Chapter 3.

Insulators, Metals, and Semiconductors

Canal
Robert Wittig

"A scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it."

Max Planck

Contents

Seal of Approval
Robert Wittig

"One of the most decisive influences on my scientific outlook came from Professor Paul Ehrenfest. He had been born and brought up in Vienna and taught at the University of Leiden in Holland. In 1929, when Born had his stroke, Ehrenfest came to Göttingen to teach for a year, and I was lucky enough to meet him. Ehrenfest took a special interest in me because I came from Vienna, and we spent a lot of time together. As a theoretical physicist, he also shunned long mathematical derivations, a trail that aroused my enthusiastic approval. He felt that mathematics often covered up essential points. At one of our first meetings he said to me, 'Don't let yourself be impressed with all the formalism you will see here in Göttingen. Don't be awed by all these great scientists. Ask questions. Don't be afraid to appear stupid. The stupid questions are usually the best and the hardest to answer. They force the speaker to think about the basic problem. Try to get to the fundamental ideas, freed from the mathematical thicket. Physics is simple but subtle.' This last remark has made the greatest impression on me and it has been one of the guidelines of my thinking."

The Joy of Insight
Victor Weisskopf

3.1. Introduction

Getting a Grip on Things
Sculpture by Robert Wittig

It is not possible in the limited time available to provide other than a brief introduction to the kinds of interactions that take place between electromagnetic fields and dielectric media. A more thorough treatment would require at least a full semester to cover the fundamentals, and probably a second semester if a representative sampling of modern applications is to be included. What *can* be done in a couple of weeks, though, is introduce and discuss important concepts and underlying principles, in particular, ones that will be encountered in later chapters. This is the goal of the present chapter. Simple models, examples, and embedded exercises serve this purpose well.

The principles apply, of course, to other areas in which the interaction of radiation with matter is important. Several will be encountered in Part V, where the quanta of the electromagnetic field, photons, are derived and used in descriptions of radiation-matter interactions.

In the present chapter, we shall deal with ionic crystals, metals, and doped semiconductors. Related systems will also be encountered: molecular crystals held together by van der Waals and/or hydrogen-bond forces, covalent crystals, and crystals with mixed covalent-ionic character. For the most part, phenomena will be explored that do not have counterparts in isolated molecules, at least not immediately obvious counterparts.

Let us begin with the ubiquitous photon. The magnitude of a photon's linear momentum in free space follows from the energy-momentum relation of special relativity: $E^2 = m^2c^4 + p^2c^2$. Just set m equal to zero. Thus, $p = E/c$, and using $E = \hbar\omega$ and $\omega = ck$ yields $p_{\text{vac}} = \hbar k_{\text{vac}}$. Now suppose a photon propagates in a medium whose optical properties are taken into account, to a high degree of accuracy, with a real index of refraction, n. Photon energy does not change in passing from vacuum to a dielectric medium, whereas $c \to c/n$ and $k_{\text{vac}} \to nk_{\text{vac}}$.

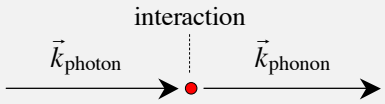
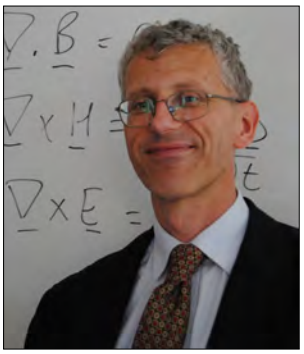


Figure 1. Photon annihilation accompanied by creation of a phonon requires $\vec{k}_{\text{photon}} = \vec{k}_{\text{phonon}}$.

If an infrared photon is annihilated through its interaction with a crystal, its wave vector, in effect, is transferred to the phonon (Fig. 1). This is what happens when an absorption spectrum is recorded. That is, the spectrum is the frequency dependence of the removal of photons from the incident field. Alternatively, if the photon becomes coupled to the phonon but is not annihilated, their wave vectors must match. Otherwise they cannot travel synchronously through the crystal.

An excellent introductory level text is the one written by Mark Fox: Optical Properties of Solids [17]. It covers many of the topics we will examine, and it is a great all around reference. Though ostensibly for a physics audience, the book is accessible to readers of varied backgrounds and levels of sophistication. You will have no trouble, as the writing is clear, and concepts are explained using simple models and straightforward arguments. The mathematics is at a level that is sufficient to get the job done, but not more. I have placed a couple of copies of this book in our course library, as well as Fox's other book: Quantum Optics: An Introduction [18]. If more are needed, they will be provided.



Mark Fox

Photons and Phonons

The diatomic ionic lattice was examined in Section 7 of Chapter 1, including atom displacements in the acoustic and optical branches throughout the Brillouin zone. For acoustic phonons with small wave vectors, the atoms move in the same direction over broad spatial expanses, as indicated in the sketch below. Consequently, the transition dipole moment is zero. Accordingly, acoustic modes do not interact with radiation.

The optical phonons of the diatomic ionic lattice differ qualitatively from the acoustic phonons in this regard. In the small wave vector part of the optical branch, the positive and negative charges in a unit cell move in opposite directions (*vide infra*, Fig. 2). To help visualize the differences, you may find it helpful to review Section 7 in Chapter 1. As mentioned earlier, a photon interacts resonantly with an ionic crystal only at small k values. The photon couples with dipoles in the optical branches, but not in the acoustic branches. Consequently, we need only to consider optical phonons.

NaCl crystal: In the small wave vector region of the acoustic branch, Na^+ and Cl^- displacements (purple and green, respectively) are in the same direction and have the same magnitude.

To see why optical phonons that couple with photons lie in the small wave vector part of the Brillouin zone, equate the magnitudes of $\vec{k}_{\text{photon}}$ and $\vec{k}_{\text{phonon}}$:

$$k_{\text{photon}} = k_{\text{phonon}} . \tag{3.1}$$

As usual, $k_{\text{photon}} = 2\pi / \lambda_{\text{photon}}$ and $k_{\text{phonon}} = 2\pi / \lambda_{\text{phonon}}$. The requirement that k_{phonon} is minuscule compared to the full range of the Brillouin zone follows when realistic numbers are introduced into eqn (3.1), as explained below.

Small Phonon Wave Vectors

The fundamental vibrational frequencies of alkali halide crystals lie in the far infrared. Let us choose, for the sake of illustration, a vibrational frequency of 200 cm^{-1} and a lattice spacing of 0.5 nm.[1] Suppose now that the frequency of an infrared photon is also 200 cm^{-1}. The magnitude of its free-space wave vector, k_{photon}, is $4\pi \times 10^{-5}$ nm^{-1}. Furthermore, we shall assume, again for the sake of illustration, that the crystal's index of refraction is unity, in which case the value of k_{photon} inside the crystal is $4\pi \times 10^{-5}$ nm^{-1}. Wave vector conservation requires that the photon and phonon wave vectors (and therefore wavelengths) match perfectly. Consequently, the phonon wave vector must have the value $4\pi \times 10^{-5}$ nm^{-1}.

For a lattice spacing of 0.5 nm, the phonon wave vector at the edge of the first Brillouin zone is $\pi / a = 2\pi$ nm^{-1}. We see that the value of $4\pi \times 10^{-5}$ nm^{-1} accounts for a mere two parts in 10^5 of the range $0 \le k \le \pi / a$, and one part in 10^5 of the full Brillouin zone. In other words, the photon wave vector constitutes a tiny fraction of the Brillouin zone. This underscores the fact that for a phonon to interact resonantly with a photon its wave vector (crystal momentum) must be minuscule. The bottom line is that only optical phonons having wave vectors close to zero can interact resonantly with radiation.

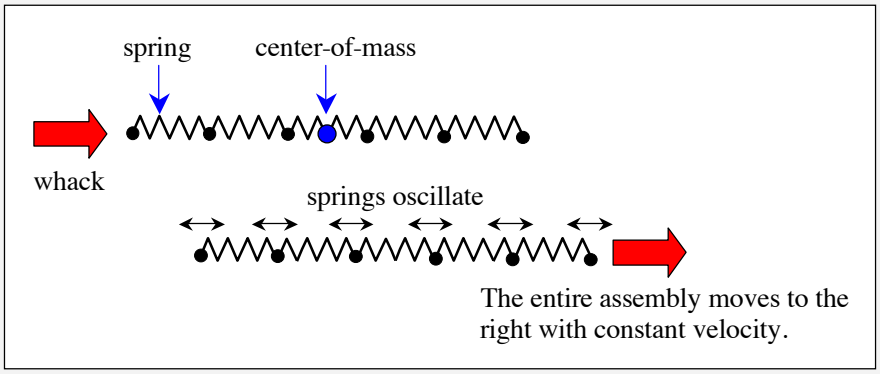
Wave Vector Conservation

The relationship between momentum conservation and wave vector conservation is subtle. Consider the 1D system shown below of equal masses connected with springs. Hitting the end transfers momentum to the system. The masses oscillate, and the system as a whole travels to the right with the imparted momentum. The oscillatory modes have no net momentum. Mathematically, the behavior of the chain of masses can be treated as a linear combination of the normal modes. The $k = 0$ mode corresponds to translation of the whole system: nonzero overall linear momentum. The other modes have zero momentum, because for every particle that is moving toward or away from its equilibrium location another particle is moving oppositely.

The question remains: why do phonons seem to carry momentum when they interact with electrons or neutrons? In these processes there are conditions on the $\vec{k}$ vectors. For example, an incident particle's $\vec{k}$ vector must be equal to its outgoing $\vec{k}$ vector plus the $\vec{k}$ vector of the created phonon. This comes from a calculation of the scattering probability, where one multiplies different waves together and integrates. If they don't add up in this way, the integral is zero. We have seen a lot of this al-

[1] These values (λ = 50 μm and a = 0.5 nm) are in rough accord with the alkali halide crystal NaCl (λ_{TO} = 61 μm and a = 0.563 nm).

ready with sums of factors such as $e^{i(k-k')na}$. The sum over n vanishes unless $k = k'$. The same idea applies here, the only difference being $e^{i(k_{in}-k_{out}-k_{phonon})na}$. It is understood that if k_{phonon} at first appears to lie outside the first Brillouin zone it is to be replaced by a new wave vector plus a reciprocal lattice vector. Thus, $\vec{k}$ vector conservation is not quite the same as conservation of momentum. Whereas momentum conservation involves the movement of particles relative to a reference frame, wave vector conservation involves the relative motions of particles.



When an electron scatters from a phonon, resulting in an electron with a different momentum, the $\vec{k} = 0$ mode must be excited. In other words the crystal as a whole changes momentum. Intuitively, it makes sense to think of this as follows. When the wavelengths of a phonon and of initial and final states of an electron match up correctly, an electron can get "kicked" from the initial to final state. This "kicking off" from the crystal lattice imparts its change in momentum to the lattice.

In a crystal the speed and kinetic energy that the crystal gains are each minuscule relative to the speed and energy of a typical phonon. Thus, we can neglect the momentum and energy transferred to the motion of the crystal. Because $\vec{k}$ conservation looks so much like momentum conservation, and because we can neglect momentum transfer to the crystal, no harm is done pretending the phonon has momentum $\hbar\vec{k}$.

Raman Scattering

Another experimental method that is used widely to obtain vibrational frequencies of condensed phase samples is Raman scattering. Here, we shall consider common-old-ordinary Raman scattering (COORS), rather than its stimulated Raman scattering (SRS) counterpart, which is also used widely. Incident photons can result in the creation of optical phonons according to the selection rules for vibrational Raman scattering in crystals, which we shall not trouble ourselves with. The wave vectors of the incident (i) and scattered (s) photons obey the relation: $\vec{k}_{\text{photon}}{}^{(i)} - \vec{k}_{\text{photon}}{}^{(s)} = \vec{k}_{\text{phonon}}$, for each optical phonon branch involved in the overall scattering process. This is similar to the criterion given by eqn (3.1), though now the criterion involves the difference between incident and scattered wave vectors.

Raman scattering situations are also encountered in which optical phonons that are already present are annihilated. In this case, the scattered photon is of higher energy than the incident photon: $\hbar\omega_s = \hbar\omega_i + \hbar\omega_{\text{phonon}}$, and the scattered photon wave vector has acquired the wave vector of the annihilated optical phonon: $\vec{k}_{\text{photon}}{}^{(s)} = \vec{k}_{\text{photon}}{}^{(i)} + \vec{k}_{\text{phonon}}$. This is referred to as anti-Stokes scattering, whereas the former case is referred to as Stokes scattering. In both cases, the phonons that are created or annihilated have extremely small wave vectors relative to the width of the Brillouin zone, because the wave vectors of the participating photons are extremely small relative to the width of the Brillouin zone. Anti-Stokes scattering has numerous applications in both SRS and COORS.[2]

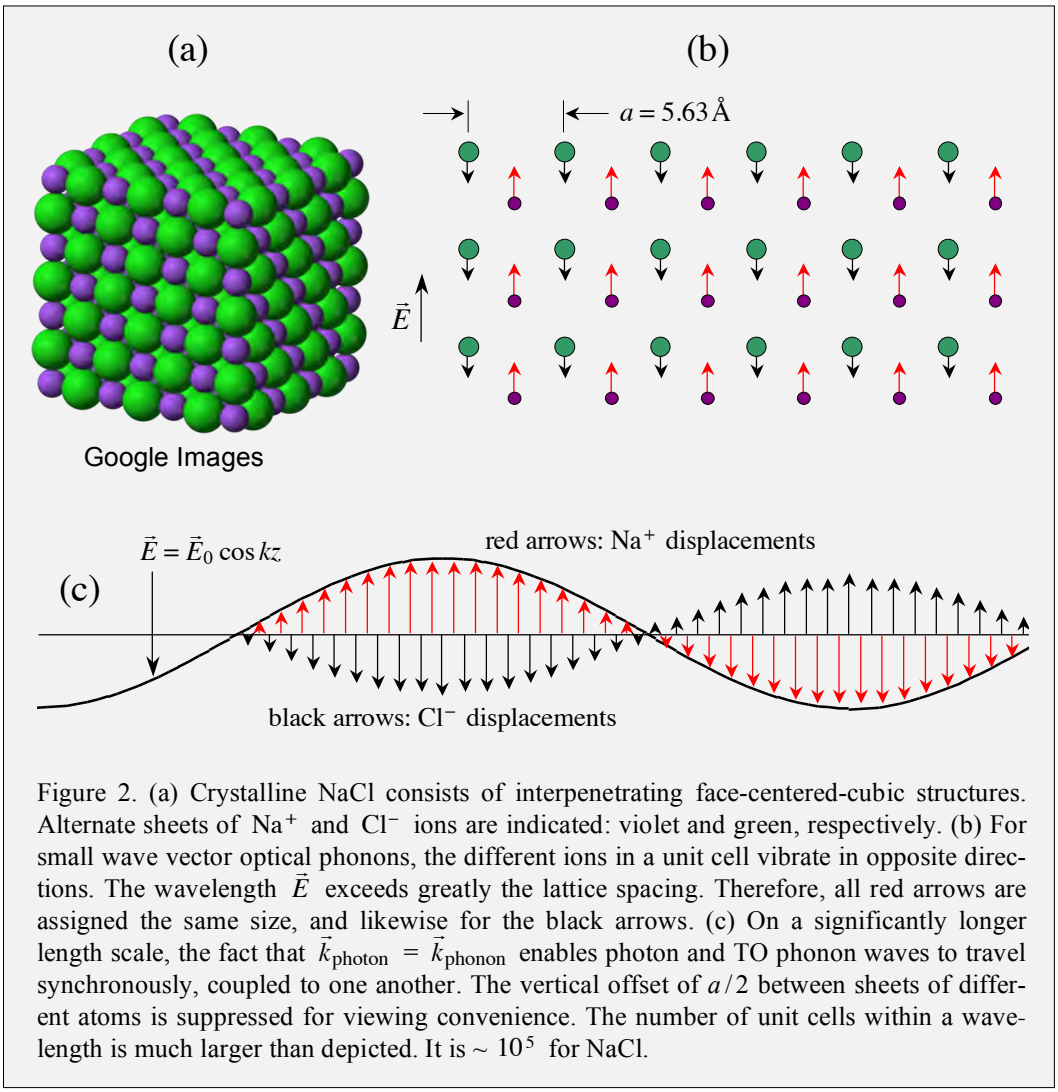


Figure 2. (a) Crystalline NaCl consists of interpenetrating face-centered-cubic structures. Alternate sheets of Na^+ and Cl^- ions are indicated: violet and green, respectively. (b) For small wave vector optical phonons, the different ions in a unit cell vibrate in opposite directions. The wavelength $\vec{E}$ exceeds greatly the lattice spacing. Therefore, all red arrows are assigned the same size, and likewise for the black arrows. (c) On a significantly longer length scale, the fact that $\vec{k}_{\text{photon}} = \vec{k}_{\text{phonon}}$ enables photon and TO phonon waves to travel synchronously, coupled to one another. The vertical offset of $a/2$ between sheets of different atoms is suppressed for viewing convenience. The number of unit cells within a wavelength is much larger than depicted. It is $\sim 10^5$ for NaCl.

[2] Raman scattering that involves acoustic phonons is referred to as Brillouin scattering. Raman scattering can be thought of in terms of vibrations within a unit cell, whereas the acoustic phonons encountered in Brillouin scattering are not so localized.

Another way to appreciate the fact that only small phonon wave vectors are involved in a lattice's interactions with electromagnetic radiation is through consideration of the spatial variation of the electromagnetic field, specifically, its relation to the optical phonon wavelength (Fig. 2). The free-space wavelength of a 200 cm^{-1} photon is 5×10^{-3} cm (50 μm). Because of the wave-vector-matching requirement, we are concerned with optical phonons that have the same wavelength as the photon. The photon wavelength inside the crystal is smaller than its free-space value according to the index of refraction, $\lambda_{\text{vacuum}} / n$. This wavelength typically spans tens of thousands of unit cells.

For an optical phonon near the center of the Brillouin zone ($k \approx 0$) of an ionic crystal, the oppositely charged ions in a unit cell move in opposite directions. If one looks at the ion displacements from one cell to the next, they are the same for like ions. For example, Na^+ in one cell moves in the same way as Na^+ in nearby cells. Only on the much larger length scale of the wavelength are the Na^+ displacements seen to differ, as illustrated in Fig. 2(c). Synchrony between oscillations of the electric field and the dipoles of the unit cells demands that the phonon and photon wave vectors match.

The NaCl crystal is depicted schematically in Fig. 2(a) using a 3D model taken from Google Images. The top layer reveals planes of like atoms that form alternate sheets of Na^+ and Cl^- ions. In Fig. 2(b), ion displacements commensurate with the electric field $\vec{E}$ are indicated. Displacement amplitudes are exaggerated greatly for viewing convenience. Because the wavelength of $\vec{E}$ is much larger than the lattice spacing, the displacements are shown as having equal magnitude for all atoms of a given type. The lattice spacing in Fig. 2(b) is 0.563 nm, whereas the TO wavelength of NaCl is approximately 6.1×10^4 nm (61 μm). Thus, a wavelength spans roughly 10^5 unit cells. Figure 2(c) illustrates the synchrony between the electromagnetic and TO phonon waves. When time dependence is included, the photon and phonon waves are seen to move together through the crystal. This facilitates the coupling that results in interesting elementary excitations called phonon polaritons, as discussed in detail in Chapter 6.

Dispersion curves depend on wave vector direction. The wave vector direction used in Fig. 2 is normal to the planes of like atoms. The dispersion curve for this direction is closest to the one derived in Chapter 1 for the ionic lattice. A 1D model was used there, and the character of transverse displacements was surmised. Figure 3 gives an example of how mode frequencies vary along different directions in *k*-space for the III-V crystal GaN, whose growth led to the 2014 Nobel Prize in Physics [49]. Such diagrams can be complicated, for example, when there are more atoms in a unit cell. Believe it or not, the curves in Fig. 3 are for a relatively simple case.

> We have two perspectives at our disposal. One starts by treating the photon and the TO phonon mode as independent (uncoupled) entities. These independent entities are the ingredients of so-called bare states. For example, *n* photons and zero quanta of the TO mode under consideration is written $|n, 0\rangle$, whereas $n - 1$ photons and one TO mode quantum reads $|n-1, 1\rangle$. When the energies of these bare states are equal, or close to it, they become strongly coupled. This coupling, which takes place through the $-\vec{\mu} \cdot \vec{E}$ interaction, yields so-called dressed states. This perspective reveals curve crossings in *k*-space and other neat stuff. Polaritons are seen to result from such avoided crossings.

In the other perspective, photons and phonons are coupled at the outset. The dielectric response function $\varepsilon(\omega)$ is derived, and phonon polaritons are taken directly from the dispersion relation. These polaritons are eigenmodes of the dielectric. Waves can only propagate in the medium as linear combinations of these modes.

In most cases, this latter perspective is preferred, as it deals directly with the dispersion relation and accounts for (and predicts) many phenomena. Coupling external radiation to the eigenmodes takes a bit of experimental skill. Chapter 6: *Phonon Polariton* is devoted to such polaritons, and Chapter 7: *Plasmon Polariton* is complementary. In the present chapter, we shall straddle these two perspectives: speaking frequently of photon absorption, which is familiar to almost everyone, while keeping the phonon polariton in mind.

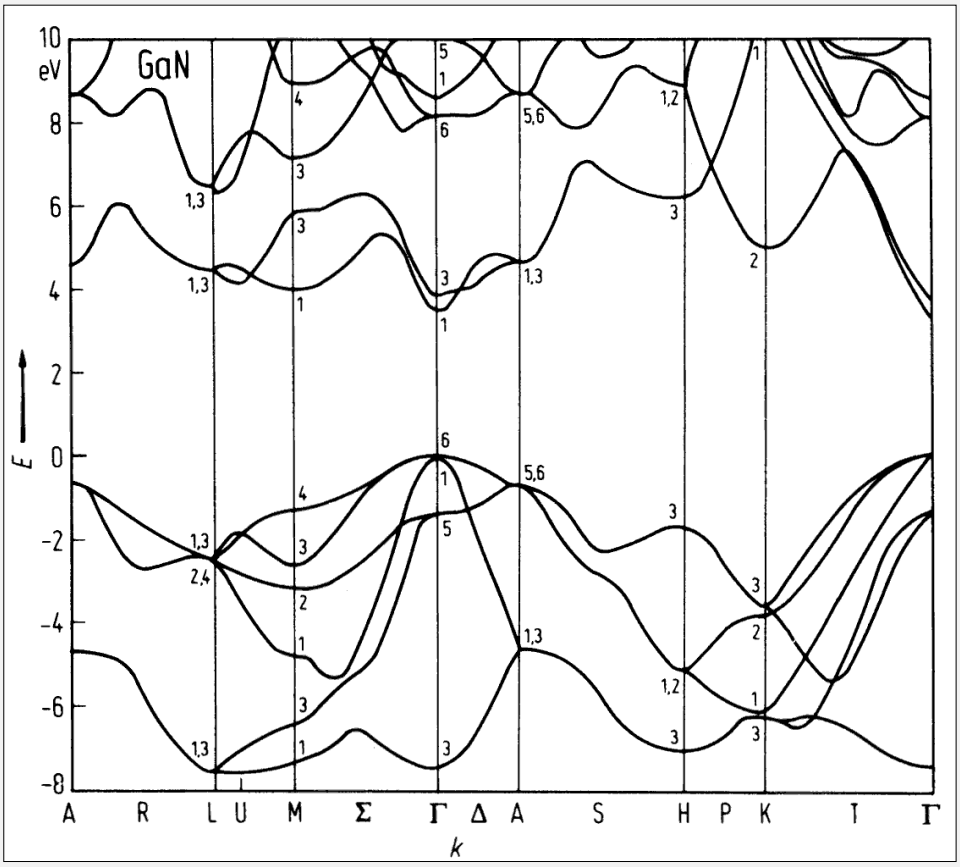


Figure 3. Band structure diagram of GaN [49]: The point is to illustrate the complexity of such diagrams for a relatively simple case. Energy versus wave vector depends on direction. There is no limit to the number of paths that can be chosen. Paths along different directions in general have different shapes. The horizontal axis labels convenient points in *k*-space.

To end this section, consider Fig. 4. Think of the optical branches shown there as representing a general case of two unlike atoms per unit cell. The curves are sketches that serve only to illustrate the restriction to small optical phonon wave vectors. The black arrow indicates an infrared photon. To keep matters simple, assume that resonant interaction takes place with a mode in which no TO phonon quanta are present. Recall that in Chapter 1 such a state was assigned the Fock space label $|0\rangle$, with the understanding that the zero in the ket stands for no quanta in the mode under consideration. All of the other modes and their respective numbers of quanta are taken as understood.

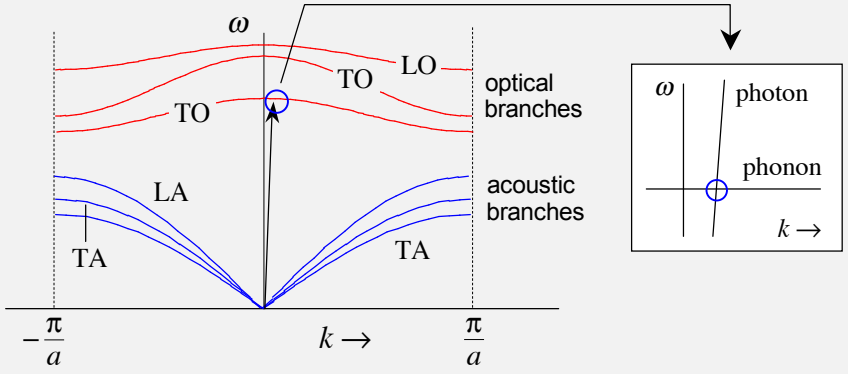


Figure 4. The near-vertical black arrow indicates the photon dispersion relation. Its angle with respect to the vertical axis is exaggerated for viewing convenience. It lies much nearer the axis than shown. Note the resonance where the photon and phonon curves cross (white box).

The assumption that there are no quanta present in the mode of interest means that, as long as the electromagnetic field is not intense, we do not have to deal with stimulated processes in which an optical phonon that is already present gives up its quantum to the field, or with transitions in which a quantum of a phonon mode is transferred to the field. The black arrow in the figure indicates photon energy and wave vector, though it exaggerates the magnitude of the photon wave vector for ease of visualization. As discussed earlier, the tip of the black arrow actually lies much closer to the axis than shown.

If acoustic phonons and other optical phonons are present, which will be the case at all temperatures of interest, their wave vectors are included in *overall* wave vector conservation. However, acoustic phonons do not interact with photons, and a single TO branch has been selected. With photoexcitation of a given optical mode, other excitations "go along for the ride." Thus, when starting from $|0\rangle$, resonance requires that $\hbar\omega_{\text{phonon}} = \hbar\omega_{\text{photon}}$, and wave vector conservation requires $\vec{k}_{\text{phonon}} = \vec{k}_{\text{photon}}$. This is indicated in the offset in Fig. 4.

3.2. TO and LO Branches: Wave Vector Conservation

In this section we shall discuss how an electromagnetic wave interacts with an ionic lattice to create optical phonons. This sounds harmless but we must tread carefully. A statement like: "how an electromagnetic wave interacts with an ionic lattice to create optical phonons," assumes a separateness in which the electromagnetic wave is one item and the optical phonon is another. This cannot be the case in an ionic crystal. Picture an optical phonon with its oscillating dipole. It is impossible for this to be separate from an electromagnetic field, as an oscillating dipole radiates photons. In other words, the photon and phonon must form a coupled entity via the matrix element of $-\vec{\mu}\cdot\vec{E}$. The combined entity is referred to as a polariton. It is the sole topic of Chapter 6.

To begin, consider a TO branch and assume a linearly polarized electromagnetic wave. The wave passes through the crystal with its electric field perpendicular to the direction of propagation, as indicated in Fig. 5(a).[3] As discussed in the previous section, the oscillating electric field is commensurate with vibration of the lattice in a direction that is perpendicular to $\vec{k}_{\text{photon}}$. The wave vector of the TO phonon lies parallel to $\vec{k}_{\text{photon}}$, so there is no difficulty with wave vector conservation when a photon and a phonon become coupled.[4]

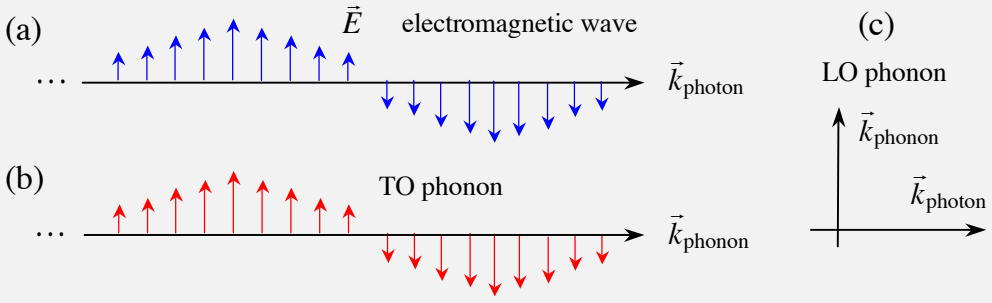


Figure 5. (a) The photon wave vector is perpendicular to its electric field. (b) A TO phonon can be excited when the transition dipole of a unit cell has a component parallel to $\vec{E}$. Phonon and photon wavelengths match in the small wave vector regime, resulting in a phonon polariton. (c) The wave vector of the LO phonon induced by a transverse electric field is perpendicular to the photon wave vector. Thus, photons cannot couple to LO phonons.

[3] Phase matching is important in many nonlinear optical processes. Energy and phase in general do not propagate in the same direction, which limits the interaction length.

[4] Cubic crystals such as NaCl have the additional simplification that a field can propagate in any direction in the crystal with its electric field experiencing the same index of refraction in any direction in the plane perpendicular to the wave vector.

Now consider the possibility of coupling between a photon and an LO phonon. Clearly, there is a problem because the wave vector of the LO phonon points in the same direction as the photon's electric field. In other words, $\vec{k}_{\text{phonon}}$ would be perpendicular to $\vec{k}_{\text{photon}}$, as seen in Fig. 5(c). This eliminates the possibility of photon-phonon coupling. For example, it is impossible to achieve overall wave vector conservation in a process that annihilates a photon and creates an LO phonon because $\vec{k}_{\text{phonon}}$ and $\vec{k}_{\text{photon}}$ would have to be perpendicular to one another. Consequently, this process is forbidden.

It turns out that the above picture of a photon coupling with a TO phonon applies to many crystals. As mentioned above, an interesting situation arises with ionic crystals. When their dispersion relation is derived and analyzed, it is seen that what propagates in regions where propagation is allowed is a combination of a photon and TO phonon: the phonon polariton. When the system is excited near its ω_{TO} resonance, energy can be transferred from the field to the crystal. In addition, it will be seen that propagation is forbidden in the region: $\omega_{\text{TO}} \leq \omega \leq \omega_{\text{LO}}$, where ω_{TO} and ω_{LO} are characteristic frequencies for transverse and longitudinal oscillations, respectively. When the system is without loss, the transitions at ω_{TO} and ω_{LO} in a plot of reflectivity versus frequency are sharp. Of course no system is completely without loss. When oscillation is damped, the sharpness of the transition the between propagating and non-propagating regions is lessened in proportion to the amount of loss.

It is useful to examine metals and semiconductors, as their highest energy electrons are labile, as opposed to the bound charges of ionic crystals. First, we shall review Maxwell's equations, which are summarized in the box on the next page. I will go through this on the board, of course calling for volunteers from time to time.

Maxwell's Equations

SI (mks) units		Gaussian (cgs) units	
microscopic	**macroscopic**	**microscopic**	**macroscopic**
$\nabla\times\vec{B}=\mu_0\vec{j}+\frac{1}{c^2}\partial_t\vec{E}$	$\nabla\times\vec{H}=\vec{j}_f+\partial_t\vec{D}$	$\nabla\times\vec{B}=\frac{4\pi}{c}\vec{j}+\partial_{ct}\vec{E}$	$\nabla\times\vec{H}=\frac{4\pi}{c}\vec{j}_f+\partial_{ct}\vec{D}$
$\nabla\times\vec{E}=-\partial_t\vec{B}$	$\nabla\times\vec{E}=-\partial_t\vec{B}$	$\nabla\times\vec{E}=-\partial_{ct}\vec{B}$	$\nabla\times\vec{E}=-\partial_{ct}\vec{B}$
$\nabla\cdot\vec{B}=0$	$\nabla\cdot\vec{B}=0$	$\nabla\cdot\vec{B}=0$	$\nabla\cdot\vec{B}=0$
$\nabla\cdot\vec{E}=\rho/\varepsilon_0$	$\nabla\cdot\vec{D}=\rho_f$	$\nabla\cdot\vec{E}=4\pi\rho$	$\nabla\cdot\vec{D}=4\pi\rho_f$
	$\vec{D}=\varepsilon_0\vec{E}+\vec{P}=\varepsilon\vec{E}$		$\vec{D}=\vec{E}+4\pi\vec{P}=\varepsilon\vec{E}$
	$\vec{B}=\mu_0(\vec{H}+\vec{M})$		$\vec{B}=\vec{H}+4\pi\vec{M}$

Subscript *f* on $\vec{j}$ and/or ρ means free; subscript *b* means bound; no subscript means total.
SI: $\rho_b=-\nabla\cdot\vec{P}$; $\vec{j}_b=\nabla\times\vec{M}+\partial_t\vec{P}$

Poynting vector: $\vec{S}$ = power per unit area

$= \frac{c}{4\pi}\vec{E}\times\vec{B}$ (Gaussian, vacuum) $= \frac{c}{8\pi}\mathrm{Re}(\vec{E}\times\vec{B}^*)$ (complex exponent)

$= \vec{E}\times\vec{H}$ (SI).

Energy density: $\frac{1}{2}\vec{E}\cdot\vec{D}+\frac{1}{2}\vec{B}\cdot\vec{H}$ (SI).

Constitutive relations: $\vec{D}=\varepsilon\vec{E}$; $\vec{B}=\mu\vec{H}$; $\vec{j}=\sigma\vec{E}$.
In general, ε, μ, and σ are second-rank tensors and functions of ω.

Linearly polarized plane waves: $\vec{E}=E_0e^{i(kz-\omega t)}\hat{x}$ and $\vec{B}=B_0e^{i(kz-\omega t)}\hat{y}$

Potentials: $\vec{B}=\nabla\times\vec{A}$ and $\vec{E}=-\nabla V-\partial_{ct}\vec{A}$ (Gaussian).
$\vec{B}=\nabla\times\vec{A}$ and $\vec{E}=-\nabla V-\partial_t\vec{A}$ (SI).

Gauge invariance: Adding the gradient of a scalar: $\vec{A}\to\vec{A}+\nabla\chi$, leaves $\vec{B}$ unchanged because $\nabla\times\nabla\chi\equiv 0$. Adding $-\partial_{ct}\chi$ (Gaussian): $V\to V-\partial_{ct}\chi$, leaves $\vec{E}$ unchanged.

Coulomb gauge: $\nabla\cdot\vec{A}=0$ (SI and Gaussian).

Lorenz gauge: $\nabla\cdot\vec{A}+(1/c)\partial_{ct}V=0$ (SI); $\nabla\cdot\vec{A}+\partial_{ct}V=0=\partial_\nu A^\nu$ (Gaussian).

3.3. Plasmas in Metals and Semiconductors

"A plasma is a quasi-neutral gas of charged and neutral particles which [sic] exhibits collective behavior."

Francis Chen
Plasma Physics and Controlled Fusion

Before examining the propagation of electromagnetic radiation in ionic crystals, let us work through the related case of propagation in the presence of what is frequently referred to as a sea of free electrons. This is an interesting construct: a vast fluid-like medium consisting of electrons that are known to interact strongly with one another, but with each electron managing to behave, to a large extent as if it were independent. It is almost as if the other electrons did not carry charge. A sea of positive charge must exist in the background to compensate the sea of electron charge. Otherwise the system would undergo what is referred to as a Coulomb explosion.

A sensible way to begin is through consideration – at a conceptual and to the extent possible non-mathematical level – of some important properties of conduction band electrons in a metal. This is the focus of the present section. Wave propagation in such media is discussed in Section 4.1 of Chapter 4, and optical properties of metals and semiconductors are discussed in Section 4.2.

It is well known that a metal's conduction band electrons travel significant distances (relative, say, to a lattice spacing) while encountering modest resistance, hence the term conduction band. Consequently, it is often said that they move about freely. Though conduction band electrons in a metal undoubtedly experience modest resistance, a term like free electron is a misnomer. A given electron is not free. The conduction band electron density is too high to justify a free electron picture, at least without taking other considerations into account. Each conduction band electron interacts with many nearby electrons, presenting a strongly correlated many-body system.

We shall see that this property plays a central role in the plasmas that are encountered in metals and doped semiconductors. Electrons near the Fermi level are responsible for many important dynamical processes. Here, the many-body nature of the electron dynamics seems inescapable, rendering the independent electron picture untenable. As fortune would have it, in place of a free electron, a "free entity" emerges to save the day. It is an elementary excitation referred to as a fermion quasiparticle or a quasi-electron.

Tradition

A traditional approach to wave propagation in plasma begins with the introduction of an expression for the current density: $\vec{j} = -en_e\vec{v}$, where $-e$ is electron charge, n_e is electron density, and $\vec{v}$ is electron velocity. The velocity is calculated under the assumption that an electron is accelerated by an electric field and simultaneously undergoes mo-

mentum relaxation that can be described by using a rate parameter, γ. Namely, the rate of momentum loss is taken to be $\gamma m\vec{v}$. The use of a single relaxation parameter is *ad hoc* and phenomenological. Though several loss mechanisms probably participate, a single parameter suffices for an elementary treatment. The current density $\vec{j}$ is put into Maxwell's equations yielding expressions that describe the plasma's dielectric properties.

One could not ask for a friendlier ansatz insofar as the math is concerned. However, the ansatz also needs to make sense in physical terms. In this regard important conceptual issues need to be addressed before commencing with the math. Most important conceptual issues enter with the introduction of a model, not the subsequent math.

For most metals, conduction band electron density ranges from a few $\times 10^{22}$ cm^{-3} to more than 10^{23} cm^{-3}. To get a rough idea of nearest-neighbor spacing, let us take, as a figure-of-merit, the quantity $n_e^{-1/3}$. This yields 2.15 Å for $n_e = 10^{23}$ cm^{-3}. For the use of a classical expression like $\vec{j} = -en_e\vec{v}$ to have even approximate validity requires that the quantum description of the particles is treated adequately using free-particle wave packets. This is implicit when writing $\hbar^2k^2/2m$ for the kinetic energy. An electron's de Broglie wavelength, λ_{deB}, provides a rough measure of the extent to which the electron can be localized with a wave packet. However, when $\hbar^2k^2/2m$ is calculated using λ_{deB} = 2.15 Å, a value of 32 eV is obtained. This extremely large value underscores the inconsistency mentioned above. Namely, without taking into account additional physics, it is impossible to enlist the classical or quasiclassical regime, because the conduction band electron density is too high to permit this.

What we have is a fluid-like medium. It is comprised of a vast sea of conduction band electrons superimposed on a positively charged, stationary medium that ensures overall electrical neutrality. These electrons must be treated as a quantum fluid. Because of fundamental issues like this, the present section is devoted to the properties of conduction band electrons in metals. Let us now begin.

Electron Density and Free Particle Character

An obvious question looms: Despite a metal's high conduction band electron density, why do these electrons seem to move about freely? An answer frequently given in elementary treatments is that the charges are screened from one another. For example, the positively charged lattice is rendered nearly neutral by the vast sea of electrons, with a given electron experiencing the positive charge only when it is quite near the lattice sites, as illustrated schematically in Fig. 6. Furthermore, it is assumed that electrons are efficiently screened from one another. This latter assertion requires a great deal of physics for its justification.

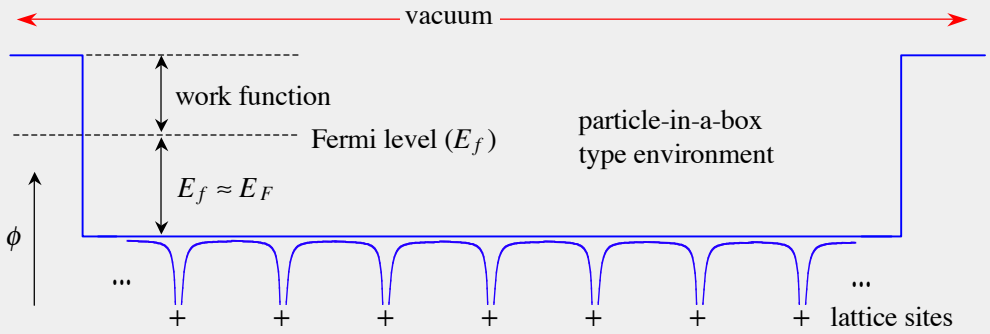


Figure 6. This schematic drawing illustrates a common rationale for using particle-in-a-box states to describe electrons in a metal. As a consequence of screening, electrons experience the positively charged lattice only when they are quite near the nuclei, whose locations are indicated with + signs. Likewise, the assumption is made that the electrons do not sense one another because of dynamical screening (yet to be described). Consequently, the conduction band electrons are assigned, as a crude first approximation, to particle-in-a-box states. For a metal at normal temperatures, the Fermi level (E_f , which is also referred to as the chemical potential, μ) and the Fermi energy E_F (the energy of the highest occupied state at $T = 0$ K) can be taken as equal to one another to a high degree of accuracy.

The above scenario was enlisted as justification for the Kronig-Penney model in Part A, where a positively charged lattice was represented mathematically using highly localized (sometimes delta function) site potentials. Figure 1 illustrates the general idea for a 1D slice through a lattice. The positively charged lattice turns out to have little influence on "nearly free electron" Bloch waves, manifesting only in the vicinities of the edges of the Brillouin zone ($k = \pm \pi/a$). The implicit assumption of effective screening of the positively charged lattice generally turns out to be valid. The electron-electron interactions are subtler, both conceptually and mathematically. For the most part this topic is avoided in elementary treatments. The positively charged lattice is fixed in space. It is not going anywhere, whereas electron-electron interactions can cause electron density to fluctuate.

If two electrons (spin up and spin down) are placed in each particle-in-a-box level, the states are occupied in a way that resembles the freshman chemistry picture of filling atomic orbitals to create ground state atoms. However, in the case of conduction band electrons, the highest occupied states are subject to thermal fluctuations, because energy differences among the states are small. As mentioned earlier, the Fermi energy, E_F , is defined as the energy of the highest occupied electron state at $T = 0$ K. This is indicated schematically in Fig. 7. It is noted that the Fermi energy E_F and Fermi level E_f , for all practical purposes, can be taken as equivalent for metals at normal temperatures. They are sometimes used interchangeably in the literature. Relative to the bottoms of their respective conduction bands, the Fermi energies of metals exceed 5 eV for the good electrical conductors: Cu, Ag, Au, and Al, specifically, 7.0, 5.5, 5.5, and 11.7 eV, respectively. Thus, even at 0 K, conduction band electrons have substantial kinetic energy.

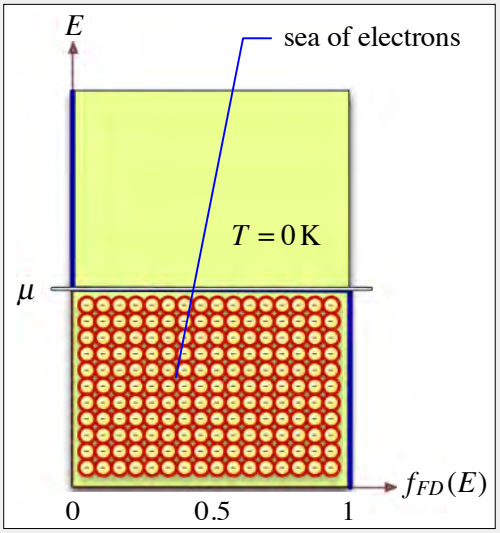


Figure 7. At $T = 0$ K, the probability of occupying a state is, by definition, unity below the Fermi energy, E_F, and zero above it. At $T = 0$ K, no electrons have $E > \mu$, where μ is the chemical potential. For a metal, the Fermi energy and chemical potential (Fermi level) are equal at 0 K. The figure uses coarse-grained energy resolution, say δE, such that 16 electrons are allowed in δE, whereas each microstate contains at most one electron. This figure is courtesy of Chris Winstead.

To get an idea of metal conduction band electron densities, let us assign to each electron (arbitrarily) the volume

$$\mathrm{V}^{\mathrm{el}} = \frac{4\pi}{3}\left(r_s a_0\right)^3. \tag{3.2}$$

The dimensionless parameter r_s is a figure-of-merit. It is given in units of Bohr radius a_0 (0.53 Å). For example, an r_s value of 2 corresponds to a sphere of radius of $2a_0$ = 1.06 Å. The parameter r_s is referred to in the literature as the characteristic radius. Sometimes it is dimensionless (as here) and sometimes it is not, in which case the right hand side of eqn (3.2) is $4\pi r_s^2/3$. The conduction band electron number density, n_e, is set equal to the inverse of V^{el}, and n_e is usually known with good accuracy. Thus, r_s values are easily obtained for many metals. The use of r_s is an interesting parameterization. Do not think of spheres packed into a solid. The metal volume could also be divided into little cubes with one electron assigned to each cube, again a figure-of-merit.

The conduction band electron density is equal to the metal atom density times the number of conduction band electrons per metal atom. For example, the monovalent metal copper has $n_e = 8.47\times10^{22}$ cm^{-3}, and using this with eqn (3.2) gives $r_s = 2.67$. The point here is not to obtain accurate n_e values, but to see qualitatively the high density of the conduction band electrons. No matter how one chooses to look at it, the conduction band electron density is too high to accommodate a free electron picture, at least without other considerations taken into account, notably screening. Of course, the overall electron density is much higher than n_e, as each Cu atom has 28 other electrons. The point is that a density like 8.47×10^{22} cm^{-3} for the conduction band electrons is remarkably high if we are to attempt to assign free electron character to these electrons. Table 1 lists a few relevant numbers.

The de Broglie wavelength of a 7 eV free electron (the energy of a copper conduction band electron at the Fermi energy) is roughly 4.2 Å. As mentioned above, a reasonable criterion for deciding whether an electron can be judged as free is based on its de Broglie wavelength, as this determines the spatial extent of wave packets that can be created. For energies lower than E_F, the wavelengths are larger than 4.2 Å. In fact, toward the bottom of the conduction band, they are considerably larger than 4.2 Å. Thus, in light of the fact that r_s = 2.67 for copper, the electron density is too high to accommodate free particle behavior without taking screening into account. We shall see that even when electron-electron screening is taken into account the electrons are still too close to one another to support a free electron picture.

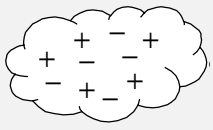
Table 1.

	$n_e / 10^{22}$	E_F / eV	r_s
Cu	8.47	7.0	2.67
Ag	5.86	5.5	3.02
Au	5.90	5.5	3.01
Al	18.07	11.7	2.07

Some properties of several plasmonic metals

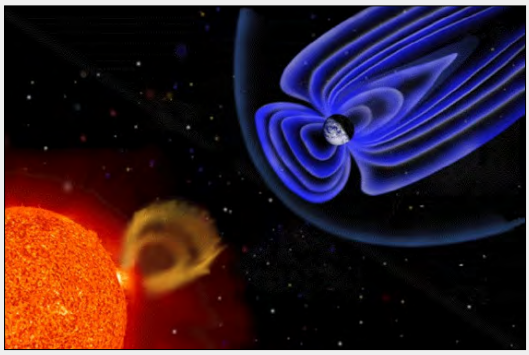
The bottom line is that it is not feasible to judge the conduction band electrons in copper as being free, and the same applies to other metals. Though efficient screening lessens considerably long-range Coulomb interactions, significant electron-electron interaction and correlation nonetheless remain.

It turns out that conduction band electrons, despite not being free, can be assigned important properties that resemble closely the counterpart properties of free electrons. It is only in this sense – *through certain analogies with truly free electrons* – that the conduction band electrons of a metal can be regarded as free. At the same time, the extent to which the treatment of electrons as being free yields reasonably accurate results is uncanny, almost too good to be true. It seems paradoxical that an electron can be only a couple of Å from its nearest neighbors and yet behave as if it were free. We shall see that reconciliation is subtle, going beyond simple screening models. It enlists the theory of Fermi liquids, which, interestingly, was introduced by Landau rather than Fermi. This theory was introduced originally to explain the fermionic liquid: ^{3}He, but it also applies, with modifications, to conduction band electrons in metals.



Plasma has overall charge neutrality.

Irrespective of the actual behavior of such electrons, there can be no doubt that the environment experienced by conduction band electrons in a metal differs qualitatively from that of the bound charges of an ionic crystal. In the present and following sections, we shall consider several regimes of the dielectric properties and electron dynamics associated with electromagnetic fields interacting with a sea of so-called free electrons superimposed on a positive ion substrate that renders the overall system electrically neutral. This situation is referred to as plasma, from the Greek word $\pi\lambda\alpha\sigma\mu\alpha$, meaning moldable substance or jelly. Not surprisingly, this situation is sometimes referred to as jellium in condensed matter physics.



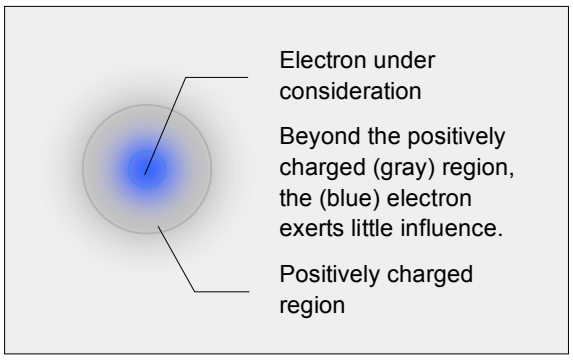
Plasma in earth's magnetosphere

We shall be concerned almost exclusively with solid materials that support plasmas, whereas there exist many different kinds of gaseous plasmas: the core of the sun, laser-driven inertial confinement, magnetic confinement, neon signs, the earth's magnetosphere, and so on. These will serve as interesting and useful examples from time to time, though they will not be belabored. The field of plasma physics deals with such plasmas, a number of which are in the relativistic regime, avoiding for the most part the solid-state counterparts of interest here: metals and doped semiconductors.

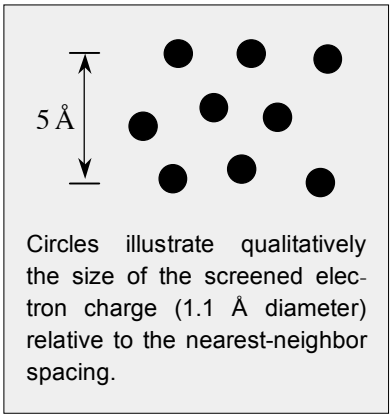
As mentioned earlier, despite the fact that a metal has overall charge neutrality, and its conduction band electrons are on average close to one another and experience strong Coulomb interaction, it is well known that they interact with electromagnetic fields in a manner that has a great deal in common with how free particles would interact. This is an interesting, important, and perhaps surprising fact. Before using it in a mathematical model of plasma, let us see how it comes about and under what conditions it is applicable.

Screening

One of the properties that distinguish a plasma is that its electrons interact collectively. This does not mean that these electrons *always* interact collectively. For example, the absorption of a photon whose energy is in slight excess of the work function of a metal can eject an electron from near the Fermi level into vacuum, as indicated in Fig. 6. In addition, interband transitions promote electrons between zero-order stationary states. What it *does* mean is that the plasma state is characterized by the collective behavior of the sea of electrons. There is no band gap within a conduction band, so the sea of electrons is highly polarizable. Plasma oscillation is that of the system's polarization.



This collective behavior goes hand-in-hand with effective screening in which the Coulomb potential emanating from a given conduction band electron in the metal falls off much more rapidly than the r^{-1} variation of an unscreened Coulomb potential. Screening occurs because a given electron repels nearby electrons, thereby exposing the positively charged, substrate, as indicated in the sketch on the right. On the one hand, this causes the potential to flatten, giving each electron a taste of freedom. On the other hand, they have nowhere to go without the cooperation of their conduction band neighbors. The electron density is simply too high to accommodate unencumbered motion of conduction band electrons that behave independently of one another.

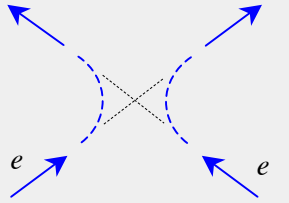


An electrical conductor like aluminum is a good example. Being trivalent, it presents a large conduction band electron density: 1.81×10^{23} cm^{-3}, which corresponds to r_s = 2.07 (radius of 1.10 Å). With so high a density, even an optimistic estimate of screening efficiency (sketch on the right) leaves enough electron-electron interaction to confound a free-electron picture. The electrons cannot get too close to one another because they would then be repelled strongly by the unscreened Coulomb potential. The environment is sufficiently dense that they are prevented from following straight-line paths.[5] Inter-electron spacing is comparable to de Broglie wavelengths at the Fermi level, with the de Broglie wavelength becoming progressively larger as energy drops below the Fermi level. It turns out that as a result, the conduction band electrons move together as an ensemble, a field of electron density, so to speak.

[5] We shall see that dynamical screening, in which the screened region accompanies the electron as it moves, helps reconcile the apparently free-particle type behavior.

Electron-Electron Scattering

Suppose two electrons scatter from one another while undergoing no additional interactions. Within their center-of-mass (c.m.) system, momentum is zero before and after interaction. Motion *of* the c.m. is unaffected, so the overall momentum of the pair is conserved. The same argument applies to 3-body and higher interactions. As long as the group of particles is taken as isolated, their overall momentum is conserved. Scattering of particles *within* a c.m. system cannot change the overall momentum *of* the c.m. system.



Direct and exchange scattering in a binary encounter. Near the center, the fact that electrons are indistinguishable enters.

Now consider a system consisting of many electrons. A given pair of electrons that scatter from one another without interacting with the other electrons do so, in general, with nonzero momentum of the pair in the overall many-body c.m. system. The momentum of the pair is preserved: $\vec{k}_1 + \vec{k}_2 = \vec{k}_3 + \vec{k}_4$, regardless of the reference frame used to define the wave vectors. And so on for the higher-order interactions that define the many-body system. Thus, electron-electron scattering can shift electrons back and forth through the Fermi surface, but without changing the overall momentum. Electrons that lie more than a few kT below the Fermi level have a hard time exchanging momenta because of the small number of unfilled levels … blah blah

Such considerations led Paul Drude (around 1900), to implicate scattering by lattice vibrations as the dominant electron momentum relaxation mechanism. Shortly after his work, scientists realized that the electrical conductivity of a metal having a defect-free stationary lattice is infinite.

In Chapter 2, thermal conductivity due to phonons was discussed, and we saw that phonon crystal momentum could be relaxed through an umklapp process. Namely, if the vector addition of the wave vectors of the interacting phonons gives a wave vector that lies outside the first Brillouin zone, a reciprocal lattice vector can be used to return the large wave vector to the first Brillouin zone. The same idea applies to the Bloch waves that characterize the electrons in a positively charged lattice. In this case, however, the wave vector of a phonon is needed. That is, electron-electron scattering creates and annihilates phonons. The details are fairly mathematical and would take more time than is available. The result is that for normal temperatures, electron-electron scattering plays a minor role.

Back to Screening

Screening is a conceptually straightforward phenomenon that arises in a broad range of environments: solid dielectric insulators, mobile charge carriers like anions and cations in solution, proteins, electrons in metals, electrons in semiconductors, and numerous oth-

er situations. Wherever there are positive and negative charges there is screening.[6] It is just a matter of its significance in a given system. Its importance and generality cannot be overstated, as it is central to a vast array of scientifically and technologically important materials, environments, and phenomena. Figure 8 gives examples from Google Images.

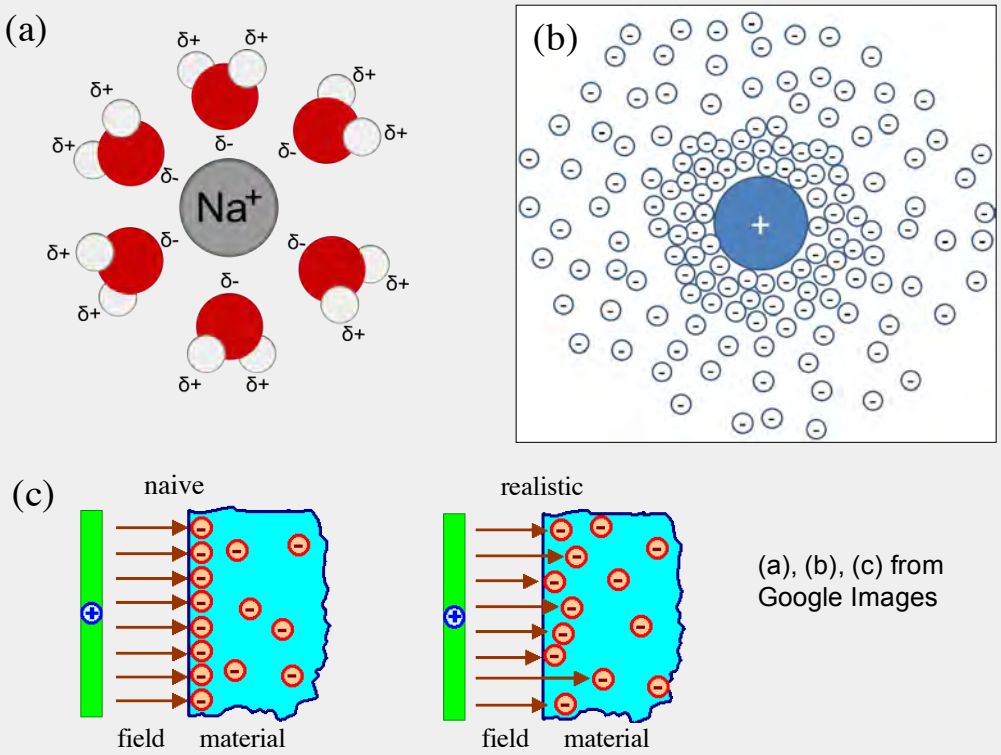


Figure 8. Examples of screening: (a) A positive ion aligns water dipoles. Imagine each water molecule rotated about its center-of-mass to the low energy configuration shown. The effect is exaggerated, as thermal fluctuations lessen dramatically the degrees of alignment and orientation. The field of the ion is diminished outside a few solvation shells. (b) A large amount of positive charge in a sea of mobile negative charges draws the negative charges toward it, thereby screening the positive charge. (c) Mobile ions in solution facing a positive voltage electrode do not pile up neatly at a surface, but arrange according to a free energy relationship.

[6] To point out an extreme example that you might find amusing: even the charge of an electron in vacuum is screened. That's right, the electron charge we deal with routinely (1.602177×10^{-19} C) is not the charge of a truly bare electron. The truly bare electron polarizes the vacuum. This polarization is represented as unstable (transient) electron-positron pairs, with the positrons pointing toward the stable electron. This phenomenon follows from quantum electrodynamics, and it has been verified experimentally.

At the same time, screening is one of those things that seem eminently reasonable at a conceptual level, though difficult to calculate with good accuracy. Professor Reisler tells me that the freshmen are comfortable with the concept. Professor Krylov tells me that, as a research topic, conduction band electrons are vexing because of the strong correlation.

In the next few subsections, simple models are introduced and discussed in order to obtain a qualitative understanding of screening mechanisms in metals and doped semiconductors. The time-honored Thomas-Fermi, Lindhard, and Debye-Hückel models are straightforward, and each will be discussed briefly. The Thomas-Fermi model yields a characteristic length, r_{TF}, that parameterizes the potential presented by a charge in a dense sea of mobile charge carriers, in our case electrons. The charge being screened can be one of the electrons, or a "test charge" that is introduced to assist visualization. We shall see that the screened potential does not display the r^{-1} variation of the Coulomb potential. Instead the potential varies as $r^{-1}\exp(-r/r_{TF})$. The Thomas-Fermi model is unsophisticated and premised on ambitious assumptions. It is an excellent pedagogical model that provides qualitative guidance.

Lindhard introduced a more accurate model. It has longer range and richer structure than Thomas-Fermi. Finally, the Debye-Hückel model is discussed. To the best of my knowledge, it is the first mathematically derived model of screening by mobile charge carriers. Helpful texts insofar as screening are Ibach and Lüth [4], Ashcroft and Mermin [1], Ziman [9,10], and Kittel [5]. A text by Coleman [29] is scheduled for publication in October 2015. In the meantime I have an e-version. The text by Amit and Verbin [28] treats statistical thermodynamics in a way that is accessible to a broad audience. The classic work of Pines and Nozières [32] is excellent, but specialized (quantum liquids) and more advanced than the others. These and others are in our course library.

Screening is a manifestation of many-body interactions. In the subsection entitled: *Quasiparticles to the Rescue*, we shall see that electron-electron interactions that manifest as screening also account for the high degree of free-electron character that emerges. Specifically, the paradoxical combination of significant free-electron character, and at the same time strong electron-electron interactions, is reconciled when the electron is replaced by its corresponding quasiparticle according to Landau's theory of Fermi liquids.

Thomas-Fermi

The Thomas-Fermi screening model is now derived. To begin, it is assumed that a positive point charge Q is placed in a vast sea of electrons superimposed on an equally vast sea of stationary positive charge. Without Q, the system has overall electrical neutrality. With Q, the system can still be treated as having overall electrical neutrality, because Q is minuscule in comparison to the vast sea of positive charge. The location of Q is defined as the origin. That is, $Q(\vec{r}) = Q\delta(0)$, where $\delta(0)$ is the delta function.

The role of Q is that of a "test charge." One of the electrons already present in the sea of electrons could be used rather than Q, but I prefer the addition of a positive test charge because it stands apart. Though a positive test charge is used here, a negative test charge could just as well be used, as either must yield the same result. Before adding Q, the equilibrium value of the electrostatic potential, $\phi(\vec{r})$, which arises from all of the positive

and negative charges, is assumed constant throughout the sample. There are significant local fluctuations. On a sufficiently large length scale, however, the system appears uniform and quiescent.

It is significant that an assumption is made here about the smoothness of the potential. It is assumed that relevant length scales are large enough to justify the assumption of a smoothly varying potential. In other words, we are focusing on the small wave vector regime. The equilibrium value of $\phi(\vec{r})$ before Q is added is conveniently set to zero, enabling us to deal only with variations. In addition, it is assumed that the positive charges including Q are immobile, fixed to a stationary lattice.

The change in electron density brought about by Q is denoted $\delta n_e(\vec{r})$, and the change in the electrostatic potential $\phi(\vec{r})$ (whose value before adding Q is zero) is denoted $\delta\phi(\vec{r})$. With the above ansatz, the microscopic Maxwell equation: $\nabla\cdot\vec{E} = 4\pi\rho$, in the electrostatic limit becomes

Dirac delta function

$$-\nabla^2\delta\phi(\vec{r}) = 4\pi\left(Q\delta(0) - e\delta n_e(\vec{r})\right). \tag{3.3}$$

Contributions from the seas of positive and negative charges cancel, leaving Q.

Solving eqn (3.3) yields the screened potential $\delta\phi(\vec{r})$. The Laplacian operating on $\delta\phi(\vec{r})$ retains just the radial part, as $\phi(\vec{r})$ has no angular variation – a welcome simplification.[7] Consequently, for the present scenario of an isotropic screened potential, $\delta n_e(r)$ and $\delta\phi(r)$ can be used rather than $\delta n_e(\vec{r})$ and $\delta\phi(\vec{r})$, namely, $\vec{r}$ is replaced with r. To relate $\delta n_e(r)$ to $\delta\phi(r)$, we shall derive an expression for $\delta n_e(r)$ in terms of $\delta\phi(r)$, and plug this expression into eqn (3.3). This requires something called the chemical potential, which we shall proceed to examine.

Chemical Potential

We shall enlist something called the chemical potential as a means of establishing the relationship between $\delta n_e(r)$ and $\delta\phi(r)$ that is needed to solve eqn (3.3). The chemical potential plays a central role in the properties of metals and semiconductors. In materials science and solid-state physics it is usually referred to as the Fermi level. It is a thermodynamic quantity that, in the present context, will be derived through consideration of the Helmholtz energy.[8]

---

[7] The radial part of the Laplacian in spherical coordinates is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi).$$

[8] The word *free* was dropped from terms like Helmholtz free energy and Gibbs free energy by IUPAC in 1988. However, tradition does not surrender easily, so one frequently hears reference to Gibbs free energy and the like.

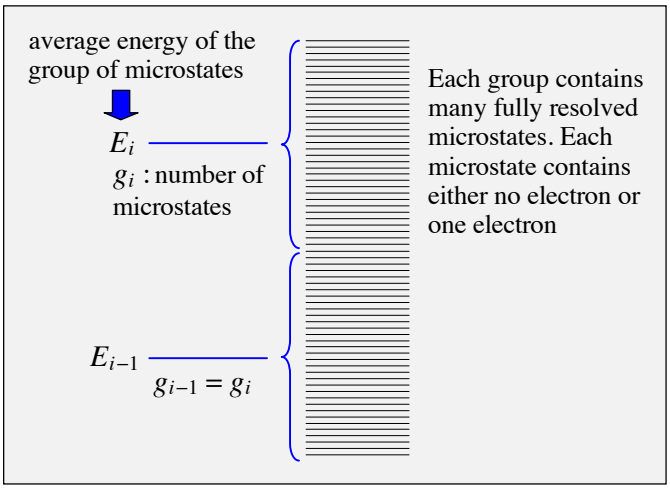
Whereas the Fermi energy is the energy of the highest occupied state at $T = 0$ K, the Fermi level is temperature dependent. We shall distinguish their symbols with subscripts: E_F and E_f, respectively. Of course, $E_f(T = 0) \equiv E_F$, and this is probably how the term Fermi level acquired its name. The chemical potential, which is usually denoted μ in chemistry and biochemistry, is discussed in numerous texts. You can undoubtedly find a more thorough discussion elsewhere, as thermodynamics is hardly my strong suit. Nonetheless, the succinct derivation presented below will suffice to get the main ideas across.

The chemical potential, being an intensive state function, has the same value at all locations in a sample, which in the present case is solid-state plasma.[9] Its fundamental defining relationship can be obtained through consideration of the Helmholtz energy, which is nearly always assigned a label of either *A* or *F*. Here we shall use the latter:

$$F = U - TS, \tag{3.4}$$

where U is the internal energy and S is the entropy. Temperature is assumed constant, and in our case the volume is also assumed constant.

Variation of F is now taken with respect to electron occupancy of the levels. To maintain generality, while at the same time facilitating the math, it is assumed that the i^{th} level has a degeneracy of g_i, and that it is possible to put n_i electrons in this level as long as $n_i \leq g_i$. Think of this degeneracy as an approximation to a group of closely spaced "microstates" (no degeneracy), as indicated in the sketch below.



Each microstate can accommodate at most one electron, meaning that occupancy is either zero or one. If we group, say, 1000 of these microstates together and call this group the i^{th} level whose degeneracy is $g_i = 1000$, we can put n_i electrons into this level, as long as $n_i \leq g_i$. This will enable us to obtain the average number of electrons in a level as a function of energy. Keep in mind the distinction between n_e (electron number density) and n_i, the number of electrons in the i^{th} degenerate level.

---

[9] The chemical potential (Fermi level) is, in general, temperature dependent, whereas the Fermi energy, E_F, is not. However, for metals and the temperatures under consideration here, the difference between the Fermi energy and the Fermi level is too small to worry about. This would not be the case for an undoped semiconductor, say, where E_F is the top of the filled valence band, whereas the Fermi level lies at the center of the band gap.

It is assumed that the system remains in equilibrium, in which case the net change in the Helmholtz energy brought about by the variation of F with respect to electron occupancy is zero. Focusing on a single pair of levels whose electron populations are allowed to vary, we write

$$\frac{\partial F}{\partial n_i}\delta n_i + \frac{\partial F}{\partial n_j}\delta n_j = 0 \,. \tag{3.5}$$

The fact that the number of particles is taken as conserved, $\delta n_i = -\delta n_j$, yields

$$\frac{\partial F}{\partial n_i} = \frac{\partial F}{\partial n_j} \,. \tag{3.6}$$

The levels were chosen arbitrarily, so it follows that $\partial F / \partial n_i$ is independent of the level chosen. Indeed, the change in Helmholtz energy with the addition of a particle is the definition of the chemical potential:

$$\mu \equiv \frac{\partial F}{\partial n_i} \,. \tag{3.7}$$

Next, an expression for $F = U - TS$ is obtained that enables eqn (3.7) to be applied. The internal energy U is the sum of $n_i E_i$ over all i. An expression for the entropy S is obtained by using its statistical properties, starting with

$$S = k \ln P \,, \tag{3.8}$$

where k is the Boltzmann constant, and P is the total number of ways of arranging fermions in all of the degenerate energy levels. Of course, P is the product of the numbers of ways of arranging particles in *each* degenerate level:

$$S = k \ln \prod_i P_i \,. \tag{3.9}$$

The symbol P_i denotes the number of ways n_i indistinguishable fermions can be placed in the i^{th} level, whose energy is E_i and whose degeneracy is g_i.

As mentioned above, the entropy, $k \ln P$, is obtained by multiplying all of the P_i together, with their statistical weight factors in place.[10] The resulting expression for P is

$$P = \prod_i \frac{g_i!}{n_i!(g_i - n_i)!}. \tag{3.10}$$

Now that we have expressions for U and S, the Helmholtz energy is expressed as

$$F = \sum_i n_i E_i - kT \ln \prod_i \frac{g_i!}{n_i!(g_i - n_i)!} \tag{3.11}$$

$$= \sum_i \left(n_i E_i - kT \left(\ln g_i! - \ln n_i! - \ln(g_i - n_i)! \right) \right). \tag{3.12}$$

Referring to eqn (3.7), the chemical potential is obtained by differentiating F with respect to any one of the n_i. That is, n_i is treated as generic. This eliminates the summation in eqn (3.12), yielding

$$\mu = E_i + kT \frac{\partial}{\partial n_i} \left(\ln n_i! + \ln(g_i - n_i)! \right). \tag{3.13}$$

This is simplified by using Stirling's approximation. For large N values, $N!$ is approximately equal to $N^N e^{-N}$, and therefore $\ln N!$ is approximately equal to $N \ln N - N$. You might find it interesting to try a few values of N in the above expression to see how large a value of N is needed to justify the use of Stirling's approximation.

Referring to eqn (3.13), differentiation with respect to n_i, and carrying out some algebraic manipulation,[11] yields

---

[10] The number of ways an object can be placed in g_i slots is g_i. This leaves $g_i - 1$ available slots. The next object to be placed in one of the accessible slots has only $g_i - 1$ options, and so on. For n_i objects, this gives, for the total number of options: $g_i(g_i - 1)...(g_i - n_i + 1)$, which is written

$$\frac{g_i!}{(g_i - n_i)!}.$$

However, the "objects" are identical fermions. Thus, there are $n_i!$ permutations that leave the system unchanged. This enters the denominator of the above expression, yielding eqn (3.10).

[11] Carrying out the differentiation yields

$$\frac{\partial}{\partial n_i} \left(\ln n_i! + \ln(g_i - n_i)! \right) = \frac{\partial}{\partial n_i} \left(n_i \ln n_i - n_i + (g_i - n_i) \ln(g_i - n_i) - (g_i - n_i) \right)$$

$$= \ln n_i + 1 - 1 - \ln(g_i - n_i) - 1 + 1 = \ln \frac{n_i}{g_i - n_i}.$$

$$\mu = E_i + kT \ln\left(\frac{n_i}{g_i - n_i}\right). \tag{3.14}$$

This is easily reorganized into the form

$$\boxed{n_i = g_i \frac{1}{e^{(E_i-\mu)/kT}+1}.} \tag{3.15}$$

When each level is counted explicitly (the levels are all taken as non-degenerate), the g_i in eqn (3.15) is equal to unity. However, in this case, we know that the number of electrons that can occupy the level is either zero or one (nothing in-between), and we must replace n_i with its average value, $\bar{n}_i = n_i / g_i$. Equation (3.15) then becomes

$$\boxed{\bar{n}_i = \frac{1}{e^{(E_i-\mu)/kT}+1}.} \tag{3.16}$$

The average value, $\bar{n}_i$, can be far smaller than unity for $E_i > \mu$. We have obtained its value by grouping microstates together into degenerate levels. This enabled Stirling's approximation to be applied. Following this, one of the degenerate levels was singled out. Note that we did not remove the degeneracy.

Maxwell-Boltzmann

As mentioned above, when the number of available states exceeds greatly the number of fermions (electrons), the average number of electrons in a given microstate is much less than unity. This can only happen if $e^{(E_i-\mu)/kT}$ is much larger than unity, in which case the +1 in the denominator can be disregarded. The availability of many unoccupied translational energy states means the electrons are free to move about insofar as their statistical properties are concerned. Of course, electrons carry charge, so in a metal they are not truly free to move about, say along straight-line paths. Rather, they interact with nearby electrons. Consequently, it is dynamically screened electrons (quasiparticles) that move about. Most importantly, electrons fill all of the low energy states, making it impossible for these electrons to move about, as this would require nearby empty states, say, within a few *kT*. The Maxwell-Boltzmann case does not apply to metals.

At the other extreme, in an intrinsic (undoped) semiconductor, $E_i - \mu >> kT$ in the conduction band, so the +1 in the denominator of eqn (3.16) can be disregarded. This is exactly the condition under which Maxwell-Boltzmann statistics prevails, and eqn (3.16) becomes

$$\bar{n}_i = e^{\mu/kT}\, e^{-E_i/kT}. \tag{3.17}$$

Summing over all states yields the total number of particles, n_0, on the left hand side. On the right hand side, the sum of all $e^{-E_i/kT}$ is the partition function, Z. Therefore, $e^{\mu/kT}$ is equal to n_0/Z, and eqn (3.17) becomes

$$\bar{n}_i = \frac{n_0}{Z} e^{-E_i/kT} . \tag{3.18}$$

This is the familiar Boltzmann distribution that we encounter frequently. We have not considered the case of bosons (Bose-Einstein statistics), where there is no limit to the number of particles that can be placed in a given state. We saw how this works with phonons. In fact, bosons *prefer* to occupy the same state. I will skip this for the time being, and, time permitting, add a box later that elaborates fermions, bosons (massive and massless), and the Maxwell-Boltzmann limit for massive particles. The mathematical difference for massive bosons is that the +1 in eqn (3.16) is replaced with –1, and the chemical potential, μ, needs further consideration. Bose-Einstein statistics applied to massive particles also yields the result given by eqn (3.18) when the limiting case of many more states than particles is taken. Note that in the above derivation it was not necessary to enlist an assumption that the particles are distinguishable, as is often done in discussions of Maxwell-Boltzmann statistics.

Back to the Chemical Potential

Let us return to Fermi-Dirac statistics, with eqn (3.16) written

$$f_{FD}(E) = \frac{1}{e^{(E-\mu)/k_B T} + 1} . \tag{3.19}$$

This is referred to as the Fermi-Dirac distribution. It applies to fermions, all of which have mass. At $T = 0$ K, eqn (3.19) describes a step function whose value is zero for $E > \mu$, and unity for $E < \mu$, as illustrated in Fig. 3.7. Pay no heed to the hype you can find nowadays in magazines like Science and Nature about massless Dirac fermions in graphene, Majorana fermions, and the like. In my opinion, this stuff falls under the heading: "playing with words."

The addition of an electron to a plasma that is in equilibrium to begin with must preserve the value of the chemical potential in the region where the electron is added if the system's equilibrium is to be maintained. This constancy of μ is general. It is true regardless of the details involved in the addition of charge to the system. The added particle need not be an electron, and it can have either positive or negative charge. An electron is special because it enters the sea of electrons near the Fermi level. In all cases, however, μ maintains its value throughout the sample.[12] There can be a transient period, for example,

[12] For the purist, this requires that the sample is infinitely large or connected to a heat bath. The derivation requires that the chemical potential be unchanged by the addition of a charged particle. The important point is that μ remains constant locally.

if a charged particle is injected instantaneously, say at $t = 0$. However, after the transient period has passed (*vide infra*, Fig. 10), the chemical potential in the region where the charged particle was injected will have returned to its original value. For the case of the positive charge Q drawing electron density toward it, the electron density increases in the region of lower electron potential energy. Figure 9 illustrates this.

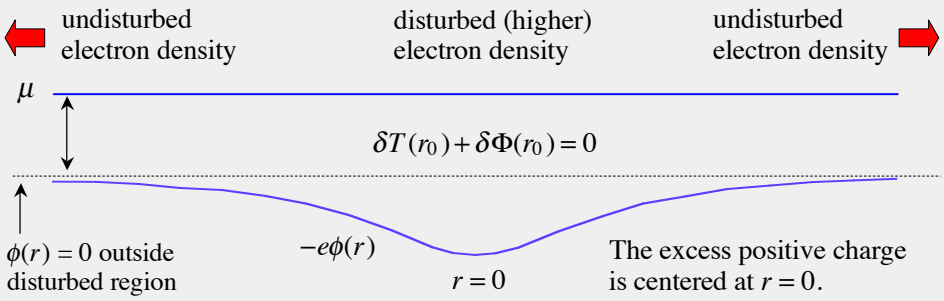


Figure 9. Assume that excess positive charge Q is distributed over a small spherical volume centered at $r = 0$. For example, this could be due to an impurity at one of the lattice sites, where the impurity has larger positive charge than the rest of the lattice sites. The potential energy of an electron, $\Phi(r) = -e\phi(r)$, in the region of the positive charge is lower than elsewhere. Consequently, the electron's kinetic energy, T, is higher than elsewhere in order to satisfy $\delta T(r) + \delta\Phi(r) = 0$.

Injecting an Electron

The above reasoning is now applied to neutral plasma in order to further understand how changes in kinetic and potential energies manifest. We shall then return to the case of the test charge $+Q$ and finish the derivation of the Thomas-Fermi screening model. Referring to Fig. 10(a), the system begins with $\phi(r) = 0$ and level occupancy that varies as the product of the density of states and the Fermi-Dirac distribution function. That is, level occupancy is proportional to $E^{1/2} f_{FD}(E)$. A sharp (vertical) cutoff at the Fermi level was used in making Fig. 10 because it was easier to draw, and, on the scale used in the figure, the sharp cutoff is not much different than the 300 K curve (*vide infra*, Fig. 13). In Fig. 10(b), we see that, because entropy is constant, injecting an electron instantaneously (say, at $t = 0$) into a local region centered at r_0 incurs subsequent changes in the kinetic energy, $T(r_0)$, and the potential energy, $\Phi(r_0)$. In writing $T(r_0)$ and $\Phi(r_0)$, it is understood that parenthetic r_0 includes the *region centered at* r_0, not just the point r_0. As indicated in Fig. 10, to maintain the value of μ, these changes must sum to zero:

$$\delta T(r_0) + \delta\Phi(r_0) = 0. \tag{3.20}$$

Because the plasma was in equilibrium before the electron was injected, the change in potential energy $\delta\Phi(r_0)$ is positive, and $\delta T(r_0)$ is negative. Figure 10 illustrates this.

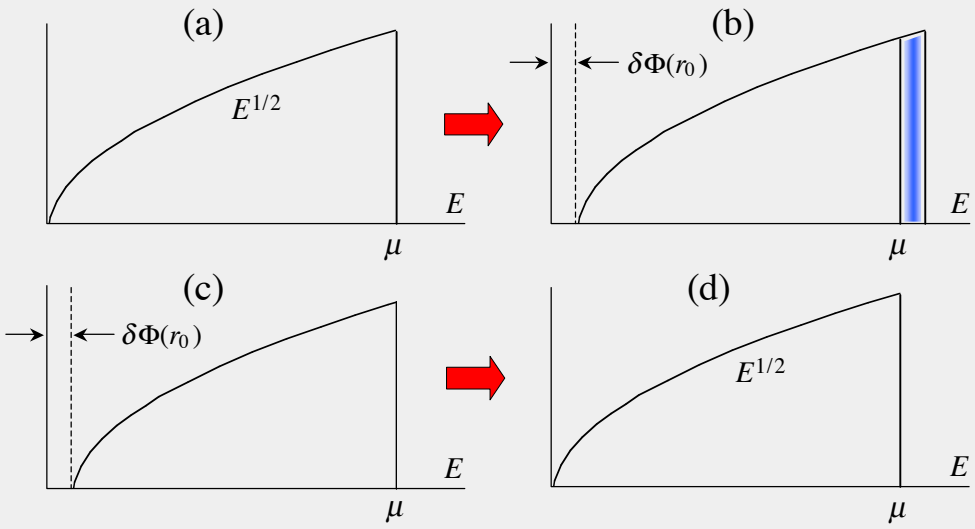


Figure 10. The system starts at (a). The curve $E^{1/2} f_{FD}(E)$ is approximated by the $T = 0$ version, which is simply $E^{1/2}$ truncated at μ. This is easier to draw, and at 300 K, variation about the Fermi level (chemical potential μ) is small. For μ to remain constant everywhere, the local change in U must be zero, as entropy is constant. In (b) the potential energy is perturbed instantaneously by the appearance of an electron. This shifts the kinetic energy of the density of states parabola to larger E. Because the chemical potential must be constant throughout, the electron density associated with the states colored blue must move out of the region of perturbed potential, yielding the situation shown in (c). This is also unstable because the added charge, being mobile, moves out of the region, yielding (d), which is identical to (a).

Recall that the equilibrium value of $\phi(r)$ was set to zero as a matter of convenience. When the injected electron appears in the region near r_0 it sees the positively charged lattice as well as the sea of electrons. Because of its presence, the local electron density exceeds the positive charge density, so the electrostatic potential $\phi(r)$ is lowered, and the electron's potential energy, $\delta\Phi(r_0) = -e\delta\phi(r)$, is raised. This displaces the entire kinetic energy parabola by an amount $\delta\Phi(r_0)$ above μ, as shown in Fig. 10(b). However, the electrons indicated with blue shading must therefore leave the region, as indicated in Fig. 10(c). In due course the electron spatial distribution recovers its original form and the density of states parabola again starts at the origin. Thus, entries (a) and (d) are the same.

Back to Thomas-Fermi

The Thomas-Fermi model is premised on the fermion nature of electrons manifesting as their occupancy of states according to the Fermi-Dirac distribution function given by eqn (3.19). In light of the fact that the electrons are not free in the sense of single particle kinetic energies: $\hbar^2 k^2 / 2m$, let us say that, rather than dealing with a true electron, something "electron-like" behaves as a free spin ½ fermion. For the quantum mechanical picture under consideration, this implies that single particle wave functions can serve as useful zero-order descriptors. It turns out that single-particle wave functions do not suf-

fice. Electrons near the Fermi level are replaced by quasiparticles. We shall continue for the time being with the single-electron picture, but keeping this in mind. Figure 11 illustrates eqn (3.19) for the cases of metals and undoped semiconductors.

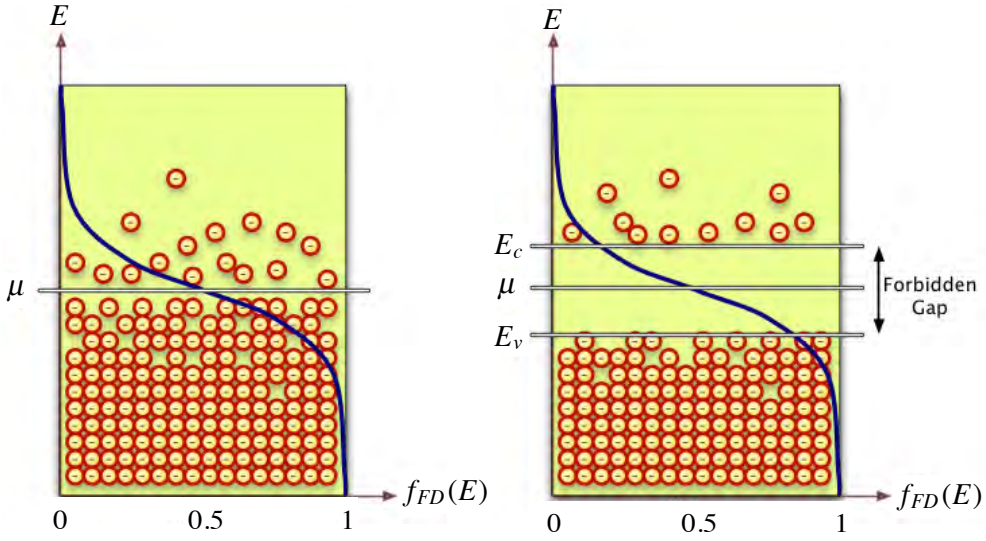


Figure 11. (a) For a metal at $T > 0$, the probability of occupying a state at energy E is given by the Fermi-Dirac distribution function [eqn (3.19)], which distributes electrons and holes about the Fermi level. The width of the spread about μ is exaggerated to facilitate visualization. (b) In the case of an undoped semiconductor, a quite small percentage of the electrons are in the conduction band. (Figure courtesy of Chris Winstead)

We have seen that the change in kinetic energy brought about through the addition (or removal) of an electron at the Fermi level is related to the corresponding change in electron number density. The relationship between n_e and kinetic energy in the disturbed region is premised on the fact that outside of this region,[13]

$$n_e = zE_f^{3/2} \tag{3.21}$$

applies, where $z = (2m/\hbar^2)^{3/2}/3\pi^2$. The changes can therefore be expressed as

[13] Quantization of translational motion is carried out using box boundary conditions. The allowed k values along one direction are $n_x\pi/L$; likewise for the other directions. The k-space volume $(\pi/L)^3$ corresponds to a single state, excluding spin. The number of states in a 3D shell in k-space is given by: the volume of the shell ($4\pi k^2 dk$), times 2 to account for spin, divided by 8 because only one quadrant of the spherical shell is used, and divided by the volume per state, $(\pi/L)^3$. This is converted to energy space using $dE = (\hbar^2/m)kdk$. The number of states per unit volume (L^3) is given by

$$\text{number of states} = \frac{1}{L^3}(4\pi k^2 dk)\,2\,\frac{1}{8}\,\frac{1}{(\pi/L)^3} = \frac{(2m/\hbar^2)^{3/2}}{2\pi^2}E^{1/2}dE\,.$$

Integrating the last term from 0 to E_f yields eqn (3.21) with $z = (2m/\hbar^2)^{3/2}/3\pi^2$.

$$\delta n_e(r) = \tfrac{3}{2}\left(zE_f^{1/2}\right)\delta E_f \,. \tag{3.22}$$

As mentioned earlier, an increase in electron density, $\delta n_e(r)$, lowers the electrostatic potential ($\delta\phi(r)$ is negative). In turn, this raises the potential energy, $\delta\Phi(r) = -e\delta\phi(r)$, of an electron in the disturbed region. Consequently, the kinetic energy at the Fermi level is lowered (δE_f is negative). Finally, replacing δE_f with $e\delta\phi(r)$ yields

$$\delta n_e(r) = \frac{3ezE_f^{1/2}}{2}\,\delta\phi(r)\,. \tag{3.23}$$

Thus, the required relationship between $\delta n_e(r)$ and $\delta\phi(r)$ has been obtained. Putting eqn (3.23) into eqn (3.3) yields the differential equation for the screened potential:

$$-\nabla^2\delta\phi(r) = 4\pi Q\delta(0) - \underbrace{6\pi e^2 zE_f^{1/2}}_{\equiv k_{TF}^2}\,\delta\phi(r)\,. \tag{3.24}$$

The expression for k_{TF}^2 ($6\pi e^2 zE_F^{1/2}$) can also be expressed in terms of the density of states at the Fermi level: $k_{TF}^2 = 4\pi e^2 D(E_F)$, where D denotes density of states: the number of states per unit energy per unit volume (see footnote on previous page). Thus, the Thomas-Fermi parameter, k_{TF}, has been identified, and a straightforward differential equation has emerged:

$$\left(\nabla^2 - k_{TF}^2\right)\delta\phi(r) = 4\pi Q\delta(0)\,. \tag{3.25}$$

This is as good a time as any to comment on a shortcoming of the model. Other than respecting the fermion nature of electrons (by using Fermi-Dirac statistics) the quantum nature of the electron has more-or-less been ignored. It has been assumed that free particle states apply from the lowest energies up to the Fermi level. For example, we have taken E_i to be $\hbar^2 k_i^2/2m - e\phi(\vec{r})$. However, this requires that $\hbar\vec{k}_i$ is identified as a momentum, which is only valid if $\phi(\vec{r})$ is a sufficiently slowly varying function of $\vec{r}$. If $\phi(\vec{r})$ varies too rapidly, the use of $\hbar\vec{k}_i$ comes into question. This quandary is what drew Landau to the quasiparticle picture of the charged Fermi liquid. This caveat notwithstanding, solution of eqn (3.25) follows standard, albeit nontrivial, mathematics,[14] yielding

[14] Because of the delta function on the right hand side of eqn (3.25), the solution is given by $4\pi Q$ times the Green's function for $\nabla^2 - k_{TF}^2$. If you look up the Green's function for the Helmholtz equation, you will find that it simply switches the minus in the above expression to a plus. A physically motivated derivation is given by Alexander Miles, which can be downloaded. All you have to do is switch his k to an imaginary one. Alternatively, you can plug the result given by eqn (3.26) into eqn (3.25) to verify consistency. If you do this, be careful to deal properly with the delta function because the function in eqn (3.26) is singular at $r = 0$. Away from $r = 0$ the delta function vanishes, but as $r = 0$ is approached the delta function accounts for the source that gives rise to the singular behavior. Note Dirac's comment: "In quantum theory whenever an improper function appears, it will be something that is to be used ultimately in an integrand."

$$\delta\phi(r) = \frac{Q}{r}\exp\left(-k_{TF}\, r\right). \tag{3.26}$$

The inverse of the parameter k_{TF} defined in eqn (3.26) is the Thomas-Fermi screening length, r_{TF}. The expression for k_{TF} in eqn (3.26) is now rearranged to yield a simple form. The algebra is carried below out step-by-step:

$$r_{TF} = {k_{TF}}^{-1} = \sqrt{\frac{1}{6\pi e^2 z {E_f}^{1/2}}} = \sqrt{\frac{1}{6\pi e^2}\frac{1}{z^{2/3}}\frac{1}{(z{E_f}^{3/2})^{1/3}}}\,. \tag{3.27}$$

Using $n_e = z{E_f}^{3/2}$ and $z = (2m/\hbar^2)^{3/2}/3\pi^2$ this becomes

$$r_{TF} = \sqrt{\frac{1}{6\pi e^2}\frac{\hbar^2}{2m}\left(3\pi^2\right)^{2/3}\frac{1}{{n_e}^{1/3}}} = {n_e}^{-1/6}\sqrt{\frac{\hbar^2}{4me^2}}\sqrt{\frac{(3\pi^2)^{2/3}}{3\pi}}$$

$$= \left(3\pi^2 n_e\right)^{-1/6}\sqrt{\frac{\pi\hbar^2}{4me^2}}\,. \tag{3.28}$$

Finally, using the Bohr radius: $a_0 = \hbar^2/me^2$, yields the compact expression

$$r_{TF} \approx \frac{1}{2}\left(\frac{{a_0}^3}{n_e}\right)^{1/6}. \tag{3.29}$$

The Thomas-Fermi model is appropriate to the regime of low-temperature and high-density, where it is necessary to treat explicitly the fermion nature of electrons using the Fermi-Dirac distribution function. The Thomas-Fermi screening length for Cu is 0.55 Å, which is quite small. On this basis, one might conclude that for distances from Q of, say, ~ 2 Å the field is screened sufficiently well that the electrons can be treated as free. This might be acceptable for some purposes, but not in the present context. As mentioned earlier, a problem with the Thomas-Fermi model is that it uses free-particle kinetic energy ($\hbar^2k^2/2m$) in the presence of screened Coulomb potentials. The use of $\hbar^2k^2/2m$ carries with it an implicit assumption that the de Broglie wavelength is smaller than the inter-particle spacing. This cannot be satisfied at the electron densities of metals.

To elaborate further, it is well known in quantum mechanics (for example, WKB approximation) that free-particle character can only be invoked if the potential energy is slowly varying. For example, it is not possible to define a de Broglie wavelength if the potential varies on a length scale comparable to, or shorter than, the de Broglie wavelength. The case of copper underscores this incompatibility. Note that a screening length of 0.55 Å is about one fourth the mean electron-electron separation (2.3 Å), and about eight times smaller than the de Broglie wavelength of copper at E_F. Figure 12 indicates

the pronounced nature of Thomas-Fermi screening. However, keep in mind that the derivation relies on an ansatz of free particle character and strong Coulomb interaction. Free particle character cannot be reconciled with correlation being present throughout the sea of electrons. It turns out that abandoning the single-electron picture in favor of a quasi-electron picture saves the day.

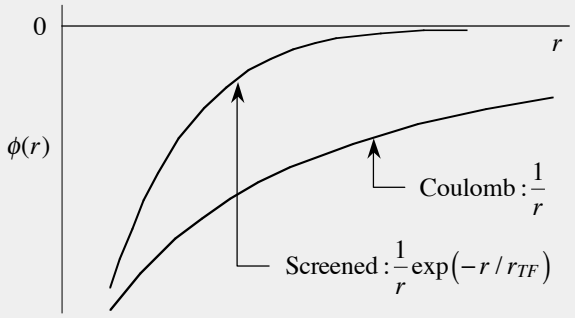


Figure 12. Screening alters markedly the long-range Coulomb potential presented by an electron. The Thomas-Fermi model yields exponential damping, with characteristic screening length r_{TF} .

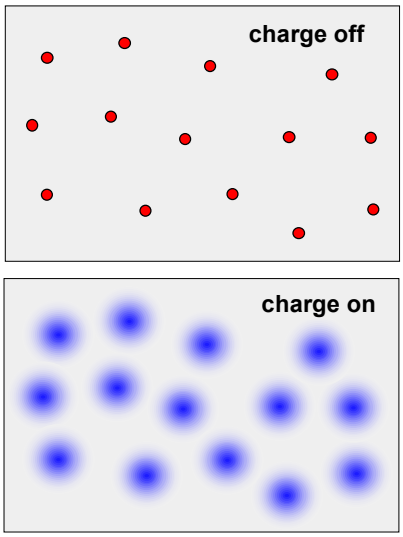
Lindhard Model

Jens Lindhard introduced a model that applies to the same regime as the Thomas-Fermi model, but is more accurate [47]. I am not going to go over it here, but if you are interested a brief description is given in Ashcroft and Mermin [1] and I will loan you this book. For the present discussion, the important result is that the long-range screened potential contains a term whose *r*-variation is

$$r^{-3}\cos\left(2k_F r\right). \tag{3.30}$$

This damped oscillation does not die out nearly as quickly as the exponential of the Thomas-Fermi model. Consequently, screening is distributed over a larger sphere and includes more electron density. Screening still occupies center stage, but the many-body aspect of the electron-electron interactions is evidently not something that can be ignored. Let us now see how this works.

Quasiparticles to the Rescue



The present subsection again begins with the ansatz of a sea of conduction band electrons. This time, however, a gedanken experiment is carried out in which electron charge is turned off at the start, with the understanding that its correct value will be recovered in due course. The electrons are still spin ½ fermions and their mass has not changed, but now they are uncharged. It is also understood that the positive charge of the immobile substrate tracks the electron charge to ensure overall charge neutrality.

As discussed earlier, at $T = 0$ K, these non-interacting electrons, being fermions, occupy all states up to the Fermi energy, E_F, and no states above it. For temperatures above 0 K (but not extremely high temperatures), the Fermi-Dirac distribution function, $f_{FD}(E)$, indicates the promotion of a modest fraction of the conduction band electrons to energies above the Fermi level (chemical potential, μ). At temperatures around 300 K, the Fermi energy, E_F, and μ are approximately equal in metals. The thermally promoted electrons have counterpart holes whose energies lie below μ, as indicated in Figs. 10 and 13. Because non-interacting electrons experience no potential, their energies are entirely kinetic, with a 3D density of states varying as $E^{1/2}$. These states are occupied according to Fermi-Dirac statistics, as indicated in Fig. 13(b).

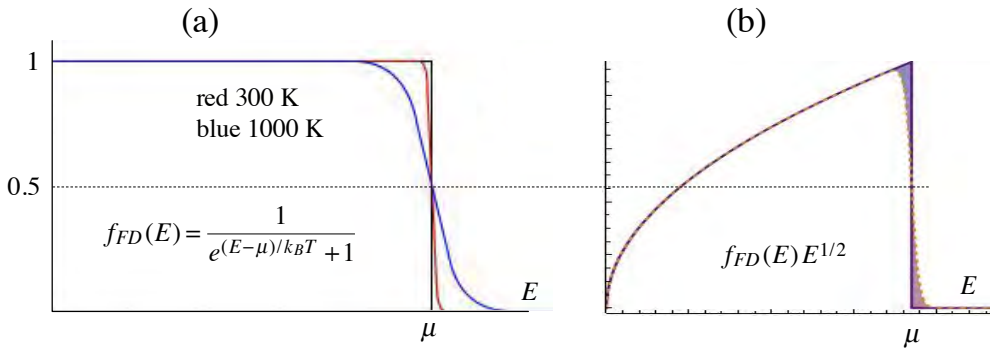


Figure 13. (a) The Fermi-Dirac function for an electron in a state with energy E is, for all practical purposes, unity at energies below μ of more than a couple of kT. (b) Combined with the density of states, which varies as $E^{1/2}$, the energy distribution differs from an $E^{1/2}$ variation that is cut off precipitously at $E = \mu$ only within a couple of kT of μ.

Electron charge is now turned on a little at a time. First, however, consider the following example to get a feel for distances and energies. The conduction band electron density of copper, 8.5×10^{22} cm^{-3}, corresponds to a nearest neighbor spacing of ~ 2.5 Å.

The unscreened Coulomb potential is large. For example, its magnitude is 6 eV at a distance of 2 Å for an unscreened electron. Moreover, even the screened potential enables interaction to extend well beyond the Thomas-Fermi characteristic screening length r_{TF}. For example, using $r_{TF} = 0.55$ Å, the exponential damping factor $\exp(-r/r_{TF})$ has a value of 0.065 at $r = 1.5$ Å. Thus the screened potential, though down considerably from its unscreened value, is nonetheless large: 0.62 eV. Let us now return to the gedanken experiment.

As electron charge is turned on, interactions do likewise, and the system develops correlation. As the magnitude of the charge is increased, the system becomes progressively more strongly interacting and correlated.[15] When the charge reaches that of a real electron, interaction is strong, inviting the provocative question: Do single-electron wave functions make any sense, or is the system so inherently many-body that they have no place?

If a single electron picture remains valid, no matter how complicated it might become, there will be no momentum relaxation, as explained in the subsection entitled *Scattering*. However, we have just seen that interaction is strong, in which case the no-momentum-relaxation argument might be inapplicable. Specifically, it is likely that it is not possible to justify holding onto a single electron picture in the face of its formidable opponents: large electron density and strong many-body interaction. If the single-electron picture is not possible, where does this leave us when it comes to applying Fermi-Dirac statistics.

Landau realized the futility of sticking with a single electron model. He abandoned it in favor of a quasielectron model, in which quasielectrons obey the exclusion principle.

Add this later.

What happens is that a single *entity* picture survives, the entity being the electron and its environment. This is the essence of the quasiparticle picture. It is valid near the Fermi energy but not at much higher energy. It turns out that the stronger the electron-electron interaction becomes, the less the quasielectron scatters from phonons.

Analogy

Analogy can provide insight through connection between understood phenomena and things less well resolved in our minds. The following scenario might prove helpful in this regard. I do not regard it as a great analogy because important points are missing. In any event, it is amusing. It involves the kinds of motions that can transpire with a large group of people, say tens of thousands.

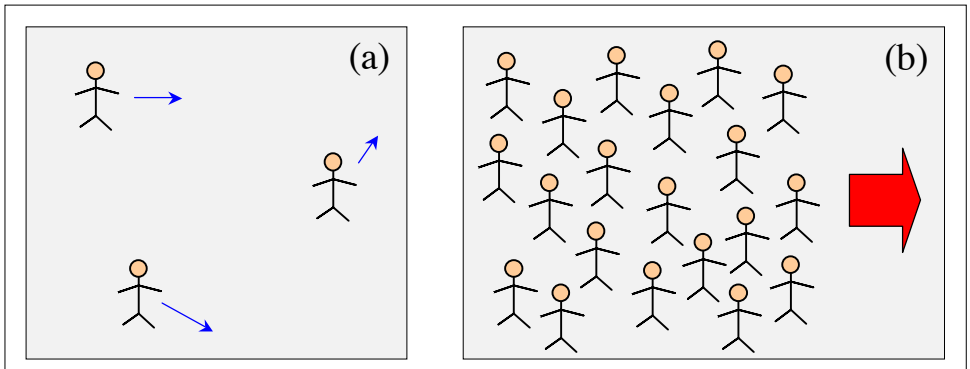
[15] Even uncharged electrons are correlated if they share space because of their fermion nature (anti-symmetrization of wave functions). The correlation of interest here is not due to this, but to the dynamics of particle interactions brought about through Coulomb forces.

When the density of the group is small, say one person every 100 square meters, each person can travel along a randomly chosen straight-line path for a significant distance before bumping into another individual. Furthermore, individuals can travel at quite different speeds. Their motions are correlated, though only over a large distance. When individuals bump into one another, they undergo a pairwise interaction and scatter elastically. An individual's velocity can change relative to that of the group as a whole, though the overall momentum of the interacting pair does not change. Of course momentum within the pair's center-of-mass system is zero.

Alternatively, what if the entire group is jammed into a small area such that the density is one person per square meter? Think of a marathon race with tens of thousands of participants. The runners are organized before the start of the race: fastest at the front of the pack, slowest at the rear. This ansatz is flawed because it assumes that the speed of each runner on the day of the race is known *a priori*. They are all good runners, so on average their speeds do not differ much from one runner to the next. However, they have good days and bad days. In any event, let us say that the above gradient is a reasonable first approximation. Besides, this is a gedanken experiment. Now consider what happens when the race begins.

Within the pack each runner interacts with nearby individuals. There is jostling, but nothing more. Motions are highly correlated on the short length scale of personal space. Each of the runners travels more slowly than if running alone or in a sparse density of runners. From afar it appears that their masses have increased. Each person moves more or less at his or her assigned speed. This situation prevails for all of the runners except those at the front of the pack. We shall leave aside what happens at the rear of the pack.

The situation is less restrained at the front of the pack, where any given runner can, in principle, overtake or fall behind. The average speed of the front of the pack changes rather little, whereas individual differences can manifest. This cannot happen within the pack, because of the high degree of correlation. I will stop here and let you ponder the extent to which this scenario is related to screening.



(a) Runners move independently, unimpeded by the presence of other runners. (b) When the runners are close to one another their movements are highly correlated, except at the front of the pack, where individuals can move ahead or fall behind.

Superconductivity

To take this idea a step further, we shall allow the positively charged lattice to distort in the direction of a given electron. As a result of this local enhancement of the positive charge density, a distant electron sees a net positive charge. It is attracted to this positive charge. Interestingly, it seems as if the two electrons are attracted to one another. In fact they are, but through a lattice distortion. A formal way of stating this is that the electrons undergo an attractive interaction mediated through the exchange of a phonon.

An analogy attributed to J. R. Schrieffer and communicated to me by Stephan Haas is that of Cooper pairs and couples on a dance floor. Imagine a large dance floor, a modern dance, and a pair of skilled partners. Throughout the dance, the partners maintain a large average distance between themselves, say, a dozen meters. They are sensitive to each other's movements, however subtle and nuanced, and they each respond in synchrony with the music that couples their motions, whirling and moving in one direction or another, as one might visualize a Cooper pair near a superconducting surface responding to magnetic and electric fields.

The dancers' respective motions are correlated, and it is clear to any observer that the two of them are dancing together as a pair. Other dancers can pass between and around our designated couple. The other dancers also are paired in couples of the same persuasion as our designated couple. In fact, the entire dance floor is filled with such pairs in which the paired individuals, on average, remain far apart, yet clearly they are dancing together. Strong correlations among the (boson) couples are due to correlations between the (fermion) individuals. The latter correlations derive from the common courtesy that an individual avoids bumping into others on the dance floor. In responding to the music, all of the couples whirl around in circles or move toward one side or another of the floor, remaining correlated throughout.

Though the ensemble of couples has strong pair-pair correlations, each couple responds to the music with its own styled nuances. They are Cooper pairs. The picture this presents is one of "quasiparticles," with individual Cooper-pair properties preserved while at the same time the pair-pair interactions serve to dress each Cooper pair.

Electron Fluid

To summarize this discussion of quasielectrons, the collective behavior of the electrons ensures that the frequency of the momentum scattering events that an electron experiences is lower than the collective oscillation frequency. The latter is referred to as the plasma frequency, ω_p, as discussed below. By definition, the electrons would be unable to act collectively were this not the case. Thus, the stage is set for treating an individual quasielectron interacting with an applied field *as if it were free*. Another way of looking at this draws upon the fact that the wavelength of the electromagnetic wave far exceeds an electron's local environment. The electron and its nearby cohort must therefore move in synchrony. It is in this sense that the plasma electrons are free. In metals, despite the fact that positive charge holds the sea of electrons in place, a given conduction quasielec-

tron hardly senses the positive charge because of screening. The electrons constitute a strongly correlated system.

Debye-Hückel Model

In the present chapter we are concerned with solids, which limits us to plasmas in metals and doped semiconductors. There are numerous other kinds of plasmas, for example in high-energy environments, both terrestrial and extraterrestrial. Some are highly energetic, like the core of the sun, whereas others are relatively benign, like a fluorescent light. Though screening is always present if the system is to be plasma, it can assume different forms depending on conditions. We have seen that the Thomas-Fermi and Lindhard models are used when the de Broglie wavelength is comparable to or larger than the spacing between electrons. Metals and heavily doped semiconductors fall into this category. In this regime the fermion nature of electrons needs to be dealt with explicitly using the Fermi-Dirac distribution function.

Before going further into metals and semiconductors, let us take a look at a model introduced by Debye and Hückel [48]. The Poisson equation in this case is

$$-\nabla^2 \delta\phi(\vec{r}) = 4\pi Q\delta(0) + 4\pi \sum_i q_i \delta n_i \ . \tag{3.31}$$

This expression appears to be very similar to eqn (3.3). Indeed, the difference lies solely in how the last term is treated.

Last terms

$-4\pi e\delta n_e(\vec{r})$ (3.3)

$4\pi \sum_i q_i \delta n_i$ (3.31)

What distinguishes the Debye-Hückel and Thomas-Fermi models is their respective treatments of the last terms in eqns (3.3) and (3.31). In the Thomas-Fermi model, this term is $-4\pi e\delta n_e(\vec{r})$, which is related to $\delta\phi(\vec{r})$ through the chemical potential and Fermi-Dirac statistics. This applies at low temperature and high concentration, where the fermion nature of electrons must be dealt with using Fermi-Dirac statistics. At the other extreme, we have the Debye-Hückel model, which applies at high temperatures and low concentrations. In this regime, kT is larger than any variation of the potential. Consequently, Boltzmann statistics are used to express changes in particle concentrations that vary with potential energy according to

$$n_i(\vec{r}) = n_i^0 \, e^{-\Phi_i(\vec{r})/kT} \ , \tag{3.32}$$

where n_i^0 is the equilibrium concentration of the i^{th} species, and the potential energy for the i^{th} species is $\Phi_i(\vec{r}) = q_i\phi(\vec{r})$. Note that the index i sums over *all* charges, not only the electrons. The last term in eqn (3.31) therefore becomes

$$4\pi \sum_i q_i \delta n_i = 4\pi \sum_i q_i \left(\delta n_i^0 - n_i^0 \, q_i \frac{\delta\phi(\vec{r})}{kT} \right) e^{-\Phi_i(\vec{r})/kT} \ . \tag{3.33}$$

The system under consideration is electrically neutral except for the test charge Q. Therefore, summing the quantities $q_i \delta n_i^0$ over all charged species except Q yields zero. Introducing this into eqn (3.31) yields the Poisson-Boltzmann equation:

$$\nabla^2 \delta\phi(\vec{r}) = 4\pi\left(-Q\delta(0) + \sum_i n_i^0 \, q_i^2 \, \frac{\delta\phi(\vec{r})}{kT} e^{-\Phi_i(\vec{r})/kT} \right). \tag{3.31}$$

The high temperature assumption means that the exponential can be replaced by unity. Thus, the last term in eqn (3.34) becomes

$$-\underbrace{\left(\sum_i \frac{4\pi q_i^2 n_i}{kT} \right)}_{k_{DH}^2} \delta\phi(\vec{r}) \; . \tag{3.35}$$

In this expression, n_i, rather than n_i^0, denotes the equilibrium concentration of the i^{th} species. Equation 3.35 identifies the Debye-Hückel wave vector, k_{DH}, and the corresponding Debye-Hückel screening length, which is simply the inverse of k_{DH}:

$$r_{DH} = \left(\frac{kT}{4\pi} \frac{1}{\sum_i q_i^2 n_i} \right)^{1/2} \; . \tag{3.36}$$

If the system is balanced in the sense that $q_+ = e$, and $n_+ = n_e$, this reduces to

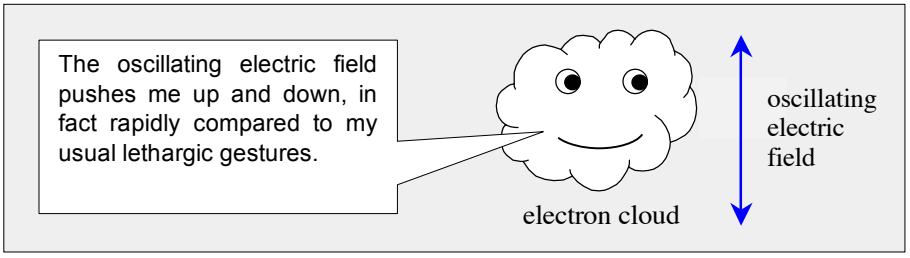
$$\boxed{r_{DH} = \sqrt{\frac{kT}{8\pi e^2 n_e}} \; .} \tag{3.37}$$

This applies to intrinsic and lightly doped semiconductors, low density gaseous plasma, ions in solution, and so on.

Metals and Doped Semiconductors

The plasma picture contrasts with that of an ionic crystal, where charges are bound securely to the lattice sites. It is appropriate to metals, where electrons move about freely in a conduction band, and to n-type semiconductors, where a relatively modest number of electrons (compared to a metal) move about freely in a conduction band. Keep in mind that the term *free* in reference to conduction band electrons is used with the understanding that a quasiparticle picture is the valid one. As mentioned above, the cases of metals and n-type semiconductors are discussed in more detail in Chapter 4.

At low frequency, the electrons perceive the driving field as quasi-static. They strive to achieve a steady state in which the momentum acquired from the electric field is balanced by momentum lost through scattering. For example, at $\omega = 0$, there exists simply the DC conductivity and static lattice polarizability of a metal or semiconductor. On the other hand, at high frequencies, say higher than the plasma oscillation frequency, an electron faces a quandary: is it bound or free? Its dynamics are treated according to the picture of a free electron. At the same time the field forces the electrons to oscillate back and forth, so they display some characteristics of being bound, though not to specific sites like the charges in ionic crystals. The distinction between bound and free is blurred.



The simple model introduced in the present section is solved without difficulty, and many of its features carry over to metals and doped semiconductors. Not surprisingly, some features carry over to ionic crystals. After all, though the charges in an ionic crystal are not mobile, its lattice vibrations are large-scale charge oscillations. We shall see that free versus bound character results in qualitatively different kinds of dielectric response to a propagating wave: $0 < \varepsilon_{re}(\omega) < 1$ for propagation through of a plasma at $\omega > \omega_p$, and $\varepsilon_{re}(\omega) > 1$ for propagation through a system of bound charges.

The positive ion substrate can be, but is not necessarily, an ordered lattice. On a scale that is large relative to atomic dimensions the positive and negative particles are each seen as distributed uniformly. It is assumed that they share the same volume and that the amounts of positive and negative charge are equal. Thus, the system is electrically neutral, and consequently there is no net macroscopic electric field caused by the charges when the system is in equilibrium. If part of the electron distribution is displaced from its equilibrium position (imagine an electron density gradient introduced at $t = 0$), oscillation in the form of wave packet propagation will occur. This will persist until relaxation deactivates the plasma oscillation.

There are a number of ways to visualize plasma oscillation. Referring to Fig. 14, imagine two slabs of equal size, one with positive charge density (black) and the other with negative charge density (blue). The slabs are displaced in the horizontal direction, so you can see that they have the same size. Each extends into the page. Their charge densities have equal magnitude. At equilibrium the black and blue slabs occupy the same volume (they overlap). Their respective mass densities: $\rho_m{}^{(+)}$ and $\rho_m{}^{(-)}$, differ markedly, with $\rho_m{}^{(+)} >> \rho_m{}^{(-)}$. Of course, we are only concerned with cases in which the negative charges are electrons. Any displacement from the equilibrium configuration will result in oscillation that persists until a loss mechanism damps the oscillation. For example, an initial ($t = 0^+$) displacement of the blue slab is shown in Fig. 14(a). The red arrow indicates the direction of its motion. The black (heavy) slab is stationary. Half a period later the blue slab has turned around and is now moving upward.

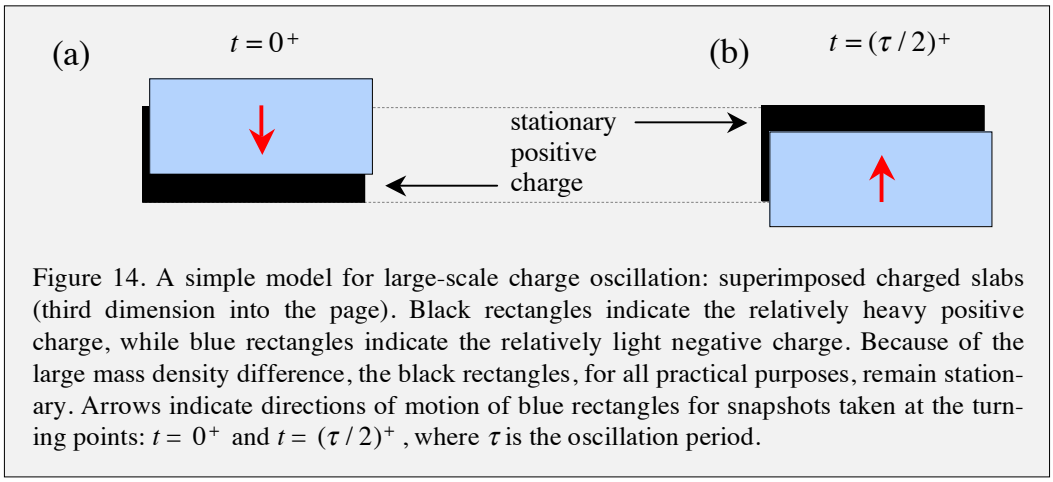


Figure 14. A simple model for large-scale charge oscillation: superimposed charged slabs (third dimension into the page). Black rectangles indicate the relatively heavy positive charge, while blue rectangles indicate the relatively light negative charge. Because of the large mass density difference, the black rectangles, for all practical purposes, remain stationary. Arrows indicate directions of motion of blue rectangles for snapshots taken at the turning points: $t = 0^+$ and $t = (\tau / 2)^+$, where τ is the oscillation period.

Chapter 4.

Plasmas in Metals and Semiconductors

Birds
Robert Wittig

Your theory is crazy, but it's not crazy enough to be true.
Niels Bohr

Contents

4.1. Wave Propagation

Bird of Prey
Robert Wittig

We shall commence our study of plasmas in metals and doped semiconductors with an introductory level examination of electromagnetic waves in a region of uniform plasma density. This is referred to frequently as bulk (or volume) plasma. A simple model of low-energy gaseous plasma is a logical starting point, as it is straightforward and gets a number of important points across. Following this warm-up exercise, we shall move on to plasmas in metals and doped semiconductors, taking additional details into account, including interactions at boundaries. These latter plasmas are the main topic of the present chapter.

To begin, the fate of a single electron subjected to a propagating electromagnetic wave is calculated, and this result is then incorporated into the macroscopic dielectric function of gaseous plasma. Here and hereafter the word electron is used with the understanding that the electrons under consideration are not truly free.[1] Electrons behave collectively in plasma; otherwise it would not be plasma. Think of the electric field that drives electron motion as a self-consistent field comprising the applied field and the fields due to all of the other charge carriers. The collective nature of electron motion is particularly clear in the solid state. We have seen that conduction band electrons in a metal are in close proximity to one another. They acquire a degree of translational freedom by virtue of screening, but they are not truly free. They are part of a highly correlated electron gas. As long as such caveats are kept in mind, using the word electron does no harm.

It is assumed that the electron responds to the force caused by only the electric field part of the electromagnetic wave. Interaction of the electron with the magnetic part of the wave, whose energy density is equal to that of the electric part, is neglected because the magnetic force is much smaller than the electric force. In this regard, note the term $\vec{v}/c$ that appears in the Lorentz force equation (Gaussian units):

$$\vec{F} = -e\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}\right). \tag{4.1}$$

The ratio $\vec{v}/c$ is a reminder that classical magnetism can be interpreted as being a consequence of relativity. This ratio arises naturally in classical special relativity. Charge that moves relative to an observer is seen by the observer to do so over lengths that are

[1] Plasma plays an important role in astrophysics, where the density of ionized species can be quite low. In the interstellar medium, the average electron density is less than one electron per cm^3. In the intergalactic medium, electron density drops to one per m^3.

contracted in proportion to a factor γ that contains v/c, namely, $\gamma = [1-(v/c)^2]^{-1/2}$. Charge, but not charge density, is conserved when it is observed from different inertial frames. As a result, electrons traveling in opposite directions in parallel wires are repelled from one another. Each sees the other's charge density as being larger than that of the positive ions. The resulting charge imbalance engenders a Coulomb force that is assigned to what is referred to as a magnetic field. The text by Purcell [34] contains an excellent discussion of these points.[2] We see from eqn (1) that $\vec{E}$ and $\vec{B}$ have the same units. Thus, small values of v/c ensure that the magnetic force is small relative to the electric force.

In the plasma case, it is assumed that the positive ions remain stationary on the plasma's oscillation time scale because they are much heavier than the electrons. This assumption eliminates up front the possibility of electron-phonon coupling, but it is appropriate in an introductory level treatment. In metals and semiconductors the positive ions are arranged in lattices that are crystalline or polycrystalline.

Maxwell's equations and related expressions are summarized in the box on the next page. The equations are given in SI and Gaussian units. A number of systems of units appear in texts and in the literature. For the material under consideration here, the most common are SI (*le Système international d'unités*, essentially mks) and Gaussian (essentially cgs). Lorentz-Heaviside units are used extensively in the relativistic theory that is the subject of Part V, but they are not useful with the material under consideration here.

When I studied electrical engineering back in the 1960's the mks (now SI) system of units was preferred. In fact, "preferred" is an understatement. Truth be told, the majority of my teachers were adamant that this was the only system of units that would ever be used for electrodynamics. Other systems of units would drift into oblivion. To this day, SI units are used almost exclusively in electrical engineering. However, it seems that Gaussian units are the more popular ones for the material in Part B, so they will be used. Moreover, they offer a number of advantages: $\vec{E}$ and $\vec{B}$ have the same units; there is no strange impedance of free space; $\vec{v}/c$ appears explicitly; and so on. You are welcome to use whichever system you are comfortable with, but be consistent, as going back and forth can be tricky. My recommendation is Gaussian.

Referring to the box entitled Maxwell's Equations on the next page, the equations listed under the headings "microscopic" treat all charges and currents explicitly. In this case, there is no separation into external versus internal charges and currents. Thus, in Gaussian units, $\vec{B} = \vec{H}$ and $\vec{D} = \vec{E}$, meaning that $\mu = \varepsilon = 1$.

The form of Maxwell's equations that we will find most useful in this chapter is the one that subsumes all microscopic details into macroscopic polarization and conductivity.

---

[2] Another way of appreciating this uses the fact that electric and magnetic fields are not separately invariant with respect to relativistic (Lorentz) transformation between inertial frames. Only an electric field is present in a frame in which charge is stationary. However, electric and magnetic fields are both present when this charge is viewed from a moving frame. Gauge field theory gets around this at the outset by establishing that the potentials ϕ and $\vec{A}$ are more fundamental than the electric and magnetic fields $\vec{E}$ and $\vec{B}$. These latter fields do not transform as four-vectors, whereas ϕ and $\vec{A}$ together constitute the four-vector A^ν, which transforms covariantly under Lorentz transformation. These more advanced topics are covered in detail in Part V.

This form is given under the headings "macroscopic." We will use microscopic treatments to derive the properties of electrons that are coupled to electric fields. In turn, these properties constitute the ingredients of plasma dielectric and conductivity functions.

Maxwell's Equations

SI (mks) units		Gaussian (cgs) units	
microscopic	**macroscopic**	**microscopic**	**macroscopic**
$\nabla\times\vec{B}=\mu_0\vec{j}+\frac{1}{c^2}\partial_t\vec{E}$	$\nabla\times\vec{H}=\vec{j}_f+\partial_t\vec{D}$	$\nabla\times\vec{B}=\frac{4\pi}{c}\vec{j}+\partial_{ct}\vec{E}$	$\nabla\times\vec{H}=\frac{4\pi}{c}\vec{j}_f+\partial_{ct}\vec{D}$
$\nabla\times\vec{E}=-\partial_t\vec{B}$	$\nabla\times\vec{E}=-\partial_t\vec{B}$	$\nabla\times\vec{E}=-\partial_{ct}\vec{B}$	$\nabla\times\vec{E}=-\partial_{ct}\vec{B}$
$\nabla\cdot\vec{B}=0$	$\nabla\cdot\vec{B}=0$	$\nabla\cdot\vec{B}=0$	$\nabla\cdot\vec{B}=0$
$\nabla\cdot\vec{E}=\rho/\varepsilon_0$	$\nabla\cdot\vec{D}=\rho_f$	$\nabla\cdot\vec{E}=4\pi\rho$	$\nabla\cdot\vec{D}=4\pi\rho_f$
	$\vec{D}=\varepsilon_0\vec{E}+\vec{P}=\varepsilon\vec{E}$		$\vec{D}=\vec{E}+4\pi\vec{P}=\varepsilon\vec{E}$
	$\vec{B}=\mu_0(\vec{H}+\vec{M})$		$\vec{B}=\vec{H}+4\pi\vec{M}$

Subscript f on $\vec{j}$ and/or ρ means free; subscript b means bound; no subscript means total.
SI: $\rho_b=-\nabla\cdot\vec{P}$; $\vec{j}_b=\nabla\times\vec{M}+\partial_t\vec{P}$

Poynting vector: $\vec{S}$ = power per unit area

$= \frac{c}{4\pi}\vec{E}\times\vec{B}$ (Gaussian, vacuum) $= \frac{c}{8\pi}\mathrm{Re}(\vec{E}\times\vec{B}^*)$ (complex exponent)

$= \vec{E}\times\vec{H}$ (SI).

Energy density: $\frac{1}{2}\vec{E}\cdot\vec{D}+\frac{1}{2}\vec{B}\cdot\vec{H}$ (SI).

Constitutive relations: $\vec{D}=\varepsilon\vec{E}$; $\vec{B}=\mu\vec{H}$; $\vec{j}=\sigma\vec{E}$.
In general, ε, μ, and σ are second-rank tensors and functions of ω.

Linearly polarized plane waves: $\vec{E}=E_0e^{i(kz-\omega t)}\hat{x}$ and $\vec{B}=B_0e^{i(kz-\omega t)}\hat{y}$

Potentials: $\vec{B}=\nabla\times\vec{A}$ and $\vec{E}=-\nabla V-\partial_{ct}\vec{A}$ (Gaussian).
$\vec{B}=\nabla\times\vec{A}$ and $\vec{E}=-\nabla V-\partial_t\vec{A}$ (SI).

Gauge invariance: Adding the gradient of a scalar: $\vec{A}\to\vec{A}+\nabla\chi$, leaves $\vec{B}$ unchanged because $\nabla\times\nabla\chi\equiv 0$. Adding $-\partial_{ct}\chi$ (Gaussian): $V\to V-\partial_{ct}\chi$, leaves $\vec{E}$ unchanged.

Coulomb gauge: $\nabla\cdot\vec{A}=0$ (SI and Gaussian).

Lorenz gauge: $\nabla\cdot\vec{A}+(1/c)\partial_{ct}V=0$ (SI); $\nabla\cdot\vec{A}+\partial_{ct}V=0=\partial_\nu A^\nu$ (Gaussian).

Gaussian units will be used exclusively throughout this chapter, with $\vec{D}=\vec{E}+4\pi\vec{P}$ and $\vec{j}=\sigma\vec{E}$ called upon frequently. We will not deal with magnetic properties, so $\mu = 1$ will be used. The tensorial nature of ε and σ is put to the side except in Example 4.2: *Faraday Rotation*. In other words, ε and σ are assumed to be scalars. Their frequency depen-

dence implies a time delay between some stimulus and the system's response to it. In addition, it is necessary to take into account the distance from the region of stimulus to where the response is registered. These aspects are explained in the box below.

Response Functions: ε(ω) and σ(ω)

The parameter ε in $\vec{D} = \varepsilon\vec{E}$ is referred to as the dielectric function, while the parameter σ in $\vec{j}_{int} = \sigma\vec{E}$ is referred to as the electrical conductivity. The expressions $\vec{D} = \varepsilon\vec{E}$ and $\vec{j}_{int} = \sigma\vec{E}$ at first sight might convey the impression that the system responds instantaneously to an electric field. Of course, this cannot be true. The response functions ε and σ must account for where and when an action takes place versus where and when something happens as a consequence. This is expressed mathematically as a convolution [19]:

$$\vec{D}(\vec{r},t) = \int d\vec{r}\,'dt'\,\varepsilon(\vec{r}-\vec{r}\,',t-t')\vec{E}(\vec{r}\,',t') \tag{1}$$

$$\vec{j}_{int}(\vec{r},t) = \int d\vec{r}\,'dt'\,\sigma(\vec{r}-\vec{r}\,',t-t')\vec{E}(\vec{r}\,',t')\,. \tag{2}$$

In eqn (1), $\varepsilon(\vec{r}-\vec{r}\,',t-t')$ is used rather than the more general mathematical expression $\varepsilon(\vec{r},\vec{r}\,';\,t,t')$, and likewise for eqn (2). In so doing, it is assumed that only space and time differences are important. Absolute space and time are unimportant. At a given spatial location $\vec{r}$ and time *t*, it is necessary to include responses due to the fields that exist at all locations $\vec{r}\,'$ and earlier times t'. Physics respects causality. When it comes to carrying out calculations, eqns (1) and (2) can prove awkward. However, Fourier transformation comes to the rescue.

The convolution theorem states that the Fourier transform of a convolution (in the present case the above integrals) is equal to the product of the Fourier transforms of the entities being convolved [35,36]. Fourier transformation takes the system from $(\vec{r},t)$ space to $(\vec{k},\omega)$ space. It decomposes the fields into plane waves. Applying the convolution theorem to eqns (1) and (2) yields

$$\vec{D}(\vec{k},\omega) = \varepsilon(\vec{k},\omega)\;\vec{E}(\vec{k},\omega) \tag{3}$$

$$\vec{j}_{int}(\vec{k},\omega) = \sigma(\vec{k},\omega)\;\vec{E}(\vec{k},\omega)\,. \tag{4}$$

The arrows on $\vec{D}$, $\vec{j}_{int}$, and $\vec{E}$ now denote vectors in $\vec{k}$-space. We can make use of the fact that with microscopic models the wavelengths of the electromagnetic fields are much larger than the molecular or unit cell dimensions. In this case, it is safe to assume that $\vec{k}$ is effectively zero. This enables us to replace $(\vec{k},\omega)$ with (ω) in eqns (3) and (4), yielding

$$\vec{D}(\omega) = \varepsilon(\omega)\;\vec{E}(\omega) \tag{5}$$

$$\vec{j}_{int}(\omega) = \sigma(\omega)\,\vec{E}(\omega)\,. \tag{6}$$

The arrows on $\vec{D}$, $\vec{j}_{int}$, and $\vec{E}$ now denote vectors in $\vec{r}$-space. It is customary to use $\varepsilon(\omega)$ when dealing with high frequencies such as those of the plasmas under consideration in this chapter, and $\sigma(\omega)$ when dealing with much lower frequencies. On the other hand, when it comes to propagation there must be $\vec{k}$ dependence. Wave propagation, by definition, includes retardation.

Plasma Dispersion Relation

The current density $\vec{j}$ is now calculated using a classical microscopic model. It is then introduced into Maxwell's equations in a manner such that all microscopic details are subsumed into the macroscopic form of Maxwell's equations, appearing as a polarization. In turn, this polarization defines a dielectric function $\varepsilon(\omega)$. It is assumed that the wavelength of the electromagnetic field exceeds greatly the spatial extent of the electron motion calculated in the microscopic model. Thus, we take $\vec{k}$ to have negligible magnitude, in which case it is safe to assume that $e^{i\vec{k}\cdot\vec{r}}$ is equal to one. Retardation enters through Maxwell's equations. This exercise yields a dispersion relation associated with the propagation of an electromagnetic wave in bulk plasma.

We begin with the ansatz that the current density $\vec{j}$ is equal to $-n_e e\vec{v}$, where n_e is the electron density and e is the magnitude of the electron charge. It is assumed here and hereafter that $\vec{v}$ varies in time as $e^{-i\omega t}$. That is, it has the same time variation as the electric field, $\vec{E} = \vec{E}_0 e^{-i\omega t}$, though with a phase difference. The velocity $\vec{v}$ of a free electron driven by $\vec{E}$ is obtained from the force equation

$$m\dot{\vec{v}} = -i\omega m\vec{v} \tag{4.2}$$

$$= -e\vec{E} - m\vec{v}\gamma\,. \tag{4.3}$$

Note that there is no restoring force in the case of $\vec{E}$ driving a free electron, as opposed to $\vec{E}$ driving the charges in an electrically insulating crystal. We shall see that this qualitative difference results in opposite signs for their polarizations. As a consequence of this difference, important propagation effects emerge in the plasma case.

As mentioned earlier, the magnetic force on the electron is negligible in the nonrelativistic regime under consideration, so it is not included in eqn (4.3). We will return to the magnetic force in Example 4.2: *Faraday Rotation*. There, an external static magnetic field is superimposed on a region of plasma. It eliminates the isotropy of space insofar as charged species are concerned, and in so doing brings about an interesting and useful effect that is referred to as Faraday Rotation. This phenomenon has widespread commercial impact. For example, Faraday isolators are used with devices ranging from CD players to sophisticated lasers.

In the present context, neglecting the force due to the magnetic component of the Lorentz force is an accurate approximation. Momentum relaxation is taken into account

through the introduction of a rate parameter γ whose units are s^{-1}. This was also done in the case of phonons. In light of the complexity of real systems and the variety of conceivable momentum relaxation mechanisms, it is not feasible that γ provides other than a crude first approximation.


Paul Drude

This model can be traced back to Paul Drude's seminal work on dielectrics around 1900. Hendrik Lorentz did something similar around the same time, so the model is sometimes referred to as the Drude-Lorentz model. Combining eqns (4.2) and (4.3) yields

$$\vec{v} = \frac{e\vec{E}}{m(i\omega - \gamma)}. \tag{4.4}$$

The dispersion relation (ω versus $\vec{k}$) is obtained by first introducing eqn (4.4) into $\vec{j} = -n_e e\vec{v}$, then introducing this $\vec{j}$ into Maxwell's equations, and finally carrying out a few manipulations. To examine propagation, it is assumed that the spatial variation of the electromagnetic wave is given by $e^{i\vec{k}\cdot\vec{r}}$, with its time dependence given by $e^{-i\omega t}$. This is a uniform plane wave, hereafter referred to simply as a plane wave. The microscopic model is thus subsumed into the fully retarded version of Maxwell's equations.

Taking the curl of the Maxwell equation: $\nabla \times \vec{E} = -\partial_{ct}\vec{B}$, and using the vector identity $\nabla \times \nabla \times \vec{C} = \nabla(\nabla \cdot \vec{C}) - \nabla^2\vec{C}$, where $\vec{C}$ is an arbitrary vector field, transforms this Maxwell equation to

$$\nabla \times \nabla \times \vec{E} = \underbrace{\nabla(\nabla \cdot \vec{E})}_{\text{vanishes when } \vec{E} \text{ is a plane wave}} - \nabla^2\vec{E} = -\nabla \times \partial_{ct}\vec{B}. \tag{4.5}$$

Because a plasma has uniform overall charge neutrality, the external charge density in Maxwell's equations is equal to zero, and therefore $\nabla \cdot \vec{D} = 0$. This will prove useful later, but right now we shall stick with the microscopic model and use the fact that for plane wave propagation $\nabla \cdot \vec{E}$ vanishes identically. You should verify this. Thus, for plane waves, $-\nabla^2\vec{E}$ is equal to $k^2\vec{E}$, and equation (4.5) becomes

$$k^2\vec{E} = -\partial_{ct}\left(\nabla \times \vec{B}\right) = -\partial_{ct}\left(\frac{4\pi}{c}\vec{j} + \partial_{ct}\vec{E}\right). \tag{4.6}$$

In writing the last parenthetic term, it is assumed that $\mu = \varepsilon = 1$. That is, it is assumed that the plasma does not exist in a dielectric environment having $\varepsilon > 1$, as would be the case with metals and semiconductors. The assumption $\mu = 1$ means that we are dealing with non-magnetic media. Cases where the lattice dielectric strength exceeds unity will be dealt with in Section 4.2. Next, insert $\vec{j} = -n_e e\vec{v}$ into eqn (4.6) and use the expression for $\vec{v}$ given by eqn (4.4). The dispersion relation is then obtained following minor manipulation:

$$c^2k^2\vec{E} = 4\pi n_e e\,\partial_t\vec{v} - \partial_t{}^2\vec{E} = 4\pi n_e e\,\partial_t\left(\frac{e\vec{E}}{m(i\omega-\gamma)}\right) - \partial_t{}^2\vec{E}$$

$$= \left(\frac{4\pi n_e e^2}{m}\right)\frac{-i\omega}{i\omega-\gamma}\vec{E} + \omega^2\vec{E}\,, \tag{4.7}$$

The electric field appears on each side of the equation so it cancels, and the last parenthetic term is defined as the square of a plasma frequency "parameter:"

$$\omega_p{}^2 = \frac{4\pi n_e e^2}{m}\,. \tag{4.8}$$

You will often find ω_p referred to as the plasma frequency. I prefer to refer to it as a plasma frequency *parameter*, because for $\gamma \neq 0$ it differs from the frequency with which plasma undergoes oscillation at its natural frequency [*vide infra*, eqn (4.39)]. Later the word parameter will be dropped to make wording less cumbersome, but you should keep in mind that ω_p is a parameter. In any event, eqn (4.7) becomes

$$c^2k^2 = \left(\omega^2 + \omega_p{}^2\frac{-i\omega}{i\omega-\gamma}\right) = \omega^2\left(1 - \frac{\omega_p{}^2}{\omega^2}\frac{\omega}{\omega+i\gamma}\right), \tag{4.9}$$

and the dispersion relation is

$$\boxed{c^2k^2 = \omega^2\left(1 - \frac{\omega_p{}^2}{\omega^2+\gamma^2} + i\frac{\omega_p{}^2}{\omega^2+\gamma^2}\frac{\gamma}{\omega}\right).} \tag{4.10}$$

This is the dispersion relation for a plasma that is present in an environment whose dielectric strength is that of vacuum, with momentum relaxation taken into account crudely using the rate parameter γ. In taking the positive square root of each side of eqn (4.10), we see that the imaginary term results in attenuation in the direction of propagation when the real part inside the parentheses is positive. When the real part is negative, there is no propagation and attenuation is extreme. Attenuation due to a negative real part has nothing to do with loss, as loss requires $\gamma \neq 0$. When γ is small enough to be neglected, eqn (4.10) becomes

$$c^2k^2 = \omega^2 - \omega_p{}^2\,. \tag{4.11}$$

Exercise: Show that the presence of the imaginary term in eqn (4.10) results in the real part of k being nonzero and positive when the real part inside the parentheses is negative with a large magnitude. Describe the effect and explain why it is unimportant. That is, there is no propagation.

An alternate approach that arrives at the same result uses $\vec{D} = \varepsilon(\omega)\vec{E} = \vec{E} + 4\pi\vec{P}$, where $\varepsilon(\omega)$ is the frequency dependent dielectric function obtained using the microscopic model. The vector field $\vec{D}$ is referred to as the electric displacement vector. It is a convenient way to express the relationship between electric field and polarization. Nothing of substance would change were $\vec{D}$ never introduced, because it is merely a definition. However, the use of $\vec{D}$ is widespread and we shall not fight tradition.[3] Leaving aside loss for the time being, the polarization $\vec{P}$ is $-n_e e\vec{r}$, where $\vec{r} = e\vec{E}/m\omega^2$. This latter expression is obtained from eqn (4.3) with $\gamma = 0$ by writing $m\dot{\vec{v}} = m\ddot{\vec{r}} = -\omega^2 m\vec{r} = -e\vec{E}$. Combining all of the above facts yields

$$4\pi\vec{P} = -\left(\frac{4\pi n_e e^2}{m}\frac{1}{\omega^2}\right)\vec{E}$$

$$= -\frac{\omega_p{}^2}{\omega^2}\vec{E}\,. \tag{4.12}$$

The relation $\vec{D} = \vec{E} + 4\pi\vec{P} = \varepsilon(\omega)\vec{E}$ enables $\varepsilon(\omega)$ to be identified:

$$\varepsilon(\omega) = 1 - \frac{\omega_p{}^2}{\omega^2}\,. \tag{4.13}$$

Using this with $c^2k^2 = \varepsilon(\omega)\omega^2$ recovers the $\gamma = 0$ dispersion relation given by eqn (4.11). When $\gamma \neq 0$, $\varepsilon(\omega)$ is complex, as we have seen in eqn (4.10). To summarize, the real and imaginary parts of $\varepsilon(\omega)$ are

$$\varepsilon_{re}(\omega) = 1 - \frac{\omega_p{}^2}{\omega^2 + \gamma^2} \tag{4.14}$$

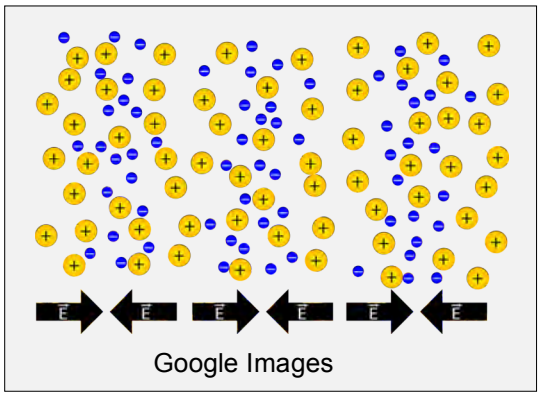
$$\varepsilon_{im}(\omega) = \frac{\omega_p{}^2}{\omega^2 + \gamma^2}\frac{\gamma}{\omega}\,. \tag{4.15}$$

As mentioned above, when $\omega^2 + \gamma^2 < \omega_p{}^2$ radiation cannot propagate in the plasma, and this has nothing to do with loss. For example, with $\gamma = 0$ (which eliminates any possibility of loss), $\varepsilon_{re}(\omega)$ is negative when $\omega < \omega_p$. In this case, k is purely imaginary, and therefore a wave incident normally at an abrupt plasma boundary is attenuated

[3] The relationship between $\vec{D}$ and $\vec{E}$, even for scalar ε, is not as straightforward as the equations might suggest. The vector field $\vec{E}$ can be visualized as a collection of vectors (say, infinitesimally close to one another) that exists throughout space. Integration of $\vec{E}\cdot d\vec{r}$ between two points returns a potential difference that is independent of path, as long as $\partial_t\vec{B}$ is negligible. On the other hand, integration of $\vec{D}$ is never carried out between two points. It is integrated as a flux through a surface. This distinction is stated succinctly in the language of differential geometry: $\vec{D}$ is dual to $\vec{E}$, and $\vec{E}$ and $\vec{D}$ integrate as 1-forms and 2-forms, respectively. Such language sounds opaque and scary, but differential geometry is really not so bad.

strongly when it tries to enter the plasma. The electric and magnetic fields are out of phase, and therefore the time average of the Poynting vector for energy flow (which is proportional to $\vec{E}\times\vec{B}$ for real fields and $\frac{1}{2}\,\mathrm{Re}(\vec{E}\times\vec{B}^{*})$ for the complex exponential fields used here) vanishes.[4]

Propagation takes place in a medium whose index of refraction is less than unity when $\omega > \omega_p$ (for $\gamma = 0$). The wave passes through the plasma without transferring energy to it, because of wave vector mismatch. From the depiction of a plasma wave on the right, it is clear that its electric field is longitudinal. This is a general property. The zeros of the dielectric function $\varepsilon(\omega)$ always give longitudinal fields. When $\omega > \omega_p$, the phase velocity v_p exceeds c. To see what is happening, consider the $\gamma = 0$ case and differentiate the dispersion relation given by eqn (4.11). This yields



$$\omega\, d\omega = c^2 k\, dk \quad \Rightarrow \quad v_p v_g = c^2 \,. \tag{4.16}$$

The product of the group and phase velocities is equal to the speed of light squared. This neat relationship applies to the bulk, lossless plasma, and to other cases that arise in applications of electromagnetic theory. When the phase velocity exceeds the speed of light, the group velocity is less than the speed of light. Do not worry about the phase velocity exceeding c. It happens all the time, and it is not a big deal. No particle travels faster than the speed of light. Phase and group velocities are discussed in Appendix 2.

---

[4] The sharp change of phase between the plane wave's electric and magnetic fields as ω goes from above ω_p to below ω_p is revealed with a few manipulations of Maxwell's equations. Assume that the plane wave's electric and magnetic fields are given by: $\vec{E} = E_x e^{i(kz-\omega t)}\hat{x}$ and $\vec{B} = B_y e^{i(kz-\omega t)}\hat{y}$. Putting these into Maxwell's equations yields

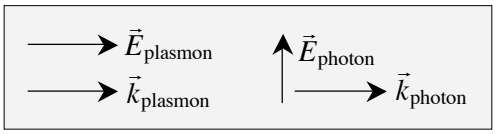
$$\nabla\times\vec{E} = ikE_x e^{i(kz-\omega t)}\hat{y} \;=\; -\partial_{ct}\vec{B} = i(\omega/c)B_y e^{i(kz-\omega t)}\hat{y}\,.$$

Taking the ratio E_x / B_y gives

$$\frac{E_x}{B_y} = \frac{\omega}{ck} = \varepsilon(\omega)^{-1/2}\,.$$

When $\varepsilon(\omega)$ becomes negative, its square root becomes imaginary. In this case the electric and magnetic fields are $\pi/2$ out of phase, and it is not possible to transport energy.

The longitudinal wave only exists at the frequency where the real part of the dielectric function vanishes ($\omega = \omega_p$ for $\gamma = 0$) and this is true for any value of k. This charge density wave, whose quantum is referred to as a plasmon, does not couple to photons. The plasmon wave vector $\vec{k}_{\text{plasmon}}$ is parallel to the electric field associated with its charge oscillation. In contrast, the wave vector of a photon is perpendicular to its electric field. Thus, wave vector conservation: $\vec{k}_{\text{photon}} = \vec{k}_{\text{plasmon}}$, cannot be satisfied, while at the same time having a component of the photon electric field in a direction that can excite the plasmon electric field.



Transverse and Longitudinal Waves

The dispersion relation for plane waves interacting with bulk plasma has been obtained. These plane waves are introduced from outside, like the radio waves that are turned around and sent back to earth by the ionosphere. Clearly, they are different than the plasma waves indicated in the picture above eqn (4.16), which are characteristic oscillations, for example, that arise from a charge disturbance. Here we shall consider the general case of any kind of electromagnetic wave whose space and time variation resides solely in its phase factor $e^{i(kz-\omega t)}$. This includes both longitudinal and transverse waves, and the waves need not be applied externally. In the next section, the difference between the dielectric function of plasma and that of a normal dielectric such as an electrically insulating crystal (the former having values of $\varepsilon(\omega)$ less than unity, the latter having values of $\varepsilon(\omega)$ greater than unity) will be revealed and discussed.

To begin, the electric field is expressed in terms of both transverse and longitudinal components, each of which propagates in the z-direction. This hybrid field is written

$$\vec{E} = E_x \hat{x} + E_z \hat{z} . \tag{4.17}$$

The transverse $E_x \hat{x}$ component is general in the sense that it could be replaced with any linear component that is perpendicular to the direction of propagation. The components E_x and E_z represent waves: $E_x^0 e^{i(kz-\omega t)}$ and $E_z^0 e^{i(kz-\omega t)}$, respectively, whose parameters E_x^0 and E_z^0 are constant in space and time. All space and time variation is contained in the phase factor $e^{i(kz-\omega t)}$. Thus, eqn (4.17) becomes

$$\vec{E} = E_x^0 e^{i(kz-\omega t)} \hat{x} + E_z^0 e^{i(kz-\omega t)} \hat{z} . \tag{4.18}$$

It turns out that Maxwell's equations yield two separate equations: one for the transverse component and one for the longitudinal component. The curl of $\vec{E}$ is nonzero for only the transverse component, whereas the transverse component has zero divergence. To see how the math works, take the curl of the Maxwell equation $\nabla \times \vec{E} = -\partial_{ct} \vec{B}$ and use the fact that $\mu = 1$. Because the curl of $E_z^0 e^{i(kz-\omega t)} \hat{z}$ is zero, this yields

$$\underbrace{\nabla\times\nabla\times\left(E_x\hat{x}\right)}_{\nabla\left(\nabla\cdot E_x\hat{x}\right)-\nabla^2\left(E_x\,\hat{x}\right)} = -\partial_{ct}\left(-\frac{4\pi n_e e}{c}\vec{v}+\partial_{ct}\vec{E}\right). \tag{4.19}$$

$$k^2E_x\hat{x}$$

On the right hand side, use $-\partial_t{}^2\vec{E} = \omega^2\vec{E}$ and insert $\vec{v}$ from eqn (4.4) to obtain

$$c^2k^2E_x\hat{x} = \left(-i\omega\frac{4\pi n_e e^2}{m(i\omega-\gamma)}+\omega^2\right)\left(E_x\hat{x}+E_z\hat{z}\right). \tag{4.20}$$

This is written as the pair of equations

$$\left(\omega^2-\omega_p{}^2\frac{i\omega}{i\omega-\gamma}\right)E_z = 0 \tag{4.21}$$

$$\left(\omega^2-\omega_p{}^2\frac{i\omega}{i\omega-\gamma}\right)E_x = c^2k^2E_x\,. \tag{4.22}$$

As expected, the transverse and longitudinal waves differ qualitatively. Referring to eqn (4.21), for negligible damping, this dispersion relation is trivial. The longitudinal field only oscillates at the frequency ω_p:

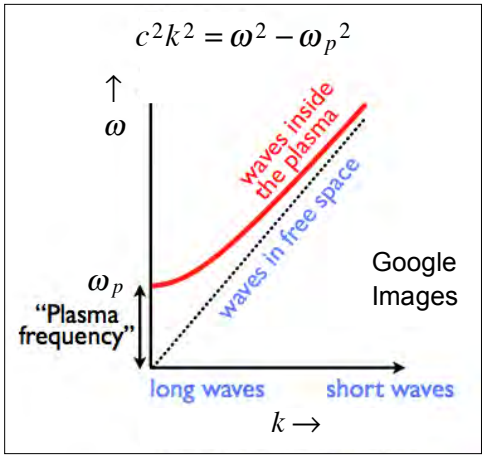
$$\boxed{\omega=\omega_p \text{ for the longitudinal field } E_z\ (\gamma=0).} \tag{4.23}$$

In other words, there is a longitudinal electric field associated with the plasma's natural oscillation.

Referring to eqn (4.22), let us now compare the dispersion relation for transverse waves and no loss ($\gamma=0$):

$$\omega^2 = \omega_p{}^2+c^2k^2\,,$$

to the dispersion relation for the toy model introduced way back in Chapter 1. It was noted there that the model was peculiar for phonons (though perfectly acceptable for quantization of the harmonically-bound discrete mass field), whereas later it would prove its worth. And here we are. The dispersion relation derived in the beginning of Chapter 1 is



$$\omega^2 = \Omega_0{}^2 + 4\Omega^2 \sin^2(ka/2) .$$

In the low-k regime, this becomes

$$\omega^2 = \Omega_0{}^2 + (\Omega a)^2 k^2 = \Omega_0{}^2 + v^2 k^2 , \tag{4.24}$$

where $v = \Omega a$ is the speed of sound. In the plasma case, ω is that of an electromagnetic wave, whereas the ω in Chapter 1 was that of a phonon. The respective speeds of c and v are those of light and sound, which differ by a factor of order 10^5. These important differences notwithstanding, the resemblance between the two models is uncanny.

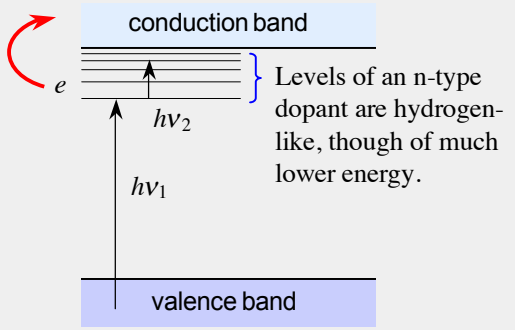
$\omega^2 = \Omega_0{}^2 + v^2 k^2$	Chapter 1, elastic limit
$\omega^2 = \omega_p{}^2 + c^2 k^2$	Bulk lossless plasma

The model presented in Chapter 1 couples each mass point to a single, well-defined reference frame. The plasma model does likewise. The sea of positive charge can be taken as stationary, and the sea of electrons then oscillates with respect to this stationary frame. A center-of-mass could be identified, and oscillation with respect to this point could be examined. In either case, the key to the correspondence between these seemingly disparate models is the use of an inertial reference frame. This will enable us in later chapters, and also in Part V, to apply the field theory model of Chapter 1 to other fields. For example, the dispersion relation: $\omega^2 = \Omega_0{}^2 + v^2 k^2$, has the same mathematical form as that of the dispersion equation that underlies the relativistic Klein-Gordon equation: $E^2 = m^2 c^4 + c^2 p^2 \implies \omega^2 = (mc^2/\hbar)^2 + c^2 k^2$. Of course, this requires exchanging constants and redefinition of the field being described. We shall get to this in Chapter 9

4.2. Optical Properties of Metals and Doped Semiconductors

Plasmas can exist in a broad range of materials: metals, semimetals, n-type and p-type semiconductors, and degenerate semiconductors. The latter are semiconductors whose extremely high dopant concentrations result in metallic properties in addition to their semiconductor properties. In semimetals and semiconductors, electrons and holes each contribute to electrical conductivity, whereas electrons alone are responsible for the electrical conductivity of metals.

Figure 1. In an n-type semiconductor dopant levels lie below the conduction band. These energy levels are hydrogen-like but of significantly lower energy because of dielectric screening and effective mass. Dopant electrons enter the conduction band with high efficiency because of the conduction band's high density of states. That is, entropy favors the conduction band.

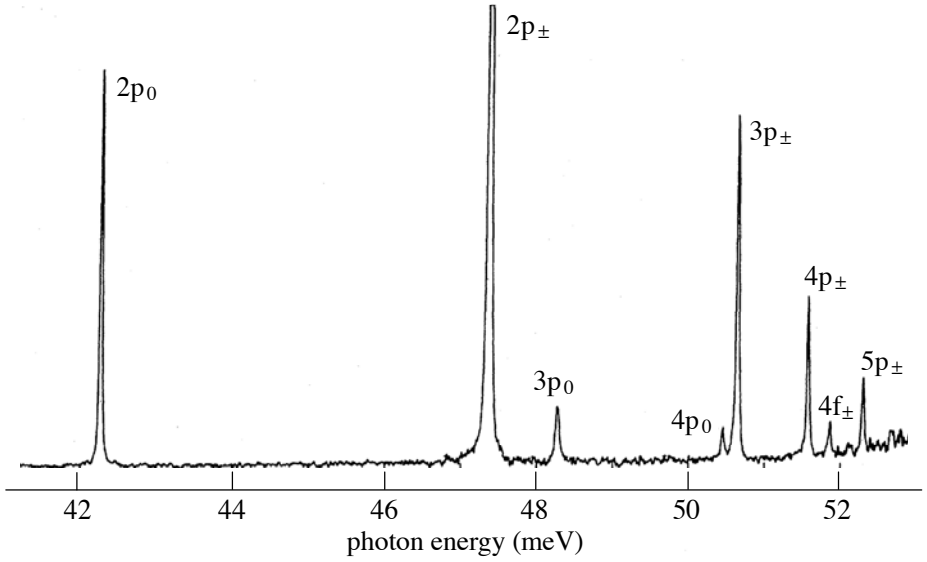


Referring to Fig. 1, in n-type doped semiconductors, electrons enter the conduction band from dopant (donor) levels that lie below the conduction band. Even if the lowest dopant level lies several kT below the bottom of the conduction band, the dopant electrons nonetheless reside primarily in the conduction band because of the latter's high density of states. In other words entropy favors the conduction band. The n-type dopant electrons account for nearly all of the dopant-inspired current. In n-type semiconductors electrons are the *majority carriers*, whereas holes are the *minority carriers*. With sufficient dopant concentration, the electrons and holes behave, to a large extent, as electrons in a metal, though with much lower concentration and with their own idiosyncratic effective masses. Electrons that enter the conduction band from the n-type dopant are referred to as *free carriers*.[5] Electrons and holes that originate in undoped semiconductor material are referred to as intrinsic. To keep matters as simple as possible, it is assumed hereafter that

[5] In p-type semiconductors, holes enter the valence band from dopant acceptor levels that lie above the valence band. They are also free carriers. Hole transport is responsible for most of the current, so holes are the majority carriers and electrons are the minority carriers. Despite use of terms like "hole transport," electrons are the carriers in this case as well. Hole transport is simply electrons that move, in the presence of an electric field, to vacant positive ion sites, exposing the positive ion sites they leave behind. In turn, the positive ion sites left behind accept other electrons, and so on, resulting in current. Mathematically, differences between electron and hole transport appear mainly in their effective masses. These are not real masses, but parameters obtained through fits to the band structure.

intrinsic electrons and holes play a negligible role, and only n-type semiconductors shall be considered.

Spectroscopic transitions that involve dopant levels are indicated in Fig. 1 with vertical arrows. Transitions from the valence band to the dopant levels ($h\nu_1$) are generally not well resolved because they can originate from a range of valence band levels. At very low temperatures (a few degrees K), the dopant electrons reside mainly in the lowest dopant level. This enables transitions to the higher hydrogen-like levels to be probed, as indicated with the shorter arrow ($h\nu_2$). An example of this is shown below [52].



This spectrum was recorded using n-type silicon (phosphorous dopant concentration of 1.2×10^{14} cm^{-3}) at 4.2 K [52]. Different subscripts are for different effective masses. Peaks labeled $2p_{\pm}$, $3p_{\pm}$, $4p_{\pm}$, and $5p_{\pm}$, form a Rydberg series. Peaks labeled with subscript zero form an even better Rydberg series.

As mentioned above, in the present section we shall leave aside semimetals, degenerate semiconductors, and p-type semiconductors, and deal just with metals and n-type semiconductors. For n-type semiconductors, the assumption that intrinsic electrons and holes play no role means that the plasma electron density is equal to the dopant density.

The positively charged lattice in a metal or doped semiconductor usually consists of crystalline regions of sufficiently large dimension to justify the assumption of a perfect crystal as a first approximation. The positively charged lattice has a dielectric property of its own that shall be referred to as ε_{lat}, with the subscript "lat" standing for lattice. It is understood that ε_{lat} is, in general, a function of frequency, so ε_{lat} could just as well be written $\varepsilon_{\text{lat}}(\omega)$.

We shall not deal explicitly with the simultaneous participation of lattice vibrations and plasma effects. We shall also avoid, for the time being, interband transitions. Thus, it is assumed that $\varepsilon_{\text{lat}}(\omega)$ is independent of frequency, and the more compact term ε_{lat} is used. Later we will discuss the role of interband transitions, which influence ε_{lat}. The frequency dependence of $\varepsilon_{\text{lat}}(\omega)$ will be considered then.

The parameter ε_{lat} is included in the model discussed below, whereas in Section 4.1, $\varepsilon = 1$ was assumed. The use of $\varepsilon = 1$ is fine for a low-density plasma such as the one in the ionosphere, but it is inappropriate for the media considered here. An aspect that was not mentioned in Section 4.1 involves the electrical conductivity, $\sigma(\omega)$, that enters Maxwell's equations via the constitutive relationship: $\vec{j}(\vec{k},\omega) = \sigma(\omega)\vec{E}(\vec{k},\omega)$. We used $\vec{j} = -n_e e\vec{v}$ to calculate $\varepsilon(\omega)$. However, $\sigma(\omega)$ is useful in some contexts, and this is discussed in the present section.

Before proceeding, you might find it useful to return to Section 4.1 and review the material from the start of that section up to eqn (4.4), which gives the electron velocity $\vec{v}$ in terms of parameters and an electric field $\vec{E} = \vec{E}_0 e^{-i\omega t}$ that drives electron motion. Equation (4.4) is reproduced below as eqn (4.25). This is where we shall begin.

$$\vec{v} = \frac{e\vec{E}}{m_e(i\omega - \gamma)} \tag{4.25}$$

Straightforward use of $\vec{j} = -n_e e\vec{v}$ yields the electrical conductivity, $\sigma(\omega)$. Specifically, $-n_e e\vec{v}$ is equated to $\sigma(\omega)\vec{E}$:

$$-n_e e\vec{v} = -\frac{n_e e^2}{m_e(i\omega - \gamma)}\vec{E} = \sigma(\omega)\vec{E}\,. \tag{4.26}$$

Notation can be simplified by using $\tau = \gamma^{-1}$. The conductivity $\sigma(\omega)$ is thus given by

$$\sigma(\omega) = \frac{n_e e^2 \tau}{m_e(1 - i\omega\tau)}\,. \tag{4.27}$$

It was mentioned earlier that, in general, linear-response terms such as $\varepsilon(\omega)$ and $\chi(\omega)$ enter the macroscopic Maxwell equations as second-rank tensors, $\vec{\vec{\varepsilon}}(\omega)$ and $\vec{\vec{\chi}}(\omega)$. Simply put, an electric field displaces electron density to different degrees in different directions. In the present treatment, scalar response functions are used, which avoids the math that accompanies tensors. In other words, the assumption of scalar response functions is consistent with the level of treatment at hand. It turns out that this assumption is accurate for nearly all of the cases we will consider.

Georg Ohm

Likewise, the conductivity is a tensor, $\vec{\vec{\sigma}}(\omega)$. It can be assigned its tensorial nature through the effective mass, or to be more accurate, the inverse of the effective mass. This results in tensor elements $(1/m_e)_{ab}$. These elements are obtained by fitting to the band structure. The appropriate equation is

$$\left(\frac{1}{m_e}\right)_{ab} = \frac{1}{\hbar^2}\frac{\partial^2 E(\vec{k})}{\partial k_a \partial k_b}. \tag{4.28}$$

As before, to keep matters simple, it is assumed that $\vec{\vec{\sigma}}(\omega)$ is a scalar, $\sigma(\omega)$. Thus, we shall deal with cases in which $\vec{j}$ and $\vec{D}$ are each parallel to $\vec{E}$. Referring to eqn (4.27), setting ω to zero gives the $\omega = 0$ (DC) conductivity:

$$\sigma_0 = \frac{n_e e^2 \tau}{m_e}. \tag{4.29}$$

You might recognize this as Ohm's law. It enables $\sigma(\omega)$ to be expressed as

$$\sigma(\omega) = \frac{\sigma_0}{1 - i\omega\tau}. \tag{4.30}$$

As with $\varepsilon(\omega) = \varepsilon_{re}(\omega) + i\varepsilon_{im}(\omega)$, it is useful to express $\sigma(\omega)$ in terms of its real and imaginary parts:

$$\sigma_{re}(\omega) = \frac{\sigma_0}{1+\omega^2\tau^2} = \frac{n_e e^2 \tau}{m_e}\frac{1}{1+\omega^2\tau^2} \tag{4.31}$$

$$\sigma_{im}(\omega) = \frac{\sigma_0\,\omega\tau}{1+\omega^2\tau^2} = \frac{n_e e^2 \tau}{m_e}\frac{\omega\tau}{1+\omega^2\tau^2}. \tag{4.32}$$

Note that, unlike $\varepsilon(\omega)$, whose imaginary part represents dissipation, it is the real part of $\sigma(\omega)$ that represents dissipation.

As the momentum relaxation time τ goes to infinity, the DC conductivity σ_0 goes to infinity. This must be the case. In the absence of momentum relaxation there can be no electrical resistance. Of course τ is never infinity.[6] Quite the contrary: in metals and doped semiconductors τ values span the approximate range $10^{-13} - 10^{-14}$ s. Indeed, measurement of σ_0 serves to determine τ. Turning to $\sigma_{im}(\omega)$, we see that the $\tau \to \infty$ limit yields $\sigma_{im}(\omega) = n_e e^2 / m_e \omega$.

Next, the Maxwell equation: $\nabla \times \vec{B} = (4\pi/c)\vec{j} + \partial_{ct}\vec{D}$, is used to identify the expression for the dielectric function $\varepsilon(\omega)$, which is then separated into its real and imaginary parts: $\varepsilon(\omega) = \varepsilon_{re}(\omega) + i\varepsilon_{im}(\omega)$. Notice that $\vec{B}$ has replaced $\vec{H}$ in the Maxwell equation because $\mu = 1$. Using $\partial_{ct}\vec{D} = (-i\omega\varepsilon_{\text{lat}}/c)\vec{E}$, and $\vec{j} = \sigma(\omega)\vec{E}$, with $\sigma(\omega)$ given by eqn (4.27), $\nabla \times \vec{B}$ becomes

[6] Superconductors have momentum relaxation rates near zero. However, the charge transport mechanism of a superconductor differs qualitatively from that of the model considered here.

$$\begin{aligned}\nabla\times\vec{B} &= \left(\frac{4\pi}{c}\frac{n_e e^2\tau}{m_e(1-i\omega\tau)} - \frac{i\omega}{c}\varepsilon_{\text{lat}}\right)\vec{E} \\ &= -\frac{i\omega}{c}\left(\varepsilon_{\text{lat}} + i\frac{4\pi n_e e^2\tau}{m_e\omega(1-i\omega\tau)}\right)\vec{E} \\ &= -\frac{i\omega}{c}\left(\varepsilon_{\text{lat}} + i\frac{4\pi n_e e^2}{m_e}\frac{\tau}{\omega}\frac{1+i\omega\tau}{1+\omega^2\tau^2}\right)\vec{E} \\ &= -\frac{i\omega}{c}\varepsilon(\omega)\vec{E} \end{aligned} \tag{4.33}$$

where

$$\varepsilon(\omega) = \underbrace{\varepsilon_{\text{lat}} - \frac{4\pi n_e e^2}{m_e}\frac{\tau^2}{1+\omega^2\tau^2}}_{\varepsilon_{re}(\omega)} + \underbrace{i\frac{4\pi n_e e^2}{m_e}\frac{\tau}{\omega}\frac{1}{1+\omega^2\tau^2}}_{\varepsilon_{im}(\omega)}. \tag{4.34}$$

It is convenient to use the screened plasma frequency parameter given by

$$\omega_p{}^2 = \frac{4\pi n_e e^2}{\varepsilon_{\text{lat}} m_e}, \tag{4.35}$$

to rewrite the expression for $\varepsilon(\omega)$ given by eqn (4.34) as

$$\varepsilon(\omega) = \underbrace{\varepsilon_{\text{lat}}\left(1 - \frac{\omega_p{}^2\tau^2}{1+\omega^2\tau^2}\right)}_{\varepsilon_{re}(\omega)} + \underbrace{i\varepsilon_{\text{lat}}\frac{\omega_p{}^2\tau}{\omega(1+\omega^2\tau^2)}}_{\varepsilon_{im}(\omega)}. \tag{4.36}$$

where $\omega_p{}^2$ is given by eqn (4.35). A similar expression was derived in Section 4.1 [eqn (4.10)]. The only difference is that ε_{lat} was taken to be unity there.

Low and High Frequency Regimes

Let us now consider the two limiting cases of very low frequency and very high frequency, starting with the low frequency regime. From eqn (4.36) we see that, as ω approaches zero, $\varepsilon(\omega)$ becomes, for all practical purposes, imaginary because the ω in the denominator of $\varepsilon_{im}(\omega)$ enables this term to overwhelm $\varepsilon_{re}(\omega)$. The real term is nearly always negative in this regime because $\omega_p\tau$ is much larger than one, so strong attenuation is assured. Thus, the ratio of $(\varepsilon_{\text{lat}} - \varepsilon_{re}(\omega))$ to $\varepsilon_{im}(\omega)$ is equal to $\omega\tau$. This is, by definition, small in the low-frequency regime, confirming the dominance of the imagin-

ary term. The bottom line is that even a significantly nonzero $\varepsilon_{re}(\omega)$ value is no match for the ω in the denominator of $\varepsilon_{im}(\omega)$. Thus, eqn (4.36) yields

$$\varepsilon(\omega) \xrightarrow{\omega \to 0} i\varepsilon_{\text{lat}} \frac{\omega_p{}^2 \tau}{\omega}. \tag{4.37}$$

Taking the square root yields the complex index of refraction. The imaginary part of the index of refraction is huge in this case, meaning that propagation is forbidden. Thus, we have arrived at the intuitive conclusion that an electric field is not permitted to propagate in a conductor at frequencies well below ω_p. This gives rise to something called the skin effect, which is the subject of one of the exercises. As ω heads toward zero, $\varepsilon_{im}(\omega)$ increases. In fact, it peaks before dropping to a smaller value at $\omega = 0$. This is referred to as the Drude peak. It is observed in all metals.

Turning now to the high frequency regime of eqn (4.36), we see that the imaginary part of $\varepsilon(\omega)$ vanishes, in which case the dielectric function $\varepsilon(\omega)$ is given by

$$\varepsilon(\omega) = \varepsilon_{\text{lat}} \left(1 - \frac{\omega_p{}^2}{\omega^2} \right). \tag{4.38}$$

Thus, we see that the plasma frequency derived in the previous section has been recovered. However, it now includes the screening provided by the lattice dielectric function, ε_{lat}. Again, keep in mind that $\omega_p{}^2$ in eqn (4.38) is the square of the screened plasma frequency parameter: $\omega_p{}^2 = 4\pi n_e e^2 / \varepsilon_{\text{lat}} m_e$.

If the frequency is sufficiently high, eqn (4.38) becomes $\varepsilon(\omega) = \varepsilon_{\text{lat}}$. This is unrealistic in metals. The ω values required to approach this limit would initiate other processes, say at ultraviolet wavelengths (and even longer in the case of Au), and ε_{lat} would no longer be constant. Consequently, eqn (4.38) is interpreted as applying to the region not too far from the plasma frequency, and even then it is on shaky ground with metals. It does much better with semiconductors because their ω_p values are relatively small.

The plasma frequency is the frequency for which the real part of the dielectric function vanishes (Fig. 2). Referring to eqn (4.36), this means that

$$\frac{\omega_p{}^2 \tau^2}{1 + \omega^2 \tau^2} = 1 \quad \Rightarrow \quad \omega^2 = \omega_p{}^2 - \gamma^2. \tag{4.39}$$

where $\gamma = \tau^{-1}$. On this basis, the ω in eqn (4.39) is the plasma frequency, that is, the frequency with which a charge disturbance oscillates on its own accord, as distinct from the parameter ω_p.

The plasma frequency is different than the ω_p given by eqn (4.35). Usually $\omega_p{}^2$ is much larger than γ^2, and in those cases the difference is not important. For example, many metals have values of ω_p of order 10^{16}, and in those cases $\omega_p{}^2 >> \gamma^2$. On the high end of possible values, we can expect that γ might be $\sim 10^{14}$. There are many cases where $\varepsilon(\omega)$ never reaches a value of zero because ε_{lat} is large and therefore the screened plasma frequency parameter is too small. It is noteworthy that the influence of the plasma and the dielectric lattice result in different signs for their contributions to the dielectric function $\varepsilon(\omega)$.

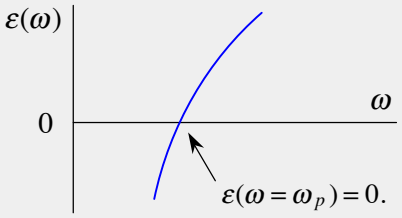


Figure 2. The plasma frequency is the value of ω for which $\varepsilon(\omega) = 0$.

Reflectivity

Monitoring the reflection of electromagnetic radiation from samples such as metals and doped semiconductors, as the frequency of the radiation is varied over a broad range, has proven to be an effective means of obtaining information about the material properties of the samples. Technological applications of considerable economic and societal impact have often followed, underscoring the importance of basic research.

Reflection
Robert Wittig

We have seen that the presence of free electrons in a system in which the free electron charge density balances the charge density of a positively charged lattice to achieve uniform, overall electrical neutrality (plasma) results in propagation being forbidden for frequencies smaller than the plasma frequency. In a plasma that experiences modest momentum relaxation ($\gamma << \omega_p$), and has minimal contributions from interband and/or intraband absorptions a reasonably sharp "edge" is present near ω_p. In other words, a plot of reflection probability, *R*, versus ω for a normally incident plane wave reveals a rapid transition from a reflection probability near unity when $\omega < \omega_p$ to modest values at frequencies above ω_p. If the assumption $\varepsilon_{\text{lat}} \approx 1$ is valid, *R* goes monotonically to zero as ω increases past ω_p.

Of course, $\varepsilon_{\text{lat}} = 1$ is almost never realistic, and when $\varepsilon_{\text{lat}} > 1$ is taken into account, *R* displays a local minimum in which it reaches a value close to zero at a frequency that is slightly higher than ω_p. We shall see that this minimum occurs when $\varepsilon_{re}(\omega) = 1$. It appears as a prominent dip in the *R* versus ω plot. The flat $R \approx 1$ region below ω_p and the dip in *R* versus ω slightly above ω_p are usually not seen in metals because of competing effects, notably interband absorption.

On the other hand, doped semiconductors offer considerable advantage over metals insofar as observing these features, because the dopant concentration (and therefore ω_p)

can be varied continuously over a broad range. The case of n-type InSb illustrates this nicely (*vide infra*, Fig. 6). The dip in a plot of R versus ω is seen to vary with dopant concentration, in good agreement with the model, as discussed below. The high infrared reflectivity of a doped semiconductor with a large bandgap is used in commercial processes to reflect heat while transmitting visible radiation.

The above results follow from elementary considerations of Maxwell's equations and electronic absorption spectroscopy. The math that underlies the reflection probability for a lossy plasma is worked out below for the case of normal incidence. This geometry gets a number of important points across with a minimum of mathematical overhead. Angles of incidence other than normal are useful for obtaining material properties, particularly in combination with the Kramers-Kronig analyses described in Example 4.1.

Complex Index of Refraction

Application of Maxwell's equations yields straightaway a wave equation and its dispersion relation for plane waves. Taking the curl of the equation: $\nabla\times\vec{E} = -\partial_{ct}\vec{B}$ yields

$$\nabla\times\nabla\times\vec{E} = -\partial_{ct}\nabla\times\vec{B} = -\partial_{ct}\left(\frac{4\pi}{c}\vec{j}_f + \varepsilon\partial_{ct}\vec{E}\right). \tag{4.40}$$

where $\mu = 1$ has been used (and therefore $\vec{H} = \vec{B}$).

The term: $\nabla\times\nabla\times\vec{E}$ $(= \nabla(\nabla\cdot\vec{E}) - \nabla^2\vec{E})$, is equal to $-\nabla^2\vec{E}$. To see why, recall that $\nabla\cdot\vec{D} = 0 = \varepsilon(\omega)\nabla\cdot\vec{E}$ when no free charge is present and ε has no $\vec{k}$ dependence. Thus, either $\varepsilon(\omega) = 0$ or $\nabla\cdot\vec{E} = 0$. The former occurs at only one frequency, and applies to fields other than plane waves. Therefore $\nabla\cdot\vec{E}$ must vanish as, of course, it does for plane waves. Likewise, it is assumed that $\vec{j}_f = 0$. Applying the above to eqn (4.40) yields the dispersion relation

$$k^2 = \varepsilon(\omega)\frac{\omega^2}{c^2}. \tag{4.41}$$

The fact that the complex index of refraction: $n = n_{re} + in_{im}$, is $\varepsilon(\omega)^{1/2}$ yields

$$k = \frac{\omega}{c}n. \tag{4.42}$$

A similar result was introduced without proof at the beginning of Chapter 3. It was pointed out there that when an electromagnetic wave passes from vacuum to a medium whose real index of refraction is n, its phase velocity and wave vector change according to $c \to c/n$ and $k_{\text{vac}} \to nk_{\text{vac}}$. Equation (4.42) says the same thing for real n, namely that $\omega/c = k_{\text{vac}}$ gives $k = nk_{\text{vac}}$. However, we see that eqn (4.42) is more general in the sense that complex n is applicable.

Next, an expression for R in terms of n is obtained by using the boundary condition for the tangential component of a plane wave that passes between different media to obtain the ratio of the electric and magnetic parts of the wave. The situation indicated in Fig. 3 illustrates why the tangential component of $\vec{E}$ is preserved across the boundary. The line integral of $d\vec{r}\cdot\vec{E}$ is taken around the dashed path that crosses the boundary between the media. Stokes' theorem applied to the surface integral of the Maxwell equation: $\nabla\times\vec{E} = -\partial_{ct}\vec{B}$ is then used to write

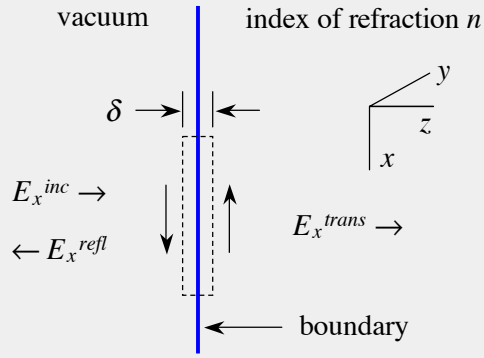


Figure 3. Integration of $\vec{E}\cdot d\vec{r}$ over the closed dashed path shows that the tangential component of $\vec{E}$ is preserved across the surface, as discussed in the text.

$$\oint d\vec{r}\cdot\vec{E} = -\partial_{ct}\iint_S d\vec{S}\cdot\vec{B}\,. \tag{4.43}$$

Referring to Fig. 3, the surface S lies in the page inside the dashed rectangle, and the right hand side of eqn (4.43) vanishes as $\delta\to 0$. Because the arrows point in opposite directions, the tangential components of the electric fields on each side of the boundary have the same value. Thus, the field's tangential component is preserved across the boundary.

The same argument applied to the Maxwell equation: $\nabla\times\vec{B} = (4\pi/c)\vec{j}_f + \partial_{ct}\vec{D}$, gives the condition for the tangential component of $\vec{B}$. It is conserved across the boundary as long as $\vec{j}_f = 0$, otherwise it is discontinuous by $(4\pi/c)\vec{j}_f{}^{surf}$.

To obtain the ratio of the electric part of the wave to the magnetic part of the wave we again refer to Fig. 3. The plane wave is incident from vacuum, with the Poynting vector

$$\vec{S} = \frac{c}{8\pi}\mathrm{Re}(\vec{E}\times\vec{B}^*) \tag{4.44}$$

giving the flow of energy. For a linearly polarized wave moving in the $+z$ direction, we can use $\vec{E} = E_x\hat{x}\,e^{i(kz-\omega t)}$ and $\vec{B} = B_y\hat{y}\,e^{i(kz-\omega t)}$. Taking the curl of $\vec{E}$ gives $ikE_x\hat{y}\,e^{i(kz-\omega t)}$, and taking $-\partial_{ct}\vec{B}$ gives $(i\omega/c)B_y\hat{y}\,e^{i(kz-\omega t)}$. Thus, the Maxwell equation: $\nabla\times\vec{E} = -\partial_{ct}\vec{B}$, gives the ratio of the electric part to the magnetic part:

$$\frac{E_x}{B_y} = \pm\frac{1}{n}\,. \tag{4.45}$$

A minus sign has appeared. For a wave moving in the $+z$ direction, the plus sign is used, whereas for a wave moving in the $-z$ direction, the minus sign is used.

Exercise: Verify the minus sign in eqn (4.45).

The n in eqn (4.45) applies to each side of the boundary in Fig. 3. In vacuum $n = 1$, while on the dielectric side the label n is used. R is obtained by applying eqn (4.45) to the boundary conditions for the tangential components. Specifically, the conditions for the tangential components of the electric and magnetic fields (assuming a surface current density of zero) that are present at the boundary are

$$E_x{}^{inc} + E_x{}^{refl} = E_x{}^{trans} \tag{4.46}$$

$$B_y{}^{inc} + B_y{}^{refl} = B_y{}^{trans} . \tag{4.47}$$

We now apply eqn (4.45) to the terms in eqn (4.47). For waves traveling from left to right, $B_y{}^{inc} = E_x{}^{inc}$ and $B_y{}^{trans} = nE_x{}^{trans}$. For the reflected wave, however, we must use $B_y{}^{refl} = -E_x{}^{refl}$, because the Poynting vector of the reflected wave points toward the left. That is, using $\vec{E}^{refl} = E_x \hat{x} e^{-i(kz+\omega t)}$ and $\vec{B}^{refl} = B_y \hat{y} e^{-i(kz+\omega t)}$ with $\nabla \times \vec{E} = -\partial_{ct} \vec{B}$ yields $kE_x = -(\omega / c)B_y$. Thus, eqn (4.47) becomes

$$E_x{}^{inc} - E_x{}^{refl} = nE_x{}^{trans} . \tag{4.48}$$

Eliminating $E_x{}^{trans}$ from eqns (4.46) and (4.48) yields the ratio

$$\frac{E_x{}^{refl}}{E_x{}^{inc}} = -\frac{n-1}{n+1} . \tag{4.49}$$

The reflection probability is given by the square of this:

$$R = \left| \frac{n-1}{n+1} \right|^2 . \tag{4.50}$$

Finally, putting the complex index of refraction, $n = n_{re} + in_{im}$, into eqn (4.50) yields

$$\boxed{R = \frac{(n_{re}-1)^2 + n_{im}{}^2}{(n_{re}+1)^2 + n_{im}{}^2} .} \tag{4.51}$$

This expression is easily evaluated once the complex dielectric function is in hand. Just take the square root: $n_{re} + in_{im} = (\varepsilon_{re} + i\varepsilon_{im})^{1/2}$. For the simplest possible (pedagogical) case: $\gamma = 0$ and $\varepsilon_{\text{lat}} = 1$, the curve corresponding to eqn (4.51) is shown in Fig. 4. This plot is intended to illustrate the influence of free carriers, not reconcile data. For example, we know that $\varepsilon_{\text{lat}} \neq 1$ with metals and semiconductors, and with metals interband absorption precludes a neat $R = 1$ region. Though γ is never equal to zero, it turns out that

momentum relaxation usually plays a modest role in a plot of R versus ω. Verification of this is left as an exercise.

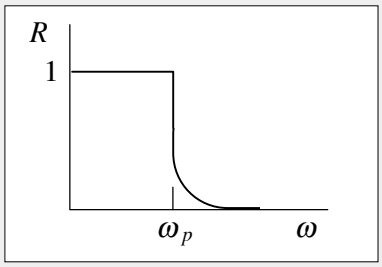
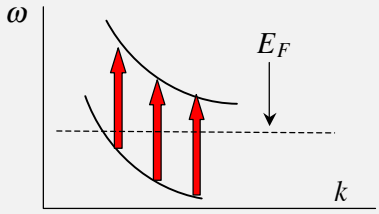


Figure 4. In the absence of momentum relaxation, interband absorption, and intraband absorption, the plasma reflection probability at normal incidence, R, is unity below the plasma frequency. It goes toward zero above ω_p. With $\varepsilon_{lat} = 1$, it goes to zero monotonically. The figure is a sketch. The correct functional form is given by eqn (4.51). Momentum relaxation does not affect the curve much, whereas $\varepsilon_{lat} > 1$, interband and intraband absorptions result in qualitative differences.

Exercise: Derive the expression for reflection probability [analogous to eqn (4.51)] that takes ε_{lat} into account. Obtain a reasonable estimate for the damping term, and provide a justification for your choice. Plot R versus ω for ε_{lat} values of 3, 6, and 12. Assume a conduction band electron density of 4×10^{22} cm^{-3}.

Metals

Metals display the phenomena of plasma reflectivity, interband absorption, momentum relaxation and (on occasion) intraband absorption in interesting ways. It almost seems that every case is a special case. Figure 5 shows reflectivity data for Al, Au, and Ag. The Al curve is rather flat, with an R value near unity at wavelengths longer than about 1 μm. The dip centered near 840 nm, and the flat R value of about 0.93 (as opposed to 1) at wavelengths to the left of this dip, are each due to interband absorptions. When a wave is reflected, even efficiently, there is penetration into the bulk. This makes it vulnerable to interband absorptions. The 840 nm dip arises because the bands involved in the transition are roughly parallel to one another in k-space for a significant wave vector range. Thus, many transitions occur at roughly the same energy (above sketch). The flat region at shorter wavelengths is due to a bunch of interband transitions that just so happen to result in a smooth, flat curve.



Lower and upper curves indicate partially filled and empty bands, respectively. Transitions from a broad range of k values have approximately the same energy.

For aluminum the wavelength λ_p that corresponds to ω_p lies below the 200 nm cutoff used in the figure. Aluminum comes closest to the idealized R versus ω plot (Fig. 4) of any of the common metals. This leaves aside the alkalis: Na, K, Rb, and Cs. These serve as excellent test cases, but they cannot be used outside laboratory environments. Because aluminum is trivalent, it is reasonable to expect its ε_{lat} value to be modest, as the atom's outermost (and therefore most polarizable) electrons are in the conduction band, leaving the less polarizable Al^{+3} lattice.

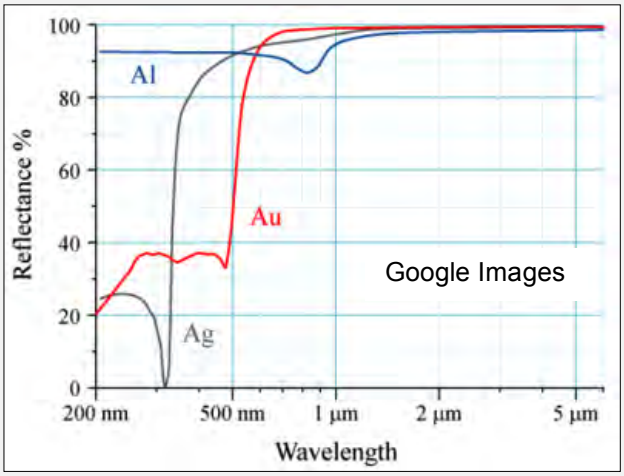


Figure 5. These ultraviolet features at first glance appear to indicate the role of a plasma edge. However, they are due to interband absorptions, their plasma edges are below 200 nm.

The Au curve drops sharply around 500 nm. However, this is unrelated to the plasma edge, as this wavelength is above the λ_p value of Au. The drop and structure are due to interband absorptions. The behavior of gold's highest energy electrons is influenced a great deal by relativistic effects in which strong orbital contraction near the core affects the outer electron density. In other words, the electronic structure is messy. The Ag curve displays a sharp drop and an enticing local dip to zero. Such behavior is expected for a case in which there is minimal contribution from interband absorption. However, it turns out that the dip is due to interband absorption. Thus, we see that for the two most important plasmonic metals, Au and Ag, the region where R is near unity is followed only until interband processes enter. Thereafter all bets are off for the simple model, at least for bulk plasma in metals.

It turns out that only a small percentage of metals support plasma character, at least to the extent that it is useful. Aside from the alkalis, this limits the field to Ag, Au, Cu, Al, and maybe a few others, with Ag and Au being the winners. The reason is that most metals, despite having large electron density, do not have a dielectric function $\varepsilon(\omega)$ that achieves the value of zero. This is due to the lattice dielectric function ε_{lat} being too large. In turn, this can be traced to interband absorptions, which enhance resonantly the lattice's dielectric response, and in so doing, endow ε_{lat} with frequency dependence. Referring to Fig. 5, notice the interesting shape of the Au curve at wavelengths below 500 nm.

Interband absorption, in general, is less important at longer wavelengths. This works to our advantage in the majority of plasmonics applications, as they involve either surfaces or nanoscale metal objects such as spheres, disks, rods, cubes, pyramids, and so on, whose plasmon resonances lie at much longer wavelengths. Chapters 5 and 7 are devoted to these topics

Semiconductors

The free carrier density of an n-type semiconductor can be varied over a broad range of values through careful control of dopant concentration. Semiconductors such as GaAs and InSb have prominent infrared-active vibrations. However, their frequencies lie in the far infrared, well beyond the band gaps of these materials. For example, InSb has a band gap of 1370 cm^{-1} at 300 K, and it is transparent down to 200 cm^{-1}. This makes InSb ideal for examining effects such as the reflectivity discussed in the previous section, of course, with the inclusion of ε_{lat}, which results in a dip to near zero reflectivity at energies slightly higher than the plasma edge. Effects due to interband absorption are eliminated by virtue of the small band gaps and ω_p values of doped semiconductors relative to metals.

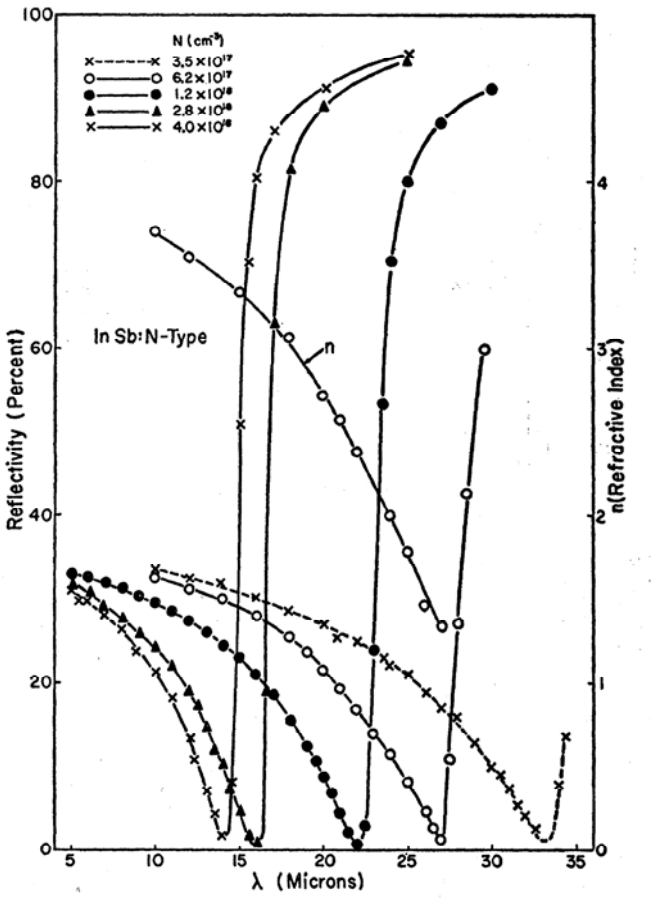


Figure 6. Reflectivity of room temperature n-type InSb for different free carrier concentrations [50].

Spitzer and Fan carried out seminal work in this area [50].[7] They examined reflectivity versus wavelength using n-type InSb samples with dopant concentration ranging from 3.5×10^{17} to 4.0×10^{18} cm^{-3}. The large infrared spectral range over which undoped InSb is transparent enabled them to isolate effects due to free carriers and ε_{lat}. The data shown in Fig. 6 illustrate clearly the predicted dip to essentially zero reflectivity just to the high-energy side of the plasma edge. This is the nicest example of this phenomenon I have yet to find in the literature.

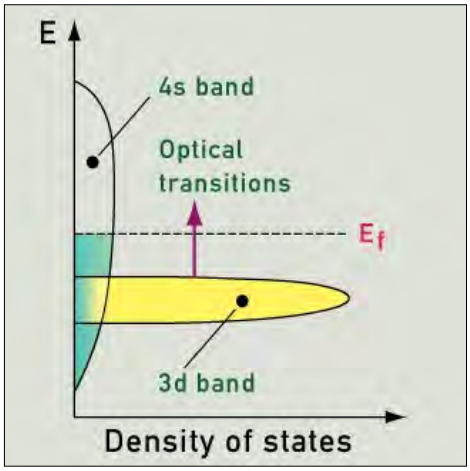
[7] Bill Spitzer established an excellent reputation in solid-state physics while at Bell Labs in the 1950's. He joined the USC faculty in 1963 in the Departments of Electrical Engineering, Materials Science, and Physics. Though he retired in the late 1990's, he remained active. I inherited a laboratory from him around 1980 on the third floor of the Seaver Science Center.

Interband Absorption

The way I think about interband absorption in crystalline media follows from analogous transitions between electronic states in isolated molecules and radicals, for example, from an electronic ground state to electronically excited states. Though isolated molecules present no counterparts to the valence and conduction band electrons of metals and semiconductors, there are no shortages of conceptual and computational challenges insofar as their photophysics and photochemistry. This is an active area of research in laser spectroscopy and dynamics and in electronic structure theory, which subsumes non-adiabatic dynamics, fragmentation, photoionization, and the like.

In crystalline media, levels merge into bands and this gets complicated. Indeed, it is far more complicated than the isolated molecule case. Still, correspondence can provide qualitative guidance. When the plasma frequency parameter ω_p is large, as it must be for metals, it is hard to imagine that interband absorption would not complicate the region of the plasma edge, and even manifest at significantly lower energies than the plasma edge. For example, the unscreened ($\varepsilon_{\text{lat}} = 1$) ω_p values for Cu, Ag, Au, and Al are 10.4, 8.65, 8.69, and 15.2 eV, respectively, and realistic ε_{lat} values do not change the fact that the characteristic frequency for plasma oscillation is quite high in these systems.

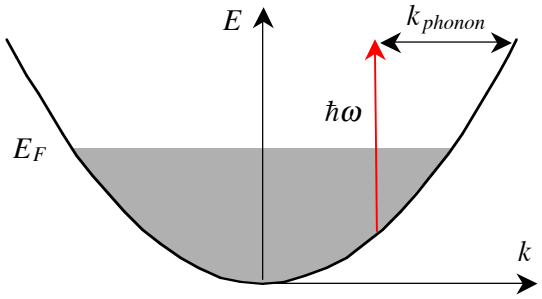
Gold is an excellent example of how interband absorption can play a central role, as seen in Fig. 5, and it is not at all the only such example. The atomic *d*-orbitals of Au, Cu, and Ag merge into bands that are confined to relatively narrow energy ranges compared to the energy ranges of their conduction bands. Interband excitations from their *d*-bands to unoccupied levels in their conduction bands account for the prominent colors of Au and Cu, and even Ag if one looks closely. The qualitative sketch on the right gives the main idea, though the real thing is infinitely more complex. These transitions enrich the spectroscopy. In addition, they are related to the real part of the dielectric function, accounting for prominent structure; see Fig. 11.17 in Ibach and Lüth [4].



From WebExhibits.org: Causes of Color

I have come across a number of articles that model interband absorption through the *ad hoc* introduction of Lorentzian functions whose location, strength, and width are treated as parameters. The fact that experimental data can be fitted nicely is amusing, though such fits mean nothing insofar as the nature of the interband absorptions are concerned, unless there is separate verification. Maybe Professors Anna Krylov and Oleg Prezhdo can shed some light on this. The discussions in Ibach and Lüth [4] and Fox [16] are definitely worth reading. I have placed two copies each of their excellent texts in our mini-library.

Intraband Absorption



Photon annihilation and concomitant phonon creation ensure wave vector conservation, though at the expense of requiring higher order interaction.

Intraband absorption, as its name implies, occurs within a band. For example, what if a photon promotes a conduction band electron to an unoccupied level in the conduction band? This is a vertical transition in *k*-space, as the electron's wave vector changes by a quite modest amount due to the small photon wave vector. Such photoexcitation cannot work by itself, however, because the dispersion relation requires a much larger change of wave vector. The only way to get around this quandary is to have another scattering event take place simultaneously. For example, if a phonon is produced, its wave vector can ensure wave vector conservation, as indicated in the above sketch.

A straightforward way to estimate the absorption is to use $k = (\omega/c)(\varepsilon_{re} + i\varepsilon_{im})^{1/2}$. The imaginary part of *k* is the absorption coefficient for exponential decay of the electric field. The absorption coefficient for exponential decay of the intensity is a factor of two larger. I will spare you most of the algebra. Using ε_{re} and ε_{im} from eqn (4.36) yields

$$k = \frac{\omega}{c}(\varepsilon_{re} + i\varepsilon_{im})^{1/2} = \frac{\omega}{c}\sqrt{\left(1 - \frac{\omega_p{}^2\tau^2}{1-\omega^2\tau^2}\right) + i\left(\frac{\omega_p{}^2\tau^2}{\omega\tau\left(1+\omega^2\tau^2\right)}\right)},$$

where $\omega_p{}^2$ is defined as $4\pi n_e e^2/m^*$. The parameter m^* is the effective electron mass for the semiconductor. The expression $4\pi n_e e^2/m^*$ looks like the one for metals, but the electron density is many orders of magnitude smaller, and the effective mass is usually larger. It is safe to assume that $\omega^2\tau^2 >> 1$. For example, using $\tau = 10^{-13}$ and $\omega = 2000$ cm^{-1} gives $\omega^2\tau^2 = 1420$. Let us now take the imaginary part of *k* and multiply by two to get the intensity absorption coefficient:

$$\alpha = \frac{\gamma}{c}\left(1 - \frac{\omega_p{}^2}{\omega^2}\right)^{-1/2}\left(\frac{\omega_p{}^2}{\omega^2}\right) = \frac{\gamma}{c\varepsilon_{re}}\frac{\omega_p{}^2}{\omega^2}$$

This shows is how momentum relaxation works its way into a region that would otherwise be transparent. I know from experience that such free carrier absorption is important in the infrared, especially as one ventures into the far infrared. Even for a crystal like intrinsic silicon it can seriously compromise transmission.

Example 4.1. Kramers-Kronig Relation

If you are unfamiliar with complex variables, do not waste your time on this example. If you *are* familiar, this example gives insight into the relationship between a material's absorption properties, notably via resonances, and its dielectric properties. The Kramers-Kronig relation is widely used in many areas of science and engineering.

We have seen that the dispersion relation for plane wave propagation in plasma can accommodate dissipation through the *ad hoc* introduction of a momentum relaxation rate parameter γ. No justification was attempted. The $\gamma m\vec{v}$ term is an easy way to introduce dissipation into a simple model. It also reconciles surprisingly well a broad range of data. Regarding terminology, you will find – in texts, articles, and discussions – that descriptors such as dissipation, loss, damping, decay, and relaxation are used interchangeably, with nuance dictated by context and whim.

Dissipation of the energy in an electromagnetic plane wave that propagates through plasma can take place via a number of routes. For example, the means of dissipation in low-density plasma such as in the ionosphere differs from that of plasma in a metal. In the former, the electrons undergo a type of acceleration that can be pictured as electrons surfing a plasma wave (Landau damping). At the other extreme, the positive charges are those of a lattice, and electron oscillatory motion can be damped through coupling to phonons, or by coupling to single-particle (electronic) excitations that in turn couple to phonons. The former is analogous to the nonadiabatic coupling of potential energy surfaces discussed in Part II.

The general case of a system whose oscillatory electron motions are damped is now considered. We are not concerned with details of the damping processes, but with how dissipation manifests in complex response functions such as the dielectric function $\varepsilon(\omega)$, polarizability $\chi(\omega)$, and conductivity $\sigma(\omega)$. We shall see that the Kramers-Kronig relation provides a mathematical basis for relating the real and imaginary parts of such response functions.

Prerequisites for application of the Kramers-Kronig relation are causality (response follows stimulus) and that the observables are bounded and have sensible asymptotic values. Here, we shall consider scalar response functions: $\vec{D} = \varepsilon(\omega)\vec{E}$, $\vec{P} = \chi(\omega)\vec{E}$ and $\vec{j} = \sigma(\omega)\vec{E}$). All physical observables satisfy causality and behave sensibly asymptotically, for example, as ω goes to infinity. Thus, the Kramers-Kronig relation has widespread applicability, of course including the material covered in the present chapter.

To begin, recall the dispersion relation that accounts for dissipation through the use of a complex dielectric function $\varepsilon(\omega)$ that is expressed in terms of a constant background lattice contribution ε_{lat} and the frequency dependent electrical conductivity $\sigma(\omega)$:

$$k^2 = \omega^2\mu\left(\varepsilon_{\text{lat}} + i\frac{4\pi}{\omega}\sigma(\omega)\right) \tag{1}$$

$$= \omega^2 \mu \varepsilon(\omega) . \tag{2}$$

The dielectric function $\varepsilon(\omega)$ is complex: $\varepsilon(\omega) = \varepsilon_{re}(\omega) + i\varepsilon_{im}(\omega)$, where $\varepsilon_{im}(\omega) = 4\pi \mathrm{Re}\,\sigma(\omega)/\omega$. As before, magnetic effects shall be put aside ($\mu = 1$), eliminating μ from eqn (2). Before proceeding, you may wish to review the material in the box entitled: *Maxwell's Equations* in Section 4.1.

Complex Response Functions

Equation (2) indicates a response to an electromagnetic field that is represented using $\varepsilon(\omega) = \varepsilon_{re}(\omega) + i\varepsilon_{im}(\omega)$, where $\vec{D}(\omega) = \varepsilon(\omega)\vec{E}(\omega)$. It can also be represented using complex susceptibility: $\chi(\omega) = \chi_{re}(\omega) + i\chi_{im}(\omega)$, where $\vec{P}(\omega) = \chi(\omega)\vec{E}(\omega)$, or complex conductivity: $\sigma(\omega) = \sigma_{re}(\omega) + i\sigma_{im}(\omega)$, where $\vec{j}(\omega) = \sigma(\omega)\vec{E}(\omega)$. In much of what follows we shall deal with complex susceptibility. No special reason underlies this choice; we could just as well deal with $\varepsilon(\omega)$ or $\sigma(\omega)$. Later we will encounter polarizability in the context of a unit cell. In this case, the polarizability is expressed as the amount of dipole moment per unit electric field in a unit cell, divided by the cell volume. As mentioned earlier, $\varepsilon(\omega)$, $\chi(\omega)$, and $\sigma(\omega)$ are in general tensors, but here they are taken as scalars.

Relations like $\vec{P}(\omega) = \chi(\omega)\vec{E}(\omega)$ are in the frequency domain except for the arrows atop the fields. Depending on your experience with such things you may or may not find expressions like this intuitive. To be on the safe side, let's see how they come about.

Convolution and Fourier Transformation

Consider the relationship between $\vec{P}(\vec{r},t)$ and $\vec{E}(\vec{r},t)$. The vector field $\vec{P}(\vec{r},t)$ acquires its time behavior from the electric field that creates the system's polarization. These fields are entirely real, as they follow from Maxwell's equations, which are entirely real. Throughout this example, sinusoidal oscillation at a single (generic) frequency is assigned to fields using the complex representation $e^{-i\omega t}$, with the understanding that real parts of functions in the $\vec{r},t$ domain are taken at the end of a calculation. This approach is fine for deriving frequency domain response functions. Many frequencies (usually a continuous distribution of ω values) are present in realistic fields. Sinusoids comprise a complete mathematical basis, so linear combinations can be used to construct any field, with superposition then giving answers.

In general, $\vec{P}(\vec{r},t)$ can be influenced by things that took place at earlier times and different locations. Thus, we write:

$$\vec{P}(\vec{r},t) = \int_{\vec{r}_0,t_0}^{\vec{r},t} d\vec{r}\,' dt'\, \chi\left(\vec{r}-\vec{r}\,', t-t'\right)\vec{E}(\vec{r}\,',t') . \tag{3}$$

The lower integration limit can be set at any convenient value that lies sufficiently in the past and far enough removed spatially that going back yet further would have little additional influence at time t and location $\vec{r}$. Equation (3) is intuitive, as there is no conceptual challenge with the fact that $\vec{P}(\vec{r},t)$ is only concerned about the past. Likewise, the propagation of an earlier disturbance manifesting in a response function is reasonable on physical grounds.

The four-dimensional Fourier transformation of eqn (3) takes the system from time to frequency and from $\vec{r}$-space space to $\vec{k}$-space. The left hand side, by definition, becomes $\vec{P}(\vec{k},\omega)$, with the arrow atop $\vec{P}$ now denoting a vector in $\vec{k}$-space. It is our good fortune that simplification occurs when the right hand side of eqn (3) is Fourier transformed. It becomes the product of the Fourier transformed quantities $\chi(\vec{k},\omega)$ and $\vec{E}(\vec{k},\omega)$. This follows from the convolution theorem, which was discussed in class. Students tend to forget about the convolution theorem, hoping it will not arise again.

Evaluation of the four-dimensional Fourier transform, though not conceptually difficult, can prove tedious. However, the transformation to $\vec{k}$-space need not be considered as long as the wavelength is much larger than a local environment such as a unit cell. In other words, the polarization and electric fields differ little from one unit cell to the next. This is the long wavelength regime that we shall work with throughout this chapter. That is, $\vec{k}$ is much smaller than a reciprocal lattice vector. Thus, spatial dependence is retained, so instead of writing $\chi(\vec{k},\omega)$ we can use $\chi(\omega)$ with the understanding that it applies locally to the electric field. Mathematically, spatial integration is carried out over a three-dimensional delta function, so it disappears. The same arguments can be used with the dielectric and conductivity functions. In all cases, frequency domain quantities are obtained and the spatial vector perspective is retained.

It is convenient to consider the response to an impulsive driving field: $\vec{E}(t) = \vec{E}_0\,\delta(t)$, which provides a convenient reference time for causality. In addition, $\vec{E}_0\,\delta(t)$ has a trivial Fourier transform: $\vec{E}(\omega) = \vec{E}_0/\sqrt{2\pi}$. As mentioned above, the time-dependent polarization $\vec{P}(t)$ is always real. It is now expressed as the Fourier transform of $\chi(\omega)\vec{E}(\omega)$, with $\vec{E}(\omega) = \vec{E}_0/\sqrt{2\pi}$:

$$\vec{P}(t) = \frac{\vec{E}_0}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\chi(\omega) \tag{4}$$

$$= \frac{\vec{E}_0}{2\pi}\int_{-\infty}^{\infty} d\omega\,(\cos\omega t - i\sin\omega t)\big(\chi_{re}(\omega)+i\chi_{im}(\omega)\big). \tag{5}$$

The fact that $\vec{P}(t)$ is always real means that only the real part of the right hand side is retained. The part of the integrand that survives is $\chi_{re}(\omega)\cos\omega t + \chi_{im}(\omega)\sin\omega t$, whereas the imaginary part: $\chi_{im}(\omega)\cos\omega t - \chi_{re}(\omega)\sin\omega t$, must integrate to zero. This reveals important symmetry properties of the real and imaginary parts of $\chi(\omega)$. Specifically, the integral of $\chi_{re}(\omega)\sin\omega t$ must vanish, and because $\sin\omega t$ is an odd function about $\omega = 0$, $\chi_{re}(\omega)$ must be an even func-

$$\chi_{re}(\omega) = \chi_{re}(-\omega)$$
$$\chi_{im}(\omega) = -\chi_{im}(-\omega)$$

tion about $\omega = 0$. Likewise, the integral of $\chi_{im}(\omega)\cos\omega t$ must vanish, and because $\cos\omega t$ is an even function about $\omega = 0$, $\chi_{im}(\omega)$ must be an odd function about $\omega = 0$.

It is important to keep in mind that these relations refer to symmetry about $\omega = 0$. The reason this is brought up is that absorption versus frequency generally (at least in simple cases) displays a symmetric Lorentzian shape centered at a resonance frequency. Thus, you might be tempted to say it is an even function. Likewise, we have seen that the real part of a dielectric function is not symmetric about a resonance frequency. It looks like an odd function. True as they are, these statements do not refer to symmetries about $\omega = 0$. To avoid confusion it might be a good idea to couch results in terms of integrals from zero to infinity, and this is done later. The bottom line here is that $\vec{P}(t)$ is well behaved.

$$\oint_C d\omega \underbrace{e^{-i\omega t}}\, \chi(\omega) = 0$$

For $t < 0$, this goes to 0 as $R \to \infty$ on the semicircle.

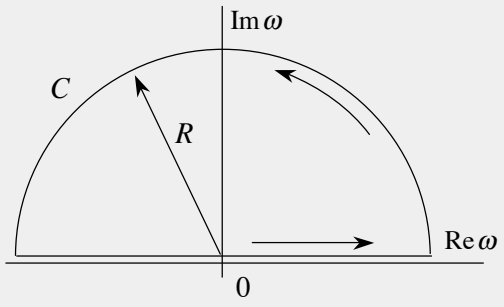


Figure 1. The factor $e^{-i\omega t}$ vanishes for $t < 0$ as $R \to \infty$ along the semicircular part of the contour C in the upper half plane. No poles are enclosed.

From eqn (4), a closed circuit in the complex ω-plane is obtained by adding a semicircular arc in the upper half plane (Fig. 1). The limit $R \to \infty$ is then taken. This works for $t < 0$ because the argument of the phase factor $e^{-i\omega t}$ has a negative real part for positive ω_{im} and $t < 0$. However, causality demands that $\vec{P}(t)$ vanishes for $t < 0$ because $\vec{E}(t)$ has not yet been turned on. This ensures that there are no poles in the upper half plane. You can verify this by checking the expressions for resonances that have been derived earlier. You will find that all of the poles lie in the lower half plane.[8] From here it is a straightforward matter to obtain the Kramers-Kronig relation. First, however, let us look further at the poles.

Complex Entities

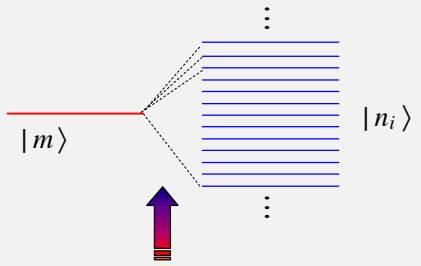
It might come as no surprise that a relationship exists between the real and imaginary parts of response functions such as $\varepsilon(\omega)$, $\chi(\omega)$, and $\sigma(\omega)$. When an incident wave drives a system of charges near one of its resonant frequencies, energy is removed efficiently from the field if the transition dipole moment is large and wave vector conservation is satisfied. The system is said to have a large absorption cross section. At the same time, the system is polarizable because electron density moves in concert with the field. In other words, electron density undergoes change (distortion) through its interaction with the field. The issue is one of establishing the relationship between the real and imaginary

[8] Be careful about which sheet of the complex plane the poles appear on.

parts: It turns out that they are related in a way that is profound and general. This is the beauty of the Kramers-Kronig relation.

The use of complex variables to describe loss provides insight and mathematical simplification. The alternative would be to treat explicitly bath degrees of freedom that accept energy from the degrees of freedom of concern. For example, were bath excitations relevant, a large Hilbert space would be required, as bath states outnumber considerably the states of interest. The idea is indicated schematically in Fig. 2. No matter how the problem is approached, treating bath states explicitly is daunting – and for little benefit.

Figure 2. A state of interest $|m\rangle$ is coupled to a dense manifold of bath states. Because the bath manifold is dense, matrix elements $\langle m|\hat{H}'|n_i\rangle$ can be taken as equal without compromising accuracy. For an *infinitely* dense manifold, $|m\rangle$ undergoes irreversible exponential decay. The effective Hamiltonian for $|m\rangle$ is then non-Hermitian, with imaginary (loss) terms on its diagonal.



Referring to Fig. 2, it is often the case that a state of interest, say $|m\rangle$, can be separated from a dense manifold of bath states, $|n_i\rangle$. The state $|m\rangle$ is deemed privileged. As mentioned above, there is usually no interest in the dynamical properties of the bath states. Fortunately, because the $|n_i\rangle$ manifold is dense, the coupling matrix elements between $|m\rangle$ and $|n_i\rangle$ (that is, $\langle m|\hat{H}'|n_i\rangle$) can each be assigned the same value. As far as $|m\rangle$ is concerned this rarely, if ever, compromises accuracy. For example, if experimental resolution is 1 cm^{-1} and the density of bath states is $10^5/\text{cm}^{-1}$, variations between individual matrix elements average out to a high degree of accuracy. Thus, an RMS matrix element will describe the system accurately.

This leads to exponential decay of the privileged state $|m\rangle$. The math falls neatly into place. It was worked out in Part III. In the limit of an infinitely dense bath manifold, the Hamiltonian becomes non-Hermitian with the decay rate Γ appearing in "imaginary energy" terms $i\hbar\Gamma/2$, where Γ is an $n \times n$ matrix. The Hilbert space of the effective Hamiltonian no longer contains the bath states, just the privileged state $|m\rangle$. Now back to Kramers-Kronig.

Relating Real and Imaginary Parts

The Kramers-Kronig relation shows how the real and imaginary parts of response functions are related to one another, subject to reasonable prerequisites. If an electric field is on for an instant: $\vec{E}(t) = \vec{E}_0\delta(t)$, response is limited to $t > 0$. When time-dependent polarization $\vec{P}(t)$ is expressed in terms of the Fourier transform of $\chi(\omega)\vec{E}(\omega)$, the fact that $\vec{E}(\omega) = \vec{E}_0/\sqrt{2\pi}$ enables us to focus on $\chi(\omega)$. Causality requires that $\chi(\omega)$ be analytic in the upper half of the ω-plane, *i.e.*, its poles lie in the lower half plane. Where $\chi(\omega)$ is analytic it satisfies conditions that relate its real and imaginary parts.[9]

The analyticity of $\chi(\omega)$ in the upper half plane is now used to derive the relationship between its real and imaginary parts. To start, recall eqn (4):

$$\vec{P}(t) = \frac{\vec{E}_0}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\,\chi(\omega) \tag{4}$$

---

[9] For a complex function $f = u + iv$, where u and v are real functions, to be analytic, it must be complex differentiable: its real and imaginary parts satisfy the Cauchy-Riemann conditions:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{i}$$

and

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{ii}$$

where $z = x + iy$ is the complex variable. This is derived in mathematics texts on complex analysis such as Churchill and Brown and in mathematical physics texts such as Wyld. In the present example, the complex variable is ω and the complex function is $\chi(\omega)$. The above equations invite an interesting geometric interpretation. Consider a two-component real vector field $\vec{g}$ (on the real space x_1 / x_2) whose elements are u and $-v$:

$$\begin{pmatrix} u \\ -v \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}. \tag{iii}$$

Equations (i) and (ii) then become:

$$\frac{\partial g_1}{\partial x_1} + \frac{\partial g_2}{\partial x_2} = 0 \tag{iv}$$

and

$$\frac{\partial g_1}{\partial x_2} - \frac{\partial g_2}{\partial x_1} = 0 \tag{v}$$

Equation (iv) indicates that $\vec{g}$ has vanishing divergence. Thus, it is not associated with a static source (charge). Equation (v) indicates that $\vec{g}$ is irrotational (zero curl in 3D). When a flux source is present, integration of $\vec{g}\cdot d\vec{r}$ over a closed circuit that encloses flux yields a nonzero value. If this flux source is confined to a strand that passes through a point in the plane, it appears as a singularity. The 2D space has been endowed with an effective curvature. Equation (v), however, does not have such singularity.

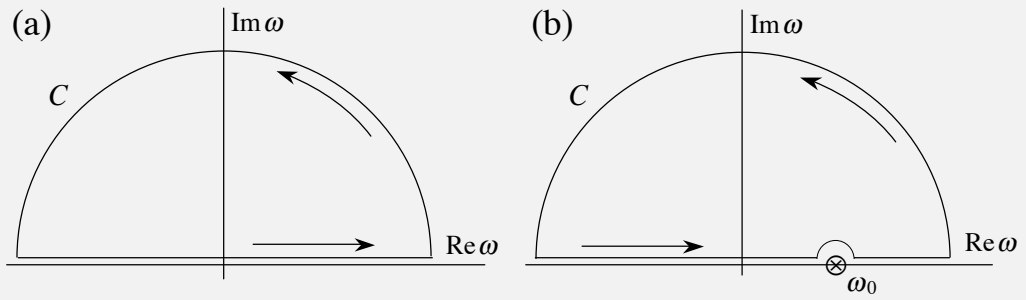


Figure 3. (a) The contour *C* encloses no singularities (simple poles) of the function $\chi(\omega)$. The function $\chi(\omega)$ is analytic within *C*. (b) The function $\chi(\omega)$ is still analytic within *C*, but now the residue at ω_0 must be taken into account.

Satisfying causality demands that $\chi(\omega)$ be analytic throughout the upper half of the complex ω-plane [Fig. 3(a)]. In addition, $\chi(\omega)$ must go to zero with increasing $|\omega|$ faster than $|\omega|^{-1}$ in order that integration on the semicircle vanishes. On physical grounds, a system cannot continue to follow an oscillating driving force as the frequency of the latter increases indefinitely, so it is natural to have the response fall off faster than $|\omega|^{-1}$.

Integration of $\chi(\omega)$ over the closed contour *C* yields zero, as no singularities are enclosed. We have seen earlier that the poles of $\chi(\omega)$ lie in the lower half plane.[10] An easy way to keep straight where the poles of $\chi(\omega)$ are located is to recall that a quantum mechanical eigenstate's time evolution is contained in the phase factor $e^{-i\omega t}$. For $t > 0$, the imaginary part of ω must be negative for the system to decay exponentially. A positive imaginary part corresponds to exponential growth in time.

To obtain the system's behavior at a specific value of (real) ω, a simple pole is placed on the real axis (included in the integrand) and the contour is taken around it, as indicated in Fig. 3(b). The contour encloses a region of analyticity, so integration over *C* still gives zero. However, the fact that the residue and principal value parts add to zero yields the relationship between the real and imaginary parts of $\chi(\omega)$. This maneuver isolates the system to the specific value ω_0. Later ω_0 will be replaced with ω, but for the time being we keep separate the pole and integration variable. Integration over *C* yields

$$\mathcal{P}\int_{-\infty}^{+\infty} d\omega\left(\frac{\chi_{re}(\omega)+i\chi_{im}(\omega)}{\omega-\omega_0}\right)=i\pi\left(\chi_{re}(\omega_0)+i\chi_{im}(\omega_0)\right). \tag{5}$$

where $\mathcal{P}$ denotes that the principal part is taken. The principal part is a balancing act, in which large negative quantities cancel large positive quantities as the pole is approached. The pole itself is not included in the principal part. It is accounted for through its residue.

[10] Sometimes poles lie on the second Riemann sheet, which lies beneath the one shown. For example, a square root can result in a branch cut that is usually taken on the positive real axis. This does not change the arguments and conclusions presented above concerning the analyticity of $\chi(\omega)$ in the upper half plane.

The *i* in eqn (5) defines the inter-relationship between the real and imaginary parts of the polarizability, in essence how they communicate with one another. Equation (5) is the Kramers-Kronig relation. Equating separately the real and imaginary parts yields the form found most often in texts:

$$\chi_{re}(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} d\omega' \frac{\chi_{im}(\omega')}{\omega'-\omega} \tag{6}$$

$$\chi_{im}(\omega) = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{+\infty} d\omega' \frac{\chi_{re}(\omega')}{\omega'-\omega}. \tag{7}$$

Referring to eqn (5), the integration variable has been changed to ω', and ω_0 has been changed to ω.

It is often useful to work only with positive frequencies, as this lessens ambiguity about how to interpret even/odd functions of ω. It is not a big deal in the present example, but in other contexts negative frequencies can be vexing. To convert everything to the positive frequency domain, start with $\chi_{re}(\omega)$ given by eqn (6) and multiply the integrand by $(\omega'+\omega)/(\omega'+\omega)$:

$$\chi_{re}(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{\chi_{im}(\omega')}{\omega'-\omega}\left(\frac{\omega'+\omega}{\omega'+\omega}\right) \tag{8}$$

$$= \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \frac{\omega'\chi_{im}(\omega')}{(\omega')^2-\omega^2}. \tag{9}$$

The denominator is an even function of ω' and $\chi_{im}(\omega')$ is an odd function of ω'. Thus, in going from eqn (8) to eqn (9), only ω' is retained in the numerator, as integration with ω in the numerator vanishes. Because the integrand in eqn (9) is an even function of ω, the intervals $(-\infty, 0)$ and $(0, \infty)$ yield the same result upon integration. This enables integration to be carried out only over positive frequencies:

$$\chi_{re}(\omega) = \frac{2}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \frac{\omega'\chi_{im}(\omega')}{(\omega')^2-\omega^2}. \tag{10}$$

Analogous steps are used for the imaginary part:

$$\chi_{im}(\omega) = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \, \frac{\chi_{re}(\omega')}{\omega'-\omega}\left(\frac{\omega'+\omega}{\omega'+\omega}\right) \tag{11}$$

$$= -\frac{\omega}{\pi}\mathcal{P}\int_{-\infty}^{\infty} d\omega' \, \frac{\chi_{re}(\omega')}{(\omega')^2-\omega^2} \tag{12}$$

$$= -\frac{2\omega}{\pi}\mathcal{P}\int_{0}^{\infty} d\omega' \, \frac{\chi_{re}(\omega')}{(\omega')^2-\omega^2}\,. \tag{13}$$

Many nuances have been sidestepped in our brief foray. For example, the electrical conductivity of a loss-free plasma has a pole at $\omega = 0$. How does one handle this? Also, methods have been developed to deal with nonlinear optical phenomena. To a non-specialist like myself, it appears that a large number of useful manipulations can be carried out. Nonetheless, the present example suffices to demonstrate the fundamental relationship that exists between the real and imaginary parts of a response function.

Example 4.2. Faraday Rotation

The ansatz that a system's dielectric properties are homogeneous and isotropic relieves us of the nuances and mathematical overhead that accompany the use of tensors. This is great as long as the ansatz has a sound physical basis. In metals and n-type semiconductors it makes perfectly good sense. The conduction band electrons move about freely except, of course, for their strong electron correlation. Consequently, they experience the same environment no matter where they are or in what direction they travel. Plasma-dielectric interfaces and nanoparticles of various sizes and shapes are altogether another matter, but even here the bulk plasma is taken as being isotropic.

This picture changes when a strong magnetic field is present in the plasma. I am not referring to the magnetic field that is part of a propagating electromagnetic wave. This magnetic field is too weak to do much. Rather, I am referring to a static magnetic field that is introduced separately, as indicated in Fig. 1. This field eliminates locally the homogeneity and isotropy of space, the latter underlying conservation of angular momentum, time-reversal symmetry, and other things. In the present example, we shall see that this field results in a tensorial environment. In turn, this environment can be used to engineer and explore effects that are not simply fascinating, but are also extremely important, ranging from astrophysics (estimation of the average magnetic field in the Milky Way galaxy) to lasers and information technologies.

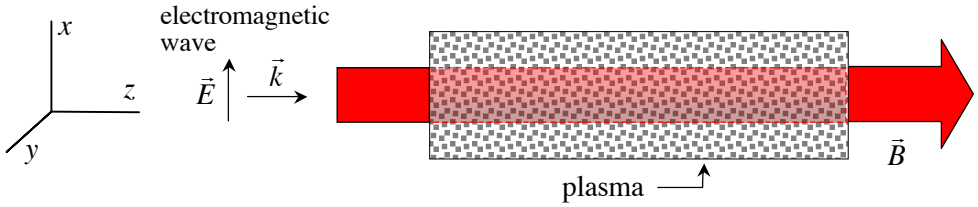


Figure 1. A static magnetic field (red) passes through the plasma. An electromagnetic wave moves in the z-direction, which is also the direction of the magnetic field lines. The magnetic part of the electromagnetic wave is unimportant.

Referring to Fig. 1, we shall start by examining an electromagnetic wave that propagates in the same direction as the (red) magnetic field lines. When an electron moves in a static magnetic field, the projection of its velocity on the direction of the field lines, $\vec{v}_{\|}$, is unaffected, because $\vec{v}_{\|} \times \vec{B} = 0$. On the other hand, the velocity's projection on a direction perpendicular to the field lines, $\vec{v}_{\perp}$, experiences a force, $\vec{F} = -e\vec{v}_{\perp} \times \vec{B}/c$. If we assume that $\vec{v}_{\perp}$ is in the y-direction and $\vec{B}$ is in the z-direction, the force is in the x-direction: $F_x = -eBv_y/c$.

The important point is that the force is perpendicular to $\vec{v}_{\perp}$, so the electron begins moving in the x-direction. It continues experiencing a force that is perpendicular to $\vec{v}_{\perp}$. This means that it circulates in a plane whose normal is aligned with the direction of the magnetic field lines. In other words, it has a helical trajectory.

This is referred to as cyclotron motion. Perhaps you have heard of particle accelerators referred to as cyclotrons, or of ion cyclotron resonance. The latter enables an ion's mass to be determined to an accuracy of order one part in 10^6. This is neat. With such high mass resolution, one easily distinguishes N_2^+ from CO^+ by their mass difference.

Exercise: Derive the cyclotron frequency for an electron, and evaluate it for a magnetic field of 1 Gauss. Provide a clear explanation of what is going on using a minimum number of equations.

The above discussion indicates that, in the presence of a static magnetic field, an electron's motion projected onto a plane perpendicular to the direction of the field lines consists of a well-defined orbit that has a well-defined frequency. Not surprisingly, this manifests in the dielectric function of the plasma-plus-magnetic-field combination, and from there the dispersion relation follows. Exactly how the ingredients fit together is not yet clear, however.

In light of the electrons' circulation, we shall represent the electric field and electron displacement in terms of circulating waves and coordinates, respectively. The idea is this. Consider a circularly polarized electromagnetic wave propagating in the *z*-direction. The tip of its electric vector makes one full revolution with every cycle of the field in the plane whose normal is aligned with the field's wave vector. This is the same kind of motion, in the sense of going around in circles, as the cyclotron circulation, though much faster. The math presented below will show how these things relate. Electron displacement follows suit, so we will also use complex coordinates: $x\,\hat{x} \pm iy\,\hat{y}$, with the usual stipulation that real parts are to be taken at the end. These complex representations facilitate the math.

Everyone likes to use linearly polarized electric fields because it is easy both computationally and experimentally. A nice thing about them is that any linearly polarized field can be expressed as a pair of circularly polarized (clockwise and counterclockwise) waves,[11] as indicated in Fig. 2. You might find it useful to think about it as follows. A photon has one unit of intrinsic (spin) angular momentum. Measurement of the spin of a single photon can only return $\pm\hbar$, never zero. A linearly polarized photon has equal probability of being found with a spin of either $+\hbar$ or $-\hbar$, as it is an equally weighted combination of these.

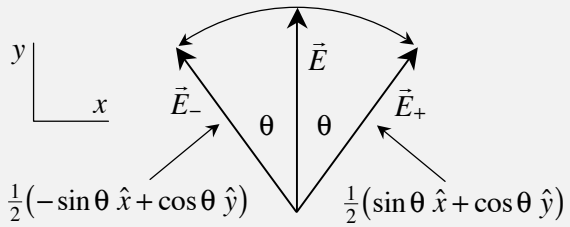


Figure 2. A linearly polarized wave (labeled $\vec{E}$) oscillates in the up and down directions with time. This is equivalent to a pair of waves (labeled $\vec{E}_+$ and $\vec{E}_-$) circulating respectively in clockwise and counterclockwise senses.

[11] These are referred to frequently as right and left circularly polarized waves. A right circularly polarized wave rotates clockwise as viewed by the receiver.

With this picture in place, we now turn to the math. The idea is to first express the Lorentz force law in terms of the complex fields and coordinates:

$$E_\pm = E_x \pm iE_y \text{ and } r_\pm = x \pm iy\,. \tag{1}$$

It is understood that $\vec{E}$ and $\vec{r}$ each vary as $e^{-i\omega t}$. To proceed, we now write the Lorentz force law:

$$m\ddot{\vec{r}} = -e\left(\vec{E} + \frac{\vec{v}}{c}\times\vec{B}\right) \Rightarrow -\omega^2\vec{r} = -\frac{e}{m}\vec{E} + i\omega\underbrace{\left(\frac{eB}{mc}\right)}_{\equiv\Omega}\underbrace{\vec{r}\times\hat{z}}_{(x\hat{x}+y\hat{y})\times\hat{z}}\,, \tag{2}$$

$x\hat{x} + y\hat{y}$; $E_x\hat{x} + E_y\hat{y}$

and make the indicated substitutions. This yields two equations:

$$-\omega^2 x = -\frac{e}{m}E_x + i\omega\Omega y \tag{3}$$

$$-\omega^2 y = -\frac{e}{m}E_y - i\omega\Omega x\,. \tag{4}$$

Now transform the Cartesian quantities to their circular counterparts:

$$x = \tfrac{1}{2}(r_+ + r_-) \quad y = -i\tfrac{1}{2}(r_+ - r_-) \quad E_x = \tfrac{1}{2}(E_+ + E_-) \quad E_y = -i\tfrac{1}{2}(E_+ - E_-)\,.$$

A modest amount of fiddling yields

$$(\omega^2 \pm \omega\Omega)r_\pm = \frac{e}{m}E_\pm\,. \tag{5}$$

From here everything falls neatly into place. The polarization: $\vec{P} = -n_e e\vec{r}$, where n_e is the electron density, is used to obtain expressions for P_+ and P_-:

$$P_\pm = -\frac{n_e e^2}{m}\frac{1}{\omega^2 \pm \omega\Omega}E_\pm\,, \tag{6}$$

and $\varepsilon(\omega) = 1 + 4\pi P/E$ is used to write

$$\varepsilon(\omega)_\pm = 1 - \frac{{\omega_p}^2}{\omega^2 \pm \omega\Omega}\,. \tag{7}$$

Thus, there are separate wave vectors, k_+ and k_-, that correspond to the expressions for the two values of the dielectric function:

$$k_\pm = \frac{\omega}{c}\sqrt{1-\frac{{\omega_p}^2}{\omega^2 \pm \omega\Omega}}\ . \tag{8}$$

The wave remains linearly polarized throughout its passage through the plasma with the applied static magnetic field. However, the wave's linear polarization rotates during this passage, as illustrated in Fig. 3. Media that display this effect are referred to as gyrotropic. It is a straightforward matter to work out the amount of rotation, which, as you might imagine, depends on the difference: $k_+ - k_-$.

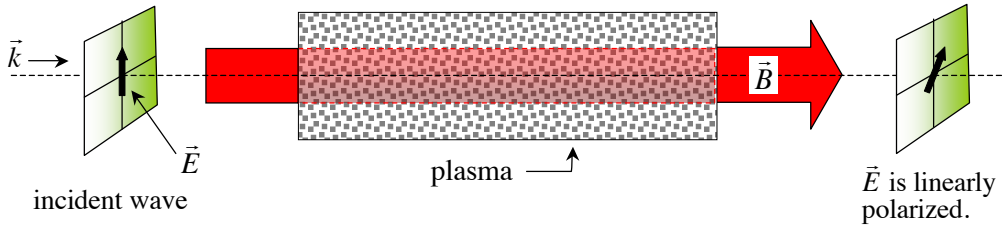


Figure 3. A polarizer (not shown) creates a linearly polarized wave (vertical $\vec{E}$) that is incident from the left. As a consequence of passing through the plasma / $\vec{B}$-field device, the wave's linear polarization has been rotated, as indicated on the right.

Exercise: Derive an expression for the amount of polarization rotation as a function of the ingredients of the model. It is safe to assume that $\omega >> \Omega$. Assume values for ω and n_e of $2\pi \times 10^{15}$ s^{-1} and 10^{12} cm^{-3}. For a magnetic field of 1 Gauss, how far must the wave travel in the plasma for it to rotate by 45°?

Faraday Isolator

Referring to Fig. 3, a linearly polarized wave is created using a polarizer located to the left of the $\vec{k}$ vector in the figure. This wave is incident on the plasma / $\vec{B}$-field device. Conditions are such that the wave's linear polarization is rotated by 45° due to its passage through the plasma, as indicated on the right side of the figure. The wave undergoes reflection somewhere to the right of the region indicated in the figure, and the reflected wave travels back (from right to left) through the device. The rotation of the wave's polarization in going from left to right is *not* reversed on its return trip. Instead, the wave rotates another 45°. Thus, when it arrives back at the polarizer its polarization is at an angle of 90° relative to that of the incident wave. The polarizer blocks the reflected wave.

Exercise: Explain why the polarization of the reflected wave rotates another 45°. Do not use equations.

The above toy model describes what is referred to as a Faraday isolator. The device prevents backscattered radiation from reaching the source of the radiation, where it can wreak havoc. For example, ultra-high-energy laser systems are used in studies of laser-induced fusion. They consist of an oscillator followed by several or more stages of amplification before the resulting intense beam is focused onto a target. A small amount of backscattered radiation, if it passes through the amplifiers, would wipe out the oscillator. Faraday isolators are used to protect the system against such effects.

On a much smaller scale, Faraday isolators are used with CD players. These isolators do not use gaseous plasmas, but ferrites instead. The ideas discussed above nonetheless still apply. Again, a small amount of reflected radiation, were it to reach the laser diode source, would wreak havoc. Historically, Faraday isolators made their debut as practical devices with microwave technology, specifically, to protect vulnerable microwave sources such as magnetrons and klystrons. The lab where I carried out my Ph.D. research was loaded with antiquated microwave equipment. Faraday isolators (using ferrites) were lying about all over the place.

Astrophysics

Faraday rotation can be used to estimate the average value of the magnetic field in our galaxy. Bursts of radiation from pulsars (rapidly rotating neutron stars) are broadband and linearly polarized. The polarization angle changes a bit from pulse to pulse, but summing results from many pulses yields a reliable value for the average polarization. This linearly polarized radiation passes through the interstellar medium's low-density plasma (less than one electron per cm^3) on its way to earth. The broadband pulsar radiation originates as an effectively instantaneous burst. Thus, its temporal width recorded on earth combined with the plasma dispersion relation, enables the path length in the plasma to be determined. There is a lot more to it, for example, how is the original polarization determined, how does one distinguish a rotation angle of θ from one of θ plus an integer times 2π, and so on? In any event, the above comments get the point across.

Careful measurements record the degree of rotation of the linear polarization, and equally careful analyses yield the average value of the projection of the magnetic field onto the line of sight. It is found that the average strength of the interstellar medium's magnetic field is a few microgauss (μG). Compare this to the earth's magnetic field, which ranges from 0.25 to 0.65 Gauss. An amusing comparison is to the neutron star that gave birth to the radiation. Neutron stars often have magnetic fields $10^{12} - 10^{15}$ Gauss.

4.3. Charge Disturbances and Their Dynamics

So far we have focused on fundamental interactions between plasma and electromagnetic waves of two types: externally applied plane waves: $\exp(i\vec{k}\cdot\vec{r}-i\omega t)$; and longitudinal waves that remind one of acoustic phonons, though with a single resonant frequency that is independent of wave vector. It was assumed that these interactions transpire within bulk plasma or at its flat surface. For example, when a wave with well-defined frequency approaches the flat surface of a metal at normal incidence, it is efficiently reflected if its frequency is less than the plasma frequency. In contrast, leaving aside interband and intraband transitions, the wave can pass through the plasma with modest attenuation due to momentum relaxation when its frequency exceeds the plasma frequency.

Wave vector conservation prevents plasmons from being created or annihilated as a plane wave passes through plasma. In solid-state plasmas, the plane wave loses energy, not through the creation of plasmons, but via electron-phonon coupling. The situation is more interesting at a plasma-dielectric interface. Under carefully chosen experimental conditions, surface plasmons can be excited efficiently. We shall see in Chapter 7 that under these conditions a photon and a surface plasmon propagate as a single unit: a *surface plasmon polariton*. It is also possible to create bulk plasmons that originate at the surface. For example, a pulsed laser can create a transient charge imbalance at the surface that propagates through the bulk, as discussed below.

Local Charge Disturbance

We shall now examine oscillations that arise from local charge disturbances, including retarded versus nonretarded regimes and how the shape of a nanoscale plasma influences the characteristic oscillation frequency of its conduction band electrons. There is an important distinction between a plasma whose characteristic dimension exceeds greatly the wavelength of the radiation versus a plasma whose characteristic dimension is significantly smaller that the wavelength of the radiation. In the former regime, retardation is important, as evidenced by repeated appearance of the phase factor $\exp(i\vec{k}\cdot\vec{r}-i\omega t)$. This is not so in the latter regime.

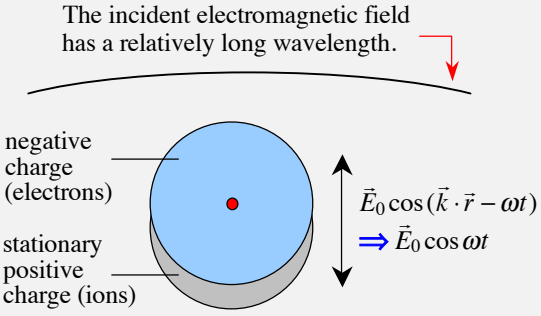


Figure 7. The field $\vec{E}_0\cos(\vec{k}\cdot\vec{r}-\omega t)$ presents a high degree of spatial uniformity over the sphere diameter, because the wavelength exceeds greatly the diameter. In this regime, retardation is unimportant, and $\vec{E}_0\cos\omega t$ can be used.

Figure 7 illustrates the small-particle regime using a spherical shape. Retardation can be neglected when $kd << 1$, where d is the sphere diameter. The small-particle regime will be discussed later in this chapter and throughout Chapter 5. The fact that retardation

does not enter in the small-particle regime can be appreciated intuitively. If the plasma is much smaller than the wavelength of the electromagnetic field, the field at any given instant in time is almost uniform over the dimension of the plasma. In a sense, the nanoscale plasma behaves like a molecule. At the same time, plasma oscillation in a nanoparticle is a coherent oscillation of the particle's electronic polarization. It has no counterpart in a large molecule, as it involves coherent oscillation of all of the electrons in the conduction band.

The highest energy electrons in a large molecule or electrically insulating crystal live in molecular orbitals. They are not free to participate in electron transport, as this would require empty states of only slightly higher energy. Generally, there are no such states. The nearest states are typically a few eV higher. In the simplest case, these electrons respond to a non-resonant applied electric field, $E_0 \cos \omega t$, as indicated in Fig. 8(a). The electrons respond quickly (almost instantaneously) to $E_0 \cos \omega t$. This results in the usual shielding of electric fields afforded by "normal dielectrics." The strength of the shielding is a function of the dielectric function of the material that makes up the sphere.

The highest energy electrons in a metal behave differently. If they move, in effect, as free particles, they obey an equation of motion: $m\ddot{\vec{r}} = -e\vec{E}$. However, $\ddot{\vec{r}}$ is then equal to $= -\omega^2 \vec{r}$, in which case $\vec{r} = (e/\omega^2)\vec{E}$. As seen in Fig. 8(b), the electron displacement relative to the applied field is opposite that shown in (a) of an electrical insulator. Near the sphere's plasmon resonance there is large field enhancement. The field inside the sphere and at its surface can be a great deal larger than the applied field.

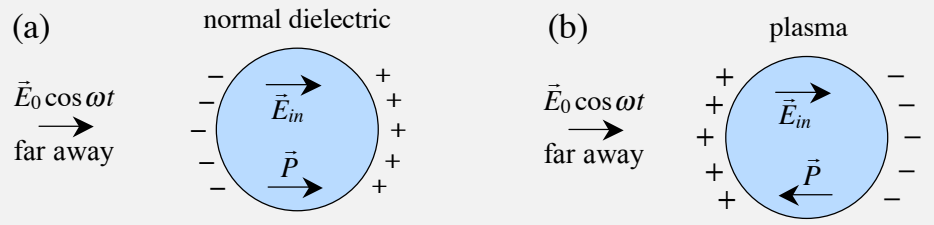


Figure 8. A field $\vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t)$ exists far from the sphere. At any instant of time the field is uniform over the sphere because $|\vec{r}_{sphere}| << \lambda$. Thus, $\vec{E}_0 \cos \omega t$ is used: the quasi-static approximation. In other words, the $\vec{k} \cdot \vec{r}$ term is not needed. (a) For a normal dielectric, the field and polarization are in the direction of the distant field. (b) The situation is different for a small metal sphere. Phases of $\vec{E}_{in}$ and $\vec{P}$ change by π as ω passes through the sphere's plasmon resonance ω_{sph}. For example, for frequencies above the sphere's plasmon resonance frequency, $\vec{E}_{in}$ and $\vec{P}$ have the phase relationship relative to $\vec{E}_0 \cos \omega t$ shown in (b), whereas for $\omega < \omega_{sph}$ their large arrows are reversed.

Charged Rectangular Slabs

Earlier it was seen that a large rectangular slab of electron density that is displaced initially from a charge-compensating slab of relatively heavy (stationary) positive charge density oscillates, in the absence of loss, at a characteristic frequency given by $\omega_p = (4\pi n_e e^2 / \varepsilon_{\text{lat}} m_e)^{1/2}$. The parameter m_e refers to electron mass, though it should be interpreted as an effective electron mass. Figure 9 gets the main idea across.

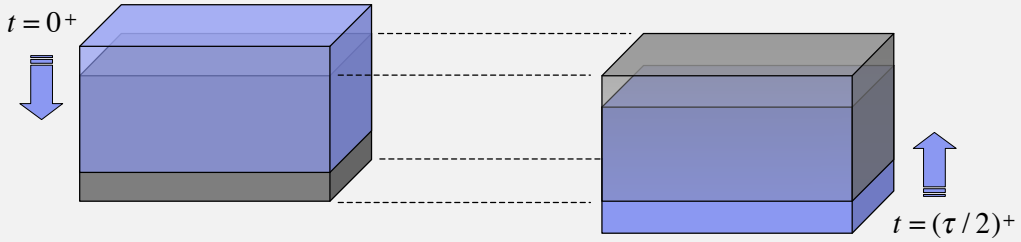
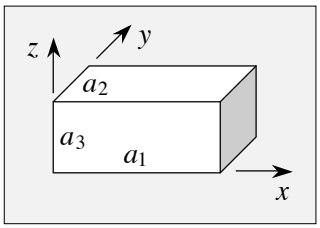


Figure 9. Gray and blue slabs indicate positive (ion) and negative (electron) charge density, respectively. The magnitudes of the positive and negative charge densities are equal. At equilibrium the slabs overlap completely, presenting an electrically neutral sample. The gray slab is taken as stationary in keeping with heavy positive ions. Dashed horizontal lines denote the fact that the gray slab is stationary, whereas the blue slab is free to move. At $t = 0$ the blue (electron) slab is displaced upward and released, as indicated on the left. It then travels downward, passing through its equilibrium location and reaching its turning point when $t = \tau / 2$, where τ is the oscillation period, as indicated on the right. This plasma oscillation persists until it is damped through a loss mechanism such as momentum relaxation.



The mathematically convenient shape of the slabs needs to be generalized as we enter the nanoscale regime. Many important features depend on plasma shape, especially in the nanoscale regime. Note that this shape dependence differs from the one encountered with resonators. For example, electron waves inside a rectangular box with conducting walls are standing waves that satisfy boundary conditions that give $k_x = n_1\pi / a_1$, $k_y = n_2\pi / a_2$, and $k_z = n_3\pi / a_3$, where the n_i are integers and a_i are box dimensions. This deals explicitly with retardation: $\vec{k} \cdot \vec{r}$ is central, not ignored. Effects due to nanoparticle shape are of a qualitatively different nature.

Charge Disturbance

Consider a spatially localized charge disturbance (also referred to as charge fluctuation or charge imbalance) that is created, for pedagogical purposes, instantaneously within a bulk plasma. Mochán treats this and several related scenarios nicely [44]. His article is a "must read" that you can easily find and download. Referring to Fig. 10, it is assumed that at the instant of its creation ($t = 0$) this charge disturbance is contained entirely within the volume labeled V_{Q0}. The total amount of charge in the disturbance introduced at $t = 0$ is labeled Q_0. The precise shape of the disturbance is not important, nor is the amount (within reason) and sign of the charge that constitutes the disturbance. As mentioned earlier, we shall assume that the positive charges are immobile. In what follows it is assumed that Q_0 is due to the injection of electrons rather than their removal. In this case, Q_0 has a negative sign.

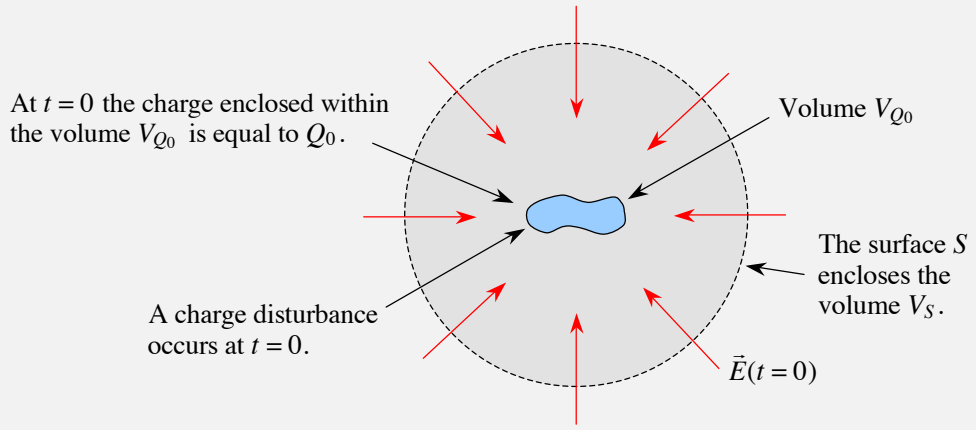


Figure 10. The completely compensated positive and negative charge densities of the plasma are disturbed through the effectively instantaneous creation of excess electron density within the volume V_{Q0}. Red arrows indicate the approximate direction of the $t = 0$ electric field. For a sufficiently small volume V_{Q0} the electric field can be taken as directed radially. Note that the picture is not to scale. Electron density initially passes out of the volume V_S, and at a later time electron density passes back into the volume V_S. The charge oscillation takes place at the plasma frequency. This figure is adapted from one in Mochán [44].

This construct enables us to examine dielectric properties and shape effects for plasma oscillation in the absence of plane waves.

The $t = 0$ charge imbalance (excess electron density) causes an electric field to be present in the plasma, and at $t = 0^+$ a net flow of electrons out of V_{Q0} commences. The charge imbalance continues to spread outward, passing through the surface S. It can be said that the system strives to find a new equilibrium. If the volume V_{Q0} is small relative to the volume enclosed by S (hereafter denoted V_S, indicated by the shaded circle in Fig. 10), the electric field $\vec{E}$ that results from the charge imbalance (red arrows in Fig. 10) is directed more-or-less radially. At $t = 0^+$ the arrows point inward because Q_0 is negative. Were Q_0 positive the arrows would point outward at $t = 0^+$.

Electrons continue to flow outward through S even after a high degree of charge neutrality within S has been achieved. In other words, the system overshoots its mark, resulting in a net positive charge within S and an excess of negative charge outside S. This eventually causes electron density outside V_S to pass back into V_S. Plasma oscillation ensues, with electron density passing back and forth through S at the plasma frequency.

Let us now see how this works mathematically. We shall leave aside the participation of thermal fluctuations. This is a good first approximation for 300 K metals, where the conduction band electrons are strongly correlated and $\hbar\omega_p >> kT$ applies. For example, the value of $\hbar\omega_p$ for silver is $\sim 73\,000\ \text{cm}^{-1}$, whereas kT at 300 K is $208\ \text{cm}^{-1}$. In the simplest case, the classical equation of motion of a plasma electron that responds to an electric field $\vec{E}$ is

$$m_e\ddot{\vec{r}} = m_e\dot{\vec{v}} = -e\vec{E}\,. \tag{4.52}$$

As discussed earlier, the electric field experienced by any given plasma electron is the *net* electric field experienced by the electron. It is a self-consistent field that is due to all the other electrons and positive charges as well as any externally applied electric field. In the present example, it is assumed that there is no externally applied electric field. Therefore, the electric field is due solely to the $t = 0$ charge disturbance and its subsequent dynamics. Again, the fact that the electron is not 100% free (dynamic screening) can be parameterized by using an effective mass, so m_e can be thought of in this way.

The usual expression for the current density: $\vec{j} = -en_e\vec{v}$, is now introduced. The expression given by eqn (4.52): $m_e\dot{\vec{v}} = -e\vec{E}$, thus becomes

$$\dot{\vec{j}} = \frac{n_e e^2}{m_e}\vec{E}\,. \tag{4.53}$$

The equation of motion for the charge disturbance is obtained by using the continuity equation that describes charge conservation:

$$-\nabla\cdot\vec{j} = \dot{\rho}\,, \tag{4.54}$$

where ρ is the excess charge density. To convert eqn (4.53) to one for $\ddot{Q}$, where Q is the net charge contained within V_S, eqn (4.54) is integrated over the volume V_S and differentiated with respect to time. Integration over V_S yields

$$-\int_{V_S} d^3r\,\nabla\cdot\vec{j} \;= \frac{d}{dt}\int_{V_S} d^3r\,\rho\,. \tag{4.55}$$

$$= \dot{Q}\,. \tag{4.56}$$

Next, differentiating eqn (4.56) with respect to time yields $\ddot{Q}$ on the right hand side and $\nabla\cdot\dot{\vec{j}}$ inside the integral on the left hand side. When the expression for $\dot{\vec{j}}$ given by eqn (4.53) is inserted on the left hand side, eqn (4.56) becomes

$$\ddot{Q} + \frac{n_e e^2}{\varepsilon_{\text{lat}}\, m_e}\int_{V_S} d^3r\,\nabla\cdot\vec{D} = 0\,. \tag{4.57}$$

Finally, replacing $\nabla\cdot\vec{D}$ with $4\pi\rho$, and replacing $4\pi n_e e^2/\varepsilon_{\text{lat}}\, m_e$ with $\omega_p{}^2$, yields the differential equation that describes the time behavior of the charge:

$$\ddot{Q} + \omega_p{}^2 Q = 0\,. \tag{4.58}$$

The charge undergoes simple harmonic oscillation at the screened plasma frequency. Plasma oscillation is seen to be a characteristic of the system's very large polarizability, which is due to the conduction band electrons responding to fields as if they were essentially free particles.

Hydrodynamics Near the Fermi Surface

Wave propagation only takes place in media that display spatial dispersion. In other words, it is necessary to have a retarded response to a disturbance that takes place elsewhere. If a charge disturbance is entirely local, as with a local charge fluctuation, the disturbance undergoes damped oscillation without propagating.

One way to appreciate how this works with a propagating longitudinal plasmon is by considering the system's hydrodynamic response. Suppose a region of plasma undergoes compression. This increases the conduction band electron density n_e and therefore the Fermi level increases (recall that $n_e = zE_f^{3/2}$; see eqn (3.27) in Chapter 3). Work must be done to achieve the compression, in which case ω_p^2 must be augmented by a term of the form $\beta^2 k^2$. That is, $\omega_p^2 \rightarrow \omega_p^2 + \beta^2 k^2$. Note that wave vector must enter this expression quadratically, as a linear (odd power) dependence would imply directional preference in an isotropic medium, and higher powers (k^4, k^6 ...) are associated with nonlinear behavior, which is of no concern here. The term β is referred to as compressibility.

We shall now obtain an expression for β by using the electrostatic limit: $\omega = 0$. Instead of the dielectric function for the longitudinal field, $1 - \omega_p^2 / \omega^2$, vanishing at $\omega = \omega_p$, we have $\omega^2 = \omega_p^2 + \beta^2 k^2 = 0$. In this case k is imaginary and the inverse of its magnitude can be equated to the Thomas-Fermi screening length, r_{TF}. Some algebra then gives an expression for β as follows:

$$\beta^2 = -\frac{\omega_p^2}{k^2} = \omega_p^2 r_{TF}^2 \tag{4.59}$$

Using the expression for r_{TF} given by eqn (3.27) yields

$$\beta^2 = \frac{4\pi n_e e^2}{m} \frac{1}{6\pi e^2 z E_f^{1/2}} = \frac{2n_e}{3mzE_f^{1/2}} \tag{4.60}$$

The fact that $n_e = zE_f^{3/2}$ yields the neat expression

$$\beta = \sqrt{\frac{2}{3}\frac{E_f}{m}} = 3^{-1/2} v_F . \tag{4.61}$$

A direct calculation of β using an ansatz of local equilibrium yields exactly the same result, $\beta = 3^{-1/2} v_F$. It is interesting that $v_F^2 / 3$ is equal to the root-mean-square value of any one of the Fermi velocity's Cartesian components. For example, the dispersion relation for the longitudinal plasmon can be written: $\omega^2 = \omega_p^2 + v_z^2 k^2$ with the understanding that the v_z in this expression is equal to $[\langle v_z \rangle^2]^{1/2}$. I do not know if any significance should be attached to this, but we recognize this dispersion relation as being a quite familiar one.

When the above result is incorporated into the dielectric function for the longitudinal plasmon, its dispersion relation is given by

$$\varepsilon(k,\omega) = 1 - \frac{\omega_p{}^2}{\omega^2 - \beta^2 k^2} . \tag{4.62}$$

Quantization

This brief subsection is important enough to warrant more than one reading. That's right: as simple as it may appear I suggest that you go over it a couple of times.

In the above classical model, it is seen that the charge Q executes simple harmonic motion. This tells us immediately that in the quantum mechanical case the quanta of the plasma field are bosons of energy $\hbar\omega_p$. These quanta are referred to as plasmons. Some math is usually required to go from the classical plasma field to its quantum mechanical counterpart. However, we have already worked through the related case of phonons in Chapter 1. It was found that the dual-use model presented there has the following dispersion relation:

$$\omega_k{}^2 = \Omega_0{}^2 + 4\Omega^2 \sin^2 \tfrac{1}{2} ka . \tag{1.39}$$

The small-k limit of this dispersion relation, which in the phonon case is the same dispersion relation as for a continuous mass density, gives

$$\omega_k{}^2 = \Omega_0{}^2 + v^2 k^2 . \tag{1.39'}$$

In going from eqn (1.39) to eqn (1.39'), use has been made of the fact that Ωa is equal to the speed of sound, v, which is preserved in the limit of a continuous mass density. Amazingly, this is the same dispersion relation, aside from the values of the constants, as for the case of the lossless plasma. Namely, using $\varepsilon = \varepsilon_{\text{lat}}(1 - \omega_p{}^2 / \omega^2)$, together with $ck = \varepsilon^{1/2}\omega$, yields:

$$\omega^2 = \omega_p{}^2 + (c/n)^2 k^2 , \tag{4.63}$$

where $n = \varepsilon_{\text{lat}}{}^{1/2}$.

The bottom line is that we have already done the work needed to quantize the plasma field. The dual-use model introduced at the outset of Chapter 1 has come to the rescue, sparing us the tedious math that would be necessary to quantize the plasma field. It was not merely a lucky coincidence that everything fell into place. The dual-use model can be extended to other field theories as well, as we shall see in due course.

Intermediate Summary

Before proceeding to other geometries, let us summarize the results obtained so far in Section 4.3. We began with the ansatz of an instantaneous local charge disturbance. This disturbance can be either an increase or decrease of electron density in V_{Q0} relative to its initial (equilibrium) value. Because of the disturbance the plasma is no longer electrically neutral in this volume.

When the electron density is raised quickly in V_{Q0} the electrons push against one another in an attempt to return the system to one of overall charge neutrality. This can only be achieved by lowering the electron density in V_{Q0}, as the positive charge density is fixed. It is assumed that the positive charges are immobile, as in a metal. Thus, electron density flows out of V_{Q0}. The excess electron density continues to expand in its attempt to restore charge neutrality, causing an electron flux to pass through a surface S. The electron density overshoots its mark, causing the electron density within S to eventually become less than the equilibrium value. In this way the electron density proceeds to oscillate about the equilibrium value at the plasma frequency until the oscillation amplitude is damped through one or more loss mechanisms. Quantization of the plasma field follows from a one-to-one correspondence with the model introduced at the beginning of Part B. Thus, plasmons and their properties enter without ado.

High-Energy Electrons Create Plasmons

The fact that plasmon quanta are created through the abrupt introduction of a charge disturbance means that a wave is launched over a sufficiently short time interval to justify calling it $t = 0$. We have seen that, to a high degree of accuracy, it is not possible for an electromagnetic wave that passes through a plasma to create plasmons because of wave vector mismatch. On the other hand, no problem arises due to wave vector mismatch for the case of a charged particle (including its dynamic screening cloud) that passes through the plasma (Fig. 11). Its wave vector is aligned with the forces it exerts on the plasma. Thus, it has no difficulty creating longitudinal oscillations.

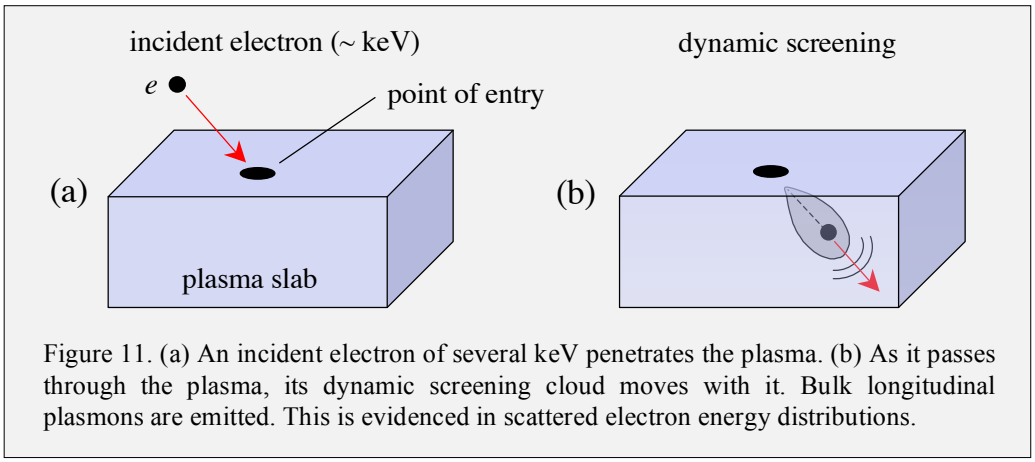


Figure 11. (a) An incident electron of several keV penetrates the plasma. (b) As it passes through the plasma, its dynamic screening cloud moves with it. Bulk longitudinal plasmons are emitted. This is evidenced in scattered electron energy distributions.

To obtain a qualitative appreciation of how this works, consider Fig. 11. In (a), an incident electron of a few keV is poised to pass through the surface of a plasma slab at the indicated location on the plasma's upper surface. In (b), the electron has entered the plasma and is passing through it. Dynamic screening results in a teardrop-shaped disturbance that accompanies the particle trajectory. The high-energy particle plus its accompanying disturbance emits plasmons along its trajectory as indicated by wave fronts. This registers as energy loss of the transmitted (scattered) electrons.

Electron energy loss spectroscopy has been used frequently to obtain energies for bulk and surface plasmons. Figure 12 shows rather nice results for the scattering of 2 keV electrons from an Al sample. The bulk plasmon energy is 15.3 eV and the surface plasmon energy is 10.3 eV. The structure to the right of the main peaks (red arrows) is due to excitation of multiple quanta. I found this figure online in a slideshow prepared by Anuradha Verma. She apparently works at a place called the Dayalbagh Educational Institute. I was unable to track down the origin of the figure in order to give proper credit. If you are able to do so, please let me know.

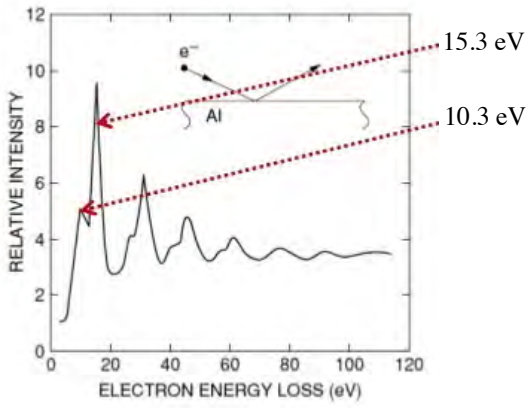


Figure 12. Electron energy loss spectroscopy: 2 keV electrons reflected from an Al sample display multiple quanta of surface and bulk plasmons; adapted from Verma [62].

Polaron

Before discussing flat and spherical plasma surfaces, let us look briefly at an interesting phenomenon that is related to the passage of a high-energy electron through plasma that was mentioned in the previous subsection.

A polaron is an electron that moves through a solid under the influence of an electric field, carrying with it a polarization cloud. There is nothing special about the polarization cloud: An electron injected into the solid simply polarizes the electron densities of the nearby atoms. This polarization can occur quickly relative to the time scale of the electron's motion (drift velocity) in the solid. In this case, the electron's polarization cloud is expected to be more spherical than is the case of a high-energy electron passing through plasma (Fig. 11).

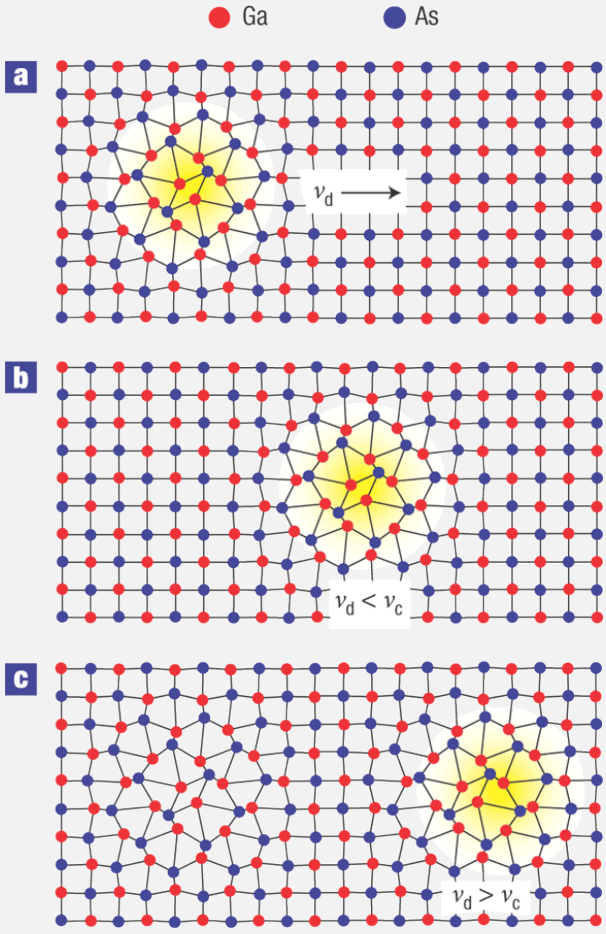


Figure 13. A polaron consists of an electron (yellow cloud) and the resulting lattice distortion [51].

Referring to Fig. 13, an electric field causes a large polaron to acquire a drift velocity. Below a critical velocity the polaron travels as a well-defined quasiparticle. However, if the field is sufficiently large, the polaron acquires enough kinetic energy that relative electron motion within the polarization cloud becomes important. Consequently, longitudinal optical phonons are emitted.

There is an important difference between the polaron and some of the other field quanta we have dealt with (and will deal with), namely: phonons, plasmons, and photons. These are bosons, whereas the polaron is a fermion. It can be said that the electron is "dressed" by the polarization cloud it carries with it. Therefore the fact that it is a fermion comes as no surprise. Nonetheless, the mathematical route to the polaron is more challenging. A fermion field is involved, as opposed to the scalar fields that yield phonons and plasmons, and the vector field that yields photons.

Charge Disturbance at a Flat Surface

Referring to Fig. 14, consider a charge disturbance that appears at $t = 0$ at the flat surface of the plasma. It is assumed that the charge disturbance is comprised of excess electrons, as was the case with Fig. 10. Electric field lines emanate from the $t = 0$ excess electron density. In the present case, current obviously cannot flow into the vacuum above the plasma surface. To see how this geometry affects the oscillation of the surface charge, let us revisit eqn (4.55):

$$-\int_{V_S} d^3r\, \nabla\cdot\vec{j} \;=\; \frac{d}{dt}\int_{V_S} d^3r\, \rho\,. \tag{4.55}$$

This equation was used earlier to describe a charge disturbance deep within the plasma. It is easily adapted for use here.

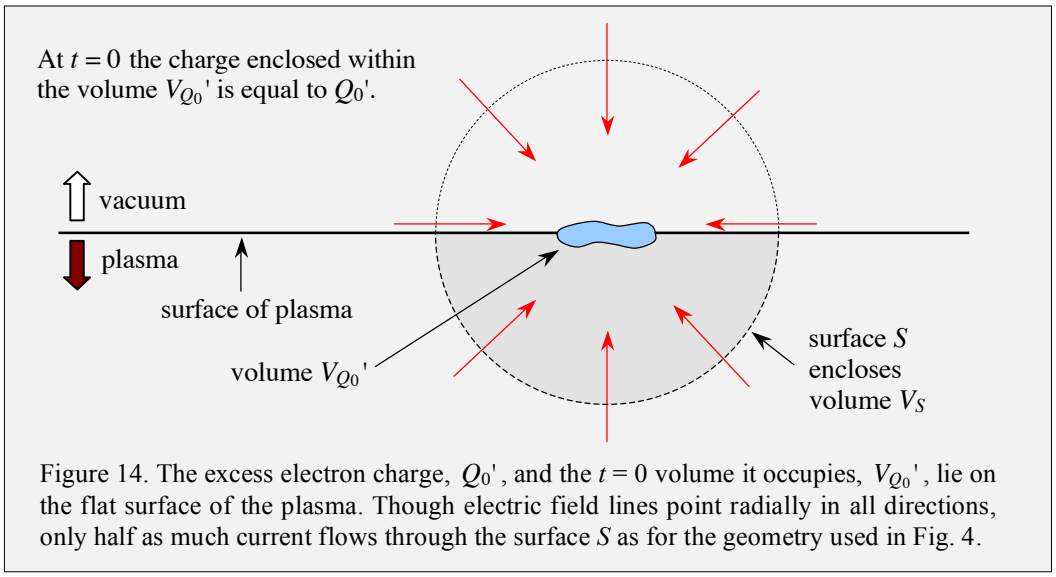


Figure 14. The excess electron charge, Q_0', and the $t = 0$ volume it occupies, V_{Q0}', lie on the flat surface of the plasma. Though electric field lines point radially in all directions, only half as much current flows through the surface S as for the geometry used in Fig. 4.

Integration over the volume V_S of the $t = 0$ charge density ρ returns Q, as this is the total amount of charge in the $t = 0$ disturbance. The left hand side of eqn (4.55) is also applicable here, as long as we take into account the fact that $\vec{j}$ is only nonzero at and below the plasma surface. In other words, because of the fact that the current density is equal to zero above the surface, integration of the electric field lines is carried out over only the lower half of the spherical surface S. This returns a value that is one-half the size of its bulk counterpart. Thus, the equation of motion given by eqn (4.58) acquires a factor of one-half, becoming

$$\ddot{Q} + \tfrac{1}{2}\omega_p{}^2 Q = 0\,. \tag{4.64}$$

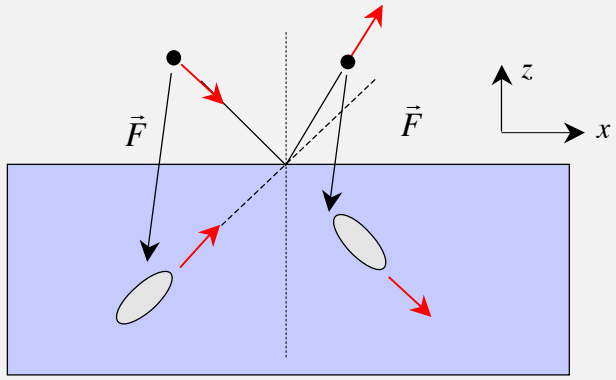
It follows that the characteristic frequency of charge oscillation is not ω_p, but instead

$$\omega_{sp} = \frac{1}{\sqrt{2}}\omega_p . \tag{4.65}$$

where ω_{sp} denotes the surface plasmon frequency. It is interesting that this is exactly the same value that was obtained earlier in the chapter through consideration of wave vector matching at the surface. In order to apply eqns (4.64) and (4.65), it is necessary that the charge disturbance remains centered at the surface. Charge will flow through the surface S, but this flow of electron density overshoots its mark, resulting in net positive charge within S. As in the case of the bulk disturbance, the system proceeds to oscillate, in the present case with frequency $\omega_p / \sqrt{2}$. As mentioned earlier, we shall see in Chapter 7 that the surface plasmon combines with a photon to yield an elementary excitation referred to as a surface plasmon polariton. This polariton propagates according to the dispersion relation for the mixed photon-surface-plasmon system.

To see how a surface plasmon can be excited using an incident charged particle such as an electron, let us enlist a scenario that is similar to the one in Fig. 11, where transmission through bulk plasma was considered. Figure 15 indicates an electron that scatters from the surface rather than penetrating it. The incident electron has much lower energy than the one that penetrates the surface in Fig. 10, for example, a few tens of eV rather than a few keV. For a positive image charge to appear within the sample requires that Thomas-Fermi type screening takes place. In other words, electron density must move out of the region that is to have a positive charge. This screening is dynamical, with the image lagging the incident electron because of the time required for the screening cloud to develop. The image charge is not a point, but instead a teardrop-shaped region. The ovals in Fig. 15 represent the teardrop-shaped regions.

Figure 15. An incident electron (few tens of eV) scatters from a metal surface. A force (black arrows) acts on the electron because of its image charge.



Referring to Fig. 15, an attractive force (black arrows) acts between the electron and its blurred image (gray cloud) throughout the electron's incoming and outgoing trajectories. This force has components in the $-x$ and $-z$ directions for both incident and scattered trajectories. Over the duration of the incoming trajectory the electron gains less energy than it would if its positive image charge could follow the electron trajectory instantaneously. On the outgoing trajectory, the integral of the decelerating force $\vec{F}$ has a larger magnitude than the corresponding integral for the accelerating force that acted over the

incoming trajectory. The net effect is that the electron loses both energy and horizontal momentum in the scattering process. When the incident angle and kinetic energy are adjusted appropriately, that is, to match the dispersion relation for a surface plasmon, wave vector matching is achieved and a surface plasmon is created. The kinetic energy of the scattered electron is reduced by the energy of the surface plasmon, and the scattered electron is not traveling in the specular direction, but instead its path leans toward the normal, as indicated in Fig. 15.

Surface and Bulk Plasmons via Pulsed Laser Radiation

It is also possible to launch surface and bulk plasmons by using intense pulsed laser radiation. Think of it this way. A laser pulse interacts with the surface on a time scale that is short relative to that of the subsequent propagation dynamics of bulk and surface plasmons. Such a scenario can be realized using femtosecond laser pulses. The radiation's electric field creates a charge disturbance at the surface, and this charge disturbance must do something. In general, a bulk plasma wave will pass downward, and a surface wave will commence propagating along the surface. From energy and momentum conservation we know that the scattering event has altered the incident radiation, giving it new energy and momentum.

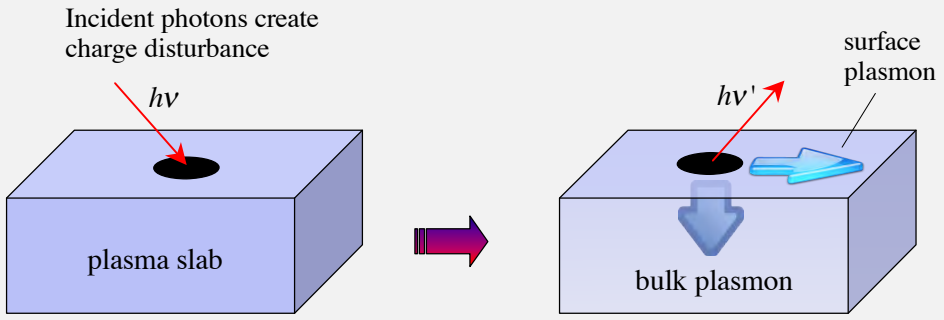


Figure 16. A p-polarized incident field can create surface and bulk plasmons. The field is of ultra-short duration. It creates a charge disturbance that subsequently propagates.

Small Metal Spheres

In the present context, a small metal sphere is defined as one whose diameter is significantly smaller that the wavelength of the electromagnetic field to which the sphere is coupled. This means that the sphere can be treated mathematically by using the electric dipole approximation in the theory of interaction of radiation with matter. Moreover, the system can be said to be quasi-static due to the fact that at any instant of time the electric field does not vary over the dimension of the sphere. In other words, the field oscillates sinusoidally in time, but there is no spatial variation over the sphere. Thus, retardation can be ignored. This sets the plasma oscillation of the small object apart from the bulk and surface plasma waves discussed earlier in this chapter.

The *dipolar surface plasma frequency* is obtained by using the expression from electrostatics for the electric field of a uniformly polarized sphere (Purcell [34], pp. 373-378). The only difference between the dielectric case discussed in Purcell and the plasma sphere is orientation of the fields. In each case, $\vec{E} = -4\pi\vec{P}/3$ applies. For the geometry indicated in Fig. 17, the mathematical relationship is

$$E_z = -\frac{4\pi}{3} P_z \,. \tag{4.66}$$

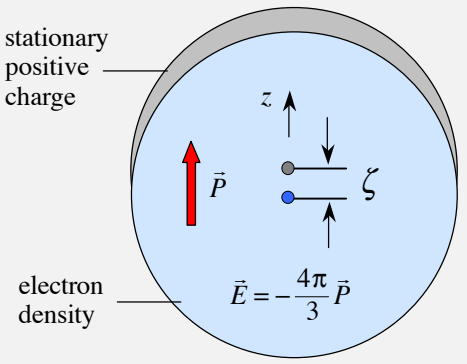


Figure 17. A sphere of uniform electron density (blue) oscillates relative to a sphere of uniform positive charge density (gray). Amplitude is exaggerated; to scale, it would hardly be noticed. The gray and blue dots near the center indicate the centers of the positive and negative charge densities, respectively.

The spherical shapes of the positive and negative charges enable them to be replaced with point charges (Fig. 17). Leaving aside the sign of the charge, the amount of charge in each sphere is $(4\pi a^3/3)n_e e$, which corresponds to a dipole moment of magnitude $(4\pi a^3/3)n_e e\zeta$ and a polarization $\vec{P}$ whose magnitude is $n_e e\zeta$. Using this with eqn (4.66) yields

$$E_z = \frac{1}{3} 4\pi n_e e \zeta \,. \tag{4.67}$$

From here, it is a straightforward matter to obtain the expression for the oscillation of the electron density sphere with respect to the stationary positively charged sphere. The electrons move in concert with one another, as they are strongly correlated. The equation of motion of the dipole displacement is $m\ddot{\zeta} = -e\vec{E}$, which becomes

$$\ddot{\zeta} = -\frac{e}{m_e}\vec{E} = -\frac{e}{m_e}\frac{1}{3} 4\pi n_e e \zeta \,. \tag{4.68}$$

Identifying the plasma frequency, this becomes

$$\ddot{\zeta} + \frac{1}{3}\omega_p{}^2 \zeta = 0 \,. \tag{4.69}$$

Thus, the characteristic plasma oscillation frequency of the small sphere is given by

$$\omega_{dsp} = \frac{1}{\sqrt{3}}\omega_p \,, \tag{4.70}$$

where ω_{dsp} denotes the dipolar surface plasmon frequency.

This exercise underscores the fact that the shape of the plasma influences strongly the characteristic frequency of the plasma oscillation. In fact, nowhere in the derivation of eqn (4.70) did the *size* of the object enter, except to require that it is small compared to a wavelength. By adjusting the shape of a metal or doped semiconductor particle one can tune the characteristic frequency of its plasma oscillation. This turns out to be one of the most attractive features of the nanoscale regime.

Dielectric Response

There is an important difference between a charge disturbance in electrically neutral plasma, in which the total electron charge and the total positive charge are conserved separately, versus the injection of charge (say electrons) into the system, or the removal of charge from it. In the case of an isolated small metal sphere subjected to an external radiation field, it is obvious that the overall system remains electrically neutral. Even if a few electrons are removed through photoionization, for all practical purposes the system can be treated as electrically neutral. For a macroscopic system such as a piece of metal connected to something, electrons might enter and leave. Nonetheless, to a good degree of approximation the system can be treated as electrically neutral.

In the absence of a noticeable amount of charge being injected into, or removed from, the otherwise neutral plasma, the system must satisfy $\nabla \cdot \vec{D} = 0$. This leaves open the possibility of local charge imbalance, specifically, of the type discussed in the previous subsections. In other words, the microscopic relation: $\nabla \cdot \vec{E} = 4\pi\rho$, can be applied locally even though $\nabla \cdot \vec{D} = 0$, where $\vec{D} = \varepsilon(\omega)\vec{E}$. The dielectric function $\varepsilon(\omega)$ does not depend on spatial coordinates, in which case $\nabla \cdot \vec{D} = 0$ requires that $\varepsilon(\omega)\nabla \cdot \vec{E} = 0$. Therefore, $\varepsilon(\omega)$ must vanish. Moreover, note that the condition $\nabla \cdot \vec{E} \neq 0$ is satisfied by a longitudinal wave, while a transverse field satisfies $\nabla \cdot \vec{E} = 0$.