Chapter 8

Electromagnetic waves

David Morin, morin@physics.harvard.edu

The waves we've dealt with so far in this book have been fairly easy to visualize. Waves involving springs/masses, strings, and air molecules are things we can apply our intuition to. But we'll now switch gears and talk about electromagnetic waves. These are harder to get a handle on, for a number of reasons. First, the things that are oscillating are electric and magnetic fields, which are much harder to see (which is an ironic statement, considering that we see with light, which is an electromagnetic wave). Second, the fields can have components in various directions, and there can be relative phases between these components (this will be important when we discuss polarization). And third, unlike all the other waves we've dealt with, electromagnetic waves don't need a medium to propagate in. They work just fine in vacuum. In the late 1800's, it was generally assumed that electromagnetic waves required a medium, and this hypothesized medium was called the "ether." However, no one was ever able to observe the ether. And for good reason, because it doesn't exist.

This chapter is a bit long. The outline is as follows. In Section 8.1 we talk about waves in an extended LC circuit, which is basically what a coaxial cable is. We find that the system supports waves, and that these waves travel at the speed of light. This section serves as motivation for the fact that light is an electromagnetic wave. In Section 8.2 we show how the wave equation for electromagnetic waves follows from Maxwell's equations. Maxwell's equations govern all of electricity and magnetism, so it is no surprise that they yield the wave equation. In Section 8.3 we see how Maxwell's equations constrain the form of the waves. There is more information contained in Maxwell's equations than there is in the wave equation. In Section 8.4 we talk about the energy contained in an electromagnetic wave, and in particular the energy flow which is described by the *Poynting vetor*. In Section 8.5 we talk about the momentum of an electromagnetic wave. We saw in Section 4.4 that the waves we've discussed so far carry energy but not momentum. Electromagnetic waves carry both.[1] In Section 8.6 we discuss polarization, which deals with the relative phases of the different components of the electric (and magnetic) field. In Section 8.7 we show how an electromagnetic wave can be produced by an oscillating (and hence accelerating) charge. Finally, in Section 8.8 we discuss the reflection and transmission that occurs when an electromagnetic wave encounters the boundary between two different regions, such as air

[1] Technically, all waves carry momentum, but this momentum is suppressed by a factor of v/c, where v is the speed of the wave and c is the speed of light. This follows from the relativity fact that energy is equivalent to mass. So a flow of energy implies a flow of mass, which in turn implies nonzero momentum. However, the factor of v/c causes the momentum to be negligible unless we're dealing with relativistic speeds.

and glass. We deal with both normal and non-normal angles of incidence. The latter is a bit more involved due to the effects of polarization.

8.1 Cable waves

Before getting into Maxwell's equations and the wave equation for light, let's do a warmup example and study the electromagnetic waves that propagate down a coaxial cable. This example should help convince you that light is in fact an electromagnetic wave.

To get a handle on the coaxial cable, let's first look at the idealized circuit shown in Fig. 1. All the inductors are L, and all the capacitors are C. There are no resistors in the circuit. With the charges, currents, and voltages labeled as shown, we have three facts:

1. The charge on a capacitor is $Q = CV \Longrightarrow q_n = CV_n$.
2. The voltage across an inductor is $V = L(dI/dt) \Longrightarrow V_{n-1} - V_n = L(dI_n/dt)$.
3. Conservation of charge gives $I_n - I_{n+1} = dq_n/dt$.

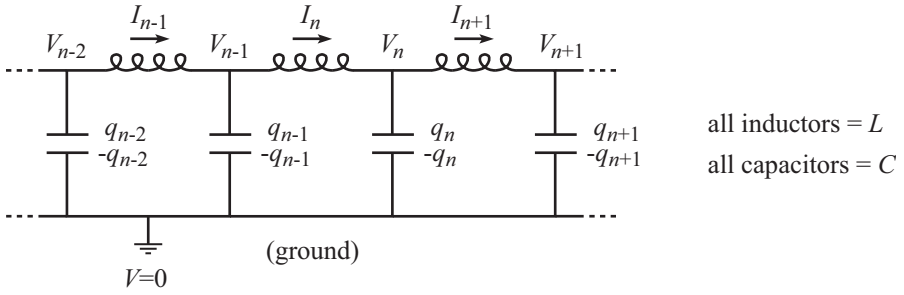


Figure 1

Our goal is to produce an equation, which will end up being a wave equation, for one of the three variables, q, I, and V (the wave equations for all of them will turn out to be the same). Let's eliminate q and I, in favor of V. We could manipulate the above equations in their present form in terms of discrete quantities, and then take the continuum limit (see Problem [to be added]). But it is much simpler to first take the continuum limit and then do the manipulation. If we let the grid size in Fig. 1 be Δx, then by using the definition of the derivative, the above three facts become

$$\begin{aligned} q &= CV, \\ -\Delta x \frac{\partial V}{\partial x} &= L\frac{\partial I}{\partial t}, \\ -\Delta x \frac{\partial I}{\partial x} &= \frac{\partial q}{\partial t}. \end{aligned} \tag{1}$$

Substituting $q = CV$ from the first equation into the third, and defining the inductance and capacitance per unit length as $L_0 \equiv L/\Delta x$ and $C_0 \equiv C/\Delta x$, the last two equations become

$$-\frac{\partial V}{\partial x} = L_0\frac{\partial I}{\partial t}, \qquad \text{and} \qquad -\frac{\partial I}{\partial x} = C_0\frac{\partial V}{\partial t}. \tag{2}$$

If we take $\partial/\partial x$ of the first of these equations and $\partial/\partial t$ of the second, and then equate the results for $\partial^2 I/\partial x\,\partial t$, we obtain

$$-\frac{1}{L_0}\frac{\partial^2 V}{\partial x^2} = -C_0\frac{\partial^2 V}{\partial t^2} \quad \Longrightarrow \quad \boxed{\frac{\partial^2 V(x,t)}{\partial t^2} = \frac{1}{L_0 C_0}\frac{\partial^2 V(x,t)}{\partial x^2}} \tag{3}$$

This is the desired wave equation, and it happens to be dispersionless. We can quickly read off the speed of the waves, which is

$$v = \frac{1}{\sqrt{L_0 C_0}} \,. \tag{4}$$

If we were to subdivide the circuit in Fig. 1 into smaller and smaller cells, L and C would depend on Δx (and would go to zero as $\Delta x \to 0$), so it makes sense to work with the quantities L_0 and C_0. This is especially true in the case of the actual cable we'll discuss below, for which the choice of Δx is arbitrary. L_0 and C_0 are the meaningful quantities that are determined by the nature of the cable.

Note that since the first fact above says that $q \propto V$, the exact same wave equation holds for q. Furthermore, if we had eliminated V instead of I in Eq. (2) by taking $\partial/\partial t$ of the first equation and $\partial/\partial x$ of the second, we would have obtained the same wave equation for I, too. So V, q, and I all satisfy the same wave equation.

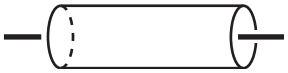


Figure 2

Let's now look at an actual coaxial cable. Consider a conducting wire inside a conducting cylinder, with vacuum in the region between them, as shown in Fig. 2. Assume that the wire is somehow constrained to be in the middle of the cylinder. (In reality, the inbetween region is filled with an insulator which keeps the wire in place, but let's keep things simple here with a vacuum.) The cable has an inductance L_0 per unit length, in the same way that two parallel wires have a mutual inductance per unit length. (The cylinder can be considered to be made up of a large number wires parallel to its axis.) It also has a capacitance C_0 per unit length, because a charge difference between the wire and the cylinder will create a voltage difference.

It can be shown that (see Problem [to be added], although it's perfectly fine to just accept this)

$$L_0 = \frac{\mu_0}{2\pi} \ln(r_2/r_1) \qquad \text{and} \qquad C_0 = \frac{2\pi\epsilon_0}{\ln(r_2/r_1)} \,, \tag{5}$$

where r_2 is the radius of the cylinder, and r_1 is the radius of the wire. The two physical constants in these equations are the *permeability of free space*, μ_0, and the *permittivity of free space*, ϵ_0. Their values are (μ takes on this value by definition)

$$\mu_0 = 4\pi \cdot 10^{-7}\,\text{H/m}\,, \qquad \text{and} \qquad \epsilon_0 \approx 8.85 \cdot 10^{-12}\,\text{F/m}\,. \tag{6}$$

H and F are the Henry and Farad units of inductance and capacitance. Using Eq. (5), the wave speed in Eq. (4) equals

$$v = \frac{1}{\sqrt{L_0 C_0}} = \frac{1}{\sqrt{\mu_0\epsilon_0}} \approx \frac{1}{\sqrt{(4\pi \cdot 10^{-7}\,\text{H/m})(8.85 \cdot 10^{-12}\,\text{F/m})}} \approx 3 \cdot 10^8\,\text{m/s}. \tag{7}$$

This is the speed of light! We see that the voltage (and charge, and current) wave that travels down the cable travels at the speed of light. And because there are electric and magnetic fields in the cable (due to the capacitance and inductance), these fields also undergo wave motion. Since the waves of these fields travel with the same speed as the original voltage wave, it is a good bet that electromagnetic waves have something to do with light. The reasoning here is that there probably aren't too many things in the world that travel with the speed of light. So if we find something that travels with this speed, then it's probably light (loosely speaking, at least; it need not be in the visible range). Let's now be rigorous and show from scratch that all electromagnetic waves travel at the speed of light (in vacuum).

8.2 The wave equation

By "from scratch" we mean by starting with Maxwell's equations. We have to start somewhere, and Maxwell's equations govern all of (classical) electricity and magnetism. There

are four of these equations, although when Maxwell first wrote them down, there were 22 of them. But they were gradually rewritten in a more compact form over the years. Maxwell's equations in vacuum in SI units are (in perhaps overly-general form):

Differential form	Integrated form
$\nabla \cdot \mathbf{E} = \dfrac{\rho_{\rm E}}{\epsilon_0}$	$\displaystyle\int \mathbf{E} \cdot d\mathbf{A} = \frac{Q_{\rm E}}{\epsilon_0}$
$\nabla \cdot \mathbf{B} = \rho_{\rm B}$	$\displaystyle\int \mathbf{B} \cdot d\mathbf{A} = Q_{\rm B}$
$\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} + \mathbf{J}_{\rm B}$	$\displaystyle\int \mathbf{E} \cdot d\mathbf{l} = -\frac{d\Phi_{\rm B}}{dt} + I_{\rm B}$
$\nabla \times \mathbf{B} = \mu_0\epsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J}_{\rm E}$	$\displaystyle\int \mathbf{B} \cdot d\mathbf{l} = \mu_0\epsilon_0 \frac{d\Phi_{\rm E}}{dt} + \mu_0 I_{\rm E}$

Table 1

If you erase the μ_0's and ϵ_0's here (which arise from the arbitrary definitions of the various units), then these equations are symmetric in $\mathbf{E}$ and $\mathbf{B}$, except for a couple minus signs. The ρ's are the electric and (hypothetical) magnetic charge densities, and the $\mathbf{J}$'s are the current densities. The Q's are the charges enclosed by the surfaces that define the $d\mathbf{A}$ integrals, the Φ's are the field fluxes through the loops that define the $d\mathbf{l}$ integrals, and the I's are the currents through these loops.

No one has ever found an isolated magnetic charge (a magnetic monopole), and there are various theoretical considerations that suggest (but do not yet prove) that magnetic monopoles can't exist, at least in our universe. So we'll set $\rho_{\rm B}$, $\mathbf{J}_{\rm B}$, and $I_{\rm B}$ equal to zero from here on. This will make Maxwell's equations appear non-symmetrical, but we'll soon be setting the analogous electric quantities equal to zero too, since we'll be dealing with vacuum. So in the end, the equations for our purposes will be symmetric (except for the μ_0, the ϵ_0, and a minus sign). Maxwell's equations with no magnetic charges (or currents) are:

Differential form	Integrated form	Known as
$\nabla \cdot \mathbf{E} = \dfrac{\rho_{\rm E}}{\epsilon_0}$	$\displaystyle\int \mathbf{E} \cdot d\mathbf{A} = \frac{Q_{\rm E}}{\epsilon_0}$	Gauss' Law
$\nabla \cdot \mathbf{B} = 0$	$\displaystyle\int \mathbf{B} \cdot d\mathbf{A} = 0$	No magnetic monopoles
$\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$	$\displaystyle\int \mathbf{E} \cdot d\mathbf{l} = -\frac{d\Phi_{\rm B}}{dt}$	Faraday's Law
$\nabla \times \mathbf{B} = \mu_0\epsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J}_{\rm E}$	$\displaystyle\int \mathbf{B} \cdot d\mathbf{l} = \mu_0\epsilon_0 \frac{d\Phi_{\rm E}}{dt} + \mu_0 I_{\rm E}$	Ampere's Law

Table 2

The last of these, Ampere's Law, includes the so-called "displacement current," $d\Phi_{\rm E}/dt$.

Our goal is to derive the wave equation for the $\mathbf{E}$ and $\mathbf{B}$ fields in vacuum. Since there are no charges of any kind in vacuum, we'll set $\rho_{\rm E}$ and $\mathbf{J}_{\rm E} = 0$ from here on. And we'll only need the differential form of the equations, which are now

$$\nabla \cdot \mathbf{E} = 0, \tag{8}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{9}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}\,, \tag{10}$$

$$\nabla \times \mathbf{B} = \mu_0\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\,. \tag{11}$$

These equations are symmetric in $\mathbf{E}$ and $\mathbf{B}$ except for the factor of $\mu_0\epsilon_0$ and a minus sign. Let's eliminate $\mathbf{B}$ in favor of $\mathbf{E}$ and see what we get. If we take the curl of Eq. (10) and

then use Eq. (11) to get ride of $\mathbf{B}$, we obtain

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) &= -\nabla \times \frac{\partial \mathbf{B}}{\partial t} \\ &= -\frac{\partial (\nabla \times \mathbf{B})}{\partial t} \\ &= -\frac{\partial}{\partial t}\left(\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right) \\ &= -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}\,. \end{aligned} \tag{12}$$

On the left side, we can use the handy "BAC-CAB" formula,

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}). \tag{13}$$

See Problem [to be added] for a derivation of this. This formula holds even if we have differential operators (such as ∇) instead of normal vectors, but we have to be careful to keep the ordering of the letters the same (this is evident if you go through the calculation in Problem [to be added]). Since both $\mathbf{A}$ and $\mathbf{B}$ are equal to ∇ in the present application, the ordering of $\mathbf{A}$ and $\mathbf{B}$ in the $\mathbf{B}(\mathbf{A} \cdot \mathbf{C})$ term doesn't matter. But the $\mathbf{C}(\mathbf{A} \cdot \mathbf{B})$ must correctly be written as $(\mathbf{A} \cdot \mathbf{B})\mathbf{C}$. The lefthand side of Eq. (12) then becomes

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) &= \nabla(\nabla \cdot \mathbf{E}) - (\nabla \cdot \nabla)\mathbf{E} \\ &= 0 - \nabla^2 \mathbf{E}, \end{aligned} \tag{14}$$

where the zero follows from Eq. (8). Plugging this into Eq. (12) finally gives

$$-\nabla^2 \mathbf{E} = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad \Longrightarrow \quad \boxed{\frac{\partial^2 \mathbf{E}}{\partial t^2} = \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{E}} \qquad (\textit{wave equation}) \tag{15}$$

Note that we didn't need to use the second of Maxwell's equations to derive this.

In the above derivation, we could have instead eliminated $\mathbf{E}$ in favor of $\mathbf{B}$. The same steps hold; the minus signs end up canceling again, as you should check, and the first equation is now not needed. So we end up with exactly the same wave equation for $\mathbf{B}$:

$$\boxed{\frac{\partial^2 \mathbf{B}}{\partial t^2} = \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{B}} \qquad (\textit{wave equation}) \tag{16}$$

The speed of the waves (both $\mathbf{E}$ and $\mathbf{B}$) is given by the square root of the coefficient on the righthand side of the wave equation. (This isn't completely obvious, since we're now working in three dimensions instead of one, but we'll justify this in Section 8.3.1 below.) The speed is therefore

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \approx 3 \cdot 10^8 \text{ m/s}. \tag{17}$$

This agrees with the result in Eq. (7). But we now see that we don't need a cable to support the propagation of electromagnetic waves. They can propagate just fine in vacuum! This is a fundamentally new feature, because every wave we've studied so far in this book (longitudinal spring/mass waves, transverse waves on a string, longitudinal sound waves, etc.), needs a medium to propagate in/on. But not so with electromagnetic waves.

Eq. (15), and likewise Eq. (16), is a vector equation. So it is actually shorthand for three separate equations for each of the components:

$$\frac{\partial^2 E_x}{\partial t^2} = \frac{1}{\mu_0 \epsilon_0}\left(\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2}\right), \tag{18}$$

and likewise for E_y and E_z. Each component undergoes wave motion. As far as the wave equation in Eq. (15) is concerned, the waves for the three components are completely independent. Their amplitudes, frequencies, and phases need not have anything to do with each other. *However*, there is more information contained in Maxwell's equations than in the wave equation. The latter follows from the former, but not the other way around. There is no reason why one equation that follows from four equations (or actually just three of them) should contain as much information as the original four. In fact, it is highly unlikely. And as we will see in Section 8.3, Maxwell's equations do indeed further constrain the form of the waves. In other words, although the wave equation in Eq. (15) gives us information about the electric-field wave, it doesn't give us *all* the information.

Index of refraction

In a dielectric (equivalently, an insulator), the vacuum quantities μ_0 and ϵ_0 in Maxwell's equations are replaced by new values, μ and ϵ. (We'll give some justification of this below, but see Sections 10.11 and 11.10 in Purcell's book for the full treatment.) Our derivation of the wave equation for electromagnetic waves in a dielectric proceeds in exactly the same way as for the vacuum case above, except with $\mu_0 \to \mu$ and $\epsilon_0 \to \epsilon$. We therefore end up with a wave velocity equal to

$$v = \frac{1}{\sqrt{\mu\epsilon}}\,. \tag{19}$$

The *index of refraction*, n, of a dielectric is defined by $v \equiv c/n$, where c is the speed of light in vacuum. We therefore have

$$v = \frac{c}{n} \quad \Longrightarrow \quad n = \frac{c}{v} = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}}\,. \tag{20}$$

Since it happens to be the case that that $\mu \approx \mu_0$ for most dielectrics, we have the approximate result that

$$n \approx \sqrt{\frac{\epsilon}{\epsilon_0}} \qquad (\text{if} \quad \mu \approx \mu_0). \tag{21}$$

And since we must always have $v \le c$, this implies $n \ge 1 \Longrightarrow \epsilon \ge \epsilon_0$.

Strictly speaking, Maxwell's equations with μ_0 and ϵ_0 work in any medium. But the point is that if we don't have a vacuum, then induced charges and currents may arise. In particular there are two types of charges. There are so-called *free* charges, which are additional charges that we can plop down in a material. This is normally what we think of when we think of charge. (The term "free" is probably not the best term, because the charges need not be free to move. We can bolt them down if we wish.) But additionally, there are *bound* charges. These are the effective charges that get produced when polar molecules align themselves in certain ways to "shield" the bound charges.

For example, if we place a positive free charge q_{free} in a material, then the nearby polar molecules will align themselves so that their negative ends form a negative layer around the free charge. The net charge inside a Gaussian surface around the charge is therefore less than q. Call it q_{net}. Maxwell's first equation is then $\nabla \cdot \mathbf{E} = \rho_{\text{net}}/\epsilon_0$. However, it is generally much easier to deal with ρ_{free} than ρ_{net}, so let's define ϵ by $\rho_{\text{net}}/\rho_{\text{free}} \equiv \epsilon_0/\epsilon < 1$.[2] Maxwell's first equation can then be written as

$$\nabla \cdot \mathbf{E} = \frac{\rho_{\text{free}}}{\epsilon}\,. \tag{22}$$

[2]The fact that the shielding is always proportional to q_{free} (at least in non-extreme cases) implies that there is a unique value of ϵ that works for all values of q_{free}.

The electric field in the material around the point charge is less than what it would be in vacuum, by a factor of ϵ_0/ϵ (and ϵ is always greater than or equal to ϵ_0, because it isn't possible to have "anti-shielding"). In a dielectric, the fact that ϵ is greater than ϵ_0 is consistent with the fact that the index of refraction n in Eq. (21) is always greater than 1, which in turn is consistent with the fact that v is always less than c.

A similar occurrence happens with currents. There can be *free* currents, which are the normal ones we think about. But there can also be *bound* currents, which arise from tiny current loops of electrons spinning around within their atoms. This is a little harder to visualize than the case with the charges, but let's just accept here that the fourth of Maxwell's equations becomes $\nabla \times \mathbf{B} = \mu\epsilon\partial\mathbf{E}/\partial t + \mu\mathbf{J}_{\text{free}}$. But as mentioned above, μ is generally close to μ_0 for most dielectrics, so this distinction usually isn't so important.

To sum up, we can ignore all the details about what's going on at the atomic level by pretending that we have a vacuum with modified μ and ϵ values. Although there certainly exist *bound* charges and currents in the material, we can sweep them under the rug and consider only the *free* charges and currents, by using the modified μ and ϵ values.

The above modified expressions for Maxwell's equations are correct if we're dealing with a single medium. But if we have two or more mediums, the correct way to write the equations is to multiply the first equation by ϵ and divide the fourth equation by μ (see Problem [to be added] for an explanation of this). The collection of all four Maxwell's equations is then

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho_{\text{free}}, \\ \nabla \cdot \mathbf{B} &= 0, \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \\ \nabla \times \mathbf{H} &= \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}_{\text{free}}, \end{aligned} \tag{23}$$

where $\mathbf{D} \equiv \epsilon\mathbf{E}$ and $\mathbf{H} \equiv \mathbf{B}/\mu$. $\mathbf{D}$ is called the *electric displacement* vector, and $\mathbf{H}$ goes by various names, including simply the "magnetic field." But you can avoid confusing it with $\mathbf{B}$ if you use the letter $\mathbf{H}$ and not the name "magnetic field."

8.3 The form of the waves

8.3.1 The wavevector k

What is the dispersion relation associated with the wave equation in Eq. (15)? That is, what is the relation between the frequency and wavenumber? Or more precisely, what is the dispersion relation for each component of $\mathbf{E}$, for example the E_x that satisfies Eq. (18)? All of the components are in general functions of four coordinates: the three spatial coordinates x, y, z, and the time t. So by the same reasoning as in the two-coordinate case we discussed at the end of Section 4.1, we know that we can Fourier-decompose the function $E_x(x, y, z, t)$ into exponentials of the form,

$$Ae^{i(k_x x + k_y y + k_z z - \omega t)} \equiv Ae^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \qquad \text{where} \quad \mathbf{k} \equiv (k_x, k_y, k_z). \tag{24}$$

Likewise for E_y and E_z. And likewise for the three components of $\mathbf{B}$. These are traveling waves, although we can form combinations of them to produce standing waves.

$\mathbf{k}$ is known as the *wavevector*. As we'll see below, the magnitude $k \equiv |\mathbf{k}|$ plays exactly the same role that k played in the 1-D case. That is, k is the wavenumber. It equals 2π times the number of wavelengths that fit into a unit length. So $k = 2\pi/\lambda$. We'll also see below that the direction of $\mathbf{k}$ is the direction of the propagation of the wave. In the 1-D case,

the wave had no choice but to propagate in the $\pm\mathbf{x}$ direction. But now it can propagate in any direction in 3-D space.

Plugging the exponential solution in Eq. (24) into Eq. (18) gives

$$-\omega^2 = \frac{1}{\mu_0\epsilon_0}(-k_x^2 - k_y^2 - k_z^2) \quad \Longrightarrow \quad \omega^2 = \frac{|\mathbf{k}|^2}{\mu_0\epsilon_0} \quad \Longrightarrow \quad \boxed{\omega = c|\mathbf{k}|} \tag{25}$$

where $c = 1/\sqrt{\mu_0\epsilon_0}$, and where we are using the convention that ω is positive. Eq. (25) is the desired dispersion relation. It is a trivial relation, in the sense that electromagnetic waves in vacuum are dispersionless.

When we go through the same procedure for the other components of $\mathbf{E}$ and $\mathbf{B}$, the "A" coefficient in Eq. (24) can be different for the $2 \cdot 3 = 6$ different components of the fields. And technically $\mathbf{k}$ and ω can be different for the six components too (as long as they satisfy the same dispersion relation). However, although we would have solutions to the six different waves equations, we wouldn't have solutions to Maxwell's equations. This is one of the cases where the extra information contained in Maxwell's equations is important. You can verify (see Problem [to be added]) that if you want Maxwell's equations to hold for *all* $\mathbf{r}$ and t, then $\mathbf{k}$ and ω must be the same for all six components of $\mathbf{E}$ and $\mathbf{B}$. If we then collect the various "A" components into the two vectors $\mathbf{E}_0$ and $\mathbf{B}_0$ (which are constants, independent of $\mathbf{r}$ and t), we can write the six components of $\mathbf{E}$ and $\mathbf{B}$ in vector form as

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \qquad \text{and} \qquad \mathbf{B} = \mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \tag{26}$$

where the $\mathbf{k}$ vector and the ω frequency are the same in both fields. The vectors $\mathbf{E}_0$ and $\mathbf{B}_0$ can be complex. If they do have an imaginary part, it will produce a phase in the cosine function when we take the real part of the above exponentials. This will be important when we discuss polarization.

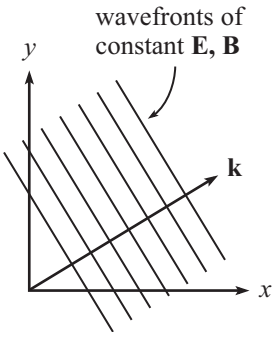


Figure 3

From Eq. (26), we see that $\mathbf{E}$ (and likewise $\mathbf{B}$) depends on $\mathbf{r}$ through the dot product $\mathbf{k}\cdot\mathbf{r}$. So $\mathbf{E}$ has the same value everywhere on the surface defined by $\mathbf{k}\cdot\mathbf{r} = C$, where C is some constant. This surface is a plane that is perpendicular to $\mathbf{k}$. This follows from the fact that if $\mathbf{r}_1$ and $\mathbf{r}_2$ are two points on the surface, then $\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2) = C - C = 0$. Therefore, the vector $\mathbf{r}_1 - \mathbf{r}_2$ is perpendicular to $\mathbf{k}$. Since this holds for any $\mathbf{r}_1$ and $\mathbf{r}_2$ on the surface, the surface must be a plane perpendicular to $\mathbf{k}$. If we suppress the z dependence of $\mathbf{E}$ and $\mathbf{B}$ for the sake of drawing a picture on a page, then for a given wavevector $\mathbf{k}$, Fig. 3 shows some "wavefronts" with common phases $\mathbf{k}\cdot\mathbf{r} - \omega t$, and hence common values of $\mathbf{E}$ and $\mathbf{B}$. The planes perpendicular to $\mathbf{k}$ in the 3-D case become lines perpendicular to $\mathbf{k}$ in the 2-D case. Every point on a given plane is equivalent, as far as $\mathbf{E}$ and $\mathbf{B}$ are concerned.

How do these wavefronts move as time goes by? Well, they must always be perpendicular to $\mathbf{k}$, so all they can do is move in the direction of $\mathbf{k}$. How fast do they move? The dot product $\mathbf{k}\cdot\mathbf{r}$ equals $kr\cos\theta$, where θ is the angle between $\mathbf{k}$ and a given position $\mathbf{r}$, and where $k \equiv |\mathbf{k}|$ and $r = |\mathbf{r}|$. If we group the product as $k(r\cos\theta)$, we see that it equals k times the projection of $\mathbf{r}$ along $\mathbf{k}$. If we rotate our coordinate system so that a new x' axis points in the $\mathbf{k}$ direction, then the projection $r\cos\theta$ simply equals the x' value of the position. So the phase $\mathbf{k}\cdot\mathbf{r} - \omega t$ equals $kx' - \omega t$. We have therefore reduced the problem to a 1-D problem (at least as far as the phase is concerned), so we can carry over all of our 1-D results. In particular, the phase velocity (and group velocity too, since the wave equation in Eq. (15) is dispersionless) is $v = \omega/k$, which we see from Eq. (25) equals $c = 1/\sqrt{\mu_0\epsilon_0}$.

REMARK: We just found that the phase velocity has magnitude

$$v = \frac{\omega}{k} \equiv \frac{\omega}{|\mathbf{k}|} = \frac{\omega}{\sqrt{k_x^2 + k_y^2 + k_z^2}}, \tag{27}$$

and it points in the $\hat{\mathbf{k}}$ direction. You might wonder if the simpler expression ω/k_x has any meaning. And likewise for y and z. It does, but it isn't a terribly useful quantity. It is the velocity at

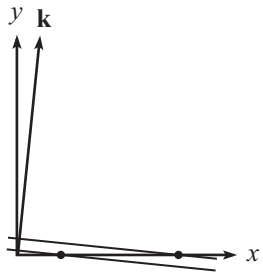


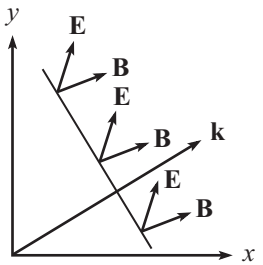
Figure 4

which a point with constant phase moves in the x direction, with y and z held constant. This follows from the fact that if we let the constant y and z values be y_0 and z_0, then the phase equals $k_x x + k_y y_0 + k_z z_0 - \omega t = k_x x - \omega t + C$, where C is a constant. So we effectively have a 1-D problem for which the phase velocity is ω/k_x.

But note that this velocity can be made arbitrarily large, or even infinite if $k_x = 0$. Fig. 4 shows a situation where $\mathbf{k}$ points mainly in the y direction, so k_x is small. Two wavefronts are shown, and they move upward along the direction of $\mathbf{k}$. In the time during which the lower wavefront moves to the position of the higher one, a point on the x axis with a particular constant phase moves from one dot to the other. This means that it is moving very fast (much faster than the wavefronts), consistent with the fact the ω/k_x is very large if k_x is very small. In the limit where the wavefronts are horizontal ($k_x = 0$), a point of constant phase moves infinitely fast along the x axis. The quantities ω/k_x, ω/k_y, and ω/k_z therefore *cannot* be thought of as components of the phase velocity in Eq. (27). The component of a vector should be smaller than the vector itself, after all.

The vector that *does* correctly break up into components is the wavevector $\mathbf{k} = (k_x, k_y, k_z)$. Its magnitude $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ represents how much the phase of the wave increases in each unit distance along the $\hat{\mathbf{k}}$ direction. (In other words, it equals 2π times the number of wavelengths that fit into a unit distance.) And k_x represents how much the phase of the wave increases in each unit distance along the x direction. This is less than k, as it should be, in view of Fig. 4. For a given distance along the x axis, the phase advances by only a small amount, compared with along the $\mathbf{k}$ vector. The phase needs the entire distance between the two dots to increase by 2π along the x axis, whereas it needs only the distance between the wavefronts to increase by 2π along the $\mathbf{k}$ vector. ♣

8.3.2 Further constraints due to Maxwell's equations



(impossible **E, B** vectors)

Figure 5

Fig. 3 tells us only that points along a given line have common values of $\mathbf{E}$ and $\mathbf{B}$. It doesn't tell us what these values actually are, or if they are constrained in other ways. For all we know, $\mathbf{E}$ and $\mathbf{B}$ on a particular wavefront might look like the vectors shown in Fig. 5 (we have ignored any possible z components). But it turns out the these vectors aren't actually possible. Although they satisfy the wave equation, they don't satisfy Maxwell's equations. So let's now see how Maxwell's equations further constrain the form of the waves. Later on in Section 8.8, we'll see that the waves are even further constrained by any boundary conditions that might exist. We'll look at Maxwell's equations in order and see what each of them implies.

- Using the expression for $\mathbf{E}$ in Eq. (26), the first of Maxwell's equations, Eq. (8), gives

$$\begin{aligned} \nabla \cdot \mathbf{E} = 0 \quad &\Longrightarrow \quad \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0 \\ &\Longrightarrow \quad ik_x E_x + ik_y E_y + ik_z E_z = 0 \\ &\Longrightarrow \quad \boxed{\mathbf{k} \cdot \mathbf{E} = 0} \end{aligned} \tag{28}$$

  This says that $\mathbf{E}$ is always perpendicular to $\mathbf{k}$. As we see from the second line here, each partial derivative simply turns into a factor if ik_x, etc.

- The second of Maxwell's equations, Eq. (9), gives the analogous result for $\mathbf{B}$, namely,

$$\boxed{\mathbf{k} \cdot \mathbf{B} = 0} \tag{29}$$

  So $\mathbf{B}$ is also perpendicular to $\mathbf{k}$.

- Again using the expression for $\mathbf{E}$ in Eq. (26), the third of Maxwell's equations, Eq. (10), gives

$$\begin{aligned} \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} &\implies \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \\ &\implies (ik_x, ik_y, ik_z) \times \mathbf{E} = -(-i\omega)\mathbf{B} \\ &\implies \boxed{\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}} \end{aligned} \tag{30}$$

  Since the cross product of two vectors is perpendicular to each of them, this result says that $\mathbf{B}$ is perpendicular to $\mathbf{E}$. And we already know that $\mathbf{B}$ is perpendicular to $\mathbf{k}$, from the second of Maxwell's equations. But technically we didn't need to use that equation, because the $\mathbf{B} \perp \mathbf{k}$ result is also contained in this $\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}$ result. Note that as above with the divergences, each partial derivative in the curl simply turns into a factor if ik_x, etc.

  We know from the first of Maxwell's equations that $\mathbf{E}$ is perpendicular to $\mathbf{k}$, so the magnitude of $\mathbf{k} \times \mathbf{E}$ is simply $|\mathbf{k}||\mathbf{E}| \equiv kE$. The magnitude of the $\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}$ relation then tells us that

$$kE = \omega B \quad \Longrightarrow \quad E = \frac{\omega}{k} B \quad \Longrightarrow \quad \boxed{E = cB} \tag{31}$$

  Therefore, the magnitudes of $\mathbf{E}$ and $\mathbf{B}$ are related by a factor of the wave speed, c. Eq. (31) is very useful, but its validity is limited to a single traveling wave, because the derivation of Eq. (30) assumed a unique $\mathbf{k}$ vector. If we form the sum of two waves with different $\mathbf{k}$ vectors, then the sum doesn't satisfy Eq. (30) for any particular vector $\mathbf{k}$. There isn't a unique $\mathbf{k}$ vector associated with the wave. Likewise for Eqs. (28) and (29).

- The fourth of Maxwell's equations, Eq. (11), can be written as $\nabla \times \mathbf{B} = (1/c^2)\partial \mathbf{E}/\partial t$, so the same procedure as above yields

$$\boxed{\mathbf{k} \times \mathbf{B} = -\frac{\omega}{c^2} \mathbf{E}} \quad \Longrightarrow \quad B = \frac{\omega}{kc^2} E \quad \Longrightarrow \quad \boxed{B = \frac{E}{c}} \tag{32}$$

  This doesn't tell us anything new, because we already know that $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{k}$ are all mutually perpendicular, and also that $E = cB$. In retrospect, the first and third (or alternatively the second and fourth) of Maxwell's equations are sufficient to derive all of the above results, which can be summarized as

$$\boxed{\mathbf{E} \perp \mathbf{k}, \qquad \mathbf{B} \perp \mathbf{k}, \qquad \mathbf{E} \perp \mathbf{B}, \qquad E = cB} \tag{33}$$

If three vectors are mutually perpendicular, there are two possibilities for how they are oriented. With the conventions of $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{k}$ that we have used in Maxwell's equations and in the exponential solution in Eq. (24) (where the $\mathbf{k} \cdot \mathbf{r}$ term comes in with a plus sign), the orientation is such that $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{k}$ form a "righthanded" triplet. That is, $\mathbf{E} \times \mathbf{B}$ points in the same direction as $\mathbf{k}$ (assuming, of course, that you're defining the cross product with the righthand rule!). You can show that this follows from the $\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}$ relation in Eq. (30) by either simply drawing three vectors that satisfy Eq. (30), or by using the determinant definition of the cross product to show that a cyclic permutation of the vectors maintains the sign of the cross product.

A snapshot (for an arbitrary value of t) of a possible electromagnetic wave is shown in Fig. 6. We have chosen $\mathbf{k}$ to point along the z axis, and we have drawn the field only for

points on the z axis. But for a given value z_0, all points of the form (x, y, z_0), which is a plane perpendicular to the z axis, have common values of $\mathbf{E}$ and $\mathbf{B}$. $\mathbf{E}$ points in the $\pm x$ direction, and $\mathbf{B}$ points in the $\pm y$ direction. As time goes by, the whole figure simply slides along the z axis at speed c. Note that $\mathbf{E}$ and $\mathbf{B}$ reach their maximum and minimum values at the same locations. We will find below that this isn't the case for standing waves.

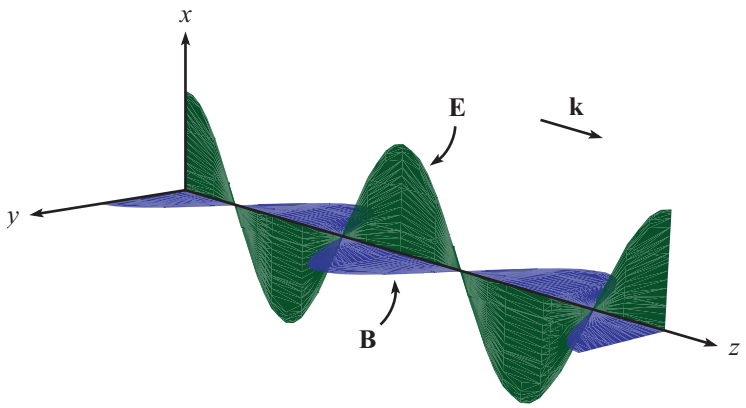


Figure 6

What are the mathematical expressions for the $\mathbf{E}$ and $\mathbf{B}$ fields in Fig. 6? We've chosen $\mathbf{k}$ to point along the z axis, so we have $\mathbf{k} = k\hat{\mathbf{z}}$, which gives $\mathbf{k} \cdot \mathbf{r} = kz$. And since $\mathbf{E}$ points in the x direction, its amplitude takes the form of $E_0 e^{i\phi}\hat{\mathbf{x}}$. (The coefficient can be complex, and we have written it as a magnitude times a phase.) This then implies that $\mathbf{B}$ points in the y direction (as drawn), because it must be perpendicular to both $\mathbf{E}$ and $\mathbf{k}$. So its amplitude takes the form of $B_0 e^{i\phi}\hat{\mathbf{y}} = (E_0 e^{i\phi}/c)\hat{\mathbf{y}}$. This is the same phase ϕ, due to Eq. (30) and the fact that $\mathbf{k}$ is real, at least for simple traveling waves. The desired expressions for $\mathbf{E}$ and $\mathbf{B}$ are obtained by taking the real part of Eq. (26), so we arrive at

$$\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(kz - \omega t + \phi), \qquad \text{and} \qquad \mathbf{B} = \hat{\mathbf{y}} \frac{E_0}{c} \cos(kz - \omega t + \phi), \tag{34}$$

These two vectors are in phase with each other, consistent with Fig. 6. And $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{k}$ form a righthanded triple of vectors, as required.

Remarks:

1. When we talk about polarization in Section 8.6, we will see that $\mathbf{E}$ and $\mathbf{B}$ don't have to point in specific directions, as they do in Fig. 6, where $\mathbf{E}$ points only along $\hat{\mathbf{x}}$ and $\mathbf{B}$ points only along $\hat{\mathbf{y}}$. Fig. 6 happens to show the special case of "linear polarization."
2. The $\mathbf{E}$ and $\mathbf{B}$ waves don't have to be sinusoidal, of course. Because the wave equation is linear, we can build up other solutions from sinusoidal ones. And because the wave equation is dispersionless, we know (as we saw at the end of Section 2.4) that any function of the form $f(z - vt)$, or equivalently $f(kz - \omega t)$, satisfies the wave equation. But the restrictions placed by Maxwell's equations still hold. In particular, the $\mathbf{E}$ field determines the $\mathbf{B}$ field.
3. A static solution, where $\mathbf{E}$ and $\mathbf{B}$ are constant, can technically be thought of as a sinusoidal solution in the limit where $\omega = k = 0$. In vacuum, we can always add on a constant field to $\mathbf{E}$ or $\mathbf{B}$, and it won't affect Maxwell's equations (and therefore the wave equation either), because all of the terms in Maxwell's equations in vacuum involve derivatives (either space or time). But we'll ignore any such fields, because they're boring for the purposes we'll be concerned with. ♣

8.3.3 Standing waves

Let's combine two waves (with equal amplitudes) traveling in opposite directions, to form a standing wave. If we add $\mathbf{E}_1 = \hat{\mathbf{x}}E_0\cos(kz - \omega t)$ and $\mathbf{E}_2 = \hat{\mathbf{x}}E_0\cos(-kz - \omega t)$, we obtain

$$\mathbf{E} = \hat{\mathbf{x}}(2E_0)\cos kz\cos\omega t. \tag{35}$$

This is indeed a standing wave, because all z values have the same phase with respect to time.

There are various ways to find the associated $\mathbf{B}$ wave. Actually, there are (at least) two right ways and one wrong way. The wrong way is to use the result in Eq. (30) to say that $\omega\mathbf{B} = \mathbf{k}\times\mathbf{E}$. This would yield the result that $\mathbf{B}$ is proportional to $\cos kz\cos\omega t$, which we will find below is incorrect. The error (as we mentioned above after Eq. (31)) is that there isn't a unique $\mathbf{k}$ vector associated with the wave in Eq. (35), because it is generated by two waves with opposite $\mathbf{k}$ vectors. If we insisted on using Eq. (30), we'd be hard pressed to decide if we wanted to use $k\hat{\mathbf{z}}$ or $-k\hat{\mathbf{z}}$ as the $\mathbf{k}$ vector.

A valid method for finding $\mathbf{B}$ is the following. We can find the traveling $\mathbf{B}_1$ and $\mathbf{B}_2$ waves associated with each of the traveling $\mathbf{E}_1$ and $\mathbf{E}_2$ waves, and then add them. You can quickly show (using $\mathbf{B} = (1/\omega)\mathbf{k}\times\mathbf{E}$ for each traveling wave separately) that $\mathbf{B}_1 = \hat{\mathbf{y}}(E_0/c)\cos(kz - \omega t)$ and $\mathbf{B}_2 = -\hat{\mathbf{y}}(E_0/c)\cos(-kz - \omega t)$. The sum of these waves give the desired associated $\mathbf{B}$ field,

$$\mathbf{B} = \hat{\mathbf{y}}(2E_0/c)\sin kz\sin\omega t. \tag{36}$$

Another method is to use the third of Maxwell's equations, Eq. (10), which says that $\nabla\times\mathbf{E} = -\partial\mathbf{B}/\partial t$. Maxwell's equations hold for *any* $\mathbf{E}$ and $\mathbf{B}$ fields. We don't have to worry about the uniqueness of $\mathbf{k}$ here. Using the $\mathbf{E}$ in Eq. (35), the cross product $\nabla\times\mathbf{E}$ can be calculated with the determinant:

$$\begin{aligned}\nabla\times\mathbf{E} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ E_x & 0 & 0 \end{vmatrix} &= \hat{\mathbf{y}}\frac{\partial E_x}{\partial z} - \hat{\mathbf{z}}\frac{\partial E_x}{\partial y} \\ &= -\hat{\mathbf{y}}(2E_0)k\sin kz\cos\omega t - 0. \end{aligned} \tag{37}$$

Eq. (30) tells us that this must equal $-\partial\mathbf{B}/\partial t$, so we conclude that

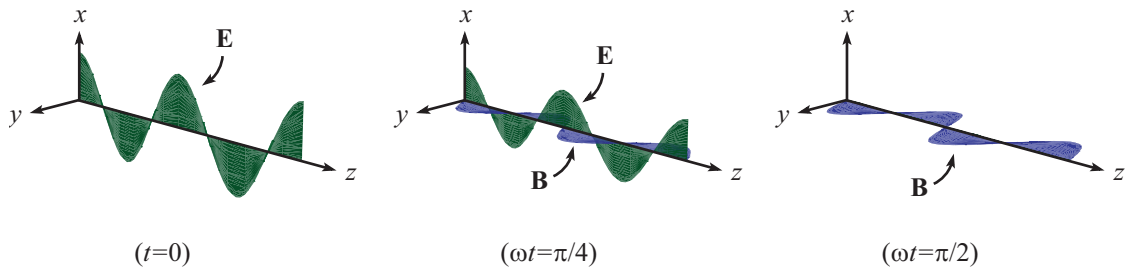
$$\mathbf{B} = \hat{\mathbf{y}}(2E_0)(k/\omega)\sin kz\sin\omega t = \hat{\mathbf{y}}(2E_0/c)\sin kz\sin\omega t. \tag{38}$$

in agreement with Eq. (36). We have ignored any possible additive constant in $\mathbf{B}$.

Having derived the associated $\mathbf{B}$ field in two different ways, we can look at what we've found. $\mathbf{E}$ and $\mathbf{B}$ are still perpendicular to each other, which makes since, because $\mathbf{E}$ is the superposition of two vectors that point in the $\pm\hat{\mathbf{x}}$ direction, and $\mathbf{E}$ is the superposition of two vectors that point in the $\pm\hat{\mathbf{y}}$ direction. But there is a major difference between standing waves and traveling waves. In traveling waves, $\mathbf{E}$ and $\mathbf{B}$ run along in step with each other, as shown above in Fig. 6. They reach their maximum and minimum values at the same times and positions. However, in standing waves $\mathbf{E}$ is maximum *when* $\mathbf{B}$ is zero, and also *where* $\mathbf{B}$ is zero (and vice versa). $\mathbf{E}$ and $\mathbf{B}$ are 90° out of phase with each other in both time and space. That is, the $\mathbf{B}$ in Eq. (36) can be written as

$$\mathbf{B} = \hat{\mathbf{y}}\frac{2E_0}{c}\cos\left(kz - \frac{\pi}{2}\right)\cos\left(\omega t - \frac{\pi}{2}\right), \tag{39}$$

which you can compare with the $\mathbf{E}$ in Eq. (35). A few snapshots of the $\mathbf{E}$ and $\mathbf{B}$ waves are shown in Fig. 7.



(t=0) (ωt=π/4) (ωt=π/2)

Figure 7

8.4 Energy

8.4.1 The Poynting vector

The energy density of an electromagnetic field is

$$\mathcal{E} = \frac{\epsilon_0}{2}E^2 + \frac{1}{2\mu_0}B^2, \tag{40}$$

where $E \equiv |\mathbf{E}|$ and $B \equiv |\mathbf{B}|$ are the magnitudes of the fields at a given location in space and time. We have suppressed the (x, y, z, t) arguments of $\mathcal{E}$, E, and B. This energy density can be derived in various ways (see Problem [to be added]), but we'll just accept it here. The goal of this section is to calculate the rate of change of $\mathcal{E}$, and to then write it in a form that allows us to determine the energy flux (the flow of energy across a given surface). We will find that the energy flux is given by the so-called *Poynting vector*.

If we write E^2 and B^2 as $\mathbf{E}\cdot\mathbf{E}$ and $\mathbf{B}\cdot\mathbf{B}$, then the rate of change of $\mathcal{E}$ becomes

$$\frac{\partial \mathcal{E}}{\partial t} = \epsilon_0 \mathbf{E}\cdot\frac{\partial \mathbf{E}}{\partial t} + \frac{1}{\mu_0}\mathbf{B}\cdot\frac{\partial \mathbf{B}}{\partial t}. \tag{41}$$

(The product rule works here for the dot product of vectors for the same reason it works for a regular product. You can verify this by explicitly writing out the components.) The third and fourth Maxwell's equations turn this into

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial t} &= \epsilon_0 \mathbf{E}\cdot\left(\frac{1}{\mu_0\epsilon_0}\nabla\times\mathbf{B}\right) + \frac{1}{\mu_0}\mathbf{B}\cdot\big(-\nabla\times\mathbf{E}\big) \\ &= \frac{1}{\mu_0}\Big(\mathbf{E}\cdot(\nabla\times\mathbf{B}) - \mathbf{B}\cdot(\nabla\times\mathbf{E})\Big). \end{aligned} \tag{42}$$

The righthand side of this expression conveniently has the same form as the righthand side of the vector identity (see Problem [to be added] for the derivation),

$$\nabla\cdot(\mathbf{C}\times\mathbf{D}) = \mathbf{D}\cdot(\nabla\times\mathbf{C}) - \mathbf{C}\cdot(\nabla\times\mathbf{D}). \tag{43}$$

So we now have

$$\frac{\partial \mathcal{E}}{\partial t} = \frac{1}{\mu_0}\nabla\cdot(\mathbf{B}\times\mathbf{E}). \tag{44}$$

Now consider a given volume V in space. Integrating Eq. (44) over this volume V yields

$$\int_V \frac{\partial \mathcal{E}}{\partial t} = \frac{1}{\mu_0}\int_V \nabla\cdot(\mathbf{B}\times\mathbf{E}) \quad\Longrightarrow\quad \frac{\partial W_V}{\partial t} = \frac{1}{\mu_0}\int_A (\mathbf{B}\times\mathbf{E})\cdot d\mathbf{A}, \tag{45}$$

where W_V is the energy contained in the volume V (we've run out of forms of the letter E), and where we have used the divergence theorem to rewrite the volume integral as a surface integral over the area enclosing the volume. $d\mathbf{A}$ is defined to be the vector perpendicular to the surface (with the positive direction defined to be outward), with a magnitude equal to the area of a little patch.

Let's now make a slight change in notation. $d\mathbf{A}$ is defined to be an outward-pointing vector, but let's define $d\mathbf{A}_{\text{in}}$ to be the inward-pointing vector, $d\mathbf{A}_{\text{in}} \equiv -d\mathbf{A}$. Eq. (45) can then be written as (switching the order of $\mathbf{E}$ and $\mathbf{B}$)

$$\frac{\partial W_V}{\partial t} = \frac{1}{\mu_0}\int_A (\mathbf{E}\times\mathbf{B})\cdot d\mathbf{A}_{\text{in}}. \tag{46}$$

We can therefore interpret the vector,

$$\boxed{\mathbf{S} \equiv \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}} \qquad (\textit{energy flux}: \text{ energy}/(\text{area}\cdot\text{time})) \tag{47}$$

as giving the flux of energy *into* a region. This vector $\mathbf{S}$ is known as the *Poynting vector*. And since $\mathbf{E}\times\mathbf{B} \propto \mathbf{k}$, the Poynting vector points in the same direction as the velocity of the wave. Integrating $\mathbf{S}$ over any surface (or rather, just the component perpendicular to the surface, due to the dot product with $d\mathbf{A}_{\text{in}}$) gives the energy flow across the surface. This result holds for any kind of wave – traveling, standing, or whatever. Comparing the units on both sides of Eq. (46), we see that the Poynting vector has units of energy per area per time. So if we multiply it (or its perpendicular component) by an area, we get the energy per time crossing the area.

The Poynting vector falls into a wonderful class of phonetically accurate theorems/results. Others are the Low energy theorem (named after S.Y. Low) dealing with low-energy photons, and the Schwarzschild radius of a black hole (kind of like a shield).

8.4.2 Traveling waves

Let's look at the energy density $\mathcal{E}$ and the Poynting vector $\mathbf{S}$ for a traveling wave. A traveling wave has $B = E/c$, so the energy density is (using $c^2 = 1/\mu_0\epsilon_0$ in the last step)

$$\mathcal{E} = \frac{\epsilon_0}{2}E^2 + \frac{1}{2\mu_0}B^2 = \frac{\epsilon_0}{2}E^2 + \frac{1}{2\mu_0}\frac{E^2}{c^2} \quad\Longrightarrow\quad \boxed{\mathcal{E} = \epsilon_0 E^2} \tag{48}$$

We have suppressed the (x, y, z, t) arguments of $\mathcal{E}$ and E. Note that this result holds only for traveling waves. A standing wave, for example, doesn't have $B = E/c$ anywhere, so $\mathcal{E}$ doesn't take this form. We'll discuss standing waves below.

The Poynting vector for a traveling wave is

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} = \frac{1}{\mu_0}E\left(\frac{E}{c}\right)\hat{\mathbf{k}}, \tag{49}$$

where we have used the facts that $\mathbf{E}\perp\mathbf{B}$ and that their cross product points in the direction of $\mathbf{k}$. Using $1/\mu_0 = c^2\epsilon_0$, arrive at

$$\boxed{\mathbf{S} = c\epsilon_0 E^2\hat{\mathbf{k}}} = c\mathcal{E}\hat{\mathbf{k}}. \tag{50}$$

This last equality makes sense, because the energy density $\mathcal{E}$ moves along with the wave, which moves at speed c. So the energy per unit area per unit time that crosses a surface

is $c\mathcal{E}$ (which you can verify has the correct units). Eq. (50) is true at all points (x, y, z, t) individually, and not just in an average sense. We'll derive formulas for the averages below.

In the case of a sinusoidal traveling wave of the form,

$$\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(kz - \omega t) \qquad \text{and} \qquad \mathbf{B} = \hat{\mathbf{y}}(E_0/c)\cos(kz - \omega t), \tag{51}$$

the above expressions for $\mathcal{E}$ and $\mathbf{S}$ yield

$$\mathcal{E} = \epsilon_0 E_0^2 \cos^2(kz - \omega t) \qquad \text{and} \qquad \mathbf{S} = c\epsilon_0 E_0^2 \cos^2(kz - \omega t)\hat{\mathbf{k}}. \tag{52}$$

Since the average value of $\cos^2(kz - \omega t)$ over one period (in either space or time) is $1/2$, we see that the average values of $\mathcal{E}$ and $|\mathbf{S}|$ are

$$\mathcal{E}_{\text{avg}} = \frac{1}{2}\epsilon_0 E_0^2 \qquad \text{and} \qquad |\mathbf{S}|_{\text{avg}} = \frac{1}{2} c\epsilon_0 E_0^2. \tag{53}$$

$|\mathbf{S}|_{\text{avg}}$ is known as the *intensity* of the wave. It is the average amount of energy per unit area per unit time that passes through (or hits) a surface. For example, at the location of the earth, the radiation from the sun has an intensity of 1360 Watts/m^2. The energy comes from traveling waves with many different frequencies, and the total intensity is just the sum of the intensities of the individual waves (see Problem [to be added]).

8.4.3 Standing waves

Consider the standing wave in Eqs. (35) and (36). With $2E_0$ defined to be A, this wave becomes

$$\mathbf{E} = \hat{\mathbf{x}} A \cos kz \cos \omega t, \qquad \text{and} \qquad \mathbf{B} = \hat{\mathbf{y}}(A/c)\sin kz \sin \omega t. \tag{54}$$

The energy density in Eq. (48) for a traveling wave isn't valid here, because it assumed $B = E/c$. Using the original expression in Eq. (40), the energy density for the above standing wave is (using $1/c^2 = \mu_0\epsilon_0$)

$$\mathcal{E} = \frac{\epsilon_0}{2}E^2 + \frac{1}{2\mu_0}B^2 = \frac{1}{2}\epsilon_0 A^2\big(\cos^2 kz \cos^2 \omega t + \sin^2 kz \sin^2 \omega t\big). \tag{55}$$

If we take the average over a full cycle in time (a full wavelength in space would work just as well), then the $\cos^2 \omega t$ and $\sin^2 \omega t$ factors turn into $1/2$'s, so the time average of $\mathcal{E}$ is

$$\mathcal{E}_{\text{avg}} = \frac{1}{2}\epsilon_0 A^2 \left(\frac{1}{2}\cos^2 kz + \frac{1}{2}\sin^2 kz\right) = \frac{\epsilon_0 A^2}{4}\,, \tag{56}$$

which is independent of z. It makes sense that it doesn't depend on z, because a traveling wave has a uniform average energy density, and a standing wave is just the sum of two traveling waves moving in opposite directions.

The Poynting vector for our standing wave is given by Eq. (47) as (again, the traveling-wave result in Eq. (50) isn't valid here):

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B} = \frac{A^2}{\mu_0 c}\hat{\mathbf{k}} \cos kz \sin kz \cos \omega t \sin \omega t. \tag{57}$$

At any given value of z, the time average of this is zero (because $\cos \omega t \sin \omega t = (1/2)\sin 2\omega t$), so there is no net energy flow in a standing wave. This makes sense, because a standing wave is made up of two traveling waves moving in opposite directions which therefore have opposite energy flows (on average). Similarly, for a given value of t, the spatial average is zero. Energy sloshes back and forth between points, but there is no net flow.

Due to the fact that a standing wave is made up of two traveling waves moving in opposite directions, you might think that the Poynting vector (that is, the energy flow) should be *identically* equal to zero, for all z and t. But it isn't, because each of the two Poynting vectors depends on z and t, so only at certain discrete times and places do they anti-align and exactly cancel each other. But on average they cancel.

8.5 Momentum

Electromagnetic waves carry momentum. However, all the other waves we've studied (longitudinal spring/mass and sound waves, transverse string waves, etc.) *don't* carry momentum. (However, see Footnote 1 above.) Therefore, it is certainly not obvious that electromagnetic waves carry momentum, because it is quite possible for waves to carry energy without also carrying momentum.

A quick argument that demonstrates why an electromagnetic wave (that is, light) carries momentum is the following argument from relativity. The relativistic relation between a particle's energy, momentum, and mass is $E^2 = p^2c^2 + m^2c^4$ (we'll just accept this here). For a massless particle ($m = 0$), this yields $E^2 = p^2c^2 \Longrightarrow E = pc$. Since photons (which is what light is made of) are massless, they have a momentum given by $p = E/c$. We already know that electromagnetic waves carry energy, so this relation tells us that they must also carry momentum. In other words, a given part of an electromagnetic wave with energy E also has momentum $p = E/c$.

However, although this argument is perfectly valid, it isn't very satisfying, because (a) it invokes a result from relativity, and (b) it invokes the fact that electromagnetic waves (light) can be considered to be made up of particle-like objects called photons, which is by no means obvious. But why should the *particle* nature of light be necessary to derive the fact that an electromagnetic *wave* carries momentum? It would be nice to derive the $p = E/c$ result by working only in terms of waves and using only the results that we have developed so far in this book. Or said in a different way, it would be nice to understand how would someone living in, say, 1900 (that is, pre-relativity) would demonstrate that an electromagnetic waves carries momentum. We can do this in the following way.

Consider a particle with charge q that is free to move around in some material, and let it be under the influence of a traveling electromagnetic wave. The particle will experience forces due to the **E** and **B** fields that make up the wave. There will also be damping forces from the material. And the particle will also lose energy due to the fact that it is accelerating and hence radiating (see Section 8.7). But the exact nature of the effects of the damping and radiation won't be important for this discussion.[3]

Assume that the wave is traveling in the z direction, and let the **E** field point along the x direction. The **B** field then points along the y direction, because $\mathbf{E} \times \mathbf{B} \propto \mathbf{k}$. The complete motion of the particle will in general be quite complicated, but for the present purposes it suffices to consider the x component of the particle's velocity, that is, the component that is parallel to **E**.[4] Due to the oscillating electric field, the particle will (mainly) oscillate back and forth in the x direction. However, we don't know the phase. In general, part of the velocity will be in phase with **E**, and part will be $\pm 90^\circ$ out of phase. The latter will turn out not to matter for our purposes,[5] so we'll concentrate on the part of the velocity that is

[3] If the particle is floating in outer space, then there is no damping, so only the radiation will extract energy from the particle.

[4] If we assume that the velocity v of the particle satisfies $v \ll c$ (which is generally a good approximation), then the magnetic force, $q\mathbf{v} \times \mathbf{B}$ is small compared with the electric force, $q\mathbf{E}$. This is true because $B = E/c$, so the magnetic force is suppressed by a factor of v/c (or more, depending on the angle between **v** and **B**) compared with the electric force. The force on the particle is therefore due mainly to the electric field.

[5] We'll be concerned with the work done by the electric field, and this part of the velocity will lead to

in phase with **E**. Let's call it $\mathbf{v}_{\rm E}$. We then have the pictures shown in Fig. 8.

You can quickly verify with the righthand rule that the magnetic force $q\mathbf{v}_{\rm E} \times \mathbf{B}$ points forward along **k** in *both* cases. $\mathbf{v}_{\rm E}$ and **B** switch sign in phase with each other, so the two signs cancel, and there is a net force forward. The particle therefore picks up some forward momentum, and this momentum must have come from the wave. In a small time dt, the magnitude of the momentum that the wave gives to the particle is

$$|d\mathbf{p}| = |\mathbf{F}_{\rm B}\,dt| = |q\mathbf{v}_{\rm E} \times \mathbf{B}|\,dt = qv_{\rm E}B\,dt = \frac{qv_{\rm E}E\,dt}{c}\,. \tag{58}$$

What is the energy that the wave gives to the particle? That is, what is the work that the wave does on the particle? (In the steady state, this work is balanced, on average, by the energy that the particle loses to damping and radiation.) Only the electric field does work on the particle. And since the electric force is qE, the amount of work done on the particle in time dt is

$$dW = \mathbf{F}_{\rm E} \cdot d\mathbf{x} = (qE)(v_{\rm E}\,dt) = qv_{\rm E}E\,dt. \tag{59}$$

(The part of the velocity that is $\pm 90^\circ$ out of phase with **E** will lead to zero net work, on average; see Problem [to be added].) Comparing this result with Eq. (58), we see that

$$|d\mathbf{p}| = \frac{dW}{c}\,. \tag{60}$$

In other words, the amount of momentum the particle gains from the wave equals $1/c$ times the amount of energy it gains from the wave. This holds for any extended time interval Δt, because any interval can be built up from infinitesimal times dt.

Since Eq. (60) holds whenever any electromagnetic wave encounters a particle, we conclude that the wave actually carries this amount of momentum. Even if we didn't have a particle in the setup, we could imagine putting one there, in which case it would acquire the momentum given by Eq. (60). This momentum must therefore be an intrinsic property of the wave.

Another way of writing Eq. (60) is

$$\frac{1}{A}\left|\frac{d\mathbf{p}}{dt}\right| = \frac{1}{c} \cdot \frac{1}{A}\frac{dW}{dt}\,, \tag{61}$$

where A is the cross-sectional area of the wave under consideration. The lefthand side is the force per area (in other words, the pressure) that the wave applies to a material. And from Eqs. (46) and (47), the righthand side is $|\mathbf{S}|/c$, where **S** is the Poynting vector. The pressure from an electromagnetic wave (usually called the *radiation pressure*) is therefore

$$\text{Radiation pressure} = \frac{|\mathbf{S}|}{c} = \frac{|\mathbf{E} \times \mathbf{B}|}{\mu_0 c} = \frac{E^2}{\mu_0 c^2}\,. \tag{62}$$

You can show (see Problem [to be added]) that the total force from the radiation pressure from sunlight hitting the earth is roughly $6{\cdot}10^8\,\text{kg}\,\text{m/s}^2$ (treating the earth like a flat coin and ignoring reflection, but these won't affect the order of magnitude). This force is negligible compared with the attractive gravitational force, which is about $3.6 \cdot 10^{22}\,\text{kg}\,\text{m/s}^2$. But for a small enough sphere, these two forces are comparable (see Problem [to be added]).

Electromagnetic waves also carry angular momentum *if* they are polarized (see Problem [to be added]).

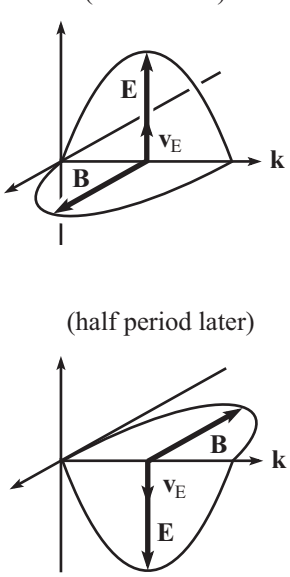


Figure 8

zero net work being done.

8.6 Polarization

8.6.1 Linear polarization

Consider the traveling wave in Eq. (34) (we'll ignore the overall phase ϕ here):

$$\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(kz - \omega t), \qquad \text{and} \qquad \mathbf{B} = \hat{\mathbf{y}} \frac{E_0}{c} \cos(kz - \omega t). \tag{63}$$

This wave has $\mathbf{E}$ always pointing in the x direction and $\mathbf{B}$ always pointing in the y direction. A wave like this, where the fields always point along given directions, is called a *linearly polarized* wave. The direction of the linear polarization is defined to be the axis along which the $\mathbf{E}$ field points.

But what if we want to construct a wave where the fields don't always point along given directions? For example, what if we want the $\mathbf{E}$ vector to rotate around in a circle instead of oscillating back and forth along a line?

Let's try making such a wave by adding on an $\mathbf{E}$ field (with the same magnitude) that points in the y direction. The associated $\mathbf{B}$ field then points in the negative x direction if we want the orientation to be the same so that the wave still travels in the same direction (that is, so that the $\hat{\mathbf{k}}$ vector still points in the $+\hat{\mathbf{z}}$ direction). The total wave is now

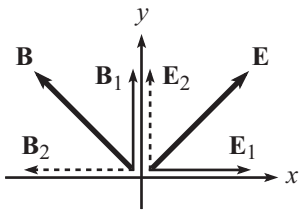


Figure 9

$$\mathbf{E} = (\hat{\mathbf{x}} + \hat{\mathbf{y}}) E_0 \cos(kz - \omega t), \qquad \text{and} \qquad \mathbf{B} = (\hat{\mathbf{y}} - \hat{\mathbf{x}}) \frac{E_0}{c} \cos(kz - \omega t). \tag{64}$$

If the two waves we added are labeled as "1" and "2" respectively, then the sum given in Eq. (64) is shown in Fig. 9. The wave travels in the positive z direction, which is out of the page. The $\mathbf{E}$ field in this wave always points along the (positive or negative) diagonal $\hat{\mathbf{x}} + \hat{\mathbf{y}}$ direction, and the $\mathbf{B}$ field always points along the $\hat{\mathbf{y}} - \hat{\mathbf{x}}$ direction. So we still have a linearly polarized wave. All we've done is rotate the fields by 45° and multiply the amplitudes by $\sqrt{2}$. Therefore, if our goal is to produce a wave that isn't linearly polarized (that is, to produce a wave where the directions of $\mathbf{E}$ and $\mathbf{B}$ change), we're going to have to come up with a more clever method than adding on fields that point in different directions.

Before proceeding further, we should note that no matter what traveling wave we have, the $\mathbf{B}$ field is completely determined by the $\mathbf{E}$ field and the $\mathbf{k}$ vector via Eq. (30). For a given $\mathbf{k}$ vector (we'll generally pick $\mathbf{k}$ to point along $\hat{\mathbf{z}}$), this means that the $\mathbf{E}$ field determines the $\mathbf{B}$ field (and vice versa). So we won't bother writing down the $\mathbf{B}$ field anymore. We'll just work with the $\mathbf{E}$ field.

We should also note that the x and y components of $\mathbf{B}$ are determined *separately* by the y and x components of $\mathbf{E}$, respectively. This follows from Eq. (30) and the properties of the cross product:

$$\begin{aligned} \mathbf{k} \times \mathbf{E} = \omega \mathbf{B} \quad &\Longrightarrow \quad (k\hat{\mathbf{z}}) \times (E_x \hat{\mathbf{x}} + E_y \hat{\mathbf{y}}) = B_x \hat{\mathbf{x}} + B_y \hat{\mathbf{y}} \\ &\Longrightarrow \quad k E_x \hat{\mathbf{y}} + k E_y (-\hat{\mathbf{x}}) = B_x \hat{\mathbf{x}} + B_y \hat{\mathbf{y}} \\ &\Longrightarrow \quad B_x = -k E_y, \quad \text{and} \quad B_y = k E_x. \end{aligned} \tag{65}$$

We see that the B_x and E_y pair of components is "decoupled" from the B_y and E_x pair. The two pairs have nothing to do with each other. Each pair can be doing whatever it feels like, independent of the other pair. So we basically have two independent electromagnetic waves. This is the key to understanding polarization.

8.6.2 Circular and elliptical polarization

Our above attempt at forming a non-linearly polarized wave failed because the two $\mathbf{E}$ fields that we added together had the *same phase*. This resulted in the two pairs of components

(E_x and B_y, and E_y and B_x) having the same phase, which in turn resulted in a simple (tilted) line for each of the total **E** and **B** fields.

Let us therefore try some different relative phases between the components. As mentioned above, from here on we'll write down only the **E** field. The **B** field can always be obtained from Eq. (30). Let's add a phase of, say, $\pi/2$ to the y component of **E**. As above, we'll have the magnitudes of the components be equal, so we obtain

$$\begin{aligned} \mathbf{E} &= \hat{\mathbf{x}} E_0 \cos(kz - \omega t) + \hat{\mathbf{y}} E_0 \cos(kz - \omega t + \pi/2) \\ &= \hat{\mathbf{x}} E_0 \cos(kz - \omega t) - \hat{\mathbf{y}} E_0 \sin(kz - \omega t). \end{aligned} \tag{66}$$

What does **E** look like as a function of time, for a given value of z? We might as well pick $z = 0$ for simplicity, in which case we have (using the facts that cosine and sine are even and odd functions, respectively)

$$\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(\omega t) + \hat{\mathbf{y}} E_0 \sin(\omega t). \tag{67}$$

This is the expression for a vector with magnitude E_0 that swings around in a counterclockwise circle in the x-y plane, as shown in Fig. 10. (And at all times, **B** is perpendicular to **E**.) This is our desired *circular polarization*. The phase difference of $\pi/2$ between the x and y components of **E** causes **E** to move in a circle, as opposed to simply moving back and forth along a line. This is consistent with the fact that the phase difference implies that E_x and E_y can't both be zero at the same time, which is a necessary property of linear polarization, because the vector passes through the origin after each half cycle.

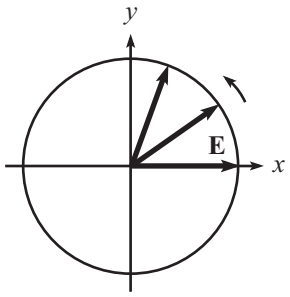


Figure 10

If we had chosen a phase of $-\pi/2$ instead of $\pi/2$, we would still have obtained circular polarization, but with the circle now being traced out in a clockwise sense (assuming that $\hat{\mathbf{z}}$ still points out of the page).

A phase of zero gives linear polarization, and a phase of $\pm\pi/2$ gives circular polarization. What about something in between? If we choose a phase of, say, $\pi/3$, then we obtain

$$\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(kz - \omega t) + \hat{\mathbf{y}} E_0 \cos(kz - \omega t + \pi/3) \tag{68}$$

As a function of time at $z = 0$, this takes the form,

$$\mathbf{E} = \hat{\mathbf{x}} E_0 \cos(\omega t) + \hat{\mathbf{y}} E_0 \cos(\omega t - \pi/3). \tag{69}$$

So the y component lags the x component by $\pi/3$. E_y therefore achieves its maximum value at a time given by $\omega t = \pi/3$ *after* E_x achieves its maximum value. A plot of **E** is shown in Fig. 11. A few of the points are labeled with their ωt values. This case is reasonably called *elliptical polarization*. The shape in Fig. 11 is indeed an ellipse, as you can show in Problem [to be added]. And as you can also show in this problem, the ellipse is always titled at $\pm 45^\circ$. The ellipse is very thin if the phase is near 0 or π, and it equals a diagonal line in either of these limits. Linear and circular polarization are special cases of elliptical polarization. If you want to produce a tilt angle other than $\pm 45^\circ$, you need to allow for E_x and E_y to have different amplitudes (see Problem [to be added]).

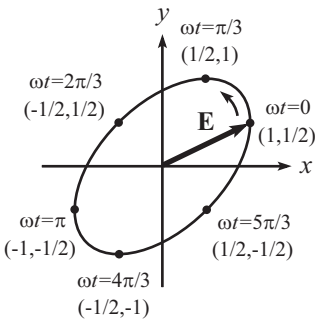


Figure 11

REMARK: We found above in Eq. (26) that the general solution for **E** is

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \tag{70}$$

In looking at this, it appears that the various components of **E** have the same phase. So were we actually justified in throwing in the above phases of $\pi/2$ and $\pi/3$, or anything else? Yes, because as we mentioned right after Eq. (26), the $\mathbf{E}_0$ vector (and likewise the $\mathbf{B}_0$ vector) doesn't have to be real. Each component can be complex and have an arbitrary phase (although the three phases in $\mathbf{B}_0$ are determined by the three phases phases in $\mathbf{E}_0$ by Maxwell's equations). For example, we

can have $E_{0,x} = A$ and $E_{0,y} = Ae^{i\phi}$. When we take the real part of the solution in Eq. (70), we then obtain

$$E_x = A\cos(\mathbf{k}\cdot\mathbf{r} - \omega t) \qquad \text{and} \qquad E_y = A\cos(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi). \tag{71}$$

So this is the source of the relative phase, which in turn is the source of the various types of polarizations. ♣

Standing waves can also have different types of polarizations. Such a wave can be viewed as the sum of polarized waves traveling in opposite directions. But there are different cases to consider, depending on the orientation of the polarizations; see Problem [to be added].

8.6.3 Wave plates

Certain anisotropic materials (that is, materials that aren't symmetric around a given axis) have the property that electromagnetic waves that are linearly polarized along one axis travel at a different speed from waves that are linearly polarized along another (perpendicular) axis. This effect is known as *birefringence*, or *double refraction*, because it has two different speeds and hence two different indices of refraction, n_x and n_y. The difference in speeds and n values arises from the difference in permittivity values, ϵ, in the two directions. This difference in speed implies, as we will see, that the electric field components in the two directions will gradually get out of phase as the wave travels through the material. If the thickness of the material is chosen properly, we can end up with a phase difference of, say, $\pi/2$ (or anything else) which implies circular polarization. Let's see how this phase difference arises.

Let the two transverse directions be x and y, and let the E_x wave travel faster than the E_y wave. As mentioned right after Eq. (65), we can consider these components to be two separate waves. Let's assume that linearly polarized light traveling in the z direction impinges on the material and that it has nonzero components in both the x and y directions. Since this single wave is driving both the E_x and E_y waves in the material, these two components will be in phase with each other at the front end of the material. The material is best described as a plate (hence the title of this subsection), because the dimension along the direction of the wave's motion is generally small compared with the other two dimensions.

What happens as the E_x and E_y waves propagate through the material? Since the same external waves is driving both the E_x and E_y waves in the material, the frequencies of these waves must be equal. However, since the speeds are different, the $\omega = vk$ relation tells us that the k values must be different. Equivalently, the relation $\lambda\nu = v$ tells us that (since ν is the same) the wavelength is proportional to the velocity. So a smaller speed means a shorter wavelength. A possible scenario is shown in Fig. 12. We have assumed that the y speed is slightly smaller than the x speed, which means that λ_y is slightly shorter than λ_x. Equivalently, k_y is slightly larger than k_x. Therefore, slightly more E_y waves fit in the material than E_x waves, as shown. What does this imply about the phase difference between the E_x and E_y waves when they exit the material?

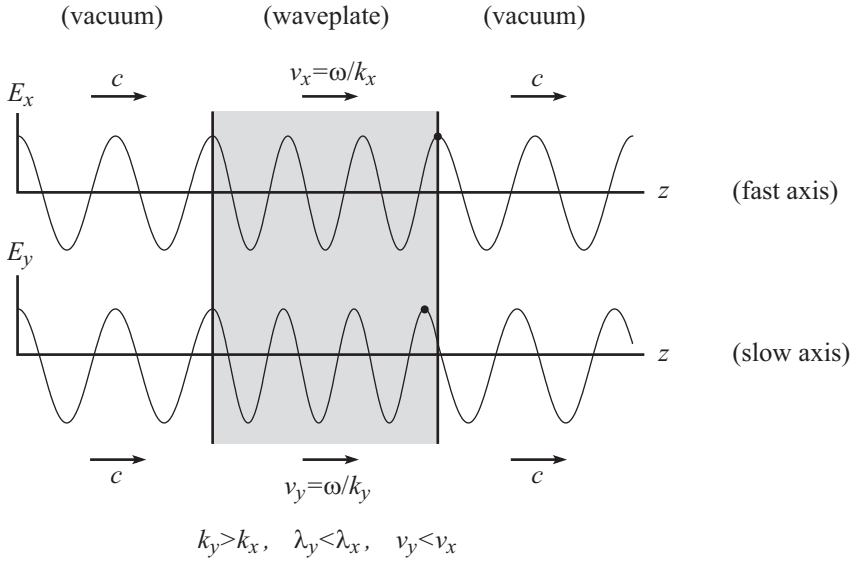


Figure 12

At the instant shown, the E_x field at the far end of the material has reached its maximum value, indicated by the dot shown. (It isn't necessary for an integral number of wavelengths to fit into the material, but it makes things a little easier to visualize.) But E_y hasn't reached its maximum quite yet. The E_y wave needs to travel a little more to the right before the crest marked by the dot reaches the far end of the material. So the phase of the E_y wave at the far end is slightly *behind* the phase of the E_x wave.[6]

By how much is the E_y phase behind the E_x phase at the far end? We need to find the phase of the E_y wave that corresponds to the extra little distance between the dots shown in Fig. 12. Let the length of the material (the thickness of the wave plate) be L. Then the number of E_x and E_y wavelengths that fit into the material are L/λ_x and L/λ_y, respectively, with the latter of these numbers being slightly larger. The number of *extra* wavelengths of E_y compared with E_x is therefore $L/\lambda_y - L/\lambda_x$. Each wavelength is worth 2π radians, so the phase of the E_y wave at the far end of the material is *behind* the phase of the E_x wave by an amount (we'll give four equivalent expressions here)

$$\Delta\phi = 2\pi L\left(\frac{1}{\lambda_y} - \frac{1}{\lambda_x}\right) = L(k_y - k_x) = \omega L\left(\frac{1}{v_y} - \frac{1}{v_x}\right) = \frac{\omega L}{c}(n_y - n_x), \qquad (72)$$

where we have used $v_i = c/n_i$. In retrospect, we could have simply written down the second of these expressions from the definition of the wavenumber k, but we have to be careful to get the sign right. The fact that a *larger* number of E_y waves fit into the material means that the phase of the E_y wave is *behind* the phase of the E_x wave. Of course, if E_y is behind by a large enough phase, then it is actually better described as being ahead. For example, being behind by $7\pi/4$ is equivalent to being ahead by $\pi/4$. We'll see shortly how we can use wave plates to do various things with polarization, including creating circularly polarized light.

[6]You might think that the E_y phase should be ahead, because it has more wiggles in it. But this is exactly backwards. Of course, if you're counting from the left end of the plate, E_y does sweep through more phase than E_x. But that's not what we're concerned with. We're concerned with the phase of the wave as it passes the far end of the plate. And since the crest marked with the dot in the E_y wave hasn't reached the end yet, the E_y phase is behind the E_x phase. So in the end, it is correct to use the simplistic reasoning of, "a given crest on the slower wave takes longer to reach the end, so the phase of the slower wave is behind."

8.6.4 Making polarized light

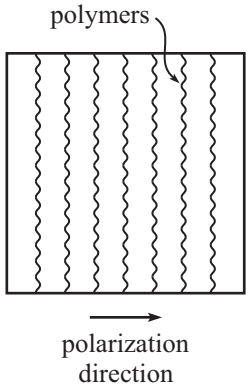


Figure 13

How can we produce polarized light? Let's look at linear polarization and then circular polarization. One way to produce linearly polarized light is to wiggle a charged particle in a certain way, so that the radiation is linearly polarized. We'll talk about how this works in Section 8.7.

Another way is to create a sheet of polymers (long molecules with repeating parts) that are stretched out parallel to each other, as shown in Fig. 13. If you have a light source that emits a random assortment of polarizations, then the sheet can filter out the light that is linearly polarized along a certain direction and leave you with only the light that is linearly polarized along the orthogonal direction. This works in the following way.

The electrons in the polymers are free to vibrate in the direction *along* the polymer, but not perpendicular to it. From the mechanism we discussed in Section 8.5, this leads to the absorption of the energy of the electric field that points along the polymer. So this component of the electric field shrinks to zero. Only the component *perpendicular* to the polymer survives. So we end up with linearly polarized light in the direction perpendicular to the polymers. You therefore can't think of the polymers as a sort of fence which lets through the component of the field that is parallel to the boards in the fence. It's the opposite of this.

Now let's see how to make circularly polarized light. As in the case of linear polarization, we can wiggle a charged particle in a certain way. But another way is to make use of the wave-plate results in the previous subsection. Let the fast and slow axes of the wave plate be the x and y axes, respectively. If we send a wave into the material that is polarized in either the x or y directions, then nothing exciting happens. The wave simply stays polarized along that direction. The phase difference we found in Eq. (72) is irrelevant if only one of the components exists.

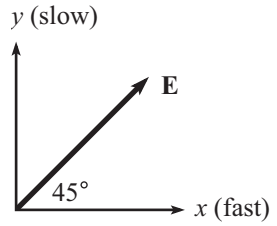


Figure 14

If we want to have anything useful come out of the phase difference caused by the plate, we need to have both components be nonzero. So let's assume that we have linearly polarized light that enters the material with a polarization direction at a 45° angle between the axes, as shown in Fig. 14. Given the frequency ω of the light, and given the two indices of refraction, n_x and n_y, let's assume that we've chosen the thickness L of the plate to yield a phase difference of $\Delta\phi = \pi/2$ in Eq. (72). Such a plate is called a *quarter-wave plate*, because $\pi/2$ is a quarter of a cycle. When the wave emerges from the plate, the E_y component is then $\pi/2$ behind the E_y component. So if we choose $t = 0$ to be the time when the wave enters the plate, we have (the phase advance of ϕ is unimportant here)

$$\begin{aligned} \text{Enter plate}: \quad \mathbf{E} &\propto \hat{\mathbf{x}}\cos\omega t + \hat{\mathbf{y}}\cos\omega t, \\ \text{Leave plate}: \quad \mathbf{E} &\propto \hat{\mathbf{x}}\cos(\omega t+\phi) + \hat{\mathbf{y}}\cos(\omega t+\phi-\pi/2) \\ &= \hat{\mathbf{x}}\cos(\omega t+\phi) + \hat{\mathbf{y}}\sin(\omega t+\phi). \end{aligned} \tag{73}$$

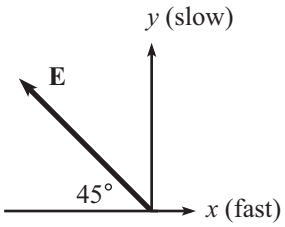


Figure 15

This wave has the same form as the wave in Eq. (67), so it is circularly polarized light with a counterclockwise orientation, as shown in above in Fig. 10.

What if we instead have an incoming wave polarized in the direction shown in Fig. 15? The same phase difference of $\pi/2$ arises, but the coefficient of the E_x part of the wave now has a negative sign in it. So we have

$$\begin{aligned} \text{Enter plate}: \quad \mathbf{E} &\propto -\hat{\mathbf{x}}\cos\omega t + \hat{\mathbf{y}}\cos\omega t, \\ \text{Leave plate}: \quad \mathbf{E} &\propto -\hat{\mathbf{x}}\cos(\omega t+\phi) + \hat{\mathbf{y}}\cos(\omega t+\phi-\pi/2) \\ &= -\hat{\mathbf{x}}\cos(\omega t+\phi) + \hat{\mathbf{y}}\sin(\omega t+\phi). \end{aligned} \tag{74}$$

This is circularly polarized light, but now with a clockwise orientation.

The thickness of a quarter-wave plate (or a half-wave plate, or anything else) depends on the wavelength of the light, or equivalently on the various other quantities in Eq. (72). Intuitively, a longer wavelength means a longer distance to get ahead by a given fraction of that wavelength. So there is no "universal" quarter-wave plate that works for all wavelengths.

8.6.5 Malus' law

If a given electromagnetic wave encounters a linear polarizer, how much of the light makes it through? Consider light that is linearly polarized in the $\hat{\mathbf{x}}$ direction, and assume that we have a polarizer whose axis (call it the $\hat{\mathbf{x}}'$ axis) makes an angle of θ with the $\hat{\mathbf{x}}$ axis, as shown in Fig. 16. (This means that the polymers are oriented at an angle $\theta \pm \pi/2$ with the $\hat{\mathbf{x}}$ axis.) In terms of the primed coordinate system, the amplitude of the electric field is

$$\mathbf{E} = E_0\hat{\mathbf{x}} = E_0(\hat{\mathbf{x}}'\cos\theta + \hat{\mathbf{y}}'\sin\theta). \tag{75}$$

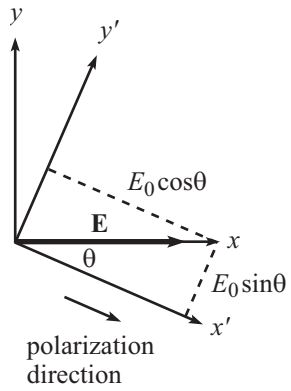


Figure 16

The y' component gets absorbed by the polarizer, so we're left with only the x' component, which is $E_0\cos\theta$. Hence, the polarizer decreases the amplitude by a factor of $\cos\theta$. The intensity (that is, the energy) is proportional to the square of the amplitude, which means that it is decreased by a factor of $\cos^2\theta$. Therefore,

$$|\mathbf{E}_{\text{out}}| = |\mathbf{E}_{\text{in}}|\cos\theta, \qquad \text{and} \qquad \boxed{I_{\text{out}} = I_{\text{in}}\cos^2\theta} \tag{76}$$

The second of these relations is known as *Malus' law*. Note that if $\theta = 90^\circ$, then $I_{\text{out}} = 0$. So two successive polarizers that are oriented at 90° with respect to each other block all of the light that impinges on them, because whatever light makes it through the first polarizer gets absorbed by the second one.

What happens if we put a third polarizer between these two at an angle of 45° with respect to each? It seems that adding another polarizer can only make things "worse" as far as the transmission of light goes, so it seems like we should still get zero light popping out the other side. However, if a fraction f of the light makes it through the first polarizer (f depends on what kind of light you shine in), then $f\cos^2 45^\circ$ makes it through the middle polarizer. And then a fraction $\cos^2 45^\circ$ of this light makes it through the final polarizer. So the total amount that makes it through all three polarizers is $f\cos^4 45^\circ = f/4$. This isn't zero! Adding the third polarizer makes things better, not worse.

This strange occurrence is due to the fact that polarizers don't act like filters of the sort where, say, a certain fraction of particles make it through a screen. In that kind of filter, a screen is always "bad" as far as letting particles through goes. The difference with actual polarizers is that the polarizer *changes* the polarization direction of whatever light makes it through. In contrast, if a particle makes it through a screen, then it's still the same particle. Another way of characterizing this difference is to note that a polarization is a vector, and vectors can be described in different ways, depending on what set of basis vectors is chosen. In short, in Fig. 17 the projection of **A** onto **C** is zero. But if we take the projection of **A** onto **B**, and then take the projection of the result onto **C**, the result isn't zero.

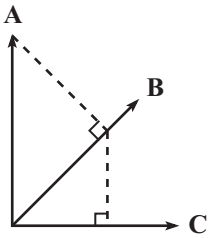


Figure 17

What happens if instead of inserting one intermediate polarizer at 45°, we insert two polarizers at angles 30° and 60°? Or three at 22.5°, 45°, and 67.5°, etc? Does more light or less light make it all the way through? The task of Problem [to be added] is to find out. You will find in this problem that something interesting happens in the case of a very large number of polarizers. The idea behind this behavior has countless applications in physics.

8.7 Radiation from a point charge

8.7.1 Derivation

In our above discussions of electromagnetic waves, we didn't worry about where the waves came from. We just studied their properties, given that they existed. Let's now look at how an electromagnetic wave can be created in the first place. We will find that an accelerating charge produces an electromagnetic wave.

(no **B** field)

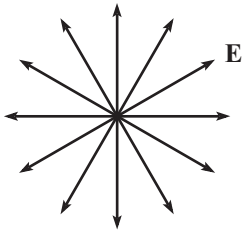


Figure 18

Consider first a stationary charge. The field is simply radial,

$$\mathbf{E} = \hat{\mathbf{r}} \frac{q}{4\pi\epsilon_0 r^2}\,, \tag{77}$$

and there is no **B** field. This is shown in Fig. 18. If we instead have a charge moving with *constant* velocity **v**, then the field is also radial, but it is bunched up in the transverse direction, as shown in Fig. 19. This can be derived with the Lorentz transformations. However, the proof isn't important here, and neither is half of the result. All we care about is the radial nature of the field. We'll be dealing with speeds that are generally much less than c, in which case the bunching-up effect is negligible. We therefore again have

$$\mathbf{E} \approx \hat{\mathbf{r}} \frac{q}{4\pi\epsilon_0 r^2} \qquad (\text{for } v \ll c). \tag{78}$$

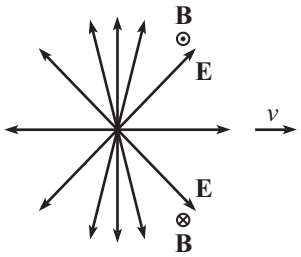


Figure 19

This is shown in Fig. 20. There is also a **B** field if the charge is moving. It points out of the page in the top half of Figs. 19 and 20, and into the page in the bottom half. This also follows from the Lorentz transformations (or simply by the righthand rule if you think of the moving charge as a current), but it isn't critical for the discussion. You can check with the righthand rule that $\mathbf{S} \propto \mathbf{E} \times \mathbf{B}$ points tangentially, which means that no power is radiated outward. And you can also check that **S** always has a forward component in the direction of the charge's velocity. This makes sense, because the field (and hence the energy) increases as the charge moves to the right, and **S** measures the flow of energy.

(for $v \ll c$)

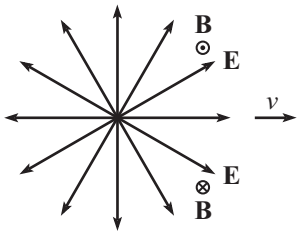


Figure 20

This radial nature of **E** for a moving charge seems reasonable (and even perhaps obvious), but it's actually quite bizarre. In Fig. 21, the field at point P points radially away from the present position of the charge. But how can P know that the charge is where it is *at this instant*? What if, for example, the charge stops shortly before the position shown? The field at P would still be directed radially away from where the charge *would have been* if it had kept moving with velocity **v**. At least for a little while. The critical fact is that the information that the charge has stopped can travel only at speed c, so it takes a nonzero amount of time to reach P. After this time, the field will point radially away from the stopped position, as expected.

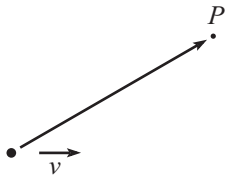


Figure 21

A reasonable question to ask is then: What happens to the field during the transition period when it goes from being radial from one point (the projected position if the charge kept moving) to being radial from another point (the stopped position)? In other words, what is the field that comes about due to the acceleration? The answer to this question will tell us how an electromagnetic field is created and what it looks like.

For concreteness, assume that the charge is initially traveling at speed v, and then let it decelerate with constant acceleration $-a$ for a time Δt (starting at $t = 0$) and come to rest. So $v = a\Delta t$, and the distance traveled during the stopping period is $(1/2)a\Delta t^2/2$. Let the origin of the coordinate system be located at the place where the deceleration starts.

Consider the situation at time T, where $T \gg \Delta t$. For example, let's say that the charge takes $\Delta t = 1\,\text{s}$ to stop, and we're looking at the setup $T = 1$ hour later. The distance $(1/2)a\Delta t^2/2$ is negligible compared with the other distances we'll be involved with, so we'll ignore it. At time T, positions with $r > cT$ have no clue that the charge has started

to decelerate, so they experience a field directed radially away from the future projected position. Conversely, positions with $r < c(T-\Delta t)$ know that the charge has stopped, so they experience a field directed radially away from the origin (or actually a position $(1/2)a\Delta t^2/2$, but this is negligible). So we have the situation shown in Fig. 22.

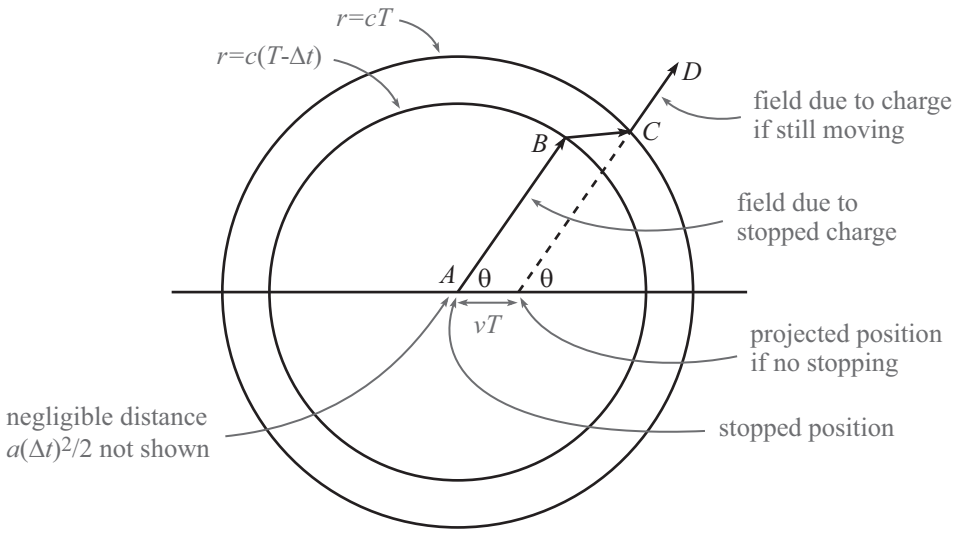


Figure 22

Since we're assuming $v \ll c$, the "exterior" field lines (the ones obtained by imagining that the charge is still moving) are essentially not compressed in the transverse direction. That is, they are spherically symmetric, as shown in Fig. 20. (More precisely, the compression effect is of order v^2/c^2 which is small compared with the effects of order v/c that we will find.) Consider the segments AB and CD in Fig. 22. These segments are chosen to make the same angle θ (which can be arbitrary) with the x axis, with AB passing through the stopped position, and the line of CD passing through the projected position. Due to the spherically symmetric nature of both the interior and exterior field lines, the surfaces of revolutions of AB and CD (which are parts of cones) enclose the same amount of flux. AB and CD must therefore be part of the same field line. This means that they are indeed connected by the "diagonal" field line BC shown.[7]

If we expand the relevant part of Fig. 22, we obtain a picture that takes the general form shown in Fig. 23. Let the radial and tangential components of the **E** field in the transition region be E_r and E_θ. From similar triangles in the figure, we have

$$\frac{E_\theta}{E_r} = \frac{vT\sin\theta}{c\Delta t}\,. \tag{79}$$

Note that the righthand side of this grows with T. Again, the units of E_θ and E_r aren't distance, so the size of the **E** vector in Fig. 23 is meaningless. But all the matters in the above similar-triangle argument is that the vector points in the direction shown.

[7]We're drawing both field lines and actual distances in this figure. This technically makes no sense, of course, because the fields don't have units of distance. But the point is to show the directions of the fields. We could always pick pick our unit size of the fields to be the particular value that makes the lengths on the paper be the ones shown.

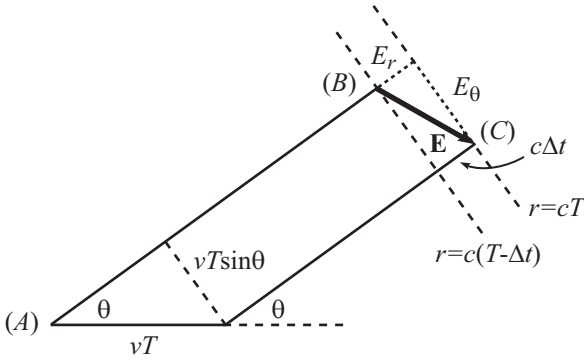


Figure 23

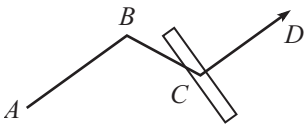


Figure 24

We now claim that E_r has the same value just outside and just inside the transition region. This follows from a Gauss's-law argument. Consider the pillbox shown in Fig. 24, which is located at the kink in the field at point C. The long sides of the box are oriented in the tangential direction, and also in the direction perpendicular to the page. The short sides are chosen to be infinitesimally small, so E_θ contributes essentially nothing to the flux. The flux is therefore due only to the radial component, so we conclude that $E_r^{\text{transition}} = E_r^{\text{outside}}$. (And likewise, a similar pillbox at point B tells us that $E_r^{\text{transition}} = E_r^{\text{inside}}$, but we won't need this. $E_r^{\text{transition}}$ varies slightly over the transition region, but the change is negligible if Δt is small.) We know that

$$E_r^{\text{outside}} = \frac{q}{4\pi\epsilon_0 r^2} = \frac{q}{4\pi\epsilon_0 (cT)^2}\,. \tag{80}$$

Eq. (79) then gives

$$\begin{aligned} E_\theta = \frac{vT\sin\theta}{c\Delta t}E_r &= \frac{vT\sin\theta}{c\Delta t}\cdot\frac{q}{4\pi\epsilon_0(cT)^2} \\ &= \frac{q\sin\theta}{4\pi\epsilon_0 c^2(cT)}\cdot\frac{v}{\Delta t} \\ &= \frac{qa\sin\theta}{4\pi\epsilon_0 rc^2} \qquad (\text{using } \ a = v/\Delta t \ \text{ and } \ r = cT) \end{aligned} \tag{81}$$

So the field inside the transition region is given by

$$\boxed{(E_r, E_\theta) = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r^2}\,,\,\frac{a\sin\theta}{rc^2}\right)} \tag{82}$$

Both of the components in the parentheses have units of $1/\text{m}^2$, as they should. We will explain below why this field leads to an electromagnetic wave, but first some remarks.

Remarks:

1. The location of the E_θ field (that is, the transition region) propagates outward with speed c.
2. E_θ is proportional to a. Given this fact, you can show that the only way to obtain the right units for E_θ using a, r, c, and θ, is to have a function of the form, $af(\theta)/rc^2$, where $f(\theta)$ is an arbitrary function of θ. And $f(\theta)$ happens to be $\sin\theta$ (times $q/4\pi\epsilon_0$).
3. If $\theta = 0$ or $\theta = \pi$, then $E_\theta = 0$. In other words, there is no radiation in the forward or backward directions. E_θ is maximum at $\theta = \pm\pi/2$, that is, in the transverse direction.

4. The acceleration vector associated with Fig. 22 points to the left, since the charge was decelerating. And we found that E_θ has a rightward component. So in general, the E_θ vector is

$$\mathbf{E}_\theta(\mathbf{r}, t) = -\frac{q}{4\pi\epsilon_0} \cdot \frac{\mathbf{a}_\perp(t')}{rc^2}\,, \tag{83}$$

where $t' \equiv t - r/c$ is the time at which the kink at point C in Fig. 22 was emmitted, and where $\mathbf{a}_\perp$ is the component of $\mathbf{a}$ that is perpendicular to the radial direction. In other words, it is the component with magnitude $a\sin\theta$ that you "see" across your vision if you are located at position $\mathbf{r}$. This is consistent with the previous remark, because $\mathbf{a}_\perp = \mathbf{0}$ if $\theta = 0$ or $\theta = \pi$. Note the minus sign in Eq. (83).

5. Last, but certainly not least, we have the extremely important fact: For sufficiently large r, E_r is negligible compared with E_θ. This follows from the fact that in Eq. (82), E_θ has only one r in the denominator, whereas E_r has two. So for large r, we can ignore the "standard" radial part of the field. We essentially have only the new "strange" tangential field. By "large r," we mean $a/rc^2 \gg 1/r^2 \Longrightarrow r \gg c^2/a$. Or equivalently $\sqrt{ra} \gg c$. In other words, ignoring relativity and using the kinematic relation $v = \sqrt{2ad}$, the criterion for large r is that (in an order-of-magnitude sense) if you accelerate something with acceleration a for a distance r, its velocity will exceed c.

   The reason why E_θ becomes so much larger than E_r is because there is a T in the numerator of Eq. (79). This T follows from the fact that in Fig. 23, the E_θ component of $\mathbf{E}$ grows with time (because vT, which is the projected position of the charge if it kept moving, grows with time), whereas E_r is always proportional to the constant quantity, $c\Delta t$.

The above analysis dealt with constant acceleration. However, if the acceleration is changing, we can simply break up time into little intervals, with the above result holding for each interval (as long as T is large enough so that all of our approximations hold). So even if $\mathbf{a}$ is changing, $\mathbf{E}_\theta(\mathbf{r}, t)$ is proportional to whatever $-\mathbf{a}_\perp(t')$ equaled at time $t' \equiv t - r/c$. In particular, if the charge is wiggling sinusoidally, then $\mathbf{E}_\theta(\mathbf{r}, t)$ is a sinusoidal wave.

The last remark above tells us that if we're far away from an accelerating charge, then the only electric field we see is the tangential one; there is essentially no radial component. There is also a magnetic field, which from Maxwell's equations can be shown to also be tangential, perpendicular to the page in Fig. 22; see Problem [to be added]. So we have electric and magnetic fields that oscillate in the tangential directions while propagating with speed c in the radial direction. But this is exactly what happens with an electromagnetic wave. We therefore conclude that an electromagnetic wave can be generated by an accelerating charge.

Of course, we know from Section 8.3 that Maxwell's equations in vacuum imply that the direction of the $\mathbf{E}$ and $\mathbf{B}$ fields must be perpendicular to the propagation direction, so in retrospect we know that this also has to be the case for whatever fields popped out of the above analysis. The main new points of this analysis are that (1) an accelerating charge can generate the electromagnetic wave (before doing this calculation, for all we know a nonzero, say, third derivative of the position is needed to generate a wave), and (2) the radial field in Eq. (82) essentially disappears due to the $1/r^2$ vs. $1/r$ behavior, leaving us with only the tangential field.

Eq. (30) gives the magnetic field as $\mathbf{k} \times \mathbf{E} = \omega\mathbf{B} \Longrightarrow \hat{\mathbf{r}} \times \mathbf{E} = c\mathbf{B}$ (using $\mathbf{k} = k\hat{\mathbf{r}}$). So in the top half of Fig. 22, $\mathbf{B}$ points into the page with magnitude $B = E/c$. And in the bottom half it points out of the page. These facts are consistent with the cylindrical symmetry of the system around the horizontal axis. If the charge is accelerating instead of decelerating as we chose above, then the $\mathbf{E}$ and $\mathbf{B}$ fields are reversed.

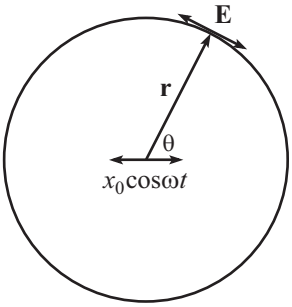


Figure 25

8.7.2 Poynting vector

What is the energy flow of a wave generated by a sinusoidally oscillating charge? Let the position of the charge be $x(t) = x_0 \cos\omega t$. The acceleration is then $a(t) = -\omega^2 x_0 \cos\omega t$. The resulting electric field at an arbitrary point is shown in Fig. 25. The energy flow at this point is given by the Poynting vector, which from Eqs. (49) and (50) is

$$\mathbf{S} = \frac{1}{\mu_0}\mathbf{E}\times\mathbf{B} = c\epsilon_0 E^2 \hat{\mathbf{r}}. \tag{84}$$

Since E is given by the E_θ in Eq. (82), the average value of the magnitude of $\mathbf{S}$ is (using $a = -\omega^2 x_0 \cos\omega t$, along with the fact that the average value of $\cos^2\omega t$ is $1/2$)

$$\begin{aligned} S_{\text{avg}} &= c\epsilon_0 \left(\frac{q}{4\pi\epsilon_0}\cdot\frac{a\sin\theta}{rc^2}\right)^2 \\ &= \frac{q^2}{16\pi^2\epsilon_0 c^3}\cdot\frac{1}{r^2}(\omega^2 x_0)^2\sin^2\theta\cdot\frac{1}{2} \\ &= \boxed{\frac{\omega^4 x_0^2 q^2\sin^2\theta}{32\pi^2\epsilon_0 c^3}\cdot\frac{1}{r^2}} \end{aligned} \tag{85}$$

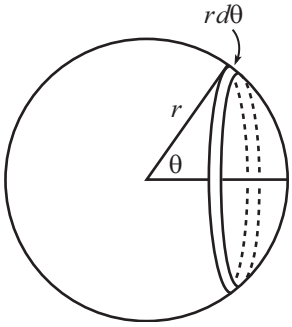


Figure 26

If you want, you can write the $1/\epsilon_0 c^3$ part of this as μ_0/c. Note that S_{avg} falls off like $1/r^2$ and is proportional to ω^4. Note also that it is zero if $\theta = 0$ or $\theta = \pi$.

The Poynting vector has units of Energy/(time $\cdot$ area). Let's integrate it over a whole sphere of radius r to find the total Energy/time, that is, the total power. If we parameterize the integral by θ, then we can slice the sphere into rings as shown in Fig. 26. The circumference of a ring is $2\pi r\sin\theta$, and the width is $r\,d\theta$. So the total power is

$$\begin{aligned} P = \int_{\text{sphere}} S_{\text{avg}} &= \int_0^\pi S_{\text{avg}}(2\pi r\sin\theta)(r\,d\theta) \\ &= \frac{\omega^4 x_0^2 q^2}{32\pi^2\epsilon_0 c^3}\cdot\frac{2\pi r^2}{r^2}\int_0^\pi \sin^3\theta\,d\theta \\ &= \boxed{\frac{\omega^4 x_0^2 q^2}{12\pi\epsilon_0 c^3}} \end{aligned} \tag{86}$$

where we have used the fact that $\int_0^\pi \sin^3\theta\,d\theta = 4/3$. You can quickly verify this by writing the integrand as $(1-\cos^2\theta)\sin\theta$. There are two important features of this result for P. First, it is independent of r. This must be the case, because if more (or less) energy crosses a sphere at radius r_1 than at radius r_2, then energy must be piling up (or be taken from) the region in between. But this can't be the case, because there is no place for the energy to go. Second, P is proportional to $(\omega^2 x_0)^2$, which is the square of the amplitude of the acceleration. So up to constant numbers and physical constants, we have

$$P \propto a_0^2 q^2, \tag{87}$$

were a_0 is the amplitude of the acceleration.

8.7.3 Blue sky

The fact that the P in Eq. (86) is proportional to ω^4 has a very noticeable consequence in everyday life. It is the main reason why the sky is blue. (The exponent doesn't have to be 4. Any reasonably large number would do the trick, as we'll see.) In a nutshell, the

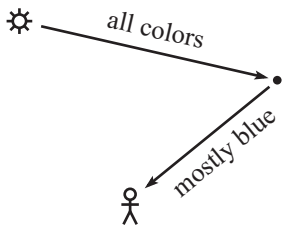


Figure 27

ω^4 factor implies that blue light (which is at the high-frequency end of visible spectrum) scatters more easily than red light (which is at the low-frequency end of visible spectrum). So if random white light (composed of many different frequencies) hits the air molecule shown in Fig. 27, the blue light is more likely to scatter and hit your eye, whereas the other colors with smaller frequencies are more likely to pass straight through. The sky in that direction therefore looks blue to you. More precisely, the intensity (power per area) of blue light that is scattered to your eye is larger than the intensity of red light by a factor of $P_{\text{blue}}/P_{\text{red}} = \omega_{\text{blue}}^4/\omega_{\text{red}}^4$. And since $\omega_{\text{blue}}/\omega_{\text{red}} \approx 1.5$, we have $P_{\text{blue}}/P_{\text{red}} \approx 5$. So 5 times as much blue light hits your eye.

This also explains why sunsets are red. When the sun is near the horizon, the light must travel a large distance through the atmosphere (essentially tangential to the earth) to reach your eye, much larger than when the sun is high up in the sky. Since blue light scatters more easily, very little of it makes it straight to your eye. Most of it gets scattered in various directions (and recall that none of it gets scattered directly forward, due to the $\sin^2\theta$ factor in Eq. (85)). Red light, on the other hand, scatters less easily, so it is more likely to make it all the way through the atmosphere in a straight line from the sun to your eye. Pollution adds to this effect, because it adds particles to the air, which strip off even more of the blue light by scattering. So for all the bad effects of pollution, cities sometimes have the best sunsets. A similar situation arises with smoke. If you look at the sun through the smoke of a forest fire, it appears as a crisp red disk (but don't look at it for too long).

The actual scattering process is a quantum mechanical one involving photons, and it isn't obvious how this translates to our electromagnetic waves. But for the present purposes, it suffices to think about the scattering process as one where a wave with a given intensity encounters a region of molecules, and the molecules grab chunks of energy and throw them off in some other direction. (The electrons in the molecules are the things that are vibrating/accelerating and creating the radiation). The point is that with blue light, the chunks of energy are 5 times as large as they are for red light.

There are, however, a number of issues that we've glossed over. The problem is rather complicated when everything is included. In particular, one issue is that in addition to the ω^4 factor, the P in Eq. (86) is also proportional to x_0^2. What if the electron's x_0 value for red light is larger than the value for blue light? It turns out that it isn't; the x_0's are all essentially the same size. This can be shown by treating the electron in the atom as an essentially undamped driven oscillator. The natural frequency ω_0 depends on the nature of the atom, and it turns out that it is much larger than the frequency ω of light in the visible spectrum (we'll just accept this fact). The driven-oscillator amplitude is given in Eq. (1.88), and when $\gamma \approx 0$ and $\omega_0 \gg \omega$, it reduces to being proportional to $1/\omega_0^2$. That is, it is independent of ω, as we wanted to show.

Other issues that complicate things are: Is there multiple scattering? (The answer is generally no.) Why is the sky not violet, in view of the fact that $\omega_{\text{violet}} > \omega_{\text{blue}}$? How does the eye's sensitivity come into play? (It happens to be peaked at green.) So there are certainly more things to consider. But the ω^4 issue we covered above can quite reasonably be called the main issue.

8.7.4 Polarization in the sky

If you look at the daytime sky with a polarizer (polarized sunglasses do the trick) and rotate it in a certain way (the polarizer, not the sky, although that would suffice too), you will find that a certain region of the sky look darker. The reason for this is that the light that makes it to your eye after getting scattered from this region is polarized. To see why, consider an electron in the air that is located at a position such that the line from it to you is perpendicular to the line from the sun to it (which is the $\hat{\mathbf{k}}$ direction of the sun's

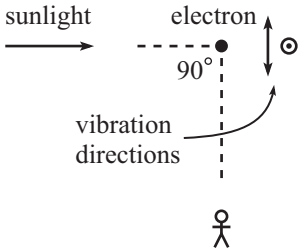


Figure 28

light); see Fig. 28. The radiation from the sun may cause this electron to vibrate. And then because it is vibrating/accelerating, it radiates light which may end up in your eye.

The electric field in the sun's light lies in the plane perpendicular to $\hat{\mathbf{k}}$. So the field is some combination of the two directions indicated in Fig. 28. The field can have a component along the line from the electron to you, and also a component perpendicular to the page, signified by the $\odot$ in the figure. The sun's light is randomly polarized, so it contains some of each of these. The electric field causes the electron to vibrate, and from the general force law $\mathbf{F} = q\mathbf{E}$, the electron vibrates in some combination of these two directions. However, due to the $\sin^2\theta$ factor in Eq. (85), you don't see any of the radiation that arises from the electron vibration along the line between it and you. Therefore, the only light that reaches your eye is the light that was created from the vibration pointing perpendicular to the page. Hence all of the light you see is polarized in this direction. So if your sunglasses are oriented perpendicular to this direction, then not much light makes it through, and the sky looks dark.

Note how the three possible directions of the resulting $\mathbf{E}$ field got cut down to one. First, the electric field that you see must be perpendicular to the $\hat{\mathbf{k}}$, due to the transverse nature of light (which is a consequence of Maxwell's equations), and due to the fact that the electron vibrates along the direction of $\mathbf{E}$. And second, the field must be perpendicular to the line from the electron to you, due to the "no forward scattering" fact that arises from the $\sin^2\theta$ factor in Eq. (85). Alternatively, we know that the $\mathbf{E}$ field that you see can't have a longitudinal component.

It's easy to see that conversely if the electron is *not* located at a position such that the line from it to you is perpendicular to the line from the sun to it, then you will receive some light that came from the vertical (on the page, in Fig. 28) oscillation of the electron. But the amount will be small if the angle is near 90°. So the overall result is that there is a reasonably thick band in the sky that looks fairly dark when viewed through a polarizer. If the sun is directly overhead, then the band is a circle at the horizon. If the sun is on the horizon, then the band is a semicircle passing directly overhead, starting and ending at the horizon.

8.8 Reflection and transmission

We'll now study the reflection and transmission that arise when light propagating in one medium encounters another medium. For example, we might have light traveling through air and then encountering a region of glass. We'll begin with the case of normal incidence and then eventually get to the more complicated case of non-normal incidence. For the case of normal incidence, we'll start off by considering only one boundary. So if light enters a region of glass, we'll assume that the glass extends infinitely far in the forward direction.

8.8.1 Normal incidence, single boundary

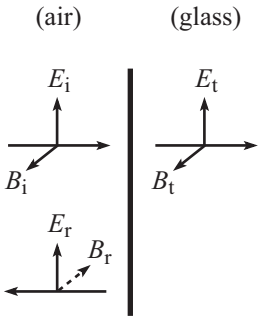


Figure 29

Consider light that travels to the right and encounters an air/glass boundary. As with other waves we've discussed, there will be a reflected wave and a transmitted wave. Because the wave equation in Eq. (15) is dispersionless, all waves travel with the same speed (in each region). So we don't need to assume anything about the shape of the wave, although we generally take it to be a simple sinusoidal one.

The incident, reflected, and transmitted waves are shown in Fig. 29 (the vertical displacement in the figure is meaningless). We have arbitrarily defined all three electric fields to be positive if they point upward on the page. A negative value of E_i, E_r, or E_t simply means that the vector points downward. We aren't assuming anything about the polarization of

the wave. No matter what kind of wave we have, the incident electric field at a given instant in time points in some direction, and we are taking that direction to be upward (which we can arrange by a suitable rotation of the axes). The waves we have drawn are understood to be the waves infinitesimally close to the boundary on either side.

Given these positive conventions for the E's, and given the known directions of the three $\hat{\mathbf{k}}$ vectors, the positive directions of the magnetic fields are determined by $\hat{\mathbf{k}} \times \mathbf{E} = \omega \mathbf{B}$, and they are shown. Note that if E_r has the same orientation as E_i (the actual sign will be determined below), then B_r must have the opposite orientation as B_i, because the $\hat{\mathbf{k}}$ vector for the reflected waves is reversed. If we define positive B to be out of the page, then the total $\mathbf{E}$ and $\mathbf{B}$ fields in the left and right regions (let's call these regions 1 and 2, respectively) are

$$\begin{aligned} E_1 &= E_i + E_r, \qquad \text{and} \qquad E_2 = E_t, \\ B_1 &= B_i - B_r, \qquad \text{and} \qquad B_2 = B_t. \end{aligned} \tag{88}$$

What are the boundary conditions? There are four boundary conditions in all: two for the components of the electric and magnetic fields parallel to the surface, and two for the components perpendicular to the surface. However, we'll need only the parallel conditions for now, because all of the fields are parallel to the boundary. The perpendicular ones will come into play when we deal with non-normal incidence in Section 8.8.3 below. The two parallel conditions are:

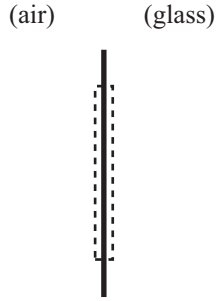


Figure 30

- Let's first find the boundary condition on $E^{\parallel}$ (the superscript "$\parallel$" stands for parallel). The third of Maxwell's equations in Eq. (23) is $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$. Consider the very thin rectangular path shown in Fig. 30. If we integrate each side of Maxwell's equation over the surface S bounded by this path, we obtain

  $$\int_S \nabla \times \mathbf{E} = -\int_S \frac{\partial \mathbf{B}}{\partial t}\,. \tag{89}$$

  By Stokes' theorem, the lefthand side equals the integral $\int \mathbf{E} \cdot d\ell$ over the rectangular path. And the right side is $-\partial \Phi_B/\partial t$, where Φ_B is the magnetic flux through the surface S. But if we make the rectangle arbitrarily thin, then the flux is essentially zero. So Eq. (89) becomes $\int \mathbf{E} \cdot d\ell = 0$. The short sides of the rectangle contribute essentially nothing to this integral, and the contribution from the long sides is $E_2^{\parallel}\ell - E_1^{\parallel}\ell$, where ℓ is the length of these sides. Since this difference equals zero, we conclude that

  $$\boxed{E_1^{\parallel} = E_2^{\parallel}} \tag{90}$$

  The component of the electric field that is parallel to the boundary is therefore continuous across the boundary. This makes sense intuitively, because the effect of the dielectric material (the glass) is to at most have charge pile up on the boundary, and this charge doesn't affect the field parallel to the boundary. (Or if it does, in the case where the induced charge isn't uniform, it affects the two regions in the same way.)

- Now let's find the boundary condition on $B^{\parallel}$. Actually, what we'll find instead is the boundary condition on $H^{\parallel}$, where the $\mathbf{H}$ field is defined by $\mathbf{H} \equiv \mathbf{B}/\mu$. We're using $\mathbf{H}$ here instead of $\mathbf{B}$ because $\mathbf{H}$ is what appears in the fourth of Maxwell's equations in Eq. (23), $\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t + \mathbf{J}_{\text{free}}$. We need to use this form because we're working with a dielectric.

  There are no free currents anywhere in our setup, so $\mathbf{J}_{\text{free}} = 0$. We can therefore apply to $\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t$ the exact same reasoning with the thin rectangle that we used

above for the electric field, except now the long sides of the rectangle are perpendicular to the page. We immediately obtain

$$\boxed{H_1^{\parallel} = H_2^{\parallel}} \tag{91}$$

The boundary condition for the $B^{\parallel}$ fields is then $B_1^{\parallel}/\mu_1 = B_2^{\parallel}/\mu_2$. However, since the μ value for most materials is approximately equal to the vacuum μ_0 value, the $B^{\parallel}$ fields also approximately satisfy $B_1^{\parallel} = B_2^{\parallel}$.

We can now combine the above two boundary conditions and solve for E_r and E_t in terms of E_i, and also H_r and H_t in terms of H_i. (The B's are then given by $B \equiv H/\mu$.) We can write the $H^{\parallel}$ boundary condition in terms of E fields by using

$$H \equiv \frac{B}{\mu} = \frac{E}{v\mu} \equiv \frac{E}{Z}, \qquad \text{where } \boxed{Z \equiv v\mu} \tag{92}$$

This expression for Z is by definition the impedance for an electromagnetic field. (We're using v for the speed of light in a general material, which equals $1/\sqrt{\mu\epsilon}$ from Eq. (19). We'll save c for the speed of light in vacuum.) Z can alternatively be written as

$$Z \equiv v\mu = \frac{\mu}{\sqrt{\mu\epsilon}} = \sqrt{\frac{\mu}{\epsilon}}. \tag{93}$$

The two boundary conditions can now be written as

$$\begin{aligned} E_1^{\parallel} = E_2^{\parallel} &\implies E_i + E_r = E_t, \\ H_1^{\parallel} = H_2^{\parallel} &\implies H_i - H_r = H_t \implies \frac{E_i}{Z_1} - \frac{E_r}{Z_1} = \frac{E_t}{Z_2}. \end{aligned} \tag{94}$$

These are two equations in the two unknowns E_r and E_t. Solving for them in terms of E_i gives

$$\boxed{E_r = \frac{Z_2 - Z_1}{Z_2 + Z_1} E_i} \qquad \text{and} \qquad \boxed{E_t = \frac{2Z_2}{Z_2 + Z_1} E_i} \tag{95}$$

These are similar (but not identical) to the reflection and transmission expressions for a transverse wave on a string; see Eq. (4.38).

We'll generally be concerned with just the E values, because the B values can always be found via Maxwell's equations. But if you want to find the H and B values, you can write the E's in the first boundary condition in terms of the H's. The boundary conditions become

$$\begin{aligned} E_1^{\parallel} = E_2^{\parallel} &\implies Z_1 H_i + Z_1 H_r = Z_2 H_t, \\ H_1^{\parallel} = H_2^{\parallel} &\implies H_i - H_r = H_t. \end{aligned} \tag{96}$$

Solving for H_r and H_t gives

$$H_r = \frac{Z_2 - Z_1}{Z_2 + Z_1} H_i, \qquad \text{and} \qquad H_t = \frac{2Z_1}{Z_2 + Z_1} H_i. \tag{97}$$

Remember that positive H_r is defined to point into the page, whereas positive H_i points out of the page, as indicated in Fig. 29. So the signed statement for the vectors is $\mathbf{H}_r = \mathbf{H}_i \cdot (Z_1 - Z_2)/(Z_1 + Z_2)$. If you want to find the B values, they are obtained via $B = \mu H$.

But again, we'll mainly be concerned with just the E values. Note that you can also quickly obtain these H's by using $E = Hz$ in the results for the E's in Eq. (95),

We can write the above expressions in terms of the index of refraction, which we defined in Eq. (20). We'll need to make an approximation, though. Since most dielectrics have $\mu \approx \mu_0$, Eq. (21) gives $n \propto \sqrt{\epsilon}$, and Eq. (93) gives $Z \propto 1/\sqrt{\epsilon}$. So we have $Z \propto 1/n$. The expressions for E_r and E_t in Eq. (95) then become

$$\boxed{E_r = \frac{n_1 - n_2}{n_1 + n_2} E_i} \quad \text{and} \quad \boxed{E_t = \frac{2n_1}{n_1 + n_2} E_i} \quad (\text{if } \mu \approx \mu_0). \tag{98}$$

In going from, say, air to glass, we have $n_1 = 1$ and $n_2 \approx 1.5$. Since $n_1 < n_2$, this means that E_r has the opposite sign of E_i. A reflection like this, where the wave reflects off a region of higher index n, is called a "hard reflection." The opposite case with a lower n is called a "soft reflection."

What is the power in the reflected and transmitted waves? The magnitude of the Poynting vector for a traveling wave is given by Eq. (49) as $S = E^2/v\mu$, where we are using v and μ instead of c and μ_0 to indicate an arbitrary dielectric. But $v\mu$ is by definition the impedance Z, so the instantaneous power is (we'll use "P" instead of "S" here)

$$\boxed{P = \frac{E^2}{Z}} \tag{99}$$

Now, it must be the case that the incident power P_i equals the sum of the reflected and transmitted powers, P_r and P_t. (All of these P's are the instantaneous values at the boundary.) Let's check that this is indeed true. Using Eq. (95), the reflected and transmitted powers are

$$\begin{aligned} P_r &= \frac{E_r^2}{Z_1} = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 \frac{E_i^2}{Z_1} = \frac{(Z_2 - Z_1)^2}{(Z_2 + Z_1)^2} P_i, \\ P_t &= \frac{E_t^2}{Z_2} = \left(\frac{2Z_2}{Z_2 + Z_1}\right)^2 \frac{E_i^2}{Z_2} = \frac{4Z_1 Z_2}{(Z_2 + Z_1)^2} \frac{E_i^2}{Z_1} = \frac{4Z_1 Z_2}{(Z_2 + Z_1)^2} P_i. \end{aligned} \tag{100}$$

We then quickly see that $P_r + P_t = P_i$, as desired.

The following topics will eventually be added:

8.8.2 Normal incidence, double boundary

8.8.3 Non-normal incidence

Huygens' principle, Snell's law

Maxwell's equations