}{\omega_0^2 - \omega^2 - i\omega\gamma}\mathbf{E}. \tag{66}$$

Clearly, $\mathbf{x}$ is resonant at ω_0; it diverges at this frequency if $\gamma = 0$.

---

[*] It is the reduced mass in fact, but the ion mass is big enough that it can be ignored. I'll take m to be the effective mass—which can be quite different from the free-electron mass—in the solid.

[†] I write $-m\omega_0^2$ rather than K because I know how it will work out, anticipating that the resonant frequency $\omega_0 = \sqrt{K/m}$. I have solved the harmonic oscillator problem in the past.

With this displacement, the dipole moment is $\mathbf{p} = -e\mathbf{x}$, and the polarizability is

$$\alpha_e = \frac{e^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma} \tag{67}$$

The dipole moment/unit volume $\mathbf{P} = n\mathbf{p}$, where n is the number of molecules per unit volume. Thus

$$\mathbf{P} = \frac{ne^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}\mathbf{E} = \chi_e \mathbf{E}$$

with χ_e the dielectric susceptibility of the semiconductor or insulator. I then write $\epsilon = 1 + 4\pi\chi_c + 4\pi\chi_e$ where χ_c is the core polarizability from page 25 ($\epsilon_c \equiv 1 + 4\pi\chi_c$) and find

$$\epsilon(\omega) = \epsilon_c + \frac{4\pi ne^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}. \tag{68}$$

Eq. 68 contains a "plasma frequency" which I'll call $\omega_{pe} \equiv \sqrt{4\pi ne^2/m}$. The e subscript stands for "electron, bound." With this definition I get the final form of the Lorentz dielectric function for dilute gases.

$$\epsilon(\omega) = \epsilon_c + \frac{\omega_{pe}^2}{\omega_0^2 - \omega^2 - i\omega\gamma}. \tag{69}$$

As $\omega \to 0$, the dielectric function becomes the static dielectric constant,

$$\epsilon(0) = \epsilon_c + \frac{\omega_{pe}^2}{\omega_0^2}$$

Unlike a metal, the dielectric constant of this nonconducting gas is positive.

Values of this dielectric constant for common gases at STP, and for visible wavelengths (which are below the first electronic transitions in the molecule and above any molecular vibration or rotational modes) are given in Table 4. The values are not much different from unity.

Table 4. "Optical" dielectric constant for some gases

Molecule	ϵ_1
He	1.000068
O_2	1.00054
HCl	1.0009
CS_2	1.003

4.4.2 The depolarizing field

As the density increases, the difference between the local field and the external field grows rapidly. I'll write

$$\mathbf{p} = \alpha_e \mathbf{E}_{\text{loc}}.$$

where α_e is the polarizability and $\mathbf{E}_{\text{loc}}$ is the value of the field at the molecule, including the external field and the field from the neighboring dipoles, in principle at distances all the way to infinity.

This problem was studied in the 19th century. The solution is known as the Lorentz-Lorenz model or the Clausius-Mossotti model. I'll start by writing the electric field at some point as a sum of three terms

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_{\text{app}} + \mathbf{E}_{\text{neighbor}} + \mathbf{E}_{\text{medium}}.$$

In this equation, $\mathbf{E}_{\text{app}} \equiv \mathbf{E}$ is the external (light) field. $\mathbf{E}_{\text{neighbor}}$ is the sum of the fields from the nearby polarized molecules, and $\mathbf{E}_{\text{medium}}$ is the contribution from the uniformly polarized medium far from the molecule.

To calculate the field from the polarization of the medium, one carves out a hollow spherical cavity in the polarized medium centered on the polarizable molecule being considered and computes the field inside the cavity. The uniform polarization $\mathbf{P}$ outside the cavity leads to a surface charge density $\Sigma = \mathbf{P} \cdot \hat{\boldsymbol{n}}$. This charge density varies as $\cos(\theta)$, and leads to a uniform field in the cavity. This field, found in your electricity and magnetism textbook,^{2} is

$$\mathbf{E}_{\text{dipole}} = \frac{4\pi}{3}\mathbf{P} \tag{70}$$

For a crystalline material, the fields of the individual dipoles in the cavity, given by Eq. 64, all contribute to the field at the center. The individual fields point in a set of specific directions which depend on the coordinates of the dipoles. Their sum depends on the crystal symmetry. For a cubic system, the fields from the neighbors add to zero. Essentially the fields from the two dipoles behind and in front of the origin are canceled by the oppositely directed and half-strength fields from the four on either side, above, and below. This cancellation continues as one moves out from the center.*

Consequently

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_{\text{app}} + \frac{4\pi}{3}\mathbf{P}$$

where

$$\mathbf{P} = \frac{1}{V}\sum_i \mathbf{p}_i = \frac{1}{V}\sum_i \alpha_{ei}\mathbf{E}_{\text{loc}} = n\alpha_e \mathbf{E}_{\text{loc}}$$

and (now writing $\mathbf{E}_{\text{app}}$ as just $\mathbf{E}$)

$$\mathbf{P} = n\alpha_e(\mathbf{E} + \frac{4\pi}{3}\mathbf{P})$$

or

$$\mathbf{P} = \frac{n\alpha_e}{1 - \frac{4\pi}{3}n\alpha_e}\mathbf{E}$$

As an aside, there is a possibility of a catastrophe here. Consider the denominator in the

---

* Clausius and Mossotti actually used a 'tight" cavity that tightly jacketed the molecule. The crystal structure was brought in by making the cavity an ellipsoid representing the anisotropy of the crystal.

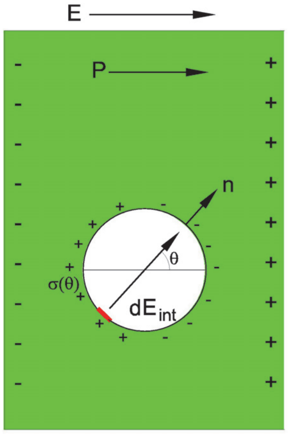

Fig. 16. A sphere is carved out from a uniformly polarized material. There is a surface charge density on the surface of the spherical cavity that is the source of a uniform depolarizing field in the cavity.[10]

equation above. It contains n, the number of atoms per unit volume. Clearly $n = 1/\Omega_0$, with Ω_0 the volume of one molecule. Now the polarizability of a metal sphere is $\alpha_e = r^3$, with r the radius of the metal sphere. Then

$$\frac{4\pi}{3}\alpha_e \sim \frac{4\pi}{3}r^3 = \Omega_0$$

and

$$\mathbf{P} = \frac{n\alpha_e}{1-1}\mathbf{E}$$

is finite when $\mathbf{E} \to 0$. Variations on such a model have been used in the theory of ferroelectrics, solids which have a spontaneous polarization.*

So now, making sure to keep α_e below r^3, I can write the dielectric function with local field corrections, and with a core polarizability, $\epsilon = \epsilon_c + 4\pi\chi_e$ as

$$\epsilon = \epsilon_c + \frac{4\pi n\alpha_e}{1 - \frac{4\pi}{3}n\alpha_e}. \tag{71}$$

This equation can be turned around. Let me take $\epsilon_c = 1$ for simplicity, and then solve for α_e in terms of ϵ.

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi}{3}n\alpha_e \tag{72}$$

in which form it is known as the Clausius-Mossotti equation.† It is approximately obeyed by simple molecular solids.

---

* It is not enough of course to have a polarizability catastrophe; the crystal symmetry must allow the polarization to exist.

† If $\epsilon_c \neq 1$ then this becomes $(\epsilon - \epsilon_c)/(\epsilon - \epsilon_c + 3) = (4\pi/3)n\alpha_e$.

4.4.3 The Lorentz dielectric function in the dense limit

My model for the polarizability, Eq. 67, is unchanged, so I can substitute that expression into Eq. 71:

$$\epsilon = \epsilon_c + \frac{\frac{4\pi ne^2/m}{\omega_0^2-\omega^2-i\omega\gamma}}{1-\frac{4\pi ne^2/3m}{\omega_0^2-\omega^2-i\omega\gamma}}$$

Using the definition of ω_{pe} and doing a little algebra, I get

$$\epsilon = \epsilon_c + \frac{\omega_{pe}^2}{\omega_0^2 - \omega^2 - i\omega\gamma - \frac{1}{3}\omega_{pe}^2} \tag{73}$$

Let me combine the two constants in the denominator into a single frequency, ω_e. This frequency is reduced from the bare resonant frequency on account of local field corrections:

$$\omega_e \equiv \sqrt{\omega_0^2 - \frac{1}{3}\omega_{pe}^2}.$$

With this definition I finally arrive at the Lorentzian dielectric function for solids:

$$\epsilon = \epsilon_c + \frac{\omega_{pe}^2}{\omega_e^2 - \omega^2 - i\omega\gamma} \tag{74}$$

4.4.4 The real and imaginary parts of ϵ

The real and imaginary parts of ϵ are obtained by multiplying the top and bottom of the second term by $\omega_e^2 - \omega^2 + i\omega\gamma$ and doing a small amount of algebra:

$$\epsilon_1 = \epsilon_c + \frac{\omega_{pe}^2(\omega_e^2 - \omega^2)}{(\omega_e^2 - \omega^2)^2 + \omega^2\gamma^2} \tag{75a}$$

$$\epsilon_2 = \frac{\omega_{pe}^2\omega\gamma}{(\omega_e^2 - \omega^2)^2 + \omega^2\gamma^2}. \tag{75b}$$

These functions are plotted in Fig. 17. The parameters are $\omega_{pe} = 50,000\,\mathrm{cm}^{-1}$ (6.2 eV), $\omega_e = 5000\,\mathrm{cm}^{-1}$ (0.62 eV), $\gamma = 300\,\mathrm{cm}^{-1}$ (38 meV) and $\epsilon_c = 3.0$. These are meant to mimic* the optical properties of Ge or perhaps AlGaAs.

The real part is positive at low frequencies. The imaginary part is very small at low frequencies. As the frequency approaches ω_e, the real part of the dielectric function has a derivative-like shape while the imaginary part shows a resonance. The full width at half maximum is $\Delta\omega = \gamma$. Just above the resonance, $\epsilon_1 < 0$. It crosses zero above the resonance. (See the green dashed line, which shows ϵ_1 scaled by 100.)

---

* The mimicry is not very good. The linewidth is too narrow and the static dielectric constant too high.

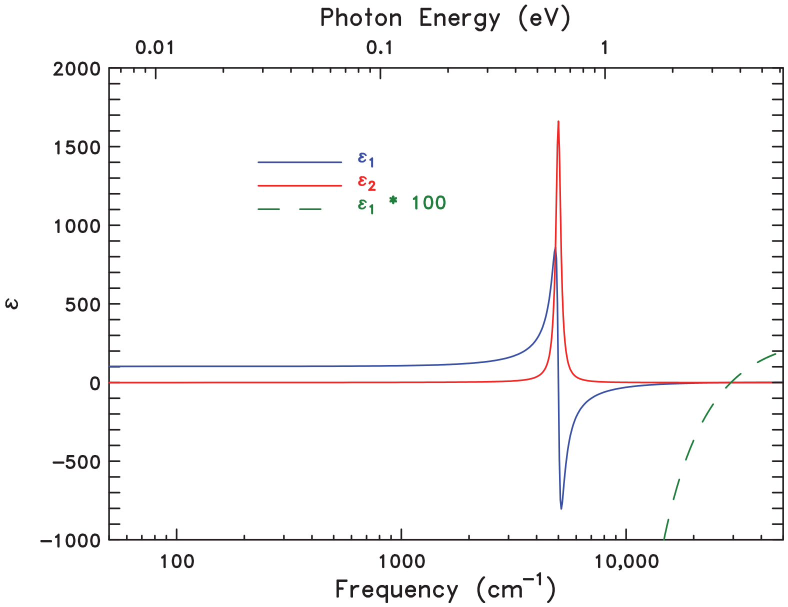

Fig. 17. Real and imaginary parts of the dielectric function of a Lorentzian oscillator.

I'll now calculate the actual "screened plasma" frequency $\tilde{\omega}_p$ where $\epsilon_1(\omega)$ equals zero. Setting the left side of Eq. 75a to zero and doing some considerable algebra, I get

$$\tilde{\omega}_p^2 = \omega_e^2 + \frac{1}{2}\left(\frac{\omega_{pe}^2}{\epsilon_c} - \gamma^2\right)\left(1 + \sqrt{1 - \frac{4\omega_e^2\gamma^2}{(\omega_{pe}^2/\epsilon_c - \gamma^2)^2}}\right).$$

This equation is not very illuminating, so let me evaluate it in the case $\omega_{pe} \gg \omega_e \gtrsim \gamma$. This is the usual case and in it

$$\tilde{\omega}_p^2 = \frac{\omega_{pe}^2}{\epsilon_c} + \omega_e^2.$$

As a general statement, one can say that the zero is pushed up by ω_e and pulled down by γ.

4.4.5 Limiting behavior

Low frequencies: The low frequency limit of Eq. 75a is $\epsilon_1 \approx \epsilon_c + \omega_p^2/\omega_e^2$. The imaginary part goes as ω at low frequencies; the conductivity σ_1 as ω^2.

Mid frequencies: At $\omega = \omega_e$, $\epsilon_1 = \epsilon_c$. The imaginary part is a maximum with $\epsilon_2 = \omega_{pe}^2/\omega\gamma$ and $\sigma_1 = \omega_{pe}^2/4\pi\gamma$. I have identified $\gamma = 1/\tau$ so that this is also: $\sigma_1 = \omega_{pe}^2\tau/4\pi$. This is the DC conductivity of a metal with plasma frequency ω_{pe} and mean free time $1/\tau$.

High frequencies: Above the resonance, ϵ_1 is negative, rising as the frequency increases, following $\epsilon_1 \approx \epsilon_c - \omega_{pe}^2/\omega^2$. At these frequencies, $\epsilon_2 \sim 1/\omega^3$.

4.4.6 Conductivity

It is easy to write $\sigma_1(\omega)$; it is related to ϵ_2 by $\sigma_1(\omega) = \omega\epsilon_2/4\pi$. Then

$$\sigma_1(\omega) = \frac{\omega_{pe}^2\omega^2\gamma/4\pi}{(\omega_e^2 - \omega^2)^2 + \omega^2\gamma^2}$$

The real and imaginary parts of σ are plotted in Fig. 18,

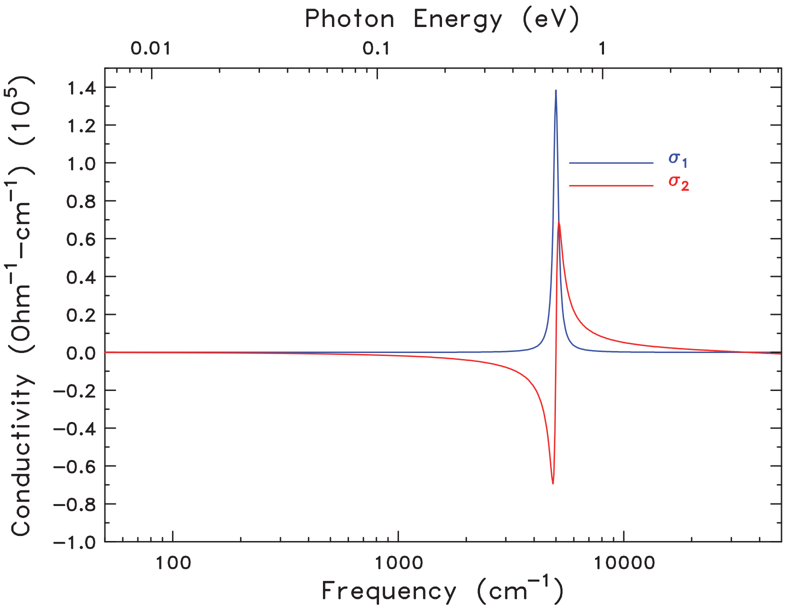

Fig. 18. Real and imaginary parts of the optical conductivity of a Lorentzian oscillator.

The shape of the Lorentzian is always the same: a peak in $\sigma_1(\omega)$ and a derivative like structure* in $\sigma_2(\omega)$. Three parameters control the details: ω_e the location, ω_{pe}^2 (as I shall show) the area under the $\sigma_1(\omega)$ curve, and γ the full width at half maximum. The value of the conductivity at the maximum is $\sigma_1(\omega_e) = \omega_{pe}^2/4\pi\gamma$.

4.4.7 The refractive index

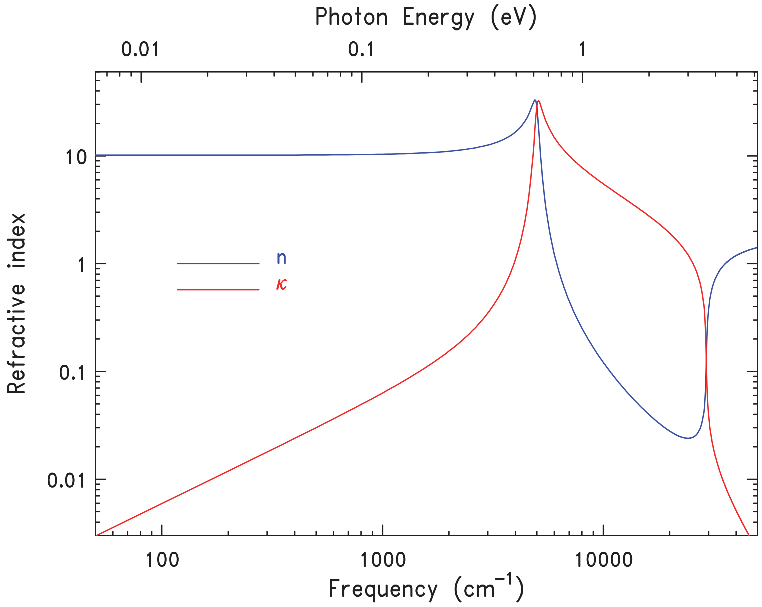

Fig. 19. Real and imaginary parts of the refractive index in the Lorentz model. Data are shown on a log-log scale.

---

[*] The negative of the derivative actually.

Figure 19 shows the real and imaginary parts of the refractive index of the Lorentz model. At low frequencies, n is much bigger than κ. κ increases with frequency. At the resonance, $n \approx \kappa$ and both are relatively large. Above the resonance, n falls well below unity but κ stays large (because $\epsilon_1 < 0$). Then, at a certain frequency ϵ_1 changes sign and κ falls rapidly while n jumps upward to a value bigger than 1. Note that both the lowest-frequency and the highest-frequency values of the refractive index n are just $n \approx \sqrt{\epsilon_1}$.

4.4.8 Reflectance

Once I have N, I can calculate the reflectance from Eq. 44. The reflectance of the Lorentz metal is shown in Fig. 20. At low frequencies, the real part of N dominates. As the frequency approaches ω_e, the reflectance increases, and above the resonance approaches that of a metal. Then, the reflectance falls rapidly from a value close to unity to a value close to zero. This drop occurs around 30,000 cm^{-1} (3.5 eV) for the data used for Fig. 20. This "plasma edge" occurs at the same frequencies where the rapid fall of κ and rise of n occur in Fig. 19. There is a minimum value of the reflectance, corresponding to the conditions* $n \approx 1$ and $\kappa \ll 1$.

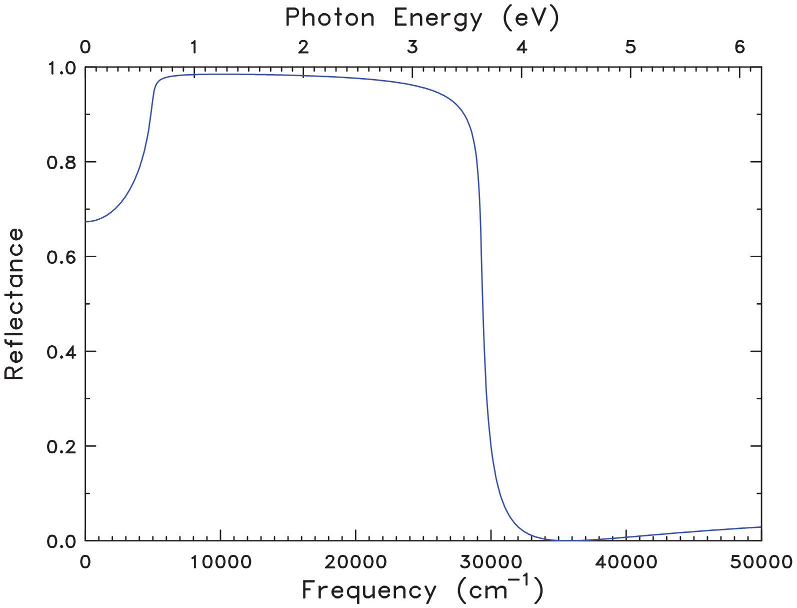

Fig. 20. Normal incidence reflectance in the Lorentz model.

4.4.9 Multiple polarizable electron levels

Let's suppose that there are a number of different electrons† with different binding energies in the solid. I'll indicate the different levels with an index j, where j runs from the

---

[*] The exact location of this "plasma minimum" is a balance of the frequency where $n - 1 \sim 0$ and, at the same time, κ has fallen to a relatively small value. It is *not* the same as the frequency where $\epsilon_1 = 0$ although these two frequencies are close to each other.

[†] Electrons are of course indistinguishable particles. However, in the multi-electron atom and in solids made from such atoms the electrons occupy different energy levels or energy bands, with the binding energy (or spring constant in the Lorentz model) different for each band.

least tightly bound electron to the most tightly bound electron I want to consider. (Even more tightly bound electrons remain of course polarizable and will contribute to ϵ_c.)

Because I am ignoring electron-electron interactions, the equation of motion for the jth carrier remains unchanged from Eq. 65, except that it needs subscripts:

$$m_j\ddot{\mathbf{x}}_j = -m_j\omega_{0j}^2\mathbf{x}_j - m_j\gamma_j\dot{\mathbf{x}}_j - e\mathbf{E}_0e^{-i\omega t},$$

where I've allowed for each level to have its own* spring constant ($m_j\omega_{0j}^2$) or resonant frequency ω_{0j}, dynamical mass m_j, and damping factor γ_j. The solution is unchanged from Eq. 66 except (still) for subscripts. The jth electron will have dipole moment $\mathbf{p}_j = -e\mathbf{x}_j$ and the total dipole moment per unit volume becomes

$$\mathbf{P} = \sum_j n_j\mathbf{p}_j$$

where n_j is the number density of type-j electrons.

Local field corrections are identical to those on page 42. The dielectric function becomes a sum of Lorentzian oscillators of the form in Eq. 74.

$$\epsilon = \epsilon_c + \sum_j \frac{\omega_{pj}^2}{\omega_j^2 - \omega^2 - i\omega\gamma_j}, \tag{76}$$

where $\omega_{pj} = \sqrt{4\pi n_j e^2/m_j}$ is the plasma frequency associated with the jth kind of electrons and ω_j is the (local-field corrected) resonant frequency for that kind.

By assumption, each kind of electron has its own resonant frequency. If these frequencies are separated by amounts large compared to the linewidth, then ϵ_2 and σ_1 will have a peak at each resonance. ϵ_1 will have a derivative like structure and will step down to a lower value as the frequency is increased past the resonance. If the separation is small compared to the linewidth, there will be a single combined peak in σ_1, with corresponding feature in ϵ_1.

Figure 21 shows one such calculation. The parameters are given in Table 5. I chose the parameters so that ω_j/γ_j was the same for all and also so that ω_{pj}/ω_j was the same for all. This choice makes the "quality factor Q of all resonances the same as well as their contributions to the static dielectric constant, $\epsilon_1(0)$.

Table 5. Parameters for five Lorentz oscillators.

Oscillator number j	ω_{pj}	ω_j	γ_j
1	5000	500	30
2	10000	1000	60
3	50000	5000	300
4	55000	5500	330
5	200000	20000	1200

$\epsilon_c = 3.0$

---

[*] But not its own charge!

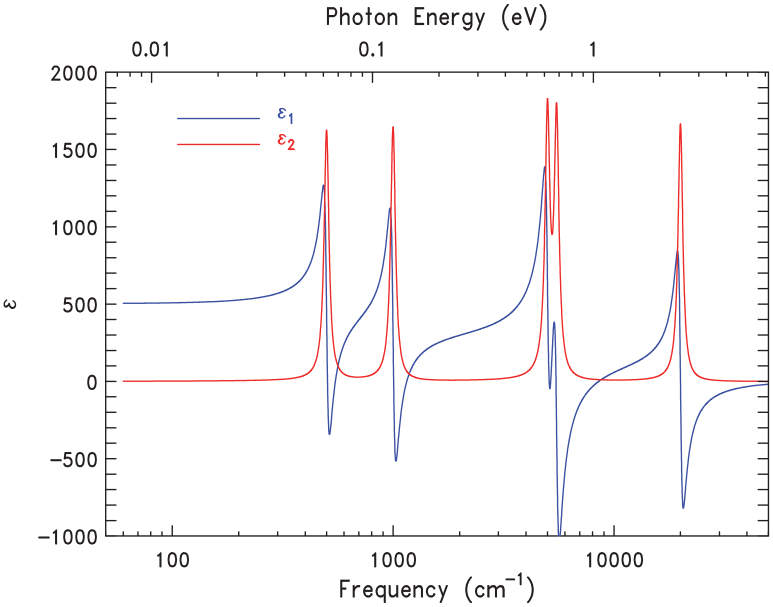

Fig. 21. Dielectric function for five Lorentz oscillators.

4.5 Vibrational absorption: phonons

The atoms in a solid may be set to vibrating either by temperature or by external forces. Most compounds, materials made up from two or more types of atoms, have a certain amount of charge transfer from one type to another.* In the extreme case of ionic solids, like NaCl, valence electrons are transferred from one atom to the other. In NaCl, the sodium atoms have lost an electron, and are present as Na$^+$ and the chlorine atoms have gained an electron, and are present as Cl$^-$. NaCl is face centered cubic, with each atom having six nearest neighbors of the opposite kind. These nearest neighbors are arranged at the corners of an octahedron surrounding the central atom. The bonding is mostly ionic, the Coulomb attraction of the ions for their opposite-sign partners. Even though there are charged ions throughout the solid, there is no net polarization because of the symmetrical arrangement of nearest neighbors. However, if an electric field is applied, the positive ions will be pushed in the direction of the field, and the negative ions will be pulled in the opposite direction, creating a net polarization. There is a natural vibrational frequency, the transverse optical phonon, at which these ions will oscillate. Here I will calculate the dielectric function contribution of these phonons from a simple harmonic oscillator model.

4.5.1 Harmonic oscillator

The model is as shown in Fig. 22. The motion of the ions can be broken up into center of mass motion and relative motion. The electrical force on the ions is $\mathbf{F}_\pm = \pm Ze\mathbf{E} = \pm Ze\mathbf{E}_{0\mathrm{loc}}e^{-i\omega t}$, affecting their relative separation x.

---

* Even if the bonding is mostly covalent, there is still a small charge difference across the bonds of dissimilar atoms.

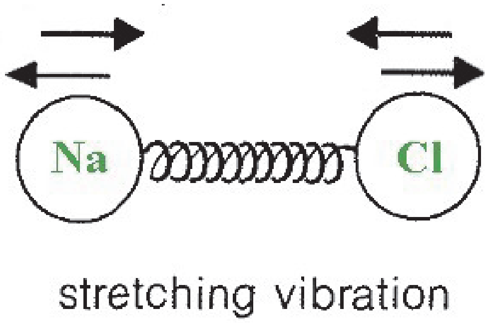

Fig. 22. The ionic force between Na^+ and Cl_- is represented by a spring. When light is shown at the NaCl crystal, the spring is stretched and shrunk at the oscillation frequency of the light. The upper (lower) arrows show the force on the ions when the field is pointing right (left).

4.5.2 Equation of motion

Newton's law is then written as

$$\mu\ddot{\mathbf{x}} = -\mu\omega_0^2\mathbf{x} - \mu\gamma\dot{\mathbf{x}} + Ze\mathbf{E}_{\mathrm{loc}}, \tag{77}$$

where μ is the reduced mass of the ions,[*] $\mu\omega_0^2 = K$, the spring constant, γ is a damping force proportional to the velocity, Z is the ionic charge, and $\mathbf{E}_{\mathrm{loc}}$ is the local field. I'll take $\mathbf{E}_{\mathrm{loc}} = \mathbf{E} + 4\pi\mathbf{P}$ as before. Equation 77 is analogous to Eq. 65 for the Lorentz model, with a change of notation. The solution is then the analog to Eq. 66. I'll write the induced dipole as $\mathbf{p} = Ze\mathbf{x}$ and define an ionic plasma frequency as

$$\Omega_p = \sqrt{\frac{4\pi n_i (Ze)^2}{\mu}},$$

where n_i is the number density of ion pairs. The next step is to make the local field correction, finding eventually the dielectric function for the optical phonons, the analog of Eq. 74:

$$\epsilon = \epsilon_e + \frac{\Omega_p^2}{\omega_T^2 - \omega^2 - i\omega\gamma} \tag{78}$$

where ϵ_e is the electronic contribution, including *both* the core electrons and the valence electrons,[†] and ω_T is the transverse optical phonon frequency, related to the dilute-limit resonance by $\omega_T = \sqrt{\omega_0^2 - \Omega_p^2/3}$.

The dielectric function and other optical function for the Lorentz oscillator have been shown in Figs. 17–20. The functions for the lattice oscillations are the same, except that the frequency scale is different. The TO mode frequencies are typically in the range of 40–600 cm^{-1} (6–100 meV or 1.6–25 THz).[‡]

---

[*] Relative motion!

[†] Here I assume—as typically is the case for ionic insulators—that the excitation energies for the electrons are much higher than the phonon energies, $\hbar\omega_T$.

[‡] The modes for very light atoms, such as C-H or O-H can be as high as 3300 cm-1 and the frequencies for vibrations in organic solids, the C-C, C=C, and C=O bonds, are in the range 800–1600 cm-1.

4.5.3 Transverse and longitudinal modes

The phonon modes may be separated into transverse and longitudinal modes.[11] The transverse modes are those that are optically active. However, not all modes lead to optical absorption. Some are Raman active and some are "silent." The selection rules can be determined by group theoretical analysis.[12–14] The infrared active modes are those that transform like the momentum (p_x, p_y, or p_z) or position (x, y, or z). In general, these are modes that are odd under inversion. (In the lingo of group theory, the infrared active modes are *ungerade* or uneven and typically carry the subscript u. Other mode are *gerade* or even. subscript g; these will be Raman active or inactive.)

The infrared-active transverse modes are the resonant frequencies of the harmonic oscillator; for each of these there is a corresponding longitudinal optical mode. In the language of transfer functions in control theory,[15] the transverse modes are the poles and the longitudinal modes are the zeros of the dielectric function.*

Let me remind you about what I said for the plasma response. If the mode is longitudinal, then $\mathbf{q}$ is parallel to $\mathbf{E}$. Thus, $\nabla \cdot \mathbf{E} \neq 0$. In the insulating solid, there is no free charge, so $\nabla \cdot \mathbf{D} = 0$. For linear materials, $\mathbf{D} = \epsilon\mathbf{E}$, so $\nabla \cdot \mathbf{D} = \epsilon\nabla \cdot \mathbf{E} = 0$. Because $\nabla \cdot \mathbf{E} \neq 0$, $\epsilon = 0$. A further implication is: $\mathbf{D} = \epsilon\mathbf{E} = 0$. But $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ so that $\mathbf{P} = -\mathbf{E}/4\pi$.

4.5.4 The longitudinal frequency

So the zero is important. Let me estimate this frequency. First, I'll write Eq. 78 on a common denominator:

$$\epsilon = \epsilon_e \frac{\omega_T^2 - \omega^2 - i\omega\gamma + \Omega_p^2/e_c}{\omega_T^2 - \omega^2 - i\omega\gamma},$$

and next I define

$$\omega_L^2 = \omega_T^2 + \Omega_p^2/e_c, \tag{79}$$

which allows me to write

$$\epsilon = \epsilon_e \frac{\omega_L^2 - \omega^2 - i\omega\gamma_L}{\omega_T^2 - \omega^2 - i\omega\gamma_T}, \tag{80}$$

where I've put subscripts on the damping factors (relaxation rates) to suggest that different damping factors may affect the transverse and longitudinal modes.

The real part of the dielectric function in Eq. 80 is zero at a frequency near ω_L. The frequency is exactly ω_L when the damping is negligible.

4.5.5 The Lyddane-Sachs-Teller relation

In the limit as $\omega \to 0$, Eq. 80 becomes

$$\epsilon(0) = \epsilon_e \left(\frac{\omega_L}{\omega_T}\right)^2. \tag{81}$$

This is the Lyddane-Sachs-Teller relation.[16] The static dielectric constant is larger than the high-frequency (electronic) dielectric constant by the square of the ratio of longitudinal to transverse frequencies. Data for a large number of insulating (and mostly ionic) solids are in Table 6, adapted from Kittel.[11]

---

* Just as the screened plasma frequency, $\omega_p/\sqrt{\epsilon_c}$, is the zero of the Drude dielectric function.

Table 6. Transverse and longitudinal frequencies in solids, and the static and high-frequency dielectric constants.

Material	Transverse mode			Longitudinal mode			$\epsilon_1(0)$	ϵ_e
	cm^{-1}	THz	meV	cm^{-1}	THz	meV		
LiH	584	17.5	72.3	1114	33.4	138.1	12.9	3.6
LiF	308	9.2	38.1	637	19.1	78.9	8.9	1.9
LiCl	191	5.7	23.7	398	11.9	49.3	12.0	2.7
LiBr	159	4.8	19.7	324	9.7	40.1	13.2	3.2
NaF	239	7.2	29.6	414	12.4	51.3	3.1	1.7
NaCl	164	4.9	20.4	265	8.0	32.9	5.9	2.25
NaBr	133	4.0	16.4	207	6.2	25.7	6.4	2.6
KF	191	5.7	23.7	324	9.7	40.1	5.5	1.5
KCl	143	4.3	17.8	212	6.4	26.3	4.85	2.1
KI	101	3.0	12.5	138	4.1	17.1	5.1	2.7
RbF	154	4.6	19.1	286	8.6	35.5	6.5	1.9
RbI	74	2.2	9.2	101	3.0	12.5	5.5	2.6
CsCl	101	3.0	12.5	164	4.9	20.4	7.2	2.6
CsI	64	1.9	7.9	85	2.5	10.5	5.65	3.0
TlCl	64	1.9	7.9	159	4.8	19.7	31.9	5.1
TlBr	43	1.3	5.3	101	3.0	12.5	29.8	5.4
AgCl	101	3.0	12.5	180	5.4	22.4	12.3	4.0
AgBr	80	2.4	9.9	133	4.0	16.4	13.1	4.6
MgO	398	11.9	49.3	743	22.3	92.1	9.8	2.95
Gap	366	11.0	45.4	403	12.1	50.0	10.7	8.5
GaAs	271	8.1	33.5	292	8.8	36.2	13.13	10.9
GaSb	228	6.8	28.3	244	7.3	30.3	15.69	14.4
InP	302	9.1	37.5	345	10.3	42.8	12.37	9.6
InAs	218	6.5	27.0	239	7.2	29.6	14.55	12.3
InSb	186	5.6	23.0	196	5.9	24.3	17.88	15.6
Sic	790	23.7	98.0	950	28.5	117.7	9.6	6.7
C	1332	39.9	165.1	1332	39.9	165.1	5.5	5.5
Si	525	15.8	65.1	525	15.8	65.1	11.7	11.7
Ge	302	9.1	37.5	302	9.1	37.5	15.8	15.8

4.5.6 Other notations.

The zero-frequency limit of Eq. 78 is

$$\epsilon_1(0) = \epsilon_c + \frac{\Omega_P^2}{\omega_T^2}.$$

This expression for the static dielectric constant is of course equivalent to Eq. 81 given the definition of ω_L in Eq. 79. Sometimes it is of interest to highlight the fact that Ω_P^2/ω_T^2 is the (additive) contribution to the static dielectric constant. Then one defines either $\Delta\epsilon$ or the so-called oscillator strength S as

$$\Delta\epsilon \equiv S \equiv \frac{\Omega_P^2}{\omega_T^2}. \tag{82}$$

The dielectric function Eq. 78 in terms of S is then

$$\epsilon = \epsilon_e + \frac{S\omega_T^2}{\omega_T^2 - \omega^2 - i\omega\gamma} \tag{83}$$

4.5.7 Multiple modes

Many solids have more than one phonon branch. As the chemical formula becomes more complex, the numbers of degrees of freedom increase. For the phonons, the key quantity is the number N of distinct atoms in the unit cell. There is one for Ag, two for NaCl, and perhaps five for Al_2O_3. With N atoms in the unit cell, there are $3N$ degrees of freedom for the normal modes. Three of these are always the acoustical phonons, in which the entire unit cell moves as single unit, leaving $3N - 3$ optical modes. The optical modes involve relative motion of the atoms within the unit cell. Although not all these modes are infrared active, many are. Each infrared active will contribute a mode to the dielectric function, which I'll model as a sum of Lorentz oscillators just as in the case multiple electronic orbitals (page 47). Equation 78 becomes

$$\epsilon = \epsilon_e + \sum_j \frac{\Omega_{pj}^2}{\omega_j^2 - \omega^2 - i\omega\gamma_j}, \tag{84}$$

while the equivalent Eq. 83 is

$$\epsilon = \epsilon_e + \sum_j \frac{S_j\omega_j^2}{\omega_j^2 - \omega^2 - i\omega\gamma_j}. \tag{85}$$

In Eqs. 84 and 85, I've written the resonant frequency as ω_j rather than ω_{Tj}. Sometimes one sees one such notation and sometimes the other.

By analogy, Eq. 80 is written

$$\epsilon = \epsilon_e \prod_j \frac{\omega_{Lj}^2 - \omega^2 - i\omega\gamma_{Lj}}{\omega_{Tj}^2 - \omega^2 - i\omega\gamma_{Tj}}, \tag{86}$$

and the Lyddane-Sachs-Teller relation, Eq. 81, becomes

$$\epsilon(0) = \epsilon_e \prod_j \left(\frac{\omega_{Lj}}{\omega_{Tj}}\right)^2. \tag{87}$$

Each oscillator contributes to the static dielectric constant, with the ones at the lowest frequencies dominant. If a mode goes "soft" (i. e., its frequency goes to zero as temperature or some other parameter is varied), the dielectric constant will diverge; such soft modes may be found in ferroelectric and almost-ferroelectric crystals.

4.6 Comments on wave propagation

I'll end this relatively long chapter with comments on how electromagnetic waves propagate in a material, using the Lorentz and the Drude models as a basis for my comments.

Electromagnetic wave propagation (and reflection/transmission at interfaces) is governed by the dielectric function, which in turn sets the refractive index. A wave traveling along x is written

$$\mathbf{E} = \mathbf{E}_0 e^{i(N\frac{\omega}{c}x - \omega t)} = \mathbf{E}_0 e^{i(n\frac{\omega}{c}x - \omega t)} e^{-\kappa\frac{\omega}{c}x}$$

with refractive index

$$N = \sqrt{\epsilon}$$

and optical absorption coefficient

$$\alpha = \frac{2\omega}{c}\kappa, \tag{88}$$

which I already wrote as Eq. 38.[*]

4.6.1 Insulators: Lorentz model

I will discuss wave propagation in the Lorentz model first. In the discussion I'll take the hierarchy of frequencies as $\omega_p \gg \omega_e \gg \gamma = 1/\tau$. Even if the ratio of these quantities is only 3 or 5, I'll consider the inequality satisfied. The dielectric function and refractive index are plotted again in Fig. 23. These functions are the same as those in Figs. 18 and 19. I have added vertical bars to separate the diagram into 4 parts, labeled A–D. Each of these has its own flavor of wave propagation.

A. Mostly transparent with weak absorption coefficient

In Region A, ϵ_1 is positive and in fact larger than unity. Moreover, $\epsilon_1 \gg \epsilon_2$ and $n \gg \kappa$. The solid is a transparent insulator with an absorption length (the inverse of α) that is longer than the wavelength, often much longer. The reflectance is governed largely by n; indeed, $\mathscr{R} \approx [(n-1)/(n+1)]^2$. Because n increases with frequency, so does the reflectance. This behavior is termed "ordinary dispersion" in the literature.

B. Strongly absorbing, near $\omega = \omega_e$ or ω_T

At the center of Region B, $\epsilon_2 \gg \epsilon_1$, so that

$$n \approx \kappa = \sqrt{\frac{\epsilon_2}{2}} \gg 1.$$

Waves are strongly damped, but currents and fields are more or less in phase. The damping length (δ or $1/\alpha$) and the wavelength differ by about 2π or π.

---

[*] Recall that the absorption coefficient describes the attenuation of the intensity, $I \sim EE^* \sim e^{-\alpha x}$, in the solid ($x$ is the distance the wave has traveled); the factor of 2 comes from the fact that the fields are squared to give the intensity.

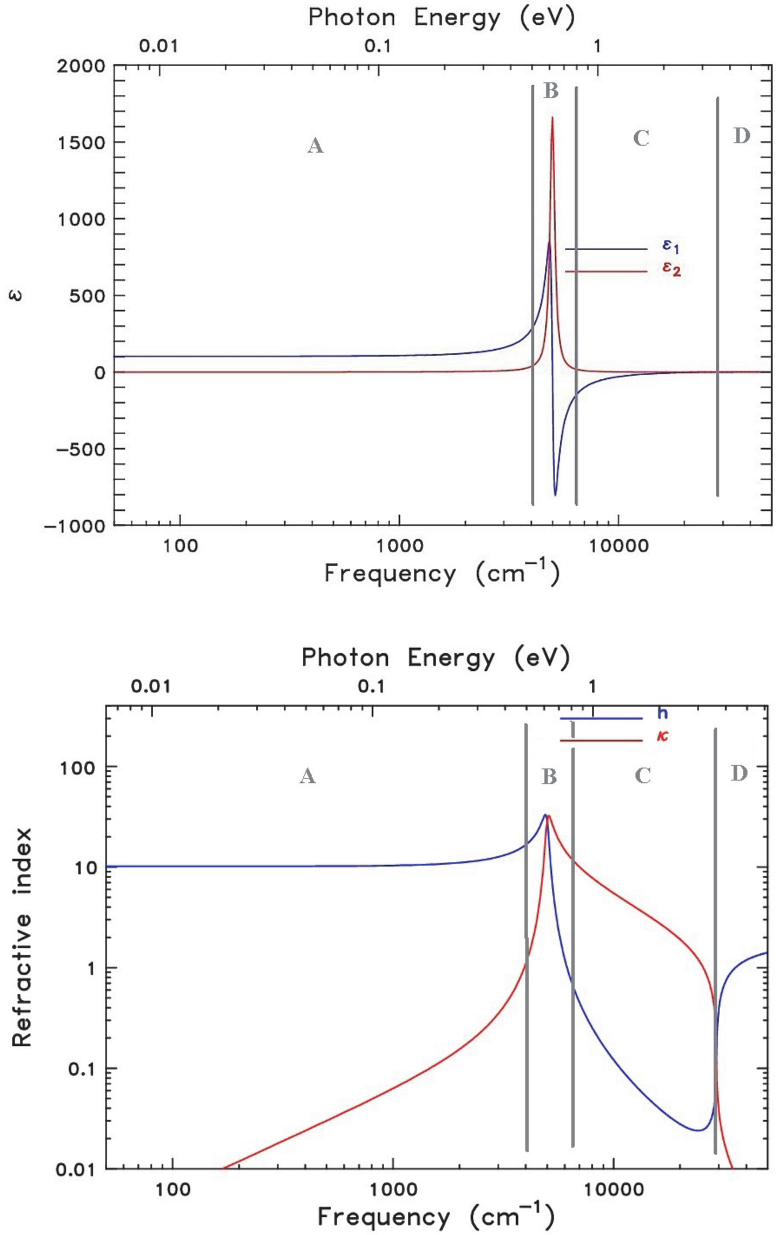

Fig. 23. (Top) Dielectric function and (bottom) refractive index for the Lorentz oscillator. The four regions A–D are shown by the vertical gray bars. The boundaries are fuzzy, and exact location is not to be taken seriously.

To estimate the reflectance, I'll start with Eq. 62:

$$\mathscr{R} = 1 - \frac{4n}{(n+1)^2 + \kappa^2},$$

which becomes, near the resonance,

$$\mathscr{R} = 1 - \frac{2}{\kappa}.$$

Exactly on resonance, $\epsilon_2 = \omega_{pe}^2/\omega\gamma$ so that the reflectance becomes

$$\mathscr{R} = 1 - \sqrt{\frac{8\omega\gamma}{\omega_{pe}^2}}.$$

This is the equivalent of Hagen Rubens, Eq. 63.

Note that (as asserted on page 2) "strongly absorbing" and "highly reflecting" go hand-in-hand here. Around the middle of region B, the refractive index begins to decrease with increasing frequency. This behavior is termed "anomalous dispersion" in the literature.

C. Mostly damped, $\omega_e \ll \omega <$ the zero of ϵ_1

In Region C, $\epsilon_1 < 0$ (negative) and $|\epsilon_1| \gg \epsilon_2$, so that n is a fraction of unity. and can be taken as close to zero. Then, $\kappa \approx \sqrt{|\epsilon_1|}$. So long as the frequency is not too close to either the resonance or the zero crossing, I may take $\epsilon_1 \approx -\omega_{pe}^2/\omega^2$ and $\kappa \approx \omega_{pe}/\omega$. The wave is strongly damped with

$$\alpha = \frac{2\omega}{c}\kappa = \frac{2\omega_{pe}}{c}$$

and a wavelength much larger than the vacuum wavelength.

To estimate the reflectivity, I need to estimate n. I will start with $2n\kappa = \epsilon_2$ and use $\epsilon_2 \approx \omega_{pe}^2\gamma/\omega^3$ and κ from above, so that $n = \omega_{pe}\gamma/2\omega^2$. Then taking $\mathscr{R}$ from Eq. 62 in the case where $\kappa \gg n$, I arrive at

$$\mathscr{R} = 1 - \frac{4n}{\kappa^2} = 1 - \frac{2\gamma}{\omega_{pe}}.$$

The reflectance is frequency independent. It is high also. Here the response is inductive, so the dissipation is not high; instead, most of the incident energy is reflected. The "anomalous dispersion continues through most of region C.

D. High frequencies, above the zero of ϵ_1

Here ϵ_1 is positive and bigger than ϵ_2; κ becomes small and $n \approx \sqrt{\epsilon_c - \omega_{pe}^2/\omega^2}$. The material is transparent again. The reflectance has a minimum very close* to the frequency where $n = 1$. The limiting high-frequency value of the reflectance is

$$\mathscr{R} = \frac{(\sqrt{\epsilon_c} - 1)^2}{(\sqrt{\epsilon_c} + 1)^2}.$$

Ordinary dispersion is recovered.

4.6.2 Metals: Drude model

A plot analogous to Fig. 23 is shown in Fig. 24. Everything I said in sections B–D above applies here too. I can rewrite the relevant equations, replacing ω_{pe} with ω_p and γ with $1/\tau$.

---

* It would be exact except for the frequency dependence of κ.

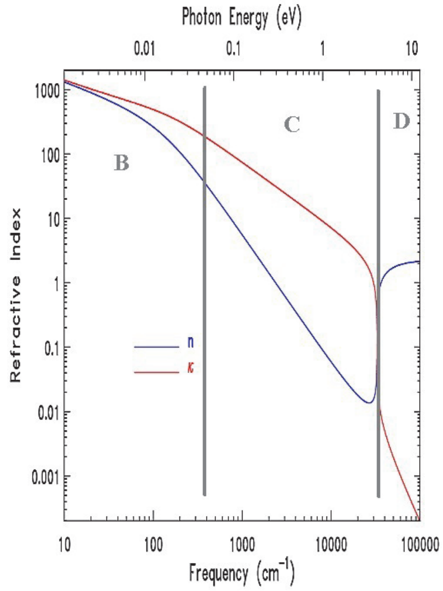

Fig. 24. Refractive index for the Drude metal. Three regions B–D are shown by the vertical grey bars. The boundaries are fuzzy, and exact location is not to be taken seriously. The image has been squashed horizontally so that it resembles the right half of the bottom panel in Fig. 23.

B. Strongly absorbing, $\omega \ll 1/\tau$

In metals, region B starts at zero frequency. At these low frequencies, $\epsilon_2 \gg \epsilon_1$ or $\sigma_1 \gg \sigma_2$. In this limit, $n \approx \kappa$. I can write

$$\epsilon_1 \approx -\omega_p^2\tau^2$$

and

$$\epsilon_2 \approx \frac{\omega_p^2\tau}{\omega}$$

and

$$n = \kappa \approx \sqrt{\frac{\omega_p^2\tau}{2\omega}} = \sqrt{\frac{2\pi\sigma_{\text{dc}}}{\omega}}$$

where the last equality comes from $\sigma_{\text{dc}} = \omega_p^2\tau/4\pi$.

The field decays as $e^{-x/\delta}$ with

$$\delta = \frac{c}{\omega\kappa} \approx \sqrt{\frac{c^2}{2\pi\omega\sigma_{\text{dc}}}} = \sqrt{\frac{2c^2}{\omega_p^2\omega\tau}}. \tag{89}$$

Eq. 89 is the classical skin depth of the metal. It correctly gives the penetration length in the low frequency limit, where $\sigma_1(\omega) \approx \sigma_{\text{dc}}$. I'll return to the skin effect in Chapter 12.

I'll start with Eq. 62 but let $n = \kappa \gg 1$ and write $\mathscr{R} = 1 - 2/\kappa$. Then the reflectance becomes

$$\mathscr{R} = 1 - \sqrt{\frac{8\omega}{\omega_p^2\tau}} = 1 - \sqrt{\frac{2\omega}{\pi\sigma_{\text{dc}}}}$$

This is what Hagen and Rubens found and what I wrote in Eq. 63.

C. Mostly damped, $1/\tau < \omega < \omega_p/sqrteps_c$

In Region C, $\epsilon_1 \approx -\omega_p^2/\omega^2$ and $\kappa \approx \omega_p/\omega$. The wave has a wavelength much larger than the vacuum wavelength and is strongly damped with

$$\delta = \frac{c}{\omega\kappa} \approx \frac{c}{\omega_p} \tag{90}$$

and

$$\alpha = \frac{2\omega}{c}\kappa \approx \frac{2\omega_p}{c}$$

To estimate the reflectivity, I need to estimate n. I will start with $2n\kappa = \epsilon_2$ and use $\epsilon_2 \approx \omega_p^2/\omega^3\tau$ and κ from above, so that $n = \omega_p/2\omega^2\tau$. Then taking $\mathscr{R}$ from Eq. 62 in the case where $\kappa \gg n$, I arrive at

$$\mathscr{R} = 1 - \frac{4n}{\kappa^2} = 1 - \frac{2}{\omega_p\tau}. \tag{91}$$

The reflectance is frequency independent.

D. High frequencies, above the zero of ϵ_1

Here ϵ_1 is positive and quickly becomes bigger than ϵ_2, i. e., $\epsilon_1 \approx \epsilon_c - \omega_p^2/\omega^2$ whereas $\epsilon_2 \approx \omega_p^2/\omega^3\tau$. The reflectance has a minimum very close* to the frequency where $n = 1$. See Fig. 14. The limiting high-frequency value of the reflectance is

$$\mathscr{R} = \frac{(\sqrt{\epsilon_c} - 1)^2}{(\sqrt{\epsilon_c} + 1)^2}. \tag{92}$$

The light that is not reflected is transmitted, as long as the sample is not too thick.†

4.7 The absorption coefficient and a common mistake

Light traveling in an absorbing medium is attenuated; the amount of attenuation is governed by the extinction coefficient κ or, almost equivalently, by the absorption coefficient α; as I wrote in Eqs. 38 and 88, the two are related by $\alpha = 2\omega\kappa/c$. κ has no dimensions, α is in cm^{-1} in cgs units.

After traveling a distance d, the intensity is reduced to

Problems

Slow glass: If you did not do sg problem in chap 2, do it now. Then, find ϵin the visible. Assume this comes from a single UV transition at 10 eV. (I want it to be high enough that I can neglect dispersion. Calculate ω_p for this. Calculate n taking $m = m_e$. Neglect damping.

Sellmeiyer (1871) and Cauchy (1836)

---

* It would be exact except for the frequency dependence of κ. Note that if $\epsilon_1 \gg \epsilon_2$ and $n = 1$ then $\epsilon_1 \approx 1$ also.

† How thick is too thick? Well, the light decays as $e^{-\alpha x}$ so too thick is a few times $\alpha = 2\omega\kappa/c = \omega\epsilon_2/nc$.

5. A LOOK AT REAL SOLIDS

Eventually the plots shown in `RealSolids.pdf` will appear in this section, along with some discussion.

Now they are just inserted on the following pages

6. INCOHERENT AND COHERENT TRANSMISSION OF A THIN FILM

Let me now think about how to do an experiment to learn about the optical properties of some material. I want to know how light of different frequencies propagates in the material. I've discussed the propagation of light in a solid, starting on page 18. The propagation is governed by the complex refractive index or, equivalently, the complex dielectric function.

A cartoon of typical experimental setups was shown in Fig. 3 (on page 4). To learn about my sample, I need to send light at wavelength λ (or frequency ω) from some optical device (laser, spectrometer, interferometer) to the sample where some if it will be reflected (see page 21) and some will cross the surface and begin to travel in the solid. If the sample is very thick, so that the light has little chance to make it to the back surface,* the reflectance (Eq. 44) is all one can measure. See section 9 for discussion of Kramers-Kronig analysis of this "single-bounce" reflectance.

Here, I want to consider right-hand side of Fig. 3, where the light transmits through the sample, with a significant fraction making it out of the rear surface, where its intensity may be measured to find the transmittance $\mathscr{T}(\omega)$. Of course some light also is reflected (from both surfaces), so I can move the detector to the other side and measure also the reflectance $\mathscr{R}(\omega)$.

6.1 Incoherent light

To begin, I'll consider the case of a thick (compared to the wavelength) slab. Moreover, it has some slight variation in thickness across the beam. The thickness variation is a few wavelengths, so that I can combine intensities and not consider coherence (interference) until later. Figure 25 shows a diagram of the paths taken by light as it passes through a slab of thickness d. The light is incident from the left.† At the first surface, some if the incident light is reflected, and some transmitted. This transmitted light passes through and some exits from the rear side. The rest is reflected, and then bounces back and forth between front and back, with a fraction leaking out of each surface each time.

6.1.1 The transmittance

To set up the problem, I'll follow the paths of the individual rays shown in Fig. 25. I will take the refractive index of the semi-infinite media on both sides to be $N = 1$. Let the incident ray have unit intensity. The reflected intensity at the front surface (sometimes

---

[*] I know the scale for "very thick": the electric field of light is attenuated as $\mathbf{E}{\sim}e^{-\omega\kappa x/c}$ an at distance x into the sample; the intensity goes as $I = I_0 e^{-\alpha x}$. So if the thickness d is large compared to $c/\omega\kappa$ or $1/\alpha$, the field or intensity is $e^{-\mathrm{large}}$ when the light reaches the back surface. In this case, the reflected light comes entirely from the first surface. In the other limit, the light that reaches the back surface will be in part reflected and some of that will pass through the front surface and become a component of the reflected field.

[†] The diagram shows the rays inside the slab at an angle $\theta$to the slab normal; I'll work out the case for normal incidence. I drew the oblique incidence to allow me to follow the various partial waves as they travel back and forth in the sample.

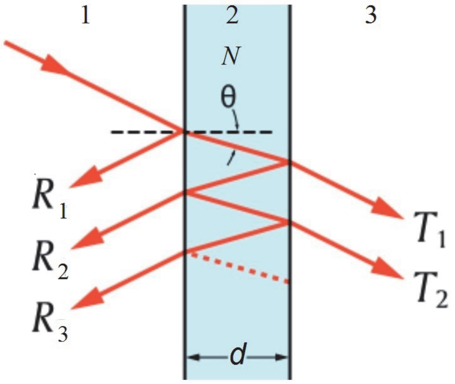

Fig. 25. Multiple internal reflections in a slab with refractive index N and thickness d.

called the prompt reflectance) is $\mathscr{R}_{12}$, where the subscripts mean "from 1 onto 2." The magnitude is given by Eq. 44,

$$\mathscr{R}_{12} = \left|\frac{1-N}{1+N}\right|^2 = \frac{(n-1)^2+\kappa^2}{(n+1)^2+\kappa^2}. \tag{93}$$

The intensity just inside the slab at the front surface is $1 - \mathscr{R}_{12}$ because there is no absorption at the interface. There is, however, absorption as the light travels through the slab, following Eq. 37, $I = I_0 e^{-\alpha x}$. When the light reaches the rear surface some is reflected. The fraction reflected is $\mathscr{R}_{23} = \mathscr{R}_{12}$. Some, $= 1 - \mathscr{R}_{12}$, transmitted. Thus, the intensity of the first ray exiting from the slab is

$$\mathscr{T}_1 = (1-\mathscr{R}_{12})e^{-\alpha x}(1-\mathscr{R}_{12}),$$

consisting of reflection losses at both surfaces and bulk absorption.

I will now follow the other rays that lead to transmission. The second ray travels three times through the sample and has intensity

$$\mathscr{T}_2 = (1-\mathscr{R}_{12})e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}(1-\mathscr{R}_{12}).$$

It is transmitted at the front, passes through the slab, is reflected, passes through, is reflected, passes through, and is transmitted. The third ray is

$$\mathscr{T}_3 = (1-\mathscr{R}_{12})e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}(1-\mathscr{R}_{12}).$$

It is transmitted at the front, passes through the slab, is reflected, passes through, is reflected, passes through, is reflected, passes through, is reflected, passes through, and is transmitted.

Now note that $\mathscr{T}_2$ and $\mathscr{T}_3$ both contain all the factors in $\mathscr{T}_1$ multiplied by $\mathscr{R}_{12}^2 e^{-2\alpha x}$ and $\mathscr{R}_{12}^4 e^{-4\alpha x}$ respectively. The next term is of course the 6th power of $\mathscr{R}_{12}e^{-\alpha x}$, and the total transmission is an infinites sum of the partial rays. The transmittance is then

$$\mathscr{T} = (1-\mathscr{R}_{12})^2 e^{-\alpha x}[1 + \mathscr{R}_{12}^2 e^{-2\alpha x} + \mathscr{R}_{12}^4 e^{-4\alpha x} + \ldots].$$

Now, I remember the expansion of $1/(1-\varepsilon) = 1 + \varepsilon + \varepsilon^2 + \ldots$, which converges as long as $\varepsilon < 1$. Clearly $\mathscr{R}_{12} < 1$ as is $e^{-2\alpha x}$, so the series is convergent. I sum it to get

$$\mathscr{T} = \frac{(1-\mathscr{R}_{12})^2 e^{-\alpha x}}{1-\mathscr{R}_{12}^2 e^{-2\alpha x}} \tag{94}$$

6.1.2 The reflectance

I can use the same approach for the reflected rays:

$$\mathscr{R}_1 = \mathscr{R}_{12},$$

and

$$\mathscr{R}_2 = (1 - \mathscr{R}_{12})e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}(1 - \mathscr{R}_{12}),$$

and

$$\mathscr{R}_3 = (1 - \mathscr{R}_{12})e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}\mathscr{R}_{12}e^{-\alpha x}(1 - \mathscr{R}_{12}).$$

Now I keep $\mathscr{R}_1$ separate and add all the partial rays for $\mathscr{R}_2$, $\mathscr{R}_3 \ldots \mathscr{R}_\infty$ together, factor out common terms, and find the same infinite series as in $\mathscr{T}$. The reflectance is:

$$\mathscr{R} = \mathscr{R}_{12} + \mathscr{R}_{12}\frac{(1 - \mathscr{R}_{12})^2 e^{-2\alpha x}}{1 - \mathscr{R}_{12}^2 e^{-2\alpha x}}. \tag{95}$$

The second term is positive, so that the slab reflectance is larger than the single-bounce reflectance $\mathscr{R}_{12}$.

6.1.3 Some approximations

A common case is the one where $\kappa \ll 1$. Why should I say that? Well, the absorption coefficient $\alpha = 2\omega\kappa/c = 4\pi\kappa/\lambda$. Suppose the slab thickness $d = 10\lambda$. This is about as thin as it can be and still be considered "thick." In this case, $\alpha d \approx 120\kappa$, where I've used the old fashioned approximation, $\pi \approx 3$. The transmission is (neglecting reflection at the surfaces)

$$\mathscr{T} \approx e^{-120\kappa}.$$

Now I want to get light through the slab, at least at the 1% level. So I take $\mathscr{T} = 0.01$ and find $-120\kappa = \ln(0.01)$. Thus $\kappa = 0.037$, and is indeed rather small.

The advantage of having small κ is that I can neglect it[*] in Eq. 93 and write $\mathscr{R}_{12} = (n-1)^2/(n+1)^2$. Substituting this into Eq. 94 and doing a little algebra, I find

$$\mathscr{T} = \frac{16n^2}{(n+1)^4 e^{\alpha d} - (n-1)^4 e^{-\alpha d}}. \tag{96}$$

Equation 96 simplifies further if instead of having small κ, I have zero κ. Then $\alpha = 0$ and the transmission simplifies to

$$\mathscr{T} = \frac{2n}{n^2 + 1}. \tag{97}$$

Eq. 97 has the correct limit when $n = 1$. Note that it is different from the transmission across the first surface, for which $\mathscr{T} = 4n/(n+1)^2$.

---

[*] Actually, I will be neglecting κ^2, which is about 0.0014 in this example.

6.2 Coherent case

If the material is thin, or if it is thick but the two surfaces are parallel to within a fraction of a wavelength, then the multiple internal reflections may add coherently; the consequences are that there are interference patterns (or standing waves effects) in the transmitted and reflected light. Layers with thicknesses corresponding to the wavelength will usually show these effects; they are responsible for the iridescent colors in an oil film on water, in the light scattered off of the wings of certain insects, and in many other cases in nature. But it is not necessary for the layer to be comparable to the wavelength. The arm cavities of LIGO[17] are 4 km in length and laser wavelength is 1 μm; nevertheless, the mirrors are held so parallel by a complex and effective control system that there is constructive interferences of the light bouncing back and forth between the mirrors even though they are separated by 4×10^9 wavelengths.

In the coherent case one must add the amplitudes of the fields[*] and, after finding the resultant field of the superposed components, can calculate the intensity or the Poynting vector in order to find out about the energy.[†] One must keep track of the phase of the fields as the light propagate; the phase advances as $e^{iq\ell}$ for travel along the path ℓ. There are also phase changes on reflection or transmission across an interface.

6.2.1 The geometry

I'll start by defining the geometry, illustrated in Fig. 26.

The layer has thickness d and refractive index N_b. It is sandwiched between media with indices N_a and N_c. I'll take the origin of the x-axis (along which the wave travels) at the front surface; the back is at $x = d$. A plane wave of unit amplitude is incident at normal incidence from the right medium: $\mathbf{E}_{\text{inc}} = \hat{\boldsymbol{e}} e^{iq_a x}$ with $q_a = N_a\omega/c$. As long as the media are isotropic, I can ignore the field direction. I'll evaluate the fields at $x = 0$ and $x = d$, either inside or outside the layer. The first surface has reflection coefficient r_{ab}, given in Eq. 41,

$$r_{ab} = \frac{N_a - N_b}{N_a + N_b},$$

and transmission coefficient t_{ab}, given in Eq. 42,

$$t_{ab} = \frac{2N_a}{N_a + N_b}.$$

There are equivalent equations for r_{bc}, r_{ba}, t_{bc}, and t_{ba}. In each of these subscripts like $_{ab}$ mean “incident from a into b.”

---

[*] One typically adds the electric field; and lets the magnetic field come along for the ride.

[†] Actually, one should *always* combine the fields coherently. The incoherent case arises because the light is in a beam wide compared to the wavelength. If the sample thickness varies over this beam dimension, then the phase differences among the combined waves will vary in the same way, with some parts adding constructively and some destructively. An average over these phases gives the resultant transmittance and reflectance.

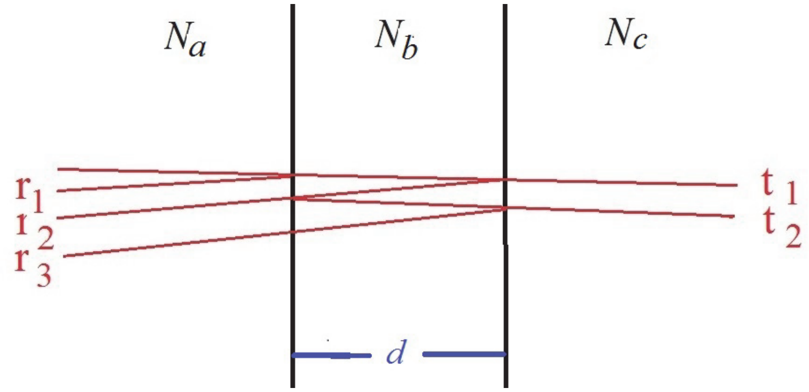

Fig. 26. Amplitude transmission and reflection of a layer with thickness d and refractive index N_b between media with indices N_a and N_c.

6.2.2 Transmission coefficient

Now, I will follow the variou partial rays to find their sum, the transmitted field. For ray t_1 the light is transmitted across the ab interface, travels a distance d, and is transmitted across the bc interface. The field is then

$$t_1 = t_{ab} e^{iq_b d} t_{bc}.$$

Some of the light is reflected at the back surface and travels to the front surface where some of it is reflected and travels to the back surface where some if it is transmitted. Starting with the incident field, I find that t_2 is

$$\begin{aligned} t_2 &= t_{ab} e^{iq_b d} r_{bc} e^{iq_b d} r_{ba} e^{iq_b d} t_{bc} = t_{ab} r_{bc} r_{ba} t_{bc} e^{3iq_b d} \\ &= t_1 r_{bc} r_{ba} e^{2iq_b d}. \end{aligned}$$

Clearly, the third partial ray is

$$t_3 = t_1 (r_{bc} r_{ba} e^{2iq_b d})^2,$$

because the ray has made 5 passes across medium b, internally reflecting twice at each surface. Thus, just as in the discussion leading to Eq. 94, the multiple internal reflections lead to an infinite series, which I can sum to get the amplitude transmission coefficient of the layer, which I'll call τ.

$$\tau = \frac{t_{ab} t_{bc} e^{iq_b d}}{1 - r_{bc} r_{ba} e^{2iq_b d}}. \tag{98}$$

6.2.3 Reflection coefficient

An identical analysis (See Eq. 95.) gives the amplitude reflection coefficient of the layer, ρ.

$$\rho = r_{ab} + \frac{t_{ab} r_{bc} t_{ba} e^{iq_b d}}{1 - r_{bc} r_{ba} e^{2iq_b d}}. \tag{99}$$

6.2.4 Intensities

The transmitted and reflected intensities are just the absolute value squared of these amplitude coefficients:

$$\mathscr{T} = \tau\tau^* \quad \text{and} \quad \mathscr{R} = \rho\rho^*. \tag{100}$$

The phase term in the numerator of the transmitted amplitude disappears when multiplied by its complex conjugate,* but the denominator contains terms like

$$e^{2iq_bd} = \sin(2n\frac{\omega}{c}d)e^{-2\kappa\frac{\omega}{c}d}.$$

The transmission has oscillations as a function of the frequency or the thickness, with period $2n\omega d/c$.

6.2.5 An example

As an illustration Figs. 27 and 28 show the transmission of a material with a strong harmonic oscillator, such as an optical phonon, in its dielectric function. The media on either side of the material are both taken to have $N_a = N_c = 1$. The parameters for layer b are chosen to resemble those of LiF. The dielectric function is the product dielectric function of Eq. 80,

$$\epsilon = \epsilon_e \frac{\omega_L^2 - \omega^2 - i\omega\gamma_L}{\omega_T^2 - \omega^2 - i\omega\gamma_T}$$

with $\omega_L = 662\,\text{cm}^{-1}$, $\omega_T = 307\,\text{cm}^{-1}$, $\gamma_L = \gamma_T = 2\,\text{cm}^{-1}$, $\epsilon_c = 1.5$ and $d = 0.01$ cm.

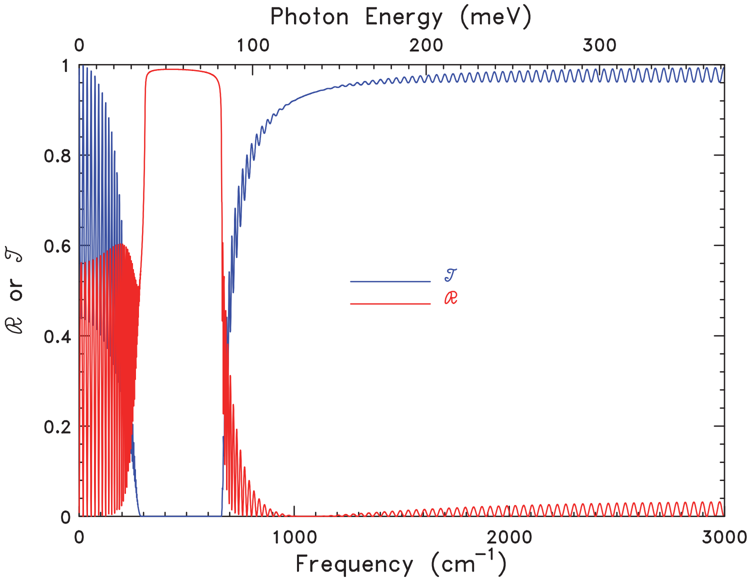

Fig. 27. Transmittance and Reflectance of a slab of “LiF” using a linear frequency scale.

---

[*] The exponential function containing κ does not dissapear of course, and the quantity $e^{-\alpha d}$ appears in the numerator of $\mathscr{T}$.

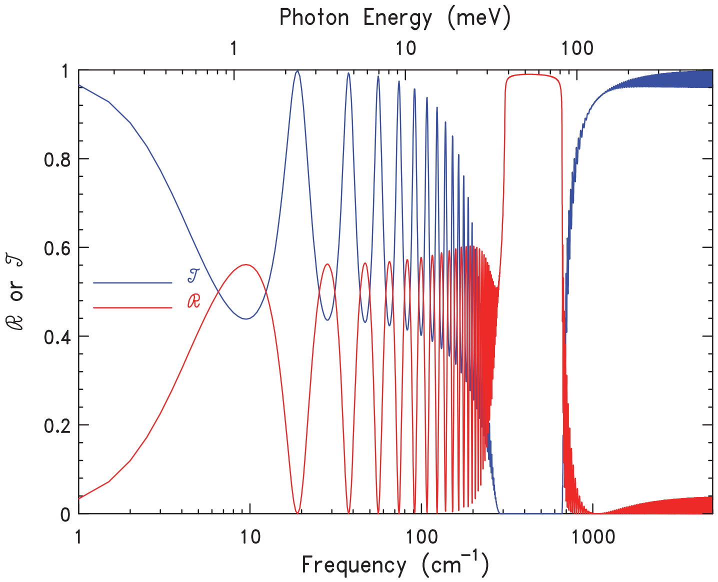

Fig. 28. Transmittance and Reflectance of a slab of “LiF” using a logarithmic frequency scale.

The oscillations in $\mathscr{T}$ and $\mathscr{R}$ are obvious. They occur below ω_T and above ω_L. Because the refractive index is larger below ω_T than above ω_L[*] the period is smaller and the amplitude larger at low frequencies than at high. In the Reststrahlen region between the TO and LO frequencies, the reflection is high and transmission zero. In the transparent regions the maxima in reflection correspond to minima in transmission.[†]

It is not obvious in the figure, but the period of the oscillations becomes shorter as the frequency is increased from zero to ω_T because the refractive index is increasing as the resonance is approached. The dielectric function changes sign at $\omega_L = 662$ cm^{-1}, but the reflection has its minimum at about 1000 cm^{-1}, the frequency where $n = 1$. Because the damping in the model is low, one can see oscillations in the transmittance and reflectance at frequencies below the reflectance minimum.

6.2.6 A second example

As a second example, I’ll calculate the transmittance of a thin metal film. The film will have to be very thin for any light to get through. In principle, the film is thin enough that electron scattering from the surface becomes important, decreasing the conductivity. In the calculations here, I will ignore this likelihood; the optical properties will be those shown in Figs. 8 (conductivity) and 10. The parameters for the model are on page 30; they are meant to mimic silver metal at room temperature. The results of the calculation are shown in Fig. 29 for thicknesses of 10, 30, and 100 Å.

The transmittance is very low at low frequencies and decreases as the square of the film thickness. The flat transmission continues until $\omega \approx 1/\tau$. Above this frequency the transmittance increases, following approximately $\mathscr{T} \sim \omega^2$. At high frequencies the transmittance approaches 100%. The thinner films reach high transmittance at lower frequencies than the thicker films do.

---

[*] $n \approx 2.6$ at low frequencies and ≈ 1.2 at high frequencies.

[†] Energy is conserved.

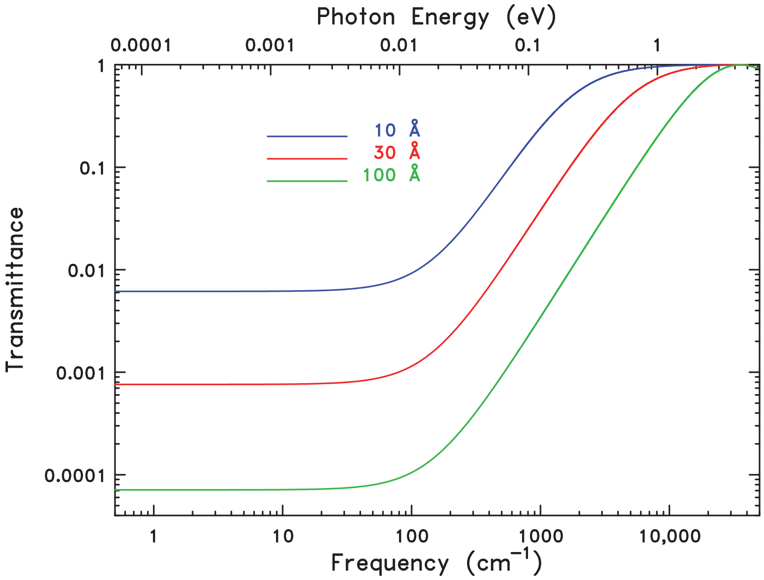

Fig. 29. Transmittance of a metal meant to represent silver. The calculated transmission curves for three thicknesses are shown versus frequency. Both axes use a logarithmic scale.

6.2.7 Layer transmittance and reflectance

It is certainly possible to calculate the transmittance and reflectance—what one measures—from Eqs. 98, 99, and 100 but it may not be profitable to slog through the algebra involved. Instead, let me make a number of simplifications. First, I'll take $N_a = N_c = 1$, i. e., put vacuum on both sides of the layer.* I can then write $N_b = N$, without subscript. Moreover, by symmetry $r_{ba} = r_{bc} \equiv r'$ and $t_{ba} = t_{bc} \equiv t'$, where the prime is to remind me that the light is incident on the vacuum from within the material. The terms where the light is incident on the film from the vacuum are unprimed: $r_{ab} \equiv r = -r_{ba}$ and $t_{ab} \equiv t$. The wavevector inside is $q = q_{ab}$. Then I have from Eqs. 41 and 42

$$\begin{aligned} r' &= \frac{N-1}{N+1} = -r \\ t' &= \frac{2N}{N+1} = 1 + r' \\ t &= \frac{2}{N+1} = 1 - r' \end{aligned}$$

and Eq. 98 becomes (after substituting and multiplying top and bottom by $(N+1)^2 e^{-iqd}$),

$$\tau = \frac{4N}{(N+1)^2 e^{-iqd} - (N-1)^2 e^{iqd}} \tag{101}$$

whereas Eq. 99 becomes (after substituting, multiplying top and bottom by $(N+1)^2$, combining both terms on a common denominator, doing a litle algebra, and then multiplying

---

* This may be too much of a simplification, as many samples consist of a thin film on a thick substrate. But that case is additionally complicated in practice by the fact that the film is thin and, hence, coherent whereas the substrate is thick, and incoherent.

top and bottom by e^{-iqd},

$$\rho = \frac{2i(1-N^2)\sin(qd)}{(N+1)^2 e^{-iqd} - (N-1)^2 e^{iqd}} \tag{102}$$

6.2.8 Non-absorbing layer

If now the layer is non-absorbing ($\kappa = 0$), the transmittance, $\tau\tau*$, becomes (starting with Eq. 101, doing a bit of algebra in the denominator, and making use of the double-angle formula)

$$\mathscr{T} = \frac{4n^2}{4n^2 + (n^2-1)^2 \sin^2(n\omega d/c)}$$

This Airy function has minimum transmittance when $\sin^2(n\omega d/c) = 1$

$$\mathscr{T}_{\min} = \left(\frac{2n}{n^2+1}\right)^2$$

and maximum transmittance when $\sin^2(n\omega d/c) = 0$

$$\mathscr{T}_{\max} = 1.$$

The reflectance is

$$\mathscr{R} = \frac{(n^2-1)^2 \sin^2(n\omega d/c)}{4n^2 + (n^2-1)^2 \sin^2(n\omega d/c)}$$

with maximum

$$\mathscr{R}_{\max} = \left(\frac{n^2-1}{n^2+1}\right)^2$$

and minimum

$$\mathscr{R}_{\min} = 0.$$

The medium is non-absorbing; consequently energy conservation tells me that $\mathscr{T}+\mathscr{R} = 1$. The period of the oscillations in $\mathscr{T}$ and $\mathscr{R}$ is the distance between the zeros of $\sin^2(n\omega d/c)$, which occur when $n\omega d/c = J\pi$ where J is an integer. Thus

$$\Delta\omega = \frac{\pi c}{nd} \tag{103}$$

Here the frequency is in radians/s, $\omega = 2\pi f$, with f the frequency in Hz. Equation 103 then becomes $\Delta f = c/(2nd)$. Finally, if the frequency is in cm^{-1}, $\nu = f/c = 1/\lambda$, and

$$\Delta\nu = \frac{1}{2nd}, \tag{104}$$

with d in cm. Each period corresponds to adding one more cycle to the standing wave inside the layer.

Observation of these multiple-internal-reflection fringes in a measurement allows for a simple, if coarse, determination of the refractive index. One picks out the transmittance maxima, and calculates

$$n(\nu) = \frac{1}{2\Delta\nu d} \tag{105}$$

where $\Delta\nu$ is the spacing between two adjacent maxima and ν is the average of their frequencies. The resolution can be improved by a factor of two by carrying out the same procedure for the minima.

Figure 30 shows a calculation of the transmittance of a non-absorbing layer with variation of the refractive index (top to bottom) from 1.4, 2, 4, 8, 16, to 32. The thicknesses were adjusted to keep nd constant: 70.7, 50, 25, 12.5, 6.25, and 3.125 μm.[*]

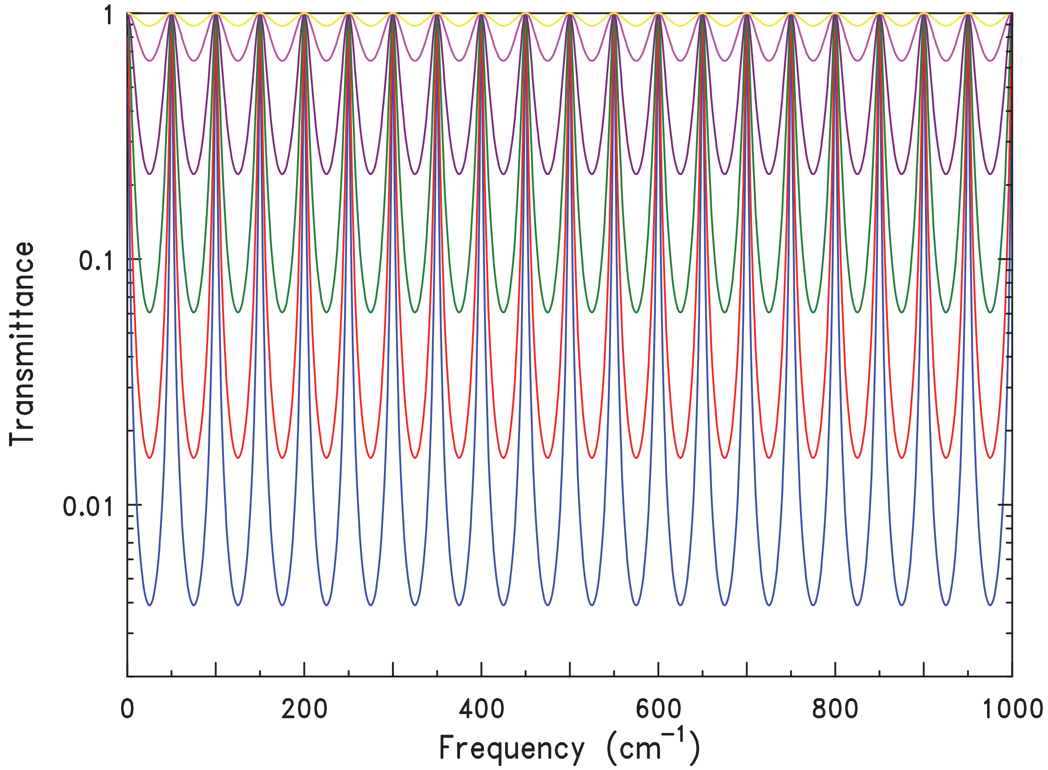

Fig. 30. Transmittance of a layer having a refractive index (top to bottom) of 1.4, 2, 4, 8, 16, and 32 and thickness of 70.7, 50, 25, 12.5, 6.25, and 3.125 μm respectively.

Note that the maximum transmittance is independent of the the refractive index or reflection coefficient of the vacuum-material interface; it is always unity[†] when the partial waves for the multiply-internally-reflected light all add constructively. (See Fig. 26.) As the reflection coefficient of the surfaces increase, the minima between the 100% transmission peaks become deeper and deeper, and the shape of the transmission spectra, which

---

[*] I note that a refractive index of 32 (dielectric constant of 1024) is uncommon. But the same effect—a high reflectance at the interfaces—can be achieved by applying multilayer dielectric coatings to the surfaces of a low index spacer.

[†] Assuming no absorption!

are close to sinusoidal for low reflection surfaces, becomes more and more a set of narrow peaks between low transmission regions. This spectrum is the transmission of a Fabry-Perot interferometer.

I'll not plot the reflectance, but I can sketch it by looking at Fig. 30 because I know (energy conservation!) that $\mathscr{R} = 1 - \mathscr{T}$. So for $n = 1.4$, the reflectance would oscillate between zero and about 10%; the reflectance maxima correspond to transmittance minima. For $n = 32$, the reflectance exceeds 99% over much of the spectrum, punctuated by narrow dips to zero when the partial waves inside interfere constructively.

The material with $n = 32$ has a single-surface reflectance $(n-1)^2/(n+1)^2 = 0.88$. So it is reasonable to ask: how can a slab with two such highly reflecting surfaces in series have transmittance equal to 1 and zero reflectance? The answer of course is that for most frequencies, it does not. But in the special case where all the partial waves interfere constructively, the energy stored in the slab grows enormously. Some of this energy leaks out of the front surface with a phase that is 180° out of phase with the light promptly reflected; the two sum to the zero reflectance. Some light also leaks out of the back and energy conservation tells me that this leakage is equal to the incident intensity, making $\mathscr{T} = 1$.

6.3 The matrix method

The use of "ABCD" matrices to calculate wave propagation is used in a variety of guises, such as to follow Gaussian laser beams through lenses, mirrors, and other optical components.[18,19] Here, following the lead of Heavens,[20] I'll discuss how it is done for multilayer thin films.

MORE TO COME

6.4 Inverting $\mathscr{R}$ and $\mathscr{T}$ to find ϵ

If my goal is to use experimental measurements of $\mathscr{R}$ and $\mathscr{T}$ to find out about the the optical constants (n, κ, ϵ, ...), I'll need to figure out how to invert the transcendental equations for the measured quantities to extract the things in which I am interested. The inversion is not trivial, and will involve some approximations or computation.

6.4.1 The Glover-Tinkham model

MORE TO COME

Problems

Write $\mathscr{T}$ for incoherent case; show that it is possible to solve quadratic for A.

Show that $\mathscr{R}$ and $\mathscr{T}$ in incoherent case may be combined to eliminate A and to extract $\mathscr{R}_{12}$

Compare $\mathscr{R}$ incoherent with and without backside

Backside reflectance of thin film on thick substrate.

Glover-Tinkham with $\sigma_1 = 0$. Show R + T = 1.

Invert GT free standing to find σ_1 and σ_2.

7. FREE-ELECTRON METALS: QUANTUM THEORY

Here I will consider the free-electron theory of metals. This theory was developed in the late 1920s and early 1930s by Sommerfeld and others. The basic idea is that one or more electrons of each atom in the solid are detached from their atomic location and free to move throughout the crystal as a free gas. The electrons are fermions, subject to the Pauli principle, and therefore follow the Fermi-Dirac distribution function. Each state is occupied by at most one electron. The electron density is high enough that the Fermi temperature is much higher than the physical temperature, so the Fermi-Dirac function is relatively sharp. The electrons are said to form a "Fermi gas" with a Fermi surface that is the constant-energy surface of the most energetic electrons.

The Fermi-Dirac statistics overwhelms other forces. Thus, the free-electron model treats the potential energy in which the electron moves (the ion potential) as a constant. (Including the potential can be done of course; this is the realm of electronic band structure. Band structure calculations have been done for many materials systems—not just elements. For many metals, the effect is mostly to modify slightly the effective mass of the charge carriers, to adjust the numerical value of the density of states at the Fermi surface, and to find bands above and below the one where the free carriers live. I discussed in the look at real solids just completed that the band structure—in particular the presence of the filled *d*-electron levels below the Fermi level—is the reason why gold, silver and copper appear different. The basic low-energy physics is however very close to what the free-electron model predicts.*)

The electron-electron interaction is also left out of the free-electron model, so that one individual electron state is independent of the other electron states. Here, I can take refuge in Landau's Fermi-liquid theory. This theory of interacting fermions describes the properties of metals in cases where the interaction between the particles of the many-body system is not small. There is a one-to-one correspondence between the elementary excitations of a Fermi gas system and a Fermi liquid system. These excitations are called "quasiparticles;" the quasiparticle states have the same quantum numbers (momentum spin, charge) as the electrons of the Fermi gas.

7.1 Schrödinger equation for free electrons

The simple free-electron model is surprisingly successful in explaining many experimental phenomena, including electrical conductivity, heat conductivity, the Wiedemann-Franz law (a relation of the ratio of electrical conductivity to thermal conductivity), the electronic contribution to the heat capacity, Hall effect, and optical properties. I'll start with the Hamiltonian, a sum of kinetic and potential energies:

$$\mathcal{H} = \frac{p^2}{2m} + \mathcal{V},$$

where p is the momentum, m the mass, and $\mathcal{V}$ the potential. In general I would have $\mathcal{V} = \mathcal{V}(\mathbf{r})$ where $\mathbf{r}$ is the location of the electron and the potential would be periodic in the lattice structure and account for he attraction of the electron to the positive ions in the

---

* In contrast, band structure is essential for understanding semiconductors and insulators.

crystal. But in the free electron model I take $\mathcal{V}$ = constant and I might as well take it to be zero.*

My quantum mechanics class told me that the momentum operator is $\mathbf{p} = -i\hbar\nabla$, so that the free electron Hamiltonian is

$$\mathcal{H} = -\frac{\hbar^2\nabla^2}{2m} \tag{106}$$

and the time independent Schrödinger equation is

$$\mathcal{H}\psi = \mathcal{E}\psi, \tag{107}$$

where $\mathcal{E}$ is the energy eigenvalue and ψ is the wave function of the single electron I am considering.

7.2 Wave function

Consider a plane-wave wave function:

$$\psi = \psi_0 e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{108}$$

with ψ_0 a complex amplitude and $\mathbf{k}$ a wave vector to be determined by solving the Schrödinger equation.

Of course (Appendix C) $\nabla^2\psi = -k^2\psi$. I need to normalize the wave function:

$$1 = \langle\psi|\psi\rangle = \int_V dV\,\psi^*(\mathbf{r})\psi(\mathbf{r}) = \int_V dV\,|\psi_0|^2 = |\psi_0|^2 V$$

where V is the entire volume of the crystal. Then,

$$\psi_0 = \frac{e^{i\phi}}{\sqrt{V}}$$

with ϕ an arbitrary phase. (The time-dependent Schrödinger equation leads to a phasor $e^{-i\omega t}$ where the frequency ω is related to the energy $\mathcal{E}$ by Planck's constant: $\mathcal{E} = \hbar\omega$.)

Substituting Eq. 108 into Eq. 107 with the Hamiltonian of Eq. 106 yields

$$+\frac{\hbar^2 k^2}{2m}\psi = \mathcal{E}\psi$$

Multiply from the left by ψ^* and integrate to find

$$\mathcal{E} = \frac{\hbar^2 k^2}{2m}. \tag{109}$$

Eq. 109 is just the kinetic energy of a free particle. The momentum operator is $\mathbf{p} = -i\hbar\nabla$ and so $\langle\psi|\mathbf{p}|\psi\rangle = \hbar\mathbf{k}$ and so I can identify $\hbar\mathbf{k}$ as the momentum. The plane-wave wave

---

[*] I need a wall at the surface to keep the electrons in the crystal, but the choice of $\mathcal{V}$'s constant value has no effect. I take the wall height to be infinite. (And then I eliminate the walls by using periodic boundary conditions.)

functions are eigenfunctions of both the Hamiltonian and of momentum. The energy is quadratic in k, with zero energy* at $k = 0$. The probability density to find the electron at any position in the crystal is constant, $P(\mathbf{r}) = \psi^*\psi = 1/V$. For conceptual clarity, let me assume a cube of edge length $L = Ja$ where a is the lattice constant of the crystal and J is a (large) integer. The volume is $V = L^3 = J^3a^3$. If L is 1 cm, then the solid contains about 10^{22} atoms and $J \approx 2.2 \times 10^7$.

7.3 Exclusion principle and boundary conditions

The electrons are fermions; only one electron can have a particular set of quantum numbers. The quantum numbers here are the three components of the wave vector $\mathbf{k}$ and the spin: $\{k_x, k_y, k_z, \sigma\}$ where $\sigma =\uparrow$ or $\downarrow$. The allowed values of the wave vector components are set by the boundary conditions. If I consider particle-in-a-box conditions, then the wave function will be zero outside the solid, and the solutions are sines and/or cosines depending on the choice of coordinate-system origin.

It is much better,† however, to use "periodic boundary conditions." I'll write these mathematically as

$$\psi(\mathbf{r} + \mathbf{L}) = \psi(\mathbf{r}), \tag{110}$$

where $\mathbf{L}$ is oriented along either the $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, or $\hat{\boldsymbol{z}}$ directions. To begin, think about a one-dimensional chain of $J = N$ atoms. Such a chain is illustrated on the left side of Fig. 31. Then one takes the chain and bends it into a circle, connecting the first and Nth atoms with a bond. Clearly, this loop satisfies Eq. 110 because the wave function must be single valued, and give the same value every time one goes once around the chain.

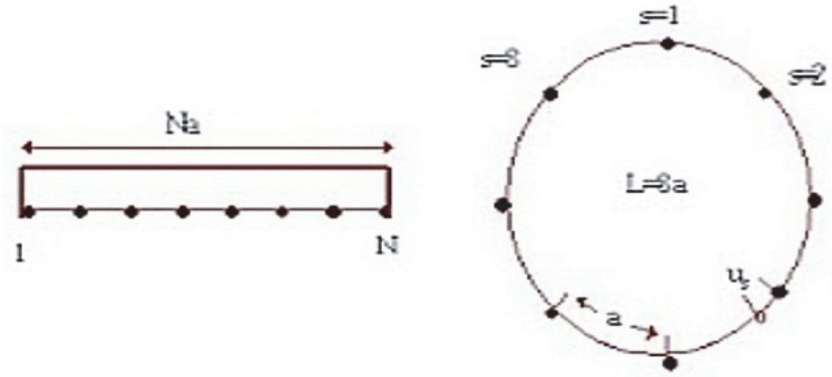

Fig. 31. A linear chain of N atoms is bent into a circle and the first and Nth atoms are connected by a bond. The system is periodic with period Na.

One may also think in a slightly different way, illustrated in Fig. 32. I take the block and make copies of it and lay them side by side. The boundary is moved off to infinity, but the fact that the copies are exact means that Eq. 110 is obeyed.

---

* Recall: the potential energy is at zero.

† The sine/cosine solutions yield standing waves in the block of material, just as the case of waves on a string with fixed ends. These waves carry no momentum and do not contribute to the current until one forms the proper linear combinations.

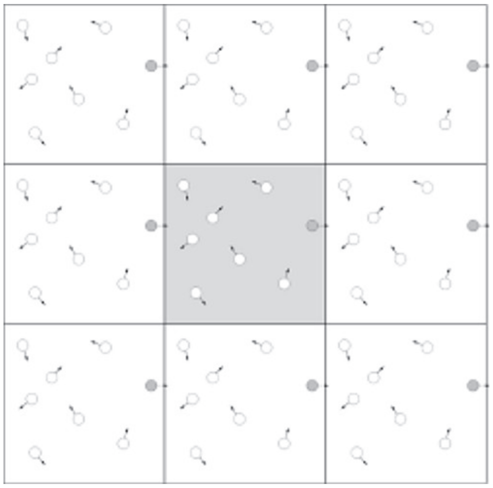

Fig. 32. A square system containing N atoms is repeated many times and the copies laid side to side. The system is periodic in both directions with period $\sqrt{N}a$.

I will apply periodic boundary conditions to the free-electron wave function of Eq. 108 for $\mathbf{L} = Ja\hat{\boldsymbol{x}}$:

$$\psi(\mathbf{r}) = \psi_0 e^{i\mathbf{k}\cdot\mathbf{r}} \quad = \quad \psi(\mathbf{r}+\mathbf{L}) = \psi_0 e^{i\mathbf{k}\cdot(\mathbf{r}+Ja\hat{\boldsymbol{x}})}$$

with a the lattice constant. Now, $\mathbf{k}\cdot\hat{\boldsymbol{x}} = k_x$. The periodic boundary conditions for this specific value of $\mathbf{L}$ then yield

$$e^{ik_xJa} = 1 \tag{111}$$

This equation has a large number of solutions; the phase of the complex exponential k_xJa must be an integer (positive or negative) times 2π radians. Hence

$$k_x = \frac{2\pi}{Ja}j_x \qquad \text{where } j_x = 0,\ \pm 1,\ \pm 2,\ \ldots$$

Applying periodic boundary conditions in the $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$ directions give equivalent quantization of k_y and k_z. Because the components of the wave vector take on essentially integer values and not continuous ones, the energy, Eq. 109 is also quantized at values:

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2k^2}{2m} = \frac{\hbar^2}{2m}\left(\frac{2\pi}{Ja}\right)^2(j_x^2 + j_y^2 + j_z^2)$$

I note that there are many degeneracies in the energy spectrum of the free electron. The two values of spin (up and down) are degenerate as well as many permutations of integers (positive and negative). These permutations are distinguishable because the momentum is

$$\mathbf{k} = (2\pi/Ja)(j_x\hat{\boldsymbol{x}} + j_y\hat{\boldsymbol{y}} + j_z\hat{\boldsymbol{z}}). \tag{112}$$

The Pauli principle says that no two electrons can have the same quantum numbers. I can call the set $\{j_x, j_y, j_z, \sigma\}$ the quantum numbers. I now pour electrons into the metal, stopping when I reach electrical neutrality.* The first few energies are

$$\begin{array}{lllll} \mathcal{E}_0 = 0 & \{0,0,0,\uparrow\} & \{0,0,0,\downarrow\} & & \\ \mathcal{E}_1 = \delta & \{1,0,0,\uparrow\} & \{0,1,0,\uparrow\} & \ldots & \{0,0,1,\downarrow\} \\ \mathcal{E}_2 = 2\delta & \{1,1,0,\uparrow\} & \{1,0,1,\uparrow\} & \ldots & \{0,1,1,\downarrow\} \end{array}$$

where $\delta = (2\pi\hbar)^2/2mJ^2a^2$ is a typical spacing between levels.

7.4 The Fermi energy

The values of $\mathbf{k}$ allowed by the boundary conditions form a grid of points in when plotted in three-dimensional space† with axes k_x, k_y, k_z. Because the energy $\mathcal{E}$ is quadratic in the wave vector k, the surfaces of constant energy are spheres of radius k. Indeed,

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2}{2m}\left(\frac{2\pi}{Ja}\right)^2 j^2$$

where $j^2 = j_x^2 + j_y^2 + j_z^2$ is an integer.

The highest energy of any of these electrons (at $T = 0$) is called the Fermi energy:

$$\mathcal{E}_F = \frac{\hbar^2}{2m}k_F^2 = \frac{\hbar^2}{2m}\left(\frac{2\pi}{Ja}\right)^2 j_F^2$$

where k_F is the radius of the sphere corresponding to the highest energy electrons and j_F is the radius of an equivalent sphere in coordinates of the integers j_x, j_y, j_z. A diagram of this Fermi sphere is shown in Fig. 33.

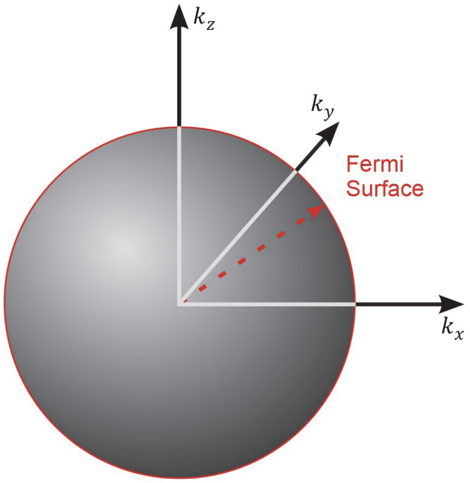

Fig. 33. The Fermi sphere in momentum space.

---

[*] One to a few electrons per atom in simple metals; one to a few per formula unit in metallic compounds.

[†] This space is called “momentum space,” “k-space,” or “reciprocal space.” It has units of inverse length.

Now let me calculate the radius of the Fermi surface. There are N electrons in the metal. Because of the spin degeneracy, two of these electrons occupy each point on the grid of points indexed by j_x, j_y, and j_z. The total number of points inside the sphere is just the volume of the sphere in "j-space." Hence,

$$\frac{N}{2} = \frac{4\pi}{3} j_F^3$$

or

$$j_F = \left(\frac{3N}{8\pi}\right)^{1/3}.$$

The Fermi wave vector is

$$k_F = \frac{2\pi}{Ja} j_F = \frac{1}{Ja}\left(3\pi^2 N\right)^{1/3}.$$

But $(Ja)^3 = V$ or $Ja = V^{1/3}$. Hence

$$k_F = \left(\frac{3\pi^2 N}{V}\right)^{1/3} = \left(3\pi^2 n\right)^{1/3} \tag{113}$$

where $n = N/V$ is the number density of electrons.

Finally, the Fermi energy is

$$\mathcal{E}_F = \frac{\hbar^2}{2m}\left(3\pi^2 n\right)^{2/3} \tag{114}$$

Related to the Fermi energy is the the Fermi temperature T_F, defined as $\mathcal{E}_F/k_B$, where k_B is Boltzmann's constant. I also define the Fermi momentum p_F and Fermi velocity v_F.

$$p_F = \hbar k_F = \sqrt{2m\mathcal{E}_F}$$

and

$$v_F = \frac{p_F}{m}.$$

These quantities correspond respectively to the momentum and group velocity of an electron at the Fermi surface. Fermi surface parameters for many metals are shown in Table 7, after Ref. 11

Fermi energies are in the range of 1.5–15 eV; Fermi temperatures are thus 18,000–180,000 K. These temperatures are much above 300 K, the temperature of the room. I did not make a bad mistake by considering (as I have done until now) $T = 0$. The Fermi velocities are surprisingly large; they are in the range 0.8–2×10^8 cm/sec. The average is about $c/200$. Our quintessential metal, silver has $\mathcal{E}_F = 5.5$ eV, $T_F = 64,000$ K and $v_F = 1.4 \times 10^8$ cm/s. I'll use these numbers when numbers are needed.

Table 7. Free electron Fermi surface parameters for simple metals

Valence	Metal	n 10^{22} cm^{-3}	k_F 10^8 cm^{-1}	v_F 10^8 cm/s	$\mathcal{E}_F$ eV	T_F 10^4 K
1	Li	4.70	1.11	1.29	4.72	5.48
	Na	2.65	0.92	1.07	3.23	3.75
	K	1.40	0.75	0.86	2.12	2.46
	Rb	1.15	0.70	0.81	1.85	2.15
	Cs	0.91	0.64	0.75	1.58	1.83
	Cu	8.45	1.36	1.57	7.00	8.12
	Ag	5.85	1.20	1.39	5.48	6.36
	Au	5.90	1.20	1.39	5.51	6.39
2	Be	24.2	1.93	2.23	14.14	16.41
	Mg	8.60	1.37	1.58	7.13	8.27
	Ca	4.60	1.11	1.28	4.68	5.43
	Sr	3.56	1.02	1.18	3.95	4.58
	Ba	3.20	0.98	1.13	3.65	4.24
	Zn	13.10	1.57	1.82	9.39	10.90
	Cd	9.28	1.40	1.62	7.46	8.66
3	Al	18.06	1.75	2.02	11.63	13.49
	Ga	15.30	1.65	1.91	10.35	12.01
	In	11.49	1.50	1.74	8.60	9.98
4	Pb	13.20	1.57	1.82	9.37	10.87
	Sn(w)	14.48	1.62	1.88	10.03	11.64

7.5 The effect of temperature

At zero temperature the Fermi-Dirac is sharp because every electron is in the lowest energy state that it can access. At finite temperatures, the probability of occupancy is given by the Fermi-Dirac distribution function $f(\mathcal{E})$,

$$f(\mathcal{E}) = \frac{1}{1 + e^{\frac{\mathcal{E}-\mu}{k_B T}}} \tag{115}$$

where μ is the chemical potential or Fermi level. At zero temperature $\mu = \mathcal{E}_F$; it has a weak temperature dependence so long as T does not approach T_F. A plot of the Fermi-Dirac distribution function is shown in Fig. 34 for a metal with $\mathcal{E}_F = 5$ eV. The effect of temperature is to broaden the distribution. Some states below the Fermi level are empty; some above it are occupied.

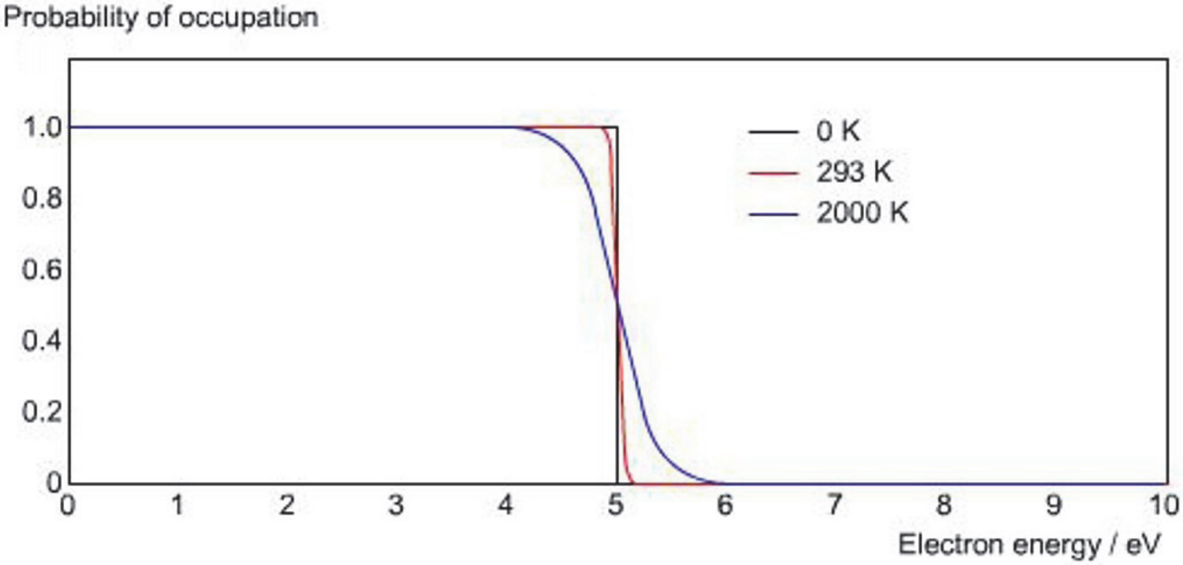

Fig. 34. The Fermi-Dirac distribution function $f(\mathcal{E})$ at three temperatures.

7.6 The density of states

I may need the density of states, so let me work it out now. I can write an equation for $N_\mathcal{E}$ the number of electrons with energy between the zero of energy and energy $\mathcal{E}$ by noting that Eq. 114 is such an equation, so

$$\mathcal{E} = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 N_\mathcal{E}}{V}\right)^{2/3}$$

so that

$$N_\mathcal{E} = \left(\frac{2m\mathcal{E}}{\hbar^2}\right)^{3/2}\frac{V}{3\pi^2}$$

I now raise $\mathcal{E}$ by just a little bit $d\mathcal{E}$ and find that the number goes up by dN.

$$N_\mathcal{E} + dN = \left(\frac{2m(\mathcal{E}+d\mathcal{E})}{\hbar^2}\right)^{3/2}\frac{V}{3\pi^2}$$

Because $d\mathcal{E}$ is infinitesimal, I can expand $(\mathcal{E}+d\mathcal{E})^{3/2} = \mathcal{E}^{3/2}(1 + 3d\mathcal{E}/2\mathcal{E} + \ldots)$ to find

$$N_\mathcal{E} + dN = \left(\frac{2m\mathcal{E}}{\hbar^2}\right)^{3/2}\frac{V}{3\pi^2}(1 + \frac{3d\mathcal{E}}{2\mathcal{E}})$$

The first term on the right is just $N_\mathcal{E}$ and the second is dN. The density of states (the number of states between $\mathcal{E}$ and $\mathcal{E}+d\mathcal{E}$) is

$$\frac{dN}{d\mathcal{E}} = \mathscr{D}(\mathcal{E}) = \left(\frac{2m}{\hbar^2}\right)^{3/2}\frac{V}{2\pi^2}\sqrt{\mathcal{E}}.$$

The density of states* grows as the energy increases as the square root of the energy.†

---

* Sometimes the symbol $N(\mathcal{E})$ is used for the density of states.

† The $\sqrt{\mathcal{E}}$ behavior is correct for three-dimensional metals, where the energy depends on three—x, y, and z—components of **k**. In two dimensions $\mathscr{D}(\mathcal{E})$ is constant and in one dimension $\mathscr{D}(\mathcal{E}) \sim 1/\sqrt{\mathcal{E}}$.

7.7 Electrical conductivity

In the absence of an external field, the metal does not carry current. The electrons are moving in all directions and the average or net motion is zero.

Consider now the effect of an electric field on these electrons. To be definite, I will specify the field as our plane wave traveling along $\hat{\boldsymbol{x}}$ with electric field oriented in the $\hat{\boldsymbol{y}}$ direction.

$$\mathbf{E} = \hat{\boldsymbol{y}} E_0 e^{i(qx-\omega t)}. \tag{116}$$

The field exerts a force on each and every charged electron:

$$\mathbf{F}_{\text{ext}} = -e\mathbf{E} = -e\hat{\boldsymbol{y}} E_0 e^{i} qxe^{-i\omega t}.$$

I can put the space variation into $\mathbf{E}_0$ and ignore the fact that the field varies in space with a length scale $\lambda = 2\pi/q$. The reason is that the wavelength λ is long compared to the effective range of the electron (the mean free path ℓ) and also large compared to the lattice constant. I also have not made a local-field correction. The electron is completely delocalized.[*] Hence the local field is the average field, which is the applied field.

Newton's second law, $F = ma$ is best written

$$\sum \mathbf{F} = \frac{d\mathbf{p}}{dt} = \hbar \frac{d\mathbf{k}}{dt} \tag{117}$$

So the external force and any internal forces will change the k-state of the electron. The resultant response is not hindered by the Pauli exclusion principle: as the electron in state k evolves to k', the one at k' has gotten out of the way by shifting to k''.

What other forces act on the electrons? Well, the electrons are not bound; indeed, the potential energy has been set to zero. But the electrons can suffer collisions with impurities, other defects, lattice vibrations (phonons), and the surface. These collisions have a mean free time τ and relax the system back towards equilibrium. In the spirit of Drude's model, I will consider that the relaxation generates an impulse $\mathbf{F}_{\text{coll}}\tau$ which changes the momentum $\hbar\mathbf{k}$ back to the equilibrium value[†] $\hbar\mathbf{k}_0$. Thus,

$$\mathbf{F}_{\text{coll}} = -\hbar \frac{\delta \mathbf{k}}{\tau} = -\hbar \frac{\mathbf{k} - \mathbf{k}_0}{\tau}. \tag{118}$$

The collisions enter the picture as abrupt changes in momentum that occur every τ seconds. Note that $\hbar\delta\mathbf{k} = m\delta\mathbf{v} = m\mathbf{v}_{\text{drift}}$ where $\mathbf{v}_{\text{drift}}$ is the average velocity acquired by the collection of electrons, known as the drift velocity.[‡]

Let me think for a moment about how the applied electric field, with its $e^{-i\omega t}$ time dependence, affects the motion of an electron. The electron has wave vector $\mathbf{k}_0$ or velocity

---

[*] ψ has amplitude everywhere in the crystal.

[†] Drude considered the equilibrium value to be zero, but I know that it is close to the Fermi momentum.

[‡] Without the electric field of the light, the average velocity is *zero*. The average speed is some fraction of the Fermi velocity, but speed and velocity are different things.

$\mathbf{v}_0 = \hbar\mathbf{k}_0/m$; this velocity will be changed (in magnitude, direction, or both) by the field. I will write the $\mathbf{k}$ of my electron in the presence of the field as

$$\mathbf{k} = \mathbf{k}_0 + \delta\mathbf{k}\, e^{-i\omega t}, \tag{119}$$

with $\delta\mathbf{k}$ in the direction of the field, $\hat{\boldsymbol{y}}$. $\delta\mathbf{k}$ represents the linear response of the electron to the external field. $\mathbf{k}_0$ is the wave vector in the absence of the field; it does not depend on the time. δk contains the amplitude and phase of the driven motion of the electron. Newton's second law becomes

$$\hbar\frac{d\mathbf{k}}{dt} = -\hbar\frac{\mathbf{k}-\mathbf{k}_0}{\tau} - e\mathbf{E}. \tag{120}$$

On substituting for $\mathbf{k}$ and $\mathbf{E}$, taking derivatives, and canceling $e^{-i\omega t}$, I get

$$-i\hbar\omega\delta\mathbf{k} = \frac{-\hbar\delta\mathbf{k}}{\tau} - e\mathbf{E}_0 \tag{121}$$

A little algebra then yields this relation between the shift of $\mathbf{k}$ and the electric field:

$$\delta k = -\frac{e\tau/\hbar}{1-i\omega\tau}E_0, \tag{122}$$

where all vectors are in the $\hat{\boldsymbol{y}}$ direction. Equation 122 looks a lot like the Drude velocity, Eq. 53, except that $\hbar$ appears. The appearance will be fleeting, however, and the theory is at best semiclassical.

The picture of what happen in the metal due to the external electric field is not intuitive. Every electron has changed its $\mathbf{k}$-vector under the influence of the electric field. This displacement occurs alike and by the same amount to the deeply buried electrons with small $\mathbf{k}$ values and low speeds and to the Fermi-surface electrons which are whizzing along at $c/200$. The Fermi surface, which is centered at $\mathbf{k} = 0$ in equilibrium is moved as a whole by $\delta\mathbf{k}$, as illustrated in Fig. 35.

Were there no scattering, the Fermi sphere would continue to displace: the electric field exerts a force on the electrons, causing acceleration. At zero frequency, both $\delta\mathbf{k}$ and the velocity of each electron would change linearly with the time, increasing for left-moving electrons and decreasing for right-moving electrons. With scattering, eventually the relaxation balances the acceleration and there is a maximum, steady-state* displacement. Because scattering events are uncorrelated and affect one or the other of the independent and non-interacting electrons, the relaxation process, shown schematically in Fig. 35, must respect the Pauli principal, taking the electron from a filled state to an empty state. If the scatting is elastic, then the energy of the initial and final states should be the same. So the relaxation process takes an electron from the leading edge of the displaced Fermi surface, where electrons occupy states that are empty in equilibrium, and deposits it just outside the trailing edge, where states that once were occupied have been emptied.

---

* But not equilibrium.

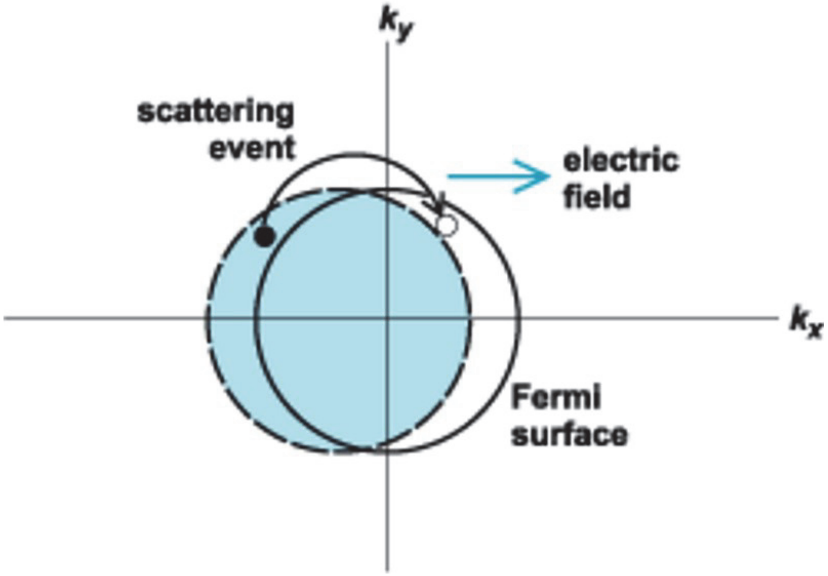

Fig. 35. The Fermi sphere in equilibrium is show uncolored and with a solid circle indicating the projected sphere. When an electric field is applied along $\hat{\boldsymbol{x}}$, the electrons are accelerated to the left. (They are negatively charged.) The displaced sphere is shown in blue, with a dashed circle. The acceleration continues until it is balanced by the scattering of electrons; one such event is shown.

The electrons that are relaxed are therefore very close to the Fermi energy and have speeds of the Fermi velocity.* Thus the relation between the mean free time τ and mean free path ℓ is

$$\ell = v_F \tau \tag{123}$$

The electrical current is $\mathbf{j} = -ne\mathbf{v}_d$ where n is the electron density and $\mathbf{v}_d$ is an average or drift velocity. This velocity is *not* the Fermi velocity, but the average of all electrons' responses to the field. The averaging process goes like:

$$\mathbf{v}_d = \langle \mathbf{v_k} \rangle = \langle \mathbf{v}_0 + \delta\mathbf{v} \rangle = \langle \delta\mathbf{v} \rangle = \delta\mathbf{v}$$

where the last equality comes from the fact that every electron feels the same force. In the above, the velocity of an electron with wave vector $\mathbf{k}$ is $\mathbf{v_k} = \mathbf{p_k}/m = \hbar\mathbf{k}/m$ and $\mathbf{v}_d = \hbar\delta\mathbf{k}/m$, making

$$\mathbf{j} = -\frac{ne\hbar}{m}\delta\mathbf{k}. \tag{124}$$

I then substitute $\delta\mathbf{k}$ from Eq. 122 to obtain

$$\mathbf{j} = -\left(\frac{ne\hbar}{m}\right)\left(\frac{-e\tau/\hbar}{1-i\omega\tau}\right)\hat{\boldsymbol{y}}E_0.$$

Using Ohm's law, $\mathbf{j} = \sigma\mathbf{E}$ I find finally

$$\sigma = \frac{ne^2\tau/m}{1-i\omega\tau}, \tag{125}$$

an equation identical to Eq. 54. Equation 125 gives a complex conductivity. The dielectric

---

[*] The realization that electron relaxation involves the fast-moving Fermi-surface electrons is where quantum mechanics (and quantum statistical mechanics) leads to new concepts in the Drude theory.

function and other optical constants follow immediately as I have shown many times before. I write $\epsilon = \epsilon_c + 4\pi i\sigma/\omega$ to allow for core polarizability and have

$$\epsilon = \epsilon_c - \frac{\omega_p^2}{\omega^2 + i\omega/\tau} \tag{126}$$

with ω_p the plasma frequency,

$$\omega_p = \sqrt{\frac{4\pi ne^2}{m}}, \tag{127}$$

as before.

7.8 Discussion of the Drude model

7.8.1 Low frequencies and the steady-state

To think a bit more about the conduction process in metals, let me first address zero frequency (or any frequency low with respect to $1/\tau$.) In this case, I can take $dv/dt \approx 0$, which is the same as saying that $d\mathbf{k}/dt \ll (\mathbf{k} - \mathbf{k}_0)/\tau$ in Eq. 120. This case is the steady state (not equilibrium!); the acceleration is zero, but only after the initial transient when the field is turned on has died away. The conductivity is

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

and Eq. 122 becomes

$$\delta k = -\frac{e\tau}{\hbar}E_0$$

The Fermi surface is displaced by an amount δk in the negative field direction as shown in Fig. 35. Now, the energy dispersion is still $\mathcal{E} = \hbar^2 k^2/2m$ so that the average energy of the electrons has increased. (The field did work on them.) Figure 36 shows the energy dispersion and Fermi energy as a function of k_y (for, say $k_x = k_z = 0$) in equilibrium (left) and with an electric field applied and corresponding steady-state current flow. Remember that $m\mathbf{v} = \hbar\mathbf{k}$; for electrons moving in the direction of electron flow, the most energetic ones have a bit higher energy than in equilibrium. In contrast, for electrons moving against the direction of electron flow, the most energetic ones have a bit smaller energy than in equilibrium. Scattering can take electrons from the first to the second direction.

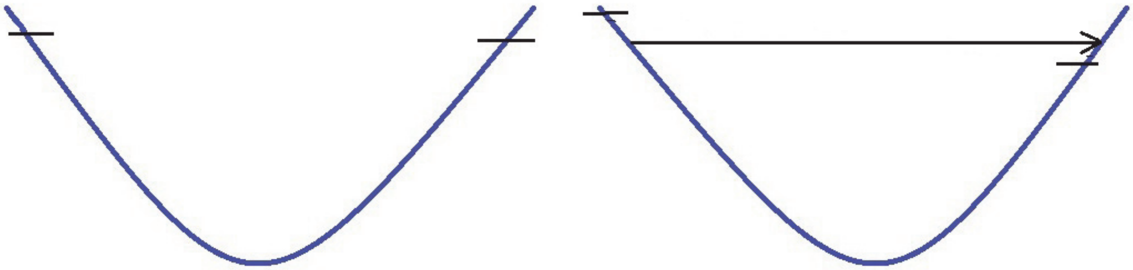

Fig. 36. Energy plotted vs k_y in equilibrium (left) and with a steady current flow (right). An elastic scattering event is also shown.

Scattering from impurities, other defects, and the surface generally is elastic. The electron momentum (k-vector) is changed but not the energy. Scattering by absorption of phonons is inelastic, but phonon energies are tens of meV whereas the Fermi energy is 1000 times higher. So on the scale of Fig. 36 the distinction is not visible.[*]

7.8.2 More numbers

Figures 35 and 36 suggest that δk is large. It is not. Let me estimate it from Eq. 124. I'll take the current density as 10^7 A/cm^2. (This is a huge current; the electrical code typically restricts current densities in copper house wiring[22] to 115 A/cm^2.) Given that e is 1.6×10^{-19} C, the current is carried by a flow of 0.6×10^{26} electrons/s cm^2. I invert Eq. 124: $\delta\mathbf{k} = -m\mathbf{j}/ne\hbar = 10^4\ \mathrm{cm}^{-1}$. Compare this to the Fermi wave vector (Table 7) of order $10^8\ \mathrm{cm}^{-1}$. Very high current densities displace the Fermi surface by only 0.01% of its radius.[†]

The energy change is also small. The energy of the Fermi-surface electrons moving in the $-y$ direction is increased to $\hbar^2(k_F + \delta k)^2/2m$. Expanding, we see that the energy goes to

$$\mathcal{E}_F + \delta\mathcal{E} = \frac{\hbar^2 k_F^2}{2m} + \frac{\hbar^2 k_F \delta k}{m}$$

or

$$\frac{\delta\mathcal{E}}{\mathcal{E}_F} = \frac{2\delta k}{k_F} \approx 10^{-4}$$

This value corresponds to the energy spread of the Fermi surface at temperatures of 1 K or so. So at most temperatures, (and most currents) the thermals spread of the Fermi distribution is much larger than the displacement due to the current.

7.8.3 The perfect conductor

Now suppose there are no collisions; the metal is a perfect conductor.[‡] The equation of motion (Eq. 120) is simplified.

$$\hbar\frac{d\mathbf{k}}{dt} = -e\mathbf{E}. \tag{128}$$

and the solution is (at finite frequencies)[§]

$$\delta k = -\frac{ie}{\hbar\omega}E_0$$

---

[*] There are phenomena where the distinction is important. Elastic vs. inelastic scattering plays a large role in localization phenomena where the phase of the wave function is critical. In the far-infrared response at low temperatures there is the Holstein effect,[21] where phonon emission affects the optical conductivity at frequencies equal to or larger than the phonon frequencies.

[†] I would be more correct if I said that very high current densities increase the speed of the fastest electrons—those at the Fermi surface—by only 0.01% of their zero-current speed.

[‡] But not a superconductor. The perfect conductor does not exhibit the Meissner effect, Josephson effect, or the energy gap in the optical conductivity. See section 11

[§] At dc, the solution is a δk that grows linearly with time: a constant acceleration.

The oscillations of k are 90° out of phase with the field. Equation 124 is unchanged and so, after a bit of algebra I find for the conductivity.

$$\sigma = i\frac{ne^2}{m\omega} \tag{129}$$

This expression is the high frequency limit of Eq. 125, and that makes sense because no scattering means that the mean free time for collisions is infinite and *any* finite frequency makes $\omega\tau$ big compared to 1. We have ($\omega > 0$)

$$\begin{aligned} \sigma_1 &= 0 \\ \sigma_2 &= \frac{ne^2}{m\omega} \end{aligned}$$

The dielectric function of the perfect metal is purely real.*

$$\epsilon = \epsilon_c - \frac{4\pi ne^2/m}{\omega^2} = \epsilon_c - \frac{\omega_p^2}{\omega^2}.$$

So both conductivity and dielectric function tell us that the perfect metal does not dissipate electromagnetic energy.

But wait! ϵ_1 is negative for frequencies below $\omega_p/\sqrt{\epsilon_c}$, which I'll call the screened plasma frequency. The fields inside are damped, and does not that damping mean attenuation and energy loss? The answer is no. The refractive index below the screened plasma frequency is indeed imaginary

$$N = \sqrt{\epsilon_c - \frac{\omega_p^2}{\omega^2}},$$

making $n = 0$ and $\kappa \approx \omega_p/\omega$ (where I have neglected ϵ_c). But in this case the reflectance (Eq. 44) is

$$\mathscr{R} = \frac{1+\kappa^2}{1+\kappa^2} \equiv 1,$$

or 100%. All the incident energy is reflected. Electromagnetic energy is conserved! Moreover, if I remember that the reflected electric field direction is 180° from the incident field direction, then $\mathbf{E}_{\text{inc}} + \mathbf{E}_{\text{refl}} = 0 = \mathbf{E}_{\text{transm}}$, making the field in the perfect conductor zero, as it should be.

7.8.4 Intraband transitions

Now, let me look qualitatively at the absorption process (σ_1 finite and frequency low) If I think of photons, the process is the absorption of a photon of energy $\hbar\omega$ and the promotion of the electron to a higher state. Because $\mathcal{E}{\sim}k^2$ and $p = \hbar k$, this process involves a change of momentum of the electron. The initial and final electron states are both on the same $\mathcal{E}$ *vs.* k curve, so this is an intraband—within the same band—transition.

---

* Except at $\omega = 0$ where the imaginary part, or conductivity, is infinite. I'll calculate the dc conductivity on page 101.

I look at the energy and momentum conservation in the process here. The Pauli principle requires that initial state be full (below the Fermi level at $T = 0$) and the final state be empty (above the Fermi level at $T = 0$).[*] I write the initial energy as[†]

$$\mathcal{E}_i = \mathcal{E}_F = \frac{\hbar^2 \mathbf{k}_F^2}{2m}$$

and the final state energy as

$$\mathcal{E}_f = \mathcal{E}_F + \hbar\omega = \frac{\hbar^2(\mathbf{k}_F^2 + 2\mathbf{k}_F \Delta k)}{2m}$$

where Δk is the change of wave vector to get to the final state energy and I have neglected $(\Delta k)^2$. Taking the difference leads to

$$\hbar\omega = \frac{\hbar^2 k_F \Delta k}{m}$$

or

$$\Delta k = \frac{m\omega}{\hbar k_F} = \frac{\omega}{v_F}.$$

The photon starts with momentum $p_\gamma = \hbar q$ where $q = \omega/c$ and ends with zero. (It is absorbed.) But

$$q \ll \Delta k \qquad \frac{\omega}{c} \ll \frac{\omega}{v_F}.$$

It is not possible to conserve momentum in this energy-conserving process. So how does the metal absorb energy? The collisions change the momentum of the electron and transfer momentum to the crystal as a whole. The process is efficient at low frequencies, when $\omega \ll 1/\tau$, and inefficient at high frequencies.

7.8.5 Uncertainty principle

Uncertainty principle arguments are always worth exploring. An electron with quantum numbers $\mathbf{k}$ travels freely between collisions for a time τ. Whether the collision is elastic or inelastic, the length of time the electron occupies state $\psi(\mathbf{k})$ is τ. The consequence is that the energy is not known better than $\Delta\mathcal{E} = \hbar/\tau$. I use Eq. 109, $\mathcal{E} = \hbar^2 k^2/2m$, to find (by differentiation) that the energy blurring of $\Delta\mathcal{E}$ means a momentum spread of Δk. $\Delta\mathcal{E} = \hbar^2 k \Delta k/m$ or (because the relevant energy is the Fermi energy)

$$\Delta k = \frac{\Delta\mathcal{E} k_F}{2\mathcal{E}_F} = \frac{\hbar k_F}{2\mathcal{E}_F \tau} = \frac{m}{\hbar k_F \tau} = \frac{1}{v_F \tau}$$

But! $v_F\tau$ is the mean free path ℓ! So the position uncertainty is the mean free path and the momentum uncertainty is $\hbar\Delta k = \hbar/\ell$. This calculation should encourage you to stop thinking about the electron as a point-like particle skating around like a billiard ball on table filled with obstacles and think of it as a quantum-mechanical wave packet that is delocalized over a range given by the mean free path.[‡]

---

[*] This statement is a zero temperature picture, but it is unchanged if T is finite. The electron still goes from occupied state to empty state.

[†] Instead of taking the initial state at $\mathcal{E}_F$, I could take at any energy between $\mathcal{E}_F - \hbar\omega$ and $\mathcal{E}_F$, but nevermind.

[‡] If the idea of the physical extent of the wave function being the mean free path bothers you, do not forget that the wave function of the plane-wave state $e^{i\mathbf{k}\cdot\mathbf{r}}$ is infinite.

7.8.6 Yet more numbers

On page 28 I have put the parameters for our standard metal: $\sigma \approx 6 \times 10^5\ \Omega^{-1}\text{cm}^{-1}$ or $\rho = 1/\sigma = 1.6\ \mu\Omega$-cm. The mean free time is $\tau = 2 \times 10^{-14}$ s. From this number and $v_F = 2 \times 10^8$ cm/sec I get

$$\ell = 4 \times 10^{-6}\ \text{cm} = 40\ \text{nm}.$$

A mean free path of 40 nm—400 Å—is a characteristic value for the noble metals at room temperature.

The resistance of pure metals is strongly temperature dependent, decreasing by large factors as the temperature is reduced. The physical reason for this decrease is that the scattering is due to interactions of the electrons with lattice vibrations (phonons) and, as T decreases both the number and the mean energy of the phonons decreases.[11,23] Figure 37 shows an example, the temperature dependence of the resistivity $\rho_{\text{dc}} = 1/\sigma_{\text{dc}}$ of silver films of various thicknesses and also that of a bulk (mm-sized) crystal.[24] The resistivity is linear at high temperatures, and then, below the Debye temperature of 220 K, falls rapidly as a power law, $\rho \sim T^n$, with n in the range 2–5, before settling down to the low temperature residual value. The bulk crystal has a residual resistivity ratio (RRR) of 250. The mean free path at low temperatures is very long, $\ell = 10\ \mu$m. This is a macroscopic length, about the diameter of a cotton fiber. It is by no means a record: crystals of Ga have been produced with mean free paths of millimeters,[25] so that electrons traveled the full thickness of the crystal before a collision; ℓ was in fact limited by the crystal size, and the collision was with its surface.

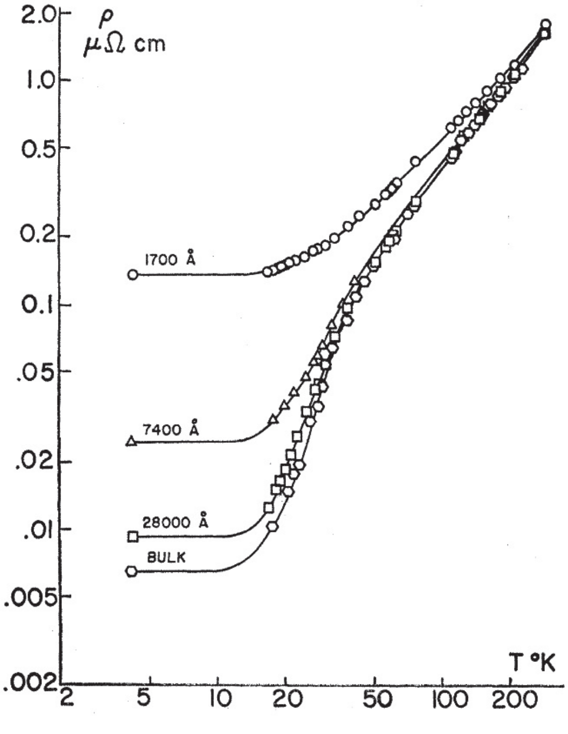

Fig. 37. Resistivity $\rho_{\text{dc}} = 1/\sigma_{\text{dc}}$ of silver at temperatures between 4 and 300 K.

8. OPTICAL EXCITATIONS: QUANTUM MECHANICS

Once I begin to consider the atomic and electronic structure of a solid, I have to move from the comfortable confines of the free-electron model. For many important solids (silicon!), the electronic structure and optical properties are given by band theory; for others (those called highly correlated or strongly interacting materials) one must go beyond band structure and consider the strong Coulomb forces among the electrons.

I'll discuss interband transitions and strongly interacting materials in subsequent chapters. Here, I want to look at the interaction of electromagnetic fields with a quantum-mechanical solid.

I will introduce the issue by considering that the solid has at least two "bands," each containing a fixed number of k states. Allowing for spin, I'll take this number as being equal to two states per atom (or per chemical formula—such as GaAs—in compound materials) in the solid. Moreover, I'll posit that there are enough electrons to fill completely the lower-energy band, called the valence band. The higher energy band—the conduction band—consists of empty states.* How is the optical response affected by this electronic structure? Well, the Pauli principle remains dominant, so there are no intraband transitions in a full band; all states are occupied.† Thus, the lowest optically-allowed excitations are from the highest (occupied) state in the valence band to the lowest (unoccupied) state in the conduction band. More important, *all* transitions are from valence to conduction band.

These optical transitions (lowest or not) must satisfy energy conservation,

$$\mathcal{E}_c = \mathcal{E}_v + \hbar\omega \tag{130}$$

where $\mathcal{E}_c$ is the unoccupied state in the conduction band, $\mathcal{E}_v$ is the occupied state in the valence band, and $\hbar\omega$ is the photon energy. The transition must also satisfy momentum conservation,

$$\mathbf{k}_c = \mathbf{k}_v + \mathbf{q}_\gamma \tag{131}$$

where $\mathbf{k}$ represents the crystal momentum of the electrons and $\mathbf{q}_\gamma$ that of the photon. The scales are quite different. The range of $\mathbf{k}$ in each band is $\pm\pi/a{\sim}\pm 5\times 10^8\ \mathrm{cm}^{-1}$. But, if $\mathcal{E}_c - \mathcal{E}_v$ is of order 1 eV, then the photon wavevector is $\mathbf{q} = 2\pi/\lambda{\sim}5\times 10^4\ \mathrm{cm}^{-1}$. The ratio of photon momentum to electron momentum is $\times 10^{-4}$. Hence, $\mathbf{k}_c = \mathbf{k}_v$; optical transitions are *vertical*, because on a plot of energies versus $\mathcal{E}_v$ and $\mathcal{E}_c$, the energy of the final state is directly above the energy of the initial state.

There is one loophole. If some other excitation is involved,

$$\mathcal{E}_c = \mathcal{E}_v + \hbar\omega \pm \hbar\Omega \tag{132}$$

where $\hbar\Omega$ is the energy if the excitation‡ (e. g., a phonon). Momentum conservation must also be satisfied:

$$\mathbf{k}_c = \mathbf{k}_v + \mathbf{q}_\gamma \pm \mathbf{k}_\Omega \tag{133}$$

---

[*] These names are historical, and a little misleading. In typical semiconductors, both valence and conduction bands are made up of linear combinations of valence orbitals. Moreover, conduction can be accomplished by either electrons in the conduction band or holes in the valence band.

[†] And none in an empty band either…

[‡] The ± is there because the excitation may either be created or (if excited by thermal processes) destroyed.

where $\mathbf{k}_\Omega$, the crystal momentum of the excitation, is typically in the range $\pm\pi/a$ (like the electrons) so the initial and final state can each be widely separated on the dispersion relation.

The selection rules given by Eqs. 130 and 131 are for direct optical transitions. Those in Eqs. 132 and 133 are for indirect optical transitions.

8.1 The solid with an electromagnetic field

I start with a system of the solid and the electromagnetic wave. The Hamiltonian is a sum of kinetic and potential energies:

$$\mathcal{H} = T + V.$$

Here, T is the kinetic energy,

$$T = \frac{1}{2m}(\mathbf{p} + \frac{e}{c}\mathbf{A}),$$

with $\mathbf{p} = -i\hbar\nabla$ the momentum and $\mathbf{A}$ the vector potential.* The potential energy is periodic in $\mathbf{T}$ the translation vector of the lattice;

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{T}).$$

The next step is to calculate the kinetic energy

$$T = \frac{1}{2m}(p^2 + \frac{e}{c}\mathbf{p}\cdot\mathbf{A} + \frac{e}{c}\mathbf{A}\cdot\mathbf{p} + \frac{e^2}{c^2}A^2)$$

---

* There is a certain mysterious character to the vector potential. It is a function from which the magnetic field and the electric field both can be calculated: In non-magnetic materials,

$$\mathbf{B} = \mathbf{H} = \nabla\times\mathbf{A} \qquad \mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}, \tag{134}$$

with the scalar potential equal to zero. (As it should be for transverse fields: if Φ is of plane-wave form then the gradient, $\mathbf{E}$ is parallel to $\mathbf{k}$.) Let me construct a vector potential consistent with our plane-wave fields:

$$\mathbf{E} = \hat{\boldsymbol{e}}E_0 e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}$$

from which Maxwell's equations lead to $q = N\omega/c$ and

$$\mathbf{H} = \frac{c}{\omega}\mathbf{q}\times\mathbf{E} = N(\hat{\boldsymbol{q}}\times\hat{\boldsymbol{e}})E_0\hat{\boldsymbol{e}}e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}$$

I write an $\mathbf{A}$ that gives $\mathbf{E}$ via Eq. 134

$$\mathbf{A} = -i\hat{\boldsymbol{e}}E_0\frac{c}{\omega}e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}.$$

(Check it. The time derivative brings down $-i\omega$) Then

$$\begin{aligned}\nabla\times\mathbf{A} &= i\mathbf{q}\times(-i\hat{\boldsymbol{e}}\frac{c}{\omega})E_0e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}\\ &= N(\hat{\boldsymbol{q}}\times\hat{\boldsymbol{e}})E_0e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}\\ &= \mathbf{H}\end{aligned}$$

with $H = NE$.

and look at the commutators*

$$[x, p_x] = xp_x - p_x x = i\hbar$$

and

$$[f(\mathbf{r}), p_x] = f(\mathbf{r})p_x - p_x f(\mathbf{r}) = i\hbar\frac{\partial}{\partial x}f(\mathbf{r})$$

so

$$[A_x, p_x] = A_x p_x - p_x A_x = i\hbar\frac{\partial}{\partial x}A_x$$

and the same for $[A_y, p_y]$ and $[A_y, p_y]$. I then can reform the vectors, with the derivative turning into a divergence

$$\mathbf{p}\cdot\mathbf{A} = \mathbf{A}\cdot\mathbf{p} - i\hbar\nabla\cdot\mathbf{A}.$$

Thus

$$T = \frac{1}{2m}(p^2 + \frac{2e}{c}\mathbf{A}\cdot\mathbf{p} - \frac{ie\hbar}{c}\nabla\cdot\mathbf{A}. + \frac{e^2}{c^2}A^2). \tag{135}$$

Now, the external field is weak, so I will keep only the first-order terms,† dropping the one in A^2.‡ I use the Coulomb gauge, $\nabla\cdot\mathbf{A} = 0$, eliminating a second term in Eq. 135. I also use $\mathbf{p} = -i\hbar\nabla$ to find

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{ie\hbar}{mc}\mathbf{A}\cdot\nabla + V(\mathbf{r}).$$

8.2 Perturbation expansion

The zero-order Hamiltonian is the Hamiltonian for the electrons in the solid

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m}\nabla^2 + V$$

and the perturbation Hamiltonian is the electromagnetic field

$$\mathcal{H}_1 = -i\frac{e\hbar}{mc}\mathbf{A}\cdot\nabla$$

I suppose that I know the solution to $\mathcal{H}_0$

$$\mathcal{H}_0 |n\rangle = E_n |n\rangle \Rightarrow \langle n|\mathcal{H}_0|n\rangle = E_n$$

where the E_n are a set of energies in the various bands and the vectors $|n\rangle$ are a complete orthonormal set, satisfying $\langle n|n'\rangle = \delta_{nn'}$. This zero-order Hamiltonian $\mathcal{H}_0$ is independent of the time. Hence I know that

$$\psi_0(t) = |n\rangle\, e^{-iEt/\hbar}$$

My perturbation, the vector potential $\mathbf{A}(t)$, is time dependent, so I must use time-dependent perturbation theory. I write the perturbed wave function as a linear combination of the $|n\rangle$s,

$$\psi = \sum_{n'} a_{n'}(t)\,|n'\rangle\, e^{-iE_{n'}t/\hbar}$$

(If the energy is in the continuum, the sum becomes an integral.)

---

[*] Obvious from $p_x = -i\hbar\frac{\partial}{\partial x}$ and $xp_x C = 0$ and $p_x x C = -i\hbar C$.

[†] In a minute I will start a perturbation theory calculation.

[‡] Were it to be kept, there would be nonlinear behavior, as $(e^{-i\omega t})^2$ becomes $e^{-2i\omega t}$, a doubled frequency.

Now, I calculate using the time-dependent Schrödinger equation

$$\mathcal{H}\psi = (\mathcal{H}_0 + \mathcal{H}_1)\psi = i\hbar\frac{\partial\psi}{\partial t}$$

Let me do this in pieces.

$$\begin{aligned}\mathcal{H}_0\psi &= \sum_{n'} a_{n'}(t)\mathcal{H}_0 \left|n'\right\rangle e^{-iE_{n'}t/\hbar} \\ &= \sum_{n'} a_{n'}(t)E_{n'} \left|n'\right\rangle e^{-iE_{n'}t/\hbar}\end{aligned}$$

and

$$\mathcal{H}_1\psi = \sum_{n'} a_{n'}(t)e^{-iE_{n'}t/\hbar}\mathcal{H}_1 \left|n'\right\rangle$$

$$i\hbar\frac{\partial\psi}{\partial t} = i\hbar\sum_{n'}\frac{\partial a_{n'}(t)}{\partial t}\left|n'\right\rangle e^{-iE_{n'}t/\hbar} + i\hbar\sum_{n'} a_{n'}(t)\left|n'\right\rangle\left(-i\frac{E_{n'}}{\hbar}\right)e^{-iE_{n'}t/\hbar}$$

I put this all together to get

$$\begin{aligned}\sum_{n'} a_{n'}(t)e^{-iE_{n'}t/\hbar}E_{n'}\left|n'\right\rangle + \sum_{n'} a_{n'}(t)e^{-iE_{n'}t/\hbar}\mathcal{H}_1\left|n'\right\rangle &= i\hbar\sum_{n'}\frac{\partial a_{n'}(t)}{\partial t}\left|n'\right\rangle e^{-iE_{n'}t/\hbar} \\ &\quad + \sum_{n'} a_{n'}(t)\left|n'\right\rangle E_{n'}e^{-iE_{n'}t/\hbar}\end{aligned}$$

The first and last terms are identical and cancel. I am left with

$$\sum_{n'} a_{n'}(t)e^{-iE_{n'}t/\hbar}\mathcal{H}_1\left|n'\right\rangle = i\hbar\sum_{n'}\dot{a}_{n'}(t)\left|n'\right\rangle e^{-iE_{n'}t/\hbar}$$

where the dot over $a_{n'}(t)$ on the right-hand side signifies a time derivative. Multiply from the left by $\langle n|$ to get

$$\sum_{n'} a_{n'}(t)e^{-iE_{n'}t/\hbar}\left\langle n\middle|\mathcal{H}_1\middle|n'\right\rangle = i\hbar\sum_{n'}\dot{a}_{n'}(t)\left\langle n\middle|n'\right\rangle e^{-iE_{n'}t/\hbar}$$

Because $\langle n|n'\rangle = \delta_{nn'}$, only one term on the right, $n' = n$, survives, so

$$i\hbar\dot{a}_n(t)e^{-iE_nt/\hbar} = \sum_{n'} a_{n'}(t)e^{-iE_{n'}t/\hbar}\left\langle n\middle|\mathcal{H}_1\middle|n'\right\rangle$$

Now I multiply by $e^{iE_nt/\hbar}$, define the matrix element as $\mathcal{H}_{1nn'} = \langle n|\mathcal{H}_1|n'\rangle$, and the energy difference (as a frequency) $\omega_{nn'} = (E_n - E_{n'})/\hbar$, and I get

$$i\hbar\dot{a}_n(t) = \sum_{n'} a_{n'}(t)\mathcal{H}_{1nn'}e^{-i\omega_{nn'}t} \tag{136}$$

Other than the omission of the A^2 term I have yet to make an approximation. Now is the time to do so. First-order perturbation theory consists of considering the field to have been turned on at time $t = 0$ and using $a_{n'}(0)$ in Eq. 136 rather than $a_{n'}(t)$. Moreover, I will assume that the system at $t = 0$ is in its ground state ($n = 0$) so that $a_0(0) = 1$ and all the others are zero. Only one term on the right of Eq. 136 remains:

$$i\hbar\dot{a}_n(t) = \mathcal{H}_{1n0}e^{-i\omega_{n0}t} \tag{137}$$

8.3 The matrix element of the perturbation

Let me take a look again at $\mathcal{H}_1$:

$$\mathcal{H}_1 = -i\frac{e\hbar}{mc}\mathbf{A}\cdot\nabla$$

with

$$\mathbf{A} = -i\frac{c}{\omega}\mathbf{E}.$$

Hence

$$\mathcal{H}_1 = -\frac{e\hbar}{m\omega}E_0 e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}\hat{\boldsymbol{e}}\cdot\nabla$$

and the matrix element I need is

$$\mathcal{H}_{1n0} = -\frac{e\hbar}{m\omega}E_0 e^{-i\omega t}\left\langle n\middle|e^{i\mathbf{q}\cdot\mathbf{r}}\hat{\boldsymbol{e}}\cdot\nabla\middle|0\right\rangle$$

making Eq. 137 become

$$i\hbar\frac{da_n}{dt} = -\frac{e\hbar}{m\omega}E_0 e^{i(\omega_{n0}-\omega)t}\left\langle n\middle|e^{i\mathbf{q}\cdot\mathbf{r}}\hat{\boldsymbol{e}}\cdot\nabla\middle|0\right\rangle. \tag{138}$$

This equation tells us how the amplitude of the state labeled by n grows with time.[*]

I can make a few simplifications here. First, the matrix element contains $e^{i\mathbf{q}\cdot\mathbf{r}}$, but I have argued that $q = 0$. Essentially, the light field varies slowly on the scale of the unit cell or the size of an electron wave packet so I may take it as constant in space (but time-varying) on that scale.[†] I can also pull $\hat{\boldsymbol{e}}$ out of the matrix element and combine with E_0 to make a vector. With a little rearrangement of order of terms, Eq. 138 becomes

$$i\hbar\dot{a}_n = -\frac{e\hbar}{m\omega}e^{i(\omega_{n0}-\omega)t}\left\langle n|\nabla|0\right\rangle\cdot\mathbf{E}_0.$$

Next, return to momentum by using $\hbar\nabla = i\mathbf{p}$ and

$$i\hbar\dot{a}_n = -i\frac{e}{m\omega}e^{i(\omega_{n0}-\omega)t}\left\langle n|\mathbf{p}|0\right\rangle\cdot\mathbf{E}_0.$$

Note that it is the component of the momentum (i. e., velocity) in the direction of the field that is important.

---

[*] The alert reader will have noticed that I started with a partial derivative in the time-dependent Schrödinger equation, went to an $\dot{a}$ notation and now have a total derivative. Because the amplitude is a function of time but not of space, this substitution is OK.

[†] The plane wave states $e^{i\mathbf{k}\cdot\mathbf{r}}$ of course have equal amplitude everywhere in the crystal. But the appropriate length scale is not the crystal size, but instead somewhere between the lattice constant and the mean free path.

Next, I manipulate the momentum matrix element

$$\langle n|\mathbf{p}|0\rangle = m\frac{d}{dt}\langle n|\mathbf{r}|0\rangle$$

To take the derivative, I need to insert the time dependence of $\langle n|$ and $|0\rangle$. These are $e^{i\omega_n t}$ and $e^{-i\omega_0 t}$ respectively.* The derivative with respect to time then works on

$$\langle n|\mathbf{p}|0\rangle = m\frac{d}{dt}e^{i\omega_n t}\langle n|\mathbf{r}|0\rangle e^{-i\omega_0 t} = im(\omega_n - \omega_0)e^{i\omega_n t}\langle n|\mathbf{r}|0\rangle e^{-i\omega_0 t}.$$

Now $\omega_n = \mathcal{E}_n/\hbar$ for each n, so $\hbar(\omega_n - \omega_0)$ is the energy difference between the initial and final states, which in turn equals (by energy conservation, Eq. 130) the photon energy $\hbar\omega$. So the momentum matrix element is equal to

$$\langle n|\mathbf{p}|0\rangle = im\omega\langle n|\mathbf{r}|0\rangle$$

and Eq. 138 then becomes

$$i\hbar\dot{a}_n = e^{i(\omega_{n0}-\omega)t}\langle n|e\mathbf{r}|0\rangle \cdot \mathbf{E}_0 \tag{139}$$

Classically the quantity $-e\mathbf{r}$ is the electric dipole moment associated with the optical transmission.

The next step is to integrate Eq. 139. But before I do this, I need to fix up a few things. I will write the electric field as a real quantity polarized along the $\hat{\boldsymbol{x}}$ direction.

$$\mathbf{E} = \hat{\boldsymbol{x}}E_0\cos(\omega t) = \frac{1}{2}\left(e^{i\omega t} + e^{-i\omega t}\right).$$

Then,

$$i\hbar\dot{a}_n = \frac{e}{2}\left(e^{i(\omega_{n0}-\omega)t} + e^{i(\omega_{n0}+\omega)t}\right)\langle n|x|0\rangle E_0.$$

I integrate both sides. The left side has

$$\int_0^t dt'\dot{a}_n = a_n(t)$$

because $a_n(0)$ is by hypothesis zero. The right side contains terms like

$$\int_0^t dt' e^{i(\Omega)t'} = \frac{e^{i(\Omega)t} - 1}{i\Omega}$$

Then

$$a_n = \frac{e}{2\hbar}\left(\frac{1 - e^{i(\omega_{n0}-\omega)t}}{\omega_{n0} - \omega} + \frac{1 - e^{i(\omega_{n0}+\omega)t}}{\omega_{n0} + \omega}\right)\langle n|x|0\rangle E_0. \tag{140}$$

The first term inside the parenthesis is resonant when $\omega = \omega_{n0}$.

---

[*] Remember that $\langle n|$ is the complex conjugate of $|n\rangle$.

8.4 Electric dipole transitions

The dipole moment* is

$$\mathfrak{p} = \langle \psi_f | -e\mathbf{r} | \psi_i \rangle + \langle \psi_i | -e\mathbf{r} | \psi_f \rangle$$

where $\psi_i = |0\rangle\, e^{-i\omega_0 t}$ is the initial (ground) state wave fuction and $\psi_f = \sum a_n \,|n\rangle\, e^{-i\omega_n t}$ with $\hbar\omega_n = \mathcal{E}_n$ is the excited state wave function. Note that I am letting the excited state be a linear combination of all the unoccupied states $|n\rangle$. Energy conservation will determine which value of n is important when the frequency is ω.

The induced dipole moment will be parallel to the electric field, the $\hat{\boldsymbol{x}}$ direction,. leading to

$$\mathfrak{p}_x = -e \sum \left[a_n^* e^{i\omega_{n0} t} \langle n|x|0\rangle + a_n e^{-i\omega_{n0} t} \langle 0|x|n\rangle \right]$$

with the energy difference $\omega_{n0} = \omega_n - \omega_0$. For shorthand notation, let $x_{n0} = \langle n|x|0\rangle$ with a corresponding formula for x_{0n}.

Now substitute for $a_n(t)$ from Eq. 140 and calculate and substitute for $a_n^*(t)$ also. Note that $\mathfrak{p}_x$ has one factor of $\langle n|x|0\rangle$ as does a_n, so there will be terms like $|\langle n|x|0\rangle|^2$ in the dipole moment. After some algebra, I get

$$\mathfrak{p}_x = \frac{2e^2}{\hbar} \sum_n \frac{\omega_{n0}}{\omega_{n0}^2 - \omega^2} |\langle n|x|0\rangle|^2 E_0 \cos\omega t = \alpha_e E$$

where ga_e is the polarizability and $E = E_0 \cos\omega t$ is the electric field.

One can add a (simulated) damping by letting $\omega_{n0} \to \omega_{n0} + i\gamma_n/2$, in which case the polarizability is

$$\alpha_e = \frac{2e^2}{\hbar} \sum_n \frac{\tilde{\omega}_{n0}}{\tilde{\omega}_{n0}^2 - \omega^2 - i\omega\gamma_n} |\langle n|x|0\rangle|^2 \tag{141}$$

The tilde in $\tilde{\omega}_{n0}$ is meant to indicate that the damping shifts the frequency. ($\tilde{\omega}_{n0} = \omega_{n0} + \gamma_{n0}/4$.) But local field corrections, as discussed on p. 42, will lead to larger shifts.

Equation 141 gives the polarizability of the site where the excitation lives. All that I have used here is that there is a ground state $|0\rangle$ and a set of excited states $|n\rangle$ at some location in the crystal. The details of the location are not specified, nor is the path taken by the electron going from $|0\rangle$ to $|n\rangle$. I may reasonably assert that the transitions occur within an atom or pair of atoms, or perhaps a molecule in the solid. They involve one electron leaving the ground state (where its absence is called a hole) and entering one or another of the excited states. There is a total of $\mathfrak{n}$ such electrons in each unit volume of the solid.[†] Now, in the ground state of the solid there are many electrons, and these electrons occupy many energy levels. The ground state, $|0\rangle$, represents the initial energy of *any* of these electrons.

---

[*] I'll use, for the moment, $\mathfrak{p}$ here because $\mathbf{p}$ is momentum.

[†] I used n as the index of the states in the unperturbed Hamiltonian because I wanted to suggest these were the principal quantum numbers of the atomic energy levels. So I will use this Gothic $\mathfrak{n}$ for the number density.

So in one case it might be an electron near the top of the valence band and in another case a core electron in the band made from the 1s orbitals of the atoms in the crystal.

With these considerations, I'll just add up all possible states, make local-field corrections (by dropping the tilde!), and obtain the dielectric function:

$$\epsilon = 1 + \frac{8\pi \mathfrak{n} e^2}{\hbar} \sum_n \frac{\omega_{n0}}{\omega_{n0}^2 - \omega^2 - i\omega\gamma_n} |\langle n|x|0\rangle|^2. \tag{142}$$

8.5 The oscillator strength

Equation 142 is of Lorentzian form, just like Eq. 74 on page 44. I can make it look a bit more familiar by defining

$$f_{n0} \equiv \frac{2m\omega_{n0}}{\hbar} |\langle n|x|0\rangle|^2 = \frac{2}{\hbar\omega_{n0} m} |\langle n|p_x|0\rangle|^2 \tag{143}$$

where m is the mass of the electron.* The quantity f_{n0} is known as the "oscillator strength" of the transition. Using it, I can write ϵ as

$$\epsilon(\omega) = 1 + \frac{4\pi \mathfrak{n} e^2}{m} \sum_n \frac{f_{n0}}{\omega_{n0}^2 - \omega^2 - i\omega\gamma_n} \tag{144}$$

and it is clear that $\mathfrak{n}$ is the the total number density of electrons in the solid, $\mathfrak{n} = ZN_{\text{atoms}}/V$, with Z the atomic number (or the average atomic number in a compound). The imaginary part, ϵ_2, has a series of peaks at the resonant frequencies $\{\omega_{n0}\}$. The widths of these peaks are set by the values of $\{\gamma_n\}$; as $\gamma_n \to 0$ the width of that peak becomes narrower and narrower until, eventually, it becomes a delta function,† $\epsilon_2 \sim \delta(\omega - \omega_{n0})$.

8.5.1 Limiting values of ϵ

I learn something by thinking about Eq. 144 in both low- and high-frequency limits. First, when $\omega \to 0$,

$$\epsilon_1(0) = 1 + \frac{4\pi \mathfrak{n} e^2}{m} \sum_n \frac{f_{n0}}{\omega_{n0}^2}. \tag{145}$$

Note that $\epsilon_2(0) = 0$; this approach does not consider free carriers. Note further that if f_{n0} is large, ω_{n0} is small, or both, that value of n contributes a lot to the static dielectric constant.

Next, I evaluate ϵ at high frequencies.

$$\epsilon_1(\infty) = 1 - \frac{4\pi \mathfrak{n} e^2}{m} \sum_n \frac{f_{n0}}{\omega^2}. \tag{146}$$

and it looks like the response of perfectly free carriers.‡

---

[*] The actual mass, 9.10938291 $\times 10^{-28}$ grams, and not some effective mass.

[†] I can prove this from Kramers-Kronig; see section 9.

[‡] As it should if the frequency is high enough.

8.5.2 Another definition of oscillator strength

Note that in many papers, especially papers about the lattice dynamics (phonons), you see the dielectric function written as I did for Eq. 83, with a factor $S_n\omega_n^2$ in the numerator:

$$\epsilon(\omega) = 1 + \sum_n \frac{S_n\omega_n^2}{\omega_n^2 - \omega^2 - i\omega\gamma_n}$$

Clearly

$$S_n\omega_n^2 = \frac{4\pi \mathfrak{n} e^2}{m} f_{n0}.$$

8.6 Oscillator strength sum rule

I turn to the sum that appears in Eq. 146, a sum over all the states n of f_{n0}.

$$\sum_n f_{n0} = \frac{2m}{\hbar}\sum_n \omega_{n0} |\langle n|x|0\rangle|^2$$

Now $|\langle n|x|0\rangle|^2 = (\langle n|x|0\rangle)^*(\langle n|x|0\rangle) = \langle 0|x|n\rangle\langle n|x|0\rangle$ and $\omega_{n0} = (\mathcal{E}_n - \mathcal{E}_0)/\hbar$. Finally, $2 = 1 + 1$ so*

$$\sum_n f_{n0} = \frac{m}{\hbar^2}\sum_n [(\langle 0|x|n\rangle (\mathcal{E}_n - \mathcal{E}_0)\langle n|x|0\rangle + \langle 0|x|n\rangle (\mathcal{E}_n - \mathcal{E}_0)\langle n|x|0\rangle]. \tag{147}$$

Of course I did not know that I wanted to write Eq. 147 as I did until I had worked the following little problems: First I look at the matrix element of the commutator of $\mathcal{H}_0$, the unperturbed Hamiltonian with x:

$$\begin{aligned}\langle n|[\mathcal{H}_0, x]|0\rangle &= \langle n|\mathcal{H}_0, x - x\mathcal{H}_0|0\rangle \\ &= \langle n|\mathcal{E}_n x - x\mathcal{E}_0|0\rangle \\ &= (\mathcal{E}_n - \mathcal{E}_0)\langle n|x|0\rangle\end{aligned}$$

where I used the fact that $\mathcal{H}_0|n\rangle = \mathcal{E}_n|n\rangle$ and went leftward in the first term of the commutator an rightward in the second. An identical process yields

$$\langle 0|[\mathcal{H}_0, x]|n\rangle = -(\mathcal{E}_n - \mathcal{E}_0)\langle 0|x|n\rangle$$

I inspect Eq. 147 and see that I can include both commutators neatly into it:

$$\sum_n f_{n0} = \frac{m}{\hbar^2}\sum_n (\langle 0|x|n\rangle\langle n|[\mathcal{H}_0, x]|0\rangle - \langle 0|[\mathcal{H}_0, x]|n\rangle\langle n|x|0\rangle). \tag{148}$$

Next, I recall the trick from quantum mechanics where the teacher "inserts a complete set of states"†

$$\sum_n \langle 0|A|n\rangle\langle n|B|0\rangle = \langle 0|A\left(\textstyle\sum_n |n\rangle\langle n|\right)B|0\rangle = \langle 0|AB|0\rangle$$

where A and B arbitrary operators and $\sum_n |n\rangle\langle n| = 1$.

---

* ☺

† Or, rather, the way I have derived it, de-inserts them.

This trick simplifies Eq. 148 a lot,

$$\begin{aligned}\sum_n f_{n0} &= \frac{m}{\hbar^2}(\langle 0|x\,[\mathcal{H}_0,x]|0\rangle - \langle 0|[\mathcal{H}_0,x]\,x|0\rangle)\\ &= \frac{m}{\hbar^2}\,\langle 0|(x\,[\mathcal{H}_0,x] - [\mathcal{H}_0,x]\,x)|0\rangle\\ &= -\frac{m}{\hbar^2}\,\langle 0|[\,[\mathcal{H}_0,x]\,,x]|0\rangle\end{aligned} \tag{149}$$

where I chose to write a minus sign in front with the double commutator in the last line because I liked its looks more than the other possibility $([\,x,[\mathcal{H}_0,x]],])$.

Equation 149 is a remarkable result: The sum over all the f_{n0} would seem to require knowledge of all the excited states $\{|n\rangle\}$ but instead it requires only the *ground state* wave function $|0\rangle$. However, the analysis can continue a bit further. I recall that

$$\mathcal{H}_0 = T + V$$

with T the kinetic energy and V the potential energy. The latter is a complicated0 function of location in the crystal, $V = V(\mathbf{r})$. But! Any function of $\mathbf{r}$ commutes with x. Thus $[V,x] = 0$ and

$$\sum_n f_{n0} = -\frac{m}{\hbar^2}\,\langle 0|[\,[T,x]\,,x]|0\rangle\,.$$

So, in addition to the ground-state wave function being the only one that matters, only the kinetic energy in the ground state, along with the dipole moment (x), appears in the sum. Of course the ground state wave function depends on the potential; the calculation must at some time find the eigenfunctions and eigenstates of $(T+V)\,|n\rangle = \mathcal{E}_n\,|n\rangle$.

8.6.1 Kinetic energy

In the Schrödinger picture, the kinetic energy os just $p^2/2m$ or

$$T - -\frac{\hbar^2}{2m}\nabla^2$$

but this is not the only possibility. In tight binding, one oftern writes a second-quantized Hamiltonian as

$$\mathcal{H} = -t\sum_{\langle i,j\rangle,\sigma}(c^\dagger_{i,\sigma}c_{j,\sigma} + h.c.),$$

where t is a hopping integral, $c^\dagger_{i\sigma}, c_{j\sigma}$ are creation and annihilation operators, $\langle i,j\rangle$ are nearest neighbor index, and σ is a spin index.

The Schrödinger picture is particularly useful in the evaluation of $\sum_n f_{n0}$. I write out the commutator:

$$\begin{aligned}[\,[T,x]\,,x] &= [(Tx-xT),x] \\ &= Tx^2-2xTx+x^2T \\ &= -\frac{\hbar^2}{2m}(\nabla^2x^2-2x\nabla^2x+x^2\nabla^2)\end{aligned}$$

Now, as I did earlier,* I write

$$[\,[T,x]\,,x]\,C = -\frac{\hbar^2}{2m}(\nabla^2x^2-2x\nabla^2x+x^2\nabla^2)C = -\frac{\hbar^2}{m}C$$

Thus,

$$\begin{aligned}\sum_n f_{n0} &= -\frac{m}{\hbar^2}\,\langle 0|-\frac{\hbar^2}{m}|0\rangle\,, \\ &= \langle 0|0\rangle \\ &= 1.\end{aligned}$$

This extremely simple result is called the Thomas-Reiche-Kuhn sum rule. It sets a limit on the values of all the oscillator strengths in the dielectric function. Moreover, the sum is independent of lots of details about what is happening in the solid.

8.6.2 Sum rule for the conductivity

I apply this sum rule to the expression I wrote for the conductivity.

$$\sigma_1(\omega) = \frac{\mathfrak{n}e^2}{m}\sum_n \frac{f_{n0}\omega^2\gamma_n}{(\omega_{n0}^2-\omega^2)^2+\omega^2\gamma_n^2}$$

I integrate both sides of this

$$\int_0^\infty d\omega\,\sigma_1(\omega) = \frac{\mathfrak{n}e^2}{m}\sum_n f_{n0}\int_0^\infty d\omega\,\frac{\omega^2\gamma_n}{(\omega_{n0}^2-\omega^2)^2+\omega^2\gamma_n^2}$$

---

[*] If you do not like this, then I will try it on an arbitrary function $f(x)$. ($f(x)=C$ is one such function.) I do it in one dimension, so $\nabla^2=\frac{\partial^2}{\partial x^2}$ and f' means derivative with respect to x.

$$\begin{aligned}[\,[T,x]\,,x]\,f(x) &= -\frac{\hbar^2}{2m}(\nabla^2x^2-2x\nabla^2x+x^2\nabla^2)f(x) \\ &= -\frac{\hbar^2}{2m}\{\frac{\partial}{\partial x}[2xf(x)+x^2f'(x)]-2x\frac{\partial}{\partial x}[f(x)+xf'(x)]+x^2f''(x)\} \\ &= -\frac{\hbar^2}{2m}\{2f(x)+2xf'(x)+x^2f''(x)+2xf'(x)-2xf'(x)-2xf'(x)-2x^2f''(x)+x^2f''(x)\} \\ &= -\frac{\hbar^2}{m}f(x)\end{aligned}$$

Most of the terms cancel, leaving only the result of taking two derivatives of x^2 multiplied by $f(x)$. Just as in the case where $f(x)=C$.

I can do some changes of variables to simplify the integral on the right, and then look it up. Or I can just look it up. This definite integral equals $\pi/2$. Therefore,

$$\int_0^\infty d\omega\,\sigma_1(\omega) = \frac{\pi}{2}\frac{\mathfrak{n}e^2}{m}\sum_n f_{n0} = \frac{\pi}{2}\frac{\mathfrak{n}e^2}{m} \tag{150}$$

because the sum on f is unity. The integral is independent of the value of the resonant frequency or of the damping.

I will check this by considering the Drude conductivity. The only charge carriers in this model are those in the Fermi sphere, and

$$\sigma_1(\omega) = \frac{\omega_p^2\tau/4\pi}{1+\omega^2\tau^2}$$

with $\omega_p^2 = 4\pi ne^2/m$.

Do the integral:

$$\begin{aligned}\int_0^\infty d\omega\,\sigma_1(\omega) &= \int_0^\infty d\omega\,\frac{\omega_p^2\tau/4\pi}{1+\omega^2\tau^2}\\ &= \frac{\omega_p^2}{4\pi}\int_0^\infty dz\,\frac{1}{1+z^2}\end{aligned}$$

where $z = \omega\tau$. The integral is $\pi/2$ so that

$$\int_0^\infty d\omega\,\sigma_1(\omega) = \frac{1}{8}\omega_p^2 = \frac{\pi}{2}\frac{ne^2}{m}. \tag{151}$$

The integral is independent of the mean free time τ. It says that if, for example, the metal is cooled to low temperatures, and the conductivity increases by the residual resistivity value, then the increase in the area under $\sigma_1(\omega)$ at very low frequencies is offset by a decease at higher frequencies. The Drude function becomes taller and narrower and the sum rule allows you to see how this happens.

9. KRAMERS-KRONIG RELATIONS AND SUM RULES

The Kramers-Kronig relation[26,27] is a consequence of our experience that observable effects are *causal*, i. e., that the cause precedes the effect.* This notion seems sensible and it is a component of most parts of physics.

This chapter will start with a couple of examples where I show that there must be some relationship between the real and imaginary parts of the complex dielectric function or the complex conductivity. Then, I'll derive the relation for a number of our favorite response functions and follow with the use of Kramers-Kronig analysis for the reflectance, a common method for estimating response functions from experiment. The chapter ends with the derivation of some sum rules from the Kramers-Kronig relations.

9.1 Necessity for a relation between absorption and dispersion

9.1.1 A notch filter

To illustrate what causality implies for the optical constants that I have defined and used, let me do a *gedankenexperiment.* I imagine the following situation: I need an optical filter that is opaque for one wavelength and transparent for a wide band of wavelengths around the stop-wavelength. The transmission of such a filter should be 100% except at the specific wavelength of opacity, where it is zero. This filter, and others like it, differencing only in the specific wavelength of operation, could be part of some wavelength multiplexing scheme where the data streams of many receivers are put onto an optical transmission line; the filter is used to select out the wavelength of a single receiver and to pass the others onward.† The design transmission of the filter is shown in Fig. 38, plotted versus frequency rather than versus wavelength.

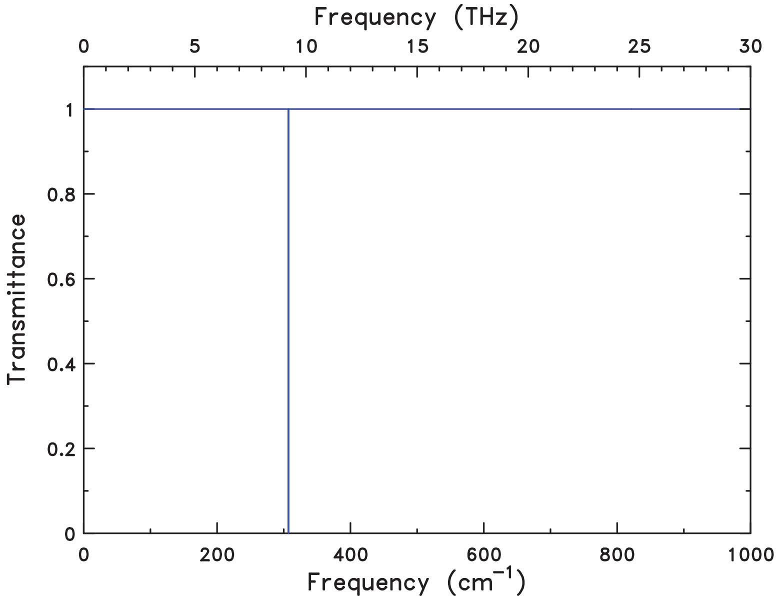

Fig. 38. A hypothetical filter that transmits 0–1000 cm^{-1} (0–30 THz) except for a band about 0.001 cm^{-1} (30 MHz) wide at $\omega_0 = 307\,\text{cm}^{-1}$ (9.2 THz).

---

* Non-causal effects are part of what is called *magic.*

† I know; there are many holes in this concept. It is, after all, only a gedankenexperiment.

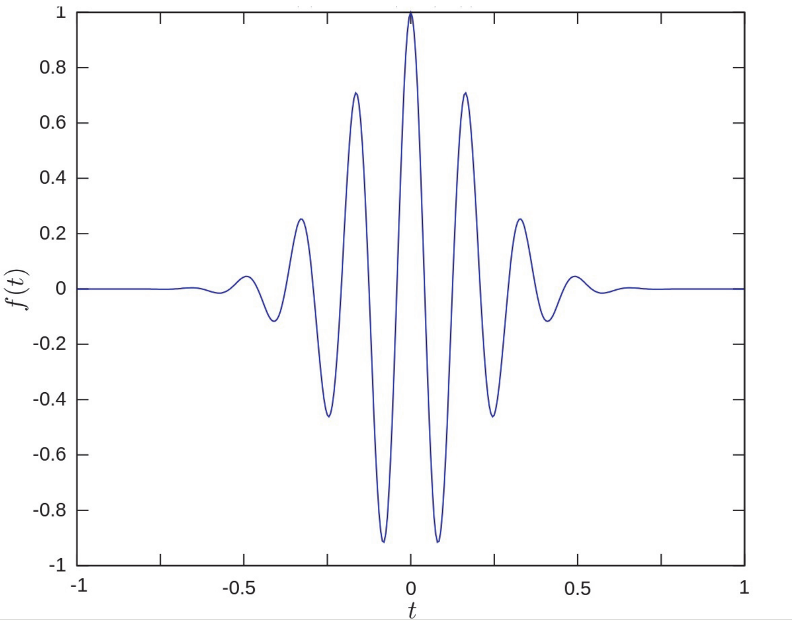

Fig. 39. A pulse with components up to above 10 THz. The zero if time is chosen to be at the peak of the pulse. The actual time when the signal arrives at the detector is determined by when it was transmitted and how long it took to get there.

Now let me imagine that my signal is a pulse such as the one in Fig. 39. If the time is in picoseconds, the pulse has oscillation frequencies (inverse of the period) of 6 to 15 THz. But it is not a sine or cosine with a single frequency. Indeed, Fourier teaches that I can write the function $f(t)$ as an integral:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{\boldsymbol{F}}(\omega)\; e^{i\omega t},$$

where $\hat{\boldsymbol{F}}(\omega)$ is the spectrum needed[*] to give $f(t)$. To give a good representation of the pulse, many frequencies will enter $\hat{\boldsymbol{F}}(\omega)$. Now, I will let the pulse be incident on my filter. The filter removes one frequency, ω_0, from the spectrum. Suppose the amplitude of the this component is A; the frequency spectrum behind the filter will be something like[†]

$$\hat{\boldsymbol{G}}(\omega) = \hat{\boldsymbol{F}}(\omega) - Ae^{i\phi}\delta(\omega - \omega_0)$$

where ϕ is the phase of the ω_0 component. I have subtracted this contribution from the spectrum to represent the effect of the filter. The pulse will be changed of course, now its

---

[*] I calculate $\hat{\boldsymbol{F}}(\omega)$ by taking the inverse transform of $f(t)$.

[†] I think of the pulse as being measured as a voltage or field; I will have to worry about specifying the real and imaginary parts correctly so that the measured signal will be real. But this is *my* gedankenexperiment and I won't think too much about this detail.

spectrum is

$$\begin{aligned} g(t) =& \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \hat{\boldsymbol{G}}(\omega)\; e^{i\omega t}, \\ =& \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, [\hat{\boldsymbol{F}}(\omega) - Ae^{i\phi}\delta(\omega-\omega_0)]\; e^{i\omega t}, \\ =& \, f(t) - A\cos(\omega_0 t + \phi) \end{aligned}$$

This result looks quite reasonable at first. My filter has done what I asked it to do: remove the ω_0 part of the pulse and transmit the rest. But then I realize that $\cos(\omega_0 t + \phi)$ is finite at all times,* including large negative times! The effect of the filter is to produce a signal at the detector long before the pulse arrived at the filter. This early appearance is a violation of causality and of course cannot happen.

The fix is exactly the mechanism used to generate the pulse in the first place: the phases of nearby Fourier components are adjusted so as to interfere destructively—indeed, perfectly so—and to cancel everything outside the times where the signal is non-zero. The filter can (and will) produce amplitudes at later times than the end of the original pulse; such ringing is an unavoidable consequence of having a narrow-band filter response. (Think of a struck bell.) But no signal will appear before the leading edge of the pulse arrives.

The filter must have a dispersive behavior, with a refractive index different from unity, especially near the absorption frequency. The Lorentzian model shown in Fig. 19 displays this behavior. The frequency-dependent refractive index will give frequency dependence to the velocity of the waves in the medium, accomplishing the required phase shifting. The amount of phase shift depends on the absorption properties of the filter, meaning that there must be a relation between the refractive index and extinction coefficient or between the real and imaginary parts of the conductivity

9.1.2 The perfect conductor

The notch filter example is striking, but gives no route to finding the relation between real and imaginary parts of the response function. Here is a second example, said to be due to Tinkham.[28] If there are no collisions, the conductivity of the Drude metal, Eqs. 54 or 125, is

$$\sigma = i\frac{ne^2}{m\omega} \tag{152}$$

i. e., zero real part and an imaginary part that diverges as $1/\omega$.

This form for σ is easy to derive. The equation of motion, Eq. 50, with no damping ($\Gamma = 0$) is

$$m\ddot{\mathbf{r}} = -e\mathbf{E}_{\text{ext}}$$

I take $e^{-i\omega t}$ time dependences for $\mathbf{E}_{\text{ext}}$ and $\mathbf{v} = \dot{\mathbf{r}}$ to find that

$$\mathbf{v} = -i\frac{e}{\omega m}\mathbf{E}$$

and then write $\mathbf{j} = -ne\mathbf{v} = \sigma\mathbf{E}$ to obtain Eq. 152.

---

[*] Except for its zeros of course

This derivation is fine for monochromatic waves, and there can be no causality issues for them because the field amplitude is constant* for $-\infty < t < \infty$. But here, I will do the opposite. Let the applied field be an *impulse,* a delta function in time† and of unit amplitude

$$\mathbf{E}(t) = \hat{\boldsymbol{e}}\delta(t)$$

This impulsive field has a frequency spectrum that is infinitely wide. I know this because I know that a representation of the delta function is

$$\delta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega t},$$

with

$$D(\omega) \equiv \int_{-\infty}^{\infty} dt\, \delta(t) e^{-i\omega t} = 1,$$

making the spectrum $\mathbf{E}(\omega)$ have unit amplitude at all frequencies.

The time dependence of the current is also the Fourier transform of its frequency spectrum. Using $\mathbf{j}(\omega) = \sigma(\omega)\mathbf{E}(\omega)$, I get

$$\mathbf{j}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \sigma(\omega)\mathbf{E}(\omega) e^{i\omega t}.$$

Now, I substitute $\mathbf{E}(\omega) = \hat{\boldsymbol{e}}$ and Eq. 152 for the conductivity to find that

$$\mathbf{j}(t) = i\frac{ne^2}{2\pi m}\hat{\boldsymbol{e}}\int_{-\infty}^{\infty} d\omega\, \frac{e^{-i\omega t}}{\omega}.$$

I look up the integral and find

$$\mathbf{j}(t) = \begin{cases} -\dfrac{ne^2}{2m}\hat{\boldsymbol{e}} & \text{if } t < 0 \\ \\ +\dfrac{ne^2}{2m}\hat{\boldsymbol{e}} & \text{if } t > 0 \end{cases}. \tag{153}$$

Here is another non-causal result. The current flows in the direction of the electric field polarization after the pulse arrives but flows in the *opposite* direction beforehand.

What is missing? Well, I can get Eq. 152 by taking the Drude conductivity, Eq. 54, and letting $\tau \to \infty$: $\sigma = (ne^2\tau/m)/(1 - i\omega\tau) \to (ne^2\tau/m)/(-i\omega\tau) = ine^2/m\omega$. This is reasonable as long as I do not take the $\omega = 0$ limit; if I do, I would start with $\sigma =$

---

[*] When I took classical mechanics and the instructor covered the forced harmonic oscillator, there was discussion of the *turn-on transient.* The oscillator responded at its resonant frequency when the force was initially applied, but then the internal damping reduced the amplitude of oscillations at this frequency to a negligible value after a few relaxation times had passed. If I turn on the field at $t = -\infty$, there is not a problem with the transient.

[†] Which might as well be centered at $t = 0$.

$(ne^2\tau/m) \to \infty$. The Drude dc conductivity is real; $\sigma_1(\omega)$ has a maximum at zero frequency and a width of $1/\tau$. (See Fig. 8.) As τ increases, the peak becomes higher and higher and the width narrower and narrower. In the limit, the real part becomes a delta function in frequency, located at the origin. Let me add this delta function to σ:

$$\sigma = A\delta(\omega) + i\frac{ne^2}{m\omega}. \tag{154}$$

where A is the weight (unknown at the moment) of the delta function.* The delta function term in σ is responsible for a current response to the impulsive electric field

$$\mathbf{j}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, A\delta(\omega)\hat{\boldsymbol{e}}e^{i\omega t} = \hat{\boldsymbol{e}}\frac{A}{2\pi}$$

If I take

$$A = \frac{ne^2\pi}{m}.$$

I can make this contribution to $\mathbf{j} = \hat{\boldsymbol{e}}ne^2/2m$, exactly canceling the current from the $ne^2/m\omega$ term for negative times and doubling it for positive times. The current is

$$\mathbf{j}(t) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{ne^2}{m}\hat{\boldsymbol{e}} & \text{if } t > 0 \end{cases}, \tag{155}$$

a step function. The impulse starts a current which, in the absence of damping, continues indefinitely.† Equation 155 makes me much happier than Eq. 153.

9.2 Kramers-Kronig integrals in linear, isotropic, local media

The Kramers-Kronig integrals are derived in a number of textbooks,[2,29–31] and have been discussed by many authors.[32–41] The subject is mostly approached by considering integrals on the complex frequency plane, although the original derivations by Kramers and Kronig relied on model dielectric functions.[41]

---

[*] Although I might guess that it is related to the area under $\sigma_1(\omega)$ in Eq. 56a.

[†] The result is easy to obtain from the equation of motion: I write $\mathbf{j} = -ne\mathbf{v}$ and $m\mathbf{a} = -e\mathbf{E}(t)$ to find

$$\frac{d\mathbf{j}}{dt} = -ne\frac{d\mathbf{v}}{dt} = \frac{ne^2}{m}\mathbf{E}(t) = \frac{ne^2}{m}\delta(t)\hat{\boldsymbol{e}}.$$

where $\hat{\boldsymbol{e}}$ is a unit vector in the $\mathbf{E}$ direction. I can integrate this equation (from $-\infty$ to t) to get

$$\mathbf{j}(t) = \begin{cases} 0 & t < 0 \\ \frac{ne^2}{m}\hat{\boldsymbol{e}} & t > 0 \end{cases}.$$

The current is zero for $t < 0$ and then jumps to a final value after the impulse has arrived.

9.2.1 Time domain response and causality

I consider the way in which the dipole moment per unit volume $\mathbf{P}$ depends on the electric field $\mathbf{E}$. These vectors are related by χ, a susceptibility,*

$$\mathbf{P} = \chi\mathbf{E}$$

where χ is the electric susceptibility and $\chi = (\epsilon - 1)/4\pi$. More generally, I can write the dipole moment per unit volume at location $\mathbf{r}$ and time t as an integral of the electric field over nearby locations and recent times,

$$\mathbf{P}(\mathbf{r}, t) = \int d^3r' \int dt'\, \chi(\mathbf{r}, \mathbf{r}', t, t')\mathbf{E}(\mathbf{r}', t').$$

Now, I want local response, so that I will be able to use local relations like $\mathbf{j} = \sigma\mathbf{E}$ and $\mathbf{D} = \epsilon\mathbf{E}$; to obtain this, I write the space part of χ as a delta function. Moreover, the time part should depend only on the difference between t and t':

$$\chi(\mathbf{r}, \mathbf{r}', t, t') = \delta(\mathbf{r} - \mathbf{r}')\chi(t - t')$$

Then

$$\mathbf{P}(\mathbf{r}, t) = \int_{-\infty}^{\infty} dt'\, \chi(t - t')\mathbf{E}(\mathbf{r}, t'). \tag{156}$$

where the polarization at $\mathbf{r}$ depends only on the field at $\mathbf{r}$. Of course, I do not make the response local in time. I know that a response can be retarded in time, with a phase difference between the sinusoidal field and the sinusoidal polarization or a ringing after an impulse force.

Causality sets a requirement on the form of $\chi(t - t')$: there can be no response before the stimulus. Suppose $\mathbf{E} = \mathbf{A}\delta(t' - t_0)$, an impulse of amplitude A at time t_0. Then $\mathbf{P}(t) = \chi(t - t_0)\mathbf{A}$. Causality requires that $\chi(t - t_0) = 0$ for $t < t_0$ so that $\mathbf{P}(t)$ is likewise zero for $t < t_0$. Equation 156 becomes

$$\mathbf{P}(\mathbf{r}, t) = \int_{-\infty}^{t} dt'\, \chi(t - t')\mathbf{E}(\mathbf{r}, t'). \tag{157}$$

Eq. 157 is a statement that the dipole moment at time t can be afected by the electric field values at all times *before* t but none after.

Equations 156 and 157 may be rewritten if I define $\tau = t - t'$. Then $dt' = -d\tau$ and

$$\mathbf{P}(\mathbf{r}, t) = \int_{0}^{\infty} d\tau\, \chi(\tau)\mathbf{E}(\mathbf{r}, t - \tau), \tag{158}$$

where the lower limit is at zero to force the response to be causal.

---

* I am going to use χ to represent the relation between dipole moment per unit volume and electric field for linear (and isotropic and homogeneous) media. For the moment, I'll consider the fields to be functions of space and time, making $\chi = \chi(\tau)$, the time difference between the time the field had a specific value and the time when the dipole moment per unit volume response exists. But later, $\chi = \chi(\omega)$, the relation between field at ω and dipole moment per unit volume at ω. I will therefore state the argument explicly in this section. Elsewhere, it will almost always be $\chi(\omega)$.

9.2.2 Fourier transformation into the frequency domain

The Kramers-Kronig integral takes place in the frequency domain. I know that the time response and frequency response are Fourier transform pairs.

$$E(\omega) = \int_{-\infty}^{\infty} dt\, E(t) e^{-i\omega t}, \qquad\qquad E(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, E(t) e^{i\omega t}. \tag{159}$$

Similarly,

$$P(\omega) = \int_{-\infty}^{\infty} dt\, P(t) e^{i\omega t}, \tag{160}$$

and (using the shift theorem)

$$\begin{aligned} \chi(\omega) &= \int_{-\infty}^{\infty} dt\, \chi(t-t') e^{i\omega(t-t')}, \\ &= \int_{0}^{\infty} d\tau\, \chi(\tau) e^{i\omega\tau}, \end{aligned} \tag{161}$$

where the second line uses the same substitution I used in going from Eqs. 156 and 157 to Eq. 158. This version is important because it shows two things.

First, I can take **P** and **E** to be real in Eq. 158. Hence $\chi(\tau)$ is also real. Of course $\chi(\omega)$ is complex on account of the complex exponential in Eq. 161. But the reality of $\chi(\tau)$ along with Eq. 161 means that $\chi(-\omega) = \chi^*(\omega)$ (ω real) or that $\chi_1(-\omega) = \chi_1(\omega)$ and $\chi_2(-\omega) = -\chi_2(\omega)$.

Second, in the next section I am going to allow ω to be a complex quantity, $\omega_1 + i\omega_2$ so that I can consider the properties of $\chi(\omega)$ as a complex function of a complex variable. Equation 161 then shows that, as long as $\chi(\tau)$ is finite for all τ, as it should be because I am considering linear materials, $\chi(\omega)$ has no poles in the upper half plane,[*] even as $\mathrm{Im}(\omega) \to +\infty$.

Now, I substitute Eq. 156 into Eq. 160 to get

$$P(\omega) = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt'\, \chi(t-t') \mathbf{E}(t') e^{i\omega t},$$

I now use the usual trick, exchanging the order of integration. I will also insert $e^{-i\omega t'} e^{i\omega t'} = 1$

$$P(\omega) = \int_{-\infty}^{\infty} dt' \mathbf{E}(t') \left[\int_{-\infty}^{\infty} dt\, \chi(t-t') e^{i\omega t} e^{-i\omega t'} \right] e^{i\omega t'}.$$

The term in square braces is $\chi(\omega)$. After moving it out of the integral, what remains is $E(\omega)$. Thus, I get a linear relaton between polarization and electric field:

$$P(\omega) = \chi(\omega) E(\omega) \tag{162}$$

Note that the convolution theorem of Fourier analysis would allow me to go from Eq. 156 to Eq. 162 directly; the former is a convolution in the time domain whereas the latter is a product in the frequency domain.

---

[*] There are of course divergences in the *lower* half plane, where $e^{-\omega_2 t}$ rapidly diverges as $-\omega_2$ or t become large.

9.2.3 The complex ω plane.

I now consider $\chi(\omega)$ to be a complex function of a complex frequency, $\omega = \omega_1 + i\omega_2$. The range of both real and imaginary parts is $-\infty \leftrightarrow \infty$. The function has these properties:

1. $\chi(\omega)$ is analytic in the upper half plane. (This arises—as discussed above—because $\chi(\tau)$ is zero for $\tau < 0$.)
2. $\chi(\omega) \to 0$ as $|\omega| \to \infty$ at least as fast as $1/\omega$.
3. $\chi(\omega)$ is Hermitian: $\chi(-\omega) = \chi^*(\omega)$.

Are these reasonable. Well, the Lorentz oscillator satisfies these conditions. It is:

$$\chi(\omega) = \frac{ne^2/m}{\omega_0^2 - \omega^2 - i\omega\gamma}$$

This function has poles when the denominator is zero. The poles are at frequencies

$$\omega = \pm\sqrt{\omega_0^2 - \gamma^2/4} - i\gamma/2$$

and lie in the lower half plane. As the frequencies become very large $\chi(\omega) = -ne^2/m\omega^2$ and so is zero as $|\omega| \to \infty$. Finally, I write it as

$$\chi(\omega) = \frac{ne^2/m}{(\omega_0^2 - \omega^2)^2 + \omega^2\gamma^2}\left[(\omega_0^2 - \omega^2) + i\omega\gamma\right].$$

By inspection, I can see that $\chi(-\omega) = \chi^*(\omega)$; the only term that is affected either by complex conjugation or by negation—and in the same way—is $i\omega\gamma$.

Free carriers have the same high-frequency limiting behavior. Indeed, all particles are free at high enough frequencies, so the $\chi(\omega)\sim 1/\omega^2$ behavior is quite general. See also Eq. 146.

Electromagnetism is unchanged as $\omega \to -\omega$ or, equivalently, whether I take an $e^{-i\omega t}$ or an $e^{i\omega t}$ form for the electric field. Let me start with the former,

$$E = E_0 e^{-i\omega t}$$

Eq. 162 can be written

$$P(\omega) = [\chi_1(\omega) + i\chi_2(\omega)][E_0\cos(\omega t) - iE_0\sin(\omega t)]$$

but the observable is the real part of this equation,

$$\mathrm{Re}(P) = \chi_1(\omega)E_0\cos(\omega t) + \chi_2(\omega)E_0\sin(\omega t).$$

I want this equation to be unchanged when ω is replaced by $-\omega$. This condition does hold, because

$$\chi_1(\omega) = \chi_1(-\omega)$$

and

$$\chi_2(\omega) = -\chi_2(-\omega).$$

$\chi_1(\omega)$ is said to be an *even* function while $\chi_2(\omega)$ is an *odd* function of ω. This statement is equivalent to saying that χ us Hermitian, so that

$$\chi(-\omega) = \chi^*(\omega).$$

9.2.4 Use of Cauchy's theorem

I will use Cauchy's integral theorem[42] to derive the Kramers-Kronig relation.[26,27] This theorem states that the line integral along a closed path of an analytic function is zero. My susceptibility $\chi(\omega')$ is such a function if I restrict myself to the upper half plane in the complex ω' space.* In addition, the function $\chi(\omega')/(\omega'-\omega)$, with ω real, is also analytic in the upper half plane. Cauchy's theorem then states that

$$\oint d\omega' \, \frac{\chi(\omega')}{\omega'-\omega} = 0 \tag{163}$$

for any closed contour in the upper half plane. I will use the contour shown in Fig. 40. It starts (*1*) at $(-\infty, 0)$, runs along the real axis until it is close to the point $(\omega, 0)$. It then (*2*) makes a semicircle around $(\omega, 0)$ in the upper half plane, and (*3*) continues along the real axis to $(\infty, 0)$. The path is closed by making a large† semicircle (*4*) of radius infinity in the upper half plane, joining the start of segment (*1*).

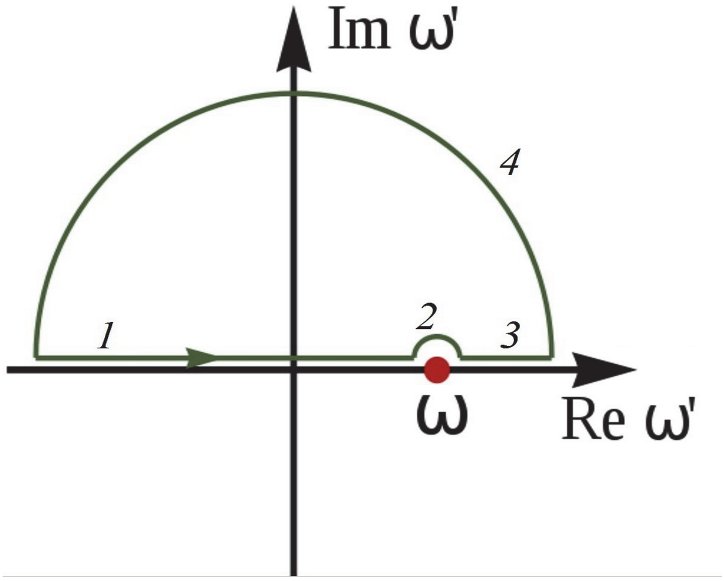

Fig. 40. Contour used for deriving the Kramers-Kronig relation. The four segments that make up the closed contour are shown.

Now, I break the integral up into its four segments

$$0 = \int_{(1)} d\omega' \, \frac{\chi(\omega')}{\omega'-\omega} + \int_{(2)} d\omega' \, \frac{\chi(\omega')}{\omega'-\omega} + \int_{(3)} d\omega' \, \frac{\chi(\omega')}{\omega'-\omega} + \int_{(4)} d\omega' \, \frac{\chi(\omega')}{\omega'-\omega},$$

and consider each of these. On segment (*4*), the length of the semicircle increases proportional to $|\omega|$, but the integrand vanishes, and the integral is zero so long as $\chi(\omega)$ falls at least as fast as $1/|\omega|$. And it does, according to our assumptions above. The line of segment

---

* I write the frequency with a prime because I want to use ω' as the dummy variable of integration.

† Very large!

(*2*) may be parameterized by writing $\omega' = \omega + ue^{i\phi}$ where ϕ is the angle the radius vector makes to the $\mathrm{Re}\,\omega'$ axis. As the path moves around the semicircle, the angle goes from π to 0. The quantity u is the (constant) radius of the semicircle. If u is small enough, the value of $\chi(\omega')$ is the same for all points on the little semicircle.* I can replace $\chi(\omega')$ with $\chi(\omega)$. Finally, $d\omega' = d\phi\, iue^{i\phi}$. Then

$$\int_{(2)} d\omega' \, \frac{\chi(\omega')}{\omega' - \omega} = iu\chi(\omega) \int_{\pi}^{0} d\phi \, e^{i\phi} \frac{1}{ue^{i\phi}} = i\chi(\omega)\phi\Big[_{\pi}^{0} = -i\pi\chi(\omega)$$

Segments (*1*) and (*3*) combine to make a principal-value integral,

$$\int_{(1)} d\omega' \, \frac{\chi(\omega')}{\omega' - \omega} + \int_{(3)} d\omega' \, \frac{\chi(\omega')}{\omega' - \omega} = \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega} \, d\omega'$$

I combine all four segments to obtain

$$\oint d\omega' \, \frac{\chi(\omega')}{\omega' - \omega} = \mathcal{P} \int_{-\infty}^{\infty} d\omega' \, \frac{\chi(\omega')}{\omega' - \omega} - i\pi\chi(\omega) = 0.$$

This equation can be rearranged to give the Kramers-Kronig relation,

$$\chi(\omega) = \frac{1}{i\pi} \mathcal{P} \int_{-\infty}^{\infty} d\omega' \, \frac{\chi(\omega')}{\omega' - \omega}. \tag{164}$$

There is an i in the denominator, so that the real part of χ is given by an integral of the imaginary part of χ and vice versa. I substitute $\chi(\omega) = \chi_1(\omega) + i\chi_2(\omega)$ on both sides; Equation 164 must hold for both its real and its imaginary parts, giving

$$\chi_1(\omega) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} d\omega' \, \frac{\chi_2(\omega')}{\omega' - \omega}, \tag{165}$$

and

$$\chi_2(\omega) = -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} d\omega' \, \frac{\chi_1(\omega')}{\omega' - \omega}, \tag{166}$$

where $\mathcal{P}$ denotes the Cauchy principal value. The real and imaginary parts of χ are not independent: the complex function can be constructed given just one of its parts.

If I had chosen segment (*2*) to be a semi-circle below instead of above the point $(\omega, 0)$, I would have had† $\int_{(2)} = +i\pi\chi(\omega)$ but the pole at ω would be *inside* the closed contour, so the contour integral will pick up the residue of the pole, $2i\pi\chi(\omega)$, making the result (of course) be the same.

---

[*] How small is small enough? In the case of the Lorentz oscillator, I must have $u \ll \gamma$.

[†] Because then $\omega' = \omega - ue^{i\phi}$

9.2.5 Eliminating negative frequencies

Equations 165 and 166 are not as useful as they could be, because the integrals cover $-\infty \to \infty$, implying that I must know the susceptibility at negative frequencies.* However, as discussed on p. 106, $\chi_1(\omega)$ is an even function of frequency and $\chi_2(\omega)$ is an odd function of frequency. I can use these properties to reduce the integration range to $0 \to \infty$. I start with Eq. 165, which gives the real part, $\chi_1(\omega)$ and transform the integral into two of specific parities by multiplying both the numerator and the denominator of the integrand by $\omega' + \omega$. Separating into two integrals yields

$$\chi_1(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \, \frac{\omega' \chi_2(\omega')}{\omega'^2 - \omega^2} + \frac{\omega}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \, \frac{\chi_2(\omega')}{\omega'^2 - \omega^2}$$

Now, $\chi_2(\omega)$ is odd so that the second integral vanishes; the part from $-\infty \to 0$ exactly cancels the part from $0 \to \infty$. Further, $\omega\chi_2(\omega)$ is even, so the part of the first integral from $-\infty \to 0$ exactly equals the part from $0 \to \infty$, So I only need to do the integral from $0 \to \infty$ and multiply by two, obtaining†

$$\chi_1(\omega) = \frac{2}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{\omega' \chi_2(\omega')}{\omega'^2 - \omega^2}. \tag{167}$$

I carry out the same derivation for the imaginary part and get

$$\chi_2(\omega) = -\frac{2\omega}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{\chi_1(\omega')}{\omega'^2 - \omega^2}. \tag{168}$$

Eqs. 167 and 168 show the Kramers-Kronig relations in the form usually used for analysis of optical functions.

9.2.6 Case when the dc conductivity is finite.

I have made a mistake (or, perhaps, have omitted something) in the previous section. Suppose the solid has a finite dc conductivity. Many solids do, including metals, superconductors, doped semiconductors, and so forth. Then at low frequencies, $\chi(\omega) = i\sigma_{\text{dc}}/\omega$, which diverges at $\omega = 0$. There is a pole at the origin and the contour of Fig. 40 runs right through it. To handle this, I will write the susceptibility for the conductor as an analytic part χ_a plus the divergent term:

$$\chi = \chi_a + i\sigma_{\text{dc}}/\omega.$$

I can then assert that $\chi_a \equiv \chi - i\sigma_{\text{dc}}/\omega$ is analytic in the upper half plane. The Kramers Kronig integral of Eq. 168 becomes

$$\chi_2(\omega) = \frac{\sigma_{\text{dc}}}{\omega} - \frac{2\omega}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{\chi_1(\omega')}{\omega'^2 - \omega^2}. \tag{169}$$

(Eq. 167 is unchanged.)

---

[*] My signal generator or spectrometer only has positive frequencies on its dial.

[†] If you want to work through this in detail, break the single integral over $-\infty \to \infty$ (Eq. 165) into two, one from $-\infty \to 0$ and a second from $0 \to \infty$. In the first, define $\omega'' = -\omega'$ and substitute for ω' everywhere, including $d\omega'$ and the limits. You will arrive a the results given here.

9.2.7 Kramers-Kronig analysis of other optical functions

There are many complex optical functions in addition to the susceptibility: dielectric function, conductivity, refractive index, etc. and there are Kramers-Kronig relations among all of them. Many can be obtained by substitution. For example, $\epsilon = 1 + 4\pi\chi$ and $\sigma_1 = \omega\chi_2 = \omega\epsilon_2/4\pi$ so that

$$\epsilon_1(\omega) = 1 + \frac{2}{\pi}\mathcal{P}\int_0^\infty d\omega'\,\frac{\omega'\epsilon_2(\omega')}{\omega'^2 - \omega^2}, \tag{170}$$

or

$$\epsilon_1(\omega) = 1 + 8\mathcal{P}\int_0^\infty d\omega'\,\frac{\sigma_1(\omega')}{\omega'^2 - \omega^2}. \tag{171}$$

Also,

$$\epsilon_2(\omega) = \frac{4\pi\sigma_{\text{dc}}}{\omega} - \frac{2\omega}{\pi}\mathcal{P}\int_0^\infty d\omega'\,\frac{[\epsilon_1(\omega') - 1]}{\omega'^2 - \omega^2}, \tag{172}$$

or, equivalently,

$$\sigma_1(\omega) = \sigma_{\text{dc}} - \frac{2\omega^2}{\pi}\mathcal{P}\int_0^\infty d\omega'\,\frac{\sigma_2(\omega')}{\omega'(\omega'^2 - \omega^2)}. \tag{173}$$

Kramers-Kronig relations are commonly used for estimating many optical functions from experiments that give one member of a complex pair. Among these are $\sigma_1 \to \sigma_2$, $\text{Im}(1/\epsilon) \to \text{Re}(1/\epsilon)$, $n \to \kappa$, κ or $\alpha = 2\omega\kappa/c \to n$, and $\mathscr{R} \to \phi$. Of these, the calculation of the phase shift on reflectance from reflectance by far the most commonly used and will be discussed in detail in the following subsection.

Table 8 shows a number of functions used in Kramers-Kronig integrals. But note also that moments of these function also are consistent with the Kramers-Kronig relations as long as the moments satisfy the convergence condition on page 106. I write the general integral as

$$\mathcal{I}(\omega) = \frac{2}{\pi}\mathcal{P}\int_0^\infty dx\,\frac{f(x)}{x^2 - \omega^2} \tag{174}$$

and the show in the left column the form $f(x)$ must take to be put in to Eq. 174. The right column shows the function that would be obtained and its dependence on $\mathcal{I}(\omega)$. The complex quaintites here are the dielectric function, $\epsilon = \epsilon_1 + i\epsilon_2$, the conductivity, $\sigma = \sigma_1 + i\sigma_2$, the susceptibility $\chi = \chi_1 + i\chi_2$, the refractive index $N = n + i\kappa$, the loss function $(1/\epsilon) = \epsilon_1/(\epsilon_1^2 + \epsilon_2^2) - i\,\epsilon_2/(\epsilon_1^2 + \epsilon_2^2)$ and the surface impedance, $Z_s = R_s + iX_s$,

Table 8. A function $f(x)$ is integrated to give $\mathcal{I}(\omega)$. The possibilities for $f(x)$ and the result in terms of $\mathcal{I}(\omega)$ are given here..

$f(x)$	$\rightarrow$	KK		transform
$x\epsilon_2(x)$	$\rightarrow$	$\epsilon_1(\omega)$	$=$	$1+\mathcal{I}(\omega)$
$\epsilon_1(x)-1$	$\rightarrow$	$\epsilon_2(\omega)$	$=$	$(4\pi\sigma_{\mathrm{dc}}/\omega)-\omega\mathcal{I}(\omega)$
$\sigma_1(x)$	$\rightarrow$	$\sigma_2(\omega)$	$=$	$-\omega\mathcal{I}(\omega)$
$\sigma_2(x)/x$	$\rightarrow$	$\sigma_1(\omega)$	$=$	$\sigma_{\mathrm{dc}}+\omega^2\mathcal{I}(\omega)$
$x\chi_2(x)$	$\rightarrow$	$\chi_1(\omega)$	$=$	$\mathcal{I}(\omega)$
$\chi_1(x)$	$\rightarrow$	$\chi_2(\omega)$	$=$	$(\sigma_{\mathrm{dc}}/\omega)-\omega\mathcal{I}(\omega)$
$x\kappa(x)$	$\rightarrow$	$n(\omega)$	$=$	$1+\mathcal{I}(\omega)$
$n-1$	$\rightarrow$	$\kappa(\omega)$	$=$	$-\omega\mathcal{I}(\omega)$
$x[-\operatorname{Im}(1/\epsilon(x)]$	$\rightarrow$	$\operatorname{Re}[1/\epsilon(\omega)]$	$=$	$1-\mathcal{I}(\omega)$
$\operatorname{Re}[1/\epsilon(x)]-1$	$\rightarrow$	$-\operatorname{Im}(1/\epsilon(\omega)]$	$=$	$\omega\mathcal{I}(\omega)$
$xX_s(x)$	$\rightarrow$	R_s	$=$	$Z_0+\mathcal{I}(\omega)$
R_s-Z_0	$\rightarrow$	X_s	$=$	$-\omega\mathcal{I}(\omega)$

$Z_0 = 377\,\Omega = 4\pi/c$ is the impedance of free space.

9.3 Kramers-Kronig analysis of reflectance

When I measure the reflectance,[*] I am taking the ratio of the reflected intensity or power to the incident intensity or power. Phase information is not available.[†] The amplitude reflectivity, Eq. 43, does have a phase; indeed I can write it as

$$r = \rho e^{i\phi} = \frac{1-N}{1+N},$$

where $\rho = \sqrt{\mathscr{R}}$ is the magnitude of the reflectivity, $\mathscr{R}$ is the single-bounce reflectance, ϕ is the phase shift on reflection, and $N = n + i\kappa$ is the complex refractive index.

It would be nice to know ϕ, because I could invert Eq. 43 to get

$$N = \frac{1-\sqrt{\mathscr{R}}e^{i\phi}}{1+\sqrt{\mathscr{R}}e^{i\phi}}, \tag{175}$$

using the known ϕ and the measured reflectance.[‡]

---

[*] Reflectance here means the *single-bounce* reflectance, $\mathscr{R}$, as discussed on p. 21 and given in Eq. 44.

[†] Measuring the phase shift on reflection is difficult. The phase of the electromagnetic field changes by 2π for every wavelength of travel of the light, so that a phase-sensitive measurement requires control of lengths to a small fraction of the wavelength.

[‡] When I looked at this a time after preparing it and tried to do the math in my head, I was sure that there was an algebra error here. But then I recalled that the phase shift for non-absorbing materials with $n > 1$ is $-\pi$ so that r is negative. The range of ϕ is $-\pi \leq \phi \leq 0$. If I wanted to have a positive phase, I could define $\theta = \phi + \pi$ so that $e^{i\theta} = -e^{i\phi}$. In this case,

Kramers-Kronig analysis is one way of estimating this phase.[30,43–45] Consider

$$\ln r = \ln \rho + i\phi.$$

Here, $\ln \rho$ is the real part and ϕ is the imaginary part. The reflectance must be causal, and hence so must be the log of the reflectance.

$$\ln \rho(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \, \frac{\phi(\omega')}{\omega' - \omega}, \tag{176}$$

and

$$\phi(\omega) = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \, \frac{\ln \rho(\omega')}{\omega' - \omega}, \tag{177}$$

The hermiticity of r, $r(-\omega) = [r(\omega)]^*$, proves that ρ is even and ϕ is odd. I use these properties as I did earlier for χ to obtain

$$\phi(\omega) = -\frac{2\omega}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{\ln \rho(\omega')}{\omega'^2 - \omega^2}. \tag{178}$$

Equation 178 is perfectly usable for numerical analysis, but there is one improvement that can be made.[30] Consider

$$\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{1}{\omega'^2 - \omega^2} = 0. \tag{179}$$

The integral is zero because the negative area for $\omega' < \omega$ cancels the positive area for $\omega' > \omega$. Thus, I can add

$$+\frac{2\omega}{\pi}\ln \rho(\omega)\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{1}{\omega'^2 - \omega^2}$$

to the right hand side of Eq. 178 without affecting the phase. Collecting terms, replacing ρ with $\sqrt{\mathscr{R}}$, and using the properties of the log,* I get

$$\phi(\omega) = -\frac{\omega}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{\ln[\mathscr{R}(\omega')/\mathscr{R}(\omega)]}{\omega'^2 - \omega^2}. \tag{180}$$

This modification has two advantages. First, if there are errors in the calibration of the reflectance measurements, so that the data for $\mathscr{R}$ are in error by a constant factor, the

---

$$N = \frac{1 + \sqrt{\mathscr{R}}e^{i\theta}}{1 - \sqrt{\mathscr{R}}e^{i\theta}}$$

and, of course,

$$\rho e^{i\theta} = \frac{N-1}{N+1}.$$

The range of θ is $0 \leq \theta \leq \pi$. The real part of $\rho e^{i\theta}$ becomes negative when $n < 1$ and κ is small, requiring $\theta > \pi/2$.

[*] $\ln\sqrt{\mathscr{R}} = (^1/_2)\ln\mathscr{R}$ and $\ln A - \ln B = \ln(A/B)$.

results for ϕ are unaffected. Second, both numerator and denominator of the integrand are zero when $\omega' = \omega$). L'Hôpital's rule shows that the ratio does not diverge; hence the pole has been removed. The second consequence is more important than the first. Removing the pole is important for simple evaluation of the integral. In contrast, even if the scale error does not affect the phase, it *does* affect $\mathscr{R}$, and N, Eq. 175, depends on both quantities.

As a final note in this section, Eq. 179 shows that the factor $[\epsilon(\omega') - 1]$ may be replaced by $\epsilon(\omega')$ in Eqs. 172 and 173. However, if I want to do a numerical integral, I will prefer to keep it as it is, so that $[\epsilon(\omega') - 1] \to 0$ as $\omega \to \infty$.

9.3.1 Extrapolations

The alert reader will have noticed that the range of the integrals in Eqs. 167–173 and 180 is 0–∞ and may wonder how one acquires data over that entire range. The answer of course that data are always limited to a finite range of frequencies. Thus, the user must use extrapolations outside the measured ranges. I will discuss these extrapolations in the context of Kramers-Kronig analysis of reflectance, but the same approaches can be used for other functions.

One must estimate the reflectance between zero and the lowest measured frequency. The best approach is to employ a model that reasonably describes the low-frequency data. Such models include Drude for metals, Lorentzian for insulators, often a sum of several Lorentzians, sometimes a Drude plus Lorentzians. Many other functions exist; some are briefly described in Appendix G. When a good fit of the model to the data is obtained, a set of reflectance points may be calculated between zero and the lowest measured frequency, using a spacing between points similar to that of the lowest-frequency data, and combined with the measured data.

Other approaches to the low-frequency extrapolations include

- Reflectance is assumed constant to DC.
- Hagen Rubens formula (metallic): $\mathscr{R} = 1 - A\omega^{1/2}$.
- Two-fluid model: $\mathscr{R} = 1 - A\omega^2$.
- Marginal Fermi Liquid: $\mathscr{R} = 1 - A\omega^1$.
- Superconducting: $\mathscr{R} = 1 - A\omega^4$.

The constant A is adjusted so the extrapolation goes through the first few points and then, using a spacing between points similar to that of the lowest-frequency data, a set of reflectance points is calculated between zero and the lowest measured frequency and combined with the measured data.

The high-frequency extrapolation can be a source of major error. It is good to use data from other experiments on identical or similar samples if these exist. Moving to the highest frequencies, one knows that in the limit as $\omega \to \infty$, the dielectric function is mostly real and slightly smaller than unity, following $\epsilon = 1 - \omega_p^2/\omega^2$, where ω_p is the plasma frequency of all the electrons in the solid and $\omega \gg \omega_p$. Then $n = 1 - \omega_p^2/2\omega^2$ and $\mathscr{R} = \omega_p^4/8\omega^4 \equiv C/\omega^4$. Typically, the region between the highest-frequency data point and the transition to ω^{-4} is filled with a weaker power law, $\mathscr{R} \sim B/\omega^s$ with $0 < s < 4$. (s does not have to be an integer.) The value of B is chosen so that the power law joins smoothly to the data at the high frequency limit; the value of C is also chosen for a smooth transition between mid- and high-frequency extrapolations. The free parameters are the exponent s and the frequency for the crossover from ω^{-s} and ω^{-4}.

9.3.2 Kramers-Kronig analysis of transmittance

The (coherent) transmittance $\mathscr{T}$ of a thin* parallel-plate slab is related to the amplitude transmission coefficient by

$$\mathscr{T} = |te^{i\phi_t}|^2$$

where t is the amplitude and ϕ the phase change on passing through the sample, including the coherent addition of multiple internal reflections.

As in reflectance, $\ln t + i\phi_t$ are causal functions that satisfy causality,[46–49]

$$\phi_t(\omega) = -\frac{\omega}{\pi}\mathcal{P}\int_0^\infty d\omega' \, \frac{\ln[\mathscr{T}(\omega')/\mathscr{T}(\omega)]}{\omega'^2 - \omega^2} + \frac{\omega}{c}d, \tag{181}$$

where d is the thickness and the last term is a correction for the phase gained on going through an identical thickness of vacuum. The equation for the transmission of a slab in terms of the refractive index or dielectric function will be discussed in section 6. Given $\mathscr{T}$ and ϕ_t, these may be inverted numerically for n and κ.

Like reflectance, Kramers-Kronig analysis of transmittance requires extrapolations. At low frequencies, I use one of these:

- Transmission is assumed constant to DC
- Metallic conductivity assumed: $\mathscr{T} = B + C\omega^2$
- Superconducting: $\mathscr{T} = A\omega^2$

Above the experimental regime, it is good to use data from other experiments on identical or similar samples if these exist. At the highest frequencies, in the x-ray range, the data can be extrapolated as $\mathscr{T} \sim 1 - C\omega^{-4}$ as appropriate for completely free electrons. Between the highest data point and the x-ray range a slower power law is typically used.

9.4 Another look at the conductivity

The conductivity σ, which appears in Ohm's law, $\mathbf{j} = \sigma\mathbf{E}$, is a well behaved function with no poles in the upper half plane. At very high frequencies, the conductivity is that of free carriers, Eq. 152. Thus it falls like $1/\omega$ and therefore (barely) satisfies the requirements on page 106. Moreover σ_1 is an even function and σ_2 is odd, so that the derivation from Eq. 163 to 168 goes through exactly, and I arrive at[31]

$$\sigma_1(\omega) = \frac{2}{\pi}\mathcal{P}\int_0^\infty d\omega' \, \frac{\omega'\sigma_2(\omega')}{\omega'^2 - \omega^2}, \tag{182}$$

and

$$\sigma_2(\omega) = -\frac{2\omega}{\pi}\mathcal{P}\int_0^\infty d\omega' \, \frac{\sigma_1(\omega')}{\omega'^2 - \omega^2}. \tag{183}$$

Equation 183 is identical to Eq. 171 but Eq. 182 differs from Eq. 173. The two integrals employ different moments of σ_1.

---

* The thickness scale is λ/n, the wavelength divided by the refractive index. If the slab is several wavelengths thick, the requirements on the variations in thickness and in the parallelism of the light passing through it become severe.

9.5 Poles and zeros

9.6 Sum rules

A sum rule is a statement that the integral of an optical function has a fixed value, is zero, or is related to a specific value of another optical function. I introduced the oscillator strength sum rule (or f sum rule) on page 95. Here I will briefly discuss what the Kramers-Kronig relation says about sum rules.[33,34,50–56]

9.6.1 The f sum rule

I start by writing the high-frequency limiting value of the real part of the dielectric function. (I have written this several times before.)

$$\epsilon_1(\omega) = 1 - \frac{\omega_p^2}{\omega^2},$$

where ω_p is a plasma frequency for every electron in the solid. In the notation of Eqs. 142 and 150,

$$\omega_p = \sqrt{\frac{4\pi \mathfrak{n} e^2}{m}},$$

where $\mathfrak{n}$ counts *every* electron in the solid.

I equate this equation to the Kramers-Kronig integral of Eq. 171, keeping in mind that ω must be very large,

$$1 - \frac{\omega_p^2}{\omega^2} = 1 + 8\mathcal{P}\int_0^\infty d\omega' \, \frac{\sigma_1(\omega')}{\omega'^2 - \omega^2}.$$

I now split the integral into two integrals, with the boundary frequency, ω_f, less than ω but high enough that $\sigma_1(\omega_f)$ is essentially zero:

$$-\frac{\omega_p^2}{8\omega^2} = \mathcal{P}\int_0^{\omega_f} d\omega' \, \frac{\sigma_1(\omega')}{\omega'^2 - \omega^2} + \mathcal{P}\int_{\omega_f}^\infty d\omega' \, \frac{\sigma_1(\omega')}{\omega'^2 - \omega^2}$$

Because $\sigma_1(\omega_f)$ is arbitrarily small, the second integral can be neglected. Because $\omega \gg \omega_f$, the term ω'^2 in the first integral can be neglected. Hence

$$-\frac{\omega_p^2}{8\omega^2} = \mathcal{P}\int_0^{\omega_f} d\omega' \, \frac{\sigma_1(\omega')}{-\omega^2}$$

I can now change sign, cancel the factors of ω^{-2}, drop the principal value symbol (there are no divergences of the integrand), and extend the upper limit to ∞ (because by ω_f I have already gotten all the area under σ_1) to get an equation identical to Eqs. 150 and 151:

$$\int_0^\infty d\omega' \, \sigma_1(\omega') = \frac{\omega_p^2}{8} = \frac{\pi}{2}\frac{\mathfrak{n} e^2}{m}. \tag{184}$$

See page 98 for an evaluation of Eq. 184 for a Drude conductivity.

Equation 184 is also a sum rule on $\omega\epsilon_2$ because $\omega\epsilon_2 = 4\pi\sigma_1$; the right hand side is 4π larger of course. A similar development of the energy loss function, $\mathrm{Im}(1/\epsilon)$ gives:

$$\int_0^\infty d\omega'\,\omega'\,\mathrm{Im}\left(-\frac{1}{\epsilon(\omega')}\right) = \frac{\pi\omega_p^2}{2} = \frac{2\pi^2 \mathfrak{n} e^2}{m}. \tag{185}$$

For a simple free-electron (Drude) metal, σ_1 has a maximum at zero frequency and rolls off with characteristic width $1/\tau$. The area under σ_1 mostly occurs at low frequencies. The loss function has a peak at ω_p, with width (in a simple model where τ is not a function of frequency) $1/\tau$. But the integrals of both functions are proportional to the plasma frequency squared.

9.6.2 The static dielectric function sum rule

Equation 171 shows that the real part (the dispersive or reactive part) of the dielectric function is related to the real part of the optical conductivity. Evaluating this equation for $\omega = 0$ yields the static dielectric function as the ω^{-2} moment of σ_1,

$$\epsilon_1(0) = 1 + 8\mathcal{P}\int_0^\infty d\omega'\,\frac{\sigma_1(\omega')}{\omega'^2}. \tag{186}$$

This equation is not useful for conductors, because there is a divergence of the integrand at $\omega' = 0$ unless σ_1 is zero there. For insulators, one can see that conductivity peaks at low frequencies are the largest factors in determining the static dielectric constant. Peaks with high values of the optical conductivity located at low frequencies, are required for high values of the the dielectric constant.* Moreover, there is a trend for materials with high dielectric constants or refractive indices in their transparent regions to have small band gaps; large bangap materials have low values of ϵ and n.

9.6.3 Dc conductivity sum rules

A third sum rule uses Eq. 182, evaluated at $\omega = 0$:

$$\int_0^\infty d\omega'\,\frac{\sigma_2(\omega')}{\omega'} = \frac{\pi}{2}\sigma_{\mathrm{dc}},$$

or, substituting for σ_2,

$$\int_0^\infty d\omega'\,[1 - \epsilon_1(\omega')] = 2\pi^2\sigma_{\mathrm{dc}}. \tag{187}$$

Eq. 187 makes it evident that it was OK to convert the principal value integral into a regular integral. This equation says that the parts of ϵ_1 that are bigger than unity are balanced by parts below unity except for the free carrier pole. Note that all solids have real parts of their dielectric functions that are slightly smaller than unity at very high energies, becoming unity asymptotically as the frequency goes to infinity.

---

* The high-temperature paraelectric phase of many ferroelectric insulators has a "soft mode," a lattice vibration where the transverse-optic frequency goes to zero as the temperature is reduced towards the Curie temperature. The dielectric constant thus diverges at the transition.

9.6.4 Sum rules for the refractive index

The real and imaginary parts of the refractive index satisfy Kramer's Kronig relations. These are given in Table 8 (and Eq. 174). I see that

$$n = 1 + \frac{2}{\pi}\int_0^\infty d\omega' \, \frac{\omega'\kappa(\omega')}{\omega'^2 - \omega^2}$$

As in the derivation for the integral of σ_1, I argue that $\kappa \to 0$ as $\omega \to \infty$ fast enough that I can stop the integral as some frequency $\omega_f \ll \omega$. I write for these large values of ω,

$$n = \sqrt{\epsilon} \approx 1 - \frac{\omega_p^2}{2\omega^2}$$

where the factor of 2 comes from expanding the square root. Then

$$-\frac{\omega_p^2}{2\omega^2} = \frac{2}{\pi}\int_0^\infty d\omega' \, \frac{\omega'\kappa(\omega')}{-\omega^2}$$

or

$$\int_0^\infty d\omega' \, \omega'\kappa(\omega') = \frac{\pi}{4}\omega_p^2. \tag{188}$$

Equation 188 is the f sum rule for the extinction coefficient. It could be written also in terms of the absorption coefficient α, because $\alpha = 2\omega\kappa/c$

Now, I can write Eq. 184 in terms of the imaginary part of the dielectric function, where it reads

$$\int_0^\infty d\omega' \, \omega'\epsilon_2(\omega') = \frac{\pi}{2}\omega_p^2,$$

so that the result of the integrations is just twice that of Eq. 188. Therefore

$$\int_0^\infty d\omega' \, \omega'\epsilon_2(\omega') = 2\int_0^\infty d\omega' \, \omega'\kappa(\omega').$$

But $\epsilon_2 = 2n\kappa$! Hence, after minimal algebra, this equation becomes

$$\int_0^\infty d\omega' \, \omega'[n(\omega') - 1]\kappa(\omega') = 0 \tag{189}$$

Finally, I can start with the Kramers-Kronig integral for κ in terms of n from Table 8 and do the usual high frequency evaluation (where $\kappa = 0$) to obtain a sum rule for the refractive index alone,

$$\int_0^\infty d\omega' \, [n(\omega') - 1] = 0, \tag{190}$$

known as the inertia sum rule.

I think that it is remarkable that the integral of n, whether weighted by $\omega\kappa$ or not, comes out to be the same as the refractive index of the vacuum. At some frequencies n is larger than one and at others smaller than one. But the average (in two senses) value is unity.

9.7 Partial sum rules

The sum rule integrals extend from zero to infinity. It is often informative to consider a partial sum rule. I can carry out the integral from zero to some frequency ω, and then plot the result as a function of the upper limit of the integration. I'll formulate the problem as follows The right hand side of Eq. 184 can be written

$$\frac{\pi \mathfrak{n} e^2}{2m} = \frac{\pi e^2}{2mV_c} N_{\mathit{eff}} \frac{m}{m^*}$$

where m^* is the effective mass, m is the free electron mass, V_c is the unit cell volume (or formula volume), and N_{eff} is the number of effective electrons within the volume V_c participating in optical transitions at frequencies lower than ω. Then I can write

$$\frac{m}{m^*} N_{\mathit{eff}}(\omega) = \frac{2mV_c}{\pi e^2} \int_0^{\omega} \sigma_1(\omega') d\omega' \qquad (191)$$

for the number* of electrons with effective mass ratio m/m^* participating in optical transitions at frequencies below ω.

Figure 41 shows $N_{\mathit{eff}}(m/m^*)$ for a a model of a metal similar to Ag. There is a Drude free carrier component, with $1/\tau \approx 130$ cm-1 and three Lorentz oscillators, at 40,000 cm^{-1} (5 eV), 54,000 cm^{-1} (5.7 eV) and 70,000 cm^{-1} (8.7 eV). The conductivity is shown in the inset. The main plot shows the partial sum rule. The integral of the Drude conductivity saturates at 1 electron/Ag atom. $N_{\mathit{eff}}(m/m^*)$ is flat over a wide frequency range and then steps up when the interband transitions begin.

Similar partial sum rules may also be written for the loss function or the extinction coefficient (absorption coefficient).

I may also write a partial sum rule based on Eq. 186,

$$\delta\epsilon_0(\omega) = 1 + 8\mathcal{P} \int_0^{\omega} d\omega' \, \frac{\sigma_1(\omega')}{\omega'^2}. \qquad (192)$$

Note that this is equivalent in some ways to writing the dielectric function as a sum of Lorentz oscillators, as in Eqs. 76 and 85, and evaluating for zero frequency:

$$\epsilon(0) = 1 + \sum_j \frac{\omega_{pj}^2}{\omega_j^2},$$

where I've set $\epsilon_c = 1$. If the oscillators are narrow compared to their separation, the integral in Eq. 192 will step up as ω passes through ω_j.

---

[*] There is occasional confusion about the value to use for the volume V_c, which is described variously as the "volume of the unit cell" or the "volume of one formula unit." Of course any carefully described volume will work, but generally one wants the number of charge carriers per atom or per primitive unit cell. And of course the conventional cell often contains several times more atoms than one would guess from the chemical formula. So the user must decide what he or she wants to determine. It may be the number of carriers per Ag atom in silver metal, the number per buckyball in C_{60}, the number per dopant in P-doped Si, or the number per copper atom in $YBa_2Cu_3O_7$. Once the decision is made, compute the volume in the crystal allocated to the desired quantity. One (almost) failsafe approach is to obtain the density ρ of the crystal and the mass M of the entity one is interested in, such as one Ag atom, 60 C atoms, one silicon atom divided by the dopant concentration, or 1/3 the mass of Y + 2Ba + 3Cu + 7O. Then $V_c = M/\rho$.

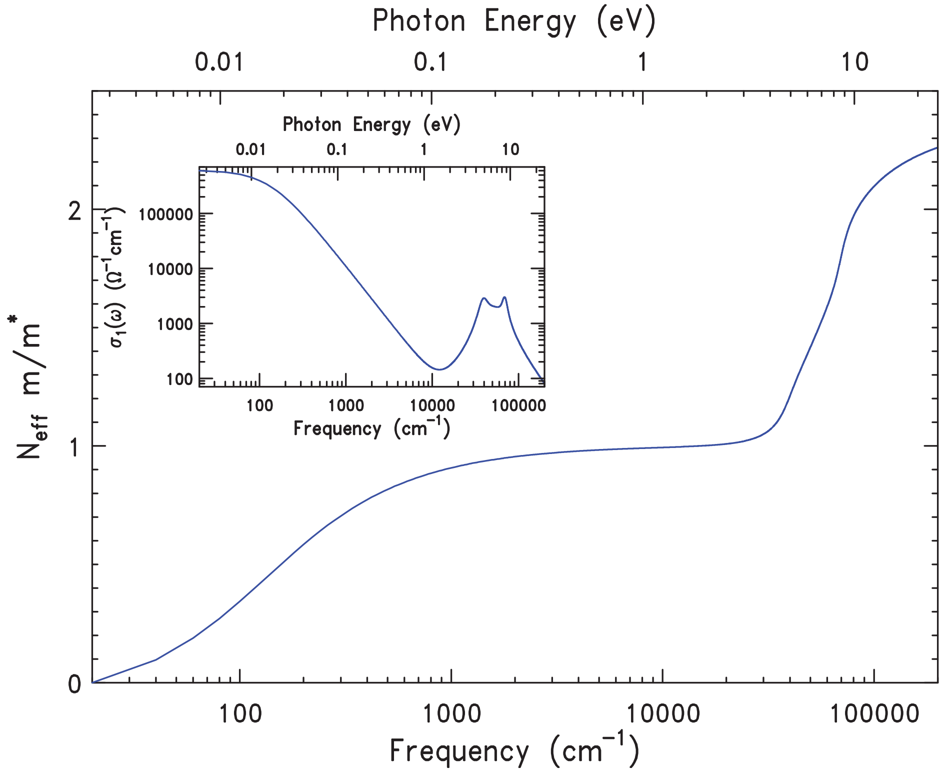

Fig. 41. $N_{eff}(m/m^*)$ for a a model of a metal similar to Ag. The inset shows the conductivity whose integral is shown in the main plot. V_c s the volume allocated to one Ag atom.

A third partial sum rule that I have seen used is based on Eq. 190. (One based in Eq. 189 is also possible.) I write

$$\delta n(\omega) = \int_0^{\omega} d\omega' \, [n(\omega') - 1].$$

This function will start positive and increase as ω increases. Indeed, because the refractive index often looks like the derivative of the extinction coefficient, it will resemble the latter function. But when n becomes smaller than unity at high frequencies, $\delta n(\omega)$ will begin to decrease with increasing ω and (if all has gone well) ramp down to zero.

10. SEMICONDUCTORS AND INSULATORS

In this Chapter, I will address interband transitions in semiconductors and insulators. All solids show these interband optical transitions. In many ways they are analogs of the electric-dipole transitions between principal quantum numbers in atoms, such as $1s \rightarrow 2p$ in hydrogen. Up to now I have treated these interband transitions in one of two ways. (1) Accumulate the low-energy contributions of these transitions into ϵ_c. (2) Include an electronic Lorentz oscillator as representing the transition between two levels separated by $\hbar\omega_e$. Neither approach is very satisfactory.*

10.1 Band structure

Before considering the optical excitations, the so-called interband transitions, I first should say something about what these "bands" really are. Free electrons do not have bands. The energy is solely the kinetic energy $\hbar^2 k^2/2m$; I take the electrons and pour them into these free-electron states until I have run out of electrons.

In reality of course, no electron is actually free. All interact with the atomic nuclei and with other electrons. The innermost ones are deeply bound in atomic-like very low-energy states. The most energetic electrons may be nearly free, with the details determined by the chemical and crystallographic structure of the solid. Even states above the Fermi energy (which are unoccupied—or empty—at low temperatures) will be affected by the band structure. The electron energies turn out to be organized into a range of quasi-continuous energies, called "bands," separated by *forbidden energy gaps* where no states exist.

10.1.1 Some hand-waving arguments

Let me start by considering an isolated atom. I know that the energies of electrons bound to this atom are quantized into discrete levels and that electromagnetic absorption (photons) will promote electrons from one level to another. Consider the ground state wave function of a simple atom, such as hydrogen. Two H atoms and their $1s$ orbitals are shown in the upper panel of Fig. 42. The wave function falls off quickly away from the central nucleus. When two hydrogens are allowed to interact, the two wave functions can add either constructively or destructively. Constructive addition leads to the bonding orbital; destructive to the antibonding orbital.† It is evident that the bonding orbital has much more electronic amplitude in the space between the two protons. This counteracts the repulsion felt by the two protons and lowers the energy of the H_2 molecule. The antibonding orbital has small amplitude in the space between, and leads to a significant increase in the energy. These effects are sketched in the bottom panels of Fig. 42. The energy levels are shown in Fig. 43. The bonding level holds the two electrons amd the antibonding level—which could hold two more—is empty.

Now consider an long chain of such atoms, containing N atoms, all separated by lattice constant a. There will be many ways to form linear combinations of the wave functions and

---

* Although if all I am interested in is the dielectric function at low frequencies—well below the energy of interband effects–they may prove adequate.

† In this discussion I am leaving out the beautiful physics that comes from the spin states of the electrons and the symmetry that the two-electron wave function must have under exchange. The bonding orbital is a singlet, $S = 0$, and the antibonding orbital a triplet, $S = 1$.

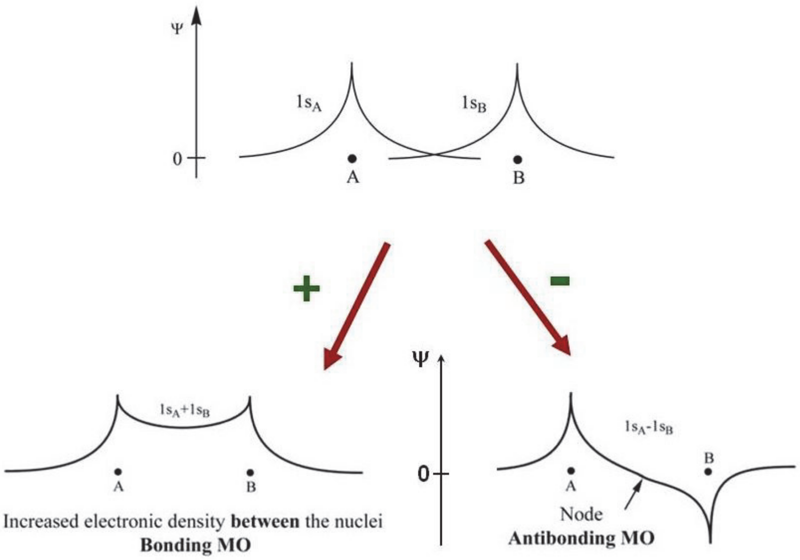

Fig. 42. Upper panel: Hydrogen ground-state wave function for two atoms: A and B. Lower panel: the two wave functions can either be added or subtracted. Addition leads to the bonding orbital; subtraction to the antibonding orbital.

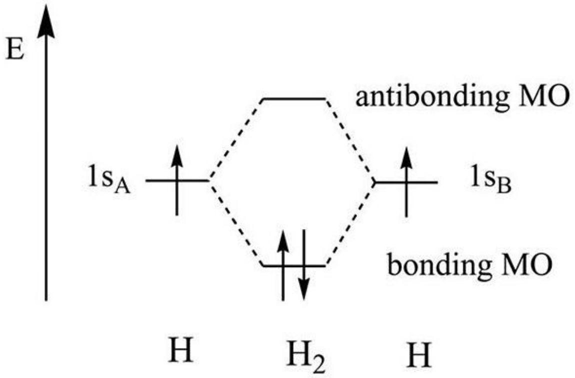

Fig. 43. Splitting of the bonding and antibonding levels in H_2.

so there will be many levels, ranging from one where all are added, so all in phase, to one where each is 180° out of phase with its two neighbors. Indeed there will be N such levels. If there is one electron/atom, half will be full and half empty. I claim that this linear chain would be a metal.*

If there were two electrons per atom, the uniform chain would have every level filled with two electrons, one spin up and one spin down. I claim this would then be an insulator.

I have gone about as far as I can without doing some calculations. I could follow the ideas above and do a *tight-binding* model. But instead, let me return to my free electron model and add, as a perturbation, a periodic potential.

---

[*] One dimensional metallic chains are susceptible to a Peierls[57] or charge-density-wave transition, where the bond lengths become alternating, short, $a - u$, and long, $a + u$. This doubles the unit cell to $2a$, and gives two electrons per dimer, e.g., a chain of H_2 molecules. This dimerized state is an insulator.

10.2 Nearly free electrons

The nucleii in the solid form a periodic lattice, so that the potential satisfies

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{T}) \tag{193}$$

where $\mathbf{T}$ is the translation vector of the crystal. In the one-dimensional case I will consder, $\mathbf{T} = na\hat{\boldsymbol{x}}$ where the lattice constant is a and n is an integer.

Now if $V = 0$, $\mathcal{E} = \hbar^2 k^2/2m$ (free electrons) with $k = 2\pi j/L$ (periodic boundary conditions, j is an integer). What is the effect of the periodic potential? Well, the free electron states are plane waves, $\psi = e^{ikx}/\sqrt{L}$ and plane waves traveling in a periodic medium are diffracted. In one dimension, the incident wave and the diffracted wave must both travel along $\hat{\boldsymbol{x}}$. Diffraction effects thus produce waves traveling in both directions along the chain of atoms.*

In general these waves just continue along forever. (There is no absorption.) But for certain values of k, I can get strong interference effects.† Consider a partial wave which travels forward for distance a, is reflected and travels backward for distance a, and is reflected again. If the total distance traveled is an integer times the wavelength, there will be constructive interference. So the condition for diffraction is $2a = \lambda$. But of course $k = 2\pi/\lambda$ so the condition is also

$$k = m\frac{\pi}{a}$$

where m, the diffraction order, is a positive or negative integer. Look at the case where $m = \pm 1$. (Other order have higher energies.) Because of diffraction and constructive interference, the wave functions become standing waves:

$$\psi_1 = \frac{e^{i\pi x/a} + e^{i\pi x/a}}{\sqrt{2L}} = \sqrt{\frac{2}{L}}\cos(\frac{\pi x}{a})$$

and

$$\psi_1 = \frac{e^{i\pi x/a} - e^{i\pi x/a}}{\sqrt{2L}} = i\sqrt{\frac{2}{L}}\sin(\frac{\pi x}{a})$$

where I have done a side calculation for the normalization. The standing waves lead to a non-uniform density,

$$|\psi_1|^2 = \frac{2}{L}\cos^2(\frac{\pi x}{a})$$

and

$$|\psi_1|^2 = \frac{2}{L}\sin^2(\frac{\pi x}{a})$$

One probability density has crests at $x = 0, \pm a, \pm 2a, \ldots$ and the other has crests half a lattice constant away.

---

[*] I can also consider this as specular reflection by each atom. Note that if there are very many atoms, the effect of very many scattering events, even very weak ones, is to have equal amplitudes flowing in both directions.

[†] Interference is of course the basis of diffraction.

Because the potential $V(x)$ is periodic, it can be written as a Fourier series; because the period is a, the longest wavelength component is a, making the lowest Fourier component $2\pi/a$. Let me consider that component* alone,

$$V(x) = -V_0 \cos(2\pi \frac{x}{a})$$

with a minimum at the lattice sites 0, $\pm a$, etc. and a maximum between the atoms. I treat this as a perturbation,

$$\begin{aligned}\mathcal{E}_1^{(1)} &= \langle \psi_1 | V | \psi_1 \rangle \\ &= -\frac{2}{L} V_0 \int_0^L dx\, \cos(2\pi \frac{x}{a}) \sin^2(\pi \frac{x}{a})\end{aligned}$$

but $\sin^2(\phi) = [1 - \cos(2\phi)]/2$ and so

$$= -\frac{1}{L} V_0 \int_0^L dx[\cos(2\pi \frac{x}{a}) - \cos^2(2\pi \frac{x}{a})$$

The first term in the integral is zero and the second is $-L/2$, so that

$$\mathcal{E}_1^{(1)} = \frac{V_0}{2}$$

and, for $\psi_2 \sim \cos(kx)$

$$\begin{aligned}\mathcal{E}_2^{(1)} &= -\frac{2}{L} V_0 \int_0^L dx\, \cos(2\pi \frac{x}{a}) \cos^2(\pi \frac{x}{a}) \\ &= -\frac{V_0}{2.}\end{aligned}$$

For the states with $k = \pm\pi/a$, there is an energy difference between the two solutions. One (the one that goes as $\sin(kx)$) is raised and the other (that goes as $\cos(kx)$) is lowered. That it goes like this is what one might have expected: the $\cos(kx)$ function has the electron located at the atoms, where the potential is a minimum while the $\sin(kx)$ function has the electron between the atoms, where the potential is maximum.

The jump in energy of magnitude V_0 at $k = \pm\pi/a$ means that there are is a range of energies for which there are no solutions to the Schrödinger equation. Note that the number of k states is unchanged, for spin-1/2 electrons there are $2\frac{2\pi/a}{2\pi/L} = 2N$ states between $k = -\pi/a$ and $k = +\pi/a$.† Moreover, there are no disallowed values of k—only disallowed values of $\mathcal{E}(k)$. These disallowed energies are call *forbidden energy gaps.* Additional gaps appear at $k = \pm 2\pi/a$, $\pm 3\pi/a$, ...

The interference/diffraction effect affects‡ $\psi(k = \pm\pi/a)$, which has $\lambda = 2a$, the most but nearby wavelengths are also affected. A calculated dispersion relation is shown in Fig. 44

---

* There is a dc component of course, but I can suppress that by choosing the zero of energy equal to the dc component.

† Plus one. The range is more properly stated as $-\pi/a < k \le \pi/a$.

‡ Sorry ...

Three bands are shown in blue, green and red; the shaded regions are the gaps. The underlying parabolic free-electron curve is close to these curves in the middle of the bands, but passes first above and then below them near the band gaps.[*]

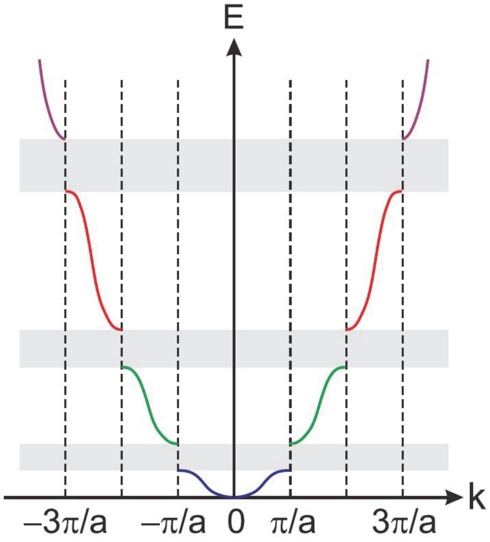

Fig. 44. Energy as a function of the wave vector for a solid with a periodic lattice potential. There are 3 bands shown completely, plus part of a fourth. The energy gaps are shaded grey.

I can now understand the reason why some materials are metals and some insulators. Suppose the linear chain is made up of Hydrogen, which has 1 electron per atom. Or suppose it is potassium or silver, with effectively one *valence* electron per atom. Then there are N electrons and $2N$ valid k values, so the Fermi wave vector is $k_F = \pm\pi/2a$; half the states in the first band (blue) are occupied at $T = 0$ and half unoccupied.

If there are two electrons per atom, this one-dimensional solid would have the first band full and a gap in energy between the highest occupied state (the most energetic state in the first band) and the lowest unoccupied state (the least energetic state in the second band). Then $k_F = \pm\pi/a$ and the electrons cannot be accelerated (with a corresponding increase in momentum) without acquiring additional energy. The solid would be an insulator.[†]

If there are 3 valence electrons per atom (aluminum), the one-dimensional Fermi wave vector is $k_F = \pm 3\pi/2a$. The first band is full but the second is half-full. Metal.

If 4 (silicon, germanium), then $k_F = \pm 2\pi/a$ and the first and second bands are both

---

[*] It typically passes through the middle of the gaps.

[†] On first hearing this, it is natural to ask why the divalent alkaline earth elements (beryllium, magnesium, calcium, ...) are in fact metals. The reason is that the electrons have dispersion in the $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ directions. Suppose I say that the symmetry of the periodic lattice is such that the diffraction occurs when $\{k_x, k_y, k_z\} = \pm\pi/a$, so that the surface in k-space where the gaps occur is a cube. For electrons with $\mathbf{k}\|\hat{\boldsymbol{x}}$ there is a gap at $\{k_x = \pi/a, k_y = 0, k_z = 0\}$ at a free-electron energy of $\hbar^2\pi^2/2ma^2$. But if I think about electrons whose wave vector is such that $k_x = k_y = k_z$, I reach the cube corner where the free electron energy is $3\hbar^2\pi^2/2ma^2$, three times higher. In many materials, this energy exceeds $\hbar^2\pi^2/2ma^2 + V_0$, so that the Fermi surface, a surface of constant energy, lies in the second band for $\mathbf{k}\|\hat{\boldsymbol{x}}$ and in the first band for k along the diagonal. The material would then be a metal.

full. The material would be an insulator.*

10.3 Bloch's theorem

Bloch's theorem is the basis for much of the picture solid-state physics has for electrons, phonons, and other quantum-mechanical objects in a crystal with translational invariance.† It states that the wave functions are of the form of a product of plane waves and a periodic function that respects the symmetry of the lattice. It reads:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \tag{194}$$

where $u_{\mathbf{k}}(\mathbf{r})$ is periodic, as in Eq. 193. So $u_{\mathbf{k}}(\mathbf{r})$ describes the variation of the wave function within the unit cell and the plane-wave exponential function gives the phase between unit cells. If $u_{\mathbf{k}} =$ constant, then Eq. 194 is the wave function for free electrons.

Equation 194 is stated for three-dimensional crystals, but for simplicity I'll discuss it only in one dimension. The Schrödinger equation is

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2} + V(x)\psi(x) = \mathcal{E}\psi(x)$$

where the potential energy function stratifies $V(x + na) = V(x)$ for n an integer and a the lattice constant. I use periodic boundary conditions, so that $\psi(x+L) = \psi(x)$. Translational symmetry implies that the probability density satisfies

$$|\psi(x + na)|^2 = |\psi(x)|^2. \tag{195}$$

Now let me translate from one unit cell to the next ($n = 1$) and see what can be the differences between the wave functions:

$$\psi(x + a) = A\psi(x) \tag{196}$$

where Eq. 195 requires that $A^*A = 1$, so A has length one and some phase. If I translate to the next neighbor along the chain. Thus

$$\psi(x + 2a) = A\psi(x + a) = A^2\psi(x)$$

Now comes the big step: translate by N lattice constants. Then

$$\psi(x + Na) = A^N\psi(x) = \psi(x) \tag{197}$$

---

* As indeed are Si and Ge. They are called *semiconductors* for historical reasons. The band gap is not so large in these materials as in, for example, Al_2O_3 or diamond. So at room temperature the Fermi-Dirac distribution function is not zero in the third band, so there are thermally excited electrons that can conduct in a semi-efficient way. There are also vacancies in the second band, so that there are also *holes* to contribute to the conductivity.

† Here, translational invariance means that the structure repeats periodically and for ever and ever as I move through space.

where the second equality comes from the periodic boundary conditions. They require that $A^N = 1$ or $A = e^{2\pi i n/N}$ so that the Nth power of A is unity. I then rewrite Eq. 196 as

$$\psi(x+a) = e^{\frac{2\pi i n}{N}}\psi(x) \tag{198}$$

I now write

$$\psi(x) = e^{ikx}u_k(x). \tag{199}$$

I make no assumptions about the values of k or the nature of the function $u(x)$. Now translate by one lattice constant to get

$$\psi(x+a) = e^{ik(x+a)}u(x+a) = e^{\frac{2\pi i n}{N}}e^{ikx}u(x)$$

where the first equality comes from Eq. 199 and the second from Eq. 198. I cancel e^{ikx} and am left with quantization of k by the periodic boundary conditions.

$$ka = \frac{2\pi n}{N} \quad \text{or} \quad k = \frac{2\pi}{a}\frac{n}{N}$$

and therefore the function u satisfies

$$u(x+a) = u(x),$$

completing the proof of Bloch's theorem.

10.4 The Brillouin zone

Bloch's theorem, the periodic lattice, and the boundary conditions all conspire to limit what values k can take. I began with it as a real number (a quantity along a continuous line) taking any value from negative to positive infinity. Then, periodic boundary conditions restricted it to a countable infinity, represented by integers. Now I will show that only a finite range of these integer-based k values has distinct physical significance.

I start by once again translating by a single lattice constant

$$\begin{aligned}\psi(x+a) =& u_k(x+a)e^{ik(x+a)} = u_k(x)e^{ikx}e^{ika}\\ =& e^{ika}\psi(x)\end{aligned} \tag{200}$$

Now, define a wave vector k' by

$$\frac{k = k' + 2\pi m}{a}$$

with m an integer. Substitute into Eq. 200 to find

$$e^{ika}\psi(x) = e^{ik'a}e^{2\pi i m}\psi(x)$$

But $e^{2\pi i m} = 1$ so

$$e^{ika}\psi(x) = e^{ik'a}\psi(x).$$

The state with wave vector k and wave vector $k + 2\pi m/a$, with m an integer, are the same. The wave function is the same as is the energy eigenvalue. k is periodic with period $2\pi/a$. Values of k in the range

$$-\frac{\pi}{a} < k \leq \frac{\pi}{a} \tag{201}$$

are all I need.

I am free to modify Fig. 44 by adding or subtracting integer times $2\pi/a$ as many times as needed to bring it into the range specified by Eq. 201 This translation produces the dispersion relations to those shown in Fig. 45.

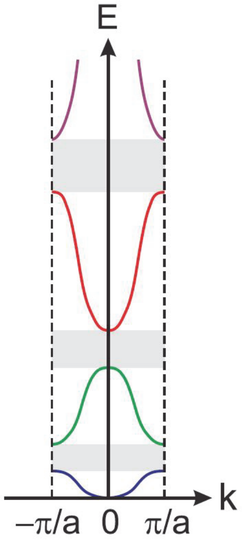

Fig. 45. Energy as a function of the wave vector in the reduced zone scheme. There are 3 bands shown completely, plus part of a fourth. The energy gaps are shaded grey.

The range of k specified by Eq. 201 is said to be in the first Brillouin zone. The bands lie one above another. Despite this drastic change, the basic discussion of page 124 as to whether a given material is metal or insulator is unchanged.

I may also apply translations by $2\pi m/a$ to extend each of the bands. The curves join smoothly and produce Fig. 46.

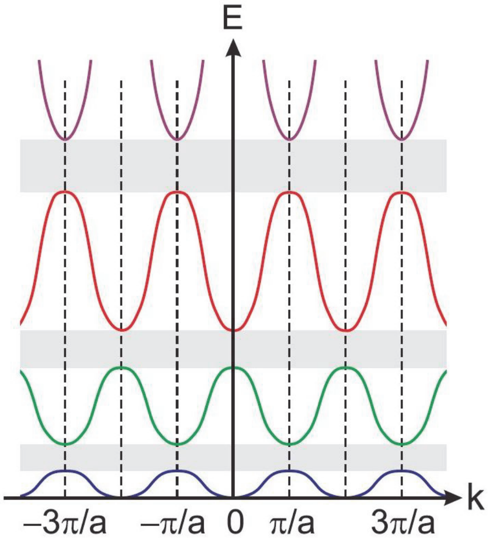

Fig. 46. Energy as a function of the wave vector in the repeated zone scheme. There are 3 bands shown completely, plus part of a fourth. The energy gaps are shaded grey.

Table 9, from wikipaedia,[58] lists a large number of semiconductors, showing the material, chemical formula, band gap and whether the gap is direct or indirect. (The band structures in Fig. 44 is direct band gap for divalent and tetravalent materials. See p. `<indirect>` for more on this.)

10.5 Band gaps in semiconductors

Table 10 shows the band gaps of a number of semiconducting materials

The data in the table of semiconductor materials band gap comes from `http://en.wikipedia.org/wiki/List\_of\_semiconductor\_materials` and includes data from a variety of sources.[11,59–67]

The following semiconducting systems can tuned to some extent, and represent not a single material but a class of materials.

10.6 Effective mass

The free electron kinentic energy depends on the electron mass,

$$\mathcal{E} = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m}$$

The kinetic energy is parabolic in the momentum, with the curvature of the function determined by m^{-1}, so

$$\frac{\partial^2 \mathcal{E}}{\partial k^2} = \frac{\hbar^2}{m},$$

or, observing (Fig. 45) that the second derivative may not be the same for all vales of k,

$$m_k^* = \frac{\hbar^2}{\frac{\partial^2 \mathcal{E}(k)}{\partial k^2}}.$$

The free-electron parabola is concave upward, and he mass is everywhere equal to the free-electron mass. Looking at Fig. 45 (or Figs. 44 or 46), I see that the curvature at the bottom of the lowest band is nearly that of the free-electron band, but that near either $k = \pm\pi/a$ or $k = 0$ the curvature is considerably larger, making m_k^* smaller than the free-electron mass. Moreover, the top half of each band is *concave downward*, changing the sign of the second derivative, so that the effective mass of these electrons is negative!

Technologically important semiconductors (Si, Ge, GaAs, InSb, ...) all have four electrons per atom (or three on one and five on the other) so the first two bands are full and the third empty. The second band is called the valence band and the third the conduction band. Now, as mentioned on page 124,[*] the gap between the top of the valence band and the bottom of the conduction band is small enough that at finite temperatures there are thermally-excited carriers in the conduction band.[†] Because the total number of electrons is fixed, empty states are left behind when thermal energies promote carriers to the conduction band. These vacant states are called *holes*. The absence of a negatively charged electron is a positive hole; the electron mass in the top half of the band is negative so the hole mass is positive. Hole energy is measured downwards from the top of the band.

---

[*] See also the footnote on p. 124.

[†] In *doped* semiconductors, the majority carriers come from impurity levels located in the forbidden band gap, but here I consider intrinsic (undoped) semiconductors and insulators.

Table 10. Semiconductor materials

Group	Material	Formula	Gap eV	Gap cm^{-1}	type
IV	Diamond	C	5.47	44,100	indirect
IV	Silicon	Si	1.11	8,900	indirect
IV	Germanium	Ge	0.67	5,400	indirect
IV	Gray tin, α-Sn	Sn	0.08	600	indirect
IV	Silicon carbide, 3C-SiC	SiC	2.3	18,500	indirect
IV	Silicon carbide, 4H-SiC	SiC	3.3	26,600	indirect
IV	Silicon carbide, 6H-SiC	SiC	3	24,100	indirect
VI	Sulfur	S_8	2.6	20,900	
VI	Gray selenium	Se	1.74	14,000	
VI	Tellurium	Te	0.33	2,600	
III-V	Boron nitride, cubic	BN	6.36	51,200	indirect
III-V	Boron nitride, hexagonal	BN	5.96	48,000	quasi-direct
III-V	Boron nitride, nanotube	BN	$\sim$5.5	$\sim$44300	
III-V	Boron phosphide	BP	2	16,100	indirect
III-V	Boron arsenide	BAs	1.5	12,000	indirect
III-V	Boron arsenide	$B_{12}As_2$	3.47	27,900	indirect
III-V	Aluminium nitride	AlN	6.28	50,600	direct
III-V	Aluminium phosphide	AlP	2.45	19,700	indirect
III-V	Aluminium arsenide	AlAs	2.16	17,400	indirect
III-V	Aluminium antimonide	AlSb	1.6	12,900	indirect
III-V	Aluminium antimonide	AlSb	2.2	17,700	direct
III-V	Gallium nitride	GaN	3.44	27,700	direct
III-V	Gallium phosphide	GaP	2.26	18,200	indirect
III-V	Gallium arsenide	GaAs	1.43	11,500	direct
III-V	Gallium antimonide	GaSb	0.726	5,800	direct
III-V	Indium nitride	InN	0.7	5,600	direct
III-V	Indium phosphide	InP	1.35	10,800	direct
III-V	Indium arsenide	InAs	0.36	2,900	direct
III-V	Indium antimonide	InSb	0.17	1,300	direct
II-VI	Cadmium selenide	CdSe	1.74	14,000	direct
II-VI	Cadmium sulfide	CdS	2.42	19,500	direct
II-VI	Cadmium telluride	CdTe	1.49	12,000	
II-VI	Zinc oxide	ZnO	3.37	27,100	direct
II-VI	Zinc selenide	ZnSe	2.7	21,700	direct
II-VI	Zinc sulfide	ZnS	3.54	28.500	direct
II-VI	Zinc telluride	ZnTe	2.25	18,100	direct

Holes contribute to the total electrical and optical properties. I write

$$\mathbf{j} = -ne\mathbf{v}_e + pe\mathbf{v}_h$$

Group	Material	Formula	Gap eV	Gap cm^{-1}	type
I-VII	Cuprous chloride	CuCl			direct
I-VI	Copper sulfide	Cu_2S	1.2	9,600	direct
IV-VI	Lead selenide	PbSe	0.27	2,100	direct
IV-VI	Lead(II) sulfide	PbS	0.37	2,900	
IV-VI	Lead telluride	PbTe	0.32	2,500	
IV-VI	Tin sulfide	SnS	1	8,000	indirect
IV-VI	Tin sulfide	SnS_2	2.2	17,700	
II-V	Cadmium arsenide	Cd_3As_2	0.14	1,100	N-type
Oxide	Titanium dioxide, anatase	TiO_2	3.2	25,800	indirect
Oxide	Titanium dioxide, rutile	TiO_2	3.02	24,300	direct
Oxide	Titanium dioxide, brookite	TiO_2	2.96	23,800	
Oxide	Copper(I) oxide	Cu_2O	2.17	17,500	
Oxide	Copper(II) oxide	CuO	1.2	9,600	P-type
Oxide	Uranium dioxide	UO_2	1.3	10,400	High
Oxide	Tin dioxide	SnO_2	3.7	29,800	Oxygen-deficient
Oxide	Barium titanate	$BaTiO_3$	3	24,100	Ferroelectric,
Oxide	Strontium titanate	$SrTiO_3$	3.3	26,600	Ferroelectric,
Oxide	Lithium niobate	$LiNbO_3$	4	32,200	Ferroelectric,
Oxide	Lanthanum copper oxide	La_2CuO_4	2	16,100	superconductive
Layered	Gallium selenide	GaSe	2.1	16,900	indirect
other	Copper indium selenide,	$CuInSe_2$	1	8,000	direct
other	Silver sulfide	Ag_2S	0.9	7,200	
other	Iron disulfide	FeS_2	0.95	7,600	
other	Copper zinc tin sulphide	Cu_2ZnSnS_4	1.49	12,000	direct

where n is the electron number density, p is the hole number density, and $\mathbf{v}_e$ $\mathbf{v}_h$ are the electron (hole) drift velocity.

10.7 Direct interband transitions

In this section I will calculate the optical properties due to interband transitions. First, I'll do the simple case of a direct ($\Delta k = 0$) transition in a solid. To start, look at Fig. 45 and consider a solid with an average of 4 electrons per atom (such as GaAs or InSb) making the first two bands fully occupied and the third band empty. A sketch of the band structure of such a solid is shown in Fig. 47. The highest filled and the lowest empty states are both at the zone center. This energy difference will be the minimum photon energy for optical absorption. There is a much wider energy difference at X; the direct interband transitions span the minimum to maximum energies

Another case is shown in Fig. 48, representing the noble metals. The fully occupied d levels are occupied 2–4 eV below the Fermi level. Direct transitions shown as the vertical arrow, can occur from these d levels to the conduction band energies above the Fermi level. Intraband transitions also occur, leading to the Drude conductivity of the conduction

Table 11. Table of semiconductor alloy systems

Group	Material	Formula	Low gap eV	High gap eV
IV	Silicon-germanium	$Si_{1-x}Ge_x$	0.67	1.11
III-V	Aluminium gallium arsenide	$Al_xGa_{1-x}As$	1.42	2.16
III-V	Indium gallium arsenide	$In_xGa_{1-x}As$	0.36	1.43
III-V	Indium gallium phosphide	$In_xGa_{1-x}P$	1.35	2.26
III-V	Aluminium indium arsenide	$Al_xIn_{1-x}As$	0.36	2.16
III-V	Gallium arsenide phosphide	GaAsP	1.43	2.26
III-V	Gallium arsenide antimonide	GaAsSb	0.7	1.42
III-V	Aluminium gallium nitride	AlGaN	3.44	6.28
III-V	Aluminium gallium phosphide	AlGaP	2.26	2.45
III-V	Indium gallium nitride	InGaN	2	3.4
II-VI	Cadmium zinc telluride,	CdZnTe	1.4	2.2
II-VI	Mercury cadmium telluride	HgCdTe	0	1.5
II-VI	Mercury zinc telluride	HgZnTe	0	2.25
other	Copper indium gallium	$Cu(In,Ga)Se_2$	1	1.7

Table 12 Effective mass in selected semiconductors

Group	Material	Electron	Hole
IV	Si (4K)	1.06	0.59
	Si (300K)	1.09	1.15
	Ge	0.55	0.37
III-V	GaAs	0.067	0.45
	InSb	0.013	0.6
II-VI	ZnO	0.29	1.21
	ZnSe	0.17	1.44

electrons.

The selection rules for these transitions are discussed on p. 87. With these considerations, and dropping the tilde, the dielectric function is given in Eq. 142, with $\gamma_{n0} \to 0$,

$$\epsilon = 1 + \frac{8\pi e^2}{\hbar} \sum_n \frac{\omega_{n0}}{\omega_{n0}^2 - \omega^2} |\langle n|x|0\rangle|^2 \tag{202}$$

As discussed on p. 94, this dielectric function has an imaginary part that is a series of delta functions,

$$\epsilon_2 = \frac{4\pi^2 e^2}{\hbar\omega} \sum_n \delta(\omega_{no} - \omega)\, \omega_{n0} |\langle n|x|0\rangle|^2 \tag{203}$$

I now convert this into an equation that uses the momentum matrix element:

$$\langle n|x|0\rangle = \frac{1}{i\omega_{n0} m} \langle n|p_x|0\rangle$$

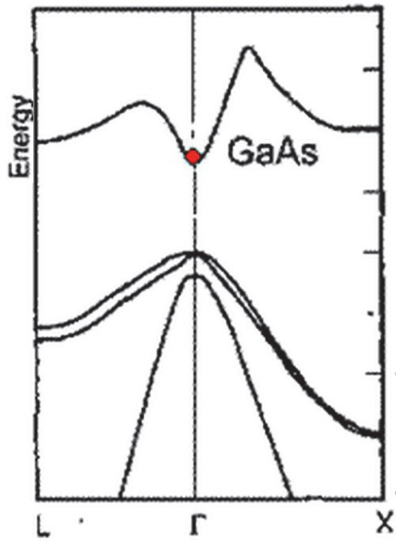

Fig. 47. Band structure, $\mathcal{E}(\mathbf{k})$, of GaAs for energies around the band gap. The horizontal axis is marked with special values of *veck* in two directions. Γ is the zone center, $\mathbf{k} = 0$ while X is the zone boundary in the k_x direction (or, on account of the cubic symmetry, the k_y or k_z direction). L is the zone boundary in the body diagonal of the cubic lattice. The three, concave down, lower bands are the (filled) valence bands and the upper band is the (empty) conduction band. The forbidden energy gap is the range from the top of the valence band (at Γ) to the bottom of the conduction band (also at Γ).

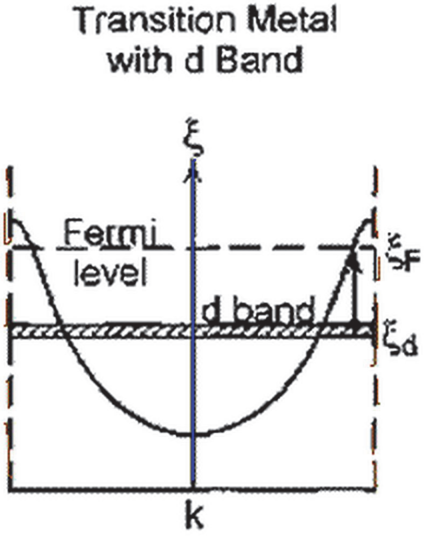

Fig. 48. Schematic band structure, $\mathcal{E}(\mathbf{k})$, of a noble metal, Ag, Cu, or Au. for energies in the conduction band. The horizontal axis is the wave vector. The Fermi energy is labeled. (Note that the Fermi surface in these metals is not a sphere, and the Fermi wave vector in some directions—the so-called "necks"—is at the zone boundary.) The d band is shown as the shaded area. Optical transitions from the d band to the conduction band occur for conduction band staes above $\mathcal{E}_F$.

so

$$\omega_{n0} |\langle n|x|0\rangle|^2 = \frac{1}{m^2} \frac{|\langle n|p_x|0\rangle|^2}{\omega_{n0}}$$

where I use p_x because the electric field is in the $\hat{\boldsymbol{x}}$ direction.

10.7.1 The matrix element

This matrix element is an integral,

$$\langle n|p_x|0\rangle = \int_V d^3r\, \psi_n^* p_x \psi_0 \tag{204}$$

I write these wave functions in Bloch form

$$\psi_{nk''} = \frac{1}{\sqrt{\Omega}}\, u_{nk''}(\mathbf{r})\, e^{i\mathbf{k}''\cdot\mathbf{r}}$$

and

$$\psi_{0k'} = \frac{1}{\sqrt{\Omega}}\, u_{0k'}(\mathbf{r})\, e^{i\mathbf{k}'\cdot\mathbf{r}}$$

with $u(\mathbf{r}) = u(\mathbf{r}+\mathbf{T})$ (periodic). Here, n is a band index, with 0 the index of the band that makes up the ground state, and $\mathbf{k}$ specifies the place in momentum space where the wave function exists. The prime and double prime have no deep meaning; they are different because the coordinate of ground and excited state could be different.[*]

Now I begin to compute the matrix element. To consider a specific model, I will do this for the semiconductor whose band structure is shown in Fig. 47. So I will replace the ground state index in Eq. 204 with v (for "valence" band) and the excited state index with c (for "conduction" band). Moreover, in the crystal I may arrive at a particular location $\mathbf{r}$ by first translating to a specific unit cell (i) by the translation vector of the lattice $\mathbf{T}_i$ and then by moving within that specific cell by a vector $\mathbf{r}_i$. This process requires

$$\mathbf{r} = \mathbf{T}_i + \mathbf{r}_i$$

and the integral in Eq. 204 can thus be written

$$\langle n|p_x|0\rangle = \sum_{\text{cells}} \int_{\text{cell}} d^3r_i\, \psi_c^* p_x \psi_v$$

Next, I install the explicit wave functions for conduction and valence bands:

$$\langle n|p_x|0\rangle = \frac{1}{\Omega} \sum_{\text{cells}} \int_{\text{cell}} d^3r_i\, u_{ck''}^*(\mathbf{T}_i+\mathbf{r}_i)\, e^{-i\mathbf{k}''\cdot(\mathbf{T}_i+\mathbf{r}_i)} p_x u_{vk'}(\mathbf{T}_i+\mathbf{r}_i)\, e^{i\mathbf{k}'\cdot(\mathbf{T}_i+\mathbf{r}_i)}$$

Now, I collect terms, bring those that are do not depend on $\mathbf{r}_i$ outside the integral, and make use of the periodicity of the Bloch states to find

$$\langle n|p_x|0\rangle = \frac{1}{\Omega} \sum_{\text{cells}} e^{i(\mathbf{k}'-\mathbf{k}'')\cdot\mathbf{T}_i} \int_{\text{cell}} d^3r_i\, e^{-ik''\cdot\mathbf{r}_i} u_{ck''}^*(\mathbf{r}_i) p_x u_{vk'}(\mathbf{r}_i)\, e^{i\mathbf{k}'\cdot\mathbf{r}_i}$$

Now I look at the pieces of this equation. First:

$$\sum_{\text{cells}} e^{i(\mathbf{k}'-\mathbf{k}'')\cdot\mathbf{T}_i} = \begin{cases} N_{\text{cells}} & \text{if } \mathbf{k}'-\mathbf{k}''=0 \\ 0 & \text{otherwise} \end{cases}$$

so I can replace $\mathbf{k}''$ with $\mathbf{k}'$ everywhere the former appears. Moreover $N_{\text{cells}}/\Omega = 1/V_{\text{cell}}$

---

[*] I allow for them to be different; the selection rule ($\mathbf{k}'' = \mathbf{k}'$) I have argued for up to now will come out of this calculation. The choice of prime and double prime rather than, for example, unprimed is designed to make the final answer come out with the notation I want.

Next, $p_x = -i\hbar\nabla_{ix}$ so

$$p_x u_{vk'}(\mathbf{r}_i)\, e^{i\mathbf{k}'\cdot\mathbf{r}_i} = -i\hbar\, e^{i\mathbf{k}'\cdot\mathbf{r}_i}[ik_x u_{vk'}(\mathbf{r}_i) + \nabla_x u_{vk'}(\mathbf{r}_i)]$$

Next,

$$e^{-ik''\cdot\mathbf{r}_i} e^{i\mathbf{k}'\cdot\mathbf{r}_i} = 1$$

because $\mathbf{k}' = \mathbf{k}''$.

Finally, because of orthogonality ($\langle c|v\rangle = 0$)

$$\int_{\text{cell}} d^3r_i\, u^*_{ck'}(\mathbf{r}_i)\, k_x\, u_{vk'}(\mathbf{r}_i) = 0.$$

So, finally, the matrix element may be written

$$\langle n|p_x|0\rangle = \frac{1}{V_{\text{cell}}} \int_{\text{cell}} d^3r_i\, u^*_{ck'}(\mathbf{r}_i)\, p_x\, u_{vk'}(\mathbf{r}_i) \equiv p_{vc}(\mathbf{k}')$$

10.7.2 The imaginary part of the dielectric function

Now I return to Eq. 203 with only small changes in notation, including

$$\omega_{n0} \to \omega_k = \frac{\mathcal{E}_c(\mathbf{k}) - \mathcal{E}_v(\mathbf{k})}{\hbar}$$

and

$$\sum_n \to 2\int \frac{d^3k}{(2\pi)^3}$$

The imaginary part of the dielectric function becomes

$$\epsilon_2 = \frac{e^2}{\pi m^2 \hbar \omega^2} \int d^3k\ |p_{vc}(\mathbf{k})|^2\, \delta(\omega_k - \omega) \tag{205}$$

10.7.3 The absorption edge

I will make a crude estimate of the absorption edge near the minimum energy. To do so, I will need to have the band structure (to know how ω_k varies with $\mathbf{k}$) and the matrix element p_{vc} (to know how it varies with $\mathbf{k}$).

For the first, I consider parabolic bands separated by $\mathcal{E}_g = \hbar\omega_g$, the minimum forbidden energy gap. See Fig. 49. The conduction band energy is

$$\mathcal{E}_c = \hbar\omega_g + \frac{\hbar^2 k^2}{2m_c}$$

where m_c is the effective mass in the conduction band. The minimum energy is at $k = 0$, where $\mathcal{E}_c = \hbar\omega_g$ The valence band energy is

$$\mathcal{E}_v = -\frac{\hbar^2 k^2}{2|m_v|}$$

where m_v is the effective mass in the valence band and I take the absolute value because the effective mass is in fact negative ($\mathcal{E}_v$ is concave downwards) and I have put the minus sign in explicitly. The zero of energy is at $k = 0$.

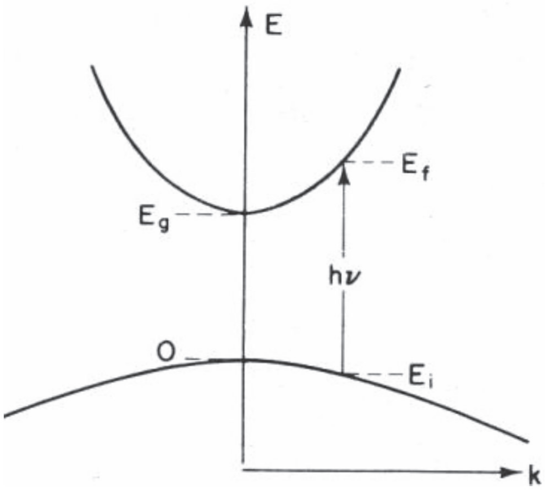

Fig. 49. Energy band structure, $\mathcal{E}(\mathbf{k})$. An optical transitions from the valence band to the conduction band is shown.

For the second, I take p_{vc} as independent of $\mathbf{k}$. It then may be taken outside the integral in Eq. 205, which becomes

$$\epsilon_2 = \frac{e^2}{\pi m^2 \hbar \omega^2} |p_{vc}|^2 \int d^3k \;\; \delta(\omega_k - \omega)$$

Now, the energy difference between valence and conduction bands depends on k according to:

$$\hbar\omega_k = \hbar\omega_g + \left(\frac{\hbar^2 k^2}{2m_c} + \frac{\hbar^2 k^2}{2|m_v|} \right).$$

Then if I define an average effective, or reduced, mass as

$$\frac{1}{m_r} = \frac{1}{m_c} + \frac{1}{m_v}$$

this becomes

$$\omega_k = \omega_g + \frac{\hbar k^2}{2m_r}.$$

Turning this around, I find

$$k = \frac{1}{2}\sqrt{\frac{2m_r}{\hbar}}\sqrt{\omega_k - \omega_g}$$

Taking the differential of this I get

$$dk = \sqrt{\frac{2m_r}{\hbar}} \frac{d\omega_k}{\sqrt{\omega_k - \omega_g}}.$$

The integral is a three dimensional integral in $d^3k = k^2\, dk\, d\Omega$ where $d\Omega$ is the angular integral over the 4π solid angle. This is easy to do as nothing depends on the angular co-ordinates Thus

$$d^3k = 4\pi k^2 dk = 2\pi \left(\frac{2m_r}{\hbar}\right)^{3/2} \sqrt{\omega_k - \omega_g}\, d\omega_k$$

The lower limit of the itegral at $k = 0$ is equivalent to $\omega_k = \omega_g$. Thus, the imaginary part of the dielectric function becomes

$$\epsilon_2 = \frac{e^2}{\pi m^2 \hbar \omega^2} |p_{vc}|^2 2\pi \left(\frac{2m_r}{\hbar}\right)^{3/2} \int_{\omega_g}^{\infty} d\omega_k \ \sqrt{\omega_k - \omega_g}\, \delta(\omega_k - \omega)$$

The integral is easy to do (thanks to the delta function). Doing it, and collecting terms, I get*

$$\epsilon_2 = \begin{cases} 0 & \omega \leq \omega_g \\ \frac{2e^2}{m^2\hbar} \left(\frac{2m_r}{\hbar}\right)^{3/2} |p_{vc}|^2 \frac{\sqrt{\omega - \omega_g}}{\omega^2} & \omega \geq \omega_g \end{cases} \tag{206}$$

I can simplify this a bit by using the oscillator strength f, where

$$f = \frac{2\langle |p_{vc}|^2 \rangle}{m\hbar\omega} = \frac{2m\omega \langle |x_{vc}|^2 \rangle}{\hbar}$$

and finally I write $\sigma_1 = \omega\epsilon_2/4\pi$. $\sigma_1 = 0$ for $\omega \leq \omega_g$ and

$$\sigma_1 = \frac{e^2}{2\pi m} \left(\frac{2m_r}{\hbar}\right)^{3/2} f \sqrt{\omega - \omega_g} \tag{207}$$

for $\omega \geq \omega_g$. See Fig. 50 for a cartoon of the absorption edge.

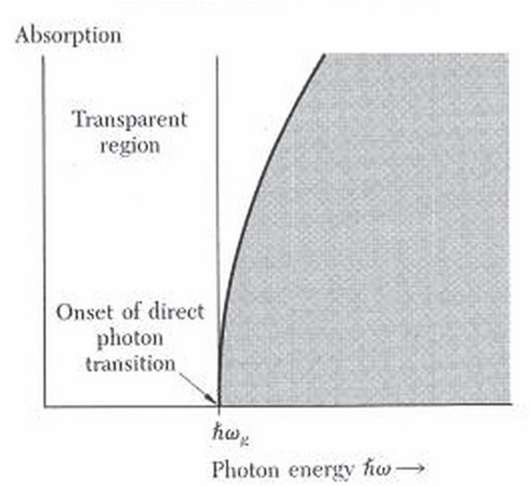

Fig. 50. Absorption edge of a direct band gap semiconductor, showing the $\sqrt{\omega - \omega_g}$ edge.

10.8 The joint density of states and critical points

---

* You often see a similar equation for α. It needs the refractive index also, because $\alpha = \omega\epsilon_2/nc$.

11. SUPERCONDUCTORS

11.1 Superconducting phenomena

All discussions of superconductivity are supposed to start in 1911, in Leiden, when and where Kamerlingh Onnes[68] and his assistants discovered that the resistance of solid mercury abruptly falls to zero at a temperature of 4.2 K. The data are shown in Fig. 51. The resistances below 21 K of Hg, Pt, and two samples of Au are shown versus the temperature. Pt and Au have resistances which level out at the residual value, similar to the data in Fig. 37. In contrast, the resistance of Hg takes a sudden drop to zero at 4.2 K.*

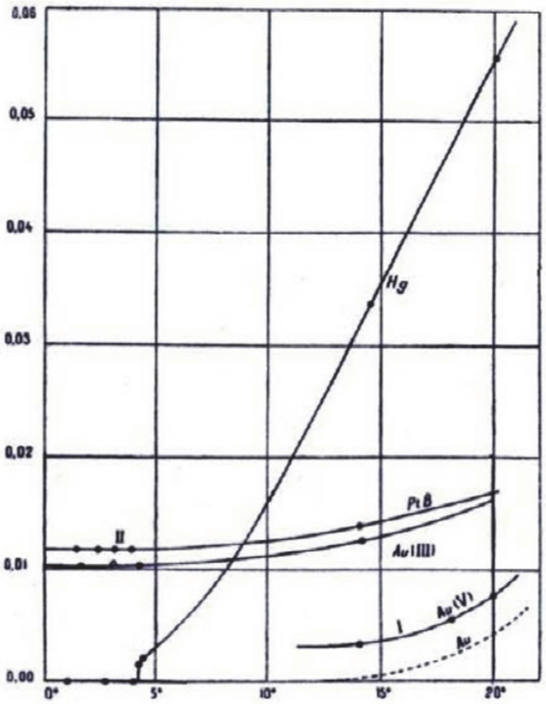

Fig. 51. Resistance (in Ohm) *vs.* temperature (in K) for Hg, Pt, and two samples of Au. The temperature range is 1 to 21 K. the hypothetical resistance of perfectly clean Au is indicated as well.

Onnes discovered other superconductors, and in the ensuing 100 or so years, superconductivity has been found to be widely distributed, occurring in elements,[68] alloys,[69] intermetallic compounds,[70] oxides,[71] ionic compounds,[72] organic charge-transfer compounds,[73] polymers,[74] and many other systems.[75–84] It seems that the only two prerequisites are (1) that the material be conductive, so that there are carriers to condense into the charged superfluid, and (2) that it not be a magnet, as strong magnetic fields destroy superconductivity.†

---

[*] Onnes reported that the resistance was smaller than $10^{-5}\ \Omega$ compared to $0.11\ \Omega$ just above T_c; it is now believed to be zero. Experiments have been done on rings of superconducting material which sustain continuous current for years with no applied voltage.

[†] Although the conductivity does not need to be that of free-electron metals and there are counterexamples to prerequisite 2.

Figure 52 shows the “records” for superconductivity over a 99 year period, 1911–2010. Several classes of materials are shown: metals, alloys and intermetallics (circles), the high-T_c cuprates (diamonds), fullerenes and intercalated graphite (triangles), and heavy-Fermion metals (squares). A list of superconductors is given in Table 13. Transition temperatures can be seen to vary from 15 mK (α-W) to to 164 K (Hg-1223 at 300 kbar).*

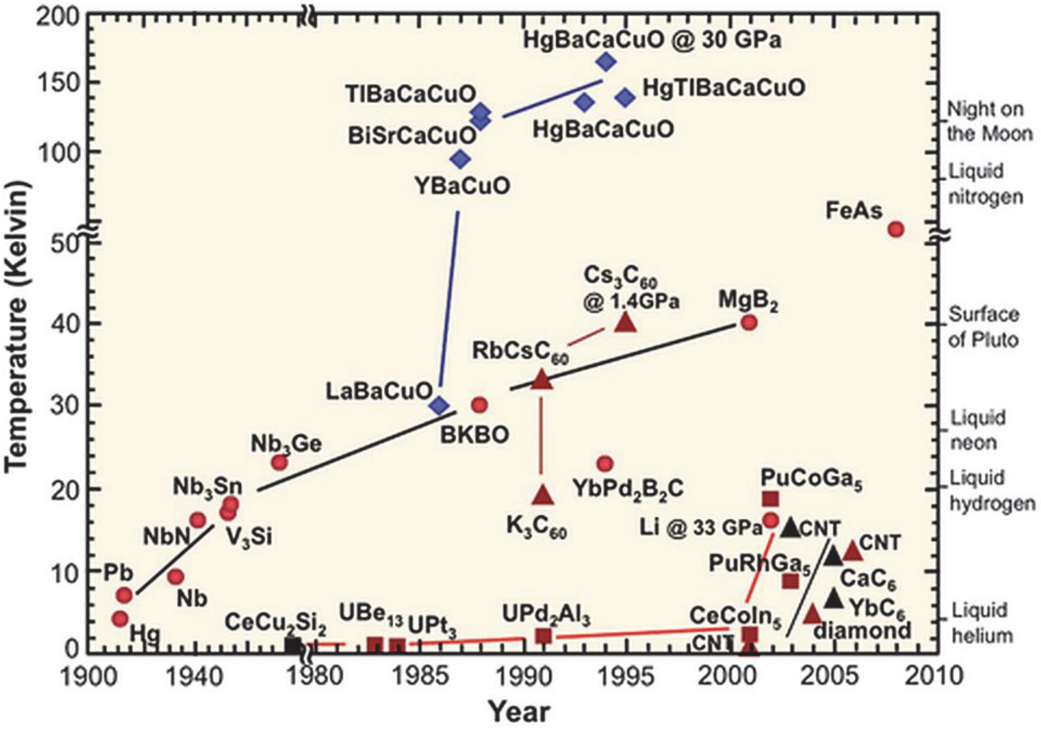

Fig. 52. Transition temperatures over a 99-year period.

The second key property of superconductors was discovered in 1929, when Meissner[85,86] observed that a superconductor expels magnetic flux, making the condition for the field inside the superconductor be that $B = 0$, where B is the magnetic field. The flux expulsion is sketched in Fig. 53. Then, from our constitutive equations, Eqs. 3 and 14

$$0 = B = \mu H = (1 + 4\pi\chi)H$$

H is not zero; hence,

$$\chi = -\frac{1}{4\pi} \tag{208}$$

In the Meissner state, the superconductor is a perfect diamagnet. The magnetic moment of the sphere is generated by currents which circulate on the surface of the sphere parallel to the equator.† The flux exclusion continues as field increase until the critical field H_c is reached and superconductivity is destroyed.

---

[*] There are T_c’s below 15 mK but none—so far—above 164 K.

[†] Actually the currents decay exponentially into the superconductor with a characteristic penetration depth. I will calculate this depth shortly.

Table 13. A list of superconducting materials. The materials are organized into elements, compounds, organics, cuprates, and pnictides.

Formula	T_c (K)	H_c (T)	Type
	Metals		
Al	1.20	0.01	I
α-Hg	4.15	0.04	
In	3.4	0.03	I
Nb	9.26	0.82	II
Pb	7.19	0.08	I
Sn	3.72	0.03	I
Ta	4.48	0.09	I
α-U	0.68		I
V	5.03	1	II
α-W	0.015	0.00012	I
Zn	0.855	0.005	I
	Compounds		
C_6K	1.5		II
$C_{60}K_3$	19.8	0.013	II
$C_{60}Rb_x$	28		II
In_2O_3	3.3		II
MgB_2	39	74	II
Nb_3Sn	18.3	24.5	II
Nb_3Ge	23.2	37	II
V_3Si	17.1	23	II
Binary alloys			
NbO	1.38		II
NbN	16	15.3	II
NbTi	10	15	II
TiN	5.6		
ZrN	10		
	Charge-transfer salts		
$(TMTSF)_2PF_6$	1.1 (at 6.5 kbar)		II
$(TMTSF)_2ReO_4$	1.2 (at 9.5 kbar)		II
$(TMTTF)_2ClO_4$	1.4		II
β-H-$(ET)_2I_3$	1.5		II
β'-$(ET)_2SF_5CH_2CF_2SO_3$	5.3		II
κ-$(ET)_2Cu(NCS)2$	10.4	10	II

There actually are two types of superconductors. Type I superconductors function as described, with perfect diamagnetic response, Eq. 208, up to the critical field. Type II superconductors are perfect diamagnets up to a (typically relatively low) lower critical field given by H_{c1}. Above this field, partial flux penetration begins, and grows until it is complete

Formula	T_c (K)	H_c (T)	Type
Cuprates (high-T_c)			
$La_{1.85}Sr_{0.15}CuO_4$	38	45	II
$Nd_{1.7}Ce_{0.3}CuO_4$	24		II
$YBa_2Cu_3O_7$ (123)	92	140	II
$Bi_2Sr_2CuO_6$ (Bi-2201)	20		II
$Bi_2Sr_2CaCu_2O_8$ (Bi-2212)	85	107	II
$Tl_2Ba_2CaCu_2O_8$ (Tl-2212)	108		II
$Tl_2Ba_2Ca_2Cu3O_{10}$ (Tl-2223)	125	75	II
$HgBa_2CuO_4$ (Hg-1201) 94		II	
$HgBa_2Ca_2Cu_3O_8$ (Hg-1223)	134	190	II
$HgBa_2Ca_2Cu_3O_8$ (Hg-1223)	164 (at 300 kbar)		II
Iron-based (pnictide)			
$LaO_{0.9}F_{0.2}FeAs$	28.5		II
$PrFeAsO_{0.89}F_{0.11}$	52		II
$GdFeAsO_{0.85}$	53.5		II
$BaFe_{1.8}Co_{0.2}As_2$	25.3		II
$SmFeAsO_{0.85}$	55		II
$Ba_{0.6}K_{0.4}Fe_2As_2$	38		II
$CaFe_{0.9}Co_{0.1}AsF$	22		II
$Sr_{0.5}Sm_{0.5}FeAsF$	56		II
FeSe	27		II

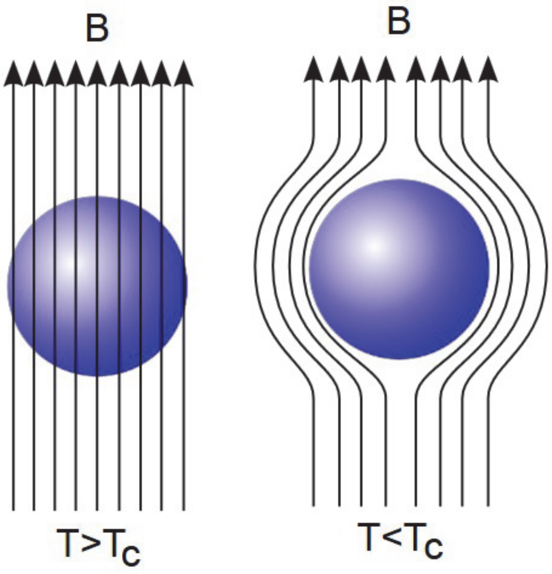

Fig. 53. A superconducting sphere in a uniform external magnetic field **B**. Above T_c, when the sphere is a normal metal, the field penetrates uniformly. Below T_c, the flux is excluded, making $\mathbf{B} = 0$ inside. At the same time, the sphere acquires a uniform magnetization.

at a (typically relatively high) upper critical field, H_{c2}. Superconductivity (zero resistance) persists up to H_{c2}. The M-H behavior of both types are sketched in Fig. 54 while Table 13 identifies whether the material is Type I or Type II. The materials with high critical fields are all Type II and some of these (NbTi, Nb_3Sn, $YBa_2Cu_3O_7$, etc.) are used in superconducting magnets.

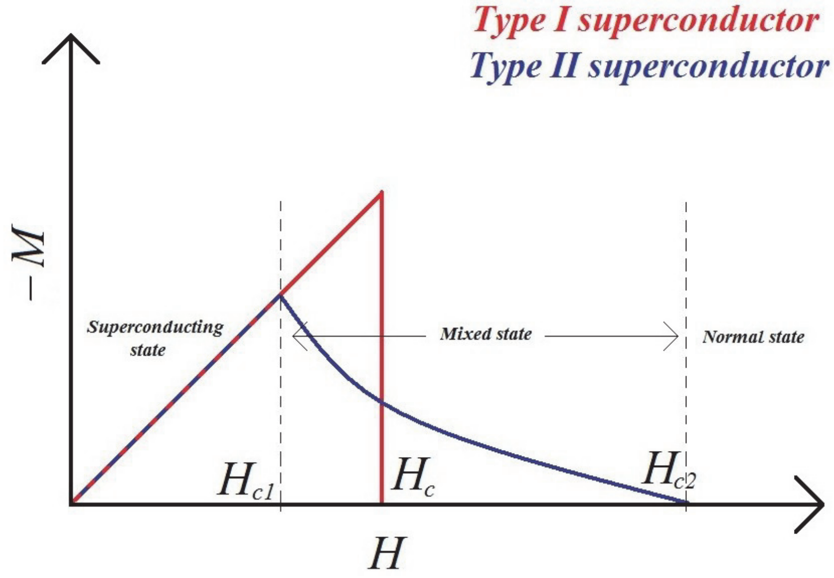

Fig. 54. Magnetization versus applied field for Type I and Type II superconductors.

Superconductivity has two big effects on optical properties. First, the infinite dc conductivity (a delta function at $\omega = 0$) gives a large inductive response to the imaginary part of the conductivity, just as in Eq. 154. Second there is a gap (called 2Δ for reasons that will become apparent) for excitations, which makes the real part of the conductivity zero for frequencies below 2Δ.*

11.2 Theoretical background

These notes are not a course in superconductivity, so I will not derive BCS theory[87] nor the phenomenological Ginzburg-Landau theory.[88] Here I will just list some of the key results.

BCS theory[87,89] is the microscopic theory of metallic superconductors. These superconductors have the following properties

- Below T_c, electrons form Cooper pairs. The Cooper state is a coherent combination of two electron states of the normal metal into a zero-momentum, singlet state $|\mathbf{k}\uparrow, -\mathbf{k}\downarrow\rangle$.
- The superconducting condensate is a coherent superposition of Cooper pairs, involving all the electrons in the metal.
- Because the pairs are made up of two electrons, the number density of Cooper pairs is $n_c = n/2$, the charge is $Q_c = 2e$, and the mass is $m_c = 2m$.
- The Cooper state exists because of an indirect attractive interaction between the two electrons.

---

* except for $\omega = 0$ where it is infinite.

- In metallic superconductors, this attraction arises through the interaction with the lattice (the phonons).
- Turning on this interaction reduces the energy in the system. In fact, the Fermi surface is unstable towards the formation of a Cooper pair as long as there is an infinitesimal attractive interaction between the electrons.
- The binding energy of Cooper pairs at the Fermi surface is

$$2\Delta = 3.5k_BT_c.$$

- The gap is small compared to the Fermi energy. The gap scale is a few meV; the Fermi energy several eV.
- Finite temperatures break Cooper pairs and, because the gap is a collective effect, reduce the gap. It eventually reaches zero at T_c. The behavior (from both experiment and theory) is shown in Fig. 55
- The excitations above the gap are called *quasiparticles*. They have many of the properties of normal electrons, being spin-half Fermions, but are affected by the BCS coherence factors and have an energy-wave vector dispersion relation that differs significantly from that of ordinary electrons in the solid.

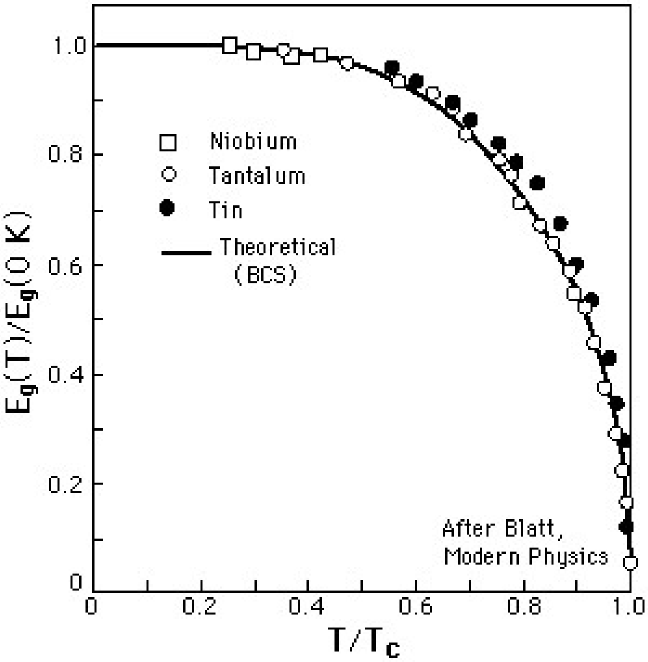

Fig. 55. Superconducting gap as a function of temperature.

Ginzburg-Landau theory[88,90,91] is a phenomenological theory that is based on the Landau theory of second-order phase transitions. The free energy of a superconductor near T_c is written in terms of a complex order parameter, ψ, which is nonzero below T_c and zero above. The theory (when folded into BCS theory[92]) makes the following identifications,

$$\psi = \sqrt{n_c(T)}e^{i\phi} = \Delta(T), \tag{209}$$

where ψ represents a coherent superposition of all Cooper pairs, the pair density is n_c, ϕ is the phase of the order parameter, and Δ is the BCS energy gap. The two-fluid model, in which there is a normal fluid (described perhaps by the Drude model) intermixed with the superfluid, is consistent with the Ginzburg-Landau theory. As n_c is reduced approaching T_c, (Fig. 55) n_n grows so that $n = 2n_c + n_n$.

11.3 The London model

11.3.1 Heuristic derivation

Dating from 1935, the London model[93] is a mix of electrodynamics and quantum mechanics. Often discussions just start by writing the London equation, but one can derive it in a relatively simple way. I'll start by reminding myself of the definition of the electrical current carried by a set of Cooper pairs. Classically is is

$$\mathbf{j} = n_c Q_c \mathbf{v} = \frac{n_c Q_c}{m_c}\mathbf{p}$$

where $\mathbf{v}$ ($\mathbf{p}$) is the average velocity (momentum) of the superconducting particles (Cooper pairs), n_c is their number density, Q_c is their charge, and m_c is their mass. Quantum mechanically, I would write something like

$$\mathbf{j}_s = \frac{Q_c}{2m_c}(\psi^* \mathbf{p}\,\psi - \psi \mathbf{p}\,\psi^*) \tag{210}$$

with $\mathbf{p}$ the momentum operator and ψ the wave function of the superconducting state, given by Eq. 209. In the presence of an electromagnetic field,

$$\mathbf{p} = -i\hbar\nabla - \frac{Q_c}{c}\mathbf{A},$$

where $\mathbf{A}$ is the vector potential. In a homogeneous superconductor, charge neutrality causes n_c to be constant. Hence,

$$-i\hbar\nabla\psi = -i\hbar\sqrt{n_c}e^{i\phi}i\nabla\phi$$

With this, Eq. 210 becomes

$$\mathbf{j}_s = \frac{n_c Q_c}{m_c}(\hbar\nabla\phi - \frac{Q_c}{c}\mathbf{A})$$

Now, for Cooper pairs,

$$Q_c = -2e \qquad m_c = 2m \qquad \text{and} \qquad n_c = \frac{n_s}{2}$$

with n_s the superfluid density in a two-fluid picture. At zero temperature in clean metals, $n_s = n$. Thus $n_c Q_c/m_c = -n_s e/2m$ and $n_c Q_c^2/m_c = n_s e^2/m$. With these replacements I arrive at

$$\mathbf{j}_s = -\frac{n_s e}{2m}\hbar\nabla\phi - \frac{n_s e^2}{mc}\mathbf{A}. \tag{211}$$

The two terms in Eq. 211 lead to different physics. The first term, proportional to the gradient of the phase, turns into the Josephson effect, pair tunneling between two superconductors. The second term controls the electromagnetic properties of homogeneous superconductors, where the phase gradient is zero. Taking constant phase to be the case for optical properties, I have

$$\mathbf{j}_s = -\frac{n_s e^2}{mc}\mathbf{A}. \tag{212}$$

Eq. 212 is known as the London equation.

11.3.2 Optical conductivity

Now, take the electromagnetic field to be $\mathbf{E} = \mathbf{E}_0 e^{-i\omega t}$, an ac field (or plane wave) of frequency ω. Recall that

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} = \mathbf{E}_0 e^{-i\omega t}$$

meaning that

$$\mathbf{A} = -\frac{ic}{w}\mathbf{E}$$

I put this vector potential into Eq. 212 to find

$$\mathbf{j}_s = i\frac{n_s e^2}{m\omega}\mathbf{E}. \tag{213}$$

As usual, Ohm's law is $\mathbf{j}_s = \sigma\mathbf{E}$. Therefore, the optical conductivity is

$$\sigma_s = i\frac{n_s e^2}{m\omega}. \tag{214}$$

I have seen this before! Look at Eqs. 129 and 152. It is the finite-frequency conductivity of the perfect metal: the Drude metal in the limit where $\tau \to \infty$. Moreover, I can define a superconducting plasma frequency $\omega_{ps} = \sqrt{4\pi n_s e^2/m}$, write $\sigma_1 = i\omega_{ps}^2/4\pi\omega$, and then (using $\epsilon_s = \epsilon_c + 4\pi i\sigma/w$) get the dielectric function:

$$\epsilon_s = \epsilon_c - \frac{\omega_{ps}^2}{\omega^2}, \tag{215}$$

Eq. 215 is the dielectric response of perfect free carriers with no scattering.

At this point I can list 4 questions that I must answer. They are

1. How do I get the infinite dc conductivity of the superconductor?
2. How does the superconductivity affect electromagnetic wave propagation and reflectance?
3. What about the Meissner effect?
4. This is cute, but why do I not take a bunch of Drude-like free carriers and just let their mean free time between collisions go to infinity?

11.3.3 Infinite dc conductivity

I already know the answer to the first question; it was discussed on page 101. The real part of the conductivity is a delta function at $\omega = 0$. I'll write it as $\sigma_1 = A\delta(\omega)$ and substitute that into the Kramers-Kronig relation, Eq. 171, so that

$$\epsilon_1(\omega) = 1 + 8\mathcal{P}\int_0^\infty d\omega' \, \frac{A\delta(\omega')}{\omega'^2 - \omega^2}. \tag{216}$$

The integral is easy to do,[*] and I get

$$\epsilon_1(\omega) = 1 - \frac{4A}{\omega^2},$$

---

[*] Because the lower limit of the integral is $\omega = 0$, which is where the delta function is located, the integral gives $1/2\omega^2$ not $1/\omega^2$.

or, equivalently, the imaginary part of the conductivity is

$$\sigma_2(\omega) = \frac{A}{\pi\omega}.$$

I compare this equation with Eq. 215; and see that $A = \omega_{ps}^2/4 = \pi n_s e^2/m$. This value for the weight of the delta function satisfies the sum rule, Eq. 184

$$\int_0^\infty d\omega'\,\sigma_1(\omega') = \frac{\omega_{ps}^2}{8} = \frac{\pi n_s e^2}{2m} \tag{217}$$

if all the electrons are part of the superconducting condensate.

11.3.4 Electromagnetic wave propagation and reflectance

The London dielectric function is purely real and negative, The refractive index in the London model is

$$N = n + i\kappa = \sqrt{\epsilon} = 0 + i\sqrt{\frac{\omega_{ps}^2}{\omega^2} - 1} \approx 0 + i\frac{\omega_{ps}}{\omega}$$

So $n = 0$ and $\kappa \approx \omega_{ps}/\omega$. (The approximation is good. The range of frequencies where the London model describes the superconducting gap ends at the gap frequency, say 5 meV or 40 cm^{-1} whereas the plasma frequency of a typical metal is, say, 5 eV or 40,000 cm^{-1}.)

The electric field in the superconductor decays as $\mathbf{E} = \mathbf{E}_0 e^{-\kappa\omega x/c} = \mathbf{E}_0 e^{-x/\lambda_L}$ with distance x away from the surface. The decay length is

$$\lambda_L = \frac{c}{\omega_{ps}} = \sqrt{\frac{mc^2}{4\pi n_s e^2}}, \tag{218}$$

where λ_L is known as the London penetration depth.*

Finally, the reflectance, given by Eq. 44, is unity. The London superconductor is a perfect reflector.

11.3.5 Meissner effect

Next, to work on the Meissner effect, I will consider a superconductor with a static magnetic field $\mathbf{B}$ applied parallel to the surface. I will use London gauge:

$$\nabla \cdot \mathbf{A} = 0$$

because, Eq. 212 implies

$$\nabla \cdot \mathbf{A} \propto \nabla \cdot \mathbf{j} = -\frac{\partial\rho}{\partial t}$$

where I've used the continuity equation. If I want to keep the static case, I want no time variation of the charge density. The London gauge enforces this requirement.

---

* The units dilemma rears its head here. ω_{ps} is an angular frequency in radians/sec or s^{-1}. But, like other plasma frequencies, it is often reported in cm^{-1}, making the London length appear to be just the vacuum wavelength (λ_0) corresponding to this plasma frequency. But if I go through the conversions correctly (Appendix A) I find that $\lambda_L = \lambda_0/2\pi$.

I take the curl of Eq. 212,

$$\nabla \times \mathbf{j}_s = -\frac{n_s e^2}{mc} \nabla \times \mathbf{A} = -\frac{n_s e^2}{mc} \mathbf{B}, \tag{219}$$

and take the curl again,

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{j}_s) =& -\frac{n_s e^2}{mc} \nabla \times \mathbf{B} \\ =& -\frac{n_s e^2}{mc} \frac{4\pi}{c} \mathbf{j}_s \\ =& \nabla(\nabla \cdot \mathbf{j}_s) - \nabla^2 \mathbf{j}_s \\ =& -\nabla^2 \mathbf{j}_s, \end{aligned}$$

where I used Maxwell's Eq. 4d to go from the first to the second line, the properties of the double cross product to get the third line, and the gauge condition to go from the third to the fourth line. Thus,

$$\nabla^2 \mathbf{j}_s = \frac{4\pi n_s e^2}{mc^2} \mathbf{j}_s = \frac{1}{\lambda_L^2} \mathbf{j}_s,$$

using Eq. 218. The physically reasonable solution is that $j_s = 0$ deep in the interior of the superconductor and that there are surface currents which decay away from the surface as

$$\mathbf{j}_s(x) = \mathbf{j}_0 e^{-x/\lambda_L}.$$

If I start with Eq. 4d and take the curl of both sides:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{B}) =& \frac{4\pi}{c} \nabla \times \mathbf{j}_s \\ =& -\frac{4\pi}{c} \frac{n_s e^2}{mc} \mathbf{B} \\ =& \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B} \\ =& -\nabla^2 \mathbf{B}, \end{aligned}$$

where I used Eq. 219 to go from the first to the second line, the properties of the double cross product to get the third line, and Eq. 4b to go from the third to the fourth line. Thus,

$$\nabla^2 \mathbf{B} = \frac{4\pi n_s e^2}{mc^2} \mathbf{B} = \frac{1}{\lambda_L^2} \mathbf{B}, \tag{220}$$

using Eq. 218. The physically reasonable solution is that $\mathbf{B} = 0$ deep in the interior of the superconductor and that that it decays away from the surface as

$$\mathbf{B}(x) = \mathbf{B}_0 e^{-x/\lambda_L},$$

as sketched in Fig. 56.

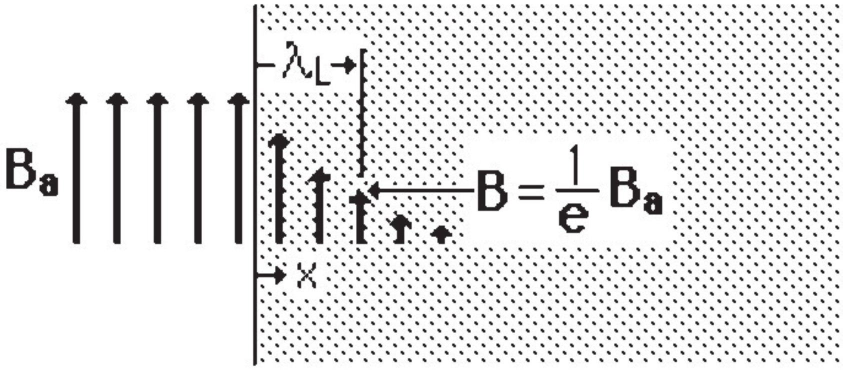

Fig. 56. The magnetic field falls from the external value to $1/e$ at a distance of λ_l below the surface of the superconductor.

Table 14. London penetration depth of superconducting metals.

Material	λ_L Å
Al	160
Nb	390
Pb	370
Sn	340
$YBa_2Cu_3O_7$	1500

So both the field and the current decay in the same way as a function of depth into the superconductor. The current can be thought of as the current required to generate a field of $-\mathbf{B}$ in the interior, so that the sum of the applied field and the generated field is zero.

11.3.6 The penetration depth

The penetration depth, $mc/4\pi n_s e^2$, it typically in the 100–2000 Å range. Some values are shown in Table 14. Metals with high electron density (Al) have shorter penetration depths than those with lower density.

The temperature dependence of the penetration depth goes as $\lambda_L(T) = 1/\sqrt{n_s(T)}$. As temperature increases towards T_c, the penetration depth increases. If I use the Ginsberg-Landau temperature dependence (and the identification of $\sqrt{n_s} = \Delta$, I get $\lambda_L(T) = \sqrt{T_c/(T_c - T}$.

The penetration depth is also the “skin depth” of the superconductor. Light* It is interesting to compare the normal-state penetration depth and the London length. The latter is

$$\lambda_L = \frac{c}{\omega_{ps}} = \sqrt{\frac{mc^2}{4\pi n_s e^2}}$$

---

* The light must be in the far-infrared range, with photon energy smaller than the energy gap, for the superconductor to respond as a superconductor.

whereas (for a Drude metal with $\sigma_{\rm dc} = ne^2\tau/m$) the penetration depth is

$$\delta = \begin{cases} \sqrt{\dfrac{mc^2}{2\pi ne^2\omega\tau}} & \omega \ll 1/\tau \\ \sqrt{\dfrac{mc^2}{2\pi ne^2}} & \omega \gg 1/\tau \end{cases}$$

As long as $n_s = n$, (which is the case for pure metals at low temperatures) the London length is the same as the high-frequency limit of the skin depth in the metal. The low-frequency skin depth of the metal is much longer that the London length, and grows as the frequency decreases.

11.3.7 The perfect conductor

The Drude conductivity is

$$\sigma_1(\omega) = \frac{ne^2\tau/m}{1+\omega^2\tau}$$

If $\tau \to \infty$, the conductivity becomes a delta function at the origin, infinite at dc and zero at all finite frequencies.*

Let me start with Maxwell's Eq. 4d, take the curl, and then the time derivative

$$\frac{d}{dt}\nabla \times (\nabla \times \mathbf{B}) = \frac{4\pi}{c}\nabla \times \frac{d\mathbf{j}}{dt}$$

where I have put the $\partial\mathbf{D}/\partial t$ term into the current. The current can be written $\mathbf{j} = -ne\mathbf{v}$. Writing out the double curl (as done above) and using $\nabla \cdot \mathbf{B} = 0$, this becomes

$$-\frac{d}{dt}\nabla^2\mathbf{B} = -\frac{4\pi}{c}\nabla \times ne\frac{d\mathbf{v}}{dt}$$

But, $d\mathbf{v}/dt = \mathbf{a}$, the acceleration, and Newton's law for these carriers is $\mathbf{a} = -e\mathbf{E}/m$, so the right hand side becomes

$$-\frac{d}{dt}\nabla^2\mathbf{B} = \frac{4\pi ne^2}{mc}\nabla \times \mathbf{E}$$

Now, I can replace $\nabla \times \mathbf{E}$ using Maxwell's Eq. 4c, interchange the operations on the left hand side, and use Eq. 218 to arrive at

$$\nabla^2\frac{d\mathbf{B}}{dt} = \frac{1}{\lambda_p^2}\frac{d\mathbf{B}}{dt},$$

where λ_p is the penetration length of the perfect conductor. The behavior of the derivative for the perfect conductor is of the same form as Eq. 220 for the superconductor, except

---

[*] At the same time $\sigma_2 \to ne^2/m\omega$, either by taking the limit or by using the Kramers-Kronig relation for the conductivity.

that it is for the derivative rather than the field itself. So for the perfect conductor, a few penetration depths from the surface, the field can be present, but

$$\frac{d\mathbf{B}}{dt} = 0,$$

i. e., cannot change. Of course, the decay length of the time-varying field is just the skin depth in the $\omega > 1/\tau$ limit.* The length is

$$\lambda_p = \frac{c}{\omega_p} = \sqrt{\frac{mc^2}{4\pi ne^2}},$$

exactly the same as Eq. 218, the equation for the London penetration depth, λ_L.

11.4 Excitations in a superconductor

The BCS theory[87,89,94] prescribes an energy gap Δ at the Fermi surface; no electron may have an energy (measured relative to the Fermi energy—aka the chemical potential) smaller than Δ. Similarly, no hole may exist with an energy† (the absence of an electron below the Fermi surface) (measured relative to the Fermi energy) smaller than Δ.

The gap is in the energy spectrum of single-particle excitations of the superconductor. It has an analog in the energy gap of a semiconductor. In the semiconductor, the lowest electron energy above the Fermi Energy is at the bottom of the conduction band; the lowest hole energy is at the top of the valence band. In the intrinsic (undoped) semiconductor, the Fermi level lies in the middle of the gap (with magnitude E_g) between valence and conduction bands. The minimum excitation energies are $E_g/2$. But superconductors are of course not semiconductors. The gap in the latter is due to the static potential of the crystal; it occurs at a crystal momentum $\mathbf{k}$ at the Brillouin zone boundary. The gap in the superconductor is tied to the Fermi surface; it occurs at a wavevector $\mathbf{k}$ whose magnitude equals the Fermi wavevector, k_F.

There is no gap in the values of the wavevectors $\{\mathbf{k}\}$! Just as in the extended zone scheme of a semiconductor, the wavevectors cover a continuous range from zero to infinity.‡ I am writing this paragraph because in many discussion of superconducting gaps and gap symmetry, a picture is drawn where the gap is shown as a shell around a Fermi sphere. The picture is wrong for two reasons: (1) The Fermi sphere is a k-space entity and there are no forbidden values of the wavevector $\mathbf{k}$. (2) The gap is most important for states near the Fermis surface, as I will discuss in a minute, but it affects (at least formally) all the quasiparticle states. In particular, hole states below the Fermi level are also gapped.

---

[*] The perfect conductor has $1/\tau = 0$ so any finite frequency is in this limit.

[†] Hole energy is measured *downwards* from the Fermi level.

[‡] In the reduced zone scheme, the wavevectors of each band can take values anywhere in the Brillouin zone.

11.4.1 The excitation spectrum of a normal metal

Before discussing the excitation spectrum of a superconductor, let me think about the excitations of a normal metal. The ground state is the $T = 0$ Fermi sphere shown on p. 75. All single-particle states between zero and $\mathcal{E}_F$ are occupied; all above are empty. There are two types of excitations out of this ground state. Electron-like excitations occupy states above $\mathcal{E}_F$ and hole-like excitations correspond to vacant states below $\mathcal{E}_F$. If the excitations are made by (low-energy) optical absorption as intraband transitions or if they are due to finite temperatures (and the broadening of the Fermi-Dirac function) then they are made by promoting an electron from below to above $\mathcal{E}_F$.[*]

As I did on page 71 I'll call the excitations, whether hole-like or electron-like, "quasiparticles." I can plot the quasiparticle excitation spectrum in the normal metal. The electron energies are

$$\mathcal{E}(k) = \frac{\hbar^2 k^2}{2m}$$

as in Eq. 109. Excitation energies are measured relative to the Fermi energy, so the energy of an electron-like or hole-like excitation

$$\mathcal{E}_n(k) = \left| \frac{\hbar^2 k^2}{2m} - \mathcal{E}_F \right| \equiv |\epsilon(k)| \,. \tag{221}$$

Here, $\epsilon(k)$ is the kinetic energy of the electron, measured relative to the Fermi energy.

I can plot $\mathcal{E}_{qp}(k)$ vs k. The curve will be almost linear, because I am interested in excitation energies on the scale of the energy gap. Superconducting gaps are typically a few meV, whereas the Fermi energy is typically more than 10 eV. [†] (See Table 7.) For meV variations around the Fermi energy, the curvature in Eq. 109 as a function of k is not noticeable. But to be sure that things are absolutely linear, I will plot the quasiparticle energy against the kinetic energy, measured relative to the Fermi energy, i. e., $\mathcal{E}_n(k)$ versus $\epsilon(k)$. Such a plot is shown in Fig. 57 as the red dashed line. The left side shows the hole-like quasiparticles and the right side the electron-like quasiparticles.

11.4.2 The excitation spectrum of a superconductor

The excitation spectrum of a superconductor is modified from that of the normal metal by the presence of the energy gap. The BCS theory[87,89,94] finds that

$$\mathcal{E}_s(k) = \sqrt{\epsilon(k)^2 + \Delta^2} \tag{222}$$

The quantity $\mathcal{E}_s(k)$ is shown in Fig. 57 as the blue solid line. The minimum quasiparticle energy is the gap Δ. Because optical absorption breaks a Cooper pair and produces a pair

---

[*] There are other options. I may inject electrons by tunneling through a barrier. Unless I withdraw them somewhere else, the metal will become charged. In small structures, this process can be observed as single electron tunneling and leads to the Coulomb blockade.[95]

[†] There are exceptions, including the cuprate superconductors and organic superconductors.

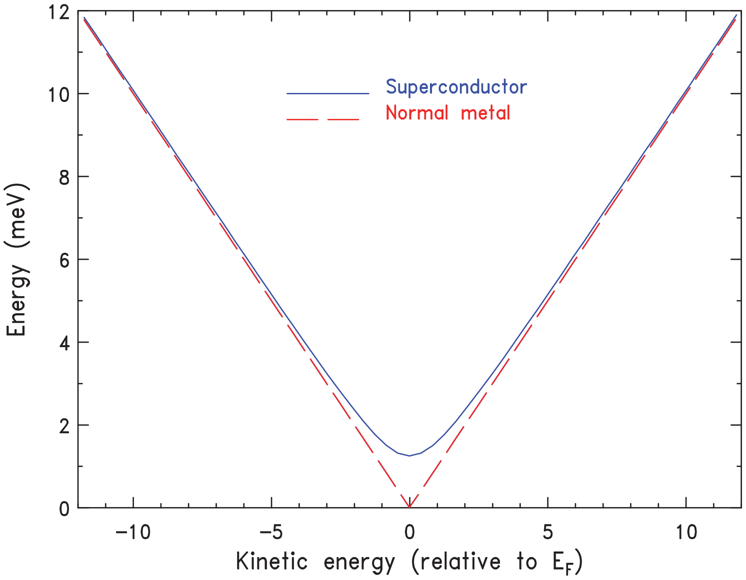

Fig. 57. The excitation spectrum of a normal metal (dashes) and a superconductor (solid). The superconducting gap was taken to be $\Delta = 1.5$ meV.

of quasiparticles, the corresponding threshold for optical absorption is 2Δ. I only need a potential difference of Δ if I tunnel electrons in* from a normal metal into the superconductor.

It is remarkable that the phenomenon of superconductivity, with such a dramatic effect on electrical transport and optical properties, affects the energy spectrum of the electrons in such a modest way. Once one moves several times the gap energy away from the Fermi surface, the energy spectrum is indistinguishable, and one might say that those deeply buried (and highly excited) states are indistinguishable from the normal metal. Yet, the London penetration depth, Eq. 218, involves all the conduction electrons.† The heat capacity also suggests that all the electrons participate in the superconducting condensation.

In some sense, I could take the same point of view about the conduction by a normal metal, shown in Fig. 35. Instead of thinking that the entire Fermi sphere is displaced, I could say that a number of electrons are excited above the Fermi surface in the direction of the electron flow, and an equal number of holes exist on the opposite side, where occupied electron states once existed.

For electrons far below $\mathcal{E}_F$, the states $|\mathbf{k},\uparrow\rangle$ and $|-\mathbf{k},\downarrow\rangle$ are both occupied with 100% probability. Are these Cooper pairs? Not really, because the BCS state is a *coherent* combination of Cooper pairs.

I can also represent the excitation energy in a semiconductor picture. In such a picture, the kinetic energy is just that of the electrons, so states below $\mathcal{E}_F$ have negative energies. I

---

[*] Or out; tunneling electrons out corresponds to putting a hole in. This *single-particle* tunneling differs from the pair tunneling of the Josephson effect. In single-particle tunneling, I must have a potential of $\pm\Delta$ between normal metal and superconductor, or I must have a potential of $\pm 2\Delta$ between two superconductors, in order to have a tunneling current. The dc Josephson effect is the tunneling of Cooper pairs between two superconductors at zero bias.

[†] In clean metals

get the picture shown in Fig. 58. The electron-like excitations have positive energies and the hole-like ones have negative energies. That the optical gap is 2Δ is obvious here, because I must promote an electron from the highest-energy state below $\mathcal{E} = 0$ (the Fermi energy) to the lowest energy state above $\mathcal{E} = 0$.

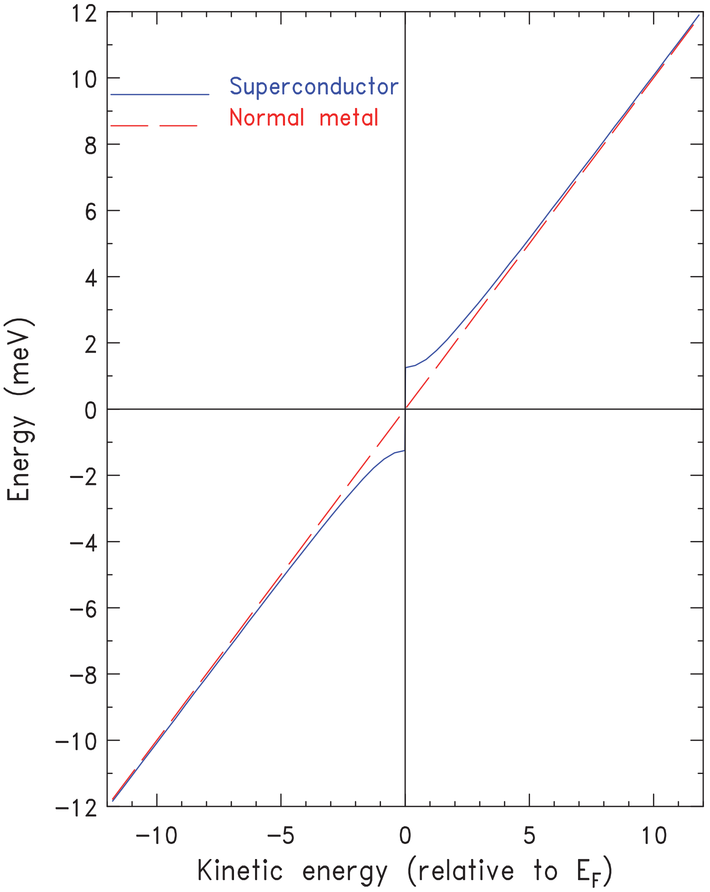

Fig. 58. The "semiconductor picture" of the excitation spectrum of a normal metal (dashes) and a superconductor (solid). The superconducting gap was taken to be $\Delta = 1.5$ meV. The Fermi energy is at $\mathcal{E} = 0$.

I like the excitation picture, Fig. 57, better. Many people do use the semiconductor picture, however.

11.5 The coherence length and density of states

11.5.1 Pippard's argument

There is a second length in addition to the penetration depth discussed already. It is the coherence length, first introduced by Pippard.[96] The coherence length also appears in the Ginsberg-Landau theory[88,90] and BCS theory.[87,94] I'll give a heuristic argument as follows. The Cooper-pair wave function extends over a finite distance; at longer distances the phase coherence is lost. This means that the wave function ψ has an envelope

$$\psi{\sim}e^{-r/\xi_0}$$

where ξ_0 is the coherence length. A consequence of the finite length is that plane waves are no longer correct for the wave function. Instead there is must be a range of wave vectors, i. e., a q-dependent theory. The range of wave vectors must be related to the superconducting binding energy Δ

$$k_{sc} = k_F \pm q = k_F \pm \frac{1}{\xi_0}.$$

This range leads to a spread of kinetic energies,

$$\frac{\hbar^2 k_{sc}^2}{2m} = \frac{\hbar^2 k_F^2}{2m} \pm \frac{\hbar^2 k_F}{m\xi_0} = \mathcal{E}_F \pm 2\Delta,$$

where I have neglected the term containing $1/\xi_0^2$. Note that this is basically an uncertainty principle argument: limiting the space part to ξ_0 leads to a spread of momenta and corresponding spread of energies. Subtracting $\mathcal{E}_F$ from both sides gives $\hbar^2 k_F/m\xi_0 = 2\Delta$ or

$$\xi_0 = \frac{\hbar^2 k_F}{2m\Delta} = \frac{\hbar v_F}{2\Delta}$$

using $\hbar k_F = m v_F$.

The BCS theory gives

$$\xi_0 = \frac{2}{\pi}\frac{\hbar v_F}{2\Delta}, \tag{223}$$

close enough. The materials parameters that appear in Eq. 223 are the Fermi velocity (or Fermi wave vector), which increase with carrier density, and the gap (or transition temperature, because in BCS $2\Delta = 3.5kT_c$). Thus high electron density, low T_c materials (think Al) will have long coherence lengths; low electron density, high T_c materials (think $YBa_2Cu_3O_7$) will have short coherence lengths. This expectation is in accord with observation. See Table 15.

Table 15. BCS coherence length of superconducting materials.

Materials	T_c (K)	ξ_0 (nm)
Aluminium	1.19	1200
Indium	3.40	330
Tin	3.72	260
Gallium	5.90	160
Lead	7.22	80
Niobium	9.25	35
PbMo	15	2.5
Nb_3Sn	17	4.0
$C_{60}K_3$	19	3.0
$C_{60}Rb_3$	31	2.3
$YBa_2Cu_3O_7$	93	1.5

11.5.2 The dirty limit

Dirty metals, alloys, and thin films often have short mean free paths. The mean free path ℓ can be of order 1–2 nm, compared to perhaps micrometers in pure materials at low temperatures. When the mean free path is shorter than the coherence length, the coherence length itself is shortened, to[94,97]

$$\frac{1}{\xi(l)} = \frac{1}{\xi_0} + \frac{1}{\ell}$$

with limits of ξ_0 in the clean limit and ℓ in the dirty limit.

The short mean free path means a broad spectrum for $\sigma_1(\omega)$. (See Eq. 56a and Fig. 8.) It also has an effect on the penetration depth. To see how, let me start with a series of inequalities:

$$\begin{aligned} \ell &< \xi_0 \\ v_F\tau &< \frac{\hbar v_F}{2\Delta} \\ 2\Delta &< \frac{\hbar}{\tau} \end{aligned} \tag{224}$$

indicating that the Drude spectrum extends beyond (often well beyond) the gap in the conductivity spectrum.

Now, I invoke the conductivity sum rule, Eq. 150:

$$\int_0^{\omega} d\omega\, \sigma_1(\omega) = \frac{1}{8}\omega_p^2$$

where I'll stop the integral at a frequency well above $1/\tau$ but below the onset of interband transitions, so that ω_p is the free-carrier plasma frequency. The sum rule is obeyed in the normal state and in the superconducting state. The normal state has the Drude free-carrier response. The superconducting state has an infinite dc conductivity [a delta function at $\omega = 0$, $A\delta(\omega)$], a gap of 2Δ and a conductivity close to the normal state at high frequencies.

An example of the conductivity of a dirty-limit superconductor is shown in Fig. 59. (The dots are data for σ_1 and σ_2 while the solid lines are a BCS theory calculation.) The conductivity is normalized by the normal-state conductivity, essentially equal to the dc conductivity. $\sigma_{\rm dc}$, so that a value of unity (at all frequencies) corresponds to the normal-state conductivity

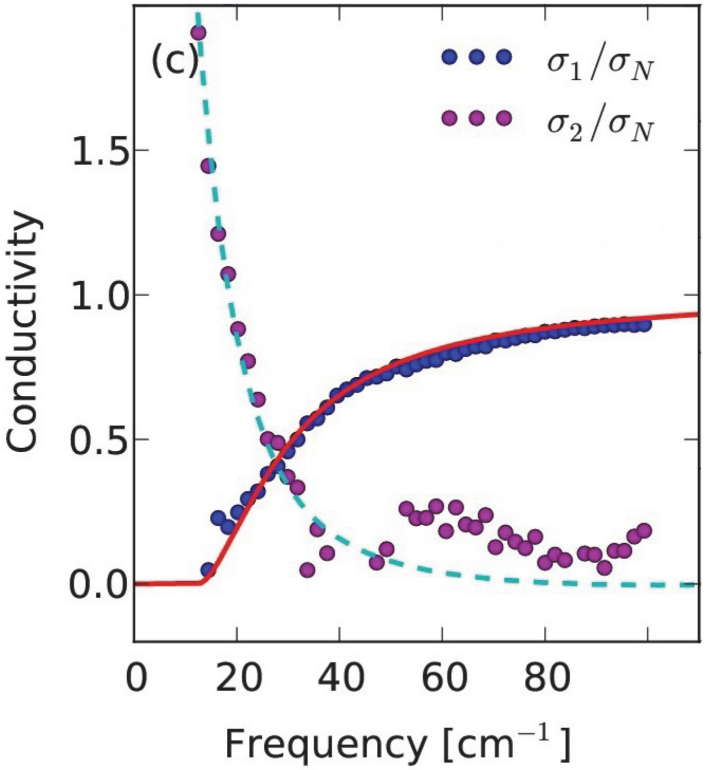

Fig. 59. Conductivity of a NbTiN film at low temperatures.

The superconducting state conductivity is zero up to the gap, 2Δ, after which it rises slowly to join the normal-state conductivity. The sum rule says that the "missing area" in σ_{1s} appears in the delta function dc conductivity. I can estimate the strength of this delta function from the sum rule. Let me calculate

$$\int_{0^+}^{\omega} d\omega\,[\sigma_{1n}(\omega) - \sigma_{1s}(\omega)] = \frac{1}{2}A = \frac{1}{8}\omega_{ps}^2$$

where I start the integral at 0^+ to miss the delta function, and the factor of $A/2$ is the integral of the delta function from 0 to 0^+. The weight of the delta function gives $\sigma_2 = \omega_{ps}^2/8\omega$. (The factor of 1/2 was discussed on page 144; it arises because the integral starting at $\omega = 0$ only picks up half of the area of the delta function.) The missing area is proportional to $\sigma_{\rm dc} = ne^2\tau/m$, to 2Δ, the width of the gap, and a factor I will call Z that represents the way the conductivity rises to reach the normal state value. Then

$$\frac{1}{8}\omega_{ps}^2 = \frac{ne^2\tau}{m}\cdot 2\Delta\cdot Z$$

and, hence,

$$\frac{1}{\lambda_L} = \frac{\omega_{ps}}{c} = \sqrt{\frac{8ne^2\tau}{m} 2\Delta Z}.$$

The penetration depth increases as τ gets shorter, following $\lambda_L \sim \tau^{-1/2}$.

11.5.3 The clean limit

It is easy to guess that the clean limit is the case opposite to that just discussed. The clean metal has a long mean free path, so that $\ell \gg \xi_0$. Then the coherence lengthy is what it is, with examples in Table 15. The inequalities in Eq. 224 point in the opposite direction, and

$$2\Delta \gg \frac{\hbar}{\tau} \tag{225}$$

The full width at half-maximum of the Drude function, Eq. 56a is equal to $1/\tau$, so that all of the Drude spectral weight is below the gap, and condenses to form the superconducting delta-function dc conductivity. The normal state conductivity evaluated at the gap frequency $\omega = 2\Delta/\hbar \equiv \omega_g$ is

$$\sigma_D = \frac{ne^2/m}{\omega_g^2 \tau} + i\frac{ne^2}{m\omega_g},$$

making σ_2 larger than σ_1 by a factor of $\omega_g \tau$.

If the normal metal is a really good conductor, then the anomalous skin effect applies in the normal metal (discussed in section 12). Similar physics dominates the optical conductivity in the superconducting state. The conductivity is wave-vector dependent, with the q value in the range of $1/\lambda_L$. The anomalous limit has been described by Mattis and Bardeen[98] and Tinkham.[94]

Other materials, especially those with lower electron density such as the cuprate superconductors and organic superconductors, have normal skin depth while being int he clean limit. Then, $\omega_{ps} = \omega_p$ and the conductivity will be

$$\sigma = \frac{\pi ne^2}{m}\delta(\omega) + i\frac{ne^2}{m\omega}. \tag{226}$$

11.5.4 The density of states

There is a one-to-one correspondence between the states of the superconductor and those of the normal metal, indexed by the quantum numbers *veck* and spin. Thus every superconducting state with energy $\mathcal{E}(k)$ is in correspondence with a normal-metal state with energy ϵ_k. The relation is of course given by Eq. 222. Thus, I have

$$\mathcal{N}_s(\mathcal{E})d\mathcal{E} = \mathcal{N}_n(\epsilon)d\epsilon$$

where $\mathcal{N}_s$ is the density of states of the superconductor and $\mathcal{N}_n$ the same for the normal metal. The free-electron density of states is $\mathcal{N}_n \sim \sqrt{\mathcal{E}_F + \epsilon}$, slowly varying for small ϵ around

$\mathcal{E}_F$. The range of ϵ in Fig. 57 is 10 meV: the Fermi energy is 10 eV. I can take $\mathcal{N}_n$ as constant, equal to the value at the Fermi level, $\mathcal{N}_0$. Then,

$$\mathcal{N}_s(\mathcal{E})d\mathcal{E} = \mathcal{N}_0\frac{d\epsilon}{d\mathcal{E}}.$$

Using Eq. 222, I find that $\epsilon(k) = \sqrt{\mathcal{E}(k)^2 - \Delta^2}$ and then

$$\frac{d\epsilon}{d\mathcal{E}} = \frac{\mathcal{E}}{\sqrt{\mathcal{E}^2 - \Delta^2}}$$

The superconductor has no states below Δ and a square root singularity located at Δ:

$$\mathcal{N}_s(\mathcal{E}) = \begin{cases} 0 & \mathcal{E} < \Delta \\ \dfrac{\mathcal{E}\mathcal{N}_0}{\sqrt{\mathcal{E}^2 - \Delta^2}} & \mathcal{E} > \Delta \end{cases} \tag{227}$$

This function is plotted in Fig. 60.

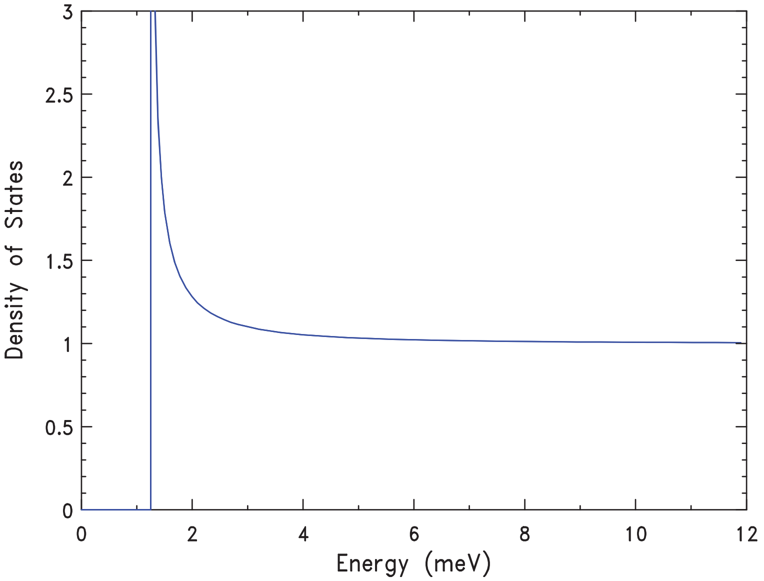

Fig. 60. The density of states of a superconductor versus the energy. $\mathcal{N}_s(\mathcal{E})$ is normalized by the Fermi surface density of states, $\mathcal{N}_0$.

Generally, when there is a peak in the density of states for excitations, there is a peak in the optical conductivity at that energy. Figure 59 shows that this peak is absent in a superconductor; indeed, the conductivity is zero exactly where the density of states diverges. It is worth asking why this happens. The answer is in the BCS coherence factors.

11.5.5 The coherence factors

The ground state of the superconductor is a linear combination of Cooper pair states where the pairs $|\mathbf{k}\uparrow, -\mathbf{k}\downarrow\rangle$ are either full [state $O(k)$] or empty [state $\bar{O}(k)$],

$$\psi_0 = v_k O(k) + u_k \bar{O}(k)$$

where v_k is the amplitude (square root of the probability) for the state $|\mathbf{k}\uparrow, -\mathbf{k}\downarrow\rangle$ to be full and u_k is the amplitude for the state $|\mathbf{k}\uparrow, -\mathbf{k}\downarrow\rangle$ to be empty. Normalization requires that $|u|^2 + |v|^2 = 1$. BCS[87,94] find that

$$v_k = \sqrt{\frac{1}{2}\left(1 - \frac{\epsilon}{\mathcal{E}}\right)} \tag{228}$$

and

$$u_k = \sqrt{\frac{1}{2}\left(1 + \frac{\epsilon}{\mathcal{E}}\right)} \tag{229}$$

with $\epsilon(k)$ defined in Eq. 221 and $\mathcal{E}(k)$ defined in Eq. 222. These functions are shown in Fig. 61.

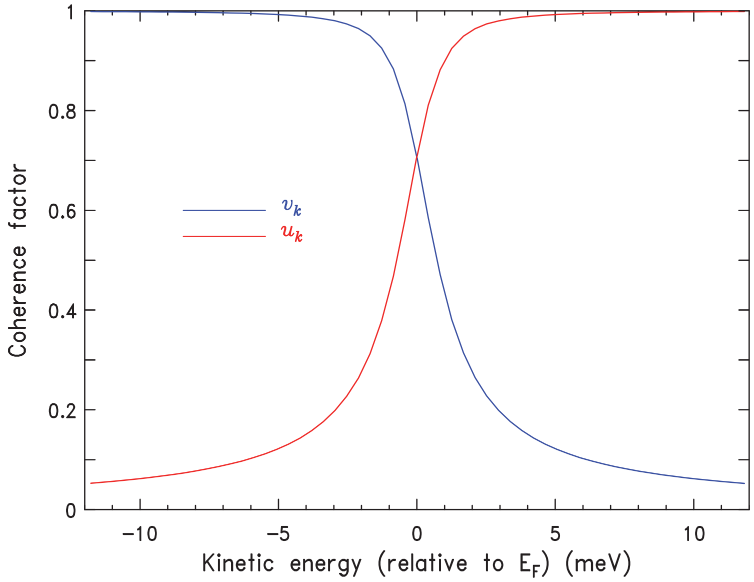

Fig. 61. The amplitude v_k for the state $|\mathbf{k}\uparrow, -\mathbf{k}\downarrow\rangle$ to be full and u_k for the state $|\mathbf{k}\uparrow, -\mathbf{k}\downarrow\rangle$ to be empty as a function of $\epsilon(k)$.

The figures shows that the superconductor (at $T = 0$ has a finite probability for Cooper-pair states with $k > k_F$ to be occupied and a finite probability for those with $k < k_F$ to be unoccupied.* This result should be contrasted with the Fermi-Dirac function of Eq. 115 which is perfectly sharp at $T = 0$. In fact, v_k qualitatively resembles $f(E)$ for $T \approx T_c$.

For states with $k = k_F$, the probability that the Cooper-pair state is empty and the probability that it is full are equal; I can think of this situation as implying that these

---

* $\epsilon(k) = 0$ for states at the Fermi energy.

Cooper pairs have 50% electron-like character and 50% hole-like character. The mixed character has important consequences for the optical properties and answers the question asked on page 157.

11.6 The optical conductivity of a superconductor

11.6.1 Mattis-Bardeen theory

The optical conductivity of a superconductor was first worked out by Mattis and Bardeen[98] and has been extended by Nam[99,100] and others.[101–103] The Mattis-Bardeen calculation took the scattering of photoexcited quasiparticles in the superconductor and of normal carriers in the metal to be the same, calculating $\sigma_{1s}(\omega)/\sigma_n$ and $\sigma_{2s}(\omega)/\sigma_n$ in the dirty limit ($\hbar/\tau \gg 2\Delta$ and the extreme anomalous limit ($\ell \gg \delta$). The result (at $T = 0$) is expressed in terms of complete elliptic integrals. The results may be written

$$\frac{\sigma_{1s}(\omega)}{\sigma_{1n}(\omega)} = \int_0^\infty d\mathcal{E}\left[1 + \frac{\Delta^2}{\mathcal{E}(\mathcal{E}+\hbar\omega)}\right]\left[\frac{\mathcal{N}_s(\mathcal{E})\mathcal{N}_s(\mathcal{E}+\hbar\omega)}{\mathcal{N}_0}\right][f(\mathcal{E}) - f(\mathcal{E}+\hbar\omega)]$$

where $f(\mathcal{E}$ is the Fermi function, enforcing the requirement that (in the semiconductor picture) the excitation is from occupied to occupied state. The density of states $\mathcal{N}_s$ is given in Eq. 227.

And a plot

11.6.2 Hand-waving calculation

Here I will make a simple-minded calculation based on the ideas of the coherence factors and the mixing of electron-like and hole-like states near the gap energy. At high frequencies the mixing is minimal and the normal-state conductivity and the superconducting-state conductivity are essentially indistinguishable. This statement agrees with observation: there is no visible change in the color of the metal on making the transition from normal to superconducting.

For frequencies near the gap, I will use a Drude-like model for the quasiparticle scattering.* Note that the scattering is essential to give a dominant real part to the conductivity; if $\omega \gg 1/\tau$ the conductivity is mostly imaginary. So, I write

$$\sigma_{1s}(\omega) = \frac{nQ_k^2\tau/m}{1+\omega^2\tau^2}.$$

where $Q_k \equiv Q\left(\mathcal{E}(k)\right)$ is the effective charge of an excitation at energy $\mathcal{E}(k)$, given by Eq. 222. This equation should be compared to the ordinary Drude conductivity of Eq. 56a,

$$\sigma_{1n}(\omega) = \frac{ne^2\tau/m}{1+\omega^2\tau^2},$$

so tnat

$$\frac{\sigma_{1s}(\omega)}{\sigma_{1n}(\omega)} = \frac{Q_k^2}{e^2}.$$

What does $Q(\mathcal{E})$ llok like. Well, remember Eqs. 228 and 229, plotted in Fig. 61, the amplitude for a state to be full (electron-like or empty (hole-like). Their squares are the

---

* As did Mattis and Bardeen.

probabilities, of course. Holes and electrons have oppsite charges, so the effective charge is the difference between the probabilities. I'll write it as

$$Q_k = -e(v_k^2 - u_k^2)$$

Using Eqs. 228 and 229, I get

$$v_k^2 - u_k^2 = \frac{\epsilon}{\mathcal{E}}$$

so that so long as the energies of both excitations* exceed the gap,

$$Q_k^2 = \frac{\epsilon^2}{\mathcal{E}^2} = \frac{\mathcal{E}^2 - \Delta^2}{\mathcal{E}^2}$$

Now, the energy of the excitations is delivered by the photon energy $\hbar\omega$. I'll take the energy of each excitation as half of this energy.† The conductivity is tben

$$\sigma_{1s}(\omega) = \begin{cases} A\delta(\omega) & \omega = 0 \\ 0 & 0 < \hbar\omega \leq 2\Delta \\ \sigma_{1n}(\omega)\,\dfrac{\omega^2 - (2\Delta_\omega)^2}{\omega^2} & \hbar\omega \geq 2\Delta \end{cases} \tag{230}$$

where $\Delta_\omega = \Delta/\hbar$. By the sum rule, A, the weight of the delta function, is given by the missing area in the superconducing-state conductivity,

The conductivity is zero right at the gap (despite the divergence of the density of states there). The reason is that the effective charge of the gap-edge excitation is zero. It then rises smoothly until it becomes equal to the normal-state conductivity at $\omega \gg 2\Delta_\omega$. This behavior is shown in Fig. 62.

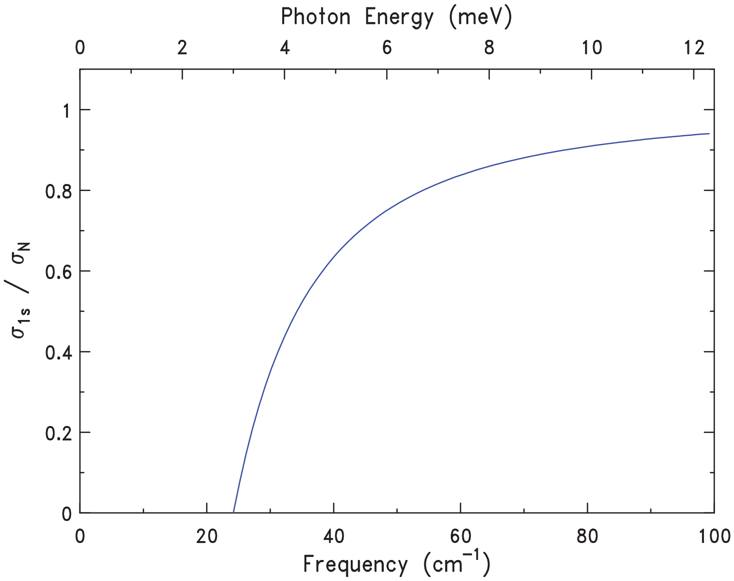

Fig. 62. The optical conductivity according to Eq. 230. The gap was taken to be $2\Delta = 3$ meV ($24\,\mathrm{cm}^{-1}$).

---

[*] Remember that the excitations are created in pairs.

[†] This division of energy is where I am waving hands the most; in principle, I should integrate over a range of energies from 0 to $\hbar\omega$ for one excitation and from $\hbar\omega$ to 0 for the other.

12. NONLOCAL EFFECTS: THE ANOMALOUS SKIN EFFECT

Ohm's law is *local:* the current $\mathbf{j}$ at location $\mathbf{r}$ depends on the electric field at location $\mathbf{r}$ and not on the field anywhere else. In pure metals at low temperatures the response to the field becomes non-local. The current at location $\mathbf{r}$ depends on fields at other locations $\mathbf{r}'$. This non-local response is not unprecedented. Anyone who has played shuffleboard (or hockey) knows that the puck will slide considerable distance away from the location where the force was applied to it.

The principal optical consequence of non-locality is the anomalous skin effect. I'll first review the normal skin effect, first introduced on page 57. Then I'll discuss the conditions necessary for the appearance of nonlocal effects and make a simple argument of how to calculate the skin depth in the nonlocal regime.

12.1 The normal skin effect

Recall Eqs. 89 and 90, which specify the penetration depth of a metal. The amplitude of the field decays with distance along the path of the light as $e^{-x/\delta}$, with $\kappa\omega/c = 1/\delta$. At low frequencies, $\omega\tau \ll 1$, I found

$$\delta = \frac{c}{\omega\kappa} \approx \sqrt{\frac{c^2}{2\pi\omega\sigma_{\rm dc}}} = \frac{c}{\omega_p}\sqrt{\frac{2}{\omega\tau}},$$

where $\sigma_{\rm dc} = \omega_p^2\tau/4\pi$. In the other limit, high frequencies, $\omega\tau \gg 1$, and the penetration depth becomes

$$\delta \approx \frac{c}{\omega_p}.$$

This length, c/ω_p is as short as the penetration depth gets.

The skin effect is not only an optical effect. It also affects the currents in wires carrying high-frequency* electrical currents. Figure 63 shows the field or current distribution in the cross-section of a circular wire. The current density is high near the surface, and falls to near zero in the center. Figure 64 shows the 300 K skin depth for four metals at frequencies from 1 Hz to 1 THz. The skin depth decreases over that range from 10 cm to 100 nm. Note that 1 THz is 33 cm^{-1} or about 4 meV. On this many-decades plot, the differences among the metals (whose conductivity varies by about a factor of 2.) is not large.

In Fig. 65 I show the skin depth for silver in the optical range, $10^{-4}\,\mathrm{cm}^{-1}$ to 40,000 cm^{-1}, or just over 10^{-8} eV to 5 eV. Note that 1 cm^{-1} is 30 GHz, so the frequency scale is also 3 MHz to 1.2 PHz. This range includes $\omega\tau = 1$, so the skin depth initially falls as $\sqrt{1/\omega}$ before flattening out in the high frequency limit. There is a significant increase in the skin depth near the plasma minimum.†

---

[*] Or any frequency. Even at 60 Hz, good conductors have a finite penetration of fields and currents. The skin depth in copper is about 1 cm at 60 Hz and 300 K.

[†] The calculation does not make an approximation for δ. The dielectric function model consists of a Drude free-carrier part, plus Lorentz oscillators above 4 eV to account for the contribution of the d electrons. I then calculated ϵ, followed by $N = n + i\kappa = \sqrt{\epsilon}$ and $\delta = c/\omega\kappa$.

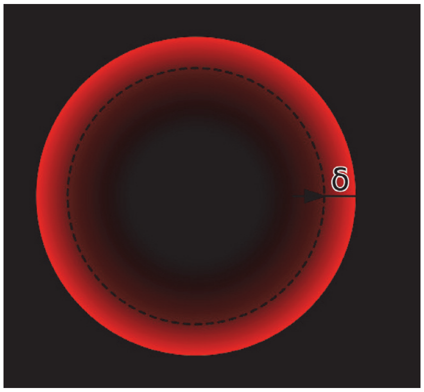

Fig. 63. The current distribution in a conductor carrying an alternating current.

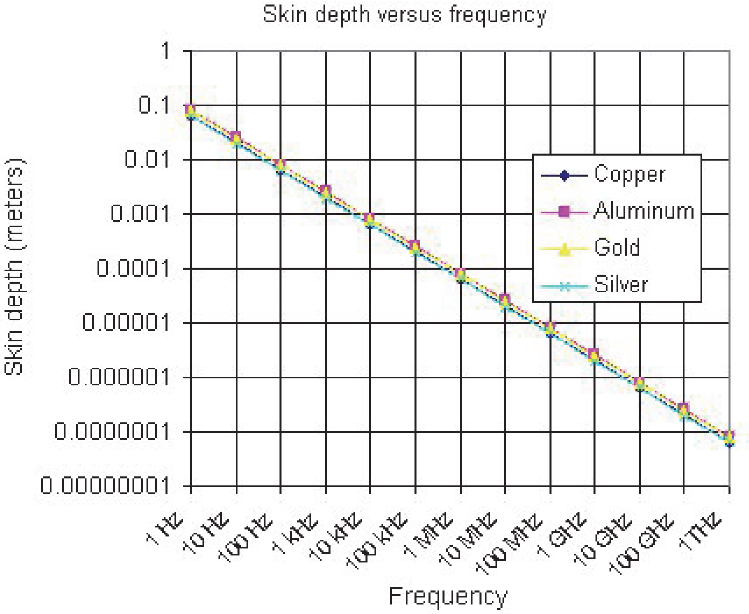

Fig. 64. The skin depth as a function of frequency for four metals.

At room temperature, the skin depth is comparable to the 40 nm mean free path at high frequencies and is much larger that this at frequencies below $1/\tau$. Now, as temperature is reduced, the conductivity improves (because the mean free path is getting longer and longer), making the skin depth (Eq. 89) shorter and shorter. Figure 65 shows the classical skin depth calculated for silver assuming the conductivity varies as in Fig. 37. It does not take long in the case of silver and other good conductors to get into a situation where the mean free path is much larger than the skin depth. The mean free path is the length scale (just as the relaxation time is the time scale) for relaxation of the current. But if the mean free path is larger than the skin depth, then those electrons moving at angles to the surface see an electric field which varies enormously over the trajectory. Electrons moving parallel to the surface see a constant field. This situation is illustrated in Fig. 66. The current is a complicated and nonlocal response to the field. The dashed curves for temperatures in the 10–100 K range in Fig. 65 are do not apply to this physical situation.

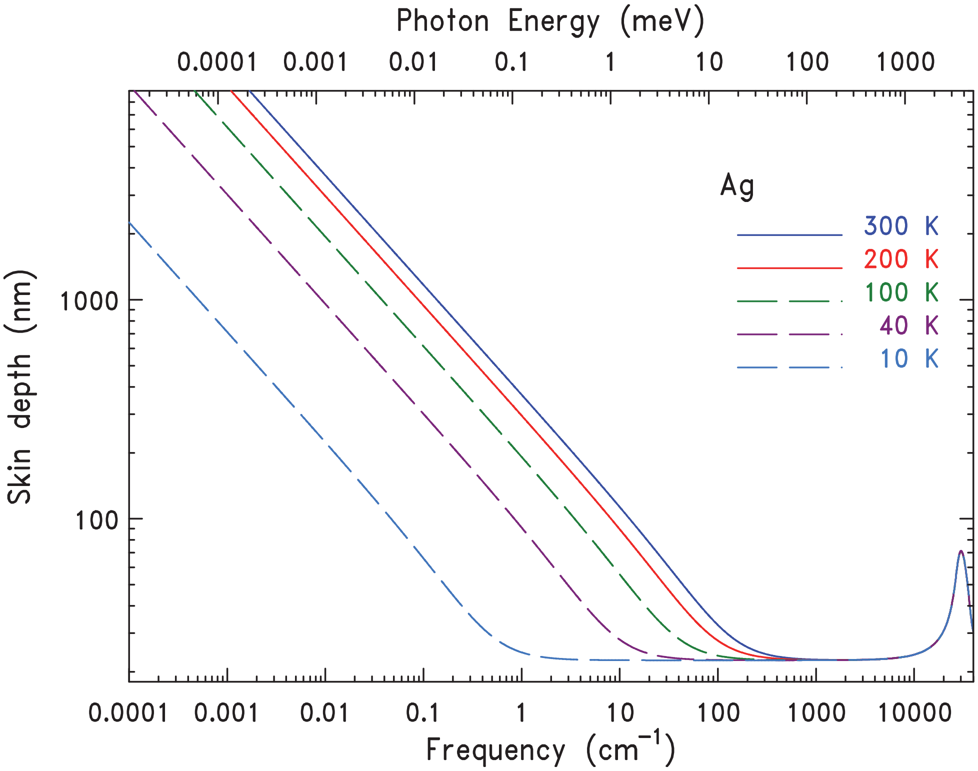

Fig. 65. Classical skin depth for silver from RF to uv frequencies at 5 temperatures. The curves shown with dashed lines are not realistic; silver would be in the anomalous skin effect regime and not the classical regime over much of the frequency range. Note that 0.0001 cm^{-1} = 3 MHz and 30,000 cm^{-1} = 900 THz.

12.2 Anomalous skin effect

This anomalous skin effect was studied in the 1950s by Pippard,[104] Chambers,[105] and many others. As I will discuss, the physical situation is:

1. $\delta \neq \delta_{\text{cl}}$.
2. $\delta_{\text{ASE}} > \delta_{\text{cl}}$.
3. The current is nonlocal, so Ohm's law becomes a convolution of the form

$$\mathbf{j}(\mathbf{r}) = \int_V \Sigma(\mathbf{r}-\mathbf{r}')\mathbf{E}(\mathbf{r}')d^3\mathbf{r}\prime$$

   where $\Sigma(\mathbf{r}-\mathbf{r}')$ is the Fourier transform of $\sigma(q,\omega)$ with $\hbar q$ a momentum.
4. The fields are q dependent with $\langle q\rangle \approx 1/\delta_{\text{ASE}}$.

The effect of an external field on a specific electron state **k** depends on how **k** is oriented relative to the surface. Two distinct electron trajectories shown in Fig. 66. Each starts immediately after a collision and travels for the mean free path before another collision occurs. The electrons are accelerated by the electric field as they move along their trajectories. The electron whose path is entirely with in the skin feels the force over the entire trajectory. (The field does work during the entire time that the electron is is flying free between collisions; the work gives energy to the electron and causes it to contribute to the electric current. At the collision, the electron motion is randomized and no longer contributes to the current.)

In contrast the electron that comes from deep with metal into the skin and scatters, perhaps departing again into the depths only sees the field for a small part of its total free path. The field therefore does much less work on this second electron.

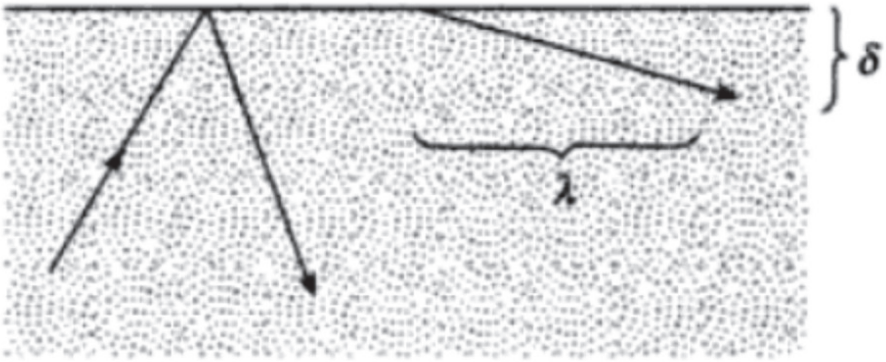

Fig. 66. Two trajectories in a clean metal. One has the electron staying within the skin and seeing the external field over its entire path. The second shows the electron coming out of the depths and then going back into them. (ℓ is the mean free path in this diagram.)

12.3 The extreme anomalous limit

The concepts in the previous paragraph form the basis of Pippard's effectiveness concept for the skin effect. The electric field amplitude decays from the surface with a length δ. The decay occurs because energy is absorbed by the electrons. To be most effective, the entire trajectory must be where the field is finite, i. e., within the skin. The path length is the mean free path ℓ. When $\ell >> \delta$, this requirement restricts the angle of the $\mathbf{k}$ to a small value relative to the surface. The "effective" number of electrons is smaller than the total number:

$$n_{\text{eff}} = \frac{\delta}{\ell} n. \tag{231}$$

The metal then acts as if the conductivity (i. e., the current) is due only to the effective electrons*

$$\sigma_{\text{eff}} = \frac{n_{\text{eff}} e^2 \tau}{m} = \frac{n_{\text{eff}} e^2 \ell}{m v_{\text{F}}}$$

where I have used $\tau = \ell / v_{\text{F}}$. Now replace n_{eff} using Eq. 231:

$$\sigma_{\text{eff}} = \frac{n e^2 \delta}{m v_{\text{F}}} \tag{232}$$

I now calculate the skin depth self consistently:

$$\delta_{\text{ASE}} = \frac{c}{\sqrt{2\pi \sigma_{\text{eff}} \omega}} = \frac{c}{\sqrt{2\pi \left(\frac{n e^2 \delta_{\text{ASE}}}{m v_{\text{F}}} \right) \omega}}$$

---

* I write the conductivity as a dc conductivity. Hence, this picture is a low frequency, $\omega \ll 1/\tau$, picture. The high frequency penetration length c/ω_p is a kinetic inductance or inertial effect and does not depend on the mean free path.

where I have substituted using Eq. 232. Square both sides, multiply by δ_{ASE}, and take cube root to find

$$\delta_{\mathrm{ASE}} = \left(\frac{c^2 m v_{\mathrm{F}}}{2\pi n e^2 \omega}\right)^{1/3}. \tag{233}$$

The skin depth in the extreme anomalous regime goes as $\omega^{-1/3}$ compared to $\omega^{-1/2}$ in the normal skin regime. If I use the definition of the plasma frequency, Eq. 127, I can recast this for comparison with Eqs. 89 and 90:

$$\delta_{\mathrm{ASE}} = \frac{c}{\omega_p}\left(\frac{2v_{\mathrm{F}}}{c}\right)^{1/3}\left(\frac{\omega_p}{\omega}\right)^{1/3}. \tag{234}$$

The plasma frequency, ω_p, appears *twice* in Eq. 234; I put the c/ω_p factor in front to allow comparison to the normal skin effect formulae. Two cube roots follow this scale factor. The first cube root is smaller than unity[*] but the second is *much* larger than unity. The plasma frequency is 70,000 cm^{-1} or 9 eV whereas the frequency is smaller than $1/\tau = 0.1\ \mathrm{cm}^{-1}$ or 10 μeV in pure metals[†] at low temperatures.

None of the terms in Eq. 233 have any significant temperature dependence; consequently, neither does δ_{ASE}. The temperature-dependent family of curves calculated for the normal skin effect all will merge into a single, frequency-depended curve at low temperatures. This behavior is illustrated in Fig. 67, where I show the skin depth for silver in the optical range, $10^{-4}\ \mathrm{cm}^{-1}$ to 40,000 cm^{-1}, or just over 10^{-8} eV to 5 eV. Note that 1 cm^{-1} is 30 GHz, so the frequency scale is also 3 MHz to 1.2 PHz. The anomalous limit (which is temperature independent) is shown as a black solid line. With increasing frequency the skin depth passes from classical to anomalous and then back to a high-frequency relaxation regime as $\omega \sim 1/\tau$. At the lowest temperatures, the crossover is at 70 kHz.

12.4 ω-τ plot

The anomalous regime has both a high-frequency and a low-frequency boundary. The low-frequency boundary occurs where the anomalous skin depth equals the normal skin depth and also equals the mean free path: $\delta_{\mathrm{ASE}} = \delta_{\mathrm{cl}} = \ell$. If $\delta = \ell$ then all electrons are effective.

I can calculate the frequency where the transition occurs by setting $\ell = \delta_{\mathrm{cl}}$ to find

$$\ell = v_{\mathrm{F}}\tau = \left(\frac{c^2}{\omega_p^2}\frac{2}{\omega\tau}\right)^{1/2}$$

or

$$\omega = \frac{2c^2}{\tau^3 v_{\mathrm{F}}^2 \omega_p^2} = \frac{2c^2 v_{\mathrm{F}}}{\ell^3 \omega_p^2} \tag{235}$$

This is also the frequency where $\delta_{\mathrm{cl}} = \delta_{\mathrm{ASE}}$.

Equation 235 represents a boundary between the normal skin effect (low ω or short τ) and the anomalous skin effect (higher ω or long τ). Both of these boundaries are computed

---

[*] Recall: $v_{\mathrm{F}} = c/200$.

[†] The anomalous skin effect *only* occurs in pure metals at low temperatures.

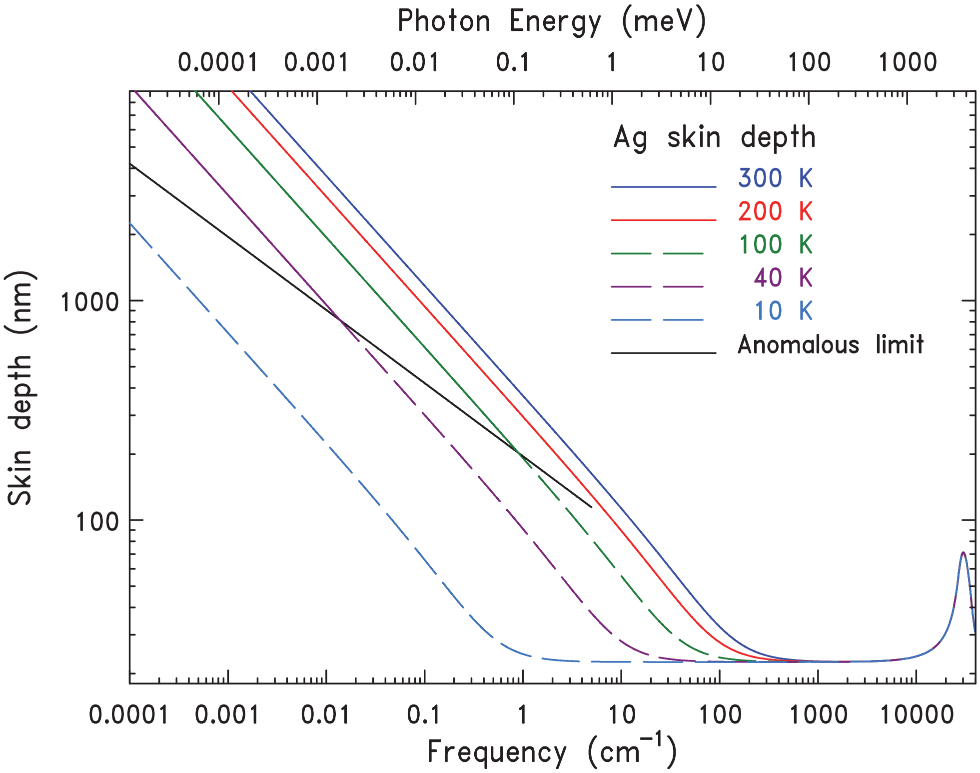

Fig. 67. Skin depth for silver from RF to uv frequencies at 5 temperatures. The extreme anomalous limit is shown as a solid line. The skin depth calculated from the classical equation is shown with dashed lines where it is incorrect. Below the frequency where the classical and anomalous skin effect curves cross, silver is in the classical regime. Note that 0.0001 cm^{-1} = 3 MHz and 30,000 cm^{-1} = 900 THz.

in the low-frequency limit, $\omega \ll 1/\tau$. Once the frequency increases enough that $\omega \gg 1/\tau$, both normal and anomalous skin effects are modified. In this high frequency region, the field oscillates many times, accelerating and decelerating the electron each cycle until it has a collision. This relaxation regime in the local limit is where $\omega \gg 1/\tau$, $\sigma_1 \ll \sigma_2$, and $\delta_{\rm cl} = c/\omega_p$, (Eq. 90). As the mean free path becomes longer than the skin depth, many electrons will enter the skin, be accelerated and decelerated each cycle of the field, collide once with an impurity or the surface, and exit the skin. The dissipation in this "anomalous relaxation" regime is less than when the mean free path is short, but the differences in the values of the skin depth in the local relaxation and anomalous relaxation regimes are small.

Figure 68 shows a roadmap[106] of the skin effects in the ω-τ plane. The horizontal axis is the frequency ω while the vertical axis is the scattering time τ. For low frequencies and short relaxation times, one is in the normal skin regime (Region A). If the mean free path and relaxation time get longer and longer, one eventually crosses the boundary given by Eq. 235 and is in the anomalous skin regime (Region D). If instead, I keep τ short but increase the frequency, I enter the relaxation regime, unless τ is really short, in which case the metal becomes partially transparent (Region C), because ϵ_1 is positive.* At high frequencies and in clean metals the relaxation regime becomes the anomalous relaxation regime (Region

---

* Write $\epsilon_1 = \epsilon_c - \omega_p^2/(\omega^2 + 1/\tau^2) = 0$. If $1/\tau$ is a large frequency, ϵ_1 crosses zero below the usual $\omega = \omega_p/\epsilon_c$.

E). Finally, if $\omega > \omega_p/\epsilon_c$ the metal is partially transparent. Figure 69 shows[106] electron trajectories in four of the five regimes. (In region C the skin depth is very large, so that the trajectories are as in A.)

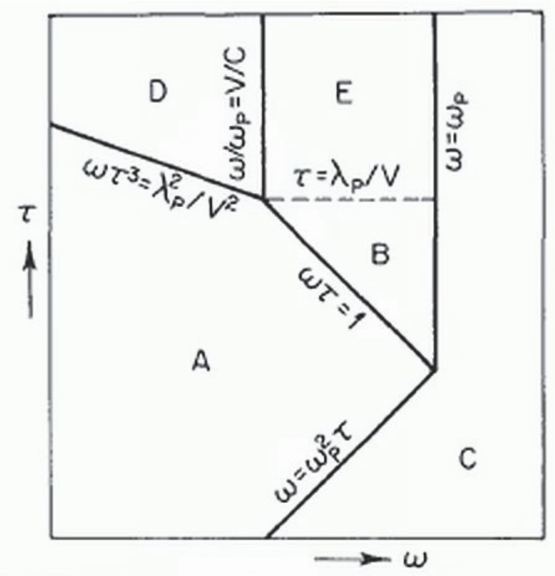

Fig. 68. Roadmap[106] of the skin effects in the ω-τ plane.

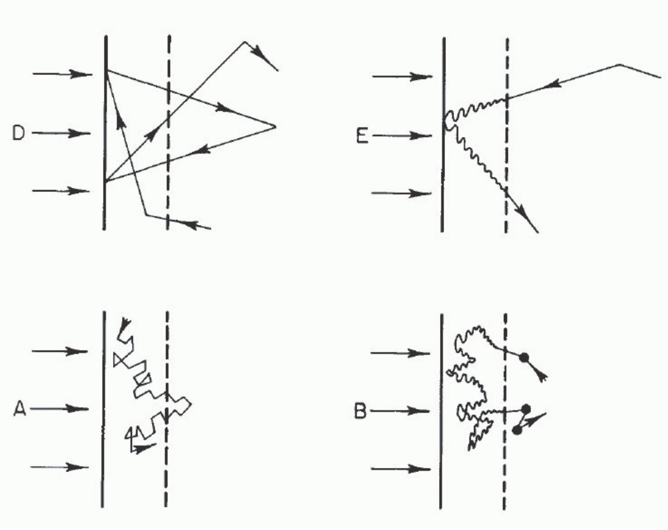

Fig. 69. Sketch[106] of electron trajectories in region A (normal skin effect), region B (relaxation regime), region D (anomalous skin effect), and region E (anomalous relaxation).

A larger version of Fig. 68, with numbers and the boundaries for three metals is shown in Fig. 70. The boundaries are sketched for (a) cesium , (b) constantan* and (c) copper. The boundaries are offset because the plasma frequencies differ with that for cesium being quite a bit smaller than the other two metals. The values of τ for copper at 4 K and at 300

---

* Constantan is an alloy containing 55% copper and 45% nickel.

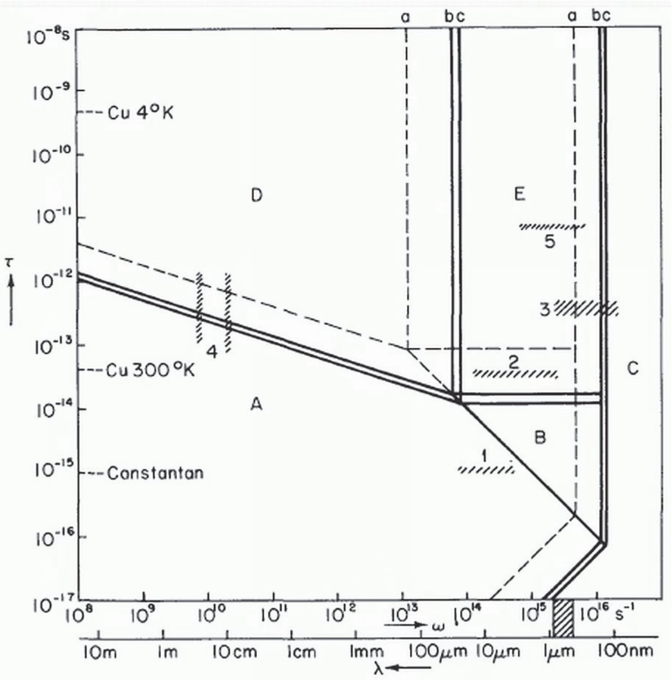

Fig. 70. Roadmap[106] of the skin effects in the ω-τ plane, calculated for (a) cesium, (b) copper, and (c) constantan.

K and for constantan[*] are shown.

12.5 The surface impedance

I want to find a material-related function to describe the optical properties of a non-local system. It of course should be also defined and reasonable for the local limit where Ohm's law is valid. Moreover is should be related to observables such as the absorption or reflectance. The skin depth is such a quantity but it specifies the behavior of the fields inside the metal, and I would like something that can be related to measured quantities outside the metal. The surface impedance $Z = R + iX$ is one such measure. As I will discuss, its real part, the surface resistance R, is related to loss in the metal and the surface reactance, X, to dispersion.

I'll start by reminding myself of the resistance and reactance of an electrical circuit. At DC, the resistance is given by freshman physics (Ohms law No. 2): $R = V/I$ where V is the voltage drop of the circuit and I is the current passing through it.[†] At finite frequencies, I must replace the real R with a complex Z. Now, if the current is organized into a surface sheet, with surface current K, then the same law is[2] $Z = E/K$. The skin is such a sheet, except that the current is not in an infinitesimally thick layer but in a layer a few times the

---

[*] The value is good at any temperature; the alloy's name is derived from "constant" because the resistivity is almost independent of temperature.

[†] #1: $V = IR$; #3: $I = V/R$. ☺

skin depth thick. Then*

$$K = \int_0^\infty j(x)dx$$

and

$$Z = \frac{E(0)}{\int_0^\infty j(x)dx}, \tag{236}$$

where $E(0)$ is the field at the surface.

Wherever the normal skin effect is in effect, I can use Ohm's law $j = \sigma E$. I also know that the field goes as $E = E(0)e^{-x/\delta}$. Then

$$Z = \frac{E(0)}{\sigma E(0) \int_0^\infty e^{-x/\delta} dx}. \tag{237}$$

The integral is easy to do; it integrates to $-(\delta)e^{-x/\delta}$. This contributes zero at the upper limit and δ at the lower limit. Hence

$$Z = \frac{1}{\sigma\delta}. \tag{238}$$

At low frequencies, $\sigma = \sigma_1 = \sigma_{\rm dc}$ and the surface impedance is real: $R = 1/\sigma_{\rm dc}\delta$ is the resistance of a slab of metal with thickness δ and equal length and width. This is often called the square resistance or the sheet resistance $R_\square$. At higher frequencies, both σ and δ are complex, and thus so is Z.

Using Eq. 89 I find (low frequencies!)

$$Z = \sqrt{\frac{2\pi\omega}{\sigma_{\rm dc}c^2}} \quad \text{when} \quad \omega \ll 1/\tau. \tag{239}$$

The surface impedance grows as $\omega^{1/2}$ and decreases if the conductivity increases.†

12.5.1 The general case (at low frequencies)

At frequencies low enough that I can neglect the inductive response (making $\sigma_1 \gg \sigma_2$) and the displacement current, Maxwell's fourth equation (Eq. 4d) is

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{j}. \tag{240}$$

Now look at the geometry of Fig. 71. The electromagnetic wave is incident from the left, normal to the surface, with **k** and **H** in the plane of the paper and **E** into the paper. As usual, I will take $\hat{\boldsymbol{k}}$, $\hat{\boldsymbol{e}}$, and $\hat{\boldsymbol{h}}$ respectively along the x, y, and z Cartesian axes. After the wave crosses the surface, it will decay with a skin depth δ (equation for and magnitude of to be specified). The current **j** (which also decays with decay length δ) is into the plane of the paper, parallel to **E**. A rectangular surface is shown in light blue. It is bounded by the 4 line segments ①, ②, ③, and ④. ③ is very many skin depths below the surface, where fields are zero.

---

* Fields and currents are vectors, but if I take the ratio of two of these, I must take the magnitudes. For isotropic and homogeneous materials, the current and electric field are of course in the same direction so I just take ratios of the amplitudes.

† But only as the square root of the conductivity. At DC the resistance decreases linearly in the conductivity.

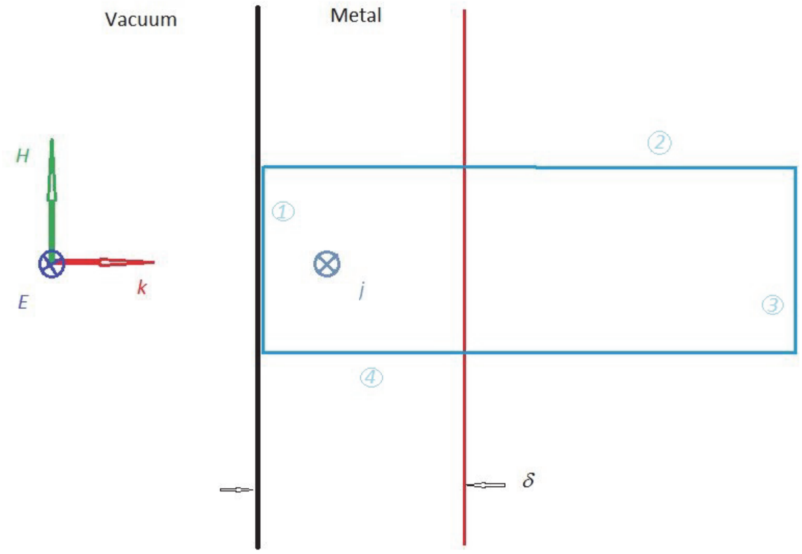

Fig. 71. The skin depth. The four line segments shown will be where I convert a surface integral to a line integral. Segment ③ is deep enough that the field is zero at its location.

So I will integrate Eq. 240 over the area of this surface using $d\mathbf{S} = \hat{\boldsymbol{y}} dz\, dx$

$$\begin{aligned}\int_S d\mathbf{S}\cdot\nabla\times\mathbf{H} &= \frac{4\pi}{c}\int_S d\mathbf{S}\cdot\mathbf{j}.\\ &= \frac{4\pi}{c}\int dz\int dx\,\hat{\boldsymbol{y}}\cdot\mathbf{j}.\end{aligned}$$

Using Eq. 237 I can evaluate easily the right-hand side:

$$\begin{aligned}\int_S d\mathbf{S}\cdot\nabla\times\mathbf{H} &= \frac{4\pi}{c}\int dz\,\frac{E(0)}{Z}\\ &= \frac{4\pi}{c}\frac{E(0)}{Z}L_{①},\end{aligned}$$

where $L_{①}$ is the length of line segment ①. But also I can transform the surface integral on the left side to a line integral around the perimeter of the surface $\mathbf{S}$ using Stokes theorem! It becomes

$$\begin{aligned}\int_S d\mathbf{S}\cdot\nabla\times\mathbf{H} &= \oint\mathbf{H}\cdot d\mathbf{l}\\ &= \int_{①} dz\,H(0) + \int_{②} dx\,\mathbf{H}\cdot\hat{\boldsymbol{x}} - \int_{③} dz\,\mathbf{H}(x\to\infty) - \int_{④} dx\,\mathbf{H}\cdot\hat{\boldsymbol{x}}\\ &= H(0)L_{①}.\end{aligned}$$

Here, the differential vector $d\mathbf{l}$ runs clockwise around the perimeter, following ① … ④.

The dot product of **H** and $d\mathbf{l}$ makes the integrals ② and ④ zero; the magnetic field **H** is zero along ③. Hence, the surface impedance is

$$Z = \frac{4\pi}{c}\frac{E(0)}{H(0)} \tag{241}$$

the ratio of the electric field at the surface to the magnetic field at the surface.

12.5.2 Impedance of free space

Suppose I replace the metal with vacuum. The fields are still defined in the y-z plane. In vacuum $E = H$ so the impedance of free space is

$$Z_0 = \frac{4\pi}{c} = 377\,\Omega \tag{242}$$

The first value, $4\pi/c = 4.189 \times 10^{-10}$ sec/cm, is the esu value; the second, $377\,\Omega$, is SI. Their ratio is $30c$.*

12.5.3 Fields at the surface

I'll start by looking back at the discussion of normal-incidence reflection on page 21. I know that the electric field is reversed on reflection (because the metal is the more dense medium). Thus, with the incident field along $\hat{\boldsymbol{y}}$, the reflected field is along $-\hat{\boldsymbol{y}}$. The reflectance is nearly 100% in good metals, so the resultant field is $1-(-r)$ with the amplitude reflectivity r given in Eq. 43.† Because the **E** fields are antiparallel, the **H** fields are parallel, adding to $1 + (-r)$, larger than then incident **H** field by almost a factor of two. Thus the surface impedance is much smaller than $377\,\Omega$. Indeed, from above I get

$$Z = \frac{4\pi}{c}\frac{1+r}{1-r}$$

with $r = (1-N)/(1+N)$. Substituting, I find

$$Z = \frac{4\pi}{c}\frac{1}{N} \tag{243}$$

But I knew this! The magnitude of the magnetic field inside the metal (or any local, nonmagnetic, linear, homogeneous, and isotropic material) is $H = NE$. The tangential components are continuous at the interface, so the fields at the surface are the same as the fields just inside the metal.

---

* See Appendix A.

† r is negative (the field is reversed) so I write it as $1-(-r)$ to indicate a difference between unity and a number just above -1, e. g., -0.99. The difference is 0.01.

12.5.4 Back to the dielectric function

Because $N = \sqrt{\epsilon}$, I can write $Z = (4\pi/c)\sqrt{1/\epsilon}$. Moreover, I can write $\epsilon = \epsilon_c + 4\pi i\sigma/\omega$ and at low frequencies (compared to ω_p) find that

$$Z = \frac{4\pi}{c}\sqrt{\frac{\omega}{4\pi i\sigma}} \tag{244}$$

Now $\sqrt{1/i} = (1-i)/\sqrt{2}$, so I can also write

$$Z = R + iX = (1-i)\sqrt{\frac{2\pi\omega}{c^2\sigma}} \tag{245}$$

At low frequencies $\sigma \approx \sigma_{\mathrm{dc}}$ and $\delta = \delta_{\mathrm{cl}}$ and*

$$R = -X = \frac{1}{\sigma_{\mathrm{dc}}\delta_{\mathrm{cl}}}.$$

12.5.5 The surface impedance in the anomalous regime

The same approach works for the anomalous skin effect:

$$R = \frac{1}{\sigma_{\mathrm{eff}}\delta_{\mathrm{ASE}}}$$

with $\sigma_{\mathrm{eff}} = ne^2\delta_{\mathrm{ASE}}/mv_{\mathrm{F}}$ and $\delta_{\mathrm{ASE}} = (c^2mv_{\mathrm{F}}/2\pi ne^2\omega)^{1/3}$. Note that δ_{ASE} comes in twice, once directly and once replacing the mean free path in σ_{eff}. Then

$$R = \frac{mv_{\mathrm{F}}}{ne^2}\left(\frac{2\pi ne^2\omega}{c^2mv_{\mathrm{F}}}\right)^{2/3}.$$

or, after some algebra

$$R = \left(\frac{(2\pi)^2mv_{\mathrm{F}}\omega^2}{ne^2c^4}\right)^{1/3}. \tag{246}$$

So the anomalous-limit surface impedance goes as $\omega^{2/3}$. It can be rewritten as:

$$R = Z_0\left(\frac{v_{\mathrm{F}}}{4c}\right)^{1/3}\left(\frac{\omega}{\omega_p}\right)^{2/3}. \tag{247}$$

with $Z_0 = 4\pi/c = 377\,\Omega$, according to whether you want cgs or SI units.

---

* Consider a thin film with thickness $d \ll \delta$. Then the currents are confined to a distance d not a distance δ, and

$$Z = \frac{1}{\sigma d} = \frac{\rho}{d} \equiv Z_\square.$$

12.5.6 Reflectance in terms of the surface impedance

According to Eq. 44 the reflectance is

$$\mathscr{R} = \left| \frac{1 - N}{1 + N} \right|^2 .$$

Now $N = Z_0/Z$ so

$$\mathscr{R} = \left| \frac{Z - Z_0}{Z + Z_0} \right|^2 . \tag{248}$$

Eq. 248 is very familiar to the discussion of energy flow in waveguides and coaxial transmission lines. It says that a discontinuity in the line impedance causes a reflection in the line. The most extreme discontinuity is a short circuit ($Z = 0$) or an open circuit $Z = \infty$); both give 100% reflection.

12.5.7 Numerical values

I use as an example, the classical surface impedance (Eq. 245) with $\sigma_{\mathrm{dc}} = 6{\times}10^5\,\Omega^{-1}\mathrm{cm}^{-1}$ at 300 K. If the frequency is $\omega = 10^{10}$ s^{-1} ($0.05\,\mathrm{cm}^{-1}$), then $R = 4$ m$\,\Omega$. Compare this very low value with the impedance of free space, $Z_0 = 377\,\Omega$. The metal looks like a short circuit. Just as a shorted transmission line reflects the signals sent down it, the metallic surface is an excellent reflector.* Indeed, I can rewrite Eq. 248 as $\mathscr{R} = 1 - 4Z/Z_0 \sim 1 - 10^{-5}$.

---

* I find this transmission line analogy very useful in understanding why a high absorber is also a high reflector. The short circuit ought to draw very large currents and dissipate lots of power. It does not. Instead, it returns the electromagnetic energy back towards the source.

Appendix A. Notes about Units

First, let me start with common units used to measure frequency or photon energy.[*]

ω, angular frequency, in s^{-1}
f, frequency, in Hz $\quad (f = \omega/2\pi)$
λ_0, vacuum wavelength, in cm[†]
ν, wave number (or "frequency"), in cm^{-1} $\quad (\nu = f/c = 1/\lambda_0 = \omega/2\pi c)$
E, photon energy. in eV $\quad (E = hf = \hbar\omega = hc\nu)$
T, temperature, in K $\quad (T = E/k_B)$

To convert:

$1\,\mathrm{cm}^{-1} = 1.44\ \mathrm{K} = 1/8066\ \mathrm{eV} = 0.124\ \mathrm{meV} = 30\ \mathrm{GHz} = 1.88\times 10^{11}\ \mathrm{s}^{-1}$
$1\ \mathrm{eV} = 8066\,\mathrm{cm}^{-1} = 11{,}600\ \mathrm{K} = 242\ \mathrm{THz} = 1.52\times 10^{15}\ \mathrm{s}^{-1} = 1/1.24\ \mu\mathrm{m}$
red $= 2$ eV$= 16{,}000\,\mathrm{cm}^{-1} = 1/(620\ \mathrm{nm})$

I am using cgs-Gaussian units for electromagnetic quantities.

t, time, in seconds (s)
ℓ, length, in cm
$c = 3 \times 10^{10}$ cm/s
E field in statvolts/cm (but irrelevant, usually)
H field in gauss or oersted (but irrelevant, usually)
m, mass, in grams (g)
h and $\hbar$, Planck's constant, in erg-s[‡]
e, charge, in esu

The dielectric function, ϵ, permeability, μ, and refractive index, N, are dimensionless. The principal *measured* quantities that have units are the absorption coefficient, α, penetration depth, δ, and conductivity, σ.

α, absorption coefficient, in cm^{-1} (even in SI, usually)
δ, penetration depth, in cm or microns or Å or even meters
σ, conductivity, is in s^{-1}, i. e., it is a *frequency*

In practical units, σ is measured in $\Omega^{-1}\mathrm{cm}^{-1}$,[§] and

$\rho = 1/\sigma$ is in Ohm-cm or Ω-cm

To convert:

$$\boxed{\sigma_{\mathrm{prac}} = \sigma_{\mathrm{esu}}/30c = \sigma_{\mathrm{esu}}/9 \times 10^{11}}$$

For copper at 300 K, $\rho \approx 1.6\ \mu\Omega$-cm, so $\sigma \approx 6 \times 10^5\ \Omega^{-1}\mathrm{cm}^{-1} \approx 6 \times 10^{17}\ \mathrm{s}^{-1}$

Note that the *equation* for conductivity is the same in both systems; in the Drude model at zero frequency it is given by Eq. 55:

$$\sigma_{dc} = ne^2\tau/m.$$

The factor of 9×10^{11} comes from using in cgs (SI) the following: esu (Coulombs) for the

---

[*] If, like the particle physicists, I set $\hbar = 1$, then (angular) frequency and energy are identical.

[†] The wavelength in a medium—even air—is of course different: $\lambda = \lambda_0/n$.

[‡] But it even more common to use electron volts, so that $\hbar = 6.58 \times 10^{-16}$ eV-s.

[§] To be fully SI-ized, one should report σ in $\mathrm{Ohm}^{-1}\mathrm{m}^{-1}$, and some do. There is a factor of 100 between $\Omega^{-1}\mathrm{cm}^{-1}$ and $\Omega^{-1}\mathrm{m}^{-1}$. I'll let the reader figure out which way it goes.

charge, grams (kg) for the mass, cm^{-3} (m^{-3}) for the carrier density, and then using the factor of 100 between between Ω^{-1}m^{-1} and Ω^{-1}cm^{-1}.

The equations for the plasma frequency are *not* the same. In cgs (see Eq. 58) it is

$$\omega_p = \sqrt{\frac{4\pi n e^2}{m}},$$

whereas in SI it is

$$\omega_p = \sqrt{\frac{n e^2}{\epsilon_0 m}},$$

with $\epsilon_0 = 8.854 \times 10^{-12}$ Farad/m the permittivity of the vacuum.*

Another quantity that one might want to calculate is the classical skin depth. According to Eq. 89, it is

$$\delta = \frac{c}{\omega\kappa} \approx \sqrt{\frac{c^2}{2\pi\omega\sigma_{\mathrm{dc}}}} = \sqrt{\frac{2c^2}{\omega_p^2\omega\tau}}.$$

where σ_{dc} is the dc conductivity, which in esu is measured in sec^{-1}. See also page 161 for more on the skin depth.

In SI, with the conductivity σ_{dc} in Ohm^{-1}m^{-1}, the skin depth is

$$\delta = \sqrt{\frac{2\epsilon_0 c^2}{\omega\sigma_{\mathrm{dc}}}}$$

where you would use $c = 3 \times 10^8$ m/s so that δ emerges in m. One may also use $\delta = \sqrt{2/\omega\sigma_{\mathrm{dc}}\mu_0}$ with $\mu_0 = 4\pi \times 10^{-7}$ N/A the permeability of free space, again with σ_{dc} in Ohm^{-1}m^{-1}.

---

[*] Persons who wonder about the reasons why the vacuum has a permittivity (and permeability) that differs from unity might want to read Casimir's discussion of the volumetric constant of empty space.[107]

Appendix B. Maxwell's equations in SI

Maxwell's equations look a bit different in SI. For example, Eq. 1 is simpler looking*

$$\nabla \cdot \mathbf{D} = \rho_{\text{ext}} \tag{249a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{249b}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{249c}$$

$$\nabla \times \mathbf{H} = \mathbf{j}_{\text{free}} + \frac{\partial \mathbf{D}}{\partial t} \tag{249d}$$

The Lorentz law (Eq. 2) is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{250}$$

where v is the particle's velocity.†

In vacuum, the SI version of Eq. 3 becomes $\mathbf{D} = \epsilon_0 \mathbf{E}$ and $\mathbf{B} = \mu_0 \mathbf{H}$, where ϵ_0 and μ_0 are the permittivity and permeability of vacuum, with values

$$\epsilon_0 = 8.854 \times 10^{-12} \quad \text{Farad/m}$$
$$\mu_0 = 4\pi \times 10^{-7} \quad \text{Henry/m.}$$

In vacuum, one finds plane-wave solutions, with velocity

$$\frac{1}{\sqrt{\epsilon_0 \mu_0}} \equiv c.$$

In macroscopic media, the auxiliary fields are

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{E} + \mathbf{P} \\ \mathbf{H} &= \frac{1}{\mu_0} \mathbf{B} - \mathbf{M} \end{aligned} \tag{251}$$

which, in local, linear, isotropic, and homogeneous media, become

$$\mathbf{D} = \epsilon_r \epsilon_0 \mathbf{E}$$
$$\mathbf{B} = \mu_r \mu_0 \mathbf{H}$$

with ϵ_r and μ_r dimensionless numbers, called respectively the relative permittivity aand relative permeability, identical to the dielectric constant ϵ and permeability μ in cgs-Gaussian units.

The SI energy density is

$$u = (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H})/2.$$

(In vacuum, $u = (\epsilon_0 E^2 + \mu_0 H^2)/2$.) The Poynting vector is

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}.$$

---

[*] Although I think that it is nice to see the speed of light appearing in the equations of electromagnetism.

[†] This one gives no indication that speeds need to approach lightspeed before magnetic forces and electric forces become comparable (for, say, fields with the same electric and magnetic energy densities).

Appendix C. Partial derivatives and vector operators acting on plane waves

Let's write a complex plane wave $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{A}_0 e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} = (A_{0x}\hat{\boldsymbol{x}} + A_{0y}\hat{\boldsymbol{y}} + A_{0z}\hat{\boldsymbol{z}})\, e^{iq_x x} e^{iq_y y} e^{iq_z z} e^{-i\omega t}$$

where $\mathbf{A}_0$ is a constant vector specifying the direction, magnitude, and phase of $\mathbf{A}$, and the wave vector $\mathbf{q} = q_x\hat{\boldsymbol{x}} + q_y\hat{\boldsymbol{y}} + q_z\hat{\boldsymbol{z}}$ is normal to the planes of constant phase.

Now take the partial with respect to time of $\mathbf{A}$. The only term containing the time in the vector is $e^{-i\omega t}$ and

$$\frac{\partial}{\partial t} e^{-i\omega t} = -i\omega e^{-i\omega t}$$

so that

$$\frac{\partial}{\partial t}\mathbf{A} = -i\omega\mathbf{A}$$

Now, I want to act on $\mathbf{A}$ with the vector operator ∇ where

$$\nabla = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}.$$

First, I'll compute the divergence $\nabla\cdot\mathbf{A}$. If I do this, and write out the dot product of ∇ with $\mathbf{A}_0$, I get

$$\nabla\cdot\mathbf{A} = (A_{0x}\frac{\partial}{\partial x} + A_{0y}\frac{\partial}{\partial y} + A_{0z}\frac{\partial}{\partial z})\, e^{iq_x x} e^{iq_y y} e^{iq_z z} e^{-i\omega t}$$

Each partial derivative only acts on one of the complex exponentials in $\mathbf{A}$, so that

$$\nabla\cdot\mathbf{A} = (iq_x A_{0x} + iq_y A_{0y} + iq_z A_{0z})\, e^{iq_x x} e^{iq_y y} e^{iq_z z} e^{-i\omega t}$$

or

$$\nabla\cdot\mathbf{A} = i\mathbf{q}\cdot\mathbf{A}$$

In a similar way, I can show that

$$\nabla\times\mathbf{A} = i\mathbf{q}\times\mathbf{A},$$

$$\nabla^2\mathbf{A} = -q^2\cdot\mathbf{A},$$

and

$$\nabla A = i\mathbf{q}A.$$

In the last of these A is a scalar.

The great thing about the complex exponential solutions to Maxwell's equations is that they allow me to replace differential equations with algebraic ones.

Appendix D. A field guide to optical "constants"

What follows is a list and brief discussion of the optical functions used in studying the optical properties of solids.

1. `Sigma_1`

The complex conductivity and dielectric function are related. In cgs the conductivity and frequency are both measured in sec^{-1} and the relation is

$$\epsilon = 1 + \frac{4\pi i \sigma}{\omega}, \tag{252}$$

whereas in SI the conductivity is in $\Omega^{-1}\text{cm}^{-1}$, and it is

$$\epsilon = 1 + \frac{i\sigma}{\omega\epsilon_0}. \tag{253}$$

with ϵ_0 the dielectric function of empty space.

In lab units, frequencies are in cm^{-1} and conductivities are in $\Omega^{-1}\text{cm}^{-1}$, and the relation is

$$\epsilon = 1 + \frac{60 i\sigma}{\omega}, \tag{254}$$

The real part of the optical conductivity (in $\Omega^{-1}\text{cm}^{-1}$) is then

$$\sigma_1 = \frac{\omega\epsilon_2}{60}, \tag{255}$$

with ω in cm^{-1}. This equation[*] can be compared to the expression in cgs:

$$\sigma_1 = \frac{\omega\epsilon_2}{4\pi}, \tag{256}$$

with ω in sec^{-1} and σ_1 also in sec^{-1}. The cgs conductivity is 9×10^{11} times larger than the SI conductivity.

2. `Sigma_2`

The imaginary part of the optical conductivity (in $\Omega^{-1}\text{cm}^{-1}$) is

$$\sigma_2 = -\frac{\omega(\epsilon_1 - 1)}{60}. \tag{257}$$

If you wish to calculate from a partial dielectric function to eliminat ϵ_c and/or other contributions:

$$\sigma_2 = -\frac{\omega(\Delta\epsilon_1)}{60}. \tag{258}$$

---

[*] The same factor of 60 appears if you calculate σ_{dc} from the plasma frequency and relaxation rate,

$$\sigma_{\text{dc}} = \frac{\omega_p^2}{60(1/\tau)}$$

with ω_p and $1/\tau$ in cm^{-1} and σ_{dc} in $\Omega^{-1}\text{cm}^{-1}$. In reality, $60 = 377\ \Omega/2\pi$.

3. `Epsilon_1`

$$\epsilon_1 = \mathrm{Re}(\epsilon) \tag{259}$$

4. `Epsilon_2`

$$\epsilon_2 = \mathrm{Im}(\epsilon) \tag{260}$$

5. `Refr Index`

N is calculated from reflectance and phase via the inversion of

$$\sqrt{\mathscr{R}}e^{i\phi} = \frac{N-1}{N+1} \tag{261}$$

and

$$n = \mathrm{Re}(N) \tag{262}$$

6. `Ext Coeff.`

The imaginary part of N is

$$\kappa = \mathrm{Im}(N) \tag{263}$$

7. `Alpha`

The absorption coefficient (in cm^{-1}) is

$$\alpha = 4\pi\omega\kappa, \tag{264}$$

correct for ω in cm^{-1}. In cgs, this is

$$\alpha = 2\omega\kappa/c. \tag{265}$$

where ω is in sec^{-1}.

8. `Skin depth`

In the skin effect the electric field goes as $e^{-x/\delta}$ but of course the wave governed by a refractive index N has a field which decays as $e^{-\kappa\omega x/c}$ (ω in rad/s). Hence:

$$\delta = \frac{10^8}{2\pi\omega\kappa}, \tag{266}$$

with ω in cm^{-1}. The factor of 10^8 makes the units of $\delta(\omega)$ in Å.

9. `Loss Fnctn`

The energy loss function governs the transmission of fast electrons through the material. These depend on the longitudinal response of the dielectric. It is typically written as $-\mathrm{Im}(1/\epsilon)$; here I use L.

$$L = \frac{\epsilon_2}{(\epsilon_1^2 + \epsilon_2^2)} = -\mathrm{Im}\,\frac{1}{\epsilon}. \tag{267}$$

10. `Surface Loss Fnctn`

The "surface" energy loss function is

$$L_s = \frac{\epsilon_2}{((\epsilon_1 - 1)^2 + \epsilon_2^2)} = -\operatorname{Im} \frac{1}{\epsilon - 1}. \tag{268}$$

Sum rules

11. `Sumrule S_1`

This is the most important sumrule function. The f-sum rule for solids can be written

$$\int_0^c \sigma_1(\omega')d\omega' = \frac{\omega_{p,tot}^2}{8}, \tag{269}$$

where $\omega_{p,tot} = \sqrt{4\pi n_{tot} e^2/m}$ is the plasma frequency for all the electrons in the solid. I may write a partial sum rule (for the conduction band, say) as

$$\int_0^\omega \sigma_1(\omega')d\omega' = \frac{\pi n e^2}{2m^*}, \tag{270}$$

where m^* is the average effective mass of the band. If the upper limit is above the free-carrier contribution and below the onset of the interband transitions, the right hand side becomes $\omega_p^2/8$ with ω_p the conduction band plasma frequency.

Next, write $n = N_{eff}/V_c$ and let m be the free electron mass. The density is in the number of effective electrons per unit cell volume (or formula volume). Then, you calculate

$$N_{eff}\frac{m}{m^*} = \frac{2mV_c}{\pi e^2}\int_0^\omega \sigma_1(\omega')d\omega' \tag{271}$$

with m the mass of an electron and m^* the effective mass of the charge carriers.

12. `Sumrule LossFn`

I may write a partial sum rule (for the conduction band, say) of the loss function as

$$\frac{1}{4\pi}\int_0^\omega \omega' L(\omega')d\omega' = \frac{\pi n e^2}{2m^*}, \tag{272}$$

where m^* is the average effective mass of the band. If the upper limit is above the free-carrier contribution and below the onset of the interband transitions, the right hand side becomes $\omega_p^2/8$ with ω_p the conduction band plasma frequency. (The factor of $1/4\pi$ is the same as appears in $\omega\epsilon_2/4\pi = \sigma_1$.) Thus,

$$N_{eff}\frac{m}{m^*} = \frac{mV_c}{2\pi^2 e^2}\int_0^\omega \omega' L(\omega')d\omega' \tag{273}$$

13. `Sumrule SurfaceLossFn`

Same as Eqs. 272 and 273, but for L_s.

Frequency-dependent scattering, self-energy, and effective mass functions

The dielectric function of perfectly free (noninteracting) carriers with density n, charge $\pm e$ and mass m is given by

$$\epsilon(\omega) = \epsilon_c - \frac{\omega_p^2}{\omega^2} \tag{274}$$

with ω_p, the plasma frequency (in cgs)

$$\omega_p = \sqrt{\frac{4\pi n e^2}{m}}. \tag{275}$$

If now I allow scattering of free carriers from impurities, defects, surfaces, and lattice oscillations,[*] with a mean free time τ or scattering rate $1/\tau$, independent of the frequency, the dielectric function has the Drude form:

$$\epsilon(\omega) = \epsilon_c - \frac{\omega_p^2}{\omega^2 + i\omega/\tau} \tag{276}$$

Now, if I turn on interactions, the scattering rate becomes a function of frequency (and probably also of temperature). However, to allow the scattering rate, $1/\tau$ to vary with frequency and remain real quantity would cause the dielectric function (and conductivity) to violate the Kramers-Kronig relation. One may accommodate this either by considering the scattering rate to be a complex function or by also having the effective mass m^* to vary with frequency.

Theorists have devised a number of such functions: the complex self-energy function Σ,[†] the complex scattering (or memory) function Γ, a complex scattering rate $1/\tau^*$ and effective mass m^*, and a complex scattering rate $1/\tau^*$ and mass enhancement factor λ. you can take the KK-derived dielectric function and compute each and all of these quantities.

`14. -Imag self-energy (one-particle)`

In some theories (such as the marginal Fermi liquid) the imaginary part of the self energy Σ plays the role of a scattering rate, with the dielectric function written as

$$\epsilon(\omega) = \epsilon_c - \frac{\omega_p^2}{\omega[\omega - \Sigma(\omega)]}. \tag{277}$$

with ω_p the (unrenormalized)[‡] plasma frequency and ϵ_c the limiting high frequency value of the dielectric function. The imaginary part of Σ is related to the quasiparticle lifetime through $1/\tau^*(\omega) = -\operatorname{Im}\Sigma(\omega)(m_b/m^*(\omega))$ with m^* the effective mass and m_b the band mass.

---

[*] In principal, coupling to phonons scatters the charge carriers by either absorption or emission of a phonon. In turn, these processes lead to a frequency and temperature-dependent scattering rate. For this discussion, the frequency dependence is ignored, as are the emission processes. Thus, I consider only the high-temperature limit.

[†] There are actually two of these.

[‡] Unrenormalized means "what it would be if there were no interactions."

You compute $-\operatorname{Im}\Sigma$ via

$$-\operatorname{Im}\Sigma(\omega) = -\frac{\omega_p^2}{\omega}\operatorname{Im}\frac{1}{\Delta\epsilon}. \tag{278}$$

with $\Delta\epsilon = \epsilon - \epsilon_c$. You must know the plasma frequency and ϵ_c. Or, subtract a model dielectric function. Subtraction removes optical phonons and interband transitions which may occur.

`15. Real self-energy (one-particle)`

The real part of Σ is related to the effective mass m^* of the interacting carriers by

$$m^*(\omega)/m_b = 1 - \operatorname{Re}\Sigma(\omega)/\omega. \tag{279}$$

Hence, you compute

$$\operatorname{Re}\Sigma(\omega) = \frac{\omega_p^2}{\omega}\operatorname{Re}\frac{1}{\Delta\epsilon} + \omega. \tag{280}$$

`16. -Imag self-energy (two-particle)`

A slightly different approach recognizes that optical excitations are created in pairs (electron and hole) and that the self-energy should be a two-particle process. Still, the imaginary part of the self energy Σ plays the role of a scattering rate, with the dielectric function written as

$$\epsilon(\omega) = \epsilon_c - \frac{\omega_p^2}{\omega[\omega - 2\Sigma(\omega/2)]}. \tag{281}$$

with ω_p the (unrenormalized) plasma frequency and ϵ_c the limiting high frequency value. The factors of 2 occur because the excitations are created in pairs (electron-hole pairs or quasiparticle pairs). (These factors do not appear in all theoretical treatments. Indeed the "one-particle" versions are much more commonly seen.) The imaginary part of Σ is related to the quasiparticle lifetime through $1/\tau^*(\omega) = -2\operatorname{Im}\Sigma(\omega/2)(m_b/m^*(\omega))$ with m^* the effective mass and m_b the band mass.

You compute $-\operatorname{Im}\Sigma$ via

$$-\operatorname{Im}\Sigma = -\frac{\omega_p^2}{2\omega}\operatorname{Im}\frac{1}{\Delta\epsilon}. \tag{282}$$

with $\Delta\epsilon = \epsilon - \epsilon_c$. You must enter the plasma frequency and ϵ_c.

After computing the right hand side of the above equation, you should divide the frequencies by 2, so that the dielectric function value at 400 cm^{-1} (for example) generates the $-\operatorname{Im}\Sigma$ value at 200 cm^{-1}.

Note that optical phonons and interband transitions may occur; you are allowed you to subtract these from ϵ.

17. `Real self-energy (two-particle)`

The real part of Σ is related to the effective mass m^* of the interacting carriers by

$$m^*(\omega)/m_b = 1 - 2\,\mathrm{Re}\,\Sigma(\omega/2)/\omega. \tag{283}$$

Hence, you compute

$$\mathrm{Re}\,\Sigma(\omega/2) = \frac{\omega_p^2}{2\omega}\,\mathrm{Re}\,\frac{1}{\Delta\epsilon} + \omega/2. \tag{284}$$

After computing the right hand side of the above equation, you divide the frequencies by 2, so that the dielectric function value at $400\,\mathrm{cm}^{-1}$ (for example) generates the $\mathrm{Re}\,\Sigma$ value at $200\,\mathrm{cm}^{-1}$.

18. `Gamma_1`

The frequency dependent and complex damping is often written in an extended Drude model by replacing $1/\tau$ by a "memory function" Γ. I write

$$\epsilon = \epsilon_c - \frac{\omega_p^2}{\omega(\omega + i\Gamma)} \tag{285}$$

with $\Gamma = \Gamma_1 + i\Gamma_2$.

$$\Gamma_1 = -\frac{\omega_p^2}{\omega}\,\mathrm{Im}\,\frac{1}{\Delta\epsilon}. \tag{286}$$

Note that $\Gamma_1 = -\Sigma_2$ (one-particle).

19. `Gamma_2`

The imaginary part of Γ is

$$\Gamma_2 = \frac{\omega_p^2}{\omega}\,\mathrm{Re}\,\frac{1}{\Delta\epsilon} + \omega \tag{287}$$

Note that $\Gamma_2 = +\Sigma_1$ (one-particle).

20. `tau^-1*`

$$1/\tau = -\omega\frac{\mathrm{Im}(\Delta\epsilon)}{\mathrm{Re}(\Delta\epsilon)} \tag{288}$$

21. `m*/m`

$$\frac{m^*}{m} = -\frac{\omega_p^2}{\omega^2}\,\mathrm{Re}(\frac{1}{\Delta\epsilon}) \tag{289}$$

`22. Mass enhancement factor` $\lambda(\omega)$

The effective mass may be written in terms of a mass enhancement factor $\lambda(\omega)$:

$$\frac{m^*}{m} = 1 + \lambda(\omega). \tag{290}$$

Hence

$$\lambda(\omega) = -\frac{\omega_p^2}{\omega^2}\,\mathrm{Re}(\frac{1}{\Delta\epsilon}) - 1 \tag{291}$$

Specialized functions, mostly for superfluids

`23. Superfluid plasma freq`

A superfluid has a dielectric function that is

$$\epsilon = \epsilon_c - \frac{\omega_{ps}^2}{\omega^2} \tag{292}$$

with ω_{ps} the superfluid plasma frequency. you calculate

$$\omega_{ps} = \sqrt{\omega^2|\,\mathrm{Re}(\Delta\epsilon)|} \tag{293}$$

where the absolute value is used instead of a minus sign to avoid numerical problems. Be sure that ϵ is negative when you invoke this function!

If ω_{ps} is (nearly) constant in frequency, the behavior is that of a superfluid.

`24. N_eff (free-el)`

Again, writing

$$\epsilon = \epsilon_c - \frac{\omega_{ps}^2}{\omega^2} \tag{294}$$

with ω_{ps} the superfluid plasma frequency. You then write $\omega_{ps}^2 = 4\pi n_s(\omega)e^2/m$ so that:

$$n_s(\omega) = \omega^2 V_c C|\,\mathrm{Re}(\Delta\epsilon)| \tag{295}$$

where C is a numerical constant.

$n_s(\omega)$ should be nearly constant as a function of frequency for the concept to be valid.

`25. London penetration depth`

The dielectric function in Eqs. 292 or 294 is that of a London superconductor, so one may also calculate λ_L as

$$\lambda_L = \frac{10^8}{2\pi\omega\sqrt{|\,\mathrm{Re}(\Delta\epsilon)|}} \tag{296}$$

with ω in cm^{-1}. The factor of 10^8 makes the units of $\lambda_L(\omega)$ in Å.

$\lambda_L(\omega)$ should be nearly constant as a function of frequency for the concept to be valid.

Other specialized functions

26. `Epsilon_1 vs 1/w^2`

If

$$\epsilon = \epsilon_\infty - \frac{\omega_{ps}^2}{\omega^2} \tag{297}$$

then a plot of ϵ vs $1/\omega^2$ should be a straight line, with slope $-\omega_{ps}^2$. So this is equivalent to function 23.

27. `1/Sigma_1 vs w^2`

If the dielectric function is Drude like, then

$$\sigma_1(\omega) = \frac{\sigma_{dc}}{1+\omega^2\tau^2} \tag{298}$$

so you can make a plot of $1/\sigma_1$ vs. ω^2. The plot function is

$$\frac{1}{\sigma_1(\omega)} = \frac{1}{\sigma_{dc}} + \frac{\tau^2}{\sigma_{dc}}\omega^2$$

This plot should be a straight line with intercept $1/\sigma_{dc}$ and slope τ^2/σ_{dc}. If needed, σ_{dc} can be read off the plot of $\sigma_1(\omega)$ assuming that the data extend well below $1/\tau$. Otherwise one can write the slope as $4\pi\tau/\omega_p^2$ and get ω_p from the plot of ϵ_1 vs. $1/\omega^2$.

28. `Resistivity (real)`

The optical resistivity is just the inverse of the optical conductivity, so the real part is

$$\rho_1 = \mathrm{Re}(\frac{1}{\sigma}). \tag{299}$$

Because σ is complex, $\rho_1 = \sigma_1/(\sigma_1^2+\sigma_2^2)$.

29. `Resistivity (imag)`

The imaginary part if the optical resistivity is

$$\rho_2 = \mathrm{Im}(\frac{1}{\sigma}). \tag{300}$$

30. `Abs(eps)`

There are examples where the magnitude of the complex dielectric function $|\epsilon|$ displays power law behavior. Hence you calculate

$$|\epsilon| = \sqrt{\epsilon_1^2+\epsilon_2^2} \tag{301}$$

Note that this should be the entire function, so ϵ_c is not automatically subtracted.

31. `Abs(sigma)`

This is

$$|\sigma| = \sqrt{\sigma_1^2 + \sigma_2^2} \tag{302}$$

32. `Tan(delta)`

The loss tangent is commonly used in the analysis of RF and microwave experiments. It is

$$\tan\delta = \frac{\epsilon_2}{\epsilon_1}. \tag{303}$$

This calculation uses the entire function, so ϵ_c is not automatically subtracted.

33. `Abs(Tan(delta))`

If you want this to be always positive, use

$$\tan\delta_{abs} = \left|\frac{\epsilon_2}{\epsilon_1}\right|. \tag{304}$$

Transmission and reflection

You use the complex dielectric function (and the thickness, if needed) to calculate reflectance and transmittance. For a sample of finite thickness, you have a choice to treat it as "thick" (i. e., incoherent) or "thin" (coherent). In the first case, you add intensities of multiple internal reflections while in the second you add amplitudes. The second case allows for interference, with fringes in the transmittance or reflectance.

34. `Reflectance`

The power reflectance (single-bounce reflectance) is

$$\mathscr{R} = \left|\frac{N-1}{N+1}\right|^2. \tag{305}$$

35. `T slab (thick)`

Add intensities.

Compute

$$\mathscr{T} = \frac{(1-\mathscr{R}_{sb})^2\, e^{-\alpha d}}{1-\mathscr{R}_{sb}^2\, e^{-2\alpha d}} \tag{306}$$

with d the thickness, α the absorption coefficient, and $\mathscr{R}_{sb}$ the single bounce reflectance. $\mathscr{R}_{sb} = [(N-1)/(N+1)]^2$.

36. `T slab (fringes)`

Add amplitudes.

To compute the coherent transmittance, you first calculate the phase gain on passing the sample (one way),

$$\delta = 2\pi\omega N d, \tag{307}$$

the amplitude transmitted *into* the slab, from vacuum,

$$t_1 = \frac{2}{N+1}, \tag{308}$$

the amplitude transmitted *out of* the slab,

$$t_2 = \frac{2N}{N+1}, \tag{309}$$

and the amplitude reflection at the interfaces, incident from within

$$r_{in} = \frac{N-1}{N+1}. \tag{310}$$

Then the formula for the transmission coefficient (amplitude) is

$$t = \frac{t_1 t_2 e^{i\delta}}{1 - r_{in}^2 e^{2i\delta}} \tag{311}$$

and to get the transmittance, you calculate

$$\mathscr{T} = tt^* \tag{312}$$

This is the Airy formula.

`37.  R slab (thick)`

Add intensities.

You compute

$$\mathscr{R} = \mathscr{R}_{sb} + \frac{\mathscr{R}_{sb}(1 - \mathscr{R}_{sb})^2\, e^{-2\alpha d}}{1 - \mathscr{R}_{sb}^2\, e^{-2\alpha d}} \tag{313}$$

with d the thickness, α the absorption coefficient, and $\mathscr{R}_{sb}$ the single bounce reflectance. $\mathscr{R}_{sb} = [(N-1)/(N+1)]^2$.

`38.  R slab (fringes)`

Add amplitudes.

To compute the coherent reflectance, you first calculate the phase gain on passing the sample (one way),

$$\delta = 2\pi\omega N d, \tag{314}$$

the amplitude transmitted *into* the slab, from vacuum,

$$t_1 = \frac{2}{N+1}, \tag{315}$$

the amplitude transmitted *out of* the slab,

$$t_2 = \frac{2N}{N+1}, \tag{316}$$

the amplitude reflection at the interfaces, incident from within

$$r_{in} = \frac{N-1}{N+1}, \tag{317}$$

and the front-surface reflectance

$$r_{fs} = \frac{1-N}{1+N}. \tag{318}$$

Here, the sign matters; these definitions make r_{fs} negative and r_{in} positive if $N > 1$.

Then the formula for the reflection coefficient (amplitude) is

$$r = r_{fs} + \frac{t_1 r_{in} t_2 e^{2i\delta}}{1 - r_{in}^2 e^{2i\delta}}. \tag{319}$$

Finally, to get the reflectance, you calculate

$$\mathscr{R} = r r^* \tag{320}$$

References

1. J. Clerk Maxwell, “A Dynamical Theory of the Electromagnetic Field,” *Phil. Trans. Royal Soc. (London)* **155**, 459–512 (1865).
2. J.D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1975).
3. Leonard I. Schiff *Quantum Mechanics* (McGraw-Hill 1968).
4. P. Drude, *Ann. Phys. (Leipzig)* **1**, 566 (1900).
5. `http://www.pdi-berlin.de/about-us/paul-drude-1/paul-drude`. His important papers and books published in German in 1890, 1894, 1900, 1902. and 1904 are listed here. See also Paul Drude, *The Theory of Optics* (Dover, 2005).
6. David C. Johnston, “The puzzle of high temperature superconductivity in layered iron pnictides and chalcogenides,” *Adv. Phys.* **59**, 803–1061 (2010).
7. ashcroft to be supplied.
8. E. Hagen and H. Rubens, “Über die Beziehung des Reflexions- und Emissionsvermögens der Metalle zu ihrem elektrischen Leitvermögen” (“On the relation of reflectivity and emissivity of metals with their electrical conductivity”), *Ann. Phys. (Leipzig)* **4**, 873 (1903).
9. `http://online.lesn.lehigh.edu/courses/mat435/lectures/metal/met1.htm`
10. Clausius-Mossotti to be supplied.
11. Charles Kittel *Introduction to Solid State Physics* 8th Edition. ISBN 978-0-471-41526-8 (Wiley, November 2004). The 3rd and 5th editions are worth looking at also.
12. Michael Tinkham *Group Theory and Quantum Mechanics* (Dover, 2003). Previously McGraw-Hill, 1964.
13. Gerald Burns *Introduction to Group Theory With Applications* (Academic Press 1977).
14. `web.mit.edu/course/6/6.734j/www/group-full02.pdf`
15. `http://en.wikipedia.org/wiki/Pole-zero_plot`
16. Teller to be supplied.
17. B.P. Abbott *et al.* (LIGO Scientific Collaboration), “LIGO: The Laser Interferometer Gravitational-wave Observatory,” *Rep. Prog. Phys.* **72**, 076901/1–25 (2009).
18. Kogelnik to be supplied.
19. Siegman to be supplied.
20. O.S. Heavens *Optical Properties of Thin Solid Films,* (Dover, 1955).
21. holstein54pr to be supplied.
22. Handbook of Electronic Tables and Formulas; `http://www.powerstream.com/Wire_Size.htm`.
23. Peter Debye, “Zur Theorie der spezifischen Waerme,” *Annalen der Physik (Leipzig)* **39**, 789 (1912).
24. D.B. Tanner and D.C. Larson, “Electrical resistivity of silver films,” *Phys. Rev.* **166**, 652–655 (1968).
25. J. Cochran and M. Yaqub, *Phys. Lett.* **5**, 307–309 (1963).
26. R. de L. Kronig, “On the theory of the dispersion of X-rays,” *J. Opt. Soc. Am.* **12**, 547–557 (1926).
27. H.A. Kramers, “La diffusion de la lumiere par les atomes,” *J. Atti Cong. Intern. Fisici, (Transactions of Volta Centenary Congress) Como* **2**, 545–557 (1927); “Die dispersion und absorption von Röntgenstrahlen,” *Phys. Z.* **30**, 522–523 (1929)

28. C. Kittel *Statistical Physics*
29. L.D. Landau and E.M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960).
30. Frederick Wooten, *Optical Properties of Solids* (Academic Press, New York, 1972).
31. Martin Dressel and George Grüner, *Electrodynamics of Solids: Optical Properties of Electrons in Matter* (Cambridge University Press, Cambridge, 2002).
32. John S. Toll, "Causality and the Dispersion Relation: Logical Foundations," *Phys. Rev.* **104**, 1760–1770 (1956).
33. Ryogo Kubo and Munetake Ichimura, "Kramers-Kronig Relations and Sum Rules," *J. Math. Phys.* **13**, 1454 (1972).
34. E. Shiles, Taizo Sasaki, Mitio Inokuti, and D.Y. Smith, "Self-consistency and sum-rule tests in the Kramers-Kronig analysis of optical data: Applications to aluminum," *Phys. Rev. B* **22**, 1612 (1980).
35. Ben Yu-Kuang Hu, "Kramers-Kronig in two lines," *Am. J. Phys.* **57**, 821–821 (1989).
36. K.-E. Peiponen and E.M. Vartiainen, "Kramers-Kronig relations in optical data inversion," *Phys. Rev. B* **44**, 8301 (1991).
37. G.W. Milton, D.J. Eyre, and J.V. Mantese, "Finite Frequency Range Kramers-Kronig Relations: Bounds on the Dispersion," *Phys. Rev. Lett.* **79**, 3062–3065 (1997).
38. Frederick W. King, "Efficient numerical approach to the evaluation of Kramers-Kronig transforms," *J. Opt. Soc. Am. B* **19**, 2427–2436 (2002).
39. Frederick W King, "Alternative approach to the derivation of dispersion relations for optical constants," *J. Phys. A: Math. Gen.* **39**, 10427 (2006).
40. K.-E. Peiponen and J.J. Saarinen, "Generalized Kramers-Kronig relations in nonlinear optical- and THz-spectroscopy," *Reports on Progress in Physics* **72**, 056401 (2009).
41. Craig F. Bohren, "What did Kramers and Kronig do and how did they do it?" *Eur. J. Phys.* **31**, 573–577 (2010).
42. A.L. Cauchy, "Oeuvres complètes," Ser. 1, 4, Paris (1890).
43. J.S. Plaskett and P.N. Schatz, "On the Robinson and Price (Kramers-Kronig) Method of Interpreting Reflection Data Taken through a Transparent Window," *J. Chem.Phys.* **38**, 612 (1963).
44. G. Andermann, A. Caron, and David A. Dows, "Kramers-Kronig Dispersion Analysis of Infrared Reflectance Bands." *J. Opt. Soc. Am.* **55**, 1210–1212 (1965).
45. D.M. Roessler, "Kramers=Kronig analysis of non-normal incidence reflection," *Br. J. Appl. Phys.* **16**, 1359 (1965).
46. K. Kozima, W. Suetaka, and P.N. Schatz, "Optical constants of thin films by a Kramers-Kronig method," *J. Opt. Soc. Am.* **56**, 181–184 (1966).
47. D.B. Romero, C.D. Porter, D.B. Tanner, L. Forro, D. Mandrus, L. Mihaly, G.L. Carr, and G.P. Williams, "Quasiparticle damping in $Bi_2Sr_2CaCu_2O_8$ and $Bi_2Sr_2CaCuO_6$," *Phys. Rev. Lett.* **68**, 1590–1593 (1992).
48. Z. Wu, Z. Chen, X. Du, J.M. Logan, J. Sippel, M. Nikolou, K. Kamarás, J.R. Reynolds, D.B. Tanner, A.F. Hebard, and A.G. Rinzler, "Transparent, conductive nanotube Films," *Science* **305**, 1273–1276 (2004).
49. F. , K. Kamarás, M. Nikolou, D.B. Tanner, Z.H. Chen, and A.G. Rinzler, "Study of charge dynamics in transparent single-walled carbon nanotube films," *Phys. Rev. B* **74**,

045431/1–6 (2006).

50. Paul C. Martin "Sum Rules, Kramers-Kronig Relations, and Transport Coefficients in Charged Systems," *Phys. Rev.* **161**, 143–155 (1967).
51. W.M. Saslow, "Two classes of Kramers-Kronig sum rules," *Phys. Lett.* **33**, 157–158 (1970).
52. M. Altarelli, D.L. Dexter, H.M. Nussenzveig, and D.Y. Smith, "Superconvergence and sum rules for the optical constants," *Phys. Rev. B* **6**, 4502–4509 (1972).
53. M. Altarelli and D.Y. Smith "Superconvergence and sum rules for the optical constants: Physical meaning, comparison with experiment, and generalization," *Phys. Rev. B* **9**, 1290–1298 (1974).
54. D.Y. Smith and E. Shiles, "Finite-energy f-sum rules for valence electrons," *Phys. Rev. B* **17**, 4689 (1978).
55. Frederick W. King, "Some bounds for the absorption coefficient of an isotropic nonconducting medium," *Phys. Rev. B* **25**, 1381 (1982).
56. G. Lévque, "Augmented partial sum rules for the analysis of optical data," *Phys. Rev. B* **34**, 5070 (1986).
57. R.E. Peierls, *Quantum Theory of Solids*, (Clarendon Press, Oxford, England, 1955).
58. `http://en.wikipedia.org/wiki/List_of_semiconductor_materials`
59. Milton Ohring *Reliability and failure of electronic materials and devices,* Academic Press, 1998 ISBN 0-12-524985-3, p. 310
60. Y. Tao, J.M. Boss, B.A. Moores, and C.L. Degen, "Single-Crystal Diamond Nanomechanical Resonators with Quality Factors exceeding one Million," arXiv:1212.1347 (2012)
61. A.K. Abass and N.H. Ahmad, "Indirect band gap investigation of orthorhombic single crystals of sulfur." *J. Phys. Chem. Solids* **47**, 143 (1986).
62. D.A. Evans, A.G. McGlynn, B.M. Towlson, M. Gunn, D. Jones, T.E. Jenkins, R. Winter, N.R.J. Poolton, "Determination of the optical band-gap energy of cubic and hexagonal boron nitride using luminescence excitation spectroscopy." *J. Phys.: Cond. Matt* **20**, 75233 (2008).
63. Claus F. Klingshirn *Semiconductor optics* (Springer, 1997), p. 127.
64. S. Banerjee et al. "Physics and chemistry of photocatalytic titanium dioxide: Visualization of bactericidal activity using atomic force microscopy," *Current Science* **90**, 1378 (2006).
65. O. Madelung, U. Rössler, M. Schulz (ed.) "Cuprous oxide (Cu_2O) band structure, band energies." in Group III Condensed Matter. Numerical Data and Functional Relationships in Science and Technology. 41C: Non-Tetrahedrally Bonded Elements and Binary Compounds I. doi:10.1007/10681727_62.
66. B. G. Yacobi Semiconductor materials: an introduction to basic principles Springer, 2003, ISBN 0-306-47361-5; C Hodes; Ebooks Corporation (8 October 2002). Chemical Solution Deposition of Semiconductor Films. CRC Press. p. 319.
67. Yasantha Nirmal Rajakarunanayake, *Optical properties of Si-Ge superlattices and wide band gap II-VI superlattices,* Dissertation (Ph.D.), California Institute of Technology, 1991
68. H.K. Onnes, Comm. Phys. Lab. Univ. Leiden **122b** (1911)

69. T.G. Berlincourt and R.R. Hake, “Superconductivity at high magnetic fields,” *Phys. Rev.* **131**, 140 (1963).
70. B.T. Matthias, T.H. Geballe, S. Geller, E. Corenzwit *Phys. Rev.* **95**, 1435 (1954).
71. G. Bednorz and K.A. Mueller, “Possible high T_C superconductivity in the Ba-La-Cu-O system,” *Z. Physik B.* **64**, 189–193 (1986).
72. J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zenitani, and J. Akimitsu, “Superconductivity at 39 K in magnesium diboride,” *Nature* **410**, 63 (2001).
73. D. Jerome, A. Mazaud, M. Ribault and K. Bechgaard, *J. Phys. Lett.* **41**, L95 (1980).
74. V.V. Walatka, Jr. M. M. Labes, and J.H. Perlstein, “Polysulfur Nitride—a one-dimensional chain with a metallic ground state,” *Phys. Rev. Lett.* **31**, 1139–1142 (1973).
75. Julian Eisenstein, “Superconducting Elements,” *Rev. Mod. Phys.* **26**, 277–291 (1954).
76. B.T. Matthias, ”Superconductivity in the periodic system.” *Prog. Low Temp. Phys.* **2**, 138–150 (1957).
77. B. Matthias, T. Geballe, and V. Compton, “Superconductivity,” *Rev. Mod. Phys.* **35**, 1–22 (1963).
78. B.W. Roberts, “Survey of superconductive materials and critical evaluation of selected properties,” *J. Phys. Chem. Ref. Data* **5**, 581 (1976).
79. D. Jerome and H.J. Schultz, *Adv. Phys.* **51**, 293–479 (2002).
80. John Singleton and Charles Mielke, “Quasi-two-dimensional organic superconductors: A review,” *Contemporary Physics* **43**, 63–96 (2002).
81. T.H. Geballe, “The never ending search for high temperature superconductivity,” J. Superconductivity Novel Magnetism, 19, 261–276 (2006).
82. A.G. Lebed (Ed.), “The Physics of Organic Superconductors and Conductors,” (Springer Series in Materials Science , Vol. 110, 2008).
83. A.A. Kordyuk, “Iron-based superconductors: Magnetism, superconductivity, and electronic structure (Review Article),” *Low Temp. Phys.* **38**, 888–899 (2012).
84. Koichi Kitazawa, “Superconductivity: 100th anniversary of its discovery and its future,“ *Japanese J. Appl.Phys.* **51**, 010001/1–14 (2012).
85. W. Meissner, *Z. fur Phys.* **58**, 570 (1929).
86. W. Meissner and R. Ochsenfeld “Ein neuer Effekt bei Eintritt der Supraleitfähigkeit,” *Naturwissenschaften* **21**, 787–788 (1933).
87. J. Bardeen, L.N. Cooper, and J.R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).
88. L.D. Landau and V.L. Ginzburg, “On the theory of superconductivity,” *Pis’ma Zh. Eksp. Teor. Fiz.* **20**, 1064 (1950).
89. J. Robert Schrieffer *Theory Of Superconductivity* (W.A. Benjamin, New York, 1964; Perseus Books, 1999)
90. P.G. de Gennes, *Superconductivity of metals and alloys* (W.A. Benjamin, New York. 1966; Perseus Books 1999).
91. L.P. Gorkov, “Microscopic derivation of the Ginzburg-Landau equations in the theory of superconductivity,” Sov. Phys. JETP, 9, 1364–1367 (1959)
92. L.P. Gor’kov, *Zh. Eksp. Teor. Fiz.* **36**, 1918 (1959).
93. F. London and H. London, “The Electromagnetic Equations of the Supraconductor,” *Proc. Roy. Soc.* **A 149**, 71-88 (1935).
94. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1980).

95. T.A. Fulton and G.J Dolan, "Observation of single-electron charging effects in small tunnel junctions," *Phys. Rev. Lett.* **59**, 109–112 (1987).
96. A.B. Pippard, "The surface energies of superconductors," *Math. Proc. Camb. Phil. Soc.* **47**, 617–625 (1951); A.B. Pippard, *Proc. Roy. Soc.* **A216**, 547 (1953).
97. D.R. Tilley and J. Tilley, *Superfluidity and Superconductivity* (Adam Hilger Ltd, Bristol, 1990).
98. D.C. Mattis and J. Bardeen, "Theory of the anomalous skin effect in normal and superconducting metals," *Phys. Rev.* **111**, 412–417 (1958).
99. Sang Boo Nam, "Theory of Electromagnetic Properties of Superconducting and Normal Systems. I," *Phys. Rev.* **156**, 470–486 (1967).
100. Sang Boo Nam, "Theory of Electromagnetic Properties of Strong-Coupling and Impure Superconductors. II," *Phys. Rev.* **156**, 487–93 (1967).
101. Piotr B. Miller, "Surface Impedance of Superconductors," *Phys. Rev.* **118**, 928–934 (1960).
102. S. Skalski, O. Betbeder-Matibet, and P.R. Weiss, "Properties of Superconducting Alloys Containing Paramagnetic Impurities," *Phys. Rev.* **136**, A1500–A1518 (1964).
103. W. Zimmermann, E.H. Brandt, M. Bauec, E. Seider, and L. Genzel, "Optical conductivity of BCS superconductors with arbitrary purity," *Physica C: Supercond.* **183**, 99–104 (1991).
104. A.B. Pippard, "The surface impedance of superconductors and normal metals at high frequencies. II. The anomalous skin effect in normal metals," *Proc. R. Soc. Lond. A* **191**, 385–399 (1947).
105. R. G. Chambers, "Anomalous Skin Effect in Metals," *Nature* **165**, 239–240 (1950).
106. H.B.G. Casimir and J. Ubbink, "The skin effect: II. The skin effect at high frequencies," *Philips Tech. Rev.* **28**, 300–315 (1967).
107. H.B.G. Casimir, "On electromagnetic units," *Helvetica Phys. Acta* **41**, 741–742 (1968). This article is reprinted in Robert L. Weber and Eric Mendoza, *A Random Walk in Science*, (Institute of Physics Publishing, Bristol, 1973).