Plane Waves and Wave Propagation

Augustin Jean Fresnel
(1788 - 1827)

October 10, 2001

Contents

In this chapter we start by considering plane waves in infinite or semi-infinite media. We shall look at their properties in both insulating and conducting materials and shall give some thought to the possible properties of materials of different kinds. We will also look at the reflection and refraction of waves at planar boundaries between different materials, a topic which forms the basis for much of physical optics. If time allows, we shall also look at some of the more abstract aspects of wave propagation having to do with causality and signal propagation.

1 Plane Waves in Uniform Linear Isotropic Non-conducting Media

1.1 The Wave Equation

One of the most important predictions of the Maxwell equations is the existence of electromagnetic waves which can transport energy. The simplest solutions are plane waves in infinite media, and we shall explore these now.

Consider a material in which

$$\mathbf{B} = \mu\mathbf{H} \qquad \mathbf{D} = \epsilon\mathbf{E} \qquad \mathbf{J} = \rho = 0. \tag{1}$$

Then the Maxwell equations read

$$\nabla\cdot\mathbf{E} = 0 \qquad \nabla\cdot\mathbf{B} = 0 \qquad \nabla\times\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} \qquad \nabla\times\mathbf{B} = \frac{\mu\epsilon}{c}\frac{\partial\mathbf{E}}{\partial t}. \tag{2}$$

Now we do several simple manipulations that will become second nature. First take the curl of one of the curl equations, e.g., Faraday's law, to find

$$\nabla\times(\nabla\times\mathbf{E}) = \nabla(\nabla\cdot\mathbf{E}) - \nabla^2\mathbf{E} = -\frac{1}{c}\frac{\partial}{\partial t}(\nabla\times\mathbf{B}) = -\frac{\mu\epsilon}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}, \tag{3}$$

where the generalized Ampère's law was employed in the last step. Because the divergence of $\mathbf{E}$ is zero, this equation may be written as

$$\left(\nabla^2 - \frac{\mu\epsilon}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{E} = 0. \tag{4}$$

Identical manipulations starting from Ampère's law rather than Faraday's law also lead to

$$\left(\nabla^2 - \frac{\mu\epsilon}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{B} = 0. \tag{5}$$

Thus any Cartesian component of $\mathbf{E}$ or $\mathbf{B}$ obeys a classical wave equation of the form

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\psi(\mathbf{x}, t) = 0, \tag{6}$$

where $v = c/\sqrt{\mu\epsilon}$.

There is a simple set of complex traveling wave solutions to this equation. They are of the form

$$u_{\mathbf{k}}(\mathbf{x}, t) = e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \tag{7}$$

where $\omega = vk$ and $\mathbf{k}$ is any real vector.[1] Notice that the derivatives of this function are

$$\begin{aligned} \nabla u_{\mathbf{k}} &= i\mathbf{k}u_{\mathbf{k}} \\ \nabla^2 u_{\mathbf{k}} &= -k^2 u_{\mathbf{k}} \\ \frac{\partial u_{\mathbf{k}}}{\partial t} &= -i\omega u_{\mathbf{k}} \\ \frac{\partial^2 u_{\mathbf{k}}}{\partial t^2} &= -\omega^2 u_{\mathbf{k}}. \end{aligned} \tag{8}$$

Hence

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)u_{\mathbf{k}} = \left(-k^2 + \frac{\omega^2}{v^2}\right)u_{\mathbf{k}} = 0, \tag{9}$$

demonstrating that we do indeed have a solution of the wave equation.

This solution is a wave "traveling" in the direction of $\mathbf{k}$ in the sense that a point of constant phase, meaning $\mathbf{k}\cdot\mathbf{x} - \omega t = constant$, moves along this direction with a speed v which is ω/k. Furthermore, we have a *plane wave*, by which we mean that a surface of constant phase is a plane; in particular, the surfaces of constant phase are just planes perpendicular to $\mathbf{k}$.

[1] This vector is real if ϵ and μ are real; they can be complex, in which case there are still solutions of this form with complex $\mathbf{k}$.

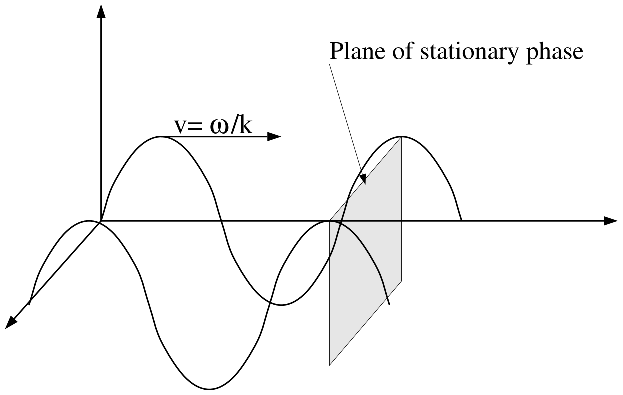


Fig.1: A point of stationary phase moves with velocity $|v| = \omega/k$

1.2 Conditions Imposed by Maxwell's Equations

Next, let us see how the electromagnetic fields can be described in terms of these scalar plane waves. Let us look for an electric field and a magnetic induction with the forms

$$\mathbf{E}(\mathbf{x},t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \qquad \mathbf{B}(\mathbf{x},t) = \mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \tag{10}$$

with the understanding that the true fields are the real parts of these complex expressions.

In addition to satisfying the wave equation, the complex fields must be solutions of the Maxwell equations. Let us see what additional constraints are thereby imposed. Consider first the divergence equations; these require that

$$0 = \nabla\cdot\mathbf{B}(\mathbf{x},t) = \nabla\cdot\left[\mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}\right] = i\mathbf{k}\cdot\mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \tag{11}$$

and

$$0 = \nabla\cdot\mathbf{E}(\mathbf{x},t) = \nabla\cdot\left[\mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}\right] = i\mathbf{k}\cdot\mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}. \tag{12}$$

Or

$$\mathbf{k}\cdot\mathbf{B}_0 = 0 \qquad \text{and} \qquad \mathbf{k}\cdot\mathbf{E}_0 = 0. \tag{13}$$

These conditions mean that $\mathbf{B}_0$ and $\mathbf{E}_0$ must be perpendicular to $\mathbf{k}$, which is to say, parallel to the surfaces of constant phase and perpendicular to the direction in which the surface of constant phase is moving. Such an electromagnetic wave is called a *transverse wave*. Notice that this nomenclature is consistent with our definition in the last chapter of a transverse vector field as one having zero divergence.

There are further conditions on the amplitudes $\mathbf{E}_0$ and $\mathbf{B}_0$ from the other Maxwell equations. From the Ampère law one has

$$\nabla \times \mathbf{B}(\mathbf{x}, t) = \frac{\mu\epsilon}{c}\frac{\partial \mathbf{E}(\mathbf{x}, t)}{\partial t} \tag{14}$$

which leads to

$$i\mathbf{k} \times \mathbf{B}_0 = -\frac{i\omega\epsilon\mu}{c}\mathbf{E}_0 \tag{15}$$

or

$$\mathbf{E}_0 = -\frac{\mathbf{k} \times \mathbf{B}_0}{k\sqrt{\mu\epsilon}} = -\frac{\mathbf{n} \times \mathbf{B}_0}{\sqrt{\epsilon\mu}} \tag{16}$$

where $\mathbf{n} = \mathbf{k}/k$ is a unit vector in the direction of propagation of the wave. From Faraday's Law and similar manipulations one finds the further, and final condition that

$$\mathbf{B}_0 = \sqrt{\mu\epsilon}(\mathbf{n} \times \mathbf{E}_0); \tag{17}$$

however, one may also find this relation from Eq. (16) and the condition that $\mathbf{n}\cdot\mathbf{B}_0 = 0$ and so it is not an additional constraint. Alternatively, one may derive Eq. (16) from Eq. (17) and the condition $\mathbf{n} \cdot \mathbf{E}_0 = 0$. As a consequence, one may, for example, write

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \tag{18}$$

where the only condition on $\mathbf{E}_0$ is $\mathbf{n} \cdot \mathbf{E}_0 = 0$. Then $\mathbf{B}(\mathbf{x}, t)$ follows from Eq. (17) and is

$$\mathbf{B}(\mathbf{x}, t) = \sqrt{\mu\epsilon}(\mathbf{n} \times \mathbf{E}_0)e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}. \tag{19}$$

Alternatively, we may start by writing

$$\mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \tag{20}$$

where $\mathbf{B}_0$ is orthogonal to $\mathbf{k}$, $\mathbf{n} \cdot \mathbf{B}_0 = 0$. Then $\mathbf{E}(\mathbf{x}, t)$ is given from Eq. (16) as

$$\mathbf{E}(\mathbf{x}, t) = -\frac{\mathbf{n} \times \mathbf{B}_0}{\sqrt{\epsilon\mu}} e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}. \tag{21}$$

From these conditions, and those obtained in the previous paragraph, we may conclude that $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{k}$ form a mutually orthogonal set.

Before leaving this section, let's look at the time-averaged energy density and Poynting vector in such electromagnetic waves. We shall write them in terms of the amplitude $\mathbf{E}_0$. First,

$$< \mathbf{S} >= \frac{c}{8\pi}\Re\left[\mathbf{E}(\mathbf{x},t)\times\mathbf{H}^*(\mathbf{x},t)\right] = \frac{c}{8\pi}\sqrt{\frac{\epsilon}{\mu}}\Re\left[\mathbf{E}_0\times(\mathbf{n}\times\mathbf{E}_0^*)\right] = \frac{c}{8\pi}\sqrt{\frac{\epsilon}{\mu}}|\mathbf{E}_0|^2\mathbf{n}. \quad (22)$$

Similarly,

$$< u >= \frac{1}{16\pi}\Re\left(\mathbf{E}(\mathbf{x},t)\cdot\mathbf{D}^*(\mathbf{x},t)+\mathbf{B}(\mathbf{x},t)\cdot\mathbf{H}^*(\mathbf{x},t)\right] = \frac{\epsilon}{8\pi}|\mathbf{E}_0|^2. \quad (23)$$

The time-averaged momentum density is:

$$< \mathbf{g} >= \frac{1}{8\pi c}\Re\left[\mathbf{E}(\mathbf{x},t)\times\mathbf{H}^*(\mathbf{x},t)\right] = \frac{\sqrt{\epsilon/\mu}}{8\pi c}|E_0|^2\mathbf{n}. \quad (24)$$

The evaluation of the time-averaged Maxwell stress tensor is left as an exercise.

2 Polarization

In this section we address the question of the most general possible monochromatic plane wave, which amounts to asking what are the possible choices of $\mathbf{E}_0$. Let us specify that $\mathbf{k} = k\boldsymbol{\epsilon_3}$ and suppose that we have an orthogonal right-handed set of real unit basis vectors $\boldsymbol{\epsilon_i}$, $i = 1,2,3$. Then it must be the case that $\mathbf{E}_0\cdot\boldsymbol{\epsilon_3} = 0$ which means that the most general amplitude $\mathbf{E}_0$ can be expanded as

$$\mathbf{E}_0 = E_{01}\boldsymbol{\epsilon_1} + E_{02}\boldsymbol{\epsilon_2}. \quad (25)$$

The scalar amplitudes in this expansion can be complex so we have in all four real amplitudes which we may choose with complete abandon. Let us write the complex scalar amplitudes in polar form,

$$E_{01} = E_1 e^{i\phi_1} \qquad\qquad E_{02} = E_2 e^{i\phi_2} \quad (26)$$

where E_i and ϕ_i, $i = 1, 2$, are real. Further, introduce

$$E_0 = (E_1^2 + E_2^2)^{1/2} \qquad \text{and} \qquad \phi = \phi_2 - \phi_1. \tag{27}$$

Then the complex field becomes

$$\mathbf{E}(\mathbf{x}, t) = E_0 \alpha_1 \left(\boldsymbol{\epsilon_1} + (\alpha_2/\alpha_1) e^{i\phi} \boldsymbol{\epsilon_2} \right) e^{i(\mathbf{k}\boldsymbol{\epsilon_3}\cdot\mathbf{x} - \omega t)} e^{i\phi_1} \tag{28}$$

where $\alpha_i = E_i/E_0$ and $\alpha_1^2 + \alpha_2^2 = 1$. In this form, the wave is seen to have just two interesting parameters, α_2/α_1 and $\phi_2 - \phi_1$; these specify the relative phase and amplitude of the two components of the vector amplitude. The other two parameters simply to set the overall magnitude of the field and its absolute phase[2].

Look at the real part of the complex wave as a function of time at a point in space which is conveniently taken to be the origin. Aside from the overall magnitude and phase, the wave looks like

$$\mathbf{E} \sim \boldsymbol{\epsilon_1} \cos(\omega t) + (\alpha_2/\alpha_1) \boldsymbol{\epsilon_2} \cos(\omega t - \phi). \tag{29}$$

If we map out the path traced by the tip of this vector in the space of ϵ_1 and ϵ_2, we find in general an ellipse. The ellipse is characterized by two parameters, equivalent to α_2/α_1 and ϕ, these being its eccentricity (the ratio of the semi-minor to the semi-major axis) and the amount by which the major axis is rotated relative to some fixed direction such as that of ϵ_1. Such a wave is said to be *elliptically polarized*, the term "polarization" referring to the behavior of the electric field at a point as a function of time. There are two limiting special cases. One is when the eccentricity is unity in which case the ellipse becomes a circle and the wave is said to be *circularly polarized*; the second is when the eccentricity becomes zero so that the ellipse reduces to a line and the wave is *linearly polarized*.

[2] These will, of course, be interesting if the wave meets another wave; but they are not interesting if there is no other wave.

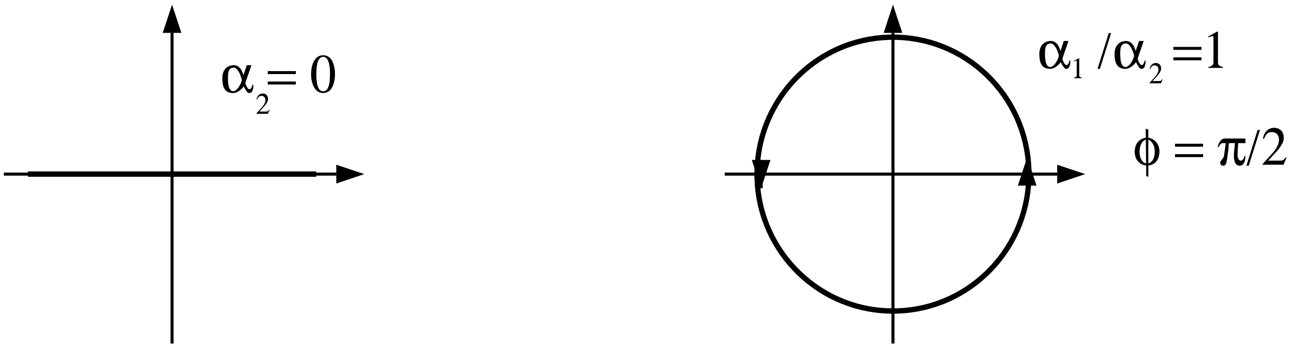


Fig.2: linearly ($\alpha_2 = 0$) and circularly ($\alpha_2/\alpha_1 = 1 \;\; \phi = \pi/2$) polarized

Often one uses a set of complex basis vectors in which $\boldsymbol{\epsilon_1}$ and $\boldsymbol{\epsilon_2}$ are replaced by vectors $\boldsymbol{\epsilon}_\pm$ defined by

$$\boldsymbol{\epsilon}_\pm \equiv \frac{1}{\sqrt{2}}(\boldsymbol{\epsilon_1} \pm i\boldsymbol{\epsilon_2}). \tag{30}$$

These have the properties

$$\boldsymbol{\epsilon}_\pm \cdot \boldsymbol{\epsilon_3} = 0 \qquad \boldsymbol{\epsilon}_\pm \cdot \boldsymbol{\epsilon}_\mp{}^* = 0 \qquad \boldsymbol{\epsilon}_\pm \cdot \boldsymbol{\epsilon}_\pm{}^* = 1, \tag{31}$$

and it is possible to write the electric field of a general plane wave as

$$\mathbf{E}(\mathbf{x}, t) = (E_+\boldsymbol{\epsilon}_+ + E_-\boldsymbol{\epsilon}_-)e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}, \tag{32}$$

where E_+ and E_- are arbitrary complex constants. If just one of these is non-zero and is written in polar form, then, aside from phase, the complex electric field at a point is

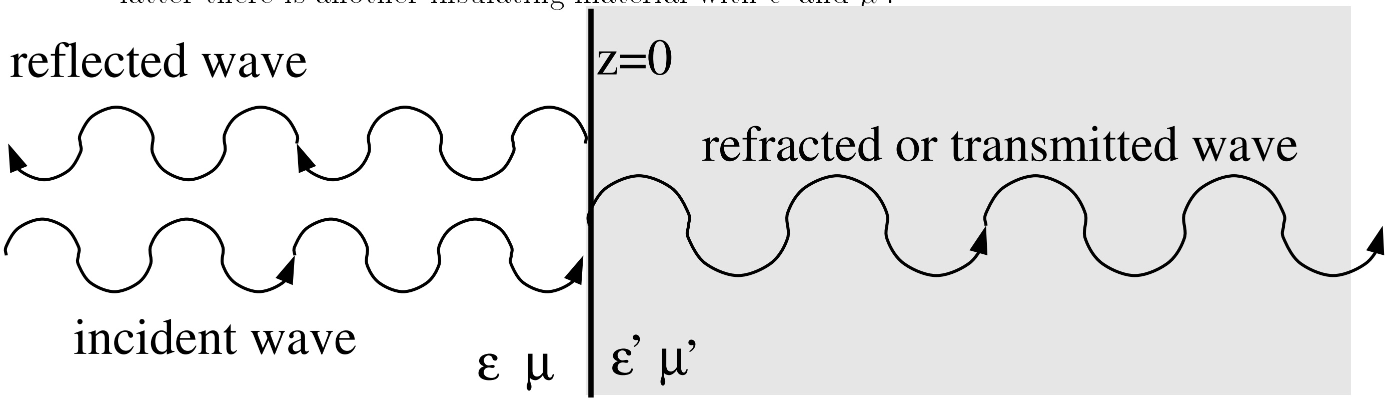
$$\mathbf{E} = |E_\pm|\frac{1}{\sqrt{2}}(\boldsymbol{\epsilon_1} \pm i\boldsymbol{\epsilon_2})e^{-i\omega t}. \tag{33}$$

The real part then varies as $\boldsymbol{\epsilon_1}\cos(\omega t) \pm \boldsymbol{\epsilon_2}\sin(\omega t)$ which is a circularly polarized wave. In the case of the upper sign, one says that the wave is *left-circularly polarized* or that it has *positive helicity*; in the case of the lower sign, it is *right-circularly polarized* or has *negative helicity*. In writing the general wave in terms of these basis vectors, we have expressed it as a superposition of positive and negative helicity waves with amplitudes E_+ and E_-, respectively.

3 Boundary Conditions; Waves at an Interface

In this section, we shall find out what plane waves must look like in semi-infinite media or when there is a planar boundary between two nonconducting materials such as air (or vacuum) and glass. We will need appropriate continuity conditions on the fields at the interface. There may be derived from general kinematic considerations, and from Maxwell equations.

The basic example from which all cases may be inferred is that of a planar interface located at $z = 0$ dividing space into two regions, $z < 0$ and $z > 0$. In the former, we assume an insulating material with dielectric constant ϵ and permeability μ; in the latter there is another insulating material with ϵ' and μ'.



Now suppose that from the left, or $z < 0$, there is an *incident* wave which has electromagnetic fields

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \qquad \mathbf{B}(\mathbf{x}, t) = \sqrt{\mu\epsilon}\frac{\mathbf{k} \times \mathbf{E}(\mathbf{x}, t)}{k}. \tag{34}$$

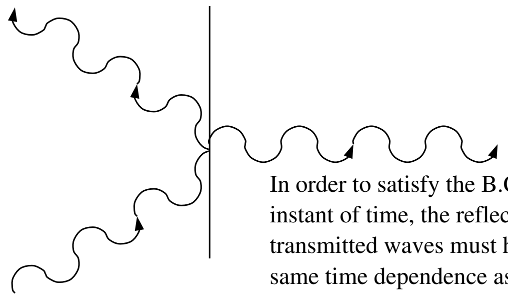
Also, $k = \omega\sqrt{\mu\epsilon}/c$, and $\mathbf{k}\cdot\hat{\mathbf{z}} > 0$ so that the wave is approaching the interface. Finally, $\mathbf{E}_0$ is such that $\mathbf{k} \cdot \mathbf{E}_0 = 0$.

The incident wave is a solution of the Maxwell equations in the region $z < 0$. At the interface, however, it is not a solution; there must be other waves present in order to satisfy the Maxwell equations (or boundary conditions) here. To phrase it another way, when the incident wave hits the interface, additional waves, called *transmitted* (or *refracted*) and *reflected* waves must be generated. The refracted waves are the ones that propagate into the medium at $z > 0$; the reflected waves are the ones that propagate back into the other medium.

3.1 Kinematic Conditions

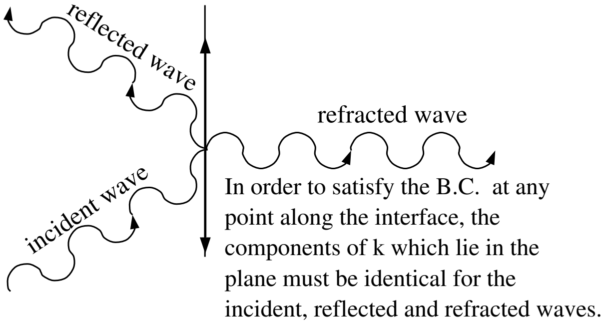
We can, from quite general considerations, learn a lot about the properties of the reflected and refracted waves.

First, in order that the continuity conditions remain satisfied at all times, given that they are satisfied at one instant of time, these waves must have the same time dependence as the incident wave. This statement follows from the linear nature of the field equations (each term in the equations is proportional to some component of one of the fields). Hence, all fields vary in time as $e^{-i\omega t}$.



In order to satisfy the B.C. at any instant of time, the reflected and transmitted waves must have the same time dependence as the incident wave (i.e. same frequency).

Second, the continuity conditions must be satisfied at all points on the interface or $z = 0$ plane. Suppose that they are satisfied at one particular point, such as $\mathbf{x} = 0$. Then, in order that they remain so for other points on the interface, each wave must vary in the same fashion as each of the other waves as one moves in the plane of the interface. This statement follows, as does the first one, from the linearity of the field equations. Now, since the dependence of a plane wave on position is $\exp(i\mathbf{k} \cdot \mathbf{x})$, this condition means that all waves (incident, reflected, and refracted) must have wave vectors whose components lying in the plane of the interface are identical.



We can express this condition as

$$\mathbf{n} \times \mathbf{k} = \mathbf{n} \times \mathbf{k}' = \mathbf{n} \times \mathbf{k}'' \tag{35}$$

where $\mathbf{k}'$ and $\mathbf{k}''$ are, respectively, the wave vectors of any refracted and reflected waves. This relation may also be written as

$$k \sin i = k'' \sin r'' = k' \sin r \tag{36}$$

where i, r'', and r are the angles between the wavevectors of the incident, reflected, and transmitted waves and the normal to the interface. They are called the *angle of incidence*, the *angle of reflection*, and the *angle of refraction*.

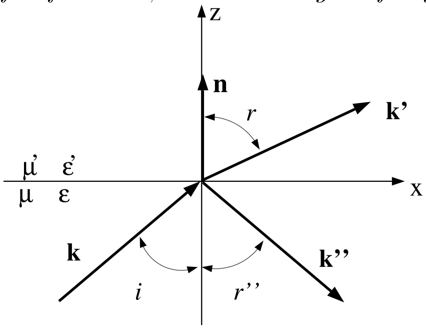


Figure 6: Definition of the angles i, r'', and r

Finally, any reflected wave is a solution of the same wave equation as the incident wave; consequently, it has a wave number $k'' = k$. Any transmitted wave, however, has wave number $k' = \omega\sqrt{\mu'\epsilon'}/c$, so $k' \neq k$. If we combine these statements with Eq. (36), we can see that $r'' = i$, *the angle of incidence equals the angle of reflection.*

For a transmitted wave, however, the wave equation is such that $k' = \omega\sqrt{\mu'\epsilon'}/c$ which is not k; in fact, $k'/n' = k/n$ where

$$n \equiv \sqrt{\mu\epsilon} \qquad \text{and} \qquad n' \equiv \sqrt{\mu'\epsilon'} \tag{37}$$

are the *indices of refraction* in the two materials. Using these definitions in Eq. (36) we find

$$n \sin i = n' \sin r \tag{38}$$

which is known in optics as *Snell's Law*.

3.2 Conditions from Maxwell's Equations

Notice that we derived Snell's law and the statement $i = r''$ without using explicitly the continuity conditions; we had only to use the fact that there **are** linear continuity conditions. Hence these properties are called kinematic properties (they don't depend on the particular dynamics of the fields which are given by the Maxwell equations) and they are applicable to much more than just electromagnetic phenomena.

To fully develop the rules of reflection and refraction for electromagnetic waves, we must use the Maxwell equations to tell us the specific relations among the fields and then must apply the continuity conditions at a specific point on the interface, such as $\mathbf{x} = 0$, and at a specific time, such as $t = 0$[3].

$$\nabla \cdot \mathbf{D} = 4\pi\rho, \ \nabla \cdot \mathbf{B} = 0, \ \nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \ \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}. \tag{39}$$

[3]For other points and times we know that the conditions will be satisfied by making sure the kinematic conditions derived above are satisfied.

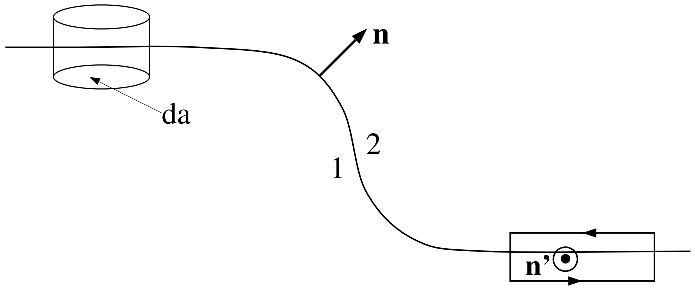


Figure 3: Integration surfaces used for the B.C.

Application of the divergence theorem to the two divergence equations using the familiar pillbox construction leads, as for the static case, to the continuity conditions

$$(\mathbf{D}_1 - \mathbf{D}_2) \cdot \mathbf{n} = 4\pi\sigma \tag{40}$$

$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n} = 0 \tag{41}$$

where $\mathbf{n}$ is a unit outward normal from material 1 and σ is the macroscopic surface-charge density. Application of Stokes' theorem to the curl equations in the "usual" way leads to

$$\mathbf{n} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \tag{42}$$

$$\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = \frac{4\pi}{c}\mathbf{K} \tag{43}$$

where $\mathbf{K}$ is the macroscopic surface-current density lying inside of the loop C to which Stokes' theorem is applied. Notice that the time derivatives in Faraday's law and Ampère's law do not contribute to the continuity conditions.[4]

For uncharged insulators, the surface sources σ and $\mathbf{K}$ are always zero; then the continuity conditions are especially simple and state that the normal components of $\mathbf{D}$ and $\mathbf{B}$ are continuous as are the tangential components of $\mathbf{H}$ and $\mathbf{E}$.

At $\mathbf{x} = 0$, $t = 0$, the fields of an incident wave, a single transmitted wave, and a single reflected wave[5] may be written as follows:

[4] We have assumed that there are no singular parts of the time derivatives localized at the interface; were there any such contributions, they would show up in the continuity conditions.

[5] We don't know at this point that we need only one reflected and one transmitted wave to obtain a solution to the boundary value problem. By construction, we will see that such is the case.

Incident wave:

$$\mathbf{E} = \mathbf{E}_0 \qquad \mathbf{B} = \frac{n}{k}(\mathbf{k} \times \mathbf{E}_0) \tag{44}$$

Reflected wave:

$$\mathbf{E} = \mathbf{E}_0'' \qquad \mathbf{B} = \frac{n}{k}(\mathbf{k}'' \times \mathbf{E}_0'') \tag{45}$$

Transmitted wave:

$$\mathbf{E} = \mathbf{E}_0' \qquad \mathbf{B} = \frac{n'}{k'}(\mathbf{k}' \times \mathbf{E}_0') = \frac{n}{k}(\mathbf{k}' \times \mathbf{E}_0'). \tag{46}$$

We suppose that we are given n, n', $\mathbf{k}$, and $\mathbf{E}_0$; we need to find $\mathbf{k}'$, $\mathbf{k}''$, $\mathbf{E}_0'$, and $\mathbf{E}_0''$. The wave vectors follow from the kinematic relations; they all lie in the plane containing the normal to the interface and the incident wave vector, called the *plane of incidence* and make angles with the normal as discussed above. As for the amplitudes, they are found from the continuity conditions:

1. D_n continuous:

$$\epsilon(\mathbf{E}_0 + \mathbf{E}_0'') \cdot \mathbf{n} = \epsilon' \mathbf{E}_0' \cdot \mathbf{n} \tag{47}$$

2. B_n continuous:

$$(\mathbf{k} \times \mathbf{E}_0 + \mathbf{k}'' \times \mathbf{E}_0'') \cdot \mathbf{n} = (\mathbf{k}' \times \mathbf{E}_0') \cdot \mathbf{n} \tag{48}$$

3. $\mathbf{E}_t$ continuous:

$$(\mathbf{E}_0 + \mathbf{E}_0'') \times \mathbf{n} = \mathbf{E}_0' \times \mathbf{n} \tag{49}$$

4. $\mathbf{H}_t$ continuous:

$$\frac{1}{\mu}(\mathbf{k} \times \mathbf{E}_0 + \mathbf{k}'' \times \mathbf{E}_0'') \times \mathbf{n} = \frac{1}{\mu'}(\mathbf{k}' \times \mathbf{E}_0') \times \mathbf{n}. \tag{50}$$

It is a messy bit of algebra to solve these equations in the general case. The task can be made simpler by writing the incident wave's electric field as a linear combination of two linearly polarized waves, which is always possible. One solves each of these cases separately. The appropriate sum of the two solutions is then the solution of the original problem. Once again, the linearity of the field equations leads to enormous simplification of the algebra. The two cases that we are going to treat are

1. polarization of $\mathbf{E}_0$ perpendicular to the plane of incidence and

2. polarization parallel to the plane of incidence.

3.2.1 Polarization of $\mathbf{E}_0$ Perpendicular to the Plane

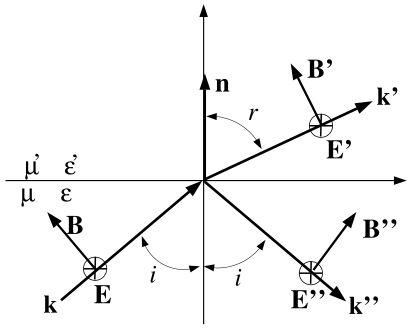


Figure 7: Polarization of $\mathbf{E}_0$ *perpendicular to the plane of incidence*

The figure sets the conventions for the first case. They are such that $\mathbf{E}_0 = E_0\hat{\mathbf{y}}$, $\mathbf{E}_0' = E_0'\hat{\mathbf{y}}$, and $\mathbf{E}_0'' = E_0''\hat{\mathbf{y}}$. Remember also that $k'' = k$, $k/n = k'/n'$, and $n \sin i = n' \sin r$. Now apply the four continuity conditions. The first gives nothing because there is no normal component of the electric displacement or electric field; the second gives $E_0 + E_0'' = E_0'$; the third gives the same constraint as the second; and the fourth results in $(k/\mu)\cos i\,(E_0 - E_0'') = (k'/\mu')\cos r\, E_0'$. Since $k' = kn'/n$ and $n = \sqrt{\mu\epsilon}$, we can write the latter as $\sqrt{\epsilon/\mu}\cos i\,(E_0 - E_0'') = \sqrt{\epsilon'/\mu'}\cos r\, E_0'$. In addition, $\cos r = \sqrt{1-\sin^2 r} = \sqrt{1-(n/n')^2\sin^2 i}$. Combining these relations we find the two conditions

$$E_0' - E_0'' = E_0 \qquad \sqrt{\frac{\epsilon'}{\mu'}}\sqrt{1-\left(\frac{n}{n'}\right)^2\sin^2 i}\,E_0' + \sqrt{\frac{\epsilon}{\mu}}\cos i\,E_0'' = \sqrt{\frac{\epsilon}{\mu}}\cos i\,E_0. \quad (51)$$

Notice that these are written entirely in terms of the angle of incidence; the angle of refraction does not appear. Their solution is easily shown to be

$$E_0' = \frac{2n\cos i}{n\cos i + (\mu/\mu')\sqrt{n'^2 - n^2\sin^2 i}}E_0$$

and

$$E_0'' = \frac{n\cos i - (\mu/\mu')\sqrt{n'^2 - n^2\sin^2 i}}{n\cos i + (\mu/\mu')\sqrt{n'^2 - n^2\sin^2 i}} E_0 \tag{52}$$

3.2.2 Polarization of $\mathbf{E}_0$ Parallel to the Plane

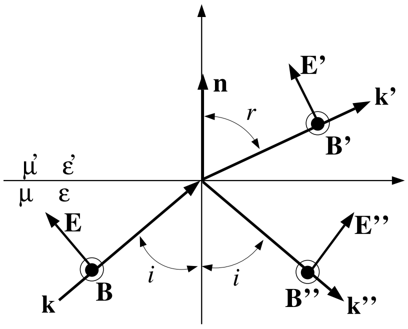


Figure 8: Polarization of $\mathbf{E}_0$ *parallel to the plane of incidence*

The second case, polarization in the plane of incidence may be similarly analyzed. The figure shows the conventions for this case. They are such that $\mathbf{E}_0 = E_0(\sin i\,\hat{\mathbf{z}} - \cos i\,\hat{\mathbf{x}})$, $\mathbf{E}_0' = E_0'(\sin r\,\hat{\mathbf{z}} - \cos r\,\hat{\mathbf{x}})$, and $\mathbf{E}_0'' = E_0''(\sin i\,\hat{\mathbf{z}} + \cos i\,\hat{\mathbf{x}})$. The first boundary condition implies that $\epsilon \sin i\,(E_0 + E_0'') = \epsilon' \sin r\, E_0'$; the second gives nothing; the third gives $\cos i\,(-E_0 + E_0'') = -\cos r\, E_0'$; and the fourth gives a condition that is redundant with the first when Snell's law is invoked. Thus we may write the two conditions, after removing all occurrences of r as in the first case, as

$$\sqrt{\frac{\epsilon}{\mu}}(E_0 + E_0'') = \sqrt{\frac{\epsilon'}{\mu'}}E_0' \qquad\qquad \cos i\,(E_0 + E_0'') = \sqrt{1 - (n/n')^2\sin^2 i}\,E_0'. \tag{53}$$

Their solution is

$$E_0' = \frac{2nn'\cos i}{(\mu/\mu')n'^2\cos i + n\sqrt{n'^2 - n^2\sin^2 i}} E_0$$

and

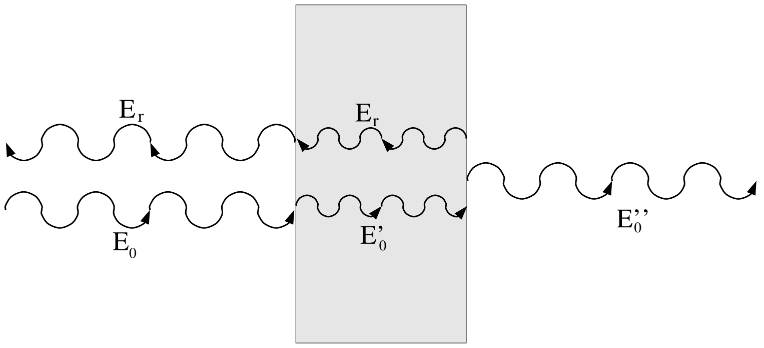
$$E_0'' = \frac{(\mu/\mu')n'^2\cos i - n\sqrt{n'^2 - n^2\sin^2 i}}{(\mu/\mu')n'^2\cos i + n\sqrt{n'^2 - n^2\sin^2 i}} E_0 \tag{54}$$

Our solutions to the reflection-refraction problem have the following characteristics by design. First, as mentioned above, they involve only the angle of incidence, the angle of refraction having been removed wherever it appeared by using Snell's law; second, the material properties enter through the permeabilities and indices of refraction as opposed to the permeabilities and dielectric constants. The reason is that for most of the materials one encounters, $\mu = \mu' = 1$ and so the permeabilities drop out of the relations. Second, one is generally more likely to be given an index of refraction than a dielectric constant and so expressing the amplitudes in terms of n makes them more readily applicable.[6]

Equations (52) and (54) are known as *Fresnel's equations*; with them we can calculate the reflection and transmission of a plane wave at a planar interface for arbitrary initial polarization. Such an incident wave gives rise to a single reflected plane wave and a single transmitted plane wave, meaning that there is just one reflected wave vector $\mathbf{k}''$ and one transmitted wave vector $\mathbf{k}'$.

3.3 Parallel Interfaces

With a little thought we may see how to generalize to the case of two (or more) parallel interfaces. Consider the figure showing two parallel interfaces separating three materials. If we follow the consequences of an incident plane wave from the first material on one side we can see that the reflection processes within the middle material of the "sandwich" generate many plane waves in here, but that these waves have just two distinct wave vectors.



[6] Of course, the relation between n and ϵ is sufficiently simple that there is really no great difference.

Figure 9: Plane wave incident on a sandwich.

Also, all waves transmitted into the third material have the same wave vector, and the "reflected" waves in the first medium all have a single wave vector. Hence one finds that in the first medium, there are just two waves with electric fields

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \qquad \mathbf{E}_r = \mathbf{E}_{r0} e^{i(\mathbf{k}_r\cdot\mathbf{x}-\omega t)}; \tag{55}$$

in the middle medium there are again just two distinct waves with fields

$$\mathbf{E}' = \mathbf{E}'_0 e^{i(\mathbf{k}'\cdot\mathbf{x}-\omega t)} \qquad \mathbf{E}'_r = \mathbf{E}'_{r0} e^{i(\mathbf{k}'_r\cdot\mathbf{x}-\omega t)} \tag{56}$$

and in the third medium there is just one plane wave with field

$$\mathbf{E}'' = \mathbf{E}''_0 e^{i(\mathbf{k}''\cdot\mathbf{x}-\omega t)}. \tag{57}$$

To find the four amplitudes $\mathbf{E}_{r0}$, $\mathbf{E}'_0$, $\mathbf{E}'_{r0}$, and $\mathbf{E}''_0$, one must apply the boundary conditions at the two interfaces, leading to four distinct linear relations involving these amplitudes and that of the incident wave, $\mathbf{E}_0$. Solving these equations, one finds the amplitudes of all waves in terms of that of the incident wave.

Returning briefly to Fresnel's equations for reflection and refraction at a single interface, let us look at the special case of normal incidence, $i = 0$. then $r = 0$ also, and the first set (polarization normal to the plane of incidence) of Fresnel equations tells us that [7]

$$E'_0 = \frac{2n}{n + (\mu/\mu')n'} E_0 \qquad E''_0 = \frac{n - (\mu/\mu')n'}{n + (\mu/\mu')n'} E_0. \tag{58}$$

These are simple results, especially when $\mu = \mu'$. They clearly tell us that when the two materials have comparable indices of refraction and permeabilities, the wave is mostly transmitted and when they have very different properties (an engineer would

[7] Actually, both sets of Fresnel equations are applicable for normal incidence. The second set, however, will produce a result with some signs switched as a consequence of the different conventions used for the directions of the electric fields in the two cases.

call that impedance mismatch), reflection is the rule. Notice also that if $n'\mu/\mu' > n$, the reflected amplitude is opposite in sign to the incident one, meaning that the electric field of the reflected wave is phase shifted by π radians relative to that of the incident one under these circumstances.

4 Reflection and Transmission Coefficients

In this section we look at the power or energy transmitted and reflected at an interface between two insulators. To do so, we must evaluate the time-averaged power in the incident, reflected, and transmitted waves which is done by calculating the Poynting vector. The energy current density toward or away from the interface is then given by the component of the Poynting vector in the direction normal to the interface. In the second medium, where there is just a single (refracted) wave, the normal component of $\mathbf{S}$ is unambiguously the transmitted power per unit area. But in the first medium, the total electromagnetic field is the sum of the fields of the incident and reflected waves. In evaluating $\mathbf{E} \times \mathbf{H}$, one finds three kinds of terms. There is one which is the cross-product of the fields in the incident wave, and its normal component gives the incident power per unit area. A second is the cross-product of the fields in the reflected wave, giving the reflected power. But there are also two cross-terms involving the electric field of one of the plane waves and the magnetic field of the other one. It turns out that the time-average of the normal component of these terms is zero, so that they may be ignored in the present context. Bearing this in mind, we have the following quantities of interest:
The time-averaged incident power per unit area:

$$\mathcal{P} = <\mathbf{S}> \cdot \mathbf{n} = \frac{c}{8\pi}\sqrt{\frac{\epsilon}{\mu}}|\mathbf{E}_0|^2 \ \frac{\mathbf{k}\cdot\mathbf{n}}{k} \tag{59}$$

The time-averaged reflected power per unit area:

$$\mathcal{P}'' = - <\mathbf{S}''> \cdot \mathbf{n} = \frac{c}{8\pi}\sqrt{\frac{\epsilon}{\mu}}|\mathbf{E}_0''|^2 \frac{\mathbf{k}''\cdot\mathbf{n}}{k} \tag{60}$$

The time-averaged transmitted power per unit area:

$$\mathcal{P}' = <\mathbf{S}'> \cdot \mathbf{n} = \frac{c}{8\pi}\sqrt{\frac{\epsilon'}{\mu'}}|\mathbf{E}'_0|^2 \frac{\mathbf{k}' \cdot \mathbf{n}}{k'} d \tag{61}$$

The *reflection coefficient* R and the *transmission coefficient* T are defined as the ratios of the reflected and transmitted power to the incident power.

We may calculate the reflection and transmission coefficients for the cases of polarization perpendicular and parallel to the plane of incidence by using the Fresnel equations. If an incident wave has general polarization so that its fields are linear combinations of these two special cases, then there is once again the possibility of cross terms in the power involving an electric field with one type of polarization and a magnetic field with the other type. Fortunately, these turn out to vanish, so that one may treat the two polarizations individually.

For the case of **polarization perpendicular to the plane of incidence**, we use the Fresnel equations (52) and (54) for the reflected and transmitted amplitudes and have

$$T = \frac{\sqrt{\frac{\epsilon'}{\mu'}} \frac{4n^2 \cos^2 i \cos r}{(n \cos i + (\mu/\mu')\sqrt{n'^2 - n^2 \sin^2 i})^2}}{\sqrt{\frac{\epsilon}{\mu}} \cos i} \tag{62}$$

Making use of the relations $n = \sqrt{\epsilon\mu}$, $n' = \sqrt{\epsilon'\mu'}$, $\sin r = (n/n') \sin i$, and $\cos i = \sqrt{1 - \sin^2 i}$, one finds that

$$T = \frac{4n(\mu/\mu') \cos i \sqrt{n'^2 - n^2 \sin^2 i}}{[n \cos i + (\mu/\mu')\sqrt{n'^2 - n^2 \sin^2 i}\,]^2}. \tag{63}$$

By similar means one can write the reflection coefficient as

$$R = \frac{[n \cos i - (\mu/\mu')\sqrt{n'^2 - n^2 \sin^2 i}\,]^2}{[n \cos i + (\mu/\mu')\sqrt{n'^2 - n^2 \sin^2 i}\,]^2} \tag{64}$$

By inspection one can see that $R + T = 1$ which expresses the conservation of energy; what is not transmitted is reflected.

The case of **polarization in the plane of incidence** is treated similarly. One finds

$$T = \frac{4nn'^2(\mu/\mu') \cos i \sqrt{n'^2 - n^2 \sin^2 i}}{[(\mu/\mu')n'^2 \cos i + n\sqrt{n'^2 - n^2 \sin^2 i}\,]^2} \tag{65}$$

and

$$R = \frac{[(\mu/\mu')n'^2\cos i - n\sqrt{n'^2 - n^2\sin^2 i}\,]^2}{[(\mu/\mu')n'^2\cos i + n\sqrt{n'^2 - n^2\sin^2 i}\,]^2}. \tag{66}$$

Once again, $R + T = 1$.

5 Examples

5.1 Polarization by Reflection

From inspection of Fresnel's equations, we can see that the relative amounts of transmitted and reflected amplitude depend on the state of polarization and are distinctly not the same for both polarizations.

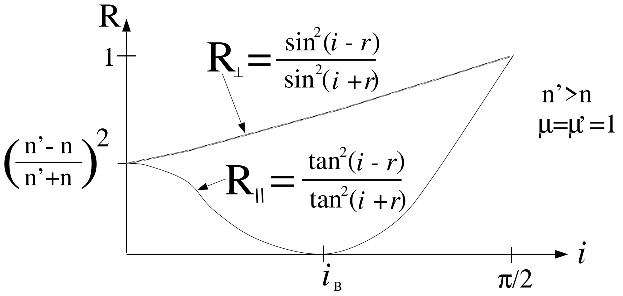


Figure 10: Reflection Coefficient when $n' > n$, and $\mu' = \mu = 1$

That means that in the general case, the polarizations of the transmitted and reflected waves will not be the same as that of the incident one. A very special case has to do with the reflected wave given incident polarization in the plane of incidence. We see that the reflected amplitude will vanish if[8]

$$n'^2\cos i = n\sqrt{n'^2 - n^2\sin^2 i}. \tag{67}$$

Squaring this relation we find

$$n'^4\cos^2 i = n^2 n'^2 - n^4\sin^2 i = n'^4(1-\sin^2 i) \text{ or } \sin^2 i = \frac{n'^2}{n'^2+n^2} \text{ or } \tan i = \frac{n'}{n}. \tag{68}$$

[8] We let $\mu = \mu'$ in this section unless explicitly stated otherwise; keeping the permeability around usually contributes nothing but extra work and obfuscation.

This special angle of incidence is called the *Brewster angle*,

$$i_B = \arctan(n'/n); \tag{69}$$

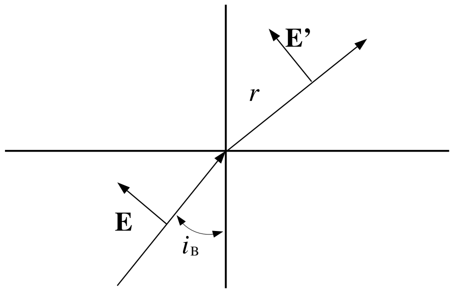


Figure 11: No reflected wave when $i = i_B$ and the field is polarized in the plane.

a wave polarized in the plane of incidence and incident on the interface at the Brewster angel is completely transmitted with no reflected wave. If a wave of general polarization is incident at the Brewster angle, then the reflected wave is completely (linearly) polarized perpendicular to the plane of incidence. Hence this phenomenon provides a method for obtaining a linearly polarized wave from an unpolarized one. More generally, if the angle of incidence is reasonably close to the Brewster angle, the reflected light is to a large degree polarized perpendicular to the plane of incidence. This fact is utilized by polarizing sun glasses which screen out most of the light polarized parallel to the surface of the earth, which is to say, most of the light reflected by the earth.

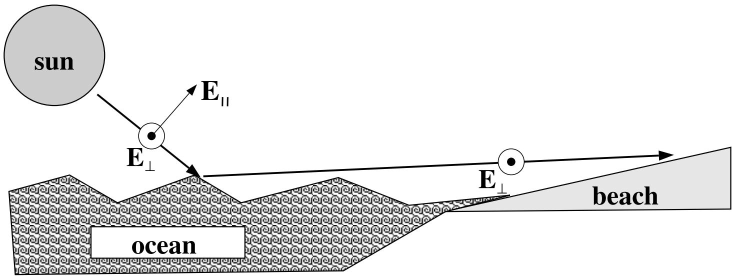


Figure 12: Light reflected from the ocean (glare) is largely polarized along the horizon, and may be removed with polarized sunglasses.

5.2 Total Internal Reflection

As a second example we look at the phenomenon of total internal reflection which is the opposite of the one just considered in that no energy is transmitted across an interface under appropriate conditions. Suppose that $n > n'$. As shown in the figure, this means that $r > i$.

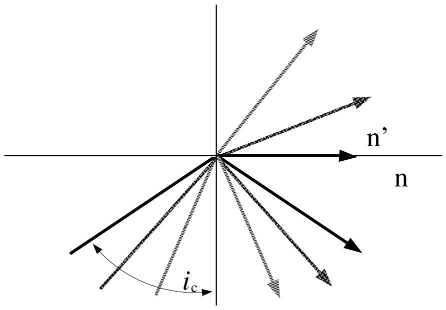


Figure 13: A series of angles when $n > n'$.

Now consider an incident wave with i large enough that $n \sin i > n'$. How can we have a refracted wave with r such that Snell's law, $n \sin i = n' \sin r$ is satisfied? Recall our argument for Snell's law; it was based on the fact that the wave vector $\mathbf{k}'$ of the refracted wave had to have a component k'_t parallel to the interface equal to the same component of the incident wave. Given that $n \sin i > n'$, this condition means that k'_t is **larger** than $\omega n'/c$ which is supposed to be the magnitude of $\mathbf{k}'$, according to the wave equation. But there is a way around this. The condition that comes from the wave equation is that, if k'_t and k'_n are respectively the components of $\mathbf{k}'$ tangential and normal to the interface, then $k'^2_t + k'^2_n = \omega^2 n'^2/c^2$. If $k'_t > \omega n'/c$, we can satisfy this condition by having k'_n be imaginary. In particular,

$$k'_n = \pm i \frac{n\omega}{c} \sqrt{\sin^2 i - (n'/n)^2}. \tag{70}$$

The choice of sign has to be such as to produce a wave that damps away to nothing in the second medium; otherwise it becomes exceedingly large (which is unphysical behavior) as one moves far away from the interface. Now that we have figured out what is $\mathbf{k}'$; that is, $k'_t = (n\omega/c) \sin i$ and k'_n is given by Eq. (70), we can see the character of the transmitted electric field. It is

$$\mathbf{E}' \sim e^{ik_t'x}e^{-|k_n'|z}e^{-i\omega t} \tag{71}$$

where $\hat{\mathbf{x}}$ is the direction of the tangential component of $\mathbf{k}$.

The Poynting vector for a wave of this sort has no component directed normal to the interface although there is one parallel to the interface. To see this, take $\mathbf{E}$ to be in the y-direction.

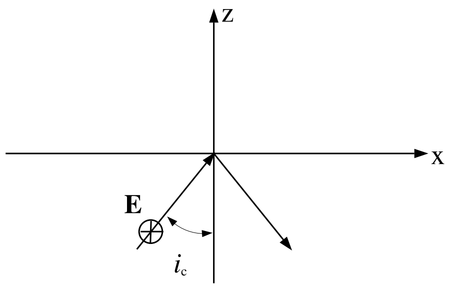


Figure 14: Polarization $\perp$ to the plane of incidence.

Then

$$\mathbf{E}' = \mathbf{E}_0' e^{i(k_t'x-\omega t)}e^{-|k_n'|z}$$

so that

$$S_z' = \frac{c}{8\pi}\Re\left(\mathbf{E}' \times \mathbf{B}'^*\right)_z = \frac{c}{8\pi}\Re\left(E_y' \times B_x'^*\right) .$$

We may use Faraday's law to relate $\mathbf{E}$ to $\mathbf{B}$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial B}{\partial t} \rightarrow \frac{i\omega}{c}B_x' = |k_n'|E_y' .$$

Thus,

$$S_z' = \frac{c}{8\pi}\Re\left\{-E_y'\frac{c|k_n'|}{-i\omega}E_y'^*\right\} = 0$$

Thus, as shown in the figure below, when $i > i_c$, the power is totally reflected.

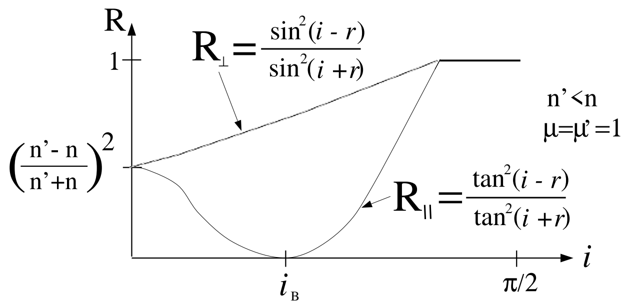


Figure 15: Reflection Coefficient when $n > n'$, and $\mu' = \mu = 1$

What we have is therefore a *surface wave* confined to the region close to the interface and transporting energy parallel to it. Moreover, by evaluating the Poynting vector of the reflected and incident waves, one finds that as much energy is reflected from the interface as is incident upon it. Hence we have the phenomenon of perfect or total reflection of the incident wave. This phenomenon is utilized in fiber optics; an electromagnetic wave is propagated inside of a thin tube of some material having a large index of refraction and surrounded by another material having a much smaller index. Wherever the wave is incident upon the wall of the tube, it is completely reflected.

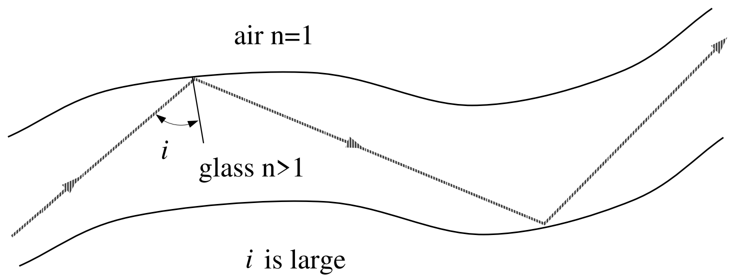


Figure 16: Total internal reflection occurs within a fiber optic tube.

There is some natural attenuation of the wave because of imperfect dielectric properties of the material itself or its coating; nevertheless, a beam of light, for example, can be transmitted long distances and around many curves (as long as they aren't too sharp) in such a "pipe."

6 Models of Dielectric Functions

The dielectric "constant" of almost any material is in fact a function of frequency, meaning that it has different values for waves of different frequencies.

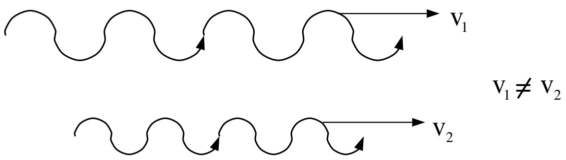


Figure 17: In a dispersive medium waves of different frequencies have different phase velocities $v = c/\sqrt{\epsilon(\omega)\mu}$.

We can make a simple model of the dielectric "function" of an insulating material as follows: Suppose that the charges which primarily respond to an electric field are electrons bound on atoms or molecules. Let one such electron be harmonically bound, meaning that the binding forces are treated as linear in the displacement of the charge from its equilibrium position. Also, let there be a damping force proportional to the velocity $\mathbf{v}$ of the electron. Then, if the mass and charge of the electron are m and $-e$, the equation of motion of the electron under the influence of an electric field $\mathbf{E}(\mathbf{x}, t)$ is

$$m\left(\frac{d^2\mathbf{x}}{dt^2} + \gamma\frac{d\mathbf{x}}{dt} + \omega_0^2\mathbf{x}\right) = -e\mathbf{E}(\mathbf{x}, t). \tag{72}$$

The harmonic restoring force is expressed through a "natural" frequency of oscillation ω_0 of the electron. We have ignored the possible influence of a magnetic induction $\mathbf{B}(\mathbf{x}, t)$ on the electron's motion. Typically this force is much smaller than the electric field force because the electron's speed is much smaller than c; there can be exceptions, however, and one of them is explored below.

Next, the typical magnitude of the electron's displacement $|\mathbf{x}|$ is on the order of an atomic size.

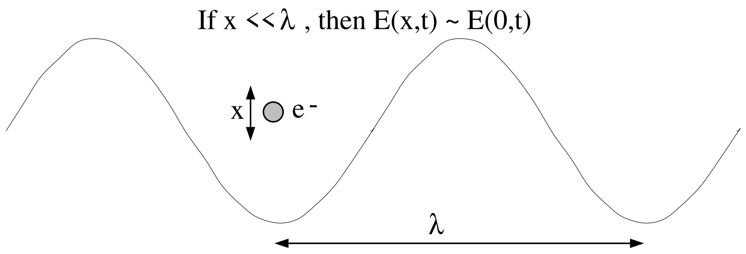


Figure 18: If the wavelength of the incident wave is much larger than the electronic displacement, then we may neglect the spacial dependence of $\mathbf{E}(\mathbf{x},t)$.

If the electric field $\mathbf{E}(\mathbf{x},t)$ is that of visible or even ultraviolet light, then the displacement is much smaller than distances over which $\mathbf{E}(\mathbf{x},t)$ varies significantly, meaning that we can approximate $\mathbf{E}(\mathbf{x},t) \approx \mathbf{E}(0,t) = \mathbf{E}_0 \exp(-i\omega t)$. In this limit, the solution we seek is of the form $\mathbf{x}(t) = \mathbf{x}_0 \exp(-i\omega t)$. Substituting into the equation of motion, we find that the equation for the amplitude of the motion is

$$m(-\omega^2 - i\omega\gamma + \omega_0^2)\mathbf{x}_0 = -e\mathbf{E}_0 \tag{73}$$

or

$$\mathbf{x}_0 = \frac{-e\mathbf{E}_0}{m(\omega_0^2 - i\omega\gamma - \omega^2)}. \tag{74}$$

The amplitude of the dipole moment associated with the motion of this electron is $\mathbf{p}_0 = -ex_0$. To find the polarization, we need to compute the dipole moments of all electrons in some finite volume of material. These electrons will not all have the same damping or natural frequencies, so let us say that there are n molecules per unit volume with z electrons each. If f_i of the electrons on each molecule have resonant frequency ω_i and damping constant γ_i, then we get a polarization or dipole moment per unit volume which varies harmonically with an amplitude

$$\mathbf{P}_0 = e^2\mathbf{E}_0 n \sum_j \left(\frac{f_j}{m(\omega_j^2 - i\omega\gamma_j - \omega^2)}\right); \tag{75}$$

this is also the relation between $\mathbf{E}(\mathbf{x},t)$ and $\mathbf{P}(\mathbf{x},t)$. If we further say that $\mathbf{E}(\mathbf{x},t)$

is the macroscopic field[9], then we can write $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \epsilon\mathbf{E}$ with the preceding expression for the polarization. The result is an expression for $\epsilon(\omega)$:

$$\epsilon(\omega) = 1 + \frac{4\pi n e^2}{m}\sum_j \left(\frac{f_j}{\omega_j^2 - i\omega\gamma_j - \omega^2}\right) \tag{76}$$

or

$$\epsilon(\omega) = 1 + \frac{4\pi n z e^2}{m}\sum_j \left(\frac{f_j}{z}\right)\left(\frac{\omega_j^2 - \omega^2 + i\omega\gamma_j}{(\omega_j^2 - \omega^2)^2 + \omega^2\gamma_j^2}\right) \equiv \epsilon_1 + i\epsilon_2 \tag{77}$$

where ϵ_1 and ϵ_2 are real.

In a typical term of the sum, different regimes of the relative sizes of ω, ω_j, and γ_j give rise to very different behaviors. The resonant frequencies are, when Planck's constant is thrown in, comparable to binding energies of electrons which are on the order of a few electron-volts, so that ω_j is of order $10^{15}\, sec^{-1}$, much the same as optical frequencies. The damping constants tend to be somewhat smaller, perhaps of order $10^{12}\, sec{-1}$ (see below). Starting from low frequencies, $\omega << \omega_j^2$ and also $\omega\gamma_j << \omega_j^2$, then we can approximate the dielectric function as

$$\epsilon(\omega) \approx 1 + \frac{4\pi n e^2}{m}\sum \frac{f_j}{\omega_j^2} \tag{78}$$

which is a constant. Now, as ω increases from a low value, the real part of ϵ will also increase (slowly at first); when it gets to within about γ_j of the smallest ω_j, there is a resonance (the electron is being "pushed" by the electric field at a frequency close to its natural frequency) which will show up in ϵ_1 as a sudden rise, fall, and rise. After this, ϵ_1 is again roughly constant. There are as many such resonances as there are distinct resonant frequencies or terms in the sum over j.

The rapid variation of the dielectric function in the vicinity of a resonance also produces a rapidly varying index of refraction, meaning that waves with relatively

[9] In this we follow Jackson, but remember the Clausius-Mossotti relation from last quarter; we argued that the electric field which produces the polarization should be the local field and not the macroscopic field. It is not difficult to make the necessary corrections to what is given here.

Figure 1: Real and imaginary parts of ϵ near resonances

close frequencies propagate with quite different speeds. The frequency regime where ϵ_1 decreases with increasing ω is known as a region of *anomalous dispersion.*

The imaginary part of ϵ also behaves in an interesting fashion near a resonance. Because the denominator of the resonant term in $\epsilon(\omega)$ gets quite small at $\omega = \omega_j$ while the numerator for the imaginary part does not get small, there is a pronounced peak in ϵ_2 here. The smaller the value of γ_j, the bigger the peak. A large imaginary part of the dielectric function produces strong damping or *absorption* of the wave, so a region of anomalous dispersion is also a region of strong absorption, termed *resonant absorption.*

Finally, for ω very large in comparison with any other frequency in the system $\omega \gg \omega_j$, the dielectric function once again becomes simple and has the form

$$\epsilon(\omega) = 1 - \frac{4\pi n z e^2}{m\omega^2} \equiv 1 - \frac{\omega_p^2}{\omega^2} \tag{79}$$

where we have introduced the *plasma frequency* of the electron system,

$$\omega_p \equiv \sqrt{\frac{4\pi n z e^2}{m}}. \tag{80}$$

For typical values of n in solids, this frequency is of order $10^{16}\,sec^{-1}$ which is as large as or larger than the frequency of visible light. Our result is interesting in that the dielectric function is smaller than unity in this regime of frequency, meaning that a point of constant phase in a harmonic wave actually travels faster than the speed of light c. Even more remarkable is the possibility that $\epsilon(\omega) < 0$ in some range of frequency. For this to occur it is necessary to have $\omega < \omega_p$ but at the same time ω must be considerably larger than any resonant frequency ω_j and also larger than the damping parameters γ_j. Such conditions can be attained in some materials; a simple example is a tenuous plasma, or gas of charged particles. Then the resonant frequencies are all zero, the plasma frequency is rather low because the density of charges is not large, and the damping is small. See the following section.

6.1 Dielectric Response of Free Electrons

Some special cases are also worthy of mention. One is the case of free electrons. For these electrons there is no restoring force and so we may set the corresponding ω_j, called ω_0, to zero. This has a profound effect on the dielectric function at low frequencies. If we extract the free-electron term from the remainder of the dielectric function and regard the latter as some constant ϵ_0 at low frequencies (see Eq. (78)), then we have

$$\epsilon = \epsilon_0 - \frac{4\pi n f_0 e^2}{m\omega(\omega + i\gamma_0)} = \epsilon_0 + i\frac{4\pi n f_0 e^2}{m\omega(\gamma_0 - i\omega)}. \tag{81}$$

This thing is singular as $\omega \to 0$, reflecting the fact that in the zero-frequency limit, the free electrons will be displaced arbitrarily far from their initial positions by any small electric field, producing a very large polarization. The singular term in ϵ in fact represents the conductivity of the free-electron material. To see how it is related to the conductivity, let us examine Ampère's law using this dielectric function and **no** macroscopic current $\mathbf{J}$, as this current will be included in the dielectric response (the polarization produced by the free electrons). From $\nabla \times \mathbf{H} = c^{-1}\partial \mathbf{D}/t$, we find

$$\nabla \times \mathbf{H} = -i\frac{\omega}{c}\left(\epsilon_0 + i\frac{4\pi n f_0 e^2}{m\omega(\gamma_0 - i\omega)}\right)\mathbf{E} = \frac{4\pi}{c}\frac{n f_0 e^2}{m(\gamma_0 - i\omega)}\mathbf{E} - i\frac{\omega}{c}\epsilon_0\mathbf{E}. \qquad (82)$$

By contrast, we may choose **not** to include the free electrons' contribution to the polarization in which case $\epsilon = \epsilon_0$. Then, however, we have to include them as macroscopic current $\mathbf{J}$; assuming linear response and isotropy, we may write $\mathbf{J} = \sigma\mathbf{E}$ where σ is the *electrical conductivity*. Using these relations, and Ampère's law, $\nabla \times \mathbf{H} = (4\pi/c)\mathbf{J} + c^{-1}\partial\mathbf{D}/\partial t$, we find

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\sigma\mathbf{E} - i\frac{\omega}{c}\epsilon_0\mathbf{E}. \qquad (83)$$

Comparison of the two preceding equations shows that by including the contribution of the free electrons in the polarization we have actually derived a simple expression for the conductivity,

$$\sigma = \frac{n f_0 e^2}{m(\gamma_0 - i\omega)} \to \frac{n f_0 e^2}{m\gamma_0}, \qquad (84)$$

the last expression holding in the zero-frequency or static limit.

Comparison of measured conductivities with this result gives one an estimate of the damping constant. In very good metallic conductors such as Cu or Ag, $\sigma \sim 10^{17}\, sec^{-1}$. The free-electron density is of order $10^{22}\, cm^{-3}$ and so one is led to $\gamma_0 \sim 10^{13}\, sec^{-1}$ which is considerably smaller than typical resonant frequencies (for bound electrons, of course).

7 A Model for the Ionosphere

The ionosphere is a region of the upper atmosphere which is ionized by solar radiation (ultraviolet, x-ray, etc.). It may be simply described as a dilute gas of charged particles, composed of electrons and protons or other heavy charged objects. The dielectric properties of this medium are mainly produced by the lighter electrons, so we shall include only them in our description. We then have just one kind of charge and it has zero resonant frequency. Because the medium is dilute, the damping is small; we shall ignore it. This is the approximation of a *collisionless* plasma and it leaves us with a very simple dielectric function,

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \tag{85}$$

For frequencies smaller than the plasma frequency, $\epsilon(\omega) < 0$, meaning that the wave number is pure imaginary since $k = \omega\sqrt{\epsilon}/c$; the corresponding wave will not propagate because its dependence on position is proportional to $\exp(i\mathbf{k}\cdot\mathbf{x})$ or $\exp(-|\mathbf{k}|z)$ given that $\mathbf{k} \parallel \hat{\mathbf{z}}$[10].

In the case of the ionosphere there is an additional complicating factor (which also makes the problem more interesting); the earth has a magnetic field which influences the motions of the electrons and hence the dielectric function. The equation of motion of the charges, including this field $\mathbf{B}_0$ is

$$m\frac{d^2\mathbf{x}}{dt^2} = -e\left[\mathbf{E} + \frac{1}{c}\left(\frac{d\mathbf{x}}{dt}\times\mathbf{B}_0\right)\right]. \tag{86}$$

We ignore the effect of the wave's magnetic induction. We shall also restrict (for simplicity) attention to the case of $\mathbf{k} \parallel \mathbf{B}_0$ and shall ignore the spatial variations of $\mathbf{E}$. In addition, and without loss of generality, we can let the wave have circular polarization. Hence we write the electric field as $\mathbf{E} = E_0\boldsymbol{\epsilon}_\pm e^{-i\omega t}$.

Under these conditions, $\mathbf{x}$ will be of the form $\mathbf{x} = \mathbf{x}_0 e^{-i\omega t}$; using this relation in the equation of motion, we find

$$-m\omega^2\mathbf{x}_0 = -e\left[E_0\boldsymbol{\epsilon}_\pm - \frac{i\omega}{c}B_0(\mathbf{x}_0\times\hat{\mathbf{z}})\right]. \tag{87}$$

The solution of this equation is $\mathbf{x}_0 = x_0\boldsymbol{\epsilon}_\pm$; one can see this easily if one realizes that $\boldsymbol{\epsilon}_\pm \times \hat{\mathbf{z}} = \pm i\boldsymbol{\epsilon}_\pm$:

$$\boldsymbol{\epsilon}_\pm \times \hat{\mathbf{z}} = \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}})\times\hat{\mathbf{z}} = \frac{1}{\sqrt{2}}(-\hat{\mathbf{y}} \pm i\hat{\mathbf{x}}) = \pm\frac{i}{\sqrt{2}}(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) = \pm i\boldsymbol{\epsilon}_\pm. \tag{88}$$

[10] For most laboratory plasmas, this occurs at microwave frequencies

Hence the equation of motion, using $\mathbf{x}_0 = x_0\boldsymbol{\epsilon}_\pm$, is

$$-m\omega^2 x_0\boldsymbol{\epsilon}_\pm = -e[E_0 \pm \frac{\omega B_0 x_0}{c}]\boldsymbol{\epsilon}_\pm. \tag{89}$$

or

$$x_0 = \frac{eE_0/m}{\omega(\omega \mp \omega_B)} \tag{90}$$

where $\omega_B \equiv eB_0/mc$ is the *cyclotron frequency.* From this point we may determine the dielectric function by repeating the arguments used in the preceding section and find

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega \mp \omega_B)}. \tag{91}$$

Our result tells us that waves with different polarization elicit different dielectric responses from the medium; such a phenomenon is known as *birefringence.* If a wave of general polarization is incident upon the plasma, it is in effect broken into its two circularly polarized components and these propagate independently. It is possible to have a wave with a frequency such that for one component $\epsilon(\omega) < 0$ and for the other, $\epsilon(\omega) > 0$. Hence, one will propagate and the other will not, providing a (not particularly practical) way of producing a circularly polarized wave.

In the specific case of the ionosphere, ω_p, ω_B, and ω can all be quite comparable. The density of electrons, which varies with the time of day and solar activity, is typically $\sim 10^5 - 10^6\, cm^{-3}$, leading to $\omega_p \sim 10^7\, sec^{-1}$. The earth's field $B_0 \sim 0.1 - 1.0\, gauss$, leading to $\omega_B \sim 10^7\, sec^{-1}$. A wave with $\omega \sim 10^7\, sec^{-1}$ is in the AM band; short-wave radio frequencies are somewhat higher, and FM radio or television have considerably higher frequencies. This means that FM and television signals are at frequencies so large that $\epsilon \approx 1$ and they propagate right through the ionosphere without significant reflection or attenuation. For this reason, the signals can be received only at locations where there is a direct path through the atmosphere from transmitter to receiver. For the lower frequency signals (short-wave and AM), however, there can be strong reflection from the ionosphere, making it possible to receive them relatively far from the transmitter. The higher the point in the ionosphere where the reflection takes place, the greater the effective range of the signal.

Figure 2: Dielectric constants vs. ω for the ionosphere

Figure 3: Electron density vs. height in the ionosphere

Because the electron density increases with height (and then decreases again), the higher frequencies tend to be reflected at greater heights (if they are reflected at all) than the lower ones, thereby giving greater range. That is why short-wave signals have longer range than AM signals, at least some of the time.

What happens if $\mathbf{k}$ is not parallel to $\mathbf{B}_0$? The medium is still birefringent so that a wave of arbitrary polarization is broken into two components that propagate independently; however, the two components are not simple circularly polarized waves. In addition, the dielectric functions and hence the indices of refraction for these two waves depend on the angle between $\mathbf{B}_0$ and $\mathbf{k}$, so the medium is not only birefringent but also anisotropic.

8 Waves in a Dissipative Medium

We have seen in the preceding sections that the dielectric function will is general be complex, reflecting the fact that a wave will be dissipated or damped under many conditions. It therefore behooves us to learn more about the properties of waves when dissipation is present. As we have seen, we can do this by employing a complex dielectric function, and we can also do it, with the same basic results, by letting ϵ be real while introducing a real conductivity and thus a macroscopic current density.

We shall do the latter, for no particular reason.

Suppose that once again we have some linear medium with $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, and $\mathbf{J} = \sigma\mathbf{E}$; ϵ, μ, and σ are taken as real. Then the Maxwell equations become

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \cdot \mathbf{E} = 0, \qquad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \tag{92}$$

and

$$\nabla \times \mathbf{B} = \frac{4\pi\mu}{c}\sigma\mathbf{E} + \frac{\epsilon\mu}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{93}$$

We have set ρ equal to zero in these equations. It may be that there is initially some macroscopic charge density within a conductor. If this is the case, that density will decay to zero with a characteristic time on the order of γ^{-1} where γ is the damping constant introduced in the section on dielectric functions; see Jackson, Problem 7.7.

Let us look for plane wave solutions to the field equations. Set $\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$ and $\mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$. The divergence equations then tell us that $\mathbf{B}_0 \cdot \mathbf{k} = 0$ and $\mathbf{E}_0 \cdot \mathbf{k} = 0$ as in a nondissipative medium. From Faraday's law we find the familiar result

$$\mathbf{B}_0 = \frac{c}{\omega}(\mathbf{k} \times \mathbf{E}_0), \tag{94}$$

and from the Ampère's law we find

$$i(\mathbf{k} \times \mathbf{B}_0) = \frac{4\pi\mu\sigma}{c}\mathbf{E}_0 - i\frac{\omega\mu\epsilon}{c}\mathbf{E}_0. \tag{95}$$

If we take the cross product of $\mathbf{k}$ with Eq. (94) and substitute Eq. (95) into the result where $\mathbf{k} \times \mathbf{B}_0$ appears, we find, after using $\mathbf{k} \times (\mathbf{k} \times \mathbf{E}_0) = -k^2\mathbf{E}_0$, that

$$-i\frac{4\pi\mu\sigma}{c}\mathbf{E}_0 - \frac{\omega\mu\epsilon}{c}\mathbf{E}_0 = -\frac{ck^2}{\omega}\mathbf{E}_0 \tag{96}$$

or

$$k^2 = \frac{\omega^2\mu\epsilon}{c^2} + i\frac{4\pi\mu\sigma\omega}{c^2}. \tag{97}$$

Taking the point of view that ω is some given real frequency, we can solve this relation for the corresponding wavenumber k, which is complex. If we write $k = k_0 + i\alpha$, then the real and imaginary parts of Eq. (97) give us two equations which may be solved for k_0 and α:

$$k_0^2 - \alpha^2 = \frac{\omega^2\epsilon\mu}{c^2} \qquad 2k_0\alpha = \frac{\omega^2\epsilon\mu}{c^2}\left(\frac{4\pi\sigma}{\epsilon\omega}\right). \tag{98}$$

The solution is

$$\left\{\begin{array}{c} k_0 \\ \alpha \end{array}\right\} = \sqrt{\mu\epsilon}\left(\frac{\omega}{c}\right)\left\{\frac{\sqrt{1+\left(\frac{4\pi\sigma}{\omega\epsilon}\right)^2}\pm 1}{2}\right\}^{1/2}. \tag{99}$$

where the + sign refers to k_0 and the - sign to α.

This expression appears somewhat impenetrable although it doesn't say anything unexpected or remarkable. It takes on simple forms in the limits of high and low conductivity. The relevant dimensionless parameter is $4\pi\sigma/\epsilon\omega$. It if is much larger than unity, corresponding to a **good conductor**, then

$$k_0 \approx \alpha \approx \frac{\sqrt{2\pi\omega\mu\sigma}}{c} \equiv \frac{1}{\delta} \quad \frac{4\pi\sigma}{\epsilon\omega} \gg 1\,. \tag{100}$$

where we have introduced the *penetration depth* δ. This is the distance that an electromagnetic wave will penetrate into a good conductor before being attenuated to a fraction $1/e$ of its initial amplitude. Since the wavelength of the wave is $\lambda = 2\pi/k_0$, δ is also a measure of the wavelength in the conductor.

For a **poor conductor**, by which we mean $4\pi\sigma/\omega\epsilon << 1$, one has

$$k_0 + i\alpha \approx \sqrt{\mu\epsilon}\frac{\omega}{c} + i\frac{2\pi}{c}\sqrt{\frac{\mu}{\epsilon}}\sigma. \tag{101}$$

Notice that in the latter case, the real part of the wavenumber is the same as in a nonconducting medium and the imaginary part is independent of frequency so that waves of all frequencies are attenuated by equal amounts over a given distance. Also, $\alpha << k_0$ which tells us that the wave travels many wavelengths before being attenuated significantly.

For a given σ, α is an increasing function of ω and saturates at high frequencies. Therefore, if one wants a wave to travel as far as possible, one wants to use as low freqency a wave as possible. Then one should be in the good-conductor limit where the attenuation varies as $\sqrt{\omega}$ and vanishes as $\omega \to 0$. A well-known application of this rule has to do with radio communication with submarines; sea water is a reasonably good conductor $\sigma \sim 10^{11}\, sec^{-1}$ and so to communicate with a submerged boat, one should send out low frequency signals which will penetrate to greater depths in the ocean than more standard signals.

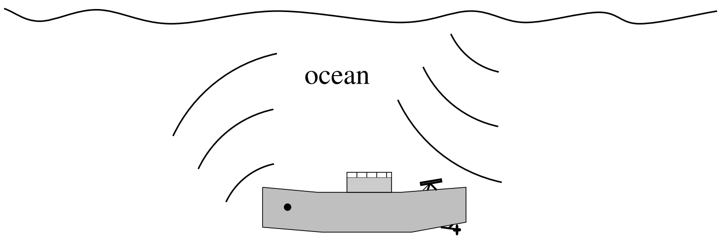


Figure 22: Low frequency waves can propagate through sea water.

Given that we have found the complex wave number, and letting $\mathbf{k}$ point in the z-direction, we have

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(k_0 z - \omega t)} e^{-\alpha z}; \tag{102}$$

the corresponding magnetic induction is found in the usual way (take $\mathbf{k} \times \mathbf{E}$):

$$\mathbf{B}(\mathbf{x}, t) = \frac{c}{\omega}(k_0 + i\alpha)(\hat{\mathbf{z}} \times \mathbf{E}_0) e^{i(k_0 z - \omega t)} e^{-\alpha z}. \tag{103}$$

Define the complex index of refraction

$$n \equiv \frac{c}{\omega} k = \frac{c}{\omega}(k_0 + i\alpha), \tag{104}$$

so that

$$\mathbf{B} = n(\hat{\mathbf{z}} \times \mathbf{E}). \tag{105}$$

Notice that because n is complex, $\mathbf{B}$ *is not in phase with* $\mathbf{E}$; to make the phase difference explicit, let us write n in polar form:

$$n = |n| e^{i\phi} \qquad \text{where} \qquad \phi = \arctan\left(\frac{\alpha}{k_0}\right). \tag{106}$$

We can find $|n|$ and ϕ in terms of other parameters; let $\gamma \equiv (4\pi\sigma/\omega\epsilon)^2$. Then

$$\phi = \arctan\left[\frac{\sqrt{1+\gamma}-1}{\sqrt{1+\gamma}+1}\right]^{1/2}. \tag{107}$$

Consider $\tan(2\phi)$:

$$\begin{aligned}\tan(2\phi) &= \frac{2\tan\phi}{1-\tan^2\phi} = 2\frac{[(\sqrt{1+\gamma}-1)/(\sqrt{1+\gamma}+1)]^{1/2}}{1-\frac{\sqrt{1+\gamma}-1}{\sqrt{1+\gamma}+1}} \\ &= [(\sqrt{1+\gamma}-1)(\sqrt{1+\gamma}+1)]^{1/2} = \sqrt{\gamma}\end{aligned} \tag{108}$$

Thus,

$$\phi = \frac{1}{2}\arctan\sqrt{\gamma} = \frac{1}{2}\arctan\left(\frac{4\pi\sigma}{\omega\epsilon}\right). \tag{109}$$

Also,

$$|n| = \frac{c}{\omega}\sqrt{k_0^2+\alpha^2} = \sqrt{\mu\epsilon}\left[1+\left(\frac{4\pi\sigma}{\omega\epsilon}\right)^2\right]^{1/4}. \tag{110}$$

Using these results in Eq. (105), we have

$$\mathbf{B}(\mathbf{x},t) = \sqrt{\mu\epsilon}\left[1+\left(\frac{4\pi\sigma}{\omega\epsilon}\right)^2\right]^{1/4} e^{\frac{i}{2}\arctan\left(\frac{4\pi\sigma}{\omega\epsilon}\right)}(\hat{\mathbf{z}}\times\mathbf{E}_0). \tag{111}$$

The amount by which $\mathbf{B}(\mathbf{x},t)$ is phase-shifted from $\mathbf{E}(\mathbf{x},t)$ is easily seen from this expression to lie between 0 and $\pi/4$; it is zero in the small σ/ω limit and $\pi/4$ in the large σ/ω limit. Another significant feature of the expression for $\mathbf{B}(\mathbf{x},t)$ is that in the small σ/ω limit, the amplitude of $\mathbf{B}$ relative to that of $\mathbf{E}$ is just $\sqrt{\mu\epsilon}$ as for insulators. But in the opposite limit, one finds that the relative amplitude is $\sqrt{4\pi\sigma\mu/\omega}$ which is much larger than unity. Here the wave has, relatively speaking, a much larger magnetic induction than electric field.

8.1 Reflection of a Wave Normally Incident on a Conductor

As an example, let us calculate the reflection of a wave normally incident on a conductor from vacuum.

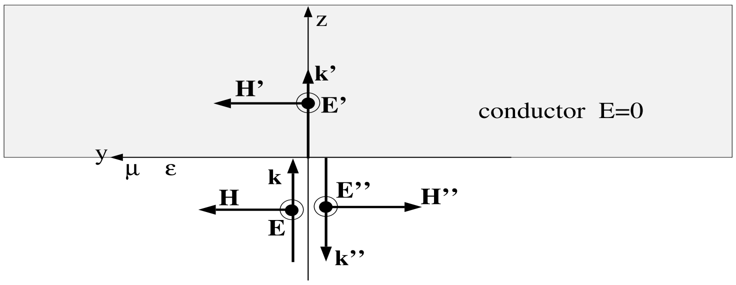


Figure 23: Wave normally incident on a conductor.

Then

$$\mathbf{k} = \frac{\omega}{c}\hat{\mathbf{z}} \qquad \mathbf{k}' = \frac{\omega}{c}n\hat{\mathbf{z}}, n = \sqrt{\mu\epsilon}(1+\gamma)^{1/4}e^{i\phi}. \tag{112}$$

The relevant boundary conditions are $\mathbf{H}_t$ and $\mathbf{E}_t$ continuous. Let $\mathbf{E}_0 = E_0\hat{\mathbf{x}}$, $\mathbf{E}_0'' = E_0''\hat{\mathbf{x}}$, and $\mathbf{E}_0' = E_0'\hat{\mathbf{x}}$. The corresponding magnetic field amplitudes are $\mathbf{H}_0 = E_0\hat{\mathbf{y}}$, $\mathbf{H}_0'' = -\mathbf{E}_0''\hat{\mathbf{y}}$, and, for the transmitted wave in the conductor,

$$\mathbf{H}_0' = \sqrt{\frac{\epsilon}{\mu}}(1+\gamma)^{1/4}e^{i\phi}E_0'\hat{\mathbf{y}}. \tag{113}$$

Our boundary conditions give immediately

$$E_0 + E_0'' = E_0' \qquad E_0 - E_0'' = \sqrt{\frac{\epsilon}{\mu}}(1+\gamma)^{1/4}e^{i\phi}E_0'. \tag{114}$$

These may be combined to yield

$$E_0' = \frac{2}{1+\sqrt{\epsilon\mu}(1+\gamma)^{1/4}e^{i\phi}}E_0 \tag{115}$$

and

$$E_0'' = \frac{1-\sqrt{\epsilon/\mu}(1+\gamma)^{1/4}e^{i\phi}}{1+\sqrt{\epsilon/\mu}(1+\gamma)^{1/4}e^{i\phi}}E_0. \tag{116}$$

Let us calculate the Poynting vector in the conductor. Its time average is

$$< \mathbf{S}' >= \frac{c}{8\pi}\Re(\mathbf{E}' \times \mathbf{H}'^*) = \frac{c}{8\pi}\Re\left\{\frac{4|E_0|^2\sqrt{\epsilon/\mu}(1+\gamma)^{1/4}e^{-i\phi}}{|1+\sqrt{\epsilon/\mu}(1+\gamma)^{1/4}e^{i\phi}|^2}\right\}e^{-2\alpha z}\hat{\mathbf{z}}. \tag{117}$$

Using the interpretation of this vector as the energy current density, we may find the power per unit area transmitted into the conductor by evaluating $<\mathbf{S}'>\cdot\hat{\mathbf{z}}$ at $z=0$,

$$\mathcal{P}' = \frac{c}{2\pi}|E_0|^2\sqrt{\frac{\epsilon}{\mu}}\,\frac{(1+\gamma)^{1/4}\cos\phi}{1+2\sqrt{\epsilon/\mu}\cos\phi\,(1+\gamma)^{1/4}+(\epsilon/\mu)(1+\gamma)^{1/2}}. \tag{118}$$

The incident power per unit area is $\mathcal{P} = c|E_0|^2/8\pi$, so the fraction of the incident power which enters the conductor, where it is dissipated as Joule heat, is

$$T = \frac{\mathcal{P}'}{\mathcal{P}} = 4\sqrt{\frac{\epsilon}{\mu}}\,\frac{(1+\gamma)^{1/4}\cos\phi}{1+2\sqrt{\epsilon/\mu}\cos\phi\,(1+\gamma)^{1/4}+(\epsilon/\mu)(1+\gamma)^{1/2}}. \tag{119}$$

This expression is much simplified in the limit of a good conductor where $\phi=\pi/4$, $\cos\phi = 1/\sqrt{2}$, and $\gamma >> 1$. Then

$$T \approx 4\sqrt{\frac{\epsilon}{\mu}}\,\frac{\gamma^{1/4}(1/\sqrt{2})}{\epsilon\gamma^{1/2}/\mu} = 2\sqrt{2}\sqrt{\frac{\mu}{\epsilon}}\sqrt{\frac{\omega\epsilon}{4\pi\sigma}} = \frac{2\mu\omega}{c}\frac{c}{\sqrt{2\pi\sigma\omega\mu}} = \frac{2\mu\omega}{c}\delta. \tag{120}$$

For a good conductor such as Cu, $\sigma \sim 10^{17}\,sec^{-1}$ and so a wave with frequency around $10^{10}\,sec^{-1}$ will have $\delta \sim 10^{-4}\,cm$ or $1\,\mu m$. Also, the better the conductor, the smaller the fraction of the incident power lost in the reflection process. For the example just given, $T\approx 10^{-4}$, meaning that the wave can be reflected some ten thousand times before becoming strongly attenuated.

9 Superposition of Waves; Pulses and Packets

No wave is truly monochromatic, although some waves, such as those produced by lasers, are exceedingly close to being so. Fortunately, in the case of linear media, the equations of motion for electromagnetic waves are completely linear and so any sum of harmonic solutions is also a solution. By making use of this superposition "principle" we can construct quite general solutions by superposing solutions of the kind we have already studied.

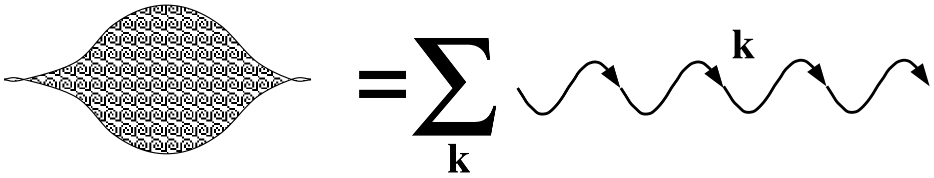


Figure 24: Any pulse in a linear media may be decomposed into a superposition of plane waves.

This procedure amounts to making a Fourier transform of the pulse. For simplicity we shall work in one spatial dimension which simply means that we will superpose waves whose wave vectors are all in the same direction (the z-direction). For the same reason, we shall also employ scalar waves; these could, for example, be the x components of the electric fields of the waves. One such wave has the form $e^{i(kz-\omega(k)t)}$ where we shall not initially restrict $\omega(k)$ to any particular form. Given a set of such waves, we can build a general solution of this kind (wave vector parallel to the z-axis) by integrating over some distribution $A(k)$ of them:

$$u(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, A(k) e^{i(kz-\omega(k)t)}. \tag{121}$$

At time $t = 0$, this function is simply

$$u(z,0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, A(k) e^{ikz} \tag{122}$$

and the inverse transform gives A in terms of the zero-time wave:

$$A(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz\, u(z,0) e^{-ikz}. \tag{123}$$

All of the standard rules of Fourier transforms are applicable to the functions $A(k)$ and $u(z,0)$. For example, if $A(k)$ is a sharply peaked function with width Δk, then the width of $u(z,0)$ must be of order $1/\Delta k$ or larger, and conversely. One may make this statement more precise by defining

$$(\Delta z)^2 \equiv < z^2 > - < z >^2 \qquad (\Delta k)^2 \equiv < k^2 > - < k >^2, \tag{124}$$

where

$$< f(k) > \equiv \frac{\int_{-\infty}^{\infty} dk\, f(k)|A(k)|^2}{\int_{-\infty}^{\infty} dk\, |A(k)|^2} \tag{125}$$

and

$$< f(z) > \equiv \frac{\int_{-\infty}^{\infty} dz\, f(z)|u(z,0)|^2}{\int_{-\infty}^{\infty} dz\, |u(z,0)|^2}. \tag{126}$$

The relation between these widths which must be obeyed is

$$\Delta z \Delta k \geq 1/2. \tag{127}$$

Now, given a "reasonable" initial wave form $u(z,0)$[11] with some Δz and a Fourier transform $A(k)$ with some Δk, the question we seek to answer is what will be the nature of $u(z,t)$? The answer is simple in principle because all we have to do is Fourier transform to find $A(k)$ and then do the integral specified by Eq. (121) to find $u(z,t)$. One can always do these integrals numerically if all else fails. Here we shall do some approximate calculations designed to demonstrate a few general points.

Suppose that we have found $A(k)$ and that it is some peaked function centered at k_0 with a width Δk. If $\omega(k)$ is reasonably well approximated by a truncated Taylor's series expansion for k within Δk of k_0, then we may write

$$\omega(k) \approx \omega_0 + \left.\frac{d\omega}{dk}\right|_{k_0} (k - k_0) \equiv \omega_0 + v_g(k - k_0) \tag{128}$$

where

$$\omega_0 \equiv \omega(k_0) \qquad \text{and} \qquad v_g = d\omega/dk|_{k_0}; \tag{129}$$

v_g is called the *group velocity* of the packet; notice that it can depend on the wave number k_0 which characterizes the typical wave numbers in the wave. In this approximation, one finds

$$u(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, A(k) e^{ik(z-v_g t)} e^{-i\omega_0 t} e^{i v_g k_0 t} = e^{i(v_g k_0 - \omega_0)t} u(z - v_g t, 0). \tag{130}$$

[11]Its time derivative $\partial u(z,t)/\partial t|_{t=0}$ must also be given to allow a unique solution of the initial value problem; our discussion is therefore incomplete but can be corrected easily.

This result tells us that the wave packet retains its initial form and translates in space at a speed v_g. It does not spread (disperse) or distort in any way. In particular, the energy carried by the wave will move with a speed v_g.

The group velocity is evidently an important quantity. We may write it in terms of the index of refraction by using the defining relation $k = \omega n(\omega)/c$. Take the derivative of this with respect to k:

$$1 = \left(\frac{n}{c} + \frac{\omega}{c}\frac{dn}{d\omega}\right)\frac{d\omega}{dk} \tag{131}$$

or

$$v_g = \frac{c}{n + \omega\frac{dn}{d\omega}}. \tag{132}$$

As an example consider the collisionless plasma relation $n = \sqrt{1 - \omega_p^2/\omega^2}$. One easily finds that

$$v_g = c\sqrt{1 - \omega_p^2/\omega^2}. \tag{133}$$

For $\omega < \omega_p$, the group velocity is imaginary which corresponds to a damped wave; for $\omega > \omega_p$, it is positive and increases from zero to c as ω increases.

Our calculations thus far have not resulted in any spreading or distortion of the wave packet because we did not include higher-order terms in the relation (called a *dispersion relation*) between ω and k. Let's treat a simple example in which $A(k)$ is a gaussian function of $k - k_0$,

$$A(k) = \left(\frac{A_0}{\delta}\right) e^{-(k-k_0)^2/2\delta^2}. \tag{134}$$

Further, let $\omega(k)$ be approximated by

$$\omega(k) = \omega_0 + v_g(k - k_0) + \alpha(k - k_0)^2. \tag{135}$$

The corresponding $u(z,t)$ is

$$u(z,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, \frac{A_0}{\delta} e^{-(k-k_0)^2/2\delta^2} e^{ikz - i[\omega_0 + v_g(k-k_0) + \alpha(k-k_0)^2]t}$$

$$= \frac{A_0}{\delta\sqrt{2\pi}} e^{i(k_0 z - \omega_0 t)} \int_{-\infty}^{\infty} dk\, e^{i(k-k_0)(z-v_g t)} e^{-(1/2\delta^2 + i\alpha t)(k-k_0)^2}$$

$$= \frac{A_0}{\sqrt{1+2i\alpha\delta^2 t}} e^{-(z-v_g t)^2\delta^2/[2(1+2i\alpha\delta^2 t)]} e^{i(k_0 z - \omega_0 t)}. \tag{136}$$

If $\alpha = 0$, this is a Gaussian-shaped packet which travels at speed v_g with a constant width equal to δ^{-1}. If $\alpha \neq 0$, it is still a Gaussian-shaped packet traveling at speed v_g; however, it does not have a constant width any longer. To make the development of the width completely clear, consider $|u(z,t)|^2$ which more nearly represents the energy density in the wave:

$$|u(z,t)|^2 = \frac{A_0^2}{\sqrt{1+4\alpha^2\delta^4 t^2}} e^{-(z-v_g t)^2\delta^2/(1+4\alpha^2\delta^4 t^2)}. \tag{137}$$

The width of this traveling Gaussian is easily seen to be

$$w(t) = \sqrt{1+4\alpha^2\delta^4 t^2}/\delta. \tag{138}$$

At short times the width increases as the square of the time, while at long times it becomes linear with t.

When the packet spreads, or disperses, in this fashion, to what extent does it make sense to think about the wave as a localized object? One measure is the width of the packet as compared with the distance it has moved. After a long time the width is approximately $2\alpha\delta t$ while the distance the packet has moved is $v_g t$. The ratio of these distances is $2\alpha\delta/v_g$, so our condition for having a localized object is

$$2\alpha\delta/v_g << 1. \tag{139}$$

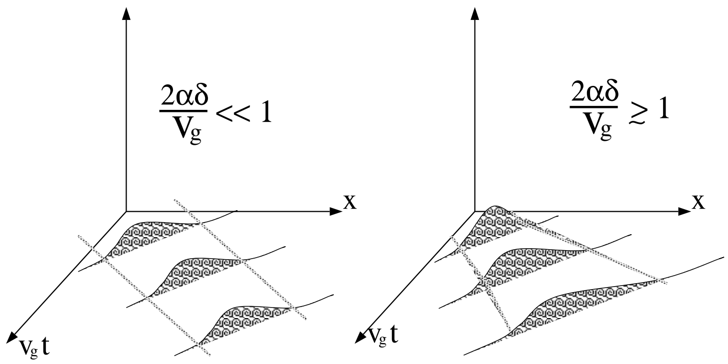


Figure 25: When δ is small, the wave is composed of few wavenumbers.

In addition, of course, the initial width of the packet must be small compared to $v_g t$ which is always possible if one waits long enough. Our inequality clearly puts a limit on the allowable size of α, for a given δ, necessary to have a well-defined pulse. For smaller δ, one can get away with larger α, a simple consequence of the fact that small δ means the width of the packet in k-space is small, leading to less dispersion.

9.1 A Pulse in the Ionosphere

Let's look also at the fate of a wave packet propagating in the ionosphere; we found in an earlier section, treating the ionosphere as a collisionless plasma and with $\mathbf{k}$ parallel to $\mathbf{B}_0$, that $\epsilon(\omega) = 1 + \omega_p^2/\omega(\omega_B - \omega)$ for one particular polarization of the wave. If ω is small enough compared to other frequencies, we may approximate in such a way that $n(\omega) = \omega_p/\sqrt{\omega\omega_B}$, which gives rise to anomalous dispersion indeed. Defining $\omega_0 \equiv \omega_p^2/\omega_B$, one finds that the group velocity of a signal is $v_g = 2c\sqrt{\omega/\omega_0}$.

Let us see how a pulse with the same $A(k)$ as in the previous example propagates. We have

$$u(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, \frac{A_0}{\delta} e^{-(k-k_0)^2/2\delta^2 + ikz - ic^2k^2t/\omega_0}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, \frac{A_0}{\delta} e^{-(k-k_0)^2/2\delta^2 + i(k-k_0)z + ik_0z - ic^2t(k-k_0)^2/\omega_0 - i2c^2k_0t(k-k_0)/\omega_0 - ic^2k_0^2t/\omega_0}$$

$$= \frac{A_0}{1 + 2i\delta^2c^2t/\omega_0)^{1/2}} e^{i(k_0z - c^2k_0^2t/\omega_0)} e^{-\frac{(z-2c^2k_0t/\omega_0)^2\delta^2}{2(1+2i\delta^2c^2t/\omega_0)}} \tag{140}$$

This is a traveling, dispersing Gaussian. Its speed is the group velocity $v_g(k_0)$. The width of the Gaussian is

$$w(t) = \sqrt{1 + 4\delta^4c^4t^2/\omega_0^2}/\delta \to 2\delta c^2 t/\omega_0 \tag{141}$$

at long times. The packet spreads at a rate given by $v_w = 2\delta c^2/\omega_0$. The ratio of this spreading rate to the group velocity is δ/k_0 and so we retain a well-defined pulse provided the spread in wavenumber is small compared to the central wavenumber.

Pulses of this general type are generated in the ionosphere by thunderstorms. They have a very broad range of frequencies ranging from very low ones up into at least the AM radio range. The electromagnetic waves tend to be guided along lines of the earth's magnetic induction, and so, if for example the storm is in the southern hemisphere, the waves travel north in the ionosphere along lines of **B** and then come back to earth in the northern hemisphere.

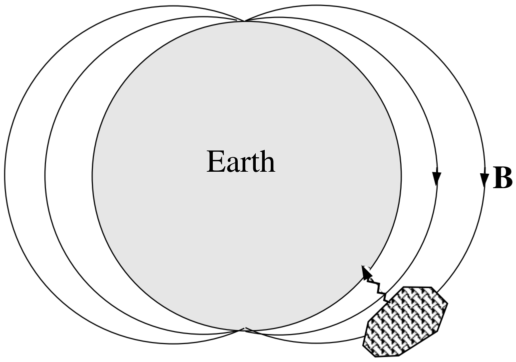


Figure 26: Lightning in the southern hemisphere yields wistlers in the north.

By this time they are much dispersed, with the higher frequency components arriving well before the lower frequency ones since $v_g = 2c\sqrt{\omega/\omega_0}$ for $\omega << \omega_0$. Frequencies in the audible range, $\omega \sim 10^2$ or $10^3\, sec^{-1}$ take one or more seconds (a long time for electromagnetic waves) to arrive. If one receives the signal and converts it directly to an audio signal at the same frequency, it sounds like a whistle, starting at high frequencies and continuing down to low ones over a time period of several seconds. This characteristic feature has caused such waves to be known as *whistlers*.

10 Causality and the Dielectric Function

A linear dispersive medium is characterized by a dielectric function $\epsilon(\omega)$ having physical origins that we have just finished exploring. One consequence of having such a relation between $\mathbf{D}(\mathbf{x},\omega)$ and $\mathbf{E}(\mathbf{x},\omega)$, that is,

$$\mathbf{D}(\mathbf{x},\omega) = \epsilon(\omega)\mathbf{E}(\mathbf{x},\omega), \tag{142}$$

is that the relation between $\mathbf{D}(\mathbf{x},t)$ and $\mathbf{E}(\mathbf{x},t)$ is **nonlocal** in time. To see this we have only to look at the Fourier transforms of $\mathbf{D}$ and $\mathbf{E}$. One has

$$\mathbf{D}(\mathbf{x},t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} d\omega\,\mathbf{D}(\mathbf{x},\omega)e^{-i\omega t} \tag{143}$$

and its inverse

$$\mathbf{D}(\mathbf{x},\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dt\,\mathbf{D}(\mathbf{x},t)e^{-i\omega t}; \tag{144}$$

similar relations hold for $\mathbf{E}(\mathbf{x},t)$ and $\mathbf{E}(\mathbf{x},\omega)$. Using the relation $\mathbf{D}(\mathbf{x},\omega) = \epsilon(\omega)\mathbf{E}(\mathbf{x},\omega)$, we have

$$\mathbf{D}(\mathbf{x},t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} d\omega\,\epsilon(\omega)\mathbf{E}(\mathbf{x},\omega)e^{-i\omega t}. \tag{145}$$

We can write $\mathbf{E}(\mathbf{x},\omega)$ here as a Fourier integral and so have

$$\begin{aligned}\mathbf{D}(\mathbf{x},t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\,\epsilon(\omega)e^{-i\omega t}\int_{-\infty}^{\infty} dt'\,e^{i\omega t'}\mathbf{E}(\mathbf{x},t')\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\,d\omega\,[\epsilon(\omega)-1+1]\mathbf{E}(\mathbf{x},t')e^{-i\omega(t-t')} =\\ \mathbf{E}(\mathbf{x},t) + \frac{1}{2\pi}&\int_{-\infty}^{\infty} dt\,d\omega\,[\epsilon(\omega)-1]\mathbf{E}(\mathbf{x},t')e^{-i\omega(t-t')} \equiv \mathbf{E}(\mathbf{x},t) + 4\pi\mathbf{P}(\mathbf{x},t).\end{aligned} \tag{146}$$

The final term, $4\pi\mathbf{P}(\mathbf{x},t)$, can be written in terms the Fourier transform[12] of $\epsilon(\omega)-1$; introduce the function

$$G(t) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\,[\epsilon(\omega)-1]e^{-i\omega t}. \tag{147}$$

Then we have

$$\mathbf{D}(\mathbf{x},t) = \mathbf{E}(\mathbf{x},t) + \int_{-\infty}^{\infty} dt'\,G(t-t')\mathbf{E}(\mathbf{x},t') \tag{148}$$

which may also be written as

$$\mathbf{D}(\mathbf{x},t) = \mathbf{E}(\mathbf{x},t) + \int_{-\infty}^{\infty} d\tau\,G(\tau)\mathbf{E}(\mathbf{x},t-\tau). \tag{149}$$

This equation makes it clear that when the medium has a frequency-dependent dielectric function, as all materials do, then the electric displacement at time t depends on

[12] Provided the order of integration can be reversed and the transform exists.

the electric field not only at time t but also at times other than t. This is somewhat disturbing because one can see that, depending on the character of G, we could get a polarization $\mathbf{P}(\mathbf{x}, t)$ that depends on values of $\mathbf{E}(\mathbf{x}, t')$ for $t' > t$, which means we get an effect arising from a cause that occurs at a time later than the effect. This behavior can be avoided if the function $G(\tau)$ vanishes when $\tau < 0$, and that is what in fact happens.

Let's look at a simple example with

$$\epsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\omega\gamma}. \tag{150}$$

Then

$$G(\tau) = \frac{\omega_p^2}{2\pi} \int_{-\infty}^{\infty} d\omega \, \frac{e^{-i\omega\tau}}{\omega_0^2 - \omega^2 - i\omega\gamma} \tag{151}$$

This integral was made for contour integration techniques. The poles of the integrand are in the lower half-plane in complex frequency space at

$$\omega_\pm = \frac{1}{2}[\pm\sqrt{4\omega_0^2 - \gamma^2} - i\gamma]; \tag{152}$$

without producing a contribution to the integral, we can close the contour in the upper (lower) half-plane when τ is smaller (larger) than zero. Because there are poles only in the lower half-plane, we can see immediately that $G(\tau)$ will be zero for $\tau < 0$. That is pleasing since we don't want the displacement (that is, the polarization) to respond at time t to the electric field at times later than t.

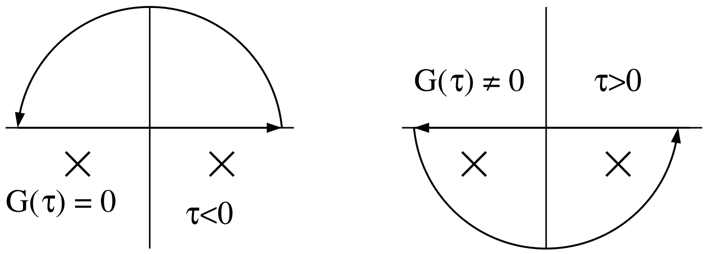


Figure 27: Because there are poles only in the lower half-plane, we can see immediately that $G(\tau)$ will be zero for $\tau < 0$.

Applying Cauchy's theorem to the case of $\tau > 0$, one finds that, for all τ,

$$G(\tau) = \omega_p^2 e^{-\gamma\tau/2} \frac{\sin(\nu_0 \tau)}{\nu_0} \theta(\tau) \tag{153}$$

where $\theta(x)$ is the step function, equal to unity for $x > 0$ and to zero otherwise, and $\nu_0 = \sqrt{\omega_0^2 - \gamma^2/4}$. The characteristic range in time of this function is γ^{-1} and hence the nonlocal (in time) character of the response is not important for frequencies smaller than about γ; it becomes important for larger ones.

One may naturally wonder whether there should also be nonlocal character of the response in space as well as in time. In fact there should and will be under some conditions. If we look back at our derivation of the model dielectric function, we see that the equation of motion of the particle was solved using $\mathbf{E}(0, t)$ instead of $\mathbf{E}(\mathbf{x}, t)$; the latter is of course the more correct choice. The difference is not important so long as the excursions of the charge from the point on which it is bound are much smaller than the wavelength of the radiation, which is the case for any kind of wave with frequencies up to those of soft X-rays. Hence the response can be expected to be local in space in insulating materials. However, if an electron is free, it can move quite far during a cycle of the field and if it does so, the response will be nonlocal in space as well as time.

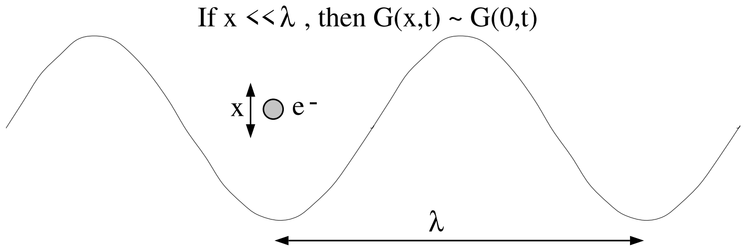


Figure 28: $G(\tau, \mathbf{x})$ will not be $\mathbf{x}$ dependent if the excursions of the charge from the point on which it is bound are much smaller than the wavelength of the radiation.

Returning to the question of causality, we have seen that the simple model dielectric function produces a function $G(t)$ which is zero for $t < 0$, as is necessary if "causality" is to be preserved, by which we mean there is no response in advance of the "cause" of that response. It is easy to see what are the features of the dielectric function that give rise to the result $G(t) = 0$ for $t < 0$. One is that there are no simple poles of the dielectric function in the upper half of the complex frequency plane. Another is that the dielectric function goes to zero for large ω fast enough that we can do the contour integral as we did it.

More generally, if one wants to have a function $G(t)$ which is consistent with the requirements of causality, this implies certain conditions on *any* $\epsilon(\omega)$. Additional conditions can be extracted from such simple things as the fact that $G(t)$ must be real so that $\mathbf{D}$ is real if $\mathbf{E}$ is. Without going into the details of the matter (see Jackson) let us make some general statements. The reality of G requires that

$$\epsilon(-\omega) = \epsilon^*(\omega^*). \tag{154}$$

That G is zero for negative times requires that $\epsilon(\omega)$ be analytic in the upper half of the frequency plane. Assuming that $G(t) \to 0$ as $t \to \infty$, one finds that $\epsilon(\omega)$ is analytic on the real axis. This last statement is actually not true for conductors which give a contribution to $\epsilon \sim i\sigma/\omega$ so that there is a pole at the origin. Finally, from the small-time behavior of $G(t)$, one can infer that at large frequencies the real part of $\epsilon(\omega) - 1$ varies as ω^{-2} while the imaginary part varies as ω^{-3}. This is accomplished by repeatedly integrating by parts

$$\epsilon(\omega) - 1 = \int_0^\infty d\tau G(\tau) e^{i\omega\tau} \approx \frac{iG(0^+)}{\omega} - \frac{G'(0^+)}{\omega^2} + \frac{iG'(0^+)}{\omega^3} + \cdots \tag{155}$$

This series is convergent for large ω. The first term vanishes if $G(\tau)$ is continuous accross $\tau = 0$. Thus

$$\Re\left(\epsilon(\omega) - 1\right) \sim \frac{1}{\omega^2} \qquad \Im\left(\epsilon(\omega) - 1\right) \sim \frac{1}{\omega^3} \tag{156}$$

From inspection, one may see that the various dielectric functions we have contrived satisfy these conditions.

Given that the dielectric function has the analyticity properties described above, it turns out that by rather standard manipulations making use of Cauchy's integral theorem, one can write the imaginary part of $\epsilon(\omega)$ in terms of an integral of the real part over real frequencies and conversely. That one can do so is important because it means, for example, that if one succeeds in measuring just the real (imaginary) part, the imaginary (real) part is then known. The downside of this apparent miracle is that one has to know the real or imaginary part for all real frequencies in order to obtain the other part.

To see how this works, notice that as a consequence of the analytic properties of the dielectric function, it obeys the relation

$$\epsilon(z) = 1 + \frac{1}{2\pi i} \oint_C d\omega' \, \frac{\epsilon(\omega') - 1}{\omega' - z} \tag{157}$$

where the contour does not enter the lower half-plane (where ϵ may have poles) anywhere and where z is inside of the contour. Let C consist of the real axis and a large semicircle which closes the path in the upper half-plane.

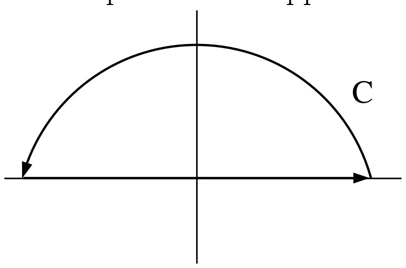


Figure 29: Contour C: $\epsilon(\omega)$ is analytic inside an on C..

Then, given that ϵ falls off fast enough, as described above, at large ω, the semicircular part of the path does not contribute to the integral. Hence we find that

$$\epsilon(z) = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega' \, \frac{\epsilon(\omega') - 1}{\omega' - z}. \tag{158}$$

At this juncture, z can be any point in the upper half-plane. Let's use $z = \omega + i\eta$ and take the limit of $\eta \to 0$, finding

$$\epsilon(\omega + i\eta) = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega' \frac{\epsilon(\omega') - 1}{\omega' - \omega - i\eta}. \tag{159}$$

The presence of the η in the denominator means that at the integration point $\omega' = \omega$, we must be careful to keep the singularity inside of, or above, the contour. Here we pick up $2\pi i$ times the residue, and the residue is just $\epsilon(\omega) - 1$. This relation shows identity but is not useful otherwise. However, one can also pull the following trick: If we integrate right across the singularity, taking the principal part (denoted P) of the integral plus an infinitesmal semicircle right below the singularity that amounts to taking $i\pi$ times the residue. Hence we can make the replacement

$$\frac{1}{\omega' - \omega - i\eta} \to P\left(\frac{1}{\omega' - \omega}\right) + i\pi\delta(\omega' - \omega) \tag{160}$$

where P stands for the principal part; this substitution leads to

$$\epsilon(\omega) = 1 + \frac{1}{\pi i} P \int_{-\infty}^{\infty} d\omega' \, \frac{\epsilon(\omega') - 1}{\omega' - \omega} \tag{161}$$

Let us write separately the real and imaginary parts of this expression:

$$\begin{aligned} \Re[\epsilon(\omega)] &= 1 + \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \, \frac{\Im[\epsilon(\omega')]}{\omega' - \omega} \\ \Im[\epsilon(\omega)] &= \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \, \frac{\Re[\epsilon(\omega') - 1]}{\omega' - \omega} \end{aligned} \tag{162}$$

These equations are known as the *Kramers-Kronig relations* for the dielectric function. They may be written as integrals over only positive frequencies because of the fact that the real part of $\epsilon(\omega)$ is an even function of ω while the imaginary part is odd. It should also be pointed out that we have assumed there is no pole in $\epsilon(\omega)$ at $\omega = 0$; if there is one (conductors have dielectric functions with this property) some modification of these expressions will be necessary.

11 Arrival of a Signal in a Dispersive Medium

Most of the wave trains one receives, such as radio signals, messages from within or without the galaxy (sent by stars, pulsars, neutron stars, etc), and so on, have to traverse dispersive media to get wherever they go. Consequently it is of considerable importance to know how the signals are distorted by the intervening material. The basic idea is this: we have seen how a pulse centered at some particular wave number or frequency tends to travel with the group velocity of the central frequency and also spreads some as a consequence of the frequency-dependence of the index of refraction or dielectric function. If the dispersion is very large, as in regions of anomalous dispersion, the pulse will not simply spread some but will be distorted beyond recognition. In addition, frequency components in this region will be strongly attenuated and so will disappear from the wave train after awhile. If a signal is initially very broad in frequency, having components ranging from very low ones, where the group velocity is roughly constant and equal to $c/\sqrt{\epsilon(0)}$, to very high ones where $\epsilon(\omega) \approx 1$ and the group velocity is about c, then the signal that arrives after traveling through a significant length of medium will be very different indeed from the initial one. All of the frequency components around the regions of anomalous dispersion will be gone. There will be some high-frequency component which travels at a speed around c and so arrives first; it is generally called the "first precursor." Then after awhile the remainder of the signal will arrive. The leading edge of this part is called the "second precursor" and it consists of those lower frequency components which have the largest group velocity and which are not appreciably attenuated. These are usually[13] the very low frequency components.

It is a straightforward matter to determine what the signal will be, using the superposition principle. Consider a pulse in one dimension with an amplitude $u(z,t)$. Given that one knows the form of this pulse and its first space derivatives as functions of time at some initial position in space[14], called z_0, then one may determine by Fourier analysis the amplitude $A(\omega)$ of the various frequency components in it. Since a frequency component ω propagates according to $\exp[i(\omega n(\omega)z/c - \omega t)]$, it is then easy in principle to find $u(z,t)$:

$$u(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega\, A(\omega) e^{i(\omega n(\omega) z/c - \omega t)}. \tag{163}$$

[13]But not always; the whistler provides a a counter example.

[14]Notice that instead of solving an initial value problem in time, we here rephrase it as an initial value problem in space.

If we can do this integral for the index of refraction of our choice, we can find the form of the wave train at all space points at any later time. Among other things, one can show by making use of the analyticity properties of the dielectric function that it is impossible for an electromagnetic signal to travel faster than the speed of light. See Jackson.

As a very simple example, consider a single-resonance dielectric function with no absorption,

$$\epsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} = n^2(\omega) \tag{164}$$

or

$$n(\omega) = \left(\frac{\omega_0^2 - \omega^2 + \omega_p^2}{\omega_0^2 - \omega^2}\right)^{1/2}. \tag{165}$$

Then

$$2n\frac{dn}{d\omega} = 2\omega\frac{\omega_p^2}{(\omega_0^2 - \omega^2)^2} \tag{166}$$

so

$$n\omega\frac{dn}{d\omega} + n^2 = \frac{\omega^2\omega_p^2}{(\omega_0^2 - \omega^2)^2} + 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} = \frac{\omega_0^4 - 2\omega_0^2\omega^2 + \omega^4 + \omega_p^2\omega_0^2}{(\omega_o^2 - \omega^2)^2}. \tag{167}$$

Hence

$$v_g = \frac{c}{\omega\frac{dn}{d\omega} + n} = c\left(\frac{\omega_0^2 - \omega^2 + \omega_p^2}{\omega_0^2 - \omega^2}\right)^{1/2}\frac{(\omega_0^2 - \omega^2)^2}{(\omega_0^2 - 2\omega_0^2\omega^2 + \omega^4 + \omega_p^2\omega_0^2)} \tag{168}$$

The first plot shows the character of v_g and of $n(\omega)$. The group velocity is largest for the largest frequencies; these will combine to provide the first precursor which may well be weak to the extent that the initial pulse does not contain many high-frequency components. The first precursor continues as lower frequency components (but still larger than $\sqrt{\omega_0^2 + \omega_p^2}$) come through. While this is going on, all of the very low frequency components arrive. This is the second precursor. Finally, if the pulse is actually a long wave train which has one predominant frequency in it, then after some time the received pulse settles down to something more or less harmonic, showing just this frequency.

A Waves in a Conductor

When we discussed the propagation of waves in an ideal dielectric, we showed that the fields were transverse to the direction of propagation. This corresponds to an isulating material, with a vanishing electrical conductivity. When we extend our discussion to include media of finite conductivity σ, there is no a priori reason that the fields will still be transverse to the direction of propagation.

Let's show that we need not worry about any longitudinal fields. Suppose that once again we have some linear medium with $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, and $\mathbf{J} = \sigma\mathbf{E}$; ϵ, μ, and σ are taken as real. Then the Maxwell equations become

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \cdot \mathbf{E} = 0, \qquad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \tag{169}$$

and

$$\nabla \times \mathbf{B} = \frac{4\pi\mu}{c}\sigma\mathbf{E} + \frac{\epsilon\mu}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{170}$$

Let's look for solutions to Maxwell's equations in the form of logitudinial waves,

$$\mathbf{E} = \hat{\mathbf{z}}E(z,t) \;\; ; \;\; \mathbf{B} = \hat{\mathbf{z}}B(z,t) \tag{171}$$

Since $\nabla \cdot \mathbf{B} = \nabla \cdot \mathbf{E} = 0$, E and B can be functions of time only. Thus $\nabla \times \mathbf{E} = \nabla \times \mathbf{B} = 0$, and the other two Maxwell's equations become

$$\frac{\partial \mathbf{B}}{\partial t} = 0 \;\; ; \;\; \frac{4\pi\sigma}{c}\mathbf{E} + \frac{\epsilon}{c}\frac{\partial \mathbf{E}}{\partial t} = 0 \tag{172}$$

The first says that $\mathbf{B}$ must be constant. The second says that $\mathbf{E}$ while uniform in space has a time dependence

$$\mathbf{E}(t) = \mathbf{E}(0)e^{-4\pi\sigma t/\epsilon} \tag{173}$$

In a conductor, $\sigma \approx 10^{16}\ \mathrm{sec}^{-1}$. Thus $\mathbf{E}(t)$ falls off very rapidly and may be neglected. Thus as worst there is a constant logitudinal $\mathbf{B}$-field as part of our wave in a conductor. Since Maxwell's equations are linear, we may drop this trivial solution and just consider the transverse fields.